The Annals of Statistics
2015, Vol. 43, No. 6, 2588–2623
DOI: 10.1214/15-AOS1353

SPECTRAL STATISTICS OF LARGE DIMENSIONAL SPEARMAN'S RANK CORRELATION MATRIX AND ITS APPLICATION

BY ZHIGANG BAO[1,*], LIANG-CHING LIN[†], GUANGMING PAN[2,*] AND WANG ZHOU[3,‡]

Nanyang Technological University[*], *National Cheng Kung University*[†] *and National University of Singapore*[‡]

Let $\mathbf{Q} = (Q_1, \dots, Q_n)$ be a random vector drawn from the uniform distribution on the set of all $n!$ permutations of $\{1, 2, \dots, n\}$. Let $\mathbf{Z} = (Z_1, \dots, Z_n)$, where Z_j is the mean zero variance one random variable obtained by centralizing and normalizing Q_j, $j = 1, \dots, n$. Assume that $\mathbf{X}_i, i = 1, \dots, p$ are i.i.d. copies of $\frac{1}{\sqrt{p}}\mathbf{Z}$ and $X = X_{p,n}$ is the $p \times n$ random matrix with $\mathbf{X}_i$ as its ith row. Then $S_n = XX^*$ is called the $p \times n$ Spearman's rank correlation matrix which can be regarded as a high dimensional extension of the classical nonparametric statistic Spearman's rank correlation coefficient between two independent random variables. In this paper, we establish a CLT for the linear spectral statistics of this nonparametric random matrix model in the scenario of high dimension, namely, $p = p(n)$ and $p/n \to c \in (0, \infty)$ as $n \to \infty$. We propose a novel evaluation scheme to estimate the core quantity in Anderson and Zeitouni's cumulant method in [*Ann. Statist.* **36** (2008) 2553–2576] to bypass the so-called joint cumulant summability. In addition, we raise a *two-step comparison approach* to obtain the explicit formulae for the mean and covariance functions in the CLT. Relying on this CLT, we then construct a distribution-free statistic to test complete independence for components of random vectors. Owing to the nonparametric property, we can use this test on generally distributed random variables including the heavy-tailed ones.

1. Introduction.

1.1. *Matrix model.* In this paper, we will consider the large dimensional Spearman's rank correlation matrices. First, we give the definition of the matrix model. Let P_n be the set consisting of all permutations of $\{1, 2, \dots, n\}$. Suppose

Received October 2014; revised June 2015.

[1]Supported in part by the Ministry of Education, Singapore, under Grant # ARC 14/11, and NSF of China, Grant No. 11371317.

[2]Supported in part by a MOE Tier 2 Grant 2014-T2-2-060 and by a MOE Tier 1 Grant RG25/14 at the Nanyang Technological University, Singapore.

[3]Supported in part by Grant R-155-000-151-112 at the National University of Singapore.

MSC2010 subject classifications. Primary 15B52; secondary 62H10.

Key words and phrases. Spearman's rank correlation matrix, nonparametric method, linear spectral statistics, central limit theorem, independence test.

that $\mathbf{Z} = (Z_1, \dots, Z_n)$ is a random vector, where

$$Z_i = \sqrt{\frac{12}{n^2-1}}\left(Q_i - \frac{n+1}{2}\right)$$

and $\mathbf{Q} := (Q_1, \dots, Q_n)$ is uniformly distributed on P_n. That is, for any permutation $(\sigma(1), \sigma(2), \dots, \sigma(n)) \in P_n$, one has $\mathbb{P}\{\mathbf{Q} = (\sigma(1), \sigma(2), \dots, \sigma(n))\} = 1/n!$. For simplicity, we will use the notation $[N] := \{1, \dots, N\}$ for any positive integer N in the sequel. Now for $m \in [n]$ we conventionally define the set of *m partial permutations of n* as

$$P_{nm} := \{(v_1, \dots, v_m) : v_1, \dots, v_m \in [n] \text{ and } v_i \neq v_j \text{ if } i \neq j\}.$$

For any mutually distinct numbers $l_1, \dots, l_m \in [n]$, it is elementary to check that $(Q_{l_1}, \dots, Q_{l_m})$ is uniformly distributed on P_{nm}. Such a fact immediately leads to the fact that $\{Z_i\}_{i=1}^n$ is strictly stationary. In addition, by setting $m = 1$ or 2, it is straightforward to check through calculations that

$$\text{(1.1)} \quad \mathbb{E}Z_i = 0, \qquad \mathbb{E}Z_i^2 = 1, \qquad \operatorname{Cov}(Z_j, Z_k) = -\frac{1}{n-1} \qquad \text{if } j \neq k.$$

Moreover, it is also easy to see that for any positive integer l,

$$\mathbb{E}|Z_1|^l \leq C_l$$

for some positive constant C_l depending on l. Besides, we note that $Z_i, i \in [n]$ are symmetric random variables.

Assuming that $\mathbf{X}_i = (x_{i1}, \dots, x_{in}), i = 1, \dots, p$ are i.i.d. copies of $\frac{1}{\sqrt{p}}\mathbf{Z}$, we set $S_n = XX^*$, where

$$X := \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_p \end{bmatrix} = (x_{ij})_{p,n}.$$

Then S_n is referred to as the *Spearman's rank correlation matrix*.

1.2. *Motivation from nonparametric statistics.* A main motivation of considering the matrix S_n is from nonparametric statistics. We consider the hypothesis testing problem on some random variable sequence $Y_1, \dots, Y_p$ as

$$\mathbf{H}_0: \qquad Y_1, \dots, Y_p \text{ are independent} \quad \text{v.s.}$$
$$\mathbf{H}_1: \qquad Y_1, \dots, Y_p \text{ are not independent.}$$

Note that since the covariance matrix and correlation matrix can capture the dependence for Gaussian variables, it is natural to compare the sample covariance or correlation matrix with diagonal matrices to detect whether $\mathbf{H}_0$ holds in the classical setting of *large n and fixed p*. Unfortunately, due to the so-called *curse of dimensionality*, it is now well known that there is no hope to approximate the population covariance or correlation matrix by sample ones in the situation of *large n and comparably large p* without any assumption imposed on the population covariance matrix. However, it is still possible to construct independence test statistics from the sample covariance or correlation matrix for Gaussian variables even in the high dimensional case. In the scenario of large n and comparably large p, there is a long list of literature devoted to studying the properties of sample covariance matrix or sample correlation matrix under the null hypothesis $\mathbf{H}_0$. For example, Bao, Pan and Zhou [4], as well as Pillai and Yin [13], studied the largest eigenvalue of sample correlation matrix; Ledoit and Wolf [11] raised a quadratic form of the spectrum of sample covariance matrix; Schott [15] considered the sum of squares of sample correlation coefficients; Jiang [9] discussed the largest off-diagonal entry of sample correlation matrix; Jiang and Yang [10] studied the likelihood ratio test statistic for the sample correlation matrix. However, for non-Gaussian variables, even in the classical *large n and fixed p* case, the idea to compare the population covariance matrix with diagonal matrices is substantially invalid for independence test for those uncorrelated but dependent variables. Moreover, for random vectors containing at least one heavy tailed component such as a Cauchy random variable, there is even no population covariance matrix.

In view of the above, we discuss a nonparametric matrix model in this paper and study its spectrum statistics under $\mathbf{H}_0$ in order to accommodate random variables with general distributions. Assume that we have n observations of the vector $(Y_1, \dots, Y_p)$. Let $Y_{11}, \dots, Y_{1n}$ be the observations of the first coefficient Y_1 and set Q_{1j} to be the rank of Y_{1j} among $Y_{11}, \dots, Y_{1n}$. We then replace $(Y_{11}, \dots, Y_{1n})$ by the corresponding normalized rank sequence $(x_{11}, \dots, x_{1n})$, where

$$x_{1j} = \sqrt{\frac{12}{p(n^2-1)}}\left(Q_{1j} - \frac{n+1}{2}\right), \qquad j \in [n].$$

Analogously, we can define the rank sequence $(x_{i1}, \dots, x_{in})$ for other $i \in [p]$. For simplicity, in this paper we only consider the case where $Y_i, i \in [p]$ are continuous random variables. In this case, the probability of a tie occurring in the sequence $Q_{i1}, \dots, Q_{in}$ for any $i \in [p]$ is zero. Then $S_n = XX^*$ with $X = (x_{ij})_{p,n}$ is the so-called Spearman's rank correlation matrix under $\mathbf{H}_0$, which can be regarded as a high dimensional extension of the classical Spearman's rank correlation coefficient between two random variables. Then we can construct statistics from S_n to tackle the above hypothesis testing problem. By contrast, the parametric models such as Pearson's sample correlation matrix and sample covariance matrix are well studied by statisticians and probabilists. However, the work on Spearman's rank correlation matrix is few and far between. In [2], Bai and Zhou proved that the limiting spectral distribution of S_n is also the famous Marchenko–Pastur law (MP law). In [21], Zhou studied the limiting behavior of the largest off-diagonal entry of S_n. Our purpose in this paper is to derive the fluctuation (a CLT) of the linear spectral statistics for S_n. As an application, we will construct a nonparametric statistic to detect dependence of components of random vectors.

1.3. *Main result.* We set $\lambda_1 \geq \cdots \geq \lambda_p$ to be the ordered eigenvalues of S_n. Our main task in this paper is to study the limiting behavior of the so-called linear spectral statistics $\mathcal{L}_n[f] = \sum_{i=1}^{p} f(\lambda_i)$ for some test function f. In this paper, we will focus on the polynomial test functions and, therefore, it suffices to study the joint limiting behavior of $\operatorname{tr} S_n^k$, $k = 1, \dots, \infty$. We state the main result as the following theorem.

THEOREM 1.1. *Assuming that both n and $p := p(n)$ tend to ∞ and*

$$n/p \to c \in (0, \infty),$$

we have

$$\{\operatorname{tr} S_n^k - \mathbb{E} \operatorname{tr} S_n^k\}_{k=2}^{\infty} \Longrightarrow \{G_k\}_{k=2}^{\infty} \qquad \text{as } n \to \infty,$$

where $\{G_k\}_{k=2}^{\infty}$ is a Gaussian process with mean zero and the covariance function given by

$$\begin{aligned}
&\operatorname{Cov}(G_{k_1}, G_{k_2}) \\
&\quad = 2c^{k_1+k_2} \sum_{j_1=0}^{k_1-1} \sum_{j_2=0}^{k_2} \binom{k_1}{1} \binom{k_2}{j_2} \left(\frac{1-c}{c}\right)^{j_1+j_2} \\
&\qquad \times \sum_{l=1}^{k_1-j_1} l \binom{2k_1 - 1 - (j_1 + l)}{k_1 - 1} \binom{2k_2 - 1 - j_2 + l}{k_2 - 1} \\
&\qquad - 2c^{k_1+k_2+1} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \binom{k_1}{j_1} \binom{k_2}{j_2} \left(\frac{1-c}{c}\right)^{j_1+j_2} \\
&\qquad \times \binom{2k_1 - j_1}{k_1 - 1} \binom{2k_2 - j_2}{k_2 - 1}.
\end{aligned} \tag{1.2}$$

Moreover, we have the following expansion for the expectation function:

$$\begin{aligned}
\mathbb{E} \operatorname{tr} S_n^k &= \frac{n^k}{(n-1)^{k-1}} \sum_{j=0}^{k-1} \frac{1}{j+1} \binom{k}{j} \binom{k-1}{j} \left(\frac{n-1}{p}\right)^j - \frac{1}{2} \sum_{j=0}^{k} \binom{k}{j}^2 c^j \\
&\quad + 2c^{1+k} \sum_{j=0}^{k} \binom{k}{j} \left(\frac{1-c}{c}\right)^j \left[\binom{2k-j}{k-1} - \binom{2k+1-j}{k-1}\right] \\
&\quad + \frac{1}{4}\left[(1-\sqrt{c})^{2k} + (1+\sqrt{c})^{2k}\right] + o(1).
\end{aligned} \tag{1.3}$$

REMARK 1.2. Note that when $k = 1$, $\operatorname{tr} S_n = n$ is deterministic. Actually, it can also be checked that the right-hand side of (1.2) is zero when $k_1 = k_2 = 1$ and the right-hand side of (1.3) equals $n + o(1)$ when $k = 1$ therein.

1.4. *Methodologies of the proof.* Roughly speaking, we will start with a *cumulant method* introduced by Anderson and Zeitouni in [1] to establish the CLT for $\{(\operatorname{tr} S_n^k - \mathbb{E}\operatorname{tr} S_n^k)/\sqrt{\operatorname{Var}(\operatorname{tr} S_n^k)}\}_{k=2}^{\infty}$. The cumulant method can be viewed as a modification of the celebrated moment method in Random Matrix Theory (RMT). Without trying to be comprehensive, we refer to [3, 6, 16, 18] for further reading. Particularly, we refer to [1, 8, 17] for the cumulant method in proving CLTs for linear spectral statistics in RMT.

As explained at the beginning, $\{Z_i\}_{i=1}^n$ is a stationary sequence, which inspires us to learn from [1]. However, in [1] the stationary sequence is required to satisfy the so-called *joint cumulant summability* property [see (3.6) below], which was used to bound the core quantity (3.5). The property of joint cumulant summability is crucial in the spectral analysis of time series; one can refer to [7, 14, 19, 20] for further reading. Unfortunately, to verify this property for a general stationary sequence is highly nontrivial. In this paper, we will not try to check whether the joint cumulant summability holds for $\mathbf{Z}$. Instead we will provide a relatively rough but crucial bound on (3.5) through a totally novel evaluation scheme; see Proposition 4.2 and Corollary 4.4 below. Such a bound will allow us to bypass the joint cumulant summability and serve as a main input to pursue the cumulant method to establish the CLT. With this CLT, what remains is therefore to evaluate the nonnegligible terms of the mean and covariance functions. It will be shown that the mean and covariance functions of $\{\operatorname{tr} S_n^k\}_{k=2}^{\infty}$ can be expressed by some sums of terms indexed by set partitions. For the explicit values of the nonnegligible terms of the mean and covariance functions, we will adopt a *two-step comparison strategy* to compare these expressions with the existing results for the sample covariance matrices.

1.5. *Organization and notation.* Our paper is organized as follows. Section 2 is devoted to the application of our CLT to the independence test on random vectors. In Section 3, we will introduce some basic notions of joint cumulants and some known results from [1]. Section 4 is our main technical part which will be devoted to providing the required bound for the sum (3.5). Specifically, we will mainly prove Proposition 4.2 therein. In Section 5, we will use the bounds obtained in Section 4 to show that all high order cumulants tend to zero when $n \to \infty$. Finally, in Section 6, we will combine the bounds in Section 4 and the aforementioned two-step comparison strategy to evaluate the main terms of the mean and covariance functions.

Throughout the paper, we will use $\#\mathbb{S}$ to represent the cardinality of a set $\mathbb{S}$. For any number set $\mathbb{A} \subset [n]$, we will use $\{j_\alpha\}_{\alpha\in\mathbb{A}}$ to denote the set of j_α with $\alpha \in \mathbb{A}$. Analogously, $(j_\alpha)_{\alpha\in\mathbb{A}}$ represents the vector obtained by deleting the components j_β with $\beta \in [n]\setminus\mathbb{A}$ from $(j_1, \dots, j_n)$. In addition, we will use the notation $\mathrm{i} = \sqrt{-1}$ to denote the imaginary unit to release i which will be frequently used as subscript or index. For any vector $\vec{\xi} = (\xi_1, \dots, \xi_N)$, we say the *position* of ξ_i in $\vec{\xi}$ is i. For example, for vector $(Z_{j_2}, Z_{j_1}, Z_{j_4}, Z_{j_3})$, the position of Z_{j_1} is 2. Moreover, we will use C to represent some positive constant independent of n whose value may be different from line to line.

2. Application on independence test. In this section, we consider an application of Theorem 1.1. We construct a nonparametric statistic to test complete independence for the components of a high dimensional random vector. Our proposed statistic is W_7 and we highlight it here at first

$$W_7 = W_7(k,\delta) := \frac{\operatorname{tr} S_n^k - \mathbb{E}\operatorname{tr} S_n^k}{\sqrt{\operatorname{Var}(G_k)}} + n^{-\delta}\left[n\left(\max_{1\le i<j\le p}\left|\frac{p}{n}s_{ij}\right|\right)^2 - 4\log p + \log\log p\right],$$

where $0 < \delta < 1$. Note that our statistic depends on the parameters k and δ. We will discuss how to choose k and δ at the end of this section. To see the performance of our statistic, we will do the simulation under various settings. We also compare our statistic with some other parametric or nonparametric statistics in the literature. To introduce these statistics, we shall define some notation at first. Let $R_n = (r_{ij})_{p\times p}$ be the Pearson's sample correlation matrix based on n independent copies of p-dimensional random vector $(Y_1, \dots, Y_p)$. We denote by $\lambda_{\max}(R)$ the largest eigenvalue of R_n. In addition, we denote by s_{ij} the (i, j)th entry of S_n. We compare the performance of our statistic W_7 with the following statistics:

(i) $$W_1 = \frac{n\lambda_{\max}(R) - (p^{1/2} + n^{1/2})^2}{(n^{1/2} + p^{1/2})(p^{-1/2} + n^{-1/2})^{1/3}}$$ (see [4] or [13]),

(ii) $$W_2 = \frac{\operatorname{tr} S_n^k - \mathbb{E}\operatorname{tr} S_n^k}{\sqrt{\operatorname{Var}(G_k)}}$$ (see Theorem 1.1),

(iii) $$W_3 = \frac{\sum_{i=2}^p \sum_{j=1}^{i-1} r_{ij}^2 - p(p-1)/(2n)}{(p/n)}$$ (see [15]),

(iv) $$W_4 = \frac{\log(|R_n|) - (p - n + 3/2)\log(1 - p/(n-1)) + (n - 2(p/(n-1)))}{\sqrt{-2[p/(n-1) + \log(1 - p/(n-1))]}}$$ (see [10]),

(v) $$W_5 = n\left(\max_{1\le i<j\le p}|r_{ij}|\right)^2 - 4\log n + \log\log n$$ (see [9]),

(vi) $$W_6 = n\left(\max_{1\le i<j\le p}\left|\frac{p}{n}s_{ij}\right|\right)^2 - 4\log p + \log\log p$$ (see [21]).

At first, from Theorem 1.1, we see that the CLT holds for the statistic W_2. By our construction, $W_7 = W_2 + n^{-\delta}W_6$, which can be regarded as a slight modification of W_2 by adding a small penalty in terms of W_6. We expect that the statistic

W_7 will take the advantages of both W_2 and W_6, and thus its performance will be better. More specifically, we illustrate the philosophy of such a construction via the following examples. Let $\mathbf{g}$ be a p-dimensional Gaussian vector with the population covariance matrix $\Sigma_{\mathbf{g}}$. Two extreme alternative hypotheses are considered below. The first case is that $\Sigma_{\mathbf{g}}$ has only one significantly large off-diagonal entry. Then the corresponding Spearman's rank correlation matrix will also have a significantly large off-diagonal entry. Since W_6 is constructed from the largest off-diagonal entry, it is sensitive to this kind of dependence structure. In contrast, one cannot tell the dependence structure in $\mathbf{g}$ by W_2 since the linear spectral statistics are relatively robust under the disturbance of a single entry of the population covariance matrix. However, W_7 has an additional penalty which is $n^{-\delta}W_6$ compared to W_2. So one can capture the dependence contained in $\mathbf{g}$ by W_7. Now, we consider the second case where $\Sigma_{\mathbf{g}}$ contains a lot of small nonzero off-diagonal entries. In this case, the statistic W_6 performs badly since the largest off-diagonal entry of $\Sigma_{\mathbf{g}}$ is close to zero. In contrast, the statistic W_7 performs as well as W_2 since the spectral statistics can accumulate all the effects caused by these small off-diagonal entries.

We below summarize the limiting null distributions of W_i, $i = 1, \dots, 7$, and the corresponding assumptions in the references [4, 9, 10, 13, 15, 21]. The null distribution of W_1 converges to the type 1 Tracy–Widom law (see [4, 13]), assuming that the variables $Y_1, \dots, Y_p$ have sub-exponential tails. The limiting null distribution of W_2 is $N(0, 1)$ by Theorem 1.1. In [15] and [10], the weak convergence of W_3 and W_4 to $N(0, 1)$ is established for the Gaussian vector $(Y_1, \dots, Y_p)$ only. If one assumes that $Y_1, \dots, Y_p$ are i.i.d. with $\mathbb{E}Y_1^{30-\varepsilon} < \infty$ for any constant $\varepsilon > 0$, the limiting distribution of W_5 derived in [9] possesses the following c.d.f.:

$$F_{W_5}(y) = \exp\{-(c^2\sqrt{8\pi})^{-1}e^{-y/2}\},$$

which is called the extreme distribution of type I. According to [21], the statistic W_6 is distribution-free, with the following asymptotic distribution:

$$F_{W_6}(y) = \exp\{-(8\pi)^{-1/2}e^{-y/2}\}.$$

Clearly, the statistic $W_7 = W_2 + n^{-\delta}W_6$ is also distribution-free and possesses the limiting distribution $N(0, 1)$ due to Slutsky's theorem.

We denote by Cauchy(α, β) the Cauchy distribution with location parameter α and scale parameter β. In addition, we denote by $t(\gamma)$ the student's t-distribution with degrees of freedom γ. We consider three null hypotheses with the nominal significance level $\alpha = 5\%$:

- $\mathbf{H}_{0,1}$: $\mathbf{Y}_j$, $j \in [n]$ are i.i.d. $N_p(\mathbf{0}, I_p)$ vectors;
- $\mathbf{H}_{0,2}$: Y_{ij}, $i \in [p]$, $j \in [n]$ are i.i.d. Cauchy$(0, 1)$ variables.;
- $\mathbf{H}_{0,3}$: $Y_{i_1,j}$ are i.i.d. $N(0, 1)$ variables; $Y_{i_2,j}$ are i.i.d. Cauchy$(0, 1)$ variables; $Y_{i_3,j}$ are i.i.d. $t(4)$ variables, where $i_1 = 1, \dots, \lfloor p/3 \rfloor$, $i_2 = \lfloor p/3 \rfloor + 1, \dots, \lfloor 2p/3 \rfloor$, $i_3 = \lfloor 2p/3 \rfloor + 1, \dots, p$ and $j \in [n]$.

For each null hypothesis, we consider two alternatives:

- $\mathbf{H}_{a,1-1}$ (one large disturbance): $\mathbf{Y}_j$, $j \in [n]$ are i.i.d. $N_p(\mathbf{0}, I_p + C)$ and $C = (c_{ik})_{p\times p}$ with $c_{ik} = 0$, $i, k \in [p]$, except $c_{12} = c_{21} = 0.8$.
- $\mathbf{H}_{a,1-2}$ (many small disturbances): $\mathbf{Y}_j$, $j \in [n]$ are i.i.d. $N_p(\mathbf{0}, I_p + D)$ and $D = (d_{ik})_{p\times p}$ with $d_{ik} = 4/p$ except $d_{ii} = 0$, $i, k \in [p]$.
- $\mathbf{H}_{a,2-1}$ (one large disturbance): X_{ij} are i.i.d. Cauchy$(0, 1)$. We set the observations $Y_{1j} = X_{1j} + 0.8X_{2j}$, $Y_{2j} = X_{2j} + 0.8X_{1j}$ and $Y_{ij} = X_{ij}$, for all $i = 3, \dots, p$ and $j \in [n]$;
- $\mathbf{H}_{a,2-2}$ (many small disturbances): X_{ij} are i.i.d. Cauchy$(0, 1)$. We set the observations $Y_{ij} = X_{ij} + (7p)^{-1} \sum_{k\neq i} X_{kj}$ for $i \in [p]$ and $j \in [n]$;
- $\mathbf{H}_{a,3-1}$ (one large disturbance): the vectors $(Y_{1,j}, \dots, Y_{\lfloor p/3\rfloor,j})$, $j \in [n]$ are i.i.d. $N_{\lfloor p/3\rfloor}(\mathbf{0}, I_{\lfloor p/3\rfloor} + C')$ and $C' = (c_{ik})_{\lfloor p/3\rfloor\times\lfloor p/3\rfloor}$ with $c_{ik} = 0$, $i, k = 1, \dots, \lfloor p/3\rfloor$, except $c_{12} = c_{21} = 0.8$. Moreover, $Y_{i,j}$, $i = \lfloor p/3\rfloor + 1, \dots, n$, $j \in [n]$ are the same as those in $\mathbf{H}_{0,3}$.
- $\mathbf{H}_{a,3-2}$ (many small disturbances): the vectors $(Y_{1,j}, \dots, Y_{\lfloor p/3\rfloor,j})$, $j \in [n]$ are i.i.d. $N_{\lfloor p/3\rfloor}(\mathbf{0}, I_{\lfloor p/3\rfloor} + D')$ and $D' = (d_{ik})_{\lfloor p/3\rfloor\times\lfloor p/3\rfloor}$ with $d_{ik} = 12/p$ except $d_{ii} = 0$, $i, k = 1, \dots, \lfloor p/3\rfloor$. Moreover, Y_{ij}, $i = \lfloor p/3\rfloor + 1, \dots, n$, $j \in [n]$ are the same as those in $\mathbf{H}_{0,3}$.

The results of sizes and powers listed in Table 1 are based on the choices of $(n, p) = (60, 40)$, $(120, 80)$, $(60, 10)$ and $(120, 160)$ and 1000 replications. The tuning parameters of W_2 and W_7 are set to be $k = 4$ and $\delta = 0.5$, which will be explained later. In the case of $(n, p) = (120, 160)$, W_4 is not defined. Hence, we ignore it in Table 1. Moreover, in the cases of $\mathbf{H}_{0,2}$ and $\mathbf{H}_{0,3}$, the distribution assumption or the moment assumption is violated for the statistics W_1, W_3, W_4 and W_5. Therefore, the corresponding values are also absent. We summarize our findings as follows.

(1) The sizes of W_2, W_3, W_4 and W_7 are close to the nominal size 5%. However, W_3 has some size distortion in the case of $(n, p) = (120, 160)$. Meanwhile, the sizes of W_1, W_5 and W_6 tend to be smaller than 5%.

(2) If the alternative hypothesis is the case of one large disturbance, W_5, W_6 and W_7 outperform the other statistics. In contrast, if the alternative hypothesis is the case of many small disturbances, W_1, W_2, W_3, W_4 and W_7 have better performance than W_5 and W_6.

Overall, the size of W_7 is close to the nominal level $\alpha = 5\%$ in our simulation study. Moreover, W_7 has higher powers than the other statistics in most cases of the alternative hypotheses.

Finally, we consider how to choose the parameters k and δ in W_2 and W_7. For illustration, we consider the case $\mathbf{H}_{0,1}$ versus $\mathbf{H}_{a,1-1}$. The parameter k's are chosen to be $2, 4, 6, 8$ and 10. The parameter δ's are chosen to be $0.3, 0.4, 0.5, 0.6, 0.7$ and 0.8, The sample size n and the dimension p are set to be $(60, 40)$, $(120, 80)$, $(60, 10)$ and $(120, 160)$. The sizes and powers of W_2 and

TABLE 1
The sizes and powers (percentage) of W_1 to W_7 for different hypotheses, sample size n and dimension p

(n, p)	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_2	W_6	W_7	W_2	W_6	W_7
	$\mathbf{H}_{0,1}$							$\mathbf{H}_{0,2}$			$\mathbf{H}_{0,3}$		
(60, 40)	0.4	4	5	4.9	2.1	3.2	4.5	4.7	1.9	4.9	4.5	2.0	4.8
(120, 80)	1.4	5	5.6	4.9	2.6	2.2	5.7	4.6	2.6	5.4	3.7	3.7	4.4
(60, 10)	0.4	3.6	2.1	3.1	3.8	3.4	4.1	3.6	2.2	4.2	3.9	3.6	4.2
(120, 160)	1.7	5.8	10.9	–	1.9	2.3	6.7	4	2.8	4.8	4.2	2.4	4.6
	$\mathbf{H}_{a,1-1}$							$\mathbf{H}_{a,2-1}$			$\mathbf{H}_{a,3-1}$		
(60, 40)	2.5	13.9	22.6	17.4	100	100	92.9	25.6	100	99.7	12.4	99.9	98.2
(120, 80)	5.6	13.2	25	20.5	100	100	97.5	26.2	100	100	12.1	100	96
(60, 10)	29.4	84.3	96.2	99.5	100	100	99.3	100	100	100	83.4	100	99.5
(120, 160)	2.2	7.7	21.5	–	100	100	99.6	11.4	100	100	6.5	100	94.6
	$\mathbf{H}_{a,1-2}$							$\mathbf{H}_{a,2-2}$			$\mathbf{H}_{a,3-2}$		
(60, 40)	99.9	99.8	99.7	67.5	15.6	18	99.8	98.2	28.7	97.8	99.9	65.4	99.9
(120, 80)	100	100	100	72.4	11.8	12.3	100	100	68.7	100	100	39	100
(60, 10)	100	100	100	100	99.8	99.5	100	91.7	45.6	91.4	100	100	100
(120, 160)	100	99.3	99.2	–	4	4.3	98.9	100	73.2	100	99.6	6.1	99.4

W_7 are given in Tables 2 and 3, respectively. Based on Tables 2 and 3, for a fixed value of δ, one can see that their sizes and powers are robust when $k \geq 4$. Therefore, we suggest to set k as 4 for both W_2 and W_7. (Of course, theoretically, any k larger than 4 is applicable, but it will increase the computational burden without significant benefit).

The parameter δ should be chosen appropriately, so that W_7 will inherit the desirable properties of W_2 and W_6. In principle, δ should not be very small or very large. If it is too close to zero, that is, $n^{-\delta}$ is close to one, the size of W_7 will be influenced since the limiting null distribution of W_7 may not be standard normal any more. On the other hand, if δ is relatively large, W_7 will not detect the dependent case where W_6 works. In other words, the power of W_7 is weak. In Tables 2 and 3, for $k = 4$, we can see that W_7 performs the best when δ is 0.5 for these four combinations of (n, p). We also conduct simulations under the other two null hypotheses $\mathbf{H}_{0,2}$ and $\mathbf{H}_{0,3}$. The results are similar to $\mathbf{H}_{0,1}$. So based on our simulations we suggest to *use* $k = 4$, $\delta = 0.5$. This empirical choice is independent of the specific distribution of $(Y_1, \dots, Y_p)$ since both W_2 and W_7 are distribution free.

3. Preliminaries and tools from Anderson and Zeitouni [1]. In this section, we will introduce some basic notions concerning cumulants and some necessary results from [1]. For some positive integer N and random variables $\xi_1, \dots, \xi_N$, the

TABLE 2
The sizes (percentage) of W_2 and W_7 for different n, p, k and δ under the null hypothesis $\mathbf{H}_{0,1}$

			W_7						
(n, p)	**k**	**W_2**	**$\delta =$**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
(60, 40)	2	4.8		22.7	14.7	10.5	7.9	7	6.4
	4	4.5		19	12.7	6	5.2	4.9	4.9
	6	4		16.4	9.6	5.8	5.4	4.7	4.2
	8	3.3		16.9	9	4.8	4.2	3.7	3.3
	10	4		16.2	6.5	4	3.6	3.8	4
(120, 80)	2	5.4		14.7	9.5	7.5	6.2	5.7	5.6
	4	5.2		14.3	8.9	5.5	5.4	5.3	5.3
	6	4.1		12.7	6.3	4.8	4.4	4.1	4.1
	8	4.2		12	6.7	5.5	4.8	4.8	4.5
	10	3.8		13.3	7.2	5	4.2	4	4
(60, 10)	2	3.9		21.3	13	5.9	5	4.8	4.3
	4	4.3		18.2	8.8	4.9	4.7	4.6	4.6
	6	4.4		17.6	7.6	5.7	4.7	4.6	4.5
	8	4.9		15.5	7	5.6	5.3	5.1	5
	10	5		13.6	6.5	5	5.2	5.1	5.1
(120, 160)	2	6		22.7	12.4	9.3	8.2	7.2	6.7
	4	5.4		13.9	8	5.8	5.6	5.5	5.4
	6	5.3		15.2	9.1	5.7	5.4	5.3	5.3
	8	5.3		13.4	8.6	5.8	5.6	5.6	5.4
	10	4.3		13.4	7.4	5.2	4.8	4.5	4.3

joint cumulant $\mathbf{C}(\xi_1, \dots, \xi_N)$ is defined by

$$\mathbf{C}(\xi_1, \dots, \xi_N) = \mathrm{i}^{-N} \frac{\partial^N}{\partial x_1 \cdots \partial x_N} \log \mathbb{E} \exp\left(\sum_{j=1}^{N} \mathrm{i} x_j \xi_j \right) \Bigg|_{x_1 = \cdots = x_N = 0}. \tag{3.1}$$

It is straightforward to check via above definition that the following properties hold:

P1: $\mathbf{C}(\xi_1, \dots, \xi_N)$ *is a symmetric function of* $\xi_1, \dots, \xi_N$.
P2: $\mathbf{C}(\xi_1, \dots, \xi_N)$ *is a multilinear function of* $\xi_1, \dots, \xi_N$.
P3: $\mathbf{C}(\xi_1, \dots, \xi_N) = 0$ *if the variables* $\xi_i, i \in [N]$ *can be split into two groups* $\{\xi_i\}_{i \in \mathbb{S}_1}$ *and* $\{\xi_i\}_{i \in \mathbb{S}_2}$ with $\mathbb{S}_1 \cap \mathbb{S}_2 = \varnothing$ *and* $\mathbb{S}_1 \cup \mathbb{S}_2 = [N]$ *such that the sigma field* $\sigma\{\xi_i\}_{i \in \mathbb{S}_1}$ *is independent of* $\sigma\{\xi_i\}_{i \in \mathbb{S}_2}$.

It is well known that the joint cumulant can be expressed in terms of moments. To state this expression, we need some notions about set partition. For some positive number N, let L_N be the lattice consisting of all the partitions of $[N]$. We say

TABLE 3
The powers (percentage) of W_2 and W_7 for different n, p, k and δ under the alternative hypothesis $\mathbf{H}_{a,1-1}$

				W_7					
(n, p)	**k**	**W_2**	**$\delta =$**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
(60, 40)	2	12.6		99.2	96.9	88.5	71.2	51.1	34.9
	4	15.9		99.4	98.4	92.9	75.4	54.1	39.3
	6	15.4		99.7	97.9	91.8	73.7	53	36.2
	8	16		99.6	98.2	90.9	70	47.5	33.3
	10	12.8		99.5	97.5	88.7	66.7	42.3	29.2
(120, 80)	2	13.4		100	100	99.9	96.5	76.6	50.9
	4	13.2		100	100	100	97.5	79.2	54.3
	6	14.6		100	100	100	96.6	77.7	52.5
	8	15.2		100	100	100	96.5	76.2	49.3
	10	18.8		100	100	100	97.2	76.4	51.4
(60, 10)	2	91.6		100	100	99.9	99.7	99.8	98.7
	4	84.9		100	100	99.9	99.5	98.4	96.9
	6	83.4		100	100	100	99.4	98	96.4
	8	80.8		100	100	99.9	99.3	97	95
	10	83.1		100	100	99.9	98.7	97	94
(120, 160)	2	5.2		100	100	99.8	87.5	51.7	26.1
	4	7.1		100	100	99.7	92	61.6	33.7
	6	7.8		100	100	99.7	91.8	61.3	34.7
	8	9.2		100	100	99.7	91.7	62.6	35.3
	10	8.9		100	100	99.9	91.2	59.1	32.8

$\pi = \{B_1, \dots, B_m\} \in L_N$ is a *partition* of the set $[N]$ if

$$\varnothing \neq B_i \subset [N], i = 1, \dots, m, \qquad \bigcup_{i=1}^{m} B_i = [N], \qquad B_i \cap B_j = \varnothing, \text{ if } i \neq j.$$

We say B_i's are *blocks* of π and m is the *cardinality* of π. We will also conventionally use the notation $\#\pi$ to denote the cardinality of a partition π all the way. L_N is a *poset* (partially ordered set) ordered by set inclusion. Specifically, given two partitions π and σ in L_N, we say $\pi \leq \sigma$ (or π is a *refinement* of σ) if every block of π is contained in a block of σ. Now given two partitions $\sigma_1, \sigma_2 \in L_N$, with the above order "$\leq$" we define $\sigma_1 \vee \sigma_2$ to be the least upper bound of σ_1 and σ_2. For example, let $N = 8$ and

$$\sigma_1 = \{\{1, 2\}, \{3, 4, 5\}, \{6\}, \{7, 8\}\}, \qquad \sigma_2 = \{\{1, 3\}, \{2, 5\}, \{4\}, \{6, 8\}, \{7\}\}.$$

Then we have $\sigma_1 \vee \sigma_2 = \{\{1, 2, 3, 4, 5\}, \{6, 7, 8\}\}$. With these notation, we have the following basic expression of joint cumulant in terms of moments,

$$\mathbf{C}(\xi_1, \dots, \xi_N) = \sum_{\pi \in L_N} (-1)^{\#\pi - 1} (\#\pi - 1)! \mathbb{E}_\pi(\xi_1, \dots, \xi_N), \tag{3.2}$$

where $\mathbb{E}_\pi(\xi_1, \dots, \xi_N) = \prod_{A\in\pi} \mathbb{E} \prod_{i\in A} \xi_i$.[4] Especially, one has

$$\mathbf{C}(\xi) = \mathbb{E}\xi, \qquad \mathbf{C}(\xi_1, \xi_2) = \mathrm{Cov}(\xi_1, \xi_2).$$

With the above concepts, we can now introduce the formula of joint cumulants of $\operatorname{tr} S_n^{k_l}, l = 1, \dots, r$ with $\sum_{l=1}^r k_l = k$, derived in [1]. To this end, we need to specify two partitions $\pi_0, \pi_1 \in L_{2k}$ as

$$\begin{aligned}\pi_0 &= \big\{\{1, 2\}, \{3, 4\}, \dots, \{2k-1, 2k\}\big\},\\ \pi_1 &= \big\{\{2, 3\}, \dots, \{\mathbf{k}_1, 1\}, \{\mathbf{k}_1 + 2, \mathbf{k}_1 + 3\}, \dots,\\ &\qquad \{\mathbf{k}_2, \mathbf{k}_1 + 1\}, \dots, \{\mathbf{k}_{r-1} + 2, \mathbf{k}_{r-1} + 3\}, \dots, \{\mathbf{k}_r, \mathbf{k}_{r-1} + 1\}\big\},\end{aligned}$$

where $\mathbf{k}_i = 2\sum_{j=1}^i k_j, i = 1, \dots, r$ and k_j's are nonnegative integers. Observe that $\#\pi_0 \vee \pi_1 = r$. Now let $L_N^{2+} \subset L_N$ be the set consisting of those partitions in which each block has cardinality larger than 2. If each block of a partition has cardinality 2, we call such a partition a *perfect matching*. For even N, we let L_N^2 ($\subset L_N^{2+}$) be the set consisting of all perfect matchings. In the sequel, we will also use the notation L_{2k}^4 to denote the set consisting of the partitions of $[2k]$ containing one 4-element block and $(k-2)$ 2-element blocks. Let $j_\alpha \in [n]$ for $\alpha \in [N]$. We use the terminology in [1] to call the index vector $\mathbf{j} := (j_1, \dots, j_N)$ a (n, N)-*word*. Moreover, we say an (n, N)-word $\mathbf{j}$ is π-measurable for some partition $\pi \in L_N$ when $j_\alpha = j_\beta$ if α, β are in the same block of π. Then by the discussions in [1] (see Proposition 5.2 therein), one has

$$\begin{aligned}&\mathbf{C}\big(\operatorname{tr} S_n^{k_1}, \dots, \operatorname{tr} S_n^{k_r}\big)\\ &\quad= \sum_{\substack{\pi\in L_{2k}^{2+}\\ \text{s.t. } \#\pi_0\vee\pi_1\vee\pi=1}} p^{-k+\#\pi_0\vee\pi} \sum_{\substack{\mathbf{j}:(n,2k)\text{-word}\\ \text{s.t. } \mathbf{j} \text{ is } \pi_1 \text{ measurable}}} \mathbf{C}_\pi(\mathbf{j}),\end{aligned} \tag{3.3}$$

where

$$\mathbf{C}_\pi(\mathbf{j}) := \mathbf{C}_\pi(Z_{j_1}, \dots, Z_{j_{2k}}) = \prod_{A\in\pi} \mathbf{C}\{Z_{j_\alpha}\}_{\alpha\in A}.$$

Equation (3.3) was derived by using Möbius inversion formula in [1]. We remind that, here we switch the roles of the parameters p and n in the setting of [1]. Moreover, $B(\mathbf{j})$ therein is always 1 for all $(n, 2k)$-words $\mathbf{j}$ in our case.

Our aim is to show that for $k_1, \dots, k_r$ with $\sum_{j=1}^r k_j = k$,

$$\mathbf{C}\big(\operatorname{tr} S_n^{k_1}, \dots, \operatorname{tr} S_n^{k_r}\big) = \begin{cases} o(1), & \text{if } r \ge 3,\\ \text{r.h.s. of (1.2)} + o(1), & \text{if } r = 2,\\ \text{r.h.s. of (1.3)}, & \text{if } r = 1.\end{cases} \tag{3.4}$$

---

[4] Here, we remind that the partition π in the notation $\mathbb{E}_\pi(\cdot)$ takes effect on the positions of the components of the vector, so does the notation $\mathbf{C}_\pi(\cdot)$ in (3.3). The reader should not confuse the positions with the subscripts of the components of the random vector. For example, for $\pi = \{\{1, 2\}, \{3, 4\}\}$, we have $\mathbb{E}_\pi(Z_{j_2}, Z_{j_3}, Z_{j_4}, Z_{j_1}) = \mathbb{E}(Z_{j_2}Z_{j_3})\mathbb{E}(Z_{j_4}Z_{j_1})$ rather than $\mathbb{E}(Z_{j_1}Z_{j_2})\mathbb{E}(Z_{j_3}Z_{j_4})$.

It is well known that (3.4) can imply Theorem 1.1 directly.

Apparently, by (3.3) we see that, to prove (3.4), the main task is to estimate the summation

$$\sum_{\substack{\mathbf{j}:(n,2k)\text{-word} \\ \text{s.t. } \mathbf{j} \text{ is } \pi_1 \text{ measurable}}} \mathbf{C}_\pi(\mathbf{j}) \tag{3.5}$$

for various π. In order to deal with the estimation in the counterpart of [1], the main assumption imposed on the stationary sequence $\{Y_k\}_{k=-\infty}^{\infty}$ considered therein is the so-called *joint cumulant summability*

$$\sum_{j_1} \cdots \sum_{j_r} \bigl|\mathbf{C}(Y_0, Y_{j_1}, \dots, Y_{j_r})\bigr| = O(1) \qquad \text{for all } r \geq 1. \tag{3.6}$$

Actually, once joint cumulant summability held for random sequence $\{Z_i\}_{i=1}^n$, one could obtain that the summation (3.5) was bounded by $O(n^{\#\pi_1 \vee \pi})$ in magnitude. However, the verification of the joint cumulant summability for a general random sequence is quite nontrivial. Without joint cumulant summability, we will provide a weaker bound on (3.5) below directly; see Proposition 4.2 and Corollary 4.4. Such a weaker bound will still make our proof strategy amenable. The following proposition proved by Anderson and Zeitouni in [1] will be crucial to ensure that our weaker bound on (3.5) still works well in the proof of (3.4).

PROPOSITION 3.1 (Proposition 3.1, [1]). *Let $\Pi \in L_{2k}^{2+}$ and $\Pi_0, \Pi_1 \in L_{2k}^2$ for some positive integer k. If $\#\Pi \vee \Pi_0 \vee \Pi_1 = 1$ and $\#\Pi_0 \vee \Pi_1 = r$ for some positive integer $r \leq k$, one has*

$$\#\Pi_0 \vee \Pi + \#\Pi_1 \vee \Pi \leq (\#\Pi + 1)\mathbf{1}_{\{r=1\}} + \min\{\#\Pi + 1, k + 1 - r/2\}\mathbf{1}_{\{r\geq 2\}}.$$[5]

At the end of this section, we state some elementary properties of the vector $\mathbf{Z}$ which will be used in the subsequent sections. We summarize them as the following lemma whose proof will be stated in the supplementary material [5].

LEMMA 3.2. *Let $\mathbf{Z} = (Z_1, \dots, Z_n)$ be the random vector defined above. Let m and α_i, $i = 1, \dots, m$ be fixed positive integers. We have the following properties of $\mathbf{Z}$*:

(i) *If $\sum_{i=1}^m \alpha_i$ is odd,*

$$\mathbb{E}\bigl(Z_1^{\alpha_1} \cdots Z_m^{\alpha_m}\bigr) = 0. \tag{3.7}$$

[5]One might note that in Proposition 3.1 of [1], the authors stated a slightly weaker bound $k + 1 - \lfloor r/2 \rfloor$ in formula (14) therein. However, according to the proof of Proposition 3.1 in [1], it is not difficult to see that one can improve it to $k + 1 - r/2$. We appreciate Professor Greg W. Anderson's confirmation on this.

(ii) *If* $\sum_{i=1}^{m}\alpha_i$ *is even*,

$$\mathbb{E}(Z_1^{\alpha_1}\cdots Z_m^{\alpha_m}) = O(n^{-\mathfrak{n}_o(\boldsymbol{\alpha})/2}), \tag{3.8}$$

where $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_m)$ *and* $\mathfrak{n}_o(\boldsymbol{\alpha}) = \#\{i \in [m] : \alpha_i$ *is odd*$\}$. *Moreover, by symmetry*, (i) *and* (ii) *still hold if we replace* $(Z_1,\dots,Z_m)$ *by* $(Z_{l_1},\dots,Z_{l_m})$ *with any mutually distinct indices* $l_1,\dots,l_m \in [n]$.

Note that (3.2) together with (i) of Lemma 3.2 implies that

$$\mathbf{C}(Z_{j_1},\dots,Z_{j_r}) = 0 \qquad \text{if } r \text{ is odd.} \tag{3.9}$$

Consequently, it suffices to consider those partition π in which each block has even cardinality. We denote L_{2k}^{even} $(\subset L_{2k}^{2+})$ to be the set of such partitions. Then we can rewrite (3.3) as

$$\begin{aligned}&\mathbf{C}(\operatorname{tr} S_n^{k_1},\dots,\operatorname{tr} S_n^{k_r}) \\ &\quad = \sum_{\substack{\pi\in L_{2k}^{\text{even}} \\ \text{s.t. } \#\pi_0\vee\pi_1\vee\pi=1}} p^{-k+\#\pi_0\vee\pi} \sum_{\substack{\mathbf{j}:(n,2k)\text{-word} \\ \text{s.t. } \mathbf{j} \text{ is } \pi_1 \text{ measurable}}} \mathbf{C}_\pi(\mathbf{j}).\end{aligned} \tag{3.10}$$

4. A rough bound on the summation (3.5).

4.1. *Factorization of the summation.* As mentioned in the last section, instead of checking the property of joint cumulant summability,[6] we will try to provide a rough bound on the summation (3.5) for any given $\pi \in L_{2k}^{\text{even}}$ directly. Now we observe by definition that

$$\mathbf{C}_\pi(\mathbf{j}) = \prod_{A\in\pi}\mathbf{C}\{Z_{j_\alpha}\}_{\alpha\in A} = \prod_{B\in\pi_1\vee\pi}\prod_{\substack{A\in\pi \\ \text{s.t. } A\subset B}}\mathbf{C}\{Z_{j_\alpha}\}_{\alpha\in A}.$$

For simplicity, we denote $q := \#\pi_1\vee\pi$ and index the blocks in $\#\pi_1\vee\pi$ in any fixed order as $B_i, i\in[q]$ for convenience. We set

$$\mathfrak{b}_i := \frac{\#B_i}{2}, \qquad m_i := \#\{A\in\pi : A\subset B_i\}, \qquad i\in[q],$$

and index the subsets of B_i in π in any fixed order by $A_i^{(\beta)}$, $\beta\in[m_i]$. Moreover, we define

$$\mathfrak{a}_i(\beta) := \frac{\#A_i^{(\beta)}}{2}.$$

[6] As mentioned above, once joint cumulant summability held, the magnitude of (3.5) could be bounded by $O(n^{\#\pi_1\vee\pi})$. To see this, one can refer to Proposition 6.1 of [1] by setting $b = p$ therein and switching the role of p by n to adapt to our notation.

Then we can write

$$\mathbf{C}_\pi(\mathbf{j}) = \prod_{i=1}^{q} \prod_{\beta=1}^{m_i} \mathbf{C}\{Z_{j_\alpha}\}_{\alpha \in A_i^{(\beta)}}. \tag{4.1}$$

To simplify the notation, we use $\mathcal{J} := \mathcal{J}(\pi_1, n, k)$ to represent the set consisting of all π_1-measurable $(n, 2k)$-word $\mathbf{j}$. Fix $\alpha \in [2k]$, we can view $\alpha : \mathcal{J} \to [n]$ defined by $\alpha(\mathbf{j}) := j_\alpha$ as a functional on $\mathcal{J}$. Then it is apparent that if $j_\alpha \equiv j_\beta$ [i.e., $\alpha(\mathbf{j}) \equiv \beta(\mathbf{j})$] on $\mathcal{J}$ for fixed $\alpha, \beta \in [2k]$, one has the fact that α and β are in the same block of π_1. We have the following lemma on the properties of the triple $(\mathcal{J}, \pi, \pi_1)$ whose proof will be put in the supplementary material [5].

LEMMA 4.1. *Regarding $\{j_\alpha\}_{\alpha \in [2k]}$ as a class of $2k$ functionals on $\mathcal{J}$, we have*:

(i) *Given $i \in [q]$, for any $\alpha \in B_i$, there exists exactly one $\gamma \in B_i$ such that $\alpha \neq \gamma$ but $j_\alpha \equiv j_\gamma$ on $\mathcal{J}$.*

(ii) *Given $i \in [q]$ with $m_i \geq 2$, for any proper subset $P \subset [m_i]$, there exists at least one $\alpha_1 \in \bigcup_{\beta \in P} A_i^{(\beta)}$, for any other $\alpha_2 \in \bigcup_{\beta \in P} A_i^{(\beta)}$ one has $j_{\alpha_1} \not\equiv j_{\alpha_2}$ on $\mathcal{J}$.*

For convenience, we use the notation

$$\mathbf{j}|_B = (j_\alpha)_{\alpha \in B} \tag{4.2}$$

for any $B \subset [2k]$. Analogously, for any partition σ and $D \subset [2k]$ we will use the notation

$$\sigma|_D = \{E \in \sigma : E \subset D\},$$

which can be viewed as the partition σ restricted on the set D. Actually, Lemma 4.1 is just a direct consequence of the fact that $\pi_1|B_i$ is a perfect matching and

$$\#\pi|_{B_i} \vee \pi_1|_{B_i} = \#(\pi \vee \pi_1)|_{B_i} = 1, \qquad i \in [q].$$

Analogous to the notation $\mathcal{J}(\pi_1, n, k)$, we denote $\mathcal{J}(\pi_1|_{B_i}, n, \mathfrak{b}_i)$ to be the set consisting of all $\pi_1|_{B_i}$-measurable words $\mathbf{j}|_{B_i}$. By (4.1) and (i) of Lemma 4.1, we see that for given π

$$(3.5) := \sum_{\mathbf{j} \in \mathcal{J}} \mathbf{C}_\pi(\mathbf{j}) = \prod_{i=1}^{q} \sum_{\mathbf{j}|_{B_i} \in \mathcal{J}(\pi_1|_{B_i}, n, \mathfrak{b}_i)} \mathbf{C}_{\pi|_{B_i}}(\mathbf{j}|_{B_i}). \tag{4.3}$$

Hence, to estimate (3.5), it suffices to provide a bound on the quantity

$$\sum_{\mathbf{j}|_{B_i} \in \mathcal{J}(\pi_1|_{B_i}, n, \mathfrak{b}_i)} \mathbf{C}_{\pi|_{B_i}}(\mathbf{j}|_{B_i}) = \sum_{\mathbf{j}|_{B_i} \in \mathcal{J}(\pi_1|_{B_i}, n, \mathfrak{b}_i)} \prod_{\beta=1}^{m_i} \mathbf{C}\{Z_{j_\alpha}\}_{\alpha \in A_i^{(\beta)}}$$

for all fixed $i \in [q]$. For simplicity, we discard the subscript i in the discussion below. Thus, we will use the notation $B, A^{(\beta)}, \mathfrak{a}(\beta), \mathfrak{b}, m$ to replace $B_i, A_i^{(\beta)}, \mathfrak{a}_i(\beta), \mathfrak{b}_i, m_i$ temporarily. Our main technical result is the following crucial proposition.

PROPOSITION 4.2. *With the above notation, we have*

$$\sum_{\mathbf{j}|_B \in \mathcal{J}(\pi_1|_B, n, \mathfrak{b})} \prod_{\beta=1}^{m} \left|\mathbf{C}\{Z_{j_\alpha}\}_{\alpha \in A^{(\beta)}}\right| = O\big(n^{1+\sum_{\beta=1}^{m}(\mathfrak{a}(\beta)-2)\mathbf{1}_{\{\mathfrak{a}(\beta)\geq 2\}}}\big). \tag{4.4}$$

REMARK 4.3. Here, we draw the attention that by (i) of Lemma 4.1, the indices in $\mathbf{j}|_B$ are equivalent in pairs on $\mathcal{J}(\pi_1|_B, n, \mathfrak{b})$. Under this constraint, the number of free $\mathbf{j}$ indices in the summation on the left-hand side of (4.4) is actually $\mathfrak{b} = \sum_{i=1}^{m} \mathfrak{a}(\beta)$.

Now we use $\mathfrak{n}_\gamma(\pi)$ to represent the number of blocks in π whose cardinalities are 2γ. We have the following corollary whose proof follows from Proposition 4.2 and (4.3) directly.

COROLLARY 4.4. *With the above notation, we have*

$$\sum_{\mathbf{j}\in\mathcal{J}} \left|\mathbf{C}_\pi(\mathbf{j})\right| = O\big(n^{\#\pi_1\vee\pi + \sum_{\gamma\geq 2}(\gamma-2)\mathfrak{n}_\gamma(\pi)}\big). \tag{4.5}$$

Our main task in this section is to prove Proposition 4.2. The tedious proof will be given in the remaining part of this section which will be further split into several subsections. In Sections 4.2–4.4, we will provide some preliminary results for our final evaluation scheme. The formal proof of Proposition 4.2 will be stated in Section 4.5.

4.2. *Bounds on joint cumulants.* In this subsection, we will provide some bounds on single joint cumulants with variables from $\mathbf{Z}$. Such bounds will help us to reduce all these joint cumulants to some products of 2-element cumulants which are more friendly for the subsequent combinatorial enumeration. Let s, t be fixed nonnegative integers. Now for $l_1, \dots, l_s, h_1, \dots, h_{2t} \in [n]$, we denote the vectors

$$\mathbf{l} := (l_1, l_1, l_2, l_2, \dots, l_s, l_s), \qquad \mathbf{h} := (h_1, \dots, h_{2t})$$

and we will use $\mathbf{l}$ *index* (resp., $\mathbf{h}$ *index*) to refer to $l_i, i \in [s]$ (resp., $h_i, i \in [2t]$). For simplicity, we will employ the notation

$$\widehat{\mathbf{lh}} = (l_1, l_1, l_2, l_2, \dots, l_s, l_s, h_1, \dots, h_{2t}),$$

which is the concatenation of $\mathbf{l}$ and $\mathbf{h}$. Note that the $\mathbf{l}$ indices appear in pairs. And in the sequel, we will use the notation

$$\mathbf{C}(\widehat{\mathbf{lh}}) := \mathbf{C}(Z_{l_1}, Z_{l_1}, \dots, Z_{l_s}, Z_{l_s}, Z_{h_1}, \dots, Z_{h_{2t}}) \tag{4.6}$$

and

$$\mathbf{C}(\mathbf{h}) := \mathbf{C}(Z_{h_1}, \dots, Z_{h_{2t}})$$

to highlight the index sequence. In this manner, for any partition $\tilde{\pi} \in L_{2s+2t}$, we denote

$$\mathbb{E}_{\tilde{\pi}}(\widehat{\mathbf{lh}}) := \mathbb{E}_{\tilde{\pi}}(Z_{l_1}, Z_{l_1}, \dots, Z_{l_s}, Z_{l_s}, Z_{h_1}, \dots, Z_{h_{2t}}).$$

Analogously, for any partition $\tilde{\sigma} \in L_{2t}$, we set

$$\mathbb{E}_{\tilde{\sigma}}(\mathbf{h}) := \mathbb{E}_{\tilde{\sigma}}(Z_{h_1}, Z_{h_2}, \dots, Z_{h_{2t}}).$$

We remind here the aforementioned convention that the partition $\tilde{\pi}$ in the notation $\mathbb{E}_{\tilde{\pi}}(\cdot)$ takes effect on the positions of components of $\widehat{\mathbf{lh}}$.

To prove Proposition 4.2, we will need the following two lemmas. Lemma 4.5 gives us an explicit order on the magnitude of 4-element cumulant, whilst Lemma 4.6 provides some rough bounds on the cumulants with more than 4 elements.

LEMMA 4.5. *Suppose that* $\mathbf{h} = (h_1, h_2, h_3, h_4)$. *Let* $d(\mathbf{h})$ *be the number of the distinct values in* $\{h_1, h_2, h_3, h_4\}$. *We have*

$$\mathbf{C}(\mathbf{h}) = O(n^{-d(\mathbf{h})+1}). \tag{4.7}$$

The proof of Lemma 4.5 will be stated in the supplementary material [5]. From Lemma 4.5, we can get the following consequences. We see that if there exists a perfect matching $\sigma = \{A_1, A_2\}$ of $\{1, 2, 3, 4\}$ such that $\{h_i\}_{i \in A_1} \cap \{h_i\}_{i \in A_2} = \varnothing$, then by using Lemma 4.5 and (1.1) we can get

$$|\mathbf{C}(\mathbf{h})| \le O(n^{-1}) |\mathbb{E}_{\sigma}(\mathbf{h})|. \tag{4.8}$$

If there is no such perfect matching, we have

$$|\mathbf{C}(\mathbf{h})| \le C |\mathbb{E}_{\sigma}(\mathbf{h})| \tag{4.9}$$

for any perfect matching $\sigma \in L_4^2$ with some positive constant C. Note that the second case occurs if and only if three or four of the indices h_1, h_2, h_3, h_4 take the same value.

The following Lemma 4.6 provides some crucial bounds on the cumulants with more than 4 underlying elements. The proof of this lemma will also be stated in the supplementary material [5].

LEMMA 4.6. *Under the above notation, we have the following bounds on the joint cumulant* $\mathbf{C}(\widehat{\mathbf{lh}})$.

(i) (*Crude bound*) *When* $s + t \ge 3$ *and* $t \ge 1$,

$$|\mathbf{C}(\widehat{\mathbf{lh}})| \le C \sum_{\sigma \in L_{2t}^2} |\mathbb{E}_{\sigma}(\mathbf{h})| \tag{4.10}$$

for some positive constant C.

(ii) *When* $s \geq 2$, $t = 1$, $l_1, \ldots, l_s$ *are mutually distinct and distinct from* h_1, h_2, *we have*

$$|\mathbf{C}(\widehat{\mathbf{lh}})| \leq O(n^{-2}). \tag{4.11}$$

(iii) *When* $s \geq 2$, $t = 1$, $l_1 = l_2$ *and* $l_2, \ldots, l_s$ *are mutually distinct and distinct from* h_1, h_2, *we have*

$$|\mathbf{C}(\widehat{\mathbf{lh}})| = O(n^{-1})|\mathbb{E}(Z_{h_1} Z_{h_2})|. \tag{4.12}$$

At the end of this subsection, we need to clear up a potential confusion which may occur when we use Lemma 4.6 in the sequel. Given an index sequence, for example, $(1, 1, 2, 2, 3, 3, 4, 4)$, we consider to use Lemma 4.6 to bound the corresponding cumulant $\mathbf{C}(Z_1, Z_1, \ldots, Z_4, Z_4)$. Obviously, we can adopt (ii) of Lemma 4.6 by setting $l_1 = 1$, $l_2 = 2$, $l_3 = 3$ and $h_1 = h_2 = 4$. Thus, $s = 3$ and $t = 1$. However, we can also say that $s = t = 2$ such that $l_1 = 1$, $l_2 = 2$ while $h_1 = h_2 = 3$ and $h_3 = h_4 = 4$. We can even say that $s = 0, t = 4$ such that $h_{2i-1} = h_{2i} = i, i = 1, \ldots, 4$. That means the determination of $\mathbf{l}$ and $\mathbf{h}$ indices as well as s and t are not substantially important. Actually in any viewpoint listed above, we can employ (ii) of Lemma 4.6. We state Lemma 4.6 with $\mathbf{l}$ and $\mathbf{h}$ in this way in order to simplify the presentation. However, when we use Lemma 4.6, we only need to check which case of (i)–(iii) is applicable to the given index sequence. Moreover, we can also represent the bounds for (ii) and (iii) in the form of the right-hand side of (4.10). In case (ii), obviously, we can find a perfect matching $\tilde{\sigma} \in L^2_{2s+2}$ such that

$$\mathbf{C}_{\tilde{\sigma}}(\widehat{\mathbf{lh}}) = \prod_{i=1}^{s} \mathbf{C}(Z_{l_i}, Z_{l_i}) \cdot \mathbf{C}(Z_{h_1}, Z_{h_2}) = \mathbb{E}(Z_{h_1}, Z_{h_2}).$$

By (1.1) and (4.11), we observe that in case (ii),

$$\begin{aligned} |\mathbf{C}(\widehat{\mathbf{lh}})| &\leq O(n^{-1})|\mathbb{E}(Z_{h_1}, Z_{h_2})| = O(n^{-1})|\mathbb{E}_{\tilde{\sigma}}(\widehat{\mathbf{lh}})| \\ &\leq O(n^{-1}) \sum_{\sigma \in L^2_{2s+2}} |\mathbb{E}_{\sigma}(\widehat{\mathbf{lh}})|. \end{aligned} \tag{4.13}$$

Note that the above bound is not as strong as (4.11) when $h_1 = h_2$. Analogously, it is easy to check that (4.13) also holds in case (iii) of Lemma 4.6.

4.3. *Cyclic product of* 2*-element cumulants.* In this subsection, we introduce the concept of cyclic product of 2-element cumulant (*cycle* in short) and the summation of this kind of products over involved components of $\mathbf{j}$ words. Such products will serve as canonical factors in the subsequent discussion on the whole product $\mathbf{C}_{\pi|_B}(\mathbf{j}|_B)$ in Proposition 4.2. Let ℓ be some positive integer and $\sigma \in L^2_{2\ell}$. As above, we use the notation $\mathcal{J}(\sigma, n, \ell)$ to denote the set of all σ-measurable $(n, 2\ell)$-words $\mathbf{j}$. Moreover, let $\sigma_0 \in L^2_{2\ell}$. Note that $\mathbf{C}_{\sigma_0}(\mathbf{j})$ is a product of ℓ 2-element cumulants. Now we define the concept of *cycle* (with respect to σ) as follows.

DEFINITION 4.7 (Cycle). Let $\sigma_0, \sigma \in L^2_{2\ell}$ and $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$. We call the cumulant product $\mathbf{C}_{\sigma_0}(\mathbf{j})$ a cycle with respect to σ if $\#\sigma \vee \sigma_0 = 1$.

REMARK 4.8. Note that actually for $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$, whether $\mathbf{C}_{\sigma_0}(\mathbf{j})$ is a cycle (with respect to σ) only depends on the perfect matchings σ and σ_0 but not on the choice of $\mathbf{j}$. However, the magnitude of a cycle $\mathbf{C}_{\sigma_0}(\mathbf{j})$ does depend on the choice of the word $\mathbf{j}$. See Lemma 4.11 below.

We can illustrate the definition of *cycle* in the following more detailed way through which the meaning of such a nomenclature can be evoked. Provided that $\#\sigma \vee \sigma_0 = 1$, it is not difficult to see that there exists a permutation ε of $[2\ell]$ such that

$$\sigma_0 = \big\{\{\varepsilon(2\alpha - 1), \varepsilon(2\alpha)\}\big\}_{\alpha=1}^{\ell}, \qquad \sigma = \big\{\{\varepsilon(2\alpha), \varepsilon(2\alpha + 1)\}\big\}_{\alpha=1}^{\ell}$$

in which we made the convention of $\varepsilon(2\ell + 1) = \varepsilon(1)$. Then for all $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$, $\mathbf{C}_{\sigma_0}(\mathbf{j})$ can be written as a product of 2-element cumulants whose indices form a cycle in the sense that if we regard $V(\mathbf{j}) := \bigcup_{\alpha=1}^{2\ell}\{j_\alpha\}$ as the set of vertices and $E(\mathbf{j}) := \bigsqcup_{\alpha=1}^{\ell}\{\overline{j_{\varepsilon(2\alpha-1)}j_{\varepsilon(2\alpha)}}\}$ as the set of edges then the multigraph $G(\mathbf{j}) = (V(\mathbf{j}), E(\mathbf{j}))$ is a cycle (i.e., closed walk). Here, the notation $\cup$ is the common union while $\sqcup$ is the disjoint union. The reader is recommended to take a look at Figure 1 for understanding the definition of a cycle. In this manner, we will also say that the word $\mathbf{j} = (j_1, \ldots, j_{2\ell})$ forms a cycle under σ_0 with respect to σ.

DEFINITION 4.9 (Product of $\mathfrak{m}$ cycles). Given $\sigma, \sigma_0 \in L^2_{2\ell}$, if $\#\sigma_0 \vee \sigma = \mathfrak{m}$ for some positive integer $\mathfrak{m} \geq 2$, we say that $\mathbf{C}_{\sigma_0}(\mathbf{j})$ with $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$ is a product of $\mathfrak{m}$ cycles with respect to σ. Actually, if $\sigma_0 \vee \sigma = \{D_1, \ldots, D_{\mathfrak{m}}\}$, then obviously $\mathbf{C}_{\sigma_0|_{D_i}}(\mathbf{j}|_{D_i})$ with $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$ is a cycle with respect to $\sigma|_{D_i}$.

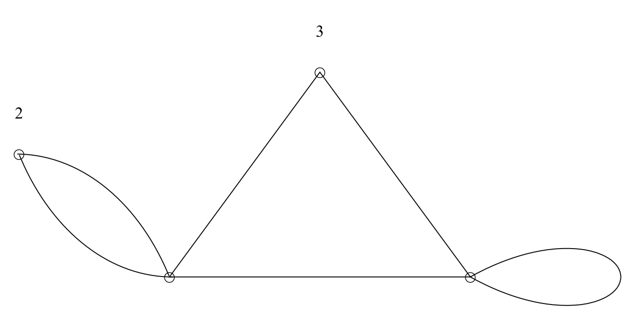

FIG. 1. *Let* $\sigma = \{\{1,2\},\{3,4\},\{5,6\},\{7,8\},\{9,10\},\{11,12\}\}$ *and* $\sigma_0 = \{\{2,3\},\{4,5\},\{6,7\},\{8,9\},\{10,11\},\{12,1\}\}$. *We take the* σ*-measurable word* $\mathbf{j}$ *to be* $(1,1,2,2,1,1,3,3,4,4,4,4)$. *Then the corresponding graph* $G(\mathbf{j})$ *for the cycle* $\mathbf{C}_{\sigma_0}(\mathbf{j})$ *is as above.*

REMARK 4.10. Note that the definition of *product of* $\mathfrak{m}$ *cycles* also only depends on σ_0 and σ. In some sense, unlike the concept of *cycle*, the *product of* $\mathfrak{m}$ *cycles* is more like a common phrase rather than a new terminology. However, we still raise it as an independent concept to emphasize the relationship between σ_0 and σ.

Actually, for some specific $\mathbf{j}$, the graphical illustration as that in the single cycle case may not evoke the name of *product of* $\mathfrak{m}$ *cycles* any more since different graphs corresponding to different cycles may have coincident vertices, and thus these cycles will be tied together if we define $G(\mathbf{j}) = (V(\mathbf{j}), E(\mathbf{j}))$ as above. To avoid this confusion, we can draw $\mathfrak{m}$ cycles separately and view the disjoint union of these components as the graph corresponding to the product of $\mathfrak{m}$ cycles. Moreover, we can use the dash line to connect coincident indices in different components. One can see the left half of Figure 2, for example. Actually, when there is some coincidence between indices from different components, we will introduce a *merge operation* to reduce the number of cycles in the product later. Before commencing this issue, we will raise a fact on the summation of single cycles. Now we have the following lemma whose proof will be stated in the supplementary material [5].

LEMMA 4.11. *Let ℓ be a fixed positive integer, and $\sigma_0, \sigma \in L^2_{2\ell}$ such that $\#\sigma_0 \vee \sigma = 1$. Assume that $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$, and thus $\mathbf{C}_{\sigma_0}(\mathbf{j})$ is a cycle with respect*

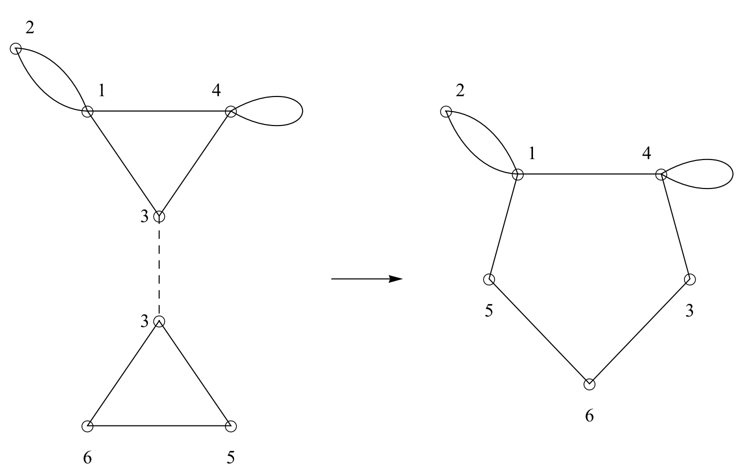

FIG. 2. *Assume that* $\sigma = \{\{2i-1, 2i\} : i = 1, \dots, 9\}$ *and* $\sigma_0 = \{\{2,3\},\{4,5\},\{6,7\},\{8,9\},\{10,11\},\{12,1\},\{13,18\},\{14,15\},\{16,17\}\}$. *We take the word* $\mathbf{j}$ *to be* $(1,1,2,2,1,1,3,3,4,4,4,4,3,3,5,5,6,6)$. *Then the figure on the left-hand side above is* $G(\mathbf{j})$ *for the product of* 2 *cycles* $\mathbf{C}_{\sigma_0}(\mathbf{j})$. *Now we fix a way to merge* $\mathbf{C}_{\sigma_0}(\mathbf{j})$ *by taking* $\tilde{\sigma}_0 = \{\{2,3\},\{4,5\},\{6,15\},\{8,9\},\{10,11\},\{12,1\},\{13,18\},\{16,17\}\}$, *then the figure on the right-hand side above is corresponding to the merged cycle.*

to σ. We have

$$|\mathbf{C}_{\sigma_0}(\mathbf{j})| = O(n^{-d(\mathbf{j})\mathbf{1}_{\{d(\mathbf{j})\geq 2\}}}), \tag{4.14}$$

where $d(\mathbf{j})$ represents the number of distinct values in the collection $\{j_\alpha\}_{\alpha=1}^{2\ell}$.

In the sequel, we call a cycle containing at least one factor $\mathbf{C}(Z_{j_{\varepsilon(2\alpha-1)}}, Z_{j_{\varepsilon(2\alpha)}})$ with $j_{\varepsilon(2\alpha-1)} \neq j_{\varepsilon(2\alpha)}$ as *in-homogeneous cycle*. Otherwise, we call it *homogeneous cycle*. With the above graphical language, a homogeneous cycle only has a single vertex and all its edges are self-loops. By contrast, an in-homogeneous cycle has at least two vertices. Following from Lemma 4.11, we have the following.

COROLLARY 4.12. *For any given positive integer ℓ, and $\sigma_0, \sigma \in L_{2\ell}^2$ such that $\#\sigma \vee \sigma_0 = 1$, we have the following corollary*:

$$\sum_{\mathbf{j}\in\mathcal{J}(\sigma,n,\ell)} |\mathbf{C}_{\sigma_0}(\mathbf{j})| = n + O(1). \tag{4.15}$$

PROOF. By using Lemma 4.11, the leading term of the left-hand side of (4.15) comes from the homogeneous cycles. Obviously, the total choice of homogeneous cycle is n. Moreover, we can see that the total contribution of the in-homogeneous cycles is $O(1)$ by (4.14). Thus, we obtain (4.15). □

Now we use Corollary 4.12 to prove a simple case of Proposition 4.2. That is $\mathfrak{a}(\beta) = 1$ for all $\beta = 1, \dots, m$. Note that in this case, by Lemma 4.1, it is not difficult to see that $\prod_{\beta=1}^m \mathbf{C}\{Z_{j_\alpha}\}_{\alpha\in A^{(\beta)}}$ is a cycle for $\mathbf{j}|_B \in \mathcal{J}(\pi_1|_B, n, \mathfrak{b})$ since $\#\pi|_B \vee \pi_1|_B = 1$. Hence, one has

$$\sum_{\mathbf{j}|_B\in\mathcal{J}(\pi_1|_B,n,\mathfrak{b})} \prod_{\beta=1}^m |\mathbf{C}\{Z_{j_\alpha}\}_{\alpha\in A^{(\beta)}}| = n + O(1). \tag{4.16}$$

We conclude this subsection by introducing the concept of *merge operation* toward the product of $\mathfrak{m}$ cycles when at least two cycles in this product have some coincident indices. Now note that if $\#\sigma \vee \sigma_0 = 2$, we see that for $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$, $\mathbf{C}_{\sigma_0}(\mathbf{j})$ is a product of two cycles by definition. In other words, there exist some permutation ε of $[2\ell]$ and $\ell_1 \in [\ell]$ such that

$$\sigma_0 = \{\{\varepsilon(2\alpha - 1), \varepsilon(2\alpha)\}\}_{\alpha=1}^{\ell}$$

and

$$\sigma = \{\{\varepsilon(2), \varepsilon(3)\}, \dots, \{\varepsilon(2\ell_1), \varepsilon(1)\}, \{\varepsilon(2\ell_1 + 2), \varepsilon(2\ell_1 + 3)\}, \dots, \\ \{\varepsilon(2\ell), \varepsilon(2\ell_1 + 1)\}\}.$$

Therefore, we have

$$\mathbf{C}_{\sigma_0}(\mathbf{j}) = \prod_{\alpha=1}^{\ell_1} \mathbf{C}(Z_{j_{\varepsilon(2\alpha-1)}}, Z_{j_{\varepsilon(2\alpha)}}) \prod_{\alpha=\ell_1+1}^{\ell} \mathbf{C}(Z_{j_{\varepsilon(2\alpha-1)}}, Z_{j_{\varepsilon(2\alpha)}}).$$

Now if for some specified $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$, there exist some $\beta \in [2\ell_1]$, $\gamma \in [2\ell] \setminus [2\ell_1]$ such that $j_{\varepsilon(\beta)}$ and $j_{\varepsilon(\gamma)}$ take the same value, we define the following *merge operation* for the two-cycle product $\mathbf{C}_{\sigma_0}(\mathbf{j})$. Without loss of generality, we let $\beta = 2\ell_1, \gamma = 2\ell_1 + 1$. Then in this case we have $j_{\varepsilon(1)} = j_{\varepsilon(2\ell_1)} = j_{\varepsilon(2\ell_1+1)} = j_{\varepsilon(2\ell)}$ since $\mathbf{j} \in \mathcal{J}(\sigma, n, \ell)$. Now by (1.1) we see that

$$
\begin{aligned}
&\left|\mathbf{C}(Z_{j_{\varepsilon(2\ell_1-1)}}, Z_{j_{\varepsilon(2\ell_1)}})\mathbf{C}(Z_{j_{\varepsilon(2\ell_1+1)}}, Z_{j_{\varepsilon(2\ell_1+2)}})\right| \\
&\quad \le \left|\mathbf{C}(Z_{j_{\varepsilon(2\ell_1-1)}}, Z_{j_{\varepsilon(2\ell_1+2)}})\right|
\end{aligned}
\tag{4.17}
$$

when $j_{\varepsilon(2\ell_1)} = j_{\varepsilon(2\ell_1+1)}$. We set

$$
\begin{aligned}
\tilde{\sigma}_0 = &\left(\sigma_0 \setminus \left\{\{\varepsilon(2\ell_1 - 1), \varepsilon(2\ell_1)\}, \{\varepsilon(2\ell_1 + 1), \varepsilon(2\ell_1 + 2)\}\right\}\right) \\
&\cup \left\{\{\varepsilon(2\ell_1 - 1), \varepsilon(2\ell_1 + 2)\}\right\}
\end{aligned}
$$

and

$$
\tilde{\sigma} = \left(\sigma \setminus \left\{\{\varepsilon(1), \varepsilon(2\ell_1)\}, \{\varepsilon(2\ell_1 + 1), \varepsilon(2\ell)\}\right\}\right) \cup \left\{\{\varepsilon(1), \varepsilon(2\ell)\}\right\}.
$$

Then we obtain that

$$
\mathbf{C}_{\tilde{\sigma}_0}(Z_{j(\varepsilon(1))} \cdots Z_{j(\varepsilon(2\ell_1-1))}, Z_{j(\varepsilon(2\ell_1+2))}, \dots, Z_{j(\varepsilon(2\ell))})
\tag{4.18}
$$

forms a cycle with respect to $\tilde{\sigma}$. We call (4.18) the *merged cycle* of $\mathbf{C}_{\sigma_0}(\mathbf{j})$ (see Figure 2, e.g.). Then by (4.17), we have

$$
\begin{aligned}
&\sum_{\substack{\mathbf{j} \in \mathcal{J}(\sigma, n, \ell) \\ \text{subject to } j_{\varepsilon(2\ell_1)} = j_{\varepsilon(2\ell_1+1)}}} \left|\mathbf{C}_{\sigma_0}(\mathbf{j})\right| \\
&\quad \le \sum_{\mathbf{j} \in \mathcal{J}(\tilde{\sigma}, n, \ell-1)} \left|\mathbf{C}_{\tilde{\sigma}_0}(Z_{j(\varepsilon(1))} \cdots Z_{j(\varepsilon(2\ell_1-1))}, Z_{j(\varepsilon(2\ell_1+2))}, \dots, Z_{j(\varepsilon(2\ell))})\right| \\
&\quad = n + O(1).
\end{aligned}
$$

Obviously, the way to do the merge operation may be not unique when there are more than one common value of the vertices from two different cycles. In this case, we can just choose one way to do the merge operation since in the sequel we only care about whether two cycles can be merged but do not care about how to merge them. Analogously, in this manner, when $\#\sigma \vee \sigma_0 \ge 3$, we can start from two cycles and use the merge operation to merge them into one cycle once there exists at least two indices (one from each cycle) taking the same value, and then we can proceed this merge operation until there is no pair of cycles can be merged into one.

4.4. *Classification of the relationships between indices.* Note that the inequalities in Lemma 4.6 rely on the relationships between the underlying indices in the joint cumulants. In order to use Lemmas 4.5 and 4.6 in the proof of Proposition 4.2, we will introduce some notation and approach to classify the relationships between the indices (components of $\mathbf{j}$).

At first, we introduce the concepts of *paired indices* and *unpaired indices* as follows. Note that for any block $\{d_1, d_2\} \in \pi_1$, we have $j_{d_1} \equiv j_{d_2}$ on $\mathcal{J}$ by definition. Now for each block $A^{(\beta)} \in \pi|_B$, we find out all π_1's blocks which are totally contained in $A^{(\beta)}$. We denote the number of such blocks by $s(\beta)$. Specifically, we find out all blocks $D_i^{(\beta)} := \{d_{i1}^{(\beta)}, d_{i2}^{(\beta)}\} \in \pi_1, i = 1, \ldots, s(\beta)$ such that $\bigcup_{i=1}^{s(\beta)} D_i^{(\beta)} \subset A^{(\beta)}$. We then set

$$l_i^{(\beta)} := j_{d_{i1}^{(\beta)}} \equiv j_{d_{i2}^{(\beta)}} \qquad \text{on } \mathcal{J}, i = 1, \ldots, s(\beta).$$

We call $l_i^{(\beta)}, i = 1, \ldots, s(\beta)$ *paired indices from* $A^{(\beta)}$ informally. The remaining indices j_α with $\alpha \in A^{(\beta)} \setminus \bigcup_{i=1}^{s(\beta)} D_i^{(\beta)}$ will be ordered (in any fixed order) and denoted by $h_i^{(\beta)}, i = 1, \ldots, 2t(\beta)$ which will be called as *unpaired indices from* $A^{(\beta)}$. Note that $h_i^{(\beta)}$ should be identical to $h_\ell^{(\gamma)}$ for some $\gamma \neq \beta$ and $\ell \in [2t(\gamma)]$. Obviously, we have $s(\beta) + t(\beta) = \mathfrak{a}(\beta)$. We remind here the word *unpaired* means that $h_i^{(\beta)}$ and $h_j^{(\beta)}$ with $i \neq j$ are not identical on $\mathcal{J}$. However, for some specific realization of $\mathbf{j} \in \mathcal{J}$, it is obvious that $h_i^{(\beta)}$ and $h_j^{(\beta)}$ may take the same value. Note that since the joint cumulant is a symmetric function of the involved variables, we can work with any specified order of these variables. Therefore, we can write

$$\mathbf{C}\{Z_{j_\alpha}\}_{\alpha \in A^{(\beta)}} = \mathbf{C}(Z_{l_1^{(\beta)}}, Z_{l_1^{(\beta)}}, \ldots, Z_{l_{s(\beta)}^{(\beta)}}, Z_{l_{s(\beta)}^{(\beta)}}, Z_{h_1^{(\beta)}}, \ldots, Z_{h_{2t(\beta)}^{(\beta)}}).$$

We mimic the notation in Section 4.2 to denote the underlying paired indices and unpaired indices sequence in $\{Z_{j_\alpha}\}_{\alpha \in A^{(\beta)}}$ by

$$\mathbf{l}^{(\beta)} := (l_1^{(\beta)}, l_1^{(\beta)}, \ldots, l_{s(\beta)}^{(\beta)}, l_{s(\beta)}^{(\beta)})$$

and

$$\mathbf{h}^{(\beta)} := (h_1^{(\beta)}, h_2^{(\beta)}, \ldots, h_{2t(\beta)}^{(\beta)}),$$

respectively. In addition, for simplicity we use the notation

$$\widehat{\mathbf{lh}}^{(\beta)} := \widehat{\mathbf{l}^{(\beta)}\mathbf{h}^{\beta}}$$

and write

$$\mathbf{C}(\widehat{\mathbf{lh}}^{(\beta)}) := \mathbf{C}\{Z_{j_\alpha}\}_{\alpha \in A^{(\beta)}}$$

by analogy with (4.6). Moreover, we will use the following notation of indices sets:

$$\{\mathbf{l}^{(\beta)}\} := \{l_1^{(\beta)}, \ldots, l_{s(\beta)}^{(\beta)}\}, \qquad \{\mathbf{h}^{(\beta)}\} := \{h_1^{(\beta)}, \ldots, h_{2t(\beta)}^{(\beta)}\}.$$

We remind here that both $\mathbf{l}^{(\beta)}$ and $\mathbf{h}^{(\beta)}$ indices are just $\mathbf{j}$ indices in different notation. We will call an index in $\mathbf{h}^{(\beta)}$, $\beta \in [m]$ as $\mathbf{h}$ index. And $\mathbf{l}$ index can be understood analogously. Moreover, when we refer to the position of an $\mathbf{l}$ or $\mathbf{h}$ index, we always mean the position of its corresponding $\mathbf{j}$ index in the word $\mathbf{j}$. In the sequel, we will also employ the notation

$$B(\mathbf{h}) = \{\alpha \in B : j_\alpha \text{ is an } \mathbf{h} \text{ index}\}. \tag{4.19}$$

Note that we can regard $h_\alpha^{(\beta)}$ and $h_\gamma^{(\beta)}$ as two different free indices in $[n]$ when we take sum over $\mathcal{J}$. However, in Lemma 4.6, the bound on the magnitude of $\mathbf{C}(\widehat{\mathbf{lh}}^{(\beta)})$ may be different according to whether $h_\alpha^{(\beta)}$ and $h_\gamma^{(\beta)}$ take the same value or not. Therefore, it is necessary to decompose the summation according to different relationships between $\mathbf{h}$ indices. For example,

$$\sum_{h_1}\sum_{h_2}\left|\mathbf{C}(Z_{h_1}, Z_{h_2})\right| = \sum_{h_1=h_2}\left|\mathbf{C}(Z_{h_1}, Z_{h_2})\right| + \sum_{h_1\neq h_2}\left|\mathbf{C}(Z_{h_1}, Z_{h_2})\right|. \tag{4.20}$$

In the above example, the terms from the first summation on the right-hand side of (4.20) (each term equals 1) are quite different from those from the second summation [each term equals $-1/(n-1)$]. For more general $\mathbf{C}(\widehat{\mathbf{lh}})$, we will introduce the following concept of *relationship matrix*.

DEFINITION 4.13 (Relationship matrix). For some positive integer N, we assume that $\ell_1, \ldots, \ell_N \in [n]$ and denote $\vec{\ell} := (\ell_1, \ldots, \ell_N)$. Let $R_{\vec{\ell}} = (\delta_{ij})_{N,N}$ with

$$R_{\vec{\ell}}(i, j) := \delta_{ij} = \begin{cases} 1, & \text{if } \ell_i = \ell_j, \\ 0, & \text{if } \ell_i \neq \ell_j. \end{cases}$$

We call $R_{\vec{\ell}}$ the relationship matrix of $\vec{\ell}$.

REMARK 4.14. Note that not all $0-1$ matrix can be a relationship matrix. For example, a matrix M with $M(1,2) = M(2,3) = 1$ while $M(1,3) = 0$ cannot be a relationship matrix.

EXAMPLE 4.1. For the vector $\mathbf{j} = (1, 1, 1, 1, 2, 3, 3, 2)$, we see the relationship matrix of $\mathbf{j}$ is the block diagonal matrix

$$R_{\mathbf{j}} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \oplus \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

4.5. *Proof of Proposition* 4.2. We recall the notation in Section 4.4 to write $\mathbf{C}\{Z_{j_\alpha}\}_{\alpha\in A^{(\beta)}}$ as

$$\mathbf{C}(\widehat{\mathbf{lh}}^{(\beta)})=\mathbf{C}(Z_{l_1^{(\beta)}},Z_{l_1^{(\beta)}},\dots,Z_{l_{s(\beta)}^{(\beta)}},Z_{l_{s(\beta)}^{(\beta)}},Z_{h_1^{(\beta)}},Z_{h_2^{(\beta)}},\dots,Z_{h_{2t(\beta)}^{(\beta)}}),$$

where $s(\beta),t(\beta)\geq 0$ are nonnegative integers and $s(\beta)+t(\beta)=\mathfrak{a}(\beta)$. Thus, our aim is to bound the following quantity:

$$\sum_{\mathbf{j}|_B\in\mathcal{J}(\pi_1|_B,n,\mathfrak{b})}\prod_{\beta=1}^m|\mathbf{C}(\widehat{\mathbf{lh}}^{(\beta)})|=\sum_{\mathbf{h}:\mathbf{j}|_B\in\mathcal{J}(\pi_1|_B,n,\mathfrak{b})}\prod_{\beta=1}^m\sum_{\mathbf{l}^{(\beta)}}|\mathbf{C}(\widehat{\mathbf{lh}}^{(\beta)})|. \tag{4.21}$$

Here, $\sum_{\mathbf{h}:\mathbf{j}|_B\in\mathcal{J}(\pi_1|_B,n,\mathfrak{b})}$ represents the summation over all choices of $\mathbf{h}$ indices along with $\mathbf{j}|_B$ running over all $\mathcal{J}(\pi_1|_B,n,\mathfrak{b})$, and

$$\sum_{\mathbf{l}^{(\beta)}}=\sum_{l_1^{(\beta)}=1}^n\cdots\sum_{l_{s(\beta)}^{(\beta)}=1}^n. \tag{4.22}$$

At first, we can handle the trivial case of $m=1$ for Proposition 4.2 as follows. Observe that in this case, $A^{(1)}=B$, and thus itself forms a block of $\pi\vee\pi_1$. Hence, $t(1)=0$ and $s(1)=\mathfrak{a}(1)$. If $\mathfrak{a}(1)\leq 2$, it is easy to see that Proposition 4.2 holds by employing (1.1) and Lemma 4.5. For the case of $\mathfrak{a}(1)\geq 3$, we need to use Lemma 4.6 by setting $h_1=h_2$ in (ii) and (iii) therein. With the aid of this setting, we can get the conclusion by noticing that except for the cases of (ii) and (iii) in Lemma 4.6 (with $h_1=h_2$ therein), the number of free indices in any other case is at most $\mathfrak{a}(1)-2$. More specifically, we can split the summation (4.22) as

$$\sum_{\mathbf{l}^{(1)}}:=\sum^{*}+\sum^{\star}+\sum^{⁂},$$

where $\sum^*$ is the summation running over the sequences $(l_1^{(1)},\dots,l_{s(1)}^{(1)})\in[n]^{s(1)}$ in which all indices are distinct from each other; $\sum^\star$ runs over the sequences in which except for one pair of coincidence indices all the others are distinct and distinct from this pair; $\sum^{⁂}$ runs over all the remaining cases. Note that the total number of the choices of indices in $\sum^{⁂}$ is $O(n^{a(1)-2})$. Then by using (ii) and (iii) of Lemma 4.6, we can actually get the stronger bound as $O(n^{\mathfrak{a}(1)-2})$ rather than $O(n^{\mathfrak{a}(1)-1})$.

Therefore, it suffices to consider the case of $m\geq 2$. The proof of this case is very complicated, so we leave it to the supplementary material [5].

4.6. *A special case*. In the sequel, we also need the following stronger bound for the special case of $\#\{\beta:\mathfrak{a}(\beta)\geq 3\}=1$ while $\#\{\beta:\mathfrak{a}(\beta)=2\}=0$.

PROPOSITION 4.15. *When* $\#\{\beta : \mathfrak{a}(\beta) \geq 3\} = 1$ *and* $\#\{\beta : \mathfrak{a}(\beta) = 2\} = 0$ *for some* $B \in \pi \vee \pi_1$, *we have*

$$\sum_{\mathbf{j}|_B \in \mathcal{J}(\pi_1|_B, n, \mathfrak{b})} \prod_{\beta=1}^{m} \left|\mathbf{C}\{Z_{j_\alpha}\}_{\alpha \in A^{(\beta)}}\right| = O\left(n^{\sum_{\beta=1}^{m}(\mathfrak{a}(\beta)-2)\mathbf{1}_{\{\mathfrak{a}(\beta)\geq 2\}}}\right).$$

The proof will be provided in the supplementary material [5]. From the above proposition, we can immediately get the following corollary.

COROLLARY 4.16. *If* $\pi \in L_{2k}^{\text{even}}$ *such that* $\mathfrak{n}_2(\pi) = 0$ *and* $\sum_{\gamma \geq 3} \mathfrak{n}_\gamma(\pi) = 1$, *we have*

$$\sum_{\mathbf{j} \in \mathcal{J}} \left|\mathbf{C}_\pi(\mathbf{j})\right| = O\left(n^{\#\pi_1 \vee \pi - 1 + \sum_{\gamma \geq 2}(\gamma-2)\mathfrak{n}_\gamma(\pi)}\right).$$

5. High order cumulants. Now with the aid of Corollaries 4.4, 4.16, and Proposition 3.1 (Proposition 3.1 of [1]) we can derive the following lemma whose proof is provided in the supplementary material [5].

LEMMA 5.1 (High order cumulants). *When* $n \to \infty$, *we have*

$$\mathbf{C}\left(\operatorname{tr} S_n^{k_1}, \ldots, \operatorname{tr} S_n^{k_r}\right) \to 0 \qquad \textit{for all } r \geq 3.$$

6. Mean and covariance functions. In this section, we prove (1.2) and (1.3). Before commencing the formal proof, we introduce some necessary notation and results on the population covariance matrix $\mathbb{E}\mathbf{Z}^*\mathbf{Z}$ at first.

6.1. *On the population matrix* $\mathbb{E}\mathbf{Z}^*\mathbf{Z}$. Let

$$T := T_{n,n} = \mathbb{E}\mathbf{Z}^*\mathbf{Z} = \begin{pmatrix} 1 & -\frac{1}{n-1} & \cdots & -\frac{1}{n-1} \\ -\frac{1}{n-1} & 1 & \cdots & -\frac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n-1} & -\frac{1}{n-1} & \cdots & 1 \end{pmatrix}_{n \times n}.$$

Note that T has one multiple eigenvalue $\frac{n}{n-1}$ with multiplicity $n-1$ and one eigenvalue 0. Roughly speaking, our aim is to find some reference sample covariance matrix of the form $\frac{1}{p}\Xi T \Xi^*$, where $\Xi := (\xi_{ij})_{p,n}$ is a random matrix with i.i.d. mean zero variance one entries, and then compare the mean and covariance functions of the spectral statistics of S_n to those of $\frac{1}{p}\Xi T \Xi^*$. For the latter, we can use the existing results from [3] and [12] to obtain the explicit formulae of the mean and covariance functions. To this end, we need to present some notions and properties on T at first. Now we denote the empirical spectral distribution of T by

$$H_n(x) := \frac{n-1}{n}\mathbf{1}\left(x \geq \frac{n}{n-1}\right) + \frac{1}{n}\mathbf{1}(x \geq 0).$$

Let $c_n = n/p$ and $m_n(z) : \mathbb{C}^+ \to \mathbb{C}^+$ satisfy

$$\begin{aligned} m_n(z) &= \int \frac{1}{t(1 - c_n - c_n z m_n(z)) - z} \, dH_n(t) \\ &= \frac{n-1}{n}\left[\frac{n}{n-1}\big(1 - c_n - c_n z m_n(z)\big) - z\right]^{-1} - \frac{1}{nz}. \end{aligned}$$

Regarding $m_n(z)$ as a Stieltjes transform, we can denote F_{c_n,H_n} as its corresponding distribution function. Moreover, we denote that $H(x) = \mathbf{1}_{\{x \geq 1\}}$. Define $m(z) : \mathbb{C}^+ \to \mathbb{C}^+$ by the equation $m(z) = [(1 - c - czm(z)) - z]^{-1}$ and set $\underline{m}(z) = -\frac{1-c}{z} + cm(z)$. In order to use the results in [12] (Theorem 1.4 therein), we need to verify the following lemma on T, which will be proved in the supplementary material [5].

LEMMA 6.1. *Under the above notation, for any fixed $z, z_1, z_2 \in \mathbb{C}^+$ we have*

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \mathbf{e}_i^* T^{1/2}\big(\underline{m}(z_1)T + I\big)^{-1} T^{1/2}\mathbf{e}_i \mathbf{e}_i^* T^{1/2}\big(\underline{m}(z_2)T + I\big)^{-1} T^{1/2}\mathbf{e}_i \\ &\qquad \to \big(\underline{m}(z_1) + 1\big)^{-1}\big(\underline{m}(z_2) + 1\big)^{-1} \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \mathbf{e}_i^* T^{1/2}\big(\underline{m}(z)T + I\big)^{-1} T^{1/2}\mathbf{e}_i \mathbf{e}_i^* T^{1/2}\big(\underline{m}(z)T + I\big)^{-2} T^{1/2}\mathbf{e}_i \\ &\qquad \to \big(\underline{m}(z) + 1\big)^{-3} \end{aligned}$$

when $n \to \infty$. Here, $\mathbf{e}_i$ is the $n \times 1$ vector with a 1 in the ith coordinate and 0's elsewhere.

6.2. *Mean function.* At first we will pursue an argument analogous to that in Section 5 to discard the negligible terms by using Corollaries 4.4, 4.16, and Proposition 3.1. And then we will evaluate the main terms by a *two-step comparison strategy* whose meaning will be clear later.

We commence with the negligible terms. Note that for the mean function, we have $r = \#\pi_0 \vee \pi_1 = 1$. Hence, we can write

$$\mathbb{E}\,\mathrm{tr}\, S_n^k = \sum_{\pi \in L_{2k}^{\mathrm{even}}} p^{-k+\#\pi_0 \vee \pi} \sum_{\mathbf{j} \in \mathcal{J}} \mathbf{C}_\pi(\mathbf{j}). \tag{6.1}$$

Now we use bound (S.54), (S.58) and (S.60) in the supplementary material [5]. Note that when $\sum_{\gamma\ge 2}\mathfrak{n}_\gamma(\pi)\ge 2$, we can easily get from (S.61) that

$$p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j})=O(n^{-1}).$$

Now we consider the case of $\sum_{\gamma\ge 2}\mathfrak{n}_\gamma(\pi)=1$. Note that, if $\mathfrak{n}_\gamma(\pi)=1$ for any $\gamma\ge 3$, we can use the improved bound in Corollary 4.16 to obtain that

$$p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j})=O(n^{-1}).$$

However, the case of $\mathfrak{n}_2(\pi)=1$, $\mathfrak{n}_\gamma=0$, $\gamma\ge 3$ does have a nonnegligible contribution to the expectation. Obviously, in this case, by (S.54) and (S.58) we see that only those terms with π satisfying

$$\#\pi\vee\pi_0+\#\pi\vee\pi_1=k \tag{6.2}$$

have $O(1)$ contribution to the total sum. Now we recall the notation L_{2k}^2 and L_{2k}^4 defined in Section 3. We can write

$$\begin{aligned}\mathbb{E}\operatorname{tr}S_n^k=&\sum_{\pi\in L_{2k}^2}p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j})\\&+\sum_{\pi\in L_{2k}^4}p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j})+o(1).\end{aligned}\tag{6.3}$$

We will not estimate the right-hand side of (6.3) by bare-handed calculation and enumeration. Instead, we will adopt a comparison approach. To this end, we need to recall some existing results on the sample covariance matrices. At first, we define a reference matrix. Let ξ,ξ_j, $j=1,\dots,n$ be i.i.d. symmetric random variables with common mean zero, variance 1 and fourth moment ν_4. Let $V=(\xi_1,\dots,\xi_n)$. Moreover, for any fixed positive integer ℓ, we assume $\mathbb{E}|\xi|^\ell\le C_\ell$ for some positive constant C_ℓ. Then we set $\mathbf{Y}=(Y_1,\dots,Y_n):=VT^{1/2}$ and let V_i, $i=1,\dots,p$ be i.i.d. copies of V. Now let Ξ be the $p\times n$ matrix with V_i as its ith row and let

$$S_n(\xi)=\frac{1}{p}\Xi T\Xi^*.$$

Actually, if we take an analogous discussion on $S_n(\xi)$ as that on S_n in the last sections, it is not difficult to see that there exists

$$\begin{aligned}\mathbb{E}\operatorname{tr}S_n^k(\xi)=&\sum_{\pi\in L_{2k}^2}p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j},\xi)\\&+\sum_{\pi\in L_{2k}^4}p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j},\xi)+o(1),\end{aligned}\tag{6.4}$$

where $\mathbf{C}_\pi(\mathbf{j},\xi)$ represents the quantity obtained by replacing Z_i by Y_i in the definition of $\mathbf{C}_\pi(\mathbf{j})$. Particularly, when ξ is Gaussian, we will write $\mathbb{E}\operatorname{tr} S_n^k(\xi)$ and $\mathbf{C}_\pi(\mathbf{j},\xi)$ as $\mathbb{E}\operatorname{tr} S_n^k(g)$ and $\mathbf{C}_\pi(\mathbf{j},g)$, respectively. Actually, since $\mathbf{Y}$ is just a linear transform of i.i.d. random sequence, the verification of Lemmas 3.2, 4.5 and 4.6 for the vector $\mathbf{Y}$ is much easier than that for $\mathbf{Z}$. The proofs of these technical results for $\mathbf{Y}$ are easily manipulated by invoking the properties P1–P3 of joint cumulants stated in Section 3. However, a more direct way to derive (6.4) is to check the property of joint cumulant summability for $\mathbf{Y}$. We sketch it as follows. At first, it is elementary check that the diagonal entries of $T^{1/2}$ are $t_d := \sqrt{(n-1)/n}$ and the off-diagonal entries are $t_o := -\sqrt{1/n(n-1)}$. Then $Y_i = t_d\xi_i + t_o\sum_{\ell\neq i}\xi_\ell$ by definition. Now let r be a fixed positive integer. Suppose that in the collection of indices $\{j_1,\dots,j_r\}\in[n]^r$, there are $r_1\geq 0$ indices taking value of 1, and the remaining $r-r_1$ indices totally take $\alpha-1$ distinct values with multiplicities $r_\beta\geq 1,\beta=2,\dots,\alpha$ such that $\sum_{\beta=1}^{\alpha} r_\beta = r$. Then by symmetry of $\mathbf{Y}$ and the properties P1–P3 of joint cumulant we have

$$\begin{aligned}
&\mathbf{C}(Y_1,Y_{j_1},\dots,Y_{j_r})\\
&\quad=\mathbf{C}(\underbrace{Y_1,\dots,Y_1}_{r_1+1},\underbrace{Y_2,\dots,Y_2}_{r_2},\dots,\underbrace{Y_\alpha,\dots,Y_\alpha}_{r_\alpha})\\
&\quad=t_d^{1+r_1}t_o^{\sum_{\beta\neq 1}r_\beta}\mathbf{C}(\underbrace{\xi_1,\dots,\xi_1}_{r+1})+\sum_{\gamma=2}^{\alpha}t_d^{r_\gamma}t_o^{1+\sum_{\beta\neq\gamma}r_\beta}\mathbf{C}(\underbrace{\xi_\gamma,\dots,\xi_\gamma}_{r+1}).
\end{aligned}$$

Obviously, the quantities $|\mathbf{C}(\underbrace{\xi_\gamma,\dots,\xi_\gamma}_{r+1})|,\gamma=1,\dots,\alpha$ are all the same and can be bounded by some positive constant from above by invoking the assumption that $\mathbb{E}|\xi|^\ell\leq C_\ell$ and the formula (3.2). Moreover, we observe that $\sum_{\beta\neq 1}r_\beta\geq\alpha-1$ and $1+\sum_{\beta\neq\gamma}r_\beta\geq\alpha-1$ for all $\gamma=2,\dots,\alpha$. In addition, we have $t_d=O(1)$, $t_o=O(n^{-1})$. Therefore,

$$\bigl|\mathbf{C}(Y_1,Y_{j_1},\dots,Y_{j_r})\bigr|=O\bigl(n^{-\alpha+1}\bigr).$$

Observe that $\alpha-1$ is the number of distinct values except for 1 in the collection $\{j_1,\dots,j_r\}$. Consequently, we have that

$$\sum_{j_1=1}^{n}\cdots\sum_{j_r=1}^{n}\bigl|\mathbf{C}(Y_1,Y_{j_1},\dots,Y_{j_r})\bigr|=O(1),$$

which implies that the joint cumulant summability holds for $\mathbf{Y}$. As explained above, we can get that the stronger bound

$$\sum_{\mathbf{j}\in\mathcal{J}}\bigl|\mathbf{C}_\pi(\mathbf{j},\xi)\bigr|=O\bigl(n^{\#\pi_1\vee\pi}\bigr)$$

holds by using the result in [1]. With the aid of this stronger bound, we can derive (6.4) by a routine discussion as before.

In addition, obviously, we have

$$\sum_{\pi\in L_{2k}^2} p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j},\xi)=\sum_{\pi\in L_{2k}^2} p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j}) \tag{6.5}$$

since this term only depends on the covariance structure T which is shared by $\mathbf{Z}$ and $\mathbf{Y}$.

For the second term on the right-hand side of (6.4), by Corollary 4.4 and (6.2) we see that it has a contribution of $O(1)$ at most. Hence, it suffices to estimate its leading term. To this end, we use (4.3) with $\mathbf{Z}$ replaced by $\mathbf{Y}$. Now we consider the sum over $\mathbf{j}|_{B_i}$ with the block B_i containing the unique 4-element block $A_i^{(\gamma)}\in\pi$. Obviously the sums over the indices with positions in the other blocks of $\pi\vee\pi_1$ are all in the case of (4.16). Without loss of generality, we can fix i and γ in the following argument. Recall the notation of paired index and unpaired index. Observe that $\mathbf{C}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(\gamma)}}$ may be in one of the following forms. When $m_i=1$, $\mathbf{C}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(\gamma)}}$ must be in the form of $\mathbf{C}(Y_{l_1^{(\gamma)}},Y_{l_1^{(\gamma)}},Y_{l_2^{(\gamma)}},Y_{l_2^{(\gamma)}})$ (case 1). When $m_i\ge 2$, it may be in the form of $\mathbf{C}(Y_{l_1^{(\gamma)}},Y_{l_1^{(\gamma)}},Y_{h_1^{(\gamma)}},Y_{h_2^{(\gamma)}})$ (case 2) or $\mathbf{C}(Y_{h_1^{(\gamma)}},Y_{h_2^{(\gamma)}},Y_{h_3^{(\gamma)}},Y_{h_4^{(\gamma)}})$ (case 3). Then we have the following lemma which will be proved in supplementary material [5].

LEMMA 6.2. *In any of the above three cases, for given $\pi|_{B_i}$ and $\pi_1|_{B_i}$, there is a unique $\sigma^{(\gamma)}\in L_4^2$ such that*

$$\prod_{\beta\neq\gamma}\mathbf{C}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(\beta)}}\cdot\mathbf{C}_{\sigma^{(\gamma)}}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(\gamma)}} \tag{6.6}$$

is a product of two cycles. The other two perfect matchings in L_4^2 will drive the above product to be only one cycle.

Note that by Proposition 4.2, we have

$$\sum_{\mathbf{j}|_{B_i}\in\mathcal{J}(\pi_1|_{B_i},n,\mathfrak{b}_i)}\prod_{\beta=1}^{m_i}\left|\mathbf{C}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(\beta)}}\right|\le O(n). \tag{6.7}$$

Our aim is to get the explicit $O(n)$ term of (6.7). To this end, we recall the discussions in Section 4. In any case of $\mathbf{C}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(\gamma)}}$, we only need to consider those $\mathbf{j}$ such that $(j_\alpha)_{\alpha\in A_i^{(\gamma)}}$ is $\sigma^{(\gamma)}$-measurable. Since we can see that all the other $\mathbf{j}|_{B_i}$ can only make an $O(1)$ contribution totally to the left-hand side of (6.7) by using Lemma 4.5 and the discussion on summations of in-homogeneous cycles in Section 4.3. Moreover, by the discussions in Section 4, we know that in any of the aforementioned three cases, the main contribution to the summation comes from the terms which can be decomposed into homogeneous cycles. In these terms, each 2-element cumulant is equal to 1. Hence, we have

$$\sum_{\mathbf{j}|_{B_i}\in\mathcal{J}(\pi_1|_{B_i},n,\mathfrak{b}_i)} \mathbf{C}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(1)}}$$

$$(6.8)\qquad = \sum_{\alpha_1,\alpha_2=1}^{n} \mathbf{C}(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2}) + O(1)$$

$$= n\mathbf{C}(Y_1, Y_1, Y_1, Y_1) + n(n-1)\mathbf{C}(Y_1, Y_1, Y_2, Y_2) + O(1).$$

Note that (6.8) also holds if we replace $\mathbf{Y}$ variables by corresponding $\mathbf{Z}$ variables.

Now for the Gaussian case we claim

$$(6.9)\qquad \sum_{\pi\in L_{2k}^4} p^{-k+\#\pi_0\vee\pi} \sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j}, g) = o(1).$$

To see (6.9), it suffices to show that for any fixed $\pi \in L_{2k}^4$, (6.7) can be strengthened to be

$$(6.10)\qquad \sum_{\mathbf{j}|_{B_i}\in\mathcal{J}(\pi_1|_{B_i},n,\mathfrak{b}_i)} \prod_{\beta=1}^{m_i} \big|\mathbf{C}\{Y_{j_\alpha}\}_{\alpha\in A_i^{(\beta)}}\big| \le O(1)$$

when ξ is Gaussian. By (6.8), it suffices to evaluate the quantity $\mathbf{C}(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2})$. Note that when $\alpha_1 = \alpha_2$,

$$(6.11)\qquad \mathbf{C}(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2}) = \mathbf{C}(Y_1, Y_1, Y_1, Y_1) = \nu_4 - 3 = 0$$

since Y_1 is Gaussian. If $\alpha_1 \ne \alpha_2$, it is not difficult to get that

$$(6.12)\qquad \mathbf{C}(Y_{\alpha_1}, Y_{\alpha_1}, Y_{\alpha_2}, Y_{\alpha_2}) = \mathbf{C}(Y_1, Y_1, Y_2, Y_2) = O(n^{-2})$$

by the definition of Y_i and propositions P1–P3 of joint cumulant. Thus, (6.10) holds, which directly implies that

$$\sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j}, g) = O(n^{\#\pi_1\vee\pi-1}), \qquad \pi \in L_{2k}^4.$$

It further yields (6.9) by combining (S.58) and the elementary fact that $\#\pi = k-1$ for $\pi \in L_{2k}^4$. Inserting (6.5) and (6.9) into (6.4), we obtain

$$\sum_{\pi\in L_{2k}^2} p^{-k+\#\pi_0\vee\pi} \sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j}) = \mathbb{E}\operatorname{tr} S_n^k(g) + o(1).$$

Therefore, for the first term on the right-hand side of (6.3), it suffices to estimate $\mathbb{E}\operatorname{tr} S_n^k(g)$. For the latter, we can use the result from [3] or [12] directly to write down

$$\sum_{\pi\in L_{2k}^2} p^{-k+\#\pi_0\vee\pi} \sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j})$$

$$(6.13)\qquad = n\int x^k\, dF_{c_n,H_n}(x) - \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{cz^k \underline{m}^3(z)/(1+\underline{m}(z))^3}{[1-c\underline{m}^2(z)/(1+\underline{m}(z))^2]}\, dz + o(1),$$

where the contour $\mathcal{C}$ is taken to enclose the interval $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ as interior. See Theorem 1.4 of [12], for instance. However, here we can further simplify (6.13) by the property of orthogonal invariance of standard Gaussian vectors. Note that when ξ is Gaussian, we have

$$\frac{1}{p}\Xi T\Xi^* \stackrel{d}{=} \frac{1}{p}\frac{n}{n-1}GG^*,$$

where $G := (g_{i,j})_{p,n-1}$ with i.i.d. $N(0,1)$ elements. Now let $\tilde{c}_n = \frac{n-1}{p}$ and $F_{\tilde{c}_n,MP}$ be Marchenko–Pastur law (MP law) with parameter $\tilde{c}_n$. Then by Theorem 1.4 of [12] and Lemma 6.1 we can rewrite (6.13) as

$$\begin{aligned}\sum_{\pi\in L_{2k}^2} p^{-k+\#\pi_0\vee\pi}\sum_{\mathbf{j}\in\mathcal{J}}\mathbf{C}_\pi(\mathbf{j}) = {}& \frac{n^k}{(n-1)^{k-1}}\int x^k\,dF_{\tilde{c}_n,MP}(x) \\ & - \frac{1}{2\pi i}\int_{\mathcal{C}}\frac{cz^k\underline{m}^3(z)/(1+\underline{m}(z))^3}{[1-c\underline{m}^2(z)/(1+\underline{m}(z))^2]}\,dz + o(1).\end{aligned}$$

Note that by (9.8.14) of [3], we see that

$$\begin{aligned}&-\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{cz^k\underline{m}^3/(1+\underline{m}(z))^3}{[1-c\underline{m}^2(z)/(1+\underline{m}(z))^2]} \\ &\quad = \frac{1}{4}[(1-\sqrt{c})^{2k}+(1+\sqrt{c})^{2k}] - \frac{1}{2}\sum_{j=0}^{k}\binom{k}{j}^2 c^j.\end{aligned}$$

Moreover, by the formula of moments of MP law (see Section 3.1.1 of [3], e.g.) one can also get that

$$\frac{n^k}{(n-1)^{k-1}}\int x^k\,dF_{\tilde{c}_n,MP}(x) = \frac{n^k}{(n-1)^{k-1}}\sum_{j=0}^{k-1}\frac{1}{j+1}\binom{k}{j}\binom{k-1}{j}\left(\frac{n-1}{p}\right)^j.$$

Now we come to estimate the second term on the right-hand side of (6.3). Now we choose ξ satisfying $\nu_4 \neq 3$. Note that (6.11) is not valid now. However, (6.12) still holds. Then in this case,

$$(6.8) = (\nu_4-3)n + O(1).$$

Note that according to (4.3), except for this B_i which containing the unique 4-element block of π, the summation over the indices with positions in $[2k]\setminus B_i$ only depends on the covariance structure since $\pi \in L_{2k}^4$. Now for $C(Z_{l_1}, Z_{l_1}, Z_{l_2}, Z_{l_2})$, we have

$$C(Z_{l_1}, Z_{l_1}, Z_{l_2}, Z_{l_2}) = \mathbf{E}Z_1^4 - 3 + o(1), \qquad l_1 = l_2$$

and

$$C(Z_{l_1}, Z_{l_1}, Z_{l_2}, Z_{l_2}) = \frac{1}{n}(1-\mathbf{E}Z_1^4) + O(n^{-2}), \qquad l_1 \neq l_2,$$

which can be checked easily by the distribution of $\mathbf{Z}$. Then we have

$$(6.8)_{[\mathbf{Y}\to\mathbf{Z}]} = -2n + O(1),$$

where $(6.8)_{[\mathbf{Y}\to\mathbf{Z}]}$ represents the quantity obtained by replacing $\mathbf{Y}$ by $\mathbf{Z}$ in (6.8). And all the other factors in (4.3) are the same as those of $\sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j},\xi)$ since they only depends on the covariance matrix T. That means

$$\sum_{\pi\in L_{2k}^4} p^{-k+\#\pi_0\vee\pi} \sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j}) = -\frac{2}{\nu_4-3} \sum_{\pi\in L_{2k}^4} p^{-k+\#\pi_0\vee\pi} \sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j},\xi) + o(1).$$

By using [12] again (see Theorem 1.4 therein), we can get that

$$\sum_{\pi\in L_{2k}^4} p^{-k+\#\pi_0\vee\pi} \sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j},\xi) = \frac{1}{\pi i}\int_{\mathcal{C}} \frac{cz^k \underline{m}^3(z)(\underline{m}(z)+1)^{-3}}{1-c\underline{m}^2(z)/(1+\underline{m}(z))^2}\,dz + o(1).$$

By (1.23) of [12], we obtain

$$\begin{aligned}\frac{1}{\pi i}\int_{\mathcal{C}} \frac{cz^k \underline{m}^3(z)(\underline{m}(z)+1)^{-3}}{1-c\underline{m}^2(z)/(1+\underline{m}(z))^2}\,dz &= 2c^{1+k}\sum_{j=0}^{k}\binom{k}{j}\left(\frac{1-c}{c}\right)^j\binom{2k-j}{k-1} \\ &\quad - 2c^{1+k}\sum_{j=0}^{k}\binom{r}{j}\left(\frac{1-c}{c}\right)^j\binom{2k+1-j}{k-1}.\end{aligned}$$

In summary, we use the Gaussian matrix as the reference matrix to obtain the value of the summation over $\pi\in L_{2k}^2$ and then use the general matrix with $\nu_4\neq 3$ as the reference one to obtain the value of the summation over $\pi\in L_{2k}^4$. We call such a comparison strategy as a *two-step comparison strategy*.

6.3. *Covariance function.* Now we estimate the covariance function. Again we start with the formula

$$\mathbf{C}(\operatorname{tr} S_n^{k_1}, \operatorname{tr} S_n^{k_2}) = \sum_{\substack{\pi\in L_{2k}^{\text{even}} \\ \text{s.t. } \#\pi_0\vee\pi_1\vee\pi=1}} p^{-k+\#\pi_0\vee\pi} \sum_{\mathbf{j}\in\mathcal{J}} \mathbf{C}_\pi(\mathbf{j}).$$

Similar to the discussion in the last subsection, by using Corollaries 4.4 and 4.16 and (S.55) again we can see that is suffices to evaluate the contribution of the summation over the partitions π satisfying $\mathfrak{n}_2(\pi)=0$ or 1 and $\mathfrak{n}_\gamma(\pi)=0$ for all $\gamma\geq 3$. Moreover, by (S.54) and (S.55) it is easy to see that $\mathbf{C}(\operatorname{tr} S_n^{k_1}, \operatorname{tr} S_n^{k_2}) = O(1)$ since $\#\pi_0\vee\pi + \#\pi_1\vee\pi \leq k$ when $r=2$. Now let

$$\begin{aligned}\tilde{L}_{2k}^2 &:= \{\pi\in L_{2k}^2 : \#\pi\vee\pi_1\vee\pi_0 = 1\},\\ \tilde{L}_{2k}^4 &:= \{\pi\in L_{2k}^4 : \#\pi\vee\pi_1\vee\pi_0 = 1\}.\end{aligned}$$

For the explicit evaluation, we adopt the aforementioned *two-step comparison strategy* again. We split the summation into the summations over $\tilde{L}_{2k}^2$ partitions and $\tilde{L}_{2k}^4$ partitions. For the first part, we compare it with that of the Gaussian case. And for the second part, we compare it with the case of $\nu_4 \neq 3$. Then it is analogous to use Theorem 1.4 of [12] and Lemma 6.1 to obtain that

$$\begin{aligned}
&\mathbf{C}(\operatorname{tr} S_n^{k_1}, \operatorname{tr} S_n^{k_2}) \\
&\quad = -\frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} \frac{z_1^{k_1} z_2^{k_2}}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\underline{m}'(z_1)\underline{m}'(z_2)\,dz_1\,dz_2 \\
&\qquad + \frac{c}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} z_1^{k_1} z_2^{k_2} \frac{d^2}{dz_1\,dz_2}\biggl(\frac{\underline{m}(z_1)\underline{m}(z_2)}{(\underline{m}(z_1)+1)(\underline{m}(z_2)+1)}\biggr)\,dz_1\,dz_2 + o(1),
\end{aligned}$$

where the contours $\mathcal{C}_1$ and $\mathcal{C}_2$ are disjoint and both enclose the interval $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ as interior. Now by (9.8.15) of [3] and (1.24) of [12], we have

$$\begin{aligned}
&-\frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} \frac{z_1^{k_1} z_2^{k_2}}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\underline{m}'(z_1)\underline{m}'(z_2)\,dz_1\,dz_2 \\
&\quad = 2c^{k_1+k_2}\sum_{j_1=0}^{k_1-1}\sum_{j_2=0}^{k_2}\binom{k_1}{j_1}\binom{k_2}{j_2}\biggl(\frac{1-c}{c}\biggr)^{j_1+j_2} \\
&\qquad \times \sum_{l=1}^{k_1-j_1} l\binom{2k_1-1-(j_1+l)}{k_1-1}\binom{2k_2-1-j_2+l}{k_2-1}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{c}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} z_1^{k_1} z_2^{k_2} \frac{d^2}{dz_1\,dz_2}\biggl(\frac{\underline{m}(z_1)\underline{m}(z_2)}{(\underline{m}(z_1)+1)(\underline{m}(z_2)+1)}\biggr)\,dz_1\,dz_2 \\
&\quad = -2c^{k_1+k_2+1}\sum_{j_1=0}^{k_1}\sum_{j_2=0}^{k_2}\binom{k_1}{j_1}\binom{k_2}{j_2}\biggl(\frac{1-c}{c}\biggr)^{j_1+j_2} \\
&\qquad \times \binom{2k_1-j_1}{k_1-1}\binom{2k_2-j_2}{k_2-1}.
\end{aligned}$$

Thus, we complete the proof of Theorem 1.1.

Acknowledgements. The authors are very grateful to Professor Greg W. Anderson for his answer to our question on Proposition 3.1 of [1]. We also thank the anonymous referees and an Associate Editor for careful reading and valuable comments.

SUPPLEMENTARY MATERIAL

Supplement to "Spectral statistics of large dimensional Spearman's rank correlation matrix and its application" (DOI: 10.1214/15-AOS1353SUPP; .pdf). This supplemental article [5] contains the proofs of Lemmas 3.2, 4.1 4.5, 4.6, 4.11, Propositions 4.2, 4.15, Lemmas 5.1, 6.1, 6.2.

REFERENCES

[1] ANDERSON, G. W. and ZEITOUNI, O. (2008). A CLT for regularized sample covariance matrices. *Ann. Statist.* **36** 2553–2576. MR2485007

[2] BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

[3] BAI, Z. D. and SILVERSTEIN, J. W. (2006). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Science Press, Beijing.

[4] BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 1–32. MR2988403

[5] BAO, Z. G., LIN, L. C., PAN, G. M. and ZHOU, W. (2015). Supplement to "Spectral statistics of large dimensional Spearman's rank correlation matrix and its application." DOI:10.1214/15-AOS1353SUPP.

[6] BOSE, A. and SEN, A. (2008). Another look at the moment method for large dimensional random matrices. *Electron. J. Probab.* **13** 588–628. MR2399292

[7] BRILLINGER, D. R. (2001). *Time Series*. *Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. MR1853554

[8] COSTIN, O. and LEBOWITZ, J. L. (1995). Gaussian fluctuation in random matrices. *Phys. Rev. Lett.* **75** 69–72. MR3155254

[9] JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906

[10] JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857

[11] LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

[12] PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244

[13] PILLAI, N. S. and YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042

[14] ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston, MA. MR0885090

[15] SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. MR2234197

[16] SINAI, Y. and SOSHNIKOV, A. (1998). Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Bras. Mat.* **29** 1–24. MR1620151

[17] SOSHNIKOV, A. (2002). Gaussian limit for determinantal random point fields. *Ann. Probab.* **30** 171–187. MR1894104

[18] WIGNER, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.* (2) **62** 548–564. MR0077805

[19] WU, W. B. and SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582

[20] XIAO, H. and WU, W. B. (2011). Asymptotic inference of autocovariances of stationary processes. Available at arXiv:1105.3423.

[21] ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans*. *Amer. Math*. *Soc*. **359** 5345–5363. MR2327033

C. BAO
G. PAN
DIVISION OF MATHEMATICAL SCIENCES
SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE 637371
SINGAPORE
E-MAIL: maomie2007@gmail.com
gmpan@ntu.edu.sg

L.-C. LIN
DEPARTMENT OF STATISTICS
NATIONAL CHENG KUNG UNIVERSITY
TAIWAN 70101
CHINA
E-MAIL: lcln@mail.ncku.edu.tw

W. ZHOU
DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
SINGAPORE
E-MAIL: stazw@nus.edu.sg