ASYMPTOTICALLY INDEPENDENT U-STATISTICS IN HIGH-DIMENSIONAL TESTING[*]

By Yinqiu He[†], Gongjun Xu[†] Chong Wu[‡] and Wei Pan[§]

University of Michigan[†], *Florida State University* [‡], *and University of Minnesota*[§]

Many high-dimensional hypothesis tests aim to globally examine marginal or low-dimensional features of a high-dimensional joint distribution, such as testing of mean vectors, covariance matrices and regression coefficients. This paper constructs a family of U-statistics as unbiased estimators of the ℓ_p-norms of those features. We show that under the null hypothesis, the U-statistics of different finite orders are asymptotically independent and normally distributed. Moreover, they are also asymptotically independent with the maximum-type test statistic, whose limiting distribution is an extreme value distribution. Based on the asymptotic independence property, we propose an adaptive testing procedure which combines p-values computed from the U-statistics of different orders. We further establish power analysis results and show that the proposed adaptive procedure maintains high power against various alternatives.

1. Introduction.

Motivation. Analysis of high-dimensional data, whose dimension p could be much larger than the sample size n, has emerged as an important and active research area [e.g., 21, 72, 25, 23]. In many large-scale inference problems, one is often interested in globally testing some overall patterns of low-dimensional features of the high-dimensional random observations. One example is genome-wide association studies (GWAS), whose primary goal is to identify single nucleotide polymorphisms (SNPs) associated with certain complex diseases of interest. A popular approach in GWAS is to perform univariate tests which examine each SNP one by one. This however may lead to low statistical power due to the weak effect size of each SNP [56] and the small statistical significance threshold ($\sim 10^{-8}$) chosen to control the multiple-comparison type I error [48]. Researchers therefore have proposed to globally test a genetic marker set with many SNPs [73, 48] in order

[*]This research is supported by NSF grants DMS-1711226, DMS-1712717, SES-1659328, CAREER SES-1846747, and NIH grants R01GM113250, R01GM126002, R01HL105397 and R01HL116720

MSC 2010 subject classifications: Primary 62F03, 62F05

Keywords and phrases: High-dimensional hypothesis test, U-statistics, Adaptive testing

to achieve higher statistical power and to better understand the underlying genetic mechanisms.

In this paper, we focus on a family of global testing problems in the high-dimensional setting, including testing of mean vectors, covariance matrices and regression coefficients in generalized linear models. These problems can be formulated as $H_0 : \mathcal{E} = \mathbf{0}$, where $\mathbf{0}$ is an all zero vector, $\mathcal{E} = \{e_l : l \in \mathcal{L}\}$ is a parameter vector with $\mathcal{L}$ being the index set, and e_l's being the corresponding parameters of interest, e.g., elements in mean vectors, covariance matrices or coefficients in generalized linear models. For the global testing problem $H_0 : \mathcal{E} = \mathbf{0}$ versus $H_A : \mathcal{E} \neq \mathbf{0}$, two different types of methods are often used in the literature. One is sum-of-squares-type statistics. They are usually powerful against "dense" alternatives, where $\mathcal{E}$ has a high proportion of nonzero elements with a large $\|\mathcal{E}\|_2 = \sum_{l\in\mathcal{L}} e_l^2$ or its weighted variants. See examples in mean testing [e.g., 4, 27, 69, 14, 13, 28, 71] and covariance testing [e.g., 3, 51, 15, 54]. The other is maximum-type statistics. They are usually powerful against "sparse" alternatives, where $\mathcal{E}$ has few nonzero elements with a large $\|\mathcal{E}\|_\infty$ [e.g., 43, 55, 32, 9, 10, 11, 67]. More recently, [22, 79] also proposed to combine these two kinds of test statistics. However, for denser or only moderately dense alternatives, neither of these two types of statistics may be powerful, as will be further illustrated in this paper both theoretically and numerically. Importantly, in real applications, the underlying truth is usually unknown, which could be either sparse, dense, or in-between. As global testing could be highly underpowered if an inappropriate testing method is used [e.g., 17], it is desired in practice to have a testing procedure with high statistical power against a variety of alternatives.

A Family of Asymptotically Independent U-Statistics. To address these issues, we propose a U-statistics framework and introduce its applications to adaptive high-dimensional testing. The U-statistics framework constructs *unbiased* and *asymptotically independent* estimators of $\|\mathcal{E}\|_a^a := \sum_{l\in\mathcal{L}} e_l^a$ for different (positive) integers a, where $a = 2$ corresponds to a sum-of-squares-type statistic, and an even integer $a \to \infty$ yields a maximum-type statistic. The adaptive testing then combines the information from different $\|\mathcal{E}\|_a^a$'s, and our power analysis shows that it is powerful against a wide range of alternatives, from highly sparse, moderately sparse to dense, to highly dense.

To illustrate our idea, suppose $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are n independent and identically distributed (i.i.d.) copies of a random vector $\mathbf{z}$. We consider the setting where each parameter e_l has an unbiased kernel function estimator $K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}})$, and γ_l is the smallest integer such that for any $1 \leq i_1 \neq \ldots \neq i_{\gamma_l} \leq n$, $\mathrm{E}[K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}})] = e_l$. This includes many testing problems on moments of low orders, such as entries in mean vectors, covariance matrices and score vectors of generalized linear models, which shall be discussed in details. The family of U-statistics can be constructed generally as follows. For integers $a \geq 1$, and $1 \leq i_1 \neq \ldots \neq i_{\gamma_l} \neq \ldots \neq i_{(a-1)\times\gamma_l+1} \cdots \neq i_{a\times\gamma_l} \leq n$, since the $\mathbf{z}$'s are i.i.d., we have $\mathrm{E}[K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}}) \cdots K_l(\mathbf{z}_{i_{(a-1)\times\gamma_l+1}}, \ldots, \mathbf{z}_{i_{a\times\gamma_l}})] = e_l^a$. Therefore, we can construct an unbiased estimator of the parameters of augmented powers e_l^a with different a. Then $\|\mathcal{E}\|_a^a$ has an unbiased estimator

$$\mathcal{U}(a) = \sum_{l\in\mathcal{L}} (P^n_{a\times\gamma_l})^{-1} \sum_{1\leq i_1\neq\ldots\neq i_{a\times\gamma_l}\leq n} \prod_{k=1}^{a} K_l(\mathbf{z}_{i_{(k-1)\times\gamma_l+1}}, \ldots, \mathbf{z}_{i_{k\times\gamma_l}}), \tag{1.1}$$

where $P^n_k = n!/(n-k)!$ denotes the number of k-permutations of n. We call a the order of the U-statistic $\mathcal{U}(a)$. If $a > b$, we say $\mathcal{U}(a)$ is of higher order than $\mathcal{U}(b)$ and vice versa.

This construction procedure can be applied to many testing problems. We give three common examples below for illustration and more detailed case-studies will be discussed in Sections 2 and 4.

EXAMPLE 1. *Consider one-sample mean testing of $H_0 : \boldsymbol{\mu} = \mathbf{0}$, where $\mathcal{E} = \boldsymbol{\mu}$ is the mean vector of a p-dimensional random vector $\mathbf{x}$. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n i.i.d. copies of $\mathbf{x}$. For each $i = 1, \ldots, n$, $j = 1, \ldots, p$, $x_{i,j}$ is a simple unbiased estimator of μ_j, then we can take the kernel function $K_j(\mathbf{x}_i) = x_{i,j}$. Following (1.1), we know the U-statistic*

$$\mathcal{U}(a) = (P^n_a)^{-1} \sum_{j=1}^{p} \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^{a} x_{i_k,j}$$

is an unbiased estimator of $\|\mathcal{E}\|_a^a = \|\boldsymbol{\mu}\|_a^a = \sum_{j=1}^p \mu_j^a$. Please see Section 4.1 for the two-sample mean testing example and related theoretical properties.

EXAMPLE 2. *Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n i.i.d. copies of a random vector $\mathbf{x}$ with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{j_1,j_2}\}_{p\times p}$. For covariance testing $H_0 : \sigma_{j_1,j_2} = 0$ for any $1 \leq j_1 \neq j_2 \leq p$, we have $\mathcal{E} = \{\sigma_l : l \in \mathcal{L}\}$ with $\mathcal{L} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$. Since $x_{i,j_1}x_{i,j_2}$ is a simple unbiased estimator of σ_{j_1,j_2}, then for each pair $l = (j_1, j_2) \in \mathcal{L}$, we can take the kernel function $K_l(\mathbf{x}_i) = x_{i,j_1}x_{i,j_2}$. Following (1.1), the U-statistic*

$$\mathcal{U}(a) = (P^n_a)^{-1} \sum_{1\leq j_1\neq j_2\leq p} \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^{a} (x_{i_k,j_1}x_{i_k,j_2})$$

is an unbiased estimator of $\|\mathcal{E}\|_a^a = \sum_{1\leq j_1\neq j_2\leq p} \sigma^a_{j_1,j_2}$. Please see Section 2 for the general case with unknown $\boldsymbol{\mu}$.

Example 3. *Consider a response variable y and its covariates $\mathbf{x} \in \mathbb{R}^p$ following a generalized linear model:* $\mathrm{E}(y|\mathbf{x}) = g^{-1}(\mathbf{x}^\intercal \boldsymbol{\beta})$, *where g is the canonical link function and $\boldsymbol{\beta} \in \mathbb{R}^p$ are the regression coefficients. Suppose that $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$, are i.i.d. copies of $(\mathbf{x}, y)$. For testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, the score vectors $(S_{i,j} = (y_i - \mu_{0,i})x_{i,j} : j = 1, \ldots, p)^\intercal$ are often used in the literature, where $\mu_{0,i} = g^{-1}(\mathbf{x}_i^\intercal \boldsymbol{\beta}_0)$. Note that $\mathrm{E}(S_{i,j}) = 0$ under H_0. Thus to test H_0, we can take $\mathcal{E} = \{\mathrm{E}(S_{i,j}) : j = 1, \ldots, p\}$ and use the U-statistic*

$$\mathcal{U}(a) = (P_a^n)^{-1} \sum_{j=1}^{p} \sum_{1 \le i_1 \ne \ldots \ne i_a \le n} \prod_{k=1}^{a} S_{i_k, j},$$

which is an unbiased estimator of $\|\mathcal{E}\|_a^a = \sum_{j=1}^p \{\mathrm{E}(S_{i,j})\}^a$. Please see Section 4.3.

Related Literature. For high-dimensional testing, some other adaptive testing procedures have recently been proposed in [61, 76, 74]. These works combine the p-values of a family of sum-of-powered statistics that are powerful against different $\|\mathcal{E}\|_a^a$'s. However in these existing works, to evaluate the p-value of the adaptive test statistic, the joint asymptotic distribution of the statistics is difficult to obtain or calculate. Accordingly computationally expensive resampling methods are often used in practice [61, 48, 78]. For some special cases such as testing means and the coefficients of generalized linear models, [76] and [74] derived the limiting distributions of the test statistics under the framework of a family of von Mises V-statistics. However, the constructed V-statistics are usually *correlated* and *biased* estimators of the target $\|\mathcal{E}\|_a^a$. It follows that in [76] and [74], numerical approximations are still needed to calculate the tail probabilities of the adaptive test statistics; see Remark 4.1 and Section 4.3. In addition, these existing adaptive testing works mainly focus on the first-order moments, and their results do not directly apply to testing second-order moments, such as covariance matrices.

To overcome these issues, this paper considers the proposed family of unbiased U-statistics. There are some other recent works providing important results on high-dimensional U-statistics [e.g., 16, 52, 82]. For instance, [82] considered testing the regression coefficients in linear models using the fourth-order U-statistic; [52] studied the limiting distributions of rank-based U-statistics; and [16] studied bootstrap approximation of the second-order U-statistics. However, these results do not directly apply to the high-order U-statistics considered in this paper.

Our Contributions. We establish the theoretical properties of the U-statistics in various high dimensional testing problems, including testing mean vectors, regression coefficients of generalized linear models, and covariance matrices. Our contributions are summarized as follows.

Under the null hypothesis, we show that the normalized U-statistics of different finite orders are jointly normally distributed. The result applies generally for any asymptotic regime with $n \to \infty$ and $p \to \infty$. In addition, we prove that all the finite-order U-statistics are asymptotically independent with each other under the null hypothesis. Moreover, we prove that U-statistics of finite orders are also asymptotically independent of the maximum-type test statistic with a limiting extreme value distribution.

Under the alternative hypothesis, we further analyze the asymptotic power for U-statistics of different orders. We show that when $\mathcal{E}$ has denser nonzero entries, $\mathcal{U}(a)$'s of lower orders tend to be more powerful; and when $\mathcal{E}$ has sparser nonzero entries, $\mathcal{U}(a)$'s of higher orders tend to be more powerful. More interestingly, we show that in the boundary case of "moderate" sparsity levels, $\mathcal{U}(a)$ with a finite $a > 2$ gives the highest power among the family of U-statistics, clearly indicating the inadequacy of both the sum-of-squares- and the maximum-type statistics.

An important application of the independence property among $\mathcal{U}(a)$'s is to construct adaptive testing procedures by combining the information of different $\mathcal{U}(a)$'s, whose univariate distributions or p-values can be easily combined to form a joint distribution to calculate the p-value of an adaptive test statistic. Compared with other existing works [e.g., 76, 74], numerical approximations of tail probabilities are no longer needed. As shown in the power analysis, an adaptive integration of information across different tests leads to a powerful testing procedure.

The rest of the paper is organized as follows. In Sections 2 and 3, we illustrate the framework by a covariance testing problem. Particularly, in Section 2.1, we study the U-statistics under null hypothesis; in Section 2.2, we analyze the power of the U-statistics; in Section 2.3, we develop an adaptive testing procedure. In Sections 3.1 and 3.2, we report simulations and a real dataset analysis. In Section 4, we study other high-dimensional testing problems, including testing means, regression coefficients and two-sample covariances. In Section 5, we discuss several extensions of the proposed framework. We give proofs and other stimulations in Supplementary Material.

2. Motivating Example: One-Sample Covariance Testing. The constructed family of U-statistics and adaptive testing procedure can be applied to various high-dimensional testing problems. In this section, we illustrate the framework with a motivating example of one-sample covariance testing. Analogous results for other high-dimensional testing problems in

Section 4 can be obtained following similar analyses. We showcase the study of one-sample covariance testing problem since this is more challenging than mean testing due to the two-way dependency structure and the one-sample problem can be used as the building block for more general cases.

Specifically, we focus on testing

$$H_0 : \sigma_{j_1,j_2} = 0 \quad \forall\ 1 \leq j_1 \neq j_2 \leq p, \tag{2.1}$$

where $\mathbf{\Sigma} = \{\sigma_{j_1,j_2} : 1 \leq j_1, j_2 \leq p\}$ is the covariance matrix of a p-dimensional real-valued random vector $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ with $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$. The observed data include n i.i.d. copies of $\mathbf{x}$, denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\intercal$. In factor analysis, testing H_0 in (2.1) can be used to examine whether $\mathbf{\Sigma}$ has any significant factor or not [2].

Global testing of covariance structure plays an important role in many statistical analysis and applications; see a review in [8]. Conventional tests include the likelihood ratio test, John's test, and Nagao's test, etc. [2, 59]. These methods, however, often fail in the high-dimensional setting when both $n, p \to \infty$. To address this issue, new procedures have been recently proposed [e.g., 3, 44, 45, 68, 66, 62, 51, 15, 43, 55, 9, 54, 67, 50]. However these methods might suffer from loss of power when the sparsity level of the alternative covariance matrix varies. In the following subsections, we introduce the general U-statistics framework, study their asymptotic properties, and develop a powerful adaptive testing procedure.

We introduce some notation. For two series of numbers $u_{n,p}$, $v_{n,p}$ that depend on n, p: $u_{n,p} = o(v_{n,p})$ denotes $\limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| = 0$; $u_{n,p} = O(v_{n,p})$ denotes $\limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| < \infty$; $u_{n,p} = \Theta(v_{n,p})$ denotes $0 < \liminf_{n,p\to\infty} |u_{n,p}/v_{n,p}| \leq \limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| < \infty$; $u_{n,p} \simeq v_{n,p}$ denotes $\lim_{n,p\to\infty} u_{n,p}/v_{n,p} = 1$. Moreover, $\xrightarrow{P}$ and $\xrightarrow{D}$ represent the convergence in probability and distribution respectively. For p-dimensional random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and $\forall j_1, \ldots, j_t \in \{1, \ldots, p\}$, we write the central moment as

$$\Pi_{j_1,\ldots,j_t} = \mathrm{E}[(x_{j_1} - \mu_{j_1}) \ldots (x_{j_t} - \mu_{j_t})]. \tag{2.2}$$

2.1. *Asymptotically Independent U-Statistics.* For testing (2.1), the set of parameters that we are interested in is $\mathcal{E} = \{\sigma_{j_1,j_2} : 1 \leq j_1 \neq j_2 \leq p\}$. Following the previous analysis of (1.1), since σ_{j_1,j_2} has a simple unbiased estimator $x_{i_1,j_1}x_{i_1,j_2} - x_{i_1,j_1}x_{i_2,j_2}$ with $1 \leq i_1 \neq i_2 \leq n$, then for integers $a \geq 1$, an unbiased U-statistic of $\|\mathcal{E}\|_a^a = \sum_{1\leq j_1\neq j_2\leq p} \sigma_{j_1,j_2}^a$ is

$$\mathcal{U}(a) = (P^n_{2a})^{-1} \sum_{1\leq j_1\neq j_2\leq p}\ \sum_{1\leq i_1\neq\ldots\neq i_{2a}\leq n}\ \prod_{k=1}^{a} (x_{i_{2k-1},j_1}x_{i_{2k-1},j_2} - x_{i_{2k-1},j_1}x_{i_{2k},j_2}).$$

This is equivalent to

$$\mathcal{U}(a) = \sum_{1\le j_1\neq j_2\le p}\sum_{c=0}^{a}(-1)^c\binom{a}{c}\frac{1}{P^n_{a+c}}\sum_{1\le i_1\neq\ldots\neq i_{a+c}\le n} \prod_{k=1}^{a-c}(x_{i_k,j_1}x_{i_k,j_2})\prod_{s=a-c+1}^{a}x_{i_s,j_1}\prod_{t=a+1}^{a+c}x_{i_t,j_2}. \tag{2.3}$$

Remark 2.1. *The U-statistics can be constructed by another method equivalently. Given $1\le j_1\neq j_2\le p$, define $\varphi_{j_1,j_2}=\sigma_{j_1,j_2}+\mu_{j_1}\mu_{j_2}$. Then*

$$\sum_{1\le j_1\neq j_2\le p}\sigma^a_{j_1,j_2} = \sum_{1\le j_1\neq j_2\le p}\sum_{c=0}^{a}\binom{a}{c}\varphi^{a-c}_{j_1,j_2}\times(-\mu_{j_1}\mu_{j_2})^c, \tag{2.4}$$

which is a polynomial function of the moments μ_j and φ_{j_1,j_2}. Since μ_j and φ_{j_1,j_2} have unbiased estimators $x_{i,j}$ and $x_{i,j_1}x_{i,j_2}$ respectively, then for $1\le i_1\neq\ldots\neq i_{a+c}\le n$, $\mathrm{E}(\prod_{k=1}^{a-c}x_{i_k,j_1}x_{i_k,j_2}\prod_{s=a-c+1}^{a}x_{i_s,j_1}\prod_{t=a+1}^{a+c}x_{i_t,j_2}) = \varphi^{a-c}_{j_1,j_2}\mu^c_{j_1}\mu^c_{j_2}$. Given this and (2.4), the U-statistics (2.3) can be obtained.

Remark 2.2. *The summed term with $c=0$ in (2.3) is*

$$\tilde{\mathcal{U}}(a) := (P^n_a)^{-1}\sum_{1\le i_1\neq\cdots\neq i_a\le n}\sum_{1\le j_1\neq j_2\le p}\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2}), \tag{2.5}$$

which has the same form as the simplified U-statistic for mean zero observations in Example 2, and is shown to be the leading term of (2.3) in proof.

We next introduce some nice properties of the U-statistics (2.3). The first one is the following location invariant property.

Proposition 2.1. *$\mathcal{U}(a)$ constructed as in (2.3) is location invariant; that is, for any vector $\mathbf{\Delta}\in\mathbb{R}^p$, the U-statistic constructed based on the transformed data $\{\mathbf{x}_i+\mathbf{\Delta}: i=1,\ldots,n\}$ is still $\mathcal{U}(a)$.*

The following proposition verifies that the constructed U-statistics are unbiased estimators of $\|\mathcal{E}\|^a_a=\sum_{1\le j_1\neq j_2\le p}\sigma^a_{j_1,j_2}$.

Proposition 2.2. *For any integer a, $\mathrm{E}[\mathcal{U}(a)]=\sum_{1\le j_1\neq j_2\le p}\sigma^a_{j_1,j_2}$. Under H_0 in (2.1), $\mathrm{E}[\mathcal{U}(a)]=0$.*

We next study the limiting properties of the constructed U-statistics under H_0 given the following assumptions on the random vector $\mathbf{x}=(x_1,\ldots,x_p)^\intercal$.

CONDITION 2.1 (Moment assumption). $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^8<\infty$ *and* $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^2>0$.

CONDITION 2.2 (Dependence assumption). *For a sequence of random variables* $\mathbf{z}=\{z_j:j\ge 1\}$ *and integers* $a<b$, *let* $\mathcal{Z}_a^b$ *be the* σ*-algebra generated by* $\{z_j:j\in\{a,\dots,b\}\}$. *For each* $s\ge 1$, *define the* α*-mixing coefficient* $\alpha_{\mathbf{z}}(s)=\sup_{t\ge 1}\{|P(A\cap B)-P(A)P(B)|:A\in\mathcal{Z}_1^t,B\in\mathcal{Z}_{t+s}^{\infty}\}$. *We assume that under* H_0, $\mathbf{x}$ *is* α*-mixing with* $\alpha_{\mathbf{x}}(s)\le M\delta^s$, *where* $\delta\in(0,1)$ *and* $M>0$ *are some constants.*

CONDITION 2.2* (Alternative dependence assumption to Condition 2.2). *Following the notation in* (2.2), *we assume that under* H_0, *for any* $j_1,j_2,j_3\in\{1,\dots,p\}$, $\Pi_{j_1,j_2,j_3}=0$; *for any* $j_1,j_2,j_3,j_4\in\{1,\dots,p\}$, $\Pi_{j_1,j_2,j_3,j_4}=\kappa_1(\sigma_{j_1,j_2}\sigma_{j_3,j_4}+\sigma_{j_1,j_3}\sigma_{j_2,j_4}+\sigma_{j_1,j_4}\sigma_{j_2,j_3})$ *for some constant* $\kappa_1<\infty$; *and for* $t=6,8$, *and any* $j_1,\cdots,j_t\in\{1,\dots,p\}$, $\Pi_{j_1,\cdots,j_t}=0$ *when at least one of these indexes appears odd times in* $\{j_1,\cdots,j_t\}$.

Condition 2.1 assumes that the eighth marginal moments of $\mathbf{x}$ are uniformly bounded from above and the second moments are uniformly bounded from below, which are true for most light-tailed distributions. Condition 2.2 assumes weak dependence among different x_j's under H_0, since the uncorrelatedness of x_j's under H_0 may not imply the independence of them, especially when x_j's are non-Gaussian. Under H_0, Condition 2.2 automatically holds when $\mathbf{x}$ is Gaussian or m-dependent. The mixing-type weak dependence is similarly considered in previous works such as [5, 13, 76] and also commonly assumed in time series and spatial statistics [26, 64]. Moreover, the variables in our motivating genome-wide association studies have a local dependence structure, with their associations often decreasing to zero as the corresponding physical distances on a chromosome increase. We note that it suffices to have Condition 2.2 hold up to a permutation of the variables.

Alternatively, we can substitute Condition 2.2 with Condition 2.2*. Condition 2.2* specifies some higher order moments of $\mathbf{x}$ and is satisfied when $\mathbf{x}$ follows an elliptical distribution with finite eighth moments and covariance $\mathbf{\Sigma}$ [see 2, 24, 59, 60]. Conditions 2.2* and 2.2 become equivalent when $\mathbf{x}$ follows a multivariate Gaussian distribution. The fourth moment condition is also assumed in other high-dimensional research [10]. In this work, the eighth moment condition is needed to establish the asymptotic joint distribution of different U-statistics.

The following theorem specifies the asymptotic variances of the finite order U-statistics and their joint limiting distribution. Since the U-statistics are degenerate under H_0, an analysis different from the asymptotic theory on non-degenerate U-statistics [e.g., 38] is needed in the proof.

THEOREM 2.1. *Under H_0 in (2.1) and Conditions 2.1 and 2.2 (or 2.2*), for $\mathcal{U}(a)$'s defined in (2.3) and any distinct finite (and positive) integers $\{a_1, \dots, a_m\}$, as $n, p \to \infty$,*

$$\Big[\frac{\mathcal{U}(a_1)}{\sigma(a_1)}, \dots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)}\Big]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m), \tag{2.6}$$

where

$$\sigma^2(a) := \mathrm{var}[\mathcal{U}(a)] \simeq \frac{a!}{P_a^n} \sum_{1 \le j_1 \ne j_2 \le p;\, 1 \le j_3 \ne j_4 \le p} (\Pi_{j_1,j_2,j_3,j_4})^a, \tag{2.7}$$

with Π_{j_1,j_2,j_3,j_4} defined in (2.2). Note that $\sigma^2(a) = \Theta(p^2 n^{-a})$.

Theorem 2.1 shows that after normalization, the finite-order U-statistics have a joint normal limiting distribution with an identity covariance matrix, which implies that they are asymptotically independent as $n, p \to \infty$. The nice independence property makes it easy to combine these U-statistics and apply our proposed adaptive testing later. Moreover, the conclusion holds on general asymptotic regime for $n, p \to \infty$, without any constraint on the relationship between n and p. We will also see in Section 4 that similar results hold generally for some other testing problems.

REMARK 2.3. *Theorem 2.1 discusses the U-statistics of finite orders, i.e., the a values do not grow with n, p. When $\{x_1, \dots, x_p\}$ are independent, Theorem 2.1 can be extended when $a = O(1)\min\{\log^\epsilon n, \log^\epsilon p\}$ for some $\epsilon > 0$. On the other hand, we will show in Section 2.2 that it is usually enough to include $\mathcal{U}(a)$'s of finite a. Therefore, we do not pursue the general case when a grows with n, p in this work.*

In the following, we further discuss the maximum-type test statistic $\mathcal{U}(\infty)$, which corresponds to the ℓ_∞-norm of the parameter vector $\mathcal{E} = \{e_l : l \in \mathcal{L}\}$, that is, $\|\mathcal{E}\|_\infty = \max_{l \in \mathcal{L}} |e_l|$. In the existing literature, there is already some corresponding established work [43, 9] on the test statistic:

$$M_n^* := \max_{1 \le j_1 \ne j_2 \le p} |\hat\sigma_{j_1,j_2} / \sqrt{\hat\sigma_{j_1,j_1}\hat\sigma_{j_2,j_2}}|, \tag{2.8}$$

where $(\hat\sigma_{j_1,j_2})_{p\times p} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\intercal / n$ and $\bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i / n$. We will take $\mathcal{U}(\infty) = M_n^*$ below. The limiting distribution of $\mathcal{U}(\infty)$ was first studied in [43] and extended by [9, 55, 67]. Next we restate the result in [9], which gives the limiting distribution of (2.8) under the following condition.

Condition 2.3. *Consider the random vector* $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ *with mean vector* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$ *and covariance matrix* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_{1,1}, \ldots, \sigma_{p,p})$. $(x_j - \mu_j)/\sqrt{\sigma_{j,j}}$ *are i.i.d. for* $j = 1, \ldots, p$. *Furthermore,* $\mathrm{E} e^{t_0(|x_1-\mu_1|/\sqrt{\sigma_{1,1}})^{\varsigma}} < \infty$ *for some* $0 < \varsigma \leq 2$ *and* $t_0 > 0$.

Theorem 2.2 (Cai and Jiang [9, Theorem 2]). *Assume Condition 2.3 and* $\log p = o(n^{\beta})$, *where* $\beta = \varsigma/(4+\varsigma)$. *Then* $P(n \times \mathcal{U}(\infty)^2 + \varpi_p \leq u) \to G(u) = e^{-(1/\sqrt{8\pi})e^{-u/2}}$, *where* $\varpi_p = -4\log p + \log\log p$ *and* $G(u)$ *is an extreme value distribution of type I.*

Theorems 2.1 and 2.2 give the limiting distributions of $\mathcal{U}(a)$ of finite orders and $\mathcal{U}(\infty)$ respectively; it is of interest to examine their joint distribution. The following theorem shows that although $\mathcal{U}(\infty)$ has limiting distribution different from $\mathcal{U}(a)$, $a < \infty$, they are still asymptotically independent.

Theorem 2.3. *Assume that Condition 2.1 is satisfied, Condition 2.3 holds for* $\varsigma = 2$, *and* $\log p = o(n^{1/7})$. *For finite integers* $\{a_1, \ldots, a_m\}$, *under* H_0, $\mathcal{U}(a_1), \ldots, \mathcal{U}(a_m)$ *and* $\mathcal{U}(\infty)$ *are mutually asymptotically independent. In specific, for any* $z_1, \ldots, z_m, y \in \mathbb{R}$, *as* $n, p \to \infty$,

$$\Big| P\Big(n\mathcal{U}(\infty)^2 + \varpi_p \geq y, \frac{\mathcal{U}(a_1)}{\sigma(a_1)} \leq z_1, \ldots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)} \leq z_m\Big) - P\Big(n\mathcal{U}(\infty)^2 + \varpi_p \geq y\Big) \times \prod_{r=1}^{m} P\Big(\frac{\mathcal{U}(a_r)}{\sigma(a_r)} \leq z_r\Big) \Big| \to 0.$$

Theorem 2.1 suggests that all the finite-order U-statistics are asymptotically independent with each other. Given this, Theorem 2.3 further shows that the maximum-type test statistic $\mathcal{U}(\infty)$ is also asymptotically mutually independent with those finite-order U-statistics. The conclusion shares similarity with some classical results on the asymptotic independence between the sum-of-squares-type and maximum-type statistics. Specifically, for random variables $w_1, \ldots, w_n$, [39, 36] proved the asymptotic independence between $\sum_{i=1}^{n} w_i^2$ and $\max_{i=1,\ldots,n}|w_i|$ for weakly dependent observations. The similar independence properties were extensively studied in literature [e.g. 57, 37, 63, 41, 76, 53]. However, there are several differences between existing literature and the results in this paper. First, we discuss a family of U-statistics $\mathcal{U}(a)$'s, which takes different a values, and $\mathcal{U}(2)$ here corresponding to the sum-of-squares-type statistic is only a special case of general $\mathcal{U}(a)$. Furthermore, we have shown not only the asymptotic independence between $\mathcal{U}(a)$ and $\mathcal{U}(\infty)$, but also the asymptotic independence among $\mathcal{U}(a)$'s of finite a values. Second, the constructed $\mathcal{U}(a)$'s are unbiased estimators, which are different from the sum-of-squares statistics usually examined in the literature. Moreover, the x's are allowed to be dependent and the theoretical development in the covariance testing involves a two-way dependence structure, which requires different proof techniques from the existing studies.

REMARK 2.4. *An alternative way to construct $\mathcal{U}(\infty)$ is to standardize $\hat{\sigma}_{j_1,j_2}$ by its variance $\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2})$. Specifically, following Cai et al. [10], we take $\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2}) = n^{-1}\sum_{i=1}^{n}\{(x_{i,j_1}-\bar{x}_{j_1})(x_{i,j_2}-\bar{x}_{j_2})-\hat{\sigma}_{j_1,j_2}\}^2$. Define $M_n^\dagger = \max_{1\le j_1\neq j_2\le p}|\hat{\sigma}_{j_1,j_2}|/\{\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2})\}^{1/2}$ and we take $\mathcal{U}(\infty)=M_n^\dagger$. Theoretically, we prove that Theorem 2.3 still holds with $\mathcal{U}(\infty)=M_n^\dagger$ in Supplementary Material Section B.11. Numerically, we provide the simulations in Supplementary Material Section C.2, which shows that M_n^* in (2.8) generally has higher power than $M_n^\dagger$.*

To apply hypothesis testing using the asymptotic results in Theorems 2.1 and 2.3, we need to estimate $\mathrm{var}\{\mathcal{U}(a)\}$. In particular, we propose the following moment estimator of (2.7):

$$\mathbb{V}_u(a) = \frac{2a!}{(P_a^n)^2}\sum_{1\le j_1\neq j_2\le p}\ \sum_{1\le i_1\neq\ldots\neq i_a\le n}\prod_{t=1}^{a}(x_{i_t,j_1}-\bar{x}_{j_1})^2(x_{i_t,j_2}-\bar{x}_{j_2})^2. \tag{2.9}$$

The next result establishes the statistical consistency of $\mathbb{V}_u(a)$.

CONDITION 2.4. *For integer a, $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^{8a}<\infty$.*

THEOREM 2.4. *Under H_0 in (2.1), assume Conditions 2.1, 2.2 and 2.4 hold. Then $\mathbb{V}_u(a)/\mathrm{var}\{\mathcal{U}(a)\}\xrightarrow{P}1$.*

Theorem 2.4 implies that the asymptotic results in Theorems 2.1 and 2.3 still hold by replacing $\mathrm{var}\{\mathcal{U}(a)\}$ with its estimator $\mathbb{V}_u(a)$. Specifically, under H_0, $[\mathcal{U}(a_1)/\sqrt{\mathbb{V}_u(a_1)},\ldots,\mathcal{U}(a_m)/\sqrt{\mathbb{V}_u(a_m)}]^\intercal \xrightarrow{D}\mathcal{N}(0,I_m)$ under Conditions 2.1, 2.2 and 2.4. Moreover, Theorem 2.3 implies that $\{\mathcal{U}(a)/\sqrt{\mathbb{V}_u(a)}\}$'s are asymptotically independent with $\mathcal{U}(\infty)$.

2.2. *Power Analysis.* In this section, we analyze the asymptotic power of the U-statistics. The power of $\mathcal{U}(2)$ has been studied in the literature. In particular, [12] studied the hypothesis testing of a high-dimensional covariance matrix with $H_0:\mathbf{\Sigma}=I_p$. The authors characterized the boundary that distinguishes the testable region from the non-testable region in terms of the Frobenius norm $\|\mathbf{\Sigma}-I_p\|_F$, and showed that the test statistic proposed by [15, 12], which corresponds to $\mathcal{U}(2)$ in this paper, is rate optimal over their considered regime. However in practice, $\mathcal{U}(2)$ may be not powerful if the alternative covariance matrix is sparse with a small $\|\mathbf{\Sigma} - I_p\|_F$. When the alternative covariance has different sparsity levels, it is of interest to further examine which $\mathcal{U}(a)$ achieves the best power performance among the constructed family of U-statistics.

To study the test power, we establish the limiting distributions of $\mathcal{U}(a)$'s under the alternative hypothesis $H_A : \mathbf{\Sigma} = \mathbf{\Sigma}_A$, where the alternative covariance matrix $\mathbf{\Sigma}_A = (\sigma_{j_1,j_2})_{p\times p}$ is specified in the following Condition 2.5. Define $J_A = \{(j_1, j_2) : \sigma_{j_1,j_2} \neq 0, 1 \leq j_1 \neq j_2 \leq p\}$, which indicates the nonzero off-diagonal entries in $\mathbf{\Sigma}_A$. The cardinality of J_A, denoted by $|J_A|$, then represents the sparsity level of $\mathbf{\Sigma}_A$.

CONDITION 2.5. *Assume $|J_A| = o(p^2)$ and for $(j_1, j_2) \in J_A$, $|\sigma_{j_1,j_2}| = \Theta(\rho)$, where $\rho = \sum_{(j_1,j_2)\in J_A} |\sigma_{j_1,j_2}|/|J_A|$.*

Here ρ represents the average signal strength of $\mathbf{\Sigma}_A$. In our following power comparison of two U-statistics $\mathcal{U}(a)$ and $\mathcal{U}(b)$, we say $\mathcal{U}(a)$ is "better" than $\mathcal{U}(b)$, if, under the same test power, $\mathcal{U}(a)$ can detect a smaller average signal strength ρ (please see the specific definition in Criterion 1 on Page 13). Condition 2.5 specifies a general family of "local" alternatives, which include banded covariance matrices, block covariance matrices, and sparse covariance matrices whose nonzero entries are randomly located.

THEOREM 2.5. *Suppose Conditions 2.1, 2.5, and A.1 (an analogous condition to Condition 2.2* under H_A) in the Supplementary Material hold. For $\mathcal{U}(a)$ in (2.3) and finite integers $\{a_1, \dots, a_m\}$, if $\rho = O(|J_A|^{-1/a_t} p^{1/a_t} n^{-1/2})$ for $t = 1, \dots, m$, then as $n, p \to \infty$,*

$$\Big[\frac{\mathcal{U}(a_1) - \mathrm{E}[\mathcal{U}(a_1)]}{\sigma(a_1)}, \dots, \frac{\mathcal{U}(a_m) - \mathrm{E}[\mathcal{U}(a_m)]}{\sigma(a_m)}\Big]^{\intercal} \xrightarrow{D} \mathcal{N}(0, I_m),$$

where for $a \in \{a_1, \dots, a_m\}$, $\mathrm{E}[\mathcal{U}(a)] = \sum_{(j_1,j_2)\in J_A} \sigma^a_{j_1,j_2}$ and $\sigma^2(a) = \mathrm{var}[\mathcal{U}(a)] \simeq 2a!\kappa_1^a n^{-a} \sum_{1\leq j_1\neq j_2\leq p} \sigma^a_{j_1,j_1}\sigma^a_{j_2,j_2}$, which is of order $\Theta(p^2 n^{-a})$.

Theorem 2.5 shows that for a single U-statistic $\mathcal{U}(a)$ of finite order a,

$$P\bigg(\frac{\mathcal{U}(a)}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}} > z_{1-\alpha}\bigg) \to 1 - \Phi\bigg(z_{1-\alpha} - \frac{\mathrm{E}[\mathcal{U}(a)]}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}}\bigg), \tag{2.10}$$

where $z_{1-\alpha}$ is the upper α quantile of $\mathcal{N}(0, 1)$ and $\Phi(\cdot)$ is the cumulative distribution function of $\mathcal{N}(0, 1)$. By Theorem 2.5, the asymptotic power of

$\mathcal{U}(a)$ of the one-sided test depends on

$$\frac{\mathrm{E}[\mathcal{U}(a)]}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}} \simeq \frac{\sum_{(j_1,j_2)\in J_A} \sigma_{j_1,j_2}^a}{\{2a!\kappa_1^a n^{-a}\sum_{1\le j_1\neq j_2\le p}(\sigma_{j_1,j_1}\sigma_{j_2,j_2})^a\}^{1/2}}. \tag{2.11}$$

By Theorem 2.5, (2.11) $= \Theta(|J_A|\rho^a p^{-1} n^{a/2})$. It follows that when $\mathrm{E}[\mathcal{U}(a)]$ is of the same order of $\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, i.e., $\mathrm{E}[\mathcal{U}(a)] = O(1)\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, the constraint of ρ in Theorem 2.5 is satisfied.

In the following power analysis, we will first compare $\mathcal{U}(a)$'s of finite a and then compare them with $\mathcal{U}(\infty)$. As we focus on studying the relationship between the sparsity level and power, we consider an ideal case where $\sigma_{j_1,j_2} = \rho > 0$ for $(j_1, j_2) \in J_A$ and $\sigma_{j,j} = \nu^2 > 0$ for $j = 1, \dots, p$. Then

$$(2.11) \simeq |J_A|\rho^a / (\sqrt{2a!\kappa_1^a}\nu^{2a} p n^{-a/2}). \tag{2.12}$$

We next show how the order of the "best" U-statistics changes when the sparsity level $|J_A|$ varies. To be specific of the meaning of "best", we compare the ρ values needed by different U-statistics to achieve the same asymptotic power. Particularly, we fix $\mathrm{E}[\mathcal{U}(a)]/\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, i.e., (2.12) to be some constant $M/\sqrt{2}$ for different a's and the asymptotic power of each $\mathcal{U}(a)$ is (2.10) $= 1 - \Phi(z_{1-\alpha} - M/\sqrt{2})$. Then by (2.12), the ρ value such that $\mathcal{U}(a)$ attains the power above is

$$\rho_a = \sqrt{\kappa_1}(a!)^{\frac{1}{2a}}\nu^2(Mp/|J_A|)^{\frac{1}{a}} n^{-\frac{1}{2}}. \tag{2.13}$$

By the definition in (2.13), we compare the power of two U-statistics $\mathcal{U}(a)$ and $\mathcal{U}(b)$ with $a \neq b$ following the Criterion 1 below.

CRITERION 1. *We say $\mathcal{U}(a)$ is "better" than $\mathcal{U}(b)$ if $\rho_a < \rho_b$.*

Given values of $n, p, |J_A|$ and M, (2.13) is a function of a. Therefore, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the smallest ρ_a value in (2.13). We then have the following proposition discussing the optimality among the U-statistics of finite orders in (2.3).

PROPOSITION 2.3. *Given $n, p, |J_A|$ and any constant $M \in (0, +\infty)$, we consider ρ_a in (2.13) as a function of integer a, then*

(i) when $|J_A| \ge Mp$, the minimum of ρ_a is achieved at $a_0 = 1$;
(ii) when $|J_A| < Mp$, the minimum of ρ_a is achieved at some a_0, which increases as $Mp/|J_A|$ increases.

By Proposition 2.3, the order a_0 that attains the smallest value of ρ_a depends on the value of $Mp/|J_A|$ and does not have a closed form solution. We use numerical plots to demonstrate the relationship between a_0 and the sparsity level. Particularly, let $|J_A| = p^{2(1-\beta)}$, where $\beta \in (0,1)$ denotes the sparsity level. To have a better visualization, we use $g(a) = \log(\rho_a n^{1/2}\kappa_1^{-1/2}\nu^{-2}) = (1/2a)\log a! + a^{-1}\log(Mp^{2\beta-1})$ instead of ρ_a. We plot $g(a)$ curves in Figure 1 for each $\beta \in \{0.1,\ldots,0.9\}$ with $M = 4$ and $p \in \{100, 10000\}$. Other values of M and p are also taken, which give similar patterns to Figure 1 and are not presented.

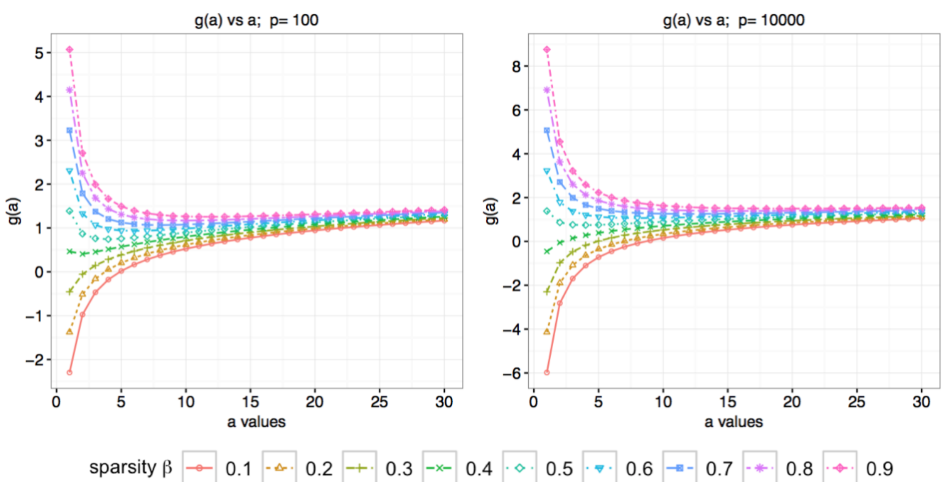

Fig 1: $g(a)$ versus a with different sparsity level β for $p = 100, 10000$

Figure 1 shows that the a_0 such that $g(a)$ attains the smallest value increases when the sparsity level β increases. In particular, when the sparsity level $\beta \le 0.3$, that is, when $|J_A|$ is "very" large and then $\mathbf{\Sigma}_A$ is "very" dense, $g(a)$ has the smallest value at $a_0 = 1$. This is consistent with the conclusion in Proposition 2.3 (i). When the sparsity level β is between 0.4 and 0.5, we note that $a_0 = 2$ achieves the minimum of $g(a)$. This shows that when $|J_A|$ is "moderately" large and $\mathbf{\Sigma}_A$ is "moderately" dense, $\mathcal{U}(2)$ is more powerful than $\mathcal{U}(1)$. When the sparsity level $\beta > 0.5$, we find that $a_0 > 2$. This implies that when $|J_A|$ becomes smaller and $\mathbf{\Sigma}_A$ becomes sparser, U-statistics of higher orders are more powerful. Additionally, we note that a_0 increases slowly as β increases, which verifies Proposition 2.3 (ii). Moreover, the curves converge as a increases and the differences of $g(a)$ for large a values ($a \ge 6$) are small. This implies that when selecting the range of considered orders of U-statistics, it suffices to select an upper bound with

$a = 6$ or 8, which gives better or similar ρ_a values to those larger a's.

In summary, when $|J_A|$ is large, i.e., $\mathbf{\Sigma}_A$ is dense, a small a tends to obtain a smaller lower bound in terms of ρ. But when $|J_A|$ decreases, i.e., $\mathbf{\Sigma}_A$ becomes sparse, a U-statistic of large finite order (or the maximum-type U-statistic as shown next) tends to obtain a smaller lower bound in ρ. This observation is consistent with the existing literature [15, 9, 12, 8].

Next, we proceed to examine the power of the maximum-type test statistic $\mathcal{U}(\infty)$, and compare it with the U-statistics $\mathcal{U}(a)$ of finite a defined in (2.3). By [9], the rejection region for $\mathcal{U}(\infty)$ with significance level α is

$$|\mathcal{U}(\infty)| \geq t_p := n^{-1/2}\sqrt{4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}}.$$

Note $t_p \simeq 2\sqrt{\log p/n}$ and under alternative, the power for $\mathcal{U}(\infty)$ is

$$P(|\mathcal{U}(\infty)| \geq t_p). \tag{2.14}$$

As discussed, we consider the alternatives satisfying Conditions 2.2* and 2.5, $\sigma_{j_1,j_2} = \rho > 0$ for $(j_1, j_2) \in J_A$, and $\sigma_{j,j} = \nu^2$ for $j = 1, \dots, p$. For simplicity, we assume $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$ and ν^2 are given, and focus on the simplified

$$\mathcal{U}(\infty) = \max_{1\leq j_1<j_2\leq p}\left|\nu^{-2}n^{-1}\sum\nolimits_{i=1}^{n}(x_{i,j_1}-\mu_{j_1})(x_{i,j_2}-\mu_{j_2})\right|. \tag{2.15}$$

We show in the following proposition when the power of $\mathcal{U}(\infty)$ asymptotically converges to 1 or is strictly smaller than 1 under alternative.

PROPOSITION 2.4. *Under the considered alternative $\mathbf{\Sigma}_A$ above, suppose $\max_{j=1,\dots,p}\mathrm{E}e^{t_0|x_j-\mu_j|^\varsigma} < \infty$ for some $0 < \varsigma \leq 2$ and $t_0 > 0$, and $\log p = o(n^\beta)$ with $\beta = \varsigma/(4+\varsigma)$. Then for (2.15), when $n, p \to \infty$,*

(i) there exists a constant $c_1 > 2$ such that if $\rho \geq c_1\sqrt{\log p/n}$, (2.14) $\to 1$;
(ii) there exists another constant $0 < c_2 < 2$ such that when $\rho \leq c_2\sqrt{\log p/n}$, Condition 2.2 holds for $\kappa_1 \leq 1$ and $|J_A| = o(1)p^{\frac{2(1-c_2/2)^2}{\kappa_1+m}}(\log p)^{\frac{1}{2}-\frac{1}{2(\kappa_1+m)}}$ for some $m > 0$, we have (2.14) $\leq \log(1-\alpha)^{-1}$.*

Recall that Proposition 2.3 shows that there exists a finite integer a_0, such that ρ_{a_0} is the minimum of (2.13), and ρ_{a_0} is a lower bound of ρ value for the finite-order U-statistics to achieve the given asymptotic power. With Propositions 2.3 and 2.4, we next compare the finite-order U-statistics defined in (2.3) with the maximum-type test statistic $\mathcal{U}(\infty)$.

PROPOSITION 2.5. *Under the conditions of Theorem 2.5 and Proposition 2.4, for any finite integer a, there exist constants c_1 and c_2 such that when p is sufficiently large,*

(i) *For any M, when $|J_A| < c_1^{-a}(a!)^{\frac{1}{2}}\kappa_1^{\frac{a}{2}}(\log p)^{-\frac{a}{2}}Mp$, $\mathcal{U}(\infty)$ has higher asymptotic power than $\mathcal{U}(a)$.*

(ii) *When M is big enough and $|J_A| > c_2^{-a}(a!)^{\frac{1}{2}}\kappa_1^{\frac{a}{2}}(\log p)^{-\frac{a}{2}}Mp$, $\mathcal{U}(a)$ has higher asymptotic power than $\mathcal{U}(\infty)$.*

From Proposition 2.3, we know when $Mp/|J_A| = O(1)$, there exists a finite a_0 such that $\mathcal{U}(a_0)$ is the "best" among all the finite-order U-statistics; in this case, Proposition 2.5 (ii) further indicates that $\mathcal{U}(a_0)$ has higher asymptotic power than $\mathcal{U}(\infty)$. Specifically, if $Mp/|J_A| < 1$, $a_0 = 1$, then $\mathcal{U}(1)$ is the "best" and its lowest detectable order of ρ is $\Theta(p|J_A|^{-1}n^{-1/2})$. More interestingly, when $\mathbf{\Sigma}_A$ is moderately dense or moderately sparse with $Mp/|J_A| > 1$ and bounded, some U-statistic of finite order $a_0 > 1$ would become the "best". By Figure 1, the value of a_0 increases as $\mathbf{\Sigma}_A$ becomes denser. On the other hand, when $\mathbf{\Sigma}_A$ is "very" sparse with $|J_A| < c_1^{-a_0}(a_0!)^{\frac{1}{2}}\kappa_1^{\frac{a_0}{2}}(\log p)^{-\frac{a_0}{2}}Mp$, $\mathcal{U}(\infty)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{\log p/n})$.

Remark 2.5. *The above power comparison results are under the constructed family of U-statistics. We note that additional formulation may further enhance the test power. For instance, [13**?**] showed that an adaptive thresholding in certain ℓ_p-type test statistics can achieve high power under the alternatives with sparse and faint signals. It is of interest to incorporate the adaptive thresholding into the constructed family of U-statistics, which is left for future study.*

Remark 2.6. *The analysis above focuses on the ideal case where the nonzero off-diagonal entries of $\mathbf{\Sigma}_A$ are the same for illustration. When these entries of $\mathbf{\Sigma}_A$ are different, similar analysis still applies by Theorem 2.5 for general covariance matrices. In specific, the asymptotic power of $\mathcal{U}(a)$ depends on the mean variance ratio (2.11) and $\rho_a = \sqrt{\kappa_1}n^{-1/2}(a!)^{1/2a} \times (M\sum_{j=1}^{p}\sigma_{j,j}^a/\sum_{1\le j_1,j_2\le p}\sigma_{j_1,j_2}^a)^{1/a}$. We can then obtain conclusions similar to Propositions 2.3–2.5. One interesting case is when $\mathbf{\Sigma}_A$ contains both positive and negative entries; the same analysis applies for even-order U-statistics, since σ_{j_1,j_2}^a's are all non-negative for even a. On the other hand, the odd-order U-statistics would have low power, since $\sum_{1\le j_1\ne j_2\le p}\sigma_{j_1,j_2}^a$ could be small due to the cancellation of positive and negative σ_{j_1,j_2}^a's. We have conducted simulations when the nonzero σ_{j_1,j_2}'s are different in Section 3.1, and the results exhibit consistent patterns as expected.*

2.3. *Application to Adaptive Testing & Computation.*

Adaptive Testing. Power analysis in Section 2.2 shows that when the sparsity level of the alternative changes, the test statistic that achieves the highest power could vary. However, since the truth is often unknown in practice, it is unclear which test statistic should be chosen. Therefore, we develop an adaptive testing procedure by combining the information from U-statistics of different orders, which would yield high power against various alternatives.

In particular, we propose to combine the U-statistics through their p-values, which is widely used in literature [58, 61, 80]. One popular method is the minimum combination, whose idea is to take the minimum p-value to approximate the maximum power [61, 80, 76]. Specifically, let Γ be a candidate set of the orders of U-statistics, which contains both finite values and ∞. We compute p-values p_a's of the U-statistics $\mathcal{U}(a)$'s satisfying $a \in \Gamma$. The minimum combination takes the statistic $T_{\text{adpUmin}} = \min\{p_a : a \in \Gamma\}$ and has the asymptotic p-value $p_{\text{adpUmin}} = 1 - (1 - T_{\text{adpUmin}})^{|\Gamma|}$, where $|\Gamma|$ denotes the size of the candidate set Γ. We reject H_0 if $p_{\text{adpUmin}} < \alpha$. Under H_0, p_a's are asymptotically independent and uniformly distributed by the theoretical results in Section 2.1. The type I error is asymptotically controlled as $P(p_{\text{adpUmin}} < \alpha) = P(\min_{a\in\Gamma} p_a < p_\alpha^*) \to \alpha$, where $p_\alpha^* = 1 - (1-\alpha)^{1/|\Gamma|}$. Since $P(\min_{a\in\Gamma} p_a < p_\alpha^*) \geq P(p_a < p_\alpha^*)$, the power of the adaptive test goes to 1 if there exists $a \in \Gamma$ such that the power of $\mathcal{U}(a)$ goes to 1. We note that the power of the adaptive test is not necessarily higher than that of all the U-statistics. This is because the power of $\mathcal{U}(a)$ is $P(p_a < \alpha)$, and is different from $P(p_a < p_\alpha^*)$ since $p_\alpha^* < \alpha$ when $|\Gamma| > 1$. Based on our extensive simulations, we find that the adaptive test is usually close to or even higher than the maximum power of the U-statistics.

REMARK 2.7. *Fisher's method [58] is another popular method for combining independent p-values. It has the test statistic* $T_{\text{adpUf}} = -2\sum_{k=1}^{|\Gamma|} \log p_k$, *which converges to* $\chi^2_{2|\Gamma|}$ *under* H_0. *By our simulations, the minimum combination and Fisher's method are generally comparable, while Fisher's method has higher power under several cases. Moreover, we can also use other methods to combine the p-values, such as higher criticism [18, 19]. We leave the study of how to efficiently combine the p-values for future research.*

We select the candidate set Γ by the power analysis in Section 2.2. We would recommend including $\{1, 2, \ldots, 6, \infty\}$, which can be powerful against a wide spectrum of alternatives. In particular, by Propositions 2.3 and 2.5, we include $a = 1, 2$ that are powerful against dense signals; $a = \infty$ that is powerful against sparse signals; and also $a = \{3, \ldots, 6\}$ for the moderately dense and moderately sparse signals. By Figure 1, it generally suffices to choose finite a up to 6–8, which often give similar/better performance to/than larger a values. The simulations in Section 3.1 confirm the good performance of this choice of Γ; and the proposed adaptive test appears to well approximate the "best" performance even when Γ may not always contain the unknown "optimal" U-statistics.

We would like to mention that the adaptive procedure can be generalized to other testing problems, as long as similar theoretical properties are given, such as the examples in Section 4.

Computation. Next we discuss the computation in the adaptive testing. A direct calculation following the form of $\mathcal{U}(a)$ in (2.3) and $\mathbb{V}(a)$ in (2.9) would be computationally expensive for large a with a cost of $O(p^2 n^{2a})$. To address this issue, we introduce a method that can reduce the cost.

We first consider a simplified setting when $\mathrm{E}(x_{i,j}) = 0$ to illustrate the idea. As discussed in Remark 2.2, we examine $\widetilde{\mathcal{U}}(a)$ defined in (2.5). Let $\mathcal{L} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$ denote the set of index tuples, and for each index tuple $l = (j_1, j_2) \in \mathcal{L}$, define $s_{i,l} = x_{i,j_1} x_{i,j_2}$. Note that $\widetilde{\mathcal{U}}(a) = (P_a^n)^{-1} \sum_{l \in \mathcal{L}} \mathcal{U}_l(a)$, where $\mathcal{U}_l(a) = \sum_{1 \leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} s_{i_k,l}$. Calculating $\mathcal{U}_l(a)$ directly is of order $O(n^a)$. We then focus on reducing the computational cost of $\mathcal{U}_l(a)$. For $l \in \mathcal{L}$ and finite integers $t_1, \ldots, t_k$, define

$$V_l^{(t_1,\ldots,t_k)} = \prod_{r=1}^{k} \Big(\sum_{i=1}^{n} s_{i,l}^{t_r} \Big), \quad U_l^{(t_1,\cdots,t_k)} = \sum_{1 \leq i_1 \neq \ldots \neq i_k \leq n} \prod_{r=1}^{k} s_{i_1,l}^{t_r}. \tag{2.16}$$

We can see that $\mathcal{U}_l(a) = U_l^{\mathbf{1}_a}$ with $\mathbf{1}_a$ being an a-dimensional vector of all ones, and $U_l^{(a)} = V_l^{(a)}$ for any finite integer a. To reduce the computational cost of $\mathcal{U}_l(a)$, the main idea is to obtain $U_l^{\mathbf{1}_a}$ from $V_l^{(t_1,\ldots,t_k)}$, whose computational cost is $O(n)$. In particular, $\mathcal{U}_l(a)$ can be attained iteratively from $V_l^{(t_1,\ldots,t_k)}$ based on the following equation

$$U_l^{(k,\mathbf{1}_{r-k})} = V_l^{(k)} \times U_l^{\mathbf{1}_{r-k}} - (r-k) \times U_l^{(k+1,\mathbf{1}_{r-k-1})}, \tag{2.17}$$

which follows from the definitions. Algorithm 1 below summarizes the steps.

We illustrate the idea of the algorithm by some examples. By definition, $U_l^{(1)} = V_l^{(1)}$, which can be computed with cost $O(n)$. Next consider in (2.17), if $r = 2$ and $k = 1$, then $U_l^{(1,1)} = V_l^{(1)} \times U_l^{(1)} - (2-1) \times U_l^{(2)} = V_l^{(1)} \times V_l^{(1)} - V_l^{(2)}$, which yields $U_l^{\mathbf{1}_2}$ with cost $O(n)$. For $U_l^{\mathbf{1}_3}$, we first take $r = 3$ and $k = 2$ in (2.17), then with cost $O(n)$, we have $U_l^{(2,1)} = V_l^{(2)} \times U_l^{(1)} - U_l^{(3)} = V_l^{(2)} \times V_l^{(1)} - V_l^{(3)}$, as $V_l^{(k)} = U_l^{(k)}$ by the definition. Given $U_l^{\mathbf{1}_2}$ and $U_l^{(2,1)}$, we obtain $U_l^{(1,\mathbf{1}_2)} = V_l^{(1)} \times U_l^{\mathbf{1}_2} - 2 \times U_l^{(2,\mathbf{1}_1)}$. Thus $U_l^{\mathbf{1}_3}$ is also computed with cost

Data: $s_{i,l}$ $(1 \le i \le n,\, l \in \mathcal{L})$.
Result: $\tilde{\mathcal{U}}(a)$.
for $l \in \mathcal{L}$ **do**
  Compute and store $V_l^{(k)} = U_l^{(k)} = \sum_{i=1}^n s_{i,l}^k$, $(k = 1, \cdots, a)$ during the algorithm;
  $U_l^{\mathbf{1}_1} = V_l^{(1)}$, $U_l^{\mathbf{1}_2} = U_l^{\mathbf{1}_1} V_l^{(1)} - U_l^{(2)}$;
  while $3 \le r \le a$ **do**
    $T_l = U_l^{(r)}$
    for $k \leftarrow r-1$ **to** 1 **do**
      $T_l = V_l^{(k)} \times U_l^{\mathbf{1}_{r-k}} - (r-k) \times T_l$
    end
    $U_l^{\mathbf{1}_r} = T_l$
  end
end
$\tilde{\mathcal{U}}(a) = (P_a^n)^{-1} \sum_{l \in \mathcal{L}} U_l^{\mathbf{1}_a}$

Algorithm 1: Iterative Computation Implementation

$O(n)$. Iteratively, for any finite integer a, we can obtain $U_l^{\mathbf{1}_a}$ from $V_l^{(t_1,\dots,t_k)}$ whose computational cost is $O(n)$. More closed form formulae representing $U_l^{\mathbf{1}_a}$ by $V_l^{(t_1,\dots,t_k)}$ are given in Section C.1.1 of Supplementary Material.

Algorithm 1 reduces the computational cost of $\tilde{\mathcal{U}}(a)$ from $O(p^2 n^a)$ to $O(p^2 n)$. Its idea is general and can be extended to compute other different U-statistics by changing the input $s_{i,l}$. In particular, the variance estimator $\mathbb{V}(a)$ can be computed with cost $O(p^2 n)$ by specifying $s_{i,l} = (x_{i,j_1} - \bar{x}_{j_1})^2 (x_{i,j_2} - \bar{x}_{j_2})^2$, for each $l \in \mathcal{L} = \{(j_1, j_2) : 1 \le j_1 \ne j_2 \le p\}$. Then $\mathbb{V}(a) = 2a!(P_a^n)^{-2} \sum_{l \in \mathcal{L}} \sum_{1 \le i_1 \ne \ldots \ne i_a \le n} \prod_{k=1}^a s_{i_k,l}$ and the Algorithm 1 can be applied. Moreover, when $\mathrm{E}(x_{i,j})$ is unknown, $\mathcal{U}(a)$ can still be computed with cost $O(p^2 n)$ using the iterative method similar to Algorithm 1. The details are provided in Section C.1.2 of Supplementary Material.

3. Simulations and Real Data Analysis.

3.1. *Simulations.* We conduct simulation studies to evaluate the performance of the proposed adaptive testing procedures, and investigate the relationship between the power and sparsity levels. For one-sample covariance testing discussed in Section 2, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \dots, n$, and consider the following five simulation settings.

Setting 1: $\mathbf{x}_i$ has p i.i.d. entries of $\mathcal{N}(0, 1)$ and Gamma$(2, 0.5)$ respectively. Under each case, we take $n = 100$ and $p \in \{50, 100, 200, 400, 600, 800, 1000\}$ to verify the theoretical results under H_0 and the validity of the adaptive test across different n and p combinations.

For the following settings 2–5, we generate $\mathbf{x}_i$ from multivariate Gaussian distributions with mean zero and different covariance matrices $\mathbf{\Sigma}_A$'s.

Setting 2: $\mathbf{\Sigma}_A = (1-\rho)I_p + \rho \mathbf{1}_{p,k_0}\mathbf{1}_{p,k_0}^\intercal$, where $\mathbf{1}_{p,k_0}$ is a p-dimensional vector with the first k_0 elements one and the rest zero. We take $(n,p) \in \{(100,300),(100,600),(100,1000)\}$, and study the power with respect to different signal sizes ρ and sparsity levels k_0.

Setting 3: The diagonal elements of $\mathbf{\Sigma}_A$ are all one and $|J_A|$ number of off-diagonal elements are ρ with random positions. We take $(n,p) \in \{(100,600),(100,1000)\}$ and let the signal size ρ and sparsity level $|J_A|$ vary to examine how the power changes accordingly.

Setting 4: The diagonal elements of $\mathbf{\Sigma}_A$ are all one and $|J_A|$ number of off-diagonal elements are uniformly generated from $(0, 2\rho)$ with random positions. We take $(n,p) = (100,1000)$ and similarly let the signal size ρ and sparsity level $|J_A|$ vary to examine how the power changes accordingly.

Setting 5: We consider the multivariate models in [15]. Specifically, for each $i = 1,\ldots,n$, $\mathbf{x}_i = \Xi \mathbf{z}_i + \boldsymbol{\mu}$, where Ξ is a matrix of dimension $p \times m$, and $\mathbf{z}_i$'s are i.i.d. Gaussian or Gamma random vectors. Under null hypothesis, $m = p$, $\Xi = I_p$ $\boldsymbol{\mu} = 2\mathbf{1}_p$; under alternative hypothesis, $m = p+1$, $\Xi = (\sqrt{1-\rho}I_p, \sqrt{2\rho}\mathbf{1}_p)$, $\boldsymbol{\mu} = 2(\sqrt{1-\rho}+\sqrt{2\rho})\mathbf{1}_p$. We also take the n and p combination in [15] with $(n,p) \in \{(40,159),(40,331),(80,159),(80,331),(80,642)\}$.

We compare several methods in the literature, including both maximum-type and sum-of-squares-type tests. In particular, the maximum-type test statistic in Jiang [43] is taken as $\mathcal{U}(\infty)$ in this framework. Since the convergence in [43] is known to be slow, we use permutation to approximate the distribution in the simulations. In addition, we consider some sum-of-squares-type methods. Specifically, we examine the identity and sphericity tests in Chen et al. [15], which are denoted as "Equal" and "Spher", respectively. We also compare the methods in Ledoit and Wolf [51] and Schott [66], which are referred to as "LW" and "Schott", respectively.

To illustrate, Figure 2 summarizes the numerical results for the setting 3 when $n = 100$ and $p = 1000$. All the results are based on 1000 simulations at the 5% nominal significance level. In Figure 2, we present the power of single U-statistics with orders in $\{1,\ldots,6,\infty\}$. "adpUmin" and "adpUf" represent the results of the adaptive testing procedure using the minimum combination and Fisher's method in Section 2.2 respectively. The simulation results show that the type I error rates of the U-statistics and adaptive test are well controlled under H_0. In addition, Figure 2 exhibits several patterns that are consistent with the power analysis in Section 2.2. First, it shows that among the U-statistics, when $|J_A|$ is very small, $\mathcal{U}(\infty)$ performs best; and when $|J_A|$ increases, the performances of some U-statistics of finite orders catch up. For instance, when $|J_A| = 100, \mathcal{U}(6)$ and $\mathcal{U}(\infty)$ are similar and are better than the other U-statistics; when $|J_A| = 400$, $\mathcal{U}(4)$ and $\mathcal{U}(5)$ are similar and better than the other U-statistics. When $\mathbf{\Sigma}_A$ is relatively dense, $\mathcal{U}(2)$ and $\mathcal{U}(1)$ become more powerful. Particularly, when $|J_A| = 1600$, $\mathcal{U}(2)$ is powerful; when $|J_A|$ becomes larger, such as when $|J_A| = 3200$, $\mathcal{U}(1)$ is overall the most powerful. Second, Figure 2 shows that "LW", "Schott", "Equal", "Spher" and $\mathcal{U}(2)$ perform similarly under various cases. In particular, these methods are not powerful when the alternative is sparse but becomes more powerful when the alternative gets denser. This is because they are all sum-of-squares-type statistics that target at dense alternatives. Third and importantly, the two adaptive tests "adpUmin" and "adpUf" maintain high power across different settings. Specifically, they perform better than most single U-statistics: their powers are usually close to or even higher than the best single U-statistic. Moreover, "adpUmin" and "adpUf" generally have higher power than the compared existing methods. We also note that "adpUf" overall performs better than "adpUmin" in this simulation setting. In summary, Figure 2 demonstrates the relationship between the sparsity levels of alternatives and the power of the tests, confirming the theoretical conclusions in Section 2.2. Notably, the proposed adaptive testing procedure is powerful against a wide range of alternatives, and thus advantageous in practice when the true alternative is unknown.

Due to the space limitation, we provide other extensive numerical studies in Supplementary Material Section C.2. The conclusions are similar to those of Figure 2, and consistent with the theoretical results in Section 2.2. In particular, the results show that the empirical sizes of the tests are close to the nominal level, suggesting the good finite-sample performance of the asymptotic approximations. Moreover, under highly dense alternatives with only non-negative entries in the covariance matrix, $\mathcal{U}(1)$ is the most powerful one among the $\mathcal{U}(a)$'s and the other tests in [51, 66, 15], in agreement with the results in Propositions 2.3 and 2.5. Furthermore, the proposed adaptive testing procedures often have higher power than most single U-statistics.

3.2. *Real Data Analysis.* Alzheimer's disease (AD) is the most prevalent neurodegenerative disease [65] and is ranked as the sixth leading cause of death in the US [77]. Every 65 seconds, someone in the US develops AD [1]. To advance our understanding of AD, the Alzheimer's Disease Neuroimaging Initiative (ADNI) was started in 2004, collecting extensive genetic data for both healthy individuals and AD patients. To gain insight into the genetic mechanisms of AD, one can test a single SNP a time. However, due to a relatively small sample size of the ADNI data, scanning across all SNPs failed to

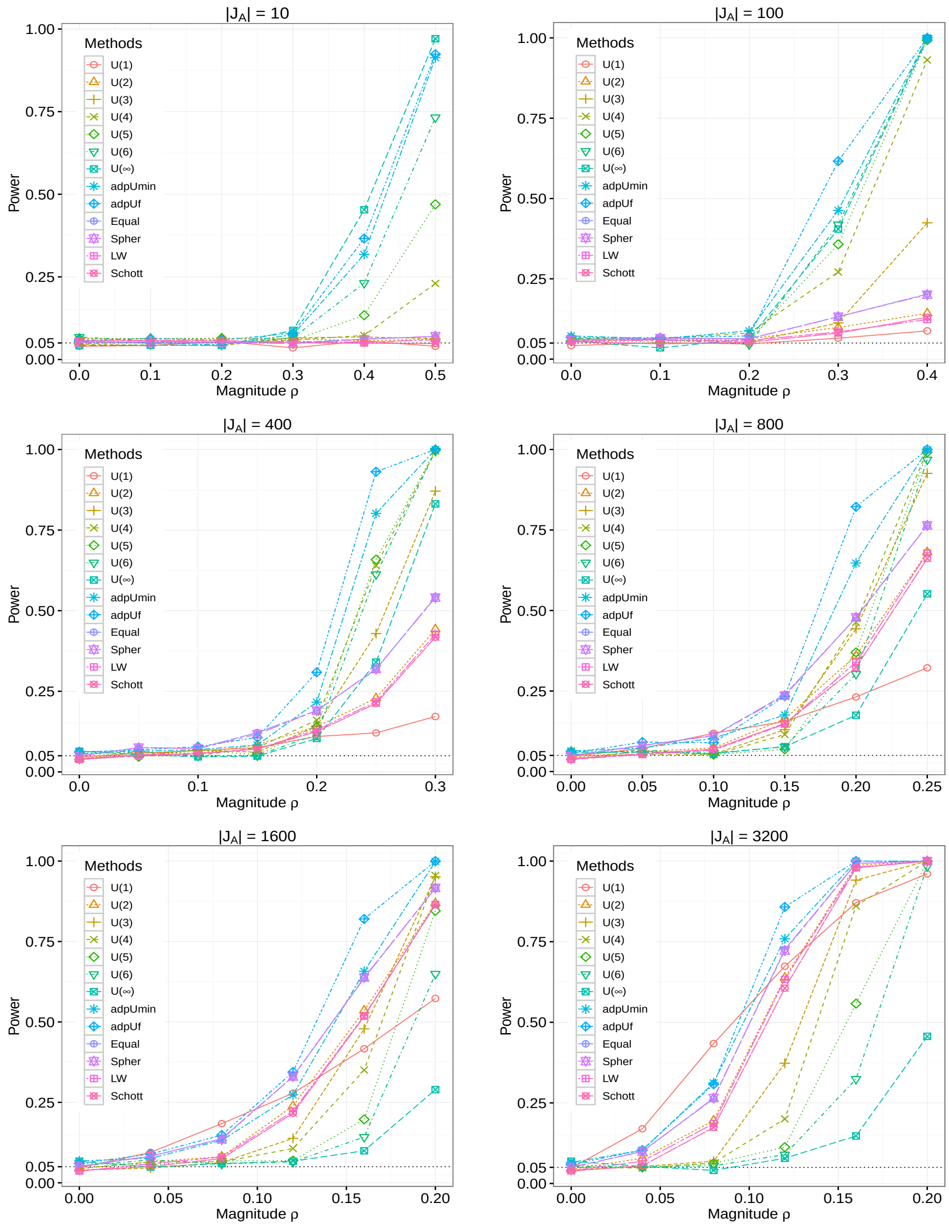

Fig 2: Power comparison.

identify any genome-wide significant SNP (with p-value $< 5 \times 10^{-8}$)[48]. To date, the largest meta-analysis of more than 600,000 individuals identified 29 significant risk loci [42] and can only explain a small proportion of AD variance. On the other hand, a group of functionally related genes as annotated in a biological pathway are often involved in the same disease susceptibility and progression [33]. Thus, pathway-based analyses, which jointly analyze a group of SNPs in a biological pathway, have become increasingly popular. We retrieve a total of 214 pathways from the KEGG database [47] for the subsequent analysis.

Although pathway-based analyses with KEGG pathways are common in real studies, formally testing the correlations of the genes in a KEGG pathway has been largely untouched. Here, we apply our method and other competing methods in [15] to test if all the genes in a pathway have correlated gene expression levels. Perhaps as expected, all methods reject the null hypothesis for all pathways with highly significant p-values, since the KEGG pathways are constructed to include only the genes with similar function into the same pathway [47], while similar function often implies co-expression (and vice versa). To compare the performance of the different tests, for each pathway we randomly select 50 subjects and restrict our analysis to pathways of at least 50 genes, leading to 103 pathways for the following analysis. Then we perturb the data by shuffling the gene expression levels of randomly selected $100(1-\alpha)\%$ genes in a pathway before applying each test. Figure 3 shows the performance of the tests with two significance cutoffs, where "$\mathcal{U}(2)$" represents the single $\mathcal{U}(2)$ statistic, "adpU" represents our proposed adaptive testing procedure using the minimum combination with candidate U-statistics of orders in $\{1,\dots,6,\infty\}$, and "Equal" and "Spher" represent the identity and sphericity tests in [15] respectively. Because all pathways are highly significant with all samples, we can treat all pathways as the true positives. Due to the adaptiveness of our proposed testing procedure, "adpU" identifies more significant pathways than the competing methods across all the levels of data perturbation (mimicking the varying sparsity levels of the alternatives).

4. Other High-Dimensional Examples. In this section, we apply the proposed U-statistics framework to other high-dimensional testing problems. Similar theoretical results to Section 2 are developed, with detailed proofs and related simulation studies provided in Supplementary Material.

4.1. *Mean Testing.* Testing mean vectors is widely used in many statistical analysis and applications [2, 59]. Under high-dimensional scenarios, e.g., in genome-wide studies, dimension of the data is often much larger than the sample size, so traditional multivariate tests such as Hotelling's T^2-test either cannot be directly applied or have low power [20]. To address this issue, several new procedures for testing high-dimensional mean vectors have been

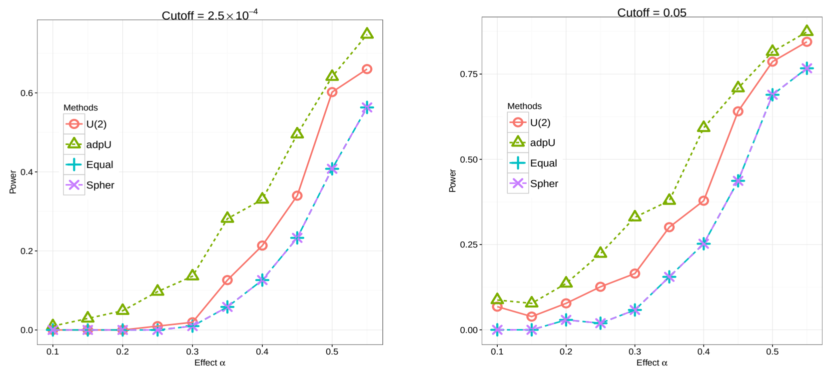

Fig 3: Power comparison of different methods with ADNI data.

proposed [4, 18, 27, 69, 14, 32, 11, 13, 28, 19, 71, 76]. However, many of the statistics only target at either sparse or dense alternatives, and suffer from loss of power for other types of alternatives. We next apply the U-statistics framework to one-sample and two-sample mean testing problems.

One-sample mean testing. We first discuss the one-sample mean vector testing. Assume that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n i.i.d. copies of a p-dimensional real-valued random vector $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$, covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{j_1,j_2} : 1 \leq j_1, j_2 \leq p\}$. We want to conduct the global test on $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0 = (\mu_{1,0}, \ldots, \mu_{p,0})^\intercal$ is given.

Similar to previous discussion, the parameter set that we are interested in is $\mathcal{E} = \{\mu_1 - \mu_{1,0}, \ldots, \mu_p - \mu_{p,0}\}$. For each $j = 1, \ldots, p$, $\mathrm{E}(x_{i,j}) = \mu_j$, so $K_j(\mathbf{x}_i) = x_{i,j} - \mu_{j,0}$ is a kernel function, which is a simple unbiased estimator of the target. Following our construction, the U-statistic for finite a is

$$\mathcal{U}(a) = \sum_{j=1}^{p} \frac{1}{P_a^n} \sum_{1 \leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} (x_{i_k,j} - \mu_{j,0}), \tag{4.1}$$

which targets at $\|\mathcal{E}\|_a^a = \sum_{j=1}^{p} (\mu_j - \mu_{j,0})^a$, and the U-statistic corresponding to $\|\mathcal{E}\|_\infty$ is $\mathcal{U}(\infty) = \max_{1\leq j \leq p} \sigma_{j,j}^{-1} (\bar{x}_j - \mu_{0,j})^2$ with $\bar{x}_j = \sum_{i=1}^{n} x_{i,j}/n$.

Given the statistics, we have the theoretical results similar to Theorems 2.1–2.3. The following Theorems 4.1–4.2 are established under similar conditions to that of Theorems 2.1–2.3. Due to the limited space, we provide the conditions and corresponding discussions in Supplementary Material.

THEOREM 4.1. *Under H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, assume Condition A.2 in Supplementary Material. Then for any finite integers $\{a_1, \ldots, a_m\}$, as $n, p \to \infty$,*

$[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *where* $\sigma^2(a) = \mathrm{var}[\mathcal{U}(a)] = \sum_{i=1}^p \sum_{j=1}^p a!\sigma_{i,j}^a / P_a^n$ *with the order of* $\Theta(a!pn^{-a})$.

THEOREM 4.2. *Under* H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, *assume Condition A.3 in Supplementary Material. Then* $\forall u \in \mathbb{R}$, $P(n\mathcal{U}(\infty) - \tau_p \le u) \to \exp\{-\pi^{-1/2}\exp(-u/2)\}$, *as* $n, p \to \infty$, *where* $\tau_p = 2\log p - \log\log p$. *In addition, for any finite integer* a, $\{\mathcal{U}(a)/\sigma(a)\}$ *and* $\{n\mathcal{U}(\infty) - \tau_p\}$ *are asymptotically independent.*

By Theorems 4.1 and 4.2, we obtain the asymptotic independence among the U-statistics and the corresponding limiting distributions of the U-statistics under H_0. Under the alternative hypothesis, since the power analysis of the one-sample mean testing is similar to that of the two-sample case, we delay the power analysis after presenting the asymptotic independence property of the proposed U-statistics in the two-sample mean testing problem.

Two-sample mean testing. Next we discuss the two-sample mean testing problem. Suppose we have two groups of p-dimensional observations $\{\mathbf{x}_i\}_{i=1}^{n_x}$ and $\{\mathbf{y}_i\}_{i=1}^{n_y}$, which are i.i.d. copies of two independent random vectors $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ and $\mathbf{y} = (y_1, \ldots, y_p)^\intercal$ respectively. Suppose $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$, $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)^\intercal$, $\mathrm{cov}(\mathbf{x}) = \boldsymbol{\Sigma}_x$ and $\mathrm{cov}(\mathbf{y}) = \boldsymbol{\Sigma}_y$. We write $n = n_x + n_y$ and assume $n_x = \Theta(n_y)$. For easy illustration, we first consider $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_y = \boldsymbol{\Sigma} = \{\sigma_{j_1,j_2} : 1 \le j_1, j_2 \le p\}$. We will then discuss the case when $\boldsymbol{\Sigma}_x \ne \boldsymbol{\Sigma}_y$, where similar analysis applies.

The two-sample mean testing examines $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$ versus $H_A : \boldsymbol{\mu} \ne \boldsymbol{\nu}$, then $\mathcal{E} = (\mu_1 - \nu_1, \ldots, \mu_p - \nu_p)^\intercal$. For $1 \le j \le p$, $1 \le k \le n_x$, $1 \le s \le n_y$, $K_j(\mathbf{x}_k, \mathbf{y}_s) = x_{k,j} - y_{s,j}$ is a simple unbiased estimator of $\mu_j - \nu_j$, and thus we construct $\mathcal{U}(a) = \sum_{j=1}^p (P_a^{n_x} P_a^{n_y})^{-1} \sum_{\substack{1 \le k_1 \ne \ldots \ne k_a \le n_x \\ 1 \le s_1 \ne \ldots \ne s_a \le n_y}} \prod_{t=1}^a (x_{k_t,j} - y_{s_t,j})$, which is also equivalent to

$$\mathcal{U}(a) = \sum_{j=1}^p \sum_{c=0}^a \binom{a}{c} \frac{(-1)^{a-c}}{P_c^{n_x} P_{a-c}^{n_y}} \sum_{\substack{1 \le k_1 \ne \cdots \ne k_c \le n_x \\ 1 \le s_1 \ne \cdots \ne s_{a-c} \le n_y}} \prod_{t=1}^c x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j}. \tag{4.2}$$

We can check that (4.2) satisfies $\mathrm{E}\{\mathcal{U}(a)\} = \sum_{j=1}^p (\mu_j - \nu_j)^a$, so $\mathcal{U}(a)$ is an unbiased estimator of $\|\mathcal{E}\|_a^a = \sum_{j=1}^p (\mu_j - \nu_j)^a$. On the other hand, for $\|\mathcal{E}\|_\infty$, following the maximum-type test statistic in Cai et al. [11], we have

$$\mathcal{U}(\infty) = \max_{1 \le j \le p} \sigma_{j,j}^{-1} (\bar{x}_j - \bar{y}_j)^2, \tag{4.3}$$

where $\bar{x}_j = \sum_{i=1}^{n_x} x_{i,j}/n_x$, $\bar{y}_j = \sum_{i=1}^{n_y} y_{i,j}/n_y$. We then obtain results similar to Theorems 2.1, 2.3 and 2.5. As the conditions are similar to those in

Section 2, we only keep the key conclusions, and the details of conditions and discussions are given in Supplementary Material Section A.8.

THEOREM 4.3. *Under Condition A.4 in Supplementary Material,* $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$ *and* $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$*, for any finite integers* $(a_1, \dots, a_m)$*, as* $n, p \to \infty$*,* $[\mathcal{U}(a_1)/\sigma(a_1), \dots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$*, where* $\sigma^2(a) \simeq a! \sum_{j_1,j_2=1}^{p} (n_x + n_y)^a \sigma_{j_1,j_2}^a/(n_x n_y)^a$ *is of the order* $\Theta(a! p n^{-a})$*.*

THEOREM 4.4. *Under Condition A.4 in Supplementary Material,* $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$ *and* $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$*,* $\forall u \in \mathbb{R}$*,* $P(\frac{n_x n_y}{n_x+n_y}\mathcal{U}(\infty) - \tau_p \le u) \to \exp\{-\pi^{-1/2}\exp(-u/2)\}$*, as* $n, p \to \infty$*, where* $\tau_p = 2\log p - \log\log p$*. Moreover,* $\{\mathcal{U}(a)/\sigma(a)\}$ *of finite integer* a *and* $\{n_x n_y \mathcal{U}(\infty)/(n_x+n_y) - \tau_p\}$ *are asymptotically independent.*

Theorems 4.3 and 4.4 provide the asymptotic properties of finite-order U-statistics and $\mathcal{U}(\infty)$ under H_0. To analyze the power of $\mathcal{U}(a)$'s, we derive the asymptotic results of $\mathcal{U}(a)$'s under the alternative hypotheses. We focus on the two-sample mean testing problem, while one-sample mean testing can be obtained similarly. Specifically, we consider the alternative $\mathcal{E}_A = \{\mu_j - \nu_j = \rho > 0 \text{ for } j = 1, \dots, k_0; \mu_j - \nu_j = 0 \text{ for } j = k_0+1, \cdots, p\}$. We then obtain similar conclusions to Theorem 2.5.

THEOREM 4.5. *Assume Condition A.4 in Supplementary Material and* $k_0 = o(p)$*. For any finite integers* $\{a_1, \dots, a_m\}$*, if* ρ *in* $\mathcal{E}_A$ *satisfies* $\rho = O(k_0^{-1/a_t} p^{1/(2a_t)} n^{-1/2})$ *for* $t = 1, \dots, m$*, then* $[\mathcal{U}(a_1) - \mathrm{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1), \dots, [\mathcal{U}(a_m) - \mathrm{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$*, as* $n, p \to \infty$*. Here* $\mathrm{E}[\mathcal{U}(a)] = \|\mathcal{E}_A\|_a^a = k_0\rho^a$ *and* $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \simeq V_a$*, with* $V_a = a! \sum_{j_1,j_2=k_0+1}^{p} (n_x + n_y)^a \sigma_{j_1,j_2}^a/(n_x n_y)^a$ *of the order* $\Theta(a! p n^{-a})$*.*

Next we compare the power of different U-statistics under alternatives with different sparsity levels. Theorem 4.5 shows that under the local alternatives, the asymptotic power of $\mathcal{U}(a)$ mainly depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. Therefore by Theorem 4.5, given constant $M > 0$, for each $\mathcal{U}(a)$, if $\rho = M^{1/a} k_0^{-1/a} V_a^{1/(2a)}$, then $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$; that is, different $\mathcal{U}(a)$'s have the same power asymptotically. For easy illustration, we consider $\sigma_{j_1,j_2} = 1$ when $j_1 = j_2 \in \{k_0+1, \dots, p\}$, and $\sigma_{j_1,j_2} = 0$ when $j_1 \neq j_2 \in \{k_0+1, \dots, p\}$, then $M^{1/a} k_0^{-1/a} V_a^{1/(2a)} \simeq \rho_a$ with

$$\rho_a := a!^{\frac{1}{2a}} (M\sqrt{p}/k_0)^{\frac{1}{a}} \{(n_x+n_y)/(n_x n_y)\}^{\frac{1}{2}}. \tag{4.4}$$

Therefore, similarly to the analysis in Section 2.2, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the minimum ρ_a in (4.4). We have the following result similar to Proposition 2.3.

PROPOSITION 4.1. *Given any constant $M \in (0, +\infty)$ and n, p, k_0, we consider ρ_a in (4.4) as a function of positive integers a, then*

(i) when $k_0 \geq M\sqrt{p}$, the minimum of ρ_a is achieved at $a_0 = 1$;
(ii) when $k_0 < M\sqrt{p}$, the minimum of ρ_a is achieved at some a_0, which increases as $M\sqrt{p}/|J_D|$ increases.

Proposition 4.1 shows that when the sparsity level k_0 is large, i.e., $\mathcal{E}_a$ is dense, a small a tends to obtain a smaller lower bound in ρ, and vice versa. As (4.4) and (2.13) are similar, we have similar patterns to that in Figure 1 when examining the corresponding numerical plots of ρ_a. In addition, [11] shows that when $\rho = \rho_\infty := C_1\sqrt{\log p/n}$ for a large C_1, the power of $\mathcal{U}(\infty)$ converges to 1, and $\sqrt{\log p/n}$ is minimax rate optimal for sparse alternatives; see also [19]. Thus, if $\rho_\infty < \rho_{a_0}$, i.e., $k_0 < MC_1^{-a_0}\sqrt{pa_0!}/\log^{a_0/2} p$, $\mathcal{U}(\infty)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{\log p/n})$. On the other hand, Proposition 4.1 shows that when $\mathcal{E}_A$ is dense with $k_0 > \sqrt{Mp}$, $\mathcal{U}(1)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{p}k_0^{-1}n^{-1/2})$. Moreover, for some large M and C_2, when $\mathcal{E}_A$ is "moderately dense" or "moderately sparse" with $C_2\sqrt{pa_0!}/\log^{a_0/2} p < k_0 < \sqrt{Mp}$, $\mathcal{U}(a_0)$ is the "best" and its lowest detectable order of ρ is $\Theta\{(\sqrt{p}/k_0)^{\frac{1}{a_0}}n^{-1/2}\}$, which is of a smaller order than the optimal detection boundary of the sparse case $\Theta(\sqrt{\log p/n})$.

More generally, when $\mathbf{\Sigma}_x \neq \mathbf{\Sigma}_y$, similar results to Theorems 4.3 and 4.5 can be obtained. In particular, we have the following corollary.

COROLLARY 4.1. *When $\mathbf{\Sigma}_x \neq \mathbf{\Sigma}_y$, under Condition A.4 in Supplementary Material, Theorem 4.3 holds with $\sigma^2(a) \simeq a!\sum_{j_1,j_2=1}^{p}(\sigma_{x,j_1,j_2}/n_x + \sigma_{y,j_1,j_2}/n_y)^a$ and Theorem 4.5 holds with $V_a = a!\sum_{j_1,j_2=k_0+1}^{p}(\sigma_{x,j_1,j_2}/n_x + \sigma_{y,j_1,j_2}/n_y)^a$.*

Corollary 4.1 shows that the asymptotic power of finite-order U-statistics depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. By the construction of finite-order U-statistics and the proof, we obtain that $\mathrm{E}\{\mathcal{U}(a)\} = k_0\rho^a$ and $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(a!pn^{-a})$. We then know that for finite-order U-statistics, similar results to Proposition 4.1 still hold by examining $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$.

The above power analysis shows that the optimal U-statistic varies when the alternative hypothesis changes. To achieve high power across various alternatives, we can develop an adaptive test similar to that in Section 2.3. Specifically, we calculate the p-values of the U-statistics (4.1) and (4.2) following the theoretical results above and the algorithm in Section 2.3. By combining the p-values as discussed in Section 2.3, the asymptotic power of the adaptive test goes to 1 if there exists one $\mathcal{U}(a)$ whose power goes to 1.

Remark 4.1. *Xu et al. [76] has also discussed the adaptive testing of two-sample mean that is powerful against various ℓ_p-norm-like sums of $\boldsymbol{\mu}-\boldsymbol{\nu}$. But [76] is under the framework of a family of von Mises V-statistics where $\mathcal{V}(a)=\sum_{j=1}^{p}(\bar{x}_j-\bar{y}_j)^a$. We note that $\mathcal{V}(a)$ is equivalent to*

$$\mathcal{V}(a)=\sum_{j=1}^{p}\sum_{c=0}^{a}(-1)^{a-c}\binom{a}{c}({n_x}^c{n_y}^{a-c})^{-1}\sum_{\substack{1\le k_1,\cdots,k_c\le n_x\\ 1\le s_1,\cdots,s_{a-c}\le n_y}}\prod_{t=1}^{c}x_{k_t,j}\prod_{m=1}^{a-c}y_{s_m,j},$$

which allows the indexes k's and s's to be the same and thus is different from the U-statistics in (4.2). [76] shows that the constructed V-statistics are biased estimators of $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_a^a$, and $\mathcal{V}(a)$ and $\mathcal{V}(b)$ are asymptotically independent if $a+b$ is odd, but are asymptotically correlated if $a+b$ is even. The constructed U-statistics in this work extend the properties of those V-statistics such that $\mathcal{U}(a)$ in (4.2) is an unbiased estimator of $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_a^a$, and all $\mathcal{U}(a)$'s are asymptotically independent with each other. Given these nice statistical properties, it becomes easier to obtain the joint asymptotic distribution of the U-statistics, and then apply the adaptive test.

4.2. *Two-Sample Covariance Testing.* The U-statistics framework can be applied similarly to testing the equality of two covariance matrices. Suppose $\{\mathbf{x}_i\}_{i=1}^{n_x}$ and $\{\mathbf{y}_i\}_{i=1}^{n_y}$ are i.i.d. copies of two independent random vectors $\mathbf{x}=(x_1,\ldots,x_p)^\intercal$ and $\mathbf{y}=(y_1,\ldots,y_p)^\intercal$ respectively. Denote $\mathrm{E}(\mathbf{x})=\boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)^\intercal$, $\mathrm{E}(\mathbf{y})=\boldsymbol{\nu}=(\nu_1,\ldots,\nu_p)^\intercal$; $\mathrm{cov}(\mathbf{x})=\boldsymbol{\Sigma}_x=\{\sigma_{x,j_1,j_2}:1\le j_1,j_2\le p\}$ and $\mathrm{cov}(\mathbf{y})=\boldsymbol{\Sigma}_y=\{\sigma_{y,j_1,j_2}:1\le j_1,j_2\le p\}$. Consider $H_0:\boldsymbol{\Sigma}_x=\boldsymbol{\Sigma}_y=\boldsymbol{\Sigma}=(\sigma_{j_1,j_2})_{p\times p}$. Given $1\le j_1,j_2\le p$, $1\le k_1\ne k_2\le n_x$, and $1\le s_1\ne s_2\le n_y$, $K_{j_1,j_2}(\mathbf{x}_{k_1},\mathbf{x}_{k_2},\mathbf{y}_{s_1},\mathbf{y}_{s_2})=(x_{k_1,j_1}x_{k_1,j_2}-x_{k_1,j_1}x_{k_2,j_2})-(y_{s_1,j_1}y_{s_1,j_2}-y_{s_1,j_1}y_{s_2,j_2})$ is a simple unbiased estimator of $\sigma_{x,j_1,j_2}-\sigma_{y,j_1,j_2}$. Therefore, for a finite positive integer a, we have the U-statistic

$$\mathcal{U}(a)=\sum_{1\le j_1,j_2\le p}\frac{1}{P_{2a}^{n_x}P_{2a}^{n_y}}\sum_{\substack{1\le k_{1,1}\ne k_{1,2}\ne\ldots\\ \ne k_{a,1}\ne k_{a,2}\le n_x}}\sum_{\substack{1\le s_{1,1}\ne s_{1,2}\ne\ldots\\ \ne s_{a,1}\ne s_{a,2}\le n_y}}\prod_{t=1}^{a}K_{j_1,j_2}(\mathbf{x}_{k_{t,1}},\mathbf{x}_{k_{t,2}},\mathbf{y}_{s_{t,1}},\mathbf{y}_{s_{t,2}}). \tag{4.5}$$

As in Remark 2.1, another formulation of $\mathcal{U}(a)$ equivalent to (4.5) is

$$
\begin{aligned}
(4.6)\qquad \mathcal{U}(a) = \sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c}(-1)^{c-b_1+b_2} \sum_{1\le j_1,j_2\le p}\ \sum_{\substack{1\le i_1\ne\ldots\ne\\ i_{2c-b_1}\le n_x}}\ \sum_{\substack{1\le w_1\ne\ldots\ne\\ w_{2(a-c)-b_2}\le n_y}} \\
C_{n_x,n_y,a,c,b_1,b_2} \times \prod_{k=1}^{b_1}(x_{i_k,j_1}x_{i_k,j_2}) \prod_{s=b_1+1}^{c} x_{i_s,j_1} \prod_{t=c+1}^{2c-b_1} x_{i_t,j_2} \\
\times \prod_{m=1}^{b_2}(y_{w_m,j_1}y_{w_m,j_2}) \prod_{l=b_2+1}^{a-c} y_{w_l,j_1} \prod_{q=a-c+1}^{2(a-c)-b_2} y_{w_q,j_2},
\end{aligned}
$$

where $C_{n_x,n_y,c,b_1,b_2} = (P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-1}a!/\{b_1!(c-b_1)!b_2!(a-c-b_2)!\}$, and (4.6) shall be used in the theoretical developments.

We next present the asymptotic results of the constructed U-statistics under the null hypothesis. Here we assume the regularity Condition A.5 or A.6, whose details and discussions are provided in Section A.13.1 of Supplementary Material due to the space limitation. We mention that Condition A.5 is a mixing-type dependence assumption similar to Condition 2.2, and Condition A.6 is a moment-type dependence assumption similar to Condition 2.2*. Particularly, Condition A.6 extends the moment assumption for second-order U-statistics in Li and Chen [54] to U-statistics of general orders; please see the detailed discussions in Section A.13.1.

THEOREM 4.6. *Under H_0 and Condition A.5 or A.6 in Supplementary Material, for finite integers $\{a_1,\ldots,a_m\}$, $[\mathcal{U}(a_1)/\sigma(a_1),\ldots,\mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0,I_m)$, where for $a\in\{a_1,\ldots,a_m\}$,*

$$
\begin{aligned}
&\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \\
&\simeq \sum_{1\le j_1,j_2,j_3,j_4\le p} a!\Big\{\frac{1}{n_x}(\Pi^x_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4}) + \frac{1}{n_y}(\Pi^y_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4})\Big\}^a
\end{aligned}
$$

with $\Pi^x_{j_1,j_2,j_3,j_4} = \mathrm{E}\{\prod_{t=1}^4(x_{1,j_t}-\mu_{j_t})\}$ and $\Pi^y_{j_1,j_2,j_3,j_4} = \mathrm{E}\{\prod_{t=1}^4(y_{1,j_t}-\nu_{j_t})\}$.

Theorem 4.6 provides the asymptotic independence and joint normality of the finite-order U-statistics, which are similar to Theorems 2.1, 4.1 and 4.3. To further study the power of these finite-order U-statistics, we next consider the alternative hypotheses where $\mathbf{\Sigma}_x \ne \mathbf{\Sigma}_y$. Let $\mathbb{J}_0$ be the largest subset of $\{1,\ldots,p\}$ such that $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2} = \sigma_{j_1,j_2}$ for any $j_1,j_2\in\mathbb{J}_0$. We then obtain the following theorem under the regularity conditions given in Section A.14 of Supplementary Material.

THEOREM 4.7. *Under Conditions A.7 and A.8 in the Supplementary Material, for finite integers* $\{a_1, \ldots, a_m\}$, $[\mathcal{U}(a_1) - \mathrm{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1), \ldots, [\mathcal{U}(a_m) - \mathrm{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *where*

$$\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \simeq a! C_{\kappa,a} \sum_{j_1,j_2,j_3,j_4 \in \mathbb{J}_0} \sigma^a_{j_1,j_2}\sigma^a_{j_3,j_4},$$

and $C_{\kappa,a} = \{(\kappa_x - 1)/n_x + (\kappa_y - 1)/n_y\}^a + 2(\kappa_x/n_x + \kappa_y/n_y)^a$ *with* κ_x *and* κ_y *given in Condition A.7.*

Given the asymptotic results under the alternatives, we next analyze the power of the finite-order U-statistics. By Theorem 4.7, the asymptotic power of $\mathcal{U}(a)$ depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. Let $J_D = \{(j_1, j_2) : \sigma_{x,j_1,j_2} \neq \sigma_{y,j_1,j_2}, 1 \leq j_1, j_2 \leq p\}$, then $\mathrm{E}\{\mathcal{U}(a)\} = \sum_{(j_1,j_2)\in J_D}(\sigma_{x,j_1,j_2} - \sigma_{y,j_1,j_2})^a$. Similarly to Section 2.2, to study the relationship between the sparsity level of $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ and the power of U-statistics, we consider the case where the non-zero differences between $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ are the same. Specifically, let $\sigma_{x,j_1,j_2} - \sigma_{y,j_1,j_2} = \rho$ for $(j_1, j_2) \in J_D$, and then $\mathrm{E}\{\mathcal{U}(a)\} = |J_D|\rho^a$. Following the analysis in Section 2.2, we compare the ρ values needed by different $\mathcal{U}(a)$'s to achieve $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ for a given constant M. In particular, for given integer a, suppose $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ is achieved when $\rho = \rho_a$. For any $a \neq b$, we compare $\mathcal{U}(a)$ and $\mathcal{U}(b)$ following Criterion 1.

We use the following example as an illustration, where $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ satisfy the conditions of Theorem 4.7. Specifically, we assume that $\mathbf{\Sigma}_x = (\sigma_{x,j_1,j_2})_{p\times p}$ has the diagonal elements $\sigma_{x,j,j} = \nu^2$; and the off-diagonal elements $\sigma_{x,j_1,j_2} = h_{|j_1-j_2|} \in (0, \nu^2)$ with $h_{|j_1-j_2|} = \Theta(\nu^2)$ when $|j_1 - j_2| \leq s$, while $\sigma_{x,j_1,j_2} = 0$ when $|j_1 - j_2| > s$. This covers the moving average covariance structure of order s, and $\mathbf{\Sigma}_x$ is a banded matrix with bandwidth s. In addition, we assume the bandwidth $s = o(p)$ and $p - |\mathbb{J}_0| = o(p)$. By the definition of $\mathbb{J}_0$, the assumption $p - |\mathbb{J}_0| = o(p)$ implies that a large square sub-matrix of $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ are the same. For simplicity, we let $n_x = n_y$ with $n = n_x + n_y$, and a similar analysis can be applied when $n_x \neq n_y$. By Theorem 4.7, $\mathrm{var}\{\mathcal{U}(a)\} \simeq (n/2)^{-a}a!\{2\kappa_1^a + \kappa_2^a\}\{p\nu^{2a} + 2\sum_{t=1}^s h_t^a(p-t)\}^2$, where $\kappa_1 = \kappa_x + \kappa_y$ and $\kappa_2 = \kappa_x + \kappa_y - 2$. Therefore we know for given finite integer a, $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ holds when $\rho = \rho_a$ defined as

$$\rho_a = \frac{(a!)^{\frac{1}{2a}}\sqrt{\kappa_1}\nu}{(n/2)^{1/2}}\Big(\frac{Mp}{|J_D|}\Big)^{1/a}\Big\{2 + \Big(\frac{\kappa_2}{\kappa_1}\Big)^a\Big\}^{\frac{1}{2a}}\Big\{1 + 2\sum_{t=1}^s\Big(\frac{h_t}{\nu^2}\Big)^a\Big(1 - \frac{t}{p}\Big)\Big\}^{\frac{1}{a}}.$$

We next compare the ρ_a's and obtain the following proposition.

PROPOSITION 4.2. *There exists $\mathbb{D}_0$ that only depends on the given $\kappa_x, \kappa_y, \nu^2, s$, and $h_t, t = 1, \ldots, s$, and satisfies $\mathbb{D}_0 = \Theta(1/s^2)$ such that*

(i) When $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, the minimum of ρ_a is achieved at $a_0 = 1$.
(ii) When $|J_D| < Mp/\sqrt{\mathbb{D}_0}$, the minimum of ρ_a is achieved at some a_0, which increases as $Mp/|J_D|$ increases.

Proposition 4.2 is similar to Propositions 2.3 and 4.1. Following the analysis in Section 2.2, Proposition 4.2 shows that when the difference $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is "very" dense with $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, $\mathcal{U}(1)$ is the most powerful U-statistic; when $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ becomes sparser as $Mp/|J_D|$ decreases, a higher order U-statistic is more powerful; when the $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is "moderately" dense or sparse, a U-statistic of finite order $a_0 > 1$ would be the most powerful one.

The power analysis above shows that the power of the U-statistics varies when the alternative changes. To maintain high power across different alternatives, we can develop an adaptive testing procedure similar to that in Section 2.3. Given the asymptotic independence in Theorem 4.6, an adaptive testing procedure using the constructed $\mathcal{U}(a)$'s is valid with the type I error asymptotically controlled. Also, the adaptive test achieves high power by combining the U-statistics as discussed in Section 2.3.

We provide simulation studies on two-sample covariance testing in Supplementary Material Section C.3. By the simulations, we first find that the type I errors of the U statistics and the adaptive test are well controlled under H_0. This verifies the theoretical results in Theorem 4.7. Second, similarly to the one-sample covariance testing, we find that generally when the difference $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is sparser, a U-statistic of higher order is more powerful, and vice versa. Moreover, under moderately sparse/dense alternatives, $\mathcal{U}(a_0)$ with $a_0 > 1$ could achieve the highest power. The results are consistent with Proposition 4.2. Third, we compare the proposed adaptive test with existing methods in literature including [66, 70, 54, 10], and find that the proposed adaptive testing procedure maintains high power across various alternatives.

REMARK 4.2. *Similarly to Section 2, we can let $\mathcal{U}(\infty)$ be the maximum-type test statistic in [10], and expect that the result similar to Theorem 2.3 holds under certain regularity conditions. However, as the dependence structure of two-sample covariance matrices is more complicated than the one-sample case, it is more challenging to establish the asymptotic joint distribution of $\mathcal{U}(\infty)$ and finite-order U-statistics. We leave this interesting problem for future study, while find in simulations that the performance of $\mathcal{U}(\infty)$ is similar to high-order U-statistics $\mathcal{U}(a)$'s.*

4.3. *Generalized Linear Model.* In this section, we consider the Example 3 of generalized linear models (on Page 4) to show that the proposed framework can be extended to other testing problems. Similarly to the results in Section 4.1, we show that the constructed U-statistics are asymptotically independent and normally distributed, and also establish the power analysis results of the U-statistics. We provide the details in Section A.16 of Supplementary Material. Recently, Wu et al. [74] also discussed the adaptive testing of generalized linear model. But similarly to [76], [74] is under the framework of a family of von Mises V-statistics, and thus is different from the current paper as discussed in Remark 4.1. Moreover, the current work provides the theoretical power analysis while [74] did not.

5. Discussion. This paper introduces a general U-statistics framework for applications to high-dimensional adaptive testing. Particularly, we focus on the examples including testing of means, covariances and regression coefficients in generalized linear models. Under the null hypothesis, we prove that the U-statistics of finite orders have asymptotic joint normality, and establish the asymptotic mutual independence among the finite-order U-statistics and $\mathcal{U}(\infty)$. Moreover, under alternative hypotheses, we analyze the power of different U-statistics and demonstrate how the most powerful U-statistic changes with the sparsity level of the alternative parameters. Based on the theoretical results, we propose an adaptive testing procedure, which is powerful against different alternatives. The superior performance of this adaptive testing is confirmed in the simulations and real data analysis.

There are several possible extensions of the U-statistics framework in this paper. First, by our current proof, the convergence rate in Theorem 2.3 is bounded by $O(\log^{-1/2} p)$, which is an upper bound and not sharp. From our extensive simulations, we find that the type I error rate of the adaptive testing is well-controlled with a relatively small p, e.g., $p = 50$. We might obtain a shaper bound of the convergence rate, but more refined concentration property of the high-dimensional and high-order U-statistics is needed. Second, the proposed framework requires that the elements in the parameter set $\mathcal{E}$ have unbiased estimates. When we can not obtain unbiased estimates easily, e.g., for the precision matrix, the proposed construction may not follow directly. Nevertheless we may use "nearly" unbiased estimators to construct "U-statistics" for hypothesis testing, such as the "nearly" unbiased estimator of the precision matrix proposed in [75]; the main challenge is then to control the accumulative bias over the parameters under high-dimensions. Third, this paper discusses the examples where the elements in $\mathcal{E}$ are comparable. When the parameters in $\mathcal{E}$ are not comparable, such as $\mathcal{E}$ containing both means and covariances parameters, the construction of U-statistics still follows but the theoretical derivation may require a careful case-by-case examination. Fourth, the construction of the U-statistics treats the parameters in $\mathcal{E}$ with equal weight. More generally, we could assign different weights to different parameter estimators. For instance, standardizing the data is one example of assigning different weights. As inappropriate weight assignments could lead to power loss, when the truth is unknown, how to effectively assign weights to maximize the test power is an interesting research question. We shall discuss these extensions in the future as a significant amount of additional work is still needed.

In addition to the examples in this paper, the proposed U-statistics framework can be applied to other high-dimensional hypothesis testing problems. For example, it can be applied to testing the block-diagonality of a covariance matrix, whose theoretical analysis would be similar to the considered one sample and two sample covariance testing problems. It can also be used to test high-dimensional regression coefficients in complex regression models other than the generalized linear models, following a similar construction based on the score functions. A key step is then to characterize the impact of nuisance parameters that are estimated under the null hypothesis, and challenges arise especially when the nuisance parameters are high-dimensional. Such interesting extensions will be further explored in our follow-up studies.

Acknowledgements. The authors thank Co-Editors Prof. Edward I. George and Prof. Richard J. Samworth, an Associate Editor, and three anonymous referees for their constructive comments. The authors also thank Prof. Ping-Shou Zhong for sharing the code of the paper [15] and Prof. Xuming He and Prof. Peter Song for helpful discussions.

SUPPLEMENTARY MATERIAL

Supplementary Material:
(doi: XXX; Supplementary.pdf). This supplementary material contains the technical proofs of the main paper and additional simulations.

REFERENCES

[1] Alzheimer's Association (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia 14*(3), 367–429.

[2] Anderson, T. W. (2009). *An introduction to multivariate statistical analysis.* John Wiley & Sons, New York.

[3] Bai, Z., D. Jiang, J.-F. Yao, and S. Zheng (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist. 37*(6B), 3822–3840.

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

[5] Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*(1), 199–227.
[6] Billingsley, P. (1995). *Probability and Measure* (3 ed.). John Wiley & Sons.
[7] Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association 106*(494), 672–684.
[8] Cai, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annual Review of Statistics and Its Application 4*(1), 423–446.
[9] Cai, T. T. and T. Jiang (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist. 39*(3), 1496–1525.
[10] Cai, T. T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association 108*(501), 265–277.
[11] Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 349–372.
[12] Cai, T. T. and Z. Ma (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli 19*(5B), 2359–2388.
[13] Chen, S. X., J. Li, and P.-S. Zhong (2014). Two-sample tests for high dimensional means with thresholding and data transformation. *arXiv preprint arXiv:1410.2848*.
[14] Chen, S. X. and Y.-L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist. 38*(2), 808–835.
[15] Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.
[16] Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *The Annals of Statistics 46*(2), 642–678.
[17] Colantuoni, C., B. K. Lipska, T. Ye, T. M. Hyde, R. Tao, J. T. Leek, E. A. Colantuoni, A. G. Elkahloun, M. M. Herman, D. R. Weinberger, and J. E. Kleinman (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature 478*(7370), 519.
[18] Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist. 32*(3), 962–994.
[19] Donoho, D., J. Jin, et al. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science 30*(1), 1–25.
[20] Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association 91*(434), 674–688.
[21] Fan, J., F. Han, and H. Liu (2014). Challenges of big data analysis. *National science review 1*(2), 293–314.
[22] Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica 83*(4), 1497–1541.
[23] Fan, J., J. Lv, and L. Qi (2011). Sparse high-dimensional models in economics.
[24] Frahm, G. (2004). *Generalized elliptical distributions: theory and applications*. Ph. D. thesis, Universität zu Köln.
[25] Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science 326*(5951), 399–403.
[26] Gaetan, C. and X. Guyon (2010). *Spatial statistics and modeling*, Volume 90. Springer.
[27] Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 477–493.

[28] Gregory, K. B., R. J. Carroll, V. Baladandayuthapani, and S. N. Lahiri (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association 110*(510), 837–849.

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

[30] Guyon, X. (1995). *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media.

[31] Hall, P. (1979). On the rate of convergence of normal extremes. *Journal of Applied Probability 16*(2), 433–439.

[32] Hall, P. and J. Jin (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist. 38*(3), 1686–1732.

[33] Heinig, M., E. Petretto, C. Wallace, L. Bottolo, M. Rotival, H. Lu, Y. Li, R. Sarwar, S. R. Langley, A. Bauerfeind, et al. (2010). A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature 467*(7314), 460.

[34] Heyde, C. C. and B. M. Brown (1970). On the departure from normality of a certain class of martingales. *Ann. Math. Statist. 41*(6), 2161–2165.

[35] Higham, N. J. (2008). *Functions of matrices: theory and computation*, Volume 104. Siam.

[36] Ho, H.-C. and T. Hsing (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *Journal of Applied Probability 33*(1), 138–145.

[37] Ho, H.-C. and W. P. McCormick (1999). Asymptotic distribution of sum and maximum for Gaussian processes. *Journal of Applied Probability 36*(4), 1031–1044.

[38] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist. 19*(3), 293–325.

[39] Hsing, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *Ann. Probab. 23*(2), 938–947.

[40] Ibragimov, I. A. and J. V. Linnik (1971). Independent and stationary sequences of random variables. *The Netherlands: Wolters-Noordhof*.

[41] James, B., K. James, and Y. Qi (2007). Limit distribution of the sum and maximum from multivariate Gaussian sequences. *Journal of multivariate analysis 98*(3), 517–532.

[42] Jansen, I. E., J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, J. Sealock, I. K. Karlsson, S. Hägg, L. Athanasiu, et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimers disease risk. *Nature Genetics 51*, 404–413.

[43] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab. 14*(2), 865–880.

[44] Jiang, T. and F. Yang (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist. 41*(4), 2029–2074.

[45] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist. 29*(2), 295–327.

[46] Kan, R. (2008). From moments of sum to moments of product. *Journal of Multivariate Analysis 99*(3), 542–554.

[47] Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research 38*(suppl 1), D355–D360.

[48] Kim, J., Y. Zhang, and W. Pan (2016). Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. *Genetics 203*(2), 715–731.

[49] Kim, T. Y. (1994). Moment bounds for non-stationary dependent sequences. *Journal of Applied Probability 31*(3), 731–742.

[50] Lan, W., R. Luo, C.-L. Tsai, H. Wang, and Y. Yang (2015). Testing the diagonality of a large covariance matrix in a regression setting. *Journal of Business & Economic Statistics 33*(1), 76–86.

[51] Ledoit, O. and M. Wolf (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist. 30*(4), 1081–1102.

[52] Leung, D. and M. Drton (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics 46*(1), 280–307.

[53] Li, D. and L. Xue (2015). Joint limiting laws for high-dimensional independence tests. *ArXiv e-prints*.

[54] Li, J. and S. X. Chen (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist. 40*(2), 908–940.

[55] Liu, W.-D., Z. Lin, and Q.-M. Shao (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab. 18*(6), 2337–2366.

[56] Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, and P. M. Visscher (2009). Finding the missing heritability of complex diseases. *Nature 461*(7265), 747.

[57] McCormick, W. and Y. Qi (2000). Asymptotic distribution for the sum and maximum of Gaussian processes. *Journal of Applied Probability 37*(4), 958–971.

[58] Mosteller, F. and R. A. Fisher (1948). Questions and answers. *The American Statistician 2*(5), 30–31.

[59] Muirhead, R. J. (2009). *Aspects of multivariate statistical theory.* John Wiley & Sons, New York.

[60] Paindaveine, D. and G. Van Bever (2014). Inference on the shape of elliptical distributions based on the MCD. *Journal of Multivariate Analysis 129*, 125–144.

[61] Pan, W., J. Kim, Y. Zhang, X. Shen, and P. Wei (2014). A powerful and adaptive association test for rare variants. *Genetics 197*(4), 1081–1095.

[62] Péché, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields 143*(3-4), 481–516.

[63] Peng, Z. and S. Nadarajah (2003). On the joint limiting distribution of sums and maxima of stationary normal sequence. *Theory of Probability & Its Applications 47*(4), 706–709.

[64] Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications 19*(2), 297–303.

[65] Prince, M., R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's & Dementia 9*(1), 63–75.

[66] Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics & Data Analysis 51*(12), 6535–6542.

[67] Shao, Q.-M. and W.-X. Zhou (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab. 42*(2), 623–648.

[68] Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics 108*(5-6),

1033–1056.

[69] Srivastava, M. S. and M. Du (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis 99*(3), 386–402.

[70] Srivastava, M. S. and H. Yanagihara (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis 101*(6), 1319–1329.

[71] Srivastava, R., P. Li, and D. Ruppert (2016). RAPTT: An exact two-sample test in high dimensions using random projections. *Journal of Computational and Graphical Statistics 25*(3), 954–970.

[72] Storey, J. D. and R. Tibshirani (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences 100*(16), 9440–9445.

[73] Wang, L., P. Jia, R. D. Wolfinger, X. Chen, and Z. Zhao (2011). Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics 98*(1), 1–8.

[74] Wu, C., G. Xu, and W. Pan (2019). An adaptive test on high-dimensional parameters in generalized linear models. *Statistica Sinica*.

[75] Xia, Y., T. Cai, and T. T. Cai (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika 102*(2), 247–266.

[76] Xu, G., L. Lin, P. Wei, and W. Pan (2016). An adaptive two-sample test for high-dimensional means. *Biometrika 103*(3), 609–624.

[77] Xu, J., S. L. Murphy, K. D. Kochanek, B. Bastian, and E. Arias (2018). Deaths: Final data for 2016. *National Vital Statistics Reports 67*(5).

[78] Xu, Z., G. Xu, and W. Pan (2017). Adaptive testing for association between two random vectors in moderate to high dimensions. *Genetic epidemiology 41*(7), 599–609.

[79] Yang, Q. and G. Pan (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *Journal of the American Statistical Association 112*(517), 188–200.

[80] Yu, K., Q. Li, A. W. Bergen, R. M. Pfeiffer, P. S. Rosenberg, N. Caporaso, P. Kraft, and N. Chatterjee (2009). Pathway analysis by adaptive combination of p-values. *Genetic epidemiology 33*(8), 700–709.

[81] Zaitsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of sn bernstein's inequality conditions. *Probability theory and related fields 74*(4), 535–566.

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

SUPPLEMENT TO "ASYMPTOTICALLY INDEPENDENT U-STATISTICS IN HIGH DIMENSIONAL ADAPTIVE TESTING"

We give proofs of the main results and additional simulations in this supplementary material. For simplicity, we use C to represent some generic positive constant, which does not change with (n, p) and may represent different values from place to place.

APPENDIX A: PROOFS AND SUPPLEMENTARY RESULTS

A.1. Proof of Proposition 2.1. To prove $\mathcal{U}(a)$ in (2.3) is location invariant, we examine the equivalent form,

$$\mathcal{U}(a) = (P^n_{2a})^{-1} \sum_{1\le j_1\neq j_2\le p} \sum_{1\le i_1\neq\ldots\neq i_{2a}\le n} \prod_{k=1}^{a} (x_{i_{2k-1},j_1} x_{i_{2k-1},j_2} - x_{i_{2k-1},j_1} x_{i_{2k},j_2}).$$

We consider $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_p)^\intercal \in \mathbb{R}^p$, and examine $a = 1$ first. For each (j_1, j_2), since

$$\begin{aligned} & (x_{i_1,j_1} + \Delta_{j_1})(x_{i_1,j_2} + \Delta_{j_2}) - (x_{i_1,j_1} + \Delta_{j_1})(x_{i_2,j_2} + \Delta_{j_2}) \\ = \ & (x_{i_1,j_1} x_{i_1,j_2} - x_{i_1,j_1} x_{i_2,j_2}) + \Delta_{j_1}(x_{i_1,j_2} - x_{i_2,j_2}), \end{aligned}$$

then it follows that

$$\begin{aligned} & \sum_{1\le i_1\neq i_2\le n} [(x_{i_1,j_1} + \Delta_{j_1})(x_{i_1,j_2} + \Delta_{j_2}) - (x_{i_1,j_1} + \Delta_{j_1})(x_{i_2,j_2} + \Delta_{j_2})] \\ & \quad - \sum_{1\le i_1\neq i_2\le n} (x_{i_1,j_1} x_{i_1,j_2} - x_{i_1,j_1} x_{i_2,j_2}) \\ = \ & \sum_{1\le i_1\neq i_2\le n} \Delta_{j_1}(x_{i_1,j_2} - x_{i_2,j_2}) + \sum_{i=1}^{n} \Delta_{j_1}(x_{i,j_2} - x_{i,j_2}) \\ = \ & \Delta_{j_1} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} (x_{i_1,j_2} - x_{i_2,j_2}) \\ = \ & 0. \end{aligned}$$

That is, $\mathcal{U}(1)$ is location invariant. For $a=2$, given (j_1, j_2), following a similar analysis to $\mathcal{U}(1)$, we have

$$\begin{aligned}
&\sum_{1\le i_1\neq\ldots\neq i_4\le n}\Big\{[(x_{i_1,j_1}+\Delta_{j_1})(x_{i_1,j_2}+\Delta_{j_2})-(x_{i_1,j_1}+\Delta_{j_1})(x_{i_2,j_2}+\Delta_{j_2})]\\
&\qquad\times[(x_{i_3,j_1}+\Delta_{j_1})(x_{i_3,j_2}+\Delta_{j_2})-(x_{i_3,j_1}+\Delta_{j_1})(x_{i_4,j_2}+\Delta_{j_2})]\Big\}\\
&-\sum_{1\le i_1\neq\ldots\neq i_4\le n}\Big\{(x_{i_1,j_1}x_{i_1,j_2}-x_{i_1,j_1}x_{i_2,j_2})\\
&\qquad\times[(x_{i_3,j_1}+\Delta_{j_1})(x_{i_3,j_2}+\Delta_{j_2})-(x_{i_3,j_1}+\Delta_{j_1})(x_{i_4,j_2}+\Delta_{j_2})]\Big\}\\
&=0.
\end{aligned}\tag{A.1}$$

Similarly, we also have

$$\begin{aligned}
&\sum_{1\le i_1\neq\ldots\neq i_4\le n}\Big\{(x_{i_1,j_1}x_{i_1,j_2}-x_{i_1,j_1}x_{i_2,j_2})\\
&\qquad\times[(x_{i_3,j_1}+\Delta_{j_1})(x_{i_3,j_2}+\Delta_{j_2})-(x_{i_3,j_1}+\Delta_{j_1})(x_{i_4,j_2}+\Delta_{j_2})]\Big\}\\
&-\sum_{1\le i_1\neq\ldots\neq i_4\le n}[(x_{i_1,j_1}x_{i_1,j_2}-x_{i_1,j_1}x_{i_2,j_2})(x_{i_3,j_1}x_{i_3,j_2}-x_{i_3,j_1}x_{i_4,j_2})]\\
&=0.
\end{aligned}\tag{A.2}$$

Combining (A.1) and (A.2), we know $\mathcal{U}(2)$ is location invariant. Following the argument above similarly, by induction, we obtain that $\mathcal{U}(a)$ is location invariant for a general integer $a \geq 3$.

A.2. Proof of Theorem 2.1. For the covariance testing example in Section 2, $\mathcal{U}(a)$ is location invariant by Proposition 2.1, and $\mathcal{U}(\infty)$ is also location invariant straightforwardly by its expression in (2.8). Then we assume without loss of generality that $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ in this section. To prove Theorem 2.1, we first derive the variances and the covariances of the U-statistics, and then prove the asymptotic joint normality of the U-statistics.

In particular, the next Lemma A.1 derives the asymptotic form of variance $\sigma^2(a)$ in (2.7).

LEMMA A.1. *Under the conditions of Theorem 2.1, for any finite integer a, following the notation in (2.2),*

$$\sigma^2(a) = \frac{a!}{P_a^n} \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} (\Pi_{j_1,j_2,j_3,j_4})^a\{1+o(1)\},$$

which is of order $\Theta(p^2 n^{-a})$. *In addition, for* $\tilde{\mathcal{U}}(a)$ *defined in* (2.5) *and* $\tilde{\mathcal{U}}^*(a) := \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$, *we have* $\mathrm{var}\{\mathcal{U}(a)\} = \mathrm{var}\{\tilde{\mathcal{U}}(a)\}\{1+o(1)\}$, $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1) \times \mathrm{var}\{\tilde{\mathcal{U}}(a)\}$, *and* $\tilde{\mathcal{U}}^*(a)/\sigma(a) \xrightarrow{P} 0$.

PROOF. *See Section B.1.1 on Page 72.* □

Moreover, the following Lemma A.2 shows that the covariances between different $\mathcal{U}(a)$'s asymptotically converge to 0.

LEMMA A.2. *Under the conditions of Theorem 2.1, for finite integers* $a \neq b$, $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} \to 0$, *as* $n, p \to \infty$.

PROOF. *See Section B.1.2 on Page 86.* □

Lemmas A.1 and A.2 together establish that the covariance matrix of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ converges to I_m asymptotically. To finish the proof of Theorem 2.1, it remains to show that the joint limiting distribution of the U-statistics is normal.

For finite integers $a_1, \ldots, a_m$, to obtain the joint asymptotic normality of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$, by the Cramér-Wold theorem, it is equivalent to prove that any fixed linear combination of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ converges to normal. Recall that Lemma A.1 shows that $\tilde{\mathcal{U}}^*(a)/\sigma(a) \xrightarrow{P} 0$ for any finite integer a. Thus by the Slutsky's theorem, it suffices to prove that any fixed linear combination of $[\tilde{\mathcal{U}}(a_1)/\sigma(a_1), \ldots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m)]^\intercal$ converges to normal. To be specific, we show that for constants $t_1, \ldots, t_m$ satisfying $\sum_{r=1}^m t_r^2 = 1$,

$$Z_n := \sum_{r=1}^{m} t_r \tilde{\mathcal{U}}(a_r)/\sigma(a_r) \xrightarrow{D} \mathcal{N}(0,1). \tag{A.3}$$

To prove (A.3), we apply the martingale central limit theorem in Heyde and Brown [34] (similar arguments can date back to Bai and Saranadasa [4]). Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_k = \sigma\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$, and $\mathrm{E}_k(\cdot)$ denote the conditional expectation given $\mathcal{F}_k$ for $k = 1, \cdots, n$. Define $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2 = \mathrm{E}_{k-1}(D_{n,k}^2)$. Note that $\mathrm{E}_0(\cdot) = \mathrm{E}(\cdot)$, and $\mathrm{E}(Z_n) = 0$ as $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. It follows that $Z_n = \sum_{k=1}^n D_{n,k}$. By martingale central limit theorem, to prove (A.3), it is sufficient to show

$$\sum_{k=1}^{n} \pi_{n,k}^2/\mathrm{var}(Z_n) \xrightarrow{P} 1, \qquad \sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4)/\mathrm{var}^2(Z_n) \to 0. \tag{A.4}$$

Here $\mathrm{var}(Z_n) \to \sum_{r=1}^m t_r^2 = 1$ by Lemmas A.1 and A.2, and $\mathrm{E}(\sum_{k=1}^n \pi_{n,k}^2) = \mathrm{var}(Z_n)$ by the following Lemma A.3.

LEMMA A.3. *Under the conditions of Theorem 2.1,* $\mathrm{E}(\sum_{k=1}^{n}\pi_{n,k}^2) = \mathrm{var}(Z_n)$.

PROOF. *See Section B.1.3 on Page 86.* □

Therefore to prove (A.4), it suffices to show

$$\mathrm{var}\left(\sum_{k=1}^{n}\pi_{n,k}^2\right)\to 0 \quad \text{and} \quad \sum_{k=1}^{n}\mathrm{E}(D_{n,k}^4)\to 0. \tag{A.5}$$

Note that $D_{n,k}$ and $\pi_{n,k}^2$ in (A.5) can be written as $D_{n,k}=\sum_{r=1}^{m}t_r A_{n,k,a_r}$ and $\pi_{n,k}^2 = \sum_{1\le r_1,r_2\le m}\mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})$, where we define $A_{n,k,a} = (\mathrm{E}_k-\mathrm{E}_{k-1})\{\tilde{\mathcal{U}}(a)/\sigma(a)\}$ for each finite integer a. The following Lemma A.4 gives the explicit form of $A_{n,k,a}$.

LEMMA A.4. *For finite integer a, when $k<a$, $A_{n,k,a}=0$; when $k\ge a$,*

$$A_{n,k,a}=\frac{a}{\sigma(a)P_a^n}\sum_{1\le i_1\ne\cdots\ne i_{a-1}\le k-1}\ \sum_{1\le j_1\ne j_2\le p}(x_{k,j_1}x_{k,j_2})\times\prod_{t=1}^{a-1}(x_{i_t,j_1}x_{i_t,j_2}).$$

PROOF. *See Section B.1.4 on Page 87.* □

With the form of $A_{n,k,a}$ in Lemma A.4, the forms of $D_{n,k}$ and $\pi_{n,k}^2$ can be obtained, and we can prove the next two Lemmas A.5 and A.6, which suggest that (A.5) holds.

LEMMA A.5. *Under the conditions of Theorem 2.1,* $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2)\to 0$. *In particular, under Condition 2.2,* $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2)=O(p^{-1}\log^3 p)$*; under Condition 2.2*,* $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2)=O(n^{-1}+p^{-2})$.

PROOF. *See Section B.1.5 on Page 89.* □

LEMMA A.6. *Under the conditions of Theorem 2.1,* $\sum_{k=1}^{n}\mathrm{E}(D_{n,k}^4) = O(1/n)$.

PROOF. *See Section B.1.6 on Page 103.* □

Finally, by Heyde and Brown [34], we have as $n,p\to\infty$,

$$\begin{aligned} &\sup_t\left|P(Z_n\le t)-\Phi(t)\right| \\ \le\ & C\left\{\mathrm{E}\left[\frac{\sum_{k=1}^{n}\mathrm{E}_{k-1}(D_{n,k}^2)}{\mathrm{var}(Z_n)}-1\right]^2+\frac{\sum_{k=1}^{n}\mathrm{E}\left(D_{n,k}^4\right)}{\mathrm{var}^2(Z_n)}\right\}^{1/5} \\ \to\ & 0, \end{aligned} \tag{A.6}$$

which proves (A.3). In summary, Theorem 2.1 is proved.

A.3. Proof of Theorem 2.3. In this section, we first introduce some notation, and then present the proof.

Notation. For $\mathcal{U}(a)$ in (2.3), by the symmetricity of covariance matrix, we can replace $\sum_{1\leq j_1\neq j_2\leq p}$ by $2\times\sum_{1\leq j_1<j_2\leq p}$. This implies that the summation over $\{(j_1,j_2) : 1 \leq j_1 \neq j_2 \leq p\}$ is equivalent to the summation over $\{(j_1,j_2) : 1\leq j_1<j_2\leq p\}$ up to a constant. Without loss of generality, we consider $j_1 < j_2$ below. We rewrite the index set $\{(j_1,j_2) : 1\leq j_1<j_2\leq p\}$ as

$$L := \Big\{(j_l^1, j_l^2) : 1 \leq l \leq q = \binom{p}{2}\Big\}, \tag{A.7}$$

where $j_l^1 = \arg\min_{1\leq k\leq p-1}\{\sum_{t=1}^{k}(p-t) \geq l\}$ and $j_l^2 = l+j_l^1 - \sum_{t=1}^{j_l^1-1}(p-t)$. For each $(j_l^1, j_l^2) \in L$, define

$$U_l^a = \sum_{1\leq i_1\neq\ldots\neq i_a\leq n}\prod_{k=1}^{a} x_{i_k,j_l^1}x_{i_k,j_l^2}. \tag{A.8}$$

Then $\tilde{\mathcal{U}}(a) = 2(P_a^n)^{-1}\sum_{l=1}^{q} U_l^a$ following the definition in (2.5). Furthermore, we define

$$\begin{aligned}
\tilde{G}_l &= \sum_{i=1}^{n} \frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}} \times \frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}, \\
M_n &= \max_{1\leq l\leq q}(\tilde{G}_l)^2, \\
\hat{G}_l &= \sum_{i=1}^{n} \frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}} \times \frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\mathbf{1}\Big\{\Big|\frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}} \times \frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\Big| \leq \tau_n\Big\} \\
&\quad - \mathrm{E}\Bigg[\sum_{i=1}^{n} \frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}} \times \frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\mathbf{1}\Big\{\Big|\frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}} \times \frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\Big| \leq \tau_n\Big\}\Bigg], \\
\hat{M}_n &= \max_{1\leq l\leq q}(\hat{G}_l)^2,
\end{aligned} \tag{A.9}$$

where we define $\sigma_{j_l^1,j_l^1} = \mathrm{var}(x_{i,j_l^1})$, $\sigma_{j_l^2,j_l^2} = \mathrm{var}(x_{i,j_l^2})$, $\tau_n = \tau\log(p+n)$ with τ being a sufficiently large positive constant and $\mathbf{1}\{\cdot\}$ represents an indicator function. In addition, we define $|\mathbf{a}|_{\min} = \min_{1\leq i\leq p}|a_i|$ for $\mathbf{a}\in\mathbb{R}^p$, and

$$y_p = 4\log p - \log\log p + y. \tag{A.10}$$

Proof. Similarly to Section A.2, since $\mathcal{U}(a)$ in (2.3) and $\mathcal{U}(\infty)$ in (2.8) are location invariant, we assume without loss of generality that $\mathrm{E}(\mathbf{x}) = \mathbf{0}$.

To prove Theorem 2.3, we first establish the asymptotic independence between $\hat{M}_n/n$ and $\tilde{\mathcal{U}}(a)/\sigma(a_r)$ for $r = 1, \ldots, m$, and then we show that $n\mathcal{U}^2(\infty)$ and $\mathcal{U}(a_r)$ are close to $\hat{M}_n/n$ and $\tilde{\mathcal{U}}(a_r)$, respectively. Specifically, the following Lemma A.7 shows that $\hat{M}_n/n$ and $\tilde{\mathcal{U}}(a_r)/\sigma(a_r)$'s are asymptotically independent.

LEMMA A.7. *Under the conditions of Theorem 2.3, when $\tau > 0$ in (A.9) is a sufficiently large constant,*

$$\Big| P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a_1)}{\sigma(a_1)} \leq z_1, \ldots, \frac{\tilde{\mathcal{U}}(a_m)}{\sigma(a_m)} \leq z_m\Big) - P\Big(\frac{\hat{M}_n}{n} > y_p\Big) \prod_{r=1}^{m} P\Big(\frac{\tilde{\mathcal{U}}(a_r)}{\sigma(a_r)} \leq z_r\Big) \Big| \to 0.$$

PROOF. *See Section B.2.1 on Page 110.* □

To show that $\hat{M}_n/n$ and $n\mathcal{U}(\infty)^2$ are close, we use M_n/n defined in (A.9) as an intermediate variable. We next prove that M_n/n and $\hat{M}_n/n$ have small difference in the sense that the conclusion in Lemma A.7 still holds by replacing $\hat{M}_n$ with M_n. This is formally stated in the following Lemma A.8.

LEMMA A.8. *Under the conditions of Theorem 2.3,*

$$\Big| P\Big(\frac{M_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a_1)}{\sigma(a_1)} \leq z_1, \ldots, \frac{\tilde{\mathcal{U}}(a_m)}{\sigma(a_m)} \leq z_m\Big) - P\Big(\frac{M_n}{n} > y_p\Big) \prod_{r=1}^{m} P\Big(\frac{\tilde{\mathcal{U}}(a_r)}{\sigma(a_r)} \leq z_r\Big) \Big| \to 0.$$

PROOF. *See Section B.2.5 on Page 130.* □

Given Lemma A.8, we further prove that M_n/n and $\tilde{\mathcal{U}}(a)/\sigma(a_r)$ are close to $n\mathcal{U}^2(\infty)$ and $\mathcal{U}(a_r)$, respectively. In particular, by the proof of Theorem 3 in Cai and Jiang [9], we know $\{n^2\mathcal{U}^2(\infty) - M_n\}/n \xrightarrow{P} 0$. In addition, Lemma A.1 proves that $\{\mathcal{U}(a_r) - \tilde{\mathcal{U}}(a_r)\}/\sigma(a_r) \xrightarrow{P} 0$. Based on these results and Lemma A.8, the following Lemma A.9 shows that the conclusion in Lemma A.8 still holds by replacing M_n/n with $n\mathcal{U}^2(\infty)$ and replacing $\tilde{\mathcal{U}}(a_r)$ with $\mathcal{U}(a_r)$.

LEMMA A.9. *Under the conditions of Theorem 2.3,*

$$\Big|P\Big(n\mathcal{U}^2(\infty) > y_p, \frac{\mathcal{U}(a_1)}{\sigma(a_1)} \le z_1, \dots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)} \le z_m\Big)$$
$$- P\Big(n\mathcal{U}^2(\infty) > y_p\Big) \prod_{r=1}^{m} P\Big(\frac{\mathcal{U}(a_r)}{\sigma(a_r)} \le z_r\Big)\Big| \to 0.$$

PROOF. *See Section B.2.6 on Page 132.* □

Lemma A.9 then proves Theorem 2.3.

A.4. Proof of Theorem 2.4. As both $\mathcal{U}(a)$ and $\mathbb{V}_u(a)$ are location invariant in the sense of Proposition 2.1, we assume $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. To prove Theorem 2.4, we decompose $\mathbb{V}_u(a) = \mathbb{V}_{u,1}(a) + \mathbb{V}_{u,2}(a)$, where we define

$$\mathbb{V}_{u,1}(a) = \frac{2a!}{(P^n_a)^2} \sum_{1\le j_1\neq j_2\le p} \sum_{1\le i_1\neq\ldots\neq i_a\le n} \prod_{t=1}^{a} x^2_{i_t,j_1} x^2_{i_t,j_2},$$

and $\mathbb{V}_{u,2}(a) = \mathbb{V}_u(a) - \mathbb{V}_{u,1}(a)$. The next Lemma A.10 shows that $\mathbb{V}_{u,1}(a)$ is of a larger order than $\mathbb{V}_{u,2}(a)$, and thus it is the leading term in $\mathbb{V}_u(a)$.

LEMMA A.10. *Under the conditions of Theorem 2.4,* $\mathbb{V}_{u,1}(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 1$ *and* $\mathbb{V}_{u,2}(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 0$.

PROOF. *See Section B.3 on Page 134.* □

Lemma A.10 implies that $\mathbb{V}_u(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 1$. As $\mathbb{V}_u(a) > 0$ with probability 1, $\mathrm{E}\{\mathbb{V}_{u,1}(a)\}/\mathbb{V}_u(a) \xrightarrow{P} 1$. In addition, note that $\mathrm{E}\{\mathbb{V}_{u,1}(a)\} = 2a!(P^n_a)^{-1}\sum_{1\le j_1\neq j_2\le p}\{\mathrm{E}(x^2_{1,j_1}x^2_{1,j_2})\}^a$. By (B.20) and (B.29) in Section B.1.1, we have $\mathrm{var}\{\mathcal{U}(a)\}/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \to 1$. Therefore,

$$\frac{\mathbb{V}_u(a)}{\mathrm{var}\{\mathcal{U}(a)\}} = \frac{\mathbb{V}_u(a)}{\mathrm{E}\{\mathbb{V}_{u,1}(a)\}} \times \frac{\mathrm{E}\{\mathbb{V}_{u,1}(a)\}}{\mathrm{var}\{\mathcal{U}(a)\}} \xrightarrow{P} 1.$$

A.5. Proof of Theorem 2.5. We first present Condition A.1 in Theorem 2.5, which is a generalized version of Condition 2.2* under H_A.

CONDITION A.1. *Following the central moment notation in (2.2), for $t \le 8$, we assume that there exists constant $\tilde{\kappa}_t$ such that $\Pi_{j_1,\ldots,j_t} = \tilde{\kappa}_t \mathrm{E}(\prod_{k=1}^{t} z_{j_k})$, where $1 \le j_1, \ldots, j_t \le p$ and $(z_1, \ldots, z_p)^\intercal \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_A)$.*

Condition A.1 generalizes Condition 2.2* to the alternative setting. Similarly to Condition 2.2*, Condition A.1 is satisfied when $\mathbf{x}$ follows an elliptical distribution with certain moment conditions [see 24**?**]. To be consistent with the notation in Condition 2.2*, we let $\kappa_1 = \tilde{\kappa}_4$ below.

We next introduce some notation, and then provide the proof.

Notation. For each given $j_1 \in \{1, \dots, p\}$, we define

$$\begin{aligned} J_{j_1} &= \{(j_1, j_2) : \sigma_{j_1,j_2} \neq 0, 1 \leq j_1 \neq j_2 \leq p\}, \\ J_{j_1}^c &= \{(j_1, j_2) : \sigma_{j_1,j_2} = 0, 1 \leq j_1 \neq j_2 \leq p\}. \end{aligned}$$

Then $J_A = \cup_{j_1=1}^p J_{j_1}$, and we correspondingly define $J_A^c = \cup_{j_1=1}^p J_{j_1}^c$, which is the set difference of $\{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$ and J_A. Moreover, we define $F(a, c) = (-1)^c \binom{a}{c} / P_{a+c}^n$, and

$$K(c, j_1, j_2) = F(a, c) \sum_{1 \leq i_1 \neq \ldots \neq i_{a+c} \leq n} \prod_{t=1}^{a-c} (x_{i_t,j_1} x_{i_t,j_2}) \prod_{t=a-c+1}^{a} x_{i_t,j_1} \prod_{t=a+1}^{a+c} x_{i_t,j_2}.$$

We decompose $\mathcal{U}(a) = T_{U,a,1,1} + T_{U,a,1,2} + T_{U,a,2}$, where

$$\text{(A.11)} \quad T_{U,a,1,1} = \sum_{(j_1,j_2) \in J_A^c} K(0, j_1, j_2), \quad T_{U,a,1,2} = \sum_{(j_1,j_2) \in J_A^c} \sum_{c=1}^{a} K(c, j_1, j_2),$$
$$T_{U,a,2} = \sum_{(j_1,j_2) \in J_A} \sum_{c=0}^{a} K(c, j_1, j_2).$$

Proof. Similarly to Section A.2, we first derive the variances and the covariances of the U-statistics, and then prove the asymptotic joint normality of the U-statistics. Particularly, the next Lemma A.11 derives the asymptotic form of $\mathrm{var}\{\mathcal{U}(a)\}$, and additionally shows that among the three terms in (A.11), $T_{U,a,1,1}$ is the leading one.

LEMMA A.11. *Under the conditions of Theorem 2.5,* $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \simeq \mathrm{var}(T_{U,a,1,1})$*, where*

$$\mathrm{var}(T_{U,a,1,1}) \simeq 2a! \kappa_1^a n^{-a} \sum_{1 \leq j_1 \neq j_2 \leq p} \sigma_{j_1,j_1}^a \sigma_{j_2,j_2}^a,$$

which is $\Theta(p^2 n^{-a})$*. Moreover,* $\mathrm{var}(T_{U,a,1,2}) = o(p^2 n^{-a})$, $\mathrm{var}(T_{U,a,2}) = o(p^2 n^{-a})$ *and* $\{\mathcal{U}(a) - T_{U,a,1,1}\}/\sigma(a) \xrightarrow{P} 0$.

PROOF. *See Section B.4.1 on Page 137.* □

The following Lemma A.12 shows that the covariance between two different U-statistics asymptotically converges to 0.

LEMMA A.12. *Under the conditions of Theorem 2.5, for two integers* $a \neq b$, $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} \to 0$.

PROOF. *See Section B.4.2 on Page 148.* □

To finish the proof, it remains to obtain the joint asymptotic normality of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$. By the Cramér-Wold theorem, it is equivalent to prove that any fixed linear combination of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ converges to a normal distribution. By Lemma A.11, $\{\mathcal{U}(a) - T_{U,a,1,1}\}/\sigma(a) \xrightarrow{P} 0$, thus by the Slutsky's theorem, it suffices to prove that any fixed linear combination of $[T_{U,a_1,1,1}/\sigma(a_1), \ldots, T_{U,a_m,1,1}/\sigma(a_m)]^\intercal$ converges to a normal distribution. Similarly to Section A.2, we redefine Z_n as below with $\sum_{r=1}^m t_r^2 = 1$, and prove that

$$Z_n := \sum_{r=1}^m t_r T_{U,a_r,1,1}/\sigma(a_r) \xrightarrow{D} \mathcal{N}(0,1). \tag{A.12}$$

We next prove (A.12) by the martingale central limit theorem, similarly to Section A.2. In particular, we define $\mathrm{E}_k(\cdot)$ in the same way as in Section A.2, and still define $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2 = \mathrm{E}_{k-1}(D_{n,k}^2)$. It follows that $D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$ and $\pi_{n,k}^2 = \sum_{1 \le r_1, r_2 \le m} t_{r_1} t_{r_2} \mathrm{E}_{k-1}(A_{n,k,a_{r_1}} A_{n,k,a_{r_2}})$, where we redefine $A_{n,k,a_r} = (\mathrm{E}_k - \mathrm{E}_{k-1})\{T_{U,a_r,1,1}/\sigma(a_r)\}$. Note that $\sigma_{j_1,j_2} = 0$ when $(j_1, j_2) \in J_A^c$, and $T_{U,a,1,1}$ is a summation over $(j_1, j_2) \in J_A^c$. Thus the proof of Lemma A.4 in Section B.1.4 applies similarly, and we obtain the explicit form of $A_{n,k,a}$. Specifically, for each finite integer a, when $k < a$, $A_{n,k,a} = 0$; when $k \ge a$,

$$A_{n,k,a} = \frac{a}{\sigma(a) P_a^n} \sum_{1 \le i_1 \neq \cdots \neq i_{a-1} \le k-1} \sum_{(j_1,j_2) \in J_A^c} (x_{k,j_1} x_{k,j_2}) \prod_{t=1}^{a-1} (x_{i_t,j_1} x_{i_t,j_2}).$$

With the form of $A_{n,k,a}$, we can obtain the explicit forms of $D_{n,k}$ and $\pi_{n,k}^2$. Then we can prove the following two Lemmas A.13 and A.14, which suggests that (A.12) holds.

LEMMA A.13. *Under the conditions of Theorem 2.5,* $\mathrm{var}(\sum_{k=1}^n \pi_{n,k}^2) \to 0$.

PROOF. *See Section B.4.3 on Page 148.* □

LEMMA A.14. *Under the conditions of Theorem 2.5 , $\sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4) \to 0$.*

PROOF. *See Section B.4.4 on Page 157.* $\square$

By Lemmas A.13 and A.14, (A.12) holds and thus Theorem 2.5 is proved.

A.6. Proof of Proposition 2.3. Consider the setting when n, p and $|J_A|$ are given and the value of M is fixed as $\Theta(1)$. We next examine ρ_a in (2.13) as a function of integer a in the following two cases.

(i) $|J_A| > Mp$. When $Mp/|J_A| < 1$, both $(Mp/|J_A|)^{1/a}$ and $(a!)^{1/(2a)}$ are increasing functions of integer a. Thus ρ_a is an increasing function of a. Since $a \in \mathbb{Z}^+$, ρ_a reaches the minimum value at $a = 1$.

(ii) $|J_A| \leq Mp$. Define $\tilde{M} = Mp/|J_A|$, and $f(a) = (a!)^{1/(2a)}(\tilde{M})^{1/a}$. Note that ρ_a and $f(a)$ only differs by a constant. To find the minimum of ρ_a, it suffices to examine the minimum of $f(a)$.

In the following, we show that when $f(a)$ starts to not decrease at some value, it will strictly increase afterwards. Specifically, we prove that $f(a+2)/f(a+1) > 1$ if $f(a+1)/f(a) \geq 1$. Note that

$$
\begin{aligned}
\frac{f(a+1)}{f(a)} &= \frac{\{(a+1)!\}^{\frac{1}{2(a+1)}}(\tilde{M})^{\frac{1}{a+1}}}{(a!)^{\frac{1}{2a}}(\tilde{M})^{\frac{1}{a}}} \\
&= \Big[\frac{\{(a+1)!\}^a \tilde{M}^{2a}}{(a!)^{a+1}\tilde{M}^{2(a+1)}}\Big]^{\frac{1}{2a(a+1)}} = \{d(a) \times \tilde{M}^{-2}\}^{\frac{1}{2a(a+1)}},
\end{aligned}
$$

where $d(a) = (a+1)^a (a!)^{-1}$. It follows that $f(a+1)/f(a) > 1$ and $f(a+1)/f(a) = 1$ are equivalent to $d(a) > \tilde{M}^2$ and $d(a) = \tilde{M}^2$, respectively. We next show that $d(a)$ is a strictly increasing function of a. In particular,

$$
\frac{d(a+1)}{d(a)} = \frac{(a+2)^{a+1}a!}{(a+1)^a(a+1)!} = \Big(\frac{a+2}{a+1}\Big)^{a+1} > 1.
$$

Therefore we have $d(a+1) > \tilde{M}^2$ if $d(a) \geq \tilde{M}^2$; and equivalently this implies that $f(a+2)/f(a+1) > 1$ if $f(a+1)/f(a) \geq 1$.

Suppose a_0 is the first integer such that $d(a_0) \geq \tilde{M}^2$, i.e., for any integer $1 \leq a < a_0$, $d(a) < \tilde{M}^2$. By the analysis above, we know $f(a)$ is decreasing when $a < a_0$, and $f(a)$ is strictly increasing when $a > a_0$. Thus a_0 achieves the minimum of $f(a)$. Since $d(a)$ is a strictly increasing function of a, we know $a_0 < \infty$ for fixed $\tilde{M}$, and a_0 increases as $\tilde{M}$ increases. Therefore the second part of proposition 2.3 is proved.

A.7. Proof of Proposition 2.4.

Proof. Consider the simplified test statistic given in (2.15). We assume $\mathrm{E}(x_{i,j})=0$ and $\mathrm{var}(x_{i,j}^2)=1$, $\forall j=1,\ldots,p$ without loss of generality. It is then equivalent to examine $\mathcal{U}(\infty)=\max_{1\le j_1<j_2\le p}|\sum_{k=1}^n x_{k,j_1}x_{k,j_2}/n|$. We next prove (i) and (ii) of Proposition 2.4 in the following Sections A.7.1 and A.7.2, respectively.

A.7.1. *Proof of (i).* Under the alternative, we consider n i.i.d. observations $(x_{k,1},x_{k,2})$, satisfying $\mathrm{E}(x_{k,1}x_{k,2})=\rho$, for $k=1,\ldots,n$. Then by Condition 2.2*, $\mathrm{var}(x_{k,1}x_{k,2})=\mathrm{E}(x_{k,1}^2x_{k,2}^2)-[\mathrm{E}(x_{k,1}x_{k,2})]^2=\kappa_1(1+2\rho^2)-\rho^2$. The power of $\mathcal{U}(\infty)$ satisfies that

$$
\begin{aligned}
(\mathrm{A.13})\qquad & P(|\,\mathcal{U}(\infty)\,|\ge t_p)\\
&= P\Big(\max_{1\le j_1<j_2\le p}\Big|\sum_{k=1}^n x_{k,j_1}x_{k,j_2}/n\Big|\ge t_p\Big)\\
&\ge P\Big(\Big|\sum_{k=1}^n x_{k,1}x_{k,2}/n\Big|\ge t_p\Big)\\
&\ge P\Big(\sum_{k=1}^n x_{k,1}x_{k,2}/n\ge t_p\Big)\\
&= P\left(\frac{\sum_{k=1}^n(x_{k,1}x_{k,2}-\rho)}{\sqrt{n}\sqrt{\mathrm{var}(x_{k,1}x_{k,2})}}\ge\frac{\sqrt{n}(t_p-\rho)}{\sqrt{\mathrm{var}(x_{k,1}x_{k,2})}}\right).
\end{aligned}
$$

We apply the central limit theorem on $x_{k,1}x_{k,2}$, $k=1,\ldots,n$, and obtain

$$
\frac{\sum_{k=1}^n(x_{k,1}x_{k,2}-\rho)}{\sqrt{n}\sqrt{\mathrm{var}(x_{k,1}x_{k,2})}}\xrightarrow{D}\mathcal{N}(0,1).
$$

Suppose Z follows a standard Gaussian distribution. As $\log p\to\infty$, $\log p/n=o(1)$, and by Berry-Esseen Theorem, we have

$$
\begin{aligned}
(\mathrm{A.13}) \quad &\ge P\left(Z\ge\frac{\sqrt{n}(t_p-\rho)}{\sqrt{\mathrm{var}(x_{k1}x_{k2})}}\right)-\frac{C\mathrm{E}|x_{k1}x_{k2}|^3}{[\mathrm{var}(x_{k1}x_{k2})]^{\frac{3}{2}}\sqrt{n}}\\
&\ge P\left(Z\ge\frac{\sqrt{n}[n^{-1/2}\sqrt{4\log p}-\rho]}{\sqrt{\kappa_1(1+2\rho^2)-\rho^2}}\right)-\frac{C\sqrt{\mathrm{E}|x_{k1}|^6\mathrm{E}|x_{k2}|^6}}{[\mathrm{var}(x_{k1}x_{k2})]^{\frac{3}{2}}\sqrt{n}}\\
&\ge P(Z\ge C(2-c_1)\sqrt{\log p})-\frac{C}{\sqrt{n}}\\
&\to 1+o(1),
\end{aligned}
$$

where the second inequality uses $t_p \le n^{-1/2}\sqrt{4\log p}$ when p is sufficiently large; the third inequality uses $\rho \ge c_1\sqrt{\log p/n}$; and the last step of convergence holds when $c_1 > 2$.

A.7.2. *Proof of (ii).* Recall the notation J_A and J_A^c in Section A.5. Under the considered alternative, when $(j_1, j_2) \in J_A$, $\mathrm{E}(x_{k,j_1}x_{k,j_2}) = \rho$; and when $(j_3, j_4) \in J_A^c$, $\mathrm{E}(x_{k,j_3}x_{k,j_4}) = 0$. We have

$$
\begin{aligned}
\text{(A.14)} \qquad & P(|\,\mathcal{U}(\infty)\,| \ge t_p) \\
\le \quad & \sum_{1\le j_1<j_2\le p} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n\Big| \ge t_p\Big) \\
\le \quad & \frac{1}{2}\sum_{(j_1,j_2)\in J_A} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n\Big| \ge t_p\Big) \\
& +\frac{1}{2}\sum_{(j_3,j_4)\in J_A^c} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n\Big| \ge t_p\Big).
\end{aligned}
$$

Next we show that under the conditions of Proposition 2.4,

$$
\text{(A.15)} \qquad \sum_{(j_1,j_2)\in J_A} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n\Big| \ge t_p\Big) \to 0,
$$

and

$$
\text{(A.16)} \qquad \frac{1}{2}\sum_{(j_3,j_4)\in J_A^c} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n\Big| \ge t_p\Big) \le \log(1-\alpha)^{-1}.
$$

Proof of (A.15). To prove (A.15), we derive an upper bound of $P(|\sum_{k=1}^n x_{k,j_1}x_{k,j_2}/n| \ge t_p)$ for each $(j_1, j_2) \in J_A$ by Lemma 6.8 in Cai and Jiang [9]. In the following, we consider a fixed index pair (j_1, j_2); and for easy presentation, we write $m_0 = \sqrt{\mathrm{var}(x_{k,j_1}x_{k,j_2})}$ and $\xi_k = (x_{k,j_1}x_{k,j_2} - \rho)/m_0$. When $(j_1, j_2) \in J_A$, we have $\mathrm{E}(\xi_k) = 0$, $\mathrm{var}(\xi_k) = 1$, and by Condition 2.2*, $m_0^2 = \kappa_1(1 + 2\rho^2) - \rho^2$. It follows that

$$
P\Big(\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n \ge t_p\Big) = P\Big(\frac{\sum_{k=1}^n \xi_k}{\sqrt{n\log p}} \ge y_n\Big),
$$

where $y_n = \sqrt{n/\log p}\, m_0^{-1}(t_p - \rho)$. We next show that y_n and $\xi_k, k = 1, \dots, n$ satisfy the conditions of Lemma 6.8 in [9]. First note that $y_n \to y = (2 - c_2)m_0^{-1}$, and $y > 0$ as $c_2 < 2$. We then show that $\mathrm{E}\{\exp(\tilde{t}_0|\xi_k|^{\vartheta})\} < \infty$ for some $\tilde{t}_0 > 0$ and $0 < \vartheta \le 1$. In particular, given ς and t_0 in Proposition 2.4, we take $\vartheta = \varsigma/2 \in (0,1]$ and $\tilde{t}_0 = t_0(2m_0)^\vartheta/2 > 0$. By Lemma B.4,

$$\begin{aligned}
&|x_{k,j_1}x_{k,j_2} - \rho|^\vartheta \le (|x_{k,j_1}x_{k,j_2}| + |\rho|)^\vartheta \le |x_{k,j_1}x_{k,j_2}|^\vartheta + |\rho|^\vartheta \\
&\le \Big(\frac{x_{k,j_1}^2 + x_{k,j_2}^2}{2}\Big)^\vartheta + |\rho|^\vartheta \le \frac{1}{2^\vartheta}(|x_{k,j_1}|^{2\vartheta} + |x_{k,j_2}|^{2\vartheta}) + |\rho|^\vartheta.
\end{aligned}$$

It follows that

$$\begin{aligned}
(\text{A.17}) \qquad & \mathrm{E}\exp(\tilde{t}_0|\xi_k|^\vartheta) \\
\le\ & \mathrm{E}\exp\Big[\frac{\tilde{t}_0}{(2m_0)^\vartheta}(|x_{k,j_1}|^{2\vartheta} + |x_{k,j_2}|^{2\vartheta}) + \frac{\tilde{t}_0}{m_0^\vartheta}|\rho|^\vartheta\Big] \\
=\ & \mathrm{E}[\exp(2^{-1}t_0|x_{k,j_1}|^\varsigma) \times \exp(2^{-1}t_0|x_{k,j_2}|^\varsigma)] \times \exp(t_0 2^{\vartheta-1}|\rho|^\vartheta) \\
\le\ & \sqrt{\mathrm{E}[\exp(t_0|x_{k,j_1}|^\varsigma)] \times \mathrm{E}[\exp(t_0|x_{k,j_2}|^\varsigma)]} \times \exp(t_0 2^{\vartheta-1}|\rho|^\vartheta),
\end{aligned}$$

where the last inequality follows from the Hölder's inequality. By the conditions in Proposition 2.4, we know $\max_{(j_1,j_2)\in J_A} \mathrm{E}(t_0|x_{k,j_1}|^\varsigma) \times \mathrm{E}(t_0|x_{k,j_2}|^\varsigma) < \infty$ and $\rho \le c_2\sqrt{\log p/n} = o(1)$. Therefore, (A.17) $< \infty$. In summary, y_n and $\xi_k, k = 1, \ldots, n$ satisfy the conditions of Lemma 6.8 in [9].

By Lemma 6.8 in [9], as $\log p = o(n^\beta)$ and $\beta = \vartheta/(2+\vartheta) = \varsigma/(4+\varsigma)$,

$$(\text{A.18}) \qquad P\Big(\frac{\sum_{k=1}^n \xi_k}{\sqrt{n\log p}} \ge y_n\Big) \simeq \frac{p^{-y_n^2/2}(\log p)^{-1/2}}{\sqrt{2\pi}y}.$$

Let $z_0 = -\log(8\pi) - 2\log\log(1-\alpha)^{-1}$, then we can write $t_p = n^{-1/2}\{4\log p - \log\log p + z_0\}^{1/2}$ and

$$\begin{aligned}
y_n^2 &= \frac{n}{\log p}(t_p - \rho)^2 \times \frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})} \\
&= \frac{n}{\log p}(t_p^2 - 2\rho t_p + \rho^2) \times \frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})} \\
&\ge \frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})} \times \Big\{\frac{1}{\log p}\Big(4\log p - \log\log p + z_0\Big) \\
&\quad - \frac{2c_2\sqrt{\log p}\sqrt{4\log p - \log\log p + z_0}}{\log p} + \frac{c_2^2\log p}{\log p}\Big\},
\end{aligned}$$

where the last inequality holds when $\rho \le c_2\sqrt{\log p/n}$ and $c_2 < 2$. Then

$$
\begin{aligned}
&p^{-y_n^2/2}\\
=\ & \exp(-(\log p)y_n^2/2)\\
\le\ & \exp\Big\{-\frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})}\Big[\frac{1}{2}\Big(4\log p-\log\log p-\log(8\pi)-2\log\log(1-\alpha)^{-1}\Big)\\
&\qquad\qquad -c_2\sqrt{\log p}\sqrt{4\log p-\log\log p+z_0}+\frac{c_2^2\log p}{2}\Big]\Big\}\\
=\ & \Big\{p^{-2}\sqrt{\log p}\times\sqrt{8\pi}\log(1-\alpha)^{-1}\times p^{-\frac{c_2^2}{2}}\\
&\quad\times\exp\Big(c_2\sqrt{\log p}\sqrt{4\log p-\log\log p+z_0}\Big)\Big\}^{1/\{\mathrm{var}(x_{k,j_1}x_{k,j_2})\}}.
\end{aligned}
$$

By Condition 2.2*, $\mathrm{var}(x_{k,j_1}x_{k,j_2}) = \kappa_1 + (2\kappa_1-1)\rho^2$; and as $\rho = o(1)$, there exists a constant $m > 0$ such that $\mathrm{var}(x_{k,j_1}x_{k,j_2}) \le \kappa_1 + m$. Thus

$$
\begin{aligned}
&p^{-y_n^2/2}(\log p)^{-1/2}\\
\le\ & (\log p)^{-1/2}\Big\{p^{-2}\sqrt{\log p}\times\sqrt{8\pi}(\log(1-\alpha)^{-1})p^{-\frac{c_2^2}{2}}\\
&\times\exp\Big(c_2\sqrt{\log p}\sqrt{4\log p-\log\log p+z_0}\Big)\Big\}^{1/\{\mathrm{var}(x_{k1}x_{k2})\}}\\
\le\ & (\log p)^{-1/2}\Big[\sqrt{8\pi}\log(1-\alpha)^{-1}\sqrt{\log p}\times p^{-2-\frac{c_2^2}{2}+2c_2}\Big]^{1/(\kappa_1+m)}.
\end{aligned}
$$

Recall that $y = (2-c_2)[\mathrm{var}(x_{k,j_1}x_{k,j_2})]^{-1/2}$. Then by (A.18),

$$
\begin{aligned}
&\frac{1}{2}\sum_{(j_1,j_2)\in J_A}P\Big(\sum_{k=1}^n x_{k,j_1}x_{k,j_2}/n\ge t_p\Big)\\
&=\frac{1}{2}\sum_{(j_1,j_2)\in J_A}P\Big(\frac{\sum_{k=1}^n\xi_k}{\sqrt{n\log p}}\ge y_n\Big)\\
&\le\frac{|J_A|}{2}\frac{(\log p)^{-1/2}}{y\sqrt{2\pi}}\Big(\sqrt{8\pi}\log(1-\alpha)^{-1}\sqrt{\log p}\times p^{-2-\frac{c_2^2}{2}+2c_2}\Big)^{\frac{1}{\kappa_1+m}}\\
&=C_\alpha\exp\Bigg(2\log\Big[p^{-\frac{(1-c_2+c_2^2/4)}{(\kappa_1+m)}}\Big\{\sqrt{|J_A|}(\log p)^{\frac{1}{4(\kappa_1+m)}-\frac{1}{4}}\Big\}\Big]\Bigg),
\end{aligned}
\tag{A.19}
$$

where $C_\alpha = \frac{1}{2y\sqrt{2\pi}}[\sqrt{8\pi}\log(1-\alpha)^{-1}]^{1/(\kappa_1+m)}$. Thus, (A.19) $\to 0$ when

$$p^{-\frac{(1-c_2/2)^2}{\kappa_1+m}}\sqrt{|J_A|}(\log p)^{\frac{1}{4(\kappa_1+m)}-\frac{1}{4}}\to 0.$$

Similarly, we have

$$\text{(A.20)}\qquad \sum_{(j_1,j_2)\in J_A} P\Big(\frac{\sum_{k=1}^n x_{k,j_1}x_{k,j_2}}{n} \le -t_p\Big)$$
$$= \sum_{(j_1,j_2)\in J_A} P\Big(\frac{\sum_{k=1}^n(-x_{k,j_1}x_{k,j_2}+\rho)}{n\sqrt{\mathrm{var}(x_{k,j_1}x_{k,j_2})}} \ge \frac{t_p+\rho}{\sqrt{\mathrm{var}(x_{k,j_1}x_{k,j_2})}}\Big),$$

and (A.20) $\to 0$ following the similar arguments as above. In summary, (A.15) holds when $J_A = o(1)p^{\frac{2(1-c_2/2)^2}{\kappa_1+m}}(\log p)^{\frac{1}{2}-\frac{1}{2(\kappa_1+m)}}$ for some $m>0$.

Proof of (A.16). Similarly to Section A.7.2, we derive an upper bound of $P(\sum_{k=1}^n x_{k,j_3}x_{k,j_4}/n \ge t_p)$ for each $(j_3,j_4)\in J_A^c$ by Lemma 6.8 in [9]. In the following, we consider a fixed index pair (j_3,j_4); and for easy presentation, we write $\tilde{\xi}_k = x_{k,j_3}x_{k,j_4}/\sqrt{\kappa_1}$, $k=1,\dots,n$. When $(j_3,j_4)\in J_A^c$, $\mathrm{E}(x_{k,j_3}x_{k,j_4}) = 0$ and $\mathrm{var}(x_{k,j_3}x_{k,j_4}) = \mathrm{E}\{(x_{k,j_3}x_{k,j_4})^2\} = \kappa_1$, then we have $\mathrm{E}(\tilde{\xi}_k)=0$ and $\mathrm{var}(\tilde{\xi}_k)=1$. To prove (A.16), we write

$$P\Big(\sum_{k=1}^n x_{k,j_3}x_{k,j_4}/n \ge t_p\Big) = P\Big(\frac{\sum_{k=1}^n \tilde{\xi}_k}{\sqrt{n\log p}} \ge \tilde{y}_n\Big),$$

where $\tilde{y}_n = \sqrt{n/\log p}\times t_p/\sqrt{\kappa_1} \to \tilde{y} = 2/\sqrt{\kappa_1}$. Similarly to Section A.7.2, we know $\tilde{y}_n$ and $\tilde{\xi}_k$, $k=1,\dots,n$ also satisfy the conditions of Lemma 6.8 in [9]. Thus by Lemma 6.8 in [9], for $z_0 = -\log(8\pi) - 2\log\log(1-\alpha)^{-1}$ and $t_p = n^{-1/2}\sqrt{4\log p - \log\log p + z_0}$,

$$\begin{aligned}
& P\Big(\frac{\sum_{k=1}^n \tilde{\xi}_k}{\sqrt{n\log p}} \ge \tilde{y}_n\Big)\\
\simeq\ & \frac{p^{-\tilde{y}_n^2/2}(\log p)^{-1/2}}{\sqrt{2\pi}\tilde{y}}\\
=\ & p^{-2/\kappa_1}(\log p)^{1/(2\kappa_1)-1/2}\frac{\exp(-z_0/(2\kappa_1))}{\sqrt{2\pi}\tilde{y}}\\
\le\ & (8\pi)^{1/(2\kappa_1)}\frac{\sqrt{\kappa_1}}{2\sqrt{2\pi}}p^{-2/\kappa_1}(\log p)^{1/(2\kappa_1)-1/2}\{\log(1-\alpha)^{-1}\}^{1/\kappa_1}.
\end{aligned}$$

Then for $\kappa_1 \le 1$ and a small $\alpha > 0$,

$$\text{(A.21)}\qquad \frac{1}{2}\sum_{(j_1,j_2)\in J_A^c} P\Big(\sum_{k=1}^n x_{k,j_3}x_{k,j_4}/n \ge t_p\Big)$$
$$\le \frac{1}{2}\frac{p(p-1)-|J_A|}{p^{2/\kappa_1}(\log p)^{-1/(2\kappa_1)+1/2}}(8\pi)^{1/(2\kappa_1)}\frac{\sqrt{\kappa_1}}{2\sqrt{2\pi}}\{\log(1-\alpha)^{-1}\}^{1/\kappa_1},$$

which attains the maximum order at $\kappa_1 = 1$, when $\kappa_1 \le 1$ and $n, p \to \infty$. Therefore asymptotically, (A.21) $\le 2^{-1}\log(1-\alpha)^{-1}$. By similar arguments, we know when $n, p \to \infty$,

$$\frac{1}{2}\sum_{(j_3,j_4)\in J_A^c} P\Big(\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n \le -t_p\Big) \le \frac{1}{2}\log(1-\alpha)^{-1}.$$

In summary, we have (A.16) holds.

Combining (A.15) and (A.16), we obtain (A.14) $\le \log(1-\alpha)^{-1}$. □

A.8. Conditions of Theorems 4.1–4.5. The conditions of Theorem 4.1 are listed in the following Condition A.2.

CONDITION A.2.

(1) $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^4 < \infty$*;* $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^2 > 0$.

(2) $\mathbf{x}$ *is* α*-mixing with* $\alpha_x(s) \le M\delta^s$*, where* $\delta \in (0,1)$ *and* $M > 0$ *are some constants. In addition,* $\sum_{j_1,j_2=1}^{p}\sigma_{j_1,j_2}^a = \Theta(p)$.

Condition A.2 is similar to Conditions 2.1 and 2.2 of Theorem 2.1. As the mean is a lower order moment function than the covariance, Condition A.2 (1) is weaker than Condition 2.1 in that only the fourth moments are needed to be uniformly bounded instead of the eighth moments. Condition A.2 (2) is a regularization condition of the structure of the covariance matrix.

The conditions of Theorem 4.2 are list in the following Condition A.3.

CONDITION A.3.

(1) There exists constant B *such that* $B^{-1} \le \lambda_{\min}(\mathbf{\Sigma}) \le \lambda_{\max}(\mathbf{\Sigma}) \le B$*, where* $\lambda_{\min}(\mathbf{\Sigma})$ *and* $\lambda_{\max}(\mathbf{\Sigma})$ *denote the minimum and maximum eigenvalues of the covariance matrix* $\mathbf{\Sigma}$*; and all correlations are bounded away from* -1 *and* 1*, i.e.,* $\max_{1\le j_1\ne j_2\le p}|\sigma_{j_1,j_2}|/(\sigma_{j_1,j_2}\sigma_{j_2,j_2})^{1/2} < 1-\eta$ *for some* $\eta > 0$.

(2) $\log p = o(n^{1/4})$*;* $\max_{1\le j\le p}\mathrm{E}[\exp(h(x_j-\mu_j)^2)] < \infty$*, for* $h \in [-M_1, M_1]$*, where* $M_1 > 0$ *is some constant.*

(3) $\{(x_{i,j}, i = 1,\dots,n) : 1 \le j \le p\}$ *is* α*-mixing with* $\alpha_x(s) \le C\delta^s$*, where* $\delta \in (0,1)$ *and* $C > 0$ *is some constant, and* $\sum_{j_1,j_2=1}^{p}\sigma_{j_1,j_2}^a = \Theta(p)$.

In Condition A.3, (1) and (2) are assumed to establish the extreme value distribution of $\mathcal{U}(\infty)$, as in Cai et al. [11] and Xu et al. [76]. Furthermore, the mixing condition in Condition (3) is used to establish the joint independence of finite order U-statistics and $\mathcal{U}(\infty)$, following the argument in Hsing [39].

The conditions of Theorems 4.3–4.5 are listed in the Condition A.4 below.

CONDITION A.4.

(1) *There exists constant B such that $B^{-1} \leq \lambda_{\min}(\mathbf{\Sigma}_x) \leq \lambda_{\max}(\mathbf{\Sigma}_x) \leq B$, where $\lambda_{\min}(\mathbf{\Sigma}_x)$ and $\lambda_{\max}(\mathbf{\Sigma}_x)$ denote the minimum and maximum eigenvalues of $\mathbf{\Sigma}_x$; and all correlations are bounded away from -1 and 1, i.e., $\max_{1\leq j_1\neq j_2\leq p} |\sigma_{x,j_1,j_2}|/(\sigma_{x,j_1,j_2}\sigma_{x,j_2,j_2})^{1/2} < 1-\eta$ for some $\eta > 0$. In addition, we assume the same assumptions hold for $\mathbf{\Sigma}_y$.*

(2) *$n, p \to \infty$, $\log p = o(1)n^{1/4}$ and $n_x/n \to \gamma \in (0,1)$. In addition, $\max_{1\leq j\leq p} \mathrm{E}[\exp(h(x_j-\mu_j)^2)] < \infty$ and $\max_{1\leq j\leq p} \mathrm{E}[\exp(h(y_j-\nu_j)^2)] < \infty$, for $h \in [-M, M]$, where M is a positive constant.*

(3) *$\{(x_{i,j}, i = 1,\ldots,n) : 1 \leq j \leq p\}$ and $\{(y_{i,j}, i = 1,\ldots,n) : 1 \leq j \leq p\}$ are α-mixing with $\alpha_x(s) \leq C\delta_x^s$ and $\alpha_y(s) \leq C\delta_y^s$, where $\delta_x, \delta_y \in (0,1)$ and C is some constant. We also assume $\sum_{j_1,j_2=1}^{p} \{\sigma_{x,j_1,j_2}/\gamma + \sigma_{y,j_1,j_2}/(1-\gamma)\}^a = \Theta(p)$.*

Condition A.4 is similar to Condition A.3. They are assumed to establish both the limiting distributions and asymptotic independence properties of $\mathcal{U}(a)$ and $\mathcal{U}(\infty)$ for testing two-sample mean.

A.9. Proof of Theorems 4.1 and 4.2.

PROOF. Under H_0, for $\mathcal{U}(a)$ in (4.1), we assume without loss of generality that $\boldsymbol{\mu}_0 = \mathbf{0}$, and then write $\mathcal{U}(a) = \sum_{j=1}^{p} (P_a^n)^{-1} \sum_{1\leq i_1\neq\cdots\neq i_a\leq n} \prod_{k=1}^{a} x_{i_k,j}$.

We start with the proof of Theorem 4.1. Similarly to Section A.2, we first derive the variances and the covariances of the U-statistics; and then prove the asymptotic joint normality of the U-statistics. In particular, for $\mathrm{var}\{\mathcal{U}(a)\}$ in Theorem 4.1, as $\mathrm{E}\{\mathcal{U}(a)\} = 0$ under H_0,

$$\mathrm{var}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}^2(a)\} = (P_a^n)^{-2} \sum_{\substack{1\leq j_1\leq p,\\ 1\leq j_2\leq p}} \sum_{\substack{1\leq i_1\neq\cdots\neq i_a\leq n,\\ 1\leq \tilde{i}_1\neq\cdots\neq \tilde{i}_a\leq n}} \mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1} x_{\tilde{i}_k,j_2}\Big).$$

Note that $\mathrm{E}(\prod_{k=1}^{a} x_{i_k,j_1} x_{\tilde{i}_k,j_2}) = 0$ when $\{i_1,\ldots,i_a\} \neq \{\tilde{i}_1,\ldots,\tilde{i}_a\}$; and $\mathrm{E}(\prod_{k=1}^{a} x_{i_k,j_1} x_{\tilde{i}_k,j_2}) = \sigma_{j_1,j_2}^a$ when $\{i_1,\ldots,i_a\} = \{\tilde{i}_1,\ldots,\tilde{i}_a\}$. Then

$$\mathrm{var}\{\mathcal{U}(a)\} = (P_a^n)^{-1} \sum_{1\leq j_1,j_2\leq p} a!\sigma_{j_1,j_2}^a. \tag{A.22}$$

By Condition A.2, $\sum_{1\le j_1,j_2\le p}\sigma_{j_1,j_2}^a=\Theta(p)$. Thus $\mathrm{var}\{\mathcal{U}(a)\}=\Theta(pn^{-a})$.

Second, we show that $\mathrm{cov}\{\mathcal{U}(a),\mathcal{U}(b)\}=0$. Note that $\mathrm{cov}\{\mathcal{U}(a),\mathcal{U}(b)\}=\mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\}$ under H_0, and

$$\mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\}=(P_a^nP_b^n)^{-1}\sum_{\substack{1\le j_1\le p,\\ 1\le j_2\le p}}\ \sum_{\substack{1\le i_1\ne\cdots\ne i_a\le n,\\ 1\le \tilde{i}_1\ne\cdots\ne\tilde{i}_a\le n}}\mathrm{E}\Big(\prod_{k=1}^a x_{i_k,j_1}\prod_{t=1}^b x_{\tilde{i}_t,j_2}\Big).$$

Since $a\ne b$, $\{i_1,\dots,i_a\}\ne\{\tilde{i}_1,\dots,\tilde{i}_b\}$. Suppose there exists an index $i\in\{i_1,\dots,i_a\}$ and $i\notin\{\tilde{i}_1,\dots,\tilde{i}_b\}$. Then under H_0,

$$\mathrm{E}\Big(\prod_{k=1}^a x_{i_k,j_1}\prod_{t=1}^b x_{\tilde{i}_t,j_2}\Big)=\mathrm{E}(x_{i,j})\mathrm{E}(\text{all the remaining terms})=0.$$

Therefore, $\mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\}=0$.

In summary, the covariance matrix of $[\mathcal{U}(a_1)/\sigma(a_1),\dots,\mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ asymptotically converges to I_m. To finish the proof of Theorem 4.1, it remains to show that the joint limiting distribution of the U-statistics is normal. By the Cramér-Wold theorem, it is sufficient to prove that any fixed linear combination of these U-statistics converges to a normal distribution. Similarly to Section A.2, we use the martingale central limit theorem [6, p.476]. Specifically, we redefine Z_n as below with $\sum_{r=1}^m t_r^2=1$, and prove that

$$Z_n:=\sum_{r=1}^m t_r\mathcal{U}(a_r)/\sigma(a_r)\xrightarrow{D}\mathcal{N}(0,1). \tag{A.23}$$

With the redefined Z_n, we define $\mathrm{E}_k(\cdot)$ in the same way as in Section A.2, and still define $D_{n,k}=(\mathrm{E}_k-\mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2=\mathrm{E}_{k-1}(D_{n,k}^2)$. Similarly to Section A.2, we have $D_{n,k}=(\mathrm{E}_k-\mathrm{E}_{k-1})Z_n=\sum_{r=1}^m t_rA_{n,k,a_r}$, where we redefine $A_{n,k,a_r}=(\mathrm{E}_k-\mathrm{E}_{k-1})\{\mathcal{U}(a_r)/\sigma(a_r)\}$. In addition, similarly to Lemma A.4, we obtain that when $k<a_r$, $A_{n,k,a_r}=0$; and when $k\ge a_r$,

$$A_{n,k,a_r}=\frac{a_r}{\sigma(a_r)P_{a_r}^n}\sum_{j=1}^p\ \sum_{1\le i_1\ne\cdots\ne i_{a_r-1}\le k-1}x_{k,j}\times\prod_{t=1}^{a_r-1}x_{i_t,j}.$$

Given the form of A_{n,k,a_r}, we can obtain the forms of $D_{n,k}$ and $\pi_{n,k}^2$. To prove (A.12), by the martingale central limit theorem, it suffices to prove the following Lemma A.15.

Lemma A.15. *Under the conditions of Theorem 4.1,* $\mathrm{var}(\sum_{k=1}^n\pi_{n,k}^2)\to 0$ *and* $\sum_{k=1}^n\mathrm{E}(D_{n,k}^4)\to 0$.

PROOF. *See Section B.5 on Page 158.* □

With Lemma A.15, the asymptotic joint normality in Theorem 4.1 is obtained by the martingale central limit theorem. For Theorem 4.2, the limiting distribution of $\mathcal{U}(\infty)$ follows from Cai et al. [11]. In addition, the asymptotic independence between $\mathcal{U}(a)/\sigma(a)$ and $n\mathcal{U}(\infty) - \tau_p$ can be obtained similarly as the proof of Theorem 4.4. We defer the details to Section A.11.

A.10. Proof of Theorem 4.3. By the following Proposition A.1, we assume that under H_0, $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$, without loss of generality.

PROPOSITION A.1. *$\mathcal{U}(a)$ constructed in (4.2) and (4.3) are location invariant; that is, for any vector $\boldsymbol{\Delta} \in \mathbb{R}^p$, the U-statistic constructed based on the transformed data $\{\mathbf{x}_i + \Delta : i = 1, \ldots, n_x\}$ and $\{\mathbf{y}_i + \Delta : i = 1, \ldots, n_y\}$ is still $\mathcal{U}(a)$.*

Proposition A.1 can be obtained straightforwardly from the definitions $\mathcal{U}(a) = \sum_{j=1}^{p} (P_a^{n_x} P_a^{n_y})^{-1} \times \sum_{\substack{1 \le k_1 \ne \ldots \ne k_a \le n_x; \\ 1 \le s_1 \ne \ldots \ne s_a \le n_y}} \prod_{t=1}^{a} (x_{k_t,j} - y_{s_t,j})$ in (4.2), and $\mathcal{U}(\infty) = \max_{1 \le j \le p} \sigma_{j,j}^{-1} \times (\bar{x}_j - \bar{y}_j)^2$ in (4.3). The proof is thus skipped.

The following proof proceeds by deriving the variances, covariances and asymptotic joint normality of the U-statistics. Particularly, the next Lemma A.16 derives the asymptotic form of $\sigma^2(a)$ in Theorem 4.3.

LEMMA A.16. *Under the conditions of Theorem 4.3,*

$$\operatorname{var}\{\mathcal{U}(a)\} \simeq \sum_{1 \le j_1, j_2 \le p} a! \Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y} \Big)^a = \Theta(pn^{-a}).$$

When $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2} = \sigma_{j_1,j_2}$, we have $\operatorname{var}[\mathcal{U}(a)] \simeq \sum_{j_1,j_2=1}^{p} a!(n_x + n_y)^a \sigma_{j_1,j_2}^a / (n_x n_y)^a$.

PROOF. *See Section B.6.1 on Page 163.* □

In addition, the following Lemma A.17 shows that different $\mathcal{U}(a)$'s of finite a are uncorrelated.

LEMMA A.17. *Under the conditions of Theorem 4.3, for finite integers $a \ne b$, $\operatorname{cov}\{\mathcal{U}(a), \mathcal{U}(b)\} = 0$.*

PROOF. *See Section B.6.2 on Page 164.* □

We then know $\mathrm{cov}\{\mathcal{U}(a_1)/\sigma(a_1),\ldots,\mathcal{U}(a_m)/\sigma(a_m)\} = I_m$ by Lemmas A.16 and A.17. The next Lemma A.18 further proves the asymptotic joint normality of the U-statistics.

LEMMA A.18. *Under the conditions of Theorem 4.3, for finite integers* $a_1,\ldots,a_m$, $\{\mathcal{U}(a_1)/\sigma(a_1),\ldots,\mathcal{U}(a_m)/\sigma(a_m)\} \xrightarrow{D} \mathcal{N}(0, I_m)$.

PROOF. *See Section B.6.3 on Page 165.* □

Combining Lemmas A.16–A.18, we finish the proof of Theorem 4.3.

A.11. Proof of Theorem 4.4. For $\mathcal{U}(\infty)$ in (4.3), the limiting distribution of $\mathcal{U}(\infty)$ is established in Cai et al. [11] and [76]. We next prove the asymptotic independence between $\mathcal{U}(\infty)$ and $\mathcal{U}(a)$ by a similar argument to that in Hsing [39], see also [76]. In this proof, we reserve the notation P for the probability measure on which $x_{i,j}$ and $y_{i,j}$ are defined, and the expectation with respect to P is denoted as E. Define $\tilde{\mathcal{U}}_c(a)/\sigma(a)$ on the conditional probability measure $\tilde{P}$, given the event $n_x n_y \mathcal{U}(\infty)/(n_x+n_y) - \tau_p \le u$ such that

$$\begin{aligned}&\tilde{P}\Big\{\tilde{\mathcal{U}}_c(a)/\sigma(a) \le u'\Big\}\\ &= P\Big\{\mathcal{U}(a)/\sigma(a) \le u' \;\Big|\; \frac{n_x n_y}{n_x+n_y}\mathcal{U}(\infty) \le \tau_p + u\Big\}.\end{aligned}$$

The expectation with respect to $\tilde{P}$ is denoted by $\tilde{\mathrm{E}}$. To show the asymptotic independence, it is sufficient to prove the following Lemma A.19.

LEMMA A.19. *Under the conditions of Theorem 4.4,* $\tilde{\mathcal{U}}_c(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$ *on the conditional measure* $\tilde{P}$.

PROOF. *See Section B.7 on Page 169.* □

A.12. Proof of Theorem 4.5. By Proposition A.1, we assume $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu} = \mathbf{0}$, without loss of generality. Then under the considered alternative $\mathcal{E}_A$, $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = \{\mu_j = \rho : j = 1,\ldots,k_0; \mu_j = 0 : j = k_0+1,\ldots,p\}$. Define $\varphi_{j_1,j_2} = \sigma_{j_1,j_2} + \mu_{j_1}\mu_{j_2}$. We have $\mathrm{E}(x_{i,j_1}x_{i,j_2}) = \varphi_{j_1,j_2}$, and under $\boldsymbol{\nu} = \mathbf{0}$, $\mathrm{E}(y_{i,j_1}y_{i,j_2}) = \sigma_{j_1,j_2}$.

Similarly to the proof of Theorem 2.5 in Section A.5, we decompose

$\mathcal{U}(a) = T_{a,1} + T_{a,2}$, where

$$T_{a,1} = \sum_{j=1}^{k_0} \sum_{c=0}^{a} \sum_{\substack{1 \le k_1 \ne \cdots \ne k_c \le n_x, \\ 1 \le s_1 \ne \cdots \ne s_{a-c} \le n_y}} G(a,c) \prod_{t=1}^{c} x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j}, \tag{A.24}$$

$$T_{a,2} = \sum_{j=k_0+1}^{p} \sum_{c=0}^{a} \sum_{\substack{1 \le k_1 \ne \cdots \ne k_c \le n_x, \\ 1 \le s_1 \ne \cdots \ne s_{a-c} \le n_y}} G(a,c) \prod_{t=1}^{c} x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j},$$

with $G(a,c) = (-1)^{a-c}\binom{a}{c}(P_c^{n_x} P_{a-c}^{n_y})^{-1}$. Then $\mathrm{E}(T_{a,1}) = \sum_{j=1}^{k_0}(\mu_j - \nu_j)^a = k_0\rho^a$ and $\mathrm{E}(T_{a,2}) = \sum_{j=k_0+1}^{p}(\mu_j - \nu_j)^a = 0$.

To prove Theorem 4.5, we derive the variances, covariances, and asymptotic joint normality of the U-statistics. Particularly, the next Lemma A.20 gives the asymptotic form of $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\}$, and shows that $T_{a,2}$ is the leading component.

Lemma A.20. *Under the conditions of Theorem 4.5,*

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq \sum_{k_0+1 \le j_1, j_2 \le p} a! \Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^a. \tag{A.25}$$

$\mathrm{var}(T_{a,2}) = \Theta(pn^{-a})$ *and* $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$. *It follows that* $\{T_{a,1} - \mathrm{E}(T_{a,1})\}/\sigma(a) \xrightarrow{P} 0$.

Proof. *See Section B.8.1 on Page 171.* □

In addition, the following Lemma A.21 shows that the covariance between two U-statistics asymptotically converges to 0.

Lemma A.21. *Under the conditions of Theorem 4.5, for two finite integers* $a \ne b$, $\{\sigma(a)\sigma(b)\}^{-1}\mathrm{cov}\{\mathcal{U}(a), \mathcal{U}(b)\} \to 0$.

Proof. *See Section B.8.2 on Page 174.* □

By the analysis above, we know that the covariance matrix of $[\{\mathcal{U}(a_1) - \mathrm{E}[\mathcal{U}(a_1)]\}/\sigma(a_1), \ldots, \{\mathcal{U}(a_m) - \mathrm{E}[\mathcal{U}(a_m)]\}/\sigma(a_m)]^\intercal$ asymptotically converges to I_m. To prove Theorem 4.5, it remains to show that the joint limiting distribution of the U-statistics is normal. By the Cramér-Wold theorem, it is equivalent to prove that any fixed linear combination of these U-statistics converges to a normal distribution. By Lemma A.20 and the Slutsky's theorem, it suffices to show that any fixed linear combination of

$[T_{a_1,2}/\sqrt{\mathrm{var}(T_{a_1,2})},\dots,T_{a_m,2}/\sqrt{\mathrm{var}(T_{a_m,2})}]^\intercal$ converges to a normal distribution for any finite m. Since $\mu_j=\nu_j$ for $j\in\{k_0+1,\dots,p\}$, and each $T_{a_t,2}$ is a summation over $j\in\{k_0+1,\dots,p\}$, we know the analysis under H_0 in Section A.10 can be applied to $T_{a_t,2}$ similarly. Given $k_0=o(p)$, we know $[T_{a_1,2}/\sqrt{\mathrm{var}(T_{a_1,2})},\dots,T_{a_m,2}/\sqrt{\mathrm{var}(T_{a_m,2})}]^\intercal$ has the joint asymptotic normality. In summary, Theorem 4.5 is proved.

A.13. Proof of Theorem 4.6. We first provide the details of the conditions of Theorem 4.6 in Section A.13.1 and then prove Theorem 4.6 in Section A.13.2.

A.13.1. *Conditions of Theorem 4.6.* Theorem 4.6 can be proved by the following Condition A.5 or Condition A.6. Note that Conditions A.5 and A.6 are assumed under H_0, where $\mathbf{\Sigma}_x=\mathbf{\Sigma}_y=\mathbf{\Sigma}=(\sigma_{j_1,j_2})_{p\times p}$.

CONDITION A.5.

(1) $n,p\to\infty$, *and* $n_x/n\to\gamma\in(0,1)$.
(2) $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^8<\infty$; $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^2>0$; $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^8<\infty$; *and* $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^2>0$.
(3) $\{(x_{i,j},i=1,\dots,n):1\le j\le p\}$ *and* $\{(y_{i,j},i=1,\dots,n):1\le j\le p\}$ *are* α*-mixing with* $\alpha_x(s)\le C\delta_x^s$ *and* $\alpha_y(s)\le C\delta_y^s$, *where* $\delta_x,\delta_y\in(0,1)$ *and* C *is some constant.*
(4) For any finite integer a, $\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a=\Theta(p^2)$.

Condition A.5 (2) is similar to Condition 2.1. Condition A.5 (3) assumes α-mixing on the two samples, which is similar to Condition 2.2. Condition A.5 (4) is a regularity condition on the covariance structure, and it is naturally satisfied for even a, given Condition A.5 (3).

Alternatively, we introduce another set of conditions similar to Condition 2.2*. We define some notation. Suppose $(z_1,\dots,z_p)^\intercal\sim\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$. Given indexes $1\le j_1,\dots,j_t\le p$, define $\Pi^0_{j_1,\dots,j_t}=\mathrm{E}(\prod_{k=1}^t z_{j_k})$. Moreover, we define $\Pi^x_{j_1,\dots,j_t}=\mathrm{E}\{\prod_{k=1}^t(x_{j_k}-\mu_{j_k})\}$ and $\Pi^y_{j_1,\dots,j_t}=\mathrm{E}\{\prod_{k=1}^t(y_{j_k}-\nu_{j_k})\}$. In addition, for given integers a and b, let $\mathbb{G}_{a,b}$ be a collection of tuples $\mathcal{G}=(g_1,g_2,\dots,g_{4(a+b)-1},g_{4(a+b)})\in\{1,\dots,8\}^{4(a+b)}$, which satisfies that $g_{2t-1}\neq g_{2t}$ for $t=1,\dots,2(a+b)$, and the number of g's equal to m is a for $m\in\{1,2,3,4\}$ and is b for $m\in\{5,6,7,8\}$. For any $\mathcal{G}\in\mathbb{G}_{a,b}$, we define $\mathbb{V}_{a,b,\mathcal{G}}=\sum_{1\le j_1,\dots,j_8\le p}\prod_{t=1}^{2(a+b)}\sigma_{j_{g_{2t-1}},j_{g_{2t}}}$, and let $S_{\mathcal{G}}$ denote the number of distinct sets among the $2(a+b)$ number of sets, $\{g_{2t-1},g_{2t}\}$, for $t=1,\dots,2(a+b)$, induced by $\mathcal{G}$. Note that generally $S_{\mathcal{G}}\ge 4$ and when $S_{\mathcal{G}} = 4$, by the symmetricity of j indexes, $\mathbb{V}_{a,b,\mathcal{G}} = \mathbb{V}_{a,b,0}$ where $\mathbb{V}_{a,b,0} := \sum_{1\le j_1,\dots,j_8\le p}\sum_{1\le j_1,\dots,j_8\le p}\sigma^a_{j_1,j_2}\sigma^a_{j_3,j_4}\sigma^b_{j_5,j_6}\sigma^b_{j_7,j_8}$.

Condition A.6.

(1) $n, p \to \infty$, *and* $n_x/n \to \gamma \in (0,1)$.
(2) $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^8 < \infty$; $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^2 > 0$; $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^8 < \infty$; *and* $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^2 > 0$.
(3) *For* $t = 3,4,6,8$, *there exist constants* $\kappa_{x,t}, \kappa_{y,t} \ge 1$ *such that* $\Pi^x_{j_1,\dots,j_t} = \kappa_{x,t}\Pi^0_{j_1,\dots,j_t}$ *and* $\Pi^y_{j_1,\dots,j_t} = \kappa_{y,t}\Pi^0_{j_1,\dots,j_t}$.
(4) *For* $a, b \in \{a_1,\dots,a_m\}$, *and any* $\mathcal{G} \in \mathbb{G}_{a,b}$ *define above, if* $S_{\mathcal{G}} > 4$, *we assume* $\mathbb{V}_{a,b,\mathcal{G}} = o(1)\mathbb{V}_{a,b,0}$.

We note that Condition A.6 (3) and (4) are alternative dependence assumptions to Condition A.5 (3) and (4). Condition A.6 (3) is an extension from Condition 2.2*, and is also satisfied when the distributions of $\mathbf{x}$ and $\mathbf{y}$ follow elliptical distributions [46]. Condition A.6 (4) implies some weak dependence structure in covariance matrix $\boldsymbol{\Sigma}$. To better illustrate the condition, we consider the case when $a = b = 2$ as an example. We note that

$$\mathbb{V}_{a,b,0} = \sum_{1\le j_1,\dots,j_8\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_5,j_6}\sigma_{j_7,j_8})^2 = \{\mathrm{tr}(\Sigma^2)\}^4,$$

and $\mathbb{V}_{a,b,0} = \mathbb{V}_{a,b,\mathcal{G}}$ when $\mathcal{G} = (1,2,3,4,5,6,7,8,1,2,3,4,5,6,7,8)$ with $S_{\mathcal{G}} = 4$. Moreover, if $\mathcal{G} = (1,3,2,4,1,2,3,4,5,6,7,8,5,6,7,8)$ with $S_{\mathcal{G}} = 6$,

$$\mathbb{V}_{a,b,\mathcal{G}} = \sum_{1\le j_1,\dots,j_8\le p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})(\sigma_{j_1,j_2}\sigma_{j_3,j_4})(\sigma_{j_5,j_6}\sigma_{j_7,j_8})^2 = \mathrm{tr}(\Sigma^4)\{\mathrm{tr}(\Sigma^2)\}^2;$$

if $\mathcal{G} = (1,3,2,4,1,2,3,4,5,7,6,8,5,6,7,8)$ with $S_{\mathcal{G}} = 8$,

$$\mathbb{V}_{a,b,\mathcal{G}} = \sum_{1\le j_1,\dots,j_8\le p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})(\sigma_{j_1,j_2}\sigma_{j_3,j_4})(\sigma_{j_5,j_7}\sigma_{j_6,j_8})(\sigma_{j_5,j_6}\sigma_{j_7,j_8}) = \{\mathrm{tr}(\Sigma^4)\}^2;$$

if $\mathcal{G} = (1,6,2,5,3,7,4,8,1,3,2,4,5,7,6,8)$ with $S_{\mathcal{G}} = 8$,

$$\mathbb{V}_{a,b,\mathcal{G}} = \sum_{1\le j_1,\dots,j_8\le p}(\sigma_{j_1,j_6}\sigma_{j_2,j_5})(\sigma_{j_3,j_7}\sigma_{j_4,j_8})(\sigma_{j_1,j_3}\sigma_{j_2,j_4})(\sigma_{j_5,j_7}\sigma_{j_6,j_8}) = \mathrm{tr}(\Sigma^8).$$

In this case, Condition A.6 (4) is equivalent to $\mathrm{tr}(\Sigma^4) = o[\{\mathrm{tr}(\Sigma^2)\}^2]$ and $\mathrm{tr}(\Sigma^8) = o[\{\mathrm{tr}(\Sigma^2)\}^4]$, which are similarly assumed in [54]. In addition, we consider another example where the $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ is of banded structure with bandwidth s and has the nonzero entries being positive constants. It follows that $\mathbb{V}_{a,b,0} = \Theta(p^4s^4)$ and $\mathbb{V}_{a,b,\mathcal{G}} = O(p^3s^5)$ when $S_{\mathcal{G}} > 4$. Therefore, in this example, Condition A.6 (4) is satisfied when $s = o(p)$.

A.13.2. *Proof of Theorem 4.6.* Since $\mathcal{U}(a)$ is location invariant, we assume $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{y}) = \mathbf{0}$, without loss of generality, in this section. We decompose $\mathcal{U}(a) = \tilde{\mathcal{U}}(a) + \tilde{\mathcal{U}}^*(a)$, where we redefine

$$\tilde{\mathcal{U}}(a) = \sum_{1\leq j_1,j_2\leq p} \frac{1}{P^{n_x}_a P^{n_y}_a} \sum_{\substack{1\leq i_1\neq\ldots\neq i_a\leq n_x;\\ 1\leq w_1\neq\ldots\neq w_a\leq n_y}} \prod_{t=1}^{a} (x_{i_t,j_1}x_{i_t,j_2} - y_{w_t,j_1}y_{w_t,j_2}),$$

and $\tilde{\mathcal{U}}^*(a) = \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$. To prove Theorem 4.6, we derive the variances, covariances, and asymptotic joint normality of the U-statistics. Particularly, the following Lemma A.22 derives the asymptotic form of $\mathrm{var}\{\mathcal{U}(a)\}$, and shows that $\tilde{\mathcal{U}}(a)$ is the leading term.

LEMMA A.22. *Under the conditions of Theorem 4.6,* $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$, $\mathcal{U}^*(a)/\sigma(a) \xrightarrow{P} 0$, *and*

$$\begin{aligned}&\mathrm{var}\{\mathcal{U}(a)\}\\ \simeq &\sum_{1\leq j_1,j_2,j_3,j_4\leq p} a!\Big\{\frac{1}{n_x}(\Pi^x_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4}) + \frac{1}{n_y}(\Pi^y_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4})\Big\}^a.\end{aligned}$$

In particular, under Condition A.5, $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(p^2n^{-a})$; *under Condition A.6,* $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(n^{-a})\sum_{1\leq j_1,j_2,j_3,j_4\leq p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a$.

PROOF. *See Section B.9.1 on Page 174.* □

Given Lemma A.22, the next Lemma A.23 shows that the covariance between two U-statistics asymptotically converges to 0.

LEMMA A.23. *Under the conditions of Theorem 4.6, for finite integers* $a \neq b$, $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} \to 0$ *as* $n, p \to \infty$.

PROOF. *See Section B.9.2 on Page 180.* □

To finish the proof, it remains to obtain the joint asymptotic normality of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ for different finite integers $a_1, \ldots, a_m$. By Cramér-Wold theorem, it is equivalent to prove that any of their fixed linear combination converges to normal. In addition, by Lemma A.22 and the Slutsky's theorem, it suffices to prove that any fixed linear combination of $[\tilde{\mathcal{U}}(a_1)/\sigma(a_1), \ldots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m)]^\intercal$ converges to normal. Specifically, similarly to Section A.2, we redefine Z_n as below with $\sum_{r=1}^m t_r^2 = 1$, and prove that

$$Z_n := \sum_{r=1}^{m} t_r\tilde{\mathcal{U}}(a_r)/\sigma(a_r) \xrightarrow{D} \mathcal{N}(0,1). \tag{A.26}$$

We next prove (A.26) following the proof of Theorem 2.1 in Section A.2 and apply the martingale central limit theorem [6, p.476].

To construct a martingale difference, we write $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})^\intercal$ and $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,p})^\intercal$; and define a new random vector

$$R_i = \mathbf{x}_i \quad \text{for } i = 1, 2, \dots, n_x; \qquad R_{n_x+j} = \mathbf{y}_j \quad \text{for } j = 1, 2, \dots, n_y.$$

We then define $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k = \sigma\{R_1, \dots, R_k\}$ for $k = 1, 2, \dots, n_x + n_y$; and let $\mathrm{E}_k(\cdot)$ denote the conditional expectation given $\mathcal{F}_k$ for $k = 1, \cdots, n_x + n_y$. Define $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2 = \mathrm{E}_{k-1}(D_{n,k}^2)$. It follows that $Z_n = \sum_{k=1}^n D_{n,k}$ as $\mathrm{E}_0(Z_n) = \mathrm{E}(Z_n) = 0$. To prove (A.26), by the martingale central limit theorem, it suffices to prove

$$\sum_{k=1}^n \pi_{n,k}^2/\mathrm{var}(Z_n) \xrightarrow{P} 1 \quad \text{and} \quad \sum_{k=1}^n \mathrm{E}(D_{n,k}^4)/\mathrm{var}^2(Z_n) \to 0. \tag{A.27}$$

To prove (A.27), we derive the explicit forms of $D_{n,k}$ and $\pi_{n,k}^2$ in Section B.9.3. Similarly to Section A.2, the following Lemma A.24 and Lemma A.25 suggest that (A.27) holds.

Lemma A.24. *Under the conditions of Theorem 4.6,* $\mathrm{var}(\sum_{k=1}^{n_x+n_y} \pi_{n,k}^2) \to 0$.

Proof. *See Section B.9.4 on Page 183.* □

Lemma A.25. *Under the conditions of Theorem 4.6,* $\sum_{k=1}^{n_x+n_y} \mathrm{E}(D_{n,k}^4) \to 0$.

Proof. *See Section B.9.5 on Page 190.* □

In summary, Theorem 4.6 is proved.

A.14. Proof of Theorem 4.7. In this section, we first provide the conditions of Theorem 4.7 in Section A.14.1 and then prove Theorem 4.7 in Section A.14.2.

A.14.1. *Conditions.* Theorem 4.7 is established under the following Conditions A.7 and A.8, where Condition A.7 is the same as Condition A.6 (1)–(3).

Condition A.7.

(1) $n, p \to \infty$, *and* $n_x/n \to \gamma \in (0, 1)$.

(2) $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^8<\infty$; $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^2>0$; $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^8<\infty$; *and* $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^2>0$.

(3) *For* $t=3,4,6,8$, *there exist* $\kappa_{x,t},\kappa_{y,t}\ge 1$ *such that* $\Pi^x_{j_1,\ldots,j_t}=\kappa_{x,t}\Pi^0_{j_1,\ldots,j_t}$ *and* $\Pi^y_{j_1,\ldots,j_t}=\kappa_{y,t}\Pi^0_{j_1,\ldots,j_t}$.

To provide Condition A.8, we first define some notation. The difference between $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ is defined as $D_{x,y}=\mathbf{\Sigma}_x-\mathbf{\Sigma}_y=(D_{j_1,j_2})_{p\times p}$. Let $\mathbb{J}_0\subseteq\{1,\ldots,p\}$ be the largest set such that for any $j_1,j_2\in\mathbb{J}_0$, $\sigma_{x,j_1,j_2}=\sigma_{y,j_1,j_2}$. Define $J_{0,D}=\{(j_1,j_2):j_1\text{ or }j_2\notin\mathbb{J}_0\}$. Given $\mathbb{J}_0$ and $a,b\in\{a_1,\ldots,a_m\}$, we define $\mathbb{V}_{a,b,0,0}=\sum_{j_1,\ldots,j_8\in\mathbb{J}_0}(\sigma_{x,j_1,j_2}\sigma_{x,j_3,j_4})^a(\sigma_{x,j_5,j_6}\sigma_{x,j_7,j_8})^b$, which also equals to $\sum_{j_1,\ldots,j_8\in\mathbb{J}_0}(\sigma_{y,j_1,j_2}\sigma_{y,j_3,j_4})^a(\sigma_{y,j_5,j_6}\sigma_{y,j_7,j_8})^b$ by the definition of $\mathbb{J}_0$. In addition, for any tuple $\mathcal{G}=(g_1,g_2,\ldots,g_{4(a+b)-1},g_{4(a+b)})\in\mathbb{G}_{a,b}$ specified in Condition A.6, we define $\mathbb{V}_{a,b,\mathcal{G},0}=\sum_{j_1,\ldots,j_8\in\mathbb{J}_0}\prod_{t=1}^{2(a+b)}\sigma_{j_{g_{2t-1}},j_{g_{2t}}}$. Note that $\mathbb{V}_{a,b,0,0}$ and $\mathbb{V}_{a,b,\mathcal{G},0}$ are defined similarly to $\mathbb{V}_{a,b,0}$ and $\mathbb{V}_{a,b,\mathcal{G}}$ in Condition A.6 by changing the range of j indexes from $\{1,\ldots,p\}$ to $\mathbb{J}_0$. Moreover, let $\mathcal{H}=\{(h_1,h_2),(h_3,h_4)\}\in\mathbb{H}$, where $\mathbb{H}$ includes $\{(1,2),(3,4)\}$, $\{(1,3),(2,4)\}$ and $\{(1,4),(2,3)\}$. For any $a\in\{a_1,\ldots,a_m\}$ and given $\mathcal{H}\in\mathbb{H}$, define

$$
\begin{aligned}
\mathbb{V}_{a,\mathcal{H},x,1}&=\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}}|\sigma_{x,j_{h_1},j_{h_2}}\sigma_{x,j_{h_3},j_{h_4}}|^a\\
\mathbb{V}_{a,\mathcal{H},x,2}&=\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}}|D_{j_{h_1},j_{h_2}}\sigma_{x,j_{h_3},j_{h_4}}|^a,\\
\mathbb{V}_{a,\mathcal{H},D,3}&=\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}}|D_{j_{h_1},j_{h_2}}D_{j_{h_3},j_{h_4}}|^a.
\end{aligned}
\tag{A.28}
$$

Similarly, we also define $\mathbb{V}_{a,\mathcal{H},y,1}$ and $\mathbb{V}_{a,\mathcal{H},y,2}$ by replacing σ_x's with σ_y's. We next present Condition A.8 of Theorem 4.7.

CONDITION A.8. *For any* $a,b\in\{a_1,\ldots,a_m\}$, $\mathcal{G}\in\mathbb{G}_{a,b}$, *and* $\mathcal{H}\in\mathbb{H}$, *we assume (A1)* $\mathbb{V}_{a,b,\mathcal{G},0}=o(1)\mathbb{V}_{a,b,0,0}$; *(A2)* $\mathbb{V}_{a,\mathcal{H},D,3}=O(n^{-a})\mathbb{V}^{1/2}_{a,a,0,0}$; *and (A3)* $\mathbb{V}_{a,\mathcal{H},x,t}=o(1)\mathbb{V}^{1/2}_{a,a,0,0}$, *for* $t=1,2$.

Equivalently we can also replace (A3) in Condition A.8 by (A3)* $\mathbb{V}_{a,\mathcal{H},y,t}=o(1)\mathbb{V}^{1/2}_{a,a,0,0}$, for $t=1,2$. This is because by $D_{j_1,j_2}=\sigma_{x,j_1,j_2}-\sigma_{y,j_1,j_2}$ and the Hölder's inequality, we know *(A2)* and *(A3)* induce *(A3)**; and *(A2)* and *(A3)** also induce *(A3)*. Thus it is equivalent to assume *(A3)* or *(A3)** in Condition A.8.

We next discuss Condition A.8. Let $\mathbf{\Sigma}_C=\{\sigma_{x,j_1,j_2}:j_1,j_2\in\mathbb{J}_0\}=\{\sigma_{y,j_1,j_2}:j_1,j_2\in\mathbb{J}_0\}$, which is the common submatrix of $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$

by the definition of $\mathbb{J}_0$. In Condition A.8, *(A1)* implies some weak dependence structure of $\mathbf{\Sigma}_C$ similar to Condition A.6 (4). We consider an example where $\mathbf{\Sigma}_x$ has the banded structure with the bandwidth s and the entries being positive constants. Then *(A1)* holds if $s = o(p)$. Moreover, under the considered example, $\mathbb{V}_{a,a,0,0}^{1/2} = (\sum_{j_1,j_2\in\mathbb{J}_0} \sigma_{x,j_1,j_2}^a)^2 \geq C|\mathbb{J}_0|^4$ and $\mathbb{V}_{a,\mathcal{H},x,1} \leq C|J_{0,D}|^2 = C_2(p-|\mathbb{J}_0|)^4$. Then *(A3)* for $t=1$ holds when $p - |\mathbb{J}_0| = o(p)$, which implies that the number of entries that are different in $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ is $o(p^2)$. In addition, *(A2)* and *(A3)* for $t=2$ are regularity conditions on the difference matrix $D_{x,y}$. For illustration, we consider an example where $D_{j_1,j_2} = \rho > 0$ for any $(j_1,j_2) \in J_{0,D}$, and $\mathbf{\Sigma}_x = I_p$. Then $\mathbb{V}_{a,a,0,0}^{1/2} = |\mathbb{J}_0|^2$, $\mathbb{V}_{a,\mathcal{H},x,2} \leq |J_{0,D}|\rho^a p$, and $\mathbb{V}_{a,\mathcal{H},D,3} \leq |J_{0,D}|^2\rho^{2a}$. Under this example, *(A2)* and *(A3)* of $t=2$ hold if $|J_{0,D}|\rho^a = O(n^{-a/2}p)$ and $|\mathbb{J}_0| \simeq p$, which are similar to the assumption in Theorem 2.5.

A.14.2. *Proof.* In this section, we prove Theorem 4.7 under Conditions A.7 and A.8. Recall that we decompose $\mathcal{U}(a) = \tilde{\mathcal{U}}(a) + \tilde{\mathcal{U}}^*(a)$ in Section A.13. We further decompose $\tilde{\mathcal{U}}(a) = T_{D,a,1} + T_{D,a,2}$, where

$$T_{D,a,1} = \sum_{j_1,j_2\in\mathbb{J}_0} \frac{1}{P_a^{n_x} P_a^{n_y}} \sum_{\substack{1\leq i_1\neq\ldots\neq i_a\leq n_x;\\ 1\leq w_1\neq\ldots\neq w_a\leq n_y}} \prod_{t=1}^{a} (x_{i_t,j_1}x_{i_t,j_2} - y_{w_t,j_1}y_{w_t,j_2}),$$

$$T_{D,a,2} = \sum_{(j_1,j_2)\in J_{0,D}} \frac{1}{P_a^{n_x} P_a^{n_y}} \sum_{\substack{1\leq i_1\neq\ldots\neq i_a\leq n_x;\\ 1\leq w_1\neq\ldots\neq w_a\leq n_y}} \prod_{t=1}^{a} (x_{i_t,j_1}x_{i_t,j_2} - y_{w_t,j_1}y_{w_t,j_2}).$$

It follows that $\mathcal{U}(a) = T_{D,a,1} + T_{D,a,2} + \tilde{\mathcal{U}}^*(a)$. To prove Theorem 4.7, we derive the variances, covariances and asymptotic joint normality of the U-statistics. In particular, the next Lemma A.26 derives the asymptotic form of $\mathrm{var}\{\mathcal{U}(a)\}$, and shows that $T_{D,a,1}$ is the leading component.

LEMMA A.26. *Under the conditions of Theorem 4.7,*

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq \sum_{1\leq j_1,j_2,j_3,j_4\in\mathbb{J}_0} a!C_{\kappa,a}\sigma_{j_1,j_2}^a\sigma_{j_3,j_4}^a,$$

where $C_{\kappa,a} = \{(\kappa_x-1)/n_x + (\kappa_y-1)/n_y\}^a + 2(\kappa_x/n_x+\kappa_y/n_y)^a$. *In addition,* $\mathrm{var}(T_{D,a,2}) = o(1)\mathrm{var}(T_{D,a,1})$ *and* $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$. *It follows that* $\{T_{D,a,2} - \mathrm{E}(T_{D,a,2})\}/\sigma(a) \xrightarrow{P} 0$ *and* $[\tilde{\mathcal{U}}^*(a) - \mathrm{E}\{\tilde{\mathcal{U}}^*(a)\}]/\sigma(a) \xrightarrow{P} 0$.

PROOF. *See Section B.10.1 on Page 192.* □

Lemma A.26 gives that $\{T_{D,a,2} - \mathrm{E}(T_{D,a,2})\}/\sigma(a) \xrightarrow{P} 0$ and $[\tilde{\mathcal{U}}^*(a) - \mathrm{E}\{\tilde{\mathcal{U}}^*(a)\}]/\sigma(a) \xrightarrow{P} 0$. Thus by the Slutsky's theorem, to prove Theorem 4.7, it suffices to prove

$$\Big[\frac{T_{D,a_1,1}}{\sqrt{\mathrm{var}(T_{D,a_1,1})}}, \ldots, \frac{T_{D,a_m,1}}{\sqrt{\mathrm{var}(T_{D,a_m,1})}}\Big] \xrightarrow{D} \mathcal{N}(\mathbf{0}, I_m). \tag{A.29}$$

Note that $T_{D,a,1}$ is a summation over j indexes in $\mathbb{J}_0$, and by the definition of $\mathbb{J}_0$, $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2}$ for any $j_1, j_2 \in \mathbb{J}_0$. Therefore the analysis under H_0 can be similarly applied to $T_{D,a,1}$. Given Condition A.7 and Condition A.8 *(A1)*, we can obtain (A.29) similarly as in Section A.13.2. In summary, Theorem 4.7 is proved.

A.15. Proof of Proposition 4.2. In this section, we prove Proposition 4.2. Under the considered example, as $p - |\mathbb{J}_0| = o(p)$, we have $\sum_{j_1,j_2,j_3,j_4\in\mathbb{J}_0} \sigma^a_{x,j_1,j_2}\sigma^a_{x,j_3,j_4} \simeq \{p\nu^{2a} + 2\sum_{t=1}^s h_t^a(p-t)\}^2$. Then by Lemma A.26, when $n_x = n_y = n/2$,

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq (n/2)^{-a} a! (2\kappa_1^a + \kappa_2^a)\Big\{p\nu^{2a} + 2\sum_{t=1}^s h_t^a(p-t)\Big\}^2, \tag{A.30}$$

where $\kappa_1 = \kappa_x + \kappa_y$ and $\kappa_2 = \kappa_x + \kappa_y - 2$.

Recall that ρ_a is defined to be the value such that when $\rho = \rho_a$ under the alternative, $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ for given M. By (A.30), ρ_a satisfies

$$|J_D|^2 \rho_a^{2a} = M^2 (n/2)^{-a} a! (2\kappa_1^a + \kappa_2^a)\Big\{p\nu^{2a} + 2\sum_{t=1}^s h_t^a(p-t)\Big\}^2.$$

We next obtain

$$\rho_a = \frac{(a!)^{\frac{1}{2a}}\sqrt{\kappa_1}\nu}{(n/2)^{1/2}}\Big(\frac{Mp}{|J_D|}\Big)^{1/a}\Big\{2 + \Big(\frac{\kappa_2}{\kappa_1}\Big)^a\Big\}^{\frac{1}{2a}}\Big\{1 + 2\sum_{t=1}^s \Big(\frac{h_t}{\nu^2}\Big)^a\Big(1 - \frac{t}{p}\Big)\Big\}^{\frac{1}{a}}.$$

Let $\tilde{M} = Mp/|J_D|$, $\tilde{h}_t = h_t/\nu^2$, $\tilde{\nu} = \sqrt{\kappa_1}\nu$, and $\tilde{\kappa}_r = \kappa_2/\kappa_1$. It follows that

$$\rho_a = \tilde{\nu}(a!)^{\frac{1}{2a}}(n/2)^{-1/2}(\tilde{M})^{\frac{1}{a}}(2 + \tilde{\kappa}_r^a)^{\frac{1}{2a}}\Big\{1 + 2\sum_{t=1}^s \tilde{h}_t^a\Big(1 - \frac{t}{p}\Big)\Big\}^{\frac{1}{a}}.$$

Similarly to Section A.6, we study ρ_a as a function of integer a and show that if ρ_a starts to not decrease at some value, it will increase afterwards.

Specifically, we show that when $\rho_{a+1}/\rho_a \geq 1$, $\rho_{a+2}/\rho_{a+1} > 1$. Note that

$$\begin{aligned}\frac{\rho_{a+1}}{\rho_a} &= \left[\frac{(a+1)!\tilde{M}^2(2+\tilde{\kappa}_r^{a+1})\Big\{1+2\sum_{t=1}^s \tilde{h}_t^{a+1}\Big(1-\frac{t}{p}\Big)\Big\}^2}{(a!)^{1+\frac{1}{a}}\tilde{M}^{2+\frac{2}{a}}(2+\tilde{\kappa}_r^a)^{1+\frac{1}{a}}\Big\{1+2\sum_{t=1}^s \tilde{h}_t^{a}\Big(1-\frac{t}{p}\Big)\Big\}^{2(1+\frac{1}{a})}}\right]^{\frac{1}{2(a+1)}} \\ &= \{\mathbb{D}(a)\tilde{M}^{-2}\}^{\frac{1}{2a(a+1)}},\end{aligned}$$

where $\mathbb{D}(a) = \mathbb{D}_1(a) \times \mathbb{D}_2(a) \times \mathbb{D}_3(a)$ with $\mathbb{D}_1(a) = (a+1)^a/a!$, $\mathbb{D}_2(a) = (2+\tilde{\kappa}_r^{a+1})^a/(2+\tilde{\kappa}_r^a)^{a+1}$ and

$$\mathbb{D}_3(a) = \Big\{1+2\sum_{t=1}^s \tilde{h}_t^{a+1}\Big(1-\frac{t}{p}\Big)\Big\}^{2a}\Big/\Big\{1+2\sum_{t=1}^s \tilde{h}_t^{a}\Big(1-\frac{t}{p}\Big)\Big\}^{2(a+1)}.$$

It follows that $\rho_{a+1}/\rho_a > 1$ and $\rho_{a+1}/\rho_a = 1$ are equivalent to $\mathbb{D}(a) > \tilde{M}^2$ and $\mathbb{D}(a) = \tilde{M}^2$, respectively.

We next show that $\mathbb{D}(a)$ is a strictly increasing functions of a as $\mathbb{D}_1(a+1)/\mathbb{D}_1(a) > 1$, $\mathbb{D}_2(a+1)/\mathbb{D}_2(a) \geq 1$ and $\mathbb{D}_3(a+1)/\mathbb{D}_3(a) \geq 1$. Particularly,

$$\frac{\mathbb{D}_1(a+1)}{\mathbb{D}_1(a)} = \frac{(a+2)^{a+1}}{(a+1)!}\frac{a!}{(a+1)^a} = \Big(1+\frac{1}{a+1}\Big)^{a+1} > 1;$$

$$\frac{\mathbb{D}_2(a+1)}{\mathbb{D}_2(a)} = \frac{(2+\tilde{\kappa}_r^{a+2})^{a+1}}{(2+\tilde{\kappa}_r^{a+1})^{a+2}} \times \frac{(2+\tilde{\kappa}_r^{a})^{a+1}}{(2+\tilde{\kappa}_r^{a+1})^{a}} = \Big\{\frac{(2+\tilde{\kappa}_r^{a+2})(2+\tilde{\kappa}_r^{a})}{(2+\tilde{\kappa}_r^{a+1})^2}\Big\}^{a+1} \geq 1,$$

where we use $2\tilde{\kappa}_r^{a+1} \leq \tilde{\kappa}_r^{a+2} + \tilde{\kappa}_r^a$ by the inequality of arithmetic and geometric means; and

$$\frac{\mathbb{D}_3(a+1)}{\mathbb{D}_3(a)} = \left[\frac{\Big\{1+2\sum_{t=1}^s \tilde{h}_t^{a+2}(1-\frac{t}{p})\Big\}\Big\{1+2\sum_{t=1}^s \tilde{h}_t^{a}(1-\frac{t}{p})\Big\}}{\Big\{1+2\sum_{t=1}^s \tilde{h}_t^{a+1}(1-\frac{t}{p})\Big\}^2}\right]^{4(a+1)} \geq 1,$$

where we use $\sum_{t=1}^s \tilde{h}_t^{a+2}(1-t/p) + \sum_{t=1}^s \tilde{h}_t^{a}(1-t/p) \geq 2\sum_{t=1}^s \tilde{h}_t^{a+1}(1-t/p)$ by the inequality of arithmetic and geometric means and $\{\sum_{t=1}^s \tilde{h}_t^{a+2}(1-t/p)\}\{\sum_{t=1}^s \tilde{h}_t^{a}(1-t/p)\} \geq \{\sum_{t=1}^s \tilde{h}_t^{a+1}(1-t/p)\}^2$ by the Hölder's inequality. In summary, $\mathbb{D}(a+1)/\mathbb{D}(a) > 1$, and thus $\mathbb{D}(a)$ is a strictly increasing function of a.

Given the monotonicity of $\mathbb{D}(a)$, we know that if $\mathbb{D}(a) \geq \tilde{M}^2$, $\mathbb{D}(a+1) > \tilde{M}^2$; equivalently this implies that if $\rho_{a+1} \geq \rho_a$, $\rho_{a+2} > \rho_{a+1}$. Suppose a_0 is the first integer such that $\mathbb{D}(a_0) \geq \tilde{M}^2$, i.e., for any integer $1 \leq a < a_0$, $\mathbb{D}(a) < \tilde{M}^2$. By the analysis above, we know ρ_a is decreasing when $a < a_0$, and ρ_a is strictly increasing when $a > a_0$. Thus a_0 achieves the minimum of ρ_a. Since $\mathbb{D}(a)$ is strictly increasing in a, we know $a_0 < \infty$ given $\tilde{M}$, and a_0 increases as $\tilde{M}$ increases.

Moreover, as $s = o(p)$, there exists some constant C such that

$$\mathbb{D}(1) = \frac{2+\tilde{\kappa}_r^2}{(2+\tilde{\kappa}_r)^2} \times \frac{\{1+2\sum_{t=1}^s \tilde{h}_t^2(1-t/p)\}^2}{\{1+2\sum_{t=1}^s \tilde{h}_t(1-t/p)\}^4} \geq \mathbb{D}_0,$$

where

$$\mathbb{D}_0 = C \times \frac{2+\tilde{\kappa}_r^2}{(2+\tilde{\kappa}_r)^2} \times \frac{\{1+2\sum_{t=1}^s \tilde{h}_t^2\}^2}{\{1+2\sum_{t=1}^s \tilde{h}_t\}^4},$$

and we have $\mathbb{D}_0 = \Theta(1/s^2)$. Therefore, when $\mathbb{D}_0 \geq \tilde{M}^2$, i.e., $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, we know $\mathbb{D}(1) \geq \tilde{M}^2$ and the minimum of $\mathbb{D}(a)$ is achieved at $a_0 = 1$. This indicates that the minimum of ρ_a is achieved at $a_0 = 1$.

A.16. Results on the Generalized Linear Model in Section 4.3.

A.16.1. *Limiting results and power analysis.* We have shown that the U-statistics framework can be used to test means and covariance matrices. Here we give an example of generalized linear models to show that the framework can be extended to other testing problems.

Consider a response variable y and covariates $\mathbf{x} = (x_1, \cdots, x_p)^\intercal$ following a generalized linear model

$$\mathrm{E}(y|\mathbf{x}) = g^{-1}(\mathbf{x}^\intercal \boldsymbol{\beta}), \tag{A.31}$$

where g is the canonical link function and $\boldsymbol{\beta}$ is the regression coefficients of interest. We are interested in testing: $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_A : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. We define the score vector $\mathbf{S} = (S_1, \ldots, S_p)^\intercal$ for $\boldsymbol{\beta}$ in (A.31), where $S_j = (y - \mu_0)x_j$, $1 \leq j \leq p$ with $\mu_0 = g^{-1}(\mathbf{x}^\intercal \boldsymbol{\beta}_0)$. Given that $E(S_j) = 0$ under H_0, the target parameters can be considered as $\mathcal{E} = \{\mathrm{E}(S_j) : j = 1, \ldots, p\}$.

Suppose that $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$, are n i.i.d. observations. Many existing tests for generalized linear models [see, e.g., 29, 74] are based on the score vectors $\mathbf{S}_i = (S_{i,1}, \ldots, S_{i,p})^\intercal$, where $S_{i,j} = (y_i - \mu_{0,i})x_{i,j}$. Note that $\mathbf{S}_i$'s are i.i.d. copies of $\mathbf{S}$ with mean $(\mathrm{E}(S_1), \ldots, \mathrm{E}(S_p))^\intercal$ and the covariance matrix denoted by $\boldsymbol{\Sigma} = \{\sigma_{j_1,j_2} : 1 \leq j_1, j_2 \leq p\}$. Therefore, $K_j(\mathbf{x}_i, y_i) = S_{i,j} = (y_i - \mu_{0,i})x_{i,j}$ provides a simple kernel function. Following (1.1), $\mathcal{U}(a) = \sum_{j=1}^p (P_a^n)^{-1} \sum_{1 \leq i_1 \neq \ldots \neq i_a \leq n} \prod_{k=1}^a S_{i_k,j}$, which is an unbiased estimator of $\|\mathcal{E}\|_a^a = \sum_{j=1}^p \{\mathrm{E}(S_j)\}^a$ for finite integers a. Moreover, we define $\mathcal{U}(\infty) = \max_{1 \leq j \leq p} \sigma_{j,j}^{-1} (\sum_{i=1}^n S_{i,j}/n)^2$, which corresponds to the $\|\mathcal{E}\|_\infty$.

Asymptotic results of the U-statistics are stated below, where we assume the conditions similar to that of Theorem 4.1.

Condition A.9.

(1) *There exists constant* B *such that* $B^{-1} \leq \lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\max}(\mathbf{\Sigma}) \leq B$, *where* $\lambda_{\min}(\mathbf{\Sigma})$ *and* $\lambda_{\max}(\mathbf{\Sigma})$ *denote the minimum and maximum eigenvalues of the covariance matrix* $\mathbf{\Sigma}$*; and all correlations are bounded away from* -1 *and* 1, *i.e.,* $\max_{1\leq j_1\neq j_2\leq p} |\sigma_{j_1,j_2}|/(\sigma_{j_1,j_2}\sigma_{j_2,j_2})^{1/2} < 1-\eta$ *for some* $\eta > 0$.

(2) $\log p = o(1)n^{1/4}$ *and* $\max_{1\leq j\leq p} \mathrm{E}[\exp\{h(S_j - \mathrm{E}(S_j))^2\}] < \infty$, *for* $h \in [-M, M]$, *where* M *is a positive constant.*

(3) *Similarly to Condition 2.2,* $\{(S_{i,j}, i = 1\dots,n) : 1 \leq j \leq p\}$ *is* α*-mixing with* $\alpha_S(s) \leq C\delta^s$, *where* $\delta \in (0,1)$ *and* C *is some constant. In addition, for finite integer* a, $\sum_{j_1,j_2=1}^{p} \sigma_{j_1,j_2}^a = \Theta(p)$.

Theorem A.27. *Under Condition A.9 and* H_0*:* $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, *for any finite integers* $(a_1,\dots,a_m)$, *as* $n,p \to \infty$, $[\mathcal{U}(a_1)/\sigma(a_1),\dots,\mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *where* $\sigma^2(a) = \sum_{i=1}^{p}\sum_{j=1}^{p} \sigma_{i,j}^a/P_a^n$, *which is of order* $\Theta(pn^{-a})$. *Besides,* $P(n\mathcal{U}(\infty) - \tau_p \leq u) \to \exp\{-\pi^{-1/2}\exp(-u/2)\}$, $\forall u \in \mathbb{R}$, *where* $\tau_p = 2\log p - \log\log p$. *In addition, for any finite integer* a, $\{\mathcal{U}(a)/\sigma(a)\}$ *and* $\{n\mathcal{U}(\infty) - \tau_p\}$ *are asymptotically independent.*

Next we compare the power of $\mathcal{U}(a)$'s under alternatives with different sparsity levels. Similarly to the mean testing problems, we consider the alternative $\mathcal{E}_A = \{\mathrm{E}(S_j) = \rho > 0 \text{ for } j = 1,\dots,k_0; \mathrm{E}(S_j) = 0 \text{ for } j = k_0+1,\cdots,p\}$, where k_0 denotes the number of nonzero entries.

Theorem A.28. *Assume Condition A.9 and* $k_0 = o(p)$. *For any finite integers* $\{a_1,\dots,a_m\}$, *if* ρ *in* $\mathcal{E}_A$ *satisfies* $\rho = O(k_0^{-1/a_t}p^{1/(2a_t)}n^{-1/2})$ *for* $t = 1,\dots,m$, *then* $[\mathcal{U}(a_1)-\mathrm{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1),\dots,[\mathcal{U}(a_m)-\mathrm{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *as* $n,p \to \infty$. *In addition,* $\mathrm{E}[\mathcal{U}(a)] = \|\mathcal{E}_A\|_a^a = k_0\rho^a$ *and*

$$\sigma^2(a) \simeq \sum_{j_1=k_0+1}^{p}\sum_{j_2=k_0+1}^{p} a!\sigma_{j_1,j_2}^a/P_a^n,$$

which is $\Theta(a!pn^{-a})$.

Theorem A.28 shows that under the considered local alternatives, the asymptotic power of $\mathcal{U}(a)$ mainly depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. Therefore, for a given constant $M > 0$, if $\rho = \rho_a$ defined as $\rho_a = M^{1/a}k_0^{-1/a}a!^{1/(2a)}\times(\sum_{j_1=k_0+1}^{p}\sum_{j_2=k_0+1}^{p}\sigma_{j_1,j_2}^a)^{1/(2a)}n^{-1/2}$, we know that different $\mathcal{U}(a)$'s asymptotically have the same power. For illustration, we further assume that $\sigma_{j,j} =$

1 when $j \in \{k_0+1,\dots,p\}$, and $\sigma_{j_1,j_2}=0$ when $j_1 \neq j_2 \in \{k_0+1,\dots,p\}$, then

$$\rho_a \simeq (M\sqrt{p}/k_0)^{\frac{1}{a}} a!^{\frac{1}{2a}} n^{-\frac{1}{2}}. \tag{A.32}$$

Therefore, following the analysis in Section 4.1, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the smallest ρ_a value in (A.32). Since (A.32) is only different from (4.4) by a constant that does not depend on the order a, Proposition 4.1 still holds. Consider $a_0 \geq 1$ as specified in Proposition 4.1; then, similar to results in the two-sample mean testing, we know when $k_0 \geq \sqrt{Mp}$, $a_0 = 1$ and $\mathcal{U}(1)$ is "better" than $\mathcal{U}(\infty)$; when $k_0 < C_1\sqrt{p}/\log^{a_0/2} p$ for some C_1, $\mathcal{U}(\infty)$ is the "best"; and when $C_2\sqrt{p}/\log^{a_0/2} p < k_0 < \sqrt{Mp}$ for some C_2, $\mathcal{U}(a_0)$ is the "best". In addition, given the similar results obtained in Theorem A.27 and power analysis, we can also develop adaptive testing procedure similar to that in Section 2.3.

REMARK A.1. *More generally, if the generalized linear model also has covariates* $\mathbf{z}$ *that we want to adjust for, the corresponding generalized linear model becomes* $\mathrm{E}(y|\mathbf{x}) = g^{-1}(\mathbf{x}^\intercal\boldsymbol{\beta} + \mathbf{z}^\intercal\boldsymbol{\alpha})$*, where* $\boldsymbol{\alpha}$ *denote the regression coefficients for* $\mathbf{z}$*. To test* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ *v.s.* $H_A : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$*, we can replace* $\mu_{0,j}$ *by* $\hat{\mu}_{0,j} = g^{-1}(\mathbf{x}_i^\intercal\boldsymbol{\beta}_0 + \mathbf{z}_i^\intercal\hat{\boldsymbol{\alpha}})$ *where* $\hat{\boldsymbol{\alpha}}$ *is an estimator of* $\boldsymbol{\alpha}$*. For instance, when* $\mathbf{z}$ *is low dimensional, we can take* $\hat{\boldsymbol{\alpha}}$ *as the maximum likelihood estimator under* H_0*. Then similar conclusion to Theorem A.27 can be derived under certain regularity conditions. We present simulation studies on generalized linear model in Supplementary Material Section C.3.1 to illustrate the good performance of the U-statistics and we leave the details of theoretical developments with nuisance parameters for future study.*

A.16.2. *Proof of Theorems A.27 and A.28 (on Page 68).* Theorem A.27 is proved following the proof of Theorem 4.1 in Section A.9. Specifically, the arguments in Section A.9 can be applied to proving Theorem A.27 by replacing $x_{i,j}$'s with $S_{i,j}$'s, and therefore the details are skipped.

The proof of Theorem A.28 is similar to the proof of Theorem 4.5 in Section A.12. In particular, we decompose $\mathcal{U}(a) = T_{a,1} + T_{a,2}$, where we redefine

$$T_{a,1} = \sum_{j=1}^{k_0} \frac{1}{P_a^n} \sum_{1\leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} S_{i_k,j}, \quad T_{a,2} = \sum_{j=k_0+1}^{p} \frac{1}{P_a^n} \sum_{1\leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} S_{i_k,j}.$$

Note that $T_{a,2}$ is a summation over $j \in \{k_0+1,\dots,p\}$ and $\mathrm{E}(S_j) = 0$ for $j \in \{k_0+1,\dots,p\}$. Thus the conclusions similar to that in Theorem A.27

hold for $T_{a,2}$. Specifically, we have $\mathrm{var}(T_{a,2}) = \Theta\{(p-k_0)n^{-a}\}$ and

$$\Big[T_{a_1,2}/\sqrt{\mathrm{var}(T_{a_1,2})}, \dots, T_{a_m,2}/\sqrt{\mathrm{var}(T_{a_m,2})}\Big] \xrightarrow{D} \mathcal{N}(0, I_m). \tag{A.33}$$

When $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$, which will be proved later, we have $\sigma^2(a) \simeq \mathrm{var}(T_{a,2})$ and $\{T_{a,1} - \mathrm{E}(T_{a,1})\}/\sigma(a) \xrightarrow{P} 0$. By the Slutsky's theorem and (A.33), Theorem A.28 is proved.

To finish the proof of Theorem A.28, it remains to prove $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$. The analysis above gives that $\mathrm{var}(T_{a,2}) = \Theta\{(p-k_0)n^{-a}\}$. As $k_0 = o(p)$, to prove $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$, it suffices to show $\mathrm{var}(T_{a,1}) = o(pn^{-a})$. Note that $\mathrm{var}(T_{a,1}) = \mathrm{E}(T_{a,1}^2) - \{\mathrm{E}(T_{a,1})\}^2$, $\mathrm{E}(T_{a,1}) = k_0\rho^a$, and

$$\mathrm{E}(T_{a,1}^2) = \frac{1}{(P_a^n)^2} \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{1\le i_1\ne\cdots\ne i_a\le n;\\ 1\le \tilde{i}_1\ne\cdots\ne\tilde{i}_a\le n}} \mathrm{E}\Big\{\prod_{k=1}^{a}(S_{i_k,j_1}S_{\tilde{i}_k,j_2})\Big\}.$$

For $0\le b\le a$, define an event $B_{S,b} = \{\{i_1,\dots,i_a\}\cap\{\tilde{i}_1,\dots,\tilde{i}_a\}$ is of size $b\}$ and correspondingly

$$G_{S,a,2,b} = (P_a^n)^{-2} \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{1\le i_1\ne\cdots\ne i_a\le n;\\ 1\le \tilde{i}_1\ne\cdots\ne\tilde{i}_a\le n}} \mathrm{E}\Big\{\prod_{k=1}^{a}(S_{i_k,j_1}S_{\tilde{i}_k,j_2}) \times \mathbf{1}_{B_{S,b}}\Big\}.$$

Then $\mathrm{E}(T_{a,1}^2) = \sum_{b=0}^{a} G_{S,a,2,b}$. To prove $\mathrm{E}(T_{a,1}^2) - \{\mathrm{E}(T_{a,1})\}^2 = o(pn^{-a})$, we show $G_{S,a,2,0} - \{\mathrm{E}(T_{a,1})\}^2 = o(pn^{-a})$ and $\sum_{b=1}^{a} G_{S,a,2,b} = o(pn^{-a})$, respectively.

When $b=0$, $\{i_1,\dots,i_a\}\cap\{\tilde{i}_1,\dots,\tilde{i}_a\} = \emptyset$, and it follows that $G_{S,a,2,0} = (P_a^n)^{-2}k_0^2P_{2a}^n\rho^{2a}$. By $\mathrm{E}(T_{a,1}) = k_0\rho^a$ and $k_0^2\rho^{2a} = O(pn^{-a})$, we have $|G_{S,a,2,0} - \{\mathrm{E}(T_{a,1})\}^2| = o(k_0^2\rho^{2a}) = o(pn^{-a})$. When $b\ge 1$,

$$G_{S,a,2,b} = C(P_a^n)^{-2} \sum_{1\le j_1,j_2\le k_0} P_{2a-b}^n(\sigma_{j_1,j_2}+\rho^2)^b\rho^{2(a-b)}.$$

The maximum order of $G_{S,a,2,b}$ is bounded by the following two quantities:

$$\sum_{1\le j_1,j_2\le k_0} \frac{P_{2a-b}^n}{(P_a^n)^2}\sigma_{j_1,j_2}^b\rho^{2(a-b)}, \tag{A.34}$$

$$\sum_{1\le j_1,j_2\le k_0} \frac{P_{2a-b}^n}{(P_a^n)^2}\rho^{2a}. \tag{A.35}$$

For (A.34), as $b \geq 1$, by Condition A.9 (3) and Lemma B.1, (A.34) $= O\{k_0 n^{-b} \rho^{2(a-b)}\}$. As $k_0 = o(p)$ and $\rho = O(k_0^{-1/a} p^{1/(2a)} n^{-1/2})$, we know (A.34) $= o(pn^{-a})$. For (A.35), when $b \geq 1$, (A.35) $= O(k_0^2 n^{-b} \rho^{2a}) = o(k_0^2 \rho^{2a}) = o(pn^{-a})$. In summary, we have $|\mathrm{var}(T_{a,1})| \leq |\{\mathrm{E}(T_{a,1})\}^2 - G_{S,a,2,0}| + \sum_{b=1}^{a} |G_{S,a,2,b}| = o(pn^{-a})$. Therefore, Theorem A.28 is proved.

APPENDIX B: ASSISTED LEMMAS

In the following Sections B.1–B.10, we provide the proofs of all the assisted lemmas used in Section A. The proofs of Remark 2.4 and Corollary 4.1 are provided in Sections B.11 and B.12, respectively. To facilitate the presentation of the proofs, we first introduce some notation and then provide four technical Lemmas B.1–B.4.

Notation. We define some notation to simplify the representation of summations in the following proofs. For $a < n$, $\mathcal{P}(n,a)$ denotes the collection of a-tuples $\mathbf{i} = (i_1, \ldots, i_a)$ satisfying $1 \leq i_1 \neq \ldots \neq i_a \leq n$. Given $\mathbf{i} \in \mathcal{P}(n,a)$, we define $\{\mathbf{i}\}$ as the corresponding set containing the elements of $\mathbf{i}$ without order, that is, $\{\mathbf{i}\} = \{i_1, \ldots, i_a\}$. We apply usual set operations on the corresponding set of $\{\mathbf{i}\}$. For example, $|\{\mathbf{i}\}|$ denotes the size of the set $\{i_1, \ldots, i_a\}$, which is a in this case. In addition, for any two integers $a, b < n$, and two tuples $\mathbf{i} \in \mathcal{P}(n,a)$ and $\mathbf{m} \in \mathcal{P}(n,b)$, the operations $\{\mathbf{i}\} \cup \{\mathbf{m}\}$ and $\{\mathbf{i}\} \cap \{\mathbf{m}\}$ give the sets that equal to the union $\{i_1, \ldots, i_a\} \cup \{m_1, \ldots, m_b\}$ and intersection $\{i_1, \ldots, i_a\} \cap \{m_1, \ldots, m_b\}$ respectively. Moreover, we write $\{\mathbf{i}\} = \{\mathbf{m}\}$ and $\{\mathbf{i}\} \neq \{\mathbf{m}\}$ to indicate that the two sets $\{i_1, \ldots, i_a\}$ and $\{m_1, \ldots, m_b\}$ contain the same elements or not respectively.

In addition, let $\mathcal{C}(n,a)$ denote the collection of a-tuples $\mathbf{i} = (i_1, \ldots, i_a)$ satisfying $1 \leq i_1, \ldots, i_a \leq n$ without constraining the elements to be different. Similarly, we define $\{\mathbf{i}\}$ as the set containing the elements of $\mathbf{i}$ without order, and the set operations also apply similarly as above. Note that $|\{\mathbf{i}\}|$ may be smaller than a under this case.

We next list four technical lemmas which shall be used in the proofs later.

LEMMA B.1. *[30, Eq. (3.5)] Under the mixing assumption in Condition 2.2, suppose Z_1 and Z_2 are $\mathcal{Z}_1^t$-measurable and $\mathcal{Z}_{t+m}^{\infty}$-measurable random variables respectively. When $\mathrm{E}(|Z_1|^{2+\epsilon}) < \infty$ and $\mathrm{E}(|Z_2|^{2+\epsilon}) < \infty$, for some constants C and $\epsilon > 0$,*

$$|\mathrm{cov}(Z_1, Z_2)| \leq C\{\alpha(m)\}^{\frac{2}{2+\epsilon}} \{\mathrm{E}(|Z_1|^{2+\epsilon})\}^{\frac{1}{2+\epsilon}} \{\mathrm{E}(|Z_2|^{2+\epsilon})\}^{\frac{1}{2+\epsilon}}.$$

The lemma above can also be obtained from Lemma 2.4 in [49] by taking $p = q = 2 + \epsilon$.

LEMMA B.2. *[**?** , Lemma 3.4.3] When $|a_i| \leq A$ and $|b_i| \leq A$, then*

$$\left|\prod_{i=1}^{q} a_i - \prod_{i=1}^{q} b_i\right| \leq \sum_{i=1}^{q} |a_i - b_i| A^{q-1}.$$

LEMMA B.3. *[10, Eq. (24)] for two series of numbers A_j and B_j for $j = 1, \ldots, p$.*

$$\left|\max_{1\leq j\leq p} A_j^2 - \max_{1\leq j\leq p} B_j^2\right| \leq 2 \max_{1\leq j\leq p} |B_j| \max_{1\leq j\leq p} |A_j - B_j| + \max_{1\leq j\leq p} |A_j - B_j|^2.$$

LEMMA B.4. *When $u, v \geq 0$ and $0 < \vartheta \leq 1$, $(u+v)^\vartheta \leq u^\vartheta + v^\vartheta$.*

PROOF. When $u \geq 0$ and $0 < \vartheta \leq 1$, $f(u) = u^\vartheta$ is concave function with $f(0) = 0$. By the subadditivity property of concave function, we have $f(u+v) \leq f(u) + f(v)$. □

B.1. Lemmas for the proof of Theorem 2.1. In this section, we prove the lemmas for the proof of Theorem 2.1 in Section A.2. We still assume without loss of generality that $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ as in Section A.2.

B.1.1. *Proof of Lemma A.1 (on Page 39, Section A.2).* To illustrate the main idea of the proof of Lemma A.1, we first consider a setting where $x_{i,j}$'s are all independent, and under this independence case we prove Lemma A.1 in Section B.1.1. Next in Section B.1.1, we prove Lemma A.1 under the dependence case with Condition 2.2. Last in Section B.1.1, we present the proof under Condition 2.2*

Proof illustration. In this section, we present the proof of Lemma A.1 by only replacing Condition 2.2 with the assumption that $x_{i,j}$'s are independent. Recall $\tilde{\mathcal{U}}(a)$ defined in (2.5) and $\tilde{\mathcal{U}}^*(a) = \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$. Then $\mathrm{var}\{\mathcal{U}(a)\} \leq \mathrm{var}\{\tilde{\mathcal{U}}(a)\} + 2\sqrt{\mathrm{var}\{\tilde{\mathcal{U}}(a)\}\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}} + \mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$. To prove Lemma A.1, we derive $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ and show $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$.

We derive $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ first. Under H_0, $\mathrm{E}(x_{i,j_1}x_{i,j_2}) = 0$ when $j_1 \neq j_2$. It follows that $\mathrm{E}\{\tilde{\mathcal{U}}(a)\} = 0$ and $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \mathrm{E}[\{\tilde{\mathcal{U}}(a)\}^2]$, and then

$$\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \frac{1}{(P_a^n)^2} \sum_{\substack{1\leq j_1\neq j_2\leq p\\ 1\leq j_3\neq j_4\leq p}} \sum_{\mathbf{i},\, \tilde{\mathbf{i}} \in \mathcal{P}(n,a)} \mathrm{E}\Big\{\prod_{k=1}^{a} (x_{i_k,j_1}x_{i_k,j_2})(x_{\tilde{i}_k,j_3}x_{\tilde{i}_k,j_4})\Big\},$$

where following the notation defined at the beginning of Section B, $\mathbf{i}$ and $\tilde{\mathbf{i}}$ represent some tuples $\mathbf{i} = (i_1, \ldots, i_a)$ satisfying $1 \leq i_1 \neq \ldots \neq i_a \leq n$; and

$\tilde{\mathbf{i}} = (i_1, \ldots, i_a)$ satisfying $1 \le i_1 \ne \ldots \ne i_a \le n$. When the corresponding two sets $\{\mathbf{i}\} \ne \{\tilde{\mathbf{i}}\}$, for example, when index $i_1 \in \{\mathbf{i}\}$ but $i_1 \notin \{\tilde{\mathbf{i}}\}$,

$$\begin{aligned}\text{(B.1)} \qquad &\mathrm{E}\Big\{\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2})(x_{\tilde{i}_k,j_3}x_{\tilde{i}_k,j_4})\Big\}\\ &=\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}) \times \mathrm{E}(\text{all the remaining terms}) = 0.\end{aligned}$$

Therefore, (B.1) $\ne 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, i.e., $\{i_1, \ldots, i_a\} = \{\tilde{i}_1, \ldots, \tilde{i}_a\}$. In particular, when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$,

$$\mathrm{E}\Big\{\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2})(x_{\tilde{i}_k,j_3}x_{\tilde{i}_k,j_4})\Big\} = \{\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})\}^a.$$

It follows that

$$\begin{aligned}\mathrm{var}\{\tilde{\mathcal{U}}(a)\} &= \frac{a!}{(P_a^n)^2}\sum_{\mathbf{i}\in\mathcal{P}(n,a)}\sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}}\{\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})\}^a\\ &= \frac{a!}{P_a^n}\sum_{1\le j_1\ne j_2\le p;\, 1\le j_3\ne j_4\le p}\{\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})\}^a.\end{aligned}$$

When $x_{i,j}$'s are independent, as $j_1 \ne j_2$ and $j_3 \ne j_4$, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) \ne 0$ only when $\{j_1, j_2\} = \{j_3, j_4\}$, which gives $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \mathrm{E}(x_{1,j_1}^2)\times \mathrm{E}(x_{1,j_2}^2)$. Therefore, $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = 2a!(P_a^n)^{-1}\sum_{1\le j_1\ne j_2\le p}\mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)$. By Condition 2.1, we have $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \Theta(p^2n^{-a})$.

We next show $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$. As $\mathrm{E}\{\tilde{\mathcal{U}}^*(a)\} = 0$, $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = \mathrm{E}[\{\tilde{\mathcal{U}}^*(a)\}^2]$. Recall the definition of $\mathcal{U}^*(a)$, then we have

$$\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = \sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}}\sum_{1\le c_1,c_2\le a}\sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a+c_1)\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}}\frac{(-1)^{c_1+c_2}\binom{a}{c_1}\binom{a}{c_2}}{P_{a+c_1}^nP_{a+c_2}^n}Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4),$$

where we correspondingly define

$$\begin{aligned}Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}\Bigg[&\prod_{k=1}^{a-c_1}x_{i_k,j_1}x_{i_k,j_2}\prod_{k=a-c_1+1}^{a}x_{i_k,j_1}\prod_{k=a+1}^{a+c_1}x_{i_k,j_2}\\ &\times\prod_{\tilde{k}=1}^{a-c_2}x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\prod_{\tilde{k}=a-c_2+1}^{a}x_{\tilde{i}_{\tilde{k}},j_3}\prod_{\tilde{k}=a+1}^{a+c_2}x_{\tilde{i}_{\tilde{k}},j_4}\Bigg].\end{aligned}$$

To evaluate $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$, we examine the value of $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$. We first note that if $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \ne 0$, the following two claims hold:

Claim 1: $\{j_1, j_2\} = \{j_3, j_4\}$; *Claim 2:* $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c_1 = c_2$.

To prove *Claim 1*, we show that if $\{j_1, j_2\} \neq \{j_3, j_4\}$, $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$. We consider $j_1 \notin \{j_3, j_4\}$ as an example. When $j_1 \notin \{j_3, j_4\}$, as $j_1 \neq j_2$, we further know $j_1 \notin \{j_2, j_3, j_4\}$ and we can write

$$Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k, j_1}\Big) \times \mathrm{E}(\text{other terms with subscripts } j_2, j_3, j_4) = 0,$$

where we use $\mathrm{E}(\prod_{k=1}^{a} x_{i_k, j_1}) = \{\mathrm{E}(x_{1,j_1})\}^a = 0$ as $\mathrm{E}(x_{1,j_1}) = 0$. In addition, to prove *Claim 2*, we show that if $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$. If $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, similarly to (B.1), suppose an index $i \in \{\mathbf{i}\}$ but $i \notin \{\tilde{\mathbf{i}}\}$. Then we can write $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1}) \times \mathrm{E}(\text{other terms}) = 0$ or $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1} x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0$. As $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ are of sizes $a + c_1$ and $a + c_2$ respectively, $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ induces $c_1 = c_2$.

Given *Claim 1* and *Claim 2*, we write $c_1 = c_2 = c$ and decompose $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ into three disjoint subsets respectively as follows:

$$\begin{aligned}
&\{\mathbf{i}\}_{(1)} = \{i_1, \ldots, i_{a-c}\},\ \{\mathbf{i}\}_{(2)} = \{i_{a-c+1}, \ldots, i_a\},\ \{\mathbf{i}\}_{(3)} = \{i_{a+1}, \ldots, i_{a+c}\},\\
&\{\tilde{\mathbf{i}}\}_{(1)} = \{\tilde{i}_1, \ldots, \tilde{i}_{a-c}\},\ \{\tilde{\mathbf{i}}\}_{(2)} = \{\tilde{i}_{a-c+1}, \ldots, \tilde{i}_a\},\ \{\tilde{\mathbf{i}}\}_{(3)} = \{\tilde{i}_{a+1}, \ldots, \tilde{i}_{a+c}\},
\end{aligned}$$

which satisfies that $\{\mathbf{i}\} = \cup_{l=1}^{3} \{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\} = \cup_{l=1}^{3} \{\tilde{\mathbf{i}}\}_{(l)}$. We next prove the following *Claim 3*: if $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$, one of the following two cases hold:

1. $j_1 = j_3$, $j_2 = j_4$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}$;
2. $j_1 = j_4$, $j_2 = j_3$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}$.

To prove *Claim 3*, we note that *Claim 1* suggests that if $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$, either $\{j_1 = j_3, j_2 = j_4\}$ or $\{j_1 = j_4, j_2 = j_3\}$ holds. We consider $j_1 = j_3$ and $j_2 = j_4$ as an example. Suppose that there exists an index $i \in \{\mathbf{i}\}_{(2)}$. Since $x_{i,j}$'s are independent with mean 0, if $i \in \{\tilde{\mathbf{i}}\}_{(1)}$, $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1}^2 x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0$; or if $i \in \{\tilde{\mathbf{i}}\}_{(3)}$, $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1} x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0$. Symmetrically, if $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) \neq 0$, we know $\{\mathbf{i}\}_{(l)} = \{\tilde{\mathbf{i}}\}_{(l)}$ for $l = 1, 2, 3$ under this case. The similar analysis also applies to the second case in *Claim 3*. Moreover, under the two cases in *Claim 3*, we have $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \{\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\}^{a-c} \{\mathrm{E}(x_{1,j_1}^2)\}^c \{\mathrm{E}(x_{1,j_2}^2)\}^c$.

In summary,

$$\begin{aligned}\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} &= \sum_{1\le j_1\ne j_2\le p}\sum_{c=1}^{a}\sum_{\mathbf{i}\in\mathcal{P}(n,a+c)}\frac{2(a-c)!c!c!}{(P^n_{a+c})^2}\{\mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)\}^a\\ &\le C\sum_{1\le j_1\ne j_2\le p}\sum_{c=1}^{a} n^{-(a+c)}\{\mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)\}^a,\end{aligned}$$

which is of order $O(p^2n^{-(a+1)})$. Since we have obtained that $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \Theta(p^2n^{-a})$, then $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ is proved.

Proof under Condition 2.2. Section B.1.1 considers the case where $x_{i,j}$'s are independent. In this section, we further prove Lemma A.1 under Condition 2.2. We first explain the proof idea intuitively. Under Condition 2.2, $x_{i,j}$'s may be no longer independent, but the dependence between x_{i,j_1} and x_{i,j_2} degenerates exponentially with their distance $|j_1 - j_2|$. We expect that when $|j_1 - j_2|$ is large enough, x_{i,j_1} and x_{i,j_2} are "asymptotically independent". Specifically, we will introduce a threshold K_0 to be defined in (B.9) below. Then we will show that the majority of (x_{i,j_1}, x_{i,j_2}) pairs satisfy $|j_1 - j_2| > K_0$, and when $|j_1 - j_2| > K_0$, x_{i,j_1} and x_{i,j_2} are weakly dependent with similar properties to those under the independence case.

We next present the detailed proof under Condition 2.2. Under H_0, similarly to Section B.1.1, we have $\mathrm{E}\{\mathcal{U}(a)\} = 0$ and $\mathrm{var}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}^2(a)\}$. Then

$$\mathrm{E}\{\mathcal{U}^2(a)\} = \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}}\ \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)\times Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4), \tag{B.2}$$

where we define $F(c_1,c_2,a) = (-1)^{c_1+c_2}\binom{a}{c_1}\binom{a}{c_2}(P^n_{a+c_1}P^n_{a+c_2})^{-1}$, and recall

$$\begin{aligned} &Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\\ &= \mathrm{E}\Big\{\prod_{k=1}^{a-c_1} x_{i_k,j_1}x_{i_k,j_2}\prod_{k=a-c_1+1}^{a} x_{i_k,j_1}\prod_{k=a+1}^{a+c_1} x_{i_k,j_2}\\ &\qquad\times\prod_{\tilde{k}=1}^{a-c_2} x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\prod_{\tilde{k}=a-c_2+1}^{a} x_{\tilde{i}_{\tilde{k}},j_3}\prod_{\tilde{k}=a+1}^{a+c_2} x_{\tilde{i}_{\tilde{k}},j_4}\Big\}.\end{aligned} \tag{B.3}$$

Similarly to Section B.1.1, to evaluate $\mathrm{var}\{\mathcal{U}(a)\}$, we next examine the value of $Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)$ under different cases.

When $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, we show (B.3) $= 0$, that is, *Claim 2* in Section B.1.1 also holds here. To see this, we assume without loss of generality that an index $i \in \{\mathbf{i}\}$ and $i \notin \{\tilde{\mathbf{i}}\}$. Then (B.3) takes one of the two following forms:

$$\begin{aligned}
&\text{(B.3)} = \mathrm{E}(x_{i,j_1}) \times \mathrm{E}(\text{all the remaining terms}) \quad (j_1 = 1, \ldots, p),\\
&\text{(B.3)} = \mathrm{E}(x_{i,j_1} x_{i,j_2}) \times \mathrm{E}(\text{all the remaining terms}) \quad (1 \leq j_1 \neq j_2 \leq p).
\end{aligned}$$

Since $\mathrm{E}(x_{i,j_1}) = 0$ and $\mathrm{E}(x_{i,j_1} x_{i,j_2}) = 0$ under H_0, we know (B.3) $= 0$ when $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$. It follows that

$$\text{(B.4)} \qquad \sum_{\substack{1 \leq j_1 \neq j_2 \leq p;\\ 1 \leq j_3 \neq j_4 \leq p}} \sum_{\substack{0 \leq c_1, c_2 \leq a;\\ \mathbf{i} \in \mathcal{P}(n, a+c_1);\\ \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c_2)}} F(c_1, c_2, a) Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}\}} = 0,$$

where $\mathbf{1}_{\{\cdot\}}$ represents an indicator function.

When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, we know $c_1 = c_2$ and we write $c_1 = c_2 = c$. If $c = 0$,

$$Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}, c=0\}} = \{\mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_3} x_{i,j_4})\}^a.$$

Then we have

$$\begin{aligned}
\text{(B.5)} \qquad & \sum_{\substack{1 \leq j_1 \neq j_2 \leq p;\\ 1 \leq j_3 \neq j_4 \leq p}} \sum_{\substack{0 \leq c \leq a;\\ \mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)}} F(c, c, a) Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}, c=0\}}\\
&= \frac{1}{(P_a^n)^2} \sum_{\mathbf{i} \in \mathcal{P}(n,a)} a! \sum_{\substack{1 \leq j_1 \neq j_2 \leq p;\\ 1 \leq j_3 \neq j_4 \leq p}} \{\mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_3} x_{i,j_4})\}^a\\
&= a! (P_a^n)^{-1} \sum_{\substack{1 \leq j_1 \neq j_2 \leq p;\\ 1 \leq j_3 \neq j_4 \leq p}} \{\mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_3} x_{i,j_4})\}^a.
\end{aligned}$$

If $c \geq 1$, for given $\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)$, we decompose the sets $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ into three disjoint sets respectively, defined as:

$$\begin{aligned}
&\{\mathbf{i}\}_{(1)} = \{i_1, \ldots, i_{a-c}\},\ \{\mathbf{i}\}_{(2)} = \{i_{a-c+1}, \ldots, i_a\},\ \{\mathbf{i}\}_{(3)} = \{i_{a+1}, \ldots, i_{a+c}\},\\
&\{\tilde{\mathbf{i}}\}_{(1)} = \{\tilde{i}_1, \ldots, \tilde{i}_{a-c}\},\ \{\tilde{\mathbf{i}}\}_{(2)} = \{\tilde{i}_{a-c+1}, \ldots, \tilde{i}_a\},\ \{\tilde{\mathbf{i}}\}_{(3)} = \{\tilde{i}_{a+1}, \ldots, \tilde{i}_{a+c}\},
\end{aligned}$$

which satisfy that $\{\mathbf{i}\} = \cup_{l=1}^3 \{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\} = \cup_{l=1}^3 \{\tilde{\mathbf{i}}\}_{(l)}$. The definitions are similarly used in Section B.1.1. We next examine the value of (B.3) by further discussing different cases.

Case 1. We consider the cases where $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, $1 \le c \le a-1$ and $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$. Then we have $\{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. Note that here $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$ is assumed, and $\{\mathbf{i}\}_{(2)}, \{\mathbf{i}\}_{(3)}, \{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$ are all nonempty as $c \ge 1$. Similarly to *Claim 3* in Section B.1.1, we next prove that if (B.3) $\neq 0$, one of the following two cases holds:

$$
\begin{aligned}
&\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}, \{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}, j_1 = j_3, j_2 = j_4;\\
&\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}, \{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}, j_1 = j_4, j_2 = j_3.
\end{aligned} \tag{B.6}
$$

We prove (B.6) by contradiction.

If $\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(2)} \neq \emptyset$ and $\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(3)} \neq \emptyset$, it means that $\{\mathbf{i}\}_{(2)}$ intersects with both $\{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$. Suppose $i_1 \in \{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(2)}$ and $i_2 \in \{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(3)}$. It follows that

$$
(\text{B.3}) = \mathrm{E}(x_{i_1,j_1} x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_1} x_{i_2,j_4}) \times \mathrm{E}(\text{all the remaining terms}).
$$

As $j_3 \neq j_4$, $\mathrm{E}(x_{i_1,j_1} x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_1} x_{i_2,j_4}) = 0$ under H_0. Therefore (B.3) $= 0$. Similarly if $\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(2)} \neq \emptyset$ and $\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(3)} \neq \emptyset$, we know (B.3) $= 0$. The analysis shows that when (B.3) $\neq 0$, $\{\mathbf{i}\}_{(2)}$ only intersects with one of $\{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$. Symmetrically, $\{\mathbf{i}\}_{(3)}$ only intersects with another one of $\{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$. Since $|\{\mathbf{i}\}_{(2)}| = |\{\mathbf{i}\}_{(3)}| = |\{\tilde{\mathbf{i}}\}_{(2)}| = |\{\tilde{\mathbf{i}}\}_{(3)}|$, it remains to consider two cases $\{\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}\}$ or $\{\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}\}$. To obtain (B.6), we next examine the two cases respectively.

If $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}$, suppose $i_1 \in \{\mathbf{i}\}_{(2)}$ and $i_2 \in \{\mathbf{i}\}_{(3)}$. Then as $\{\mathbf{i}\}_{(2)} \cap \{\mathbf{i}\}_{(3)} = \emptyset$,

$$
(\text{B.3}) = \mathrm{E}(x_{i_1,j_1} x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_2} x_{i_2,j_4}) \times \mathrm{E}(\text{all the remaining terms}),
$$

which is nonzero only when $j_1 = j_3$ and $j_2 = j_4$. Similarly, if $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}$, (B.3) $\neq 0$ only when $j_1 = j_4$ and $j_2 = j_3$. In summary, if (B.3) $\neq 0$, (B.6) is obtained, and

$$
\begin{aligned}
&Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}, \{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)}, 1\le c\le a-1\}}\\
={}& Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{\{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)}, 1\le c\le a-1\}}\\
&\times \left(\mathbf{1}_{\left\{ \substack{\{\mathbf{i}\}_{(2)}=\{\tilde{\mathbf{i}}\}_{(2)},\, j_1=j_3,\\ \{\mathbf{i}\}_{(3)}=\{\tilde{\mathbf{i}}\}_{(3)},\, j_2=j_4} \right\}} + \mathbf{1}_{\left\{ \substack{\{\mathbf{i}\}_{(2)}=\{\tilde{\mathbf{i}}\}_{(3)},\, j_1=j_4,\\ \{\mathbf{i}\}_{(3)}=\{\tilde{\mathbf{i}}\}_{(2)},\, j_2=j_3} \right\}} \right).
\end{aligned}
$$

In addition, under the two cases in (B.6), we have $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \{\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^{a-c}\{\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)\}^c$. Therefore,

$$\begin{aligned}
&\sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\\ \{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)},\\ 1\le c\le a-1}\right\}} \\
&= \sum_{\substack{1\le c\le a-1;\\ \mathbf{i}\in\mathcal{P}(n,a+c);\\ 1\le j_1\ne j_2\le p}} \frac{\binom{a}{c}^2 2(a-c)!c!c!}{(P_{a+c}^n)^2}\{\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^{a-c}\{\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)\}^c \\
&= \sum_{c=1}^{a-1} O(p^2 n^{-(a+c)}),
\end{aligned} \tag{B.7}$$

where the last equation uses Condition 2.1.

Case 2. We consider the cases when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, $1 \le c \le a-1$, $\{\mathbf{i}\}_{(1)} \ne \{\tilde{\mathbf{i}}\}_{(1)}$ and $\{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)} \ne \emptyset$. Suppose that there exists an index $i_1 \in \{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)}$. Since $\{\mathbf{i}\}_{(1)} \ne \{\tilde{\mathbf{i}}\}_{(1)}$ and $|\{\mathbf{i}\}_{(1)}| = |\{\tilde{\mathbf{i}}\}_{(1)}|$, there exists another index $i_2 \in \{\mathbf{i}\}_{(1)}$ and $i_2 \notin \{\tilde{\mathbf{i}}\}_{(1)}$. As $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, we know $i_2 \in \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. Without loss of generality, we assume $i_2 \in \{\tilde{\mathbf{i}}\}_{(2)}$, then

$$(\text{B.3}) = \mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4})\mathrm{E}(x_{i_2,j_1}x_{i_2,j_2}x_{i_2,j_3})\mathrm{E}(\text{other terms}). \tag{B.8}$$

As $j_1 \ne j_2$ and $j_3 \ne j_4$ in summation, it suffices to discuss four sub-cases $\{j_1 = j_3 \text{ and } j_2 = j_4\}$, $\{j_1 = j_4 \text{ and } j_2 = j_3\}$, $\{j_1 \ne j_3 \text{ and } j_1 \ne j_4\}$ and $\{j_2 \ne j_3 \text{ and } j_2 \ne j_4\}$ under Case 2.

Case 2.1 If $j_1 = j_3$ and $j_2 = j_4$, (B.8) gives

$$(\text{B.3}) = \mathrm{E}(x_{i_1,j_1}^2 x_{i_1,j_2}^2) \times \mathrm{E}(x_{i_2,j_1}^2 x_{i_2,j_2}) \times \mathrm{E}(\text{all the remaining terms}).$$

When $x_{i,j}$'s are independent as in Section B.1.1, we know $\mathrm{E}(x_{i_2,j_1}^2 x_{i_2,j_2}) = \mathrm{E}(x_{i_2,j_1}^2)\mathrm{E}(x_{i_2,j_2}) = 0$ and thus (B.3) $= 0$. Alternatively, under Condition 2.2, (B.3) may no longer be 0 due to the dependence of $x_{i,j}$'s. But as discussed at the beginning of Section B.1.1, we expect that x_{i,j_1} and x_{i,j_2} are "asymptotically independent" as $|j_1 - j_2|$ increases, and thus we expect that (B.3) is close to 0 when $|j_1 - j_2|$ is large. To quantitatively evaluate (B.3) based on $|j_1 - j_2|$, we introduce a threshold K_0 below, and discuss the value of (B.3) when $|j_1 - j_2| > K_0$ and $|j_1 - j_2| \le K_0$, respectively.

Specifically, given δ in Condition 2.2 and positive constants μ and ϵ, we define

$$K_0 = -(2+\epsilon)(4+\mu)(\log p)/(\epsilon \log \delta). \tag{B.9}$$

When $|j_1 - j_2| > K_0$, by Conditions 2.1 and 2.2, we have

$$\begin{aligned}|(\text{B.3})| \le C \times |\mathrm{E}(x_{i_2,j_1}^2 x_{i_2,j_2})| &= C \times |\mathrm{cov}(x_{i_2,j_1}^2, x_{i_2,j_2})| \\ &\le C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(1)p^{-(4+\mu)},\end{aligned}$$

where $|\mathrm{cov}(x_{i_2,j_1}^2, x_{i_2,j_2})| \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}}$ holds by the α-mixing inequality in Lemma B.1. When $|j_1 - j_2| \le K_0$, by the uniform boundedness of moments from Condition 2.1, we have (B.3) $= O(1)$. To summarize, we define an event $S_{nem} = \{\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}, 1 \le c \le a-1, \{\mathbf{i}\}_{(1)} \ne \{\tilde{\mathbf{i}}\}_{(1)}, \{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)} \ne \emptyset\}$. Then

$$\begin{aligned}&Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4\}} \\ &= Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \left(\mathbf{1}_{\left\{ \substack{S_{nem}, j_1=j_3, j_2=j_4, \\ |j_1-j_2|>K_0} \right\}} + \mathbf{1}_{\left\{ \substack{S_{nem}, j_1=j_3, j_2=j_4, \\ |j_1-j_2|\le K_0} \right\}} \right).\end{aligned}$$

The analysis above gives $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4, |j_1-j_2|>K_0\}} = O(1)p^{-(4+\mu)}$ and $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4, |j_1-j_2|\le K_0\}} = O(1)$, respectively. Moreover, the total number of (j_1, j_2) pairs satisfying $|j_1 - j_2| \le K_0$ and $|j_1 - j_2| > K_0$ are $O(p^2)$ and $O(pK_0)$, respectively. Therefore,

$$\begin{aligned}&\left| \sum_{\substack{1\le j_1 \ne j_2 \le p; \\ 1 \le j_3 \ne j_4 \le p}} \sum_{\substack{0 \le c \le a; \\ \mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)}} F(c,c,a) Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4\}} \right| \\ &\le \sum_{\substack{1\le j_1 \ne j_2 \le p; \\ 1 \le j_3 \ne j_4 \le p}} \sum_{\substack{0 \le c \le a; \\ \mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)}} \Big| F(c,c,a) \Big| \times \mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4\}} \\ &\qquad \times \Big\{ O(p^{-(4+\mu)}) \mathbf{1}_{\{|j_1-j_2|>K_0\}} + C \times \mathbf{1}_{\{|j_1-j_2|\le K_0\}} \Big\} \\ &= \sum_{c=1}^{a-1} n^{-(a+c)} \Big\{ O(1)p^2 p^{-(4+\mu)} + O(1)pK_0 \Big\} = o(p^2 n^{-a}).\end{aligned} \tag{B.10}$$

Case 2.2 If $j_1 = j_4$ and $j_2 = j_3$, similarly to *Case 2.1*, we have

$$\begin{aligned}&\left| \sum_{\substack{1\le j_1 \ne j_2 \le p; \\ 1 \le j_3 \ne j_4 \le p}} \sum_{\substack{0 \le c \le a; \\ \mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)}} F(c,c,a) Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{S_{nem}, j_1=j_4, j_2=j_3\}} \right| \\ &= o(p^2 n^{-a}).\end{aligned} \tag{B.11}$$

Case 2.3 We discuss the cases where $j_1 \ne j_3$ and $j_1 \ne j_4$. If $x_{i,j}$'s are independent as in Section B.1.1, we know $\mathrm{E}(x_{i_1,j_1} x_{i_1,j_2} x_{i_1,j_3} x_{i_1,j_4}) =$

$\mathrm{E}(x_{i_1,j_1})\mathrm{E}(\text{other terms}) = 0$; thus by (B.8), (B.3) $= 0$ under this setting. Similarly to *Case 2.1*, under Condition 2.2, (B.3) may be no longer 0, and we will discuss the value of (B.3) using the threshold K_0 in (B.9).

To evaluate (B.3), by (B.8), we examine $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4})$. Let $(\tilde{j}_1,\tilde{j}_2,\tilde{j}_3,\tilde{j}_4)$ be the ordered version of (j_1,j_2,j_3,j_4) satisfying $\tilde{j}_1 \le \tilde{j}_2 \le \tilde{j}_3 \le \tilde{j}_4$, then $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4}) = \mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})$. Under the considered cases where $j_1 \neq j_3$ and $j_1 \neq j_4$, at least one of the two equations, $\mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}) = 0$ and $\mathrm{E}(x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4}) = 0$, holds. It follows that $\mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4}) = \mathrm{cov}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}\ ,\ x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})$. We thus can write

$$
\begin{aligned}
|\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4})| &= |\mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})| \\
&= |\mathrm{cov}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}\ ,\ x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})| \\
&= |\mathrm{cov}(x_{i_1,\tilde{j}_1}\ ,\ x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})| \\
&= |\mathrm{cov}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}\ ,\ x_{i_1,\tilde{j}_4})|.
\end{aligned} \tag{B.12}
$$

We next discuss the value of (B.12) based on the the maximum distance between the indexes in $(\tilde{j}_1,\tilde{j}_2,\tilde{j}_3,\tilde{j}_4)$, which is defined as

$$
\kappa_m = \max\{|\tilde{j}_2 - \tilde{j}_1|, |\tilde{j}_3 - \tilde{j}_2|, |\tilde{j}_4 - \tilde{j}_3|\}. \tag{B.13}
$$

We evaluate (B.12) when $\kappa_m > K_0$ and $\kappa_m \le K_0$, respectively. First, if $\kappa_m > K_0$, by $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, Conditions 2.1, 2.2, and Lemma B.1, we have (B.12) $\le C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(p^{-(4+\mu)})$. If $\kappa_m \le K_0$, by Condition 2.1, (B.12) $= O(1)$. It follows that $Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{S_{nem},j_1\neq j_3,j_1\neq j_4,\kappa_m>K_0\}} = O(p^{-(4+\mu)})$, and $Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{S_{nem},j_1\neq j_3,j_1\neq j_4,\kappa_m\le K_0\}} = O(1)$, where the event S_{nem} is defined in *Case 2.1*. Note that the total number of (j_1,j_2,j_3,j_4) tuples satisfying $\kappa_m > K_0$ and $\kappa_m \le K_0$ are $O(p^4)$ and $O(pK_0^3)$, respectively. Thus

$$
\begin{aligned}
&\Bigg| \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{S_{nem},j_1\neq j_3,j_1\neq j_4\}} \Bigg| \\
\le\ & \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} |F(c,c,a)| \times \mathbf{1}_{\{S_{nem},j_1\neq j_3,j_1\neq j_4\}} \\
&\qquad \times \Big[O(p^{-(4+\mu)})\mathbf{1}_{\{\kappa_m>K_0\}} + C \times \mathbf{1}_{\{\kappa_m\le K_0\}} \Big] \\
=\ & \sum_{c=1}^{a-1} n^{-(a+c)}\{p^2 O(p^{-(4+\mu)}) + O(1)pK_0^3\} = o(p^2n^{-a}).
\end{aligned} \tag{B.14}
$$

Case 2.4 If $j_2 \neq j_3$ and $j_2 \neq j_4$, similarly to *Case 2.3*, we have

$$\left| \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{S_{nem},j_2\neq j_3,j_2\neq j_4\}}\right| = o(p^2n^{-a}). \tag{B.15}$$

By (B.10), (B.11), (B.14), (B.15), and the definition of S_{nem}, we obtain

$$\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4) \times\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},1\le c\le a-1,\{\mathbf{i}\}_{(1)}\neq\{\tilde{\mathbf{i}}\}_{(1)},\{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}\neq\emptyset\}} = o(p^2n^{-a}). \tag{B.16}$$

Case 3. We consider $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, $1 \le c \le a-1$, and $\{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)} = \emptyset$. Here $\{\mathbf{i}\}_{(1)}$ and $\{\tilde{\mathbf{i}}\}_{(1)}$ are not empty as $c \le a-1$. Suppose there exist $i_1 \in \{\mathbf{i}\}_{(1)}$ and $i_2 \in \{\tilde{\mathbf{i}}\}_{(1)}$ with $i_1 \neq i_2$. Since $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)} = \emptyset$, we know $i_1 \in \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$ and $i_2 \in \{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)}$. Without loss of generality, we assume $i_1 \in \{\tilde{\mathbf{i}}\}_{(2)}$ and $i_2 \in \{\mathbf{i}\}_{(2)}$, then

$$\text{(B.3)} = \mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1}) \times \mathrm{E}(\text{other terms}).$$

To evaluate (B.3), we examine $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3})\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1})$. As $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, we can write

$$\begin{aligned}\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}) = \mathrm{cov}(x_{i_1,j_1}\,,\, x_{i_1,j_2}x_{i_1,j_3}) &= \mathrm{cov}(x_{i_1,j_2}\,,\, x_{i_1,j_1}x_{i_1,j_3})\\ &= \mathrm{cov}(x_{i_1,j_3}\,,\, x_{i_1,j_1}x_{i_1,j_2}),\end{aligned}$$

and similarly,

$$\begin{aligned}\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1}) = \mathrm{cov}(x_{i_2,j_3}\,,\, x_{i_2,j_4}x_{i_2,j_1}) &= \mathrm{cov}(x_{i_2,j_4}\,,\, x_{i_2,j_3}x_{i_2,j_1})\\ &= \mathrm{cov}(x_{i_2,j_1}\,,\, x_{i_2,j_3}x_{i_2,j_4}).\end{aligned}$$

Recall κ_m in (B.13) and K_0 in (B.9). If $\kappa_m > K_0$, by Conditions 2.1 and 2.2, and Lemma B.1, we have

$$\left|\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3})\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1})\right| \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(1)p^{-(4+\mu)}. \tag{B.17}$$

If $\kappa_m \le K_0$, by Condition 2.1, $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3})\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1}) = O(1)$. Note that the total number of (j_1, j_2, j_3, j_4) tuples satisfying $\kappa_m > K_0$ and

$\kappa_m \le K_0$ are $O(p^4)$ and $O(pK_0^3)$, respectively. Therefore,

$$\begin{aligned}
(\text{B.18}) \quad & \Bigg| \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\};\\ 1\le c\le a-1;\\ \{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}=\emptyset}\right\}} \Bigg| \\
\le & \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} \Big|F(c,c,a)\Big| \times \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\};1\le c\le a-1;\\ \{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}=\emptyset}\right\}} \\
& \qquad \times \Big[Cp^{-(4+\mu)}\mathbf{1}_{\{\kappa_m>K_0\}} + C\mathbf{1}_{\{\kappa_m\le K_0\}}\Big] \\
= & \sum_{c=1}^{a-1} n^{-(a+c)}\{O(1)p^4p^{-(4+\mu)} + O(1)pK_0^3\} = o(p^2n^{-a}).
\end{aligned}$$

Case 4. When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c = a$, we know $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)} = \emptyset$ and $\{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. Then similarly *Case 1*, we have

$$\begin{aligned}
(\text{B.19}) \quad & \Bigg| \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},c=a\}} \Bigg| \\
= & \; o(p^2n^{-a}).
\end{aligned}$$

In summary, by (B.2), (B.4)–(B.7), (B.16), (B.18), and (B.19),

$$(\text{B.20}) \qquad \operatorname{var}\{\mathcal{U}(a)\} = \frac{a!}{P_a^n} \sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} \{\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\}^a + o(p^2n^{-a}).$$

Note that we assume $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. For the general case with $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$, by Proposition 2.1, it is equivalent to replace $x_{i,j}$ by $x_{i,j} - \mu_j$ in (B.20).

We next show that $\operatorname{var}\{\tilde{\mathcal{U}}(a)\} =$ (B.5) and $\operatorname{var}[\tilde{\mathcal{U}}^*(a)] = o(p^2n^{-a})$. First note that $\mathrm{E}\{\tilde{\mathcal{U}}(a)\} = \mathrm{E}\{\tilde{\mathcal{U}}^*(a)\} = 0$ under H_0 as $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. Then it suffices to show $\mathrm{E}\{\{\tilde{\mathcal{U}}(a)\}^2\} =$ (B.5) and $\mathrm{E}\{\{\tilde{\mathcal{U}}^*(a)\}^2\} = o(p^2n^{-a})$. By the definition of $\tilde{\mathcal{U}}(a)$ in (2.5), we know

$$\begin{aligned}
(\text{B.21}) \qquad & \mathrm{E}\{\tilde{\mathcal{U}}^2(a)\} \\
= & \sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4) \times \mathbf{1}_{\{c_1=c_2=0\}}.
\end{aligned}$$

Therefore, $\mathrm{E}\{\tilde{\mathcal{U}}^2(a)\} = $ (B.5) from previous discussion. Moreover, as $\tilde{\mathcal{U}}^*(a) = \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$, we know

$$\begin{aligned}\tilde{\mathcal{U}}^*(a) = &\sum_{c=0}^{a} \mathbf{1}_{\{c\geq 1\}} \sum_{1\leq j_1\neq j_2\leq p} (-1)^c \binom{a}{c} \frac{1}{P^n_{a+c}} \sum_{\mathbf{i}\in\mathcal{P}(n,a+c)} \\ &\times \prod_{k=1}^{a-c}(x_{i_k,j_1}x_{i_k,j_2}) \prod_{k=a-c+1}^{a} x_{i_k,j_1} \prod_{k=a+1}^{a+c} x_{i_k,j_2}.\end{aligned} \tag{B.22}$$

It follows that

$$\begin{aligned}&\mathrm{E}[\{\tilde{\mathcal{U}}^*(a)\}^2] \\ = &\sum_{\substack{1\leq j_1\neq j_2\leq p\\ 1\leq j_3\neq j_4\leq p}} \sum_{\substack{0\leq c_1,c_2\leq a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\times \mathbf{1}_{\{c_1\geq 1,c_2\geq 1\}}.\end{aligned} \tag{B.23}$$

Also by previous discussion, we know $\mathrm{E}[\{\tilde{\mathcal{U}}^*(a)\}^2] = o(p^2n^{-a})$.

To finish the proof of Lemma A.1, it remains to show $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = $ (B.5) $= \Theta(p^2n^{-a})$, and it suffices to prove

$$\sum_{\substack{1\leq j_1\neq j_2\leq p;\\ 1\leq j_3\neq j_4\leq p}} \{\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\}^a = \Theta(p^2). \tag{B.24}$$

To prove (B.24), we examine $\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})$. Similarly to *Case 2* above, as $j_1\neq j_2$ and $j_3\neq j_4$ in summation, it suffices to discuss four cases $\{j_1 = j_3 \text{ and } j_2 = j_4\}$, $\{j_1 = j_4 \text{ and } j_2 = j_3\}$, $\{j_1\neq j_3 \text{ and } j_1\neq j_4\}$, and $\{j_2\neq j_3 \text{ and } j_2\neq j_4\}$.

If $j_1 = j_3$, $j_2 = j_4$, and $|j_1 - j_2| > K_0$, then by Conditions 2.1, 2.2, and Lemma B.1, we have

$$\begin{aligned}|\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})| =&\mathrm{E}(x^2_{i,j_1}x^2_{i,j_2}) = \mathrm{cov}(x^2_{i,j_1},x^2_{i,j_2}) + \mathrm{E}(x^2_{i,j_1})\mathrm{E}(x^2_{i,j_2})\\ \geq&\Theta(1) - |\mathrm{cov}(x^2_{i,j_1},x^2_{i,j_2})| \geq \Theta(1) - C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = \Theta(1).\end{aligned}$$

If $j_1 = j_3$, $j_2 = j_4$, and $|j_1 - j_2|\leq K_0$, by Condition 2.1, $\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4}) = O(1)$. Note that (j_1,j_2) pairs satisfying $|j_1-j_2| > K_0$ and $|j_1-j_2|\leq K_0$ are

$O(p^2)$ and $O(pK_0)$, respectively. Thus,

$$\begin{aligned}
&\sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_1=j_3,j_2=j_4\}} \\
&= \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^4 x_{i,j_t}\Big)\Big]^a \mathbf{1}_{\left\{\substack{j_1=j_3,\\ j_2=j_4}\right\}} [\mathbf{1}_{\{|j_1-j_2|>K_0\}} + \mathbf{1}_{\{|j_1-j_2|\le K_0\}}] \\
&= \Theta(p^2) + O(pK_0) = \Theta(p^2).
\end{aligned} \tag{B.25}$$

If $j_1 = j_4$ and $j_2 = j_3$, similarly to (B.25), we have

$$\sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_1=j_4,j_2=j_3\}} = \Theta(p^2). \tag{B.26}$$

If $j_1 \ne j_3$ and $j_1 \ne j_4$, we know (B.12) holds. Recall K_0 in (B.9) and κ_m in (B.13). Similarly to the analysis of (B.14), we have

$$\begin{aligned}
&\sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_1\ne j_3,j_1\ne j_4\}} \\
&= \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^4 x_{i,j_t}\Big)\Big]^a \mathbf{1}_{\{j_1\ne j_3,j_1\ne j_4\}} \Big[\mathbf{1}_{\{\kappa_m>K_0\}} + \mathbf{1}_{\{\kappa_m\le K_0\}}\Big] \\
&= o(p^2).
\end{aligned} \tag{B.27}$$

If $j_2 \ne j_3$ and $j_2 \ne j_4$, similarly to (B.27), we have

$$\sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_2\ne j_3,j_2\ne j_4\}} = o(p^2). \tag{B.28}$$

In summary, combining (B.25)–(B.28), we have

$$\sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^4 x_{i,j_t}\Big)\Big]^a \simeq 2 \sum_{1\le j_1\ne j_2\le p} \{\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^a. \tag{B.29}$$

Combining (B.20), (B.21) and (B.29), Lemma A.1 is proved.

Proof under Condition 2.2.* In this section, we prove Lemma A.1 by substituting Condition 2.2 with Condition 2.2*. Following the notation in Section B.1.1, we have

$$\operatorname{var}\{\mathcal{U}(a)\} = \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)\times Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4).$$

When $\{\mathbf{i}\}\neq\{\tilde{\mathbf{i}}\}$, under H_0, we know (B.3) $=0$ and (B.4) holds similarly. As $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ are of sizes $a+c_1$ and $a+c_2$ respectively, in the following we consider $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$, which induces $c_1=c_2$ and we write $c_1=c_2=c$.

When $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$ and $c=0$, we know (B.5) also holds similarly, and $\operatorname{var}\{\tilde{\mathcal{U}}(a)\}=$ (B.5) by (B.21). By Condition 2.2*,

$$\begin{aligned}
\text{(B.30)}\qquad & \mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\\
&=\kappa_1\Big\{\mathrm{E}(x_{i,j_1}x_{i,j_2})\mathrm{E}(x_{i,j_3}x_{i,j_4})+\mathrm{E}(x_{i,j_1}x_{i,j_3})\mathrm{E}(x_{i,j_2}x_{i,j_4})\\
&\qquad+\mathrm{E}(x_{i,j_1}x_{i,j_4})\mathrm{E}(x_{i,j_2}x_{i,j_3})\Big\}.
\end{aligned}$$

Since $j_1\neq j_2$ and $j_3\neq j_4$, we know under H_0, (B.30) $\neq 0$ only when $\{j_1=j_3, j_2=j_4\}$ or $\{j_1=j_4, j_2=j_3\}$; and then (B.30) $=\kappa_1\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)$. Thus

$$\text{(B.5)} = 2a!(P_a^n)^{-1}\sum_{1\le j_1\neq j_2\le p}\{\kappa_1\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)\}^a=\Theta(p^2n^{-a}),$$

where the second equation follows from Condition 2.1.

When $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$ and $c\ge 1$, $|\{\mathbf{i}\}_{(2)}|=|\{\mathbf{i}\}_{(3)}|=|\{\tilde{\mathbf{i}}\}_{(2)}|=|\{\tilde{\mathbf{i}}\}_{(3)}|>0$. Without loss of generality, we first consider an index $i\in\{\mathbf{i}\}_{(2)}$, and discuss four cases.

Case 1.1 If $i\notin\{\tilde{\mathbf{i}}\}$, since $\mathrm{E}(\mathbf{x})=\mathbf{0}$, we know

$$\text{(B.3)} = \mathrm{E}(x_{i,j_1})\times\mathrm{E}(\text{all the remaining terms})=0.$$

Case 1.2 If $i\in\{\tilde{\mathbf{i}}\}_{(2)}$,

$$\text{(B.3)} = \mathrm{E}(x_{i,j_1}x_{i,j_3})\times\mathrm{E}(\text{all the remaining terms}),$$

which is nonzero when $j_1=j_3$.

Case 1.3 If $i\in\{\tilde{\mathbf{i}}\}_{(3)}$,

$$\text{(B.3)} = \mathrm{E}(x_{i,j_1}x_{i,j_4})\times\mathrm{E}(\text{all the remaining terms})=0,$$

which is nonzero when $j_1 = j_4$.

Case 1.4 If $i \in \{\tilde{\mathbf{i}}\}_{(1)}$, this suggests $\{\mathbf{i}\}_{(1)} \neq \emptyset$ and thus $c \leq a-1$. By Condition 2.2*,

$$\text{(B.31)} \qquad \text{(B.3)} = \mathrm{E}(x_{i,j_1}x_{i,j_3}x_{i,j_4}) \times \mathrm{E}[\text{all the remaining terms}] = 0.$$

When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c \leq a-1$, we have $\{\mathbf{i}\}_{(1)} \neq \emptyset$. We assume without loss of generality that an index $i \in \{\mathbf{i}\}_{(1)}$, and then discuss two cases.

Case 2.1 If $i \in \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$, symmetrically, (B.3) takes a form similarly to that in (B.31), which is 0 under H_0 by Condition 2.2*.

Case 2.2 If $i \notin \{\tilde{\mathbf{i}}\}$, by $j_1 \neq j_2$, we know under H_0,

$$\text{(B.3)} = \mathrm{E}(x_{i,j_1}x_{i,j_2}) \times \mathrm{E}(\text{all the remaining terms}) = 0.$$

In summary, (B.3) $\neq 0$ only when one of the following two cases holds:

1. $j_1 = j_3$, $j_2 = j_4$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}$;
2. $j_1 = j_4$, $j_2 = j_3$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}$.

Under these two cases, (B.3) $= \{\kappa_1 \mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^{a-c}\{\mathrm{E}(x_{i,j_1}^2)\}^c\{\mathrm{E}(x_{i,j_2}^2)\}^c$. It follows that when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c \geq 1$,

$$\text{(B.32)} \qquad \sum_{\substack{1\leq j_1\neq j_2\leq p;\\ 1\leq j_3\neq j_4\leq p}} \sum_{\substack{0\leq c\leq a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},c\geq 1\}}$$

$$= \sum_{\substack{1\leq c\leq a;\\ 1\leq j_1\neq j_2\leq p}} \binom{a}{c}^2 \frac{2}{P^n_{a+c}} \{\kappa_1\mathrm{E}(x_{i,j_1}^2x_{i,j_2}^2)\}^{a-c}\{\mathrm{E}(x_{i,j_1}^2)\}^c\{\mathrm{E}(x_{i,j_2}^2)\}^c$$

$$= \sum_{c=1}^{a} O(p^2 n^{-(a+c)}) = o(pn^{-a}),$$

where the last two equations use Condition 2.1. Similarly to Section B.1.1, by (B.4) and (B.23), we know $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = \text{(B.32)} = o(pn^{-a}) = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$.

Remark B.1. *κ_1 is assumed to be a constant in Condition 2.2*. But the similar arguments apply in the proof if κ_1 changes with n, p but converges to a constant.*

B.1.2. *Proof of Lemma A.2 (on Page 40, Section A.2).* Note that for two integers $a \neq b$, $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \mathrm{E}[\mathcal{U}(a)\mathcal{U}(b)/\{\sigma(a)\sigma(b)\}]$, and by Lemma A.1, $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$. Recall $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\}$ from definition. Then by Cauchy-Schwarz inequality, we have

$$\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \mathrm{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\}/\{\sigma(a)\sigma(b)\} + o(1).$$

In addition,

$$\mathrm{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\} = \sum_{\substack{1\le j_1\neq j_2\le p,\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a),\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,b)}} \mathrm{E}\Big\{\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2})\prod_{\tilde{k}=1}^{b}(x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4})\Big\}.$$

Since $a\neq b$, we know the two sets $\{i_1,\dots,i_a\}$ and $\{\tilde{i}_1,\dots,\tilde{i}_b\}$ can not be the same. Following similar analysis to that of (B.1), as $\mathrm{E}(x_{i,j_1}x_{i,j_2})=0$ under H_0, we have $\mathrm{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\}=0$, and thus $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a),\mathcal{U}(b)/\sigma(b)\}=o(1)$.

In particular, we note that given Lemma A.1, the argument does not depend on whether Condition 2.2 or 2.2* is specified.

B.1.3. *Proof of Lemma A.3 (on Page 41, Section A.2).* We first show for $1\le k_1\neq k_2\le n$, $\mathrm{E}(D_{n,k_1}D_{n,k_2})=0$. Without loss of generality, we consider $k_1<k_2$. Then $\mathrm{E}_{k_1}Z_n\in\mathcal{F}_{k_2}$, and

$$\begin{aligned}
&\mathrm{E}(D_{n,k_1}D_{n,k_2})\\
=&\mathrm{E}\left[(\mathrm{E}_{k_1}Z_n-\mathrm{E}_{k_1-1}Z_n)(\mathrm{E}_{k_2}Z_n-\mathrm{E}_{k_2-1}Z_n)\right]\\
=&\mathrm{E}[\mathrm{E}_{k_1}Z_n\times\mathrm{E}_{k_2}Z_n-\mathrm{E}_{k_1-1}Z_n\times\mathrm{E}_{k_2}Z_n-\mathrm{E}_{k_1}Z_n\times\mathrm{E}_{k_2-1}Z_n\\
&\quad+\mathrm{E}_{k_1-1}Z_n\times\mathrm{E}_{k_2-1}Z_n]\\
=&\mathrm{E}[(\mathrm{E}_{k_1}Z_n)Z_n]-\mathrm{E}[(\mathrm{E}_{k_1-1}Z_n)Z_n]-\mathrm{E}[(\mathrm{E}_{k_1}Z_n)Z_n]+\mathrm{E}[(\mathrm{E}_{k_1-1}Z_n)Z_n]\\
=&0.
\end{aligned}$$

It follows that

$$\mathrm{E}\left(\sum_{k=1}^{n}\pi_{n,k}^2\right)=\sum_{k=1}^{n}\mathrm{E}\left(D_{n,k}^2\right)=\mathrm{E}\left(\sum_{k=1}^{n}D_{n,k}\right)^2=\mathrm{var}(Z_n),$$

where the last equation uses the fact that $\mathrm{E}(D_{n,k})=0$ and $Z_n=\sum_{k=1}^{n}D_{n,k}$ from construction.

In particular, we note that the argument does not depend on whether Condition 2.2 or 2.2* is specified.

B.1.4. *Proof of Lemma A.4 (on Page 41, Section A.2).* For given finite integer a, we derive the expression of $(\mathrm{E}_k-\mathrm{E}_{k-1})[\tilde{\mathcal{U}}(a)/\sigma(a)]$. The form of A_{n,k,a_r} for a general finite integer a_r in Lemma A.4 follows similarly.

By the definition in (2.5), we know

$$(\mathrm{E}_k-\mathrm{E}_{k-1})\tilde{\mathcal{U}}(a)=(P_a^n)^{-1}\sum_{\substack{1\le j_1\neq j_2\le p;\\ \mathbf{i}\in\mathcal{P}(n,a)}}(\mathrm{E}_k-\mathrm{E}_{k-1})\Big[\prod_{t=1}^{a}x_{i_t,j_1}x_{i_t,j_2}\Big]. \tag{B.33}$$

To derive (B.33), we next examine the value of

$$(\mathrm{E}_k - \mathrm{E}_{k-1})\Big[\prod_{t=1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big]. \tag{B.34}$$

We claim (B.34) $\neq 0$ only when $k \in \{i_1, \dots, i_a\}$. If $k \notin \{i_1, \dots, i_a\}$, we assume without loss of generality that $i_1, \dots, i_m < k$ and $i_{m+1}, \dots, i_a > k$. Then

$$\begin{aligned}
&(\mathrm{E}_k - \mathrm{E}_{k-1})\left[\prod_{t=1}^{a} x_{i_t,j_1}x_{i_t,j_2}\right] \\
&= \Big(\prod_{t=1}^{m} x_{i_t,j_1}x_{i_t,j_2}\Big)\Big[\mathrm{E}_k\Big(\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) - \mathrm{E}_{k-1}\Big(\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big)\Big] \\
&= 0.
\end{aligned}$$

Thus if (B.34) $\neq 0$, we know $k \in \{i_1, \dots, i_a\}$. In addition, we next show (B.34) $\neq 0$ only when $i_1, \dots, i_a \leq k$. Suppose that if there exist some indexes in $\{i_1, \dots, i_a\}$ that are greater than k, we assume without loss of generality that $i_m = k$, $i_1, \dots, i_{m-1} < k$, and $i_{m+1}, \dots, i_a > k$. Then

$$\begin{aligned}
\mathrm{E}_k\Big(\prod_{t=1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) &= \Big(\prod_{t=1}^{m} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}_k\Big(\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) \\
&= \Big(\prod_{t=1}^{m} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}\Big(\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) = 0,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}_{k-1}\Big(\prod_{t=1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) &= \Big(\prod_{t=1}^{m-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}_{k-1}\Big(x_{k,j_1}x_{k,j_2}\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) \\
&= \Big(\prod_{t=1}^{m-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}(x_{k,j_1}x_{k,j_2})\prod_{t=m+1}^{a}\mathrm{E}(x_{i_t,j_1}x_{i_t,j_2}) = 0.
\end{aligned}$$

Therefore, we know (B.34) $\neq 0$ when $k \in \{i_1, \dots, i_a\}$ and $i_1, \dots, i_a \leq k$.

When $k < a$, there exist some indexes in $\{i_1, \dots, i_a\} > k$. Thus (B.34) $= 0$, and (B.33) $= 0$. When $k \geq a$, assume without loss of generality that $i_a = k$ and $i_1, \cdots, i_{a-1} \leq k-1$, then

$$\mathrm{E}_{k-1}\Big[\Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)x_{k,j_1}x_{k,j_2}\Big] = \Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}(x_{k,j_1}x_{k,j_2}) = 0,$$

and

$$\mathrm{E}_k\Big[\Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)x_{k,j_1}x_{k,j_2}\Big] = \Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)x_{k,j_1}x_{k,j_2}.$$

In summary, for $k \geq a$,

$$\begin{aligned}
&(\mathrm{E}_k - \mathrm{E}_{k-1})\,\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)}\\
&= \frac{1}{\sigma(a)P_a^n}\sum_{\substack{1\leq i_1\neq\cdots\neq i_{a-1}\leq k-1;\\ 1\leq j_1\neq j_2\leq p}}\binom{a}{1}\times(\mathrm{E}_k - \mathrm{E}_{k-1})\Big[\Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)x_{k,j_1}x_{k,j_2}\Big]\\
&= \frac{a}{\sigma(a)P_a^n}\sum_{1\leq i_1\neq\cdots\neq i_{a-1}\leq k-1}\ \sum_{1\leq j_1\neq j_2\leq p}(x_{k,j_1}x_{k,j_2})\times\prod_{t=1}^{a-1}(x_{i_t,j_1}x_{i_t,j_2}).
\end{aligned}$$

In particular, we note that the argument does not depend on whether Condition 2.2 or 2.2* is specified.

B.1.5. *Proof of Lemma A.5 (on Page 41, Section A.2).* By Lemma A.4, we know the explicit form of $D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$, and it follows that $\pi_{n,k}^2 = \sum_{1\leq r_1,r_2\leq m} t_{r_1}t_{r_2}\mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})$. Note that by Cauchy-Schwarz inequality, for some constant C,

$$\mathrm{var}\Big(\sum_{k=1}^n \pi_{n,k}^2\Big) \leq Cn^2 \max_{1\leq k\leq n;\,1\leq r_1,r_2\leq m}\mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}),$$

where we define $c(n,a_r) = [a_r\times\{\sigma(a_r)P_{a_r}^n\}^{-1}]^2$ and

$$\begin{aligned}
\mathbb{T}_{k,a_{r_1},a_{r_2}} &= \mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})\\
&= \sum_{\substack{\mathbf{i}\in\mathcal{P}(k-1,a_{r_1}-1),\\ \tilde{\mathbf{i}}\in\mathcal{P}(k-1,a_{r_2}-1)}}\ \sum_{\substack{1\leq j_1\neq j_2\leq p,\\ 1\leq j_3\neq j_4\leq p}}\{c(n,a_{r_1})c(n,a_{r_2})\}^{1/2}\\
&\qquad\times\mathrm{E}\Big(\prod_{t=1}^4 x_{k,j_t}\Big)\times\Big(\prod_{t=1}^{a_{r_1}-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\times\Big(\prod_{t=1}^{a_{r_2}-1} x_{\tilde{i}_t,j_3}x_{\tilde{i}_t,j_4}\Big).
\end{aligned}$$

Therefore to prove Lemma A.5, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}) = o(n^{-2})$ for every $1\leq k\leq n$ and $1\leq r_1,r_2\leq m$.

Without loss of generality, we prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$ for any fixed constants a_1 and a_2 and $1\leq k\leq n$. Similarly to Section B.1.1, for illustration, we first consider a simple setting where $x_{i,j}$'s are independent in

Section B.1.5. Next in Section B.1.5, we prove that under Condition 2.2, $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-1}\log^3 p) = o(n^{-2})$. Last in Section B.1.5, we prove that under Condition 2.2*, $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-2}+n^{-3}) = o(n^{-2})$. Then Lemma A.5 is proved.

Proof illustration. In this section, we assume $x_{i,j}$'s are independent and prove $\mathbb{T}_{k,a_1,a_2} = o(n^{-2})$.

When $x_{i,j}$'s are independent, since $j_1 \neq j_2$ and $j_3 \neq j_4$, we know that $\mathrm{E}(x_{k,j_1}x_{k,j_2}x_{k,j_3}x_{k,j_4}) \neq 0$ only when $\{j_1, j_2\} = \{j_3, j_4\}$; and it follows that $\mathrm{E}(x_{k,j_1}x_{k,j_2}x_{k,j_3}x_{k,j_4}) = \mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)$. Thus $\mathbb{T}_{k,a_1,a_2} = 2c(n,a) \times T_{k,a_1,a_2}$, where we define

$$T_{k,a_1,a_2} = \sum_{\substack{\mathbf{i}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{1\le j_1\neq j_2\le p} \prod_{t=1}^{2} \mathrm{E}(x_{1,j_t}^2)\Big(\prod_{t=1}^{a_1-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\Big(\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2}\Big).$$

We note that $c(n,a)$ is of order $\Theta(p^{-2}n^{-a})$ by Lemma A.1. To prove $\mathrm{var}(\mathbb{T}_{k,a,a}) = o(n^{-2})$, it suffices to show that $\mathrm{var}(T_{k,a_1,a_2}) = o(n^{a_1+a_2-2}p^4)$. If $a_1 = a_2 = 1$, T_{k,a_1,a_2} is not random and thus $\mathrm{var}(T_{k,a_1,a_2}) = 0$. It remains to consider $a_1 \ge 1$ or $a_2 \ge 1$ below. To examine $\mathrm{var}(T_{k,a_1,a_2})$, we will first consider $\mathrm{E}(T_{k,a_1,a_2})$ and $\mathrm{E}(T_{k,a_1,a_2}^2)$, then $\mathrm{var}(T_{k,a_1,a_2}) = \mathrm{E}(T_{k,a_1,a_2}^2) - \{\mathrm{E}(T_{k,a_1,a_2})\}^2$.

For $\mathrm{E}(T_{k,a_1,a_2})$, note that $\mathrm{E}\{(\prod_{t=1}^{a_1-1} x_{i_t,j_1}x_{i_t,j_2})(\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2})\} \neq 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ for given $\mathbf{i} \in \mathcal{P}(k-1,a_1-1)$ and $\tilde{\mathbf{i}} \in \mathcal{P}(k-1,a_2-1)$. Therefore, if $a_1 \neq a_2$, $\mathrm{E}(T_{k,a_1,a_2}) = 0$. If $a_1 = a_2 = a$ for some a, we have

$$\mathrm{E}(T_{k,a_1,a_2}) = \sum_{\substack{\mathbf{i}\in\mathcal{P}(k-1,a-1),\\ \tilde{\mathbf{i}}\in\mathcal{P}(k-1,a-1)}} \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}\}} \sum_{1\le j_1\neq j_2\le p} \{\mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)\}^a, \tag{B.35}$$

where $\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}\}}$ represents an indicator such that the two sets $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$; and we write

$$\{\mathrm{E}(T_{k,a_1,a_2})\}^2 = \sum_{\substack{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a-1),\\ \tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a-1)}} \sum_{\substack{1\le j_1\neq j_2\le p,\\ 1\le j_3\neq j_4\le p}} \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}\}} \prod_{t=1}^{4}\{\mathrm{E}(x_{1,j_t}^2)\}^a.$$

where $\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}\}}$ represents an indicator such that $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$ hold at the same time.

For $\mathrm{E}(T_{k,a_1,a_2}^2)$, we have

$$\mathrm{E}(T_{k,a_1,a_2}^2) = \sum_{\substack{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{\substack{1\le j_1\neq j_2\le p,\\ 1\le j_3\neq j_4\le p}} \tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}), \tag{B.36}$$

where for the simplicity of notation, we define

$$\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) = \mathrm{E}\Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2}x_{m_t,j_3}x_{m_t,j_4}x_{\tilde{m}_t,j_3}x_{\tilde{m}_t,j_4}\Big)\prod_{t=1}^{4}\mathrm{E}(x_{1,j_t}^2).$$

We decompose $\mathrm{E}(T_{k,a_1,a_2}^2) = \mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} + \mathrm{E}(T_{k,a_1,a_2}^2)_{(2)}$, where

$$\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} = \sum_{\substack{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}}\ \sum_{\substack{1\le j_1\neq j_2\le p,\\ 1\le j_3\neq j_4\le p}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\\ \{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}}\right\}}\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}),$$

$$\mathrm{E}(T_{k,a_1,a_2}^2)_{(2)} = \sum_{\substack{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}}\ \sum_{\substack{1\le j_1\neq j_2\le p,\\ 1\le j_3\neq j_4\le p}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}\neq\{\tilde{\mathbf{i}}\}\text{ or}\\ \{\mathbf{m}\}\neq\{\tilde{\mathbf{m}}\}}\right\}}\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}),$$

where the two indicators $\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\,\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}\}}$ and $\mathbf{1}_{\{\{\mathbf{i}\}\neq\{\tilde{\mathbf{i}}\}\text{ or }\{\mathbf{m}\}\neq\{\tilde{\mathbf{m}}\}\}}$ represent that $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}$ hold at the same time or not, respectively. To prove $\mathrm{var}(T_{k,a_1,a_2}) = o(n^{a_1+a_2-2}p^4)$, since $|\mathrm{var}(T_{k,a_1,a_2})| \le |\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| + |\mathrm{E}(T_{k,a_1,a_2})_{(2)}|$, we show $|\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| = o(n^{2(a-1)}p^4)$ and $\mathrm{E}(T_{k,a_1,a_2})_{(2)} = o(n^{a_1+a_2-2}p^4)$, respectively below.

Part I: $|\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| = o(n^{a_1+a_2-2}p^4)$. By the analysis above, $\mathrm{E}(T_{k,a_1,a_2}) = 0$ if $a_1\neq a_2$. Also we know $\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} = 0$ if $a_1\neq a_2$, since $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}$ will not happen. Thus it remains to consider $a_1=a_2=a$ for some a below. By the forms of $\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)}$ and $\{\mathrm{E}(T_{k,a_1,a_2})\}^2$, we consider $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}$. If $\{\mathbf{i}\}\cap\{\mathbf{m}\}=\emptyset$,

$$\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) = \prod_{t=1}^{4}\{\mathrm{E}(x_{1,j_t}^2)\}^a, \tag{B.37}$$

where we use the independence between $x_{i,j}$'s and $j_1\neq j_2$ and $j_3\neq j_4$. If $\{j_1,j_2\}\cap\{j_3,j_4\}=\emptyset$, (B.37) also holds similarly by the independence between $x_{i,j}$'s. In summary, when $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}$, we know that $|\mathrm{E}\{\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j})\} - \prod_{t=1}^{4}\{\mathrm{E}(x_{1,j_t}^2)\}^a| = 0$, if $\{\mathbf{i}\}\cap\{\mathbf{m}\}=\emptyset$ or $\{j_1,j_2\}\cap$

$\{j_3, j_4\} = \emptyset$. It follows that

$$
\begin{aligned}
\text{(B.38)} \qquad & |\mathrm{E}(T^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| \\
&\le \sum_{\substack{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{\substack{1\le j_1\ne j_2\le p,\\ 1\le j_3\ne j_4\le p}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\,\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\},\,\{\mathbf{i}\}\cap\{\mathbf{m}\}\ne\emptyset,\\ \{j_1,j_2\}\cap\{j_3,j_4\}\ne\emptyset}\right\}} \\
&\qquad \times \left|\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) - \prod_{t=1}^{4}\{\mathrm{E}(x^2_{1,j_t})\}^a\right| \\
&\le Cn^{a_1+a_2-3}p^{4-1} = o(n^{a_1+a_2-2}p^4),
\end{aligned}
$$

where we use the boundedness of moments in Condition 2.1 and the facts:

$$
\sum_{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a_1-1);\,\tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)} \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\,\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\},\,\{\mathbf{i}\}\cap\{\mathbf{m}\}\ne\emptyset\}} \le Cn^{a_1+a_2-3},
$$
$$
\sum_{1\le j_1\ne j_2\le p;\,1\le j_3\ne j_4\le p} \mathbf{1}_{\{\{j_1,j_2\}\cap\{j_3,j_4\}\ne\emptyset\}} \le Cp^{4-1}.
$$

Part II: $\mathrm{E}(T_{k,a_1,a_2})_{(2)} = o(n^{a_1+a_2-2}p^4)$. We claim that $\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) = 0$ when $|\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}\cup\{\mathbf{m}\}\cup\{\tilde{\mathbf{m}}\}| > a_1 + a_2 - 2$, i.e., one of the index only appears once in the four index sets. To see this, we assume, without loss of generality, $i_1 \in \{\mathbf{i}\}$ but $i_1 \notin \{\tilde{\mathbf{i}}\}\cup\{\mathbf{m}\}\cup\{\tilde{\mathbf{m}}\}$, then

$$
\text{(B.39)} \qquad \tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) = \mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}) \times \mathrm{E}(\text{the remaining terms}) = 0.
$$

Thus when $\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) \ne 0$, the union of the four sets satisfies

$$
\text{(B.40)} \qquad |\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}\cup\{\mathbf{m}\}\cup\{\tilde{\mathbf{m}}\}| \le a_1 + a_2 - 2.
$$

In addition, note that we need to consider $\{\mathbf{i}\} \ne \{\tilde{\mathbf{i}}\}$ or $\{\mathbf{m}\} \ne \{\tilde{\mathbf{m}}\}$ when analyzing $\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}$. Assume, without loss of generality, that there exists an index $i_1 \in \{\mathbf{i}\}$ but $i_1 \notin \{\tilde{\mathbf{i}}\}$. Similarly to (B.39), we have $\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) \ne 0$ only when $i_1 \in \{\mathbf{m}\}\cup\{\tilde{\mathbf{m}}\}$. If $i_1 \in \{\mathbf{m}\}$ and $i_1 \in \{\tilde{\mathbf{m}}\}$,

$$
\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) = \mathrm{E}(x_{1,j_1}x_{1,j_3}x_{1,j_4}) \times \mathrm{E}(\text{all the remaining terms}) = 0,
$$

as $j_3 \ne j_4$ and $x_{i,j}$'s are independent; if i_1 is only in one of $\{\mathbf{m}\}$ and $\{\tilde{\mathbf{m}}\}$, for example, $i_1 \in \{\mathbf{m}\}$ but $i_1 \notin \{\tilde{\mathbf{m}}\}$, then

$$
\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) = \mathrm{E}(x_{1,j_1}x_{1,j_3}) \times \mathrm{E}(\text{all the remaining terms}),
$$

which is nonzero only when $j_1 = j_3$. By analyzing the indexes in $\{\tilde{\mathbf{i}}\}$ symmetrically, we further know $\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) \neq 0$ only when $\{j_1, j_2\} = \{j_3, j_4\}$. Therefore,

$$|\{j_1, j_2, j_3, j_4\}| = 2. \tag{B.41}$$

Combining (B.40) and (B.41), and by the boundedness of moments in Condition 2.1, we have

$$|\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}| = O(n^{a_1+a_2-2}p^2). \tag{B.42}$$

In summary, combining (B.38) and (B.42), we have

$$\begin{aligned}
|\mathrm{var}(T_{k,a_1,a_2})| &= |\mathrm{E}(T^2_{k,a_1,a_2}) - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| \\
&\leq |\mathrm{E}(T^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| + |\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}| \\
&= O(n^{a_1+a_2-3}p^3) + O(n^{a_1+a_2-2}p^2).
\end{aligned}$$

which is $o(n^{a_1+a_2-2}p^4)$.

Proof under Condition 2.2.

Proof idea. Section B.1.5 assumes that $x_{i,j}$'s are independent. In this section, we further prove Lemma A.5 under Condition 2.2. Similarly to Section B.1.1, we know that under Condition 2.2, $x_{i,j}$'s may be no longer independent, but the dependence between x_{i,j_1} and x_{i,j_2} degenerates exponentially with their distance $|j_1 - j_2|$. To quantitatively examine $|j_1 - j_2|$, we will introduce a threshold of distance D_0 to be defined in (B.46) below, which is similar to K_0 in (B.9). Intuitively, when $|j_1 - j_2| > D_0$, x_{i,j_1} and x_{i,j_2} are "asymptotically independent" with similar properties to those under the independence case in Section B.1.5. The following proof will provide comprehensive discussions based on D_0.

Recall that as argued at the beginning of Section B.1.5, to prove Lemma A.5, it suffices to show $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-1}\log^3 p) = o(n^{-2})$ for any fixed integers a_1 and a_2. To facilitate the discussion, we define some notation to be used in the proof.

Notation. For given tuples $\mathbf{i}^{(l)} = (i_1, \ldots, i_{a_l-1}) \in \mathcal{P}(k-1, a_l-1)$ with $l = 1, 2$, we define $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) = (i^{(1)}_1, \ldots, i^{(1)}_{a_1-1}, i^{(2)}_1, \ldots, i^{(2)}_{a_2-1})$, and let $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)})$ be a collection of tuples $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)})$ where $\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l-1)$ for $l = 1, 2$. Moreover, we define $\mathcal{J} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p)$. Then

$$\mathbb{T}_{k,a_1,a_2} = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)});\\ (j_1,j_2),(j_3,j_4)\in\mathcal{J}}} \Big\{\prod_{l=1}^{2} c(n, a_l)\Big\}^{1/2} \times \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2),$$

where we recall that $c(n,a) = [a \times \{\sigma(a) P_a^n\}^{-1}]^2$ and we define

$$\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2) = \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big) \prod_{l=1}^{2} \prod_{t=1}^{a_l-1} (x_{i_t^{(l)}, j_{2l-1}} x_{i_t^{(l)}, j_{2l}}).$$

In addition, for easy representation, we define $a_3 = a_1$ and $a_4 = a_2$. Then for given tuples $\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1)$ with $l = 1,2,3,4$, we define the tuple

$$(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}) = (i_1^{(1)}, \dots, i_{a_1-1}^{(1)}, i_1^{(2)}, \dots, i_{a_2-1}^{(2)}, i_1^{(3)}, \dots, i_{a_1-1}^{(3)}, i_1^{(4)}, \dots, i_{a_2-1}^{(4)}),$$

and let $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$ be a collection of $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$ where $\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1)$ with $l = 1,2,3,4$. Then we can write

$$\mathbb{T}_{k,a_1,a_2}^2 = \sum_{\substack{S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}); \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in \mathcal{J}}} \prod_{l=1}^{2} c(n, a_l)\, \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4),$$

where we define

$$\begin{aligned} &\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4) \\ &= \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big) \mathrm{E}\Big(\prod_{t=5}^{8} x_{k,j_t}\Big) \prod_{l=1}^{4} \prod_{t=1}^{a_l-1} x_{i_t^{(l)}, j_{2l-1}} x_{i_t^{(l)}, j_{2l}}. \end{aligned}$$

Recall the definitions at the beginning of Section B. $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$ represents that the two tuples have the same elements without order. We next decompose $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)})$ into two parts: the collection $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 1)$ contains the tuples $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)})$ satisfying $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$, and the collection $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 2)$ contains the tuples $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)})$ satisfying $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$. Then we can write $\mathbb{T}_{k,a_1,a_2} = \sum_{v=1}^{2} \mathbb{T}_{k,a_1,a_2,v}$, where

$$\mathbb{T}_{k,a_1,a_2,v} = \sum_{\substack{S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, v); \\ (j_1,j_2),(j_3,j_4) \in \mathcal{J}}} \Big\{\prod_{l=1}^{2} c(n, a_l)\Big\}^{1/2} \times \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2).$$

In addition, for $v = 1,2$, we let the collection $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, v, v)$ contain the tuples $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$ such that $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, v)$ and $(\mathbf{i}^{(3)}, \mathbf{i}^{(4)}) \in S(\mathbf{i}^{(3)}, \mathbf{i}^{(4)}, v)$. It follows that for $v = 1,2$, we can write

$$\begin{aligned} \mathbb{T}_{k,a_1,a_2,v}^2 = &\sum_{\substack{S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, v, v); \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in \mathcal{J}}} \prod_{l=1}^{2} c(n, a_l) \\ &\times \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4). \end{aligned} \tag{B.43}$$

We next define some notation on the j indexes. Given a tuple $(j_{t_1}, j_{t_2}, j_{t_3}, j_{t_4})$, we write its corresponding ordered version as

$$(\tilde{j}_{t_1}, \tilde{j}_{t_2}, \tilde{j}_{t_3}, \tilde{j}_{t_4}) \quad \text{satisfying} \quad \tilde{j}_{t_1} \le \tilde{j}_{t_2} \le \tilde{j}_{t_3} \le \tilde{j}_{t_4}. \tag{B.44}$$

Given the ordered indexes, we define the maximum distance between indexes in the given tuple as $\mathbb{D}_M(j_{t_1}, j_{t_2}, j_{t_3}, j_{t_4}) = \max\{\tilde{j}_{t_2} - \tilde{j}_{t_1}, \tilde{j}_{t_3} - \tilde{j}_{t_2}, \tilde{j}_{t_4} - \tilde{j}_{t_3}\}$. For the simplicity of presentation later, for tuples $(j_1, j_2), (j_3, j_4), (j_5, j_6), (j_7, j_8) \in \mathcal{J}$, we further define

$$\begin{aligned} \kappa_1 &= \mathbb{D}_M(j_1, j_2, j_3, j_4), \qquad & \kappa_2 &= \mathbb{D}_M(j_5, j_6, j_7, j_8), \\ \kappa_3 &= \mathbb{D}_M(j_1, j_2, j_5, j_6) \qquad & \kappa_4 &= \mathbb{D}_M(j_1, j_2, j_7, j_8). \end{aligned} \tag{B.45}$$

In the following discussion, to quantitatively evaluate the distances in (B.45), we introduce a threshold D_0 below. In particular, given small positive constants μ and ϵ, and δ in Condition 2.2, we define

$$D_0 = \frac{-(2+\epsilon)(8+\mu)\log p}{\epsilon \log \delta}, \tag{B.46}$$

which will be used as discussed at the beginning of this section on Page 92.

Proof. We present the proof of $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-1}\log^3 p)$ based on the notation above. Note that we can write $\mathbb{T}_{k,a_1,a_2} = \sum_{v=1}^{2} \mathbb{T}_{k,a_1,a_2,v}$. By the Cauchy-Schwarz inequality, we know it suffices to show $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,v}) = O(n^{-2}p^{-1}\log^3 p)$ for $v = 1, 2$ respectively.

Step I: $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,1}) = O(n^{-2}p^{-1}\log^3 p)$. By the definition of $\mathbb{T}_{k,a_1,a_2,1}$, we have $\{\mathbf{i}^{(1)}\} \ne \{\mathbf{i}^{(2)}\}$ for $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 1)$. Suppose, without loss of generality, that index $i \in \{\mathbf{i}^{(1)}\}$ but $i \notin \{\mathbf{i}^{(2)}\}$. Then under H_0,

$$\begin{aligned} & \mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\} \\ = \quad & \mathrm{E}(x_{i,j_1} x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0. \end{aligned} \tag{B.47}$$

Therefore $\mathrm{E}(\mathbb{T}_{k,a_1,a_2,1}) = 0$ and $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,1}) = \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,1})$.

By (B.43), we have

$$\begin{aligned} \mathbb{T}^2_{k,a_1,a_2,1} = \sum_{\substack{S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 1, 1); \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in \mathcal{J}}} & \prod_{l=1}^{2} c(n, a_l) \\ & \times \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4). \end{aligned}$$

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,1}) = O(n^{-2}p^{-1}\log^3 p)$, we will next show that for given $(j_1, j_2), (j_3, j_4), (j_5, j_6), (j_7, j_8) \in \mathcal{J}$,

$$\mathrm{E}\Big\{ \sum_{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},1,1)} \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4) \Big\} = O(n^{a_1+a_2-2}); \tag{B.48}$$

and for given $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 1, 1)$,

$$\mathrm{E}\Big\{ \sum_{\substack{(j_1,j_2),(j_3,j_4),\\ (j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4) \Big\} = O(p^3 \log^3 p). \tag{B.49}$$

Given (B.48) and (B.49), since $c(n, a_l) = \Theta(p^{-2}n^{-a_l})$, we can obtain $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,1}) = O(n^{-2}p^{-1}\log^3 p)$. Thus to finish the proof, it remains to prove (B.48) and (B.49).

To prove (B.48), we claim that $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\} = 0$ when $| \cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| > a_1 + a_2 - 2$, i.e., there exists one index only appears once in the four index sets $\{\mathbf{i}^{(l)}\}$, $l = 1, \dots, 4$. Too see this, suppose an index $i \in \{\mathbf{i}^{(1)}\}$ but $i \notin \{\mathbf{i}^{(2)}\}$, $i \notin \{\mathbf{i}^{(3)}\}$ and $i \notin \{\mathbf{i}^{(4)}\}$, then (B.47) holds. Therefore, $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\} \neq 0$ only when

$$\Big| \cup_{l=1}^4 \{\mathbf{i}^{(l)}\} \Big| \leq a_1 + a_2 - 2. \tag{B.50}$$

By the boundedness of moments from Condition 2.1, we know (B.48) holds.

We next prove (B.49). For given $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 1, 1)$, we know $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\} \neq \{\mathbf{i}^{(4)}\}$. Suppose, without loss of generality, there exists an index $i \in \{\mathbf{i}^{(3)}\}$ and $i \notin \{\mathbf{i}^{(4)}\}$. If $i \notin \{\mathbf{i}^{(1)}\}$ and $i \notin \{\mathbf{i}^{(2)}\}$, similarly, (B.47) holds. Then we consider $i \in \{\mathbf{i}^{(1)}\}$ or $i \in \{\mathbf{i}^{(2)}\}$ in the following three cases.

Case 1: When $i \in \{\mathbf{i}^{(1)}\}$ and $i \notin \{\mathbf{i}^{(2)}\}$, we know

$$\begin{aligned} &\mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\Big\} \\ &= \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=5}^{8} x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=1,2,5,6} x_{i,j_t}\Big) \times \mathrm{E}(\text{other terms}). \end{aligned} \tag{B.51}$$

If $x_{i,j}$'s are independent as in Section B.1.5, we know (B.51) $\neq 0$ only when $\{j_1, j_2\} = \{j_3, j_4\} = \{j_5, j_6\} = \{j_7, j_8\}$, which induces $|\{j_1, \dots, j_8\}| = 2$ and $\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} \mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\} = O(p^2)$, i.e., (B.49) is obtained. Under Condition 2.2, $x_{i,j}$'s may be no longer independent, but as discussed at the beginning of Section B.1.5, we can still prove (B.49) similarly to the independence case. In particular, based on D_0 in (B.46), we evaluate (B.51) by discussing the following three sub-cases (a)–(c).

(a) When both (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contain only two distinct indexes within each tuple, i.e., $|\{j_1, j_2, j_3, j_4\}| = |\{j_5, j_6, j_7, j_8\}| = 2$, we consider without loss of generality that $j_1 = j_3$, $j_2 = j_4$, $j_5 = j_7$, and $j_6 = j_8$. Then

$$\text{(B.51)} = \mathrm{E}(x_{k,j_1}^2 x_{k,j_2}^2)\mathrm{E}(x_{k,j_5}^2 x_{k,j_6}^2)\mathrm{E}(x_{k,j_1}x_{k,j_2}x_{k,j_5}x_{k,j_6})\mathrm{E}(\text{other terms}).$$

(a.1) If (j_1, j_2, j_5, j_6) contains two distinct indexes, i.e., $|\{j_1, j_2, j_5, j_6\}| = 2$, we assume without loss of generality that $j_1 = j_5$ and $j_2 = j_6$. Then $|\{j_1, \ldots, j_8\}| = 2$ and in this case, the total number of distinct j indexes is $O(p^2)$.

(a.2) If (j_1, j_2, j_5, j_6) contains at least three distinct indexes, that is, $|\{j_1, j_2, j_5, j_6\}| \geq 3$, we have $|\{\tilde{j}_1, \tilde{j}_2, \tilde{j}_5, \tilde{j}_6\}| \geq 3$, where $(\tilde{j}_1, \tilde{j}_2, \tilde{j}_5, \tilde{j}_6)$ denotes the ordered version of (j_1, j_2, j_5, j_6) following the notation in (B.44). Then we have $\mathrm{E}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2})\mathrm{E}(x_{k,\tilde{j}_5}x_{k,\tilde{j}_6}) = 0$. Together with $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, we can write

$$\begin{aligned} |\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_5}x_{1,j_6})| &= |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}\ ,\ x_{k,\tilde{j}_5}x_{k,\tilde{j}_6})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1}\ ,\ x_{k,\tilde{j}_2}x_{k,\tilde{j}_5}x_{k,\tilde{j}_6})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}x_{k,\tilde{j}_5}\ ,\ x_{k,\tilde{j}_6})|. \end{aligned} \tag{B.52}$$

Recall that κ_3 in (B.45) represents the maximum distance between (j_1, j_2, j_5, j_6). If $\kappa_3 > D_0$, by Conditions 2.1 and 2.2, and the α-mixing inequality in Lemma B.1, we know

$$|\text{(B.51)}| \leq C \times \text{(B.52)} \leq C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).$$

If $\kappa_3 \leq D_0$, the total number of distinct j indexes is $O(pD_0^3)$.

(b) When both (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) have at least 3 distinct elements, i.e., $|\{j_1, j_2, j_3, j_4\}| \geq 3$ and $|\{j_5, j_6, j_7, j_8\}| \geq 3$, following the notation in (B.44), similarly to (B.52), we can write

$$\begin{aligned} |\mathrm{E}(x_{k,j_1}x_{k,j_2}x_{k,j_3}x_{k,j_4})| &= |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}\ ,\ x_{k,\tilde{j}_3}x_{k,\tilde{j}_4})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1}\ ,\ x_{k,\tilde{j}_2}x_{k,\tilde{j}_3}x_{k,\tilde{j}_4})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}x_{k,\tilde{j}_3}\ ,\ x_{k,\tilde{j}_4})|, \end{aligned} \tag{B.53}$$

and

$$|\mathrm{E}(x_{k,j_5}x_{k,j_6}x_{k,j_7}x_{k,j_8})| = |\mathrm{cov}(x_{k,\tilde{j}_5}x_{k,\tilde{j}_6}\ ,\ x_{k,\tilde{j}_7}x_{k,\tilde{j}_8})| = |\mathrm{cov}(x_{k,\tilde{j}_5}\ ,\ x_{k,\tilde{j}_6}x_{k,\tilde{j}_7}x_{k,\tilde{j}_8})| = |\mathrm{cov}(x_{k,\tilde{j}_5}x_{k,\tilde{j}_6}x_{k,\tilde{j}_7}\ ,\ x_{k,\tilde{j}_8})|. \tag{B.54}$$

When $\max\{\kappa_1,\kappa_2\} > D_0$ in this case, by Conditions 2.1 and 2.2, and the α-mixing inequality,

$$|(\text{B.51})| \leq C \times (\text{B.53}) \times (\text{B.54}) \leq C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}). \tag{B.55}$$

When $\max\{\kappa_1,\kappa_2\} \leq D_0$, by the definitions in (B.45), we know under this case, the indexes in (j_1,j_2,j_3,j_4) are close to each other within the distance D_0, and the indexes in (j_5,j_6,j_7,j_8) are also close to each other within the distance D_0. Then the total number of distinct indexes is $O(pD_0^3 \times pD_0^3) = O(p^2D_0^6)$.

(c) If only one of (j_1,j_2,j_3,j_4) and (j_5,j_6,j_7,j_8) contains at least 3 distinct indexes, without loss of generality, we assume $|\{j_1,j_2,j_3,j_4\}| \geq 3$ and $|\{j_5,j_6,j_7,j_8\}| = 2$. When $\kappa_1 \leq D_0$, the indexes in (j_1,j_2,j_3,j_4) are close within distance D_0. As (j_5,j_6,j_7,j_8) only contains 2 distinct indexes, the total number of distinct j indexes is $O(p^3D_0^3)$. When $\kappa_1 > D_0$, by Conditions 2.1 and 2.2, and the α-mixing inequality, we know

$$|(\text{B.51})| \leq C \times (\text{B.53}) \leq C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}). \tag{B.56}$$

Case 2: When $i \notin \{\mathbf{i}^{(1)}\}$ and $i \in \{\mathbf{i}^{(2)}\}$, we know similar conclusion holds by symmetricity.

Case 3: When $i \in \{\mathbf{i}^{(1)}\}$ and $i \in \{\mathbf{i}^{(2)}\}$, we have

$$\begin{aligned} &\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\} \\ = \ &\mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=5}^{8} x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=1}^{6} x_{k,j_t}\Big) \times \mathrm{E}(\text{other terms}) \end{aligned} \tag{B.57}$$

Similarly to **Case 1** above, to evaluate (B.57), we next discuss two sub-cases with D_0 in (B.46).

(a) When both (j_1,j_2,j_3,j_4) and (j_5,j_6,j_7,j_8) only contain 2 distinct indexes within each tuple, i.e., $|\{j_1,j_2,j_3,j_4\}| = |\{j_5,j_6,j_7,j_8\}| = 2$, we assume $j_1 = j_3$, $j_2 = j_4$, $j_5 = j_7$ and $j_6 = j_8$ without loss of generality. Then

$$(\text{B.57}) = \mathrm{E}(x_{k,j_1}^2 x_{k,j_2}^2)\mathrm{E}(x_{k,j_5}^2 x_{k,j_6}^2)\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2 x_{i,j_5} x_{i,j_6})\mathrm{E}(\text{other terms}).$$

Following the notation in (B.44), when $\tilde{k}_3^* := \min\{\tilde{j}_2 - \tilde{j}_1, \tilde{j}_5 - \tilde{j}_2, \tilde{j}_6 - \tilde{j}_5\} < D_0$, the total number of distinct j indexes is $O(p^3 D_0)$. When $\tilde{k}_3^* > D_0$, by Conditions 2.1, 2.2, and the α-mixing inequality,

$$\begin{aligned}
&|\mathrm{E}(x_{1,\tilde{j}_1}^2 x_{1,\tilde{j}_2}^2 x_{1,\tilde{j}_5} x_{1,\tilde{j}_6})| \\
&= |\mathrm{cov}(x_{1,\tilde{j}_1}^2 \,,\, x_{1,\tilde{j}_2}^2 x_{1,\tilde{j}_5} x_{1,\tilde{j}_6}) + \mathrm{E}(x_{1,\tilde{j}_1}^2)\mathrm{cov}(x_{1,\tilde{j}_2}^2 \,,\, x_{1,\tilde{j}_5} x_{1,\tilde{j}_6}) \\
&\quad + [\mathrm{E}(x_{1,\tilde{j}_1}^2)]^2 \mathrm{cov}(x_{1,\tilde{j}_5} \,,\, x_{1,\tilde{j}_6})| \\
&\le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).
\end{aligned}$$

(b) If at least one of (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) has at least 3 distinct indexes within the tuple, it means that $|\{j_1, j_2, j_3, j_4\}| \ge 3$ or $|\{j_5, j_6, j_7, j_8\}| \ge 3$. Similarly to (B.55) and (B.56), we know that when $\max\{\kappa_1, \kappa_2\} > D_0$, $|(\text{B.57})| = O(p^{-(8+\mu)})$; when $\max\{\kappa_1, \kappa_2\} \le D_0$, the total number of distinct j indexes is $O(p^3 D_0^3)$.

Combining Cases 1–3 discussed above, we obtain

$$\begin{aligned}
&\mathrm{E}\Big\{ \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4) \Big\} \\
&= O(p^3 D_0^3) + \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} O(p^{-(8+\mu)}) \\
&= O(p^3 \log^3 p) + p^8 O(p^{-(8+\mu)}) = O(p^3 \log^3 p),
\end{aligned}$$

where we use $\mu > 0$ and $D_0 = O(\log p)$ by (B.46). Thus (B.49) is proved.

Step II: $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,2}) = O(n^{-2}p^{-1}\log^3 p)$. Recall that $\mathbb{T}_{k,a_1,a_2,2}$ is constructed from $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 2)$, where $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$. As $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$ happens only when $a_1 = a_2$, so it remains to consider $a_1 = a_2 = a$ for some integer a below. It follows that $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2)\} = \{\mathrm{E}(\prod_{t=1}^4 x_{1,j_t})\}^a$, then

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2}) = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},2);\\(j_1,j_2),(j_3,j_4)\in\mathcal{J}}} \Big\{ \prod_{l=1}^2 c(n, a_l) \Big\}^{1/2} \times \Big\{ \mathrm{E}\Big(\prod_{t=1}^4 x_{1,j_t} \Big) \Big\}^a,$$

and

$$\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2})\}^2 = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2);\\(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \prod_{l=1}^2 c(n, a_l) \Big\{ \mathrm{E}\Big(\prod_{t=1}^4 x_{1,j_t} \Big) \mathrm{E}\Big(\prod_{t=5}^8 x_{1,j_t} \Big) \Big\}^a.$$

Moreover, by (B.43), we know $\mathbb{T}^2_{k,a_1,a_2,2}$ is a summation over $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{4)}) \in S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2)$, where $\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}$ by the construction. We further define $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q)$ to be the collection of tuples $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{4)})$ such that $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}|=q$, where $0\le q\le a-1$. Then we write $\mathbb{T}^2_{k,a_1,a_2,2}=\sum_{q=0}^{a-1}\mathbb{T}^2_{k,a_1,a_2,2,(q)}$, where we define

$$\mathbb{T}^2_{k,a_1,a_2,2,(q)} = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q);\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \prod_{l=1}^{2} c(n,a_l) \times \mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4).$$

In particular, when $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}|=q$,

$$\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\Big\} = \Big\{\mathrm{E}\Big(\prod_{t=1}^{4}x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8}x_{1,j_t}\Big)\Big\}^{a-q}\Big\{\prod_{t=1}^{8}x_{1,j_t}\Big\}^{q}.$$

Therefore, for $a_1=a_2=a$,

$$\begin{aligned}\mathrm{var}(\mathbb{T}_{k,a_1,a_2,2}) &= \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,2}) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2})\}^2\\ &= \sum_{q=0}^{a-1}\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,2,(q)}) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2})\}^2\\ &= \sum_{q=1}^{a-1}\sum_{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q)}\prod_{l=1}^{2}c(n,a_l)\times\mathbb{D}_{k,a,a,2,q},\end{aligned}$$

where we define

$$\mathbb{D}_{k,a,a,2,q} = \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4}x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8}x_{1,j_t}\Big)\Big\}^{a-q} \times\Bigg[\Big\{\mathrm{E}\Big(\prod_{t=1}^{8}x_{1,j_t}\Big)\Big\}^{q} - \Big\{\mathrm{E}\Big(\prod_{t=1}^{4}x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8}x_{1,j_t}\Big)\Big\}^{q}\Bigg],$$

and use $\mathbb{D}_{k,a,a,2,q}=0$ when $q=0$. By the construction, we know the total number of tuples in the collection $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q)$ is bounded by $Cn^{2(a-1)-q}$, that is, for some constant C,

$$\sum_{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q)} 1 \le Cn^{2(a-1)-q}. \tag{B.58}$$

Since $c(n,a) = \Theta(p^{-2}n^{-a})$, to prove $\mathrm{var}(\mathbb{T}^2_{k,a_1,a_2,2}) = O(n^{-2}p^{-1}\log^3 p)$, it suffices to show for given tuple $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$, $\mathbb{D}_{k,a_1,a_2,2,q} = O(p^3 \log^3 p)$ for $1 \le q \le a-1$.

By Condition 2.1 and Lemma B.2 (on Page 71), for $1 \le q \le a-1$,

$$
\begin{aligned}
|\mathbb{D}_{k,a,a,2,q}| \le C \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} & \left|\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\right| \times \left|\mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big)\right| \\
& \times \left|\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big) - \mathrm{E}\Big(\prod_{t=1}^{8} x_{1,j_t}\Big)\right|.
\end{aligned}
$$

To evaluate $\mathbb{D}_{k,a,a,2,q}$, we next discuss several cases, based on the notation $\kappa_1, \ldots, \kappa_4$ in (B.45), and D_0 in (B.46).

(a) When both tuples (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contain only two distinct indexes, i.e., $|\{j_1, j_2, j_3, j_4\}| = |\{j_5, j_6, j_7, j_8\}| = 2$, we assume without loss of generality that $j_1 = j_3$, $j_2 = j_4$, $j_5 = j_7$ and $j_6 = j_8$. Then $\mathrm{E}(\prod_{t=1}^4 x_{1,j_t}) = \mathrm{E}(x^2_{1,j_1} x^2_{1,j_2})$, $\mathrm{E}(\prod_{t=5}^8 x_{1,j_t}) = \mathrm{E}(x^2_{1,j_5} x^2_{1,j_6})$ and $\mathrm{E}(\prod_{t=1}^8 x_{1,j_t}) = \mathrm{E}(x^2_{1,j_1} x^2_{1,j_2} x^2_{1,j_5} x^2_{1,j_6})$. Following the notation in (B.44), let $(\tilde{j}_1 \le \tilde{j}_2 \le \tilde{j}_5 \le \tilde{j}_6)$ be the ordered version of (j_1, j_2, j_5, j_6). When $\min\{\tilde{j}_2 - \tilde{j}_1, \tilde{j}_5 - \tilde{j}_2, \tilde{j}_6 - \tilde{j}_5\} \le D_0$, the total number of distinct j indexes is $O(p^3 D_0)$. When $\min\{\tilde{j}_2 - \tilde{j}_1, \tilde{j}_5 - \tilde{j}_2, \tilde{j}_6 - \tilde{j}_5\} > D_0$, by Conditions 2.1 and 2.2, and the α-mixing inequality in Lemma B.1,

$$
\begin{aligned}
& \left|\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big) - \mathrm{E}\Big(\prod_{t=1}^{8} x_{1,j_t}\Big)\right| \\
= & \left|\mathrm{E}(x^2_{1,j_1} x^2_{1,j_2})\mathrm{E}(x^2_{1,j_5} x^2_{1,j_6}) - \mathrm{E}(x^2_{1,j_1} x^2_{1,j_2} x^2_{1,j_5} x^2_{1,j_6})\right| \\
\le & \, C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).
\end{aligned}
$$

(b) When both (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contain at least 3 distinct indexes, i.e., $|\{j_1, j_2, j_3, j_4\}| \ge 3$ and $|\{j_5, j_6, j_7, j_8\}| \ge 3$, we know similarly (B.53) and (B.54) hold. When $\max\{\kappa_1, \kappa_2\} > D_0$, by Conditions 2.1 and 2.2, and the α-mixing inequality in Lemma B.1, we obtain

$$
|\mathbb{D}_{k,a_1,a_2,2,q}| \le C(\text{B.53}) \times (\text{B.54}) \le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O\{p^{-(8+\mu)}\}.
$$

When $\max\{\kappa_1, \kappa_2\} \le D_0$, by the definitions in (B.45), we know under this case the indexes in (j_1, j_2, j_3, j_4) are close to each other within the distance D_0, and the indexes in (j_5, j_6, j_7, j_8) are also close to each other within the distance D_0. Then the total number of distinct j indexes is $O(pD_0^3 \times pD_0^3) = O(p^2 D_0^6)$.

(c) When only one of (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contains at least 3 distinct indexes, without loss of generality, we assume $|\{j_1, j_2, j_3, j_4\}| \geq 3$ and $|\{j_5, j_6, j_7, j_8\}| = 2$. Recall κ_1 defined in (B.45). When $\kappa_1 \leq D_0$, the indexes in (j_1, j_2, j_3, j_4) are close within distance D_0. As (j_5, j_6, j_7, j_8) only contains 2 distinct indexes, the total number of distinct j indexes is $O(p^3 D_0^3)$. When $\kappa_1 > D_0$, by Conditions 2.1 and 2.2, and the α-mixing inequality in Lemma B.1, we know similarly (B.53) holds, and

$$|\mathbb{D}_{k,a_1,a_2,2,q}| \leq C(\text{B.53}) \leq C\delta^{\frac{D_0 \epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).$$

In summary,

$$|\mathbb{D}_{k,a_1,a_2,2,q}| = p^8 \times O(p^{-(8+\mu)}) + O(p^3 D_0^3) = O(p^3 \log^3 p). \tag{B.59}$$

Thus we obtain that for given $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$, $\mathbb{D}_{k,a_1,a_2,2,q} = O(p^3 \log^3 p)$. Combined with (B.58), $\mathrm{var}(\mathbb{T}^2_{k,a_1,a_2,2}) = O(n^{-2} p^{-1} \log^3 p)$ follows.

Combining the results in *Step I* and *Step II* above, we obtain $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2} p^{-1} \log^3 p)$, and thus Lemma A.5 is proved under Condition 2.2.

*Proof under Condition 2.2**. In this section, we prove Lemma A.5 by substituting Condition 2.2 with Condition 2.2*. Note that although the independence between $x_{i,j}$'s is assumed in Section B.1.5, it is only used to specify certain joint moments of $x_{i,j}$'s. Alternatively, Condition 2.2* is assumed to obtain similar properties on the joint moments, and the proof follows similarly to that in Section B.1.5.

In particular, we will prove that $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-3} + n^{-2}p^{-2})$ for two given finite integers a_1 and a_2 below. Under H_0 and given Condition 2.2*, as $j_1 \neq j_2$ and $j_3 \neq j_4$, we have $\mathrm{E}(x_{1,j_1} x_{1,j_2} x_{1,j_3} x_{1,j_4}) \neq 0$ only when $\{j_1, j_2\} = \{j_3, j_4\}$, and then $\mathrm{E}(x_{1,j_1} x_{1,j_2} x_{1,j_3} x_{1,j_4}) = \kappa_1 \mathrm{E}(x^2_{1,j_1}) \mathrm{E}(x^2_{1,j_2})$. It follows that $\mathbb{T}_{k,a_1,a_2} = 2c(n, a) \times \tilde{T}_{k,a_1,a_2}$, where $\tilde{T}_{k,a,a} = \kappa_1 T_{k,a,a}$ with $T_{k,a,a}$ defined in Section B.1.5. To prove $\mathrm{var}(\mathbb{T}_{k,a,a}) = o(n^{-2})$, it suffices to show that $\mathrm{var}(\tilde{T}_{k,a_1,a_2}) = n^{a_1+a_2-2} p^4 O(n^{-1} + p^{-2})$ as argued in Section B.1.5.

Similarly to Section B.1.5, to show $\mathrm{var}(\tilde{T}_{k,a_1,a_2}) = n^{a_1+a_2-2} p^4 O(n^{-1} + p^{-2})$, we examine $\{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2$ and $\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})$ respectively. For $\mathrm{E}(\tilde{T}_{k,a_1,a_2})$, under Condition 2.2*, similarly to (B.35), we know $\mathrm{E}\{(\prod_{t=1}^{a_1-1} x_{i_t,j_1} x_{i_t,j_2}) \times (\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1} x_{\tilde{i}_t,j_2})\} \neq 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$. When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, we write $a_1 = a_2 = a$ for some a and then $\mathrm{E}\{(\prod_{t=1}^{a_1-1} x_{i_t,j_1} x_{i_t,j_2}) \times (\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1} x_{\tilde{i}_t,j_2})\} = \{\kappa_1 \mathrm{E}(x^2_{1,j_1}) \mathrm{E}(x^2_{1,j_2})\}^{a-1}$. We thus have $\{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2 = \{\kappa_1^a \mathrm{E}(T_{k,a_1,a_2})\}^2$

with T_{k,a_1,a_2} defined in Section B.1.5. Moreover, following (B.36) in Section B.1.5, we have

$$\mathrm{E}(\tilde{T}^2_{k,a_1,a_2}) = \sum_{\substack{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a_1-1);\\ \tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}} \tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}).$$

We further decompose $\mathrm{E}(\tilde{T}^2_{k,a_1,a_2}) = \mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(1)} + \mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(2)}$, where $\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(1)}$ and $\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(2)}$ are defined with the same forms as $\mathrm{E}(T^2_{k,a_1,a_2})_{(1)}$ and $\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}$ in Section B.1.5, respectively. To prove $\mathrm{var}(\tilde{T}_{k,a_1,a_2}) = n^{a_1+a_2-2}p^4O(n^{-1}+p^{-2})$, similarly to Section B.1.5, we derive $|\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2|$ and $\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}$ respectively.

Step I: $|\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2|$. By the forms of $\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(1)}$ and $\mathrm{E}(\tilde{T}_{k,a_1,a_2})$, we consider $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$ below. If $\{\mathbf{i}\}\cap\{\mathbf{m}\} = \emptyset$, $|\mathrm{E}\{\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j})\} - \kappa_1^{2a}\prod_{t=1}^4\{\mathrm{E}(x^2_{1,j_t})\}^a| = 0$ by Condition 2.2*; if $\{\mathbf{i}\}\cap\{\mathbf{m}\}\ne\emptyset$, $|\{\mathbf{i}\}\cup\{\mathbf{m}\}| \le a_1+a_2-2-1$, thus $|\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2| = O(n^{a_1+a_2-3}p^4)$ by Condition 2.1.

Step II: $\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}$. We note that for $j_1\ne j_2$, $\mathrm{E}(x_{1,j_1}x_{1,j_2}) = 0$; and for any additional index j_3, we have $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}) = 0$ under Condition 2.2*. Thus (B.41) and (B.42) still hold here, and we obtain $\mathrm{E}(T^2_{k,a_1,a_2})_{(2)} = O(n^{a_1+a_2-2}p^2)$.

In summary,

$$\begin{aligned}|\mathrm{var}(T_{k,a_1,a_2})| &\le |\mathrm{E}(\tilde{T}^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2| + |\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}|\\ &= n^{a_1+a_2-2}p^4O(n^{-1}+p^{-2}).\end{aligned}$$

It follows that $\mathrm{var}(\sum_{k=1}^n \pi^2_{n,k}) = O(n^{-1}+p^{-2})$ by the argument at the beginning of Section B.1.5. Therefore Lemma A.5 is proved.

B.1.6. *Proof of Lemma A.6 (on Page 41, Section A.2).* By Lemma A.4,

$$\sum_{k=1}^n \mathrm{E}(D^4_{n,k}) = \sum_{k=1}^n \sum_{1\le r_1,r_2,r_3,r_4\le m} \prod_{l=1}^4 t_{r_l} \times \mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_{r_l}}\Big). \tag{B.60}$$

To prove Lemma A.6, it suffices to show that for given $1\le k\le n$ and $1\le r_1,r_2,r_3,r_4\le m$, we have $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_{r_l}}) = O(n^{-2})$.

Similarly to Sections B.1.1 and B.1.5 above, we first illustrate the proof of Lemma A.6, when $x_{i,j}$'s are independent. Then in Section B.1.6, we prove Lemma A.6 under Condition 2.2. Last in Section B.1.6, we prove Lemma A.6 under Condition 2.2*.

Proof illustration. In this section, we assume that $x_{i,j}$'s are independent and prove $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_l}) = O(n^{-2})$ for given integers a_l, $l = 1, \dots, 4$. By Lemma A.4, when $k < a_l$, $A_{n,k,a_l} = 0$. We next focus on $\max_{1\le l\le 4} a_l \le k \le n$. By Lemma A.4, we have

$$\mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_l}\Big) = \Big\{\prod_{l=1}^4 c(n,a_l)\Big\}^{1/2} \sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,\dots,4;\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8), \tag{B.61}$$

where $\mathbf{i}^{(l)} = (i_1^{(l)}, \dots, i_{a_l-1}^{(l)})$, $l = 1, \dots, 4$ represent the tuples satisfying $1 \le i_1^{(l)} \neq \dots \neq i_{a_l-1}^{(l)} \le n$; $\mathcal{J} = \{(j_1, j_2) : 1 \le j_1 \neq j_2 \le p)$; $\mathbf{j}_8$ represents the tuple $(j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8)$; and we define

$$Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8) = \mathrm{E}\Big(\prod_{r=1}^8 x_{k,j_r}\Big)\,\mathrm{E}\Big(\prod_{t=1}^{a_l-1}\prod_{l=1}^4 x_{i_t^{(l)},j_{2l-1}} x_{i_t^{(l)},j_{2l}}\Big).$$

We claim that $\mathrm{E}(\prod_{r=1}^8 x_{k,j_r}) \neq 0$ only when

$$|\{j_t : t = 1, \dots, 8\}| \le 4. \tag{B.62}$$

If $|\{j_t : t = 1, \dots, 8\}| \ge 5$, it implies that one of the j index in $\{j_t : t = 1, \dots, 8\}$ only appears once. We assume without loss of generality that j_1 only appears once, i.e., $j_1 \notin \{j_t : t = 2, \dots, 8\}$. Since $x_{k,j}$'s are independent, $\mathrm{E}(\prod_{r=1}^8 x_{k,j_r}) = \mathrm{E}(x_{k,j_1})\mathrm{E}(\text{all the remaining terms}) = 0$. Thus (B.62) is proved. Similarly to (B.39) and (B.40), we further know $Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8) \neq 0$ only when

$$\Big|\bigcup_{l=1}^4 \{\mathbf{i}^{(l)}\}\Big| \le \sum_{l=1}^4 (a_l - 1)/2. \tag{B.63}$$

In summary, combining (B.62) and (B.63), we have

$$\mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_l}\Big) = O(p^{-4} n^{-\frac{1}{2}\sum_{l=1}^4 a_l} n^{\frac{1}{2}\sum_{l=1}^4 (a_l-1)} p^4) = O(n^{-2}).$$

Proof under Condition 2.2. Section B.1.6 proves Lemma A.6 when $x_{i,j}$'s are independent. In this section, we further prove Lemma A.6 under Condition 2.2. We first illustrate the proof idea intuitively, which is similar to Sections B.1.1 and B.1.5. Under Condition 2.2, $x_{i,j}$'s may be no longer independent, but the dependence between x_{i,j_1} and x_{i,j_2} degenerates exponentially with their distance $|j_1-j_2|$. To quantitatively examine $|j_1-j_2|$, we use the threshold of distance D_0 defined in (B.46). Intuitively, when $|j_1-j_2|>D_0$, x_{i,j_1} and x_{i,j_2} are "asymptotically independent" with similar properties to those under the independence case in Section B.1.6. The following proof will provide comprehensive discussions based on D_0.

We next present the detailed proof of Lemma A.6. Note that to prove Lemma A.6, by the analysis at the beginning of Section B.1.6, it suffices to show $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_l}) = O(n^{-2})$. Recall that we can write (B.61) and we have $\prod_{l=1}^4 c^{1/2}(n,a_l) = \Theta(p^{-4}n^{-\frac{1}{2}\sum_{l=1}^4 a_l})$. It remains to show

$$\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,\dots,4;\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8) = O(p^4 n^{\frac{1}{2}\sum_{l=1}^4(a_l-1)}). \tag{B.64}$$

To prove (B.64), we show the order of (B.64) in n and p respectively in the following two steps.

Step I: order of n. We show for any fixed $\mathbf{j}_8=(j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8)$,

$$\left|\sum_{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,\dots,4} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\right| = O(n^{\frac{1}{2}\sum_{l=1}^4(a_l-1)}). \tag{B.65}$$

We note that $Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\neq 0$ only if (B.63) holds. Too see this, suppose one index i_1 only appears once in the four sets $\{\mathbf{i}^{(1)}\}$, $\{\mathbf{i}^{(2)}\}$, $\{\mathbf{i}^{(3)}\}$, $\{\mathbf{i}^{(4)}\}$. For example $i_1\in\{\mathbf{i}^{(1)}\}$, but $i_1\notin\cup_{l=2}^4\{\mathbf{i}^{(l)}\}$. Then

$$Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8) = \mathrm{E}(x_{i_1,j_1}x_{i_1,j_2})\times\mathrm{E}(\text{the remaining terms}) = 0,$$

Therefore by (B.63) and Condition 2.1,

$$\text{(B.65)} = O(n^{\frac{1}{2}\sum_{l=1}^4(a_l-1)}). \tag{B.66}$$

Step II: order of p. To prove (B.64), it remains to show that for given $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)})$,

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8) = O(p^4). \tag{B.67}$$

Let μ be a positive constant same as in (B.46). Define an event $B_J^c = \{Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8) = O(p^{-(8+\mu)})\}$ and let B_J represent the complement set of B_J^c correspondingly. Note that

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times\mathbf{1}_{B_J^c} = O(p^8p^{-(8+\mu)}) = o(1).$$

Moreover by Condition 2.1, $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(1)$ always holds. Thus to prove (B.67), it remains to show

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} \mathbf{1}_{B_J} = O(p^4). \tag{B.68}$$

We write the ordered version of $\mathbf{j}_8 = (j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8)$ as $\tilde{\mathbf{j}}_8 = (\tilde{j}_1, \tilde{j}_2, \tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6, \tilde{j}_7, \tilde{j}_8)$, which satisfies $\tilde{j}_1 \le \tilde{j}_2 \le \tilde{j}_3 \le \tilde{j}_4 \le \tilde{j}_5 \le \tilde{j}_6 \le \tilde{j}_7 \le \tilde{j}_8$. To facilitate the proof, we first introduce three claims below, which will be proved later. In particular, for given $\mathbf{j}_8$, if $\mathbf{1}_{B_J} = 1$, the corresponding ordered tuple $\tilde{\mathbf{j}}_8$ of $\mathbf{j}_8$ satisfies the following three claims with D_0 defined in (B.46).

Claim 1 : For any index $\tilde{j}_k \in \tilde{\mathbf{j}}_8$, if it has two neighbors $\tilde{j}_{k-1}$ and $\tilde{j}_{k+1}$, its distances with the two neighbors $\tilde{j}_{k-1}$ and $\tilde{j}_{k+1}$ can not be bigger than D_0 together. That is, at least one of $|\tilde{j}_{k-1} - \tilde{j}_k| \le D_0$ and $|\tilde{j}_k - \tilde{j}_{k+1}| \le D_0$ is true. For $\tilde{j}_1$ and $\tilde{j}_8$ with only one neighbor, they satisfy $|\tilde{j}_1 - \tilde{j}_2| \le D_0$ and $|\tilde{j}_7 - \tilde{j}_8| \le D_0$.

Claim 2 : For a pair of indexes $(\tilde{j}_{k-1}, \tilde{j}_k)$ in $\tilde{\mathbf{j}}_8$, when $\tilde{j}_{k-1} \ne \tilde{j}_k$, if it has two neighbors $\tilde{j}_{k-2}$ and $\tilde{j}_{k+1}$, the distances of the pair with the two neighbors can not be bigger than D_0 together. That is, at least one of $|\tilde{j}_{k-2} - \tilde{j}_{k-1}| \le D_0$ and $|\tilde{j}_k - \tilde{j}_{k+1}| \le D_0$ holds. For the pairs $(\tilde{j}_1, \tilde{j}_2)$ and $(\tilde{j}_7, \tilde{j}_8)$ with only one neighbor, when $\tilde{j}_1 \ne \tilde{j}_2$ and $\tilde{j}_7 \ne \tilde{j}_8$, they satisfy $|\tilde{j}_2 - \tilde{j}_3| \le D_0$ and $|\tilde{j}_6 - \tilde{j}_7| \le D_0$.

Claim 3 :

(a) For given $\{\tilde{j}_4, \tilde{j}_5, \tilde{j}_6, \tilde{j}_7, \tilde{j}_8\}$,

$$\sum_{\tilde{j}_1,\tilde{j}_2,\tilde{j}_3} \mathbf{1}_{B_J \cap \{\tilde{j}_1 = \tilde{j}_2\}} = O(p^2), \quad \sum_{\tilde{j}_1,\tilde{j}_2,\tilde{j}_3} \mathbf{1}_{B_J \cap \{\tilde{j}_1 \ne \tilde{j}_2\}} = O(pD_0^2).$$

(b) For given $\{\tilde{j}_1, \tilde{j}_2, \tilde{j}_3, \tilde{j}_4, \tilde{j}_5\}$,

$$\sum_{\tilde{j}_6,\tilde{j}_7,\tilde{j}_8} \mathbf{1}_{B_J \cap \{\tilde{j}_7 = \tilde{j}_8\}} = O(p^2), \quad \sum_{\tilde{j}_6,\tilde{j}_7,\tilde{j}_8} \mathbf{1}_{B_J \cap \{\tilde{j}_7 \ne \tilde{j}_8\}} = O(pD_0^2).$$

Given three claims above, we show (B.68) by discussing different cases.

1. When both $\tilde{j}_1 \ne \tilde{j}_2$ and $\tilde{j}_7 \ne \tilde{j}_8$, by Claim 3, we know the summation over indexes $(\tilde{j}_1, \tilde{j}_2, \tilde{j}_3)$ is of order pD_0^2 and the summation over indexes $(\tilde{j}_6, \tilde{j}_7, \tilde{j}_8)$ is also of order pD_0^2. Then we consider $(\tilde{j}_4, \tilde{j}_5)$. When $|\tilde{j}_4 - \tilde{j}_5| \le D_0$, the summation is of order $(pD_0^2) \times pD_0 \times pD_0^2 = p^3 D_0^5 = p^4$. When $|\tilde{j}_4 - \tilde{j}_5| > D_0$, applying Claim 1 on $\tilde{j}_4$ and $\tilde{j}_5$ respectively, we

know $|\tilde{j}_3-\tilde{j}_4|\le D_0$ and $|\tilde{j}_5-\tilde{j}_6|\le D_0$ hold. Therefore, the summation is of order $pD_0^2\times D_0\times p\times D_0\times pD_0^2=p^3D_0^6=p^4$. In summary,

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}\mathbf{1}_{B_J\cap\{\tilde{j}_1\neq\tilde{j}_2,\tilde{j}_7\neq\tilde{j}_8\}}=O(p^4).$$

2. When only one of $\tilde{j}_1\neq\tilde{j}_2$ and $\tilde{j}_7\neq\tilde{j}_8$ holds, without loss of generality, we consider $\tilde{j}_1=\tilde{j}_2$ and $\tilde{j}_7\neq\tilde{j}_8$.
   (a) When $|\tilde{j}_2-\tilde{j}_3|>D_0$, applying Claim 1 on $\tilde{j}_3$, we know $|\tilde{j}_3-\tilde{j}_4|\le D_0$. Then consider the pair $(\tilde{j}_3,\tilde{j}_4)$. If $\tilde{j}_3=\tilde{j}_4$, by Claim 1, $|\tilde{j}_5-\tilde{j}_4|\le D_0$ or $|\tilde{j}_5-\tilde{j}_6|\le D_0$ holds. As $\tilde{j}_7\neq\tilde{j}_8$, by Claim 3, the summation over $(\tilde{j}_6,\tilde{j}_7,\tilde{j}_8)$ is of order pD_0^2. Therefore, the total summation order is $O(p\times p\times D_0\times pD_0^2)=O(p^4)$. If $\tilde{j}_3\neq\tilde{j}_4$, applying Claim 2 on the pair $(\tilde{j}_3,\tilde{j}_4)$, we know $|\tilde{j}_4-\tilde{j}_5|\le D_0$ as we discuss $|\tilde{j}_2-\tilde{j}_3|>D_0$. Also, as $\tilde{j}_7\neq\tilde{j}_8$, by Claim 3, the summation order over $(\tilde{j}_6,\tilde{j}_7,\tilde{j}_8)$ is $O(pD_0^2)$. Thus the total order of summation is $O(pD_0pD_0^2pD_0^2)=O(p^4)$. In summary,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}}\mathbf{1}_{\left\{B_J\cap\{\text{one of }\tilde{j}_1\neq\tilde{j}_2\text{ or }\tilde{j}_7\neq\tilde{j}_8,\,|\tilde{j}_2-\tilde{j}_3|>D_0\}\right\}}=O(p^4).$$

   (b) When $|\tilde{j}_2-\tilde{j}_3|\le D_0$, the summation over $\tilde{j}_1,\tilde{j}_2,\tilde{j}_3$ is of order pD_0. Then we consider $\tilde{j}_4,\tilde{j}_5$. If $|\tilde{j}_4-\tilde{j}_5|\le D_0$, the summation over $\tilde{j}_1,\tilde{j}_2,\tilde{j}_3,\tilde{j}_4,\tilde{j}_5$ is of order $pD_0pD_0=p^2D_0^2$. As $\tilde{j}_7\neq\tilde{j}_8$, by Claim 3, we know the summation order of $\tilde{j}_6,\tilde{j}_7,\tilde{j}_8$ is pD_0^2. Then the total summation order of this case is $O(1)p^2D_0^2pD_0^2=O(p^4)$. If $|\tilde{j}_4-\tilde{j}_5|>D_0$, applying Claim 1 on $\tilde{j}_4$ and $\tilde{j}_5$ respectively, we have $|\tilde{j}_3-\tilde{j}_4|\le D_0$ and $|\tilde{j}_5-\tilde{j}_6|\le D_0$. Also, as $\tilde{j}_7\neq\tilde{j}_8$, by Claim 3, we know the summation order of $\tilde{j}_6,\tilde{j}_7,\tilde{j}_8$ is $O(pD_0^2)$. Then the total summation order is $O(1)pD_0\times D_0pD_0\times pD_0^2=O(p^4)$. In summary,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}}\mathbf{1}_{\left\{B_J\cap\{\text{one of }\tilde{j}_1\neq\tilde{j}_2\text{ or }\tilde{j}_7\neq\tilde{j}_8,|\tilde{j}_2-\tilde{j}_3|\le D_0\}\right\}}=O(p^4).$$

3. When both $\tilde{j}_1=\tilde{j}_2$ and $\tilde{j}_7=\tilde{j}_8$, then we consider $(\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6)$.
   (a) If the number of distinct elements in $\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}$ is smaller and equal to 2, the order of summation over $\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6$ is $O(p^2)$. We

use $|\{\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6\}| \leq 2$ to represent this case, then

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{\left\{B_J\cap\{\tilde{j}_1=\tilde{j}_2,\,\tilde{j}_7=\tilde{j}_8,\,|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|\leq 2\}\right\}} = O(p^4).$$

(b) If the number of distinct elements in $\{\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6\}$ is 3, we use $|\{\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6\}| = 3$ to represent this case. Then two of $\tilde{j}_3 \neq \tilde{j}_4$, $\tilde{j}_4 \neq \tilde{j}_5$ and $\tilde{j}_5 \neq \tilde{j}_6$ hold. We consider without loss of generality $\tilde{j}_3 \neq \tilde{j}_4$, $\tilde{j}_4 \neq \tilde{j}_5$ and $\tilde{j}_5 = \tilde{j}_6$. We apply Claim 2 on the pair $(\tilde{j}_3, \tilde{j}_4)$ and Claim 1 on $\tilde{j}_3$. Then at least two of $|\tilde{j}_2 - \tilde{j}_3| \leq D_0$, $|\tilde{j}_3 - \tilde{j}_4| \leq D_0$ and $|\tilde{j}_4 - \tilde{j}_5| \leq D_0$ holds. Thus the summation order is $O(pD_0^2p^2) = O(p^4)$. In summary,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{\left\{B_J\cap\{\tilde{j}_1=\tilde{j}_2,\,\tilde{j}_7=\tilde{j}_8,\,|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|=3\}\right\}} = O(p^4).$$

(c) If the number of distinct elements in $\{\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6\}$ is 4, we use $|\{\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6\}| = 4$ to represent this case, and we know $\tilde{j}_3 \neq \tilde{j}_4$, $\tilde{j}_4 \neq \tilde{j}_5$ and $\tilde{j}_5 \neq \tilde{j}_6$. Applying Claim 2 on the pair $(\tilde{j}_3, \tilde{j}_4)$, and applying Claim 1 on the two single indexes $\tilde{j}_3$ and $\tilde{j}_4$ respectively, we know at least two of $|\tilde{j}_2 - \tilde{j}_3| \leq D_0$, $|\tilde{j}_3 - \tilde{j}_4| \leq D_0$ and $|\tilde{j}_4 - \tilde{j}_5| \leq D_0$ hold. Therefore the summation over $(\tilde{j}_1, \tilde{j}_2, \tilde{j}_3, \tilde{j}_4, \tilde{j}_5)$ is of order $O(p \times pD_0^2) = O(p^2D_0^2)$. Then applying Claim 1 on $\tilde{j}_6$, we know at least one of $|\tilde{j}_5 - \tilde{j}_6| \leq D_0$ and $|\tilde{j}_6 - \tilde{j}_7| \leq D_0$ holds. Then the total order of summation for this part is $O(p^2D_0^2 \times pD_0) = O(p^4)$, that is,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{B_J\cap\{\tilde{j}_1=\tilde{j}_2,\,\tilde{j}_7=\tilde{j}_8,\,|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|=4\}} = O(p^4).$$

Combining the results obtained, we know (B.68) is proved. Thus to prove (B.67), it remains to prove the three claims above.

By the definition of $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8)$ in Section B.1.6,

$$\left|Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8)\right| \leq C\left|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\right|.$$

Then it is sufficient to show that for given $\mathbf{j}_8$, when the ordered version $\tilde{\mathbf{j}}_8$ of $\mathbf{j}_8$ does not follow the three claims,

$$\left|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\right| = O(p^{-(8+\mu)}). \tag{B.69}$$

Proof of Claim 1.

(1) When the index $\tilde{j}_k$ has two neighbors, we give the proof by an example of $k = 3$. All the other cases can be obtained following similar analysis without loss of generality. Suppose $\tilde{j}_3$'s distances between its neighbors $\tilde{j}_2$ and $\tilde{j}_4$ are both bigger than D_0, i.e., $|\tilde{j}_2 - \tilde{j}_3| > D_0$ and $|\tilde{j}_3 - \tilde{j}_4| > D_0$. Then by Conditions 2.1, 2.2, and the α-mixing inequality in Lemma B.1,

$$\begin{aligned}
&\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\
=\ & \Big|\mathrm{cov}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}, \prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\Big) \times \mathrm{E}\Big(\prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\
\leq\ & C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + C \times |\mathrm{cov}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2}\ ,\ x_{k,\tilde{j}_3}) + \mathrm{E}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2})\mathrm{E}(x_{k,\tilde{j}_3})| \\
=\ & O(p^{-(8+\mu)}) + C \times |\mathrm{cov}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2}\ ,\ x_{k,\tilde{j}_3})| \\
=\ & O(p^{-(8+\mu)}).
\end{aligned}$$

Thus (B.69) holds.

(2) For $\tilde{j}_1$ and $\tilde{j}_8$ with only one neighbor, we give the proof on $\tilde{j}_1$, while $\tilde{j}_8$ can be proved similarly. By Conditions 2.1, 2.2, and Lemma B.1,

$$\begin{aligned}
\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| &= \Big|\mathrm{cov}\Big(x_{k,\tilde{j}_1}, \prod_{t=2}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}(x_{k,\tilde{j}_1}) \times \mathrm{E}\Big(\prod_{t=2}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\
&\leq C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + 0 \qquad (\ \mathrm{E}(x_{k,\tilde{j}_1}) = 0\ ) \\
&= O(p^{-(8+\mu)}).
\end{aligned}$$

Thus (B.69) also holds.

Proof of Claim 2:.

(1) When the pair $(\tilde{j}_{k-1}, \tilde{j}_k)$ has two neighbors, we give the proof by the example when $k = 5$, i.e., we consider the pair $(\tilde{j}_4, \tilde{j}_5)$. The other cases can be proved similarly without loss of generality. Suppose $\tilde{j}_4 \neq \tilde{j}_5$ with $|\tilde{j}_3 - \tilde{j}_4| > D_0$ and $|\tilde{j}_5 - \tilde{j}_6| > D_0$. As $\mathrm{E}(x_{k,\tilde{j}_4} x_{k,\tilde{j}_5}) = 0$ under H_0, by

Conditions 2.1 and 2.2, and Lemma B.1, we have

$$
\begin{aligned}
& \Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\
= \;& \Big|\mathrm{cov}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\,,\,\prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\Big) \times \mathrm{E}\Big(\prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\
\leq \;& C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + \Big|\mathrm{E}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\Big) \times \Big\{\mathrm{cov}\Big(\prod_{t=4}^{5} x_{k,\tilde{j}_t}\,,\,\prod_{t=6}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=4}^{5} x_{\tilde{j}_t}\Big)\mathrm{E}\Big(\prod_{t=6}^{8} x_{k,\tilde{j}_t}\Big)\Big\}\Big| \\
= \;& C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + \Big|\mathrm{E}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2} x_{\tilde{j}_3}) \times \{\mathrm{cov}(x_{k,\tilde{j}_4} x_{k,\tilde{j}_5}\,,\, x_{k,\tilde{j}_6} x_{k,\tilde{j}_7} x_{k,\tilde{j}_8}) + 0\}\Big| \\
\leq \;& C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).
\end{aligned}
$$

Thus (B.69) holds.

(2) For the pairs $(\tilde{j}_1, \tilde{j}_2)$ and $(\tilde{j}_7, \tilde{j}_8)$ with only one neighbor, we give the proof on $(\tilde{j}_1, \tilde{j}_2)$, while the proof on $(\tilde{j}_7, \tilde{j}_8)$ can be obtained similarly. If $\tilde{j}_1 \neq \tilde{j}_2$ and $|\tilde{j}_2 - \tilde{j}_3| > D_0$, as $\mathrm{E}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2}) = 0$ under H_0, by Conditions 2.1 and 2.2, and the α-mixing inequality in Lemma B.1, we have

$$
\begin{aligned}
\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| \;&=\; \Big|\mathrm{cov}\Big(\prod_{t=1}^{2} x_{k,\tilde{j}_t}\,,\,\prod_{t=3}^{8} x_{\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=1}^{2} x_{k,\tilde{j}_t}\Big)\mathrm{E}\Big(\prod_{t=3}^{8} x_{\tilde{j}_t}\Big)\Big| \\
&\leq\; C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(Cp^{-(8+\mu)}).
\end{aligned}
$$

Thus (B.69) holds.

Proof of Claim 3:. The Claim 3 (a) is obtained by applying Claim 1 on the $\tilde{j}_1$ and Claim 2 on the pair $(\tilde{j}_1, \tilde{j}_2)$ when $\tilde{j}_1 \neq \tilde{j}_2$. The Claim 3 (b) is also obtained similarly.

Proof under Condition 2.2.* In this section, we prove Lemma A.6 by substituting Condition 2.2 with Condition 2.2*. Similarly to Section B.1.5, the proof under Condition 2.2* follows similarly to the proof under the independence case in Section B.1.6. In particular, we note that Condition 2.2* implies that if one of the indexes in $\{j_1, \ldots, j_8\}$ only appears once, $\mathrm{E}(\prod_{r=1}^{8} x_{k,j_r}) = 0$. Therefore when $\mathrm{E}(\prod_{r=1}^{8} x_{k,j_r}) \neq 0$, (B.62) holds. Also following similar analysis, we know (B.63) holds by Condition 2.2* and $\mathrm{E}(x_{1,j_1} x_{1,j_2}) = 0$ for $j_1 \neq j_2$. Combining (B.62) and (B.63), Lemma A.6 is proved.

B.2. Lemmas for the proof of Theorem 2.3.

B.2.1. *Proof of Lemma A.7 (on Page 43, Section A.3).* For easy illustration, we first prove Lemma A.7 when $m = 1$ in Section B.2.1, and next present the proof for $m > 1$ in Section B.2.1.

Proof for $m = 1$. Specifically, in this section, we prove

$$\left| P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \leq 2z\Big) - P\Big(\frac{\hat{M}_n}{n} > y_p\Big) P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \leq 2z\Big) \right| \to 0.$$

Note that by definitions in (A.8) and (A.9),

$$\begin{aligned} \text{(B.70)} \qquad & P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \leq 2z\Big) \\ = \; & P\Big(\max_{1\leq l\leq q} (\hat{G}_l)^2 > ny_p, (\sigma(a)P_a^n)^{-1} \sum_{m=1}^{q} U_m^a \leq z\Big) \\ = \; & P\Big(\Big\{ \cup_{l=1}^{q} \{(\hat{G}_l)^2 > ny_p\}\Big\} \cap \Big\{(\sigma(a)P_a^n)^{-1} \sum_{m=1}^{q} U_m^a \leq z\Big\}\Big). \end{aligned}$$

Define the events $E_l = \{(\hat{G}_l)^2 > ny_p\} \cap \{(\sigma(a)P_a^n)^{-1} \sum_{m=1}^{q} U_m^a \leq z\}$, then we have

$$\text{(B.71)} \qquad \text{(B.70)} = P(\cup_{l=1}^{q} E_l).$$

We next examine the upper and lower bounds of (B.71). Particularly, using the Bonferroni's inequality, for any even number $d < [q/2]$, we obtain

$$\begin{aligned} \text{(B.72)} \qquad \sum_{s=1}^{d} (-1)^{s-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P(\cap_{t=1}^{s} E_{l_t}) &\leq P(\ \cup_{l=1}^{q} E_l\ ) \\ &\leq \sum_{s=1}^{d-1} (-1)^{s-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P(\cap_{t=1}^{s} E_{l_t}). \end{aligned}$$

We consider $d = O(\log^{1/5} p)$ below. The following proof proceeds by examining the upper and lower bounds of $P(\cap_{t=1}^{s} E_{l_t})$ first and combining them based on (B.72).

To facilitate the discussion, we define some notation. Let

$$H_d = \sum_{s=1}^{d} (-1)^{s-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P(\cap_{t=1}^{s} \{(\hat{G}_{l_t})^2 > ny_p\}).$$

By the Bonferroni's inequality, we have

$$H_d \le P(\ \cup_{l=1}^{q}\{(\hat{G}_l)^2 > ny_p\}\ ) \le H_{d-1}. \tag{B.73}$$

Given $l_1, \ldots, l_s$, we define two index sets: $I_s = \{(j_{l_t}^1, j_{l_t}^2), 1 \le t \le s\}$ and correspondingly

$$L_{I_s} = \{(j_1, j_2) :\ (j_1, j_2) \cap (u, t) \ne \emptyset, (u, t) \in I_s \text{ and } (j_1, j_2) \in L\}, \tag{B.74}$$

where L is defined in (A.7). (B.74) suggests that L_{I_s} contains all the index pairs that have overlap with the index pairs in I_s. Note that the definitions of I_s and L_{I_s} depend on the given indexes $l_1, \ldots, l_s$; for the simplicity of notation, we write I_s and L_{I_s} in this proof without ambiguity. It follows that

$$\sum_{m=1}^{q} U_m^a = \sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a + \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^a. \tag{B.75}$$

The cardinality of L_{I_s} is no greater than $2ps$ by construction. Furthermore, $2ps \le 2pd$ as $s \le d$. Note that the indexes in I_s and $L \setminus L_{I_s}$ have no intersection. By this construction and the independence assumption in Condition 2.3, for any finite integers $a_1, a_2 \ge 1$, we know

$$\{U_l^{a_1},\ (j_l^1, j_l^2) \in I_s\} \quad \text{and} \quad \{U_l^{a_2},\ (j_l^1, j_l^2) \in L \setminus L_{I_s}\}$$

are independent.

We next examine the upper bound of $P(\cap_{t=1}^{s} E_{l_t})$. By the definition of E_l and (B.75),

$$\begin{aligned} & P(\cap_{t=1}^{s} E_{l_t}) \\ &= P\Big(\bigcap_{t=1}^{s}\Big\{\Big\{(\sigma(a)P_a^n)^{-1}\sum_{m=1}^{q} U_m^a \le z\Big\}\bigcap\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\Big) \\ &= P\Big(\bigcap_{t=1}^{s}\Big\{\Big\{(\sigma(a)P_a^n)^{-1}\Big[\sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a + \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^a\Big] \le z\Big\} \\ &\qquad\qquad \bigcap\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\Big). \end{aligned} \tag{B.76}$$

Let Γ_p represent a number of order $\Theta\{(\log p)^{-1/2}\}$ and we have

$$\begin{aligned} & \Big\{(\sigma(a)P_a^n)^{-1}\Big(\sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a + \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^a\Big) \le z\Big\} \\ \subseteq & \Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a\Big| \ge \Gamma_p\Big\}\bigcup\Big\{(\sigma(a)P_a^n)^{-1}\sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^a \le \Gamma_p + z\Big\}. \end{aligned}$$

Thus (B.76) has the following upper bound,

$$
\begin{aligned}
\text{(B.76)} \leq\ & P\Big(\Big\{ \cap_{t=1}^{s} \{(\hat{G}_{l_t})^2 > ny_p\}\Big\} \bigcap \Big\{(\sigma(a)P_a^n)^{-1}\Big| \sum_{(j_l^1, j_l^2)\in L_{I_s}} U_l^a \Big| \geq \Gamma_p\Big\}\Big) \\
& + P\Big(\Big\{ \cap_{t=1}^{s} \{(\hat{G}_{l_t})^2 > ny_p\}\Big\} \bigcap \Big\{(\sigma(a)P_a^n)^{-1} \sum_{(j_l^1, j_l^2)\in L\backslash L_{I_s}} U_l^a \leq \Gamma_p + z\Big\}\Big).
\end{aligned}
$$

In addition, we note that $\{\hat{G}_l,\ (j_l^1, j_l^2) \in I_s\}$ and $\{U_l^a,\ (j_l^1, j_l^2) \in L\backslash L_{I_s}\}$ are independent, because of $I_s \cap (L\backslash L_{I_s}) = \emptyset$ by the construction and the independence assumption in Condition 2.3. It follows that

$$
\text{(B.76)} \leq P_s + P_{ys}P_{+z}, \tag{B.77}
$$

where for simplicity we define

$$
\begin{aligned}
P_s &= P\Big(\Big\{(\sigma(a)P_a^n)^{-1}\Big| \sum_{(j_l^1, j_l^2)\in L_{I_s}} U_l^a \Big| \geq \Gamma_p\Big\}\Big), \\
P_{ys} &= P\Big(\bigcap_{t=1}^{s} \{(\hat{G}_{l_t})^2 > ny_p\}\Big), \\
P_{+z} &= P\Big(\Big\{(\sigma(a)P_a^n)^{-1} \sum_{(j_l^1, j_l^2)\in L\backslash L_{I_s}} U_l^a \leq \Gamma_p + z\Big\}\Big).
\end{aligned} \tag{B.78}
$$

Note that although the notation P_{ys}, P_{+z} and P_s in (B.78) suppress their dependence on the specific choice of $(l_1, \ldots, l_s)$, this will not influence the proof due to the i.i.d. assumption in Condition 2.3.

Similarly we examine the lower bound of $P(\cap_{t=1}^{s} E_{l_t})$. In particular,

$$
\begin{aligned}
& \Big\{(\sigma(a)P_a^n)^{-1} \sum_{(j_l^1, j_l^2)\in L\backslash L_{I_s}} U_l^a \leq z - \Gamma_p\Big\} \\
\subseteq\ & \Big\{(\sigma(a)P_a^n)^{-1}\Big| \sum_{(j_l^1, j_l^2)\in L_{I_s}} U_l^a \Big| \geq \Gamma_p\Big\} \bigcup \Big\{(\sigma(a)P_a^n)^{-1} \sum_{m=1}^{q} U_m^a \leq z\Big\}.
\end{aligned}
$$

Then (B.76) has the following lower bound,

$$
\begin{aligned}
\text{(B.76)} \geq\ & -P\Big(\Big\{ \cap_{t=1}^{s} \{(\hat{G}_{l_t})^2 > ny_p\}\Big\} \bigcap \Big\{(\sigma(a)P_a^n)^{-1}\Big| \sum_{(j_l^1, j_l^2)\in L_{I_s}} U_l^a \Big| \geq \Gamma_p\Big\}\Big) \\
& + P\Big(\Big\{ \cap_{t=1}^{s} \{(\hat{G}_{l_t})^2 > ny_p\}\Big\} \bigcap \Big\{(\sigma(a)P_a^n)^{-1} \sum_{(j_l^1, j_l^2)\in L\backslash L_{I_s}} U_l^a \leq z - \Gamma_p\Big\}\Big).
\end{aligned}
$$

Similarly to (B.77), by the independence between $\{\hat{G}_l,\ (j_l^1, j_l^2) \in I_s\}$ and $\{U_l^a,\ (j_l^1, j_l^2) \in L \backslash L_{I_s}\}$, we obtain

$$\text{(B.76)} \geq P_{ys} \times P_{-z} - P_s, \tag{B.79}$$

where P_{ys} and P_s are defined same as in (B.78), and we define

$$P_{-z} = P\Big((\sigma(a)P_a^n)^{-1} \sum_{(j_l^1, j_l^2) \in L \backslash L_{I_s}} U_l^a \leq z - \Gamma_p\Big).$$

We have obtained the upper and lower bounds of $P(\cap_{t=1}^s E_{l_t})$ in (B.77) and (B.79) respectively. We next prove that P_{+z} in (B.77) and P_{-z} in (B.79) are close in the sense that there exists some constant $C > 0$,

$$|P_{+z} - P_z| \leq C \times \Gamma_p \quad \text{and} \quad |P_{-z} - P_z| \leq C \times \Gamma_p, \tag{B.80}$$

where we define $P_z = P((\sigma(a)P_a^n)^{-1} \sum_{m=1}^q U_m^a \leq z)$. To obtain (B.80), we note that $\sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a$ is a summation over index pairs in L_{I_s}, and L_{I_s} is of size $2ps$, which is $o(p^2)$ as $s \leq d$ and $d = O(\log^5 p)$. Following similar analysis of $\tilde{\mathcal{U}}^*(a)/\sigma(a) \xrightarrow{P} 0$ in Lemma A.1, we know $(\sigma(a)P_a^n)^{-1} \sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a \xrightarrow{P} 0$. Moreover, by $\tilde{\mathcal{U}}(a) = 2(P_a^n)^{-1} \sum_{l=1}^q U_l^a$ in (A.8), $\Gamma_p = \Theta(\log^{-1/2} p)$ and the convergence result in (A.6), we have for given z,

$$|P_{+z} - \Phi(2z + 2\Gamma_p)| \leq C\Gamma_p,\ |P_{-z} - \Phi(2z - 2\Gamma_p)| \leq C\Gamma_p,\ |P_z - \Phi(2z)| \leq C\Gamma_p.$$

As $|\Phi(2z + 2\Gamma_p) - \Phi(2z)| \leq C\Gamma_p$ for given z, $|P_{+z} - P_z| \leq |P_{+z} - \Phi(2z + 2\Gamma_p)| + |\Phi(2z + 2\Gamma_p) - \Phi(2z)| + |P_z - \Phi(2z)| \leq C\Gamma_p$. Similarly, as $|\Phi(2z - 2\Gamma_p) - \Phi(2z)| \leq C\Gamma_p$, $|P_{-z} - P_z| \leq C\Gamma_p$. Therefore (B.80) is obtained.

In summary, given (B.77), (B.79) and (B.80), we have

$$|P(\cap_{t=1}^s E_{l_t}) - P_{ys} \times P_z| \leq P_s + C \times \Gamma_p \times P_{ys}.$$

Given the above property of $P(\cap_{t=1}^s E_{l_t})$, we next derive an upper bound of (B.71) based on the relationship in (B.72). Specifically,

$$\begin{aligned}
& P(\cup_{l=1}^q E_l) \\
& \leq \sum_{s=1}^{d-1} (-1)^{s-1} \sum_{1 \leq l_1 < \ldots < l_s \leq q} P(\cap_{t=1}^s E_{l_t}) \\
& \leq \sum_{s=1}^{d-1} (-1)^{s-1} \sum_{1 \leq l_1 < \ldots < l_s \leq q} \{P_{ys} P_z + (-1)^{s-1} \times [C\Gamma_p \times P_{ys} + P_s]\} \\
& \leq H_{d-1} \times P_z + \sum_{s=1}^{d-1} \sum_{1 \leq l_1 < \ldots < l_s \leq q} (C \times \Gamma_p \times P_{ys} + P_s), \qquad \text{(B.81)}
\end{aligned}$$

where the last inequality uses the notation in (B.73), i.e.,

$$H_{d-1} = \sum_{s=1}^{d-1} (-1)^{s-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P_{ys}, \tag{B.82}$$

and the fact that P_z does not depend on $l_1, \ldots, l_s$ in summation. From (B.73), we know $H_{d-1} \le P_y + |H_{d-1} - H_d|$, where we define

$$P_y = P\Big(\bigcup_{l=1}^{q} \{(\hat{G}_l)^2 > n y_p\} \Big). \tag{B.83}$$

As a result, we have

$$\text{(B.81)} \le P_y \times P_z + |H_{d-1} - H_d| \times P_z + \sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} (C\Gamma_p P_{ys} + P_s).$$

Next we prove $|H_{d-1} - H_d| \times P_z \to 0$, $\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} \Gamma_p \times P_{ys} \to 0$ and $\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P_s \to 0$ by the following three Lemmas B.5–B.7, respectively.

Lemma B.5. *Under the conditions of Theorem 2.3, when $s = O(\log^{1/5} p)$,*

$$\begin{aligned} &\sum_{1 \le l_1 < \ldots < l_s \le q} P\Big(\bigcap_{t=1}^{s} \{(\hat{G}_{l_t})^2 / n \ge 4 \log p - \log\log p + y\} \Big) \\ = \;& \frac{1}{s!} \Big(\frac{1}{2\sqrt{2\pi}} e^{-\frac{y}{2}} \Big)^s (1 + o(1)) + o(1). \end{aligned}$$

Proof. *See Section B.2.2 on Page 119.* □

Lemma B.6. *Under the conditions of Theorem 2.3, when $d = O(\log^{1/5} p)$,*

$$\begin{aligned} &\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P\Big(\bigcap_{t=1}^{s} \{(\hat{G}_{l_t})^2 / n \ge 4 \log p - \log\log p + y\} \Big) \\ = \;& \sum_{s=1}^{d-1} \frac{1}{s!} \Big(\frac{1}{2\sqrt{2\pi}} e^{-y/2} \Big)^s \{1 + o(1)\} + o(1). \end{aligned}$$

Proof. *See Section B.2.3 on Page 124.* □

LEMMA B.7. *Under the conditions of Theorem 2.3,*

$$\sum_{s=1}^{d-1} \sum_{1\le l_1<\ldots<l_s\le q} P\Big(\Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a\Big| \ge \Gamma_p\Big\}\Big) \to 0,$$

where L_{I_s} is defined in (B.74), $d = O(\log^{1/5} p)$, $q = \binom{p}{2}$ and $\Gamma_p = \Theta(\log^{-1/2} p)$.

PROOF. *See Section B.2.4 on Page 125.* □

First, we show $|H_{d-1} - H_d| \times P_z \to 0$. By Lemma B.5, when $d \to \infty$,

$$\begin{aligned} |H_{d-1} - H_d| &= \sum_{1\le l_1<\ldots<l_d\le q} P\Big(\bigcap_{t=1}^{d}\{(\hat{G}_{l_t})^2 > ny_p\}\Big) \\ &\le C\frac{1}{d!}\Big(\frac{1}{2\sqrt{2\pi}}e^{-y/2}\Big)^d \le Ce \times \Big(\frac{e^{1-y/2}}{2\sqrt{2\pi}d}\Big)^d \to 0, \end{aligned}$$

where the last inequality follows from $d! \ge e(d/e)^d$. Second, we show that $\sum_{s=1}^{d-1}\sum_{1\le l_1<\ldots<l_s\le q}\Gamma_p P_{ys} \to 0$. By the definition of P_{ys} in (B.78), and Lemma B.5, $\sum_{s=1}^{d-1}\sum_{1\le l_1<\ldots<l_s\le q}\Gamma_p P_{ys} = \Gamma_p\sum_{s=1}^{d-1}\frac{1}{s!}(\frac{1}{2\sqrt{2\pi}}e^{-y/2})^s + o(1) \to 0$, where we use $\Gamma_p = \Theta(\log^{-1/2} p) \to 0$ and $\sum_{s=1}^{d-1}\frac{1}{s!}(\frac{1}{2\sqrt{2\pi}}e^{-y/2})^s < \infty$ from $s! \ge e(s/e)^s$. Third, we obtain $\sum_{s=1}^{d-1}\sum_{1\le l_1<\ldots<l_s\le q} P_s \to 0$ directly from Lemma B.7 following the notation P_s in (B.78).

In summary, the analysis above shows that $P(\cup_{l=1}^q E_l) \le P_y \times P_z + o(1)$. On the other hand, following similar arguments, we can obtain $P(\cup_{l=1}^q E_l) \ge P_y \times P_z + o(1)$. Therefore, $|P(\cup_{l=1}^q E_l) - P_y \times P_z| \to 0$ is obtained, that is,

$$\Big|P(\cup_{l=1}^q E_l) - P(\ \cup_{l=1}^q\{(\hat{G}_l)^2 > ny_p\}\ )P\Big(\Big\{(\sigma(a)P_a^n)^{-1}\sum_{m=1}^{q} U_m^a \le z\Big\}\Big)\Big| \to 0.$$

Recall the notation in (B.70) and (B.71). We then know Lemma A.7 is proved for $m = 1$.

Proof for $m > 1$. We still use the notation defined in Section A.3, where $U_l^{a_r}$ and $\tilde{\mathcal{U}}(a_r)$ for $r = 1, \ldots, m$ follow the definitions in (A.8) and (2.5) respectively. To prove Lemma A.7 for $m > 1$, we note that similarly to (B.71), we can write

$$P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a_1)}{\sigma(a_1)} \le 2z_1, \ldots, \frac{\tilde{\mathcal{U}}(a_m)}{\sigma(a_m)} \le 2z_m\Big) = P(\cup_{l=1}^q E_l), \tag{B.84}$$

where we redefine the events

$$E_l = \bigcap_{r=1}^{m} \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \sum_{v=1}^{q} U_v^{a_r} \le z_r \Big\} \cap \{ (\hat{G}_l)^2 > n y_p \}.$$

It follows that (B.72) and (B.73) still hold. For given $l_1, \dots, l_s$, we define I_s and L_{I_s} same as in (B.74). Then for $r = 1, \dots, m$, we write

$$\sum_{v=1}^{q} U_v^{a_r} = \sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^{a_r} + \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^{a_r}.$$

By the construction of L_{I_s} and the independence assumption in Condition 2.3, we know

$$\cup_{r=1}^{m} \{ U_l^{a_r},\ (j_l^1, j_l^2) \in I_s \} \quad \text{and} \quad \cup_{r=1}^{m} \{ U_l^{a_r},\ (j_l^1, j_l^2) \in L \setminus L_{I_s} \}$$

are independent.

Similarly to (B.76), given $l_1, \dots, l_s$, we have

$$\begin{aligned} \text{(B.85)} \quad & P(\cap_{t=1}^{s} E_{l_t}) \\ &= P\bigg(\bigcap_{r=1}^{m} \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \Big[\sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^{a_r} + \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^{a_r} \Big] \le z_r \Big\} \\ &\qquad \cap \{ \cap_{t=1}^{s} \{ (\hat{G}_{l_t})^2 > n y_p \} \} \bigg). \end{aligned}$$

We take Γ_p same as in Section B.2.1 with $\Gamma_p = \Theta\{(\log p)^{-1/2}\}$. Then for each $r = 1, \dots, m$, we have

$$\begin{aligned} & \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \Big[\sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^{a_r} + \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^{a_r} \Big] \le z_r \Big\} \\ & \subseteq \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \Big| \sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^{a_r} \Big| \ge \Gamma_p \Big\} \bigcup \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^{a_r} \le \Gamma_p + z_r \Big\}, \end{aligned}$$

and

$$\begin{aligned} & \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \sum_{(j_l^1, j_l^2) \in L \setminus L_{I_s}} U_l^{a_r} \le z_r - \Gamma_p \Big\} \\ & \subseteq \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \Big| \sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^{a_r} \Big| \ge \Gamma_p \Big\} \bigcup \Big\{ (\sigma(a_r) P_{a_r}^n)^{-1} \sum_{v=1}^{q} U_v^{a_r} \le z_r \Big\}. \end{aligned}$$

Therefore similarly to (B.77) and (B.79), we know

$$\text{(B.86)}\qquad \text{(B.85)} \le P_{ys}P_{+z} + \sum_{r=1}^{m} P_{s_r}, \quad \text{(B.85)} \ge P_{ys}P_{-z} - \Big(\sum_{r=1}^{m} P_{s_r}\Big),$$

where P_{ys} is defined in (B.78) and we further define

$$\begin{aligned}
P_{+z} &= P\Big(\bigcap_{r=1}^{m}\Big\{(\sigma(a_r)P_{a_r}^{n})^{-1}\sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r + \Gamma_p\Big\}\Big),\\
P_{s_r} &= P\Big((\sigma(a_r)P_{a_r}^{n})^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^{a_r}\Big| \ge \Gamma_p\Big),\\
P_{-z} &= P\Big(\bigcap_{r=1}^{m}\Big\{(\sigma(a_r)P_{a_r}^{n})^{-1}\sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r - \Gamma_p\Big\}\Big).
\end{aligned}$$

We note that the cardinality of L_{I_s} is no greater than $2ps$ which is $o(p^2)$. Similarly to Section B.2.1, we know $(\sigma(a_r)P_{a_r}^{n})^{-1} \times \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \xrightarrow{P} 0$ for $r = 1,\dots,m$. Combined with Theorem 2.1, we know $\{(\sigma(a_r)P_{a_r}^{n})^{-1} \times \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} : r = 1,\dots,m\}$ converges to $\mathcal{N}(0, I_m)$ and thus are asymptotically independent. We then have

$$\text{(B.87)}\qquad \Big|P_{+z} - \prod_{r=1}^{m} P_{+z_r}\Big| \to 0, \quad \Big|P_{-z} - \prod_{r=1}^{m} P_{-z_r}\Big| \to 0,$$

where we define

$$\begin{aligned}
P_{+z_r} =& P\Big((\sigma(a_r)P_{a_r}^{n})^{-1}\sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r + \Gamma_p\Big),\\
P_{-z_r} =& P\Big((\sigma(a_r)P_{a_r}^{n})^{-1}\sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r - \Gamma_p\Big).
\end{aligned}$$

Similarly to (B.80), for each $r = 1,\dots,m$, we have

$$\text{(B.88)}\qquad |P_{+z_r} - P_{z_r}| \le C\Gamma_p \quad \text{and} \quad |P_{-z_r} - P_{z_r}| \le C\Gamma_p,$$

where we define $P_{z_r} = P((\sigma(a_r)P_{a_r}^{n})^{-1}\sum_{v=1}^{q} U_v^{a_r} \le z_r)$. Combining (B.87) and (B.88), we have

$$\Big|P_{+z} - \prod_{r=1}^{m} P_{z_r}\Big| \to 0 \quad \text{and} \quad \Big|P_{-z} - \prod_{r=1}^{m} P_{z_r}\Big| \to 0.$$

By (B.86) and (B.88),

$$\left|(\text{B.85}) - P_{ys}\prod_{r=1}^{m} P_{z_r}\right| \leq o(1)P_{ys} + \sum_{r=1}^{m} P_{s_r}. \tag{B.89}$$

Given (B.89), similarly to (B.81), we have

$$\begin{aligned}
&P(\cup_{l=1}^{q} E_l) \\
&\leq \sum_{s=1}^{d-1}(-1)^{s-1}\sum_{1\leq l_1<\ldots<l_s\leq q}\left\{P_{ys}\prod_{r=1}^{m}P_{z_r} + (-1)^{s-1}\times\left[o(1)P_{ys}+\sum_{r=1}^{m}P_{s_r}\right]\right\} \\
&\leq H_{d-1}\prod_{r=1}^{m}P_{z_r} + \sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q}\left\{o(1)P_{ys}+\sum_{r=1}^{m}P_{s_r}\right\} \\
&\leq P_y\prod_{r=1}^{m}P_{z_r} + |H_{d-1}-H_d|\prod_{r=1}^{m}P_{z_r} + \sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q}\left\{o(1)P_{ys}+\sum_{r=1}^{m}P_{s_r}\right\},
\end{aligned}$$

where H_{d-1} follows the definition in (B.82) and we use (B.73) and the definition (B.83) in the last inequality. By Lemma B.5, $|H_{d-1} - H_d| \to 0$; by Lemma B.6, $o(1)\sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q} P_{ys} \to 0$; by Lemma B.7, $\sum_{r=1}^{m}\sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q} P_{s_r} =\to 0$.

In summary, we have shown that $P(\cup_{l=1}^{q} E_l) \leq P_y \times \prod_{r=1}^{m} P_{z_r} + o(1)$. Moreover, following similar arguments, we have $P(\cup_{l=1}^{q} E_l) \geq P_y \times \prod_{r=1}^{m} P_{z_r} + o(1)$. Therefore, $|P(\cup_{l=1}^{q} E_l) - P_y \times \prod_{r=1}^{m} P_{z_r}| \to 0$ is obtained, that is,

$$\left|P(\cup_{l=1}^{q} E_l) - P(\cup_{l=1}^{q}\{(\hat{G}_l)^2 > ny_p\})\prod_{r=1}^{m} P((\sigma(a_r)P_{a_r}^{n})^{-1}\sum_{v=1}^{q} U_v^{a_r} \leq z_r)\right| \to 0.$$

Since (B.84) $= P(\cup_{l=1}^{q} E_l)$, $\{\hat{M}_n/n > y_p\} = \cup_{l=1}^{q}\{(\hat{G}_l)^2 > ny_p\}$ and $\tilde{\mathcal{U}}(a_r) = 2(P_{a_r}^{n})^{-1}\sum_{v=1}^{q} U_v^{a_r}$, we know Lemma A.7 is proved for $m > 1$.

B.2.2. *Proof of Lemma B.5 (on Page 115, Section B.2.1).* In this section, we prove Lemma B.5. The proof will use Lemmas B.2.2 and B.2.2, which will be presented and proved in Sections B.2.2 and B.2.2, respectively.

PROOF. Following the definitions in (A.9), $\hat{G}_l$ will not change if $x_{i,j}$ is scaled by its standard deviation $\sigma_{j,j}$. Thus in the discussion below, we assume without loss of generality that $\sigma_{j,j} = 1$, $j = 1, \ldots, p$ for the simplicity of representation.

Given i and $1 \leq l_1 < \ldots < l_s \leq q$, we define $\check{\mathcal{X}}_{i,j_{l_t}^1,j_{l_t}^2} = x_{i,j_{l_1}^1} x_{i,j_{l_1}^2} \times \mathbf{1}\{|x_{i,j_{l_t}^1} x_{i,j_{l_t}^2}| \leq \tau_n\}$ for $t = 1, \ldots, s$, $\mathbf{W}_i = (\check{\mathcal{X}}_{i,j_{l_1}^1,j_{l_1}^2}, \ldots, \check{\mathcal{X}}_{i,j_{l_s}^1,j_{l_s}^2})^\intercal$, and let

$|\mathbf{W}_i|_{\min}$ denote the minimum absolute value of the entries in the vector $\mathbf{W}_i$. It follows that $P(\cap_{t=1}^{s}\{(\hat{G}_{l_t})^2/n \geq 4\log p - \log\log p + y\}) = P(|\sum_{i=1}^{n}\mathbf{W}_i|_{\min} \geq \sqrt{n}y_p^{1/2})$, where y_p is defined in (A.10).

We prove Lemma B.5 through examining $\mathbf{W}_i$, $i = 1, \ldots, n$. Since $\mathbf{W}_i$'s are independent and identically distributed random vectors, $\mathrm{cov}(\sum_{i=1}^{n}\mathbf{W}_i) = n \times \mathrm{cov}(\mathbf{W}_1)$. We apply Theorem 1.1 in [81] and obtain

$$
\begin{aligned}
\text{(B.90)} \qquad & P\Big(\Big|\sum_{i=1}^{n}\mathbf{W}_i\Big|_{\min} \geq \sqrt{n}y_p^{1/2}\Big) \\
\leq \;& P\Big(|\mathbf{N}_s|_{\min} \geq \sqrt{n}y_p^{1/2} - \epsilon\sqrt{n}(\log p)^{-1/2}\Big) + \\
& c_1 s^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon}{c_2 s^{5/2}\tau_n(\log p)^{1/2}}\Big),
\end{aligned}
$$

where c_1 and c_2 are positive constants; $\epsilon \to 0$, which will be specified later; and $\mathbf{N}_s := (N_{l_1}, \ldots, N_{l_s})^\intercal$ follows multivariate normal distribution with $\mathrm{E}(\mathbf{N}_s) = 0$ and $\mathrm{cov}(\mathbf{N}_s) = \mathrm{cov}(\sum_{i=1}^{n}\mathbf{W}_i) = n \times \mathrm{cov}(\mathbf{W}_1)$. Moreover, we apply Theorem 1.1 in [81] in terms of lower bound and obtain

$$
\begin{aligned}
& P\left(\Big|\sum_{i=1}^{n}\mathbf{W}_i\Big|_{\min} \geq \sqrt{n}y_p^{1/2}\right) \\
\geq \;& P\Big(|\mathbf{N}_s|_{\min} \geq \sqrt{n}y_p^{1/2} + \epsilon\sqrt{n}(\log p)^{-1/2}\Big) - c_1 s^{5/2}\exp\left(-\frac{n^{1/2}\epsilon}{c_2 s^{5/2}\tau_n(\log p)^{1/2}}\right).
\end{aligned}
$$

As $s = O(\log^{1/5} p)$, $\log p = o(n^{1/7})$, and $\tau_n = \tau\log(p+n)$, when $\epsilon \to 0$ sufficiently slow, there exists a constant $M > 0$ such that

$$
c_1 s^{5/2}\exp\Big(-\frac{\epsilon n^{1/2}}{c_2 s^{5/2}\tau_n(\log p)^{1/2}}\Big) = O(1)e^{-Mn^{3/14}}.
$$

Therefore, for $s = O(\log^{1/5} p)$,

$$
\begin{aligned}
\text{(B.91)} \qquad & \sum_{1\leq l_1<\ldots<l_s\leq q} c_1 s^{5/2}\exp\left(-\frac{\epsilon n^{1/2}}{c_2 s^{5/2}\tau_n(\log p)^{1/2}}\right) \\
= \;& O(1)q^s \times e^{-Mn^{3/14}} = O(1)e^{-Mn^{3/14}+2s\log p} = o(1).
\end{aligned}
$$

In summary, by (B.91) and Lemma B.8 in Section B.2.2 below, Lemma B.5 is proved. □

Lemma B.8 and its proof.

LEMMA B.8. *For $s = O(\log^{1/5} p)$ and $\mathbf{N}_s$ in (B.90),*

$$\sum_{1\le l_1<\ldots<l_s\le q} P\Big[|\mathbf{N}_s|_{\min} \ge \sqrt{n}\{y_p^{1/2} \pm \epsilon(\log p)^{-1/2}\}\Big] \simeq \frac{1}{s!}\Big\{\frac{1}{2\sqrt{2\pi}}\exp\Big(-\frac{y}{2}\Big)\Big\}^s.$$

PROOF. We write $v_p = y_p^{1/2} \pm \epsilon(\log p)^{-1/2}$, which represents two numbers in this proof. Since the proof below will be the same for the two numbers respectively, we abuse the use of notation v_p below.

We define $\mathbf{U}_s = \mathrm{cov}(\mathbf{W}_1)$, where $\mathbf{W}_1$ is defined in Section B.2.2. By the density of multivariate normal,

$$\begin{aligned}
& P\Big(|\mathbf{N}_s|_{\min} \ge \sqrt{n}(y_p^{1/2} \pm \epsilon(\log p)^{-1/2})\Big) \\
= \;& P\left(\frac{1}{\sqrt{n}}|\mathbf{N}_s|_{\min} \ge y_p^{1/2} \pm \epsilon(\log p)^{-1/2}\right) \\
= \;& \frac{1}{(2\pi)^{s/2}|\mathbf{U}_s|^{1/2}}\int_{|\mathbf{y}_{\min}|\ge v_p}\exp\Big(-\frac{1}{2}\mathbf{y}^\intercal(\mathbf{U}_s)^{-1}\mathbf{y}\Big)d\mathbf{y} \\
= \;& \frac{1}{(2\pi)^{s/2}}\int_{|\mathbf{U}_s^{1/2}\mathbf{z}_{\min}|\ge v_p}\exp\Big(-\frac{1}{2}\mathbf{z}^\intercal\mathbf{z}\Big)d\mathbf{z}. \qquad \text{(B.92)}
\end{aligned}$$

We note that $\mathbb{Z}_{P,1} \le$ (B.92) $\le \mathbb{Z}_{P,1} + \mathbb{Z}_{P,2}$, where we define

$$\begin{aligned}
\mathbb{Z}_{P,1} &= \frac{1}{(2\pi)^{s/2}}\int_{|\mathbf{U}_s^{1/2}\mathbf{z}|_{\min}\ge v_p, |\mathbf{z}|_{\max}\le 4\sqrt{s\log p}}\exp\Big(-\frac{1}{2}\mathbf{z}^\intercal\mathbf{z}\Big)d\mathbf{z}, \\
\mathbb{Z}_{P,2} &= \frac{1}{(2\pi)^{s/2}}\int_{|\mathbf{z}|_{\max}> 4\sqrt{s\log p}}\exp\Big(-\frac{1}{2}\mathbf{z}^\intercal\mathbf{z}\Big)d\mathbf{z}.
\end{aligned}$$

To prove Lemma B.8, we show $\mathbb{Z}_{P,2} = o(1)\{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\}^s$ and $\mathbb{Z}_{P,1} \simeq \{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\}^s$ respectively in the following.

We first prove $\mathbb{Z}_{P,2} = o(1)\{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\}^s$. Let $z \sim \mathcal{N}(0,1)$. By the property of standard normal distribution, we have

$$P(z > t) \simeq (\sqrt{2\pi}t)^{-1}e^{-t^2/2} \text{ as } t \to +\infty. \qquad \text{(B.93)}$$

It follows that

$$\begin{aligned}
\mathbb{Z}_{P,2} &= s \times P(|z| > 4\sqrt{s\log p}) \\
&\simeq s \times \frac{2}{\sqrt{2\pi}\times 4\sqrt{s\log p}}\exp(-8s\log p) \\
&= \frac{1}{2\sqrt{2\pi}}\sqrt{\frac{s}{\log p}} \times p^{-8s} = o(1)\Big\{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\Big\}^s. \qquad \text{(B.94)}
\end{aligned}$$

Next we prove $\mathbb{Z}_{P,1} \simeq \{\frac{1}{\sqrt{2\pi}p^2} e^{-y/2}\}^s$. Note that

$$\begin{aligned}\mathbb{Z}_{P,1} &= \frac{1}{(2\pi)^{s/2}} \int_{|\mathbf{z}+(\mathbf{U}_s^{1/2}-I_s)\mathbf{z}|_{\min}\geq v_p, |\mathbf{z}|_{\max}\leq 4\sqrt{s\log p}} \exp\Big(-\frac{\mathbf{z}^\intercal\mathbf{z}}{2}\Big) d\mathbf{z} \\ &\leq \frac{1}{(2\pi)^{s/2}} \int_{|\mathbf{z}|_{\min}\geq v_p - |(\mathbf{U}_s^{1/2}-I_s)\mathbf{z}|_{\max}; |\mathbf{z}|_{\max}\leq 4\sqrt{s\log p}} \exp\Big(-\frac{\mathbf{z}^\intercal\mathbf{z}}{2}\Big) d\mathbf{y},\end{aligned}$$

where I_s represents an identity matrix of size $s\times s$. When $|\mathbf{z}|_{\max} \leq 4\sqrt{s\log p}$, we have $|(\mathbf{U}_s^{1/2}-I_s)\mathbf{z}|_{\max} \leq 4Cs\sqrt{s\log p}(p+n)^{-c_0\tau}$ by Lemma B.9 in Section B.2.2 below. It follows that

$$\mathbb{Z}_{P,1} \leq \frac{1}{(2\pi)^{s/2}} \int_{|\mathbf{z}|_{\min}\geq \tilde{v}_p} \exp\Big(-\frac{\mathbf{z}^\intercal\mathbf{z}}{2}\Big) d\mathbf{y}, \tag{B.95}$$

where we define $\tilde{v}_p = v_p - 4Cs\sqrt{s\log p}(p+n)^{-c_0\tau}$. We set τ as a sufficiently large constant such that $s\sqrt{s\log p} = o\{(p+n)^{c_0\tau}\}$, then $\tilde{v}_p = 2\sqrt{\log p}\{1 + o(1)\}$. By (B.93) and (B.95),

$$\begin{aligned}\mathbb{Z}_{P,1} &\leq \Big\{\frac{2}{\sqrt{2\pi}\tilde{v}_p}\exp(-\tilde{v}_p^2/2)\Big\}^s \\ &= \Big\{2\frac{1+o(1)}{\sqrt{2\pi}\sqrt{4\log p}}\exp\Big(-2\log p + (\log\log p)/2 - y/2 + o(1)\Big)\Big\}^s \\ &= \Big\{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\Big\}^s\{1+o(1)\}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathbb{Z}_{P,1} &\geq \frac{1}{(2\pi)^{s/2}} \int_{|\mathbf{z}|_{\min}\geq v_p + |(\mathbf{U}_s^{1/2}-I_s)\mathbf{z}|_{\max}, |\mathbf{z}|_{\max}\leq 4\sqrt{s\log p}} \exp\Big(-\frac{\mathbf{z}^\intercal\mathbf{z}}{2}\Big) d\mathbf{y} \\ &\geq \frac{1}{(2\pi)^{s/2}} \int_{|\mathbf{z}|_{\min}\geq v_p + 4Cs\sqrt{s\log p}(p+n)^{-\tau/2}} \exp\Big(-\frac{\mathbf{z}^\intercal\mathbf{z}}{2}\Big) d\mathbf{y} - \mathbb{Z}_{P,2} \\ &= \Big\{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\Big\}^s\{1+o(1)\}.\end{aligned}$$

We therefore obtain $\mathbb{Z}_{P,1} \simeq \{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\}^s$.

Since $\mathbb{Z}_{P,1} \leq$ (B.92) $\leq \mathbb{Z}_{P,1} + \mathbb{Z}_{P,2}$, $\mathbb{Z}_{P,1} \simeq \{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\}^s$ and $\mathbb{Z}_{P,2} = o(1)\{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\}^s$, we obtain (B.92) $\simeq \{\frac{1}{\sqrt{2\pi}p^2}e^{-y/2}\}^s$. It follows that as

$p \to \infty$ and $s = O(\log^{1/5} p)$,

$$\begin{aligned}
&\sum_{1 \le l_1 < \ldots < l_s \le q} P\Big(|\mathbf{N}_s|_{\min} \ge \sqrt{n}(y_p^{1/2} \pm \epsilon(\log p)^{-1/2})\Big) \\
&= \binom{q}{s}\Big\{\frac{1}{\sqrt{2\pi}p^2}\exp(-y/2)\Big\}^s\{1+o(1)\} \qquad \Big(q = \frac{p(p-1)}{2}\Big) \\
&= \frac{1}{s!}\Big\{\frac{1}{2\sqrt{2\pi}}\exp(-y/2)\Big\}^s\{1+o(1)\}.
\end{aligned}$$

$\square$

Lemma B.9 and its proof.

LEMMA B.9. *For $\mathbf{U}_s$ in Section B.2.2, there exist some positive constants C and c_0 such that $|\mathbf{U}_s^{1/2} - I_s|_{\max} \le C(p+n)^{-c_0\tau}$, where $|\cdot|_{\max}$ represents the element-wise maximum absolute value, and τ is the constant satisfying $\tau_n = \tau\log(p+n)$ from (A.9).*

PROOF. Recall that $\mathbf{U}_s = \mathrm{cov}(\mathbf{W}_1)$ and $\mathbf{W}_1 = (\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}, \ldots, \check{\mathcal{X}}_{1,j_{l_s}^1,j_{l_s}^2})$ for given $1 \le l_1 < \ldots < l_s \le q$, which is defined at the beginning of Section B.2.2. To prove Lemma B.9, we prove $|\mathbf{U}_s - I_s|_{\max} \le C(p+n)^{-c_0\tau}$ first. Specifically, we show the diagonal and off-diagonal elements of $\mathrm{cov}(\mathbf{W}_1) - I_s$ are bounded by $C(p+n)^{-c_0\tau}$ respectively.

First we show for given (j_l^1, j_l^2), $|\mathrm{var}(\check{\mathcal{X}}_{1,j_l^1,j_l^2}) - 1| \le C(p+n)^{-c_0\tau}$. By the independence assumption in Condition 2.3 and $\sigma_{j,j} = 1$ for $j = 1, \ldots, p$, we know $\mathrm{var}(x_{1,j_l^1}x_{1,j_l^2}) = 1$; by $\mathrm{E}(x_{1,j_l^1}x_{1,j_l^2}) = 0$, we have $\mathrm{var}(x_{1,j_l^1}x_{1,j_l^2}) = \mathrm{E}\{(x_{1,j_l^1}x_{1,j_l^2})^2\}$. It follows that

$$\begin{aligned}
\Big|\mathrm{var}(\check{\mathcal{X}}_{1,j_l^1,j_l^2}) - 1\Big| &= \Big|\mathrm{var}(\check{\mathcal{X}}_{1,j_l^1,j_l^2}) - \mathrm{var}(x_{1,j_l^1}x_{1,j_l^2})\Big| \\
&= \Big|\mathrm{E}\Big\{(\check{\mathcal{X}}_{1,j_l^1,j_l^2})^2\Big\} - \Big\{\mathrm{E}(\check{\mathcal{X}}_{1,j_l^1,j_l^2})\Big\}^2 - \mathrm{E}\Big\{(x_{1,j_l^1}x_{1,j_l^2})^2\Big\}\Big| \\
&\le \Big|\mathrm{E}\Big\{(\check{\mathcal{X}}_{1,j_l^1,j_l^2})^2\Big\} - \mathrm{E}\Big\{(x_{1,j_l^1}x_{i,j_l^2})^2\Big\}\Big| + \Big|\mathrm{E}(\check{\mathcal{X}}_{1,j_l^1,j_l^2})\Big|^2, \qquad \text{(B.96)}
\end{aligned}$$

where we use $\mathrm{var}(x_{1,j_l^1}x_{1,j_l^2}) = 1$ in the first equation; and we use the definition of $\mathrm{var}(\check{\mathcal{X}}_{1,j_l^1,j_l^2})$ and $\mathrm{var}(x_{1,j_l^1}x_{1,j_l^2}) = \mathrm{E}\{(x_{1,j_l^1}x_{1,j_l^2})^2\}$ in the second equation. Recall the definition $\check{\mathcal{X}}_{1,j_l^1,j_l^2} = x_{1,j_l^1}x_{1,j_l^2} \times \mathbf{1}\{|x_{1,j_l^1}x_{1,j_l^2}| \le \tau_n\}$. We then have

$$\begin{aligned}
&\Big|\mathrm{E}\Big\{(x_{1,j_l^1}x_{1,j_l^2})^2\Big\} - \mathrm{E}\Big\{(\check{\mathcal{X}}_{1,j_l^1,j_l^2})^2\Big\}\Big| \qquad \text{(B.97)} \\
&= \Big|\mathrm{E}\Big[(x_{1,j_l^1}x_{1,j_l^2})^2\mathbf{1}\{|x_{1,j_l^1}x_{1,j_l^2}| > \tau_n\}\Big]\Big|,
\end{aligned}$$

and $|\mathrm{E}(\check{\mathcal{X}}_{1,j_l^1,j_l^2})| = |\mathrm{E}(x_{1,j_l^1}x_{1,j_l^2} \times \mathbf{1}\{|x_{1,j_l^1}x_{1,j_l^2}| > \tau_n\})|$ as $\mathrm{E}(x_{1,j_l^1}x_{1,j_l^2}) = 0$. Since $\mathbf{1}\{|x_{1,j_l^1}x_{1,j_l^2}| > \tau_n\}) \leq \mathbf{1}\{|x_{1,j_l^1}| > \sqrt{\tau_n}\} + \mathbf{1}\{|x_{1,j_l^2}| > \sqrt{\tau_n}\}$, and x_{1,j_l^1} and x_{1,j_l^2} are i.i.d. by Condition 2.3, by Hölder's inequality, we know

$$
\begin{aligned}
\text{(B.98)} \qquad (\text{B.97}) &\leq C \times \mathrm{E}\Big(x_{1,j_l^1}^2 \mathbf{1}\{|x_{1,j_l^1}| > \sqrt{\tau_n}\}\Big) \times \mathrm{E}(x_{1,j_l^2}^2) \\
&\leq C \times \{\mathrm{E}(x_{1,j_l^1}^4) P(|x_{1,j_l^1}| > \sqrt{\tau_n}\})\}^{1/2} \times \mathrm{E}(x_{1,j_l^2}^2),
\end{aligned}
$$

and also

$$
\text{(B.99)} \qquad \Big|\mathrm{E}(\check{\mathcal{X}}_{1,j_l^1,j_l^2})\Big| \leq C \times \{\mathrm{E}(x_{1,j_l^1}^2) P(|x_{1,j_l^1}| > \sqrt{\tau_n}\})\}^{1/2} \times \mathrm{E}(|x_{1,j_l^2}|).
$$

By Markov's inequality, $P(|x_{1,j_l^1}| > \sqrt{\tau_n}\}) \leq \mathrm{E}\{\exp(t_0 x_{1,j_l^1}^2)\}\exp(-t_0\tau_n)$, where t_0 is given in Condition 2.3. Combining (B.96)–(B.99), we obtain that there exists some positive constants C and c_0 such that

$$
(\text{B.96}) \leq C \times \{\mathrm{E}(\exp(t_0 x_{1,j_l^1}^2))\exp(-t_0\tau_n)\}^{1/2} \leq C(p+n)^{-c_0\tau},
$$

where we use the assumption that x_{1,j_l^1} and x_{1,j_l^2} are i.i.d. and $\mathrm{E}\{\exp(t_0 x_{1,j_l^1}^2)\} < \infty$ as Condition 2.3 holds for $\vartheta = 2$.

Second, we prove that for given $l_1 \neq l_2$, there exist some positive constants C and c_0 such that $|\mathrm{cov}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}, \check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2})| \leq C(p+n)^{-c_0\tau}$. We note that under H_0, $\mathrm{cov}(x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}, x_{1,j_{l_2}^1}x_{1,j_{l_2}^2}) = \mathrm{E}(x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}x_{1,j_{l_2}^1}x_{1,j_{l_2}^2}) = 0$ as $j_{l_1}^1 \neq j_{l_1}^2$ and $j_{l_2}^1 \neq j_{l_2}^2$. It follows that

$$
\begin{aligned}
&\Big|\mathrm{cov}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}, \check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2})\Big| \\
&= \Big|\mathrm{cov}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}, \check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2}) - \mathrm{E}(x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}x_{1,j_{l_2}^1}x_{1,j_{l_2}^2})\Big| \\
&\leq \Big|\mathrm{E}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}\check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2}) - \mathrm{E}(x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}x_{1,j_{l_2}^1}x_{1,j_{l_2}^2})\Big| + \Big|\mathrm{E}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}) \times \mathrm{E}(\check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2})\Big|.
\end{aligned}
$$

By the definition of $\check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2}$,

$$
\begin{aligned}
&\Big|\mathrm{E}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}\check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2}) - \mathrm{E}(x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}x_{1,j_{l_2}^1}x_{1,j_{l_2}^2})\Big| \\
&\leq \Big|\mathrm{E}\Big[|x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}x_{1,j_{l_2}^1}x_{1,j_{l_2}^2}|\Big(\mathbf{1}\{|x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}| > \tau_n\} + \mathbf{1}\{|x_{1,j_{l_2}^1}x_{1,j_{l_2}^2}| > \tau_n\}\Big)\Big]\Big|.
\end{aligned}
$$

Similarly to (B.98) and (B.99), by Hölder's inequality, we know that there exist some positive constants C and c_0 such that

$$
\begin{aligned}
\Big|\mathrm{E}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}\check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2}) - \mathrm{E}(x_{1,j_{l_1}^1}x_{1,j_{l_1}^2}x_{1,j_{l_2}^1}x_{1,j_{l_2}^2})\Big| &\leq C(p+n)^{-c_0\tau}, \\
\Big|\mathrm{E}(\check{\mathcal{X}}_{1,j_{l_1}^1,j_{l_1}^2}) \times \mathrm{E}(\check{\mathcal{X}}_{1,j_{l_2}^1,j_{l_2}^2})\Big| &\leq C(p+n)^{-c_0\tau}.
\end{aligned}
$$

It follows that $|\mathrm{cov}(\check{\mathcal{X}}_{1,j^1_{l_1},j^2_{l_1}}, \check{\mathcal{X}}_{1,j^1_{l_2},j^2_{l_2}})| \leq C(p+n)^{-c_0\tau}$.

In summary, $|\mathbf{U}_s - I_s|_{\max} \leq C(p+n)^{-c_0\tau}$ is obtained. By the matrix version taylor expansion of $\mathbf{U}_s^{1/2}$ at I_s [see, e.g., 35], the element wise differences between $\mathbf{U}_s^{1/2}$ and I_s are also bounded by $C(p+n)^{-c_0\tau}$. □

B.2.3. *Proof of Lemma B.6 (on Page 115, Section B.2.1).* By the proof of Lemma B.5 in Section B.2.2, we have

$$\begin{aligned}
&\sum_{s=1}^{d-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P\Big[\bigcap_{t=1}^{s}\Big\{(\hat{G}_{l_t})^2/n \geq 4\log p - \log\log p + y\Big\}\Big] \\
=\;& \sum_{s=1}^{d-1}\left[\frac{1}{s!}\Big(\frac{1}{2\sqrt{2\pi}}e^{-y/2}\Big)^s\{1+o(1)\} + O(1)e^{-Mn^{3/14}+2s\log p}\right].
\end{aligned}$$

Since $\log p = o(1)n^{1/7}$ and $d = O(\log^{1/5} p)$, we know $Mn^{3/14} - 2d\log p - \log d \to \infty$ and $\sum_{s=1}^{d-1} O(1)e^{-Mn^{3/14}+2s\log p} \leq O(1)e^{-Mn^{3/14}+2d\log p+\log d} = o(1)$. It follows that

$$\begin{aligned}
&\sum_{s=1}^{d-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P(\cap_{t=1}^{s}\{(\hat{G}_{l_t})^2/n \geq 4\log p - \log\log p + y\}) \\
=\;& \sum_{s=1}^{d-1}\frac{1}{s!}\Big(\frac{1}{2\sqrt{2\pi}}e^{-y/2}\Big)^s\{1+o(1)\} + o(1).
\end{aligned}$$

B.2.4. *Proof of Lemma B.7 (on Page 115, Section B.2.1).*

Proof. Recall the definition of U_l^a in (A.8), and we write $U^a_{(j_l^1,j_l^2)} = U_l^a$. By Lemma A.1, we know $\sigma(a)P^n_a = \Theta(pn^{a/2})$. Then for given $l_1,\ldots,l_s$,

$$\begin{aligned}
&P\Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a\Big| \geq C\Gamma_p\Big\} \\
&\leq P\Big\{\Big|n^{-a/2}\sum_{(j_l^1,j_l^2)\in L_{I_s}} U^a_{(j_l^1,j_l^2)}\Big| \geq Cp\Gamma_p\Big\} \leq P_{U,+} + P_{U,-},
\end{aligned} \tag{B.100}$$

where we define

$$\begin{aligned}
P_{U,+} &= P\Big(n^{-a/2}\sum_{(j_l^1,j_l^2)\in L_{I_s}} U^a_{(j_l^1,j_l^2)} \geq Cp\Gamma_p\Big), \\
P_{U,-} &= P\Big(n^{-a/2}\sum_{(j_l^1,j_l^2)\in L_{I_s}} U^a_{(j_l^1,j_l^2)} \leq -Cp\Gamma_p\Big).
\end{aligned}$$

By $q = \binom{p}{2}$,

$$\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P\Big(\Big\{(\sigma(a) P_a^n)^{-1} \Big| \sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a \Big| \ge \Gamma_p \Big\}\Big) \le dp^{2d} \max_{1 \le s \le d-1;\, 1 \le l_1 < \ldots < l_s \le q} (P_{U,+} + P_{U,-}). \tag{B.101}$$

To prove Lemma B.7, it suffices to prove that $P_{U,+}$ and $P_{U,-}$ are $o(d^{-1}p^{-2d})$ for each given s and $l_1, \ldots, l_s$.

We show $P_{U,+} = o(d^{-1}p^{-2d})$ in the following and the same conclusion holds for $P_{U,-}$ by applying similar analysis. By the construction of L_{I_s} in (B.74) and the i.i.d. assumption in Condition 2.3, we know that there exists an integer $D \le 2s$ such that

$$\begin{aligned} P_{U,+} &\le \sum_{k=1}^{D} P\Big(\sum_{m=k+1}^{p} n^{-a/2} U_{(k,m)}^a \ge Cp\Gamma_n / D \Big) \\ &\le D \max_{1 \le k \le D} \mathrm{E}\Big[P_k \Big(\sum_{m=k+1}^{p} n^{-a/2} U_{(k,m)}^a \ge Cp\Gamma_p / D \Big) \Big], \end{aligned} \tag{B.102}$$

where P_k represents the probability measure conditioning on $\{x_{1,k}, \ldots, x_{n,k}\}$ with $k \in \{1, \ldots, p\}$. To prove $P_{U,+} = o(d^{-1}p^{-2d})$, in the following we show that $\mathrm{E}[P_k(\sum_{m=k+1}^{p} n^{-a/2} U_{(k,m)}^a \ge C \times p\Gamma_p / D)] = o(D^{-1}d^{-1}p^{-2d})$ for $k = 1$; and the same conclusion holds for $k \ge 2$ by similar analysis given the i.i.d. assumption in Condition 2.3 and $k \le D = O(\log^{1/5} p)$. Specifically, we next prove that $\mathrm{E}[P_1(\{\sum_{m=2}^{p} n^{-a/2} U_{(1,m)}^a \ge Cp\Gamma_p / D\})] = o(D^{-1}d^{-1}p^{-2d})$.

Define

$$\bar{U}_x = n^{-a} \sum_{1 \le i_1 \ne \ldots \ne i_a \le n} x_{i_1,1}^2 \ldots x_{i_a,1}^2, \tag{B.103}$$

then $\mathrm{E}(\bar{U}_x) \le \{\mathrm{E}(x_{11}^2)\}^a = \Theta(1)$. Given a constant $t > 0$, we define an event $T_{t,1} = \{|\bar{U}_x - \mathrm{E}(\bar{U}_x)| \le t\}$, and let $\mathbf{1}_{T_{t,1}}$ denote the indicator function of the event $T_{t,1}$. It follows that

$$\begin{aligned} &\mathrm{E}\Big[P_1 \Big(\Big\{ \sum_{m=2}^{p} n^{-a/2} U_{(1,m)}^a \ge Cp\Gamma_p / D \Big\} \Big) \Big] \\ = \ &\mathrm{E}\Big[P_1 \Big(\Big\{ \sum_{m=2}^{p} n^{-a/2} U_{(1,m)}^a \ge Cp\Gamma_p / D \Big\} \Big) \times (\mathbf{1}_{T_{t,1}} + \mathbf{1}_{T_{t,1}^c}) \Big] \\ \le \ &\mathrm{E}(P_{T_{t,1}}) + P(T_{t,1}^c), \end{aligned} \tag{B.104}$$

where $\mathbf{1}_{T_{t,1}^c} = 1 - \mathbf{1}_{T_{t,1}}$; $T_{t,1}^c$ denotes the complement set of the event $T_{t,1}$; and $P_{T_{t,1}} = P_1\{\sum_{m=2}^p n^{-a/2}U_{(1,m)}^a \geq Cp\Gamma_p/D\} \times \mathbf{1}_{T_{t,1}}$. It remains to prove that $\mathrm{E}(P_{T_{t,1}})$ and $P(T_{t,1}^c)$ are $o(D^{-1}d^{-1}p^{-2d})$ respectively.

Part 1: $\mathbf{E}(\boldsymbol{P_{T_{t,1}}})$ Given an integer a, define $h_p = C(p/\log^2 p)^{a/(a+1)}$. For easy presentation, we let $\mathbf{1}_H$ denote an indicator function of the event $\{|n^{-a/2}U_{(1,m)}^a| \leq h_p\}$. We next decompose $n^{-a/2}U_{(1,m)} = z_{m,1} + z_{m,2}$, where

$$
\begin{aligned}
z_{m,1} &= n^{-a/2}\Big[U_{(1,m)}^a\mathbf{1}_H - \mathrm{E}_1\{U_{(1,m)}^a\mathbf{1}_H\}\Big],\\
z_{m,2} &= n^{-a/2}\Big[\mathrm{E}_1\{U_{(1,m)}^a\mathbf{1}_H\} + U_{(1,m)}^a(1-\mathbf{1}_H)\Big]\\
&= n^{-a/2}\Big[-\mathrm{E}_1\{U_{(1,m)}^a(1-\mathbf{1}_H)\} + U_{(1,m)}^a(1-\mathbf{1}_H)\Big];
\end{aligned}
\tag{B.105}
$$

in (B.105), E_1 denotes the expectation conditioning on $\{x_{1,1},\ldots,x_{n,1}\}$, and we use $\mathrm{E}_1\{U_{(1,m)}^a\mathbf{1}_H\} = -\mathrm{E}_1\{U_{(1,m)}^a(1-\mathbf{1}_H)\}$ as $\mathrm{E}_1\{U_{(1,m)}^a\} = 0$. Given $n^{-a/2}U_{(1,m)}^a = z_{m,1} + z_{m,2}$, we have $P_{T_{t,1}} \leq P_{z,1} + P_{z,2}$, where we define

$$
P_{z,1} = P_1\Big(\sum_{m=2}^p z_{m,1} \geq Cp\Gamma_p/D\Big)\mathbf{1}_{T_{t,1}},\;\; P_{z,2} = P_1\Big(\sum_{m=2}^p z_{m,2} \geq Cp\Gamma_p/D\Big)\mathbf{1}_{T_{t,1}}.
$$

To evaluate $\mathrm{E}(P_{T_1})$, we examine $\mathrm{E}(P_{z,1})$ and $\mathrm{E}(P_{z,2})$ respectively below.

Part 1.1: $\mathbf{E}(\boldsymbol{P_{z,1}})$ When conditioning on $\{x_{1,1},\ldots,x_{n,1}\}$, since $z_{m,1}$'s are independent and bounded random variables, by Bernstein inequality,

$$
P_{z,1} \leq C\exp\Big(-\frac{Cp^2\Gamma_p^2/D^2}{\sum_{m=2}^p \mathrm{E}_1(z_{m,1}^2) + Ch_p p\Gamma_p/D}\Big)\mathbf{1}_{T_1}.
\tag{B.106}
$$

Note that $0 \leq \mathrm{E}_1(z_{m,1}^2) \leq \mathrm{E}_1[\{n^{-a/2}U_{(1,m)}^a\}^2]$ and

$$
\begin{aligned}
\mathrm{E}_1\Big[\{n^{-a/2}U_{(1,m)}^a\}^2\Big] &= n^{-a}\sum_{\substack{1\leq i_1\neq\ldots\neq i_a\leq n;\\ 1\leq \tilde{i}_1\neq\ldots\neq\tilde{i}_a\leq n}}\Big(\prod_{r=1}^a x_{i_r,1}x_{\tilde{i}_r,1}\Big)\times\mathrm{E}\Big(\prod_{r=1}^a x_{i_r,m}x_{\tilde{i}_r,m}\Big)\\
&= a!n^{-a}\sum_{1\leq i_1\neq\ldots\neq i_a\leq n}\Big(\prod_{r=1}^a x_{i_r,1}^2\Big)\times\{\mathrm{E}(x_{1,m}^2)\}^a\\
&= a!\bar{U}_x\times\{\mathrm{E}(x_{1,m}^2)\}^a,
\end{aligned}
$$

where from the first equation to the second equation, we use the fact that $\mathrm{E}(\prod_{r=1}^a x_{i_r,m}x_{\tilde{i}_r,m}) \neq 0$ only when $\{i_1,\ldots,i_a\} = \{\tilde{i}_1,\ldots,\tilde{i}_a\}$. It follows that

$\mathrm{E}_1(z_{m,1}^2) \leq C \times \bar{U}_x$. As $\mathbf{1}_{T_{t,1}}$ indicates the event $\{|\bar{U}_x - \mathrm{E}(\bar{U}_x)| \leq t\}$ and $\mathrm{E}(\bar{U}_x) = \Theta(1)$, it suffices to consider $\mathrm{E}_1(z_{m,1}^2) = \Theta(1)$ in (B.106) and then

$$\text{(B.107)} \qquad \mathrm{E}(P_{z,1}) \leq \exp\{-Cp\Gamma_p/(Dh_p)\}.$$

Part 1.2: $\mathbf{E}(\boldsymbol{P_{z,2}})$ By the definition of $z_{m,2}$ in (B.105),

$$\text{(B.108)} \quad \mathrm{E}(P_{z,2}) \leq P\Big(\max_{2\leq m\leq p} |n^{-a/2}U^a_{(1,m)}| > h_p\Big) \leq pP(|n^{-a/2}U^a_{(1,2)}| > h_p),$$

where the last inequality follows from the i.i.d. assumption in Condition 2.3. By the result in Section C.1.1, we know $U^a_{(1,2)} = \sum_{1\leq i_1\neq\ldots\neq i_a\leq n}\prod_{k=1}^a x_{i_k,1}x_{i_k,2}$ can be written as a linear combination of $\prod_{k=1}^{\iota}\{\sum_{i=1}^n (x_{i,1}x_{i,2})^{a_k}\}$, where $a_1,\ldots,a_\iota$ are positive integers such that $a_1+\ldots+a_\iota = a$. It follows that for finite integer a,

$$\begin{aligned}
&P(|n^{-a/2}U^a_{(1,2)}| > h_p)\\
&\leq \sum_{a_1+\ldots+a_\iota=a} P\Big(n^{-a/2}\prod_{k=1}^{\iota}\Big|\sum_{i=1}^n (x_{i,1}x_{i,2})^{a_k}\Big| > Ch_p\Big).\\
&\leq \sum_{a_1+\ldots+a_\iota=a}\sum_{k=1}^{\iota} P\Big(\sum_{i=1}^n |x_{i,1}x_{i,2}/\sqrt{n}|^{a_k} > Ch_p^{a_k/a}\Big).
\end{aligned}$$

Case 1: If $a_k = 1$, since Condition 2.3 holds for $\varsigma = 2$ in Theorem 2.2, we know $x_{i,1}x_{i,2}$, $i = 1,\ldots,n$, are i.i.d. sub-exponential random variables. By the Bernstein-type inequality of sub-exponential random variables, we have

$$\text{(B.109)} \quad P\Big(\sum_{i=1}^n |x_{i,1}x_{i,2}| > C\sqrt{n}h_p^{1/a}\Big) \leq C\exp(-C\min\{Ch_p^{2/a}, C\sqrt{n}h_p^{1/a}\}).$$

Case 2: If $2 \leq a_k \leq a$, we let $B_p = Cn^{-1/6}h_p^{2/(3a)}$. We then decompose $|x_{i,1}x_{i,2}/\sqrt{n}|^{a_k} = s_i + t_i$, where we define

$$\begin{aligned}
&s_i = |x_{i,1}x_{i,2}/\sqrt{n}|^{a_k}\mathbf{1}_{H_{B_p}} - \mu_i, \quad t_i = |x_{i,1}x_{i,2}/\sqrt{n}|^{a_k}(1-\mathbf{1}_{H_{B_p}}) + \mu_i,\\
&\mathbf{1}_{H_{B_p}} = \mathbf{1}_{\{|x_{i,1}x_{i,2}/\sqrt{n}|\leq B_p\}}, \quad \mu_i = \mathrm{E}\{|x_{i,1}x_{i,2}/\sqrt{n}|^{a_k}\mathbf{1}_{H_{B_p}}\}.
\end{aligned}$$

It follows that

$$P\Big(\sum_{i=1}^n |x_{i,1}x_{i,2}/\sqrt{n}|^{a_k} > Ch_p^{a_k/a}\Big) \leq P\Big(\sum_{i=1}^n s_i > Ch_p^{a_k/a}\Big) + P\Big(\sum_{i=1}^n t_i > Ch_p^{a_k/a}\Big).$$

Since $|s_i| \leq C \times B_p^{a_k}$ from construction, by Bernstein inequality,

$$P\Big(\sum_{i=1}^{n} s_i > Ch_p^{a_k/a}\Big) \leq C \exp\Big(-\frac{Ch_p^{2a_k/a}}{\sum_{i=1}^n \mathrm{E}(s_i^2) + CB_p^{a_k} h_p^{a_k/a}}\Big). \tag{B.110}$$

As $2 \leq a_k \leq a$, by Condition 2.3, we have

$$\sum_{i=1}^{n} \mathrm{E}(s_i^2) \leq \sum_{i=1}^{n} \mathrm{E}\Big\{\Big(\frac{x_{i,1}x_{i,2}}{\sqrt{n}}\Big)^{2a_k}\Big\} \leq \frac{\mathrm{E}\{(x_{1,1}x_{1,2})^{2a_k}\}}{n^{a_k-1}} \leq \mathrm{E}[(x_{1,1}x_{1,2})^{2a_k}] < \infty.$$

Since $h_p^{1/a}/B_p \to \infty$, from (B.110), we have

$$P\Big(\sum_{i=1}^{n} s_i > Ch_p^{a_k/a}\Big) \leq \exp(-Ch_p^{2/a}/B_p^2). \tag{B.111}$$

In addition, by the definition of t_i,

$$P\Big(\sum_{i=1}^{n} t_i > Ch_p^{a_k/a}\Big) \leq P\Big\{\sum_{i=1}^{n} |x_{i,1}x_{i,2}/\sqrt{n}|^{a_k}(1-\mathbf{1}_{H_{B_p}}) > Ch_p^{a_k/a} - \Big|\sum_{i=1}^{n}\mu_i\Big|\Big\}.$$

We note that $\sum_{i=1}^n \mu_i \leq n^{-1} \times \sum_{i=1}^n [\mathrm{E}\{(x_{1,1}x_{1,2})^{2a_k}\}]^{1/2} < \infty$ by Hölder's inequality and Condition 2.3. As $h_p \to \infty$, $Ch_p^{a_k/a} - |\sum_{i=1}^n \mu_i| > 0$ when n and p are sufficiently large. Since $1 - \mathbf{1}_{H_{B_p}}$ indicates $|x_{i,1}x_{i,2}/\sqrt{n}| > B_p$,

$$\begin{aligned} P\Big(\sum_{i=1}^{n} t_i > Ch_p^{a_k/a}\Big) &\leq P\Big(\max_{1\leq i\leq n} |x_{i,1}x_{i,2}/\sqrt{n}|^{a_k} > B_p^{a_k}\Big) \\ &\leq n \times P(|x_{i,1}x_{i,2}/\sqrt{n}| > B_p) \\ &\leq n \times \mathrm{E}\{\exp(t_0|x_{1,1}x_{1,2}|)\}/\exp\{t_0(\sqrt{n}B_p)\} \\ &\leq \exp(-C\sqrt{n}B_p + \log n), \end{aligned} \tag{B.112}$$

where we use $\mathrm{E}\{\exp(t_0|x_{1,1}x_{1,2}|)\} \leq \mathrm{E}\{\exp(t_0(x_{1,1}^2 + x_{1,2}^2)/2)\} < \infty$ as Condition 2.3 holds for $\varsigma = 2$. By (B.108), (B.109), (B.111) and (B.112),

$$\begin{aligned} \mathrm{E}(P_{z,2}) \leq Cp \times \Big[&\exp\Big(-C\min\{Ch_p^{2/a}, C\sqrt{n}h_p^{1/a}\}\Big) \\ &+ \exp(-Ch_p^{2/a}/B_p^2) + \exp(-C\sqrt{n}B_p + \log n)\Big]. \end{aligned} \tag{B.113}$$

Part 2: $\boldsymbol{P(T_{t,1}^c)}$ By the definition in (B.104), $P(T_{t,1}^c) = P(|\bar{U}_x - \mathrm{E}(\bar{U}_x)| > t)$. Moreover, by the definition in (B.103), $\mathrm{E}(\bar{U}_x) = \Theta(1)$ and $\bar{U}_x \geq 0$.

Therefore we know there exist large positive constants C and t such that $\{|\bar{U}_x - \mathrm{E}(\bar{U}_x)| > t\} \subseteq \{\bar{U}_x > Ct\}$ and $P(T_{t,1}^c) \leq P(\bar{U}_x > Ct)$. Since $\bar{U}_x \leq (\sum_{i=1}^n x_{i,1}^2/n)^a$ and $x_{i,1}^2$ are i.i.d. sub-exponential random variables, we have

$$P(T_{t,1}^c) \leq P\Big\{\Big(\sum_{i=1}^n x_{i1}^2/n\Big)^a \geq Ct\Big\} = P\Big(\sum_{i=1}^n x_{i1}^2/n \geq Ct^{1/a}\Big) \leq C\exp(-Cn), \tag{B.114}$$

where the last inequality is obtained by the Bernstein-type inequality of sub-exponential random variables.

By the analysis above, (B.104) $\leq \mathrm{E}(P_{z,1}) + \mathrm{E}(P_{z,2}) + P(T_1^c)$. Recall that $h_p = C(p/\log^2 p)^{a/(a+1)}$, $\log p = o(n^{1/7})$, $\Gamma_p = \Theta(\log^{-1/2} p)$, $D = O(\log^{1/5} p)$ and $B_p = Cn^{-1/6}h_p^{2/(3a)}$. Then combining (B.107), (B.113) and (B.114), we have (B.104) $= o(D^{-1}d^{-1}p^{-2d})$. Therefore Lemma B.7 is proved. □

B.2.5. *Proof of Lemma A.8 (on Page 43, Section A.3).* Similarly to Section B.2.1, we first prove Lemma A.8 for $m = 1$ in Section B.2.5 and then for $m > 1$ in Section B.2.5.

Proof for $m = 1$. Specifically, in this section, we prove for finite integer a,

$$\left|P\Big(\frac{M_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \leq z\Big) - P\Big(\frac{M_n}{n} > y_p\Big)P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \leq z\Big)\right| \to 0. \tag{B.115}$$

To prove (B.115), we start by proving the following two conclusions (B.116) and (B.117), which suggest that M_n and $\hat{M}_n$ have small difference in probability. To be specific, as $n, p \to \infty$,

$$|P(M_n/n > y_p) - P(\hat{M}_n/n > y_p)| \to 0, \tag{B.116}$$

and

$$\begin{aligned}|P(M_n/n > y_p\,, \tilde{\mathcal{U}}(a)/\sigma(a) \leq z) \\ - P(\hat{M}_n/n > y_p\,, \tilde{\mathcal{U}}(a)/\sigma(a) \leq z)| \to 0.\end{aligned} \tag{B.117}$$

To prove (B.116) and (B.117), recall that in (A.9), M_n and $\hat{M}_n$ are defined using $\tilde{G}_l$ and $\hat{G}_l$ respectively. We next focus on the difference between $\tilde{G}_l$ and $\hat{G}_l$. Since $\tilde{G}_l$ and $\hat{G}_l$ will not change if the data $x_{i,j}$ is scaled by its standard deviation, then we assume, without loss of generality, $\sigma_{j,j} = 1$, $j = 1, \ldots, p$ in the following discussion.

By the definitions in (A.9), we have

$$P\Big(\max_{1\le l\le q}|\tilde G_l-\hat G_l|\ge(\log p)^{-1}\Big)\le P\Big(\max_{1\le l\le q}\max_{1\le i\le n}|x_{i,j_l^1}x_{i,j_l^2}|\ge\tau_n\Big).$$

Note that $|x_{i,j_l^1}x_{i,j_l^2}|\le(x_{i,j_l^1}^2+x_{i,j_l^2}^2)/2$. Then

$$\begin{aligned}
&P\Big(\max_{1\le l\le q}\max_{1\le i\le n}|x_{i,j_l^1}x_{i,j_l^2}|\ge\tau_n\Big)\\
&\le P\Big(\max_{1\le l\le q}\max_{1\le i\le n}(x_{i,j_l^1}^2+x_{i,j_l^2}^2)\ge 2\tau_n\Big)\\
&\le P\Big(\max_{1\le l\le q}\max_{1\le i\le n}x_{i,j_l^1}^2\ge\tau_n\Big)+P\Big(\max_{1\le l\le q}\max_{1\le i\le n}x_{i,j_l^2}^2\ge\tau_n\Big) \qquad \text{(B.118)}\\
&\le 2P\Big(\max_{1\le j\le p}\max_{1\le i\le n}x_{i,j}^2\ge\tau_n\Big) \qquad \text{(B.119)}\\
&\le 2np\max_{1\le j\le p}P(|x_{1,j}^2|\ge\tau_n).
\end{aligned}$$

From (B.118) to (B.119), we use $\max_{1\le l\le q}x_{i,j_l^k}^2=\max_{1\le j\le p}x_{i,j}^2$ for each i and $k=1,2$. To see this, recall the notation defined in Section A.3 (on Page 42). In particular, subscript l is defined to indicate a pair of indexes (j_l^1,j_l^2) with $1\le j_l^1<j_l^2\le p$. Since j_l^1 and j_l^1 only take values from the range $\{1,\dots,p\}$, we know $\{j_l^k:1\le l\le q\}\subseteq\{1,\dots,p\}$ for $k=1,2$, and then $\max_{1\le l\le q}x_{i,j_l^1}^2=\max_{1\le j\le p}x_{i,j}^2$. Moreover, by Condition 2.3 with $\varsigma=2$,

$$np\max_{1\le j\le p}P(|x_{1,j}^2|\ge\tau_n)\le Cnp(n+p)^{-\tau}\mathrm{E}\exp(x_{1,1}^2)\to 0.$$

It follows that $P(\max_{1\le l\le q}|\tilde G_l-\hat G_l|\ge(\log p)^{-1})\to 0$. Conditioning on $\max_{1\le l\le q}|\tilde G_l-\hat G_l|\le(\log p)^{-1}$, by Lemma B.3 and $|\hat G_l|\le\tau_n$,

$$\begin{aligned}
|M_n-\hat M_n| &= \Big|\max_{1\le l\le q}(\tilde G_l)^2-\max_{1\le l\le q}(\hat G_l)^2\Big|\\
&\le 2\max_{1\le l\le q}|\hat G_l|\max_{1\le l\le q}|\tilde G_l-\hat G_l|+\max_{1\le l\le q}|\tilde G_l-\hat G_l|^2\\
&\le 2\tau_n/\log p+(\log p)^{-2}.
\end{aligned}$$

Recall that $\tau_n=O(\log(p+n))$, then $|M_n/n-\hat M_n/n|\xrightarrow{P}0$. Therefore (B.116) and (B.117) are obtained.

Given (B.116) and (B.117), we next prove (B.115). In particular, we write

$$P\Big(\frac{M_n}{n}>y_p,\frac{\tilde{\mathcal U}(a)}{\sigma(a)}\le z\Big)-P\Big(\frac{M_n}{n}>y_p\Big)P\Big(\frac{\tilde{\mathcal U}(a)}{\sigma(a)}\le z\Big)=\Delta_{P,1}+\Delta_{P,2}+\Delta_{P,3},$$

where we define

$$\Delta_{P,1} = P\Big(M_n/n > y_p, \tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big) - P\Big(\hat{M}_n/n > y_p, \tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big),$$
$$\Delta_{P,2} = P\Big(\hat{M}_n/n > y_p, \tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big) - P\Big(\hat{M}_n/n > y_p) \times P(\tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big),$$
$$\Delta_{P,3} = P\Big(\hat{M}_n/n > y_p\Big) \times P\Big(\tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big) - P\Big(M_n/n > y_p\Big) \times P\Big(\tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big).$$

Note that the left hand side of (B.115) $\le |\Delta_{p,1}| + |\Delta_{p,2}| + |\Delta_{p,3}|$. By Lemma A.7, $|\Delta_{p,2}| \to 0$; by (B.117), $|\Delta_{p,1}| \to 0$; by $|\Delta_{p,3}| \le |P(\hat{M}_n/n > y_p) - P(M_n/n > y_p)|$ and (B.116), $|\Delta_{p,3}| \to 0$. In summary, (B.115) is proved.

Proof for $m > 1$. Following the proof in Section B.2.5, we know that (B.116) still holds and similarly to (B.117),

$$|P(M_n/n > y_p, \tilde{\mathcal{U}}(a_1)/\sigma(a_1) \le z_1, \dots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m) \le z_m) - P(\hat{M}_n/n > y_p, \tilde{\mathcal{U}}(a_1)/\sigma(a_1) \le z_1, \dots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m) \le z_m)| \to 0.$$

Given these results and Lemma A.7, we know that Lemma A.8 holds for $m > 1$, following the arguments in Section B.2.5 similarly.

B.2.6. *Proof of Lemma A.9 (on Page 43, Section A.3).* Similarly to Section B.2.5, we first prove Lemma A.9 for $m = 1$ in Section B.2.6, and then discuss the case for $m > 1$ in Section B.2.6.

Proof for $m = 1$. Specifically, in this section, we prove for finite integer a and given z,

$$\Big|P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} \le z, n\mathcal{U}^2(\infty) > y_p\Big) - P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} \le z\Big)P\Big(n\mathcal{U}^2(\infty) > y_p\Big)\Big| \to 0. \tag{B.120}$$

To prove this, we use M_n/n as an intermediate variable and first show

$$\Big|P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z, \frac{M_n}{n} > y_p\Big) - P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z\Big)P\Big(\frac{M_n}{n} > y_p\Big)\Big| \to 0. \tag{B.121}$$

To facilitate the proof, we define some notation. Given small constant $\epsilon > 0$,

$$
\begin{aligned}
&P_{uz} = P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z\Big), \quad P_{zy} = P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z, \frac{M_n}{n} > y_p\Big),\\
&P_{uz+\epsilon} = P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z+\epsilon\Big), \quad P_{z+\epsilon} = P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z+\epsilon, \frac{M_n}{n} > y_p\Big),\\
&P_{uz-\epsilon} = P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z-\epsilon\Big), \quad P_{z-\epsilon} = P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z-\epsilon, \frac{M_n}{n} > y_p\Big),\\
&P_{y_p} = P\Big(\frac{M_n}{n} > y_p\Big),
\end{aligned}
$$

$\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution, and $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$. Then

$$
\begin{aligned}
(\text{B.121}) &= |P_{zy} - P_{uz} \times P_{y_p}|\\
&\le |P_{zy} - P_{z+\epsilon}| + |P_{z+\epsilon} - P_{uz+\epsilon}P_{y_p}| + |P_{uz+\epsilon}P_{y_p} - P_{uz}P_{y_p}|.
\end{aligned}
$$

We next show (B.121) $\to 0$ by proving the three parts above all converges to 0 respectively.

First we show $|P_{zy} - P_{z+\epsilon}| \to 0$. Note that $P_{z+\epsilon} \le P_{zy} \le P_{z-\epsilon}$, then $|P_{zy} - P_{z+\epsilon}| \le |P_{z-\epsilon} - P_{z+\epsilon}|$. In addition,

$$
\begin{aligned}
&|P_{z-\epsilon} - P_{z+\epsilon}|\\
\le\ & |P_{z-\epsilon} - P_{uz-\epsilon} \times P_{y_p}| + |P_{uz-\epsilon} \times P_{y_p} - P_{uz+\epsilon} \times P_{y_p}| + |P_{uz+\epsilon} \times P_{y_p} - P_{z+\epsilon}|\\
\le\ & o(1) + |P_{uz+\epsilon} - P_{uz-\epsilon}|,
\end{aligned}
$$

where we use (B.115) in the last inequality. Moreover, by the proof of Theorem 2.1 in Section A.2, we know $\tilde{\mathcal{U}}(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$. Thus when $n, p \to \infty$ and $\epsilon \to 0$,

$$
\begin{aligned}
&|P_{uz+\epsilon} - P_{uz-\epsilon}|\\
\le\ & |P_{uz+\epsilon} - \bar{\Phi}(z+\epsilon)| + |\bar{\Phi}(z+\epsilon) - \bar{\Phi}(z-\epsilon)| + |P_{uz-\epsilon} - \bar{\Phi}(z-\epsilon)| + o(1)\\
\to\ & 0.
\end{aligned}
$$

Second, we know $|P_{z+\epsilon} - P_{uz+\epsilon}P_{y_p}| \to 0$ by (B.115). Last, we show $|P_{uz+\epsilon}P_{y_p} - P_{uz}P_{y_p}| \to 0$. By the proof of Theorem 2.1 in Section A.2, we know $\tilde{\mathcal{U}}(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$, $\{\mathcal{U}(a) - \tilde{\mathcal{U}}(a)/\sigma(a)\} \xrightarrow{P} 0$, and $\mathcal{U}(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$. Thus when $n, p \to \infty$ and $\epsilon \to 0$,

$$
\begin{aligned}
&|P_{uz+\epsilon}P_{y_p} - P_{uz}P_{y_p}|\\
\le\ & |P_{uz+\epsilon} - P_{uz}|\\
\le\ & |P_{uz+\epsilon} - \bar{\Phi}(z+\epsilon)| + |\bar{\Phi}(z+\epsilon) - \bar{\Phi}(z)| + |P_{uz} - \bar{\Phi}(z)| + o(1)\\
\to\ & 0.
\end{aligned}
$$

In summary (B.121) is proved.

We next prove (B.120) similarly to the proof of (B.121). Specifically, we write

$$\begin{aligned}&\left|P\Big(n\mathcal{U}^2(\infty) > y_p, \frac{\mathcal{U}(a)}{\sigma(a)} \le z\Big) - P\Big(n\mathcal{U}^2(\infty) > y_p\Big)P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} \le z\Big)\right| \\ &= |P_{z0} - P_{y0} \times P_{uz}|,\end{aligned}$$

where we define $P_{z0} = P(n\mathcal{U}^2(\infty) > y_p, \frac{\mathcal{U}(a)}{\sigma(a)} > z)$ and $P_{y0} = P(n\mathcal{U}^2(\infty) > y_p)$. Note that

$$|P_{z0} - P_{y0}P_{uz}| \le |P_{z0} - P_{zy-\epsilon}| + |P_{zy-\epsilon} - P_{y-\epsilon}P_{uz}| + |P_{y-\epsilon}P_{uz} - P_{y0}P_{uz}|,$$

where

$$\begin{aligned}P_{zy-\epsilon} &= P\Big(\frac{M_n}{n} > y_p - \epsilon, \frac{\mathcal{U}(a)}{\sigma(a)} > z\Big), \quad P_{y-\epsilon} = P\Big(\frac{M_n}{n} > y_p - \epsilon\Big), \\ P_{zy+\epsilon} &= P\Big(\frac{M_n}{n} > y_p + \epsilon, \frac{\mathcal{U}(a)}{\sigma(a)} > z\Big), \quad P_{y+\epsilon} = P\Big(\frac{M_n}{n} > y_p + \epsilon\Big).\end{aligned}$$

To prove (B.120), we will show $|P_{z0}-P_{zy-\epsilon}|$, $|P_{zy-\epsilon}-P_{y-\epsilon}P_{uz}|$, and $|P_{y-\epsilon}P_{uz}-P_{y0}P_{uz}|$ all converge to 0 respectively.

First we show $|P_{z0} - P_{zy-\epsilon}| \to 0$. Note that $W_n \xrightarrow{P} 0$ where $W_n = (n^2\mathcal{U}^2(\infty) - M_n)/n$ by the proof of Theorem 3 in [9]. Then for any $\epsilon > 0$, $P(|W_n| > \epsilon) \to 0$. Since $P_{zy+\epsilon} - P(|W_n| > \epsilon) \le P_{z0} \le P_{zy-\epsilon} + P(|W_n| > \epsilon)$, we have $|P_{z0} - P_{zy-\epsilon}| \le |P_{zy-\epsilon} - P_{zy+\epsilon}| + o(1)$. Furthermore,

$$\begin{aligned}&|P_{zy-\epsilon} - P_{zy+\epsilon}| \\ &\le |P_{zy-\epsilon} - P_{y-\epsilon}P_{uz}| + |P_{y-\epsilon}P_{uz} - P_{y+\epsilon}P_{uz}| + |P_{y+\epsilon}P_{uz} - P_{zy+\epsilon}| \to 0,\end{aligned}$$

where the last equation follows from (B.121) and $|P_{y-\epsilon} - P_{y+\epsilon}| \to 0$ when $\epsilon \to 0$. Second we know $|P_{zy-\epsilon} - P_{y-\epsilon}P_{uz}| \to 0$ by (B.121). Last we show $|P_{y-\epsilon}P_{uz} - P_{y0}P_{uz}| \to 0$. In particular, as $P_{y+\epsilon} - P(|W_n| > \epsilon) \le P_{y0} \le P_{y-\epsilon} + P(|W_n| > \epsilon)$ and $P(|W_n| > \epsilon) \to 0$, we have

$$|P_{y-\epsilon}P_{uz} - P_{y0}P_{uz}| \le |P_{y-\epsilon} - P_{y0}| \le |P_{y-\epsilon} - P_{y+\epsilon}| + o(1) \to 0.$$

In summary, Lemma A.9 is proved.

Proof for $m > 1$. Note that $W_n = \{n^2\mathcal{U}^2(\infty) - M_n\}/n \xrightarrow{P} 0$ and $\tilde{\mathcal{U}}^*(a_r) = \mathcal{U}(a_r) - \tilde{\mathcal{U}}(a_r) \xrightarrow{P} 0$ for each $r = 1, \ldots, m$ as argued in Section A.3. Therefore when m is finite, the arguments above can be applied to prove Lemma A.9 for $m > 1$ similarly.

B.3. Lemmas for the proof of Theorem 2.4.

B.3.1. *Proof of Lemma A.10 (on Page 44, Section A.4).* We first prove $\mathbb{V}_{u,1}(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 1$, and it suffices to prove $\mathrm{var}\{\mathbb{V}_{u,1}(a)\}/\mathrm{E}^2\{\mathbb{V}_{u,1}(a)\} \to 0$. By the notation defined at the beginning of Section B, we have

$$
\begin{aligned}
&\mathrm{var}\{\mathbb{V}_{u,1}(a)\}\\
&= \mathrm{E}\{\mathbb{V}_{u,1}^2(a)\} - \mathrm{E}^2\{\mathbb{V}_{u,1}(a)\}\\
&= \frac{(2a!)^2}{(P_a^n)^4} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a);\\ 1\le j_1\neq j_2\le p,\\ 1\le j_3\neq j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^a x_{i_t,j_1}^2 x_{i_t,j_2}^2 x_{\tilde{i}_t,j_3}^2 x_{\tilde{i}_t,j_4}^2\Big) - \Big\{\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\mathrm{E}(x_{1,j_3}^2 x_{1,j_4}^2)\Big\}^a\Big].
\end{aligned}
$$

To evaluate $\mathrm{var}\{\mathbb{V}_{u,1}(a)\}$, we consider the summed term in $\mathrm{var}\{\mathbb{V}_{u,1}(a)\}$, that is,

$$
\mathrm{E}\Big(\prod_{t=1}^a x_{i_t,j_1}^2 x_{i_t,j_2}^2 x_{\tilde{i}_t,j_3}^2 x_{\tilde{i}_t,j_4}^2\Big) - \{\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\}^a\{\mathrm{E}(x_{1,j_3}^2 x_{1,j_4}^2)\}^a. \tag{B.122}
$$

When $\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\} = \emptyset$, (B.122) $= 0$. We then know that (B.122) $\neq 0$ only when $|\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}| \le 2a-1$. Along with Condition 2.1, we have

$$
|\mathrm{var}\{\mathbb{V}_{u,1}(a)\}| \le Cp^4 n^{-4a} n^{2a-1},
$$

which induces $\mathrm{var}\{\mathbb{V}_{u,1}(a)\} = O(p^4 n^{-2a-1})$. By (B.24) and (B.29), we know $\mathrm{E}\{\mathbb{V}_{u,1}(a)\} = \Theta(p^2 n^{-a})$. It follows that $\mathrm{var}\{\mathbb{V}_{c,1}(a)\}/\mathrm{E}^2\{\mathbb{V}_{c,1}(a)\} \to 0$ as $n\to\infty$.

We next prove $\mathbb{V}_{u,2}(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 0$. By the Markov's inequality, it suffices to prove $\mathrm{E}\{\mathbb{V}_{u,2}^2(a)\} = o(1)[\mathrm{E}\{\mathbb{V}_{u,1}(a)\}]^2$. As $\mathrm{E}\{\mathbb{V}_{u,1}(a)\} = \Theta(p^2 n^{-a})$, it is sufficient to prove $\mathrm{E}\{\mathbb{V}_{u,2}^2(a)\} = o(p^4 n^{-2a})$ below.

We first derive the form of $\mathbb{V}_{u,2}(a)$. In particular, when $a = 1$,

$$
\begin{aligned}
\mathbb{V}_{u,2}(1) &= \mathbb{V}_u(1) - \mathbb{V}_{u,1}(1)\\
&= \frac{1}{n^2}\sum_{1\le j_1\neq j_2\le p}\sum_{i\in\mathcal{P}(n,1)}\Big\{(x_{i,j_1}-\bar{x}_{j_1})^2(x_{i,j_2}-\bar{x}_{j_2})^2 - x_{i,j_1}^2 x_{i,j_2}^2\Big\}\\
&= \frac{1}{n^2}\sum_{1\le j_1\neq j_2\le p}\sum_{1\le i\le n}\sum_{\substack{s_1+r_1=1,\\ s_2+r_2=1}} C_{s_1,r_1,s_2,r_2}\prod_{k=1}^2\Big\{(-x_{i,j_k}\bar{x}_{j_k})^{s_k}(\bar{x}_{j_k}^2)^{r_k}\Big\},
\end{aligned}
$$

where C_{s_1,r_1,s_2,r_2} is some constant and we use

$$\begin{aligned} & (x_{i,j_1}-\bar{x}_{i,j_1})^2(x_{i,j_2}-\bar{x}_{i,j_2})^2 - x_{i,j_1}^2 x_{i,j_2}^2 \\ = \; & (x_{i,j_1}^2 - 2x_{i,j_1}\bar{x}_{j_1} + \bar{x}_{j_1}^2)(x_{i,j_2}^2 - 2x_{i,j_2}\bar{x}_{j_2} + \bar{x}_{j_2}^2) - x_{i,j_1}^2 x_{i,j_2}^2 \\ = \; & \sum_{s_1+r_1=1,\, s_2+r_2=1} \Big\{ (-2x_{i,j_1}\bar{x}_{j_1})^{s_1} (\bar{x}_{j_1}^2)^{r_1} \Big\} \times \Big\{ (-2x_{i,j_2}\bar{x}_{j_2})^{s_2} (\bar{x}_{j_2}^2)^{r_2} \Big\}. \end{aligned}$$

Following this example, we similarly give the form of $\mathbb{V}_{u,2}(a)$ for general $a \geq 1$. Given tuple $\mathbf{i} \in \mathcal{P}(n,a)$, for $k = 1,2$, let $\mathbf{i}_{(a-r_k)}^{(k)}$ represent a sub-tuple of $\mathbf{i}$ with length $a - r_k$, and define $\mathcal{S}(\mathbf{i}, a - r_k)$ to be the collection of sub-tuples of $\mathbf{i}$ with length $a - r_k$. Then for $a \geq 1$, we write $\mathbb{V}_{u,2}(a) = \sum_{1\leq s_1+r_1\leq a, 1\leq s_2+r_2 \leq a} T_{s_1,r_1,s_2,r_2}$, where

$$\begin{aligned} T_{s_1,r_1,s_2,r_2} \;=\; & \frac{a!}{(P_a^n)^2} \sum_{1\leq j_1\neq j_2\leq p} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a); \\ \mathbf{i}_{(a-r_k)}^{(k)} \in \mathcal{S}(\mathbf{i},a-r_k):\, k=1,2}} C_{s_1,r_1,s_2,r_2} \\ & \times \prod_{k=1}^{2} \Big\{ (-\bar{x}_{j_k})^{s_k+2r_k} \prod_{t_k=1}^{s_k} x_{i_{t_k}^{(k)},j_k} \prod_{t_k=s_k+1}^{a-r_k} (x_{i_{t_k}^{(k)},j_k})^2 \Big\}. \end{aligned}$$

When a is finite, it suffices to prove $\mathrm{E}(T_{s_1,r_1,s_2,r_2}^2) = o(p^4 n^{-2a})$. Note that

$$\begin{aligned} & \mathrm{E}(T_{s_1,r_1,s_2,r_2}^2) \\ = \; & \frac{(a!)^2}{(P_a^n)^4} \sum_{\substack{1\leq j_1\neq j_2\leq p \\ 1\leq \tilde{j}_1\neq \tilde{j}_2\leq p}} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a); \\ \mathbf{i}_{(a-r_k)}^{(k)} \in \mathcal{S}(\mathbf{i},a-r_k):\, k=1,2; \\ \tilde{\mathbf{i}}_{(a-r_k)}^{(k)} \in \mathcal{S}(\tilde{\mathbf{i}},a-r_k):\, k=1,2}} C_{s_1,r_1,s_2,r_2}^2 \\ & \times \mathrm{E}\Big\{ \prod_{k=1}^{2} (\bar{x}_{j_k}\bar{x}_{\tilde{j}_k})^{s_k+2r_k} \prod_{t_k=1}^{s_k} (x_{i_{t_k}^{(k)},j_k} x_{\tilde{i}_{t_k}^{(k)},\tilde{j}_k}) \prod_{t_k=s_k+1}^{a-r_k} (x_{i_{t_k}^{(k)},j_k} x_{\tilde{i}_{t_k}^{(k)},\tilde{j}_k})^2 \Big\}. \end{aligned}$$

Recall that $\bar{x}_j = \sum_{i=1}^n x_{i,j}/n$. We have

$$\begin{aligned} & \mathrm{E}(T_{s_1,r_1,s_2,r_2}^2) \\ = \; & \frac{(a!)^2}{(P_a^n)^4 n^{\sum_{k=1}^{2}(2s_k+4r_k)}} \sum_{\substack{1\leq j_1\neq j_2\leq p \\ 1\leq \tilde{j}_1\neq \tilde{j}_2\leq p}} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a); \\ \mathbf{i}_{(a-r_k)}^{(k)} \in \mathcal{S}(\mathbf{i},a-r_k):\, k=1,2; \\ \tilde{\mathbf{i}}_{(a-r_k)}^{(k)} \in \mathcal{S}(\tilde{\mathbf{i}},a-r_k):\, k=1,2}} C_{s_1,r_1,s_2,r_2}^2 \\ & \times \sum_{\mathbf{m}^{(k)},\tilde{\mathbf{m}}^{(k)} \in \mathcal{C}(n,s_k+2r_k);\, k=1,2} T\{\mathbf{i}_{(a-r_k)}^{(k)}, \tilde{\mathbf{i}}_{(a-r_k)}^{(k)}, \mathbf{m}^{(k)}, \tilde{\mathbf{m}}^{(k)}; k=1,2\}, \end{aligned}$$

where $\mathcal{C}(n, s_k + 2r_k)$ follows the notation at the beginning of Section B and

$$
\begin{aligned}
& T\{\mathbf{i}^{(k)}_{(a-r_k)}, \tilde{\mathbf{i}}^{(k)}_{(a-r_k)}, \mathbf{m}^{(k)}, \tilde{\mathbf{m}}^{(k)}; k=1,2\} \\
= \ & \mathrm{E}\Big\{ \prod_{k=1}^{2} \prod_{\tilde{t}_k=1}^{s_k+2r_k} (x_{m_{\tilde{t}_k}, j_k} x_{\tilde{m}_{\tilde{t}_k}, \tilde{j}_k}) \prod_{t_k=1}^{s_k} (x_{i^{(k)}_{t_k}, j_k} x_{\tilde{i}^{(k)}_{t_k}, \tilde{j}_k}) \prod_{t_k=s_k+1}^{a-r_k} (x_{i^{(k)}_{t_k}, j_k} x_{\tilde{i}^{(k)}_{t_k}, \tilde{j}_k})^2 \Big\}.
\end{aligned}
$$

Since $\mathrm{E}(x_{i,j}) = 0$, $T\{\mathbf{i}^{(k)}_{(a-r_k)}, \tilde{\mathbf{i}}^{(k)}_{(a-r_k)}, \mathbf{m}^{(k)}, \tilde{\mathbf{m}}^{(k)}; k=1,2\} \neq 0$ only when

$$
\begin{aligned}
& \left| \bigcup_{k=1}^{2} \{\mathbf{m}^{(k)}\} \cup \{\tilde{\mathbf{m}}^{(k)}\} \cup \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\} \right| - \left| \bigcup_{k=1}^{2} \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\} \right| \\
\leq \ & \sum_{k=1}^{2} (s_k + 2r_k).
\end{aligned}
$$

Since $\mathbf{i}^{(k)}_{(a-r_k)}$ and $\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}$ are sub-tuples of $\mathbf{i}$ and $\tilde{\mathbf{i}} \in \mathcal{P}(n,a)$, $|\cup_{k=1}^{2} \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\}| \leq |\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a$. Therefore,

$$
\left| \bigcup_{k=1}^{2} \{\mathbf{m}^{(k)}\} \cup \{\tilde{\mathbf{m}}^{(k)}\} \cup \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\} \right| \leq 2a + \sum_{k=1}^{2} (s_k + 2r_k). \tag{B.123}
$$

By (B.123) and the boundedness of moments in Condition 2.4, we have

$$
\begin{aligned}
\mathrm{E}(T^2_{s_1,r_1,s_2,r_2}) &= O\Big(p^4 n^{-4a - \sum_{k=1}^{2}(2s_k+4r_k) + 2a + \sum_{k=1}^{2}(s_k+2r_k)} \Big) \\
&= O(p^4 n^{-2a - \sum_{k=1}^{2}(s_k+2r_k)}) = o(p^4 n^{-2a}),
\end{aligned}
$$

where we use $\sum_{k=1}^{2}(s_k + 2r_k) \geq 1$.

B.4. Lemmas for the proof of Theorem 2.5.

B.4.1. *Proof of Lemma A.11 (on Page 45, Section A.5).* To show $\mathrm{var}\{\mathcal{U}(a)\} \simeq \mathrm{var}(T_{U,a,1,1})$, it suffices to prove $\mathrm{var}(T_{U,a,1,1}) = \Theta(p^2 n^{-a})$, $\mathrm{var}(T_{U,a,1,2}) = o(p^2 n^{-a})$ and $\mathrm{var}(T_{U,a,2}) = o(p^2 n^{-a})$. The following three sections B.4.1–B.4.1 prove the three results respectively.

$\mathrm{var}(T_{U,a,1,1}) = \Theta(p^2 n^{-a})$. As $\mathrm{E}(T_{U,a,1,1}) = 0$, $\mathrm{var}(T_{U,a,1,1}) = \mathrm{E}(T^2_{U,a,1,1})$, and we have

$$
\mathrm{var}(T_{U,a,1,1}) = \sum_{(j_1,j_2),(j_3,j_4) \in J_A^c} (P^n_a)^{-2} \sum_{\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n,a)} \mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1} x_{i_k,j_2} x_{\tilde{i}_k,j_3} x_{\tilde{i}_k,j_4} \Big).
$$

Similarly to Section B.1.1, $\mathrm{E}(\prod_{k=1}^{a} x_{i_k,j_1}x_{i_k,j_2}x_{\tilde{i}_k,j_3}x_{\tilde{i}_k,j_4}) \neq 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$. Therefore,

$$\mathrm{var}(T_{U,a,1,1}) = \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} (P_a^n)^{-1} a! \times \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^a.$$

By Condition A.1, as $(j_1, j_2), (j_3, j_4) \in J_A^c$,

$$\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}). \tag{B.124}$$

We next evaluate (B.124) by discussing three cases on (j_1, j_2, j_3, j_4). First, if $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 2$, (B.124) $= \kappa_1\sigma_{j_1,j_1}\sigma_{j_2,j_2} = \Theta(1)$ by Condition 2.1.

$$\sum_{\substack{(j_1,j_2),\\(j_3,j_4)\in J_A^c}} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^a \times \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=2\}} = 2\sum_{(j_1,j_2)\in J_A^c} (\kappa_1\sigma_{j_1,j_1}\sigma_{j_2,j_2})^a.$$

Second, if $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 1$, we assume without loss of generality $j_1 = j_3$ and $j_2 \neq j_4$, (B.124) $= \kappa_1\sigma_{j_1,j_1}\sigma_{j_2,j_4}$, which is nonzero only when $(j_2, j_4) \in J_A$, and then (B.124) $= O(\rho^a)$. By the symmetricity of the indexes,

$$\begin{aligned}&\sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^a \times \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=1\}}\\ \leq\ & C\sum_{1\leq j\leq p;\,(j_2,j_4)\in J_A} \rho^a = O(1)p|J_A|\rho^a.\end{aligned}$$

Third, if $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 0$, we know $j_1 \neq j_2 \neq j_3 \neq j_4$, and (B.124) $\neq 0$ only if $(j_1, j_3), (j_2, j_4) \in J_A$ or $(j_1, j_4), (j_2, j_3) \in J_A$. Then (B.124) $= O(\rho^{2a})$. By the symmetricity of the indexes,

$$\begin{aligned}&\sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^a \times \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=0\}}\\ \leq\ & C\sum_{(j_1,j_3),(j_2,j_4)\in J_A} \rho^{2a} = O(1)|J_A|^2\rho^{2a}.\end{aligned}$$

In summary, we know

$$\begin{aligned}\mathrm{var}(T_{U,a,1,1}) &= 2a!\kappa_1^a(P_a^n)^{-1}\sum_{(j_1,j_2)\in J_A^c}\sigma_{j_1,j_1}^a\sigma_{j_2,j_2}^a\\ &\quad + O(1)p|J_A|\rho^a n^{-a} + O(1)|J_A|^2\rho^{2a}n^{-a}.\end{aligned}$$

Since we assume $|J_A|\rho^a = O(pn^{-a/2})$, $|J_A| = o(p^2)$ and $|J_A^c| = \Theta(p^2)$,

$$\mathrm{var}(T_{U,a,1,1}) \simeq 2a!\kappa_1^a (P_a^n)^{-1} \sum_{1\le j_1\neq j_2\le p} (\sigma_{j_1,j_1}\sigma_{j_2,j_2})^a,$$

which is of order $\Theta(p^2 n^{-a})$.

$\mathrm{var}(T_{U,a,1,2}) = o(p^2n^{-a})$. In this section, we prove $\mathrm{var}(T_{U,a,1,2}) = o(p^2n^{-a})$. As $T_{U,a,1,2} = \sum_{(j_1,j_2)\in J_A^c}\sum_{c=1}^a K(c,j_1,j_2)$, by the Cauchy-Schwarz inequality,

$$\mathrm{var}(T_{U,a,1,2}) \le C \times \sum_{c=1}^{a} \mathrm{var}\Big\{\sum_{(j_1,j_2)\in J_A^c} K(c,j_1,j_2)\Big\},$$

where C is some constant. As a is finite, to prove $\mathrm{var}(T_{U,a,1,2}) = o(p^2n^{-a})$, it suffices to prove $\mathrm{var}\{\sum_{(j_1,j_2)\in J_A^c} K(c,j_1,j_2)\} = o(p^2n^{-a})$, for each $1\le c\le a$. Note that $\mathrm{E}\{K(c,j_1,j_2)\} = 0$ and then

$$\begin{aligned}\mathrm{var}\Big\{\sum_{(j_1,j_2)\in J_A^c} K(c,j_1,j_2)\Big\} &= \mathrm{E}\Big[\Big\{\sum_{(j_1,j_2)\in J_A^c} K(c,j_1,j_2)\Big\}^2\Big]\\ &= F^2(a,c)\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\(j_1,j_2),(j_3,j_4)\in J_A^c}} Q_c(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4),\end{aligned}$$

where we define

$$\begin{aligned}Q_c(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4) = \mathrm{E}\Big[&\prod_{t=1}^{a-c} x_{i_t,j_1}x_{i_t,j_2}\prod_{t=a-c+1}^{a} x_{i_t,j_1}\prod_{t=a+1}^{a+c} x_{i_t,j_2}\\ &\times\prod_{\tilde{t}=1}^{a-c} x_{\tilde{i}_{\tilde{t}},j_3}x_{\tilde{i}_{\tilde{t}},j_4}\prod_{\tilde{t}=a-c+1}^{a} x_{\tilde{i}_{\tilde{t}},j_3}\prod_{\tilde{t}=a+1}^{a+c_2} x_{\tilde{i}_{\tilde{t}},j_4}\Big].\end{aligned}$$

As $F^2(a,c) = O(n^{-2(a+c)})$, to finish the proof, it remains to prove

$$\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\(j_1,j_2),(j_3,j_4)\in J_A^c}} Q_c(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4) = o(n^{2(a+c)-a}p^2). \tag{B.125}$$

We note that $\mathrm{E}(x_{1,j}) = 0$ and $\mathrm{E}(x_{1,j_1}x_{1,j_2}) = \mathrm{E}(x_{1,j_3}x_{1,j_4}) = 0$ for $(j_1,j_2),(j_3,j_4)\in J_A^c$. Similarly to Section B.4.1, $Q_c(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4) = 0$ if $\{\mathbf{i}\}\neq\{\tilde{\mathbf{i}}\}$, and

$$\sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)} \mathbf{1}_{\{Q_c(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\neq 0\}} = \sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)} \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}\}} = O(n^{a+c}). \tag{B.126}$$

To prove (B.125), it remains to prove for given $\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)$,

$$\Big| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \Big| = O(p^2). \tag{B.127}$$

We next prove (B.127) by discussing the value of $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$. To facilitate the discussion, for given $\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)$, we decompose the sets $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ into three disjoint sets respectively, defined as

$$\{\mathbf{i}\}_{(1)} = \{i_1, \ldots, i_{a-c}\},\ \{\mathbf{i}\}_{(2)} = \{i_{a-c+1}, \ldots, i_a\},\ \{\mathbf{i}\}_{(3)} = \{i_{a+1}, \ldots, i_{a+c}\},$$
$$\{\tilde{\mathbf{i}}\}_{(1)} = \{\tilde{i}_1, \ldots, \tilde{i}_{a-c}\},\ \{\tilde{\mathbf{i}}\}_{(2)} = \{\tilde{i}_{a-c+1}, \ldots, \tilde{i}_a\},\ \{\tilde{\mathbf{i}}\}_{(3)} = \{\tilde{i}_{a+1}, \ldots, \tilde{i}_{a+c}\},$$

which satisfy that $\{\mathbf{i}\} = \cup_{l=1}^3 \{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\} = \cup_{l=1}^3 \{\tilde{\mathbf{i}}\}_{(l)}$.

When $c \le a-1$, $\{\mathbf{i}\}_{(1)} \neq \emptyset$. We consider an index $i \in \{\mathbf{i}\}_{(1)}$, and discuss four different cases. First, if $i \notin \{\tilde{\mathbf{i}}\}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_2}) \mathrm{E}(\text{other terms}) = 0,$$

where the last equation follows from $\mathrm{E}(x_{i,j_1} x_{i,j_2}) = 0$ when $(j_1, j_2) \in J_A$. Second, if $i \in \{\tilde{\mathbf{i}}\}_{(2)}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_3}) \mathrm{E}(\text{other terms}) = 0$$

where the last equation is obtained by Condition A.1. Third, if $i \in \{\tilde{\mathbf{i}}\}_{(3)}$, similarly by Condition A.1, we also know

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_4}) \mathrm{E}(\text{other terms}) = 0. \tag{B.128}$$

Fourth, if $i \in \{\tilde{\mathbf{i}}\}_{(1)}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_3} x_{i,j_4}) \mathrm{E}(\text{other terms}). \tag{B.129}$$

Under Condition A.1, as $\mathrm{E}(x_{i,j_1} x_{i,j_2}) = \mathrm{E}(x_{i,j_3} x_{i,j_4}) = 0$ when (j_1, j_2) and $(j_3, j_4) \in J_A^c$,

$$\mathrm{E}\Big(\prod_{t=1}^4 x_{1,j_t} \Big) = \kappa_1 \Big\{ \mathrm{E}(x_{i,j_1} x_{i,j_3}) \mathrm{E}(x_{i,j_2} x_{i,j_4}) + \mathrm{E}(x_{i,j_1} x_{i,j_4}) \mathrm{E}(x_{i,j_2} x_{i,j_3}) \Big\}.$$

In addition, when $c = a$, $\{\mathbf{i}\}_{(1)} = \emptyset$ but $\{\mathbf{i}\}_{(1)}$ and $\{\mathbf{i}\}_{(3)} \neq \emptyset$. We next consider an index $i \in \{\mathbf{i}\}_{(1)}$ without loss of generality. Following similar analysis, we know $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$ when $i \notin \{\tilde{\mathbf{i}}\}$.

By symmetrically analyzing the indexes in $\mathbf{i}$ and $\tilde{\mathbf{i}}$ similarly as above, we know that $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$ only when $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$ and $\{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. When $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$, suppose $r = |\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(2)}|$ then $|\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = c-r$, $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(2)}| = c-r$, and $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = r$. It follows that

$$
\begin{aligned}
\text{(B.130)} \qquad & Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \\
&= \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^{a-c} \{\mathrm{E}(x_{1,j_1}x_{1,j_3})\mathrm{E}(x_{1,j_2}x_{1,j_4})\}^r \\
&\quad \times \{\mathrm{E}(x_{1,j_1}x_{1,j_4})\mathrm{E}(x_{1,j_2}x_{1,j_3})\}^{c-r} \\
&= \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^{a-c} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^r (\sigma_{j_1,j_4}\sigma_{j_2,j_3})^{c-r}.
\end{aligned}
$$

To prove (B.127), we next examine the value of (B.130) with respect to three different cases of (j_1, j_2, j_3, j_4).

Case (1) If $|\{j_1, j_2\}| \cap |\{j_3, j_4\}| = 2$, it means that $\{j_1, j_2\} = \{j_3, j_4\}$. Assume, without loss of generality, that $j_1 = j_3$ and $j_2 = j_4$. Then (B.130) $= O(1)(\sigma_{j_1,j_1}\sigma_{j_2,j_2})^{a-c+r}(\sigma^2_{j_1,j_2})^{c-r}$, which is nonzero only when $r = c$ as $\sigma_{j_1,j_2} = 0$. By the symmetricity of j indexes and the boundedness of moments in Condition 2.1,

$$
\Bigg| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{|\{j_1,j_2\}|\cap|\{j_3,j_4\}|=2\}} \Bigg| \leq Cp^2.
$$

Case (2) If $|\{j_1, j_2\}| \cap |\{j_3, j_4\}| = 1$, we assume without loss of generality that $j_1 = j_3$ but $j_2 \neq j_4$. Then (B.130) $= O(1)(\sigma_{j_1,j_1}\sigma_{j_2,j_4})^{a-c+r}(\sigma_{j_1,j_4}\sigma_{j_1,j_2})^{c-r}$, which is also nonzero only when $r = c$. By the symmetricity of j indexes and Condition 2.1, we have

$$
\begin{aligned}
&\Bigg| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{|\{j_1,j_2\}|\cap|\{j_3,j_4\}|=1\}} \Bigg| \\
\leq\ & C\Bigg| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} (\sigma_{j_1,j_1}\sigma_{j_2,j_4})^a \Bigg| \leq \sum_{1\leq j\leq p,\ (j_2,j_4)\in J_A} O(\rho^a) = O(p|J_A|\rho^a),
\end{aligned}
$$

where we use Condition 2.5 that $\sigma_{j_2,j_4} = \rho$ when $(j_2, j_4) \in J_A$ and $\sigma_{j_2,j_4} = 0$ when $(j_2, j_4) \notin J_A$.

Case (3) If $|\{j_1, j_2\}| \cap |\{j_3, j_4\}| = 0$, it means that $j_1 \neq j_2 \neq j_3 \neq j_4$. Then

$$(\text{B.130}) = O(1)(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3})^{a-c}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^r(\sigma_{j_1,j_4}\sigma_{j_2,j_3})^{c-r},$$

which nonzero only when $(j_1, j_3), (j_2, j_4) \in J_A^c$ or $(j_1, j_4), (j_2, j_3) \in J_A^c$. By the symmetricity of j indexes, Condition 2.1 and Condition 2.5,

$$\begin{aligned} &\left| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{|\{j_1,j_2\}|\cap|\{j_3,j_4\}|=0\}} \right| \\ \leq\ & C \sum_{(j_1,j_3),(j_2,j_4)\in J_A^c} \rho^{2a} = O(|J_A|^2\rho^{2a}). \end{aligned}$$

In summary,

$$\left| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \right| = O(p^2 + p|J_A|\rho^a + |J_A|^2\rho^{2a}) = o(p^2),$$

as we assume $|J_A|\rho^a = O(pn^{-a/2})$.

$\mathrm{var}(T_{U,a,2}) = o(p^2n^{-a})$. Similarly to Section B.4.1, by the Cauchy-Schwarz inequality,

$$\mathrm{var}(T_{U,a,2}) \leq C \sum_{c=0}^{a} \mathrm{var}(T_{U,a,2,c}), \tag{B.131}$$

where $T_{U,a,2,c} = \sum_{(j_1,j_2)\in J_A} K(c, j_1, j_2)$. To prove $\mathrm{var}(T_{U,a,2}) = o(p^2n^{-a})$, it suffices to prove $\mathrm{var}(T_{U,a,2,c}) = o(p^2n^{-a})$ for $0 \leq c \leq a$. Following the notation in Section B.4.1, we have

$$\mathrm{E}(T^2_{U,a,2,c}) = F^2(a, c) \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\ (j_1,j_2),(j_3,j_4)\in J_A}} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4).$$

When $1 \leq c \leq a$, $\mathrm{E}(T_{U,a,2,c}) = 0$; when $c = 0$, $\mathrm{E}(T_{U,a,2,0}) = \sum_{(j_1,j_2)\in J_A} \sigma^a_{j_1,j_2}$. Then

$$\mathrm{var}(T_{U,a,2,c}) = F^2(a, c) \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\ (j_1,j_2),(j_3,j_4)\in J_A}} \tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4), \tag{B.132}$$

where we define $\tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$ when $1 \leq c \leq a$; and $\tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) - (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a$ when $c = 0$.

To prove $\mathrm{var}(T_{U,a,2,c}) = o(p^2 n^{-a})$ for $1 \le c \le a$, we next examine the value of $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$. For given $\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)$, we define $\{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\}_{(l)}$ for $l = 1, 2, 3$ same as in Section B.4.1. Consider an index $i \in \{\mathbf{i}\}_{(2)}$. If $i \notin \{\tilde{\mathbf{i}}\}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1})\mathrm{E}(\text{other terms}) = 0.$$

If $i \in \{\tilde{\mathbf{i}}\}_{(1)}$, by Condition A.1,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_3} x_{i,j_4})\mathrm{E}(\text{other terms}) = 0.$$

Similarly, for an index $i \in \{\mathbf{i}\}_{(3)}$, we have $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$ if $i \notin \{\tilde{\mathbf{i}}\}$ or $i \in \{\tilde{\mathbf{i}}\}_{(1)}$. Analyzing the indexes in $\{\tilde{\mathbf{i}}\}$ symmetrically, we know that $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$ only when $\{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. Suppose $|\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(2)}| = r$, then $|\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = c - r$, $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(2)}| = c - r$, and $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = r$. Moreover, we let $|\{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)}| = t_c$ then $0 \le t_c \le a - c$. It follows that

$$\begin{aligned}
&Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \\
&= \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{i,j_t}\Big)\Big\}^{t_c} \Big\{\mathrm{E}(x_{i,j_1}x_{i,j_2})\mathrm{E}(x_{i,j_3}x_{i,j_4})\Big\}^{a-c-t_c} \\
&\quad \times \{\mathrm{E}(x_{i,j_1}x_{i,j_3})\mathrm{E}(x_{i,j_2}x_{i,j_4})\}^{r} \{\mathrm{E}(x_{i,j_1}x_{i,j_4})\mathrm{E}(x_{i,j_2}x_{i,j_3})\}^{c-r}.
\end{aligned} \tag{B.133}$$

To examine (B.125), we next analyze (B.133) with respect to different c and t_c values, where $0 \le c \le a$, $0 \le r \le c$ and $0 \le t_c \le a - c$.

When $c = 0$ and $t_c = t_0 = 0$, it means that $\{\mathbf{i}\} = \{\mathbf{i}\}_{(1)}$, $\{\tilde{\mathbf{i}}\} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\} \cap \{\tilde{\mathbf{i}}\} = \emptyset$ and $Q_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a$. Then

$$\begin{aligned}
&\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{t_0=0\}} \\
&= \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \Big\{Q_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) - (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a\Big\}\mathbf{1}_{\{t_0=0\}} = 0.
\end{aligned} \tag{B.134}$$

In the following, it remains to consider the cases when $c \ge 1$ or $t_c \ge 1$ in (B.133), which are examined by discussing three cases (j_1, j_2, j_3, j_4) below.

Case (1) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 2$, we assume without loss of generality that $j_1 = j_3$ and $j_2 = j_4$. Then by Condition A.1, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2})$, and

$$(\text{B.133}) = \{\kappa_1(2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2})\}^{t_c} \sigma_{j_1,j_2}^{2(a-c-t_c)} (\sigma_{j_1,j_1}\sigma_{j_2,j_2})^r (\sigma_{j_1,j_2})^{2(c-r)}.$$

Case (1.1) For $c = 0$ and $1 \le t_c = t_0 \le a$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \le 2a - t_0$, and

$$\left| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{c=0, 1 \le t_0 \le a, |\{j_1,j_2\} \cap \{j_3,j_4\}|=2\}} \right| \tag{B.135}$$
$$\le C \sum_{t_0=1}^{a} n^{2a-t_0} \sum_{(j_1,j_2) \in J_A} |\sigma_{j_1,j_2}|^{2(a-t_0)} |2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2}|^{t_0} + |\sigma_{j_1,j_2}|^{2a}$$
$$= \sum_{t_0=1}^{a} O(1) n^{2a-t_0} |J_A| \times (\rho^{2a} + \rho^{2(a-t_0)}),$$

where we use Condition 2.5.

Case (1.2) For $1 \le c \le a$ and $0 \le t_c \le a - c$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \le 2a - t_c$, and for each c given,

$$\left| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{1 \le t_c \le a-c, |\{j_1,j_2\} \cap \{j_3,j_4\}|=2\}} \right| \tag{B.136}$$
$$\begin{aligned} &\le C \sum_{\substack{0 \le r \le c; \\ 0 \le t_c \le a-c}} n^{2a-t_c} \sum_{(j_1,j_2) \in J_A} |\sigma_{j_1,j_2}|^{2(a-c-t_c)} \\ &\qquad \times |2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2}|^{t_c} |\sigma_{j_1,j_1}\sigma_{j_2,j_2}|^r |\sigma_{j_1,j_2}|^{2(c-r)} \\ &= \sum_{\substack{0 \le r \le c; \\ 0 \le t_c \le a-c}} O(1) n^{2a-t_c} |J_A| \{\rho^{2(a-r)} + \rho^{2(a-t_c-r)}\}. \end{aligned}$$

Case (2) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 1$, we assume without loss of generality that $j_1 = j_3$ and $j_2 \ne j_4$. Then by Condition A.1, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(2\sigma_{j_1,j_2}\sigma_{j_1,j_4} + \sigma_{j_1,j_1}\sigma_{j_2,j_4})$.We then know

$$\begin{aligned} \text{(B.133)} &= \{\kappa_1(2\sigma_{j_1,j_2}\sigma_{j_1,j_4} + \sigma_{j_1,j_1}\sigma_{j_2,j_4})\}^{t_c} (\sigma_{j_1,j_2}\sigma_{j_1,j_4})^{a-c-t_c} \\ &\quad \times (\sigma_{j_1,j_1}\sigma_{j_2,j_4})^r (\sigma_{j_1,j_4}\sigma_{j_1,j_2})^{c-r}. \end{aligned}$$

Case (2.1) For $c = 0$ and $1 \le t_c = t_0 \le a$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \le 2a - t_0$, and (B.133) $\neq 0$ at least when $(j_1,j_2),(j_1,j_4)\in J_A$. Then

$$\text{(B.137)}\quad \Bigg|\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a);\\(j_1,j_2),(j_3,j_4)\in J_A}} \tilde{Q}_0(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{c=0,1\le t_0\le a,|\{j_1,j_2\}\cap\{j_3,j_4\}|=1\}}\Bigg|$$
$$\le C\sum_{t_0=1}^{a} n^{2a-t_0}\sum_{(j_1,j_2),(j_1,j_4)\in J_A}\Big(|\sigma_{j_1,j_2}\sigma_{j_1,j_4}|^a+|\sigma_{j_2,j_4}|^{t_0}|\sigma_{j_1,j_2}\sigma_{j_1,j_4}|^{a-t_0}\Big)$$
$$=\sum_{t_0=1}^{a}O(1)n^{2a-t_0}\max_{1\le j_1\le p}|J_{j_1}|\times|J_A|(\rho^{2a}+\rho^{2a-t_0}).$$

Case (2.2) For $c\ge 1$ and $0\le t_c\le a-c$, we have $|\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}|\le 2a-t_c$. (B.133) $\neq 0$ when $(j_1,j_2),(j_1,j_4)\in J_A$ or $(j_2,j_4)\in J_A$. For given c, the range of (B.133) is between $O(\rho^{2a-t_c-r})$ and $O(\rho^{2a-r})$.

$$\text{(B.138)}\quad \Bigg|\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a);\\(j_1,j_2),(j_3,j_4)\in J_A}} \tilde{Q}_c(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{0\le t_c\le a-c,|\{j_1,j_2\}\cap\{j_3,j_4\}|=1\}}\Bigg|$$
$$=\sum_{\substack{0\le r\le c;\\0\le t_c\le a-c}}O(1)n^{2a-t_c}\max_{1\le j_1\le p}|J_{j_1}|\times|J_A|(\rho^{2a-t_c-r}+\rho^{2a-r}).$$

Case (3) If $|\{j_1,j_2\}\cap\{j_3,j_4\}|=0$, we know $j_1\neq j_2\neq j_3\neq j_4$. Then by Condition A.1 and 2.5, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})=\kappa_1(\sigma_{j_1,j_2}\sigma_{j_3,j_4}+\sigma_{j_1,j_3}\sigma_{j_2,j_4}+\sigma_{j_1,j_4}\sigma_{j_2,j_3})=O(\rho^2)$. Therefore, (B.133) $=O(\rho^{2a})$.

Case (3.1) For $c=0$ and $1\le t_c=t_0\le a$, we have $|\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}|\le 2a-t_0$.

$$\text{(B.139)}\quad \Bigg|\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a);\\(j_1,j_2),(j_3,j_4)\in J_A}} \tilde{Q}_0(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{c=0,1\le t_0\le a,|\{j_1,j_2\}\cap\{j_3,j_4\}|=0\}}\Bigg|$$
$$\le C\sum_{t_0=1}^{a}\sum_{(j_1,j_2),(j_3,j_4)\in J_A}|\sigma_{j_1,j_2}\sigma_{j_3,j_4}|^a=\sum_{t_0=1}^{a}n^{2a-t_0}|J_A|^2O(\rho^{2a}).$$

Case (3.2) For $1\le c\le a$ and $0\le t_c\le a$, we have $|\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}|\le 2a-t_c$.

Then for given $c \geq 1$,

$$\left| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{0 \leq t_c \leq a-c, |\{j_1,j_2\} \cap \{j_3,j_4\}|=0\}} \right| \tag{B.140}$$

$$\leq C \sum_{\substack{0 \leq r \leq c; \\ 0 \leq t_c \leq a-c}} n^{2a-t_c} \sum_{(j_1,j_2),(j_3,j_4) \in J_A} |\sigma_{j_1,j_2}\sigma_{j_3,j_4}|^a$$

$$= \sum_{t_c=0}^{a-c} n^{2a-t_c} |J_A|^2 O(\rho^{2a}),$$

where we use the symmetricity of indexes.

Combining (B.135)–(B.140) above, and by (B.131) and (B.132) and $F(a,c) = O(n^{-(a+c)})$, we know

$$\begin{aligned} &\mathrm{var}(T_{1,a,2}) \\ = &\sum_{t_0=1}^{a} O(1) \frac{1}{n^{t_0}} |J_A| \times \{\rho^{2a} + \rho^{2(a-t_0)}\} \\ &+ \sum_{c=1}^{a} \sum_{t_c=0}^{a-c} \sum_{r=0}^{c} O(1) |J_A| \frac{1}{n^{2c+t_c}} \{\rho^{2(a-r)} + \rho^{2(a-t_c-r)}\} \\ &+ \sum_{t_0=1}^{a} O(1) \frac{1}{n^{t_0}} \max_{1 \leq j_1 \leq p} |J_{j_1}| \times |J_A| (\rho^{2a} + \rho^{2a-t_0}) \\ &+ \sum_{c=1}^{a} \sum_{t_c=0}^{a-c} \sum_{r=0}^{c} O(1) \frac{1}{n^{2c+t_c}} \max_{1 \leq j_1 \leq p} |J_{j_1}| |J_A| (\rho^{2a-t_c-r} + \rho^{2a-r}) \\ &+ \sum_{t_0=1}^{a} O(1) \frac{1}{n^{t_0}} |J_A|^2 \rho^{2a} + \sum_{c=1}^{a} \sum_{t_c=0}^{a-c} O(1) \frac{1}{n^{2c+t_c}} |J_A|^2 \rho^{2a}. \end{aligned} \tag{B.141}$$

We then examine the six summed terms in the right hand side of (B.141) and show that they are $o(p^2 n^{-a})$ respectively.

(1) For the first term in (B.141), as $|J_A|\rho^a = O(pn^{-a/2})$,

$$n^{-t_0}|J_A|\rho^{2a} = n^{-t_0}|J_A|^{-1}|J_A|^2\rho^{2a} = o(p^2 n^{-a}),$$

and

$$\begin{aligned} n^{-t_0}|J_A|\rho^{2(a-t_0)} =& n^{-t_0}|J_A|^{1-2(a-t_0)/a}(|J_A|\rho^a)^{2(a-t_0)/a} \\ =& O(1) n^{-t_0}|J_A|^{-1+2t_0/a}(pn^{-a/2})^{2(a-t_0)/a} \\ =& O(1) p^2 n^{-a}|J_A|^{-1+t_0/a}(|J_A|/p^2)^{t_0/a} = o(p^2 n^{-a}), \end{aligned}$$

where we use $1 \le t_0 \le a$ and $|J_A| = o(p^2)$ in the last equation.

(2) For the second term in (B.141), as $r \le c \le a$ and $|J_A| = o(p^2)$,

$$\begin{aligned} n^{-(2c+t_c)}|J_A|\rho^{2(a-r)} =& n^{-(2c+t_c)}|J_A|^{1-2(a-r)/a}(|J_A|\rho^a)^{2(a-r)/a} \\ =& O(1)p^2 n^{-a+r-2c-t_c}|J_A|^{-1+r/a}(|J_A|/p^2)^{r/a} \\ =& o(p^2 n^{-a}), \end{aligned}$$

and similarly as $r \le c \le a$, $t_c + r \le a$ and $c \ge 1$,

$$\begin{aligned} & n^{-(2c+t_c)}|J_A|\rho^{2(a-t_c-r)} \\ =& O(1)p^2 n^{-a+t_c+r-2c-t_c}|J_A|^{-1+(t_c+r)/a}(|J_A|/p^2)^{(t_c+r)/a} \\ =& o(p^2 n^{-a}). \end{aligned}$$

(3) For the third term in (B.141), as $1 \le t_0 \le a$, and $|J_A|\rho^a = O(pn^{-a/2})$,

$$n^{-t_0} \max_{1\le j_1\le p} |J_{j_1}| \times |J_A|\rho^{2a} = \frac{\max_{1\le j_1\le p}|J_{j_1}|}{n^{t_0}|J_A|}|J_A|^2\rho^{2a} = o(p^2n^{-a}),$$

and

$$\begin{aligned} & n^{-t_0} \max_{1\le j_1\le p} |J_{j_1}| \times |J_A|\rho^{2a-t_0} \\ =& \ n^{-t_0} \max_{1\le j_1\le p} |J_{j_1}| \times |J_A|^{1-(2a-t_0)/a}(|J_A|\rho^a)^{(2a-t_0)/a} \\ =& \ O(1)p^2 n^{-a-t_0/2} \max_{1\le j_1\le p} |J_{j_1}| \times |J_A|^{-1+t_0/(a)}p^{-t_0/a} \\ =& \ O(1)\frac{p^2}{n^{a+t_0/2}}\frac{\max_{1\le j_1\le p}|J_{j_1}|}{|J_A|}\Big(\frac{|J_A|}{\max_{1\le j_1\le p}|J_{j_1}|}\frac{\max_{1\le j_1\le p}|J_{j_1}|}{p}\Big)^{t_0/a} \\ =& \ O(1)\frac{p^2}{n^{a+t_0/2}}\Big(\frac{\max_{1\le j_1\le p}|J_{j_1}|}{|J_A|}\Big)^{1-t_0/a}\Big(\frac{\max_{1\le j_1\le p}|J_{j_1}|}{p}\Big)^{t_0/a} = o(p^2n^{-a}), \end{aligned}$$

where in the last equation, we use $1 \le t_0 \le a$, $\max_{1\le j_1\le p}|J_{j_1}| \le |J_A|$ and $\max_{1\le j_1\le p}|J_{j_1}| \le p$.

(4) For the fourth term in (B.141),

$$\begin{aligned} & n^{-(2c+t_c)} \max_{1\le j_1\le p} |J_{j_1}||J_A|\rho^{2a-t_c-r} \\ =& \ n^{-(2c+t_c)} \max_{1\le j_1\le p} |J_{j_1}||J_A|^{1-(2a-t_c-r)/a}(|J_A|\rho^a)^{(2a-t_c-r)/a} \\ =& \ O(1)\frac{p^2}{n^a}\frac{1}{n^{2c+t_c/2-r/2}}\frac{\max_{1\le j_1\le p}|J_{j_1}|}{|J_A|}\Big(\frac{|J_A|}{p}\Big)^{(t_c+r)/a} \\ =& \ O(1)\frac{p^2}{n^a}\frac{1}{n^{2c+t_c/2-r/2}}\Big(\frac{\max_{1\le j_1\le p}|J_{j_1}|}{|J_A|}\Big)^{1-(t_c+r)/a}\Big(\frac{\max_{1\le j_1\le p}|J_{j_1}|}{p}\Big)^{(t_c+r)/a} \\ =& \ o(p^2n^{-a}), \end{aligned}$$

where we obtain the last equation by noting that $t_c+r \leq a$, $r \leq c$ and $c \geq 1$. Similarly, we have

$$
\begin{aligned}
& n^{-(2c+t_c)} \max_{1\leq j_1\leq p} |J_{j_1}||J_A|\rho^{2a-r} \\
=& O(1)\frac{p^2}{n^a}\frac{1}{n^{2c+t_c-r/2}}\Big(\frac{\max_{1\leq j_1\leq p}|J_{j_1}|}{|J_A|}\Big)^{1-r/a}\Big(\frac{\max_{1\leq j_1\leq p}|J_{j_1}|}{p}\Big)^{r/a} \\
=& o(p^2n^{-a}).
\end{aligned}
$$

(5) For the fifth and sixth terms in (B.141), as $|J_A|\rho^a = O(pn^{-a/2})$, $t_0 \geq 1$ and $c \geq 1$, we know

$$
\frac{1}{n^{t_0}}|J_A|^2\rho^{2a} = o(p^2n^{-a}), \text{ and } \frac{1}{n^{2c+t_c}}|J_A|^2\rho^{2a} = o(p^2n^{-a}).
$$

B.4.2. *Proof of Lemma A.12 (on Page 46, Section A.5).* The proof is similar to Section B.1.2. In particular, Lemma A.12 shows that $\mathrm{var}\{\mathcal{U}(a)\} \simeq \mathrm{var}(T_{U,a,1,1})$. By the Cauchy-schwarz inequality,

$$
\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \mathrm{E}\{T_{U,a,1,1}T_{U,b,1,1}\}/\{\sigma(a)\sigma(b)\} + o(1),
$$

where we use $\mathrm{E}(T_{U,a,1,1}) = \mathrm{E}(T_{U,b,1,1}) = 0$. For two integers $a \neq b$, we next prove $\mathrm{E}(T_{U,a,1,1}T_{U,b,1,1})$=0. Specifically,

$$
\begin{aligned}
& \mathrm{E}(T_{U,a,1,1}T_{U,b,1,1}) \\
= & (P^n_aP^n_b)^{-1} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a),\tilde{\mathbf{i}}\in\mathcal{P}(n,b);\\ (j_1,j_2),(j_3,j_4)\in J_A^c}} \mathrm{E}\Big(\prod_{k=1}^a x_{i_k,j_1}x_{i_k,j_2}\prod_{\tilde{k}=1}^b x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\Big).
\end{aligned}
$$

Since $a \neq b$, $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$. Assume without loss of generality that $a < b$ and index $i \in \{\tilde{\mathbf{i}}\}$ but $i \notin \{\mathbf{i}\}$. Then

$$
\mathrm{E}\Big(\prod_{k=1}^a x_{i_k,j_1}x_{i_k,j_2}\prod_{\tilde{k}=1}^b x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\Big) = \mathrm{E}(x_{1,j_3}x_{1,j_4}) \times \mathrm{E}(\text{other terms}) = 0,
$$

where we use the $\sigma_{j_1,j_2} = \sigma_{j_3,j_4} = 0$ for $(j_1,j_2),(j_3,j_4) \in J_A^c$. Therefore $\mathrm{cov}(T_{U,a,1,1}, T_{U,b,1,1}) = 0$ and the lemma is proved.

B.4.3. *Proof of Lemma A.13 (on Page 46, Section A.5).* We prove Lemma A.13 similarly as in Section B.1.5. By the Cauchy-Schwarz inequality, for some constant C,

$$\mathrm{var}\Big(\sum_{k=1}^{n}\pi_{n,k}^2\Big) \leq Cn^2 \max_{1\leq k\leq n;\, 1\leq r_1,r_2\leq m} \mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}),$$

where $c(n,a) = [a \times \{\sigma(a)P_a^n\}^{-1}]^2$ and for two finite integers a_1 and a_2, $\mathbb{T}_{k,a_1,a_2} = \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$. In particular, when $k < \max\{a_1,a_2\}$, $\mathbb{T}_{k,a_1,a_2} = 0$; when $k \geq \max\{a_1,a_2\}$,

$$\begin{aligned}\mathbb{T}_{k,a_1,a_2} &= \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})\\ &= \sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2;\\ (j_1,j_2),(j_3,j_4)\in J_A^c}} \Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}^{1/2} \mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2)\end{aligned}$$

with

$$\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2) \quad = \quad \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big)\prod_{l=1}^{2}\prod_{t=1}^{a_l-1}(x_{i_t^{(l)},j_{2l-1}}x_{i_t^{(l)},j_{2l}}).$$

To prove $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2) \to 0$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}) = o(n^{-2})$ for any $1 \leq r_1, r_2 \leq m$. Without loss of generality, we consider two finite integers a_1 and a_2, and prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$ when $\max\{a_1,a_2\} \leq k \leq n$.

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, we decompose $\mathbb{T}_{k,a_1,a_2} = \sum_{M=2}^{4}\mathbb{T}_{k,a_1,a_2,(M)}$, where

$$\begin{aligned}\mathbb{T}_{k,a_1,a_2,(M)} = &\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2;\\ (j_1,j_2),(j_3,j_4)\in J_A^c}} \mathbf{1}_{\{|\{j_1,j_2\}\cup\{j_3,j_4\}|=M\}}\\ &\times\Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}^{1/2}\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2).\end{aligned}$$

Here $2 \leq M \leq 4$ because $2 \leq |\{j_1,j_2\}\cup\{j_3,j_4\}| \leq 4$ when $(j_1,j_2),(j_3,j_4) \in J_A^c$. By the Cauchy-Schwarz inequality, to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,(M)}) = o(n^{-2})$ for $M = 2,3,4$. For easy presentation, we let $a_3 = a_1$ and $a_4 = a_2$, and then

$$\begin{aligned}\mathbb{T}_{k,a_1,a_2,(M)}^2 \quad = \quad &\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4;\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{\left\{\substack{|\{j_1,j_2\}\cup\{j_3,j_4\}|=M,\\ |\{j_5,j_6\}\cup\{j_7,j_8\}|=M}\right\}}\\ &\times\Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}\times\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2,3,4),\end{aligned}$$

where

$$\begin{aligned}&\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\\&=\mathrm{E}\Big(\prod_{t=1}^{4}x_{k,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8}x_{k,j_t}\Big)\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1}x_{i_t^{(l)},j_{2l-1}}x_{i_t^{(l)},j_{2l}}\Big).\end{aligned}$$

By $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}=\mathrm{E}\{\mathbb{T}^2_{k,a_1,a_2,(M)}\}-\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,(M)})\}^2$,

$$\begin{aligned}&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}\\&=\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2,3,4;\\(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}}\mathbf{1}_{\left\{\substack{|\{j_1,j_2\}\cup\{j_3,j_4\}|=M,\\|\{j_5,j_6\}\cup\{j_7,j_8\}|=M}\right\}}\Big\{\prod_{l=1}^{2}c(n,a_l)\Big\}\\&\quad\times\Big[\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\Big\}\\&\quad-\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\Big\}\times\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=3,4)\Big\}\Big],\end{aligned}$$

where we similarly define

$$\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=3,4)=\mathrm{E}\Big(\prod_{t=5}^{8}x_{k,j_t}\Big)\prod_{l=3}^{4}\prod_{t=1}^{a_l-1}(x_{i_t^{(l)},j_{2l-1}}x_{i_t^{(l)},j_{2l}}).$$

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,(M)})=o(n^{-2})$, we examine the value of

$$\begin{aligned}&\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\Big\}\\&-\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\Big\}\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=3,4)\Big\}.\end{aligned}\tag{B.142}$$

We next show that when (B.142) $\neq 0$, the following two claims hold:

$$\begin{aligned}&\textit{Claim 1:}\ \ (\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\})\neq\emptyset,\\&\textit{Claim 2:}\ |\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|\le a_1+a_2-2.\end{aligned}\tag{B.143}$$

Claim 1 can be straightforwardly seen from the definition (B.142). We then prove *Claim 2*. Note that $\mathrm{E}\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\}\neq 0$ only when $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|\le a_1+a_2-2$ following similar analysis to Section B.1.5. In addition, as $\sigma_{j_1,j_2}=\sigma_{j_3,j_4}=0$ when $(j_1,j_2),(j_3,j_4)\in J_A^c$, we know that $\mathrm{E}\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\}\neq 0$ only when $\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}$; as $\sigma_{j_5,j_6}=\sigma_{j_7,j_8}=0$, we similarly know that $\mathrm{E}\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=3,4)\}\neq 0$ only when $\{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}$. It follows that if $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|>a_1+a_2-2$, (B.142) $=0$.

Thus to evaluate $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}$, it remains to consider (B.142) under the cases when $(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\})\neq\emptyset$ and $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|\leq a_1+a_2-2$.

Given the two claims above, we examine $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}$ for $M=2,3,4$ respectively. To facilitate the discussion, we decompose $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\} = \mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} + \mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)}$, where

$$\begin{aligned}&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)}\\ &= \sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4;\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{\left\{\substack{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|=a_1+a_2-2;\\ |\{j_1,j_2\}\cup\{j_3,j_4\}|=M;\\ |\{j_5,j_6\}\cup\{j_7,j_8\}|=M}\right\}} \prod_{l=1}^2 c(n,a_l)\times(\text{B.142}),\end{aligned}$$

and

$$\begin{aligned}&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)}\\ &= \sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4;\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{\left\{\substack{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|<a_1+a_2-2;\\ |\{j_1,j_2\}\cup\{j_3,j_4\}|=M;\\ |\{j_5,j_6\}\cup\{j_7,j_8\}|=M}\right\}} \prod_{l=1}^2 c(n,a_l)\times(\text{B.142}).\end{aligned}$$

We next consider $M=2,3,4$ in the following **Cases (1)–(3)**, respectively. We assume without loss of generality that $a_1\leq a_2$ in the following.

Case (1): When $M=2$, by the definition of $\mathbb{T}_{k,a_1,a_2,(M)}$, we know $\{j_1,j_2\}=\{j_3,j_4\}$, $\{j_5,j_6\}=\{j_7,j_8\}$, and $|\{j_t: t=1,\dots,8\}|\leq 4$. It follows that $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)} = O\{\prod_{l=1}^2 c(n,a_l)p^4n^{a_1+a_2-3}\} = o(n^{-2})$ by the boundedness of moments in Condition 2.1 and the definition of $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)}$.

We next prove $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)}=o(n^{-2})$. Recall that we consider $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|=a_1+a_2-2$ here by the construction of $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)}$. Suppose $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(2)}\}|=s$, where $s\leq a_1-1$. Then symmetrically $|\{\mathbf{i}^{(3)}\}\cap\{\mathbf{i}^{(4)}\}|=s$. Further assume $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}|=s_1$, then $|\{\mathbf{i}^{(2)}\}\cap\{\mathbf{i}^{(3)}\}|=a_1-1-s-s_1$, $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(4)}\}|=a_1-1-s-s_1$ and $|\{\mathbf{i}^{(2)}\}\cap\{\mathbf{i}^{(4)}\}|=a_2-a_1+s_1$. It follows that $|(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\})|=a_1+a_2-2-2s$. Note that (B.142) $=0$ if $a_1+a_2-2-2s=0$, which can only be achieved when $a_1=a_2$ and $s=a_1-1$. It remains to consider $a_1+a_2-2-2s\geq 1$, that is,

$0 \le s \le A_0$, where $A_0 = (a_1 + a_2 - 3)/2$. Given s and s_1, we have

$$\text{(B.142)} = \Big\{\mathrm{E}\Big(\prod_{t=1,2,5,6} x_{1,j_t}\Big)\Big\}^{s_1}\Big\{\mathrm{E}\Big(\prod_{t=3,4,7,8} x_{1,j_t}\Big)\Big\}^{a_2-a_1+s_1} \times \Big\{\mathrm{E}\Big(\prod_{t=3,4,5,6} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=1,2,7,8} x_{1,j_t}\Big)\Big\}^{a_1-1-s-s_1} \times \Big\{\mathrm{E}\Big(\prod_{t=1,2,3,4} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5,6,7,8} x_{1,j_t}\Big)\Big\}^{s+1}. \tag{B.144}$$

Under the considered **Case (1)**, $\{j_1, j_2\} = \{j_3, j_4\}$ and $\{j_5, j_6\} = \{j_7, j_8\}$. If $|\{j_t : t = 1, \ldots, 8\}| \le 3$, we know by Condition 2.1,

$$\Bigg|\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \text{(B.142)} \times \mathbf{1}_{|\{j_t:t=1,\ldots,8\}|\le 3}\Bigg| = O(p^3). \tag{B.145}$$

If $|\{j_t : t = 1, \ldots, 8\}| = 4$, $\{j_1, j_2\} \cap \{j_5, j_6\} = \emptyset$. By Conditions 2.1, A.1 and 2.5, we know $\mathrm{E}(\prod_{t=1}^4 x_{1,j_t}) = \kappa_1(\sigma_{j_1,j_1} + \sigma_{j_2,j_2}) = O(1)$ and similarly $\mathrm{E}(\prod_{t=5}^8 x_{1,j_t}) = O(1)$. By (B.144), (B.142) $\neq 0$ only if $\mathrm{E}(\prod_{t=1,2,5,6} x_{1,j_t}) \neq 0$. This induces (j_1, j_5), $(j_2, j_6) \in J_A$ or (j_1, j_6), $(j_2, j_5) \in J_A$, and then (B.142) $= O(\rho^{2(a_1+a_2-2s)})$. By the symmetricity of j indexes, we have

$$\begin{aligned}&\Bigg|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} \text{(B.142)} \times \mathbf{1}_{\{|\{j_t:t=1,\ldots,8\}|=4\}}\Bigg| \\ &\quad \le C \sum_{(j_1,j_5),(j_2,j_6)\in J_A} \rho^{2(a_1+a_2-2-2s)} \le C|J_A|^2\rho^{2(a_1+a_2-2-2s)}.\end{aligned} \tag{B.146}$$

By (B.145) and (B.146),

$$\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} = \sum_{s=0}^{A_0} O\Big\{p^3 + |J_A|^2\rho^{2(a_1+a_2-2-2s)}\Big\} n^{a_1+a_2-2}\prod_{l=1}^{2} c(n, a_l).$$

Note that $O(p^3 n^{a_1+a_2-2})\prod_{l=1}^2 c(n, a_l) = o(n^{-2})$, and

$$\begin{aligned}&|J_A|^2\rho^{2(a_1+a_2-2-2s)} n^{a_1+a_2-2} c(n, a_1)c(n, a_2) \\ &\quad = O(1)p^{-4}n^{-2}|J_A|^{2-\frac{2(a_1+a_2-2-2s)}{a_1+a_2}}\big(|J_A|\rho^{a_1} \times |J_A|\rho^{a_2}\big)^{\frac{2(a_1+a_2-2-2s)}{a_1+a_2}} \\ &\quad = O(1)|J_A|^{2-\frac{2(a_1+a_2-2-2s)}{a_1+a_2}} p^{\frac{2(a_1+a_2-2-2s)}{a_1+a_2}-4} n^{-(a_1+a_2-2-2s)-2} \\ &\quad = O(1)|J_A|^{-\frac{a_1+a_2-2-2s}{a_1+a_2}}(|J_A|/p^2)^{2-\frac{a_1+a_2-2-2s}{a_1+a_2}} n^{-(a_1+a_2-2-2s)-2} \\ &\quad = o(n^{-2}).\end{aligned} \tag{B.147}$$

Therefore $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} = o(n^{-2})$.

Case (2): When $M = 3$, we assume without loss of generality that $j_1 = j_3$ and $j_5 = j_7$, then

$$\{j_1, j_2, j_3, j_4\} = \{j_1, j_2, j_4\} \quad \text{and} \quad \{j_5, j_6, j_7, j_8\} = \{j_5, j_6, j_8\}. \tag{B.148}$$

It follows that $\mathrm{E}(\prod_{t=1}^4 x_{1,j_t}) = \kappa_1\sigma_{j_1,j_1}\sigma_{j_2,j_4}$ and $\mathrm{E}(\prod_{t=5}^8 x_{1,j_t}) = \kappa_1\sigma_{j_5,j_5}\sigma_{j_6,j_8}$, which are 0 when (j_2, j_4) and $(j_6, j_8) \in J_A^c$; and are $O(\rho)$ when (j_2, j_4) and $(j_6, j_8) \in J_A$. This suggests that if (B.142) $\neq 0$, (j_2, j_4) and $(j_6, j_8) \in J_A$.

We first examine $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(1)}$, which is the part of summation in $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}$ when $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| = a_1 + a_2 - 2$. Recall that the two claims in (B.143) also hold here. Similarly to **Case (1)** above, we still assume $|\{\mathbf{i}^{(1)}\} \cap \{\mathbf{i}^{(2)}\}| = s$, and $|\{\mathbf{i}^{(1)}\} \cap \{\mathbf{i}^{(3)}\}| = s_1$, then (B.144) holds. We next discuss several sub-cases based on the size of the set $\{j_t : t = 1, \dots, 8\}$.

Case (2.1): When $|\{j_t : t = 1, \dots, 8\}| = 6$, we know $\{j_1, j_2, j_3, j_4\} \cap \{j_5, j_6, j_7, j_8\} = \emptyset$ by (B.148). Then by (B.144), we know if (B.142) $\neq 0$, then $(j_2, j_4), (j_6, j_8), (j_1, j_5), (j_2, j_6) \in J_A$ or $(j_2, j_4), (j_6, j_8), (j_1, j_6), (j_2, j_5) \in J_A$. Thus by the symmetricity of the j indexes, we have

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{(\text{B.142})\neq 0} \times \mathbf{1}_{|\{j_t:t=1,\dots,8\}|=6} \leq C \sum_{\substack{(j_1,j_5),(j_2,j_6),\\(j_2,j_4),(j_6,j_8)\in J_A}} 1 \leq C|J_A|^3.$$

By Conditions A.1 and 2.5, (B.142) $= O(\rho^{\tilde{A}_1})$, where $\tilde{A}_1 = 2(a_1 + a_2 - 2 - 2s) + 2(s+1) = 2(a_1 + a_2) - 2(s+1)$. Thus

$$\left| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.142})\mathbf{1}_{|\{j_t:t=1,\dots,8\}|=6} \right| = O(|J_A|^3 \rho^{\tilde{A}_1}).$$

Case (2.2): When $|\{j_t : t = 1, \dots, 8\}| = 5$, recall that we assume (B.148), where $j_1 = j_3$ and $j_5 = j_7$ without loss of generality. If we further assume $j_1 = j_5$, $\{j_t : t = 1, \dots, 8\} = \{j_1, j_2, j_4, j_6, j_8\}$. Then for (B.142) $\neq 0$, $\mathrm{E}(\prod_{t=1,2,3,4} x_{1,j_t}) \times \mathrm{E}(\prod_{t=5,6,7,8} x_{1,j_t}) \neq 0$, then $(j_2, j_4), (j_6, j_8) \in J_A$ holds. In addition, under this case, (B.142) $= O\{\rho^{(a_1+a_2-2-2s)+2(s+1)}\} = O(\rho^{a_1+a_2})$, and we have

$$\left| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} \mathbf{1}_{(\text{B.142})=O(\rho^{a_1+a_2}),\, |\{j_t:t=1,\dots,8\}|=5} \right| = O(p|J_A|^2).$$

If given $j_1 = j_3$ and $j_5 = j_7$, instead, assume $j_1 \neq j_5$. We have $j_1 \neq j_2$, $j_1 \neq j_4$ and $j_1 \neq j_5$. Then for (B.142) $\neq 0$, by discussing different cases of j indexes, we know that (B.142) achieves the order between $O(\rho^{\tilde{A}_1})$ and $O(\rho^{\tilde{A}_2})$ where $\tilde{A}_1$ is defined as above and $\tilde{A}_2 = 2(s+1) + (1+2) \times (a_1 + a_2 - 2s - 2)/2 = 3(a_1 + a_2)/2 - (s+1)$. Moreover, we have

$$\left| \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{\{(\text{B.142})=O(\rho^u),\, \tilde{A}_2 \le u \le \tilde{A}_1,\, |\{j_t:t=1,\dots,8\}|=5\}} \right| = O(D_{\max}|J_A|^2).$$

In summary,

$$\begin{aligned}
&\left| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.142}) \times \mathbf{1}_{|\{j_t:t=1,\dots,8\}|=5} \right| \\
&= O(D_{\max}|J_A|^2\rho^{\tilde{A}_1}) + O(D_{\max}|J_A|^2\rho^{\tilde{A}_2}) + O(p|J_A|^2\rho^{a_1+a_2}).
\end{aligned}$$

Case (2.3): When $|\{j_t : t = 1, \dots, 8\}| = 4$, similarly as case (2.3), we can discuss $j_1 = j_5$ and $j_1 \neq j_5$ respectively. When $j_1 = j_5$, we note that (B.142) can achieve the orders between $O(\rho^{a_1+a_2})$ and $O(\rho^{\tilde{A}_3})$ with $\tilde{A}_3 = (a_1 + a_2 - 2 - 2s)/2 + 2(s+1) = (a_1 + a_2)/2 + s + 1$. Moreover,

$$\left| \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{(\text{B.142})=O(\rho^u),\tilde{A}_3 \le u \le a_1+a_2,\, |\{j_t:t=1,\dots,8\}|=4} \right| = O(pD_{\max}|J_A|).$$

In addition, when $j_1 \neq j_5$, we note that (B.142) can achieve the order between $O(\rho^{a_1+a_2})$ and $O(\rho^{\tilde{A}_1})$. Under this case,

$$\left| \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{(\text{B.142})=O(\rho^u),\tilde{A}_4 \le u \le a_1+a_2,\, |\{j_t:t=1,\dots,8\}|=4} \right| = O(|J_A|^2).$$

In summary, by $|J_A| \le pD_{\max}$,

$$\begin{aligned}
&\left| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.142}) \times \mathbf{1}_{|\{j_t:t=1,\dots,8\}|=4} \right| \\
&= O(pD_{\max}|J_A|\rho^{\tilde{A}_3}) + O(pD_{\max}|J_A|\rho^{a_1+a_2}) + O(|J_A|^2\rho^{\tilde{A}_1}).
\end{aligned}$$

Case (2.4): When $|\{j_t : t = 1, \dots, 8\}| \le 3$, we know by Condition 2.1,

$$\left| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.142}) \times \mathbf{1}_{|\{j_t:t=1,\dots,8\}|\le 3} \right| = O(p^3).$$

In summary, combining Cases (2.1)–(2.4) above, we know

$$\begin{aligned}
&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(1)} \\
&= \prod_{l=1}^{2} c(n,a_l) n^{a_1+a_2-2} \sum_{s=0}^{A_0} \Big\{ O(p^3) + O(|J_A|^3 \rho^{\tilde{A}_1}) \\
&\quad + O(D_{\max}|J_A|^2 \rho^{\tilde{A}_1}) + O(D_{\max}|J_A|^2 \rho^{\tilde{A}_2}) + O(p|J_A|^2 \rho^{a_1+a_2}) \\
&\quad + O(pD_{\max}|J_A| \rho^{\tilde{A}_3}) + O(pD_{\max}|J_A| \rho^{a_1+a_2}) + O(|J_A|^2 \rho^{\tilde{A}_1}) \Big\},
\end{aligned} \tag{B.149}$$

where $\tilde{A}_1 = 2(a_1+a_2) - 2(s+1)$, $\tilde{A}_2 = 3(a_1+a_2)/2 - (s+1)$ and $\tilde{A}_3 = (a_1+a_2)/2 + s + 1$.

Note that

$$\begin{aligned}
&\prod_{l=1}^{2} c(n,a_l) \times n^{a_1+a_2-2} |J_A|^3 \rho^{\tilde{A}_1} \\
&= p^{-4} n^{-2} |J_A|^3 \rho^{2(a_1+a_2-s-1)} \\
&= p^{-4} n^{-2} (|J_A|\rho^{a_1} \times |J_A|\rho^{a_2})^{\frac{2(a_1+a_2-s-1)}{a_1+a_2}} |J_A|^{3-\frac{4(a_1+a_2-s-1)}{a_1+a_2}} \qquad \text{(B.150)} \\
&= O(1) n^{-2} p^{\frac{4(a_1+a_2-s-1)}{a_1+a_2}-4} n^{-(a_1+a_2-s-1)} |J_A|^{-1+\frac{4(s+1)}{a_1+a_2}} \qquad \text{(B.151)} \\
&= O(1) n^{-2} p^{-\frac{4(s+1)}{a_1+a_2}} n^{-(a_1+a_2-s-1)} |J_A|^{-1+\frac{4(s+1)}{a_1+a_2}} \\
&= O(1) n^{-2} (|J_A|/p^2)^{\frac{2(s+1)}{a_1+a_2}} |J_A|^{-1+\frac{2(s+1)}{a_1+a_2}} \\
&= o(n^2),
\end{aligned}$$

where from (B.150) to (B.151), we use $|J_A|\rho^a = O(pn^{-a/2})$, and in the last equation, we use $2(s+1) \le a_1 + a_2 - 1$. Following similar analysis, we know that all the terms in (B.149) are $o(n^{-2})$ and $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(1)} = o(n^{-2})$.

We next examine $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)}$. Note that if (B.142) $\neq 0$, (j_2,j_4) and $(j_6,j_8) \in J_A$. We can discuss different cases of $\{j_1,\ldots,j_8\}$ similarly as above. Then by Conditions 2.5 and A.1, as $\rho = O(|J_A|^{-1/a_t} p^{1/a_t} n^{-1/2})$ for $t=1,2$, we have $\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}$ (B.142) $= O(p^4)$. Given that $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| < a_1+a_2-2$ in $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)}$, we obtain $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)} = \prod_{l=1}^2 c(n,a_l) \times O\{p^4 n^{a_1+a_2-3}\} = o(n^{-2})$.

In summary, we have $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\} = o(n^{-2})$.

Case (3): When $M = 4$, we consider $j_1 \neq j_2 \neq j_3 \neq j_4$ and $j_5 \neq j_6 \neq$

$j_7 \neq j_8$ under this case. Since $\sigma_{j_1,j_2} = \sigma_{j_3,j_4} = \sigma_{j_5,j_6} = \sigma_{j_7,j_8} = 0$,

$$\begin{aligned}
\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) &= \kappa_1(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}),\\
\mathrm{E}(x_{1,j_5}x_{1,j_6}x_{1,j_7}x_{1,j_8}) &= \kappa_1(\sigma_{j_5,j_7}\sigma_{j_6,j_8} + \sigma_{j_5,j_8}\sigma_{j_6,j_7}),
\end{aligned}$$

which are $O(\rho^2)$. Following similar analysis to **Case (2)**, we can examine the different cases when $|\{j_t : t = 1, \dots, 8\}|$ is between 4 and 8, and obtain,

$$\begin{aligned}
&\operatorname{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(1)}\\
&= O(1)\prod_{l=1}^{2} c(n,a_l) \times n^{a_1+a_2-2}\sum_{s=0}^{A_0}\Big[|J_A|^2\rho^{4(s+1)}\\
&\quad + D_{\max}|J_A|^2\rho^{4(s+1)}\Big(\rho^{a_1-1-s} + \rho^{a_2-1-s}\Big)\\
&\quad + \max\{|J_A|, D_{\max}^2\} \times |J_A|^2\rho^{4(s+1)}\Big(\rho^{2(a_1-1-s)} + \rho^{2(a_2-1-s)}\Big)\\
&\quad + D_{\max}|J_A|^3\Big(\rho^{2(a_1+a_2)-(a_1-1-s)} + \rho^{2(a_1+a_2)-(a_2-1-s)}\Big)\\
&\quad + |J_A|^4\rho^{2(a_1+a_2)}.
\end{aligned} \tag{B.152}$$

Note that $\prod_{l=1}^{2} c(n,a_l)n^{a_1+a_2-2}|J_A|^4\rho^{2(a_1+a_2)} = O(1)p^{-4}n^{-2}p^4n^{-(a_1+a_2)} = o(n^{-2})$. Moreover,

$$\begin{aligned}
&\prod_{l=1}^{2} c(n,a_l) \times n^{a_1+a_2-2}D_{\max}|J_A|^2\rho^{4(s+1)}\Big(\rho^{a_1-1-s} + \rho^{a_2-1-s}\Big)\\
&= p^{-4}n^{-2}D_{\max}|J_A|^2\Big(\rho^{a_1+3(s+1)} + \rho^{a_2+3(s+1)}\Big).
\end{aligned} \tag{B.153}$$

To show (B.153) $= o(n^{-2})$ by symmetricity, it suffices to show for any integer a_1, $p^{-4}D_{\max}|J_A|^2\rho^{a_1+3(s+1)} = o(1)$.

$$p^{-4}D_{\max}|J_A|^2\rho^{a_1+3(s+1)}$$

$$= p^{-4}D_{\max}(|J_A|\rho^{a_1})^{\frac{a_1+3(s+1)}{a_1}}|J_A|^{2-\frac{a_1+3(s+1)}{a_1}} \tag{B.154}$$

$$= O(1)p^{-4}D_{\max}(pn^{-a_1/2})^{\frac{a_1+3(s+1)}{a_1}}|J_A|^{2-\frac{a_1+3(s+1)}{a_1}} \tag{B.155}$$

$$\begin{aligned}
&= O(1)n^{-\frac{a_1+3(s+1)}{2}}(|J_A|/p^2)^{1-\frac{(s+1)}{a_1}}\\
&\quad \times (D_{\max}/p)^{1-\frac{s+1}{a_1}}(D_{\max}/|J_A|)^{\frac{s+1}{a_1}}|J_A|^{-\frac{s+1}{a_1}},\\
&= o(1),
\end{aligned}$$

where from (B.154) to (B.155), we use $|J_A|\rho^{a_1} = O(pn^{-a_1/2})$, and in the last equation we use $|J_A| = o(p^2)$, $D_{\max} \le p$ and $D_{\max} \le |J_A|$. For other terms in (B.152), similar analysis can be applied and we have $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(1)} = o(n^{-2})$.

In addition, similarly to the analysis of $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)}$, by Conditions 2.5 and A.1, we still have $\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}$ (B.142) $= O(p^4)$. Since $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| < a_1 + a_2 - 2$ in $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(2)}$ by construction, we obtain $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(2)} = \prod_{l=1}^2 c(n,a_l) \times O\{p^4 n^{a_1+a_2-3}\} = o(n^{-2})$. In summary, $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\} = o(n^{-2})$ is proved.

B.4.4. *Proof of Lemma A.14 (on Page 47, Section A.5).* Similarly to Section B.1.6,

$$\sum_{k=1}^n \mathrm{E}(D_{n,k}^4) = \sum_{k=1}^n \sum_{1\le r_1,r_2,r_3,r_4\le m} \prod_{l=1}^4 t_{r_l} \times \mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_{r_l}}\Big),$$

where we use the redefined notation in Section A.5. To prove Lemma A.6, it suffices to show that for given $1 \le k \le n$ and $1 \le r_1, r_2, r_3, r_4 \le m$, we have $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_{r_l}}) = o(n^{-1})$. Moreover by the Cauchy-Schwarz inequality, it suffices to show $\mathrm{E}(A_{n,k,a}^4) = o(n^{-1})$ for $a \in \{a_1, \dots, a_m\}$. Following (B.61), we have $A_{n,k,a} = 0$ when $k < a$; and when $k \ge a$,

$$\mathrm{E}(A_{n,k,a}^4) = c^2(n,a) \sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a-1),\, l=1,2,3,4;\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8),$$

where $\mathbf{i}^{(l)} = (i_1^{(l)}, \dots, i_a^{(l)})$ represents tuples $1 \le i_1^{(l)} \ne \dots \ne i_a^{(l)} \le n$, and

$$Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = \mathrm{E}\Big(\prod_{r=1}^8 x_{k,j_r}\Big)\,\mathrm{E}\Big(\prod_{l=1}^4\prod_{t=1}^{a-1} x_{i_t^{(l)},j_{2l-1}} x_{i_t^{(l)},j_{2l}}\Big).$$

As $c(n,a) = \Theta(p^{-1}n^{-a/2})$, to prove $\mathrm{E}(A_{n,k,a}^4) = o(n^{-1})$, it suffices to show

$$\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a-1),\, l=1,2,3,4;\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = o(p^4 n^{2a-1}).$$

Since $\sigma_{j_1,j_2} = 0$ if $(j_1,j_2) \in J_A^c$, then similarly to Section B.1.6, we have $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) \ne 0$ only when $|\bigcup_{l=1}^4 \{\mathbf{i}^{(l)}\}| \le 2(a-1)$, and similarly to (B.65),

$$\sum_{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a-1),\, l=1,\dots,4} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(n^{2a-2}).$$

It then remains to show

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)=O(p^4). \tag{B.156}$$

We next prove by discussing $|\{j_t : t=1,\dots,8\}|$ and the corresponding value of $Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)$. By Condition A.1, $Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)$ can be written as certain linear combination of $\prod_{t=1}^{4a}(\sigma_{j_{g_{2t-1}},j_{g_{2t}}})$, where $g_{2t-1}\neq g_{2t}$ and $(g_1,\dots,g_{8a})$ contain a number of $1,\dots,8$ respectively. If $|\{j_t : t=1,\dots,8\}|\le 4$, by Condition 2.1,

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\dots,8\}|\le 4\}}=O(p^4).$$

If $|\{j_t : t=1,\dots,8\}|=5$, note that for $j_1\neq j_2$, $\sigma_{j_1,j_2}\neq 0$ only when $(j_1,j_2)\in J_A$, then

$$\begin{aligned}&\Big|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\dots,8\}|=5\}}\Big|\\ &\le C\sum_{\substack{1\le j_1,j_2,j_5\le p,\\ (j_6,j_8)\in J_A}}\sigma^a_{j_1,j_1}\sigma^a_{j_2,j_2}\sigma^a_{j_5,j_5}\sigma^a_{j_6,j_8}=O(p^3|J_A|\rho^a)=o(p^4),\end{aligned}$$

where in the last equation, we use $|J_A|\rho^a=O(pn^{-a/2})$. In addition, similarly, if $|\{j_t : t=1,\dots,8\}|=6$,

$$\begin{aligned}&\Big|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\dots,8\}|=6\}}\Big|\\ &\le C\sum_{\substack{1\le j_1,j_2\le p,\\ (j_5,j_7),(j_6,j_8)\in J_A}}\sigma^a_{j_1,j_1}\sigma^a_{j_2,j_2}\sigma^a_{j_5,j_7}\sigma^a_{j_6,j_8}=O(p^2|J_A|^2\rho^{2a})=o(p^4).\end{aligned}$$

If $|\{j_t : t=1,\dots,8\}|=7$,

$$\begin{aligned}&\Big|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\dots,8\}|=7\}}\Big|\\ &\le C\sum_{\substack{1\le j_1\le p,\\ (j_2,j_4),(j_5,j_7),(j_6,j_8)\in J_A}}\sigma^a_{j_1,j_1}\sigma^a_{j_2,j_4}\sigma^a_{j_5,j_7}\sigma^a_{j_6,j_8}=O(p|J_A|^3\rho^{3a})=o(p^4).\end{aligned}$$

If $|\{j_t : t=1,\dots,8\}|=8$,

$$\begin{aligned}&\Big|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\dots,8\}|=8\}}\Big|\\ &\le C\sum_{(j_1,j_3),(j_2,j_4),(j_5,j_7),(j_6,j_8)\in J_A}\sigma^a_{j_1,j_3}\sigma^a_{j_2,j_4}\sigma^a_{j_5,j_7}\sigma^a_{j_6,j_8}=O(|J_A|^4\rho^{4a})=o(p^4).\end{aligned}$$

In summary, (B.156) is obtained and Lemma A.14 is proved. □

B.5. Lemmas for the proof of Theorem 4.1. In this section, we prove Lemma A.15 on Page 55, where we prove $\mathrm{var}(\sum_{k=1}^n \pi_{n,k}^2) \to 0$ and $\sum_{k=1}^n \mathrm{E}(D_{n,k}^4) \to 0$ in the following Sections B.5.1 and B.5.1, respectively.

B.5.1. *Proof of Lemma A.15 (on Page 55, Section A.9).*

Proof of $\mathrm{var}(\sum_{k=1}^n \pi_{n,k}^2) \to 0$. Similarly to Section B.1.5, $D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$, and then $\pi_{n,k}^2 = \sum_{1\leq r_1,r_2\leq m} t_{r_1}t_{r_2}\mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})$. Note that by the Cauchy-Schwarz inequality, for some constant C,

$$\mathrm{var}\Big(\sum_{k=1}^n \pi_{n,k}^2\Big) \leq Cn^2 \max_{1\leq k\leq n;\, 1\leq r_1,r_2\leq m} \mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}),$$

where $c(n,a) = [a \times \{\sigma(a)P_a^n\}^{-1}]^2$ and for two integers a_1 and a_2 we still define $\mathbb{T}_{k,a_1,a_2} = \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$. In particular, when $k < \max\{a_1,a_2\}$, $\mathbb{T}_{k,a_1,a_2} = 0$; when $k \geq \max\{a_1,a_2\}$,

$$\begin{aligned}\mathbb{T}_{k,a_1,a_2} &= \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2}) \\ &= \sum_{\substack{1\leq j_1,j_2\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2}} \{c(n,a_1)c(n,a_2)\}^{1/2}\sigma_{j_1,j_2}\prod_{l=1}^{2}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}.\end{aligned}$$

To prove Lemma $\mathrm{var}(\sum_{k=1}^n \pi_{n,k}^2) \to 0$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, where $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = \mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2$. We consider without loss of generality that $k \geq \max\{a_1,a_2\}$.

When $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$, $\mathrm{E}(\prod_{l=1}^2\prod_{t=1}^{a_l} x_{i_t^{(l)},j_t}) = 0$; and when $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$, it induces $a_1 = a_2$ and $\mathrm{E}(\prod_{l=1}^2\prod_{t=1}^{a_l} x_{i_t^{(l)},j_t}) = \sigma_{j_1,j_2}^a$ where we write $a_1 = a_2 = a$. It follows that when $a_1 \neq a_2$, $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = 0$; when $a_1 = a_2 = a$,

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = \sum_{\substack{1\leq j_1,j_2\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2}} \mathbf{1}_{\{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}\}} \times \{c(n,a_1)c(n,a_2)\}^{1/2}\sigma_{j_1,j_2}^a.$$

Then

$$\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = \sum_{\substack{1\leq j_1,j_2,j_3,j_4\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}}\right\}} \prod_{l=1}^{2} c(n,a_l) \times (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a.$$

In addition, we obtain

$$\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) = \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \Big\{\prod_{l=1}^{2} c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big),$$

where for simplicity of representation, we set $a_3 = a_1$ and $a_4 = a_2$. Define

$$G_{k,a_1,a_2,1} = \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\},\\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}=\emptyset}\right\}} \times \Big\{\prod_{l=1}^{2} c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big).$$

Since $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2| \le |\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{k,a_1,a_2,1}| + |G_{k,a_1,a_2,1} - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2|$, we next prove $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{k,a_1,a_2,1} = o(n^{-2})$ and $G_{k,a_1,a_2,1} - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = o(n^{-2})$ respectively.

Step I: $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{k,a_1,a_2,1} = o(n^{-2})$. When $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$, $\{\mathbf{i}^{(3)}\} = \{\mathbf{i}^{(4)}\}$ and $\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\} = \emptyset$, it implies that $a_1 = a_2 = a$, $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| \le a_1 + a_2 - 3$, and

$$\Big(\prod_{l=1}^{2}\sigma_{j_{2l-1},j_{2l}}\Big) \times \mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big) = (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a.$$

It follows that if $a_1 \ne a_2$, $\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{k,a_1,a_2,1} = 0$; if $a_1 = a_2 = a$,

$$\begin{aligned}&\Big|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{k,a_1,a_2,1}\Big| \\ &= c(n,a_1)c(n,a_2)O(n^{a_1+a_2-3})\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a\Big| = o(n^{-2})\end{aligned}$$

where we use $c(n,a) = \Theta(p^{-1}n^{-a})$ and by Condition A.2,

$$\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a = O(p^2). \tag{B.157}$$

Step II: $G_{k,a_1,a_2,1} - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = o(n^{-2})$. We write $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{k,a_1,a_2,1} = G_{k,a_1,a_2,2} + G_{k,a_1,a_2,3}$, where

$$G_{k,a_1,a_2,2} = \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\},\\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}\ne\emptyset}\right\}} \times \Big\{\prod_{l=1}^{2} c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big),$$

and

$$G_{k,a_1,a_2,3} = \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}\neq\{\mathbf{i}^{(2)}\}\text{ or }\\ \{\mathbf{i}^{(3)}\}\neq\{\mathbf{i}^{(4)}\}}\right\}} \times \Big\{\prod_{l=1}^{2} c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big).$$

For $G_{k,a_1,a_2,2}$, it is a summation over the indexes satisfying $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}, \{\mathbf{i}^{(3)}\} = \{\mathbf{i}^{(4)}\}$ and $\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}\neq\emptyset$. Thus $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| \le a_1+a_2-3$, and by $c(n,a)=\Theta(p^{-1}n^{-a})$ and (B.157),

$$|G_{k,a_1,a_2,2}| \le Cp^{-2}n^{-(a_1+a_2)}n^{a_1+a_2-3}\sum_{1\le j_1,j_2,j_3,j_4\le p}\sigma_{j_1,j_2}\sigma_{j_3,j_4} = o(n^{-2}).$$

For $G_{k,a_1,a_2,3}$, it is a summation over the indexes satisfying $\{\mathbf{i}^{(1)}\}\neq\{\mathbf{i}^{(2)}\}$ or $\{\mathbf{i}^{(3)}\}\neq\{\mathbf{i}^{(4)}\}$. We assume without loss of generality that $\{\mathbf{i}^{(1)}\}\neq\{\mathbf{i}^{(2)}\}$ and there exists an index $m\in\{\mathbf{i}^{(1)}\}$ but $m\notin\{\mathbf{i}^{(2)}\}$. Similarly to Section B.1.5, we know

$$\Big(\prod_{l=1}^{2}\sigma_{j_{2l-1},j_{2l}}\Big)\times\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1}x_{i_t^{(l)},j_l}\Big) \tag{B.158}$$

is nonzero only when $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| \le a_1+a_2-2$, that is, each index appears at least twice among the four sets $\{\mathbf{i}^{(l)}\}, l=1,2,3,4$. Therefore, we know if (B.158) $\neq 0$, $m\in\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\}$. If $m\in\{\mathbf{i}^{(3)}\}$ but $m\notin\{\mathbf{i}^{(4)}\}$, (B.158) $=\sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_1,j_3}\mathrm{E}(\text{other terms})$. Under this case, we define $\tilde{K}_0 = -(2+\epsilon)(4+\gamma)\log p/(\epsilon\log\delta)$, where γ and ϵ are some positive constants and δ is from Condition A.2. Then we have

$$\begin{aligned}\sum_{1\le j_1,j_2,j_3,j_4\le p}(\text{B.158}) &\le C\sum_{1\le j_1,j_2,j_3,j_4\le p}\sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_1,j_3}\\ &\le C\sum_{\substack{|j_1-j_2|\le\tilde{K}_0,\\ |j_3-j_4|\le\tilde{K}_0,\\ |j_1-j_3|\le\tilde{K}_0}} 1 + C\sum_{|j_1-j_2|\ge\tilde{K}_0}\delta^{|j_1-j_2|\epsilon/(2+\epsilon)}\\ &= O(p\tilde{K}_0^2)+O(p^4p^{-(4+\gamma)}),\end{aligned} \tag{B.159}$$

where in the second inequality, we use the symmetricity of j indexes and also use Lemma B.1 similarly as in Section A.9. If $m\in\{\mathbf{i}^{(4)}\}$ but $m\notin\{\mathbf{i}^{(s)}\}$, (B.159) also holds similarly. If $m\in\{\mathbf{i}^{(3)}\}$ and $m\in\{\mathbf{i}^{(4)}\}$, (B.158) $=$

$\sigma_{j_1,j_2}\sigma_{j_3,j_4}\mathrm{E}(x_{m,j_1}x_{m,j_3}x_{m,j_4})\mathrm{E}(\text{other terms})$. Similarly to (B.159), as $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, if $|j_1 - j_3| > \tilde{K}_0$ and $|j_1 - j_4| > \tilde{K}_0$, (B.158) $\leq C\delta^{|j_1-j_2|\epsilon/(2+\epsilon)}$. Thus under this case, we also have $\sum_{1\leq j_1,j_2,j_3,j_4\leq p}$ (B.158) $= O(p\tilde{K}_0^2) + O(p^{-\gamma})$. Recall that (B.158) $\neq 0$ only when $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| \leq a_1+a_2-2$. By $c(n,a) = \Theta(p^{-1}n^{-a})$ and $\tilde{K}_0 = O(\log p)$,

$$
\begin{aligned}
|G_{k,a_1,a_2,3}| &\leq Cp^{-2}n^{-(a_1+a_2)}n^{a_1+a_2-2}\sum_{1\leq j_1,j_2,j_3,j_4\leq p}|(\text{B.158})| \\
&= n^{-2}p^{-2}\Big\{O(p\tilde{K}_0^2) + O(p^{-\gamma})\Big\} = o(n^{-2}).
\end{aligned}
$$

In summary,

$$
\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) \leq |\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{k,a_1,a_2,1}| + |G_{k,a_1,a_2,2}| + |G_{k,a_1,a_2,3}| = o(n^{-2}),
$$

and then $\mathrm{var}(\sum_{k=1}^n \pi_{n,k}^2) \to 0$ is proved.

Proof of $\sum_{k=1}^n \mathrm{E}(D_{n,k}^4) \to 0$. Similarly to Section B.1.6,

$$
\sum_{k=1}^n \mathrm{E}(D_{n,k}^4) = \sum_{k=1}^n \sum_{1\leq r_1,r_2,r_3,r_4\leq m}\prod_{l=1}^4 t_{r_l} \times \mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_{r_l}}\Big).
$$

To prove $\sum_{k=1}^n \mathrm{E}(D_{n,k}^4) \to 0$, it suffices to show that for given $1 \leq k \leq n$ and finite integers (a_1, a_2, a_3, a_4), we have $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_l}) = o(n^{-1})$.

In particular,

$$
\begin{aligned}
\mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_l}\Big) \quad &= \quad \Big\{\prod_{l=1}^4 c(n,a_l)\Big\}^{1/2} \sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,\ldots,4;\\ 1\leq j_1,j_2,j_3,j_4\leq p}} \\
&\qquad \mathrm{E}\Big(\prod_{l=1}^4 x_{k,j_l}\Big)\mathrm{E}\Big(\prod_{l=1}^4\prod_{t=1}^{a_l-1} x_{i_t,j_l}\Big).
\end{aligned}
$$

Similarly to Section B.1.6, we have $\mathrm{E}(\prod_{l=1}^4\prod_{t=1}^{a_l-1} x_{i_t,j_l}) \neq 0$ only when $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| \leq \sum_{l=1}^4 (a_l - 1)/2$. We will prove that

$$
\sum_{1\leq j_1,j_2,j_3,j_4\leq p} \mathrm{E}\Big(\prod_{l=1}^4 x_{k,j_l}\Big) = O(p^2). \tag{B.160}
$$

Then as $c(n,a) = \Theta(p^{-1}n^{-a})$,

$$
\mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_l}\Big) = O(1)p^{-2}n^{-\sum_{l=1}^4 a_l/2}n^{\sum_{l=1}^4(a_l-1)/2}p^2 = o(n^{-1}).
$$

To finish the proof, it remains to show (B.160). When $|\{j_1, j_2, j_3, j_4\}| \le 2$,

$$\sum_{1\le j_1,j_2,j_3,j_4\le p} \mathrm{E}\Big(\prod_{l=1}^{4} x_{k,j_l}\Big)\mathbf{1}_{\{|\{j_1,j_2,j_3,j_4\}|\le 2\}} = O(p^2).$$

When $|\{j_1, j_2, j_3, j_4\}| \ge 3$, we assume without loss of generality that $j_1 \le j_2 \le j_3 \le j_4$. For $\tilde{K}_0$ defined in Section B.5.1, if $|j_1 - j_2| > \tilde{K}_0$ or $|j_3 - j_4| > \tilde{K}_0$, $|\mathrm{E}(\prod_{l=1}^{4} x_{k,j_l})| \le C\delta^{|j_1-j_2|\epsilon/(2+\epsilon)} = O(p^{-(4+\gamma)})$. If $|j_1 - j_2| \le \tilde{K}_0$ and $|j_3 - j_4| \le \tilde{K}_0$, but $|j_2 - j_3| > K_0$, by Lemma B.1,

$$\Big|\mathrm{E}\Big(\prod_{l=1}^{4} x_{k,j_l}\Big)\Big| \le \sigma_{j_1,j_2}\sigma_{j_3,j_4} + C\delta^{|j_1-j_2|\epsilon/(2+\epsilon)} = \sigma_{j_1,j_2}\sigma_{j_3,j_4} + O(p^{-(4+\gamma)}).$$

Therefore

$$\begin{aligned}&\sum_{1\le j_1,j_2,j_3,j_4\le p} \mathrm{E}\Big(\prod_{l=1}^{4} x_{k,j_l}\Big)\mathbf{1}_{\{|\{j_1,j_2,j_3,j_4\}|\ge 3\}}\\ &= O(p\tilde{K}_0^3) + O(p^4 p^{-(4+\gamma)}) + \sum_{1\le j_1,j_2,j_3,j_4\le p} \sigma_{j_1,j_2}\sigma_{j_3,j_4} = O(p^2),\end{aligned}$$

where in the last equation, we use Condition A.2 (2). In summary, (B.160) is proved and the proof is finished.

B.6. Lemmas for the proof of Theorem 4.3.

B.6.1. *Proof of Lemma A.16 (on Page 56, Section A.10).* Under H_0 : $\boldsymbol{\mu} = \boldsymbol{\nu}$, we assume $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$ without loss of generality by Proposition A.1. To derive $\mathrm{var}\{\mathcal{U}(a)\}$, we write $\mathcal{U}(a) = \sum_{j=1}^{p} \mathcal{U}^{(j)}(a)$, where we define $G(a, c) = (-1)^{a-c}\binom{a}{c}(P_c^{n_x})^{-1}(P_{a-c}^{n_y})^{-1}$, and

$$\mathcal{U}^{(j)}(a) = \sum_{c=0}^{a} G(a, c) \sum_{\substack{\mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}} \prod_{t=1}^{c} x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j}. \tag{B.161}$$

Since $\mathrm{E}\{\mathcal{U}(a)\} = 0$ under H_0,

$$\mathrm{var}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}^2(a)\} = \sum_{1\le j_1,j_2\le p} \mathrm{E}\{\mathcal{U}^{(j_1)}(a) \times \mathcal{U}^{(j_2)}(a)\}. \tag{B.162}$$

Note that for given $1 \le j_1, j_2 \le p$,

$$\mathrm{E}\{\mathcal{U}^{(j_1)}(a)\mathcal{U}^{(j_2)}(a)\} = \sum_{\substack{0\le c\le a,\\ \mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}} \sum_{\substack{0\le \tilde{c}\le a,\\ \tilde{\mathbf{k}}\in\mathcal{P}(n_x,\tilde{c}),\\ \tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-\tilde{c})}} G(a, c)G(a, \tilde{c})Q(\mathbf{k}, \mathbf{s}, \tilde{\mathbf{k}}, \tilde{\mathbf{s}}, \mathbf{j}).$$

where we define

$$Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = \mathrm{E}\Big(\prod_{t=1}^{c} x_{k_t,j_1} \prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}\Big)\mathrm{E}\Big(\prod_{m=1}^{a-c} y_{s_m,j_1} \prod_{\tilde{m}=1}^{a-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}\Big).$$

Since we assume the $n = n_x + n_y$ copies are independent from each other and $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$, then $Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}}) = 0$ if $\{\mathbf{k}\} \neq \{\tilde{\mathbf{k}}\}$ or $\{\mathbf{s}\} \neq \{\tilde{\mathbf{s}}\}$. If $\{\mathbf{k}\} = \{\tilde{\mathbf{k}}\}$ and $\{\mathbf{s}\} = \{\tilde{\mathbf{s}}\}$, it induces $c = \tilde{c}$ and $Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = \sigma^{c}_{x,j_1,j_2}\sigma^{a-c}_{y,j_1,j_2}$. It follows that

$$\begin{aligned}
\text{(B.163)} \quad \mathrm{E}\{\mathcal{U}^{(j_1)}(a)\mathcal{U}^{(j_2)}(a)\} &= \sum_{c=0}^{a} G^2(c) P^{n_x}_{c} P^{n_y}_{a-c} c!(a-c)!\sigma^{c}_{x,j_1,j_2}\sigma^{a-c}_{y,j_1,j_2} \\
&= a!\sum_{c=0}^{a}\binom{a}{c}(P^{n_x}_{c})^{-1}(P^{n_y}_{a-c})^{-1}\sigma^{c}_{x,j_1,j_2}\sigma^{a-c}_{y,j_1,j_2} \\
&\simeq a!\Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^{a}.
\end{aligned}$$

Combining (B.162) and (B.163), we obtain $\mathrm{var}\{\mathcal{U}(a)\}$. By Condition A.4, $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(pn^{-a})$.

B.6.2. *Proof of Lemma A.17 (on Page 56, Section A.10).* Since under H_0, $\mathrm{E}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}(b)\} = 0$, we have $\mathrm{cov}\{\mathcal{U}(a),\mathcal{U}(b)\} = \mathrm{E}\{\mathcal{U}(a)\times\mathcal{U}(b)\}$. Following (B.161),

$$\text{(B.164)} \qquad \mathrm{E}\{\mathcal{U}(a)\times\mathcal{U}(b)\} = \sum_{1\le j_1,j_2\le p} \mathrm{E}\{\mathcal{U}^{(j_1)}(a)\times\mathcal{U}^{(j_2)}(b)\},$$

where

$$\begin{aligned}
\mathrm{E}\{\mathcal{U}^{(j_1)}(a)\times\mathcal{U}^{(j_2)}(b)\} = &\sum_{\substack{0\le c\le a,\\ \mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}} \sum_{\substack{0\le \tilde{c}\le b,\\ \tilde{\mathbf{k}}\in\mathcal{P}(n_x,c),\\ \tilde{\mathbf{s}}\in\mathcal{P}(n_y,b-c)}} G(a,c)G(b,\tilde{c}) \\
&\times \mathrm{E}\Big(\prod_{t=1}^{c} x_{k_t,j_1}\prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}\Big)\mathrm{E}\Big(\prod_{m=1}^{a-c} y_{s_m,j_1}\prod_{\tilde{m}=1}^{b-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}\Big).
\end{aligned}$$

As $a \neq b$, $\{\mathbf{k}\} \neq \{\tilde{\mathbf{k}}\}$ and $\{\mathbf{s}\} \neq \{\tilde{\mathbf{s}}\}$ always hold. Then as $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$, $\mathrm{E}(\prod_{t=1}^{c} x_{k_t,j_1}\prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}) = 0$ and $\mathrm{E}(\prod_{m=1}^{a-c} y_{s_m,j_1}\prod_{\tilde{m}=1}^{b-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}) = 0$, similarly to Section B.1.2. It follows that (B.164) $= 0$ and the lemma is proved.

B.6.3. *Proof of Lemma A.18 (on Page 57, Section A.10).* By the Cramér-Wold Theorem, to prove the asymptotic joint normality of the U-statistics, it suffices to prove that any of their fixed converges to normal. For illustration, we first prove the asymptotic normality for each $\mathcal{U}(a)$ of finite a. The similar arguments can be applied to the linear combination of finite U-statistics and then the joint normality is obtained.

Recall $\mathcal{U}(a) = \sum_{j=1}^{p} \mathcal{U}^{(j)}(a)$ from (B.161). To derive the limiting distribution of $\mathcal{U}(a)$, we use Bernstein's block method in [40, page 338]; see also [13, 76]. Specifically, we partition the sequence, $\sigma^{-1}(a)\times\mathcal{U}^{(j)}(a)$, $j=1,\dots,p$, into r blocks, where each block contains b variables such that $rb \le p < (r+1)b$. For each $1 \le k \le r$, we partition the kth block into two sub-blocks with a larger one $A_{k,1}$ and a smaller one $A_{k,2}$. Suppose each $A_{k,1}$ has b_1 variables and each $A_{k,2}$ has $b_2 = b - b_1$ variables. We require $r \to \infty$, $b_1 \to \infty$, $b_2 \to \infty$, $rb_1/p \to 1$ and $rb_2/p \to 0$ as $p \to \infty$. We write

$$A_{k,1}(a) = \sum_{i=1}^{b_1} \mathcal{U}^{(k-1)b+i}(a), \quad A_{k,2}(a) = \sum_{i=1}^{b_2} \mathcal{U}^{(k-1)b+b_1+i}(a),$$

and further define $\mathcal{U}_1 = \sigma^{-1}(a)\sum_{k=1}^{r} A_{k,1}(a)$, $\mathcal{U}_2 = \sigma^{-1}(a)\sum_{k=1}^{r} A_{k,2}(a)$, and $\mathcal{U}_3 = \sigma^{-1}(a)\sum_{j=rb+1}^{p} \mathcal{U}^{(j)}(a)$. Thus we have the decomposition: $\sigma^{-1}(a)\times \mathcal{U}(a) = \mathcal{U}_1 + \mathcal{U}_2 + \mathcal{U}_3$.

The Bernstein's block method makes $A_{k,1}$ "almost" independent, thus the study of $\mathcal{U}_1$ may be related to the cases of sums of independent random variables. In addition, since b_2 is small compared with b_1, we will show that the sums $\mathcal{U}_2$ and $\mathcal{U}_3$ will be small compared with the total sum of variables in the sequence, i.e., $\sigma^{-1}(a)\times\mathcal{U}(a)$. In particular, we first show

$$\sigma^{-1}(a)\times\mathcal{U}(a) = \mathcal{U}_1 + o_p(1),$$

where $o_p(1)$ represents that the remaining term converges to 0 in probability. Since $\mathrm{E}(\mathcal{U}_2) = \mathrm{E}(\mathcal{U}_3) = 0$, it suffices to prove that $\mathrm{var}(\mathcal{U}_2) = \mathrm{var}(\mathcal{U}_3) = o(1)$.

For $\mathcal{U}_2$, note that $\mathcal{U}_2 = \sigma^{-1}(a)\sum_{k=1}^{r} A_{k,2}(a)$. Then

$$\text{(B.165)} \quad \mathrm{var}(\mathcal{U}_2) \le \sigma^{-2}(a) \sum_{\substack{1\le k_1,k_2\le r;\\ 1\le i_1,i_2\le b_2}} \left| \mathrm{cov}\left\{ \mathcal{U}^{((k_1-1)b+b_1+i_1)}(a),\ \mathcal{U}^{((k_2-1)b+b_1+i_2)}(a) \right\} \right|.$$

Recall $\alpha_x(s)$ and $\alpha_y(s)$ in Condition A.4. Define $\alpha(s) = \alpha_x(s) + \alpha_y(s)$, then $\alpha(s) \le C\delta^s$, where $\delta = \max\{\delta_x, \delta_y\} \in (0,1)$. By the α-mixing inequality in Lemma B.1,

$$\left|\mathrm{cov}\left\{ n^{a/2}\mathcal{U}^{(i)}(a), n^{a/2}\mathcal{U}^{(j)}(a) \right\}\right| \le 8\{\alpha(|i-j|)\}^{\frac{\epsilon}{2+\epsilon}} \max_{1\le j\le p} \left[\mathrm{E}\left| n^{a/2}\mathcal{U}^{(j)}(a)\right|^{2+\epsilon}\right]^{\frac{2}{2+\epsilon}}.$$

We take $\epsilon = 2$, and by Lemma B.10 (on Page 168, Section B.6.4), we have $\max_{1\le j\le p} \mathrm{E}\{n^{a/2}\mathcal{U}^{(j)}(a)\}^{2+\epsilon} < \infty$. It follows that

$$
\begin{aligned}
\text{(B.166)}\quad & \left|\operatorname{cov}\left\{\mathcal{U}^{((k_1-1)b+b_1+i_1)}(a), \mathcal{U}^{((k_2-1)b+b_1+i_2)}(a)\right\}\right| \\
= \;& n^{-a}\left|\operatorname{cov}\left\{n^{a/2}\mathcal{U}^{((k_1-1)b+b_1+i_1)}(a), n^{a/2}\mathcal{U}^{((k_2-1)b+b_1+i_2)}(a)\right\}\right| \\
\le \;& Cn^{-a}\alpha\left\{|((k_1-1)b+b_1+i_1)-((k_2-1)b+b_1+i_2)|\right\}^{\frac{2}{4}} \\
\le \;& Cn^{-a}\delta^{|k_1b+i_1-k_2b-i_2|/2}.
\end{aligned}
$$

By (B.165), (B.166) and $\sigma^2(a) = \Theta(pn^{-a})$ from Lemma A.16,

$$
\begin{aligned}
& \operatorname{var}(\mathcal{U}_2) \\
\le \;& \sigma^{-2}(a)\sum_{\substack{1\le k_1,k_2\le r;\\ 1\le i_1,i_2\le b_2}}\left|\operatorname{cov}\left\{\mathcal{U}^{((k_1-1)b+b_1+i_1)}(a), \mathcal{U}^{((k_2-1)b+b_1+i_2)}(a)\right\}\right| \\
\le \;& \sigma^{-2}(a)\sum_{\substack{1\le k_1,k_2\le r;\\ 1\le i_1,i_2\le b_2}} n^{-a}C\delta^{|k_1b+i_1-k_2b-i_2|/2} \\
= \;& O(1)p^{-1}n^a rb_2 n^{-a} = O(1)rb_2p^{-1},
\end{aligned}
$$

which converges to 0 by our construction, i.e., $rb_2/p \to 0$. This shows that $\operatorname{var}(\mathcal{U}_2) = o(1)$. Next we exmaine $\mathcal{U}_3 = \sigma^{-1}(a)\sum_{j=rb+1}^{p}\mathcal{U}^{(j)}(a)$. Similarly, by Lemmas B.1 and B.10, and $\epsilon = 2$,

$$
\begin{aligned}
\operatorname{var}(\mathcal{U}_3) &= \sigma^{-2}(a)n^{-a}\sum_{i=rb+1}^{p}\sum_{j=rb+1}^{p}\operatorname{cov}\left\{n^{a/2}\mathcal{U}^{(i)}(a), n^{a/2}\mathcal{U}^{(j)}(a)\right\} \\
&\le O(1)p^{-1}n^a n^{-a}\sum_{i=rb+1}^{p}\sum_{j=rb+1}^{p} C\alpha\left(|i-j|\right)^{\frac{\epsilon}{2+\epsilon}} \\
&\le O(1)p^{-1}\sum_{i=rb+1}^{p}\sum_{j=rb+1}^{p}\delta^{|i-j|/2} \\
&\le O(1)p^{-1}(p-rb-1) \\
&\le O(1)p^{-1}b.
\end{aligned}
$$

Since $b/p \to 0$, $\operatorname{var}(\mathcal{U}_3) = o(1)$.

Given $\operatorname{var}(\mathcal{U}_2) = o(1)$ and $\operatorname{var}(\mathcal{U}_3) = o(1)$ above, next we focus on $\mathcal{U}_1$. By the α-mixing assumption in Condition A.4, and following the similar arguments in [40, page 338], we have for properly chosen r and b_2,

$$
\left|\mathrm{E}\left\{\exp(it\mathcal{U}_1)\right\} - \prod_{k=1}^{r}\mathrm{E}\left[\exp\left\{it\sigma^{-1}(a)A_{k,1}(a)\right\}\right]\right| \le 16r\alpha(b_2) \to 0.
$$

This suggests there exist independent random variables $\{\xi_k : k = 1, \cdots, r\}$ such that ξ_k and $A_{k,1}(a)$ are identically distributed and $\mathcal{U}_1$ has the same asymptotic distribution as $\sigma^{-1}(a)\sum_{k=1}^{r}\xi_k$. To prove the asymptotic normality of $\sigma^{-1}(a)\mathcal{U}_1$, now it remains to show that central limit theorem holds for $\sigma^{-1}(a)\sum_{k=1}^{r}\xi_k$. Then we check the Lyapunov condition, i.e., check that the moments of ξ_k satisfy

$$s_r^{-4}\sum_{k=1}^{r}\mathrm{E}\left\{\sigma^{-1}(a)|\xi_k|\right\}^4 \to 0, \tag{B.167}$$

where we define $s_r^2 = \sum_{k=1}^{r}\mathrm{var}\{\sigma^{-1}(a)\xi_k\}$. By Lemma B.10, for even $\epsilon > 0$,

$$\mathrm{M}_{4+\epsilon} := \max_{1\le j\le p}\left\{\left\|n^{a/2}\left\{\mathcal{U}^{(j)}(a)\right\}\right\|_{4+\epsilon}\right\} < \infty. \tag{B.168}$$

Then by the moment bounds in [49, Theorem 1], and the α-mixing assumption in Condition A.4, for $g(2,\epsilon) = \epsilon/(4+\epsilon)$,

$$\mathrm{E}\left(\left[\sum_{j=1}^{b_1} n^{a/2}\left\{\mathcal{U}^{(j)}(a)\right\}\right]^4\right) \le Cb_1^2\Big\{C + M_{4+\epsilon}^4\sum_{j=1}^{b_1} j^{2-1}\alpha(j)^{g(2,\epsilon)}\Big\}$$

As $\delta \in (0,1)$ and $0 < g(2,\epsilon) < 1$,

$$\sum_{j=1}^{\infty} j\alpha(j)^{g(2,\epsilon)} \leqslant C\sum_{j=1}^{\infty} j \times (\delta^{g(2,\epsilon)})^j < \infty.$$

It follows that

$$\begin{aligned}
\mathrm{E}\left\{\sigma^{-1}(a)A_{1,1}(a)\right\}^4 &= \sigma^{-4}(a)n^{-2a}\mathrm{E}\left[\sum_{j=1}^{b_1} n^{a/2}\left\{\mathcal{U}^{(j)}(a)\right\}\right]^4 \\
&\le O(1)p^{-2}n^{2a}n^{-2a} \times b_1^2\left\{C + M_{4+\epsilon}^4\sum_{j=1}^{b_1} j^{2-1}\alpha(j)^{g(2,\epsilon)}\right\} \\
&= O(1)p^{-2} \times b_1^2.
\end{aligned}$$

Similarly, for other $k > 1$, $\mathrm{E}\left\{\sigma^{-1}(a)A_{k,1}(a)\right\}^4$ have the same bound. Thus,

$$\sum_{k=1}^{r}\sigma^{-4}(a)E|\xi_k|^4 = O(1)rp^{-2}b_1^2. \tag{B.169}$$

In addition,

$$\begin{aligned}\text{var}\{\sigma^{-1}(a)\xi_k\} &= \sigma^{-2}(a)\text{var}\left\{\sum_{i=1}^{b_1}\mathcal{U}^{((k-1)b+i)}(a)\right\}\\ &= \sigma^{-2}(a)\sum_{1\le i_1,i_2\le b_1}\text{cov}\left\{\mathcal{U}^{((k-1)b+i_1)}(a),\mathcal{U}^{((k-1)b+i_2)}(a)\right\}\\ &= \sigma^{-2}(a)\sum_{1\le i_1,i_2\le b_1}(\text{B.163}).\end{aligned}$$

By Condition A.4 and $rb_1/p \to 1$, we have

$$\begin{aligned}s_r^4 &= \Big[\sum_{j=1}^{r}\text{var}\left\{\xi_j/\sigma(a)\right\}\Big]^2\\ &= \Theta(1)p^{-2}n^{2a}(r\times b_1 n^{-a})^2 = \Theta(1)p^{-2}r^2b_1^2.\end{aligned} \tag{B.170}$$

Combine (B.169) and (B.170), (B.167) is proved as $r\to\infty$.

In summary, for any finite integer a, we prove the asymptotic normality of $\mathcal{U}(a)/\sigma(a)$. For any linear combination of U-statistics $Z_n := \sum_{r=1}^{m} t_r\mathcal{U}(a_r)/\sigma(a_r)$, we can similarly decompose Z_n into three parts and apply the analysis above. The similar conclusion holds for finite m and the asymptotic joint normality is obtained by the Cramér-Wold Theorem.

B.6.4. *Proof of Lemma B.10 (on Page 168, Section B.6.3).*

LEMMA B.10. *For $\forall$ finite even $\omega > 0$ any $\forall$ finite integer $a > 0$,*

$$\max_{1\le j\le p}\text{E}\left\{n^{a/2}\mathcal{U}^{(j)}(a)\right\}^{\omega} < \infty.$$

PROOF. Recall the definition of $\mathcal{U}^{(j)}(a)$ in (B.161). For positive even ω,

$$\begin{aligned}&\text{E}[\{\mathcal{U}^{(j)}(a)\}^{\omega}]\\ &= \sum_{l=1}^{\omega}\sum_{\substack{0\le c_l\le a,\\ \mathbf{k}^{(l)}\in\mathcal{P}(n_x,c_l),\\ \mathbf{s}^{(l)}\in\mathcal{P}(n_y,a-c_l)}} G(c_l)\text{E}\left(\prod_{l=1}^{\omega}\prod_{t_l=1}^{c_l}x_{k_{t_l}^{(l)},j}\right)\text{E}\left(\prod_{l=1}^{\omega}\prod_{m_l=1}^{a-c_l}y_{s_{m_l}^{(l)},j}\right).\end{aligned} \tag{B.171}$$

Define the index tuple $(\mathbf{k}^{(1)},\ldots,\mathbf{k}^{(\omega)}) = (k_1^{(1)},\ldots,k_{c_1}^{(1)},\ldots,k_1^{(\omega)},\ldots,k_{c_\omega}^{(\omega)})$. When $|\{(\mathbf{k}^{(1)},\ldots,\mathbf{k}^{(\omega)})\}| > \sum_{l=1}^{\omega}c_l/2$, it means that one of the index appears only once. Suppose index $i\in\{(\mathbf{k}^{(1)},\ldots,\mathbf{k}^{(\omega)})\}$ only appears once, then under H_0,

$$\mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{t_l=1}^{c_l} x_{k_{t_l}^{(l)},j}\right) = \mathrm{E}(x_{i,j}) \times \mathrm{E}(\text{other terms}) = 0. \tag{B.172}$$

Thus (B.172) $\neq 0$ only when $|\{(\mathbf{k}^{(1)},\dots,\mathbf{k}^{(\omega)})\}| \leq \sum_{l=1}^{\omega} c_l/2$. By the boundedness of moments in Condition A.4,

$$\max_{1\leq j\leq p} \sum_{0\leq c_l\leq a,\, \mathbf{k}^{(l)}\in\mathcal{P}(n_x,c_l)} \mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{t_l=1}^{c_l} x_{k_{t_l}^{(l)},j}\right) = O\left(n_x^{\sum_{l=1}^{\omega} c_l/2}\right).$$

Similarly, we have

$$\max_{1\leq j\leq p} \sum_{0\leq c_l\leq a,\, \mathbf{s}^{(l)}\in\mathcal{P}(n_y,a-c_l)} \mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{m_l=1}^{a-c_l} y_{s_{m_l}^{(l)},j}\right) = O\left(n_y^{\sum_{l=1}^{\omega} (a-c_l)/2}\right).$$

As $G(a,c) = \Theta(n_x^{-c} n_y^{-(a-c)})$, by (B.171), $\max_{1\leq j\leq p} \mathrm{E}[\{n^{a/2}\mathcal{U}^{(j)}(a)\}^{\omega}] < \infty$. □

B.7. Lemmas for the proof of Theorem 4.4.

B.7.1. *Proof of Lemma A.19 (on Page 57, Section A.11).* Recall $\mathcal{U}^{(j)}(a)$ defined in (B.161). Similarly to $\tilde{\mathcal{U}}_c(a)$, we define $\tilde{\mathcal{U}}_c^{(j)}(a)$ as the sequence of random variables on the conditional probability measure $\tilde{P}$, given the event $n_x n_y \mathcal{U}(\infty)/(n_x+n_y) - \tau_p \leq u$ such that

$$\begin{aligned}&\tilde{P}\Big\{\tilde{\mathcal{U}}_c^{(j)}(a) \leq u_j : 1 \leq j \leq p\Big\}\\ &= P\Big\{\mathcal{U}^{(j)}(a) \leq u_j : 1\leq j\leq p \;\Big|\; \frac{n_x n_y}{n_x+n_y}\mathcal{U}(\infty) \leq \tau_p + u\Big\}.\end{aligned}$$

Then $\sigma^{-1}(a)\tilde{\mathcal{U}}_c(a) = \sigma^{-1}(a)\sum_{j=1}^{p}\tilde{\mathcal{U}}_c^{(j)}(a)$, and we prove the asymptotic normality of $\sigma^{-1}(a)\tilde{\mathcal{U}}_c(a)$ similarly to Section B.6.3. In particular, we partition the sequence $\{\sigma^{-1}(a)\times\tilde{\mathcal{U}}_c^{(j)}(a) : 1\leq j\leq p\}$ into r blocks, where each block contains b variables such that $rb \leq p < (r+1)b$. For each $1\leq k\leq r$, we further partition the kth block into two sub-blocks such that a larger one $\tilde{A}_{k,1}$ contains the first b_1 variables and a smaller one $\tilde{A}_{k,2}$ contains the last $b_2 = b - b_1$ variables. Similarly, for $1 \leq k \leq r$, we write

$$\tilde{A}_{k,1}(a) = \sum_{i=1}^{b_1}\tilde{\mathcal{U}}_c^{(k-1)b+i}(a), \quad \tilde{A}_{k,2}(a) = \sum_{i=1}^{b_2}\tilde{\mathcal{U}}_c^{(k-1)b+b_1+i}(a).$$

Correspondingly, define $\tilde{\mathcal{U}}_1 = \sigma^{-1}(a)\sum_{k=1}^{r}\tilde{A}_{k,1}(a), \tilde{\mathcal{U}}_2 = \sigma^{-1}(a)\sum_{k=1}^{r}\tilde{A}_{k,2}(a)$ and $\tilde{\mathcal{U}}_3 = \sigma^{-1}(a)\sum_{j=rb+1}^{p}\tilde{\mathcal{U}}_c^{(j)}(a)$. Then we have the decomposition: $\sigma^{-1}(a)\times \tilde{\mathcal{U}}_c(a) = \tilde{\mathcal{U}}_1 + \tilde{\mathcal{U}}_2 + \tilde{\mathcal{U}}_3$. To show that $\sigma^{-1}(a)\times\tilde{\mathcal{U}}_c(a)$ satisfies the central limit theorem, we first show that $\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_2^2) = o(1)$ and $\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_3^2) = o(1)$.

$$
\begin{aligned}
\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_2^2) &= \sigma^{-2}(a)\tilde{\mathrm{E}}\Big\{\Big(\sum_{k=1}^{r}\tilde{A}_{k,2}(a)\Big)^2\Big\} \\
&\le \sigma^{-2}(a)\Bigg(\sum_{1\le k_1,k_2\le r}\Big[\tilde{\mathrm{E}}\Big\{\tilde{A}_{k_1,2}^2(a)\Big\}\Big]^{1/2}\Big[\tilde{\mathrm{E}}\Big\{\tilde{A}_{k_2,2}^2(a)\Big\}\Big]^{1/2}\Bigg) \\
&\le \sigma^{-2}(a)\Big[P\Big\{\frac{n_x n_y}{n_x+n_y}\mathcal{U}(\infty) < \tau_p\Big\}\Big]^{-1} \\
&\quad \times\Bigg(\sum_{1\le k_1,k_2\le r}\Big[\mathrm{E}\Big\{A_{k_1,2}^2(a)\Big\}\Big]^{1/2}\Big[\mathrm{E}\Big\{A_{k_2,2}^2(a)\Big\}\Big]^{1/2}\Bigg),
\end{aligned}
$$

where in the last inequality we use the fact that

$$
\begin{aligned}
\tilde{\mathrm{E}}\Big\{\tilde{A}_{k,2}^2(a)\Big\} &= \frac{\mathrm{E}\{A_{k,2}^2(a)\mathbf{1}_{\{n_x n_y\mathcal{U}(\infty)/(n_x+n_y)<\tau_p+u\}}\}}{P\{n_x n_y\mathcal{U}(\infty)/(n_x+n_y) < \tau_p+u\}} \\
&\le \frac{\mathrm{E}\{A_{k,2}^2(a)\}}{P\{n_x n_y\mathcal{U}(\infty)/(n_x+n_y) < \tau_p+u\}}.
\end{aligned}
$$

The upper bound above converges to 0 under the α-mixing condition by choosing proper convergence rate b_2; see Eq. (18.4.8) of [40]. Similarly, we can also show $\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_3^2) = o(1)$. It remains to examine the $\tilde{\mathcal{U}}_1$. Define $\alpha(s)$ as the mixing coefficient of $\{(x_{1,j},\ldots,x_{n_x,j},y_{1,j},\ldots,y_{n_y,j} : j = 1,\ldots,p)\}$ and define $\tilde{\alpha}(s)$ as the corresponding mixing coefficient on the conditional probability measure. Following a similar argument to that in [39, Lemma 2.2], we have

$$
\tilde{\alpha}(d) \le 4\frac{\max_{1\le h\le p-d} P\{U_{h,d}^0(\infty) > \tau_p+u\} + \alpha(d)}{[P\{n_x n_y\mathcal{U}(\infty)/(n_x+n_y) < \tau_p+u\}]^3},
$$

where $U_{h,d}^0(\infty) = \max_{h\le j\le h+d} U^{(j)}(\infty)$, $U^{(j)}(\infty) = \sigma_{j,j}^{-1}\times(\bar{x}_j-\bar{y}_j)^2\times n_x n_y/(n_x+n_y)$, and recall $\tau_p = 2\log p - \log\log p$. Since $x_{i,j}$ and $y_{i,j}$ are sub-gaussian random variables by Condition A.4 [**?** , Proposition 2.5.2], we know $\sigma_{j,j}^{-1/2}\times(\bar{x}_j-\bar{y}_j)\times\sqrt{n_x n_y}/\sqrt{n_x+n_y}$ is a sub-gaussian variable with variance 1. Therefore, $\max_{1\le h\le p-d} P\{U_{h,d}^0(\infty) > \tau_p+u\} \le d\max_{1\le j\le p} P\{U^{(j)}(\infty) > \tau_p+u\} \le Cd\exp\{-(\tau_p+u)/2\} \le Cdp^{-1}\sqrt{\log p}$. Then similarly to [40, page

338], we have

$$
\begin{aligned}
&\Big|\tilde{\mathrm{E}}\Big\{\exp(it\tilde{\mathcal{U}}_1)\Big\} - \prod_{k=1}^{r} \tilde{\mathrm{E}}\Big[\exp\Big\{it\sigma^{-1}(a)\tilde{A}_{k,1}(a)\Big\}\Big]\Big| \\
\leq\ & 16r\tilde{\alpha}(b_2) \\
\leq\ & 64r\frac{\max_{1\leq h\leq p-b_2} P\{U^0_{h,b_2}(\infty) > \tau_p + u\} + \alpha(b_2)}{[P\{n_x n_y \mathcal{U}(\infty)/(n_x+n_y) < \tau_p + u\}]^3},
\end{aligned}
$$

which converges to 0 for properly chosen r and b_2 such that $rb_2\sqrt{\log p}/p \to 0$. Thus there exist independent $\{\tilde{\xi}_k : k = 1,\ldots,r\}$ such that $\tilde{\xi}_k$ and $\tilde{A}_{k1}(a)$ are identically distributed on probability measure $\tilde{P}$. Similarly to [39, Lemma 2.4, Lemma 2.5], we have $\tilde{\mathrm{E}}\{\sigma^{-1}(a)\sum_{k=1}^r \tilde{\xi}_k\} \to 0$ and $\tilde{\mathrm{E}}[\{\sigma^{-1}(a)\sum_{k=1}^r \tilde{\xi}_k\}^2] \to 1$. To show the asymptotic normality on the conditional probability measure, it remains to check the Lyapunov condition that

$$
\sum_{k=1}^{r} \tilde{\mathrm{E}}\left\{\sigma^{-1}(a)|\tilde{\xi}_k|\right\}^4 \leq \sigma^{-4}(a)\frac{\sum_{k=1}^r \mathrm{E}(\xi_k^4)}{P\{n_x n_y\mathcal{U}(\infty)/(n_x+n_y) < \tau_p + u\}} \to 0,
$$

where ξ_k are define same as in Appendix Section B.6.3, and the convergence result follows from (B.167). This implies the asymptotic normality of conditional distribution given $\{n_x n_y \mathcal{U}(\infty)/(n_x+n_y) < \tau_p + u\}$. Thus we obtain the asymptotic independence between $\mathcal{U}(a)/\sigma(a)$ and $\mathcal{U}(\infty)$.

B.8. Lemmas for the proof of Theorem 4.5.

B.8.1. *Proof of Lemma A.20.* Recall the definitions in (A.24). $T_{a,2}$ is the summation over j indexes in the set $\{k_0,\ldots,p\}$ such that $\mu_j = \nu_j = 0$. Then $\mathrm{E}(T_{a,2}) = 0$. Following the argument in Section B.6.1, we obtain

$$
\mathrm{var}(T_{a,2}) \simeq \sum_{k_0+1\leq j_1,j_2\leq p} a!\Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^a.
$$

Let $\mathcal{V}_{a,j_1,j_2} = \{\sigma_{x,j_1,j_2}/\gamma + \sigma_{y,j_1,j_2}/(1-\gamma)\}^a$. By the mixing assumption in Condition A.4 and Lemma B.1, we know there exist some constants C and $\tilde{\delta}$ such that $|\mathcal{V}_{a,j_1,j_2}| \leq C\tilde{\delta}^{|j_1-j_2|}$. Note that

$$
\begin{aligned}
&\Big|\sum_{1\leq j_1,j_2\leq p}\mathcal{V}_{a,j_1,j_2} - \sum_{k_0+1\leq j_1,j_2\leq p}\mathcal{V}_{a,j_1,j_2}\Big| \\
&= \Big|\Big(\sum_{1\leq j_1,j_2\leq k_0} + \sum_{1\leq j_1\leq k_0,\, k_0+1\leq j_2\leq p} + \sum_{1\leq j_2\leq k_0,\, k_0+1\leq j_1\leq p}\Big)\mathcal{V}_{a,j_1,j_2}\Big| \\
&\leq C\Big(\sum_{1\leq j_1,j_2\leq k_0} + \sum_{1\leq j_1\leq k_0,\, k_0+1\leq j_2\leq p} + \sum_{1\leq j_2\leq k_0,\, k_0+1\leq j_1\leq p}\Big)\tilde{\delta}^{|j_1-j_2|} = O(k_0).
\end{aligned}
$$

Since $k_0 = o(p)$ and Condition A.4 assumes that $\sum_{1\le j_1,j_2\le p} \mathcal{V}_{a,j_1,j_2} = \Theta(p)$, then $\sum_{k_0+1\le j_1,j_2\le p} \mathcal{V}_{a,j_1,j_2} = \Theta(p)$. It follows that $\mathrm{var}(T_{a,2}) = \Theta(p^2 n^{-a})$.

It remains to prove $\mathrm{var}(T_{a,1}) = o(pn^{-a})$. Note that $\mathrm{var}(T_{a,1}) = \mathrm{E}(T_{a,1}^2) - \{\mathrm{E}(T_{a,1})\}^2$, and $\mathrm{E}(T_{a,1}) = k_0\rho^a$. Following the definition in (A.24),

$$\mathrm{E}(T_{a,1}^2) = \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{0\le c\le a,\\ \mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}} \sum_{\substack{0\le \tilde{c}\le a,\\ \tilde{\mathbf{k}}\in\mathcal{P}(n_x,\tilde{c}),\\ \tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-\tilde{c})}} G(a,c)G(a,\tilde{c})Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}),$$

where similarly to Section B.6.1,

$$Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = \mathrm{E}\Big(\prod_{t=1}^{c} x_{k_t,j_1} \prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}\Big)\mathrm{E}\Big(\prod_{m=1}^{a-c} y_{s_m,j_1} \prod_{\tilde{m}=1}^{a-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}\Big).$$

Since $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu} = \mathbf{0}$, if $\{\mathbf{s}\} \neq \{\tilde{\mathbf{s}}\}$, $Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = 0$. If $\{\mathbf{s}\} = \{\tilde{\mathbf{s}}\}$, it induces $c = \tilde{c}$. When $\{\mathbf{s}\} = \{\tilde{\mathbf{s}}\}$, let $b = |\{\mathbf{k}\} \cap \{\tilde{\mathbf{k}}\}|$, then $0 \le b \le c$,

$$\mathrm{E}\{Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j})\} = \mu_{j_1}^{c-b}\mu_{j_2}^{c-b}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c} = \rho^{2(c-b)}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c},$$

and

$$\mathrm{E}(T_{a,1}^2) = \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{0\le c\le a,\\ \mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\\ \mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)}} G^2(a,c) \times \rho^{2(c-b)}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c} \times \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}\}}.$$

We next decompose $\mathrm{E}(T_{1,a}^2) = G_{t,1,a,1} + G_{t,1,a,2} + G_{t,1,a,3}$, where

$$G_{t,1,a,1} = \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{0\le c\le a,\\ \mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\\ \mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)}} G^2(a,c)\rho^{2(c-b)}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c}\mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},c=a,b=0\}},$$

$$G_{t,1,a,2} = \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{0\le c\le a,\\ \mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\\ \mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)}} G^2(a,c)\rho^{2(c-b)}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c}\mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},c\le a-1,b=0\}},$$

and

$$G_{t,1,a,3} = \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{0\le c\le a,\\ \mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\\ \mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)}} G^2(a,c)\rho^{2(c-b)}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c}\mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},1\le b\le c\}}.$$

Note that $|\mathrm{var}(T_{a,1})| \leq |G_{t,1,a,1} - \{\mathrm{E}(T_{a,1})\}^2| + |G_{t,1,a,2}| + |G_{t,1,a,3}|$. To prove $\mathrm{var}(T_{a,1}) = o(pn^{-a})$, we will next show $|G_{t,1,a,1} - \{\mathrm{E}(T_{a,1})\}^2|$, $|G_{t,1,a,2}|$ and $|G_{t,1,a,3}|$ are $o(pn^{-a})$ respectively.

First, as $\sum_{\mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,a);\,\mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},c=a,b=0\}} = P^{n_x}_{2a}$ and $G(a,a) = (P^{n_x}_a)^{-1}$,

$$G_{t,1,a,1} = \sum_{1\leq j_1,j_2\leq k_0} \sum_{\substack{0\leq c\leq a,\\ \mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\\ \mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)}} G^2(a,c)\rho^{2a}\mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},c=a,b=0\}} = \frac{P^{n_x}_{2a}}{(P^{n_x}_a)^2}k_0^2\rho^{2a}.$$

Then $|G_{t,1,a,1} - \{\mathrm{E}(T_{a,1})\}^2| = o(1)k_0^2 n^{-2a}n^{2a}\rho^{2a} = o(pn^{-a})$, where we use $\mathrm{E}(T_{a,1}) = k_0\rho^a$. In addition, as $\sum_{\mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},c\leq a-1,b=0\}} = O(n^{2c+a-c})$ and $G(a,c) = \Theta(n^{-a})$, we have

$$|G_{t,1,a,2}| \leq C \sum_{1\leq j_1,j_2\leq k_0} \sum_{c=0}^{a-1} n^{-(a-c)}\rho^{2c}\sigma^{a-c}_{j_1,j_2}.$$

Since $\sum_{1\leq j_1,j_2\leq k_0} \sigma_{j_1,j_2} = O(k_0)$ by Condition A.4 and Lemma B.1, we further know $|G_{t,1,a,2}| = \sum_{c=0}^{a-1} O(k_0\rho^{2c}n^{-(a-c)})$. As $\rho = O(k_0^{-1/a}p^{1/(2a)}n^{-1/2})$ and $k_0 = o(p)$, we obtain $|G_{t,1,a,2}| = o(pn^{-a})$. Moreover, as $G(a,c) = \Theta(n^{-a})$, $\varphi_{j_1,j_2} = \rho^2 + \sigma_{j_1,j_2}$, and $\sum_{\mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},b\geq 1\}} = O(n^{2c-b+a-c})$,

$$|G_{t,1,a,3}| \leq C \sum_{\substack{0\leq c\leq a,\\ 1\leq b\leq c}} \sum_{1\leq j_1,j_2\leq k_0} n^{-(b+a-c)}\rho^{2(c-b)}(\sigma_{j_1,j_2}+\rho^2)^b_{j_1,j_2}\sigma^{a-c}_{j_1,j_2}.$$

For given c and b, the maximum order of $\sum_{1\leq j_1,j_2\leq k_0} n^{-(b+a-c)}\rho^{2(c-b)}(\sigma_{j_1,j_2}+\rho^2)^b_{j_1,j_2}\sigma^{a-c}_{j_1,j_2}$ is bounded by the following two quantities:

$$\sum_{1\leq j_1,j_2\leq k_0} Cn^{-(b+a-c)}\rho^{2c}\sigma^{a-c}_{j_1,j_2}, \tag{B.173}$$

$$\sum_{1\leq j_1,j_2\leq k_0} Cn^{-(b+a-c)}\sigma^{b+a-c}_{j_1,j_2}\rho^{2(c-b)}. \tag{B.174}$$

For (B.173), when $c = a$, (B.173) $= O(k_0^2 n^{-b}\rho^{2a}) = o(pn^{-a})$. When $c \leq a-1$, since $\sum_{1\leq j_1,j_2\leq k_0} \sigma_{j_1,j_2} = O(k_0)$ by Condition A.4 and Lemma B.1, then (B.173) $= O(k_0 n^{-(b+a-c)}\rho^{2c}) = o(pn^{-a})$. For (B.174), as $b \geq 1$, $b + a - c \geq 1$. Then similarly by Condition A.4 and Lemma B.1, (B.174) $= O(k_0 n^{-(b+a-c)}\rho^{2(c-b)}) = o(pn^{-a})$.

In summary, we obtain $\mathrm{var}(T_{a,1}) = o(pn^{-a}) = o(1)\mathrm{var}(T_{a,2})$. Then

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq \mathrm{var}(T_{a,2}) \simeq \sum_{k_0+1\le j_1,j_2\le p} a!\Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^a.$$

By the Markov's inequality, $\{T_{a,1} - \mathrm{E}(T_{a,1})\}/\sigma(a) \xrightarrow{P} 0$.

B.8.2. *Proof of Lemma A.21.* Note that

$$\{\sigma(a)\sigma(b)\}^{-1}\mathrm{cov}\{\mathcal{U}(a), \mathcal{U}(b)\} = \{\sigma(a)\sigma(b)\}^{-1} \times \sum_{1\le l_1,l_2\le 2} \mathrm{cov}(T_{a,l_1}, T_{b,l_2}).$$

Lemma A.20 suggests that $\mathrm{var}(T_{a,1}) = o(1)\sigma^2(a)$. By the Cauchy-Schwarz inequality, $\{\sigma(a)\sigma(b)\}^{-1}\mathrm{cov}\{\mathcal{U}(a), \mathcal{U}(b)\} = \{\sigma(a)\sigma(b)\}^{-1}\mathrm{cov}(T_{a,2}, T_{b,2}) + o(1)$. To finish the proof, it suffices to show $\mathrm{cov}(T_{a,2}, T_{b,2}) = 0$. Note that $T_{a,2}$ and $T_{b,2}$ are summation over j indexes in the set $\{k_0, \ldots, p\}$ such that $\mu_j = \nu_j = 0$. Then the proof in Section B.6.2 applies similarly and we have $\mathrm{cov}(T_{a,2}, T_{b,2}) = 0$.

B.9. Lemmas for the proof of Theorem 4.6.

B.9.1. *Proof of Lemma A.22 (on Page 61, Section A.13).* In the following, we will first derive the form of $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ and then prove that $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$.

As we assume $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = \mathbf{0}$, then $\mathrm{cov}(x_{1,j_1}, x_{1,j_2}) = \mathrm{E}(x_{1,j_1}x_{1,j_2})$ and $\mathrm{cov}(y_{1,j_1}, y_{1,j_2}) = \mathrm{E}(y_{1,j_1}y_{1,j_2})$. It follows that $\mathrm{E}\{\tilde{\mathcal{U}}(a)\} = 0$ and $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \mathrm{E}\{\tilde{\mathcal{U}}^2(a)\}$. By definition,

$$\tilde{\mathcal{U}}(a) = (P_a^{n_x}P_a^{n_y})^{-1} \sum_{1\le j_1,j_2\le p} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n_x,a);\\ \mathbf{w}\in\mathcal{P}(n_y,a)}} \mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2),$$

where we define $\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2) = \prod_{t=1}^a (x_{i_t,j_1}x_{i_t,j_2} - y_{w_t,j_1}y_{w_t,j_2})$. Then

$$\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \frac{1}{(P_a^{n_x}P_a^{n_y})^2} \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a);\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \mathrm{E}\Big\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}}, \tilde{\mathbf{w}}, j_3, j_4)\Big\}.$$

Under H_0, $\Sigma_x = \Sigma_y = \Sigma = (\sigma_{j_1,j_2})_{p\times p}$, then $\mathrm{E}(x_{1,j_1}x_{1,j_2} - \sigma_{j_1,j_2}) = 0$ and $\mathrm{E}(y_{1,j_1}y_{1,j_2} - \sigma_{j_1,j_2}) = 0$. If $|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}| + |\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}| < a$, it means that the common indexes between $(\mathbf{i}, \mathbf{w})$ and $(\tilde{\mathbf{i}}, \tilde{\mathbf{w}})$ is smaller than a, then we know $\mathrm{E}\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}}, \tilde{\mathbf{w}}, j_3, j_4)\} = 0$. If $|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}| + |\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}| \ge$

a, we know $\mathrm{E}\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},j_3,j_4)\}$ is a linear combination of $(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}$, where $a-|\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}|\le m\le|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}|$ and

$$\begin{aligned}\mathbf{X}_{j_1,j_2,j_3,j_4} &= \mathrm{E}\{(x_{1,j_1}x_{1,j_2}-\sigma_{j_1,j_2})(x_{1,j_3}x_{1,j_4}-\sigma_{j_3,j_4})\},\\ \mathbf{Y}_{j_1,j_2,j_3,j_4} &= \mathrm{E}\{(y_{1,j_1}y_{1,j_2}-\sigma_{j_1,j_2})(y_{1,j_3}y_{1,j_4}-\sigma_{j_3,j_4})\}.\end{aligned}$$

And if $|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}|+|\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}|=t_0$,

$$\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a);\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}}\mathbf{1}_{\{|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}|+|\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}|=t_0\}}=O(n^{4a-t_0}),$$

which achieves the largest order at $t_0=a$ when $t_0\ge a$. Therefore,

$$\begin{aligned}\mathrm{var}\{\tilde{\mathcal{U}}(a)\} \simeq\ & \frac{1}{(P_a^{n_x}P_a^{n_y})^2}\sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a);\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}}\mathbf{1}_{\{|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}|+|\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}|=a\}}\\ & \times\mathrm{E}\Big\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},j_3,j_4)\Big\}.\end{aligned}$$

It follows that

$$\begin{aligned}&\mathrm{var}\{\tilde{\mathcal{U}}(a)\}\\ \simeq\ &\sum_{1\le j_1,j_2,j_3,j_4\le p}\sum_{m=0}^{a}\frac{P_{2a-m}^{n_x}P_{a+m}^{n_y}}{(P_a^{n_x}P_a^{n_y})^2}\binom{a}{m}^2\binom{a-m}{a-m}^2\\ &\times m!(a-m)!(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m},\end{aligned} \tag{B.175}$$

and then (B.175) $\simeq\sum_{1\le j_1,j_2,j_3,j_4\le p}a!(\mathbf{X}_{j_1,j_2,j_3,j_4}/n_x+\mathbf{Y}_{j_1,j_2,j_3,j_4}/n_y)^a$.

We next prove $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ under Conditions A.5 and A.6 in the following Sections B.9.1 and B.9.1 respectively.

Under Condition A.5. To prove $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ under Condition A.5, we will first show $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=\Theta(p^2n^{-a})$. Note that $P_{2a-m}^{n_x}P_{a+m}^{n_y}/(P_a^{n_x}P_a^{n_y})^2\simeq Cn^a$. By (B.175), it remains to show that for any $m\in\{0,1,\ldots,a\}$,

$$\sum_{1\le j_1,j_2,j_3,j_4\le p}(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}=\Theta(p^2). \tag{B.176}$$

We next prove (B.176) by discussing different cases of $\{j_1,j_2,j_3,j_4\}$, and using $K_0=-(2+\epsilon)(8+2\mu)(\log p)/(\epsilon\log\delta)$ similarly to (B.46), where ϵ and μ are positive constants and $\delta=\max\{\delta_x,\delta_y\}$ from Condition A.5.

Case 1: If $|j_1 - j_2| \le K_0$ and $|j_3 - j_4| \le K_0$, we define a distance $\kappa_d = \min\{|j_1-j_3|, |j_1-j_4|, |j_2-j_3|, |j_2-j_4|\}$, and discuss when $\kappa_d > K_0$ and $\kappa_d \le K_0$ respectively. For the simplicity of notation, define two indicator functions $I_1 = \mathbf{1}_{\{|j_1-j_2|\le K_0, |j_3-j_4|\le K_0, \kappa_d > K_0\}}$ and $I_2 = \mathbf{1}_{\{|j_1-j_2|\le K_0, |j_3-j_4|\le K_0, \kappa_d \le K_0\}}$. By definition, we have $\mathbf{X}_{j_1,j_2,j_3,j_4} = \text{cov}(x_{1,j_1}x_{1,j_2}, x_{1,j_3}x_{1,j_4})$ and $\mathbf{Y}_{j_1,j_2,j_3,j_4} = \text{cov}(y_{1,j_1}y_{1,j_2}, y_{1,j_3}y_{1,j_4})$. When $\kappa_d > K_0$, we know $\mathbf{X}_{j_1,j_2,j_3,j_4} \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}}$ by Condition A.5 (2) and (3) and Lemma B.1. It follows that

$$\begin{aligned}\text{(B.177)} \qquad &\Big| \sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_1 \Big| \\ &\le C p^4 \delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(1) p^4 \times p^{-(8+2\mu)} = o(1).\end{aligned}$$

In addition, note that $\sum_{1\le j_1,j_2,j_3,j_4\le p} I_2 = O(pK_0^3) = O(p\log^3 p)$. By Condition A.5 (2), we know

$$\Big| \sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_2 \Big| = O(p\log^3 p).$$

Case 2: If $|j_1 - j_2| > K_0$ or $|j_3 - j_4| > K_0$, by Lemma B.1, we know that $|\sigma_{j_1,j_2}\sigma_{j_3,j_4}| \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}}$. We consider $|j_1 - j_2| > K_0$ without loss of generality and discuss the following cases (i)–(iv).

(i) When $|j_2 - j_3| > K_0/2$ and $|j_2 - j_4| > K_0/2$,

$$|\mathbf{X}_{j_1,j_2,j_3,j_4}| = |\text{cov}(x_{1,j_1}x_{1,j_3}x_{1,j_4}\,,\, x_{1,j_2}) - \sigma_{j_1,j_2}\sigma_{j_3,j_4}| \le C\delta^{\frac{K_0\epsilon}{2(2+\epsilon)}}.$$

(ii) When $|j_2 - j_3| \le K_0/2$ and $|j_2 - j_4| \le K_0/2$, we know that $|j_1 - j_3| \ge |j_1 - j_2| - |j_2 - j_3| > K_0/2$ and $|j_1 - j_4| \ge |j_1 - j_2| - |j_2 - j_4| > K_0/2$. Then

$$\text{(B.178)} \quad |\mathbf{X}_{j_1,j_2,j_3,j_4}| = |\text{cov}(x_{1,j_1}\,,\, x_{1,j_2}x_{1,j_3}x_{1,j_4}) - \sigma_{j_1,j_2}\sigma_{j_3,j_4}| \le C\delta^{\frac{K_0\epsilon}{2(2+\epsilon)}}.$$

(iii) When $|j_2 - j_3| \le K_0/2$ and $|j_2 - j_4| > K_0/2$, as we know $|j_1 - j_2| > K_0$, then $|j_1 - j_3| > K_0/2$. We next discuss three sub-cases.

(iiia) If $|j_1 - j_4| > K_0/2$, we know (B.178) also holds.

For easy presentation, let I_3 be an indicator function when $\{j_1, j_2, j_3, j_4\}$ satisfies the sub-cases (i), (ii) and (iiia) above. Then similarly to (B.177),

$$\Big| \sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_3 \Big| = o(1).$$

(iiib) If $|j_1 - j_4| \le K_0/2$, and $|j_3 - j_4| \le K_0/2$, we know under this case $|j_2 - j_3|, |j_1 - j_4|, |j_3 - j_4| \le K_0$. Let $I_4 = \mathbf{1}_{\{|j_2-j_3|,|j_1-j_4|,|j_3-j_4|\le K_0\}}$. We have $\sum_{1\le j_1,j_2,j_3,j_4\le p} I_4 = O(pK_0^3)$. By Condition A.5 (2), we know

$$\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_4\Big| = O(p\log^3 p).$$

(iiic) If $|j_1 - j_4| \le K_0/2$, and $|j_3 - j_4| > K_0/2$, we know

$$\mathbf{X}_{j_1,j_2,j_3,j_4} \ge \mathrm{E}(x_{1,j_1}x_{1,j_4})\mathrm{E}(x_{1,j_2}x_{1,j_3}) - C\delta^{\frac{K_0\epsilon}{2(2+\epsilon)}}.$$

Let I_5 be an indicator function of the sub-case (iiic) above. Then

$$\begin{aligned}
&\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_5\Big| \\
&= \Big|\sum_{1\le j_1,j_2,j_3,j_4\le p} (\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a \times I_5\Big| + O(p^4p^{-(4+\mu)}) \\
&= \Big|\sum_{|j_1-j_4|\le K_0/2,\, |j_2-j_3|\le K_0/2} (\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a\Big| + o(1) \\
&= \Big|\sum_{1\le j_1,j_2,j_3,j_4\le p} (\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a - \sum_{|j_1-j_4|>K_0 \text{ or } |j_2-j_3|>K_0} (\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a\Big| + o(1) \\
&= \Theta(p^2).
\end{aligned}$$

where the last equation uses Conditions A.5 (3) and (4) and Lemma B.1.

(iv) When $|j_2 - j_3| > K_0/2$ and $|j_2 - j_4| \le K_0/2$, this is symmetric to the sub-case (iii) discussed above. Define an indicator function $I_6 = \mathbf{1}_{\{|j_2-j_3|>K_0/2,|j_2-j_4|\le K_0/2\}}$. We then have

$$\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_6\Big| = \Theta(p^2).$$

In summary, (B.176) is proved and thus $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \Theta(p^2n^{-a})$ is obtained. To prove $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$, it remains to show that $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(p^2n^{-a})$.

We write $\mathcal{U}(a) = \sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c} C_{a,c,b_1,b_2}T_{b_1,b_2,c}$, where we define

$C_{a,c,b_1,b_2} = (-1)^{c-b_1+b_2} a!/\{b_1!b_2!(c-b_1)!(a-c-b_2)!\}$, and

$$T_{b_1,b_2,c} = \sum_{1\le j_1,j_2\le p} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\in\mathcal{P}(n_y,2(a-c)-b_2)}} (P^{n_x}_{2c-b_1} P^{n_y}_{2(a-c)-b_2})^{-1} \times \prod_{k=1}^{b_1}(x_{i_k,j_1}x_{i_k,j_2} - \sigma_{j_1,j_2}) \prod_{k=b_1+1}^{c} x_{i_k,j_1} \prod_{k=c+1}^{2c-b_1} x_{i_k,j_2} \times \prod_{m=1}^{b_2}(y_{w_m,j_1}y_{w_m,j_2} - \sigma_{j_1,j_2}) \prod_{l=b_2+1}^{a-c} y_{w_l,j_1} \prod_{q=a-c+1}^{2(a-c)-b_2} y_{w_q,j_2}. \tag{B.179}$$

Then $\tilde{\mathcal{U}}(a) = \sum_{c=0}^{a}(-1)^{a-c}T_{c,a-c,c}$ and $\tilde{\mathcal{U}}^*(a) = \sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c} C_{a,c,b_1,b_2} \times T_{b_1,b_2,c}\mathbf{1}_{b_1+b_2\le a-1}$. Note that $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} \le C \max_{b_1,b_2,c;b_1+b_2\le a-1}\{\mathrm{var}(T_{b_1,b_2,c})\}$, where C is some constant. When a is finite, to prove $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(p^2n^{-a})$, it suffices to show that $\mathrm{var}(T_{b_1,b_2,c}) = o(p^2n^{-a})$ for each (b_1,b_2,c) satisfying $b_1+b_2 \le a-1$. Note that $\mathrm{E}(T_{b_1,b_2,c}) = 0$ under H_0, then $\mathrm{var}(T_{b_1,b_2,c}) = \mathrm{E}(T^2_{b_1,b_2,c})$ and

$$\mathrm{var}(T_{b_1,b_2,c}) = (P^{n_x}_{2c-b_1} P^{n_y}_{2(a-c)-b_2})^{-2} \sum_{\substack{1\le j_1,j_2\le p;\\ 1\le \tilde{j}_1,\tilde{j}_2\le p}} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,2(a-c)-b_2)}} \mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2), \tag{B.180}$$

where we let

$$\begin{aligned} &\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2) \\ &= \mathrm{E}\Big\{\prod_{k=1}^{b_1}(x_{i_k,j_1}x_{i_k,j_2} - \sigma_{j_1,j_2})(x_{\tilde{i}_k,\tilde{j}_1}x_{\tilde{i}_k,\tilde{j}_2} - \sigma_{\tilde{j}_1,\tilde{j}_2}) \prod_{k=b_1+1}^{c}(x_{i_k,j_1}x_{\tilde{i}_k,\tilde{j}_1}) \\ &\quad \times \prod_{k=c+1}^{2c-b_1}(x_{i_k,j_2}x_{\tilde{i}_k,\tilde{j}_2})\Big\}\mathrm{E}\Big\{\prod_{m=1}^{b_2}(y_{w_m,j_1}y_{w_m,j_2} - \sigma_{j_1,j_2})(y_{\tilde{w}_m,\tilde{j}_1}y_{\tilde{w}_m,\tilde{j}_2} - \sigma_{j_1,j_2}) \\ &\quad \times \prod_{m=b_2+1}^{a-c}(y_{w_m,j_1}y_{\tilde{w}_m,\tilde{j}_1}) \prod_{m=a-c+1}^{2(a-c)-b_2}(y_{w_m,j_2}y_{\tilde{w}_m,\tilde{j}_2})\Big\}. \end{aligned}$$

Since we assume without loss of generality that $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = \mathbf{0}$, then $\mathrm{E}(x_{1,j_1}x_{1,j_2} - \sigma_{j_1,j_2}) = \mathrm{E}(y_{1,j_1}x_{1,j_2} - \sigma_{j_1,j_2}) = 0$. It follows that when $\{\mathbf{i}\} \ne \{\tilde{\mathbf{i}}\}$ or $\{\mathbf{w}\} \ne \{\tilde{\mathbf{w}}\}$, $\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2) = 0$. When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{w}\} = \{\tilde{\mathbf{w}}\}$, we have $|\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}| + |\{\mathbf{w}\}\cup\{\tilde{\mathbf{w}}\}| = 2c - b_1 + 2(a-c) - b_2$. By Condition

A.5 (1) and (2), for any given $\{j_1, j_2, \tilde{j}_1, \tilde{j}_2\}$,

$$(B.181)\quad (P^{n_x}_{2c-b_1} P^{n_y}_{2(a-c)-b_2})^{-2} \sum_{\substack{\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n_x, 2c-b_1); \\ \mathbf{w}\, \tilde{\mathbf{w}} \in \mathcal{P}(n_y, 2(a-c)-b_2)}} \mathbb{T}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, j_1, j_2, \tilde{j}_1, \tilde{j}_2)$$
$$= O(n^{-2(2a+b_1+b_2)} \times n^{2a-b_1-b_2}) = O(n^{-2a+b_1+b_2}) = o(n^{-a-1})$$

where in the last equation, we use $b_1 + b_2 \leq a - 1$. In addition, similarly to (B.176), we have that for any given $(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}})$,

$$(B.182)\quad \sum_{1 \leq j_1, j_2, \tilde{j}_1, \tilde{j}_2 \leq p} \mathbb{T}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, j_1, j_2, \tilde{j}_1, \tilde{j}_2) = O(p^2).$$

In summary, by (B.181) and (B.182), we know $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = O(p^2 n^{-a-1}) = o(p^2 n^{-a})$.

Under Condition A.6. In this section, we prove that $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ under Condition A.6. Recall that we have already obtained $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ in (B.175). By Condition A.6 (3), we have

$$(B.183)\quad \begin{aligned} \mathbf{X}_{j_1,j_2,j_3,j_4} &= \kappa_x(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}) + (\kappa_x - 1)\sigma_{j_1,j_2}\sigma_{j_3,j_4}, \\ \mathbf{Y}_{j_1,j_2,j_3,j_4} &= \kappa_y(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}) + (\kappa_y - 1)\sigma_{j_1,j_2}\sigma_{j_3,j_4}. \end{aligned}$$

Then by Condition A.6 (1) and (4), we know $(\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}$ is a linear combination of

$$(B.184)\quad \prod_{t=1}^{a} \Big\{ \sigma_{j_{g_1^{(t)}}, j_{g_2^{(t)}}} \times \sigma_{j_{g_3^{(t)}}, j_{g_4^{(t)}}} \Big\},$$

where $\{(g_1^{(t)}, g_2^{(t)}), (g_3^{(t)}, g_4^{(t)}) : t = 1, \ldots, a\}$ are a allocations of the set $\{1, 2, 3, 4\}$ into 2 (unordered) pairs. When the a allocations are the same, by the symmetricity of j indexes,

$$\sum_{1 \leq j_1, j_2, j_3, j_4 \leq p} \prod_{t=1}^{a} \sigma_{j_{g_1^{(t)}}, j_{g_2^{(t)}}} \sigma_{j_{g_3^{(t)}}, j_{g_4^{(t)}}} = \sum_{1 \leq j_1, j_2, j_3, j_4 \leq p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a.$$

When the a allocations are different, by Condition A.6 (4),

$$(B.185)\quad \sum_{1 \leq j_1, j_2, j_3, j_4 \leq p} \prod_{t=1}^{a} \sigma_{j_{g_1^{(t)}}, j_{g_2^{(t)}}} \sigma_{j_{g_3^{(t)}}, j_{g_4^{(t)}}} = o(1) \sum_{1 \leq j_1, j_2, j_3, j_4 \leq p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a,$$

which can be obtained by taking square of both sides of (B.185) and using Condition A.6 (4). It follows that by (B.175), Condition A.6 (1) and (4) and the symmetricity of j indexes,

$$\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \Theta(n^{-a}) \sum_{1\le j_1,j_2,j_3,j_4\le p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a. \tag{B.186}$$

We next show $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$. Similarly to Section B.9.1, we know it suffices to prove $\mathrm{var}(T_{b_1,b_2,c}) = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ for $0 \le c \le a$, $0 \le b_1 \le c$, $0 \le b_2 \le a-c$ and $b_1 + b_2 \le a-1$. Note that (B.180) still holds here, and when $\{\mathbf{i}\} \ne \{\tilde{\mathbf{i}}\}$ or $\{\mathbf{w}\} \ne \{\tilde{\mathbf{w}}\}$, $\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2) = 0$. Therefore, (B.181) also holds. By Condition A.6 (3) and (4), similarly to the analysis of (B.186), we have for any given $(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}})$,

$$\begin{aligned}\sum_{1\le j_1,j_2,j_3,j_4\le p} & \mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2) \\ &= O(1) \sum_{1\le j_1,j_2,j_3,j_4\le p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a.\end{aligned} \tag{B.187}$$

Combining (B.181) and (B.187),

$$\mathrm{var}(T_{b_1,b_2,c}) = O(n^{-a-1}) \sum_{1\le j_1,j_2,j_3,j_4\le p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}.$$

B.9.2. *Proof of Lemma A.23 (on Page 61, Section A.13).* Since $\mathrm{E}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}(b)\} = 0$ under H_0, $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\}/\{\sigma(a)\sigma(b)\}$. Recall that $\mathcal{U}(a) = \tilde{\mathcal{U}}(a) + \tilde{\mathcal{U}}^*(a)$ and $\mathcal{U}(b) = \tilde{\mathcal{U}}(b) + \tilde{\mathcal{U}}^*(b)$. Then

$$\begin{aligned}\mathrm{E}\Big\{\frac{\mathcal{U}(a)}{\sigma(a)} \times \frac{\mathcal{U}(b)}{\sigma(b)}\Big\} &= \mathrm{E}\Big\{\frac{\tilde{\mathcal{U}}(a) + \tilde{\mathcal{U}}^*(a)}{\sigma(a)} \times \frac{\tilde{\mathcal{U}}(b) + \tilde{\mathcal{U}}^*(b)}{\sigma(b)}\Big\} \\ &= \mathrm{E}\Big\{\frac{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)}{\sigma(a)\sigma(b)}\Big\} + o(1),\end{aligned} \tag{B.188}$$

where the last equation follows by Lemma A.22. By the definition and notation in Section B.9.1,

$$\tilde{\mathcal{U}}(a) = \tilde{C}_a \sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,a);\\ \mathbf{w}\in\mathcal{P}(n_y,a)}} \mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2), \quad \tilde{\mathcal{U}}(b) = \tilde{C}_b \sum_{\substack{1\le \tilde{j}_1,\tilde{j}_2\le p;\\ \tilde{\mathbf{i}}\in\mathcal{P}(n_x,b);\\ \tilde{\mathbf{w}}\in\mathcal{P}(n_y,b)}} \mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2),$$

where we let $\tilde{C}_a = (P^{n_x}_a P^{n_y}_a)^{-1}$, $\tilde{C}_b = (P^{n_x}_b P^{n_y}_b)^{-1}$, $\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2) = \prod_{t=1}^a (x_{i_t,j_1}x_{i_t,j_2} - y_{w_t,j_1}y_{w_t,j_2})$ and $\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2) = \prod_{t=1}^b (x_{\tilde{i}_t,\tilde{j}_1}x_{\tilde{i}_t,\tilde{j}_2} -$

$y_{\tilde{w}_t,\tilde{j}_1}y_{\tilde{w}_t,\tilde{j}_2}$). It follows that

$$\mathrm{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\} = \tilde{C}_a\tilde{C}_b \sum_{\substack{1\le j_1,j_2,\tilde{j}_1,\tilde{j}_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,a);\,\tilde{\mathbf{i}}\in\mathcal{P}(n_x,b)\\ \mathbf{w}\in\mathcal{P}(n_y,a);\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,b)}} \mathrm{E}\Big\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2)\Big\}.$$

As $a \neq b$, we know $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{w}\} \neq \{\tilde{\mathbf{w}}\}$. It follows that similarly to Section B.1.2, $\mathrm{E}\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2)\} = 0$. Therefore $\mathrm{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\} = 0$ and $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a),\mathcal{U}(b)/\sigma(b)\} = o(1)$.

B.9.3. *Derivation of $D_{n,k}$ and $\pi_{n,k}^2$.* To prove Lemmas A.24 and A.25, we derive the forms of $D_{n,k}$ and $\pi_{n,k}^2$ in this section. By construction, $D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$, where $A_{n,k,a_r} = (\mathrm{E}_k - \mathrm{E}_{k-1})[\tilde{\mathcal{U}}(a_r)/\sigma(a_r)]$. In addition, $\pi_{n,k}^2 = \sum_{1\le r_1,r_2\le m} t_{r_1}t_{r_2}\mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})$. It then suffices to derive the form of $A_{n,k,a}$ for a given integer a, and also derive $\mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$ for two given integers a_1 and a_2.

For easy presentation, we define $\mathcal{X}_{i,j_1,j_2} = x_{i,j_1}x_{i_t,j_2} - \sigma_{j_1,j_2}$ and $\mathcal{Y}_{i,j_1,j_2} = y_{i,j_1}y_{i_t,j_2} - \sigma_{j_1,j_2}$ in the following. Then under H_0,

$$\tilde{\mathcal{U}}(a) = (P_a^{n_x}P_a^{n_y})^{-1} \sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,a);\,\mathbf{w}\in\mathcal{P}(n_y,a)}} \prod_{t=1}^{a}(\mathcal{X}_{w_t,j_1,j_2} - \mathcal{Y}_{i_t,j_1,j_2}).$$

Part I: $1 \le k \le n_x$. When $1 \le k \le n_x$, similarly to Section B.1.4, as $\mathrm{E}(\mathcal{X}_{1,j_1,j_2}) = 0$ under H_0, we have

$$(\mathrm{E}_k - \mathrm{E}_{k-1})\Big\{\prod_{t=1}^{a}(\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2})\Big\} = (\mathrm{E}_k - \mathrm{E}_{k-1})\Big(\prod_{t=1}^{a}\mathcal{X}_{i_t,j_1,j_2}\Big),$$

which is nonzero only when $i_1,\ldots,i_a \le k$ and $k \in \{i_1,\ldots,i_a\}$. Then we know when $k < a$, $A_{n,k,a} = 0$ and when $k \ge a$,

$$A_{n,k,a} = c_1(n,a) \sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(k-1,a-1)}} \Big(\prod_{t=1}^{a-1}\mathcal{X}_{i_t,j_1,j_2}\Big)\mathcal{X}_{k,j_1,j_2}, \tag{B.189}$$

where $c_1(n,a) = a!/\{P_a^{n_x}\sigma(a)\}$. For two integers a_1 and a_2,

$$\begin{aligned}&\mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})\\ &= \prod_{l=1}^{2}c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2}} \mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2),\end{aligned}$$

where

$$\begin{aligned}&\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\\&=\prod_{l=1}^{2}\Big(\prod_{t=1}^{a_l-1}\mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}}\Big)\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4}).\end{aligned}$$

Part II: $n_x+1\le k\le n_x+n_y$. When $n_x+1\le k\le n_x+n_y$, we have

$$\prod_{t=1}^{a}(\mathcal{X}_{i_t,j_1,j_2}-\mathcal{Y}_{i_t,j_1,j_2})=\sum_{s=0}^{a}\sum_{\substack{\mathbf{i}^*\in\mathcal{S}(\mathbf{i},s);\\ \mathbf{w}^*\in\mathcal{S}(\mathbf{w},a-s)}}\Big(\prod_{t=1}^{s}\mathcal{X}_{i_t^*,j_1,j_2}\Big)\Big(\prod_{\tilde{t}=1}^{a-s}\mathcal{Y}_{w_{\tilde{t}}^*,j_1,j_2}\Big),$$

where $\mathcal{S}(\mathbf{i},s)$ represents the collection of sub-tuples of $\mathbf{i}$ with length s and $\mathcal{S}(\mathbf{w},a-s)$ represents the collection of sub-tuples of $\mathbf{w}$ with length $a-s$, which is similarly used in Section B.3.1. When $n_x+1\le k\le n_x+n_y$, similarly to Section B.1.4, $(\mathrm{E}_k-\mathrm{E}_{k-1})\{\prod_{t=1}^{s}(x_{i_t^*,j_1}x_{i_t^*,j_2}-\sigma_{j_1,j_2})\prod_{\tilde{t}=1}^{a-s}(y_{w_{\tilde{t}}^*,j_1}y_{w_{\tilde{t}}^*,j_2}-\sigma_{j_1,j_2})\}\neq 0$ only when $w_1^*,\dots,w_{a-s}^*\le k-n_x$ and $k-n_x\in\{w_1^*,\dots,w_{a-s}^*\}$, and then

$$(\mathrm{E}_k-\mathrm{E}_{k-1})\Big(\prod_{t=1}^{s}\mathcal{X}_{i_t^*,j_1,j_2}\prod_{\tilde{t}=1}^{a-s}\mathcal{Y}_{w_{\tilde{t}}^*,j_1,j_2}\Big)=\mathcal{Y}_{k-n_x,j_1,j_2}\prod_{t=1}^{s}\mathcal{X}_{i_t^*,j_1,j_2}\prod_{\tilde{t}=1}^{a-s-1}\mathcal{Y}_{w_{\tilde{t}}^*,j_1,j_2}.$$

It follows that

$$A_{n,k,a}=\sum_{s=L_k}^{a-1}\sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,s);\\ \mathbf{w}\in\mathcal{P}(k-n_x-1,a-s-1)}}c_2(n,a,s)\mathcal{Y}_{k-n_x,j_1,j_2}\prod_{t=1}^{s}\mathcal{X}_{i_t,j_1,j_2}\prod_{\tilde{t}=1}^{a-s-1}\mathcal{Y}_{w_{\tilde{t}},j_1,j_2},$$

where $L_k=\max\{n_x-k+a,0\}$ and $c_2(n,a,s)=P_{a-s}^{n_x-s}P_s^{n_y-a+s}\{P_a^{n_x}P_a^{n_y}\sigma(a)\}^{-1}$. Thus for two constants a_1 and a_2,

$$\begin{aligned}&\mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})\\&=\sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ L_k\le s_l\le a_l:\,l=1,2;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\,l=1,2;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\,l=1,2}}\prod_{l=1}^{2}c_2(n,a_l,s_l)\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2),\end{aligned}$$

where

$$\begin{aligned}&\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\\&=\prod_{l=1}^{2}\Big(\prod_{t=1}^{s_l}\mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}}\prod_{\tilde{t}=1}^{a_l-s_l-1}\mathcal{Y}_{w_{\tilde{t}}^{(l)},j_{2l-1},j_{2l}}\Big)\mathrm{E}(\mathcal{Y}_{k-n_x,j_1,j_2}\mathcal{Y}_{k-n_x,j_3,j_4}).\end{aligned}$$

B.9.4. *Proof of Lemma A.24 (on Page 62, Section A.13).* Note that by the Cauchy-Schwarz inequality, for some constant C,

$$\mathrm{var}\Big(\sum_{k=1}^{n} \pi_{n,k}^2\Big) \leq Cn^2 \max_{1\leq k\leq n;\, 1\leq r_1,r_2\leq m} \mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}),$$

where for two integers a_1 and a_2, $\mathbb{T}_{k,a_1,a_2} = \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$ is given in Section B.9.3. Therefore to prove Lemma A.24, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}) = o(n^{-2})$ for every $1 \leq k \leq n$ and $1 \leq r_1, r_2 \leq m$. We next prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$ when $a \leq k \leq n_x$ and $n_x + 1 \leq k \leq n_x + n_y$ in the following Parts I and II respectively.

Part I: $a \leq k \leq n_x$. We first derive the form of $\mathrm{var}(\mathbb{T}_{k,a_1,a_2})$ when $a \leq k \leq n_x$. As $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2$, we next derive $\mathrm{E}(\mathbb{T}_{k,a_1,a_2})$ and $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2})$. In particular,

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = \prod_{l=1}^{2} c(n,a_l) \sum_{\substack{1\leq j_1,j_2,j_3,j_4\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2}} \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2)\Big\}.$$

For easy presentation, we let $a_3 = a_1$ and $a_4 = a_2$, and have

$$\begin{aligned}&\Big\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\Big\}^2\\ =& \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\leq j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2)\Big\}\\ &\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=3,4)\Big\}.\end{aligned}$$

In addition, we have

$$\begin{aligned}\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) = &\prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\leq j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}}\\ &\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\},\end{aligned}$$

where we define

$$\begin{aligned}&\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\\ =& \prod_{l=1}^{4}\Big(\prod_{t=1}^{a_l-1} \mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}}\Big)\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4})\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8}).\end{aligned}$$

Let $\mathbf{1}_E$ be an indicator function of the event that $(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\})=\emptyset$. Then define

$$G_{a_1,a_2,1} = \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \times \mathbf{1}_E \\ \times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2,3,4)\Big\}.$$

We also note that

$$\begin{aligned}\text{(B.190)}\qquad & \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2,3,4)\Big\}\times\mathbf{1}_E \\ &= \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2)\Big\} \\ &\quad \times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=3,4)\Big\}\times\mathbf{1}_E.\end{aligned}$$

Since $|\mathrm{var}(\mathbb{T}_{k,a_1,a_2})| \le |\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1}| + |\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}|$, to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, we will next show that $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$ and $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1}| = o(n^{-2})$. In particular, we present the proof under Conditions A.5 and A.6 in the following Sections B.9.4 and B.9.4, respectively.

Proof under Condition A.5.

Step I: $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$. If $a_1\ne a_2$, we have $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = G_{a_1,a_2,1} = 0$. It remains to consider $a_1=a_2$ below. Note that

$$\begin{aligned}\text{(B.191)}\qquad & \mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2)\} \\ & \times\mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=3,4)\}\end{aligned}$$

satisfies that (B.191) $\ne 0$ only if $\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}$. Thus,

$$\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}}\right\}} \times \text{(B.191)}.$$

Similarly, $\mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2,3,4)\}\times\mathbf{1}_E \ne 0$ only when $\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}$. Therefore, by (B.190),

$$G_{a_1,a_2,1} = \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\},\\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}=\emptyset}\right\}} \times \text{(B.191)},$$

and then

$$|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| \le \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\},\\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}\neq\emptyset}\right\}} \times |(\text{B.191})|. \tag{B.192}$$

Note that

$$\sum_{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\},\\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}\neq\emptyset}\right\}} = O(n^{a_1+a_2-3}). \tag{B.193}$$

In addition, by Condition A.5 (2),

$$\begin{aligned} &\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p} |(\text{B.191})| \\ &\le C \sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p} \Big|\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4})\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8})\Big|. \end{aligned} \tag{B.194}$$

Recall that $\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4}) = \mathbf{X}_{j_1,j_2,j_3,j_4}$ and $\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8}) = \mathbf{X}_{j_5,j_6,j_7,j_8}$ following the notation in Section B.9.1. Following the similar analysis for the proof of (B.176), we obtain $\sum_{1\le j_1,j_2,j_3,j_4\le p} |\mathbf{X}_{j_1,j_2,j_3,j_4}| = O(p^2)$ and $\sum_{1\le j_5,j_6,j_7,j_8\le p} |\mathbf{X}_{j_5,j_6,j_7,j_8}| = O(p^2)$. It follows that (B.194) $= O(p^4)$. Note that $c(n,a) = \Theta(p^{-1}n^{-a/2})$ by Lemma A.22. Combining (B.193) and (B.194), we obtain $\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1} = o(n^{-2})$.

Step II: $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1}| = o(n^{-2})$. By construction, we have

$$\begin{aligned} \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1} = \prod_{l=1}^{4} c(n,a_l) &\sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} (1-\mathbf{1}_E) \\ &\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2,3,4)\Big\}. \end{aligned} \tag{B.195}$$

When $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| > a_1+a_2-2$, which means that there exists one index that only appears once among the four sets $\{\mathbf{i}^{(l)}\}$, $l=1,2,3,4$, then similarly to Section B.1.5,

$$\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2)\Big\} \times (1-\mathbf{1}_E) \tag{B.196}$$

satisfies that (B.196) $=0$. When $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|<a_1+a_2-2$,

$$\sum_{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2,3,4}\mathbf{1}_{\{|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|<a_1+a_2-2\}}=O(n^{a_1+a_2-3}).\tag{B.197}$$

Similarly to the analysis of (B.194) above, by Condition A.5, we have

$$\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p}(\text{B.196})=O(p^4).\tag{B.198}$$

Therefore, by (B.197), (B.198) and $c(n,a)=\Theta(p^{-1}n^{-a/2})$,

$$\begin{aligned}&\prod_{l=1}^{4}c(n,a_l)\sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2,3,4}}(1-\mathbf{1}_E)\mathbf{1}_{\{|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|<a_1+a_2-2\}}\\ &\qquad\qquad\times\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\Big\}\\ =\;&O(1)n^{-a_1-a_2}p^{-4}n^{a_1+a_2-3}p^4=o(n^{-2}).\end{aligned}$$

Last, we consider $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|=a_1+a_2-2$. Note that $1-\mathbf{1}_E\neq 0$ indicates that $(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\})\neq\emptyset$ under this case. By the symmetricity of the j indexes, we have

$$\begin{aligned}&\Big|\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p}(\text{B.196})\Big|\\ &\le C\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p}\Big|\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4})\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8})\\ &\qquad\times\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_5,j_6})\mathrm{E}(\mathcal{X}_{k,j_3,j_4}\mathcal{X}_{k,j_7,j_8})\Big|.\end{aligned}\tag{B.199}$$

Following similar arguments to that in Sections B.1.5 and B.9.1, by discussing different cases of j indexes, we have (B.199) $=o(p^4)$. Thus,

$$\begin{aligned}&\prod_{l=1}^{4}c(n,a_l)\sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2,3,4}}(1-\mathbf{1}_E)\mathbf{1}_{\{|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}|=a_1+a_2-2\}}\\ &\qquad\qquad\times\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\Big\}\\ =\;&o(1)n^{-a_1-a_2}p^{-4}n^{a_1+a_2-2}p^4=o(n^{-2}).\end{aligned}$$

In summary, we obtain $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2})-G_{a_1,a_2,1}=o(n^{-2})$.

Proof under Condition A.6. Similarly to Section B.9.4, we next prove $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$ and $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1}| = o(n^{-2})$.

Step I: $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$. Following the same analysis in Section B.9.4, we obtain (B.192) and (B.193). By Condition A.6 (2) and (4), we have

$$\begin{aligned}(\text{B.200})\qquad & \sum_{1\le j_1,j_2,j_3,j_4 j_5,j_6,j_7,j_8\le p} (\text{B.191}) \\ &= O(1)\Big\{\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^{a_1}\Big\}\Big\{\sum_{1\le j_5,j_6,j_7,j_8\le p}(\sigma_{j_5,j_6}\sigma_{j_7,j_8})^{a_2}\Big\}.\end{aligned}$$

Note that $\sigma^2(a) = \Theta(n^{-a}) \times \sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a$ by Lemma A.22, and $c(n,a) = \Theta(1)\{n^a\sigma(a)\}^{-1}$. Combining (B.193) and (B.200), we have $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2}| = o(n^{-2})$.

Step II: $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1}| = o(n^{-2})$. Similarly to Section B.9.4, we have (B.195) and $\mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\} \neq 0$ only when $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| \le a_1+a_2-2$.

When $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| < a_1+a_2-2$, (B.197) still holds. By Condition A.6 (2) and (4), similarly to (B.200), we have

$$\begin{aligned}&\sum_{1\le j_1,j_2,j_3,j_4 j_5,j_6,j_7,j_8\le p} \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\} \\ &= O(1)\Big\{\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^{a_1}\Big\}\Big\{\sum_{1\le j_5,j_6,j_7,j_8\le p}(\sigma_{j_5,j_6}\sigma_{j_7,j_8})^{a_2}\Big\}.\end{aligned}$$

Note that $\sigma^2(a) = \Theta(n^{-a}) \times \sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a$ by Lemma A.22, and $c(n,a) = \Theta(1)\{n^a\sigma(a)\}^{-1}$. Then we have

$$\begin{aligned}&\prod_{l=1}^4 c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathbf{1}_{\{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|<a_1+a_2-2\}} \\ &\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\} = o(n^{-2}).\end{aligned}$$

When $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| = a_1+a_2-2$, by the construction of $\mathbf{1}_E$, we know

$$(\text{B.201})\qquad \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\} \times (1-\mathbf{1}_E)$$

satisfies that (B.201) $\neq 0$ if $(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\}) \cap (\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\}) \neq \emptyset$. Then by Condition A.6 (3) and (4), we know (B.201) is a linear combination of $\sum_{1\le j_1,\dots,j_8\le p}\prod_{t=1}^{a+b}\sigma_{j_{g_{2t-1}},j_{g_{2t}}}$ with $S_{\mathcal{G}} > 4$, where we recall that $S_{\mathcal{G}}$ is the number of distinct sets among $\{g_{2t-1}, g_{2t}\}, t = 1, \dots, a+b$, induced by $\mathcal{G} = (g_1, \dots, g_{2(a+b)})$. Therefore,

$$\begin{aligned}
&\prod_{l=1}^{4} c(n, a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} (1-\mathbf{1}_E)\times \mathbf{1}_{\{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|=a_1+a_2-2\}}\\
&\quad\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\Big\}\\
&\le C\Big\{\prod_{l=1}^{4} c(n, a_l)\Big\}\times n^{a_1+a_2-2}\sum_{\mathcal{G}:S_{\mathcal{G}}>4}\Big|\sum_{1\le j_1,\dots,j_8\le p}\prod_{t=1}^{a+b}\sigma_{j_{g_{2t-1}},j_{g_{2t}}}\Big|\\
&= o(n^{-2}).
\end{aligned}$$

where the last equation follows by Condition A.6 (4), $\sigma^2(a) = \Theta(n^{-a}) \times \sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a$, and $c(n, a) = \Theta(1)\{n^a\sigma(a)\}^{-1}$. In summary, we obtain $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1} = o(n^{-2})$.

Part II: $n_x \le k \le n_x + n_y$. In this section, we prove that when $n_x \le k \le n_x + n_y$, $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$. Recall the form derived in Section B.9.3. We have $\mathbb{T}_{k,a_1,a_2} = \sum_{L_1\le s_1\le a_1, L_2\le s_2\le a_2}\mathbb{T}_{k,a_1,a_2,s_1,s_2}$, where

$$\begin{aligned}
\mathbb{T}_{k,a_1,a_2,s_1,s_2} = &\sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\, l=1,2;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\, l=1,2}} \prod_{l=1}^{2} c_2(n, a_l, s_l)\\
&\times \mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2).
\end{aligned}$$

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2}) = o(n^{-2})$. In particular, for easy presentation, we set $a_3 = a_1$, $a_4 = a_2$ $s_3 = s_1$ and $s_4 = s_2$, and then have

$$\begin{aligned}
\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})\}^2 = &\sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\, l=1,2,3,4;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\, l=1,2,3,4}} \prod_{l=1}^{4} c_2(n, a_l, s_l)\\
&\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2)\Big\}\\
&\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 3,4)\Big\}.
\end{aligned}$$

In addition, we have

$$\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,s_1,s_2}) = \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\,l=1,2,3,4;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\,l=1,2,3,4}} \prod_{l=1}^4 c_2(n,a_l,s_l)$$
$$\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\Big\},$$

where we define

$$\begin{aligned}&\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\\ &=\prod_{l=1}^4\Big(\prod_{t=1}^{s_l}\mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}}\prod_{\tilde{t}=1}^{a_l-s_l-1}\mathcal{Y}_{w_{\tilde{t}}^{(l)},j_{2l-1},j_{2l}}\Big)\\ &\quad \mathrm{E}(\mathcal{Y}_{k-n_x,j_1,j_2}\mathcal{Y}_{k-n_x,j_3,j_4})\times\mathrm{E}(\mathcal{Y}_{k-n_x,j_5,j_6}\mathcal{Y}_{k-n_x,j_7,j_8}).\end{aligned}$$

Therefore $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2}) = \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,s_1,s_2}) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})\}^2$ is derived. We note that the form of $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})$ is very similar to the $\mathrm{var}(\mathbb{T}_{k,a_1,a_2})$ in Section B.9.4. In particular, we can write $\mathcal{Z}_{i,j_1,j_2} = \mathcal{X}_{i,j_1,j_2}$ if $i \le n_x$ and $\mathcal{Z}_{i,j_1,j_2} = \mathcal{Y}_{i-n_x,j_1,j_2}$ if $i > n_x$. Then we let $\mathbf{q}^{(l)} = (\mathbf{i}^{(l)}, \tilde{\mathbf{w}}^{(l)})$ to be a joint index tuple of $\mathbf{i}^{(l)}$ and $\mathbf{w}^{(l)}$, where $\tilde{\mathbf{w}}^{(l)}$ is transformed from $\mathbf{w}^{(l)}$ by adding each index with n_x. Also let $\mathbf{1}_{\tilde{E}}$ be an indicator function of the event that $(\{\mathbf{q}^{(1)}\}\cup\{\mathbf{q}^{(2)}\})\cap(\{\mathbf{q}^{(3)}\}\cup\{\mathbf{q}^{(4)}\})=\emptyset$. Then define

$$G_{a_1,a_2,2} = \prod_{l=1}^4 c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\,l=1,2,3,4;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\,l=1,2,3,4}} \times \mathbf{1}_{\tilde{E}}$$
$$\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\Big\}.$$

Similarly to Section B.9.4, we also note that

$$\begin{aligned}&\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\Big\}\times\mathbf{1}_{\tilde{E}}\\ &=\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\Big\}\\ &\quad\times\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=3,4)\Big\}\times\mathbf{1}_{\tilde{E}}.\end{aligned}$$

Given Conditions A.5 and A.6, we know that similarly to Section B.9.4, we can show $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})\}^2 - G_{a_1,a_2,2}| = o(n^{-2})$ and $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,s_1,s_2}) - G_{a_1,a_2,2}| = o(n^{-2})$ respectively. Finally we obtain $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2}) = o(n^{-2})$. The proof is very similar and the details is thus skipped.

B.9.5. *Proof of Lemma A.25 (on Page 62, Section A.13).* Recall the form of $D_{n,k}$ derived in Section B.9.3:

$$\sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4) = \sum_{k=1}^{n} \sum_{1 \le r_1, r_2, r_3, r_4 \le m} \prod_{l=1}^{4} t_{r_l} \times \mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a_{r_l}} \Big).$$

To prove Lemma A.25, it suffices to show that for given $1 \le k \le n$ and $1 \le r_1, r_2, r_3, r_4 \le m$, we have $\mathrm{E}(\prod_{l=1}^{4} A_{n,k,a_{r_l}}) = o(n^{-1})$. In addition, by the Cauchy-Schwarz inequality, it suffices to show $\mathrm{E}(A_{n,k,a}^4) = o(n^{-1})$ for each given finite a.

Part I: $1 \le k \le n_x$. We consider without loss of generality that $k \ge a$ and

$$\mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a}^4 \Big) = c^4(n,a) \sum_{\substack{1 \le j_1, \dots, j_8 \le p; \\ \mathbf{i}^{(l)} \in \mathcal{P}(k-1, a-1),\, l=1,\dots,4}} \mathrm{E}\Big(\prod_{l=1}^{4} \prod_{t_l=1}^{a-1} \mathcal{X}_{i_{t_l}^{(l)}, j_{2l-1}, j_{2l}} \Big) \times \mathrm{E}\Big(\prod_{l=1}^{4} \mathcal{X}_{j_{2l-1}, j_{2l}} \Big).$$

As $\mathrm{E}(\mathcal{X}_{j_1,j_2}) = 0$ under H_0, we know

$$\mathrm{E}\Big(\prod_{l=1}^{4} \prod_{t_l=1}^{a-1} \mathcal{X}_{i_{t_l}^{(l)}, j_{2l-1}, j_{2l}} \Big) \neq 0$$

only when $| \cup_{l=1}^{4} \{\mathbf{i}^{(l)}\} | \le 2(a-1)$. Note that $c(n,a) = \Theta(1)\{n^a \sigma(a)\}^{-1}$. To finish the proof, it suffices to show that for given $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$, we have

$$\sum_{1 \le j_1, \dots, j_8 \le p} \mathrm{E}\Big(\prod_{l=1}^{4} \prod_{t_l=1}^{a-1} \mathcal{X}_{i_{t_l}^{(l)}, j_{2l-1}, j_{2l}} \Big) \mathrm{E}\Big(\prod_{l=1}^{4} \mathcal{X}_{j_{2l-1}, j_{2l}} \Big) = O(n^{2a}) \sigma^4(a). \tag{B.202}$$

We next prove (B.202) under Conditions A.5 and A.6 in the following Sections B.9.5 and B.9.5, respectively.

Under Condition A.5. Recall that $\mathcal{X}_{i,j_1,j_2} = x_{i,j_1} x_{i,j_2} - \sigma_{j_1,j_2}$. By the symmetricity of the j indexes, we have

$$\begin{aligned} \sum_{1 \le j_1, \dots, j_8 \le p} \Big| \mathrm{E}\Big(\prod_{l=1}^{4} \mathcal{X}_{j_{2l-1}, j_{2l}} \Big) \Big| \le C & \sum_{1 \le j_1, \dots, j_8 \le p} \Bigg\{ \Big| \mathrm{E}\Big(\prod_{l=1}^{8} x_{1,j_l} \Big) \Big| \\ & + \Big| \mathrm{E}\Big(\prod_{l=1}^{6} x_{1,j_l} \Big) \sigma_{j_7,j_8} \Big| + \Big| \mathrm{E}\Big(\prod_{l=1}^{4} x_{1,j_l} \Big) \sigma_{j_5,j_6} \sigma_{j_7,j_8} \Big| + \Big| \prod_{l=1}^{4} \sigma_{j_{2l-1}, j_{2l}} \Big| \Bigg\}. \end{aligned}$$

Under Condition A.5 with the mixing-type assumption, following similar analysis in Sections B.1.5 and B.1.6, we know $\sum_{1\leq j_1,\dots,j_8\leq p}|\mathrm{E}(\prod_{l=1}^{8} x_{1,j_l})|$, $\sum_{1\leq j_1,\dots,j_8\leq p}|\mathrm{E}(\prod_{l=1}^{6} x_{1,j_l})\sigma_{j_7,j_8}|$, $\sum_{1\leq j_1,\dots,j_8\leq p}|\mathrm{E}(\prod_{l=1}^{4} x_{1,j_l})\sigma_{j_5,j_6}\sigma_{j_7,j_8}|$ and $\sum_{1\leq j_1,\dots,j_8\leq p}|\prod_{l=1}^{4}\sigma_{j_{2l-1},j_{2l}}|$ are all $O(p^4)$. It follows that

$$\sum_{1\leq j_1,\dots,j_8\leq p}\Big|\mathrm{E}\Big(\prod_{l=1}^{4}\mathcal{X}_{j_{2l-1},j_{2l}}\Big)\Big| = O(p^4), \tag{B.203}$$

Recall that Lemma A.22 shows that $\sigma^2(a) = \Theta(p^2 n^{-a})$. By (B.203) and Condition A.5 (2), we have (B.202) holds and $\mathrm{E}(A_{n,k,a}^4) = o(n^{-1})$.

Under Condition A.6. By Condition A.6 (3), we know that $\mathrm{E}(\prod_{l=1}^{4}\prod_{t_l=1}^{a-1}\mathcal{X}_{i_{t_l}^{(l)},j_{2l-1},j_{2l}})\times$ $\mathrm{E}(\prod_{l=1}^{4}\mathcal{X}_{j_{2l-1},j_{2l}})$ is a linear combination of $\mathrm{E}(\prod_{t=1}^{4a}\sigma_{j_{g_{2t-1}},j_{g_{2t}}})$, where $\mathcal{G} = (g_1,\dots,g_{8a}) \in \{1,\dots,8\}^{8a}$ satisfies that $g_{2t-1}\neq g_{2t}$ for $t = 1,\dots,4a$ and the number of g's equal to m is a for each $m\in\{1,\dots,8\}$. By Condition A.6 (4), for given $\mathcal{G}$ satisfying the constraints, $\sum_{1\leq j_1,\dots,j_8\leq p}\sigma_{j_{g_{2t-1}},j_{g_{2t}}} = O(1)\sum_{1\leq j_1,\dots,j_8\leq p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_5,j_6}\sigma_{j_7,j_8})^a$. Then we have

$$\begin{aligned}
&\sum_{1\leq j_1,\dots,j_8\leq p}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t_l=1}^{a-1}\mathcal{X}_{i_{t_l}^{(l)},j_{2l-1},j_{2l}}\Big)\times \mathrm{E}(\prod_{l=1}^{4}\mathcal{X}_{j_{2l-1},j_{2l}})\\
&= O(1)\sum_{1\leq j_1,\dots,j_8\leq p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_5,j_6}\sigma_{j_7,j_8})^a = O(1)\Big(\sum_{1\leq j_1,j_2\leq p}\sigma_{j_1,j_2}^a\Big)^4.
\end{aligned}$$

Recall that Lemma A.22 shows that $\sigma^2(a) = \Theta(n^{-a})(\sum_{1\leq j_1,j_2\leq p}\sigma_{j_1,j_2}^a)^2$. Therefore, (B.202) is obtained and Lemma A.25 is proved.

Part II: $n_x+1\leq k\leq n_x+n_y$. Section B.9.3 derives that $A_{n,k,a} = \sum_{s=L_k}^{a-1}A_{n,k,a,s}$, where

$$A_{n,k,a,s} = \sum_{\substack{1\leq j_1,j_2\leq p;\\ \mathbf{i}\in\mathcal{P}(n_x,s);\\ \mathbf{w}\in\mathcal{P}(k-n_x-1,a-s-1)}} c_2(n,a,s)\mathcal{Y}_{k-n_x,j_1,j_2}\prod_{t=1}^{s}\mathcal{X}_{i_t,j_1,j_2}\prod_{\tilde{t}=1}^{a-s-1}\mathcal{Y}_{w_{\tilde{t}},j_1,j_2}.$$

Similarly to Section B.9.5, it suffices to show that for given finite integers a and s, $\mathrm{E}(A_{n,k,a,s}^4) = o(n^{-1})$. Following the arguments in Section B.9.4, we know $A_{n,k,a,s}$ takes a similar form to $A_{n,k,a}$ in Section B.9.5. Therefore the proof in Section B.9.5 can be applied similarly to show $\mathrm{E}(A_{n,k,a,s}^4) = o(n^{-1})$ in this section. The proof will be very similar and the details are thus skipped.

B.10. Lemmas for the proof of Theorem 4.7.

B.10.1. *Proof of Lemma A.26 (on Page 64, Section A.14).* In this section, to prove Lemma A.26, we study $\mathrm{var}(T_{D,a,1})$, $\mathrm{var}(T_{D,a,2})$ and $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ respectively.

Part I: $\mathrm{var}(T_{D,a,1})$. We first derive $\mathrm{var}(T_{D,a,1})$. Note that $T_{D,a,1}$ is a summation over j indexes in $\mathbb{J}_0$, and $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2}$ for $j_1, j_2 \in \mathbb{J}_0$. Following the arguments in Section B.9.1, similarly to (B.175), we have

$$\mathrm{var}(T_{D,a,1}) \simeq \sum_{1\leq j_1,j_2,j_3,j_4\in\mathbb{J}_0} a!(\mathbf{X}_{j_1,j_2,j_3,j_4}/n_x + \mathbf{Y}_{j_1,j_2,j_3,j_4}/n_y)^a.$$

By Condition A.7 (3), (B.183) still holds. Then by Condition A.8 and the symmetricity of j indexes,

$$\mathrm{var}(T_{D,a,1}) \simeq C_{\kappa,a} \sum_{1\leq j_1,j_2,j_3,j_4\in\mathbb{J}_0} a!\sigma^a_{j_1,j_2}\sigma^a_{j_3,j_4}, \tag{B.204}$$

where $C_{\kappa,a} = \{(\kappa_x - 1)/n_x + (\kappa_y - 1)/n_y\}^a + 2(\kappa_x/n_x + \kappa_y/n_y)^a$, and $\mathrm{var}(T_{D,a,1})$ is of order $\Theta(n^{-a}\mathbb{V}^{1/2}_{a,a,0,0})$ with $\mathbb{V}^{1/2}_{a,a,0,0} = \sum_{j_1,\dots,j_4\in\mathbb{J}_0}(\sigma_{x,j_1,j_2}\sigma_{x,j_3,j_4})^a$ defined on Page 63.

Part II: $\mathrm{var}(T_{D,a,2})$. We show $\mathrm{var}(T_{D,a,2}) = o(1)\mathrm{var}(T_{D,a,1})$. Particularly,

$$T_{D,a,2} = \sum_{(j_1,j_2)\in J_{0,D}} \frac{1}{P^{n_x}_a P^{n_y}_a} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n_x,a),\\ \mathbf{w}\in\mathcal{P}(n_y,a)}} \prod_{t=1}^{a}(\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2}),$$

where we redefine $\mathcal{X}_{i,j_1,j_2} = x_{i,j_1}x_{i,j_2} - \sigma_{y,j_1,j_2}$ and $\mathcal{Y}_{i,j_1,j_2} = y_{i,j_1}y_{i,j_2} - \sigma_{y,j_1,j_2}$. Moreover, we define

$$G_{D,a} = \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P^{n_x}_a P^{n_y}_a)^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}=\emptyset\}}(D_{j_1,j_2}D_{j_3,j_4})^a.$$

To prove $\mathrm{var}(T_{D,a,2}) = \mathrm{E}(T^2_{D,a,2}) - \{\mathrm{E}(T_{D,a,2})\}^2$ is $o(1)\mathrm{var}(T_{D,a,1})$, we next show $|\mathrm{E}(T^2_{D,a,2}) - G_{D,a}|$ and $|\{\mathrm{E}(T_{D,a,2})\}^2 - G_{D,a}|$ are both $o(1)\mathrm{var}(T_{D,a,1})$.

Note that $\mathrm{E}(\mathcal{X}_{i,j_1,j_2}) = D_{j_1,j_2}$ and $\mathrm{E}(\mathcal{Y}_{i,j_1,j_2}) = 0$. We have

$$\{\mathrm{E}(T_{D,a,2})\}^2 = \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P^{n_x}_a P^{n_y}_a)^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} (D_{j_1,j_2}D_{j_3,j_4})^a.$$

Then

$$
\begin{aligned}
&|\{\mathrm{E}(T_{D,a,2})\}^2 - G_{D,a}| \\
\leq &\Big| \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P_a^{n_x} P_a^{n_y})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}\neq\emptyset\}} (D_{j_1,j_2} D_{j_3,j_4})^a \Big| \\
\leq &Cn^{-1} \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |D_{j_1,j_2} D_{j_3,j_4}|^a,
\end{aligned}
$$

where we use $\sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)} \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}\neq\emptyset\}} = O(n^{4a-1})$. In addition,

$$
\begin{aligned}
&|\mathrm{E}(T_{D,a,2}^2) - G_{D,a}| \\
\leq &\ C \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P_a^{n_x} P_a^{n_y})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \\
&\Bigg(\mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}=\emptyset\}} \Big| \mathrm{E}\Big\{ \prod_{t=1}^a (\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2})(\mathcal{X}_{\tilde{i}_t,j_3,j_4} - \mathcal{Y}_{\tilde{w}_t,j_3,j_4}) \Big\} - (D_{j_1,j_2} D_{j_3,j_4})^a \Big| \\
&\quad + \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}\neq\emptyset\}} \Big| \mathrm{E}\Big\{ \prod_{t=1}^a (\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2})(\mathcal{X}_{\tilde{i}_t,j_3,j_4} - \mathcal{Y}_{\tilde{w}_t,j_3,j_4}) \Big\} \Big| \Bigg).
\end{aligned}
$$

We redefine $\mathbf{X}_{j_1,j_2,j_3,j_4} = \mathrm{E}(\mathcal{X}_{i,j_1,j_2}\mathcal{X}_{i,j_3,j_4})$ and $\mathbf{Y}_{j_1,j_2,j_3,j_4} = \mathrm{E}(\mathcal{Y}_{i,j_1,j_2}\mathcal{Y}_{i,j_3,j_4})$. Then

$$
\begin{aligned}
&|\mathrm{E}(T_{D,a,2}^2) - G_{D,a}| \\
\leq &\ C \sum_{1\leq m_1+m_2\leq a} n^{-m_1-m_2} \\
&\times \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} \Big| \mathbf{X}_{j_1,j_2,j_3,j_4}^{m_1} \mathbf{Y}_{j_1,j_2,j_3,j_4}^{m_2} (D_{j_1,j_2} D_{j_3,j_4})^{a-m_1-m_2} \Big|.
\end{aligned}
$$

Note that $\mathbf{Y}_{j_1,j_2,j_3,j_4} = \sigma_{y,j_1,j_3}\sigma_{y,j_2,j_4} + \sigma_{y,j_1,j_4}\sigma_{y,j_2,j_3}$ and $\sigma_{y,j_1,j_2} = \sigma_{x,j_1,j_2} - D_{j_1,j_2}$. By Conditions A.7 and A.8, the Hölder's inequality and definitions in (A.28), we have

$$
\mathrm{var}(T_{D,a,2}) \leq C \max_{\substack{\mathcal{H}\in\mathbb{H},\\ t=1,2}} \Big\{ \sum_{m=1}^a (n^{-a}\mathbb{V}_{a,\mathcal{H},x,t})^{m/a} (\mathbb{V}_{a,\mathcal{H},D,3})^{1-m/a},\ n^{-1}\mathbb{V}_{a,\mathcal{H},D,3} \Big\}.
$$

Therefore by Condition A.8 and (B.204), $\mathrm{var}(T_{D,a,2}) = o(1)n^{-a}\mathbb{V}_{a,a,0,0}^{1/2} = o(1)\mathrm{var}(T_{D,a,1})$.

Part III: $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$. Last, we prove $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}(T_{D,a,1})$. Similarly to Section B.9.1, we write $\tilde{\mathcal{U}}^*(a) = \sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c} C_{a,c,b_1,b_2} \times T_{b_1,b_2,c}\mathbf{1}_{b_1+b_2\le a-1}$, where $T_{b_1,b_2,c}$ is defined in (B.179). For finite a, to prove $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}(T_{D,a,1})$, it suffices to prove $\mathrm{var}(T_{b_1,b_2,c}) = o(1)\mathrm{var}(T_{D,a,1})$ for $0 \le c \le a$ and $b_1 + b_2 \le a-1$. As $\mathrm{E}(\mathcal{Y}_{i,j_1,j_2}) = 0$ and $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = 0$, we know that if $b_1 + b_2 \le a-1$, $\mathrm{E}(T_{b_1,b_2,c}) = 0$. Then $\mathrm{var}(T_{b_1,b_2,c}) = \mathrm{E}(T^2_{b_1,b_2,c})$, which takes a similar form to (B.180). Specifically, we can write $\mathrm{var}(T_{b_1,b_2,c}) = \mathrm{var}(T_{b_1,b_2,c})_{(1)} + \mathrm{var}(T_{b_1,b_2,c})_{(2)}$, where

$$\mathrm{var}(T_{b_1,b_2,c})_{(1)} = \sum_{j_1,j_2,\tilde{j}_1,\tilde{j}_2\in\mathbb{J}_0} (P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,2(a-c)-b_2)}} \mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2),$$

and

$$\mathrm{var}(T_{b_1,b_2,c})_{(2)} = \sum_{\substack{(j_1,j_2),\\ (\tilde{j}_1,\tilde{j}_2)\in J_{0,D}}} (P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,2(a-c)-b_2)}} \mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2),$$

and $\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2)$ is defined same as in (B.180).

Note that $\mathrm{var}(T_{b_1,b_2,c})_{(1)}$ is a summation over j indexes in $\mathbb{J}_0$, and $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2}$ for $j_1, j_2 \in \mathbb{J}_0$. Therefore the arguments under H_0 in Section B.9.1 can be applied similarly to $\mathrm{var}(T_{b_1,b_2,c})_{(1)}$. Then we have $\mathrm{var}(T_{b_1,b_2,c})_{(1)} = o(n^{-a})(\sum_{j_1,j_2\in\mathbb{J}_0}\sigma^a_{j_1,j_2})^2$ which is $o(1)\mathrm{var}(T_{D,a,1})$. We next consider $\mathrm{var}(T_{b_1,b_2,c})_{(2)}$. As $\mathrm{E}(\mathcal{Y}_{i,j_1,j_2}) = 0$ and $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = 0$, by the definition in (B.180), we know $\mathrm{E}\{\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2)\} \ne 0$ only when $\{i_{b_1+1},\dots,i_{2c-b_1}\} = \{\tilde{i}_{b_1+1},\dots,\tilde{i}_{2c-b_1}\}$ and $\{\mathbf{w}\} = \{\tilde{\mathbf{w}}\}$. Let $m_0 = b_1 - |\{i_1,\dots,i_{b_1}\}\cap\{\tilde{i}_1,\dots,\tilde{i}_{b_1}\}|$. By Condition A.7 (3) and the Hölder's inequality,

$$\begin{aligned}
&\mathrm{var}(T_{b_1,b_2,c})_{(2)}\\
&\le Cn_x^{-(c-b_1)}n_y^{-(a-c-b_2)} \max_{\substack{\mathcal{H}\in\mathbb{H}_0,\\ 0\le m_0\le b_1}} \Bigg\{\Bigg(n_y^{-a}\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |\sigma_{y,j_{h_1},j_{h_2}}\sigma_{y,j_{h_3},j_{h_4}}|^a\Bigg)^{\frac{a-c}{a}}\\
&\qquad\times\Bigg(n_x^{-a}\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |\sigma_{x,j_{h_1},j_{h_2}}\sigma_{x,j_{h_3},j_{h_4}}|^a\Bigg)^{\frac{c-m_0}{a}}\\
&\qquad\times\Bigg(\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |D_{j_{h_1},j_{h_2}}D_{j_{h_3},j_{h_4}}|^a\Bigg)^{\frac{m_0}{a}}\Bigg\}\\
&\le Cn^{-(a-b_1-b_2)} \max_{\mathcal{H}\in\mathbb{H}_0,t=1,2}\{n^{-a}\mathbb{V}_{a,\mathcal{H},x,t},\mathbb{V}_{a,\mathcal{H},D,3}\},
\end{aligned}$$

where the last inequality uses $\sigma_{y,j_1,j_2} = \sigma_{x,j_1,j_2} - D_{j_1,j_2}$. As $b_1 + b_2 \le a - 1$, $\mathrm{var}(T_{b_1,b_2,c})_{(2)} \le Cn^{-1}\max_{\mathcal{H}\in\mathbb{H}_0;t=1,2}\{n^{-a}\mathbb{V}_{a,\mathcal{H},x,t}, \mathbb{V}_{a,\mathcal{H},D,3}\}$. By Condition A.8 and (B.204), we know $\mathrm{var}(T_{b_1,b_2,c})_{(2)} = o(1)\mathrm{var}(T_{D,a,1})$.

B.11. Proof of Remark 2.4. In this section, we prove the conclusion in Remark 2.4. To be specific, we prove in the following that under the conditions of Theorem 2.3,

$$\left|P\Big(n(M_n^\dagger)^2 > y_p, \frac{\mathcal{U}(a_1)}{\sigma(a_1)} \le 2z_1, \dots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)} \le 2z_m\Big) - P\Big(n(M_n^\dagger)^2 > y_p\Big)\prod_{r=1}^m P\Big(\frac{\mathcal{U}(a_r)}{\sigma(a_r)} \le 2z_r\Big)\right| \to 0. \tag{B.205}$$

Note that we already know M_n/n and $\mathcal{U}(a_r)/\sigma(a_r)$'s for $r = 1, \dots, m$ are asymptotically independent by the proof of Lemmas A.8 and A.9. In this section, the proof idea is that we show the difference between $n(M_n^\dagger)^2$ and M_n/n is $o_p(1)$ and then obtain (B.205). To prove that $n(M_n^\dagger)^2 - M_n/n$ is $o_p(1)$, we introduce an intermediate variable $\tilde{M}_n/n$ defined below, and show that $\tilde{M}_n/n - n(M_n^\dagger)^2 = o_p(1)$ and $\tilde{M}_n/n - M_n/n = o_p(1)$ respectively.

Specifically, we define

$$\tilde{M}_n/n = \max_{1\le j_1\ne j_2\le p} |n\hat{\sigma}^2_{j_1,j_2}/\theta_{j_1,j_2}|,$$

where $\hat{\sigma}_{j_1,j_2} = \sum_{i=1}^n\{(x_{i,j_1} - \bar{x}_{j_1})(x_{i,j_2} - \bar{x}_{j_2})\}/n$ and $\theta_{j_1,j_2} = \mathrm{var}\{(x_{i,j_1} - \mu_{j_1})(x_{i,j_2} - \mu_{j_2})\}$. Moreover, by (A.9), we have

$$M_n/n = \max_{1\le j_1\ne j_2\le p} |n\tilde{\sigma}^2_{j_1,j_2}/\theta_{j_1,j_2}|,$$

where we use the fact that $\theta_{j_1,j_2} = \sigma_{j_1,j_1}\sigma_{j_2,j_2}$ by Condition 2.3 and define $\tilde{\sigma}_{j_1,j_2} = \sum_{i=1}^n\{(x_{i,j_1} - \mu_{j_1})(x_{i,j_2} - \mu_{j_2})\}/n$. In addition, we have

$$M_n^\dagger = \max_{1\le j_1\ne j_2\le p} |\hat{\sigma}_{j_1,j_2}|/(\hat{\theta}_{j_1,j_2})^{1/2},$$

where we let $\hat{\theta}_{j_1,j_2} = \widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2}) = n^{-1}\sum_{i=1}^n\{(x_{i,j_1}-\bar{x}_{j_1})(x_{i,j_2}-\bar{x}_{j_2})-\hat{\sigma}_{j_1,j_2}\}^2$. In the following, we will first compare $\tilde{M}_n/n$ and $n(M_n^\dagger)^2$, and then compare $\tilde{M}_n/n$ and M_n/n. Also for simplicity, we assume without loss of generality that $\mu_j = 0$ and $\sigma_{j,j} = 1$.

Note that $n(M_n^\dagger)^2 = \max_{1\le j_1\ne j_2\le p}|n\hat{\sigma}^2_{j_1,j_2}/\hat{\theta}_{j_1,j_2}|$, which differs from $\tilde{M}_n/n$ only by replacing θ_{j_1,j_2} with $\hat{\theta}_{j_1,j_2}$. By the proof of Lemma 3 in [10], we know that for any $C_2 > 0$, there exists some constant C_1 such that

$$P\Big(\max_{1\le j_1\ne j_2\le p} |\hat{\theta}_{j_1,j_2} - \theta_{j_1,j_2}|/\theta_{j_1,j_2} \ge C_1\sqrt{\log p/n}\Big) = O(p^{-C_2}).$$

Under the event $|\hat{\theta}_{j_1,j_2}/\theta_{j_1,j_2} - 1| \leq C_1\sqrt{\log p/n}$, we have

$$\begin{aligned}
&|\tilde{M}_n/n - n(M_n^{\dagger})^2| \\
=& \Big| \max_{1\leq j_1\neq j_2\leq p} n\hat{\sigma}^2_{j_1,j_2}/\theta_{j_1,j_2} - \max_{1\leq j_1\neq j_2\leq p} n\hat{\sigma}^2_{j_1,j_2}/\hat{\theta}_{j_1,j_2}\Big| \\
\leq& \max_{1\leq j_1\neq j_2\leq p} |n\hat{\sigma}^2_{j_1,j_2}/\theta_{j_1,j_2}| \times \max_{1\leq j_1\neq j_2\leq p} |1-\theta_{j_1,j_2}/\hat{\theta}_{j_1,j_2}| \\
\leq& \max_{1\leq j_1\neq j_2\leq p} |n\hat{\sigma}^2_{j_1,j_2}/\theta_{j_1,j_2}| C_1\sqrt{\log p/n}.
\end{aligned}$$

It follows that $n(M_n^{\dagger})^2 = \tilde{M}_n/n\{1+O(\sqrt{\log p/n})\}$. Since $\log p/n \to 0$ and $\tilde{M}_n/n$ has a limit by Theorem 3 in Cai and Jiang [9], then $|\tilde{M}_n/n - n(M_n^{\dagger})^2| = o_p(1)$.

We next compare $\tilde{M}_n/n$ and M_n/n. by Lemma B.3,

$$\begin{aligned}
&|\tilde{M}_n/n - M_n/n| \\
\leq& C \max_{1\leq j_1\neq j_2\leq p} \Big| \sum_{i=1}^n (x_{i,j_1}-\bar{x}_{j_1})(x_{i,j_2}-\bar{x}_{j_2}) - \sum_{i=1}^n x_{i,j_1}x_{i,j_2}\Big|^2 \Big/ n \\
&+ C\sqrt{M_n/n} \max_{1\leq j_1\neq j_2\leq p} \Big| \sum_{i=1}^n (x_{i,j_1}-\bar{x}_{j_1})(x_{i,j_2}-\bar{x}_{j_2}) - \sum_{i=1}^n x_{i,j_1}x_{i,j_2}\Big| \Big/ \sqrt{n} \\
\leq& C \max_{1\leq j\leq p} n\bar{x}_j^4 + Cn^{1/2}\sqrt{M_n/n} \max_{1\leq j\leq p} \bar{x}_j^2,
\end{aligned}$$

where in the last inequality we use $\max_{1\leq j_1\neq j_2\leq p} \bar{x}_{j_1}\bar{x}_{j_2} \leq \max_{1\leq j_1\neq j_2\leq p}(\bar{x}_{j_1}^2 + \bar{x}_{j_2}^2)/2 \leq \max_{1\leq j\leq p}\bar{x}_j^2$. By Eq. (27) in Lemma 2 of Cai and Liu [7], we know that $\max_{1\leq j\leq p}|\bar{x}_j| = O_p(\sqrt{\log p/n})$. Since we assume $\log p = o(n^{1/7})$, and Proposition 6.3 in [9] shows that M_n/n has a limit, we know $|\tilde{M}_n/n - M_n/n| = o_p(1)$.

In summary, $|M_n/n - n(M_n^{\dagger})^2| \leq |M_n/n - \tilde{M}_n/n| + |\tilde{M}_n/n - n(M_n^{\dagger})^2| = o_p(1)$. Since $|M_n/n - n(M_n^{\dagger})^2| = o_p(1)$ and M_n/n and $\mathcal{U}(a_r)/\sigma(a_r)$'s for $r = 1,\dots,m$ are asymptotically independent, similarly to the proof of Lemma A.9, we know (B.205) is proved.

B.12. Proof of Corollary 4.1. Since the proofs in Sections A.10 and A.12 do not rely on $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$, the proof of Corollary 4.1 follows from Sections A.10 and A.12 directly. We also obtain $\mathrm{var}\{\mathcal{U}(a)\}$ under the null and alternative hypotheses by Lemma A.16 (on Page 56) and Lemma A.20 (on Page A.20), respectively.

APPENDIX C: COMPUTATION & SUPPLEMENTARY SIMULATIONS

C.1. Computation.

C.1.1. *Formulae for (2.15).* Note that $\mathcal{U}_l(a) = U_l^{\mathbf{1}_a}$ by the definitions in (2.16), and for different l's, the computation methods of $U_l^{\mathbf{1}_a}$'s are the same. Therefore in the following, for simplicity, we give the formulae of $U_l^{\mathbf{1}_a}$ without the subscript l:

$$\begin{aligned}
U^{\mathbf{1}_1} =& V^{(1)},\\
U^{\mathbf{1}_2} =& V^{(1,1)} - V^{(2)},\\
U^{\mathbf{1}_3} =& V^{\mathbf{1}_3} - 3V^{(2,1)} + 2V^{(3)},\\
U^{\mathbf{1}_4} =& V^{\mathbf{1}_4} - 6V^{(2,1,1)} + 8V^{(3,1)} + 3V^{(2,2)} - 6V^{(4)},\\
U^{\mathbf{1}_5} =& V^{\mathbf{1}_5} - 10V^{(2,\mathbf{1}_3)} + 20V^{(3,\mathbf{1}_2)} + 15V^{(2,2,1)} - 30V^{(4,1)}\\
& - 20V^{(2,3)} + 24V^{(5)},\\
U^{\mathbf{1}_6} =& V^{\mathbf{1}_6} - 15V^{(\mathbf{1}_4,2)} + 40V^{(3,\mathbf{1}_3)} + 45V^{(1,1,2,2)},\\
& - 90V^{(1,1,4)} - 120V^{(1,2,3)} + 144V^{(1,5)} - 15V^{(2,2,2)}\\
& + 90V^{(2,4)} + 40V^{(3,3)} - 120V^{(6)},
\end{aligned}$$

where $U^{\mathbf{1}_a}$ and $V^{(t_1,\dots,t_k)}$ are defined as in (2.16).

C.1.2. *Computation with unknown mean.* In this section, we provide the details of the computation of $\mathcal{U}(a)$ when $\mathrm{E}(x_{i,j})$ is unknown. We note that $\mathcal{U}(a)$ is some linear combination of

$$\sum_{1\le i_1\neq\ldots\neq i_k\le n} \prod_{t=1}^{k} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}, \tag{C.1}$$

where $a \le k \le 2a$, $r_{t,1}, r_{t,2} \ge 0$ and $r_{t,1} + r_{t,2} \ge 1$. A direct calculation of (C.1) has computational cost $O(n^k)$, which is large when k is large. But following the discussion in Section 2.3, we can similarly reduce the computational cost of (C.1) to order $O(n)$ with an iterative method. In particular, we note that

$$\begin{aligned}
&\sum_{1\le i_1\neq\ldots\neq i_k\le n} \prod_{t=1}^{k} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}\\
&=\Big(\sum_{1\le i_1\neq\ldots\neq i_{k-1}\le n} \prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}\Big)\Big(\sum_{i=1}^{n} x_{i,j_1}^{r_{k,1}} x_{i,j_2}^{r_{k,2}}\Big)\\
&\quad - \sum_{m=1}^{k-1} \sum_{1\le i_1\neq\ldots\neq i_{k-1}\le n} \Big(\prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}\Big) x_{i_m,j_1}^{r_{k,1}} x_{i_m,j_2}^{r_{k,2}}.
\end{aligned} \tag{C.2}$$

Suppose we can compute $\sum_{1\le i_1\neq\ldots\neq i_{k-1}\le n}\prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}$ with cost $O(n)$ for any $(r_{t,1}, r_{t,2}), t = 1,\ldots,k-1$. Then by the relationship in (C.2), we can obtain (C.1) with cost $O(n)$ iteratively.

We then illustrate the iterative method with some examples. When $k = 1$, for any given $(r_{1,1}, r_{1,2})$, we know $\sum_{i=1}^n x_{i,j_1}^{r_{1,1}} x_{i,j_2}^{r_{1,2}}$ can be computed with cost $O(n)$. When $k = 2$, by (C.2), we have $\sum_{1\le i_1\neq i_2\le n}\prod_{t=1}^{2} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}} = (\sum_{i=1}^n x_{i,j_1}^{r_{1,1}} x_{i,j_2}^{r_{1,2}})(\sum_{i=1}^n x_{i,j_1}^{r_{2,1}} x_{i,j_2}^{r_{2,2}}) - \sum_{i=1}^n x_{i,j_1}^{r_{1,1}+r_{2,1}} x_{i,j_2}^{r_{1,2}+r_{2,2}}$, which can be computed with cost $O(n)$. For a general k, suppose for any given $(r_{t,1}, r_{t,2}), t = 1,\ldots,k-1$, we can compute $\sum_{1\le i_1\neq\ldots\neq i_{k-1}\le n}\prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}$ with cost $O(n)$. Then by (C.2), we can obtain (C.1) with computational cost $O(n)$.

Given the iterative method discussed above, we can compute $\mathcal{U}(a)$ with cost $O(p^2 n)$. For example, we can write $\mathcal{U}(1)$ as

$$\sum_{1\le j_1\neq j_2\le p}\Big\{n^{-1}\sum_{i=1}^n x_{i,j_1}x_{i,j_2} - (P_2^n)^{-1}\Big(\sum_{i_1=1}^n x_{i_1,j_1}\sum_{i_2=1}^n x_{i_2,j_2} - \sum_{i=1}^n x_{i,j_1}x_{i,j_2}\Big)\Big\}.$$

For $a = 2$, similar analysis holds. Note that

$$\mathcal{U}(2) = \sum_{1\le j_1\neq j_2\le p}\Big\{(P_2^n)^{-1}\mathcal{U}_1(2) - 2(P_3^n)^{-1}\mathcal{U}_2(2) + (P_4^n)^{-1}\mathcal{U}_3(2)\Big\},$$

where

$$\begin{aligned}
\mathcal{U}_1(2) &= \sum_{1\le i_1\neq i_2\le n}\prod_{t=1}^{2} x_{i_t,j_1}x_{i_t,j_2},\\
\mathcal{U}_2(2) &= \sum_{1\le i_1\neq i_2\neq i_3\le n}(x_{i_1,j_1}x_{i_1,j_2})(x_{i_2,j_1})(x_{i_3,j_2}),\\
\mathcal{U}_3(2) &= \sum_{1\le i_1\neq i_2\neq i_3\neq i_4\le n}\prod_{t=1}^{2} x_{i_t,j_1}\prod_{t=3}^{4} x_{i_t,j_2}.
\end{aligned}$$

We then find that $\mathcal{U}_1(2)$,$\mathcal{U}_2(2)$ and $\mathcal{U}_3(2)$ can be computed with cost $O(n)$ using the following formulae.

$$\mathcal{U}_1(2) = \Big(\sum_{i=1}^n x_{i,j_1}x_{i,j_2}\Big)^2 - \sum_{i=1}^n (x_{i,j_1}x_{i,j_2})^2.$$

$$\begin{aligned}
\mathcal{U}_2(2) =&\Big(\sum_{i=1}^n x_{i,j_1}x_{i,j_2}\Big)\Big(\sum_{1\le i_1\neq i_2\le n} x_{i,j_1}x_{i,j_2}\Big)\\
&- \sum_{1\le i_1\neq i_2\le n}(x_{i_1,j_1}^2 x_{i_1,j_2})x_{i_2,j_2} - \sum_{1\le i_1\neq i_2\le n}(x_{i_1,j_1}x_{i_1,j_2}^2)x_{i_2,j_1},
\end{aligned}$$

where we use $\sum_{1\leq i_1\neq i_2\leq n} x_{i,j_1}x_{i,j_2} = (\sum_{i=1}^n x_{i,j_1})(\sum_{i=1}^n x_{i,j_2}) - \sum_{i=1}^n x_{i,j_1}x_{i,j_2}$, and $\sum_{1\leq i_1\neq i_2\leq n}(x_{i_1,j_1}^2 x_{i_1,j_2})x_{i_2,j_2} = (\sum_{i=1}^n x_{i,j_1}^2 x_{i,j_2})(\sum_{i=1}^n x_{i,j_2}) - \sum_{i=1}^n x_{i,j_1}^2 x_{i,j_2}^2$.

$$\mathcal{U}_3(2) = \Big(\sum_{1\leq i_1\neq i_2\leq n} x_{i_1,j_1}x_{i_2,j_1}\Big)\Big(\sum_{1\leq i_3\neq i_4\leq n} x_{i_3,j_2}x_{i_4,j_2}\Big) - 2\mathcal{U}_1(2) - 4\mathcal{U}_3(2),$$

where we use $\sum_{1\leq i_1\neq i_2\leq n} x_{i_1,k}x_{i_2,k} = (\sum_{i=1}^n x_{i,k})^2 - \sum_{i=1}^n x_{i,k}^2$ for $k = j_1, j_2$.

When $a \geq 3$, the similar iterative method can be applied. But the closed form for computation might be hard to derive directly. Alternatively, we introduce a simplified form of U-statistics: $\mathcal{U}_c(a) = (P_a^n)^{-1}\sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \sum_{1\leq j_1\neq j_2\leq p}\prod_{t=1}^a (x_{i_t,j_1} - \bar{x}_{j_1})(x_{i_t,j_2} - \bar{x}_{j_2})$. We note that $\mathcal{U}_c(a)$ takes a similar form to $\tilde{\mathcal{U}}(a)$ in (2.5), but replacing each observation $x_{i,j}$ with the centered correspondence $x_{i,j} - \bar{x}_j$. Therefore, $\mathcal{U}_c(a)$ can be computed with cost $O(n)$ using Algorithm 1, if we set $s_{i,l} = (x_{i,j_1} - \bar{x}_{j_1})(x_{i,j_2} - \bar{x}_{j_2})$ in Algorithm 1 for $l \in \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$. We then show that we can substitute $\mathcal{U}(a)$ with $\mathcal{U}_c(a)$ when $a \geq 3$ in computation under certain conditions.

PROPOSITION C.1. *Under the Conditions of Theorem 2.4, consider $a \geq 3$. If a is odd, $p = o(n^{1+a/2})$; if a is even, $p = o(n^{a/2})$. Then $\{\mathcal{U}(a) - \mathcal{U}_c(a)\}/\sigma(a) \xrightarrow{P} 0$.*

Proposition C.1 is proved in the following Section C.1.3. It implies that the results in Theorem 2.4 sill hold by replacing $\mathcal{U}(a)$ with $\mathcal{U}_c(a)$. As discussed above, we recommend including U-statistics of orders $\{1, 2, 3, \ldots, 6, \infty\}$ in the adaptive testing procedure. Then Proposition C.1 requires that $p = o(n^2)$, which suits a wide range of applications. Combining Theorem 2.4 and Proposition C.1, we can conduct the test with quick computation of cost $O(p^2 n)$.

On the other hand, we can conduct the test more generally without Condition 2.4 and the requirement $p = o(n^2)$. Specifically, we compute $\tilde{\mathcal{U}}(a)$ in (2.5) with cost $O(p^2 n)$. Then $[\tilde{\mathcal{U}}(a) - \mathrm{E}\{\tilde{\mathcal{U}}(a)\}]/\sqrt{\mathrm{var}\{\tilde{\mathcal{U}}(a)\}} \xrightarrow{D} \mathcal{N}(0,1)$ by Lemma A.1 in Supplementary Material and Theorem 2.4. To test H_0 in (2.1), it suffices to estimate $\mathrm{E}\{\tilde{\mathcal{U}}(a)\}$ and $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ with permutation. This may have higher computational cost than the method above due to permutation, but is computationally more efficient than estimating p-values directly via permutation or bootstrap, especially when evaluating small p-values.

C.1.3. *Proof of Proposition C.1 (on Page 199).* In this section, we prove Proposition C.1. As both $\mathcal{U}_c(a)$ and $\mathcal{U}(a)$ are location invariant in the sense of Proposition 2.1, similarly to the proof of Theorem 2.4, we assume $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ in the proofs in this section.

Let $\mathcal{U}_{c,1} = \tilde{\mathcal{U}}(a)$ in (2.5), and $\mathcal{U}_{c,2}(a) = \mathcal{U}_c(a) - \mathcal{U}_{c,1}(a)$. By the proof of Theorem 2.1, we know $\{\mathcal{U}(a) - \mathcal{U}_{c,1}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \xrightarrow{P} 0$. To finish the proof of Proposition C.1, it suffices to prove $\mathcal{U}_{c,2}(a)/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \xrightarrow{P} 0$. By Lemma A.1, $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(p^2 n^{-a})$. Then it suffices to prove $\mathrm{E}\{\mathcal{U}_{c,2}^2(a)\} = o(p^2 n^{-a})$ by the Markov's inequality. To derive $\mathcal{U}_{c,2}(a)$, we similarly use the notation in Section B.3. Specifically, given tuple $\mathbf{i} \in \mathcal{P}(n,a)$, let $\mathbf{i}_{(s_1+s_2+s_3)}$ represent a sub-tuple of $\mathbf{i}$ with length $s_1+s_2+s_3$, and define $\mathcal{S}(\mathbf{i}, s_1+s_2+s_3)$ to be the collection of sub-tuples of $\mathbf{i}$ with length $s_1+s_2+s_3$. Then we write

$$
\begin{aligned}
&\mathcal{U}_{c,2}(a)\\
&= \sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a);\\ 1\le j_1\neq j_2\le p}} \sum_{\substack{0\le s_1,s_2\le a;\\ 0\le s_3<a}} \sum_{\mathbf{i}_{(s_1+s_2+s_3)}\in\mathcal{S}(\mathbf{i},s_1+s_2+s_3)} (\bar{x}_{j_1}\bar{x}_{j_2})^{a-s_1-s_2-s_3}\\
&\quad \times \Big\{(-\bar{x}_{j_2})^{s_1}\prod_{t=1}^{s_1} x_{i_t,j_1}\Big\}\Big\{(-\bar{x}_{j_1})^{s_2}\prod_{t=s_1+1}^{s_1+s_2} x_{i_t,j_2}\Big\}\Big\{\prod_{t=s_1+s_2+1}^{s_1+s_2+s_3} x_{i_t,j_1}x_{i_t,j_2}\Big\}\\
&= \sum_{0\le s_1,s_2\le a;\, 0\le s_3<a} C_{s_1,s_2,s_3} T_{s_1,s_2,s_3},
\end{aligned}
$$

where C_{s_1,s_2,s_3} are some constants that only depend on s_1, s_2, s_3 and a, and

$$
\begin{aligned}
T_{s_1,s_2,s_3} = &\sum_{1\le j_1\neq j_2\le p;\, \mathbf{i}\in\mathcal{P}(n,s_1+s_2+s_3)} \frac{1}{P^n_{s_1+s_2+s_3}} \times (\bar{x}_{j_1}\bar{x}_{j_2})^{a-s_1-s_2-s_3}\\
&\times \Big\{(-\bar{x}_{j_2})^{s_1}\prod_{t=1}^{s_1} x_{i_t,j_1}\Big\}\Big\{(-\bar{x}_{j_1})^{s_2}\prod_{t=s_1+1}^{s_1+s_2} x_{i_t,j_2}\Big\}\Big\{\prod_{t=s_1+s_2+1}^{s_1+s_2+s_3} x_{i_t,j_1}x_{i_t,j_2}\Big\}.
\end{aligned}
$$

When a is finite, it suffices to prove $\mathrm{E}(T^2_{s_1,s_2,s_3}) = o(p^2 n^{-a})$.

Particularly,

$$
\begin{aligned}
\text{(C.3)}\qquad & \mathrm{E}(T^2_{s_1,s_2,s_3}) \\
= & \sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le \tilde{j}_1\ne \tilde{j}_2\le p}} \sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,s_1+s_2+s_3)} \left(\frac{1}{P^n_{s_1+s_2+s_3}}\right)^2 \\
& \times\mathrm{E}\Bigg[(\bar{x}_{j_1}\bar{x}_{j_2})^{a-s_1-s_2-s_3}\Big\{(-\bar{x}_{j_2})^{s_1}\prod_{t=1}^{s_1}x_{i_t,j_1}\Big\}\Big\{(-\bar{x}_{j_1})^{s_2}\prod_{t=s_1+1}^{s_1+s_2}x_{i_t,j_2}\Big\} \\
& \quad\times\Big\{\prod_{t=s_1+s_2+1}^{s_1+s_2+s_3}x_{i_t,j_1}x_{i_t,j_2}\Big\}(\bar{x}_{\tilde{j}_1}\bar{x}_{\tilde{j}_2})^{a-s_1-s_2-s_3}\Big\{(-\bar{x}_{\tilde{j}_2})^{s_1}\prod_{t=1}^{s_1}x_{\tilde{i}_t,\tilde{j}_1}\Big\} \\
& \quad\times\Big\{(-\bar{x}_{\tilde{j}_1})^{s_2}\prod_{t=s_1+1}^{s_1+s_2}x_{\tilde{i}_t,\tilde{j}_2}\Big\}\Big\{\prod_{t=s_1+s_2+1}^{s_1+s_2+s_3}x_{\tilde{i}_t,\tilde{j}_1}x_{\tilde{i}_t,\tilde{j}_2}\Big\}\Bigg] \\
= & \sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le \tilde{j}_1\ne \tilde{j}_2\le p}} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,s_1+s_2+s_3);\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{C}(n,2a-s_1-s_2-2s_3)}} C_{n,s_1,s_2,s_3}M(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},\mathbf{j}),
\end{aligned}
$$

where we define on Page 71 that $\mathbf{w}\in\mathcal{C}(n,s)$ represents tuples $i_1,\dots,i_s$ satisfying $1\le i_1,\dots,i_s\le n$, and $C_{n,s_1,s_2,s_3}=(P^n_{s_1+s_2+s_3}n^{2a-s_1-s_2-s_3})^{-2}$ and

$$
\begin{aligned}
\text{(C.4)}\qquad & M(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},\mathbf{j}) \\
& = \prod_{t=1}^{s_1}x_{i_t,j_1}x_{\tilde{i}_t,\tilde{j}_1}\prod_{t=s_1+1}^{s_1+s_2}x_{i_t,j_2}x_{\tilde{i}_t,\tilde{j}_2}\prod_{t=s_1+s_2+1}^{s_1+s_2+s_3}(x_{i_t,j_1}x_{i_t,j_2})(x_{\tilde{i}_t,\tilde{j}_1}x_{\tilde{i}_t,\tilde{j}_2}) \\
& \qquad\times\prod_{k=1}^{a-s_1-s_3}x_{w_k,j_1}x_{\tilde{w}_k,\tilde{j}_1}\prod_{k=a-s_1-s_3+1}^{2a-s_1-s_2-2s_3}x_{w_k,j_2}x_{\tilde{w}_k,\tilde{j}_2}.
\end{aligned}
$$

We write $M(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},\mathbf{j})=M_{j_1}M_{j_2}M_{\tilde{j}_1}M_{\tilde{j}_2}$, where

$$
\begin{aligned}
M_{j_1}&=\prod_{t=1}^{s_1}x_{i_t,j_1}\prod_{t=s_1+s_2+1}^{s_1+s_2+s_3}x_{i_t,j_1}\prod_{k=1}^{a-s_1-s_3}x_{w_k,j_1},\quad M_{j_2}=\prod_{t=s_1+1}^{s_1+s_2+s_3}x_{i_t,j_2}\prod_{k=a-s_1-s_3+1}^{2a-s_1-s_2-2s_3}x_{w_k,j_2},\\
M_{\tilde{j}_1}&=\prod_{t=1}^{s_1}x_{\tilde{i}_t,\tilde{j}_1}\prod_{t=s_1+s_2+1}^{s_1+s_2+s_3}x_{\tilde{i}_t,\tilde{j}_1}\prod_{k=1}^{a-s_1-s_3}x_{\tilde{w}_k,\tilde{j}_1},\quad M_{\tilde{j}_2}=\prod_{t=s_1+1}^{s_1+s_2+s_3}x_{\tilde{i}_t,\tilde{j}_2}\prod_{k=a-s_1-s_3+1}^{2a-s_1-s_2-2s_3}x_{\tilde{w}_k,\tilde{j}_2}.
\end{aligned}
$$

As $\mathrm{E}(\mathbf{x})=\mathbf{0}$, when $a=1$, $\mathrm{E}(M_{j_1})=\mathrm{E}(M_{j_2})=\mathrm{E}(M_{\tilde{j}_1})=\mathrm{E}(M_{\tilde{j}_2})=0$. We then consider $a\ge 2$. As $\mathrm{E}(\mathbf{x})=\mathbf{0}$, $i_1\ne\ldots\ne i_{s_1+s_2+s_3}$ and $\tilde{i}_1\ne\ldots\ne$

$\tilde{i}_{s_1+s_2+s_3}$, we know that $\mathrm{E}(M_{j_1}) \neq 0$ only when $\{i_1, \ldots, i_{s_1}, i_{s_1+s_2+1}, \ldots, i_{s_1+s_2+s_3}\} \subseteq \{w_1, \ldots, w_{a-s_1-s_3}\}$ and

$$|S_{j_1}| \leq s_1 + s_3 + \lfloor (a - 2s_1 - 2s_3)/2 \rfloor = \lfloor a/2 \rfloor, \tag{C.5}$$

where $S_{j_1} = \{i_1, \ldots, i_{s_1}, i_{s_1+s_2+1}, \ldots, i_{s_1+s_2+s_3}, w_1, \ldots, w_{a-s_1-s_3}\}$. Similarly, when $\mathrm{E}(M_{j_2}) \neq 0$, we know $\{i_{s_1+1}, \ldots, i_{s_1+s_2+s_3}\} \subseteq \{w_{a-s_1-s_3+1}, \ldots, w_{2a-s_1-s_2-2s_3}\}$, and

$$|S_{j_2}| \leq s_2 + s_3 + \lfloor (a - 2s_2 - 2s_3)/2 \rfloor = \lfloor a/2 \rfloor, \tag{C.6}$$

where $S_{j_2} = \{i_{s_1+1}, \ldots, i_{s_1+s_2+s_3}, w_{a-s_1-s_3+1}, \ldots, w_{2a-s_1-s_2-2s_3}\}$. As $|S_{j_1} \cap S_{j_2}| = s_3$, combining (C.5) and (C.6), we know that if $\mathrm{E}(M_{j_1}) \neq 0$ and $\mathrm{E}(M_{j_2}) \neq 0$,

$$|S_{j_1} \cup S_{j_2}| \leq 2\lfloor a/2 \rfloor - s_3 \tag{C.7}$$

Similarly, if $\mathrm{E}(M_{\tilde{j}_1}) \neq 0$, we know

$$|S_{\tilde{j}_1}| \leq \lfloor a/2 \rfloor, \tag{C.8}$$

where $S_{\tilde{j}_1} = \{\tilde{i}_1, \ldots, \tilde{i}_{s_1}, \tilde{i}_{s_1+s_2+1}, \ldots, \tilde{i}_{s_1+s_2+s_3}, \tilde{w}_1, \ldots, \tilde{w}_{a-s_1-s_3}\}$. If $\mathrm{E}(M_{\tilde{j}_2}) \neq 0$, we know

$$|S_{\tilde{j}_2}| \leq \lfloor a/2 \rfloor, \tag{C.9}$$

where $S_{\tilde{j}_2} = \{\tilde{i}_{s_1+1}, \ldots, \tilde{i}_{s_1+s_2+s_3}, \tilde{w}_{a-s_1-s_3+1}, \ldots, \tilde{w}_{2a-s_1-s_2-2s_3}\}$. If $\mathrm{E}(M_{\tilde{j}_1}) \neq 0$ and $\mathrm{E}(M_{\tilde{j}_2}) \neq 0$, we know

$$|S_{\tilde{j}_1} \cup S_{\tilde{j}_2}| \leq 2\lfloor a/2 \rfloor - s_3. \tag{C.10}$$

To evaluate $\mathrm{E}(T^2_{s_1,s_2,s_3})$ in (C.3), for the simplicity of representation, in the following we write

$$\sum_{\text{ALL SUM}} = \sum_{1 \leq j_1 \neq j_2 \leq p;\, 1 \leq \tilde{j}_1 \neq \tilde{j}_2 \leq p} \quad \sum_{\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, s_1+s_2+s_3);\, \mathbf{w}, \tilde{\mathbf{w}} \in \mathcal{C}(n, 2a-s_1-s_2-2s_3)}.$$

We next evaluate $\mathrm{E}(T^2_{s_1,s_2,s_3})$ by discussing the indexes $\{j_1, j_2, \tilde{j}_1, \tilde{j}_2\}$. We first consider $|\{j_1, j_2, \tilde{j}_1, \tilde{j}_2\}| = 4$, and the summation

$$\sum_{\text{ALL SUM}} \mathbf{1}_{\{|\{j_1, j_2, \tilde{j}_1, \tilde{j}_2\}| = 4\}} \times C_{n,s_1,s_2,s_3} \times \mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\}.$$

Note that $|\{j_1, j_2, \tilde{j}_1, \tilde{j}_2\}| = 4$ implies that $j_1 \neq j_2 \neq \tilde{j}_1 \neq \tilde{j}_2$. Without loss of generality, we assume $j_1 < j_2 < \tilde{j}_1 < \tilde{j}_2$, while the other cases can follow similar analysis. Define $\kappa_1 = j_2 - j_1$, $\kappa_2 = \tilde{j}_1 - j_2$ and $\kappa_3 = \tilde{j}_2 - \tilde{j}_1$. In addition, for some small positive constants μ and ϵ and δ in Condition 2.2, define $K_0 = -(2+\epsilon)(4+\mu)(\log p)/(\epsilon \log \delta)$. If $\kappa_m = \max\{\kappa_1, \kappa_2, \kappa_3\} \geq K_0$, we can write

$$|\mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\}| \leq C\delta^{K_0\epsilon/(2+\epsilon)} + \Delta_{j,\tilde{j}}.$$

We next evaluate $\Delta_{j,\tilde{j}}$ by discussing the following cases (a)–(c).

Case (a) If all three $\kappa_1, \kappa_2, \kappa_3 > K_0$, we have

$$\Delta_{j,\tilde{j}} = |\mathrm{E}(M_{j_1})\mathrm{E}(M_{j_2})\mathrm{E}(M_{\tilde{j}_1})\mathrm{E}(M_{\tilde{j}_2})|.$$

Then if $\Delta_{j,\tilde{j}} \neq 0$, we know $\mathrm{E}(M_{j_1}), \mathrm{E}(M_{j_2}), \mathrm{E}(M_{\tilde{j}_1})$ and $\mathrm{E}(M_{\tilde{j}_2}) \neq 0$, which implies that (C.7) and (C.10) hold. By Condition 2.4, we know that

$$\sum_{\text{ALL SUM}} \Delta_{j,\tilde{j}} \mathbf{1}_{\{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4, \kappa_1,\kappa_2,\kappa_3 > K_0\}} = O(1)p^4 n^{4\lfloor a/2 \rfloor - 2s_3}.$$

In addition, $\mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\} \neq 0$ only if $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\} \cup \{\mathbf{w}\} \cup \{\tilde{\mathbf{w}}\}| \leq 2a - s_3$. It follows that

$$\begin{aligned}
&\left| \sum_{\text{ALL SUM}} C_{n,s_1,s_2,s_3} \mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\} \mathbf{1}_{\left\{\substack{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4; \\ \kappa_1,\kappa_2,\kappa_3 > K_0}\right\}} \right| \\
&\leq C \sum_{s_3=0}^{a-1} n^{-2(2a-s_3)} n^{2a-s_3} p^4 C\delta^{K_0\epsilon/(2+\epsilon)} \\
&\quad + \sum_{\text{ALL SUM}} C n^{-2(2a-s_3)} \Delta_{j,\tilde{j}} \mathbf{1}_{\{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4, \kappa_1,\kappa_2,\kappa_3 > K_0\}}, \\
&= o(n^{-(a+1)}) + O(1)p^4 n^{4\lfloor a/2 \rfloor - 4a},
\end{aligned} \tag{C.11}$$

where we use $\sum_{\text{ALL SUM}} \mathbf{1}_{\{\mathrm{E}\{M(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},\mathbf{j})\} \neq 0\}} = \sum_{s_3=0}^{a-1} n^{2a-s_3} p^4$, $\delta^{K_0\epsilon/(2+\epsilon)} = O(1)p^{-(4+\mu)}$, and $C_{n,a,s_1,s_2,s_3} = \Theta(1)n^{-2(2a-s_3)}$. If a is even, (C.11) $= O(1)p^4 n^{-2a} = o(1)p^2 n^{-a}$. If a is odd, (C.11) $= O(1)p^4 n^{-2a-2} = o(1)p^2 n^{-a}$.

Case (b.1) If $\kappa_1 \leq K_0$, $\kappa_2 > K_0$ and $\kappa_3 > K_0$,

$$\Delta_{j,\tilde{j}} = |\mathrm{E}(M_{j_1} M_{j_2})\mathrm{E}(M_{\tilde{j}_1})\mathrm{E}(M_{\tilde{j}_2})|$$

If $\mathrm{E}(M_{\tilde{j}_1})$ and $\mathrm{E}(M_{\tilde{j}_2}) \neq 0$, we know (C.10) holds. We then consider $\mathrm{E}(M_{j_1} M_{j_2})$ with $j_1 \neq j_2$. Note that

$$
\begin{aligned}
& M_{j_1} M_{j_2} \\
= & \prod_{t=1}^{s_1} x_{i_t,j_1} \prod_{t=s_1+1}^{s_1+s_2} x_{i_t,j_2} \prod_{t=s_1+s_2+1}^{s_1+s_2+s_3} (x_{i_t,j_1} x_{i_t,j_2}) \prod_{k=1}^{a-s_1-s_3} x_{w_k,j_1} \prod_{k=a-s_1-s_3+1}^{2a-s_1-s_2-2s_3} x_{w_k,j_2}.
\end{aligned}
$$

As $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ and $\mathrm{E}(x_{1,j_1} x_{1,j_2}) = 0$ under H_0 when $j_1 \neq j_2$, we know $\mathrm{E}(M_{j_1} M_{j_2}) \neq 0$ only when $\{i_1, \ldots, i_{s_1+s_2+s_3}\} \subseteq \{w_1, \ldots, w_{2a-s_1-s_2-2s_3}\}$ and

$$
|S_{j_1} \cup S_{j_2}| \leq \lfloor (2a - s_3)/2 \rfloor \tag{C.12}
$$

We then know $\Delta_{j,\tilde{j}} \neq 0$ only when (C.10) and (C.12) hold, and thus

$$
\begin{aligned}
& \sum_{\text{ALL SUM}} \Delta_{j,\tilde{j}} \times \mathbf{1}_{\{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4, \kappa_1 \leq K_0, \kappa_2, \kappa_3 > K_0\}} \\
& = \sum_{s_3=0}^{a-1} O(1) p^3 K_0 n^{2\lfloor a/2 \rfloor - s_3 + \lfloor (2a-s_3)/2 \rfloor}.
\end{aligned}
$$

Then similarly to (C.11), we have

$$
\begin{aligned}
& \Big| \sum_{\text{ALL SUM}} C_{n,a,s_1,s_2,s_3} \mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\} \mathbf{1}_{\left\{ \substack{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4; \\ \kappa_1 \leq K_0; \kappa_2, \kappa_3 > K_0} \right\}} \Big| \\
& \leq o(n^{-(a+1)}) + \sum_{\text{ALL SUM}} C_{n,a,s_1,s_2,s_3} \Delta_{j,\tilde{j}} \mathbf{1}_{\left\{ \substack{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4; \\ \kappa_1 \leq K_0; \kappa_2, \kappa_3 > K_0} \right\}} \\
& = o(n^{-(a+1)}) + \sum_{s_3=0}^{a-1} O(1) p^3 K_0 n^{2\lfloor a/2 \rfloor - s_3 + \lfloor (2a-s_3)/2 \rfloor - 4a + 2s_3}.
\end{aligned} \tag{C.13}
$$

If a is even, we use $2\lfloor a/2 \rfloor - s_3 + \lfloor (2a - s_3)/2 \rfloor - 4a + 2s_3 \leq -2a + s_3/2 \leq -a - (a+1)/2$ as $s_3 \leq a - 1$. Then (C.13) $= O(1) p^3 K_0 n^{-a-(a+1)/2} = o(1) p^2 n^{-a}$. If a is odd, we use $2\lfloor a/2 \rfloor - s_3 + \lfloor (2a - s_3)/2 \rfloor - 4a + 2s_3 \leq -2a + s_3/2 \leq -a - (a+3)/2$ as $2\lfloor a/2 \rfloor = a - 1$ and $s_3 \leq a - 1$. Then (C.13) $= O(1) p^3 K_0 n^{-a-(a+3)/2} = o(1) p^2 n^{-a}$.

Case (b.2) If $\kappa_1 > K_0$, $\kappa_2 > K_0$ and $\kappa_3 \leq K_0$, similarly to Case (b.1), by symmetricity, we know

$$
\begin{aligned}
& \Big| \sum_{\text{ALL SUM}} C_{n,a,s_1,s_2,s_3} \mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\} \mathbf{1}_{\left\{ \substack{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4; \\ \kappa_1, \kappa_2 > K_0; \kappa_3 \leq K_0} \right\}} \Big| \\
& = o(n^{-(a+1)}) + \sum_{s_3=0}^{a-1} O(1) p^3 K_0 n^{2\lfloor a/2 \rfloor - s_3 + \lfloor (2a-s_3)/2 \rfloor - 4a + 2s_3}.
\end{aligned} \tag{C.14}
$$

Then (C.16) $= o(1)p^2 n^{-a}$.

Case (b.3) If $\kappa_1 > K_0$, $\kappa_2 \le K_0$ and $\kappa_3 > K_0$,

$$\Delta_{j,\tilde{j}} = |\mathrm{E}(M_{j_1})\mathrm{E}(M_{j_2}M_{\tilde{j}_1})\mathrm{E}(M_{\tilde{j}_2})|.$$

If $\mathrm{E}(M_{j_1}), \mathrm{E}(M_{\tilde{j}_2}) \neq 0$, we know (C.5) and (C.8) hold. We then consider $\mathrm{E}(M_{j_2}M_{\tilde{j}_1})$. Note that

$$\begin{aligned} & M_{j_2}M_{\tilde{j}_1} \\ &= \prod_{t=s_1+1}^{s_1+s_2+s_3} x_{i_t,j_2} \prod_{t=a-s_1-s_3+1}^{2a-s_1-s_2-2s_3} x_{w_t,j_2} \prod_{t=1}^{s_1} x_{\tilde{i}_t,\tilde{j}_1} \prod_{t=s_1+s_2+1}^{s_1+s_2+s_3} x_{\tilde{i}_t,\tilde{j}_1} \prod_{t=1}^{a-s_1-s_3} x_{\tilde{w}_t,\tilde{j}_1}. \end{aligned}$$

If $\mathrm{E}(M_{j_2}M_{\tilde{j}_1}) \neq 0$, we know that $|S_{j_2} \cup S_{\tilde{j}_1}| \le a$. As $|(S_{j_2} \cup S_{\tilde{j}_1}) \cap (S_{j_1} \cup S_{\tilde{j}_2})| = 2s_3$, we have $|S_{j_1} \cup S_{j_2} \cup S_{\tilde{j}_1} \cup S_{\tilde{j}_2}| \le a + 2\lfloor a/2 \rfloor - 2s_3$. We then know

$$\sum_{\text{ALL SUM}} \Delta_{j,\tilde{j}} \mathbf{1}_{\{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4, \kappa_1,\kappa_3 > K_0, \kappa_2 \le K_0\}} = \sum_{s_3=0}^{a-1} O(1)p^3 K_0 n^{a+2\lfloor a/2 \rfloor - 2s_3}.$$

Then similarly to (C.13), we have

$$\begin{aligned} \text{(C.15)} \qquad & \Bigg| \sum_{\text{ALL SUM}} C_{n,a,s_1,s_2,s_3} \mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\} \mathbf{1}_{\left\{\substack{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4; \\ \kappa_1,\kappa_3 > K_0; \kappa_2 \le K_0}\right\}} \Bigg| \\ &= o(n^{-(a+1)}) + O(1)p^3 K_0 n^{2\lfloor a/2 \rfloor - 3a}. \end{aligned}$$

If a is even, we know (C.15) $= p^3 K_0 n^{-2a} = o(1)p^2 n^{-a}$. If a is odd, we know (C.15) $= p^3 K_0 n^{-2a-1} = o(1)p^2 n^{-a}$.

Case (c) If two of $\kappa_1, \kappa_2, \kappa_3 \le K_0$, we know

$$\sum_{j_1,j_2,\tilde{j}_1,\tilde{j}_2} \mathbf{1}_{\{\text{two of } \kappa_1,\kappa_2,\kappa_3 \le K_0\}} = O(p^2 K_0^2).$$

Following definition in (C.4), we know $\mathrm{E}\{M(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, \mathbf{j})\} \neq 0$ only when $|S_{j_1} \cup S_{j_2} \cup S_{\tilde{j}_1} \cup S_{\tilde{j}_2}| \le 2a - s_3$. It implies that

$$\sum_{\text{ALL SUM}} \Delta_{j,\tilde{j}} \mathbf{1}_{\{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4, \text{ two of } \kappa_1,\kappa_2,\kappa_3 \le K_0\}} = O(1)p^2 K_0^2 n^{2a-s_3}.$$

Similarly to (C.15), we have

$$\text{(C.16)}\quad \Big|\sum_{\text{ALL SUM}} C_{n,a,s_1,s_2,s_3}\mathrm{E}\{M(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},\mathbf{j})\}\mathbf{1}_{\left\{\substack{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=4;\\ \text{two of } \kappa_1,\kappa_2,\kappa_3\le K_0}\right\}}\Big| = o(n^{-(a+1)}) + \sum_{s_3=0}^{a-1} O(1)p^2K_0^2 n^{-2a+s_3}.$$

As $s_3 \le a-1$ and $K_0 = O(\log p)$, we know (C.16) $= O(1)p^2K_0^2n^{-a-1} = o(1)p^2n^{-a}$.

Case (d) If $|\{j_1, j_2, j_3, j_4\}| = 3$ or 2, similar analysis can be applied, and we know that

$$\text{(C.17)}\quad \Big|\sum_{\text{ALL SUM}} C_{n,a,s_1,s_2,s_3}\mathrm{E}\{M(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},\mathbf{j})\}\mathbf{1}_{\{|\{j_1,j_2,\tilde{j}_1,\tilde{j}_2\}|=2 \text{ or } 3\}}\Big| = o(n^{-(a+1)}) + o(1)p^2n^{-2a}.$$

Summarizing Cases (a)–(d) above, we obtain $\mathrm{E}(T^2_{s_1,s_2,s_3}) = o(p^2n^{-a})$.

C.2. Simulations on One-Sample Covariance Testing. In this section, we provide extensive simulation studies for the one-sample covariance testing discussed in Section 2. We present the results of the five simulation settings introduced in Section 3.1 in the following Sections C.2.1–C.2.5.

C.2.1. *Study 1: Empirical Size.* In this study, we verify the theoretical results under H_0 in Section 2 and the show validity of the adaptive testing procedure across different n and p values. In particular, we fix $n = 100$ and take $p \in \{50, 100, 200, 400, 600, 800, 1000\}$. Then we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \ldots, n$, and each $\mathbf{x}_i$ has i.i.d. entries of $\mathcal{N}(0, 1)$ and Gamma$(2, 0.5)$ respectively. The results are summarized in the following Tables 1 and 2 respectively.

In Tables 1 and 2, we provide the simulation results of all the single U-statistics with orders in $\{1, \ldots, 6\}$. For $\mathcal{U}(\infty)$, we first use the test statistic (2.8) same as in Jiang [43], which is denoted as "$\mathcal{U}(\infty)$ 1" below. Since the convergence in [43] is slow, we use permutation to approximate the distribution in the simulations. We also use the standardized version $M_n^\dagger$ given in Remark 2.4, which is denoted as "$\mathcal{U}(\infty)$ 2" below. Given "$\mathcal{U}(\infty)$ 1" and "$\mathcal{U}(\infty)$ 2", we apply the adaptive testing with minimum combination and Fisher's method respectively. The results are denoted as "adpUmin1", "adpUf1", "adpUmin2" and "adpUf2" respectively below. In addition, we also compare several methods in the literature. The identity and sphericity

TABLE 1
Empirical Type I errors under Guassian distribution; $n = 100$.

p	50	100	200	400	600	800	1000
$\mathcal{U}(1)$	0.054	0.055	0.045	0.053	0.048	0.052	0.036
$\mathcal{U}(2)$	0.058	0.058	0.066	0.050	0.071	0.048	0.063
$\mathcal{U}(3)$	0.057	0.066	0.061	0.055	0.051	0.063	0.052
$\mathcal{U}(4)$	0.054	0.067	0.052	0.080	0.053	0.041	0.056
$\mathcal{U}(5)$	0.049	0.054	0.059	0.070	0.045	0.049	0.053
$\mathcal{U}(6)$	0.039	0.057	0.063	0.061	0.056	0.057	0.074
$\mathcal{U}(\infty)$ 1	0.046	0.055	0.049	0.067	0.064	0.042	0.044
$\mathcal{U}(\infty)$ 2	0.040	0.047	0.045	0.056	0.048	0.050	0.048
adpUmin 1	0.056	0.066	0.067	0.064	0.067	0.056	0.051
adpUf 1	0.065	0.083	0.069	0.079	0.063	0.058	0.060
adpUmin 2	0.054	0.069	0.065	0.060	0.062	0.055	0.057
adpUf 2	0.069	0.082	0.065	0.065	0.058	0.057	0.062
Identity	0.055	0.053	0.058	0.053	0.061	0.049	0.053
Sphericity	0.053	0.050	0.058	0.053	0.062	0.049	0.054
LW	0.058	0.051	0.053	0.045	0.067	0.048	0.058
Schott	0.052	0.055	0.050	0.052	0.050	0.044	0.051

TABLE 2
Empirical Type I errors under Gamma distribution; $n = 100$.

p	50	100	200	400	600	800	1000
$\mathcal{U}(1)$	0.043	0.049	0.054	0.048	0.050	0.049	0.043
$\mathcal{U}(2)$	0.057	0.075	0.062	0.054	0.057	0.055	0.061
$\mathcal{U}(3)$	0.054	0.064	0.050	0.041	0.057	0.051	0.056
$\mathcal{U}(4)$	0.047	0.056	0.061	0.056	0.052	0.053	0.045
$\mathcal{U}(5)$	0.043	0.043	0.054	0.052	0.050	0.053	0.049
$\mathcal{U}(6)$	0.032	0.035	0.059	0.045	0.046	0.053	0.044
$\mathcal{U}(\infty)$ 1	0.052	0.045	0.048	0.053	0.045	0.049	0.055
$\mathcal{U}(\infty)$ 2	0.044	0.052	0.052	0.053	0.044	0.051	0.045
adpUmin 1	0.051	0.054	0.069	0.062	0.049	0.058	0.065
adpUf 1	0.055	0.060	0.075	0.067	0.054	0.058	0.067
adpUmin 2	0.049	0.055	0.068	0.063	0.049	0.059	0.066
adpUf 2	0.063	0.067	0.070	0.058	0.047	0.057	0.061
Identity	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Sphericity	0.088	0.065	0.071	0.056	0.060	0.059	0.050
LW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Schott	0.051	0.063	0.053	0.053	0.055	0.046	0.060

tests in Chen et al. [15] are denoted as "Equal" and "Spher" below; the methods in Ledoit and Wolf [51] and Schott [66], which are referred to as "LW" and "Schott" respectively.

C.2.2. *Study 2.* In this section, we provide the simulation results for the second setting in Section 3. In particular, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \dots, n$, and $\mathbf{x}_i$ follows multivariate Gaussian distribution with mean zero and covariance $\mathbf{\Sigma}_A = (1-\rho)I_p + \rho \mathbf{1}_{p,k_0}\mathbf{1}_{p,k_0}^\intercal$.

Similarly to Figure 2, we conduct simulations on the adaptive procedure with U-statistics of orders in $\{1, \dots, 6, \infty\}$. We provide the simulation results of all the single U-statistics and the adaptive procedure, and also compare with some other methods in the literature. We take $(n, p) \in \{(100, 300), (100, 600), (100, 1000)\}$, and provide the results in the following Figures 4–6 respectively.

In Figure 4, the first 7 plots are simulated with $k_0 \in \{2, 5, 7, 10, 13, 20, 50\}$. Particularly, we include results of $\mathcal{U}(a)$ for $a \in \{1, \dots, 6, \infty\}$; the adaptive procedure "adpU" by minimum combination of these single U-statistics; identity and sphericity tests in [15], which are denoted as 'Equal" and "Sh-per", respectively. We can see that when $k_0 \in \{7, 10, 13\}$, the results of "adpU" are better than all the other test statistics. For other cases, the results of "adpU" are close to the best results of single U-statistics. In addition, we also examine the case when the nonzero off-diagonal elements of $\mathbf{\Sigma}_A$, i.e., σ_{j_1,j_2} with $1 \leq j_1 \neq j_2 \leq k_0$, have same absolute value $|\rho|$, but can be positive or negative with equal probability. The results of powers versus different $|\rho|$ values are given by 8th plot in Figure 4, which is consistent with Remark 2.6 in Section 2.2.

In Figures 5 and 6, the meanings of the legends are the same as in Tables 1 and 2, and are already explained in Section C.2.1. We can find similar patterns to that in Figure 4.

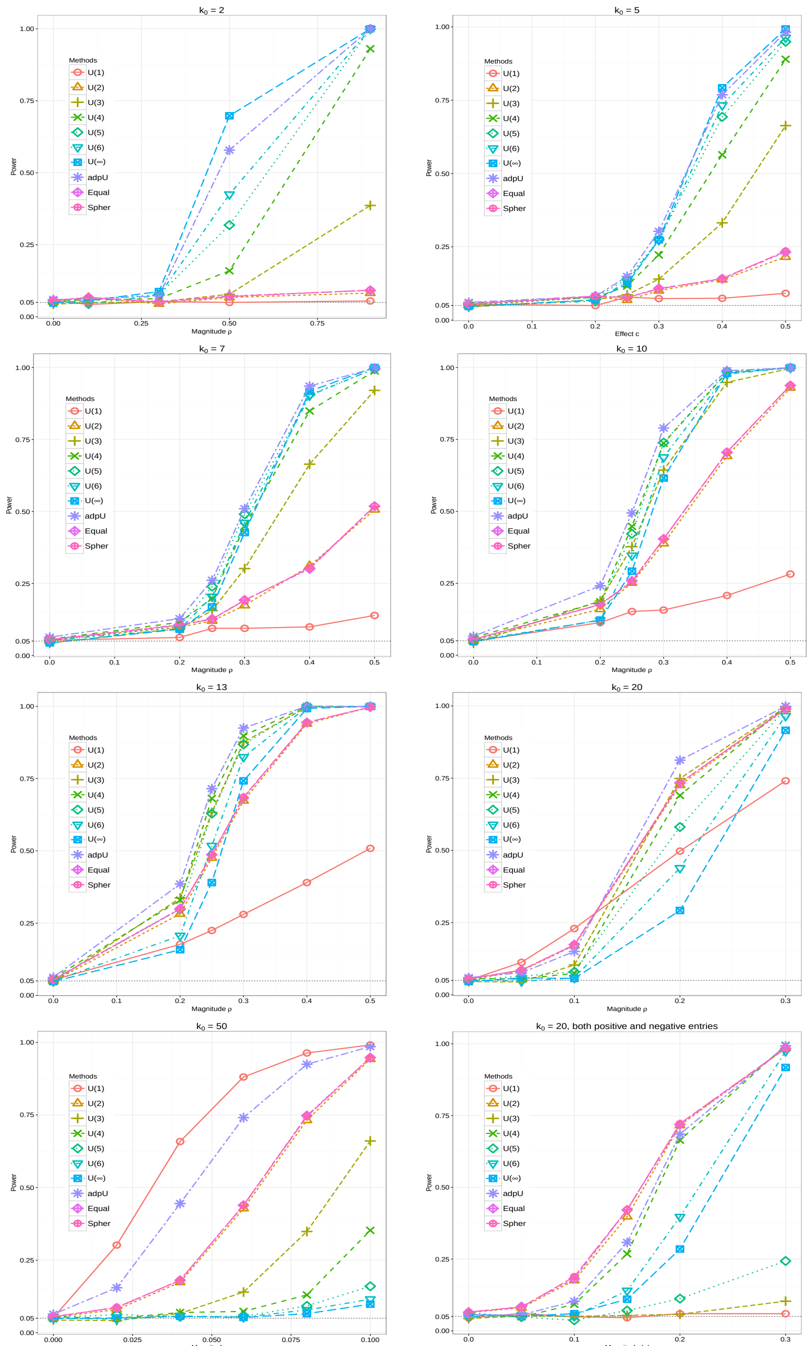

Fig 4: Study 2: $n = 100, p = 300$.

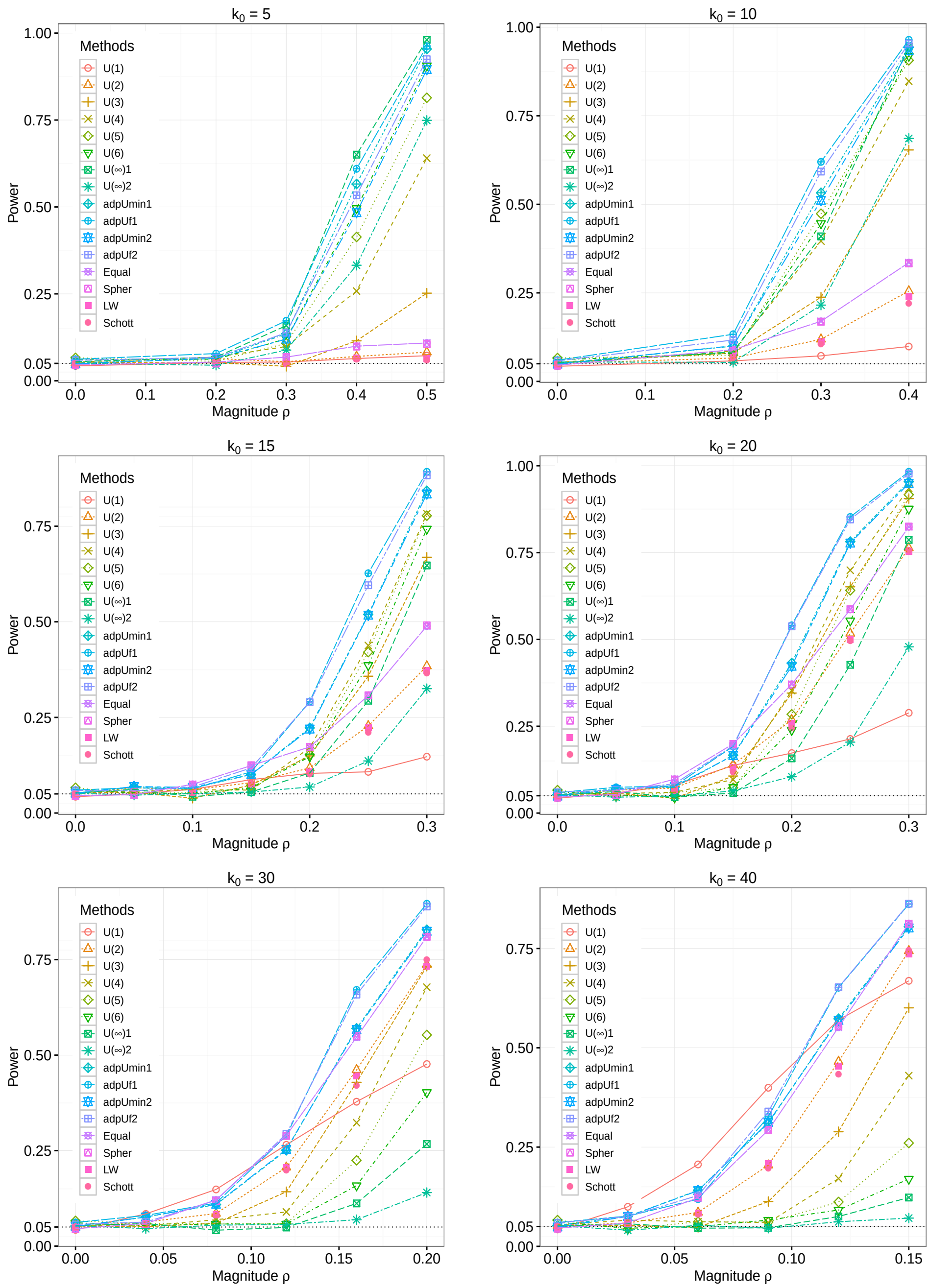

Fig 5: Study 2: $n = 100, p = 600$.

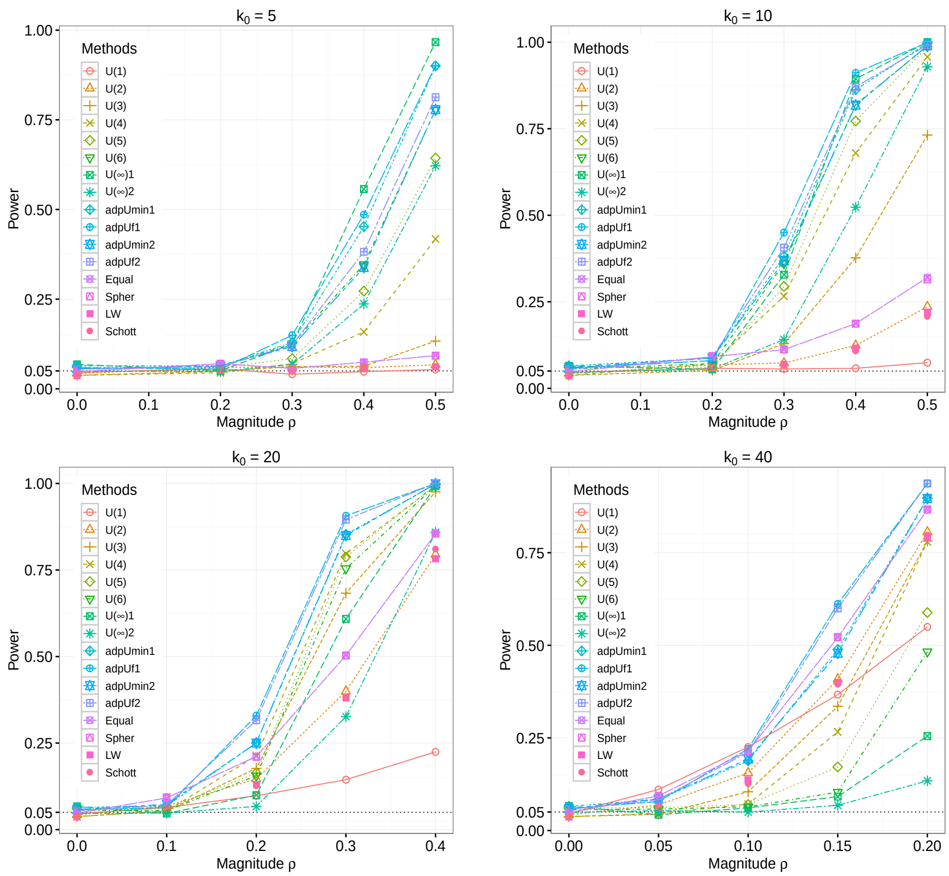

Fig 6: Study 2: $n = 100, p = 1000$.

C.2.3. *Study 3.* We provide supplementary simulations for the third setting in Section 3.1. In particular, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \dots, n$, and $\mathbf{x}_i$ follows multivariate Gaussian distribution with mean zero and covariance $\mathbf{\Sigma}_A$. In this case, $\mathbf{\Sigma}_A$ is symmetric and positive definite, and has the diagonal being all one and only $|J_A|$ random positions being nonzero with value ρ. Note that here ρ represents the magnitude of the alternative signal; and $|J_A|$ represents its sparsity level with a larger value indicating a denser alternative, and vice versa. We let $|J_A|$ and ρ vary to examine how the power changes correspondingly. We take $(n, p) \in \{(100, 600), (100, 1000)\}$, and provide the results in the following Figures 7–8 respectively. The meanings of the legends are the same as in Tables 1 and 2, and are already explained in Section C.2.1. We observe similar patterns to that in the figures in Section C.2.2.

C.2.4. *Study 4.* In this section, we provide the simulation results of the fourth setting in Section 3.1. In particular, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \dots, n$, and $\mathbf{x}_i$ follows multivariate Gaussian distribution with mean zero and covariance $\mathbf{\Sigma}_A$. Under this setting, $\mathbf{\Sigma}_A$ is symmetric and positive definite and has the diagonal being all one and $|J_A|$ random positions taking values uniformly in the range $(0, 2\rho)$. Therefore, the nonzero off-diagonal elements in $\mathbf{\Sigma}_A$ are different. Figure 9 below presents the power versus ρ when $n = 100$ and $p = 1000$. The meanings of the legends are the same as in Tables 1 and 2, and are already explained in Section C.2.1. We observe similar patterns to that in the figures in Section C.2.2.

C.2.5. *Study 5.* In this section, we compare our methods with the methods in Chen et al. [15] following their multivariate models. Specifically, for each $i = 1, \dots, n$, $\mathbf{x}_i = \Xi \mathbf{z}_i + \boldsymbol{\mu}$, where Ξ is a matrix of dimension $p \times m$ with $m \geq p$. Under null hypothesis, $m = p$, $\Xi = I_p$ $\boldsymbol{\mu} = \mu_0 \mathbf{1}_p$ with $\mu_0 = 2$; under alternative hypothesis, $m = p + 1$, $\boldsymbol{\mu} = 2(\sqrt{1-\rho} + \sqrt{2\rho})\mathbf{1}_p$, $\Xi = (\sqrt{1-\rho} I_p, \sqrt{2\rho}\mathbf{1}_p)$, thus $\mathbf{\Sigma} = (1-\rho)I_p + 2\rho \mathbf{1}_p \mathbf{1}_p^\intercal$. Two settings are examined: first, $\mathbf{z}_i$'s are i.i.d. multivariate Gaussian random vectors with mean $\mathbf{0}$ and covariance I_p; second, $\mathbf{z}_i = (z_{i,1}, \dots, z_{i,m})^\intercal$ consists of i.i.d. random variables $z_{i,j}$ which are standardized Gamma$(4, 0.5)$ random variables so that $\mathbf{z}_i$ has mean $\mathbf{0}$ and covariance I_p.

To mimic "large p, small n" situation, [15] sets dimension $p = c_1 \exp(n^\eta) + c_2$, where $\eta = 0.4$, for $(c_1, c_2) = (1, 10)$ and $(c_1, c_2) = (2, 0)$ respectively. In particular, we consider $(n, p) \in \{(40, 159), (40, 331), (80, 159), (80, 331), (80, 642)\}$. The results are based on 1000 simulations and the nominal significance level of the tests is 5%.

In the tables 3–10, results outside and inside parentheses are calculated

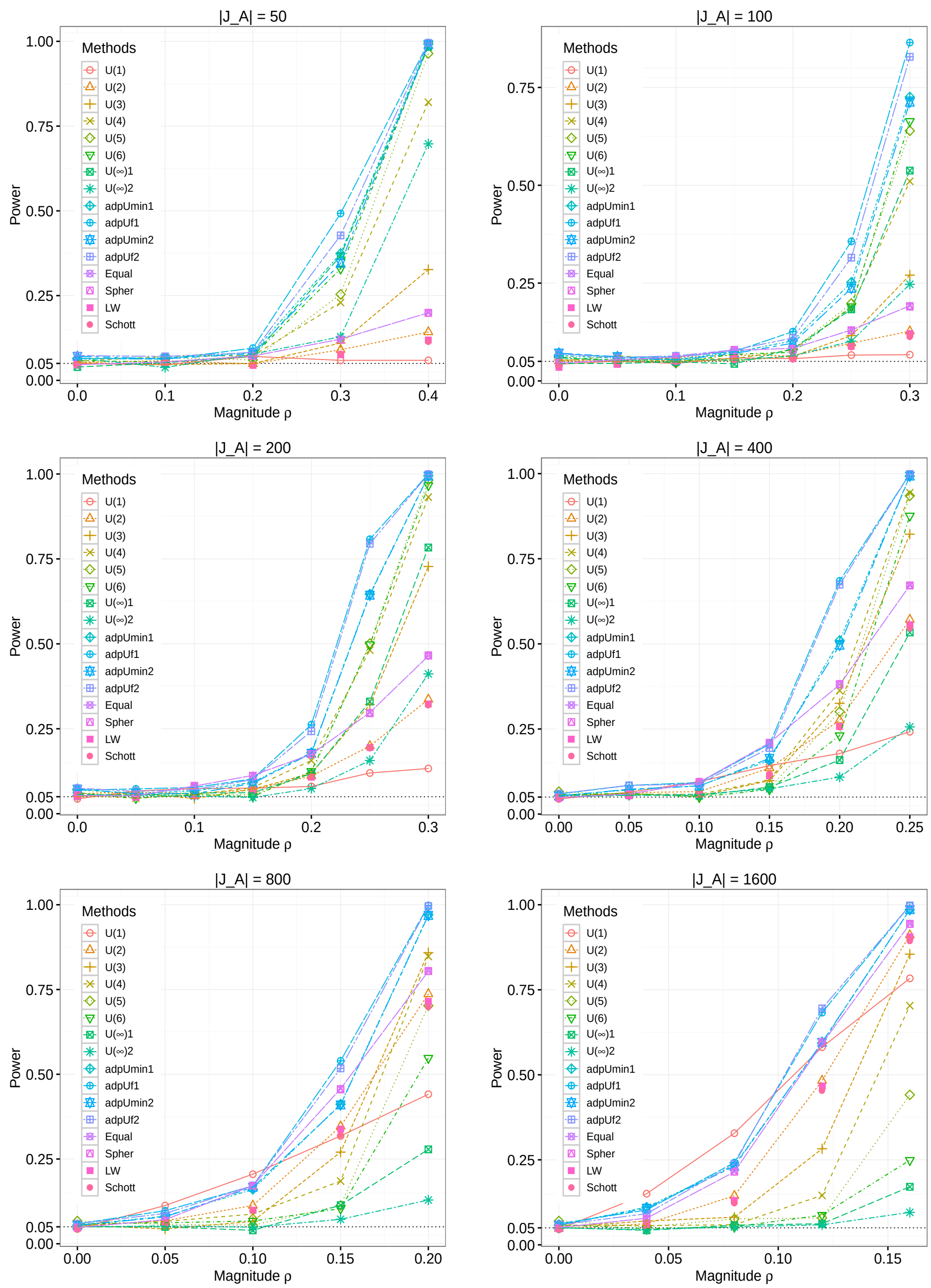

Fig 7: Study 3: $n = 100, p = 600$.

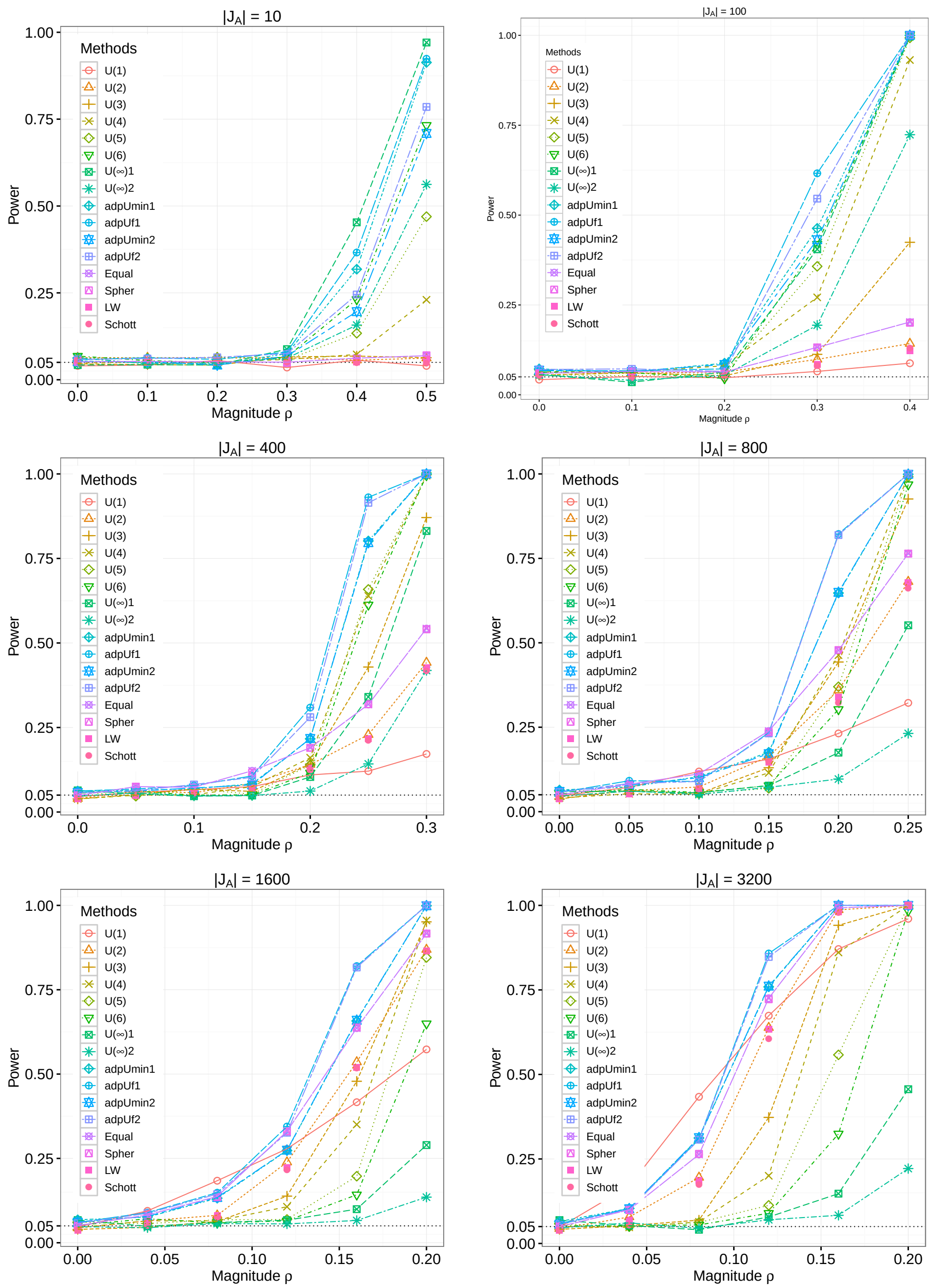

Fig 8: Study 3: $n = 100, p = 1000$.

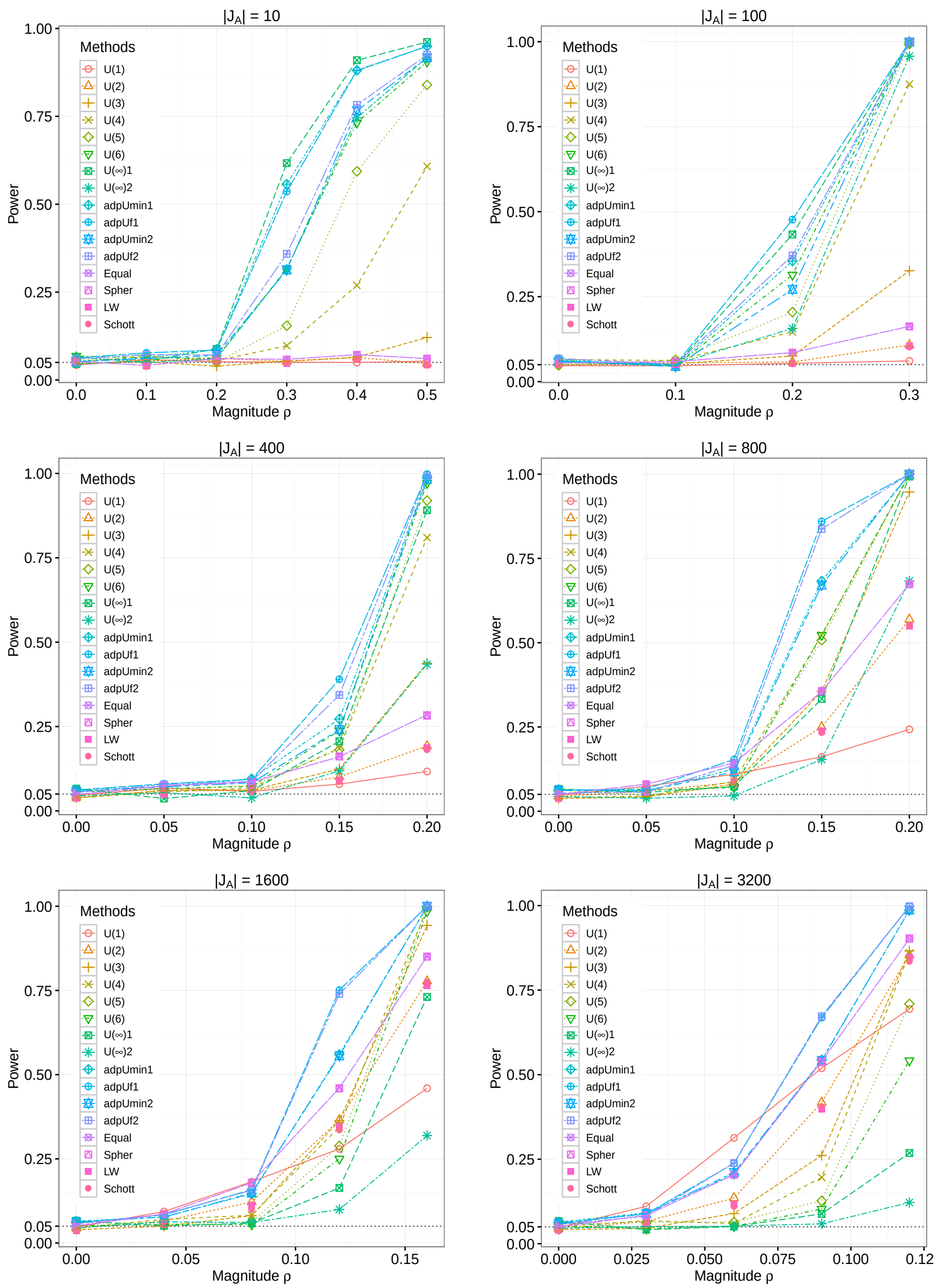

Fig 9: Study 4: $n = 100, p = 1000$.

from parametric-permutation- and asymptotics-based methods, respectively. To be specific, psarametric-permutation-based method means estimating p-values or powers by permutation; and asymptotic-based method uses the asymptotic theoretical results and is described in Section 2.3. For each $a \in \{1, \ldots, 6, \infty\}$, the row of "$\mathcal{U}(a)$" has results using the single test statistic $\mathcal{U}(a)$; and the row of "adpU" is obtained by the adaptive testing procedure which combines all single candidate U-statistics in the tables using the minimum combination. In addition, "Ident" and "Spher" rows denote the identity and sphericity tests in [15] separately.

In the tables 3–8, we find that the empirical sizes of most tests are close to the nominal level, except $\mathcal{U}(\infty)$ due to the slow convergence to extreme value distribution as pointed out in [31]. "Ident" and "Spher" tests perform similarly to $\mathcal{U}(2)$ in both settings. This is reasonable because they are all sum-of-squares-type statistics. Moreover, for the ρ's examined, $\mathcal{U}(1)$ has higher power than $\mathcal{U}(2)$, as the constructed alternative is very dense and only has positive entries. In addition, "adpU" achieves high power for different cases, and its power converges to 1, as one of the test statistics has power converging to 1. In Tables 9 and 10, data are standardized with sample mean and variance. It can be seen that methods in [15] perform poorly in this case. Other than this, the results follow similar patterns to results in other tables.

TABLE 3
Empirical Type I errors and power (%) under simulation setting 1. $n = 80, p = 331$.

ρ	0	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	4.4 (4)	93.4 (90.6)	100 (99.9)	100 (100)	100 (100)
$\mathcal{U}(2)$	5 (5.6)	5.5 (6)	7.2 (5.9)	13.1 (10.2)	19.7 (14.4)
$\mathcal{U}(3)$	5.4 (6.1)	4.5 (4)	6.3 (5.4)	6.9 (4.5)	9 (5.4)
$\mathcal{U}(4)$	4.7 (5.1)	6 (5.4)	3.7 (4.6)	4.2 (5.3)	6 (4.8)
$\mathcal{U}(5)$	5.4 (6.3)	4.9 (4.7)	5.3 (5.6)	6 (5.7)	6.1 (5.1)
$\mathcal{U}(6)$	4.6 (4.9)	5.8 (5.4)	4.9 (4.5)	5.2 (4.8)	4.8 (5)
$\mathcal{U}(\infty)$	4.7 (0.3)	5 (0.6)	5.5 (0.7)	5.1 (0.4)	5.9 (0.8)
aSPU	5 (5.4)	81 (81.8)	99.4 (99.4)	100 (100)	100 (100)
Ident	5.5	5.7	8.2	14.4	21.8
Spher	5.6	5.7	8.1	14.2	21.4

TABLE 4
Empirical Type I errors and power (%) under simulation setting 2; $n = 80, p = 331$.

ρ	0	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	5.3 (4.6)	56.7 (50.3)	92.5 (89.3)	99.3 (99.1)	100 (99.8)
$\mathcal{U}(2)$	5.4 (5)	5.5 (5.7)	6.9 (5.4)	7.7 (5.8)	11.4 (7.3)
$\mathcal{U}(3)$	5.6 (5.4)	4.5 (3.5)	5.7 (4)	5.8 (4.8)	7.2 (5.1)
$\mathcal{U}(4)$	4.8 (3.9)	4.9 (4.1)	4.9 (5)	6.5 (6.8)	4.9 (5.1)
$\mathcal{U}(5)$	6.1 (5.1)	5.6 (6.1)	5.1 (5.2)	5.5 (5.7)	5.2 (5.5)
$\mathcal{U}(6)$	6.4 (5.6)	5.4 (4.1)	5.1 (5.3)	5.1 (5.4)	5.8 (5.3)
$\mathcal{U}(\infty)$	5.5 (3)	5.3 (2.5)	6 (2.8)	5.5 (2.8)	6.8 (3.1)
adpU	6.4 (6.5)	35 (36.3)	78.7 (79.2)	96.1 (96.1)	99.5 (99.6)
Ident	6.7	6.5	7.4	9.2	13.5
Spher	6.2	6.2	7	9.1	12.9

TABLE 5
Empirical Type I errors and power (%) under simulation setting 1; $n = 40, p = 159$.

ρ	0	0.0005	0.001	0.0015	0.002	0.0025
$\mathcal{U}(1)$	5.8 (4.6)	16.6 (13.6)	36.5 (32.3)	57.4 (51.3)	69.2 (65.1)	83.3 (80)
$\mathcal{U}(2)$	5.2 (4.9)	4.6 (3.1)	4.6 (5.6)	5.3 (4.5)	5.5 (4.8)	5.9 (4.8)
$\mathcal{U}(3)$	4.9 (4.8)	5.8 (5.4)	5.6 (5.6)	5.6 (4.9)	4.6 (4.7)	5.6 (5)
$\mathcal{U}(4)$	4.6 (5.7)	4.2 (4.1)	5.6 (4.6)	4.7 (4.6)	4.5 (5.1)	5.3 (4.9)
$\mathcal{U}(5)$	5.5 (5.6)	5.3 (6.2)	5.7 (4.9)	3.1 (3.1)	4.7 (4.4)	5.5 (5.4)
$\mathcal{U}(6)$	4.4 (4.3)	4.8 (4.6)	4.4 (4.7)	4.3 (4.3)	4.8 (4.6)	5 (4.2)
$\mathcal{U}(\infty)$	5.1 (0.1)	5.1 (0.1)	4.2 (0)	4.6 (0.1)	4.6 (0)	5.5 (0.1)
adpU	5.7 (5.8)	8.9 (10.6)	18.5 (21.1)	31.5 (34.2)	47.4 (50.8)	63.2 (66.2)
Ident	5.8	5.3	5.9	6.8	6.8	7.1
Spher	5.8	5.1	5.7	6.5	6.5	7.2

TABLE 6
Empirical Type I errors and power (%) under simulation setting 1; $n = 40, p = 331$.

ρ	0	0.0025	0.005	0.01	0.015	0.02
$\mathcal{U}(1)$	5.9 (5.4)	99.4 (99.3)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	5.1 (4.4)	7 (6.3)	15.5 (10.7)	65.8 (60)	95.1 (93.1)	99.3 (98.7)
$\mathcal{U}(3)$	5.4 (5.5)	7.6 (4.6)	13 (7.5)	26.3 (19.7)	53.9 (44.1)	76.9 (68.9)
$\mathcal{U}(4)$	4.8 (5.1)	4.9 (5.4)	6.8 (5.6)	6.3 (6.6)	11.4 (7.7)	14.4 (11.7)
$\mathcal{U}(5)$	5.9 (4.8)	5.5 (4.9)	7 (6.6)	5.6 (4.9)	8.6 (7.3)	8.5 (8.2)
$\mathcal{U}(6)$	4.1 (4.9)	3.4 (4.5)	6.8 (4.6)	4.8 (6.5)	5.5 (6.6)	8 (8.6)
$\mathcal{U}(\infty)$	4.2 (0)	4.1 (0)	6.1 (0)	4.9 (0)	6.6 (0)	7.3 (0.1)
adpU	5.2 (5.8)	97.5 (98.5)	100 (100)	100 (100)	100 (100)	100 (100)
Ident	6.2	8.3	19.2	68	95.5	99.3
Spher	6.3	8.2	18.6	67.6	95.4	99.3

TABLE 7
Empirical Type I errors and power (%) under simulation setting 1; $n = 80, p = 159$.

ρ	0	0.0025	0.005	0.01	0.015	0.02
$\mathcal{U}(1)$	5.7 (4.7)	98.1 (97)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	6.2 (5.1)	6.8 (5.5)	16.5 (11.4)	68.4 (60.6)	96.7 (94.7)	100 (99.9)
$\mathcal{U}(3)$	6 (4.7)	6.2 (5.5)	7.4 (5.9)	15.2 (9.2)	34.8 (26.2)	69.2 (61.4)
$\mathcal{U}(4)$	5.4 (5.6)	4 (3.8)	4.7 (4.2)	7.6 (7.1)	10.6 (9)	18.2 (15.7)
$\mathcal{U}(5)$	4.5 (4.9)	4.6 (4.2)	4.8 (4.5)	5.3 (5.3)	9.6 (7.6)	13.1 (13)
$\mathcal{U}(6)$	5.6 (5.3)	3.9 (4.7)	4 (3.3)	5.3 (4.9)	8.7 (8)	12 (12.4)
$\mathcal{U}(\infty)$	4.5 (0.8)	6.1 (1.1)	4.9 (1.4)	5.4 (1.7)	8 (1.5)	10.7 (3.3)
adpU	5.7 (7)	91.8 (92.6)	99.8 (99.8)	100 (100)	100 (100)	100 (100)
Ident	6.7	7.8	18.5	71.1	97.3	100
Spher	6.7	7.2	18	69.6	97	100

TABLE 8
Empirical Type I errors and power (%) under simulation setting 1; $n = 80, p = 642$.

ρ	0	0.0025	0.005	0.01	0.015	0.02
$\mathcal{U}(1)$	5.8 (4.8)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	6.4 (6.2)	17.9 (12.7)	71.2 (63.4)	99.8 (99.8)	100 (100)	100 (100)
$\mathcal{U}(3)$	5.2 (5.6)	6.2 (3.6)	19.3 (13.3)	68.4 (57.3)	96.4 (94)	99.8 (99.6)
$\mathcal{U}(4)$	5.2 (5.2)	6.2 (6.4)	5.2 (5.2)	8.5 (6.4)	25 (18.3)	57.9 (51.7)
$\mathcal{U}(5)$	6.4 (4.6)	5 (5.2)	6.4 (5.4)	7.8 (7.2)	11.7 (9.9)	21.1 (16.9)
$\mathcal{U}(6)$	4 (4.2)	5.8 (6.4)	6 (6)	4.2 (5.2)	9.3 (10.3)	13.1 (15.3)
$\mathcal{U}(\infty)$	4.4 (0.6)	5 (0.2)	5.6 (0.4)	7 (0.8)	9.3 (0.8)	15.3 (0.6)
adpU	6 (4.2)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
Ident	6.8	18.9	72.6	100	100	100
Spher	6.6	18.7	72.6	100	100	100

TABLE 9
Empirical Type I errors and power (%) under simulation setting 2; $n = 80, p = 159$.

ρ	0	0.0005	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	4.9 (4.2)	26.1 (20.4)	57.1 (49.7)	95.2 (93.1)	99.9 (99.8)	100 (99.9)
$\mathcal{U}(2)$	4.9 (4.4)	3.9 (5.3)	5.9 (5.2)	6.7 (4.8)	8.3 (5.6)	12.2 (7.7)
$\mathcal{U}(3)$	5.4 (5.2)	4.7 (5.3)	4.3 (4.1)	6 (4)	5.9 (5.1)	7 (5)
$\mathcal{U}(4)$	5.4 (4.9)	5.5 (5.2)	4.8 (4.8)	5.9 (6.3)	6.7 (7.2)	4.6 (4.6)
$\mathcal{U}(5)$	7.3 (6.2)	5.4 (5.6)	5.8 (6.5)	5.3 (6.3)	5.8 (5.5)	5.6 (5.6)
$\mathcal{U}(6)$	6.5 (5.6)	4.9 (5)	5.5 (5.3)	4.9 (5.2)	5.5 (5.4)	4.2 (4.7)
$\mathcal{U}(\infty)$	5.9 (3)	5.7 (2.1)	5.8 (2.5)	5.7 (2.6)	5.5 (2.9)	6.7 (3.3)
adpU	5.7 (5)	12.1 (13.1)	34.8 (34.6)	81.9 (82.6)	98.1 (98.1)	99.9 (99.8)
Ident	0.2	0.1	0.1	0.2	0.1	0.1
Spher	0.2	0.1	0.1	0.2	0	0.1

TABLE 10

Empirical Type I errors and power (%) under simulation setting 2; $n = 80, p = 642$.

ρ	0	0.0005	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	2.8 (2.2)	94.2 (93)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	5.8 (4.2)	4.2 (4.8)	6 (5.6)	11.9 (7.2)	22.3 (14.5)	45.9 (36.2)
$\mathcal{U}(3)$	3.6 (3.8)	5.4 (5.2)	7.2 (5)	6 (3.6)	11.9 (7.6)	15.1 (9.3)
$\mathcal{U}(4)$	4.4 (4.4)	4.6 (4.4)	6.4 (6.2)	4.8 (3.8)	5.4 (5.2)	7 (6.2)
$\mathcal{U}(5)$	7 (5.6)	6 (5)	6.2 (5.4)	7 (6.2)	6.6 (5.4)	7.4 (5.6)
$\mathcal{U}(6)$	7 (5.4)	5 (4.6)	4.6 (5.6)	6.8 (7.2)	5.4 (4.6)	5.6 (5.8)
$\mathcal{U}(\infty)$	4.8 (2.2)	6.2 (2.4)	4.8 (0.8)	6.2 (3)	6.4 (2.6)	5.2 (1.6)
adpU	5 (4)	84.5 (85.9)	100 (100)	100 (100)	100 (100)	100 (100)
Ident	0	0.4	0.2	0.4	2.4	8.3
Spher	0	0.4	0.2	0.4	2.4	7.8

C.3. Simulations on Other Testing Examples. In this section, we provide the simulation results on other testing examples discussed in Section 4. We present simulations on generalized linear model in Section C.3.1. In addition, we provide simulations on two-sample covariance testing to examine the empirical type I error and power in Sections C.3.2 and C.3.3, respectively.

C.3.1. *Study 6: GLM.* In this study, we conduct simulations for generalized linear model considering the following model

$$y_i = \mathbf{z}_i^\intercal \boldsymbol{\alpha} + \mathbf{x}_i^\intercal \boldsymbol{\beta} + \epsilon_i, \tag{C.18}$$

for $i = 1, \dots, n$. We generate i.i.d. $\mathbf{x}_i$ from the multivariate normal distribution $\mathcal{N}(0, \Sigma)$. We show the results with an equal variance and a first-order autoregressive correlation matrix case, that is, $\Sigma = (0.4^{|i-j|})$. We further generate $\mathbf{z}_i$ of two covariates with entries i.i.d. from standard normal distribution $\mathcal{N}(0, 1)$, and ϵ_i are the random errors following i.i.d. normal distribution $\mathcal{N}(0, 0.5)$. In (C.18), we take $\boldsymbol{\alpha} = (0.3, 0.3)^\intercal$, $\boldsymbol{\beta} = \mathbf{0}$ or $\neq \mathbf{0}$ corresponded to the null hypothesis H_0 and the alternative hypothesis H_A, respectively. Under H_A, $\lfloor ps \rfloor$ elements in $\boldsymbol{\beta}$ are set to be non-zero, where $s \in [0, 1]$ controls signal sparsity. We vary s to mimic varying sparsity situations, from sparse to dense signals with $s \in \{0.001, 0.1, 0.3, 0.7, 0.9\}$. The positions of non-zero elements in $\boldsymbol{\beta}$ are assumed to be uniformly distributed in $\{1, 2, \dots, p\}$, and their values are constant c, where c is the effect of signals that vary in the simulations. The results are based on 1000 simulations with 5% nominal significance level, $n = 500$ and $p = 1000$. We summarized the results in Figure 10. It shows similar patterns as in Study I.

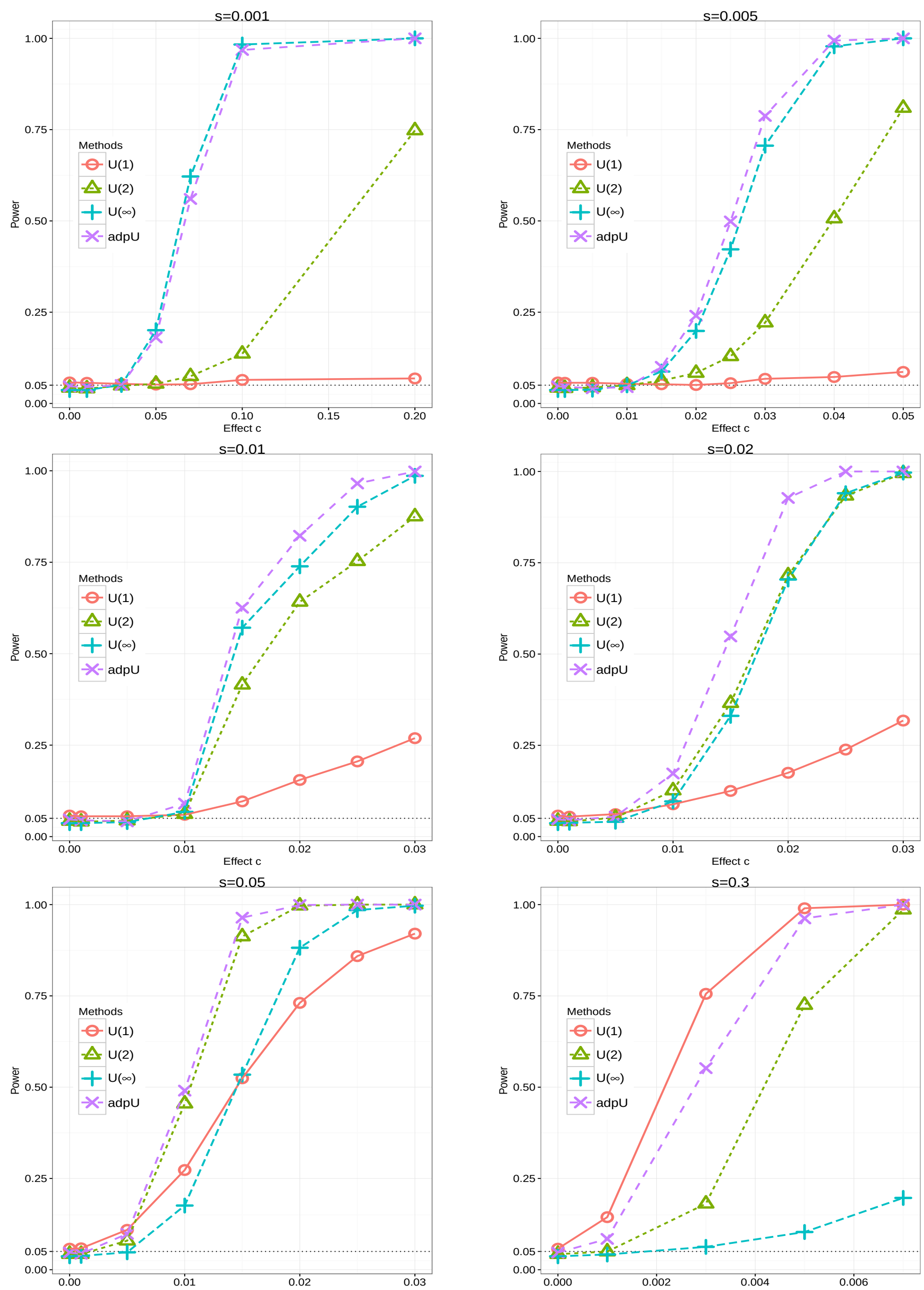

Fig 10: Power comparison under generalized linear model simulation setting.

C.3.2. *Study 7: Two-sample covariance testing under H_0.* In this section, we examine the empirical Type I errors of the proposed the adaptive testing procedure and compare it with the other methods.

We follow the simulation settings in Yang and Pan [79]. In particular, let $A(s)$ be the $s \times s$ covariance matrix of MA(1) model with the parameter $\theta_1 = 0.4$. In addition, $B = 0.7I_{p-s}$ is a $(p-s) \times (p-s)$ scaled identity matrix. We then define the matrix $Q(s) = \mathrm{BlkDiag}(A(s), B)$, where "BlkDiag" indicates a block diagonal matrix. We take $s = p^{1/2}$ and $n = 100$, and consider $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = Q(s)$. The results are presented in Table 11.

In Table 11, we provide the simulation results of the single U-statistics $\mathcal{U}(a)$ with $a \in \{1, \ldots, 6\}$. In addition, we provide the simulation results of $\mathcal{U}(\infty)$ using permutation and the asymptotic distribution in Cai et al. [10], which are denoted as "$\mathcal{U}(\infty)$ permutation" and "$\mathcal{U}(\infty)$ Tony" respectively. Given the results of $\mathcal{U}(1), \ldots, \mathcal{U}(6)$ and "$\mathcal{U}(\infty)$ (permutation)", "adpUmin 1" and "adpUf 1" represent the results of the adaptive testing procedure using minimum combination and Fisher's method respectively. Similarly, given the results of $\mathcal{U}(1), \ldots, \mathcal{U}(6)$ and "$\mathcal{U}(\infty)$ (Tony)", "adpUmin 2" and "adpUf 2" represent the results of the adaptive testing procedure using minimum combination and Fisher's method respectively. Moreover, "Schott", "Sriva" and "Chen" represent the methods in Schott [66], Srivastava and Yanagihara [70] and Li and Chen [54], respectively. In addition, we denote the tests without and with Micro term in Yang and Pan [79] as "Pan1" and "Pan2" respectively. The tests in [79] are time-consuming. Therefore we only provide the simulation results at $p = 50$, which takes about 100 times the time of the proposed adaptive testing procedure.

Based on our simulation results, we find that the empirical Type I errors of the single U-statistics are close the nominal levels, which verifies the theoretical results of Theorem 4.6. Moreover, comparing "$\mathcal{U}(\infty)$ (permutation)" and "$\mathcal{U}(\infty)$ (Tony)", we find that using the asymptotic distribution in Cai et al. [10] gives conservative Type I errors that are smaller than the nominal levels. In addition, by examining the results of minimum combination and Fisher's method, we find that both of the two methods give empirical Type I errors that are close to the nominal level, while the Fisher's method may have slight size inflation compared to the minimum combination.

TABLE 11
Empirical Type-I errors under $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = Q(s)$; $n = 100$, $s = p^{1/2}$

p	50	100	200	300
$\mathcal{U}(1)$	0.052	0.055	0.040	0.039
$\mathcal{U}(2)$	0.051	0.060	0.053	0.047
$\mathcal{U}(3)$	0.048	0.061	0.054	0.054
$\mathcal{U}(4)$	0.039	0.059	0.067	0.053
$\mathcal{U}(5)$	0.056	0.046	0.041	0.066
$\mathcal{U}(6)$	0.045	0.044	0.041	0.044
$\mathcal{U}(\infty)$ (permutation)	0.047	0.042	0.049	0.052
adpUmin 1	0.043	0.057	0.059	0.053
adpUf 1	0.076	0.081	0.060	0.076
$\mathcal{U}(\infty)$ (Tony)	0.018	0.024	0.016	0.013
adpUmin 2	0.044	0.056	0.059	0.051
adpUf 2	0.051	0.056	0.040	0.050
Chen	0.050	0.049	0.049	0.050
Sriva	0.166	0.002	0.000	0.000
Schott	0.074	0.119	0.236	0.418
Pan1	0.055	NA	NA	NA
Pan2	0.058	NA	NA	NA

C.3.3. *Study 8: Two-sample covariance testing power.* In this section, we examine the power of the two-sample covariance testing.

We follow the covariance matrix models in Yang and Pan [79]. In particular, let $H(\tau_0, \tau_1, r) = (h_{i,j})_{p\times p}$, where $h_{i,j} = 0$ except $h_{i,i} = \tau_0$, $i = 1, \dots, r$ and $h_{i,i+1} = h_{i,i-1} = \tau_1$, $i = 1, \dots, r-1$. Here τ_0 and τ_1 are used to measure the level of faint alternatives and r is used to measure the sparsity level of alternative. We fix $\mathbf{\Sigma}_x = I_p$, the $p \times p$ identity matrix, and examine the following three representative covariance matrix models of $\mathbf{\Sigma}_y$.

Model 1: (Extreme faint, $\tau_0 = 0.04, \tau_1 = 0.2, r = p$). $\mathbf{\Sigma}_y = I_p + H(0.04, 0.2, p)$. This matrix can also be considered as the covariance matrix of MA(1) model with the parameter $\theta_1 = 0.2$, which is also used in Li and Chen [54].

Model 2: (Extreme sparse, $\tau_0 = 1, \tau_1 = 1.5, r = 2$). $\mathbf{\Sigma}_y = I_p + H(1, 1.5, 2)$. This model only has four large disturbances compared with $\mathbf{\Sigma}_x$, which is regarded as the extreme sparse (ES) alternative.

Model 3: (Reasonable faint and sparse, $\tau_0 = 0.3, \tau_1 = 0.3, r = p/10$) $\mathbf{\Sigma}_y = I_p + H(0.3, 0.3, p/10)$. The value of r here is between 2 (in Model 2) and p (in Model 1), which is regarded as a moderately sparse setting.

Under each model above, we take $n = 100$, $p \in \{50, 100, 200, 300\}$, and provide the simulation results of the Models 1–3 in the Tables 12–14 respectively. The explanation of each row are the same as in Table 11, which is given in Section C.3.2. Similarly, we note that the tests in Yang and Pan [79] are very time-consuming. Therefore for "Pan 1" and "Pan 2", we only provide the simulation results at $p = 50$, which takes about 100 times the time of the proposed adaptive testing procedure.

TABLE 12
Empirical Power under Model 1 (Extreme faint); $n = 100$.

p	50	100	200	300
$\mathcal{U}(1)$	0.397	0.389	0.408	0.416
$\mathcal{U}(2)$	0.445	0.458	0.456	0.484
$\mathcal{U}(3)$	0.290	0.309	0.354	0.371
$\mathcal{U}(4)$	0.197	0.211	0.199	0.205
$\mathcal{U}(5)$	0.244	0.397	0.752	0.855
$\mathcal{U}(6)$	0.054	0.052	0.054	0.091
$\mathcal{U}(\infty)$ (permutation)	0.066	0.062	0.044	0.029
adpUmin 1	0.478	0.511	0.692	0.783
adpUf 1	0.600	0.648	0.843	0.886
$\mathcal{U}(\infty)$ (Tony)	0.091	0.072	0.087	0.072
adpUmin 2	0.480	0.513	0.691	0.781
adpUf 2	0.619	0.669	0.855	0.903
Chen	0.573	0.574	0.569	0.623
Sriva	0.513	0.586	0.598	0.569
Schott	0.667	0.731	0.888	0.956
Pan1	0.640	NA	NA	NA
Pan2	0.669	NA	NA	NA

TABLE 13
Empirical Power under Model 2 (Extreme sparse); $n = 100$.

p	50	100	200	300
$\mathcal{U}(1)$	0.068	0.056	0.048	0.049
$\mathcal{U}(2)$	0.725	0.364	0.122	0.086
$\mathcal{U}(3)$	0.993	0.960	0.850	0.660
$\mathcal{U}(4)$	1.000	0.997	0.988	0.956
$\mathcal{U}(5)$	0.934	0.874	0.803	0.682
$\mathcal{U}(6)$	0.972	0.960	0.935	0.914
$\mathcal{U}(\infty)$ (permutation)	0.966	0.919	0.852	0.772
adpUmin 1	1.000	0.992	0.984	0.959
adpUf 1	1.000	0.996	0.989	0.970
$\mathcal{U}(\infty)$ (Tony)	0.999	1.000	0.997	1.000
adpUmin 2	1.000	0.997	0.993	0.995
adpUf 2	1.000	0.999	0.992	0.992
Chen	0.800	0.457	0.196	0.127
Sriva	0.787	0.433	0.166	0.101
Schott	0.864	0.640	0.550	0.654
Pan1	0.673	NA	NA	NA
Pan2	0.694	NA	NA	NA

We then analyze the simulation results. Model 1 is the extreme faint case and $\mathbf{\Sigma}_y - \mathbf{\Sigma}_x$ is dense. We find that under this case, the U-statistics of

TABLE 14
Empirical Power under Model 3 (Reasonable faint and sparse); $n = 100$.

p	50	100	200	300
$\mathcal{U}(1)$	0.072	0.067	0.069	0.070
$\mathcal{U}(2)$	0.090	0.096	0.096	0.083
$\mathcal{U}(3)$	0.155	0.151	0.152	0.145
$\mathcal{U}(4)$	0.175	0.162	0.162	0.154
$\mathcal{U}(5)$	0.347	0.582	0.868	0.946
$\mathcal{U}(6)$	0.308	0.494	0.732	0.854
$\mathcal{U}(\infty)$ (permutation)	0.028	0.034	0.027	0.018
adpUmin 1	0.337	0.496	0.797	0.901
adpUf 1	0.355	0.535	0.802	0.910
$\mathcal{U}(\infty)$ (asymptotic)	0.254	0.319	0.409	0.403
adpUmin 2	0.348	0.508	0.798	0.901
adpUf 2	0.426	0.620	0.862	0.940
Chen	0.138	0.149	0.153	0.144
Sriva	0.092	0.096	0.097	0.100
Schott	0.189	0.283	0.486	0.712
Pan1	0.167	NA	NA	NA
Pan2	0.186	NA	NA	NA

small orders, e.g., $\mathcal{U}(1)$ and $\mathcal{U}(2)$ are powerful. The tests based on the sum-of-squares type statistics including "Chen", "Sriva" and "Schott" are also powerful under this case. Our proposed adaptive testing procedure using Fisher's method has comparable power performance to "Pan 1" and "Pan 2", and is computationally more efficient. Model 2 is the extreme sparse case. Under this case, we find that generally U-statistics of higher orders, e.g., $\mathcal{U}(4)$ and $\mathcal{U}(\infty)$, are more powerful than the U-statistics of smaller orders, e.g., $\mathcal{U}(1)$ and $\mathcal{U}(2)$. Model 3 is the moderately faint and sparse case. Under this case, we can see that a finite-order U-statistic $\mathcal{U}(5)$ is the most powerful one. Neither the maximum-type test statistic $\mathcal{U}(\infty)$ and the sum-of-squares type test statistic $\mathcal{U}(2)$, "Chen", "Sriva" and "Schott" are very powerful. Tests in [79] considering only faint or sparse alternatives are not very powerful under this case. On the other hand, the proposed adaptive testing procedure maintains high power under this case.

REFERENCES

[1] Alzheimer's Association (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia 14*(3), 367–429.

[2] Anderson, T. W. (2009). *An introduction to multivariate statistical analysis.* John Wiley & Sons, New York.

[3] Bai, Z., D. Jiang, J.-F. Yao, and S. Zheng (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist. 37*(6B), 3822–3840.

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

[5] Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*(1), 199–227.

[6] Billingsley, P. (1995). *Probability and Measure* (3 ed.). John Wiley & Sons.

[7] Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association 106*(494), 672–684.

[8] Cai, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annual Review of Statistics and Its Application 4*(1), 423–446.

[9] Cai, T. T. and T. Jiang (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist. 39*(3), 1496–1525.

[10] Cai, T. T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association 108*(501), 265–277.

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

[12] Cai, T. T. and Z. Ma (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli 19*(5B), 2359–2388.

[13] Chen, S. X., J. Li, and P.-S. Zhong (2014). Two-sample tests for high dimensional means with thresholding and data transformation. *arXiv preprint arXiv:1410.2848*.

[14] Chen, S. X. and Y.-L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist. 38*(2), 808–835.

[15] Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

[16] Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *The Annals of Statistics 46*(2), 642–678.

[17] Colantuoni, C., B. K. Lipska, T. Ye, T. M. Hyde, R. Tao, J. T. Leek, E. A. Colantuoni, A. G. Elkahloun, M. M. Herman, D. R. Weinberger, and J. E. Kleinman (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature 478*(7370), 519.

[18] Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist. 32*(3), 962–994.

[19] Donoho, D., J. Jin, et al. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science 30*(1), 1–25.

[20] Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association 91*(434), 674–688.

[21] Fan, J., F. Han, and H. Liu (2014). Challenges of big data analysis. *National science review 1*(2), 293–314.

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

[23] Fan, J., J. Lv, and L. Qi (2011). Sparse high-dimensional models in economics.

[24] Frahm, G. (2004). *Generalized elliptical distributions: theory and applications*. Ph. D. thesis, Universität zu Köln.

[25] Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science 326*(5951), 399–403.

[26] Gaetan, C. and X. Guyon (2010). *Spatial statistics and modeling*, Volume 90. Springer.

[27] Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006). Testing

against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 477–493.

[28] Gregory, K. B., R. J. Carroll, V. Baladandayuthapani, and S. N. Lahiri (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association 110*(510), 837–849.

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

[30] Guyon, X. (1995). *Random fields on a network: modeling, statistics, and applications.* Springer Science & Business Media.

[31] Hall, P. (1979). On the rate of convergence of normal extremes. *Journal of Applied Probability 16*(2), 433–439.

[32] Hall, P. and J. Jin (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist. 38*(3), 1686–1732.

[33] Heinig, M., E. Petretto, C. Wallace, L. Bottolo, M. Rotival, H. Lu, Y. Li, R. Sarwar, S. R. Langley, A. Bauerfeind, et al. (2010). A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature 467*(7314), 460.

[34] Heyde, C. C. and B. M. Brown (1970). On the departure from normality of a certain class of martingales. *Ann. Math. Statist. 41*(6), 2161–2165.

[35] Higham, N. J. (2008). *Functions of matrices: theory and computation*, Volume 104. Siam.

[36] Ho, H.-C. and T. Hsing (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *Journal of Applied Probability 33*(1), 138–145.

[37] Ho, H.-C. and W. P. McCormick (1999). Asymptotic distribution of sum and maximum for Gaussian processes. *Journal of Applied Probability 36*(4), 1031–1044.

[38] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist. 19*(3), 293–325.

[39] Hsing, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *Ann. Probab. 23*(2), 938–947.

[40] Ibragimov, I. A. and J. V. Linnik (1971). Independent and stationary sequences of random variables. *The Netherlands: Wolters-Noordhof*.

[41] James, B., K. James, and Y. Qi (2007). Limit distribution of the sum and maximum from multivariate Gaussian sequences. *Journal of multivariate analysis 98*(3), 517–532.

[42] Jansen, I. E., J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, J. Sealock, I. K. Karlsson, S. Hägg, L. Athanasiu, et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimers disease risk. *Nature Genetics 51*, 404–413.

[43] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab. 14*(2), 865–880.

[44] Jiang, T. and F. Yang (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist. 41*(4), 2029–2074.

[45] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist. 29*(2), 295–327.

[46] Kan, R. (2008). From moments of sum to moments of product. *Journal of Multivariate Analysis 99*(3), 542–554.

[47] Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research 38*(suppl 1), D355–D360.

[48] Kim, J., Y. Zhang, and W. Pan (2016). Powerful and adaptive testing for multi-

trait and multi-SNP associations with GWAS and sequencing data. *Genetics 203*(2), 715–731.

[49] Kim, T. Y. (1994). Moment bounds for non-stationary dependent sequences. *Journal of Applied Probability 31*(3), 731–742.

[50] Lan, W., R. Luo, C.-L. Tsai, H. Wang, and Y. Yang (2015). Testing the diagonality of a large covariance matrix in a regression setting. *Journal of Business & Economic Statistics 33*(1), 76–86.

[51] Ledoit, O. and M. Wolf (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist. 30*(4), 1081–1102.

[52] Leung, D. and M. Drton (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics 46*(1), 280–307.

[53] Li, D. and L. Xue (2015). Joint limiting laws for high-dimensional independence tests. *ArXiv e-prints*.

[54] Li, J. and S. X. Chen (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist. 40*(2), 908–940.

[55] Liu, W.-D., Z. Lin, and Q.-M. Shao (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab. 18*(6), 2337–2366.

[56] Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, and P. M. Visscher (2009). Finding the missing heritability of complex diseases. *Nature 461*(7265), 747.

[57] McCormick, W. and Y. Qi (2000). Asymptotic distribution for the sum and maximum of Gaussian processes. *Journal of Applied Probability 37*(4), 958–971.

[58] Mosteller, F. and R. A. Fisher (1948). Questions and answers. *The American Statistician 2*(5), 30–31.

[59] Muirhead, R. J. (2009). *Aspects of multivariate statistical theory.* John Wiley & Sons, New York.

[60] Paindaveine, D. and G. Van Bever (2014). Inference on the shape of elliptical distributions based on the MCD. *Journal of Multivariate Analysis 129*, 125–144.

[61] Pan, W., J. Kim, Y. Zhang, X. Shen, and P. Wei (2014). A powerful and adaptive association test for rare variants. *Genetics 197*(4), 1081–1095.

[62] Péché, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields 143*(3-4), 481–516.

[63] Peng, Z. and S. Nadarajah (2003). On the joint limiting distribution of sums and maxima of stationary normal sequence. *Theory of Probability & Its Applications 47*(4), 706–709.

[64] Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications 19*(2), 297–303.

[65] Prince, M., R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's & Dementia 9*(1), 63–75.

[66] Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics & Data Analysis 51*(12), 6535–6542.

[67] Shao, Q.-M. and W.-X. Zhou (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab. 42*(2), 623–648.

[68] Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics 108*(5-6), 1033–1056.

[69] Srivastava, M. S. and M. Du (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis 99*(3), 386–402.

[70] Srivastava, M. S. and H. Yanagihara (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis 101*(6), 1319–1329.

[71] Srivastava, R., P. Li, and D. Ruppert (2016). RAPTT: An exact two-sample test in high dimensions using random projections. *Journal of Computational and Graphical Statistics 25*(3), 954–970.

[72] Storey, J. D. and R. Tibshirani (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences 100*(16), 9440–9445.

[73] Wang, L., P. Jia, R. D. Wolfinger, X. Chen, and Z. Zhao (2011). Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics 98*(1), 1–8.

[74] Wu, C., G. Xu, and W. Pan (2019). An adaptive test on high-dimensional parameters in generalized linear models. *Statistica Sinica*.

[75] Xia, Y., T. Cai, and T. T. Cai (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika 102*(2), 247–266.

[76] Xu, G., L. Lin, P. Wei, and W. Pan (2016). An adaptive two-sample test for high-dimensional means. *Biometrika 103*(3), 609–624.

[77] Xu, J., S. L. Murphy, K. D. Kochanek, B. Bastian, and E. Arias (2018). Deaths: Final data for 2016. *National Vital Statistics Reports 67*(5).

[78] Xu, Z., G. Xu, and W. Pan (2017). Adaptive testing for association between two random vectors in moderate to high dimensions. *Genetic epidemiology 41*(7), 599–609.

[79] Yang, Q. and G. Pan (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *Journal of the American Statistical Association 112*(517), 188–200.

[80] Yu, K., Q. Li, A. W. Bergen, R. M. Pfeiffer, P. S. Rosenberg, N. Caporaso, P. Kraft, and N. Chatterjee (2009). Pathway analysis by adaptive combination of p-values. *Genetic epidemiology 33*(8), 700–709.

[81] Zaitsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of sn bernstein's inequality conditions. *Probability theory and related fields 74*(4), 535–566.

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

Yinqiu He and Gongjun Xu
Department of Statistics
University of Michigan
E-mail: yqhe@umich.edu
gongjun@umich.edu

Chong Wu
Department of Statistics
Florida State University
E-mail: cwu3@fsu.edu

Wei Pan
Division of Biostatistics
School of Public Health
University of Minnesota
E-mail: panxx014@umn.edu