Title
A unified framework for testing high-dimensional parameters: A data-adaptive approach

Authors
Zhou, Cheng
Zhang, Xinsheng
Zhou, Wenxin
et al.

Publication Date
2018-08-08

Peer reviewed

Submitted to the Annals of Statistics

A UNIFIED FRAMEWORK FOR TESTING HIGH DIMENSIONAL PARAMETERS: A DATA-ADAPTIVE APPROACH

By Cheng Zhou[‡], Xinsheng Zhang[‡], Wenxin Zhou[§] and Han Liu[§]

Department of Statistics, Fudan University[‡] and Department of Operation Research and Financial Engineering, Princeton University[§]

High dimensional hypothesis test deals with models in which the number of parameters is significantly larger than the sample size. Existing literature develops a variety of individual tests. Some of them are sensitive to the dense and small disturbance, and others are sensitive to the sparse and large disturbance. Hence, the powers of these tests depend on the assumption of the alternative scenario. This paper provides a unified framework for developing new tests which are adaptive to a large variety of alternative scenarios in high dimensions. In particular, our framework includes arbitrary hypotheses which can be tested using high dimensional U-statistic based vectors. Under this framework, we first develop a broad family of tests based on a novel variant of the L_p-norm with $p \in \{1, \ldots, \infty\}$. We then combine these tests to construct a data-adaptive test that is simultaneously powerful under various alternative scenarios. To obtain the asymptotic distributions of these tests, we utilize the multiplier bootstrap for U-statistics. In addition, we consider the computational aspect of the bootstrap method and propose a novel low cost scheme. We prove the optimality of the proposed tests. Thorough numerical results on simulated and real datasets are provided to support our theory.

1. Introduction. Modern data acquisition routinely produces massive datasets in many scientific areas, e.g. genomics, astronomy, functional Magnetic Resonance Imaging (fMRI), and image processing. Effective analysis of such data requires us to test high dimensional parameters ([47, 58, 71, 76, 29, 18]). Though specific methods have been developed to infer high dimensional mean and covariance parameters. It is unclear how to choose the best test when the parameter of interest has a complex structure and the pattern of possible alternative hypothesis is unknown. In particular, we need a unified framework for constructing tests of high dimensional parameters which are simultaneously powerful under a large variety of alternative assumptions. This paper provides such a framework.

Keywords and phrases: High dimensional hypothesis tests, U-statistics, multiplier bootstrap methods, data-adaptive tests

1.1. *General setup.* Our framework considers a generic setup for high dimensional inference. More specifically, let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ be two d-dimensional random vectors independent of each other. $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}$ are independent and identically distributed (i.i.d.) random samples from $\boldsymbol{X}$ with $\boldsymbol{X}_k = (X_{k1}, X_{k2}, \ldots, X_{kd})^\top$. Similarly, $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2}$ are i.i.d. random samples from $\boldsymbol{Y}$ with $\boldsymbol{Y}_k = (Y_{k1}, Y_{k2}, \ldots, Y_{kd})^\top$. We set $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}\}$, $\mathcal{Y} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2}\}$, and

$$
\begin{aligned}
\widehat{u}_{1,s} &= \binom{n_1}{m}^{-1} \sum_{1 \le k_1 < \cdots < k_m \le n_1} \Phi_s(\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m}), \\
\widehat{u}_{2,s} &= \binom{n_2}{m}^{-1} \sum_{1 \le k_1 < \cdots < k_m \le n_2} \Phi_s(\boldsymbol{Y}_{k_1}, \ldots, \boldsymbol{Y}_{k_m}),
\end{aligned} \tag{1.1}
$$

where $s = 1, \ldots, q$, and Φ_s is a m-order symmetric kernel function. We assume that Φ_s is symmetric and that each kernel function is of the same order m only for notational simplicity.[1,2]

We then define two U-statistic based vectors as

$$
\widehat{\boldsymbol{u}}_1 := (\widehat{u}_{1,1}, \widehat{u}_{1,2}, \ldots, \widehat{u}_{1,q})^\top \quad \text{and} \quad \widehat{\boldsymbol{u}}_2 := (\widehat{u}_{2,1}, \widehat{u}_{2,2}, \ldots, \widehat{u}_{2,q})^\top. \tag{1.2}
$$

We use $\boldsymbol{u}_\gamma$ to denote the expectation of $\widehat{\boldsymbol{u}}_\gamma$, i.e., $\boldsymbol{u}_\gamma = (u_{\gamma,1}, u_{\gamma,2}, \ldots, u_{\gamma,q})^\top$ with $u_{\gamma,s} = \mathbb{E}[\widehat{u}_{\gamma,s}]$ for $\gamma = 1, 2$ and $s = 1, \ldots, q$. We are interested in testing the hypotheses:

(i) (One-sample problem) For a given $\boldsymbol{u}_0 \in \mathbb{R}^q$,

$$
\mathbf{H}_0 : \boldsymbol{u}_1 = \boldsymbol{u}_0 \qquad \text{v.s.} \qquad \mathbf{H}_1 : \boldsymbol{u}_1 \neq \boldsymbol{u}_0; \tag{1.3}
$$

(ii) (Two-sample problem)

$$
\mathbf{H}_0 : \boldsymbol{u}_1 = \boldsymbol{u}_2 \qquad \text{v.s.} \qquad \mathbf{H}_1 : \boldsymbol{u}_1 \neq \boldsymbol{u}_2. \tag{1.4}
$$

We consider the high dimensional setting that d/n (or q/n) does not necessarily go to zero. These two kinds of hypotheses are quite general and include most existing studies as special cases.

[1] If Φ_s is an asymmetric kernel function, it gives a U-statistic $\widehat{u}_{1,s} = \frac{1}{m!}\binom{n_1}{m}^{-1} \sum \Phi_s(\boldsymbol{X}_{\ell_1}, \ldots, \boldsymbol{X}_{\ell_m})$, where the summation is over all permutations of distinct elements $\{\ell_1, \ldots, \ell_m\}$ from $\{1, \ldots, n_1\}$. By setting $\Phi_s^0(\mathbf{x}_1, \ldots, \mathbf{x}_m) = (m!)^{-1} \sum \Phi_s(\boldsymbol{x}_{k_1}, \ldots, \boldsymbol{X}_{k_m})$, where the summation is over all permutations of $\{1, \ldots, m\}$, we rewrite $\widehat{u}_{1,s}$ as a U-statistic with a symmetric kernel Φ_s^0. For $\boldsymbol{Y}$, we can rewrite $\widehat{u}_{2,s}$ as a U-statistic with a symmetric kernel similarly.

[2] If $\{\Phi_s\}_{s=1,\ldots,q}$ have different kernel orders, we require that the kernel orders are uniformly bounded.

1.2. *Special cases and applications.* In this section, we provide several special cases of the above general testing problem.

- Matrix-based one-sample test:

$$\mathbf{H}_0 : \mathbf{U}_1 = \mathbf{I}_d \qquad \text{v.s.} \qquad \mathbf{H}_1 : \mathbf{U}_1 \neq \mathbf{I}_d, \tag{1.5}$$

  where $\mathbf{U}_1$'s entries are estimated by U-statistics and $\mathbf{I}_d$ is an identity matrix of size d. The hypothesis (1.5) is often used to infer the independence of random variables. This problem plays a fundamental role in many fields including multiple testing ([9]), naive Bayes classification ([69, 32]), and independent component analysis([25]). Under the Gaussian setting, testing (1.5) with $\mathbf{U}_1$ as covariance matrix is well studied both in low ([60, 57, 1]) and high ([45, 42, 10, 62, 3, 22, 12, 43, 17]) dimensions. Moreover, [42, 48, 75, 51, 16, 12, 65] consider the high dimensional independence test under more general distribution. Considering robustness, rank-based U-statistics such as Kendall' s tau and Spearman's rho are introduced to describe the dependence of random variables. As for their definitions and basic theoretical properties, we refer to the book [46]. Recently, [35, 6] study how to utilyze general U-statistics for high dimensional independence test.
- Matrix-based two-sample test:

$$\mathbf{H}_0 : \mathbf{U}_1 = \mathbf{U}_2 \qquad \text{v.s.} \qquad \mathbf{H}_1 : \mathbf{U}_1 \neq \mathbf{U}_2, \tag{1.6}$$

  where $\mathbf{U}_1$ and $\mathbf{U}_2$ are matrices such that their entries are estimated by U-statistics. The hypothesis (1.6) is often used before the discriminant analysis ([1, 64, 13, 55, 33, 54, 36]) to simplify the test statistics. For low dimensional two-sample covariance matrix test, we refer its theoretical properties to [1]. In recent years, [63, 68, 49, 14, 21] study how to perform the two-sample covariance matrix test in high dimensions. Moreover, [46, 35, 6, 74] consider how to use general U-statistics to replace covariance coefficients.
- Means test:
  (i) (One-sample problem)

$$\mathbf{H}_0 : \boldsymbol{\mu}_1 = \mathbf{0} \qquad \text{v.s.} \qquad \mathbf{H}_1 : \boldsymbol{\mu}_1 \neq \mathbf{0}; \tag{1.7}$$

  (ii) (Two-sample problem)

$$\mathbf{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \qquad \text{v.s.} \qquad \mathbf{H}_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2, \tag{1.8}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are mean vectors of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Testing the mean vector is a special case of (1.7) and (1.8). The testing of mean values is very fundamental. We refer their low dimensional properties to [1]. Recently, a large amount of literature work on high dimensional means test ([4, 67, 66, 22, 15, 20]).

For (1.5) and (1.6), we can convert the matrix into a column vector by vectorization to obtain equivalent tests with the same form as (1.7) or (1.8). Therefore, (1.5) and (1.6) fall in our framework.

Testing high dimensional U-statistic parameters also has many important practical applications. For example, in gene selection, we use it to detect gene differences [37, 39, 38, 14, 15] or rare variants [8, 50, 70, 47, 58] between the diseased and non-diseased population. In finance, we use it to detect anomalies ([19]) and test the market efficiency ([28, 30, 31]).

1.3. *Background and existing work.* In the low dimensional setting with $d < n$ fixed, the Hotelling's T^2 test enjoys certain kind of optimality and has been widely used. To test two-sample mean vectors, the Hotelling's T^2 is defined as

$$\frac{n_1 n_2}{n_1 + n_2}(\overline{\boldsymbol{X}} - \overline{\boldsymbol{Y}})^\top S_{1,2}^{-1}(\overline{\boldsymbol{X}} - \overline{\boldsymbol{Y}}),$$

where $\overline{\boldsymbol{X}} = n_1^{-1}\sum_{k=1}^{n_1}\boldsymbol{X}_k$, $\overline{\boldsymbol{Y}} = n_2^{-1}\sum_{k=1}^{n_2}\boldsymbol{Y}_k$, and

$$S_{1,2} = \frac{1}{n_1 + n_2 - 2}\Big(\sum_{k=1}^{n_1}(\boldsymbol{X}_k - \overline{\boldsymbol{X}})(\boldsymbol{X}_k - \overline{\boldsymbol{X}})^\top + \sum_{k=1}^{n_2}(\boldsymbol{Y}_k - \overline{\boldsymbol{Y}})(\boldsymbol{Y}_k - \overline{\boldsymbol{Y}})^\top\Big).$$

As for the limiting distribution, large and moderate deviations of Hotelling's T^2, we refer to [1, 27, 52].

In the high dimensional setting, many tests have been proposed to test high dimensional vectors and matrices. These tests fall in two categories: the L_2-type versus L_∞-type tests. Specifically, for (1.7) and (1.8), the L_2-type tests are based on $\|\mathbf{A}(\boldsymbol{u}_1 - \boldsymbol{u}_0)\|_2$ or $\|\mathbf{A}(\boldsymbol{u}_1 - \boldsymbol{u}_2)\|_2$, and the L_∞-type tests are based on $\|\mathbf{A}(\boldsymbol{u}_1 - \boldsymbol{u}_0)\|_\infty$ or $\|\mathbf{A}(\boldsymbol{u}_1 - \boldsymbol{u}_2)\|_\infty$ for some operator $\mathbf{A}$. On one hand, the L_2-type tests [4, 63, 66, 68, 22, 49] aim to detect relatively dense signals, as the L_2-norm accumulates small deviations of all entries. On the other hand, the L_∞-type tests [14, 15] are more sensitive to sparse signals, where some strong perturbations exist on a small number of entries. [52, 14, 15] illustrate that the L_∞-type tests are reasonably more powerful than the L_2-type tests and enjoy certain kind of optimality when the alternative is sparse.

1.4. *Our contributions.* Theoretically, there is no uniformly most powerful test under different scenarios of the alternatives ([26]). Therefore, depending on the unknown truth of alternatives, a given and fixed test may or may not be powerful. In this paper, we aim to develop a broad family of tests such that at least one of them is powerful enough in a given situation. We then combine these tests to obtain a data-adaptive test that will maintain high power across a wide range of alternative scenarios. We develop our family of tests based on a new family of adjusted L_p-norms with $p = 1, 2, \ldots, \infty$, so that there is at least one test in our family is powerful no matter the signal is dense or sparse. The limiting distribution of the data-adaptive test is very complex that we cannot obtain its explicit form. Therefore, we use the bootstrap method to approximate the limiting distribution, so that we can obtain the critical value and valid P-value of the test. More specifically, to obtain a better approximation in the high dimensional setting, we adjust L_p-norm while building the test statistics. In detail, we introduce it as follows.

Definition 1.1. For $\mathbf{v} = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, we define $\|\mathbf{v}\|_{(s_0,p)} := \big(\sum_{j=d-s_0+1}^{d} (v^{(j)})^p\big)^{1/p}$, where $v^{(1)}, v^{(2)}, \ldots, v^{(d)}$ are the order statistics of $|v_1|, \ldots, |v_d|$ with $0 \le v^{(1)} \le v^{(2)} \le \ldots \le v^{(d)}$.

By this definition, for any positive integer s_0, we have $\|\mathbf{v}\|_{(s_0,\infty)} = \|\mathbf{v}\|_\infty$, where $\|\mathbf{v}\|_\infty = \max_{j=1,\ldots,d} |v_d|$. Moreover, the following proposition shows that $\|\cdot\|_{(s_0,p)}$ is a norm for any $1 \le p \le \infty$.

Proposition 1. For any $1 \le p \le \infty$, $\|\cdot\|_{(s_0,p)}$ is a norm on $\mathbb{R}^d$.

The detailed proof of Proposition 1 is in Appendix B.1 of supplementary materials. In this paper, we assume $1 \le p \le \infty$ to make $\|\cdot\|_{(s_0,p)}$ a norm. Therefore, similarly to L_p-norm, we can call $\|\mathbf{v}\|_{(s_0,p)}$ the (s_0, p)-norm of $\mathbf{v}$ in this paper. To construct the above family of tests, we use the (s_0, p)-norm as the adjusted L_p-norm. More details on this testing procedure is in Section 2. This paper has four major contributions:

- First, we introduce a new family of tests based on the (s_0, p)-norm. As is shown in the simulation experiment of Section 4, the power of traditional L_p-norm based test decreases tremendously (especially for small p) as $q \to \infty$. The reason is that the L_p-norm with small p is easy to accumulate the noise of all entries. Therefore, we introduce s_0 to increase the signal-noise ratio of test statistics. The introduction of s_0 is also crucial in establishing our theoretical results for high dimensional multiplier bootstrap. Moreover, we obtain the required scaling between s_0, p, q, and n for the proposed bootstrap methods.

- Secondly, as it is hard to obtain the joint distribution of test statistics with various (s_0, p)-norm, we use the multiplier bootstrap method to obtain its asymptotic distribution. In low dimensions, this bootstrap method is well studied for both the sum of random variables ([61, 53, 59, 7]) and U-statistics ([44, 2, 56, 41, 40, 34]). In high dimensions, the multiplier bootstrap is also useful for approximating the sum of random vectors ([23]). Motivated by these results, we generalize multiplier bootstrap method for U-statistics to the high dimensional setting with theoretical guarantees.
- Thirdly, for adapting to the possible alternatives, we propose a new approach to combine these (s_0, p)-norm based tests. Our combined test automatically chooses the most powerful test within the chosen combination according to the data. Therefore, we call this test the data-adaptive combined test. However, to obtain the P-value for the combined test, we originally need a double-loop bootstrap procedure, which suffers from high computational cost. To avoid this, we propose a novel computationally efficient scheme which generates nonindependent bootstrap samples. We also provide theoretical guarantees for this new bootstrap scheme in the high dimensional setting.
- Finally, combining the developed theory for the proposed methods and exiting lower bounds in the literature, we present that our methods are rate-optimal in many settings.

1.5. *Notation.* We set $\|\mathbf{v}\|_p$ as the L_p-norm of a vector $\mathbf{v} = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$. We denote the spherical surface in $\mathbb{R}^d$ by $\mathbb{S}^{d-1} := \{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\|_2 = 1\}$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if there exists a constant C such that $|a_n| \leq C|b_n|$ holds for all sufficiently large n, write $a_n = o(b_n)$ if $a_n/b_n \to 0$, and write $a_n \asymp b_n$ if there exist constants $C \geq c > 0$ such that $c|b_n| \leq |a_n| \leq C|b_n|$ for all sufficiently large n. For a sequence of random variables $\{\xi_1, \xi_2, \ldots\}$, we use $\xi_n \to \xi$ to denote that the sequence $\{\xi_n\}$ converges in probability towards ξ as $n \to \infty$. For simplicity, we also use $\xi_n = o_p(1)$ to denote $\xi_n \to 0$.

1.6. *Paper organization.* The rest of this paper is organized as follows. In Section 2 we propose the new testing procedures: the individual (s_0, p)-norm based test and the data-adaptive combined test. In Section 3, we develop a theory to analyze the size and power of the proposed tests. Section 4 provides some numerical results on simulated data to justify our proposed methods' size and power. In Section 5, we discuss some potential future work. Supplementary materials provide both proofs and additional numerical results on both simulated and real data.

2. Methodology. This section introduces the (s_0, p)-norm based individual tests and the data-adaptive combined test for testing high dimensional U-statistic based parameters. We also introduce how to exploit the multiplier bootstrap method to obtain the critical values and P-values for both individual and combined tests. In the following, we introduce individual tests based on the (s_0, p)-norm in Section 2.1 and the data-adaptive combined test in Section 2.2.

2.1. *Individual tests based on the* (s_0, p)*-norm.* We introduce the (s_0, p)-norm based tests which are basic components of the data-adaptive combined test. First, we explain the construction motivation in Section 2.1.1 and describe the test statistics in Section 2.1.2. We then introduce bootstrapping scheme for U-statistics in high dimensions in Section 2.1.3 and use it to obtain critical values and P-values for the proposed tests.

2.1.1. *Motivation of the construction of the* (s_0, p)*-norm.* We first introduce the motivation of the proposed individual tests. In the existing literature, there are two types of tests (L_2-type and L_∞-type tests) to test high dimensional vectors or matrices. The L_2-type tests are sensitive to dense signals and the L_∞-type tests are sensitive to sparse signals. Therefore, the performance of these tests depends on the pattern of possible alternatives. If such pattern is unknown, it is more desirable to construct a data-adaptive test which is simultaneously powerful under various alternative scenarios. For this, we need to construct a family of versatile tests so that for a given alternative at least one test wiithin the family is powerful. Inspired by the existing L_2-type and L_∞-type tests, we build the test family based on the L_p-norm. Importantly, as p increases, the L_p-norm puts more weight on the larger entries while gradually ignoring the remaining smaller entries. As $p \to \infty$, we have $\|\mathbf{v}\|_p \to \|\mathbf{v}\|_\infty$ for any $\mathbf{v} \in \mathbb{R}^d$, where $\|\mathbf{v}\|_\infty$'s value only depends on the largest entry of $\mathbf{v}$. More generally, as p increases, we put more weight on the larger entries, eventually realizing the L_∞-type test. Hence, by properly choosing p from the proposed test family, there exists at least one test within the family that is powerful in each alternative situation.

However, it is problematic to directly use the L_p-norm ($p < \infty$) to construct the test statistics in high dimensions. For example, when $d/n \not\to 0$, Hotelling's T^2 test (L_2-type) performs poorly, as the Pearson's sample covariance matrices no longer converge to their population counterparts under the spectral norm ([5]). For high dimensional testing problems, we need to adjust the test statistics or make structured assumptions on the population covariance matrix to obtain better asymptotic distributions of the test statistics. We face the same problem while using L_p-norm ($p < \infty$) to construct the test statistics. Hence, to avoid making unnecessary assumptions on the covariance structure of the random vector, we introduce the (s_0, p)-norm to adjust the original L_p-norm. As is shown by numerical simulations in Section 4, the L_p-norm based test with small p has significant power loss when the dimension of the parameter of interest $q \to \infty$. The introduction of s_0 can boost the power of L_p-norm based test especially for small p. More specifically, when p is small, the L_p-norm accumulates noise from all the entries, which leads to significant power loss. By exploiting the (s_0, p)-norm, we can enhance the signal-noise ratio for the obtained test statistics. When p is large, the choice of s_0 becomes less critical. In theory, for the bootstrap scheme to work properly under any $1 \le p \le \infty$, we require that $s_0^2 \log(qn) = O(n^\delta)$ holds for some $0 < \delta < 1/7$. Therefore, s_0 can also go to the infinity as $n \to \infty$. By simulation, s_0 close to s, which is the true unknown number of entries violating $\mathbf{H}_0$, is preferable. More details on the choice of s_0 are provided in Section 3 and 4.

2.1.2. *The (s_0, p)-norm based test statistics.* Before presenting the test statistics, we first introduce the following jackknife variance estimator for the U-statistic $\widehat{u}_{\gamma,s}$ defined in (1.1) with $\gamma = 1, 2$ and $s = 1, 2 \ldots, q$. As $m \ge 2$, we define

$$\widehat{v}_{1,s} = m^2 n_1^{-1} \sum_{k=1}^{n_1} (Q_{1k,s} - \widehat{u}_{1,s})^2, \ \widehat{v}_{2,s} = m^2 n_2^{-1} \sum_{k=1}^{n_2} (Q_{2k,s} - \widehat{u}_{2,s})^2, \tag{2.1}$$

where we set

$$\begin{aligned} Q_{1k,s} &:= \binom{n_1-1}{m-1}^{-1} \sum_{\substack{1\le \ell_1 < \cdots < \ell_{m-1} \le n_1 \\ \ell_j \ne k, j=1,\ldots,m-1}} \Phi_s(\boldsymbol{X}_k, \boldsymbol{X}_{\ell_1}, \ldots, \boldsymbol{X}_{\ell_{m-1}}), \\ Q_{2k,s} &:= \binom{n_2-1}{m-1}^{-1} \sum_{\substack{1\le \ell_1 < \cdots < \ell_{m-1} \le n_2 \\ \ell_j \ne k, j=1,\ldots,m-1}} \Phi_s(\boldsymbol{Y}_k, \boldsymbol{Y}_{\ell_1}, \ldots, \boldsymbol{Y}_{\ell_{m-1}}). \end{aligned} \tag{2.2}$$

We use $\widehat{v}_{\gamma,s}$ to estimate the variance of $\sqrt{n_\gamma}\widehat{u}_{\gamma,s}$. Therefore, $\widehat{v}_{\gamma,s}/n_\gamma$ is the variance estimator for $\widehat{u}_{\gamma,s}$. As $m = 1$, $\widehat{u}_{\gamma,s}$ and $\widehat{v}_{\gamma,s}$ are reduced to

$$\left\{ \begin{aligned} \widehat{u}_{1,s} &= n_1^{-1} \sum_{k=1}^{n_1} \Phi_s(\boldsymbol{X}_k), \\ \widehat{u}_{2,s} &= n_2^{-1} \sum_{k=1}^{n_2} \Phi_s(\boldsymbol{Y}_k), \end{aligned} \right. \quad \left\{ \begin{aligned} \widehat{v}_{1,s} &= n_1^{-1} \sum_{k=1}^{n_1} (\Phi_s(\boldsymbol{X}_k) - \widehat{u}_{1,s})^2, \\ \widehat{v}_{2,s} &= n_2^{-1} \sum_{k=1}^{n_2} (\Phi_s(\boldsymbol{X}_k) - \widehat{u}_{2,s})^2. \end{aligned} \right. \tag{2.3}$$

After introducing these notations, we present our (s_0, p)-norm based test statistics. For this, we define $\boldsymbol{W} = (W_1, \dots, W_q)^\top$ and $\boldsymbol{N} = (N_1, \dots, N_q)^\top$, where we set W_s and N_s as

$$
\begin{aligned}
W_s &:= (\widehat{u}_{1,s} - u_{0,s})/\sqrt{\widehat{v}_{1,s}/n_1}, \\
N_s &:= (\widehat{u}_{1,s} - \widehat{u}_{2,s})/\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}.
\end{aligned} \tag{2.4}
$$

For the one-sample problem in (1.7), we propose the test statistic $W_{(s_0,p)} := \|\boldsymbol{W}\|_{(s_0,p)}$. Similarly, for the two-sample problem in (1.8), we propose the test statistic $N_{(s_0,p)} := \|\boldsymbol{N}\|_{(s_0,p)}$. Throughout this paper, if not specially specified, we require $1 \le p \le \infty$ to make $\|\cdot\|_{(s_0,p)}$ a norm, which is also required by the theory.

2.1.3. *Bootstrap procedure for the asymptotic distribution.* In the high dimensional setting, [23] introduce the multiplier bootstrap method for the sum of independent random vectors. In detail, let $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ be independent random vectors in $\mathbb{R}^d$ with $\boldsymbol{Z}_k = (Z_{k1}, \dots, Z_{kd})^\top$ and $\mathbb{E}[\boldsymbol{Z}_k] = \boldsymbol{0}$ for $k = 1, \dots, n$. Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ be independent standard normal random variables, the multiplier bootstrap sample for $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ is $\varepsilon_1\boldsymbol{Z}_1, \dots, \varepsilon_n\boldsymbol{Z}_n$. The bootstrap sample for the sample mean $n^{-1}\sum_{k=1}^n \boldsymbol{Z}_k$ then becomes $n^{-1}\sum_{k=1}^n \varepsilon_k\boldsymbol{Z}_k$. To fully utilyze this result, we use multiplier bootstrap scheme for for high dimensional U-statistics. In detail, we generate independent samples $\varepsilon^b_{1,1}, \dots, \varepsilon^b_{1,n_1}$ and $\varepsilon^b_{2,1}, \dots, \varepsilon^b_{2,n_2}$ from $\varepsilon \sim N(0,1)$ for $b = 1, \dots, B$ and set

$$
\begin{aligned}
\widehat{u}^b_{1,s} &= \binom{n_1}{m}^{-1} \sum_{1 \le k_1 < \dots < k_m \le n_1} (\varepsilon^b_{1,k_1} + \dots + \varepsilon^b_{1,k_m})\big(\Phi_s(\boldsymbol{X}_{k_1}, \dots, \boldsymbol{X}_{k_m}) - \widehat{u}_{1,s}\big), \\
\widehat{u}^b_{2,s} &= \binom{n_2}{m}^{-1} \sum_{1 \le k_1 < \dots < k_m \le n_2} (\varepsilon^b_{2,k_1} + \dots + \varepsilon^b_{2,k_m})\big(\Phi_s(\boldsymbol{Y}_{k_1}, \dots, \boldsymbol{Y}_{k_m}) - \widehat{u}_{2,s}\big).
\end{aligned}
$$

Correspondingly, we set $\widehat{\boldsymbol{u}}^b_\gamma := (\widehat{u}^b_{\gamma,1}, \dots, \widehat{u}^b_{\gamma,q})^\top$ for $\gamma = 1, 2$. After introducing $\widehat{\boldsymbol{u}}^b_\gamma$, we define $\boldsymbol{W}^b = (W^b_1, \dots, W^b_q)^\top$ and $\boldsymbol{N}^b = (N^b_1, \dots, N^b_q)^\top$, where

$$
W^b_s = \widehat{u}^b_{1,s}/\sqrt{\widehat{v}_{1,s}/n_1}, \; N^b_s = (\widehat{u}^b_{1,s} - \widehat{u}^b_{2,s})/\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}. \tag{2.5}
$$

Bootstrap samples become $\{N^b_{(s_0,p)}\}_{b=1,\dots,B}$ and $\{W^b_{(s_0,p)}\}_{b=1,\dots,B}$ with

$$
W^b_{(s_0,p)} = \|\boldsymbol{W}^b\|_{(s_0,p)} \quad \text{and} \quad N^b_{(s_0,p)} = \|\boldsymbol{N}^b\|_{(s_0,p)}. \tag{2.6}
$$

Given the significance level α and the bootstrap samples, we set the critical values of $W_{(s_0,p)}$ and $N_{(s_0,p)}$ as

$$\widehat{t}^{W}_{\alpha,(s_0,p)} = \inf\Big\{t \in \mathbb{R} : \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\{W^b_{(s_0,p)} \le t\} > 1-\alpha\Big\},$$
$$\widehat{t}^{N}_{\alpha,(s_0,p)} = \inf\Big\{t \in \mathbb{R} : \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\{N^b_{(s_0,p)} \le t\} > 1-\alpha\Big\}.$$

Therefore, we obtain the (s_0, p)-norm based tests for (1.7) and (1.8) as

$$T^{W}_{\alpha,(s_0,p)} := \mathbb{I}\left\{W_{(s_0,p)} \ge \widehat{t}^{W}_{\alpha,(s_0,p)}\right\},\; T^{N}_{\alpha,(s_0,p)} := \mathbb{I}\left\{N_{(s_0,p)} \ge \widehat{t}^{N}_{\alpha,(s_0,p)}\right\}. \tag{2.7}$$

We reject $\mathbf{H}_0$ of (1.7) if and only if $T^{W}_{\alpha,(s_0,p)} = 1$ and reject $\mathbf{H}_0$ of (1.8) if and only if $T^{N}_{\alpha,(s_0,p)} = 1$. Accordingly, we estimate $W_{(s_0,p)}$ and $N_{(s_0,p)}$'s oracle P-values $P^{W}_{(s_0,p)}$ and $P^{N}_{(s_0,p)}$ by

$$\begin{aligned}\widehat{P}^{W}_{(s_0,p)} &= (B+1)^{-1}\sum_{b=1}^{B} \mathbb{I}\{W^b_{(s_0,p)} > W_{(s_0,p)}\}\\ \widehat{P}^{N}_{(s_0,p)} &= (B+1)^{-1}\sum_{b=1}^{B} \mathbb{I}\{N^b_{(s_0,p)} > N_{(s_0,p)}\}.\end{aligned} \tag{2.8}$$

Therefore, given a significance level α, we reject $\mathbf{H}_0$ of (1.7) if and only if $\widehat{P}^{W}_{(s_0,p)} \le \alpha$ and reject $\mathbf{H}_0$ of (1.8) if and only if $\widehat{P}^{N}_{(s_0,p)} \le \alpha$.

2.2. *Data-adaptive combined test .* We now introduce the data-adaptive combined test. In Section 2.2.1, we present the test procedure. In Section 2.2.2, we introduce a double-loop bootstrap procedure to obtain the P-value of the data-adaptive test. To reduce the expensive computation cost of the double-loop bootstrap procedure, in Section 2.2.3 we introduce a low cost bootstrap procedure which obtains nonindependent bootstrap samples. The theory of this new low cost bootstrap procedure is provided in Section 3.3.

2.2.1. *Test statistics.* $W_{(s_0,p)}$ and $N_{(s_0,p)}$ have different powers for different p and alternative scenarios. For example, $W_{(s_0,\infty)}$ and $N_{(s_0,\infty)}$ are sensitive to large perturbations on a small number of entries of $\boldsymbol{u}_1 - \boldsymbol{u}_0$ and $\boldsymbol{u}_1 - \boldsymbol{u}_2$. Moreover, $W_{(s_0,2)}$ and $N_{(s_0,2)}$ are sensitive to small perturbations on a large number of entries of $\boldsymbol{u}_1 - \boldsymbol{u}_0$ and $\boldsymbol{u}_1 - \boldsymbol{u}_2$. We aim to combine these tests to construct a data-adaptive test which is simultaneously powerful under different alternatives.

For the one-sample problem, as small P-values of $W_{(s_0,p)}$ lead to the rejection of $\mathbf{H}_0$ in (1.7), we construct the data-adaptive test statistic W_{ad} by taking the minimum of P-values of all individual tests, i.e.,

$$W_{\mathrm{ad}} = \min_{p\in\mathcal{P}} \widehat{P}^{W}_{(s_0,p)}, \tag{2.9}$$

where $\mathcal{P} \subset \{1,2,\ldots,\infty\}$ is a candidate set of p. A bootstrap procedure to obtain W_{ad} is described in Algorithm 1. For the two-sample problem in (1.8), we construct the data-adaptive test statistic N_{ad} as

$$N_{\mathrm{ad}} = \min_{p\in\mathcal{P}} \widehat{P}^{N}_{(s_0,p)}. \tag{2.10}$$

Throughout this paper, we require that $\#(\mathcal{P}) < \infty$ is a fixed constant, which is also required by the theory and discussed in Section 3.3. If the alternative pattern is unknown, we recommend using the balanced $\mathcal{P}$ including both small and large values of $p \in [1,\infty]$. For example, $\mathcal{P} = \{1,2,\ldots,5,\infty\}$ is used in the later simulation experiments. If the alternative pattern is known, we can boost the power of the data-adaptive combined test by choosing $\mathcal{P}$ accordingly. For example, for possible sparse alternatives, $\mathcal{P}$ should consist of large values of p.

Algorithm 1 A bootstrap procedure to obtain W_{ad}

Input: $\mathcal{X}$.
Output: $W^1_{(s_0,p)},\ldots,W^B_{(s_0,p)}$ with $p\in\mathcal{P}$, and W_{ad}.
procedure
  $W_{(s_0,p)} = \|\boldsymbol{W}\|_{(s_0,p)}$ with $\boldsymbol{W} = (W_1,\ldots,W_q)^\top$ and $W_s = (\widehat{u}_{1,s} - u_{0,s})/\sqrt{\widehat{v}_{1,s}/n_1}$.
  for $b \leftarrow 1$ **to** B **do**
    Sample independent standard normal random variables $\{\varepsilon^b_{1,1},\ldots,\varepsilon^b_{1,n_1}\}$.
    $\widehat{u}^b_{1,s} = \binom{n_1}{m}^{-1} \sum_{1\le k_1<\cdots<k_m\le n_1} (\varepsilon^b_{1,k_1}+\cdots+\varepsilon^b_{1,k_m})\big(\Phi_s(\boldsymbol{X}_{k_1},\ldots,\boldsymbol{X}_{k_m}) - \widehat{u}_{1,s}\big)$.
    $W^b_s = \widehat{u}^b_{1,s}/\sqrt{\widehat{v}_{1,s}/n_1}$ for $s = 1,\ldots,q$.
    for p **in** $\mathcal{P}$ **do**
      $W^b_{(s_0,p)} = \|\boldsymbol{W}^b\|_{(s_0,p)}$ with $\boldsymbol{W}^b = (W^b_1,\ldots,W^b_q)^\top$.
    end for
  end for
  $\widehat{P}^W_{(s_0,p)} = \sum_{b=1}^B \mathbb{1}\{W^b_{(s_0,p)} > W_{(s_0,p)}\}/(B+1)$ for $p\in\mathcal{P}$.
  $W_{\mathrm{ad}} = \min_{p\in\mathcal{P}} \widehat{P}^W_{(s_0,p)}$.
end procedure

2.2.2. *Double-loop bootstrap procedure.* We present how to obtain P-value of W_{ad}. By setting $F_{W,\mathrm{ad}}(x)$ as the distribution function of W_{ad}, W_{ad}'s oracle P-value becomes $F_{W,\mathrm{ad}}(W_{\mathrm{ad}})$. As $F_{W,\mathrm{ad}}(x)$ is unknown, we need to use the bootstrap method to estimate it, which leads to a double-loop bootstrap procedure. In the outer loop, by Algorithm 1 we obtain the bootstrap samples for $W_{(s_0,p)}$, i.e, $\{W^1_{(s_0,p)}, \ldots, W^B_{(s_0,p)}\}$. In the inner loop, for each $b \in \{1, \ldots, B\}$, we use Algorithm 2 to obtain bootstrap samples for $W^b_{(s_0,p)}$, i.e., $\{W^{b,1}_{(s_0,p)}, \ldots, W^{b,L}_{(s_0,p)}\}$, and construct the bootstrap samples for W_{ad} as

$$W^b_{\mathrm{ad}} = \min_{p \in \mathcal{P}} \frac{\sum_{\ell=1}^{L} \mathbb{I}\{W^{b,\ell}_{(s_0,p)} > W^b_{(s_0,p)}\}}{L+1} \qquad \text{for} \qquad b = 1, \ldots, B.$$

With the bootstrap samples, we can estimate the oracle P-value of W_{ad} by

$$\frac{1}{B+1}\left(\Big(\sum_{b=1}^{B} \mathbb{I}\{W^b_{\mathrm{ad}} \le W_{\mathrm{ad}}\}\Big) + 1\right).$$

Figure 1 illustrates this double-loop bootstrap method. By this double-loop bootstrap procedure, to guarantee the independence of $W^1_{\mathrm{ad}}, \ldots, W^B_{\mathrm{ad}}$, we totally need $LB + B$ samples from (2.1.3), which is computationally expensive when L and B are large.

Algorithm 2 A double-loop bootstrap procedure to obtain bootstrap samples of W_{ad}

Input: $\mathcal{X}$ and $W^1_{(s_0,p)}, \ldots, W^B_{(s_0,p)}$ for $p \in \mathcal{P}$.
Output: $W^1_{\mathrm{ad}}, \ldots, W^B_{\mathrm{ad}}$.

procedure
  for $b \leftarrow 1$ **to** B **do**
    for $\ell \leftarrow 1$ **to** L **do**
      Sample independent standard normal random variables $\{\varepsilon^{b,\ell}_{1,1}, \ldots, \varepsilon^{b,\ell}_{1,n_1}\}$.
      $\widehat{u}^{b,\ell}_{1,s} = \binom{n_1}{m}^{-1} \sum_{1 \le k_1 < \cdots < k_m \le n_1} (\varepsilon^{b,\ell}_{1,k_1} + \cdots + \varepsilon^{b,\ell}_{1,k_m})\big(\Phi_s(\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m}) - \widehat{u}_{1,s}\big)$.
      $W^{b,\ell}_s = \widehat{u}^{b,\ell}_{1,s} / \sqrt{\widehat{v}_{1,s}/n_1}$ for $s = 1, \ldots, q$.
      for p **in** $\mathcal{P}$ **do**
        $W^{b,\ell}_{(s_0,p)} = \|\boldsymbol{W}^{b,\ell}\|_{(s_0,p)}$ with $\boldsymbol{W}^{b,\ell} = (W^{b,\ell}_1, \ldots, W^{b,\ell}_q)^\top$.
      end for
    end for
    $\widehat{P}_{W^b,(s_0,p)} = \sum_{\ell=1}^{L} \mathbb{I}\{W^{b,\ell}_{(s_0,p)} > W^b_{(s_0,p)}\}/(L+1)$ for $p \in \mathcal{P}$.
    $W^b_{\mathrm{ad}} = \min_{p \in \mathcal{P}} \widehat{P}_{W^b,(s_0,p)}$
  end for
end procedure

2.2.3. *A low cost bootstrap procedure.* To handle the computational bottleneck of the double-loop bootstrap, we propose to replace Algorithm 2 with Algorithm 3, which is computationally more efficient but obtains non-independent bootstrap samples for W_{ad}, denoted as $\{W^1_{\mathrm{ad}'}, \ldots, W^B_{\mathrm{ad}'}\}$.

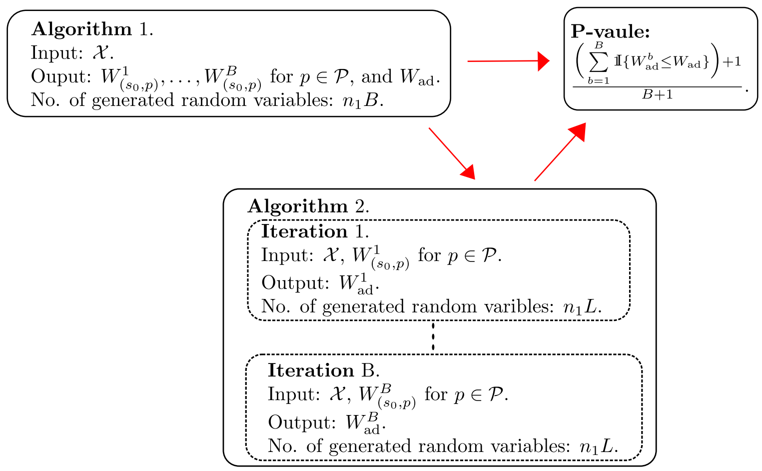

Fig 1. *Flowchart for the double-loop bootstrap procedure and total number of generated standard normal random variables.*

In detail, in Algorithm 1 by (2.1.3), (2.5), and (2.6) we generate bootstrap samples for $W_{(s_0,p)}$, i.e., $W^1_{(s_0,p)}, \dots, W^B_{(s_0,p)}$. To avoid the double-loop bootstrap procedure, we need to more effectively utilize the generated bootstrap samples $W^1_{(s_0,p)}, \dots, W^B_{(s_0,p)}$. For this, we set

$$\widehat{P}^{b,W}_{(s_0,p)} = \frac{\sum_{b_1 \neq b} \mathbb{I}\{W^{b_1}_{(s_0,p)} > W^b_{(s_0,p)}\}}{B} \text{ for } b = 1, \dots, B \quad \text{and} \quad p \in \mathcal{P}.$$

We use $W^b_{\text{ad}'} = \min_{p \in \mathcal{P}} \widehat{P}^{b,W}_{(s_0,p)}$ as the bootstrap sample for W_{ad}, and estimate the oracle P-value by

$$\widehat{P}^W_{\text{ad}} = \frac{\left(\sum_{b=1}^{B} \mathbb{I}\{W^b_{\text{ad}'} \le W_{\text{ad}}\}\right) + 1}{B+1}. \tag{2.11}$$

The samples $W^1_{\text{ad}'}, \dots, W^B_{\text{ad}'}$ are nonindependent. However, we can prove that they are asymptotically independent as $n_1, B \to \infty$, which plays a pivotal role in proving the consistency of $\widehat{P}^W_{\text{ad}}$.

Figure 2 illustrates the process of the low cost bootstrap procedure. To obtain the P-value of W_{ad}, we don't need to generate new bootstrap samples. In total, to perform the data-adaptive test we only need to generate B bootstrap samples from (2.1.3).

We similarly deal with the two-sample problem. By generating bootstrap samples for $N_{(s_0,p)}$, i.e., $N^1_{(s_0,p)}, \dots, N^B_{(s_0,p)}$ and setting

$$\widehat{P}^{b,N}_{(s_0,p)} = \frac{\sum_{b_1 \neq b} \mathbb{I}\{N^{b_1}_{(s_0,p)} > N^b_{(s_0,p)}\}}{B} \text{ for } b = 1, \dots, B \text{ and } p \in \mathcal{P}, \tag{2.12}$$

we use $N^b_{\text{ad}'} = \min_{p \in \mathcal{P}} \widehat{P}^{b,N}_{(s_0,p)}$ as the bootstrap sample of N_{ad}. Therefore, we can similarly estimate the oracle P-value of N_{ad} by

$$\widehat{P}^N_{\text{ad}} = \frac{\left(\sum_{b=1}^B \mathbb{I}\{N^b_{\text{ad}'} \le N_{\text{ad}}\}\right) + 1}{B+1}. \tag{2.13}$$

With the estimated P-values of the data-adaptive tests W_{ad} and N_{ad}, given significance level α, we reject $\mathbf{H}_0$ of (1.7) if and only if $\widehat{P}^W_{\text{ad}} \le \alpha$ and reject $\mathbf{H}_0$ of (1.8) if and only if $\widehat{P}^N_{\text{ad}} \le \alpha$. Therefore, we set

$$T^W_{\text{ad}} = \mathbb{I}\{\widehat{P}^W_{\text{ad}} \le \alpha\} \quad \text{and} \quad T^N_{\text{ad}} = \mathbb{I}\{\widehat{P}^N_{\text{ad}} \le \alpha\}. \tag{2.14}$$

Algorithm 3 A low cost bootstrap procedure

Input: $\mathcal{X}$ and $W^1_{(s_0,p)}, \dots, W^B_{(s_0,p)}$ for $p \in \mathcal{P}$.
Output: $W^1_{\text{ad}'}, \dots, W^B_{\text{ad}'}$.

```
procedure
    for b ← 1 to B do
        for p in 𝒫 do
            P̂^{b,W}_{(s_0,p)} = Σ_{b_1≠b} 𝕀{W^{b_1}_{(s_0,p)} > W^b_{(s_0,p)}}/B
        end for
        W^b_{ad'} = min_{p∈𝒫} P̂^{b,W}_{(s_0,p)}.
    end for
end procedure
```

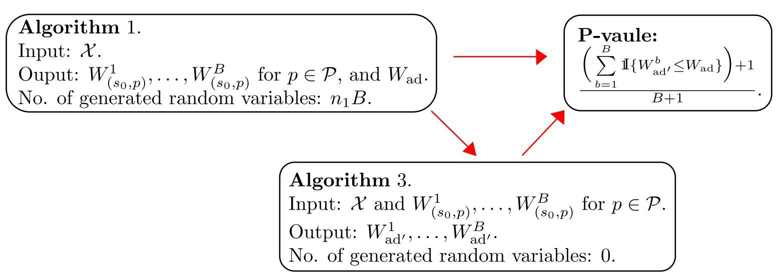

FIG 2. *Flowchart for the low cost bootstrap procedure with low computation cost and total number of generated standard normal random variables.*

REMARK 2.1. To construct test statistics $W_{(s_0,p)}$ and $N_{(s_0,p)}$, we normalize $\widehat{u}_{1,s} - u_{1,s}$ and $\widehat{u}_{1,s} - \widehat{u}_{2,s}$ by dividing their standard deviation estimators. If we assume that U-statistics have the same variance under the null hypothesis (homogeneity assumption), we can build $W_{(s_0,p)}$ and $N_{(s_0,p)}$ without the normalization to avoid introducing unnecessary estimation error. Therefore, W_s and N_s become

$$W_s := \widehat{u}_{1,s} - u_{0,s} \qquad \text{and} \qquad N_s := \widehat{u}_{1,s} - \widehat{u}_{2,s}.$$

For the same reason, we set $W_s^b = \widehat{u}_{1,s}^b$ and $N_s^b = \widehat{u}_{1,s}^b - \widehat{u}_{2,s}^b$ when performing bootstrap procedure of Sections 2.1.3 and 2.2. As the proof is similar for the test statistics without normalization, in Section 3 we only analyze the theoretical properties of the test statistics with normalization.

3. Theoretical properties.

In this section, we discuss the theoretical properties of the proposed testing methods including the (s_0, p)-norm based test and data-adaptive combined test. We first introduce several assumptions in Section 3.1. We then analyze the asymptotic size and power of the (s_0, p)-norm based test in Section 3.2. At last, we analyze the data-adaptive combined test in Section 3.3.

3.1. *Assumptions.*

Before presenting the theoretical properties, we introduce the assumptions that are needed in this paper. We also explain the intuitions of these assumptions. Throughout this paper, for the two-sample problem, we assume $n_1 \asymp n_2 \asymp n := \max(n_1, n_2)$, which means that n_1, n_2, and n are of the same order. We then introduce some other assumptions. Assumption **(A)** characterizes the scaling of s_0, q, and n. Assumptions **(E)**, **(M1)** and **(M2)** specify the requirements of the kernel functions. In detail, we introduce Assumption **(A)** as follows.

- **(A)** For the one-sample problem in (1.7), we assume that there is some $0 < \delta < 1/7$ such that $s_0^2 \log(q) = O(n_1^{\delta})$ holds. For the two-sample problem in (1.8), we similarly assume that there is some $0 < \delta < 1/7$ such that $s_0^2 \log q = O(n^{\delta})$ holds.

Assumptions **(A)** also allows q and s_0 to go to the infinity, as long as $s_0^2 \log(qn) = o(n^{\delta})$ holds with some $0 < \delta < 1/7$.

We then introduce the assumptions on the kernel functions of the U-statistics. For $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^d$, define

$$\boldsymbol{\Psi}(\mathbf{x}_1, \ldots, \mathbf{x}_m) := \big(\Psi_1(\mathbf{x}_1, \ldots, \mathbf{x}_m), \ldots, \Psi_q(\mathbf{x}_1, \ldots, \mathbf{x}_m)\big)^\top$$
$$\boldsymbol{h}(\mathbf{x}) := \big(h_1(\mathbf{x}), \ldots, h_q(\mathbf{x})\big)^\top,$$

where Ψ_s and h_s are

$$\begin{aligned} \Psi_s(\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m}) &= \Phi_s(\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m}) - u_{1,s} \\ h_s(\boldsymbol{X}_k) &= \mathbb{E}[\Psi_s(\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m})|\boldsymbol{X}_k]. \end{aligned} \tag{3.1}$$

Also, set $\mathcal{V}_{s_0} := \{\mathbf{v} \in \mathbb{S}^{q-1} : \|\mathbf{v}\|_0 \leq s_0\}$. With these introduced notations, by setting $0 < K, b < \infty$ as some positive constants, we are now ready to state Assumptions **(E)**, **(M1)**, and **(M2)**.

- **(E)** For different indexes $0 < i_1, \ldots, i_m < n_1$ and $0 < j_1, \ldots, j_m < n_2$, we require

$$\max_{1\leq s\leq q} \mathbb{E}\big[\exp\big(|\Psi_s(\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_m})|/K\big)\big] \leq 2,$$
$$\max_{1\leq s\leq q} \mathbb{E}\big[\exp\big(|\Psi_s(\boldsymbol{Y}_{j_1}, \ldots, \boldsymbol{Y}_{j_m})|/K\big)\big] \leq 2.$$

- **(M1)** $\mathbb{E}[|\mathbf{v}^\top \boldsymbol{h}(\boldsymbol{X})|^2] \geq b$ and $\mathbb{E}[|\mathbf{v}^\top \boldsymbol{h}(\boldsymbol{Y})|^2] \geq b$ hold for any $\mathbf{v} \in \mathcal{V}_{s_0}$.
- **(M2)** For $\ell = 1, 2$, we require

$$\max_{1\leq s\leq q} \mathbb{E}[|h_s(\boldsymbol{X})|^{2+\ell}] \leq K^\ell, \quad \max_{1\leq s\leq q} \mathbb{E}[|h_s(\boldsymbol{Y})|^{2+\ell}] \leq K^\ell.$$

Assumption **(E)** requires that $\Psi_s(\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_m})$ and $\Psi_s(\boldsymbol{Y}_{j_1}, \ldots, \boldsymbol{Y}_{j_s})$ follow the sub-exponential distribution. Especially, bounded Ψ_s including useful rank-based U-statistics such as Kendall's tau and Spearman's rho satisfy this condition. Assumption **(M1)** excludes degenerate U-statistics. Moreover, it also requires that the inner product of $\boldsymbol{h}(\boldsymbol{X})$ (or $\boldsymbol{h}(\boldsymbol{Y})$) and any $\mathbf{v} \in \mathcal{V}_{s_0}$ is not degenerated. The distribution assumptions **(E)**, **(M1)**, and **(M2)** are useful for applying high-dimensional central limiting theorem (CLT) in Lemma A.1. These assumptions are also justified by [24].

3.2. *Theoretical properties of* (s_0, p)*-norm based test statistics.* After introducing the assumptions in Section 3.1, we now state the theoretical properties of the (s_0, p)-norm based test. Firstly, we consider the asymptotic size. The following theorem justifies the multiplier bootstrap for $W_{(s_0,p)}$ and $N_{(s_0,p)}$, which is crucial for the size control.

THEOREM 3.1. Suppose all assumptions in Section 3.1 hold. Under $\mathbf{H}_0$ of (1.7), we have

$$\sup_{z\in(0,\infty)} \Big|\mathbb{P}(W_{(s_0,p)} \leq z) - \mathbb{P}(W^b_{(s_0,p)} \leq z|\mathcal{X})\Big| = o_p(1), \text{ as } n_1 \to \infty. \tag{3.2}$$

Similarly, under $\mathbf{H}_0$ of (1.8) we have

$$\sup_{z\in(0,\infty)} \Big|\mathbb{P}(N_{(s_0,p)} \leq z) - \mathbb{P}(N^b_{(s_0,p)} \leq z|\mathcal{X}, \mathcal{Y})\Big| = o_p(1), \text{ as } n \to \infty. \tag{3.3}$$

PROOF. The proof of (3.2) is similar to that of (3.3). For simplicity, we only present the proof of (3.3), which consists of three steps. We first analyze the approximate distribution of $\boldsymbol{N}$. We then obtain the distribution of the bootstrap sample $\boldsymbol{N}^b$ given $\mathcal{X}$ and $\mathcal{Y}$. At last, we analyze the approximation error between $\boldsymbol{N}$ and $\boldsymbol{N}^b|\mathcal{X},\mathcal{Y}$ to yield (3.3). We only sketch the proof here. More detailed proof is presented in Appendix B.2 of supplementary materials.

Step (i) (Sketch). In this step, we aim to obtain the approximate distribution of $\boldsymbol{N}$ under the null hypothesis. Under the null hypothesis we have $u_{1,s} = u_{2,s}$. Therefore, we rewrite N_s as

$$N_s = (\widetilde{u}_{1,s} - \widetilde{u}_{2,s})/\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2},$$

where $\widetilde{u}_{\gamma,s} := \widehat{u}_{\gamma,s} - u_{\gamma,s}$ is the centered version of $\widehat{u}_{\gamma,s}$. As $\widetilde{u}_{\gamma,s}$ is also a U-statistic, by the Hoeffding's decomposition we can approximate $\widetilde{u}_{\gamma,s}$ by a sum of independent random variables. In detail, we use $(m/n_1)\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k)$ and $(m/n_2)\sum_{k=1}^{n_2} h_s(\boldsymbol{Y}_k)$ to approximate $\widetilde{u}_{1,s}$ and $\widetilde{u}_{2,s}$. By setting

$$\sigma_{1,st} = \mathbb{E}\big(h_s(\boldsymbol{X})h_t(\boldsymbol{X})\big) \quad \text{and} \quad \sigma_{2,st} = \mathbb{E}\big(h_s(\boldsymbol{Y})h_t(\boldsymbol{Y})\big) \tag{3.4}$$

for $1 \le s,t \le q$, as $n \to \infty$ we have $\widehat{v}_{\gamma,s} \to m^2\sigma_{\gamma,ss}$, which motivates us to define

$$H_s^N = \Big(\frac{1}{n_1}\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k) - \frac{1}{n_2}\sum_{k=1}^{n_2} h_s(\boldsymbol{Y}_k)\Big)\Big/\sqrt{\sigma_{1,ss}/n_1 + \sigma_{2,ss}/n_2}. \tag{3.5}$$

By setting $\boldsymbol{H}^N = (H_1^N,\dots,H_q^N)^\top$, we use $\boldsymbol{H}^N$ as an approximation of $\boldsymbol{N}$. However, we don't know the exact distribution of $\boldsymbol{H}^N$. As $\boldsymbol{H}^N$ is a sum of independent random vectors with zero mean, by the central limit theorem we can use a normal random vector to further approximate $\boldsymbol{H}^N$.

Let $\boldsymbol{G}^N$ be a Gaussian random vector with the same mean vector and covariance matrix as $\boldsymbol{H}^N$. By setting $\boldsymbol{\Sigma}_1 := (\sigma_{1,st}), \boldsymbol{\Sigma}_2 := (\sigma_{2,st}) \in \mathbb{R}^{q\times q}$, we have

$$\boldsymbol{G}^N \sim N(\boldsymbol{0}, \mathbf{R}_{12}) \quad \text{with} \quad \mathbf{R}_{12} := \mathbf{D}_{12}^{-1/2}\boldsymbol{\Sigma}_{12}\mathbf{D}_{12}^{-1/2}, \tag{3.6}$$

where we set

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_1/n_1 + \boldsymbol{\Sigma}_2/n_2 \quad \text{and} \quad \mathbf{D}_{12} = \mathrm{Diag}(\boldsymbol{\Sigma}_{12}). \tag{3.7}$$

We then use the distribution of $\boldsymbol{G}^N$ to approximate that of $\boldsymbol{N}$.

Step (ii) (Sketch). In this step, we aim to obtain the distribution of $\boldsymbol{N}^b|\mathcal{X},\mathcal{Y}$. We rewrite $\widehat{u}^b_{1,s}$ and $\widehat{u}^b_{2,s}$ in (2.1.3) as

$$(3.8)\qquad \widehat{u}^b_{1,s} = \frac{m}{n_1}\sum_{k=1}^{n_1}(Q_{1k,s} - \widehat{u}_{1,s})\varepsilon^b_{1,k},\ \ \widehat{u}^b_{2,s} = \frac{m}{n_2}\sum_{k=1}^{n_2}(Q_{2k,s} - \widehat{u}_{2,s})\varepsilon^b_{2,k},$$

where $Q_{1k,s}$ and $Q_{2k,s}$ are defined in (2.2). Considering that $\varepsilon^b_{\gamma,1},\ldots,\varepsilon^b_{\gamma,n_1}$ are independent standard normal random variables, by (3.8) we have $\widehat{\boldsymbol{u}}^b_\gamma := (\widehat{u}^b_{\gamma,1},\ldots,\widehat{u}^b_{\gamma,q})|\mathcal{X},\mathcal{Y} \sim N(\mathbf{0}, m^2\widehat{\boldsymbol{\Sigma}}_\gamma/n_\gamma)$ with $\widehat{\boldsymbol{\Sigma}}_\gamma := (\widehat{\sigma}_{\gamma,st}) \in \mathbb{R}^{q\times q}$ and

$$(3.9)\qquad \widehat{\sigma}_{\gamma,st} = \frac{1}{n_\gamma}\sum_{k=1}^{n_1}(Q_{\gamma k,s} - \widehat{u}_{\gamma,s})(Q_{\gamma k,t} - \widehat{u}_{\gamma,t}),$$

for $\gamma = 1, 2$. Apparently, by the definition of $\widehat{v}_{\gamma,s}$ in (2.1) we have $\widehat{v}_{\gamma,s} = m^2\widehat{\sigma}_{\gamma,ss}$. By setting

$$\widehat{\boldsymbol{\Sigma}}_{12} = \widehat{\boldsymbol{\Sigma}}_1/n_1 + \widehat{\boldsymbol{\Sigma}}_2/n_2 \quad \text{and} \quad \widehat{\mathbf{D}}_{12} = \mathrm{Diag}(\widehat{\boldsymbol{\Sigma}}_{12}),$$

given $\mathcal{X}$ and $\mathcal{Y}$ we have

$$(3.10)\qquad \boldsymbol{N}^b = m^{-1}\widehat{\mathbf{D}}_{12}^{-1/2}(\widehat{\boldsymbol{u}}^b_1 - \widehat{\boldsymbol{u}}^b_2) \sim N(\mathbf{0}, \widehat{\mathbf{R}}_{12}),$$

where we set $\widehat{\mathbf{R}}_{12} = \widehat{\mathbf{D}}_{12}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\mathbf{D}}_{12}^{-1/2}$.

Step (iii) (Sketch). In this step, we aim to obtain the approximation error between $\boldsymbol{N}$ and $\boldsymbol{N}^b|\mathcal{X},\mathcal{Y}$. For this, we analyze the estimation error between $\widehat{\mathbf{R}}_{12}$ and $\mathbf{R}_{12}$. We then combine results from **Steps (i)** and **(ii)** to finish the proof of (3.3). The detailed proof is in Appendix B.2 of supplementary materials. □

REMARK 3.2. Assumption **(A)** requires that $s_0^\zeta\log(q) = O(n^\delta)$ holds with $\zeta = 2$ and $0 < \delta < 1/7$. However, $\zeta = 2$ is not optimal for each individual p. By the proof of Theorem 3.1, ζ depends on the facet number of a polytope to approximate $B_{(s_0,p)}(x) = \{\mathbf{v} \in \mathbb{R}^q : \|\mathbf{v}\|_{(s_0,p)} \le x\}$. If $p = 1$, $B_{(s_0,p)}(x)$ itself is a polytope, which makes $\zeta = 1$ is enough for obtaining (3.2) and (3.3). Similarly, If $p = \infty$, $\zeta = 0$ is sufficient. To make (3.2) and (3.3) hold for any $p \in [1,\infty]$, by Lemma A.3 in Appendix A, we set $\zeta = 2$ in Theorem 3.1.

As an implication of Theorems 3.1, the following corollary shows that under mild moment conditions on the kernel functions of U-statistics, by using the multiplier bootstrap introduced in Section 2.1.3, the size of (s_0,p)-norm based test is asymptotically α, as desired.

Corollary 3.1. Suppose all assumptions in Section 3.1 hold. For the one-sample problem in (1.7), under $\mathbf{H}_0$ of (1.7) we have

$$\mathbb{P}_{\mathbf{H}_0}\big(T^W_{\alpha,(s_0,p)} = 1\big) \to \alpha \ \text{ and } \ \widehat{P}^W_{(s_0,p)} - P^W_{(s_0,p)} \to 0, \text{ as } n_1,\, B \to \infty. \tag{3.11}$$

as $n_1, B \to \infty$. Similarly, for the two-sample problem in (1.8), under $\mathbf{H}_0$ of (1.8) we have

$$\mathbb{P}_{\mathbf{H}_0}\big(T^N_{\alpha,(s_0,p)} = 1\big) \to \alpha \ \text{ and } \widehat{P}^N_{(s_0,p)} - P^N_{(s_0,p)} \to 0, \text{ as } n,\, B \to \infty. \tag{3.12}$$

The detailed proof of Corollary 3.1 is in Appendix B.3 of supplementary materials. After analyzing the asymptotic size of the (s_0, p)-norm based test, we now turn to the analysis of its power. For this, we need the following notations: $\boldsymbol{D}_1 = (D_{1,1}, \ldots, D_{1,q})^\top$ and $\boldsymbol{D}_2 = (D_{2,1}, \ldots, D_{2,q})^\top$ with

$$\begin{aligned} D_{1,s} &= |u_{1,s} - u_{0,s}| / \sqrt{m^2\sigma_{1,ss}/n_1}, \\ D_{2,s} &= |u_{1,s} - u_{2,s}| / \sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}, \end{aligned} \tag{3.13}$$

where $\sigma_{\gamma,ss}$ is defined in (3.4). We need new Assumption $\mathbf{(A)}'$ to describe the scaling between s_0, q, and n for test statistics $W_{(s_0,p)}$ and $N_{(s_0,p)}$ to reject with overwhelming probability under the alternative.

- $\mathbf{(A)}'$ For the one-sample problem in (1.7), we assume $\log q = o(n_1^{1/3})$ and $n_1 = O(q^{\delta_1})$ with some $\delta_1 > 0$, as $n_1, q \to \infty$. For the two-sample problem in (1.8), we assume $\log q = o(n^{1/3})$ and $n = O(q^{\delta_1})$ with some $\delta_1 > 0$, as $n, q \to \infty$. Moreover, we also assume that there is a constant $\delta_2 > 0$ such that $s_0 = O(\log^{\delta_2}(q))$ holds for both problems.

After the introduction of Assumption $\mathbf{(A)}'$, we then state the theorem that characterizes the power of $W_{(s_0,p)}$ and $N_{(s_0,p)}$.

Theorem 3.3. Suppose Assumptions $\mathbf{(A)}'$, $\mathbf{(E)}$, $\mathbf{(M1)}$, and $\mathbf{(M2)}$ hold. For the one-sample problem in (1.7), we assume $\varepsilon_{n_1} = o(1)$ with $\varepsilon_{n_1}\sqrt{\log q} \to \infty$ as $n_1, q \to \infty$. If $\mathbf{H}_1$ of (1.7) holds with

$$\|\boldsymbol{D}_1\|_{(s_0,p)} \geq s_0(1 + \varepsilon_{n_1})\big(\sqrt{2\log q} + \sqrt{2\log(1/\alpha)}\big), \tag{3.14}$$

we have $\mathbb{P}_{\mathbf{H}_1}\big(T^W_{(s_0,p)} = 1\big) \to 1$ as $n_1, q, B \to \infty$. Similarly, for the two-sample problem in (1.8) we assume $\varepsilon_n = o(1)$, and $\varepsilon_n\sqrt{\log q} \to \infty$ as $n, q \to \infty$. If $\mathbf{H}_1$ of (1.8) holds with

$$\|\boldsymbol{D}_2\|_{(s_0,p)} \geq s_0(1 + \varepsilon_n)\big(\sqrt{2\log q} + \sqrt{2\log(1/\alpha)}\big), \tag{3.15}$$

we have $\mathbb{P}_{\mathbf{H}_1}\big(T^N_{(s_0,p)} = 1\big) \to 1$ as $n, q, B \to \infty$.

The detailed proof of Theorem 3.3 is presented in Appendix B.4 of supplementary materials. The scaling of q and n in Theorem 3.3 is weaker than Assumption **(A)**, allowing larger q for proposed tests to correctly reject the null hypothesis. Moreover, by the proof of Theorem 3.3, for $m = 1$, we can further relax the conditions $\log q = o(n_1^{1/3})$ and $\log q = o(n^{1/3})$ by $\log q = o(n_1^{1/2})$ and $\log q = o(n^{1/2})$ in Assumption $\mathbf{(A)}'$.

3.3. *Theoretical properties of* $W_{\rm ad}$ *and* $N_{\rm ad}$. In Section 2.2, we introduce the data-adaptive test by combining the (s_0, p)-norm based tests with $p \in \mathcal{P}$, where $\mathcal{P} \subset \{1, 2 \ldots, \infty\}$ is a finite fixed set specified by users. Intuitively, by combining tests with various norms, the data-adaptive test enjoys high power across various alternative hypothesis scenarios. In (2.9) and (2.10), we introduce the data-adaptive tests as

$$W_{\rm ad} = \min_{p \in \mathcal{P}} \widehat{P}^W_{(s_0,p)} \qquad \text{and} \qquad N_{\rm ad} = \min_{p \in \mathcal{P}} \widehat{P}^N_{(s_0,p)}, \tag{3.16}$$

where $\widehat{P}^W_{(s_0,p)}$ and $\widehat{P}^N_{(s_0,p)}$ are defined in (2.8). By setting $F_{W,(s_0,p)}(z) := \mathbb{P}(W_{(s_0,p)} \le z)$ and $F_{N,(s_0,p)}(z) := \mathbb{P}(N_{(s_0,p)} \le z)$, we have that the oracle P-values of $W_{(s_0,p)}$ and $N_{(s_0,p)}$ are

$$P^W_{(s_0,p)} := 1 - F_{W,(s_0,p)}(W_{(s_0,p)}) \quad \text{and} \quad P^N_{(s_0,p)} := 1 - F_{N,(s_0,p)}(N_{(s_0,p)}).$$

By the definitions of $\widehat{P}^W_{(s_0,p)}$ and $\widehat{P}^N_{(s_0,p)}$ in (2.8), $\widehat{P}^W_{(s_0,p)}$ and $\widehat{P}^N_{(s_0,p)}$ estimate $P^W_{(s_0,p)}$ and $P^N_{(s_0,p)}$. Therefore, by (3.16) $W_{\rm ad}$ and $N_{\rm ad}$ estimate

$$\widetilde{W}_{\rm ad} = \min_{p \in \mathcal{P}} P^W_{(s_0,p)} \qquad \text{and} \qquad \widetilde{N}_{\rm ad} = \min_{p \in \mathcal{P}} P^N_{(s_0,p)}. \tag{3.17}$$

By setting $\widetilde{F}_{W,\rm ad}(z) := \mathbb{P}(\widetilde{W}_{\rm ad} \le z)$ and $\widetilde{F}_{N,\rm ad}(z) := \mathbb{P}(\widetilde{N}_{\rm ad} \le z)$, considering that the small values of $W_{\rm ad}$ and $N_{\rm ad}$ yield the rejection of the null hypotheses, we have that the oracle P-values of $W_{\rm ad}$ and $N_{\rm ad}$ are $\widetilde{F}_{W\rm ad}(\widetilde{W}_{\rm ad})$ and $\widetilde{F}_{N,\rm ad}(\widetilde{N}_{\rm ad})$.

After introducing these notations, we aim to justify the bootstrap procedure in Section 2.2 by showing that $\widehat{P}^W_{\rm ad}$ and $\widehat{P}^N_{\rm ad}$ (defined in (2.11) and (2.13)) are consistent estimators of the oracle P-values $\widetilde{F}_{W\rm ad}(\widetilde{W}_{\rm ad})$ and $\widetilde{F}_{N,\rm ad}(\widetilde{N}_{\rm ad})$. For this, we introduce Assumption $\mathbf{(A)}''$ to specify the scaling between s_0, q and n for the data-adaptive combined test.

To state Assumption $\mathbf{(A)}''$, we need some additional notations. For the two-sample problem, we introduce $\boldsymbol{G}^N \sim N(\mathbf{0}, \mathbf{R}_{12}) \in \mathbb{R}^q$ in (3.6) to approximate $\boldsymbol{N}$. We set $f_{\boldsymbol{G}^N,(s_0,p)}(x)$ and $c_{\boldsymbol{G}^N,(s_0,p)}(\alpha)$ as the probability density function and the α-quantile of $\|\boldsymbol{G}^N\|_{(s_0,p)}$. We then define $h_{q,N}(\epsilon)$ as

$$h_{q,N}(\epsilon) = \max_{p\in\mathcal{P}} \max_{x\in I^N_{(s_0,p)}(\epsilon)} f^{-1}_{\boldsymbol{G}^N,(s_0,p)}(x),$$

where $I^N_{(s_0,p)}(\epsilon) = [c_{\boldsymbol{G}^N,(s_0,p)}(\epsilon), c_{\boldsymbol{G}^N,(s_0,p)}(1-\epsilon)]$. For the one-sample problem, we define $h_{q,W}(\epsilon)$ similarly for $\boldsymbol{G}^W \sim N(\mathbf{0}, \mathbf{R}_1) \in \mathbb{R}^q$, where

$$\mathbf{R}_1 = (r_{1,st}) \in \mathbb{R}^{q\times q} \quad \text{with} \quad r_{1,st} = \text{Corr}(h_s(\boldsymbol{X}), h_t(\boldsymbol{X})).$$

By definition, $\mathbf{R}_1$ and $\mathbf{R}_{12}$ are the asymptotic correlation matrices of $\boldsymbol{W}$ and $\boldsymbol{N}$, where $\boldsymbol{W}$ and $\boldsymbol{N}$ are defined in (2.4). With these additional notations, we then state Assumption **(A)**$''$ as follows.

- **(A)**$''$ Under (1.7), as $n_1 \to \infty$, we assume that $h^{0.6}_{q,W}(\epsilon)s_0^2 \log q = o(n_1^{1/10})$ holds for any $0 < \epsilon < 1$. Under (1.8), as $n \to \infty$, we assume that $h^{0.6}_{q,N}(\epsilon)s_0^2 \log q = o(n^{1/10})$ holds for any $0 < \epsilon < 1$.

Compared to Assumption **(A)**, the required scaling in Assumption **(A)**$''$ is more stringent. This is because when analyzing the combined test, we need not only the convergence of distribution functions of the test statistics but also their uniform convergence of the quantile functions on $[\epsilon, 1-\epsilon]$.

REMARK 3.4. Let $1 \le s_0, \#(\mathcal{P}) < \infty$. If there are $0 < C_0 < \infty$ and $0 < \eta < 1$ such that $C_0^{-1} < \lambda_{\min}(\mathbf{R}_{12}) \le \lambda_{\max}(\mathbf{R}_{12}) < C_0$ and $\max_{i\neq j}|r_{ij}| < \eta$, we have $h_{q,N}(\epsilon) = O(1)$ for any $\epsilon \in (0,1)$, as $q \to \infty$. Similarly, if $C_0^{-1} < \lambda_{\min}(\mathbf{R}_1) \le \lambda_{\max}(\mathbf{R}_1) < C_0$ and $\max_{i\neq j}|r_{ij}| < \eta$, we also have $h_{q,W}(\epsilon) = O(1)$ for any $\epsilon \in (0,1)$, as $q \to \infty$. The detailed proof is in Appendix B.6 of supplementary materials.

The detailed proof of Remark 3.4 is in Appendix B.6 of supplementary materials, in which we obtain a joint asymptotic distribution for the order statistics of nonindependent Gaussian random variables. This result is nontrivial and of independent technical interest. After introducing additional assumptions, we then justify the data-adaptive combined test by the following theorem.

THEOREM 3.5. Suppose Assumptions **(A)**$''$, **(E)**, **(M1)** and **(M2)** hold. For the one-sample problem, under $\mathbf{H}_0$ of (1.7) we have

$$\mathbb{P}_{\mathbf{H}_0}\big(T^W_{\text{ad}} = 1\big) \to \alpha \text{ and } \widetilde{F}_{W,\text{ad}}(\widetilde{W}_{\text{ad}}) - \widehat{P}^W_{\text{ad}} \to 0 \text{ as } n_1, B \to \infty. \tag{3.18}$$

Similarly, for the two-sample problem, under $\mathbf{H}_0$ of (1.8) we have

$$\mathbb{P}_{\mathbf{H}_0}\big(T^N_{\text{ad}} = 1\big) \to \alpha \text{ and } \widetilde{F}_{N,\text{ad}}(\widetilde{N}_{\text{ad}}) - \widehat{P}^N_{\text{ad}} \to 0 \text{ as } n, B \to \infty. \tag{3.19}$$

The detailed proof of Theorem 3.5 is in Appendix B.5 of supplementary materials.

REMARK 3.6. To prove Theorem 3.5, we first show that for any fixed $0 < \epsilon < 1$, not only the distribution function of N_{ad} but also its quantile function on $[\epsilon, 1-\epsilon]$ converge to those of $\widetilde{N}_{\text{ad}}$. By choosing ϵ sufficiently small, we then prove that the probability of $\widetilde{N}_{\text{ad}} \in (0, \epsilon)$ is negligible to finish the proof. If $\#(\mathcal{P}) \to \infty$, we cannot guarantee $\widetilde{N}_{\text{ad}} \in (0, \epsilon)$ is negligible any more. Moreover, it is also very hard to prove the convergence of quantile functions on $(\epsilon, 1-\epsilon)$ for N_{ad} with $\epsilon \to 0$. Hence, when constructing the combined test, we require $0 < \#(\mathcal{P}) < \infty$. By simulation, we recommend using $\mathcal{P} = \{1, 2, 3, 4, 5, \infty\}$. The simulation also shows that there is no significant power advantage to add more elements to $\mathcal{P}$ (see Appendix F.3). Therefore, the assumption of finite $\#(\mathcal{P})$ is enough for the practical usage.

We now turn to the analysis of the power of the combined test. For this, we have the following result.

THEOREM 3.7. Suppose Assumptions **(A)′**, **(E)**, **(M1)**, and **(M2)**) hold. For the one-sample problem in (1.7), we assume $\log q = o(n_1^{1/2})$, $\varepsilon_{n_1} = o(1)$, and $\varepsilon_{n_1}\sqrt{\log q} \to \infty$ as $n_1, q \to \infty$. If $\mathbf{H}_1$ of (1.7) holds with

$$\|\boldsymbol{D}_1\|_{(s_0,p)} \geq s_0(1+\varepsilon_{n_1})\big(\sqrt{2\log q} + \sqrt{2\log(\#\{\mathcal{P}\}/\alpha)}\big), \tag{3.20}$$

we have $\mathbb{P}_{\mathbf{H}_1}\big(T_{\text{ad}}^W = 1\big) = 1$ as $n_1, q, B \to \infty$. Similarly, for the two-sample problem in (1.8) we assume $\log q = o(n^{1/2})$, $\varepsilon_n = o(1)$, and $\varepsilon_n\sqrt{\log q} \to \infty$ as $n, q \to \infty$. If $\mathbf{H}_1$ of (1.8) holds with

$$\|\boldsymbol{D}_2\|_{(s_0,p)} \geq s_0(1+\varepsilon_n)\big(\sqrt{2\log q} + \sqrt{2\log(\#\{\mathcal{P}\}/\alpha)}\big), \tag{3.21}$$

we have $\mathbb{P}_{\mathbf{H}_1}\big(T_{\text{ad}}^N = 1\big) \to 1$ as $n, q, B \to \infty$.

The detailed proof of Theorem 3.7 is presented in Appendix B.7 of supplementary materials.

REMARK 3.8. On one hand, by Theorems 3.3 and 3.7, we require $\|\boldsymbol{u}_1 - \boldsymbol{u}_0\|_{(s_0,p)} \succeq s_0\sqrt{\log(q)/n_1}$ or $\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{(s_0,p)} \succeq s_0\sqrt{\log(q)/n}$ for our proposed methods to reject the null hypothesis with overwhelming probability. On the other hand, by Theorem 3 in [14, 15], Theorem 4.3 in [35], and Theorem 3.5 in [74], for both vector-based and matrix-based high dimensional tests, any α-level test is unable to reject the null hypothesis correctly uniformly over $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|_\infty \geq c_0\sqrt{\log(d)/n}$ or $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|_\infty \geq c_0\sqrt{\log(d)/n}$ with c_0 sufficiently small. Therefore, we have that our proposed methods with finite s_0 are rate-optimal for these sparse alternatives.

4. Simulation results. The goal of this section is to investigate the numerical performance of the proposed tests. For this, we compare our methods with several existing methods from the literature. In this section, we only consider the high dimensional mean test under different settings. We put additional simulation results for testing high-dimensional covariance/correlation coefficients in Appendix F to illustrate the proposed methods' generality. Apart from simulated datasets, Appendix F also includes the experimental results on real world fMRI datasets.

In the context of high dimensional mean test, we compare the proposed tests with four existing methods: Hotelling's T^2 test, the L_2-type tests given in [5] and [67], and the L_∞-type test give in [15]. We refer these four tests as T^2, BY, SD, and CLX. For simplicity, we only consider the two-sample problem. We generate synthetic data from a wide range of covariance structure including both sparse and non-sparse settings. We also consider a wide range of alternative scenarios including both sparse and dense settings to investigate the power of the proposed methods.

Under the null hypothesis, we sample $n_1 + n_2$ data points from the following models.

- **Model 1.** (Gaussian distribution with block diagonal $\mathbf{\Sigma}$) We set $\mathbf{\Sigma}^\star = (\sigma^\star_{ij}) \in \mathbb{R}^{d\times d}$ with $\sigma^\star_{ii} \stackrel{\text{i.i.d.}}{\sim} \mathrm{U}(1,2)$, $\sigma^\star_{ij} = 0.5$ for $5(k-1)+1 \le i \ne j \le 5k$, where $k = 1, \ldots, \lfloor d/5 \rfloor$, and $\sigma^\star_{ij} = 0$ otherwise. In this model, under the null hypothesis we generate $n_1 + n_2$ random vectors from $N(\mathbf{0}, \mathbf{\Sigma}^\star)$.
- **Model 2.** (Gaussian distribution with banded $\mathbf{\Sigma}$) We set $\mathbf{\Sigma}' = (\sigma'_{ij}) \in \mathbb{R}^{d\times d}$ with $\sigma'_{ij} = 0.4^{|i-j|}$ for $1 \le i, j \le d$. In this model, under the null hypothesis we generate $n_1 + n_2$ random vectors from $N(\mathbf{0}, \mathbf{\Sigma}')$.
- **Model 3.** (Gaussian distribution with non-sparse $\mathbf{\Sigma}$) We set $\mathbf{F} = (f_{ij}) \in \mathbb{R}^{d\times d}$ with $f_{ii} = 1$, $f_{ii+1} = f_{i+1i} = 0.5$, and $f_{ij} = 0$ otherwise. We also set that $\mathbf{U} \sim \mathrm{U}(\Lambda_{d,k})$ follows the uniform distribution on the Stiefel manifold $\Lambda_{d,k}$ (i.e., $\Lambda_{d,k} = \{\mathbf{H} \in \mathbb{R}^{d\times k} : \mathbf{H}^\top\mathbf{H} = \mathbf{I}_k\}$). After introducing $\mathbf{F}$ and $\mathbf{U}$, we then set the correlation matrix as $\mathbf{R} = (\mathbf{D}^f)^{-1/2}(\mathbf{F} + \mathbf{U}\mathbf{U}^\top)(\mathbf{D}^f)^{-1/2}$ with $\mathbf{D}^f = \mathrm{Diag}(\mathbf{F} + \mathbf{U}\mathbf{U}^\top)$. By setting $\mathbf{D} = (d_{ij}) \in \mathbb{R}^{d\times d}$ as a diagonal matrix with $d_{ii} \sim U(1,2)$, we generate $n_1 + n_2$ random vectors from $N(0, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$.
- **Model 4.** (Multivariate t distribution) We generate $n_1 + n_2$ random vectors from the multivariate t distribution $t(\nu, \boldsymbol{\mu}, \mathbf{\Sigma})$ according to $\boldsymbol{\mu} + \boldsymbol{Z}/\sqrt{W/\nu}$, where we have $W \sim \chi^2(\nu)$ and $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbf{\Sigma})$ with W and $\boldsymbol{Z}$ independent of each other. In the simulation, we set $\boldsymbol{\mu} = \mathbf{0}$, $\nu = 5$, and $\mathbf{\Sigma} = \mathbf{\Sigma}^\star$.

We use the above models to show that the proposed methods are valid given a fixed size α under various covariance structures and distributions. To present the empirical power of the proposed methods, we introduce a random vector $\boldsymbol{V} \in \mathbb{R}^d$ with exactly s nonzero entries, which are selected randomly from d coordinates. Each nonzero entry follows an independent uniform distribution $\mathrm{U}(u_1, u_2)$. Under the alternative hypothesis, we set $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = \boldsymbol{V}$. By choosing different s, u_1, and u_2, we compare the power of the proposed methods with that the existing methods under both the sparse and non-sparse settings.

TABLE 1
Empirical sizes for **Model 1** *with* $\alpha = 0.05$, $B = 300$, *and* $n_1 = n_2 = 100$ *based on 2000 replications.*

		Empirical size (%)										
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	T^2	BY	SD	CLX
75	5	5.50	5.85	6.15	6.35	6.30	6.60	6.50	5.05	5.85	4.65	5.25
	30	4.20	4.45	4.90	5.30	5.70	6.90	5.90	5.05	5.85	4.65	5.25
	75	3.70	3.95	4.75	5.10	5.65	6.75	5.50	5.05	5.85	4.65	5.25
200	10	4.75	4.50	4.85	5.20	5.25	6.55	5.75	-	4.85	3.85	5.35
	50	2.80	2.90	3.55	3.80	4.35	6.45	4.75	-	4.85	3.85	5.35
	100	1.90	2.25	2.50	3.60	3.85	6.45	4.80	-	4.85	3.85	5.35
	150	2.35	2.45	2.75	3.70	4.15	6.85	4.90	-	4.85	3.85	5.35
	200	2.30	2.35	2.95	3.65	4.35	7.10	5.15	-	4.85	3.85	5.35
400	10	4.20	4.30	4.70	5.30	5.40	7.60	5.90	-	5.35	4.55	6.65
	50	2.45	2.50	2.80	3.45	4.25	8.25	5.00	-	5.35	4.55	6.65
	100	2.05	2.30	2.25	2.65	3.95	7.90	4.75	-	5.35	4.55	6.65
	200	1.45	1.60	1.90	2.70	3.60	7.75	4.55	-	5.35	4.55	6.65
	400	1.40	1.40	1.75	2.70	3.80	7.85	4.70	-	5.35	4.55	6.65
800	10	4.75	4.95	5.20	5.50	5.95	9.10	6.30	-	5.65	4.65	7.45
	100	0.75	1.20	1.40	1.80	2.65	8.85	4.45	-	5.65	4.65	7.45
	200	0.40	0.50	0.75	1.40	2.00	8.85	4.45	-	5.65	4.65	7.45
	400	0.55	0.45	0.70	1.20	2.10	8.15	3.95	-	5.65	4.65	7.45
	600	0.40	0.35	0.80	1.20	2.00	8.70	4.00	-	5.65	4.65	7.45
	800	0.40	0.55	0.75	1.35	1.85	8.65	3.65	-	5.65	4.65	7.45

In Table 1, we present the empirical sizes of introduced methods for **Model 1**. We set $n_1 = n_2 = n = 100$ and $q = d = 75, 200, 400, 800$. The nominal significance level is 0.05. We compare our methods with four other tests: T^2, BY, SD, and CLX. Moreover, T^2, BY, and SD are L_2-type and CLX is L_∞-type. The T^2 test requires $d < n$, so that we don't perform T^2 test as $d > n$. In the current setting, the four existing methods can control the size correctly, except that CLX test suffers a size distortion as d is significantly larger ($d = 800$) than n. For the (s_0, p)-norm based tests, when s_0 is significantly smaller ($s_0 = 5, 10$) than d, they can control the size correctly, except that the (s_0, ∞)-norm based test suffers a size distortion as d is significantly large ($d = 800$). As s_0 increases, the empirical size of

Table 2
Empirical power of **Model 1** *with* $\alpha = 0.05$, $B = 300$, *and* $n_1 = n_2 = 100$ *based on 2000 replications.*

Empirical power (%) with $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = \boldsymbol{V}$ with $s = 5$, $u_1 = 0$, and $u_2 = 4\sqrt{\log(d)/n}$												
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	T^2	BY	SD	CLX
75	5	82.10	84.35	85.70	86.50	86.80	85.10	86.90	73.7	67.85	66.45	83.5
	30	49.10	69.20	78.75	83.80	85.40	85.50	84.50	73.7	67.85	66.45	83.5
	75	32.70	64.25	78.20	83.25	85.15	85.00	83.70	73.7	67.85	66.45	83.5
200	10	75.85	81.05	83.85	84.95	86.10	86.40	85.65	-	55.65	53.90	85.25
	50	36.20	59.65	75.35	81.50	84.00	86.05	84.40	-	55.65	53.90	85.25
	100	23.60	48.90	72.65	80.75	84.20	86.45	84.35	-	55.65	53.90	85.25
	150	18.70	45.40	72.10	81.15	84.45	86.45	84.35	-	55.65	53.90	85.25
	200	17.55	45.20	72.40	81.00	84.20	86.15	84.25	-	55.65	53.90	85.25
400	10	77.90	82.15	85.20	87.25	88.05	87.80	87.90	-	44.25	42.75	87.05
	50	34.25	56.40	71.85	79.65	84.05	87.90	85.55	-	44.25	42.75	87.05
	100	18.45	40.45	65.15	77.30	83.50	87.60	85.55	-	44.25	42.75	87.05
	200	9.85	29.35	60.25	76.25	83.35	88.10	85.45	-	44.25	42.75	87.05
	400	6.95	25.60	60.00	75.90	83.60	87.50	85.05	-	44.25	42.75	87.05
Empirical power (%) with $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = \boldsymbol{V}$ with $s = 100$, $u_1 = 0$, and $u_2 = 3\sqrt{1/n}$												
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	T^2	BY	SD	CLX
200	10	87.55	86.05	85.00	83.50	81.75	51.25	82.30	-	96.20	96.05	47.45
	50	93.25	93.80	93.40	92.05	89.40	51.05	90.20	-	96.20	96.05	47.45
	100	92.75	93.85	93.80	92.55	89.80	51.55	91.35	-	96.20	96.05	47.45
	150	90.95	93.55	94.35	92.35	89.90	52.75	90.65	-	96.20	96.05	47.45
	200	90.05	93.75	93.80	92.45	89.55	51.70	91.10	-	96.20	96.05	47.45
400	10	70.40	69.95	68.75	67.90	66.00	42.70	64.00	-	85.10	84.25	38.70
	50	72.60	73.95	74.95	74.85	73.35	42.30	70.25	-	85.10	84.25	38.70
	100	69.85	72.80	73.95	74.45	73.30	42.35	69.45	-	85.10	84.25	38.70
	200	61.55	68.25	73.05	74.45	73.90	42.45	68.30	-	85.10	84.25	38.70
	400	54.40	67.15	73.00	74.50	73.40	43.15	67.15	-	85.10	84.25	38.70

(s_0, p)-norm based tests decreases dramatically especially for small p, making the (s_0, p)-norm based tests with small p overly conservative. Although the (s_0, p)-norm based tests perform differently with different s_0 and p, the data-adaptive combined test T_{ad}^N can control the size correctly under various settings of d and n.

In Table 2, we compare these methods under different alternative scenarios. In the sparse alternative setting, we set $\boldsymbol{\mu}_2 = \boldsymbol{V}$ with $s = 5$ nonzero entries. Each entry follows independent uniform distribution $\mathrm{U}(0, 4\sqrt{\log(d)/n})$. In this setting, the L_∞-type test achieves a higher empirical power than the L_2-type tests. In the dense alternative setting, we set $\boldsymbol{\mu}_2 = \boldsymbol{V}$ with $s = 100$ nonzero entries of the magnitude $\mathrm{U}(0, 3n^{-1/2})$. In this setting, the L_2-type tests are more powerful. This similar pattern also appears in the (s_0, p)-norm based tests. As p increases, the (s_0, p)-norm based test is more sensitive to the sparse alternative. The influence of s_0 is more complicated. However, by choosing s_0 close to s, the tests always enjoy good performance. For the data-adaptive combined test T_{ad}^N, we choose a balanced $\mathcal{P}$ including both small and large values of p. Hence, in various settings of the alternative scenarios, d, and n, it always has a high power. Although T_{ad}^N with balanced $\mathcal{P}$ may not be the most powerful option for some alternatives, T_{ad}^N is adaptive to the alternative setting and powerful enough in various kinds of alternative scenarios. Theoretically, there is no uniformly most powerful test in all the alternative scenarios [26]. If the alternative pattern is unknown, the data-adaptive test with balanced $\mathcal{P}$ (including small and large p) is a good choice. If the alternative pattern is known, by choosing $\mathcal{P}$ accordingly we can still construct a powerful test. For the choice of s_0, similarly to the (s_0, p)-norm based tests, T_{ad}^N with s_0 close to s is always powerful.

We put the numerical results of **Models 2-4** in Appendix F of supplementary materials. Their experimental results are similar to **Model 1** and indicate that the proposed methods work well in various settings.

5. Summary and discussion. This paper considers the problem of testing high dimensional U-statistic based vectors. We construct a family of tests based on the (s_0, p)-norm. By the introduction of s_0, when q is large, we can increase the power compared to the tradition L_p-norm based test (especially for small p). Moreover, by choosing p properly, we can further enhance the power under different alternatives. We also introduce a data-adaptive combined test, which is simultaneously powerful under a wide variety of alternatives. Moreover, We also develop a trick for avoiding the high computational cost of the double-loop bootstrap for the data-adaptive combined test with theoretical guarantee in high dimensions.

We then discuss the choice of s_0 and $\mathcal{P}$. Theoretically, for individual (s_0, p)-norm tests we generally require that $s_0^\zeta \log q = o(n^\delta)$ holds with $\zeta = 2$ and $0 < \delta < 1/7$ for all $p \in [1, \infty]$. We also point out that it is possible to reduce ζ for some specified p. For combined tests, we require $0 < \#(\mathcal{P}) < \infty$ to prevent the test statistic from going to 0. By simulation, we also see that the proposed tests with s_0 close to s (true unknown number of entries violating $\mathbf{H}_0$) enjoy high power, which makes s a good candidate for s_0. In practice, we recommend choosing s_0 and s as close as possible without violating theoretical conditions.

There are several possible future directions of this work. For instance, how to generalize the idea to the k-sample testing problems ($k > 2$) has been for a future investigation. This may require a nontrivial extension of the theoretical analysis. Moreover, our theory is based on the Gaussian approximation for the sum of high dimensional independent random vectors from [23]. [73] and [72] further study Gaussian approximations for high dimensional time series, which allow to generalize our methods for dependent data. As a significant amount of additional work is still needed, we shall report the results elsewhere in the future.

References.

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis (3rd).* Wiley New York.

[2] Arcones, M. A. and Gine, E. (1993). Limit theorems for U-processes. *Annals of Probability* **21** 1494–1542.

[3] Bai, Z., Jiang, D., Yao, J. and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Annals of Statistics* **37** 3822–3840.

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

[5] Bai, Z. and Yin, Y. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Annals of Probability* **21** 1275–1294.

[6] Bao, Z., Lin, L., Pan, G. and Zhou, W. (2015). Spectral statistics of large dimensional Spearman's rank correlation matrix and its application. *Annals of Statistics* **43** 2588–2623.

[7] Barbe, P. and Bertail, P. (2012). *The Weighted Bootstrap*, vol. 98. Springer Science and Business Media.

[8] Basu, S. and Pan, W. (2011). Comparison of statistical tests for disease association with rare variants. *Genetic Epidemiology* **35** 606–619.

[9] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 289–300.

[10] Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. *Statistics and Probability Letters* **74** 281–289.

[11] Bonnéry, D., Breidt, F. J. and Coquet, F. (2012). Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population. *Bernoulli* **18** 1361–1385.

[12] Cai, T. and Jiang, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *Journal of Multivariate Analysis* **107** 24–39.

[13] Cai, T. and Liu, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *Journal of the American Statistical Association* **106** 1566–1577.

[14] Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association* **108** 265–277.

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

[16] Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Annals of Statistics* **39** 1496–1525.

[17] Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388.

[18] Cakici, N. (2015). The five-factor Fama-French model: International evidence. *Available at SSRN 2601662* .

[19] Castagna, J. P., Sun, S. and Siegfried, R. W. (2003). Instantaneous spectral

analysis: Detection of low-frequency shadows associated with hydrocarbons. *The Leading Edge* **22** 120–127.

[20] Chang, J., Zhou, W. and Zhou, W. (2014). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *arXiv preprint arXiv:1406.1939* .

[21] Chang, J., Zhou, W. and Zhou, W. (2015). Bootstrap tests on high dimensional covariance matrices with applications to understanding gene clustering. *arXiv preprint arXiv:1505.04493* .

[22] Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Annals of Statistics* **38** 808–835.

[23] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* **41** 2786–2819.

[24] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Central limit theorems and bootstrap in high dimensions. *arXiv preprint arXiv:1412.3661* .

[25] Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing* **36** 287–314.

[26] Cox, D. R. and Hinkley, D. V. (1979). *Theoretical Statistics*. CRC Press.

[27] Dembo, A. and Shao, Q. (2006). Large and moderate deviations for Hotelling's T^2-statistic. *Electronic Communications in Probability* **11** 149–159.

[28] Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33** 3–56.

[29] Fama, E. F. and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics* **105** 457–472.

[30] Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116** 1–22.

[31] Fama, E. F. and French, K. R. (2016). International tests of a five-factor asset pricing model. *Journal of Financial Economics* **123** 441–463.

[32] Fan, J. and Fan, Y. (2008). High dimensional classification using features annealed independence rules. *Annals of statistics* **36** 2605.

[33] Fan, J., Feng, Y. and Tong, X. (2012). A road to classification in high dimensional space: The regularized optimal affine discriminant. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74** 745–771.

[34] Gombay, E. and Horvath, L. (2002). Rates of convergence for U-statistic processes and their bootstrapped versions. *Journal of Statistical Planning and Inference* **102** 247–272.

[35] Han, F. and Liu, H. (2014). Distribution-free tests of independence with applications to testing more structures. *arXiv preprint arXiv:1410.4179* .

[36] Han, F., Zhao, T. and Liu, H. (2013). CODA: High dimensional copula discriminant analysis. *Journal of Machine Learning Research* **14** 629–671.

[37] Ho, J. W., Stefani, M., dos Remedios, C. G. and Charleston, M. A. (2008). Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics* **24** 390–398.

[38] Hu, R., Qiu, X. and Glazko, G. (2010). A new gene selection procedure based on the covariance distance. *Bioinformatics* **26** 348–354.

[39] Hu, R., Qiu, X., Glazko, G., Klebanov, L. and Yakovlev, A. (2009). Detecting intergene correlation changes in microarray analysis: A new approach to gene selection. *BMC Bioinformatics* **10** 20.

[40] Huskova, M. and Jansen, P. (1993). Generalized bootstrat for studentized U-statistics: A rank statistic approach. *Statistics and Probability Letters* **16** 225–233.

[41] Huskova, M. and Janssen, P. (1993). Consistency of the generalized bootstrap for degenerate U-statistics. *Annals of Statistics* **21** 1811–1823.

[42] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Annals of Applied Probability* **14** 865–880.

[43] Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Annals of Statistics* **41** 2029–2074.

[44] Korolyuk, V. S. and Borovskich, Y. V. (2013). *Theory of U-Statistics*, vol. 273. Springer Science and Business Media.

[45] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Annals of Statistics* **30** 1081–1102.

[46] Lee, A. (1990). *U-Statistics: Theory and Practice*. Statistics: A Series of Textbooks and Monographs, Taylor and Francis.

[47] Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., Team, E. L. P., Christiani, D. C., Wurfel, M. M. and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *The American Journal of Human Genetics* **91** 224–237.

[48] Li, D. and Rosalsky, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices. *Annals of Applied Probability* **16** 423–447.

[49] Li, J. and Chen, S. (2012). Two sample tests for high-dimensional covariance matrices. *Annals of Statistics* **40** 908–940.

[50] Lin, D. and Tang, Z. (2011). A general framework for detecting disease associations with rare variants in sequencing studies. *The American Journal of Human Genetics* **89** 354–367.

[51] Liu, W., Lin, Z. and Shao, Q. (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Annals of Applied Probability* **18** 2337–2366.

[52] Liu, W. and Shao, Q. (2013). A cramér moderate deviation theorem for Hotellings T^2-statistic with applications to global tests. *Annals of Statistics* **41** 296–322.

[53] Lo, A. Y. (1987). A large sample study of the Bayesian bootstrap. *Annals of Statistics* **15** 360–375.

[54] Mai, Q. and Zou, H. (2013). Semiparametric sparse discriminant analysis in ultra-high dimensions. *Biometrika* **99** 29–42.

[55] Mai, Q., Zou, H. and Yuan, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika* **99** 29–42.

[56] Mason, D. M. and Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Annals of Statistics* **20** 1611–1624.

[57] Nagao, H. (1973). On some test criteria for covariance matrix. *Annals of Statistics* **1** 700–709.

[58] Pan, W., Kim, J., Zhang, Y., Shen, X. and Wei, P. (2014). A powerful and adaptive association test for rare variants. *Genetics* **197** 1081–1095.

[59] Parzen, M., Wei, L. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81** 341–350.

[60] Roy, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York.

[61] Rubin, D. B. et al. (1981). The Bayesian bootstrap. *Annals of Statistics* **9** 130–134.

[62] Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956.

[63] Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics and Data*

Analysis **51** 6535–6542.
[64] Shao, J., Wang, Y., Deng, X. and Wang, S. (2011). Sparse linear discriminant analysis by thresholding for high dimensional data. *Annals of statistics* **39** 1241–1265.
[65] Shao, Q. and Zhou, W. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Annals of Probability* **42** 623–648.
[66] Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis* **100** 518–532.
[67] Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis* **99** 386–402.
[68] Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis* **101** 1319–1329.
[69] Tibshirani, R., Hastie, T., Narasimhan, B. and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences* **99** 6567–6572.
[70] Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* **89** 82–93.
[71] Zang, Y., He, Y., Zhu, C., Cao, Q., Sui, M., Liang, M., Tian, L., Jiang, T. and Wang, Y. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development* **29** 83–91.
[72] Zhang, D. and Wu, W. (2015). Gaussian approximation for high dimensional time series. *arXiv preprint arXiv:1508.07036* .
[73] Zhang, X. and Cheng, G. (2014). Bootstrapping high dimensional time series. *arXiv preprint arXiv:1406.1037* .
[74] Zhou, C., Han, F., Zhang, X. and Liu, H. (2015). An extreme-value approach for testing the equality of large U-statistic based correlation matrices. *arXiv preprint arXiv:1502.03211* .
[75] Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society* **359** 5345–5363.
[76] Zou, Q., Zhu, C., Yang, Y., Zuo, X., Long, X., Cao, Q., Wang, Y. and Zang, Y. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *Journal of Neuroscience Methods* **172** 137–141.

SUPPLEMENT MATERIALS TO "A UNIFIED FRAMEWORK FOR TESTING HIGH DIMENSIONAL PARAMETERS: A DATA-ADAPTIVE APPROACH"

By Cheng Zhou[‡], Xinsheng Zhang[‡], Wenxin Zhou[§] and Han Liu[§]

Department of Statistics, Fudan University[‡] *and Department of Operation Research and Financial Engineering, Princeton University*[§]

ABSTRACT

The supplementary materials contain additional details of the paper "A Unified Framework for Testing High Dimensional Parameters: A Data-Adaptive Approach" authored by Cheng Zhou, Xinsheng Zhang, Wen-Xin Zhou, and Han Liu. After introducing some useful lemmas in Appendix A, We prove main results in Appendix B. In Appendices C and D, we prove lemmas required by the proofs in Appendix B. In Appendix E, we prove lemmas introduced in Appendix A. In Appendix F, we present additional numerical experimental results. Throughout supplementary materials, we use C, C_1, $C_2, \dots$ to denote constants which do not depend on n, d, and q. These constants can vary from place to place.

APPENDIX A: USEFUL LEMMAS

In Appendix A, we introduce some useful lemmas that will be used many times for proving main results. We put their proof in Appendix E. To present these lemmas, we need some additional notations. Let $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ be independent random vectors in $\mathbb{R}^d$ with $\boldsymbol{Z}_k = (Z_{k1}, \dots, Z_{kd})^\top$ and $\mathbb{E}[\boldsymbol{Z}_k] = \mathbf{0}$ for $k = 1, \dots, n$. Let $\boldsymbol{W}_1, \dots, \boldsymbol{W}_n$ be independent Gaussian random vectors in $\mathbb{R}^d$ such that $\boldsymbol{W}_k$ has the same mean vector and covariance matrix as $\boldsymbol{Z}_k$. By setting $\mathcal{V}_{s_0} := \{\mathbf{v} \in \mathbb{S}^{d-1} : \|\mathbf{v}\|_0 \le s_0\}$, we require the following conditions:

- **(M1)**$'$ $n^{-1}\sum_{k=1}^n \mathbb{E}\big[(\mathbf{v}'\boldsymbol{Z}_k)^2\big] \ge b > 0$ for any $\mathbf{v} \in \mathcal{V}_{s_0}$;
- **(M2)**$'$$n^{-1}\sum_{k=1}^n \mathbb{E}\big[|Z_{kj}|^{2+\ell}\big] \le K^\ell$ for $\ell = 1, 2$ and $j = 1, \dots, d$.
- **(E)**$'$ $\mathbb{E}\big[\exp(|Z_{kj}|/K)\big] \le 2$ for $j = 1, \dots, d$ and $k = 1, \dots, n$.

Lemma A.1. Assume $s_0^2 \log(dn) = O(n^\zeta)$ with $0 < \zeta < 1/7$. If $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ satisfy **(M1)**$'$, **(M2)**$'$, and **(E)**$'$. By setting $S_n^{\boldsymbol{Z}} = n_1^{-1/2}\sum_{k=1}^n \boldsymbol{Z}_k$ and $S_n^{\boldsymbol{W}} = n_1^{-1/2}\sum_{k=1}^n \boldsymbol{W}_k$, for $1 \le p \le \infty$ and sufficiently large n, there is a constant $\zeta_0 > 0$ such that

$$\sup_{z\in(0,\infty)}\Big|\mathbb{P}\Big(\|S_n^{\boldsymbol{Z}}\|_{(s_0,p)} \le z\Big) - \mathbb{P}\Big(\|S_n^{\boldsymbol{W}}\|_{(s_0,p)} \le z\Big)\Big| \le Cn^{-\zeta_0}, \tag{A.1}$$

where C depends on b and K.

Lemma A.2. (Corollary 1.2 in [2]) For any compact and symmetric convex set $\mathcal{C} \in \mathbb{R}^d$ with non-empty interior and $\gamma > e/4\sqrt{2}$, there exist a polytope $\mathcal{P} \in \mathbb{R}^d$ and a constant $\epsilon_\gamma > 0$ such that for any $0 < \epsilon < \epsilon_\gamma$, we have

$$\mathcal{P} \subset \mathcal{C} \subset (1+\epsilon)\mathcal{P} \qquad \text{and} \qquad V < \Big(\frac{\gamma}{\sqrt{\epsilon}} \ln \frac{1}{\epsilon}\Big)^d,$$

where V is the vertex number of $\mathcal{P}$.

We call a set A^m is m-generated if it is the intersection of m half-spaces. Therefore, A^m is a polytope with at least m facets. We then set $\mathcal{V}(A^m)$ as the set of m unit vectors that are outward normal to the facets of A^m. For $\epsilon > 0$, we then define

$$A^{m,\epsilon} := \cap_{v \in \mathcal{V}(A^m)} \{w \in \mathbb{R}^d : w^\top v \le \mathcal{S}_{A^m}(v) + \epsilon\},$$

where $\mathcal{S}_{A^m}(v) := \sup\{w^\top v : w \in A^m\}$.

Lemma A.3. Let $\mathcal{E}^{R,d} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \le R\}$ and $V^{z,d}_{(s_0,p)} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_{(s_0,p)} \le z\}$. For any $\gamma > e/4\sqrt{2}$, there is a m-generated convex set $A^m \in \mathbb{R}^d$ and a constant ϵ_γ such that for any $0 < \epsilon < \epsilon_\gamma$, we have

$$A^m \subset \mathcal{E}^{R,d} \cap V^{z,d}_{(s_0,p)} \subset A^{m,R\epsilon} \qquad \text{and} \qquad m \le d^{s_0}\Big(\frac{\gamma}{\sqrt{\epsilon}} \ln \frac{1}{\epsilon}\Big)^{s_0^2}.$$

Lemma A.4. (Nazarovs inequality in [11]) Let $\boldsymbol{W} = (W_1, \ldots, W_d)^\top \in \mathbb{R}^d$ be centered Gaussian random vector with $\inf_{k=1,\ldots,d} E[W_k^2] \ge b > 0$. For any $\mathbf{x} \in \mathbb{R}^d$ and $a > 0$, we then have

$$\mathbb{P}(\boldsymbol{W} \le \mathbf{x} + a) - \mathbb{P}(\boldsymbol{W} \le \mathbf{x}) \le Ca\sqrt{\log d},$$

where C only depends on b.

Lemma A.5. $\boldsymbol{W} = (W_1, \ldots, W_d)^\top$ is a random vector with the marginal distribution $N(0, \sigma^2)$. For any $t > 0$, we have

$$\mathbb{E}\Big[\max_{1 \le i \le d} |W_i|\Big] \le \frac{\log(2d)}{t} + \frac{t\sigma^2}{2}. \tag{A.2}$$

To estimate the covariance matrix of U-statistic based vector, we introduce $\sigma_{\gamma,st}$ and $\widehat{\sigma}_{\gamma,st}$ in (3.4) and (3.9). The following lemma then analyzes the estimation error of $\widehat{\sigma}_{\gamma,st}$. To analyze the correlation matrix, we also provide the approximation error of $\widehat{r}_{\gamma,st}$, where

$$r_{\gamma,st} = \sigma_{\gamma,st}/\sqrt{\sigma_{\gamma,ss}\sigma_{\gamma,tt}} \qquad \text{and} \qquad \widehat{r}_{\gamma,st} = \widehat{\sigma}_{\gamma,st}/\sqrt{\widehat{\sigma}_{\gamma,ss}\widehat{\sigma}_{\gamma,tt}}. \tag{A.3}$$

LEMMA A.6. Assumptions **(E)**, **(M1)**, and **(M2)** hold. For $\log(qn) = o(n^{1/3})$ and $m > 1$, when n is sufficiently large,

$$\max_{\substack{1\le s,t\le q\\ \gamma=1,2}} \max\Big(\big|\widehat{\sigma}_{\gamma,st} - \sigma_{\gamma,st}\big|, \big|\widehat{r}_{\gamma,st} - r_{\gamma,st}\big|\Big) \le C\frac{\log^{3/2}(qn)}{\sqrt{n}}, \tag{A.4}$$

holds with probability $1 - C_1 n^{-1}$. For $\log(qn) = o(n^{1/2})$ and $m = 1$, when n is sufficiently large,

$$\max_{\substack{1\le s,t\le q\\ \gamma=1,2}} \max\Big(\big|\widehat{\sigma}_{\gamma,st} - \sigma_{\gamma,st}\big|, \big|\widehat{r}_{\gamma,st} - r_{\gamma,st}\big|\Big) \le C\sqrt{\frac{\log(qn)}{n}} + C\frac{\log^2(qn)}{n}, \tag{A.5}$$

holds with probability $1 - C_1 n^{-1}$.

APPENDIX B: PROOF OF MAIN RESULTS

In Appendix B, we present the detailed proofs of main results including Proposition 1, Theorems 3.1, 3.3, 3.5 3.7, Remarks 3.4, and Corollary 3.1.

B.1. Proof of Proposition 1 .

PROOF. We need to prove that for any $1 \le p \le \infty$, $a \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have (i) $\|a\mathbf{x}\|_{(s_0,p)} = |a|\|\mathbf{x}\|_{(s_0,p)}$; (ii) $\|\mathbf{x}+\mathbf{y}\|_{(s_0,p)} \le \|\mathbf{x}\|_{(s_0,p)} + \|\mathbf{y}\|_{(s_0,p)}$; (iii) $\|\mathbf{x}\|_{(s_0,p)} = 0$ implies $\mathbf{x} = \mathbf{0}$. By Definition 1.1, for $\mathbf{x} = (x_1, \dots, x_d)^\top$ we have

$$\|\mathbf{x}\|_{(s_0,p)} = \Big(\sum\nolimits_{j=d-s_0+1}^{d} (x^{(j)})^p\Big)^{1/p},$$

We use $\boldsymbol{k}_1$ to denote the index of $x^{(d-s_0+1)}, x^{(d-s_0+2)}, \dots, x^{(d)}$. Therefore, we have $\|\mathbf{x}\|_{(s_0,p)} = \|\mathbf{x}_{\boldsymbol{k}_1}\|_p$, where $\mathbf{x}_{\boldsymbol{k}_1} \in \mathbb{R}^{s_0}$. We then separately prove (i), (ii), and (iii). For (i), we have

$$\|a\mathbf{x}\|_{(s_0,p)} = \|a\mathbf{x}_{\boldsymbol{k}_1}\|_p = |a|\|\mathbf{x}_{\boldsymbol{k}_1}\|_p = |a|\|\mathbf{x}\|_{(s_0,p)}.$$

For (iii), from $\|\mathbf{x}\|_{(s_0,p)} = 0$, we have $x^{(d)} = 0$, which implies $\mathbf{x} = \mathbf{0}$. Therefore, to prove Proposition 1, we only need to prove

$$\|\mathbf{x}+\mathbf{y}\|_{(s_0,p)} \le \|\mathbf{x}\|_{(s_0,p)} + \|\mathbf{y}\|_{(s_0,p)}.$$

Similarly to the definition of $\boldsymbol{k}_1$, we define $\boldsymbol{k}_2$, $\boldsymbol{k}_{12}$ for $\mathbf{y}$ and $\mathbf{x}+\mathbf{y}$. We then have

$$\|\mathbf{y}\|_{(s_0,p)} = \|\mathbf{y}_{\boldsymbol{k}_2}\|_p \quad \text{and} \quad \|\mathbf{x}+\mathbf{y}\|_{(s_0,p)} = \|(\mathbf{x}+\mathbf{y})_{\boldsymbol{k}_{12}}\|_p. \tag{B.1}$$

For $1 \leq p \leq \infty$, $\|\cdot\|_p$ is a norm. Hence, we have

$$\|(\mathbf{x}+\mathbf{y})_{\boldsymbol{k}_{12}}\|_p = \|\mathbf{x}_{\boldsymbol{k}_{12}} + \mathbf{y}_{\boldsymbol{k}_{12}}\|_p \leq \|\mathbf{x}_{\boldsymbol{k}_{12}}\|_p + \|\mathbf{y}_{\boldsymbol{k}_{12}}\|_p. \tag{B.2}$$

By the definition of $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$, we have

$$\|\mathbf{x}_{\boldsymbol{k}_{12}}\|_p \leq \|\mathbf{x}_{\boldsymbol{k}_1}\|_p = \|\mathbf{x}\|_{(s_0,p)} \quad \text{and} \quad \|\mathbf{y}_{\boldsymbol{k}_{12}}\|_p \leq \|\mathbf{y}_{\boldsymbol{k}_2}\|_p = \|\mathbf{y}\|_{(s_0,p)}. \tag{B.3}$$

Combining (B.1), (B.2), and (B.3), we have (iii), which finishes the proof. □

B.2. Proof of Theorem 3.1.

PROOF. In Theorem 3.1, we aim to prove (3.2) and (3.3). For simplicity, we only present the detailed proof of (3.3). According to the proof sketch in Section 3.2, the proof proceeds in three steps. In the first step, we obtain the approximate distribution of $\boldsymbol{N}$. In the second step, given $\mathcal{X}$ and $\mathcal{Y}$ we obtain the bootstrap sample $\boldsymbol{N}^b$'s distribution. In the last step, we analyze the approximation error between $\boldsymbol{N}$ and $\boldsymbol{N}^b|\mathcal{X},\mathcal{Y}$ to yield (3.3).

Step (i). In this step, we aim to obtain the approximate distribution of $\boldsymbol{N}$. As $\widehat{u}_{\gamma,s}$ is a U-statistic, by the Hoeffding decomposition we approximate $\boldsymbol{N}$ by a sum of independent random vectors. Hence, we can further approximate the sum by its Gaussian counterpart. In detail, under the null hypothesis we have $u_{1,s} = u_{2,s}$. Therefore, we rewrite N_s as

$$N_s = (\widetilde{u}_{1,s} - \widetilde{u}_{2,s})/\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}, \tag{B.4}$$

where $\widetilde{u}_{\gamma,s} := \widehat{u}_{\gamma,s} - u_{\gamma,s}$ is the centralized version of $\widehat{u}_{\gamma,s}$. For introducing Hoeffding decomposition, we define

$$h_s(\boldsymbol{X}_k) = \mathbb{E}[\Psi_s(\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m})|X_k],$$

where Ψ_s are defined in (3.1). Hence, by the Hoeffding decomposition, we decompose $\widetilde{u}_{\gamma,s}$ as

$$\begin{aligned} \widetilde{u}_{1,s} &= \frac{m}{n_1}\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k) + \binom{n_1}{m}^{-1}\Delta_{n_1,s}, \\ \widetilde{u}_{2,s} &= \frac{m}{n_2}\sum_{k=1}^{n_2} h_s(\boldsymbol{Y}_k) + \binom{n_2}{m}^{-1}\Delta_{n_2,s}, \end{aligned} \tag{B.5}$$

where we define $\Delta_{n_1,s}$ and $\Delta_{n_1,s}$ as

$$\Delta_{n_1,s} = \sum_{1\leq k_1<k_2<\ldots<k_m\leq n_1} \Big(\Psi_s(\boldsymbol{X}_{k_1}, \ldots, \boldsymbol{X}_{k_m}) - \sum_{\ell=1}^{m} h_s(\boldsymbol{X}_{k_\ell})\Big),$$

$$\Delta_{n_2,s} = \sum_{1\leq k_1<k_2<\ldots<k_m\leq n_2} \Big(\Psi_s(\boldsymbol{Y}_{k_1}, \ldots, \boldsymbol{Y}_{k_m}) - \sum_{\ell=1}^{m} h_s(\boldsymbol{Y}_{k_\ell})\Big).$$

We then use $m\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k)/n_1$ and $m\sum_{k=1}^{n_2} h_s(\mathbf{Y}_k)/n_2$ to approximate $\widetilde{u}_{1,s}$ and $\widetilde{u}_{2,s}$. By setting $\mathbf{\Sigma}_1 := (\sigma_{1,st}), \mathbf{\Sigma}_2 := (\sigma_{2,st}) \in \mathbb{R}^{q\times q}$ with

$$\sigma_{1,st} = \mathbb{E}\big(h_s(\boldsymbol{X})h_t(\boldsymbol{X})\big) \qquad \text{and} \qquad \sigma_{2,st} = \mathbb{E}\big(h_s(\boldsymbol{Y})h_t(\boldsymbol{Y})\big), \tag{B.6}$$

considering $\widehat{v}_{\gamma,s} = m^2\widehat{\sigma}_{\gamma,ss}$, as $n \to \infty$ we have $\widehat{v}_{\gamma,s} \to m^2\sigma_{\gamma,ss}$, which motivates us to define

$$H_s^N = \Big(\frac{1}{n_1}\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k) - \frac{1}{n_2}\sum_{k=1}^{n_2} h_s(\mathbf{Y}_k)\Big)\Big/\sqrt{\sigma_{1,ss}/n_1 + \sigma_{2,ss}/n_2}. \tag{B.7}$$

Moreover, by setting $\boldsymbol{H}^N = (H_1^N, \dots, H_q^N)^\top$, we have that $\boldsymbol{H}^N$ approximates $\boldsymbol{N}$, and the approximation error is characterized by the following lemma.

Lemma B.1. Assumptions **(A)**, **(E)**, **(M1)**, and **(M2)** hold. Under $\mathbf{H}_0$ of (1.8) there is a constant $C > 0$ such that as $n \to \infty$, we have

$$\mathbb{P}\Big(\|\boldsymbol{N} - \boldsymbol{H}^N\|_{(s_0,p)} > \varepsilon\Big) = o(1), \tag{B.8}$$

where $\varepsilon = Cs_0 \log^2(qn)n^{-1/2}$.

The proofs of Lemma B.1 is in Appendix C.1 of supplementary materials. By the definition of H_s^N in (B.7), $\boldsymbol{H}^N$ is a sum of random vectors with zero mean and covariance matrix $\mathbf{R}_{12}$, where we set

$$\mathbf{R}_{12} := \mathbf{D}_{12}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{D}_{12}^{-1/2} \tag{B.9}$$

with $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_1/n_1 + \mathbf{\Sigma}_2/n_2$ and $\mathbf{D}_{12} = \mathrm{Diag}(\mathbf{\Sigma}_{12})$. Therefore, by the central limit theorem, we can use the Gaussian random vector $\boldsymbol{G}^N \sim N(\mathbf{0}, \boldsymbol{R}_{12})$ to approximate $\boldsymbol{H}^N$. To characterize the approximation error, considering $A_z := \{\mathbf{v}, \|\mathbf{v}\|_{(s_0,p)} \le z\} \in \mathcal{A}_{s_0}$, by Lemma A.1, there is $\zeta_0 > 0$ such that

$$\sup_z \Big|\mathbb{P}(\|\boldsymbol{H}^N\|_{(s_0,p)} \le z) - \mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} \le z)\Big| \le Cn^{-\zeta_0} \tag{B.10}$$

where the constant C only depends on K and b. We then use $\boldsymbol{G}^N$ as the approximation for $\boldsymbol{N}$.

Step (ii). In this step, we aim to obtain the distribution of $\boldsymbol{N}^b|\mathcal{X}, \mathcal{Y}$. For this, we rewrite $\widehat{u}_{1,s}^b$ and $\widehat{u}_{2,s}^b$ in (2.1.3) as

$$\widehat{u}_{1,s}^b = \frac{m}{n_1}\sum_{k=1}^{n_1}(Q_{1k,s} - \widehat{u}_{1,s})\varepsilon_{1,k}^b, \ \ \widehat{u}_{2,s}^b = \frac{m}{n_2}\sum_{k=1}^{n_2}(Q_{2k,s} - \widehat{u}_{2,s})\varepsilon_{2,k}^b, \tag{B.11}$$

where $Q_{1k,s}$ and $Q_{2k,s}$ are defined in (2.2). Considering that $\varepsilon^b_{\gamma,1},\ldots,\varepsilon^b_{\gamma,n_\gamma}$ are i.i.d. standard normal random variables, therefore given $\mathcal{X}$ and $\mathcal{Y}$, $\widehat{\boldsymbol{u}}^b_\gamma := (\widehat{u}^b_{\gamma,1},\ldots,\widehat{u}^b_{\gamma,q})$ follows $N(\mathbf{0}, m^2\widehat{\boldsymbol{\Sigma}}_\gamma/n_\gamma)$ with $\widehat{\boldsymbol{\Sigma}}_\gamma := (\widehat{\sigma}_{\gamma,st}) \in \mathbb{R}^{q\times q}$ where

$$\widehat{\sigma}_{\gamma,st} = \frac{1}{n_1}\sum_{k=1}^{n_1}(Q_{\gamma k,s} - \widehat{u}_{\gamma,s})(Q_{\gamma k,t} - \widehat{u}_{\gamma,t}). \tag{B.12}$$

Apparently, by the definition of $\widehat{v}_{\gamma,s}$ in (2.1) we have $\widehat{v}_{\gamma,s} = m^2\widehat{\sigma}_{\gamma,ss}$. Therefore, by setting $\widehat{\boldsymbol{\Sigma}}_{12} = \widehat{\boldsymbol{\Sigma}}_1/n_1 + \widehat{\boldsymbol{\Sigma}}_2/n_2$ and $\widehat{\mathbf{D}}_{12} = \mathrm{Diag}(\widehat{\boldsymbol{\Sigma}}_{12})$, we have

$$\boldsymbol{N}^b|\mathcal{X},\mathcal{Y} = m^{-1}\widehat{\mathbf{D}}_{12}^{-1/2}(\widehat{\boldsymbol{u}}^b_1 - \widehat{\boldsymbol{u}}^b_2)|\mathcal{X},\mathcal{Y} \sim N(\mathbf{0}, \widehat{\mathbf{R}}_{12}),$$

where $\widehat{\mathbf{R}}_{12} = \widehat{\mathbf{D}}_{12}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\mathbf{D}}_{12}^{-1/2}$.

Step (iii). In this step, we combine results from previous two steps to justify the bootstrap procedure, i.e., we aim to prove

$$\sup_{z\in(0,\infty)}\left|\mathbb{P}(N_{(s_0,p)} > z) - \mathbb{P}(N^b_{(s_0,p)} > z|\mathcal{X},\mathcal{Y})\right| = o_p(1).$$

For this, we need both the lower and upper bounds of $\mathbb{P}\big(N_{(s_0,p)} > z\big) - \mathbb{P}\big(N^b_{(s_0,p)} > z|\mathcal{X},\mathcal{Y}\big)$. We first presents how to obtain the upper bounds. By the triangle inequality, we have

$$\mathbb{P}(\|\boldsymbol{N}\|_{(s_0,p)} > z) \le \mathbb{P}(\|\boldsymbol{H}^N\|_{(s_0,p)} > z-\varepsilon) + \underbrace{\mathbb{P}(\|\boldsymbol{N}-\boldsymbol{H}^N\|_{(s_0,p)} > \varepsilon)}_{\rho_1}. \tag{B.13}$$

By Lemmas B.1, we have $\rho_1 = o(1)$. We then bound $\mathbb{P}(\|\boldsymbol{H}^N\|_{(s_0,p)} > z-\varepsilon)$. For this, we have

$$\mathbb{P}(\|\boldsymbol{H}^N\|_{(s_0,p)} > z-\varepsilon) \le \rho_2 + \mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > z-\varepsilon), \tag{B.14}$$

where $\rho_2 = \sup_{x>0}\left|\mathbb{P}(\|\boldsymbol{H}^N\|_{(s_0,p)} > x) - \mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > x)\right|$. By (B.10), we have $\rho_2 \le Cn^{-\zeta_0}$ which yields

$$\mathbb{P}(\|\boldsymbol{N}\|_{(s_0,p)} > z) \le \underbrace{\mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > z-\varepsilon)}_{\rho_3} + o(1), \tag{B.15}$$

as $n\to\infty$. We then decompose ρ_3 as $\rho_3 = \mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > z) + \rho_4$ with

$$\rho_4 = \mathbb{P}(z-\varepsilon < \|\boldsymbol{G}^N\|_{(s_0,p)} \le z).$$

To control ρ_4, by utilizing the anti-concentration inequality for the the Gaussian random vector in Lemma A.4, we introducing the following lemma.

Lemma B.2. Assumptions **(A)** and **(M1)** hold. For any $z > 0$ and $\varepsilon = O(s_0 \log^2(qn) n^{-1/2})$, we have $\mathbb{P}(z - \varepsilon < \|\boldsymbol{G}^N\|_{(s_0,p)} \le z) = o(1)$ as $n \to \infty$.

The proof of Lemma B.2 is in Appendix C.2 of supplementary materials. By Lemma B.2, we then have

$$\mathbb{P}(\|\boldsymbol{N}\|_{(s_0,p)} > z) \le \mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > z) + o(1),$$

as $n \to \infty$. As is shown in **Step (ii)**, under the null hypothesis we have $\boldsymbol{N}^b | \mathcal{X}, \mathcal{Y} \sim N(\mathbf{0}, \widehat{\mathbf{R}}_{12})$. Considering $\boldsymbol{G}^N \sim N(\mathbf{0}, \mathbf{R}_{12})$, we have

$$\mathbb{P}(\|\boldsymbol{N}\|_{(s_0,p)} > z) - \mathbb{P}(\|\boldsymbol{N}^b\|_{(s_0,p)} > z | \mathcal{X}, \mathcal{Y}) \le \widehat{D}_5 + o(1). \tag{B.16}$$

with $\widehat{D}_5 = \sup_{z>0} \Big| \mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > z) - \mathbb{P}(\|\boldsymbol{N}^b\|_{(s_0,p)} > z | \mathcal{X}, \mathcal{Y}) \Big|$. The following lemma presents the upper bound of $\widehat{D}_5$.

Lemma B.3. Assumptions **(A)**, **(E)**, **(M1)** and **(M2)** hold. With probability at least $1 - C_1 n^{-1}$, we have $\widehat{D}_5 = o_p(1)$ as $n \to \infty$.

The proof of Lemma B.3 is in Appendix C.3 of supplementary materials. Therefore, we have

$$\sup_{z>0} \Big(\mathbb{P}(\|\boldsymbol{N}\|_{(s_0,p)} > z) - \mathbb{P}(\|\boldsymbol{N}^b\|_{(s_0,p)} > z | \mathcal{X}, \mathcal{Y}) \Big) = o_p(1), \tag{B.17}$$

uniformly for any $z > 0$. We can similarly construct the lower bound and obtain

$$\sup_{z>0} \Big| \mathbb{P}(\|\boldsymbol{N}\|_{(s_0,p)} > z) - \mathbb{P}(\|\boldsymbol{N}^b\|_{(s_0,p)} > z | \mathcal{X}, \mathcal{Y}) | = o_p(1),$$

which finishes the proof of (3.3) in Theorem 3.1. □

B.3. Proof of Corollary 3.1 .

Proof. In Corollary 3.1, we aim to prove (3.11) and (3.12). As the proof of (3.11) is similar, we only prove (3.12). As $\widehat{P}^N_{(s_0,p)} - P^N_{(s_0,p)} \to 0$ implies $\mathbb{P}_{\mathbf{H}_0}(T^N_{\alpha,(s_0,p)} = 1) \to \alpha$, for proving (3.12) we only need to prove that as $n, B \to \infty$, we have

$$\widehat{P}^N_{(s_0,p)} - P^N_{(s_0,p)} \to 0, \tag{B.18}$$

where $\widehat{P}^N_{(s_0,p)}$ is defined in (2.8) and $P^N_{(s_0,p)}$ is the oracle P-value of $N_{(s_0,p)}$. By introducing

$$\text{(B.19)}\qquad \begin{aligned} F_{N,(s_0,p)}(z) &= \mathbb{P}\big(\|\boldsymbol{N}\|_{(s_0,p)} \le z\big) \\ \widehat{F}_{N^b,(s_0,p)}(z) &= (B+1)^{-1}\Big(\sum_{b=1}^{B} \mathbb{I}\left\{N^b_{(s_0,p)} \le z|\mathcal{X},\mathcal{Y}\right\} + 1\Big), \end{aligned}$$

consider the definitions of $\widehat{P}^N_{(_s0,p)}$ and $P^N_{(s_0,p)}$, we have

$$\text{(B.20)}\qquad \widehat{P}^N_{(s_0,p)} = 1 - \widehat{F}_{N^b,(s_0,p)}\big(N_{(s_0,p)}\big),\ P^N_{(s_0,p)} = 1 - F_{N,(s_0,p)}\big(N_{(s_0,p)}\big).$$

According to Theorems 3.1, under Assumptions **(A)**, **(S)**, **(E)**, **(M1)**, and **(M2)**, by setting $T_1 = \big|1 - F_{N^b,(s_0,p)}\big(N_{(s_0,p)}\big) - P^N_{(s_0,p)}\big|$ with

$$\text{(B.21)}\qquad F_{N^b,(s_0,p)}(z) := \mathbb{P}\big(\|\boldsymbol{N}^b\|_{(s_0,p)} \le z\big|\mathcal{X},\mathcal{Y}\big),$$

we have $T_1 \to 0$ as $n \to \infty$. Considering (B.19) and (B.20), we use the triangle inequality to obtain $\left|P^N_{(s_0,p)} - \widehat{P}^N_{(s_0,p)}\right| \le T_1 + T_2$ with

$$T_2 = \left|F_{N^b,(s_0,p)}\big(N_{(s_0,p)}\big) - \widehat{F}_{N^b,(s_0,p)}\big(N_{(s_0,p)}\big)\right|$$

By Massart's inequality (see Section 1.5 in [10]), we have

$$\text{(B.22)}\qquad \sup_{z\in\mathbb{R}}\left|\widehat{F}_{N^b,(s_0,p)}(z) - F_{N^b,(s_0,p)}(z)\right| \to 0, \qquad \text{as } n, B \to \infty.$$

Therefor, as n, $B \to \infty$, we have $T_2 \to 0$, which finishes the proof.

□

B.4. Proof of Theorem 3.3.

Proof. For simplicity, we only consider the two-sample problem. The proof proceeds in two steps. In the first step, we give an upper bound of the oracle critical value

$$t^N_{\alpha,(s_0,p)} = \inf\Big\{t \in \mathbb{R} : \mathbb{P}\big(\|\boldsymbol{N}^b\|_{(s_0,p)} \le t|\mathcal{X},\mathcal{Y}\big) > \alpha\Big\}.$$

In the second step, with the obtained upper bound of $t^N_{\alpha,(s_0,p)}$, we construct a lower bound of $\mathbb{P}\big(N_{(s_0,p)} > t^N_{\alpha,(s_0,p)}\big)$. By showing that this lower bound goes to 1 under (3.15), we have

$$\mathbb{P}\big(N_{(s_0,p)} > t^N_{\alpha,(s_0,p)}\big) \to 1,$$

as $n, q \to \infty$. Considering that $\widehat{t}^N_{\alpha,(s_0,p)}$ is a bootstrap estimator for $t^N_{\alpha,(s_0,p)}$, under (3.15) we then have $\mathbb{P}\big(N_{(s_0,p)} > \widehat{t}^N_{\alpha,(s_0,p)}\big) \to 1$, as $n, B \to \infty$.

Step (i). In this step, we give an upper bound of $t^N_{\alpha,(s_0,p)}$. By the definition of $\boldsymbol{N}^b$ in (2.5), $\boldsymbol{N}^b|\mathcal{X},\mathcal{Y}$ is a q-dimensional Gaussian random vector with standard normal entries. According to Lemma A.5, by setting $\sigma = 1$ and $t = \sqrt{2\log q}$ we have

$$\mathbb{E}\big[\|\boldsymbol{N}^b\|_\infty|\mathcal{X},\mathcal{Y}\big] \le \sqrt{2\log q} + \frac{1}{\sqrt{2\log q}} = \sqrt{2\log q}\big(1+\{2\log q\}^{-1}\big). \tag{B.23}$$

By Theorem 5.8 of [4], we have

$$\mathbb{P}\Big(\|\boldsymbol{N}^b\|_\infty \ge \mathbb{E}\big[\|\boldsymbol{N}^b\|_\infty|\mathcal{X},\mathcal{Y}\big] + u\Big|\mathcal{X},\mathcal{Y}\Big) < \exp(-u^2/2). \tag{B.24}$$

By setting c_α as the α-quantile of $\|\boldsymbol{N}^b\|_\infty|\mathcal{X},\mathcal{Y}$, combining (B.23) and (B.24), we have

$$c_{1-\alpha} \le \sqrt{2\log q}\big(1+\{2\log q\}^{-1}\big) + \sqrt{2\log(1/\alpha)}. \tag{B.25}$$

Considering that $t^N_{\alpha,(s_0,p)}$ is the $1-\alpha$ quantile of $\|\boldsymbol{N}^b\|_{(s_0,p)}|\mathcal{X},\mathcal{Y}$, by the inequality $\|\boldsymbol{N}^b\|_{(s_0,p)} \le s_0^{1/p}\|\boldsymbol{N}^b\|_\infty$, we then have $t^N_{\alpha,(s_0,p)} \le s_0^{1/p}c_{1-\alpha}$. Therefore, by (B.25) we have

$$t^N_{\alpha,(s_0,p)} \le s_0^{1/p}\Big(\sqrt{2\log q}\big(1+\{2\log q\}^{-1}\big) + \sqrt{2\log(1/\alpha)}\Big). \tag{B.26}$$

Step (ii) In this step, we aim to obtain an lower bound of $\mathbb{P}\big(N_{(s_0,p)} > t^N_{\alpha,(s_0,p)}\big)$. By (B.26), we have $\mathbb{P}(N_{(s_0,p)} > t^N_{\alpha,(s_0,p)}) \ge L^N_1$, where

$$L^N_1 = \mathbb{P}\bigg(N_{(s_0,p)} > s_0^{1/p}\Big(\sqrt{2\log q}\big(1+\{2\log q\}^{-1}\big) + \sqrt{2\log(1/\alpha)}\Big)\bigg). \tag{B.27}$$

To obtain the lower bound of L^N_1, we need some additional notations. By setting N_s as $N_s = (\widehat{u}_{1,s} - \widehat{u}_{2,s})/\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}$, in (2.4), we define $N_{(s_0,p)} = \|\boldsymbol{N}\|_{(s_0,p)}$, where $\boldsymbol{N} = (N_1, \dots, N_q)^\top$. Under the alternative hypothesis, $u_{1,s} = u_{2,s}$ cannot hold for all $s \in \{1, \dots, q\}$, which motivates us to define

$$N^1_s = \frac{\widehat{u}_{1,s} - \widehat{u}_{2,s} - u_{1,s} + u_{2,s}}{\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}} \quad \text{and} \quad \boldsymbol{N}^1 = (N^1_1, \dots, N^1_q)^\top. \tag{B.28}$$

Considering that $\widehat{v}_{\gamma,s}$ is the variance estimator for $\sqrt{n_\gamma}\widehat{u}_{\gamma,s}$ and that $\widehat{v}_{\gamma,s}$ has the limit $m^2\sigma_{\gamma,ss}$ as $n_\gamma \to \infty$, we introduce $\boldsymbol{D}_2 = (D_{2,1}, \dots, D_{2,q})^\top$ and $\widehat{\boldsymbol{D}}_2 = (\widehat{D}_{2,1}, \dots, \widehat{D}_{2,q})^\top$, where

$$\begin{aligned} D_{2,s} &= |u_{1,s} - u_{2,s}| / \sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2} \\ \widehat{D}_{2,s} &= |u_{1,s} - u_{2,s}| / \sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}. \end{aligned} \tag{B.29}$$

Without loss of generality, we assume that largest s_0 entries of $\boldsymbol{D}_2$ is k_1, k_2, $\dots$, k_{s_0}. Therefore, by setting $\boldsymbol{k} = (k_1, \dots, k_{s_0})^\top$ under (3.15) we have

$$\|\boldsymbol{D}_2\|_{(s_0,p)} = \|(\boldsymbol{D}_2)_{\boldsymbol{k}}\|_p \geq s_0(1+\varepsilon_n)\Big(\sqrt{2\log q} + \sqrt{2\log(1/\alpha)}\Big), \tag{B.30}$$

where we set $\varepsilon_n \to 0$ and $\varepsilon_n\sqrt{\log q} \to \infty$ as $n \to \infty$. By the definition of (s_0, p) distance and the triangle inequality, we have

$$N_{(s_0,p)} \geq \|\boldsymbol{N}_{\boldsymbol{k}}\|_p \geq \|(\widehat{\boldsymbol{D}}_2)_{\boldsymbol{k}}\|_p - \|\boldsymbol{N}^1_{\boldsymbol{k}}\|_p. \tag{B.31}$$

As we impose conditions on $\boldsymbol{D}_2$ not on $\widehat{\boldsymbol{D}}_2$, by the definitions of $\boldsymbol{D}_2$ and $\widehat{\boldsymbol{D}}_2$ in (B.29) we need the estimation error of $\widehat{v}_{\gamma,s}$. By Lemme A.6, considering Assumption **(M1)**, for $m > 1$, with probability at least $1 - C_1 n^{-1}$, we have

$$\max_{\substack{\gamma=1,2 \\ s=1,\dots,q}} \Big|\sqrt{\frac{\widehat{v}_{\gamma,s}}{m^2\sigma_{\gamma,ss}}} - 1\Big| \leq C\frac{\log^{3/2}(qn)}{\sqrt{n}}, \tag{B.32}$$

when n is sufficiently large. Similarly, for $m = 1$ and sufficiently large n with probability at least $1 - C_1 n^{-1}$ we have

$$\max_{\substack{\gamma=1,2 \\ s=1,\dots,q}} \Big|\sqrt{\frac{\widehat{v}_{\gamma,s}}{\sigma_{\gamma,ss}}} - 1\Big| \leq C\sqrt{\frac{\log(qn)}{n}} + C\frac{\log^2(qn)}{n}. \tag{B.33}$$

Therefore, we introduce the event $\mathcal{E}_0(x)$ as

$$\mathcal{E}_0(x) = \left\{ \max_{\substack{\gamma=1,2 \\ s=1,\dots,q}} \Big|\sqrt{\frac{\widehat{v}_{\gamma,s}}{m^2\sigma_{\gamma,ss}}} - 1\Big| \leq x \right\}.$$

We set $x \asymp \log^{3/2}(qn)/\sqrt{n}$ for $m > 1$ and $x \asymp \sqrt{\log(qn)/n} + \log^2(qn)/n$ for $m = 1$. We then have $\mathbb{P}(\mathcal{E}_0(x)^c) \preceq n^{-1}$. By (B.31), under $\mathcal{E}_0(x)$ we have

$$N_{(s_0,p)} \geq \underbrace{\frac{1}{1+x}\|\boldsymbol{D}_2\|_{(s_0,p)} - \|\boldsymbol{N}^1_{\boldsymbol{k}}\|_p}_{L_2^N}. \tag{B.34}$$

Therefore, by partitioning the event based on $\mathcal{E}_0(x)$, we use (B.34) to obtain

$$L_1^N \geq \mathbb{P}\Big(L_2^N > s_0\big(1+\{2\log q\}^{-1}\big)\Big(\sqrt{2\log q}+\sqrt{2\log(1/\alpha)}\Big), \mathcal{E}_0(x)\Big).$$

Considering (B.30), by choosing u satisfying $(1+x)\big(1+u+\{2\log q\}^{-1}\big) = (1+\varepsilon_n)$ we have

$$L_1^N \geq \mathbb{P}\Big(\|\boldsymbol{N}_{\boldsymbol{k}}^1\|_p < s_0 u\sqrt{2\log q}\Big). \tag{B.35}$$

By the triangle inequality, for $i \in \{1,\dots,s_0\}$ we have

$$\mathbb{P}\Big(\|\boldsymbol{N}_{\boldsymbol{k}}^1\|_p \geq s_0 u\sqrt{2\log q}\Big) \leq s_0 \max_{1\leq i\leq s_0} \mathbb{P}\Big(|N_{k_i}^1| \geq u\sqrt{2\log q}\Big). \tag{B.36}$$

Therefore, combining (B.35) and (B.36) we have

$$L_1^N \geq 1-\mathbb{P}\Big(\|\boldsymbol{N}_{\boldsymbol{k}}^1\|_p \geq s_0 u\sqrt{2\log q}\Big) \geq 1 - s_0 \max_{1\leq i\leq s_0} \mathbb{P}\Big(|N_{k_i}^1| \geq u\sqrt{2\log q}\Big).$$

By the definition of L_1 in (B.27), to prove $L_1 \to 1$ we only need to obtain

$$s_0\mathbb{P}\big(|N_{k_i}^1| \geq u\sqrt{2\log q}\big) \to 0, \tag{B.37}$$

uniformly as $n, q \to \infty$. For this, we introduce the following lemma.

LEMMA B.4. Under Assumptions **(A)**$'$, **(E)**, **(M1)**, and **(M2)**, as $n, q \to \infty$, we have

$$s_0 \max_{s=1,\dots,q} \mathbb{P}\big(|N_s^1| \geq u\sqrt{2\log q}\big) \to 0. \tag{B.38}$$

The detailed proof of Lemma B.4 is in Appendix C.4 of supplementary materials. By Lemma B.4, we finish the proof. □

B.5. Proof of Theorem 3.5.

PROOF. In Theorem 3.5, we aim to prove (3.18) and (3.19). As the proof of (3.18) is similar, we only prove (3.19). The proof proceeds in two steps. In the first step, by setting $F_{N,\mathrm{ad}}(z) = \mathbb{P}(N_{\mathrm{ad}} \leq z|\mathcal{X},\mathcal{Y})$ and $\widetilde{F}_{N,\mathrm{ad}}(z) = \mathbb{P}(\widetilde{N}_{\mathrm{ad}} \leq z)$, we prove that as $n, B \to \infty$, we have

$$\widetilde{F}_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}}) - F_{N,\mathrm{ad}}(N_{\mathrm{ad}}) \to 0, \tag{B.39}$$

where N_{ad} and $\widetilde{N}_{\mathrm{ad}}$ are defined in (3.16) and (3.17). In the second step, we prove that

$$F_{N,\mathrm{ad}}(N_{\mathrm{ad}}) - \widehat{P}^{N}_{\mathrm{ad}} \to 0, \tag{B.40}$$

as $n, B \to \infty$. Combining (B.39) and (B.40), we can easily obtain (3.19).

Step (i). In this step, we aim to prove (B.39). For this, we need the following lemma to analyze the difference between the cumulative distribution functions of N_{ad} and $\widetilde{N}_{\mathrm{ad}}$.

LEMMA B.5. Assumptions **(A)**$''$, **(E)**, **(M1)** and **(M2)** hold. Under $\mathbf{H}_0$ of (1.8), we have that for any $\epsilon > 0$

$$\sup_{z\in[\epsilon,1-\epsilon]} \Big| F_{N,\mathrm{ad}}(z) - \widetilde{F}_{N,\mathrm{ad}}(z) \Big| = 0, \tag{B.41}$$

as $n, B \to \infty$.

The proof of Lemma B.5 is in Appendix C.5 of supplementary materials. After introducing Lemma B.5, we then prove (B.39). In detail, we aim to prove that for any $\delta, \epsilon' > 0$ we have

$$\underbrace{\mathbb{P}\Big(|\widetilde{F}_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}}) - F_{N,\mathrm{ad}}(N_{\mathrm{ad}})| \geq \delta\Big)}_{\Delta_1} < \epsilon' \tag{B.42}$$

as $n \to \infty$. By plugging in $F_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}})$, we use the triangle inequality to obtain $\Delta_1 \leq \Delta_2 + \Delta_3$, where

$$\begin{aligned} \Delta_2 =& \mathbb{P}\Big(|\widetilde{F}_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}}) - F_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}})| \geq \delta/2\Big), \\ \Delta_3 =& \mathbb{P}\Big(|F_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}}) - F_{N,\mathrm{ad}}(N_{\mathrm{ad}})| \geq \delta/2\Big). \end{aligned} \tag{B.43}$$

We then separately bound Δ_2 and Δ_3. To prove (B.42). We only need to show both $\Delta_2 < \epsilon'/2$ and $\Delta_3 < \epsilon'/2$ hold as n and B are sufficiently large. For Δ_2, by setting $\mathcal{E}_{\widetilde{N},\mathrm{ad}}(\epsilon) := \{\widetilde{N}_{\mathrm{ad}} \in [\epsilon, 1-\epsilon]\}$, we can bound Δ_2 by

$$\Delta_2 \leq \mathbb{P}\Big(|\widetilde{F}_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}}) - F_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}})| \geq \delta/2 \cap \mathcal{E}_{\widetilde{N},\mathrm{ad}}(\epsilon)\Big) + \Delta_4, \tag{B.44}$$

where $\Delta_4 = \mathbb{P}\big((\mathcal{E}_{\widetilde{N},\mathrm{ad}}(\epsilon))^c\big)$. By the definition of $\widetilde{N}_{\mathrm{ad}}$ in (3.17), by choosing ϵ small enough, we have $\Delta_4 \leq \epsilon'/4$. By Lemma B.5 and the definition of $\mathcal{E}_{\widetilde{N},\mathrm{ad}}(\epsilon)$, we also have

$$\mathbb{P}\Big(|\widetilde{F}_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}}) - F_{N,\mathrm{ad}}(\widetilde{N}_{\mathrm{ad}})| \geq \delta/2 \cap \mathcal{E}_{\widetilde{N},\mathrm{ad}}(\epsilon)\Big) \leq \epsilon'/4, \tag{B.45}$$

for sufficiently large n and B. Hence, we have $\Delta_2 \leq \epsilon'/2$ holds as n and B are sufficiently large. After the proof foe Δ_2, we then bound Δ_3. By the definition of $\widetilde{N}_{\rm ad}$ in (3.17) and Corollary 3.1, we have

$$|\widetilde{N}_{\rm ad} - N_{\rm ad}| \to 0, \qquad \text{as } n,\, B \to \infty. \tag{B.46}$$

Therefore, we obtain that $f_{N,{\rm ad}}(z) = F'_{N,{\rm ad}}(z)$ is uniformly bounded for sufficiently large n, B. Hence, there is a constant C such that

$$|F_{N,{\rm ad}}(\widetilde{N}_{\rm ad}) - F_{N,{\rm ad}}(N_{\rm ad})| \leq C|\widetilde{N}_{\rm ad} - N_{\rm ad}|. \tag{B.47}$$

Combining (B.43), (B.46), and (B.47), we have $\Delta_3 \leq \epsilon'/2$ for sufficiently large n and B. Therefore, we finish the proof of (B.39).

Step (ii). In this step, we aim to prove (B.40). For this, we introduce

$$\begin{aligned} &F_{N^b,(s_0,p)} = \mathbb{P}\Big(N^b_{(s_0,p)} \leq z|\mathcal{X},\mathcal{Y}\Big),\\ &N^b_{\rm ad} = \min_{p\in\mathcal{P}}\Big(1 - F_{N^b,(s_0,p)}(N^b_{(s_0,p)})\Big), \end{aligned} \tag{B.48}$$

where $N^b_{(s_0,p)}$ is defined in (2.6). Therefore, we define the cumulation distribution function of $N^b_{\rm ad}|\mathcal{X},\mathcal{Y}$ as

$$F_{N^b,{\rm ad}}(z) = \mathbb{P}(N^b_{\rm ad} \leq z|\mathcal{X},\mathcal{Y}). \tag{B.49}$$

Considering the definition of $\widehat{P}^N_{\rm ad}$ in (2.13), by setting

$$\widehat{F}_{N,{\rm ad}'}(z) = \left(\sum_{b=1}^{B} \mathbb{I}\{N^b_{{\rm ad}'} \leq z|\mathcal{X},\mathcal{Y}\} + 1\right)/(B+1), \tag{B.50}$$

we have $\widehat{P}^N_{\rm ad} = \widehat{F}_{N,{\rm ad}'}(N_{\rm ad})$.

To prove $F_{N,{\rm ad}}(N_{\rm ad}) - \widehat{P}^N_{\rm ad} \to 0$, by plugging in $F_{N^b,{\rm ad}}(N_{\rm ad})$ and using the triangle inequality, it is sufficient to prove

$$F_{N,{\rm ad}}(N_{\rm ad}) - F_{N^b,{\rm ad}}(N_{\rm ad}) \to 0 \quad \text{and} \quad F_{N^b,{\rm ad}}(N_{\rm ad}) - \widehat{P}^N_{\rm ad} \to 0, \tag{B.51}$$

as $n,\, B \to \infty$. To prove (B.51), we introduce the following two lemmas.

LEMMA B.6. Assumptions **(A)**$''$, **(E)**, **(M1)**, and **(M2)** hold. Under $\mathbf{H}_0$ of (1.8), by setting $F_{N,{\rm ad}}(z) = \mathbb{P}(N_{\rm ad} \leq z|\mathcal{X},\mathcal{Y})$ and $F_{N^b,{\rm ad}}(z) = \mathbb{P}(N^b_{\rm ad} \leq z|\mathcal{X},\mathcal{Y})$, we have

$$\sup_{z\in[\epsilon,1-\epsilon]} |F_{N,{\rm ad}}(z) - F_{N^b,{\rm ad}}(z)| \to 0, \qquad \text{as } n, B \to \infty, \tag{B.52}$$

for any $\epsilon > 0$.

LEMMA B.7. For any $\epsilon > 0$, we have that as $n, B \to \infty$,

$$\sup_{z\in[\epsilon,1-\epsilon]} |F_{N^b,\mathrm{ad}}(z) - \widehat{F}_{N,\mathrm{ad}'}(z)| \to 0, \tag{B.53}$$

where $\widehat{F}_{N,\mathrm{ad}'}(z)$ is defined in (B.50).

The proofs of Lemmas B.6 and B.7 are in Appendices C.6 and C.7 of supplementary materials. Let $\mathcal{E}_{N,\mathrm{ad}}(\epsilon) = \{N_{\mathrm{ad}} \in [\epsilon, 1-\epsilon]\}$. Considering Lemmas B.6 and B.7, by replacing $\mathcal{E}_{\widetilde{N},\mathrm{ad}}(\epsilon)$ with $\mathcal{E}_{N,\mathrm{ad}}$, similarly to (B.44) and (B.45) we can prove (B.51), which finishes the proof of Theorem 3.5.

□

B.6. Proof of Remark 3.4 .

PROOF. For $\boldsymbol{G} \sim N(\mathbf{0}, \mathbf{R}) \in \mathbb{R}^q$ with $q \geq 1$ fixed, the distribution of $\|\boldsymbol{G}\|_{(s_0,p)}$ is absolutely continuous with respect to the Lebesgue measure and its density function $f^{\boldsymbol{G}}_{(s_0,p)}$ is positive everywhere. This implies that for any $\epsilon > 0$, $\min_{c^{\boldsymbol{G}}_{\epsilon,(s_0,p)} \leq z \leq c^{\boldsymbol{G}}_{1-\epsilon,(s_0,p)}} f^{\boldsymbol{G}}_{(s_0,p)}(z) > 0$. To prove the result after taking infimum over all positive integers q, it suffices to show that as long as $\mathbf{R} \in \mathcal{R}$, the limiting distribution of $\|\boldsymbol{G}\|_{(s_0,p)}$ as $q \to \infty$ exists with an absolutely continuous density function. For this, we prove a stronger result, which characterizes the joint asymptotic distribution of the top s_0 order statistics of weakly dependent standard normal random variables. In detail, let $v^{(1)}, v^{(2)}, \ldots, v^{(q)}$ be an ascending sequence of the magnitudes of the coordinates of $\mathbf{v} \in \mathbb{R}^q$ such that $0 \leq v^{(1)} \leq v^{(2)} \leq \ldots \leq v^{(q)}$. Set $\boldsymbol{G} = (G_1, \ldots, G_q)^\top \sim N(\mathbf{0}, \mathbf{R})$ with $\mathbf{R} \in \mathcal{R}$ and $\boldsymbol{G}^I = (G^I_1, \ldots, G^I_q)^\top \sim N(\mathbf{0}, \mathbf{I}_q)$. Moreover, let $\varphi_j(\boldsymbol{G}) = G^{(q-j+1)}$ for $j = 1, \ldots, q$ and $a_q = 2\log q - \log(\log q)$. For any $\mathbf{x} = (x_1, x_2, \ldots, x_{s_0})$ with $x_1 > x_2 > \cdots > x_{s_0} > 0$, by setting $f_{\mathrm{ext}}(t_1, \ldots, t_{s_0}) = \exp\big(-\frac{1}{2}\sum_{j=1}^{s_0-1} t_j\big) g(t_{s_0}) I(t_1 > t_2 > \cdots > t_{s_0})$, where $g(t) = 2^{-1}\pi^{-1/2}\exp(-t/2 - \pi^{-1/2}e^{-t/2})$, we shall prove that as $q \to \infty$,

$$\begin{aligned} &\mathbb{P}\Big(\varphi_1^2(\boldsymbol{G}) \leq x_1 + a_q, \ldots, \varphi_{s_0}^2(\boldsymbol{G}) \leq x_{s_0} + a_q\Big) \\ &\longrightarrow \left(\frac{1}{2\sqrt{\pi}}\right)^{s_0-1} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{s_0}} f_{\mathrm{ext}}(t_1, \ldots, t_{s_0})\, dt_{s_0} \cdots dt_1, \end{aligned} \tag{B.54}$$

holds uniformly for $\mathbf{R} \in \mathcal{R}$.

For simplicity, we only prove (B.54) for $s_0 = 2$, as the general case can be dealt with similarly. Let $y_{jq} = \sqrt{x_j + a_q}$ for $j = 1, 2$, and note

$$\Big\{\varphi_1(\boldsymbol{G}) > \sqrt{x_1 + a_q}, \varphi_2(\boldsymbol{G}) > \sqrt{x_2 + a_q}\Big\}$$
$$= \bigcup_{1\le i\ne j\le q} \Big\{(|G_i|, |G_j|) > (y_{1q}, y_{2q})\Big\} = \bigcup_{k=1}^{2\bar{q}} \Big\{(|G_{i_k}|, |G_{j_k}|) > (y_{1q}, y_{2q})\Big\},$$

where $\{(i_k, j_k)\}_{k=1}^{2\bar{q}} = \{(1,2), (1,3)\ldots,(1,q), (2,1), (2,3)\ldots,(2,q),\ldots,(q,q-1)\}$ and $\bar{q} = q(q-1)/2$. By the Bonferroni inequality, for any fixed $k < \bar{q}$, we have

$$\sum_{\ell=1}^{2k}(-1)^{\ell-1}E_\ell \le \mathbb{P}\Big(\varphi_1(\boldsymbol{G}) > \sqrt{x_1 + a_q}, \varphi_2(\boldsymbol{G}) > \sqrt{x_2 + a_q}\Big) \le \sum_{\ell=1}^{2k-1}(-1)^{\ell-1}E_\ell, \tag{B.55}$$

where

$$E_\ell = \sum_{1\le k_1<\cdots<k_\ell\le 2\bar{q}} \mathbb{P}\Big(|G_{i_{k_1}}| > y_{1q}, |G_{j_{k_1}}| > y_{2q}, \ldots, |G_{i_{k_\ell}}| > y_{1q}, |G_{j_{k_\ell}}| > y_{2q}\Big).$$

Moreover, for every $2 \le t \le 2\ell$, define

$$E_{\ell,t} = \sum_{\substack{1\le k_1<\cdots<k_\ell\le 2\bar{q} \\ \#\{i_{k_1}, j_{k_1},\ldots,i_{k_\ell}, j_{k_\ell}\}=t}} \underbrace{\mathbb{P}\Big(\min_{1\le\nu\le\ell}|G_{i_{k_\nu}}| > y_{1q}, \min_{1\le\nu\le\ell}|G_{j_{k_\nu}}| > y_{2q}\Big)}_{P_{k_1,\ldots,k_\ell}}. \tag{B.56}$$

We define index sets $\mathcal{I}^c$, $\mathcal{I}$, and $\mathcal{I}_k$ in the same way as in the proof of Lemma 6 in [6]. Therefore, we have $\mathcal{I} = \cup_{k=1}^{t-1}\mathcal{I}_k$. Further, for $1 \le i_1 < \cdots < i_t \le q$, by defining

$$Q(i_1,\ldots,i_t) = \Big\{1 \le k_1 < \cdots < k_\ell \le 2\bar{q} : \{i_{k_1}, j_{k_1}, \ldots, i_{k_\ell}, j_{k_\ell}\} = \{i_1,\ldots,i_t\}\Big\},$$

with $\#Q(i_1,\ldots,i_t) \le \binom{t(t-1)}{\ell}$, we have

$$E_{\ell,t} = \underbrace{\sum_{(i_1,\ldots,i_t)\in\mathcal{I}^c} \sum_{\substack{(k_1,\ldots,k_\ell) \\ \in Q(i_1,\ldots,i_t)}} P_{k_1,\ldots,k_\ell}}_{M_1(\ell,t)} + \underbrace{\sum_{(i_1,\ldots,i_t)\in\mathcal{I}} \sum_{\substack{(k_1,\ldots,k_\ell) \\ \in Q(i_1,\ldots,i_t)}} P_{k_1,\ldots,k_\ell}}_{M_2(\ell,t)}.$$

For $(k_1,\ldots,k_\ell) \in Q(i_1,\ldots,i_t)$ with $(i_1,\ldots,i_t) \in \mathcal{I}^c$, a straightforward adaptation of the arguments used to prove (20) in [6] yields that, as $q \to \infty$,

$$P_{k_1,\ldots,k_\ell} = \{1 + o(1)\}P^I_{k_1,\ldots,k_\ell}, \tag{B.57}$$

where $P^{I}_{k_1,\ldots,k_\ell}$ is defined in the same way as $P_{k_1,\ldots,k_\ell}$ in (B.56) by replacing G_i with G^{I}_i.

For $(k_1,\ldots,k_\ell)\in Q(i_1,\ldots,i_t)$ with $(i_1,\ldots,i_t)\in\mathcal{I}_k$ for some $1\le k\le t-1$, considering $y_{1q}>y_{2q}$, we have

$$P_{k_1,\ldots,k_\ell}\le\mathbb{P}(|G_{i_1}|>y_{2q},\ldots,|G_{i_t}|>y_{2q}):=\widetilde{P}_{i_1,\ldots,i_t}.$$

Now, it follows from (21) in [6] with slight modification that, as $q\to\infty$,

(B.58)
$$M_2(\ell,t)\le\sum_{(i_1,\ldots,i_t)\in\mathcal{I}}\binom{t(t-1)}{\ell}\widetilde{P}_{i_1,\ldots,i_t}=\binom{t(t-1)}{\ell}\sum_{(i_1,\ldots,i_t)\in\mathcal{I}}\widetilde{P}_{i_1,\ldots,i_t}\to 0.$$

We define $M^{I}_2(\ell,t)$ by replacing entries of $\boldsymbol{G}$ with the corresponding entries of $\boldsymbol{G}^{I}$ in $M_2(\ell,t)$. Similarly to (B.58), we have $M^{I}_2(\ell,t)=o(1)$ as $q\to\infty$. Therefore, as $q\to\infty$, we have

$$E_\ell=\sum_{t=2}^{2\ell}E_{\ell,t}=\{1+o(1)\}\sum_{1\le k_1<\cdots<k_\ell\le 2\bar{q}}P^{I}_{k_1,\ldots,k_\ell}+o(1).$$

This, together with (B.55) implies that, as $q\to\infty$,

(B.59)
$$\begin{aligned}&\{1+o(1)\}\sum_{\ell=1}^{2k}(-1)^{\ell-1}\sum_{1\le k_1<\cdots<k_\ell\le 2\bar{q}}P^{I}_{k_1,\ldots,k_\ell}+o(1)\\ \le&\mathbb{P}\Big(\varphi_1(\boldsymbol{G})>\sqrt{x_1+a_q},\varphi_2(\boldsymbol{G})>\sqrt{x_2+a_q}\Big)\\ \le&\{1+o(1)\}\sum_{\ell=1}^{2k-1}(-1)^{\ell-1}\sum_{1\le k_1<\cdots<k_\ell\le 2\bar{q}}P^{I}_{k_1,\ldots,k_\ell}+o(1).\end{aligned}$$

On the other hand, observing

(B.60)
$$\begin{aligned}&\mathbb{P}\Big(\varphi_1(\boldsymbol{G}^{I})>\sqrt{x_1+a_q},\varphi_2(\boldsymbol{G}^{I})>\sqrt{x_2+a_q}\Big)\\ &=\lim_{k\to\infty}\sum_{\ell=1}^{2k}(-1)^{\ell-1}\sum_{1\le k_1<\cdots<k_\ell\le 2\bar{q}}P^{I}_{k_1,\ldots,k_\ell},\end{aligned}$$

and $a_q=2\log q-\log(\log q)$, by [8], the bivariate vector $(\varphi_1^2(\boldsymbol{G}^{I})-a_q,\varphi_2^2(\boldsymbol{G}^{I})-a_q)$ has a limiting distribution with joint density function

$$g_2(t_1,t_2)=\frac{g(t_1)g(t_2)}{G(t_1)}=\frac{e^{-t_1/2}}{2\sqrt{\pi}}g(t_2),\qquad\text{for}\quad t_1>t_2,$$

where $G(t)=\exp(-\pi^{-1/2}e^{-t/2})$ and $g(t)=G'(t)$. Therefore, the limit in (B.60) is equal to $\int_{x_1}^{\infty}\int_{x_2}^{\infty}g_2(t_1,t_2)I(t_1>t_2)\,dt_2\,dt_1$, which together with (B.59) proves (B.54) by letting $q\to\infty$ first and then $k\to\infty$. □

B.7. Proof of Theorem 3.7.

Proof. For simplicity, we only consider the two-sample problem. In detail, we aim to prove

$$\mathbb{P}_{\mathbf{H}_1}\big(T_{\mathrm{ad}}^N = 1\big) \to 1, \qquad \text{as } n,\, B \to \infty. \tag{B.61}$$

under (3.21) and some assumptions. By the definition of T_{ad}^N in (2.14), for proving (B.61), it is equivalent to prove

$$\mathbb{P}_{\mathbf{H}_1}\big(\widehat{P}_{\mathrm{ad}}^N \le \alpha\big) \to 1, \qquad \text{as } n,\, B \to \infty. \tag{B.62}$$

By the definition of $\widehat{P}_{\mathrm{ad}}^N$ and $\widehat{F}_{N,\mathrm{ad}'}(z)$ in (2.13) and (B.50), (B.62) becomes

$$\mathbb{P}_{\mathbf{H}_1}\big(\widehat{F}_{N,\mathrm{ad}'}(N_{\mathrm{ad}}) < \alpha\big) \to 1, \qquad \text{as } n,\, B \to \infty. \tag{B.63}$$

Therefore, to obtain (B.61), it is sufficient to prove (B.63). By setting $\alpha' = \alpha/\#\{\mathcal{P}\}$, we can prove that α is also an upper bound of $\widehat{F}_{N,\mathrm{ad}'}(\alpha')$, i.e.,

$$\mathbb{P}_{\mathbf{H}_1}\Big(\widehat{F}_{N,\mathrm{ad}'}\big(\alpha'\big) \le \alpha\Big) \to 1, \qquad \text{as } n,\, B \to \infty. \tag{B.64}$$

By (B.64), to obtain (B.63) it is sufficient to prove

$$\mathbb{P}_{\mathbf{H}_1}\big(N_{\mathrm{ad}} \le \alpha'\big) \to 1, \qquad \text{as } n,\, B \to \infty. \tag{B.65}$$

By the definition of N_{ad} in (2.10), we have

$$\mathbb{P}_{\mathbf{H}_1}\big(\widehat{P}_{(s_0,p)}^N \le \alpha'\big) \le \mathbb{P}_{\mathbf{H}_1}\big(N_{\mathrm{ad}} \le \alpha'\big), \tag{B.66}$$

for any $p \in \mathcal{P}$. By Theorem 3.3, under (3.21) we have

$$\mathbb{P}_{\mathbf{H}_1}\big(\widehat{P}_{(s_0,p)}^N \le \alpha'\big) \to 1, \qquad \text{as } n,\, B \to \infty. \tag{B.67}$$

Combining (B.66) and (B.67), we prove (B.65).

To complete the proof, we now prove (B.64). By Lemma B.7, for any $0 < \alpha < 1$, we have

$$\mathbb{P}\Big(\widehat{F}_{N,\mathrm{ad}'}\big(\alpha'\big) \le \alpha\Big) = \mathbb{P}\Big(F_{N^b,\mathrm{ad}}(\alpha') \le \alpha\Big), \qquad \text{as } n,\, B \to \infty. \tag{B.68}$$

Moreover, by the definition of $F_{N^b,\mathrm{ad}}(z)$ in (B.49), we have that $F_{N^b,\mathrm{ad}}(\alpha') \le \alpha$ holds with probability 1, which yields (B.64). □

APPENDIX C: PROOF OF LEMMAS IN APPENDIX B

C.1. Proof of Lemma B.1.

PROOF. To prove Lemma B.1, we need to bound $\mathbb{P}\Big(\|\boldsymbol{N}-\boldsymbol{H}^N\|_{(s_0,p)} > \varepsilon\Big)$, where $\varepsilon = Cs_0\log^2(qn)n^{-1/2}$. We first prove for $m > 1$. For this, we set $\widehat{\boldsymbol{H}}^N = (\widehat{H}_1^N, \ldots, \widehat{H}_q^N)^\top$ with

$$\widehat{H}_s^N = \Big(\frac{1}{n_1}\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k) - \frac{1}{n_2}\sum_{k=1}^{n_2} h_s(\boldsymbol{Y}_k)\Big)\Big/\sqrt{\widehat{\sigma}_{1,ss}/n_1 + \widehat{\sigma}_{2,ss}/n_2}. \tag{C.1}$$

By plugging $\widehat{\boldsymbol{H}}^N$, we have $\mathbb{P}(\|\boldsymbol{N}-\boldsymbol{H}^N\|_{(s_0,p)} > \varepsilon) \le D_1 + D_2$ with

$$D_1 = \mathbb{P}\Big(\|\boldsymbol{N}-\widehat{\boldsymbol{H}}^N\|_{(s_0,p)} > \varepsilon/2\Big),\ D_2 = \mathbb{P}\Big(\|\widehat{\boldsymbol{H}}^N-\boldsymbol{H}^N\|_{(s_0,p)} > \varepsilon/2\Big).$$

Therefore, we only need to separately prove $D_1 = o(1)$ and $D_2 = o(1)$ as $n \to \infty$.

For proving $D_1 = o(1)$, by setting $\mathcal{E}_{12} := \{\min_{s,\gamma}\widehat{\sigma}_{\gamma,ss} > b/2\}$, we have

$$D_1 \le \underbrace{\mathbb{P}\Big(\big\|\boldsymbol{N}-\widehat{\boldsymbol{H}}^N\|_{(s_0,p)} > \varepsilon/2 \cap \mathcal{E}_{12}\Big)}_{I_1} + \mathbb{P}\big(\mathcal{E}_{12}^c\big). \tag{C.2}$$

Considering Assumptions **(A)** and **(M1)**, by Lemma A.6, we have $\mathbb{P}\big(\mathcal{E}_{12}^c\big) = o(1)$ as $n \to \infty$. Hence, we only need to prove $I_1 = o(1)$ as $n \to \infty$. By the Hoeffding's decomposition, considering $\widehat{v}_{\gamma,s} = m^2\widehat{\sigma}_{\gamma,ss}$ and $\|\mathbf{v}\|_{(s_0,p)} \le s_0^{1/p}\|\mathbf{v}\|_\infty$, we have

$$I_1 \le \mathbb{P}\left(\max_{1\le s\le q}\left|\binom{n_1}{m}^{-1}\Delta_{n_1,s} - \binom{n_2}{m}^{-1}\Delta_{n_2,s}\right| > \frac{mb^{1/2}\varepsilon}{\sqrt{2}s_0^{1/p}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right),$$

where $\Delta_{n_1,s}$ and $\Delta_{n_2,s}$ are residuals of the Hoeffding's decomposition. For bounding the residuals, we threshold the kernel by $B_n = C\log(qn)$. For this, we introduce

$$\begin{aligned} V_{1,s}^{i_1,\ldots,i_m} &= \Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})\,\mathbb{I}\{|\Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})| \le B_n\}, \\ E_{1,s} &= \mathbb{E}\Big(\Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})\,\mathbb{I}\{|\Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})| \le B_n\}\Big), \end{aligned} \tag{C.3}$$

and denote the thresholded kernel and Hoeffding's projection by

$$\begin{aligned} \widehat{\Psi}_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m}) &= V_{1,s}^{i_1,\ldots,i_m} - E_{1,s}, \\ \widehat{h}_s(\boldsymbol{X}_i) &= \mathbb{E}\Big(\widehat{\Psi}_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})|\boldsymbol{X}_i\Big). \end{aligned} \tag{C.4}$$

Hence, the corresponding residuals become

$$\widehat{\Delta}_{n_1,s} = \sum_{1\le i_1<\ldots<i_m\le n_1} \Big(\widehat{\Psi}_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m}) - \sum_{\ell=1}^{m} \widehat{h}_s(\boldsymbol{X}_{i_\ell})\Big),$$

By the definitions of both $\Delta_{n_1,s}$ and $\widehat{\Delta}_{n_1,s}$, we then have

$$|\Delta_{n_1,s} - \widehat{\Delta}_{n_1,s}| \le \Bigg|\Delta_{n_1,s} - \Big(\sum_{1\le i_1<\ldots<i_m\le n_1} V_{1,s}^{i_1,\ldots,i_m} - \sum_{\ell=1}^{m} \mathbb{E}(V_{1,s}^{i_1,\ldots,i_m}|\boldsymbol{X}_{i_\ell})\Big)\Bigg| + (m-1)\binom{n_1}{m}|E_{1,s}|.$$

Considering that $\varepsilon = Cs_0\log^2(qn)n^{-1/2}$, we have

$$\frac{mb^{1/2}\varepsilon}{\sqrt{2}s_0^{1/p}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = O(\log^2(qn)/n).$$

By choosing a proper constant C in B_n, considering Assumption **(E)**, we have $\max_s(|E_{1,s}|+|E_{2,s}|) \prec \log^2(qn)/n$. Hence, when n is sufficiently large, we use the triangle inequality to get $I_1 \le I_{1,1} + I_{1,2}$, where

$$I_{1,1} = P\left(\max_{1\le s\le q}\left|\binom{n_1}{m}^{-1}\widehat{\Delta}_{n_1,s} - \binom{n_2}{m}^{-1}\widehat{\Delta}_{n_2,s}\right| > C\frac{\log^2(qn)}{n}\right)$$

$$I_{1,2} = Cqn^m \max_{\substack{s,i_\ell,j_\ell \\ \ell=1,\ldots,m}} \big(\mathbb{P}(|\Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})|>B_n)+\mathbb{P}(|\Psi_s(\boldsymbol{Y}_{j_1},\ldots,\boldsymbol{Y}_{j_m})|>B_n)\big)$$

By choosing a proper constant C in B_n, considering Assumption **(E)**, we have $I_{1,2} = o(1)$. For $I_{1,1}$, by Proposition 2.3 (c) in [1], we obtain

$$I_{1,1} \le Cq\exp\big(-C_1 n^{1-\frac{2}{m}}\log^{\frac{2}{m}}(qn)\big). \tag{C.5}$$

Considering $m \ge 2$ and Assumption **(A)**, we then have $I_{1,1} = o(1)$. Therefore, we prove that $D_1 = o(1)$, as $n \to \infty$.

After the proof for D_1, we then prove that $D_2 = o(1)$. Considering

$$\|\widehat{\boldsymbol{H}}^N - \boldsymbol{H}^N\|_{(s_0,p)} \le s_0^{1/p}\|\widehat{\boldsymbol{H}}^N - \boldsymbol{H}^N\|_\infty, \tag{C.6}$$

we have $D_2 \le \mathbb{P}(\|\widehat{\boldsymbol{H}}^N - \boldsymbol{H}^N\|_\infty > 0.5s_0^{-1/p}\varepsilon)$. By the definitions of $\boldsymbol{H}^N$ and $\widehat{\boldsymbol{H}}^N$ in (B.7) and (C.1), we have $\|\widehat{\boldsymbol{H}}^N - \boldsymbol{H}^N\|_\infty \le I_2 I_3$ with

$$\begin{aligned} I_2 &= \max_{1\le s\le q}\frac{|\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k) - \rho\sum_{k=1}^{n_2} h_s(\boldsymbol{Y}_k)|}{\sqrt{n_1\sigma_{1,ss}+\rho^2 n_2\sigma_{2,ss}}}, \\ I_3 &= \max_{1\le s\le q}\left|1 - \frac{\sqrt{\sigma_{1,ss}+\rho\sigma_{2,ss}}}{\sqrt{\widehat{\sigma}_{1,ss}+\rho\widehat{\sigma}_{2,ss}}}\right|, \end{aligned} \tag{C.7}$$

where $\rho = n_1/n_2$. By Assumption **(E)** and exponential inequality, we have that $I_2 \leq C\sqrt{\log(qn)}$ holds with probability $1 - C_1 n^{-1}$.

For bounding I_3, we introduce the following lemma.

LEMMA C.1. $\xi_1, \ldots, \xi_s \in \mathbb{R}$ are positive random variables with $\xi_s > 0$. For $y \in (0, 1]$, we have

$$\mathbb{P}\Big(\max_{1\leq s\leq q} |1 - \xi_s| \leq y/2\Big) \leq \mathbb{P}\Big(\max_{1\leq s\leq q} |1 - \xi_s^{-1}| \leq y\Big). \tag{C.8}$$

The detailed proof of Lemma C.1 is in Appendix D.1. Motivated by Lemma C.1, we introduce

$$I_3' := \max_{1\leq s\leq q} \Big|1 - \frac{\sqrt{\widehat{\sigma}_{1,ss} + \rho\widehat{\sigma}_{2,ss}}}{\sqrt{\sigma_{1,ss} + \rho\sigma_{2,ss}}}\Big|.$$

By Assumption **(M1)**, considering $(a + b)(a - b) = a^2 - b^2$, we use the triangle inequality to obtain

$$\begin{aligned} I_3' &\leq \max_{1\leq s\leq q} (\sigma_{1,ss} + \rho\sigma_{2,ss})^{-1} |\widehat{\sigma}_{1,ss} + \rho\widehat{\sigma}_{2,ss} - \sigma_{1,ss} - \rho\sigma_{2,ss}| \\ &\leq (1 + \rho)^{-1} b^{-1} \Big(\max_{1\leq s\leq q} |\widehat{\sigma}_{1,ss} - \sigma_{1,ss}| + \rho \max_{1\leq s\leq q} |\widehat{\sigma}_{2,ss} - \sigma_{2,ss}|\Big). \end{aligned}$$

Therefore, by Lemma A.6, $I_3' \leq C \log^{3/2}(qn) n^{-1/2}$ holds with probability $1 - C_1 n^{-1}$. By Lemma C.1, we then have that $I_3 \leq C \log^{3/2}(qn) n^{-1/2}$ holds with probability $1 - C_1 n^{-1}$ for sufficiently large n. Combining (C.6) and the bound for I_2 and I_3, we then have

$$\|\widehat{\boldsymbol{H}}^N - \boldsymbol{H}^N\|_{(s_0,p)} \leq C s_0 \frac{\log^2(qn)}{\sqrt{n}}$$

with probability $1 - C_1 n^{-1}$ for sufficiently large n. Therefore, we have $D_2 = o(1)$ as $n \to \infty$, which finishes the proof for $m > 1$.

By Lemma A.6 and similar proof, we can also prove for $m = 1$. As the proof is much easier and similar to the proof for $m > 1$, we omit the proof here. □

C.2. Proof of Lemma B.2.

PROOF. For notational simplicity, we set

$$L_{z,\varepsilon} := \mathbb{P}(\|\boldsymbol{G}\|_{(s_0,p)} \leq z + \varepsilon) - \mathbb{P}(\|\boldsymbol{G}\|_{(s_0,p)} \leq z),$$

where $z > 0$ and $\varepsilon = O(s_0 \log^2(qn) n^{-1/2})$. Let $\mathcal{E}^{R,q} = \{\mathbf{x} \in \mathbb{R}^q : \|\mathbf{x}\| \le R\}$ and $V^{z,q}_{(s_0,p)} = \{\mathbf{x} \in \mathbb{R}^q : \|\mathbf{x}\|_{(s_0,p)} \le z\}$. We then have

$$L_{z,\varepsilon} \le \underbrace{\mathbb{P}\big(\boldsymbol{G} \in \mathbb{R}^q \backslash \mathcal{E}^{R,q}\big)}_{L_1} + \underbrace{\mathbb{P}\big(\boldsymbol{G} \in V^{z+\varepsilon,q}_{(s_0,p)} \cap \mathcal{E}^{R,q}\big) - \mathbb{P}\big(\boldsymbol{G} \in V^{z,q}_{(s_0,p)} \cap \mathcal{E}^{R,q}\big)}_{L_2}.$$

By the tail probability of Gaussian distribution, we have

$$L_1 \le q(2\pi R^2 q^{-1})^{-1/2} \exp(-R^2 q^{-1}/2).$$

For L_2, by Lemma A.3, there is a m-generated convex set $A^m \in \mathbb{R}^q$ such that

$$A^m \subset V^{z,q}_{(s_0,p)} \cap \mathcal{E}^{R,q} \subset A^{m,R\epsilon} \qquad \text{and} \qquad m \le q^{s_0}\Big(\frac{\gamma}{\sqrt{\epsilon}} \ln \frac{1}{\epsilon}\Big)^{s_0^2}. \tag{C.9}$$

Hence, there is a constant C such that

$$V^{z+\varepsilon}_{(s_0,p)} \cap \mathcal{E}^{R,q} \subset A^{m,R\epsilon + C\varepsilon} \tag{C.10}$$

By setting $R = qn$ and $\epsilon = (qn)^{-2}$, we have $R\epsilon \prec \varepsilon$ and $L_1 = o(1)$. By Lemma A.4, we combine (C.9) and (C.10) to obtain $L_2 \le C\varepsilon s_0 \sqrt{\log(qn)} = O(s_0^2 \log^{5/2}(qn) n^{-1/2})$. By Assumption **(A)**, we have $L_2 = o(1)$, which finishes the proof. □

C.3. Proof of Lemma B.3.

Proof. In Lemma B.3, we aim to bound $\widehat{D}_5$, where

$$\widehat{D}_5 := \sup_{z>0} \Big| \mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > z) - \mathbb{P}(\|\boldsymbol{N}^b\|_{(s_0,p)} > z | \mathcal{X}, \mathcal{Y}) \Big|.$$

To bound $\widehat{D}_5$, we need to analyze the distributions of $\boldsymbol{G}^N$ and $\boldsymbol{N}^b | \mathcal{X}, \mathcal{Y}$. Considering the definitions of $\boldsymbol{\Sigma}_\gamma$ and $\widehat{\boldsymbol{\Sigma}}_\gamma$ in (3.4) and (3.9), by setting

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_1/n_1 + \boldsymbol{\Sigma}_2/n_2 \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{12} = \widehat{\boldsymbol{\Sigma}}_1/n_1 + \widehat{\boldsymbol{\Sigma}}_2/n_2,$$

we have $\boldsymbol{G}^N \sim N(\mathbf{0}, \mathbf{R}_{12})$ and $\boldsymbol{N}^b | \mathcal{X}, \mathcal{Y} \sim N(\mathbf{0}, \widehat{\mathbf{R}}_{12})$, where $\mathbf{R}_{12}$ and $\widehat{\mathbf{R}}_{12}$ are defined as

$$\begin{aligned} \mathbf{R}_{12} &= \mathrm{Diag}(\boldsymbol{\Sigma}_{12})^{-1/2} \boldsymbol{\Sigma}_{12} \mathrm{Diag}(\boldsymbol{\Sigma}_{12})^{-1/2} = (r_{12,ij})_{1 \le i,j \le q}, \\ \widehat{\mathbf{R}}_{12} &= \mathrm{Diag}(\widehat{\boldsymbol{\Sigma}}_{12})^{-1/2} \widehat{\boldsymbol{\Sigma}}_{12} \mathrm{Diag}(\widehat{\boldsymbol{\Sigma}}_{12})^{-1/2} = (\widehat{r}_{12,ij})_{1 \le i,j \le q}. \end{aligned} \tag{C.11}$$

Lemma C.2.

After analyzing the distributions of $\boldsymbol{G}^N$ and $\boldsymbol{N}^b|\mathcal{X},\mathcal{Y}$, we then bound $\widehat{D}_5$. For this, we rewrite $\widehat{D}_5$ as $\widehat{D}_5 = \max\big(\sup_{z\in(0,\widetilde{R}]} I_z, \sup_{z\in(\widetilde{R},\infty)} I_z\big)$, where

$$I_z = \big|\mathbb{P}(\|\boldsymbol{G}^N\|_{(s_0,p)} > z) - \mathbb{P}(\|\boldsymbol{N}^b\|_{(s_0,p)} > z|\mathcal{X},\mathcal{Y})\big|,$$

and $\widetilde{R} = Cs_0\sqrt{n}$. For $\sup_{z\in(\widetilde{R},\infty)} I_z$, considering $\|\mathbf{v}\|_{(s_0,p)} \le s_0^{1/p}\|\mathbf{v}\|_\infty \le s_0\|\mathbf{v}\|_\infty$, we have

$$\text{(C.12)} \qquad \sup_{z\in(R,\infty)} I_z \le \mathbb{P}(\|\boldsymbol{G}^N\|_\infty > C\sqrt{n}) + \mathbb{P}(\|\boldsymbol{N}^b\|_\infty > C\sqrt{n}|\mathcal{X},\mathcal{Y}).$$

Considering $r_{12,ii} = \widehat{r}_{12,ii} = 1$, by the tail probability of Gaussian distribution, we further have

$$\text{(C.13)} \qquad \sup_{z\in(\widetilde{R},\infty)} I_z \le Cq\exp(-C_1 n) = o(1).$$

We now bound $\sup_{z\in(0,\widetilde{R}]} I_z^D$. Let $\mathcal{E}^{\widetilde{R},q} = \{\mathbf{x}\in\mathbb{R}^q : \|\mathbf{x}\| \le \widetilde{R}\}$ and $V_{(s_0,p)}^{z,q} = \{\mathbf{x}\in\mathbb{R}^q : \|\mathbf{x}\|_{(s_0,p)} \le z\}$. Hence, considering $\|\mathbf{x}\| \le q^{1/2}\|\mathbf{x}\|_\infty \le q^{1/2}\|\mathbf{x}\|_{(s_0,p)}$, we have $V_{(s_0,p)}^{z,q} \subset \mathcal{E}^{\widetilde{R}q^{1/2},q}$ for $z < \widetilde{R}$. Therefore, Considering Lemma A.3, there is a m-generated convex set A^m and $\epsilon > 0$ such that

$$A^m \subset V_{(s_0,p)}^{z,d} \subset A^{m,\widetilde{R}q^{1/2}\epsilon} \qquad \text{and} \qquad m \le d^{s_0}\Big(\frac{\gamma}{\sqrt{\epsilon}}\ln\frac{1}{\epsilon}\Big)^{s_0^2}.$$

Let $\varepsilon' = Rq^{1/2}\epsilon$. By setting $\epsilon = (qn)^{-3/2}$, we have $\varepsilon' = s_0(qn)^{-1}$. We then have $I_z \le L_{z,1} + L_{z,2}$ with

$$\text{(C.14)} \qquad \begin{aligned} L_{z,1} &= \max\big(\mathbb{P}(\boldsymbol{G}^N \in A^{m,\varepsilon'}\backslash A^m), \mathbb{P}(\boldsymbol{N}^b \in A^{m,\varepsilon'}\backslash A^m)\big) \\ L_{z,2} &= \max\Big(\big|\mathbb{P}(\boldsymbol{G}^N \in A^{m,\varepsilon'}) - \mathbb{P}(\boldsymbol{N}^b \in A^{m,\varepsilon'}|\mathcal{X},\mathcal{Y})\big|, \\ &\qquad \big|\mathbb{P}(\boldsymbol{G}^N \in A^m) - \mathbb{P}(\boldsymbol{N}^b \in A^m|\mathcal{X},\mathcal{Y})\big|\Big), \end{aligned}$$

for $z < \widetilde{R}$. We then separately bound $L_{z,1}$ and $L_{z,2}$. For $L_{z,1}$, by Lemma A.4 and Assumption **(A)**, we have

$$\text{(C.15)} \qquad L_{z,1} \le C\varepsilon'\sqrt{\log(m)} = Cs_0^2(qn)^{-1}\sqrt{\log(qn)} = o(1).$$

Considering $\mathcal{V}_{s_0} := \{\mathbf{v}\in\mathbb{S}^{q-1} : \|\mathbf{v}\|_0 \le s_0\}$, we have

$$\begin{aligned} \sup_{\mathbf{v}_1,\mathbf{v}_2\in\mathcal{V}_{s_0}} |\mathbf{v}_1^\top(\widehat{\mathbf{R}}_{12} - \mathbf{R}_{12})\mathbf{v}_2| &\le \|\widehat{\mathbf{R}}_{12} - \mathbf{R}_{12}\|_\infty\|\mathbf{v}_1\|_1\|\mathbf{v}_2\|_1 \\ &\le s_0\|\widehat{\mathbf{R}}_{12} - \mathbf{R}_{12}\|_\infty. \end{aligned}$$

Therefore, combining Theorem 4.1 and Remark 4.1 in [7], by Lemma C.2, with probability at least $1 - C_1 n^{-1}$, we have

$$L_{z,2} \le C\Big(s_0 \frac{\log^{3/2}(qn)}{\sqrt{n}}\Big)^{1/3} \log^{2/3}(mn) \le C\Big(\frac{s_0^{10}\log^7(qn)}{n}\Big)^{1/6}. \tag{C.16}$$

From Assumption **(A)**, we have $L_{z,2} = o(1)$, which finishes the proof. □

C.4. Proof of Lemma B.4.

PROOF. We first prove for $m > 1$. By the definition of N_s^1 in (B.28), we have

$$N_s^1 = \underbrace{\frac{\widehat{u}_{1,s} - \widehat{u}_{2,s} - u_{1,s} + u_{2,s}}{\sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}}}_{\widetilde{N}_s^1} \cdot \frac{\sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}}{\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}}. \tag{C.17}$$

By Lemma A.6 and Lemma C.1, for sufficiently large n with probability at least $1 - C_1 n^{-1}$ we have

$$\left|1 - \frac{\sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}}{\sqrt{\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2}}\right| \le C\frac{\log^{3/2}(qn)}{\sqrt{n}}. \tag{C.18}$$

By setting

$$\mathcal{E}(z) = \big\{\big|1 - (\widehat{v}_{1,s}/n_1 + \widehat{v}_{2,s}/n_2)^{-1/2}(m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2)^{1/2}\big| \le z\big\}$$

and $z \asymp \log^{3/2}(qn)/\sqrt{n}$, we can bound $\mathbb{P}(|N_s^1| \ge x)$ by

$$\mathbb{P}(|N_s^1| \ge x) \le \mathbb{P}\big(|N_s^1| \ge x, \mathcal{E}(z)\big) + Cn^{-1}. \tag{C.19}$$

By the definition of $\mathcal{E}(z)$ and (C.17), we then have

$$\mathbb{P}\big(|N_s^1| \ge x, \mathcal{E}(z)\big) \le \mathbb{P}\bigg(\frac{|\widehat{u}_{1,s} - \widehat{u}_{2,s} - u_{1,s} + u_{2,s}|}{\sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}} \ge (1+z)^{-1}x\bigg)$$

Considering $z = o(1)$ and $x = u\sqrt{2\log q}$ in Lemma B.4, to prove (B.38), we only need to prove that as $n, q \to \infty$, we have

$$\underbrace{s_0\,\mathbb{P}\bigg(\frac{|\widehat{u}_{1,s} - \widehat{u}_{2,s} - u_{1,s} + u_{2,s}|}{\sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}} \ge C\sqrt{\log q}\bigg)}_{A_1} \to 0,$$

uniformly for s. By triangle and Hoeffding's inequalities, we have

$$s_0 A_1 \le s_0 \underbrace{\mathbb{P}\Big(\frac{|\frac{1}{n_1}\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k) - \frac{1}{n_2}\sum_{k=1}^{n_2} h_s(\boldsymbol{Y}_k|}{\sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}} \ge \frac{C}{2}\sqrt{\log q}\Big)}_{A_2} + s_0 \underbrace{\mathbb{P}\Big(\frac{\big|\binom{n_1}{m}^{-1}\Delta_{n_1,s} - \binom{n_2}{m}^{-1}\Delta_{n_2,s}\big|}{\sqrt{m^2\sigma_{1,ss}/n_1 + m^2\sigma_{2,ss}/n_2}} \ge \frac{C}{2}\sqrt{\log q}\Big)}_{A_3}.$$

By the exponential inequality for sub-exponential distribution, considering Assumption $\mathbf{(A)}'$ we have $s_0 A_2 \le s_0 \exp(-C\log^{1/2}(q)) \to 0$. As A_3 does not exist for $m = 1$, we only need to deal with $m > 1$. Similarly to (C.3), we threshold the kernel of $\widehat{u}_{\gamma,s} - u_{\gamma,s}$ by $B_n = C\log(q)$ and construct the threshold residual $\widehat{\Delta}_{n_\gamma,s}$. Similarly to the proof of bounding $|\binom{n_1}{m}^{-1}\Delta_{n_1,s} - \binom{n_2}{m}^{-1}\Delta_{n_2,s}|$ in Lemma B.1, by setting

$$A_{3,1} = \mathbb{P}\Big(\Big|\binom{n_1}{m}^{-1}\widehat{\Delta}_{n_1,s} - \binom{n_2}{m}^{-1}\widehat{\Delta}_{n_2,s}\Big| > C\sqrt{\frac{\log q}{n}}\Big)$$
$$A_{3,2} = \max_{\substack{i_\ell, j_\ell \\ \ell=1,\ldots,m}} \big(\mathbb{P}(|\Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})| > B_n) + \mathbb{P}(|\Psi_s(\boldsymbol{Y}_{j_1},\ldots,\boldsymbol{Y}_{j_m})| > B_n)\big),$$

For proving $s_0 A_3 \to 0$, we only need to prove $s_0 A_{3,1} \to 0$, $s_0 n^m A_{3,2} \to 0$, and $|E_{1,s}| + |E_{2,s}| \prec \sqrt{\log q/n}$, where $E_{\gamma,s}$ is defined in (C.3). By by Proposition 2.3 (c) in [1], under Assumption $\mathbf{(A)}'$, we have

$$s_0 A_{3,1} \le C_1 s_0 \exp\Big(-(n^{\frac{m-1}{2}}\log^{-\frac{1}{2}}(q))^{\frac{2}{m}}\Big) \to 0.$$

As $\Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})|$ and $\Psi_s(\boldsymbol{X}_{i_1},\ldots,\boldsymbol{X}_{i_m})|$ have sub-exponential tails from Assumption $\mathbf{(E)}$, similarly to the proof in Lemma B.1, under Assumption $\mathbf{(A)}'$ we have $s_0 n^m A_{3,2} \to 0$, and $|E_{1,s}| + |E_{2,s}| \prec \sqrt{\log q/n}$, which finishes the proof.

□

C.5. Proof of Lemma B.5.

Proof. In Lemma B.5, we aim to prove (B.41). For this, we need to bound

$$\sup_{z\in[\epsilon,1-\epsilon]} \Big|1 - \widetilde{F}_{N,\mathrm{ad}}(z) - \mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})\Big|.$$

By the definition of N_{ad} in (3.16), we have $N_{\mathrm{ad}} = \min_{p\in\mathcal{P}} \widehat{P}^N_{(s_0,p)}$, where $\mathcal{P}$ is a finite set. Therefore, without loss of generality, we assume $\mathcal{P} = \{p_1, p_2\}$

with $1 \le p_1 \ne p_2 \le \infty$. We then have $N_{\mathrm{ad}} = \min\big(\widehat{P}^N_{(s_0,p_1)}, \widehat{P}^N_{(s_0,p_2)}\big)$. We then have $\mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y}) = \mathbb{P}\Big(\big\{\widehat{P}^N_{(s_0,p_1)} > z\big\} \cap \big\{\widehat{P}^N_{(s_0,p_2)} > z\big\}\Big|\mathcal{X},\mathcal{Y}\Big)$. In (B.19) and (B.21), we introduce $F_{N^b,(s_0,p_\ell)}(z)$ and $\widehat{F}_{N^b,(s_0,p_\ell)}(z)$ as

$$\text{(C.20)} \qquad \begin{aligned} F_{N^b,(s_0,p_\ell)}(z) &= \mathbb{P}(\|\boldsymbol{N}^b\|_{(s_0,p_\ell)} \le z|\mathcal{X},\mathcal{Y}), \\ \widehat{F}_{N^b,(s_0,p_\ell)}(z) &= \frac{\sum_{b=1}^B \mathbb{I}\left\{N^b_{(s_0,p_\ell)} \le z|\mathcal{X},\mathcal{Y}\right\} + 1}{B+1}, \end{aligned}$$

for $\ell = 1, 2$. By the definition of $\widehat{P}^N_{(s_0,p)}$ in (2.8), we then have $\widehat{P}^N_{(s_0,p_\ell)} = 1 - \widehat{F}_{N^b,(s_0,p_\ell)}\big(N_{(s_0,p_\ell)}\big)$. Therefore, we can rewrite $\mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})$ as

$$\text{(C.21)} \quad \mathbb{P}\Big(\widehat{F}_{N^b,(s_0,p_1)}(N_{(s_0,p_1)}) < 1-z, \widehat{F}_{N^b,(s_0,p_2)}(N_{(s_0,p_2)}) < 1-z|\mathcal{X},\mathcal{Y}\Big).$$

Similarly, by setting $F_{N,(s_0,p_\ell)}(z) = \mathbb{P}\big(N_{(s_0,p_\ell)} \le z\big)$ we can also rewrite $1 - \widetilde{F}_{N,\mathrm{ad}}(z)$ as

$$\text{(C.22)} \qquad \mathbb{P}\Big(F_{N,(s_0,p_1)}\big(N_{(s_0,p_1)}\big) < 1-z, F_{N,(s_0,p_2)}\big(N_{(s_0,p_2)}\big) < 1-z\Big).$$

Combining (C.21) and (C.22), by setting

$$\begin{aligned} D_1(z) &= \mathbb{P}\Big(F_{N,(s_0,p_1)}\big(N_{(s_0,p_1)}\big) < 1-z, F_{N,(s_0,p_2)}\big(N_{(s_0,p_2)}\big) < 1-z\Big), \\ D_2(z) &= \mathbb{P}\Big(\widehat{F}_{N^b,(s_0,p_1)}(N_{(s_0,p_1)}) < 1-z, \widehat{F}_{N^b,(s_0,p_2)}(N_{(s_0,p_2)}) < 1-z\Big), \end{aligned}$$

we have $\Big|1 - \widetilde{F}_{N,\mathrm{ad}}(z) - \mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})\Big| = \Big|D_1(z) - D_2(z)\Big|$. By Glivenko-Cantelli Theorem, we have $\lim_{B\to\infty} \sup_{z\in\mathbb{R}} |\widehat{F}_{N^b,(s_0,p_\ell)}(z) - F_{N^b,(s_0,p_\ell)}(z)| = 0$ almost surely, which motives us to introduce

$$D_3(z) = \mathbb{P}\Big(F_{N^b,(s_0,p_1)}(N_{(s_0,p_1)}) < 1-z, F_{N^b,(s_0,p_2)}(N_{(s_0,p_2)}) < 1-z\Big).$$

We then use the triangle inequality to bound $\Big|1 - \widetilde{F}_{N,\mathrm{ad}}(z) - \mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})\Big|$ by

$$\text{(C.23)} \quad \Big|1 - \widetilde{F}_{N,\mathrm{ad}}(z) - \mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})\Big| \le \big|D_1(z) - D_3(z)\big| + \big|D_3(z) - D_2(z)\big|.$$

By (C.23), to prove (B.41), it is sufficient to prove that as $n, B \to \infty$, we have

$$\text{(C.24)} \qquad \sup_{z\in[\epsilon,1-\epsilon]} |D_1(z) - D_3(z)| \to 0 \text{ and } \sup_{z\in[\epsilon,1-\epsilon]} |D_3(z) - D_2(z)| \to 0,$$

for any fixed $\epsilon > 0$.

By Lemma 5 in [3], we can prove

$$\lim_{B\to\infty} \sup_{z\in[\epsilon,1-\epsilon]} |D_3(z) - D_2(z)| = 0. \tag{C.25}$$

Hence, we only need to prove $\lim_{n\to\infty} \sup_{z\in[\epsilon,1-\epsilon]} |D_1(z) - D_3(z)| = 0$. For this, we introduce the following lemma.

Lemma C.3. Assumptions **(A)**$''$, **(E)**, **(M1)**, and **(M2)** hold. Under $\mathbf{H}_0$ of (1.8) for any $\epsilon > 0$ we have

$$\sup_{z\in[\epsilon,1-\epsilon]} |D_1(z) - D_3(z)| \to 0, \qquad \text{as } n \to \infty.$$

The proof of Lemma C.3 is in Appendix D.2 of supplementary materials. Combining (C.25) and Lemma C.3, we prove (C.24), which finishes the proof of Lemma B.5. □

C.6. Proof of Lemma B.6.

Proof. In Lemma B.6, we aim to prove (B.52). We set

$$F_{N,\mathrm{ad}}(z) = \mathbb{P}(N_{\mathrm{ad}} \le z|\mathcal{X},\mathcal{Y}) \quad \text{and} \quad F_{N^b,\mathrm{ad}}(z) = \mathbb{P}(N^b_{\mathrm{ad}} \le z|\mathcal{X},\mathcal{Y}),$$

where N_{ad} and N^b_{ad} are defined in (2.10) and (B.48). Hence, to prove (B.52), it is sufficient to prove

$$\sup_{z\in[\epsilon,1-\epsilon]} \Big|\mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y}) - \mathbb{P}(N^b_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})\Big| \to 0 \text{ as } n, B \to \infty. \tag{C.26}$$

Without loss of generality, we assume $\mathcal{P} = \{p_1, p_2\}$ with $1 \le p_1 \ne p_2 \le \infty$. We can then rewrite $\mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})$ as

$$\mathbb{P}\Big(\widehat{F}_{N^b,(s_0,p_1)}(N_{(s_0,p_1)}) < 1 - z, \widehat{F}_{N^b,(s_0,p_2)}(N_{(s_0,p_2)}) < 1 - z|\mathcal{X},\mathcal{Y}\Big), \tag{C.27}$$

where $\widehat{F}_{N^b,(s_0,p_\ell)}(z)$ is defined in (C.20). Similarly, we can rewrite $\mathbb{P}(N^b_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})$

$$\mathbb{P}\Big(F_{N^b,(s_0,p_1)}(N^b_{(s_0,p_1)}) < 1 - z, F_{N^b,(s_0,p_2)}(N^b_{(s_0,p_2)}) < 1 - z|\mathcal{X},\mathcal{Y}\Big), \tag{C.28}$$

where $F_{N^b,(s_0,p_\ell)}(z)$ is defined in (B.48). Let

$$L = \mathbb{P}\Big(F_{N^b,(s_0,p_1)}(N_{(s_0,p_1)}) < 1 - z, F_{N^b,(s_0,p_2)}(N_{(s_0,p_2)}) < 1 - z|\mathcal{X},\mathcal{Y}\Big).$$

By Massart's inequality (see Section 1.5 in [10]) and Lemma 5 in [11], under Assumptions **(A)**$''$, **(E)**, **(M1)**, and **(M2)**, for any fix $\epsilon > 0$, we have

$$\sup_{z\in[\epsilon,1-\epsilon]} \Big|\mathbb{P}(N_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y}) - L\Big| \to 0 \qquad \text{as } n,\, B\to\infty. \tag{C.29}$$

Similarly to the proof of Theorems 3.1, considering (C.28), we also have

$$\sup_{z\in[0,1]} \Big|\mathbb{P}(N^b_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y}) - L\Big| \to 0, \qquad \text{as } n,\, B\to\infty. \tag{C.30}$$

Combining (C.29) and (C.30), we use the triangle inequality to obtain (C.26), which finishes the proof of Lemma B.6.

□

C.7. Proof of Lemma B.7.

Proof. In Lemma B.7, we aim to prove (B.53). By the definitions of $F_{N^b,\mathrm{ad}}(z)$ and $\widehat{F}_{N,\mathrm{ad}'}(z)$ in (B.49) and (B.50), we have

$$\begin{aligned} 1 - F_{N^b,\mathrm{ad}}(z) &= \mathbb{P}(N^b_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y}) \\ 1 - \widehat{F}_{N,\mathrm{ad}'}(z) &= \sum_{b=1}^{B} \mathbb{I}\{N^b_{ad'} > z|\mathcal{X},\mathcal{Y}\}/(B+1), \end{aligned} \tag{C.31}$$

where N^b_{ad} and $N^b_{\mathrm{ad}'}$ are defined in (B.48) and (2.12). Therefore, for (B.53) it is sufficient to prove

$$\sup_{z\in[\epsilon,1-\epsilon]} \left|\mathbb{P}\Big(N^b_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y}\Big) - \Big(1-\widehat{F}_{N,\mathrm{ad}'}(z)\Big)\right| \to 0, \tag{C.32}$$

as $n, B \to \infty$. Without loss of generality, we assume $\mathcal{P} = \{p_1, p_2\}$ with $1 \le p_1 \ne p_2 \le \infty$, which yields

$$N^b_{\mathrm{ad}'} = \min\Big(\widehat{P}^{b,N}_{(s_0,p_1)}, \widehat{P}^{b,N}_{(s_0,p_2)}\Big), \tag{C.33}$$

where $\widehat{P}^{b,N}_{(s_0,p)}$ is defined in (2.12). Combining (C.31) and (C.33), we then have

$$1 - \widehat{F}_{N,\mathrm{ad}'}(z) = \frac{\sum_{b=1}^{B} \mathbb{I}\{\widehat{P}^{b,N}_{(s_0,p_1)} > z, \widehat{P}^{b,N}_{(s_0,p_2)} > z|\mathcal{X},\mathcal{Y}\}}{B+1}. \tag{C.34}$$

By setting $\widehat{F}^{b,N}_{(s_0,p)}(z) = B^{-1}\Big(\sum_{b_1\neq b}\mathbb{I}\{N^{b_1}_{(s_0,p)} \le z|\mathcal{X},\mathcal{Y}\}+1\Big)$, considering the definition of $\widehat{P}^{b,N}_{(s_0,p)}$ in (2.12), we have $\widehat{P}^{b,N}_{(s_0,p)} = 1-\widehat{F}^{b,N}_{(s_0,p)}(N^b_{(s_0,p)})$. Therefore, by (C.34), we rewrite $1-\widehat{F}_{N,\mathrm{ad}'}(z)$ as

$$\text{(C.35)}\quad \frac{\sum_{b=1}^{B}\mathbb{I}\left\{\widehat{F}^{b,N}_{(s_0,p_1)}(N^b_{(s_0,p_1)}) < 1-z, \widehat{F}^{b,N}_{(s_0,p_2)}(N^b_{(s_0,p_2)}) < 1-z|\mathcal{X},\mathcal{Y}\right\}}{B+1},$$

As $\widehat{F}^{b,N}_{(s_0,p)}(z) \to F_{N^b,(s_0,p)}(z)$, to approximate $1-\widehat{F}_{N,\mathrm{ad}'}(z)$ we introduce $S(z)$ as

$$\text{(C.36)}\quad \frac{\sum_{b=1}^{B}\mathbb{I}\left\{F_{N^b,(s_0,p_1)}(N^b_{(s_0,p_1)})<1-z, F_{N^b,(s_0,p_2)}(N^b_{(s_0,p_2)})<1-z|\mathcal{X},\mathcal{Y}\right\}}{B+1},$$

where $F_{N^b,(s_0,p_\ell)}$ is defined in (B.48). To analyze the difference between $1-\widehat{F}_{N,\mathrm{ad}'}(z)$ and $S(z)$, we introduce the following lemma.

LEMMA C.4. Let ϵ be any positive real number, we have

$$\sup_{z\in[\epsilon,1-\epsilon]}\left|1-\widehat{F}_{N,\mathrm{ad}'}(z)-S(z)\right| = 0, \quad \text{as } n,\ B\to\infty.$$

The proof of Lemma C.4 is in Appendix D.3. Considering $N^b_{\mathrm{ad}} = \min_{p\in\mathcal{P}}\Big(1-F_{N^b,(s_0,p)}(N^b_{(s_0,p)})\Big)$ from (B.48), we can rewrite $S(z)$ as

$$S(z) = (B+1)^{-1}\sum_{b=1}^{B}\mathbb{I}\{N^b_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y}\}.$$

By Massart's inequality (see Section 1.5 in [10]), we have

$$\text{(C.37)}\qquad \sup_{z\in[0,1]}|S(z)-\mathbb{P}(N^b_{\mathrm{ad}} > z|\mathcal{X},\mathcal{Y})| \to 0,$$

as $n, B\to\infty$. Combining Lemma C.4 and (C.37), we plug in $S(z)$ and use the triangle inequality to obtain (C.32), which finishes the proof of Lemma B.7. □

APPENDIX D: PROOFS OF LEMMAS IN APPENDIX C

D.1. Proof of Lemma C.1.

PROOF. Under the event $\{\max_{1\le s\le q}|1-\xi_s|\le y/2\}$, we have $|1-\xi_s|\le y/2$ for any $s\in\{1,\dots,q\}$. Considering $y\in(0,1]$, by the simple calculation, $|1-\xi_s|\le y/2$ implies

$$|1-\xi_s^{-1}|\le \max\left(\frac{y}{2+y},\frac{y}{2-y}\right)\le y,$$

for any $s\in\{1,\dots,q\}$. Therefore, we have

$$\Big\{\max_{1\le s\le q}|1-\xi_s|\le y/2\Big\}\subseteq\Big\{\max_{1\le s\le q}|1-\xi_s^{-1}|\le y\Big\},$$

which implies (C.8). Hence, we finish the proof of Lemma C.1. □

D.2. Proof of Lemma C.3.

PROOF. Without loss of generality, we assume $\mathcal{P}=\{p_1,p_2\}$ with $1\le p_1\ne p_2\le\infty$. We set

$$\begin{aligned}D_1(z)&=\mathbb{P}\Big(F_{N,(s_0,p_1)}\big(N_{(s_0,p_1)}\big)<1-z,F_{N,(s_0,p_2)}\big(N_{(s_0,p_2)}\big)<1-z\Big),\\ D_3(z)&=\mathbb{P}\Big(F_{N^b,(s_0,p_2)}(N_{(s_0,p_1)})<1-z,F_{N^b,(s_0,p_2)}(N_{(s_0,p_2)})<1-z\Big),\end{aligned}\tag{D.1}$$

where $F_{N,(s_0,p_\ell)}(z)$ and $F_{N^b,(s_0,p_\ell)}(z)$ are defined in (B.19) and (B.21). In Lemma C.3, we aim to prove

$$\lim_{n\to\infty}\sup_{z\in[\epsilon,1-\epsilon]}|D_1(z)-D_3(z)|=0.\tag{D.2}$$

By following the proof of Theorem 3.1, under Assumptions **(A)**$''$, **(E)**, **(M1)**, and **(M2)**, by setting

$$\begin{aligned}F_{N,12}(z_1,z_2)&=\mathbb{P}\big(N_{(s_0,p_1)}\le z_1,N_{(s_0,p_2)}\le z_2\big),\\ F_{G,12}(z_1,z_2)&=\mathbb{P}\big(\|\boldsymbol{G}^N\|_{(s_0,p_1)}\le z_1,\|\boldsymbol{G}^N\|_{(s_0,p_2)}\le z_2\big)\Big|,\end{aligned}$$

with $\boldsymbol{G}^N\sim N(\mathbf{0},\mathbf{R}_{12})$ with $\mathbf{R}_{12}$ defined in (B.9), we have

$$\sup_{z_1,z_2\in(0,\infty)}\Big|F_{N,12}(z_1,z_2)-F_{G,12}(z_1,z_2)\Big|\to 0,\ \text{ as } n\to\infty.\tag{D.3}$$

(D.3) motives us to introduce

$$\begin{aligned}D_4(z)&=\mathbb{P}\Big(F_{N,(s_0,p_1)}\big(\|\boldsymbol{G}^N\|_{(s_0,p_1)}\big)<1-z,F_{N,(s_0,p_2)}\big(\|\boldsymbol{G}^N\|_{(s_0,p_2)}\big)<1-z\Big),\\ D_5(z)&=\mathbb{P}\Big(F_{N^b,(s_0,p_1)}\big(\|\boldsymbol{G}^N\|_{(s_0,p_1)}\big)<1-z,F_{N^b,(s_0,p_2)}\big(\|\boldsymbol{G}^N\|_{(s_0,p_2)}\big)<1-z\Big).\end{aligned}$$

Combining (D.1) and (D.3), we then have

$$\sup_{z\in(0,1)} |D_1(z) - D_4(z)| \to 0 \qquad \text{and} \qquad \sup_{z\in(0,1)} |D_3(z) - D_5(z)| \to 0,$$

as $n \to \infty$. Therefore, by using the triangle inequality, to prove (D.2) we only need to prove

$$\sup_{z\in[\epsilon,1-\epsilon]} |D_4(z) - D_5(z)| \to 0, \qquad \text{as } n \to \infty. \tag{D.4}$$

By Assumption **(A)**$''$, considering Theorems 3.1, for any $\epsilon > 0$ and sufficiently large n, we have

$$\sup_{z\in[\epsilon,1-\epsilon]} \left| F^{-}_{N,(s_0,p)}(z) - F^{-}_{N^b,(s_0,p)}(z) \right| \le h_{q,N}(\epsilon) \sup_{t\in\mathbb{R}} |F_{N,(s_0,p)}(t) - F_{N^b,(s_0,p)}(t)|.$$

Moreover, by the proof of Lemma B.2 we have

$$\sup_{z\in[\epsilon,1-\epsilon]} |D_4(z) - D_5(z)| \le C h_{q,N}(\epsilon) s_0 \sqrt{\log(nq)} \sup_{t\in\mathbb{R}} \left| F_{N,(s_0,p)}(t) - F_{N^b,(s_0,p)}(t) \right|,$$

for sufficiently large n. By the proof of Lemma A.1, B.2, B.3 and Theorem 3.1, under Assumption **(A)**$''$, **(E)**, **(M1)**, and **(M2)**, we have

$$\sup_{z\in[\epsilon,1-\epsilon]} |D_4(z) - D_5(z)| \le C h_{q,N}(\epsilon) s_0 \sqrt{\log(qn)} \Big(\frac{s_0^{14} \log^7(qn)}{n} \Big)^{1/6}.$$

In Assumption **(A)**$''$, we set $h^{0.6}_{q,N}(\epsilon) s_0^2 \log(qn) = o(n^{1/10})$. Therefore, we have

$$\sup_{z\in[\epsilon,1-\epsilon]} |D_4(z) - D_5(z)| \to 0, \qquad \text{as } n \to \infty,$$

which finishes the proof.

□

D.3. Proof of Lemma C.4.

Proof. In Lemma C.4, we aim to prove $\sup_{z\in[\epsilon,1-\epsilon]} \big|1 - \widehat{F}_{N,\mathrm{ad}'}(z) - S(z)\big| \to 0$, as $n, B \to \infty$. For this, we need to prove that for any $\delta, \widetilde{\varepsilon} > 0$

$$\mathbb{P}\left(\sup_{z\in[\epsilon,1-\epsilon]} \big|1 - \widehat{F}_{N,\mathrm{ad}'}(z) - S(z)\big| > \delta \right) < \widetilde{\varepsilon}, \tag{D.5}$$

holds for sufficient large n and B. By setting

$$\widehat{F}^{b,N}_{(s_0,p)}(z) = B^{-1}\big(\sum_{b_1\neq b} \mathbb{I}\{N^{b_1}_{(s_0,p)} \le z|\mathcal{X},\mathcal{Y}\} + 1\big),$$

considering Massart's inequality (Section 1.5 in [10]), we have

$$\sup_{\substack{1\le b\le B\\ z\in\mathbb{R}}} \Big|\widehat{F}^{b,N}_{(s_0,p)}(z) - F_{N^b,(s_0,p)}(z)\Big| \to 0, \qquad \text{as } n,\ B\to\infty.$$

Considering Lemma 5 in [3], for any fixed $\epsilon, \delta' > 0$, by setting

$$\mathcal{A}(\delta') = \left\{ \sup_{\substack{1\le b\le B\\ z\in[\epsilon,1-\epsilon]}} \Big|\widehat{F}^{b,N-}_{(s_0,p)}(z) - F^{-}_{N^b,(s_0,p)}(z)\Big| \le \delta' \right\},$$

as n and B are sufficiently large, we have $\mathbb{P}(\mathcal{A}(\delta')^c) \le \widetilde{\varepsilon}/2$. Therefore, considering that $F^{-}_{N^b,(s_0,p)}(z)$ is Lipschitz continuous on $z\in[\epsilon, 1-\epsilon]$, by the definitions of $1-\widehat{F}_{N,\mathrm{ad}'}$ and $S(z)$ in (C.35) and (C.36), under $\mathcal{A}(\delta')$ there is a constant C such that $S(z+C\delta') \le 1-\widehat{F}_{N,\mathrm{ad}'}(z) \le S(z-C\delta')$ holds for any $z\in[\epsilon,1-\epsilon]$ and sufficiently large n and B. Hence, under $\mathcal{A}(\delta')$ we have $\sup_{z\in[\epsilon,1-\epsilon]}\Big|1-\widehat{F}_{N,\mathrm{ad}'}(z)-S(z)\Big| \le \mathcal{L}$ with

$$\mathcal{L} = \max\left(\sup_{z\in[\epsilon,1-\epsilon]}\Big|S(z+C\delta')-S(z)\Big|, \sup_{z\in[\epsilon,1-\epsilon]}\Big|S(z-C\delta')-S(z)\Big|\right). \tag{D.6}$$

Therefore, to prove (D.5), we only need to prove

$$\mathbb{P}\big(\mathcal{L} > \delta, \mathcal{A}(\delta')\big) \le \widetilde{\varepsilon}/2, \tag{D.7}$$

for sufficiently large n and B. By Massart's inequality (Section 1.5 in [10]) and the definition of $S(z)$ in (C.36), we have

$$\sup_{z\in[0,1]} |S(z) - F_{N^b,\mathrm{ad}}(z)| \to 0, \qquad \text{as } n,\ B\to\infty, \tag{D.8}$$

where $F_{N^b,\mathrm{ad}}(z)$ is defined in (B.49). By (D.8), the limit of $\mathcal{L}$ is

$$\max\Big(\sup_{z\in[\epsilon,1-\epsilon]}\Big|F_{N^b,\mathrm{ad}}(z+C\delta')-F_{N^b,\mathrm{ad}}(z)\Big|, \sup_{z\in[\epsilon,1-\epsilon]}\Big|F_{N^b,\mathrm{ad}}(z-C\delta')-F_{N^b,\mathrm{ad}}(z)\Big|\Big).$$

As $F_{N^b,\mathrm{ad}}(z)$ is uniformly Lipschitz contentious on $[\epsilon, 1-\epsilon]$, there is a constant C_1 such that

$$0 \le \mathcal{L} \le C_1\delta', \tag{D.9}$$

holds for sufficiently large n and B. By setting δ' small enough and (D.9), we obtain (D.7), which finishes the proof of Lemma C.4. □

APPENDIX E: PROOF OF USEFUL LEMMAS IN APPENDIX A

E.1. Proof of Lemma A.1. By setting $\mathcal{E}^{R,d} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \le R\}$, from Assumption **(E)′**, we have

$$\mathbb{P}(S_n^{\boldsymbol{Z}} \in (\mathcal{E}^{R,d})^c) \vee \mathbb{P}(S_n^{\boldsymbol{W}} \in (\mathcal{E}^{R,d})^c) = C_1 d \exp(-C_2 R d^{-1/2}).$$

By setting $V_{(s_0,p)}^{z,d} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_{(s_0,p)} \le z\}$, we then have

$$\sup_z \Big| \mathbb{P}\big(S_n^{\boldsymbol{Z}} \in V_{(s_0,p)}^{z,d}\big) - P\big(S_n^{\boldsymbol{W}} \in V_{(s_0,p)}^{z,d}\big) \big| \le A_1 + A_2, \tag{E.1}$$

where $A_1 = C_1 d \exp(-C_2 R d^{-1/2})$ and $A_2 = \sup_z P_z$ with

$$P_z = |\mathbb{P}\big(S_n^{\boldsymbol{Z}} \in \mathcal{E}^{R,d} \cap V_{(s_0,p)}^{z,d}\big) - P\big(S_n^{\boldsymbol{W}} \in \mathcal{E}^{R,d} \cap V_{(s_0,p)}^{z,d}\big)|.$$

We then approximate $\mathcal{E}^{R,d} \cap V_{(s_0,p)}^{z,d}$ with m-generated convex set. According to Lemmas A.3 and A.4, by setting

$$\bar{\rho} = |\mathbb{P}(S_n^{\boldsymbol{Z}} \in A^m) - \mathbb{P}(S_n^{\boldsymbol{W}} \in A^m)| \vee |\mathbb{P}(S_n^{\boldsymbol{Z}} \in A^{m,R\epsilon}) - \mathbb{P}(S_n^{\boldsymbol{W}} \in A^{m,R\epsilon})|,$$

we have $P_z \le CR\epsilon \log^{1/2}(m) + \bar{\rho}$, where C only depends on b. By high dimensional CLT for Hyperreactangles in [7], we have

$$\bar{\rho} \le C\Big(\frac{\log^7(mn)}{n}\Big)^{1/6},$$

where C only depends on b. Considering (E.1), we then have

$$\begin{aligned}
&\sup_z \Big| \mathbb{P}\big(S_n^{\boldsymbol{Z}} \in V_{(s_0,p)}^{z}\big) - P\big(S_n^{\boldsymbol{W}} \in V_{(s_0,p)}^{z}\big) \big| \\
&\qquad \le CR\epsilon \log^{1/2}(m) + C\Big(\frac{\log^7(mn)}{n}\Big)^{1/6} + C_1 d \exp(-C_2 R d^{-1/2}).
\end{aligned}$$

By setting $\epsilon = (dn)^{-3/2}$ and $R = (dn)^{1/2}$, considering $s_0^2 \log(dn) = O(n^\zeta)$ with $0 < \zeta < 1/7$, we have

$$R\epsilon \log^{1/2}(m) \preceq \Big(\frac{\log^7(mn)}{n}\Big)^{1/6}, \ d\exp(-C_2 R d^{-1/2}) \preceq \Big(\frac{\log^7(mn)}{n}\Big)^{1/6},$$

which yields (A.1).

E.2. Proof of Lemma A.3.

Proof. By the definition of m-generated convex sets A^m and $A^{m,\epsilon}$, Lemma A.3 is an immediate corollary of Lemma A.2. □

E.3. Proof of Lemma A.5.

PROOF. By the Jensen's inequality, we have

$$\exp\Big(t\mathbb{E}\Big[\max_{1\le i\le d}|W_i|\Big]\Big) \le \mathbb{E}\Big[\exp\Big(t\max_{1\le i\le d}|W_i|\Big)\Big] \le d\mathbb{E}[\exp(t|W_i|)]. \tag{E.2}$$

By (23) of [12], we have

$$\mathbb{E}[\exp(t|W_i|)] = 2e^{\frac{\sigma^2 t^2}{2}}[1-\Phi(-\sigma t)] \le 2e^{\frac{\sigma^2 t^2}{2}}. \tag{E.3}$$

Combining (E.2) and (E.3), we have

$$\exp\Big(t\mathbb{E}\Big[\max_{1\le i\le d}|W_i|\Big]\Big) \le 2de^{\frac{\sigma^2 t^2}{2}},$$

which yields (A.2).

□

E.4. Proof of Lemma A.6.

PROOF. We first prove for $m>1$. For simplicity, we only present the proof for $\boldsymbol{X}$. In (3.4), we set $\boldsymbol{\Sigma}_1 = (\sigma_{1,st})$ with

$$\sigma_{1,st} = \mathbb{E}[h_s(\boldsymbol{X})h_t(\boldsymbol{X})], \tag{E.4}$$

where h_s is defined in (3.1). To estimate $\boldsymbol{\Sigma}_1$, in (3.9) we introduce $\widehat{\boldsymbol{\Sigma}}_1 := (\widehat{\sigma}_{1,st}) \in \mathbb{R}^{q\times q}$, where $\widehat{\sigma}_{1,st} = {n_1}^{-1}\sum_{k=1}^{n_1}(Q_{1k,s}-\widehat{u}_{1,s})(Q_{1k,t}-\widehat{u}_{1,t})$. By setting $\widetilde{u}_{1,s} = \widehat{u}_{1,s} - u_{1,s}$ and $\widetilde{Q}_{1k,s} = Q_{1k,s} - u_{1,s}$, we rewrite $\widehat{\sigma}_{1,st}$ as

$$\widehat{\sigma}_{1,st} = n_1^{-1}\sum_{k=1}^{n_1}\widetilde{Q}_{1k,s}\widetilde{Q}_{1k,t} - \widetilde{u}_{1,s}\widetilde{u}_{1,t}. \tag{E.5}$$

To provide an upper bound for $\max_{1\le s,t\le q}\left|\widehat{\sigma}_{1,st}-\sigma_{1,st}\right|$, by combining (E.4) and (E.5), we use the triangle inequality to obtain

$$\max_{1\le s,t\le q}\left|\widehat{\sigma}_{1,st}-\sigma_{1,st}\right| \le \underbrace{\max_{1\le s,t\le q}\left|n_1^{-1}\Big(\sum_{k=1}^{n_1}\widetilde{Q}_{1k,s}\widetilde{Q}_{1k,t}\Big) - \mathbb{E}[h_s(\boldsymbol{X})h_t(\boldsymbol{X})]\right|}_{L_1} + \underbrace{\max_{1\le s,t\le q}\left|\widetilde{u}_{1,s}\widetilde{u}_{1,t}\right|}_{L_2}.$$

We then bound L_1 and L_2 separately. For bounding L_2, we introduce

$$\widetilde{u}'_{1,s} = \binom{n_1}{m}^{-1} \sum_{1 \le i_1 < \ldots, < i_m \le n_1} V_{1,s}^{i_1,\ldots,i_m} - E_{1,s},$$

where $V_{1,s}^{i_1,\ldots,i_m}$ and $E_{1,s}$ are defined in (C.3) with threshold $B_n = C\log(qn)$. For any $\delta > 0$, by choosing proper C, we have $E_{1,s} \prec (qn)^{-\delta}$. We then have

$$|\widetilde{u}_{1,s} - \widetilde{u}'_{1,s}| \le \underbrace{\left|\widetilde{u}_{1,s} - \binom{n_1}{m}^{-1} \sum_{1 \le i_1 < \ldots, < i_m \le n_1} V_{1,s}^{i_1,\ldots,i_m}\right|}_{L_{2,s}} + E_{1,s}.$$

By setting $z \succ (qn)^{-\delta}$, we have

$$\max_{1 \le s \le q} \mathbb{P}(|\widetilde{u}_{1,s}| > z) \le \max_{1 \le s \le q} \Big(\mathbb{P}(|\widetilde{u}'_{1,s}| > z/3) + \mathbb{P}(L_{2,s} > z/3)\Big). \tag{E.6}$$

By using the exponential inequality for bounded U-statistics we have

$$\max_{1 \le s \le q} \mathbb{P}(|\widetilde{u}'_{1,s}| > z/3) \le C\exp(-C_1 n z^2 / B_n^2). \tag{E.7}$$

By Assumption **(E)**, we also have

$$\max_{1 \le s \le q} \mathbb{P}(L_{2,s} > z/3) \le C n_1^m \exp(-C_1 B_n) \tag{E.8}$$

Combining (E.6), (E.7), and (E.8), we then have

$$\begin{aligned} \mathbb{P}(L_2 > y) &\le q^2 \max_{1 \le s,t \le q} \mathbb{P}\big(|\widetilde{u}_{1,s}\widetilde{u}_{1,t}| > y\big) \le 2q^2 \max_{1 \le s \le q} \mathbb{P}\big(|\widetilde{u}_{1,s}| > \sqrt{y}\big) \\ &\le Cq^2\exp(-C_1 n y / B_n^2) + Cq^2 n_1^m \exp(-C_1 B_n). \end{aligned} \tag{E.9}$$

Therefore, for sufficiently large n_1 with probability $1 - Cn_1^{-1}$ we have $L_2 \le \log^3(qn)/n$.

We now bound L_1. Considering that $n_1^{-1}\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k)h_t(\boldsymbol{X}_k)$ approximates $\mathbb{E}[h_s(\boldsymbol{X}_k)h_t(\boldsymbol{X}_k)]$, we use triangle inequality again to bound L_1 by

$$\begin{aligned} L_1 \le\ & \underbrace{\max_{1 \le s,t \le q}\left|n_1^{-1}\big(\textstyle\sum_{k=1}^{n_1} \widetilde{Q}_{1k,s}\widetilde{Q}_{1k,t}\big) - n_1^{-1}\displaystyle\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k)h_t(\boldsymbol{X}_k)\right|}_{L_3} \\ & + \underbrace{\max_{1 \le s,t \le q}\left|n_1^{-1}\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k)h_t(\boldsymbol{X}_k) - \mathbb{E}[h_s(\boldsymbol{X})h_t(\boldsymbol{X})]\right|}_{L_4}. \end{aligned} \tag{E.10}$$

By Assumption **(E)**, $h_s(\boldsymbol{X})$ has sub-exponential tails. Therefore, by Theorem 6 in [9], we have

$$\mathbb{P}(L_4 > z) \le Cq^2 \exp(-C_1 n_1 z^2) + Cq^2 \exp\big(-C_2(n_1 z)^{1/2}\big). \tag{E.11}$$

Therefore, for sufficiently large n_1, with probability $1 - Cn_1^{-1}$, we have

$$L_4 \le C\sqrt{\frac{\log(qn)}{n}} + C_1\frac{\log^2(qn)}{n}.$$

After bounding L_4, we now deal with L_3. For this, we decompose $\widetilde{Q}_{1k,s}$ as

$$\widetilde{Q}_{1k,s} = \binom{n_1 - 1}{m-1}^{-1}\Big(Ah_s(\boldsymbol{X}_k) + BS_{1,s} + \Upsilon_{1,s}^{(k)}\Big), \tag{E.12}$$

with $A = \binom{n_1-1}{m_1-1} - \binom{n_1-2}{m-2}$, $B = \binom{n_1-1}{m-2}$, $S_{1,s} := \sum_{\beta=1}^{n_1} h_s(\boldsymbol{X}_\beta)$ and

$$\Upsilon_{1,s}^{(k)} = \sum_{\substack{1\le \ell_1<\ldots<\ell_{m-1}\le n_1 \\ \ell_j\neq k, j=1,\ldots,m-1}}^{n_1} \Gamma_{1,s}^{k,\ell_1\ell_2\ldots\ell_{m-1}}, \tag{E.13}$$

with $\Gamma_{1,s}^{k,\ell_1\ell_2\ldots\ell_{m-1}} = \Psi_s(\boldsymbol{X}_k, \boldsymbol{X}_{\ell_1}\ldots,\boldsymbol{X}_{\ell_m-1}) - \big(h_s(\boldsymbol{X}_k) + \sum_{i=1}^{m-1} h_s(\boldsymbol{X}_{\ell_i})\big)\Big)$. $\Psi_s(\boldsymbol{X}_{k_1},\ldots,\boldsymbol{X}_{k_m})$, the centralized version of $\Phi_s(\boldsymbol{X}_{k_1},\ldots,\boldsymbol{X}_{k_m})$, is defined in (3.1). For notational simplicity, by setting

$$V_{1,st}^2 := \sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k)h_t(\boldsymbol{X}_k),\ \Lambda_{1,s} := \sum_{k=1}^{n_1}\Upsilon_{1,s}^{(k)},\ \Lambda_{1,st}^2 := \sum_{k=1}^{n_1}\Upsilon_{1,s}^{(k)}\Upsilon_{1,t}^{(k)}, \tag{E.14}$$

and $D = \binom{n_1-1}{m-1}$. we have $L_3 = \max_{1\le s,t\le q} L_{3,st}$, where $L_{3,st}$ is defined as

$$\Big|\frac{1}{n_1}\Big(\frac{1}{D^2}\sum_{k=1}^{n_1}\big(Ah_s(\boldsymbol{X}_k) + BS_{1,s} + \Upsilon_{1,s}^{(k)}\big)\big(Ah_t(\boldsymbol{X}_k) + BS_{1,t} + \Upsilon_{1,t}^{(k)}\big)$$
$$-\sum_{k=1}^{n_1} h_s(\boldsymbol{X}_k)h_t(\boldsymbol{X}_k)\Big)\Big|,$$

After introducing these notations, we can expand $L_{3,st}$ as

$$\begin{aligned} L_{3,st} =& \Big|\frac{A^2 - D^2}{n_1 D^2}V_{1,st}^2 + \frac{1}{n_1 D^2}(2AB + n_1 B^2)S_{1,s}S_{1,t} + \frac{1}{n_1 D^2}\Lambda_{1,st}^2 \\ &+ \frac{A}{n_1 D^2}\sum_{k=1}^{n_1}(\Upsilon_{1,s}^{(k)}h_t(\boldsymbol{X}_k) + \Upsilon_{1,t}^{(k)}h_s(\boldsymbol{X}_k)) + \frac{B}{n_1 D^2}(\Lambda_{1,s}S_{1,t} + \Lambda_{1,t}S_{1,s})\Big|. \end{aligned}$$

By using the triangle inequality on $L_{3,st}$, we have $L_{3,st} \le J_{1,st} + J_{2,st} + J_{3,st} + J_{4,st} + J_{5,st}$, where

$$J_{1,st} := \Big|\frac{A^2 - D^2}{n_1 D^2} V_{1,st}^2\Big|,\ J_{2,st} := \Big|\frac{2AB + n_1 B^2}{n_1 D^2} S_{1,s} S_{1,t}\Big|,$$
$$J_{3,st} := \Big|\frac{1}{n_1 D^2}\Lambda_{1,st}^2\Big|, J_{4,st} := \Big|\frac{A}{n_1 D^2}\sum_{k=1}^{n_1}\big(\Upsilon_{1,s}^{(k)} h_t(\boldsymbol{X}_k) + \Upsilon_{1,t}^{(k)} h_s(\boldsymbol{X}_k)\big)\Big|,$$
$$J_{5,st} := \Big|\frac{B}{n_1 D^2}(\Lambda_{1,s} S_{1,t} + \Lambda_{1,t} S_{1,s})\Big|.$$

We now bound $J_{1,st}, \ldots, J_{5,st}$ separately. By the definitions of A and D, we obtain

$$A = O(n_1^{m-1}),\ D = O(n_1^{m-1}) \text{ and } D - A = \binom{n_1 - 2}{m - 2} = O(n_1^{m-2}).$$

Thus, for $J_{1,st}$, by the definition of $V_{1,st}$ in (E.14), by Assumption **(M2)** we easily have that $\max_{1\le s,t\le q} J_{1,st} = O_p(n_1^{-1})$. For $J_{2,st}$, considering $B = O(n_1^{m-2})$, we use the exponential inequality to have

$$\text{(E.15)} \qquad \mathbb{P}(J_{2,st} > y) = \mathbb{P}\Big(\frac{S_{1,s}S_{1,t}}{n_1^2} \ge Cy\Big) \le C_1 \exp(-C_2 n_1 \min(y, \sqrt{y})).$$

With probability $1 - Cn_1^{-1}$, we then have $\max_{1\le s,t\le q} J_{2,st} \le \log(qn_1)n_1^{-1}$ for sufficiently large n_1. We then bound $J_{3,st}$. Recalling $\Lambda_{1,st}^2 := \sum_{k=1}^{n_1}\Upsilon_{1,s}^{(k)}\Upsilon_{1,t}^{(k)}$ in (E.14), we have

$$\mathbb{P}(J_{3,st} > y) = \mathbb{P}\Big(\frac{\Lambda_{1,st}^2}{n_1^{2m-1}} \ge Cy\Big) = \mathbb{P}\Big(\sum_{k=1}^{n_1}\Upsilon_{1,s}^{(k)}\Upsilon_{1,t}^{(k)} \ge Cn_1^{2m-1}y\Big)$$
$$\le \sum_{k=1}^{n_1}\mathbb{P}\Big(\Upsilon_{1,s}^{(k)}\Upsilon_{1,t}^{(k)} \ge Cn_1^{2m-2}y\Big).$$

By the definition of $\Upsilon_{1,s}^{(k)}$ in (E.13), given $\boldsymbol{X}_k$, we can treat

$$\Psi_s(\boldsymbol{X}_k, \boldsymbol{X}_{\ell_1}\ldots, \boldsymbol{X}_{\ell_m-1}) - \big(h_{ij}(\boldsymbol{X}_k) + \sum_{r=1}^{m-1} h_{ij}(\boldsymbol{X}_{\ell_r})\big),$$

as a symmetric kernel function. Therefore, $\Upsilon_{1,s}^{(k)}/D|\boldsymbol{X}_k$ is a U-statistic with a kernel function of zero mean and $m-1$ order. Hence, similarly to L_2, we threshold the kernel with $C\log(qn_1)$ and use the exponential inequality for U-statistics to obtain that for sufficiently large n_1 with probability with $1-C_1n_1^{-1}$, we have

$$\max_{1\le s,t\le q} J_{3,st} \le C\frac{\log^2(qn)}{n}.$$

We now bound $J_{4,st}$ and $J_{5,st}$. For $J_{4,st}$, we use the Cauchy-Swartz inequality on $\sum_{k=1}^{n_1}\Upsilon_{1,s}^{(k)}h_t(\boldsymbol{X}_k)$ and $\sum_{k=1}^{n_1}\Upsilon_{1,t}^{(k)}h_s(\boldsymbol{X}_k)$ to obtain

$$J_{4,st} \le \Big|\frac{A}{n_1D^2}(\Lambda_{1,ss}V_{1,tt}+\Lambda_{1,tt}V_{1,ss})\Big|. \tag{E.16}$$

For $J_{5,st}$, by using the Cauchy-Swartz inequality on $\Lambda_{1,s}$ and $S_{1,s}$, we have

$$J_{5,st} \le \Big|\frac{B}{D^2}(\Lambda_{1,ss}V_{1,tt}+\Lambda_{1,tt}V_{1,ss})\Big|. \tag{E.17}$$

Combining (E.16) and (E.17), we have

$$J_{4,st}+J_{5,st} \le \underbrace{\Big|\frac{A+n_1B}{n_1D^2}(\Lambda_{1,ss}V_{1,tt}+\Lambda_{1,tt}V_{1,ss})\Big|}_{J_{6,st}}. \tag{E.18}$$

Considering $A=O(n_1^{m-1})$, $B=O(n_1^{m-2})$, and $D=O(n_1^{m-1})$, by the triangle inequality we have

$$\max_{1\le s,t\le q} J_{6,st} \le C\max_{1\le s,t\le q}\frac{\Lambda_{1,ss}V_{1,tt}}{n_1^m} = \Big(\max_{1\le s\le q}\underbrace{\frac{\Lambda_{1,ss}^2}{n_1^{2m-3/2}}}_{J'_{6,s}}\max_{1\le s\le q}\underbrace{\frac{V_{1,ss}^2}{n_1^{3/2}}}_{J''_{6,s}}\Big)^{1/2}.$$

Similarly to L_4, from Assumption **(M2)**, we have $\max_{1\le s\le q} J''_{6,s}=O_p(n_1^{-1/2})$. For $J'_{6,s}$, we have

$$\mathbb{P}\Big(\frac{\Lambda_{1,ss}^2}{n_1^{2m-3/2}}\ge y\Big) \le \sum_{k=1}^{n_1}\mathbb{P}\Big(\frac{|\Upsilon_{1,ss}^{(k)}|}{n_1^{m-1}}\ge Cn_1^{-1/4}y^{1/2}\Big). \tag{E.19}$$

By thresholding kernel with $C\log(qn)$ and exponential inequality for U-statistics, for sufficiently large n_1, $\max_{1\le s\le q} J'_{6,s} \le \log^3(qn_1)n_1^{-1/2}$ holds with probability $1-C_1n_1^{-1}$. Therefore, we have

$$\max_{1\le s,t\le q} J_{6,st} \le C\log^{3/2}(qn_1)n_1^{-1/2}.$$

From all above results, for sufficiently large n_1, with probability $1 - C_1 n_1^{-1}$, we have

$$\max_{1\leq ,s,t\leq q} |\widehat{\sigma}_{1,st} - \sigma_{1,st}| \leq C \frac{\log^{3/2}(qn_1)}{\sqrt{n_1}}. \tag{E.20}$$

After analyzing the approximation error of $\widehat{\sigma}_{1,st}$, we then prove for $\widehat{r}_{1,st}$. By (A.3), we have $\widehat{r}_{1,st} = \widehat{\sigma}_{1,st}/\sqrt{\widehat{\sigma}_{1,ss}\widehat{\sigma}_{1,tt}}$ and $r_{1,st} = \sigma_{1,st}/\sqrt{\sigma_{1,ss}\sigma_{1,tt}}$. Therefore, we have

$$\begin{aligned}
|\widehat{r}_{1,st} - r_{1,st}| &= \Big| \frac{\widehat{\sigma}_{1,st}}{\sqrt{\widehat{\sigma}_{1,ss}\widehat{\sigma}_{1,tt}}} - \frac{\sigma_{1,st}}{\sqrt{\sigma_{1,ss}\sigma_{1,tt}}} \Big| \\
&\leq \underbrace{\Big| \frac{\widehat{\sigma}_{1,st}}{\sqrt{\widehat{\sigma}_{1,ss}\widehat{\sigma}_{1,tt}}} - \frac{\widehat{\sigma}_{1,st}}{\sqrt{\sigma_{1,ss}\sigma_{1,tt}}} \Big|}_{A_1} + \underbrace{\Big| \frac{\widehat{\sigma}_{1,st}}{\sqrt{\sigma_{1,ss}\sigma_{1,tt}}} - \frac{\sigma_{1,st}}{\sqrt{\sigma_{1,ss}\sigma_{1,tt}}} \Big|}_{A_2}.
\end{aligned}$$

Hence, to bound $|\widehat{r}_{1,st} - r_{1,st}|$ we bound A_1 and A_2 separately. For A_1, we rewrite it as

$$A_1 = \Big| \frac{\widehat{\sigma}_{1,st}}{\sqrt{\widehat{\sigma}_{1,ss}\widehat{\sigma}_{1,tt}}} \Big| \Big| 1 - \frac{\sqrt{\widehat{\sigma}_{1,ss}\widehat{\sigma}_{1,tt}}}{\sqrt{\sigma_{1,ss}\sigma_{1,tt}}} \Big|.$$

Considering $|\widehat{r}_{1,st}| \leq 1$ and $a^2 - b^2 = (a+b)(a-b)$, we have

$$A_1 \leq \sigma_{1,ss}^{-1}\sigma_{1,tt}^{-1} \big| \widehat{\sigma}_{1,ss}\widehat{\sigma}_{1,tt} - \sigma_{1,ss}\sigma_{1,tt} \big|.$$

By Assumption **(M1)** and **(M2)**, there are constants b and B, such that $0 < b \leq \sigma_{1,ss} \leq B < \infty$ for $s = 1, \ldots, q$. Hence, we have

$$A_1 \leq b^{-2} \max_{1\leq s\leq q} |\widehat{\sigma}_{1,ss} - \sigma_{1,ss}|^2 + 2Bb^{-2} \max_{1\leq s\leq q} |\widehat{\sigma}_{1,ss} - \sigma_{1,ss}|. \tag{E.21}$$

For A_2, by $\sigma_{1,ss} \geq b > 0$ from Assumption **(M1)** we have

$$A_2 \leq b^{-1} \max_{1\leq s,t\leq q} |\widehat{\sigma}_{1,st} - \sigma_{1,st}|. \tag{E.22}$$

Combining (E.20), (E.21), and (E.22), we then have that

$$\max_{1\leq ,s,t\leq q} |\widehat{r}_{1,st} - r_{1,st}| \leq C \frac{\log^{3/2}(qn_1)}{\sqrt{n_1}},$$

holds with the overwhelming probability, which finishes the proof for $m > 1$.

We then prove for $m = 1$. We decompose $\widehat{\sigma}_{1,st}$ as

$$\widehat{\sigma}_{1,st} = n_1^{-1} \sum_{k=1}^{n_1} \Psi_s(\boldsymbol{X}_k)\Psi_t(\boldsymbol{X}_k) - \overline{\Psi}_{1,s}\overline{\Psi}_{1,t},$$

where $\Psi_s(\boldsymbol{X}_k) = \Phi_s(\boldsymbol{X}_k) - u_{1,s}$ and $\overline{\Psi}_{1,s} = n_1^{-1}\sum_{k=1}^{n_1}\Psi_s(\boldsymbol{X}_k)$. Considering $\sigma_{1,st} = \mathbb{E}[\Psi_s(\boldsymbol{X})\Psi_t(\boldsymbol{X})]$, by setting

$$B_1 = \mathbb{P}\Big(\max_{1\le s,t\le q}\big|{n_1}^{-1}\sum_{k=1}^{n_1}\Psi_s(\boldsymbol{X}_k)\Psi_t(\boldsymbol{X}_k) - \mathbb{E}[\Psi_s(\boldsymbol{X})\Psi_t(\boldsymbol{X})]\big| > x/2\Big)$$

$$B_2 = \mathbb{P}\Big(\max_{1\le s,t\le q}\overline{\Psi}_{1,s}\overline{\Psi}_{1,t} > x/2\Big)$$

we then have

$$\mathbb{P}\Big(\max_{1\le s,t\le q}\big|\widehat{\sigma}_{\gamma,st} - \sigma_{\gamma,st}\big| > x\Big) \le B_1 + B_2, \tag{E.23}$$

By Theorem 6 in [9], we can bound B_1 by

$$B_1 \le Cq^2\exp(-C_1 n_1 x^2) + Cq^2\exp\big(-C_2(n_1 x)^{1/2}\big). \tag{E.24}$$

Similarly, for the term B_2 in (E.23), we use the same argument to obtain

$$\mathbb{P}\Big(\max_{1\le s,t\le q}\overline{\Psi}_{1,s}\overline{\Psi}_{1,t} > x/2\Big) \le Cq^2\exp(-C_1 n_1 x) + Cq^2\exp\big(-C_2(n_1\sqrt{x})\big). \tag{E.25}$$

Combining (E.23), (E.24), and (E.25), for sufficiently large n_1, with probability $1 - C_1 n_1^{-1}$, we have

$$\max_{1\le, s,t\le q}|\widehat{\sigma}_{1,st} - \sigma_{1,st}| \le C\sqrt{\frac{\log(qn_1)}{n_1}} + C\frac{\log^2(qn_1)}{n_1}. \tag{E.26}$$

Similarly to $m > 1$, we also have that

$$\max_{1\le, s,t\le q}|\widehat{r}_{1,st} - r_{1,st}| \le C\sqrt{\frac{\log(qn_1)}{n_1}} + C\frac{\log^2(qn_1)}{n_1},$$

holds with the overwhelming probability for $m = 1$. □

APPENDIX F: MORE SIMULATION RESULTS

This section consists of three parts. Firstly, we present the empirical size for high dimensional mean tests based on **Models 2-4**, which are introduced in Section 4. Secondly, we apply our methods to test high dimensional covariance/correlation coefficients to illustrate the generality of proposed methods. At last, we apply our methods to analyze resting-state functional magnetic resonance imaging (fMRI) data.

Table 3
Empirical sizes of **Models 2, 3** *and* **4** *with* $\alpha = 0.05$, $B = 300$, *and* $n_1 = n_2 = 100$ *based on 2000 replications.*

						Model 2						
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	T^2	BY	SD	CLX
75	5	6.20	6.50	6.55	6.85	7.00	6.65	7.10	5.25	6.50	5.40	5.05
	30	4.30	4.75	5.35	6.00	6.35	6.75	6.25	5.25	6.50	5.40	5.05
	75	4.55	4.75	5.60	6.00	6.25	6.50	6.30	5.25	6.50	5.40	5.05
200	10	5.20	5.45	5.75	5.65	6.20	6.65	6.30	-	5.35	4.60	6.10
	50	3.30	3.40	3.80	4.50	5.30	6.25	5.30	-	5.35	4.60	6.10
	100	2.85	3.05	3.35	3.95	4.75	7.10	5.10	-	5.35	4.60	6.10
	150	3.00	3.10	3.55	4.50	5.10	7.00	5.50	-	5.35	4.60	6.10
	200	2.70	2.90	3.40	4.20	5.05	7.10	5.15	-	5.35	4.60	6.10
400	10	4.85	5.00	5.45	5.45	5.95	0.71	6.90	-	5.10	4.10	6.25
	50	1.90	2.15	2.60	3.30	3.90	7.40	5.45	-	5.10	4.10	6.25
	100	1.35	1.50	1.85	2.80	3.85	7.20	4.75	-	5.10	4.10	6.25
	200	1.05	1.15	1.70	2.65	3.70	7.00	4.45	-	5.10	4.10	6.25
	400	1.30	1.65	1.75	2.70	3.55	7.10	4.50	-	5.10	4.10	6.25
						Model 3						
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	T^2	BY	SD	CLX
75	5	5.25	5.65	6.25	6.15	6.30	6.90	6.75	5.30	6.10	5.40	5.90
	30	4.70	4.70	5.35	5.75	6.20	6.95	5.65	5.30	6.10	5.40	5.90
	75	4.25	4.80	5.05	5.10	5.75	7.00	5.75	5.30	6.10	5.40	5.90
200	10	3.75	4.05	4.65	5.20	5.35	7.05	5.85	-	5.70	4.90	5.50
	50	2.80	2.60	3.20	3.50	4.15	6.70	4.65	-	5.70	4.90	5.50
	100	2.45	2.50	2.75	3.50	4.35	6.60	4.20	-	5.70	4.90	5.50
	150	2.40	2.55	2.75	3.70	4.40	7.05	4.50	-	5.70	4.90	5.50
	200	2.15	2.30	2.75	3.60	4.35	6.70	4.65	-	5.70	4.90	5.50
400	10	3.95	4.30	4.80	4.85	5.30	7.35	6.05	-	5.25	3.95	6.25
	50	1.40	1.80	2.15	2.55	3.70	7.15	4.75	-	5.25	3.95	6.25
	100	1.10	1.20	1.65	2.25	3.05	7.05	4.45	-	5.25	3.95	6.25
	200	0.90	0.95	1.25	1.95	3.20	7.10	4.35	-	5.25	3.95	6.25
	400	0.95	0.75	1.30	2.10	3.20	7.15	3.80	-	5.25	3.95	6.25
						Model 4						
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	T^2	BY	SD	CLX
75	5	4.10	4.05	4.05	4.70	4.95	5.50	5.05	4.10	3.90	3.60	4.40
	30	3.05	3.00	3.20	3.55	3.90	5.15	5.00	4.10	3.90	3.60	4.40
	75	2.75	3.15	3.30	3.75	4.10	5.60	4.65	4.10	3.90	3.60	4.40
200	10	2.45	2.75	2.80	3.10	3.30	5.30	4.20	-	1.75	1.50	4.35
	50	1.05	1.05	1.30	1.75	2.35	5.50	3.30	-	1.75	1.50	4.35
	100	1.10	1.10	1.20	1.65	2.35	5.60	3.00	-	1.75	1.50	4.35
	150	0.85	0.90	1.10	1.45	2.25	5.65	3.35	-	1.75	1.50	4.35
	200	1.00	1.10	1.10	1.65	1.95	5.65	2.75	-	1.75	1.50	4.35
400	10	2.85	3.05	3.35	3.40	4.15	5.70	4.20	-	0.85	0.45	4.20
	50	0.95	0.95	1.05	1.30	1.80	5.65	3.20	-	0.85	0.45	4.20
	100	0.45	0.65	0.60	0.75	1.20	5.45	2.80	-	0.85	0.45	4.20
	200	0.30	0.30	0.35	1.00	1.60	5.40	2.70	-	0.85	0.45	4.20
	400	0.30	0.30	0.50	0.70	1.50	5.50	2.45	-	0.85	0.45	4.20

F.1. Additional simulation results of testing high dimensional mean values. In Section 4, we introduce **Models 1-4** for high dimensional mean tests. In this section, we show the numerical results for **Models 2-4** in Table 3.

F.2. Simulation results of testing high dimensional covariance and correlation coefficients . In this section, we carry out the simulation of the marginal test using the Pearson's covariance and Kendall's tau correlation matrices. For simplicity, we consider the one-sample problem. In the simulation, Z and $\boldsymbol{X} \in \mathbb{R}^d$ are the response variable and the explanatory vector. We generate n_1 data points of $(Z, \boldsymbol{X}^\top)^\top$ from the following models.

- **Model 5.** Let $\boldsymbol{\Sigma}_0^L, \boldsymbol{\Sigma}_1^L \in \mathbb{R}^{(d+1)\times(d+1)}$ to be

$$\boldsymbol{\Sigma}_0^L = \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & (\mathbf{D}^\star)^{-1/2}\boldsymbol{\Sigma}^\star(\mathbf{D}^\star)^{-1/2} \end{bmatrix}, \ \boldsymbol{\Sigma}_1^L = \begin{bmatrix} 1 & \boldsymbol{V}^\top \\ \boldsymbol{V} & (\mathbf{D}^\star)^{-1/2}\boldsymbol{\Sigma}^\star(\mathbf{D}^\star)^{-1/2} \end{bmatrix},$$

  where $\boldsymbol{V} \in \mathbb{R}^d$ has s nonzero entries with the magnitude $\mathrm{U}(u_1, u_2)$. Under the null hypothesis, we generate n_1 random vectors from $t(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\nu = 5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0^L$ as the samples of $(Z, \boldsymbol{X}^\top)^\top$. Under the alternative hypothesis, we generate the samples of $(Z, \boldsymbol{X}^\top)^\top$ from $t(5, \mathbf{0}, \boldsymbol{\Sigma}_1^L + \delta \boldsymbol{I}_{d+1})$ with $\delta = |\lambda_{\min}(\boldsymbol{\Sigma}_1^L)| + 0.5$.

The experimental results of **Model 5** are in Table 4. In **Model 5** we compare the proposed tests based on Pearson's covariance and Kendall' tau correlation matrices. The pattern of empirical size and power for **Model 5** is similar to **Models 1-4**. Moreover, the experiment shows that Kendall's tau based test is more powerful than the Pearson's covariance based one for distributions with the heavy tails and strong tail dependence.

F.3. Simulation results of increasing $\#(\mathcal{P})$. In this section, we discuss the impact of $\#(\mathcal{P})$ by simulation. In Sections 2.2 and 3.3, we require fixed $\mathcal{P}$ for the data-adaptive combined test. In Remark 3.6, we discuss theoretical difficulties of increasing $\#(\mathcal{P})$. In this section, we present the performance of proposed methods under various $\mathcal{P}$.

For this we generate the data based on **Model 1** in Section 4. We consider various $\mathcal{P}$. In detail, we set $\mathcal{P}_1 = \{1, 2\}$, $\mathcal{P}_2 = \{1, 2, \infty\}$, $\mathcal{P}_3 = \{1, 2, 3, 4, 5\}$, $\mathcal{P}_4 = \{1, 2, 3, 4, 5, \infty\}$, $\mathcal{P}_5 = \{1, 2, \ldots, 10, \infty\}$, and $\mathcal{P}_6 = \{1, 2, \ldots, 20, \infty\}$. We also consider various alternatives with $s = 5, 50, 100$, from sparse to dense. The simulation results are in Table 5.

From Table 5, we recommend using $\mathcal{P}_4 = \{1, 2, 3, 4, 5, \infty\}$. It has good performance for both sparse and dense alternatives. Table 5 also shows that there is no power advantage to add more elements to $\mathcal{P}_4$.

TABLE 4
Empirical size and power of **Model 5** *with $\alpha = 0.05$, $B = 300$, and $n_1 = 200$ based on 2000 replications.*

		Empirical size (%)													
		Pesrson's sample covariance							Kendall's tau						
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N
200	10	0.00	0.00	0.00	0.15	0.15	2.65	1.50	3.75	4.30	4.75	5.40	6.05	8.20	6.85
	50	0.00	0.00	0.00	0.00	0.05	2.60	1.30	1.20	1.75	1.90	3.35	4.35	8.75	5.25
	100	0.00	0.00	0.00	0.00	0.00	2.10	1.35	0.60	0.85	1.70	2.55	3.95	8.60	5.65
	150	0.00	0.00	0.00	0.00	0.00	2.65	1.25	0.60	0.85	1.50	2.65	3.75	8.35	4.75
	200	0.00	0.00	0.00	0.00	0.00	2.40	1.20	0.60	0.90	1.40	2.80	3.55	8.20	5.25
		Empirical power (%) with $s=5$, $u_1=0$, and $u_2 = 4\sqrt{\log(d)/n_1}$													
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N
200	10	13.70	20.7	27.95	34.00	38.85	55.20	50.60	73.20	78.75	81.40	83.45	84.05	84.20	84.00
	50	0.40	1.50	5.50	14.00	24.40	55.95	49.05	26.90	52.75	70.55	78.50	81.80	84.05	81.85
	100	0.05	0.40	3.20	11.55	23.65	55.65	48.85	11.85	39.20	66.35	77.05	81.90	84.15	82.00
	150	0.05	0.30	2.80	12.00	22.90	55.20	48.25	8.30	35.00	65.65	77.25	81.55	84.25	82.15
	200	0.05	0.40	3.05	11.85	23.30	55.20	47.55	6.95	34.95	65.55	76.55	81.70	84.05	81.45
		Empirical power (%) with $s=5$, $u_1=0$, and $u_2 = 3\sqrt{1/n_1}$													
d	s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	T_{ad}^N
200	10	10.25	10.55	11.35	11.65	12.80	16.80	13.95	75.85	75.05	74.15	72.90	70.65	46.55	68.20
	50	4.90	5.55	6.85	8.40	9.6	16.25	12.60	78.30	79.80	80.60	79.50	77.10	47.15	74.50
	100	2.95	4.05	4.85	6.45	8.2	17.20	11.15	73.80	78.65	80.90	80.35	77.60	46.85	75.10
	150	2.70	4.00	5.15	6.50	8.60	16.90	11.15	69.55	78.00	80.75	79.70	77.90	47.45	73.75
	200	2.75	3.65	5.35	6.85	8.60	16.45	11.15	68.00	78.45	81.15	80.70	77.60	46.95	74.25

TABLE 5
Empirical size and power of T_{ad}^N under **Model 1** *with $\alpha = 0.05$, $B = 300$, $d = 400$, and $n_1 = n_2 = 200$ based on 1000 replications.*

	Empirical size(%) with						Empirical power (%) with $s=5$, $u_1=0$, $u_2 = 4\sqrt{\log(d)/n_1}$					
s_0	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$
10	5.1	5.6	5.3	5.2	5.3	5.4	82.8	86.3	86.6	86.4	86.5	86.5
50	3.7	4.6	4.8	4.6	5.1	4.6	60.4	84.0	82.4	84.7	85.0	85.1
100	2.9	3.9	3.7	4.5	4.6	4.5	44.2	83.7	81.6	84.5	84.8	85.0
150	2.6	3.6	3.5	3.8	4.1	4.2	33.6	83.3	80.8	83.6	84.6	84.5
200	2.3	4.0	3.5	3.9	4.0	4.1	29.1	83.5	81.1	84.1	84.2	84.8
	Empirical power (%) with $s=50$, $u_1=0$,$u_2 = 4\sqrt{1/n_1}$						Empirical power (%)with $s=100$, $u_1=0$, $u_2 = 3\sqrt{1/n_1}$					
s_0	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$
10	76.0	75.2	78.6	77.3	74.8	75.1	71.1	65.1	0.1	69.5	64.1	64.8
50	75.8	75.2	79.0	78.2	78.0	78.0	79.6	73.8	77.9	76.3	72.7	74.3
100	70.1	71.1	79.4	78.3	77.3	76.1	78.6	72.0	77.0	76.0	73.1	74.2
150	65.0	67.5	78.1	77.2	75.3	75.3	75.8	68.5	76.9	75.8	73.1	73.7
200	60.2	65.9	76.8	76.5	74.6	74.0	74.5	68.5	76.0	75.1	73.8	73.8

F.4. Real data example. In this section, we apply our methods to analyze resting-state functional magnetic resonance imaging (fMRI) data. We aim to compare the resting-sate fMRI scans between the attention deficit hyperactivity disorder (ADHD) and normal children. For each subject, the resting-state fMRI scan is a high dimensional time series. Instead of dealing with the time series directly, we alternatively use an index named amplitude of low frequency fluctuation (ALFF) to yield a high dimensional vector for each subject. Each entry of ALFF is defined as the total power within the frequency range between 0.01 and 0.1 Hz of the corresponding entry of the original fMRI time series, which reflects the slow fluctuation. In general, ALFF reflects the intensity of regional spontaneous brain activity. As for the detailed definition of ALFF, we refer to [13]. Existing literature [13, 14] utilizes univariate two-sample t-tests to detect differentially experessed brain areas between the diseased and control groups based on ALFF. Before we conduct the univariate two-sample tests, it is a common practice to perform a global test to verify that there is significant difference of ALFFs between two groups. By the definition of ALFF, we utilize the high dimensional mean test to perform the global test.

Our experiment is based on the first dataset of Peking University from the ADHD-200 sample.[3] The sample consists of 85 subjects, in which 24 subjects have ADHD. Therefore, the control group has 61 subjects. ALFF analysis is performed by using the C-PAC software. The C-PAC software preprocesses the data by registering each person's fMRI scan to the standard MN152 template. To increase the signal-noise ratio, the software also performs slice timing correction, body motion correction, nuisance signal correction, and temporal filtering. Because of the difference of individual brain baseline activity, we standardize the ALFF for each subject. We then use the Gaussian kernel to perform the spatial smoothing for each subject. Moreover, existing literature and psychological knowledge suggest that the ALFF of brain's gray matter is related to the mental disease. Hence, we restrict the testing area to the gray matter of the brain. For detailed description of the processing procedure, we refer to [13], [14], and the user guide of C-PAC software.[4]

Figure 3 illustrates P-values of univariate two-sample t-tests. Figure 3(A) illustrates the P-value map to the standard MN152 brain template with the slice thickness 3mm at the given threshold (P-value < 0.2). Moreover, Figure 3(B) illustrates the estimated density of these P-values. Figure 3 shows there are significant ALFF differences between the diseased and control groups in

[3] The website for ADHD-200 sample is http://fcon_1000.projects.nitrc.org/indi/adhd200/.

[4] The website for the C-PAC software is http://fcp-indi.github.io/.

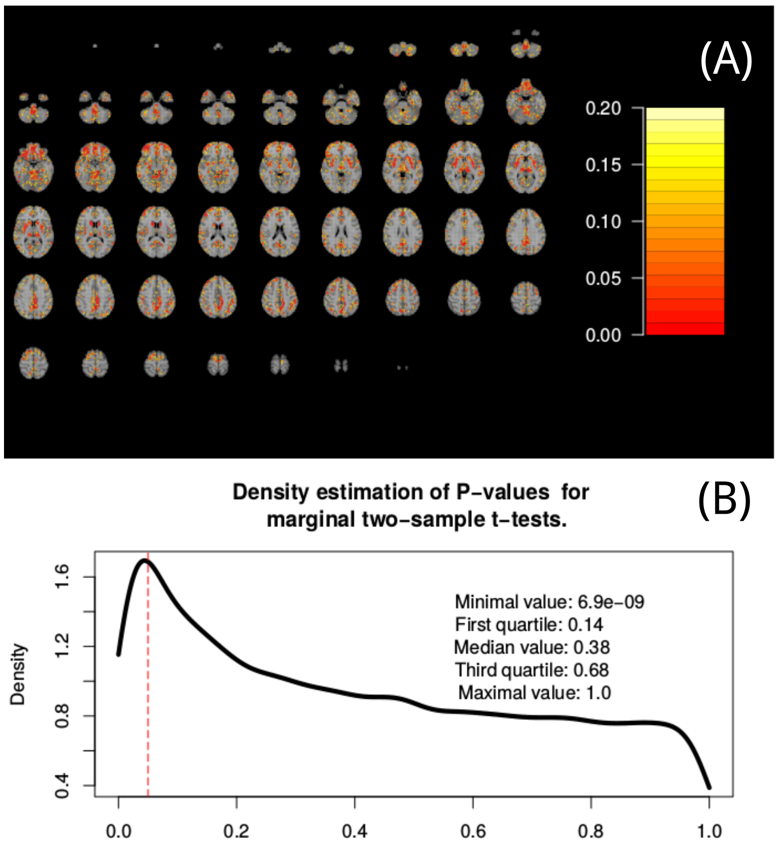

Fig 3. *P-values of the marginal two-sample t-tests on ALFFs between ADHD and control groups. (A) The P-value map on the standard MN152 brain template with the slice thickness 3mm at the given threshold (P-value < 0.2). (B) The estimated density of the P-values and some summary statistics.*

some brain areas.

TABLE 6
P-values of the (s_0, p)-norm tests and data-adaptive combined test with $s_0 = 40, 400, 4000, 8000$ and $B = 1000$ on the ALFF data.

P-values of global tests between the ADHD and control groups							
s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	$T_{\rm ad}^N$
40	0.001	0.001	0.001	0.001	0.001	0.001	0.000
400	0.013	0.013	0.012	0.011	0.010	0.000	0.000
4000	0.016	0.015	0.015	0.015	0.013	0.000	0.000
8000	0.016	0.015	0.013	0.011	0.008	0.000	0.000
P-values of global tests within the control group							
s_0	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=\infty$	$T_{\rm ad}^N$
40	0.192	0.192	0.193	0.195	0.196	0.254	0.237
400	0.301	0.295	0.290	0.288	0.284	0.299	0.355
4000	0.373	0.362	0.352	0.337	0.323	0.273	0.354
8000	0.406	0.394	0.387	0.375	0.360	0.282	0.370

We then use both the individual (s_0, p)-norm test and data-adaptive combined test with balanced $\mathcal{P} = \{1, \ldots, 5, \infty\}$ to perform the global test. We also randomly split the sample for the control group into two subsamples with 30 and 31 subjects. We then perform the global mean test between the two subsamples of the control group to confirm the validity of our proposed methods. As is shown in Figure 3, at most 20% of the gray matter is potentially different between the diseased and control groups. Therefore, considering that the voxel size is about 40000, we set $s_0 = 40, 400, 4000, 8000$ in the experiment. The experiment result is presented in Table 6, which shows that our proposed methods are quite powerful to distinguish the ADHD and control groups.

REFERENCES

[1] ARCONES, M. A. and GINE, E. (1993). Limit theorems for U-processes. *Annals of Probability* **21** 1494–1542.

[2] BARVINOK, A. (2014). Thrifty approximations of convex bodies by polytopes. *International Mathematics Research Notices* **2014** 4341–4356.

[3] BONNÉRY, D., BREIDT, F. J. and COQUET, F. (2012). Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population. *Bernoulli* **18** 1361–1385.

[4] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press.

[5] CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106** 672–684.

[6] CAI, T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means

under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** 349–372.

[7] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Central limit theorems and bootstrap in high dimensions. *arXiv preprint arXiv:1412.3661* .

[8] David, H. A. and Nagaraja, H. N. (2003). *Order Statistics (3rd)*. John Wiley.

[9] Delaigle, A., Hall, P. and Jin, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on student's t-statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73** 283–301.

[10] Dudley, R. M. (2014). *Uniform Central Limit Theorems (2rd)*. Cambridge University Press.

[11] Nazarov, F. (2003). On the maximal perimeter of a convex set in R^n with respect to a Gaussian measure. *Geometric Aspects of Functional Analysis* Lecture Notes in Mathematics Volume 1807, 169–187.

[12] Tsagris, M., Beneki, C. and Hassani, H. (2014). On the folded normal distribution. *Mathematics* **2** 12–28.

[13] Zang, Y., He, Y., Zhu, C., Cao, Q., Sui, M., Liang, M., Tian, L., Jiang, T. and Wang, Y. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development* **29** 83–91.

[14] Zou, Q., Zhu, C., Yang, Y., Zuo, X., Long, X., Cao, Q., Wang, Y. and Zang, Y. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *Journal of Neuroscience Methods* **172** 137–141.