# Limiting spectral distribution of large dimensional Spearman's rank correlation matrices

Zeyu Wu, Cheng Wang\*

School of Mathematical Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, China.

## **Abstract**

In this paper, we study the empirical spectral distribution of Spearman's rank correlation matrices, under the assumption that the observations are independent and identically distributed random vectors and the features are correlated. We show that the limiting spectral distribution is the generalized Marčenko-Pastur law with the covariance matrix of the observation after standardized transformation. With these results, we compare several classical covariance/correlation matrices including the sample covariance matrix, Pearson's correlation matrix, Kendall's correlation matrix and Spearman's correlation matrix.

Keywords: Kendall's correlation, Limiting spectral distribution, Random matrix theory, Spearman's correlation.

2020 MSC: 62G30, 62H20.

### 1. Introduction

Statistical inference for covariance or correlation matrices are fundamental problems in high dimensional data analysis [7, 10]. Recently, rank-based correlation matrices, e.g., Kendall's tau and Spearman's rho, have drawn increasing attention in a variety of fields. Due to the robustness of non-parametric statistics, these rank-based correlations have appealing properties in theory and have the potential to solve problems for high dimensional data with complex structure such as heavy-tailed distributions [12].

In high dimensional data analysis, Liu et al. [19] and Xue and Zou [27] firstly used rank-based correlation matrices to conduct sparse estimation for the covariance matrix and the precision matrix (i.e., the inverse of the covariance matrix). For Gaussian distribution, there exists an explicit relationship between rank-based correlation and Pearson's correlation. Exploiting this neat property, Liu et al. [20] extended the Gaussian graphical model to the non-paranormal distribution. Further, there are many works on hypothesis testing of covariance/correlation matrices, which are based on Kendall's tau correlation matrix  $\tau_n$  or Spearman's rho correlation matrix  $\rho_n$ . In details, Bao et al. [6] derived the asymptotic normal distributions of  $\text{tr}(\rho_n^k)$ ,  $k \in \{2, 3, ...\}$  and conducted several statistics for testing the complete independence. Han et al. [11] studied the maximum norm of off-diagonal elements which converges weakly to Gumbel distribution and proposed test statistics based on Kendall's tau and Spearman's rho. Leung and Drton [16] considered the test statistics based on the squared Frobenius norm and see also Mao [22].

Technically, Liu et al. [19] and Xue and Zou [27] focused on the sparse estimation of Kendall's tau and Spearman's rho. To guarantee the consistency, they need to bound the maximum norm between the empirical and population rank-based correlation matrices, i.e.,  $\|\rho_n - \rho\|_{\infty}$  and  $\|\tau_n - \tau\|_{\infty}$ . To control the maximum norm, the crucial issue is to bound the tail probability of each element uniformly. The main tools are the Hoeffding decomposition for U-statistics and some classical concentration inequalities such as Hoeffding and McDiarmid inequalities. As far as the hypothesis testing, to derive the asymptotic distribution of the test statistics, we need more refined results. Under the null hypothesis of independence, all the features are independent and it is doable to derive the distribution of the test statistics. Under

Preprint submitted to Elsevier May 31, 2022

 $<sup>{\</sup>rm *Corresponding\ author.\ Email\ address:\ chengwang@sjtu.edu.cn}$ 

the alternative case, it is challenging to analyze Kendall's tau or Spearman's rho since rank variables depend on all the samples. Thus, the existing works [e.g., 6, 11, 16, 22] only derived the distribution under the null case.

In random matrix theory, the same challenges also exist. Bai and Zhou [2] and Bandeira et al. [3] studied the limiting spectral distribution of Spearman's rho and Kendall's tau, respectively. Bao et al. [6] and Li et al. [18] considered the asymptotic distribution of the test statistics involved Spearman's rho and Kendall's tau, respectively. Bao [5] and Bao [4] derived the Tracy–Widom limits for Spearman's rho and Kendall's tau, respectively. All these works are based on the independent assumption. For the data with a common covariance structure, the data matrix after ranking does not have independent columns or rows anymore and in random matrix theory, it is challenging to analyze such matrices and see Bai and Zhou [2] for more details. Recently, we derived the limiting spectral distribution (LSD) of Kendall's rank correlation matrix with dependence in Li et al. [17] and it is the first result on rank correlation matrices with dependence. In this paper, as a companion work, we study the limiting spectral distribution of Spearman's rank correlation matrices. Under mild conditions, we show that the LSD follows the generalized Marčenko-Pastur law with a conditional covariance matrix. With these results, we can look insider into Spearman's correlation matrix, and also its connections with other covariance/correlation matrices.

The rest of the paper is organized as follows. In Section 2, we revisit the Spearman's rank correlation matrix and provide some trivial results when the data is correlated. In Section 3, we formulate Spearman's correlation matrix in the form of U-statistics and we present the main results on LSD in Section 4. Finally, we provide discussions on several covariance/correlation matrices and also compare their LSDs through numerical simulations in Section 5. All technical proofs are relegated to Appendix.

## 2. Spearman's rank correlation matrix

We first introduce some necessary notation. Throughout the paper,  $\mathbf{I}_m$  is a  $m \times m$  identity matrix and  $\mathbf{1}_m$  is a  $m \times 1$  vector with all ones.  $\|\cdot\|_2$  denotes the Frobenius norm of a vector or matrix.  $\|\cdot\|$  is the spectral norm and  $\|\cdot\|_{\infty}$  is the element-wise maximum norm of a vector or matrix, respectively.

Given the independent and identically distributed (i.i.d.) observations  $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$ , we have the data matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{X}_n^{\mathsf{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = (x_{ij})_{n \times p},$$

where each row is an observation and each column is a feature. Transforming each column of the data matrix into the order statistics  $(r_{ij})_{n \times p}$  and normalizing these order statistics, we get the standardized ranking matrix,

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{R}_{\mathsf{T}}^{\mathsf{T}} \end{pmatrix} \stackrel{\text{def}}{=} \left( \sqrt{\frac{12}{n^2 - 1}} (r_{ij} - \frac{n+1}{2}) \right)_{n \times p}.$$

With the ranking matrix, Spearman [25] introduced the famous Spearman's rank correlation matrix

$$\rho_n = \frac{1}{n} \mathbf{R}^{\mathsf{T}} \mathbf{R} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{R}_i \mathbf{R}_i^{\mathsf{T}}.$$
 (1)

From the statistical point of view, inducing the order statistics makes the correlation more robust to heavy-tailed distributions. From the technical point of view, the ranking violates the independent structure among the rows of the observations. In particular, we have the following proposition.

**Proposition 1.** Assuming 
$$\mathbf{X}_1, \dots, \mathbf{X}_n$$
, i.i.d.  $\sim N(0, \mathbf{I}_p)$ , we have  $\mathbb{E}\{\mathbf{R}^\top \mathbf{R}/n\} = \mathbf{I}_p$  and  $\mathbb{E}\{\mathbf{R}\mathbf{R}^\top/p\} = (n\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^\top)/(n-1)$ .

From Proposition 1, we can see that when the raw data is with i.i.d. rows, the ranking matrix **R** does not have independent rows anymore. Since the columns of the data matrix **X** is also i.i.d., the ranking matrix **R** has i.i.d. columns and in special, each column  $(r_{1j}, \ldots, r_{nj})$  is uniformly distributed from the permutations of  $\{1, \ldots, n\}$ . Bai and Zhou [2] studied this case and they proved that the LSD of  $\rho_n$  is also the Marčenko-Pastur law.

On the other hand, although we assume  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , *i.i.d.* ~  $N(0, \mathbf{I}_p)$ , it actually covers a large class of distributions since the Spearman's correlation is rank-based. For example, when all the features are independent and continuous, we can transform each feature into a standard normal distribution where the ranking matrix is invariant. It is referred to the monotonic invariance of rank-based correlation matrices introduced by Weihs et al. [26] and the non-paranormal model proposed by Liu et al. [20]. Thus, the independent case of previous works on Spearman's correlation matrix [e.g., 2, 5, 6] can be formulated as  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , *i.i.d.* ~  $N(0, \mathbf{I}_p)$ .

Assuming that the data matrix is with i.i.d. entries is very limited. A natural extension is to consider general covariance structures. For the multivariate normal distribution with general  $\Sigma$ , we have the following result.

**Proposition 2.** Assuming  $X_1, ..., X_n$ , i.i.d.  $\sim N(\mathbf{0}, \Sigma)$ , where  $\Sigma_{ii} = 1$ ,  $i \in \{1, ..., p\}$ , we have

$$\mathbb{E}\left\{\frac{1}{n}\mathbf{R}^{\top}\mathbf{R}\right\} = \frac{3}{n+1}\boldsymbol{\Sigma}_{1} + \frac{3(n-2)}{n+1}\boldsymbol{\Sigma}_{2}, \ \mathbb{E}\left\{\frac{1}{p}\mathbf{R}\mathbf{R}^{\top}\right\} = \frac{n}{n-1}(\mathbf{I}_{n} - \frac{1}{n}\mathbf{1}_{n}\mathbf{1}_{n}^{\top}),$$

where

$$\Sigma_1 = \left(\frac{2}{\pi}\arcsin(\Sigma_{ij})\right)_{p \times p} \stackrel{\text{def}}{=} \frac{2}{\pi}\arcsin(\Sigma), \ \Sigma_2 = \left(\frac{2}{\pi}\arcsin(\Sigma_{ij}/2)\right)_{p \times p} \stackrel{\text{def}}{=} \frac{2}{\pi}\arcsin(\Sigma/2).$$

From Proposition 2, we can see that the rows and columns of the ranking matrix  $\mathbf{R}$  are both dependent if the data has a general covariance structure  $\Sigma$ . From the perspective of random matrix theory, it is challenging to analyze such matrices.

## 3. U-statistic of Spearman's correlation

For each feature  $j \in \{1, \dots, p\}$ , we invoke the empirical cumulative distribution function

$$\hat{F}_j(x) = \frac{1}{n} \sum_{i=1}^n I(x_{ij} \le x),$$

and the order statistics are

$$r_{ij} = n \cdot \hat{F}_j(x_{ij}) = 1 + \sum_{k \neq i} I(x_{kj} \le x_{ij}), \ i \in \{1, \dots, n\}, \ j \in \{1, \dots, p\}.$$

Throughout the paper, we consider the continuous population distribution and assume that there is no ties in data matrix.

Defining the sign vector

$$\mathbf{A}_{ik} = \operatorname{sign} \left( \mathbf{X}_i - \mathbf{X}_k \right) = \left( \operatorname{sign} \left( x_{i1} - x_{k1} \right), \dots, \operatorname{sign} \left( x_{ip} - x_{kp} \right) \right)^{\mathsf{T}},$$

we have

$$r_{ij} - \frac{n+1}{2} = 1 + \sum_{k \neq i} \frac{1 + \operatorname{sign}(x_{ij} - x_{kj})}{2} - \frac{n+1}{2} = \frac{1}{2} \sum_{k \neq i} \operatorname{sign}(x_{ij} - x_{kj}).$$

In vector form, we can get

$$\mathbf{R}_{i} = \sqrt{\frac{12}{n^{2} - 1}} \begin{pmatrix} r_{i1} - \frac{n+1}{2} \\ \vdots \\ r_{in} - \frac{n+1}{2} \end{pmatrix} = \sqrt{\frac{3}{n^{2} - 1}} \sum_{k \neq i} \mathbf{A}_{ik}.$$
 (2)

With these notations, we can rewrite the Spearman's rank correlation matrix as follows

$$\rho_n = \frac{3}{n(n^2 - 1)} \sum_{i=1}^n \left\{ \sum_{k_1, k_2 \neq i} \mathbf{A}_{ik_1} \mathbf{A}_{ik_2}^\top \right\}.$$
 (3)

Following the definition of U-statistics, we can divide the summation into two parts and each part is a U-statistic.

**Proposition 3.** Assuming the observations  $X_1, ..., X_n$  have no ties, the Spearman's correlation matrix can be defined by

$$\rho_n = \frac{3}{n(n^2 - 1)} \sum_{i,j}^* \mathbf{A}_{ij} \mathbf{A}_{ij}^{\mathsf{T}} + \frac{3}{n(n^2 - 1)} \sum_{i,j,k}^* \mathbf{A}_{ij} \mathbf{A}_{ik}^{\mathsf{T}}, \tag{4}$$

where  $\Sigma^*$  denotes summation over mutually different indices.

When  $X_1, ..., X_n$  are i.i.d. from a population with absolutely continuous densities, by Proposition 3, we can see

$$\mathbb{E}(\boldsymbol{\rho}_n) = \frac{3}{n(n^2 - 1)} n(n - 1) \cdot \cos(\mathbf{A}_{12}) + \frac{3}{n(n^2 - 1)} n(n - 1)(n - 2) \cdot \cos(\mathbf{A}_1) = \frac{3}{n + 1} \cos(\mathbf{A}_{12}) + \frac{3(n - 2)}{n + 1} \cos(\mathbf{A}_1),$$

where

$$\mathbf{A}_{i} \stackrel{\text{def}}{=} \mathbb{E}\{\mathbf{A}_{ik} \mid \mathbf{X}_{i}\}. \tag{5}$$

The population Spearman's correlation matrix consists of two covariance matrices  $cov(\mathbf{A}_{12})$  and  $cov(\mathbf{A}_{1})$ . For the classical sample covariance matrix

$$\mathbf{S}_n = \frac{1}{2n(n-1)} \sum_{i,j}^* (\mathbf{X}_i - \mathbf{X}_j) (\mathbf{X}_i - \mathbf{X}_j)^{\mathsf{T}},$$

we know  $cov(\mathbf{X}_i - \mathbf{X}_j) = 2cov(\mathbf{X}_i)$ . Thus, the sign function introduces a non-linear correlation into rank-based correlation matrices and these two covariance matrices are different, e.g.,

$$cov\left(\mathbf{A}_{12}\right) = \mathbf{\Sigma}_1 = \frac{2}{\pi}\arcsin(\mathbf{\Sigma}), \text{ and } cov\left(\mathbf{A}_1\right) = \mathbf{\Sigma}_2 = \frac{2}{\pi}\arcsin(\mathbf{\Sigma}/2),$$

for the multivariate normal distribution  $N(\mathbf{0}, \mathbf{\Sigma})$ .

Since both  $cov(\mathbf{A}_{12})$  and  $cov(\mathbf{A}_{1})$  can describe correlations among features, an intuitive way is to conduct U-statistics for these two covariance matrices separately, e.g.,

$$\boldsymbol{\tau}_n = \frac{1}{n(n-1)} \sum_{i,j}^* \mathbf{A}_{ij} \mathbf{A}_{ij}^\top, \tag{6}$$

and

$$\tilde{\boldsymbol{\rho}}_n = \frac{3}{n(n-1)(n-2)} \sum_{i,j,k}^* \mathbf{A}_{ij} \mathbf{A}_{ik}^{\mathsf{T}}.$$
 (7)

Interestingly,  $\tau_n$  is exactly Kendall's rank correlation matrix proposed by [15] and  $\tilde{\rho}_n$  is an improved Spearman's rank correlation matrix proposed by [13]. See also Example 3 of [11]. Our recent work [17] studied the LSD of  $\tau_n$  and in this work, we consider Spearman's correlation matrices  $\rho_n$  and  $\tilde{\rho}_n$ .

## 4. Limiting spectral distribution

For an  $n \times n$  Hermitian matrix  $\mathbf{H}_n$  whose eigenvalues are  $\lambda_1, \dots, \lambda_n$ , the empirical spectral distribution of  $\mathbf{H}_n$  is defined as

$$F^{\mathbf{H}_n}(x) = \frac{1}{n} \sum_{i=1}^n I(\lambda_i \le x).$$

The limit of  $F^{\mathbf{H}_n}$  is called the limiting spectral distribution of  $\mathbf{H}_n$ . In random matrix theory, the LSD is usually defined by its Stieltjes transform

$$s_F(z) = \int \frac{1}{x-z} dF(x), \ z \in \mathbb{C}^+,$$

where  $\mathbb{C}^+$  denotes the upper complex plane. With a Stieltjes transform s(z), the distribution function can be obtained by the inversion formula

$$F(b) - F(a) = \lim_{\nu \to 0^+} \frac{1}{\pi} \int_a^b \Im s_F(x + i\nu) dx,$$

where  $\mathfrak{I}(\cdot)$  is the imaginary part and i is the imaginary unit.

Note

$$\tilde{\boldsymbol{\rho}}_n = \frac{3}{n} \sum_{i=1}^n \left\{ \frac{1}{(n-1)(n-2)} \sum_{j,k\neq i}^* \mathbf{A}_{ij} \mathbf{A}_{ik}^\top \right\}$$

and for a given i,

$$\frac{1}{(n-1)(n-2)} \sum_{i,k\neq i}^* \mathbf{A}_{ij} \mathbf{A}_{ik}^{\mathsf{T}}$$

is also a U-statistic for  $\mathbf{A}_i \mathbf{A}_i^{\mathsf{T}}$ . Intuitively, the improved Spearman's rank correlation matrix  $\tilde{\rho}_n$  is close to the random matrix

$$\mathbf{W}_n = \frac{3}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{A}_i^{\mathsf{T}}.$$
 (8)

The following result shows that  $\tilde{\rho}_n$  and  $\mathbf{W}_n$  share the same limiting spectral distribution.

**Theorem 1.** Assuming  $X_1, ..., X_n$ , i.i.d.  $\sim N(\mathbf{0}, \Sigma)$  where  $\Sigma_{ii} = 1$ ,  $i \in \{1, ..., p\}$ , and if

- (i)  $p/n \to y \in (0, \infty)$ ;
- (ii)  $tr(\Sigma^2) = o(p^2);$

we have

$$L(F^{\tilde{\rho}_n}, F^{\mathbf{W}_n}) \to 0$$
, in probability, (9)

where  $L(\cdot, \cdot)$  is the Levy distance between two distribution functions.

For Spearman's rank correlation matrix  $\rho_n$ , by Proposition 3, we know

$$\boldsymbol{\rho}_n = \frac{3}{n+1}\boldsymbol{\tau}_n + \frac{n-2}{n+1}\tilde{\boldsymbol{\rho}}_n = \frac{3}{n+1}\boldsymbol{\tau}_n - \frac{3}{n+1}\tilde{\boldsymbol{\rho}}_n + \tilde{\boldsymbol{\rho}}_n,$$

where  $\tau_n$  is the Kendall's rank correlation matrix (6). We claim that  $\rho_n$  and  $W_n$  also share the same limiting spectral distribution. The following result establishes this claim.

Proposition 4 (Weak convergence). Under the assumptions of Theorem 1, we have

$$L(F^{\rho_n}, F^{\mathbf{W}_n}) \to 0$$
, in probability. (10)

The proofs of Theorem 1 and Proposition 4 are based on Corollary A.41 of Bai and Silverstein [1] and we need to show  $\mathbb{E} \|\tilde{\boldsymbol{\rho}}_n - \mathbf{W}_n\|_2^2 = o(p)$  and  $\mathbb{E} \|\boldsymbol{\rho}_n - \mathbf{W}_n\|_2^2 = o(p)$ . To further refine the conclusion, e.g., to show the strong convergence, we need to prove  $\|\tilde{\boldsymbol{\rho}}_n - \mathbf{W}_n\|_2^2/p \to 0$ , a.s. which is challenging since the calculation of higher order moments of  $\mathbf{A}_i$  is hard. Here, we turn to the ranking statistics  $\mathbf{R}$ . Intuitively,

$$\mathbf{R} = \left(\sqrt{\frac{12}{n^2 - 1}}(r_{ij} - \frac{n+1}{2})\right)_{n \times p} = \sqrt{\frac{12n^2}{n^2 - 1}} \left(\hat{F}_j(x_{ij}) - \frac{n+1}{2n}\right)_{n \times p} \approx 2\sqrt{3} \left(F_j(x_{ij}) - \frac{1}{2}\right)_{n \times p} = \sqrt{3}(\mathbf{A}_1, \dots, \mathbf{A}_n)^{\top}.$$

Thus, we can control the difference between  $F^{\rho_n}$  and  $F^{\mathbf{W}_n}$  by bounding  $\mathbf{R} - \sqrt{3}(\mathbf{A}_1, \dots, \mathbf{A}_n)^{\mathsf{T}}$ . Specifically, we will use Corollary A.42 of Bai and Silverstein [1], i.e.,

$$L^{4}\left(F^{\boldsymbol{\rho}_{n}}, F^{\mathbf{W}_{n}}\right) \leq \frac{2\operatorname{tr}\left(\boldsymbol{\rho}_{n} + \mathbf{W}_{n}\right)}{np^{2}} \|\mathbf{R} - \sqrt{3}(\mathbf{A}_{1}, \dots, \mathbf{A}_{n})^{\mathsf{T}}\|_{2}^{2}$$

and then we can show the strong convergence as follows.

**Theorem 2** (Strong convergence). Assuming  $X_1, \ldots, X_n$  are i.i.d. continuous random vectors and  $p/n \to y \in (0, \infty)$ , we have

$$L(F^{\rho_n}, F^{\mathbf{W}_n}) \to 0$$
, almost surely. (11)

It is noted that Theorem 2 provides a stronger conclusion with a weaker condition and the key technical tool is the Dvoretzky–Kiefer–Wolfowitz inequality for empirical cumulative distribution functions. By Proposition 4 or Theorem 2, to study the LSD of  $\rho_n$ , we can consider the matrix  $\mathbf{W}_n$  which is the sample covariance matrix of the random vectors  $\sqrt{3}(\mathbf{A}_1,\ldots,\mathbf{A}_n)$ . It is noted that  $\mathbf{A}_1,\ldots,\mathbf{A}_n$  are i.i.d., and the LSD can be derived from classical results on sample covariance matrices, e.g., Theorem 1 of Bai and Zhou [2]. Assuming  $\mathbf{X}_1,\ldots,\mathbf{X}_n$ , i.i.d.  $\sim N(\mathbf{0},\mathbf{\Sigma})$  where  $\Sigma_{ii}=1$ , we have

$$\mathbf{A}_i = \mathbb{E}\{\mathbf{A}_{ij} \mid \mathbf{X}_i\} = \begin{pmatrix} 2\Phi(x_{i1}) - 1 \\ \vdots \\ 2\Phi(x_{ip}) - 1 \end{pmatrix}.$$

Here  $\Phi(\cdot)$  is the cumulative distribution function of N(0, 1). Therefore, each entry of  $(\mathbf{A}_1, \dots, \mathbf{A}_n)$  follows the uniform distribution U[-1, 1] and

$$cov(\mathbf{A}_i) = \mathbf{\Sigma}_2 = \frac{2}{\pi} arcsin(\mathbf{\Sigma}/2).$$

Based on the main result of Bai and Zhou [2], we can derive the LSD of Spearman's rank correlation matrix as follows.

**Theorem 3.** Assume  $X_1, \ldots, X_n$ , i.i.d.  $\sim N(\mathbf{0}, \Sigma)$  where  $\Sigma_{ii} = 1$ ,  $i \in \{1, \ldots, p\}$ , and

- (i)  $p/n \to y \in (0, \infty)$  as  $n \to \infty$ ;
- (ii) the spectral norm of  $\Sigma$  is uniformly bounded by a constant C:
- (iii) the empirical spectral distribution of 6  $\arcsin(\Sigma/2)/\pi$  tends to a non-random probability distribution H.

Then, with probability 1,  $F^{\rho_n}$  tends to a probability distribution, whose Stieltjes transform m = m(z),  $z \in \mathbb{C}^+$  satisfies

$$m = \int \frac{1}{t(1 - y - yzm) - z} dH(t). \tag{12}$$

Noting the Spearman's correlation matrix is rank-based, Theorem 3 actually holds for the non-paranormal distribution proposed by Liu et al. [20]. In particular, for a random vector  $\mathbf{Z} = (z_1, \dots, z_p)^{\mathsf{T}} \in \mathbb{R}^p$ , there exist monotone

functions  $\{f_j\}_{j=1}^p$  such that  $(f_1(z_1), \ldots, f_p(z_p)) \sim N(\mathbf{0}, \mathbf{\Sigma})$  where  $\Sigma_{ii} = 1, i \in \{1, \ldots, p\}$ . Denoting the cumulative distribution function of  $z_j$  as  $F_j$ , Liu et al. [20] show that  $f_j(t) = \Phi^{-1}(F_j(t))$  and the non-paranormal distribution model actually assumes

$$\left(\Phi^{-1}(F_1(z_1)),\ldots,\Phi^{-1}(F_p(z_p))\right) \sim N(\mathbf{0},\mathbf{\Sigma}).$$

We can see that the non-paranormal distribution extends the multivariate normal distribution and more discussions can be found in Lu et al. [21]. More generally, we can further extend the assumption to general distributions with some moment conditions and it is referred to Li et al. [17] for more details. To illustrate the robustness of Spearman's rank correlation, we consider a toy example where each element follows the Cauchy distribution.

## **Example 1.** For the data matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{X}_n^{\mathsf{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = (x_{ij})_{n \times p},$$

where  $x_{ij}$  follows the standard Cauchy distribution, we consider the monotone transformation

$$x_{ij} \mapsto \Phi^{-1}\left(\frac{1}{2} + \frac{1}{\pi}\arctan(x_{ij})\right) \stackrel{\text{def}}{=} y_{ij}.$$

As we know, the transformation does not change the ranking matrix **R** and to analyze **R**, we can make assumptions on  $y_{ij}$  without loss of generality. For the independent case, i.e.,  $x_{i1}, \ldots, x_{ip}$  are independent, we have  $(y_{i1}, \ldots, y_{ip})^{\top} \sim N(0, \mathbf{I})$ . For the dependent case, we need further assumptions to define the correlation structure. Noting  $y_{ij} \sim N(0, 1)$ , a natural way is to assume

$$(y_{i1},\ldots,y_{ip}) \sim N(\mathbf{0},\mathbf{\Sigma}) \tag{13}$$

where  $\Sigma_{ii} = 1$ ,  $i \in \{1, ..., p\}$ . This is exactly the non-paranormal distribution and  $\Sigma$  is called the latent generalized correlation matrix by Lu et al. [21]. In summary, we have the relationships

Spearman's correlation  $(x_{ij}, x_{ik})$  = Spearman's correlation  $(y_{ij}, y_{ik}) = \frac{2}{\pi} \arcsin(\text{Pearson's correlation } (y_{ij}, y_{ik}))$ ,

where the second equality is due to the non-paranormal distribution assumption (13).

# 5. Discussions and simulations

In this section, we compare several important covariance/correlation matrices. Given the data  $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$ , there are four classical covariance/correlation matrices in statistical applications:

• Sample covariance matrix:

$$\mathbf{S}_{n} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_{i} - \bar{\mathbf{X}}) (\mathbf{X}_{i} - \bar{\mathbf{X}})^{\top} = \frac{1}{2n(n-1)} \sum_{i,j}^{*} (\mathbf{X}_{i} - \mathbf{X}_{j}) (\mathbf{X}_{i} - \mathbf{X}_{j})^{\top} = \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{X}_{i} \mathbf{X}_{i}^{\top} - \frac{n}{n-1} \bar{\mathbf{X}} \bar{\mathbf{X}}^{\top},$$

where  $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_{i}$ ;

• Pearson's correlation matrix [24]:

$$\mathbf{P}_n = \{\operatorname{diag}(\mathbf{S}_n)\}^{-1/2} \mathbf{S}_n \{\operatorname{diag}(\mathbf{S}_n)\}^{-1/2},$$

where  $diag(\cdot)$  is a diagonal matrix with the diagonal entries of the matrix;

• Kendall's correlation matrix [15]:

$$\boldsymbol{\tau}_n = \frac{1}{n(n-1)} \sum_{i,j}^* \mathbf{A}_{ij} \mathbf{A}_{ij}^{\mathsf{T}},$$

where  $\mathbf{A}_{ij} = \operatorname{sign}(\mathbf{X}_i - \mathbf{X}_j)$ ;

• Spearman's correlation matrix [25]

$$\boldsymbol{\rho}_n = \frac{3}{n(n^2 - 1)} \sum_{i,i}^* \mathbf{A}_{ij} \mathbf{A}_{ij}^{\mathsf{T}} + \frac{3}{n(n^2 - 1)} \sum_{i,i,k}^* \mathbf{A}_{ij} \mathbf{A}_{ik}^{\mathsf{T}},$$

and the improved Spearman's correlation matrix [13]

$$\tilde{\boldsymbol{\rho}}_n = \frac{3}{n(n-1)(n-2)} \sum_{i,j,k}^* \mathbf{A}_{ij} \mathbf{A}_{ik}^{\mathsf{T}}.$$

For comparison purposes, we assume that  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , *i.i.d.* ~  $N(\mathbf{0}, \mathbf{\Sigma})$ . Since  $\mathbf{P}_n, \tau_n$  and  $\rho_n, \tilde{\rho}_n$  are all correlation matrices which are invariant to the scale of features, we assume  $\Sigma_{ii} = 1$  to make a fair comparison on the sample covariance matrix.

When  $\Sigma = I$ , Marčenko and Pastur [23] showed that the LSD of  $S_n$  is the standard Marčenko-Pastur law which has the density function

$$p(x) = \frac{1}{2\pi xy} \sqrt{(x_+ - x)(x - x_-)} I(x_- \le x \le x_+),$$

for  $y \le 1$  and has a point mass 1-1/y at the origin if y > 1. Here  $p/n \to y \in (0, \infty)$ ,  $x_+ = (1+\sqrt{y})^2$  and  $x_- = (1-\sqrt{y})^2$ . Jiang [14] proved that the LSD of  $\mathbf{P}_n$  is also the Marčenko-Pastur law. Bai and Zhou [2] established the Marčenko-Pastur law for the Spearman's rank correlation matrix. Bandeira et al. [3] proved that the LSD of Kendall's correlation matrix is an affine transformation of the Marčenko-Pastur law, i.e.,

$$\frac{2}{3}Y + \frac{1}{3},$$

where *Y* follows the Marčenko-Pastur law. Fig 1 shows the theoretical LSDs and the empirical distributions based on 100 replications for these four matrices.

For general  $\Sigma$ , Marčenko and Pastur [23] derived the LSD of the sample covariance matrix whose Stieltjes transform m is given by the Marčenko-Pastur equation

$$m = \int \frac{1}{t(1 - y - yzm) - z} dH(t), \ H(t) = \lim_{p \to \infty} F^{\Sigma}(t).$$

This distribution is called the generalized Marčenko-Pastur law. El Karoui [8] proved that the Marčenko-Pastur equation also holds for the Pearson's correlation matrix. For the Spearman's correlation matrix, our Theorem 1 shows that

$$\boldsymbol{\rho}_n \stackrel{\text{LSD}}{=} \tilde{\boldsymbol{\rho}}_n \stackrel{\text{LSD}}{=} \mathbf{W}_n = \frac{3}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{A}_i^{\mathsf{T}},$$

where  $\stackrel{\text{LSD}}{=}$  denotes the two matrices share the same LSD. Our Theorem 3 further proves that the LSD is also the generalized Marčenko-Pastur law whose Stieltjes transform m is

$$m = \int \frac{1}{t(1-y-yzm)-z} dH(t), \ H(t) = \lim_{p \to \infty} F^{3\Sigma_2}(t).$$

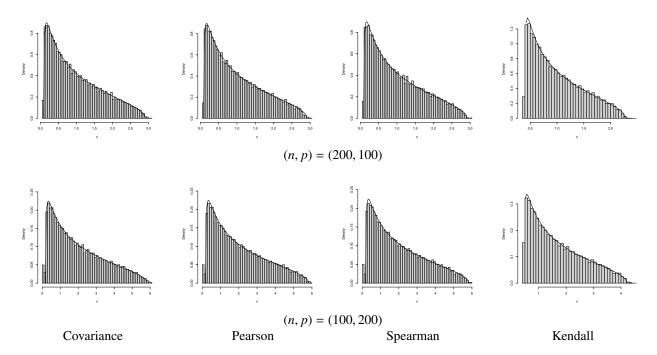


Fig. 1: The limiting spectral distributions of sample covariance matrix, Pearson's correlation matrix, Spearman's correlation matrix and Kendall's correlation matrix where  $X_1, \ldots, X_n$ , *i.i.d.*  $\sim N(0, \mathbf{I})$ .

For the Kendall's correlation matrix, Li et al. [17] established that

$$\boldsymbol{\tau}_n \stackrel{\text{\tiny LSD}}{=\!\!\!=} \frac{2}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{A}_i^{\top} + \boldsymbol{\Sigma}_3,$$

and they presented the LSD in two equations where the LSD is not the generalized Marčenko-Pastur law anymore. Here

$$\Sigma_1 = \frac{2}{\pi} \arcsin(\Sigma), \ \Sigma_2 = \frac{2}{\pi} \arcsin(\Sigma/2), \ \Sigma_3 = \Sigma_1 - 2\Sigma_2.$$

In particular, we consider a specific tridiagonal covariance matrix

$$\Sigma(\rho) = \begin{pmatrix} 1 & \rho & & & \\ \rho & 1 & \rho & & \\ & \ddots & \ddots & \ddots & \\ & & \rho & 1 & \rho \\ & & & \rho & 1 \end{pmatrix},$$

where  $\rho \neq 0$  and  $|\rho| \leq 1/2$ . By Szegö Theorem, we have

$$\lim_{p\to\infty} F^{\Sigma}(t) = 1 - \frac{1}{\pi} \arccos\frac{t-1}{2\rho}, \ |t-1| \le 2|\rho|,$$

and

$$\lim_{p \to \infty} F^{3\Sigma_2}(t) = 1 - \frac{1}{\pi} \arccos \frac{t-1}{2\rho_1}, \ |t-1| \le 2|\rho_1|, \ \rho_1 = \frac{6}{\pi} \arcsin \frac{\rho}{2}.$$

Therefore, the Marčenko-Pastur equation of the sample covariance matrix and the Pearson's correlation matrix is

$$m = \int_{1-2|\rho|}^{1+2|\rho|} \frac{1}{t(1-y-yzm)-z} d\left(1-\frac{1}{\pi}\arccos\frac{t-1}{2\rho}\right) = \frac{1}{\sqrt{(1-y-yzm-z)^2-4\rho^2(1-y-yzm)^2}},$$

and similarly, the Marčenko-Pastur equation of the Spearman's correlation matrix is

$$m = \frac{1}{\sqrt{(1 - y - yzm - z)^2 - 4\rho_1^2(1 - y - yzm)^2}}.$$

Solving the equation and using the inversion formula of the Stiejtjes transformation, we can get the LSDs. The LSD of Kendall's correlation matrix can be found in the Proposition 4.2 of Li et al. [17]. Fig 2 shows the theoretical LSDs and the empirical distributions based on 100 replications for  $\Sigma = \Sigma(0.5)$ .

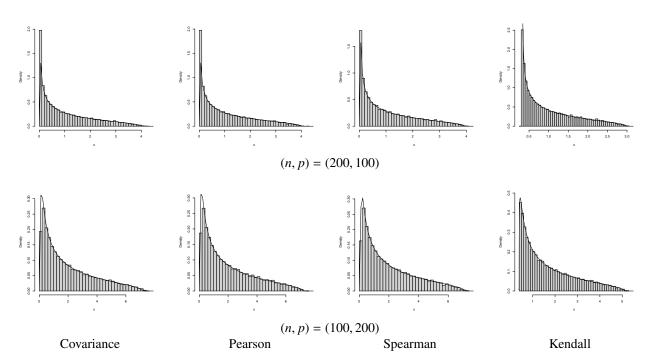


Fig. 2: The limiting spectral distributions of sample covariance matrix, Pearson's correlation matrix, Spearman's correlation matrix and Kendall's correlation matrix where  $\mathbf{X}_1, \ldots, \mathbf{X}_n$ , *i.i.d.*  $\sim N(0, \mathbf{\Sigma}(0.5))$ .

In summary, under mild conditions, the Pearson's correlation matrix shares the same properties of the sample covariance matrix. For rank-based correlation matrix, the Spearman's rho has the generalized Marčenko-Pastur law with the population covariance  $6/\pi \arcsin(\Sigma/2)$  and the Kendall's tau is equivalent to a sample covariance matrix with the population covariance matrix  $\Sigma_2$  plus a deterministic covariance matrix  $\Sigma_3$ .

## Acknowledgments

We thank the Editor, an Associate Editor, and anonymous reviewers for their insightful comments. Wang's research is supported by National Natural Science Foundation of China (11971017, 12031005) and NSF of Shanghai (21ZR1432900).

## 6. Appendix

We first collect several important lemmas.

**Lemma 1** (Grothendieck's identity). *Consider a bi-variate normal distribution:* 

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\},$$

where  $\rho \in [-1, 1]$ . Then,  $\mathbb{E}\{sign(z_1)sign(z_2)\} = 4\mathbb{E}\{I(z_1, z_2 > 0)\} - 1 = 2\arcsin\rho/\pi$ .

Lemma 2 (Esscher [9]). Consider a multivariate normal distribution

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1/2 & \rho & \rho/2 \\ 1/2 & 1 & \rho/2 & \rho \\ \rho & \rho/2 & 1 & 1/2 \\ \rho/2 & \rho & 1/2 & 1 \end{pmatrix}$$

where  $\rho \in (-1, 1)$ . Then,

$$\mathbb{E} \prod_{j=1}^{4} sign(z_j) = (\frac{2}{\pi} \arcsin \rho)^2 - (\frac{2}{\pi} \arcsin \rho/2)^2 + \frac{1}{9}.$$
 (14)

**Lemma 3.** For any  $x \in [0, 1]$ ,  $2\arcsin(x/2) \le \arcsin(x/2) \le \arcsin(x/2)$  and  $2x/\pi \le 2\arcsin(x)/\pi \le x$ .

**Lemma 4.** Assuming  $X_1, \ldots, X_n$ , i.i.d.  $\sim N(0, \Sigma)$ , where  $\Sigma_{ii} = 1$ , we have  $\mathbb{E}(A_{ij}) = \mathbb{E}(A_i) = 0$  and

$$cov(\mathbf{A}_{ij}) = \mathbf{\Sigma}_1 = \frac{2}{\pi} \arcsin(\mathbf{\Sigma}), \ cov(\mathbf{A}_i) = \mathbf{\Sigma}_2 = \frac{2}{\pi} \arcsin(\mathbf{\Sigma}/2).$$

**Lemma 5.** Assume  $X_1, ..., X_n$ , i.i.d. ~  $N(0, \Sigma)$ , where  $\Sigma_{ii} = 1$ , we have

$$var(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{13}) \ge var(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{1}) \ge var(\mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1})$$

and

$$var(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{13}) = tr(\mathbf{\Sigma}_{1}^{2}) - tr(\mathbf{\Sigma}_{2}^{2}).$$
 (15)

**Proof of Lemma 5**. Firstly, it is easy to see  $\mathbb{E}(\mathbf{A}_{12}\mathbf{A}_{13}^{\mathsf{T}}) = \mathbb{E}(\mathbf{A}_{12}\mathbf{A}_{1}^{\mathsf{T}}) = \mathbb{E}(\mathbf{A}_{1}\mathbf{A}_{1}^{\mathsf{T}}) = \mathbf{\Sigma}_{2}$ , which yields  $\mathbb{E}(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{13}) = \mathbb{E}(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{13}) = \mathbb{E$ 

$$\begin{aligned} \operatorname{var}(\mathbf{A}_{12}^{\top}\mathbf{A}_{13}) &= \mathbb{E}\{\mathbf{A}_{12}^{\top}\mathbf{A}_{13} - \operatorname{tr}(\boldsymbol{\Sigma}_{2})\}^{2} = \mathbb{E}\{\mathbf{A}_{12}^{\top}\mathbf{A}_{13} - \mathbf{A}_{12}^{\top}\mathbf{A}_{1} + \mathbf{A}_{12}^{\top}\mathbf{A}_{1} - \operatorname{tr}(\boldsymbol{\Sigma}_{2})\}^{2} \\ &= \mathbb{E}(\mathbf{A}_{12}^{\top}\mathbf{A}_{13} - \mathbf{A}_{12}^{\top}\mathbf{A}_{1})^{2} + \operatorname{var}(\mathbf{A}_{12}^{\top}\mathbf{A}_{1}) + 2\mathbb{E}(\mathbf{A}_{12}^{\top}\mathbf{A}_{13} - \mathbf{A}_{12}^{\top}\mathbf{A}_{1})(\mathbf{A}_{12}^{\top}\mathbf{A}_{1} - \operatorname{tr}(\boldsymbol{\Sigma}_{2})) \\ &= \mathbb{E}(\mathbf{A}_{12}^{\top}\mathbf{A}_{13} - \mathbf{A}_{12}^{\top}\mathbf{A}_{1})^{2} + \operatorname{var}(\mathbf{A}_{12}^{\top}\mathbf{A}_{1}) \geq \operatorname{var}(\mathbf{A}_{12}^{\top}\mathbf{A}_{1}), \end{aligned}$$

and similarly

$$\begin{aligned} \operatorname{var}(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{1}) &= \mathbb{E}(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{1} - \mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1})^{2} + \operatorname{var}(\mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1}) + 2\mathbb{E}(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{1} - \mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1})(\mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1} - \operatorname{tr}(\boldsymbol{\Sigma}_{2})) \\ &= \mathbb{E}(\mathbf{A}_{12}^{\mathsf{T}}\mathbf{A}_{1} - \mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1})^{2} + \operatorname{var}(\mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1}) \geq \operatorname{var}(\mathbf{A}_{1}^{\mathsf{T}}\mathbf{A}_{1}). \end{aligned}$$

The calculation of  $var(\mathbf{A}_{12}^{\top}\mathbf{A}_{13})$  can be found in the proof of Lemma A 3.2 of Li et al. [17]. The proof of Lemma 5 is completed.

**Proof of Proposition 2**. Here we prove Proposition 2 and Proposition 1 is a special case of Proposition 2 by setting  $\Sigma = I$ .

Noting  $\mathbf{R}^{\top} = (\mathbf{R}_1, \dots, \mathbf{R}_n)$ , we have

$$\mathbb{E}\{\frac{1}{n}\mathbf{R}^{\top}\mathbf{R}\} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{R}_{i}\mathbf{R}_{i}^{\top} = \mathbb{E}\{\mathbf{R}_{1}\mathbf{R}_{1}^{\top}\}, \ \mathbb{E}\{\frac{1}{p}\mathbf{R}\mathbf{R}^{\top}\} = \frac{1}{p}\mathbb{E}\left(\mathbf{R}_{i}^{\top}\mathbf{R}_{j}\right)_{n\times n}.$$

By the formula (2),

$$\mathbf{R}_i = \sqrt{\frac{3}{n^2 - 1}} \sum_{k \neq i} \mathbf{A}_{ik},$$

and we have

$$\mathbb{E}(\mathbf{R}_{1}\mathbf{R}_{1}^{\mathsf{T}}) = \frac{3}{n^{2} - 1} \sum_{l,m=2}^{n} \mathbb{E}\{\mathbf{A}_{1l}\mathbf{A}_{1m}^{\mathsf{T}}\} = \frac{3}{n^{2} - 1}\{(n - 1)\operatorname{cov}(\mathbf{A}_{12}) + (n - 1)(n - 2)\operatorname{cov}(\mathbf{A}_{1})\}$$
$$= \frac{3}{n + 1} (\mathbf{\Sigma}_{1} + (n - 2)\mathbf{\Sigma}_{2}).$$

Next, we calculate  $\mathbb{E}(\mathbf{R}_i^{\mathsf{T}}\mathbf{R}_j)$ . When j = i,

$$\mathbb{E}(\mathbf{R}_i^{\mathsf{T}}\mathbf{R}_i) = \operatorname{tr}\left\{\mathbb{E}(\mathbf{R}_1\mathbf{R}_1^{\mathsf{T}})\right\} = \frac{3}{n+1}\left(\operatorname{tr}(\mathbf{\Sigma}_1) + (n-2)\operatorname{tr}(\mathbf{\Sigma}_2)\right) = \frac{3}{n+1}\left(p + \frac{n-2}{3}p\right) = p.$$

For  $i \neq j$ ,

$$\mathbb{E}(\mathbf{R}_{i}^{\mathsf{T}}\mathbf{R}_{j}) = \frac{3}{n^{2} - 1} \mathbb{E}\left(\sum_{l \neq i} \mathbf{A}_{il}\right)^{\mathsf{T}} \left(\sum_{m \neq j} \mathbf{A}_{jm}\right) = \frac{3}{n^{2} - 1} \mathbb{E}\sum_{l \neq i, m \neq j} \mathbf{A}_{il}^{\mathsf{T}} \mathbf{A}_{jm}$$

$$= \frac{3}{n^{2} - 1} \left(\mathbb{E}\sum_{m \neq j} \mathbf{A}_{ij}^{\mathsf{T}} \mathbf{A}_{jm} + \mathbb{E}\sum_{m \neq i, j} \mathbf{A}_{im}^{\mathsf{T}} \mathbf{A}_{jm} + \mathbb{E}\sum_{l \neq i, l \neq j} \mathbf{A}_{il}^{\mathsf{T}} \mathbf{A}_{ji}\right)$$

$$= \frac{3}{n^{2} - 1} \left(-(n - 2) \operatorname{tr}(\mathbf{\Sigma}_{2}) - \operatorname{tr}(\mathbf{\Sigma}_{1}) + (n - 2) \operatorname{tr}(\mathbf{\Sigma}_{2}) - (n - 2) \operatorname{tr}(\mathbf{\Sigma}_{2})\right) = -\frac{3}{n^{2} - 1} \left(p + \frac{n - 2}{3}p\right) = -\frac{p}{n - 1}.$$

Thus

$$\mathbb{E}\{\frac{1}{p}\mathbf{R}\mathbf{R}^{\top}\} = \frac{n}{n-1}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\top}).$$

The proof is completed.

Proof of Theorem 1. Recall that

$$\tilde{\boldsymbol{\rho}}_n - \mathbf{W}_n = \frac{3}{n(n-1)(n-2)} \sum_{i,i,k}^* \left( \mathbf{A}_{ij} \mathbf{A}_{ik}^{\mathsf{T}} - \mathbf{A}_i \mathbf{A}_i^{\mathsf{T}} \right).$$

For i, j, k, defining the kernel function  $\phi(i, j, k) = \mathbf{A}_{ij} \mathbf{A}_{ik}^{\mathsf{T}} - \mathbf{A}_i \mathbf{A}_i^{\mathsf{T}}$  and the symmetric kernel function

$$\psi(i, j, k) = \frac{1}{6} \sum_{(i', j', k') = \pi(i, j, k)} \phi(i', j', k'),$$

we have

$$\tilde{\boldsymbol{\rho}}_n - \mathbf{W}_n = \frac{18}{n(n-1)(n-2)} \sum_{i < j < k} \psi(i, j, k).$$

By the symmetric properties of U-statistics, we can get

$$\frac{1}{9}\mathbb{E}\{(\tilde{\boldsymbol{\rho}}_{n} - \mathbf{W}_{n})^{\top} (\tilde{\boldsymbol{\rho}}_{n} - \mathbf{W}_{n})\} = \frac{6}{n(n-1)(n-2)} \sum_{i < j < k} \mathbb{E}\{\psi(1,2,3)^{\top} \psi(i,j,k)\} 
= \frac{6}{n(n-1)(n-2)} \mathbb{E}\{\psi(1,2,3)^{\top} \psi(1,2,3)\} + \frac{18(n-3)}{n(n-1)(n-2)} \mathbb{E}\{\psi(1,2,3)^{\top} \psi(1,2,4)\} 
+ \frac{18(n-3)(n-4)}{2n(n-1)(n-2)} \mathbb{E}\{\psi(1,2,3)^{\top} \psi(1,4,5)\}.$$
(16)

Next, we bound these terms. For the first term, we have

$$\mathbb{E} \|\phi(1,2,3)\|_2^2 = \mathbb{E} \text{tr} \left(\mathbf{A}_{12} \mathbf{A}_{13}^\top - \mathbf{A}_1 \mathbf{A}_1^\top \right) \left(\mathbf{A}_{13} \mathbf{A}_{12}^\top - \mathbf{A}_1 \mathbf{A}_1^\top \right) = \mathbb{E} \{ (\mathbf{A}_{12}^\top \mathbf{A}_{12}) (\mathbf{A}_{13}^\top \mathbf{A}_{13}) \} - \mathbb{E} (\mathbf{A}_1^\top \mathbf{A}_1)^2 \le p^2$$

and then

$$\mathbb{E}\|\psi(1,2,3)\|_{2}^{2} = \mathbb{E}\left\|\frac{1}{6}\sum_{(i,j,k)=\pi(1,2,3)}\phi(i,j,k)\right\|_{2}^{2} \leq \frac{1}{6}\sum_{(i,j,k)=\pi(1,2,3)}\mathbb{E}\left\|\phi(i,j,k)\right\|_{2}^{2} = \mathbb{E}\left\|\phi(1,2,3)\right\|_{2}^{2} \leq p^{2}.$$
 (17)

For the second term,

$$\mathbb{E}\{\psi(1,2,3)^{\top}\psi(1,2,4)\} \le \left\{ \mathbb{E}\|\psi(1,2,3)\|_{2}^{2} \right\}^{1/2} \left\{ \mathbb{E}\|\psi(1,2,4)\|_{2}^{2} \right\}^{1/2} = \mathbb{E}\|\psi(1,2,3)\|_{2}^{2} \le p^{2}.$$
 (18)

For the third term, we have the conditional expectation

$$\mathbb{E}\{\phi(1,2,3) \mid \mathbf{X}_1\} = \mathbb{E}\{\phi(1,3,2) \mid \mathbf{X}_1\} = \mathbf{0}, \ \mathbb{E}\{\phi(2,1,3) \mid \mathbf{X}_1\} = \mathbb{E}\{\phi(3,1,2) \mid \mathbf{X}_1\} = \mathbb{E}\{\mathbf{A}_{21}\mathbf{A}_{2}^{\top} - \mathbf{\Sigma}_{2} \mid \mathbf{X}_{1}\}, \\ \mathbb{E}\{\phi(2,3,1) \mid \mathbf{X}_1\} = \mathbb{E}\{\phi(3,2,1) \mid \mathbf{X}_1\} = \mathbb{E}\{\mathbf{A}_{2}\mathbf{A}_{21}^{\top} - \mathbf{\Sigma}_{2} \mid \mathbf{X}_{1}\},$$

and then

$$\mathbb{E}\{\psi(1,2,3)\mid \mathbf{X}_1\} = \frac{1}{3}\mathbb{E}\{(\mathbf{A}_2\mathbf{A}_{21}^\top - \boldsymbol{\Sigma}_2) + (\mathbf{A}_{21}\mathbf{A}_2^\top - \boldsymbol{\Sigma}_2) \mid \mathbf{X}_1\}.$$

Thus,

$$\begin{split} \mathbb{E}\{\psi(1,2,3)^\top \psi(1,4,5)\} &= \frac{1}{9} \mathbb{E} tr\{(\mathbf{A}_2 \mathbf{A}_{21}^\top - \boldsymbol{\Sigma}_2) + (\mathbf{A}_{21} \mathbf{A}_{2}^\top - \boldsymbol{\Sigma}_2)\}\{(\mathbf{A}_3 \mathbf{A}_{31}^\top - \boldsymbol{\Sigma}_2) + (\mathbf{A}_{31} \mathbf{A}_{3}^\top - \boldsymbol{\Sigma}_2)\} \\ &= \frac{2}{9} \{\mathbb{E}(\mathbf{A}_{21}^\top \mathbf{A}_3)(\mathbf{A}_{31}^\top \mathbf{A}_2) + \mathbb{E}(\mathbf{A}_{21}^\top \mathbf{A}_{31})(\mathbf{A}_{2}^\top \mathbf{A}_3) - 4tr(\boldsymbol{\Sigma}_2^2)\} \\ &\leq \frac{1}{9} \{\mathbb{E}(\mathbf{A}_{21}^\top \mathbf{A}_3)^2 + \mathbb{E}(\mathbf{A}_{31}^\top \mathbf{A}_2)^2 + 2\mathbb{E}(\mathbf{A}_{21}^\top \mathbf{A}_{31} - tr(\boldsymbol{\Sigma}_2))(\mathbf{A}_2^\top \mathbf{A}_3)\} - \frac{8}{9}tr(\boldsymbol{\Sigma}_2^2) \\ &\leq \frac{1}{9} \{\mathbb{E}(\mathbf{A}_{21}^\top \mathbf{A}_3)^2 + \mathbb{E}(\mathbf{A}_{31}^\top \mathbf{A}_2)^2\} + \frac{2}{9} \{var(\mathbf{A}_{12}^\top \mathbf{A}_{13})\}^{1/2} \{var(\mathbf{A}_2^\top \mathbf{A}_3)\}^{1/2} - \frac{8}{9}tr(\boldsymbol{\Sigma}_2^2) \\ &= \frac{2}{9}tr(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2) + \frac{2}{9} \{tr(\boldsymbol{\Sigma}_1^2) - tr(\boldsymbol{\Sigma}_2^2)\}^{1/2} \{tr(\boldsymbol{\Sigma}_2^2)\}^{1/2} - \frac{8}{9}tr(\boldsymbol{\Sigma}_2^2). \end{split}$$

Together with Lemmas 3 and 5, we can get

$$\mathbb{E}\{\psi(1,2,3)^{\top}\psi(1,4,5)\} \le \{\frac{6}{9} + \frac{2\sqrt{8}}{9} - \frac{8}{9}\}\operatorname{tr}(\Sigma_{2}^{2}) \le \frac{1}{2}\operatorname{tr}(\Sigma_{2}^{2}).$$
(19)

Finally, combining (16)- (19), we conclude that

$$\frac{1}{9}\mathbb{E}\{(\tilde{\boldsymbol{\rho}}_n - \mathbf{W}_n)^\top (\tilde{\boldsymbol{\rho}}_n - \mathbf{W}_n)\} \le \frac{6 + 18(n - 3)}{n(n - 1)(n - 2)}p^2 + \frac{9(n - 3)(n - 4)}{n(n - 1)(n - 2)}\{\operatorname{tr}(\boldsymbol{\Sigma}_2^2)\}.$$

By Corollary A.41 of Bai and Silverstein [1],

$$L^{3}\left(F^{\tilde{\rho}_{n}}, F^{\mathbf{W}_{n}}\right) \leq \frac{1}{p} \|\tilde{\boldsymbol{\rho}}_{n} - \mathbf{W}_{n}\|_{2}^{2}$$

which yields  $\mathbb{E}L^3\left(F^{\tilde{\rho}_n},F^{\mathbf{W}_n}\right)\to 0$  and then

$$L(F^{\tilde{\rho}_n}, F^{\mathbf{W}_n}) \to 0$$
, in probability.

The proof is completed.

## **Proof of Proposition 4**. Since

$$\rho_n = \frac{3}{n+1}\tau_n - \frac{3}{n+1}\tilde{\rho}_n + \tilde{\rho}_n,$$

by Corollary A.41 of Bai and Silverstein [1],

$$L^{3}\left(F_{n}^{\rho}, F^{\tilde{\rho}_{n}}\right) \leq \frac{1}{p} \|\rho_{n} - \tilde{\rho}_{n}\|_{2}^{2} \leq \frac{18}{n^{2}p} \|\tau_{n}\|_{2}^{2} + \frac{18}{n^{2}p} \|\tilde{\rho}_{n}\|_{2}^{2}.$$

Noting  $\tau_n$  is a correlation matrix, we have  $\|\tau_n\|_{\infty} \le 1$  and then  $\|\tau_n\|_2^2 \le p^2$ . Similarly, we can show  $\|\tilde{\rho}_n\|_2^2 \le 9p^2$ . Thus  $L\left(F^{\rho_n}, F^{\tilde{\rho}_n}\right) \to 0$ . Together with Theorem 1, the claim follows. 

## Proof of Theorem 2. Recall

$$\mathbf{R} = \sqrt{\frac{12}{n^2 - 1}} \left( r_{ij} - \frac{n+1}{2} \right)_{n \times p} = \sqrt{\frac{12n^2}{n^2 - 1}} \left( \hat{F}_j(x_{ij}) - \frac{n+1}{2n} \right)_{n \times p}$$

and  $\rho_n = \frac{1}{n} \mathbf{R}^{\mathsf{T}} \mathbf{R}$ . Writing

$$\tilde{\mathbf{R}} = \sqrt{\frac{12n^2}{n^2 - 1}} \left( F_j(x_{ij}) - \frac{n+1}{2n} \right)_{n \times n}$$

by Corollary A.42 of Bai and Silverstein [1], we have

$$L^{4}\left(F^{\rho_{n}}, F^{\frac{1}{n}\tilde{\mathbf{R}}^{\top}\tilde{\mathbf{R}}}\right) \leq \frac{2}{p^{2}}\left\{\operatorname{tr}\left(\rho_{n} + \frac{1}{n}\tilde{\mathbf{R}}^{\top}\tilde{\mathbf{R}}\right)\right\}\left\{\frac{1}{n}\|\mathbf{R} - \tilde{\mathbf{R}}\|_{2}^{2}\right\} \leq \frac{20}{np}\|\mathbf{R} - \tilde{\mathbf{R}}\|_{2}^{2},$$

where we use the facts  $tr(\rho_n) = p$  and

$$\operatorname{tr}\left(\frac{1}{n}\tilde{\mathbf{R}}^{\top}\tilde{\mathbf{R}}\right) \leq p||\tilde{\mathbf{R}}||_{\infty}^{2} \leq \frac{12n^{2}p}{n^{2}-1}\left(\frac{n+1}{2n}\right)^{2} \leq 9p, \ (n \geq 2).$$

Noting

$$\frac{1}{np} \|\mathbf{R} - \tilde{\mathbf{R}}\|_{2}^{2} \leq \frac{12n^{2}}{n^{2} - 1} \max_{i,j} |\hat{F}_{j}(x_{ij}) - F_{j}(x_{ij})|^{2} \leq \frac{12n^{2}}{n^{2} - 1} \max_{j} \left\{ \sup_{x \in \mathbb{R}} |\hat{F}_{j}(x) - F_{j}(x)|^{2} \right\} = \frac{12n^{2}}{n^{2} - 1} \left\{ \max_{j} \|\hat{F}_{j}(x) - F_{j}(x)\|_{\infty} \right\}^{2},$$

we need to control  $\|\hat{F}_j(x) - F_j(x)\|_{\infty}$ ,  $j \in \{1, ..., p\}$  simultaneously. Thus, we shall use the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality which provides a refined result of the classical Glivenko-Cantelli Theorem. By DKW inequality,  $\forall \epsilon > 0$ ,

$$\Pr\left(\|\hat{F}_j(x) - F_j(x)\|_{\infty} \ge \epsilon\right) \le 2e^{-2n\epsilon^2}$$

which yields

$$\Pr\left(\max_{j} \|\hat{F}_{j}(x) - F_{j}(x)\|_{\infty} \ge \epsilon\right) \le \sum_{j=1}^{p} \Pr\left(\|\hat{F}_{j}(x) - F_{j}(x)\|_{\infty} \ge \epsilon\right) \le 2pe^{-2n\epsilon^{2}}.$$

Since  $p/n \to y$ , we have  $2pe^{-n\epsilon^2} \le 1$  for large enough n and then

$$\sum_{n=1}^{\infty} 2pe^{-2n\epsilon^2} = \sum_{n=1}^{\infty} \{2pe^{-n\epsilon^2}\} \cdot e^{-n\epsilon^2} < \infty.$$

By the Borel-Cantelli Lemma, we obtain

$$\max_{j} ||\hat{F}_{j}(x) - F_{j}(x)||_{\infty} \to 0$$
, almost surely,

and then

$$L(F^{\rho_n}, F^{\frac{1}{n}\tilde{\mathbf{R}}^{\mathsf{T}}\tilde{\mathbf{R}}}) \to 0$$
, almost surely. (20)

Next, we bound the difference between  $\tilde{\mathbf{R}}^{\mathsf{T}}\tilde{\mathbf{R}}/n$  and  $\mathbf{W}_n=3\sum_{i=1}^n\mathbf{A}_i\mathbf{A}_i^{\mathsf{T}}/n$ . Noting

$$\tilde{\mathbf{R}} = \sqrt{\frac{12n^2}{n^2 - 1}} \left( F_j(x_{ij}) - \frac{n+1}{2n} \right)_{n \times n}$$

and

$$\sqrt{3}(\mathbf{A}_1,\ldots,\mathbf{A}_n)^{\mathsf{T}} = \sqrt{12}\left(F_j(x_{ij}) - \frac{1}{2}\right)_{n \times p},$$

we can get

$$\|\tilde{\mathbf{R}} - \sqrt{3}(\mathbf{A}_1, \dots, \mathbf{A}_n)^{\mathsf{T}}\|_{\infty} \le 2\sqrt{3} \left| \sqrt{\frac{n^2}{n^2 - 1}} - 1 \right| + \sqrt{3} \left| \frac{n + 1}{\sqrt{n^2 - 1}} - 1 \right| \to 0.$$

Using Corollary A.42 of Bai and Silverstein [1] again, we have  $L^4\left(F^{\mathbf{W}_n}, F^{\frac{1}{n}\tilde{\mathbf{R}}^T\tilde{\mathbf{R}}}\right) \to 0$ . Combined with (20), we conclude that

$$L(F^{\rho_n}, F^{\mathbf{W}_n}) \to 0$$
, almost surely.

The proof is completed.

**Proof of Theorem 3.** Since  $A_1, ..., A_n$  are i.i.d., we can apply Theorem 1.1 of Bai and Zhou [2]. It is sufficient to check assumption 1 of their theorem, i.e., for any non-random  $p \times p$  matrix **B**,

$$\frac{1}{p^2} \operatorname{var} \left( \mathbf{A}_1^{\top} \mathbf{B} \mathbf{A}_1 \right) \to 0.$$

Noting  $\mathbf{A}_i = 2\Phi(\mathbf{X}_i) - 1$  where  $\Phi(\cdot)$  is differentiable, the Gaussian Poincaré inequality together with  $\|\mathbf{\Sigma}\| \le C$  can yield the above conclusion. A detailed proof can be found in Lemma 3.3 of [17]. The proof is completed.

#### References

- [1] Z. Bai, J. W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, Springer, New York, 2010.
- [2] Z. Bai, W. Zhou, Large sample covariance matrices without independence structures in columns, Statistica Sinica (2008) 425-442.
- [3] A. S. Bandeira, A. Lodhia, P. Rigollet, Marčenko-Pastur law for Kendall's tau, Electronic Communications in Probability 22 (2017).
- [4] Z. Bao, Tracy-Widom limit for Kendall's tau, Annals of Statistics 47 (2019) 3504-3532.
- [5] Z. Bao, Tracy-Widom limit for Spearman's rho, 2019. Available at http://mazgbao.people.ust.hk.
- [6] Z. Bao, L.-C. Lin, G. Pan, W. Zhou, Spectral statistics of large dimensional Spearman's rank correlation matrix and its application, Annals of Statistics 43 (2015) 2588–2623.
- [7] T. T. Cai, Global testing and large-scale multiple testing for high-dimensional covariance structures, Annual Review of Statistics and Its Application 4 (2017) 423–446.
- [8] N. El Karoui, Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond, Annals of Applied Probability 19 (2009) 2362–2405.
- [9] F. Esscher, On a method of determining correlation from the ranks of the variates, Scandinavian Actuarial Journal 1924 (1924) 201-219.
- [10] J. Fan, Y. Liao, H. Liu, An overview of the estimation of large covariance and precision matrices. The Econometrics Journal 1 (2016) 1–32.
- [11] F. Han, S. Chen, H. Liu, Distribution-free tests of independence in high dimensions, Biometrika 104 (2017) 813–828.
- [12] T. P. Hettmansperger, J. W. McKean, Robust Nonparametric Statistical Methods, CRC Press, Boca Raton, 2010.
- [13] W. Hoeffding, A class of statistics with asymptotically normal distributions, Annals of Mathematical Statistics 19 (1948) 293-325.
- [14] T. Jiang, The limiting distributions of eigenvalues of sample correlation matrices, Sankhyā: The Indian Journal of Statistics 66 (2004) 35-48.
- [15] M. G. Kendall, A new measure of rank correlation, Biometrika 30 (1938) 81-93.
- [16] D. Leung, M. Drton, Testing independence in high dimensions with sums of rank correlations, Annals of Statistics 46 (2018) 280-307.
- [17] Z. Li, C. Wang, Q. Wang, On eigenvalues of a high dimensional Kendall's rank correlation matrix with dependences, arXiv:2109.13624 (2021) 1–28.
- [18] Z. Li, Q. Wang, R. Li, Central limit theorem for linear spectral statistics of large dimensional Kendall's rank correlation matrices and its applications, Annals of Statistics 49 (2021) 1569–1593.
- [19] H. Liu, F. Han, M. Yuan, J. Lafferty, L. Wasserman, High-dimensional semiparametric Gaussian copula graphical models, Annals of Statistics 40 (2012) 2293–2326.
- [20] H. Liu, J. Lafferty, L. Wasserman, The nonparanormal: Semiparametric estimation of high dimensional undirected graphs, Journal of Machine Learning Research 10 (2009) 2295–2328.
- [21] J. Lu, F. Han, H. Liu, Robust scatter matrix estimation for high dimensional distributions with heavy tail, IEEE Transactions on Information Theory 67 (2021) 5283–5304.
- [22] G. Mao, Robust test for independence in high dimensions, Communications in Statistics-Theory and Methods 46 (2017) 10036–10050.
- [23] V. A. Marčenko, L. A. Pastur, Distribution of eigenvalues for some sets of random matrices, Mathematics of the USSR-Sbornik 1 (1967) 457.
- [24] K. Pearson, Determination of the coefficient of correlation, Science 30 (1909) 23–25.
- [25] C. Spearman, The proof and measurement of association between two things, The American Journal of Psychology 15 (1904) 72–101.
- [26] L. Weihs, M. Drton, N. Meinshausen, Symmetric rank covariances: a generalized framework for nonparametric measures of dependence, Biometrika 105 (2018) 547–562.
- [27] L. Xue, H. Zou, Regularized rank-based estimation of high-dimensional nonparanormal graphical models, Annals of Statistics 40 (2012) 2541–2571.