Bernoulli **25**(2), 2019, 1472–1503
https://doi.org/10.3150/18-BEJ1027

An extreme-value approach for testing the equality of large U-statistic based correlation matrices

CHENG ZHOU[1,*], FANG HAN[2,**], XIN-SHENG ZHANG[1,†] and HAN LIU[3,‡]

[1]*Department of Statistics, Management School, Fudan University, Shanghai, China. E-mail: *chengzhmike@gmail.com; †xszhang@fudan.edu.cn*
[2]*Department of Statistics, University of Washington, Seattle, WA 98195, USA. E-mail: **fanghan@uw.edu*
[3]*Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. E-mail: ‡hanliu@northwestern.edu*

There has been an increasing interest in testing the equality of large Pearson's correlation matrices. However, in many applications it is more important to test the equality of large rank-based correlation matrices since they are more robust to outliers and nonlinearity. Unlike the Pearson's case, testing the equality of large rank-based statistics has not been well explored and requires us to develop new methods and theory. In this paper, we provide a framework for testing the equality of two large U-statistic based correlation matrices, which include the rank-based correlation matrices as special cases. Our approach exploits extreme value statistics and the Jackknife estimator for uncertainty assessment and is valid under a fully nonparametric model. Theoretically, we develop a theory for testing the equality of U-statistic based correlation matrices. We then apply this theory to study the problem of testing large Kendall's tau correlation matrices and demonstrate its optimality. For proving this optimality, a novel construction of least favorable distributions is developed for the correlation matrix comparison.

Keywords: extreme value type I distribution; hypothesis testing; Jackknife variance estimator; Kendall's tau; U-statistics

1. Introduction

Let $\boldsymbol{X} = (X_1, \dots, X_d)^T$ and $\boldsymbol{Y} = (Y_1, \dots, Y_d)^T$ be two d-dimensional random vectors. We denote $\boldsymbol{X}_1, \dots, \boldsymbol{X}_{n_1}$ with $\boldsymbol{X}_k = (X_{k1}, \dots, X_{kd})^T$ to be n_1 independent samples of $\boldsymbol{X}$ and $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{n_2}$ with $\boldsymbol{Y}_k = (Y_{k1}, \dots, Y_{kd})^T$ to be n_2 independent samples of $\boldsymbol{Y}$. Letting $n := \max\{n_1, n_2\}$, we aim to test the equality of U-statistic based correlation matrices (e.g., Kendall's tau or Spearman's rho) of $\boldsymbol{X}$ and $\boldsymbol{Y}$. We consider the high dimensional regime that $d, n \to \infty$ and d/n does not necessarily go to zero as $n \to \infty$. This problem has important applications, including portfolio selection (Markowitz [32]), high dimensional discriminant analysis (Han, Zhao and Liu [18], Mai and Zou [31]) and gene selection (Ho et al. [19], Hu et al. [22], Hu, Qiu and Glazko [21]).

When $d/n \to 0$, Anderson [1] and Muirhead [33] study the problem of testing the equality of two Pearson's correlation matrices. Major test criteria include the likelihood ratio (Anderson [1]), spectral norm of difference (Roy [37]) and Frobenius norm of difference (Nagao [34]). When $d/n \nrightarrow 0$, the likelihood ratio test and the tests in Roy [37] and Nagao [34] perform poorly, as

1350-7265

Pearson's sample correlation matrices no longer converge to their population counterparts under the spectral norm (Bai and Yin [5]). A line of research aims to correct the aforementioned tests or proposing new methods. For the likelihood ratio test, Bai et al. [3] introduce a corrected LRT test which works when $d/n \to c \in (0, 1)$, and Jiang, Jiang and Yang [23] generalize it to the case when $d < n$ and $c = 1$. Based on the spectral norm of difference, Han, Xu and Zhou [17] use the bootstrap method to generalize Roy's test in high dimension. As a generalization of Nagao's proposal, Schott [38] and Li and Chen [27] propose new test statistics based on an unbiased estimator of the Frobenius norm of the matrix difference, and Srivastava and Yanagihara [42] propose another test statistic based on the difference of two Frobenius norms. Recently, Cai, Liu and Xia [10] propose a method based on the sup-norm of the matrix difference and prove its rate optimality under a sparse alternative.

In many applications, it is more meaningful to test the equality of two rank-based correlation matrices but instead of the Pearson's correlation matrices. In particular, Embrechts, Lindskog and McNeil [13] point out that the Pearson's correlation coefficient "might prove very misleading" in measuring the dependence and advocate the usage of rank correlation coefficients, such as Kendall's tau (Kendall [24]) or Spearman's rho (Spearman [40]). Though testing the equality of high dimensional rank-based correlation matrices is of fundamental importance, there has been very little work in this area. To bridge this gap, this paper proposes a unified framework for testing the equality of two large U-statistic based correlation matrices $\mathbf{U}_1$ and $\mathbf{U}_2$, which include rank-based correlation matrices as special examples. More specifically, let $\mathbf{U}_1 = (u_{1,ij})$ be a type of correlation matrix of $\boldsymbol{X}$ and all the elements of $\mathbf{U}_1$ can be estimated by U-statistics.[1] Similarly to $\mathbf{U}_1$, we define $\mathbf{U}_2 = (u_{2,ij})$ to be the same kind of U-statistic based correlation matrix of $\boldsymbol{Y}$. In this paper, we aim to test the hypothesis

$$\mathbf{H}_0 : \mathbf{U}_1 = \mathbf{U}_2 \quad \text{v.s.} \quad \mathbf{H}_1 : \mathbf{U}_1 \neq \mathbf{U}_2. \tag{1.1}$$

Testing (1.1) plays an important role in many fields. For example, testing the equality of two Kendall's tau correlation matrices $\mathbf{U}_1^\tau$ and $\mathbf{U}_2^\tau$,

$$\mathbf{H}_0^\tau : \mathbf{U}_1^\tau = \mathbf{U}_2^\tau \quad \text{v.s.} \quad \mathbf{H}_1^\tau : \mathbf{U}_1^\tau \neq \mathbf{U}_2^\tau, \tag{1.2}$$

can be used to test the model of copula discriminant analysis (Han, Zhao and Liu [18], Mai and Zou [31]).

There are 4 major contributions of this paper. First, for the first time in the literature, we develop a unified framework for testing the equality of two large U-statistic based correlation matrices. This framework builds upon a fully nonparametric model and enables us to conduct homogeneity tests using a wide range of correlation measures. Secondly, as a special example, we examine the problem of testing the equality of two large Kendall's tau matrices and prove the minimax optimality of the proposed method. Thirdly, we further propose alternative approaches for testing $\mathbf{U}_1^\tau = \mathbf{U}_2^\tau$, which attain better empirical performance than the Jackknife based one. Finally, to develop a theory of testing the equality of general U-statistic based correlation matrices,

[1]Such U-statistic based correlation measures are quite general. For example, $u_{1,ij}$ can represent the Kendall's tau correlation coefficient between X_i and X_j.

we develop an upper bound of Jackknife variance estimation error, which enables us to obtain the explicit rate of convergence. For Kendall's tau matrices, we prove an upper bound of the traditional plug-in variance estimation error and an upper bound of the variance difference between two Kendall's tau correlation coefficients. These upper bounds allow us to exploit the extreme value theory under the dependent setting to prove theorems in this paper. Their constructions are nontrivial and are of independent technical interest. To prove the optimality of the proposed testing methods for Kendall's tau matrices, we construct a collection of least favorable distributions with regard to the test hypothesis. This construction technique is novel and tailored for testing the equality of correlation matrices. In contrast, the construction in Cai, Liu and Xia [10] only perturbs the diagonal elements of covariance matrices, which does not affect the resulting correlation matrices.

1.1. More related works

Apart from the Pearson's correlation coefficient and general U-statistic based correlation measurements studied in this paper, existing literature also considers other measures of dependence. These include the distance correlation (Székely, Rizzo and Bakirov [43]) and randomized dependence coefficient (Lopez-Paz, Hennig and Schölkopf [29]). To the best of our knowledge, there is no work discussing testing the equality of dependence structure with regard to these dependent measures.

Our work is closely related to the random matrix theory on rank correlation matrices. Bai and Zhou [4], Zhou [47], Bao et al. [6], and Han, Chen and Liu [16] study the theoretical properties of large rank-based correlation matrices. Specifically, for these random matrices, Bai and Zhou [4] prove the Marchenko–Pastur law for the limiting spectral distribution, Zhou [47] and Han, Chen and Liu [16] prove the extreme value type I distribution for the entry-wise maximum, and Bao et al. [6] derive the limiting distributions of traces of all higher moments. Most of these results hold only under the independence setting, that is, the entries of $\boldsymbol{X}$ are independent of each other. In contrast, our work focuses on the dependent setting.

Our work is also related to the robust testing, where the test statistics are robust estimators of the Pearson's covariance/correlation coefficients. These include S-estimators and some robust dispersion estimators. We refer to O'Brien [35], Aslam and Rocke [2] and the references therein for details. Our work is also related to the adaptive estimation of a large correlation/covariance matrix (Cai and Liu [9]) or a large Gaussian (copula) graphical model. See, for example, Bickel and Levina [8], Zhao, Roeder and Liu [45], Ravikumar et al. [36] and Liu et al. [28].

1.2. Notation

We denote $\|\mathbf{v}\|_2 = (\sum_{j=1}^d v_j^2)^{1/2}$ as the Euclidean norm of a vector $\mathbf{v} = (v_1, \dots, v_d)^T \in \mathbb{R}^d$. For a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{d\times q}$, we define its spectral norm $\|\mathbf{A}\|_2 := \sup_{\|\boldsymbol{x}\|_2\le 1} \|\mathbf{A}\boldsymbol{x}\|_2$ and Frobenius norm $\|\mathbf{A}\|_F := \sqrt{\sum_{i,j} a_{ij}^2}$. We define the matrix entrywise sup-norm as $\|\mathbf{A}\|_{\max} := \max\{|a_{ij}|\}$. We use $\mathrm{Rank}(\mathbf{A})$ to denote the rank of $\mathbf{A}$. If $\mathbf{A}$ is a square matrix, we define $\mathrm{Diag}(\mathbf{A})$ to be a diagonal matrix with the same main diagonal as $\mathbf{A}$. We use $\mathbf{I}_d$ to denote an identity matrix of size d. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if there exists a constant C such that $|a_n| \le C|b_n|$ holds for all sufficiently large n, write $a_n = o(b_n)$ if $a_n/b_n \to 0$, and write $a_n \asymp b_n$ if there exist constants $C \ge c > 0$ such that $c|b_n| \le |a_n| \le C|b_n|$ for all sufficiently large n. For a square matrix $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$, we use $\lambda_{\min}(\mathbf{\Sigma})$ and $\lambda_{\max}(\mathbf{\Sigma})$ to denote the minimal and maximal eigenvalues of $\mathbf{\Sigma}$. For a set B, we use $|B|$ to denote its cardinality.

1.3. Paper organization

The rest of this paper is organized as follows. Section 2 formalizes the problem, describes a general testing procedure and analyzes the theoretical properties (e.g., size and power) of the proposed test. In Section 3, we focus on testing large Kendall's tau matrices, for which we consider two models: a fully nonparametric model and a semiparametric Gaussian copula model. Under certain modelling assumptions, for Kendall's tau matrices we propose additional tests which have better empirical performance compared to the general testing procedure. Section 4 provides thorough numerical results on both simulated and real data. In Section 5, we discuss potential future work. Appendix A contains the proof of the main theorem. We put the proofs of all other results in Supplementary Material (Zhou et al. [46]) of this paper.

2. A general procedure for testing U-statistic based matrices

This section presents a generic testing method for U-statistic based matrix comparison. In Section 2.1, we describe the proposed testing procedure. In Section 2.2, we analyze its asymptotic size and power. In Section 2.3, we consider comparing a row or column of U-statistic based matrices.

Before presenting the testing procedure, we introduce some notations for U-statistics. For $i, j = 1, \dots, q$, let Φ_{ij} be a U-statistic's kernel function defined as

$$\Phi_{ij} : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{m} \to \mathbb{R} \qquad \text{with the symmetric property: } \Phi_{ij} = \Phi_{ji}, \tag{2.1}$$

where m is the kernel order. Thus, we have a family of functions $\{\Phi_{ij}, 1 \le i, j \le q\}$. Furthermore, each Φ_{ij} is a symmetric Borel measurable function with the kernel order m fixed.[2] We assume that Φ_{ij} is uniformly bounded. Many useful U-statistics satisfy these conditions. We set

$$\widehat{u}_{1,ij} := \binom{n_1}{m}^{-1} \sum_{1 \le \ell_1 < \cdots < \ell_m \le n_1} \Phi_{ij}(\boldsymbol{X}_{\ell_1}, \dots, \boldsymbol{X}_{\ell_m}),$$

$$\widehat{u}_{2,ij} := \binom{n_2}{m}^{-1} \sum_{1 \le \ell_1 < \cdots < \ell_m \le n_2} \Phi_{ij}(\boldsymbol{Y}_{\ell_1}, \dots, \boldsymbol{Y}_{\ell_m}).$$

[2]We assume each Φ_{ij} has the same fixed kernel order m for presentation clearness. It is straightforward to extend to the setting that m's are uniformly bounded.

We then define the following U-statistic based matrices $\widehat{\mathbf{U}}_a \in \mathbb{R}^{q\times q}$ for $a = 1, 2$:

$$\widehat{\mathbf{U}}_1 := (\widehat{u}_{1,ij})_{1\le i,j\le q} \quad \text{and} \quad \widehat{\mathbf{U}}_2 := (\widehat{u}_{2,ij})_{1\le i,j\le q}. \tag{2.2}$$

Correspondingly, we use $\mathbf{U}_a := (u_{a,ij})_{1\le i,j\le q}$ to denote the expectation of $\widehat{\mathbf{U}}_a$, i.e., $u_{a,ij} = \mathbb{E}[\widehat{u}_{a,ij}]$. We can view $\mathbf{U}_1$ and $\mathbf{U}_2$ as a type of correlation matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$. We are interested in testing the equality of $\mathbf{U}_1$ and $\mathbf{U}_2$, which includes testing the equality of two large Kendall's tau or Spearman's rho correlation matrices.

We note that q is the row and column number of $\mathbf{U}_a$ and $\widehat{\mathbf{U}}_a$, while d is the dimension of $\boldsymbol{X}$ and $\boldsymbol{Y}$. q and d can be different. Therefore, the framework considered in this paper is quite general. For example, it allows $\mathbf{U}_a$ to represent the dependence structure on a dimension reduced data, where the dimension reduction step is incorporated in the kernel function $\{\Phi_{ij}, 1 \le i, j \le q\}$.

Remark 2.1. We can relax Φ_{ij} to be an asymmetric kernel function without loss of generality. Specifically, an asymmetric kernel $\Phi(\cdot)$ gives a U-statistic

$$\widehat{u} = \frac{1}{m!}\binom{n_1}{m}^{-1}\sum \Phi(\boldsymbol{X}_{\ell_1}, \dots, \boldsymbol{X}_{\ell_m}),$$

where the summation is taken over all combinations of distinct elements $\{\ell_1, \dots, \ell_m\}$ from $\{1, \dots, n_1\}$. Using the Hoeffding's method (Hoeffding [20]), $\widehat{u}$ is also a U-statistic of the symmetric kernel $\Phi^0(\cdot)$:

$$\Phi^0(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \frac{1}{m!}\sum \Phi(\mathbf{x}_{\alpha_1}, \dots, \mathbf{x}_{\alpha_m}),$$

where the summation is taken over all permutations of $\{1, \dots, m\}$. For example, to construct an unbiased[3] estimator for Spearman's rho, El Maache and Lepage [12] recommends to use the U-statistic with the kernel

$$\Phi_{ij}(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3) = 2^{-1}\sum\sum\sum_{1\le\alpha\ne\beta\ne\gamma\le 3} \operatorname{sign}(X_{\alpha i} - X_{\beta i}) \operatorname{sign}(X_{\alpha j} - X_{\gamma j}).$$

The Kendall's tau matrix is an example of the U-statistic based matrix defined in (2.2). More specifically, we set

$$\Phi_{ij}(\boldsymbol{X}_k, \boldsymbol{X}_\ell) = \operatorname{sign}(X_{ki} - X_{\ell i}) \operatorname{sign}(X_{kj} - X_{\ell j}),$$
$$\Phi_{ij}(\boldsymbol{Y}_k, \boldsymbol{Y}_\ell) = \operatorname{sign}(Y_{ki} - Y_{\ell i}) \operatorname{sign}(Y_{kj} - Y_{\ell j}),$$

[3] The Spearman rank-order correlation, i.e., the sample correlation between the rank values of two variables, is a biased estimator of the population Spearman's rho.

and $q = d$. The Kendall's tau sample correlation coefficients $\widehat{\tau}_{1,ij}$ and $\widehat{\tau}_{2,ij}$ are then defined as

$$\widehat{\tau}_{1,ij} := \frac{2}{n_1(n_1-1)} \sum_{1\le k<\ell\le n_1} \operatorname{sign}(X_{ki}-X_{\ell i})\operatorname{sign}(X_{kj}-X_{\ell j}),$$

$$\widehat{\tau}_{2,ij} := \frac{2}{n_2(n_2-1)} \sum_{1\le k<\ell\le n_2} \operatorname{sign}(Y_{ki}-Y_{\ell i})\operatorname{sign}(Y_{kj}-Y_{\ell j}).$$

Their population counterparts are $\tau_{a,ij} := \mathbb{E}[\widehat{\tau}_{a,ij}]$ for $a = 1, 2$. We then write sample and population Kendall's tau matrices as

$$\widehat{\mathbf{U}}_a^\tau = (\widehat{\tau}_{a,ij}) \quad \text{and} \quad \mathbf{U}_a^\tau = (\tau_{a,ij}), \tag{2.3}$$

where $a = 1, 2$. In Section 3, we consider testing the large Kendall's tau matrices.

2.1. A general testing procedure

For testing (1.1) in high dimensions, we use the sup-norm criterion. Such a choice is motivated by the fact that the sup-norm is very sensitive to perturbations on a small number of entries compared to the null hypothesis. We then propose the test statistic:

$$M_n := \max_{1\le i,j\le q} M_{ij} \qquad \text{with } M_{ij} := \frac{(\widehat{u}_{1,ij}-\widehat{u}_{2,ij})^2}{\widehat{\sigma}^2(\widehat{u}_{1,ij})+\widehat{\sigma}^2(\widehat{u}_{2,ij})} \text{ for } 1\le i, j\le q. \tag{2.4}$$

In (2.4), $\widehat{\sigma}^2(\widehat{u}_{1,ij})$ is a Jackknife estimator of $\widehat{u}_{1,ij}$'s variance and is defined as

$$\widehat{\sigma}^2(\widehat{u}_{1,ij}) := \frac{m^2(n_1-1)}{n_1(n_1-m)^2} \sum_{\alpha=1}^{n_1} (q_{1\alpha,ij}-\widehat{u}_{1,ij})^2, \tag{2.5}$$

with

$$q_{1\alpha,ij} := \binom{n_1-1}{m-1}^{-1} \sum_{\substack{1\le\ell_1<\cdots<\ell_{m-1}\le n_1\\ \ell_j\ne\alpha,\, j=1,\dots,m-1}} \Phi_{ij}(\boldsymbol{X}_\alpha, \boldsymbol{X}_{\ell_1}, \dots, \boldsymbol{X}_{\ell_{m-1}}).$$

The definition of $\widehat{\sigma}^2(\widehat{u}_{2,ij})$ is similar for $\boldsymbol{Y}$.

For a given significance level $0<\alpha<1$, we construct the test to be

$$\mathrm{T}_\alpha := \mathbb{1}\{M_n \ge G^-(\alpha) + 4\log q - \log(\log q)\}, \tag{2.6}$$

where $G^-(\alpha) := -\log(8\pi) - 2\log(-\log(1-\alpha))$. We reject $\mathbf{H}_0$ in (1.1) if and only if $\mathrm{T}_\alpha = 1$.

In some applications, our interest is to compare a particular row or column of matrices, that is, we aim at testing the hypothesis:

$$\mathbf{H}_{0,i} : \mathbf{u}_{1,i\star} = \mathbf{u}_{2,i\star} \quad \text{v.s.} \quad \mathbf{H}_{1,i} : \mathbf{u}_{1,i\star} \ne \mathbf{u}_{2,i\star}, \tag{2.7}$$

where $\mathbf{u}_{1,i\star}$ and $\mathbf{u}_{2,i\star}$ are the ith rows of $\mathbf{U}_1$ and $\mathbf{U}_2$. To test this hypothesis, we construct a similar test statistic $M_{n,i} = \max_{1 \le j \le q} M_{i,j}$, and the according test is

$$\mathrm{T}_{\alpha,i} = \mathbb{1}\{M_{n,i} > G'^{-}(\alpha) + 2\log q - \log\log q\}, \tag{2.8}$$

where $G'^{-}(\alpha) := -\log(\pi) - 2\log(-\log(1-\alpha))$. We reject $\mathbf{H}_{0,i}$ if and only if $\mathrm{T}_{\alpha,i} = 1$.

2.2. Theoretical properties

Our main theoretical result is to characterize the limiting null distribution of M_n. We further analyze the power of the proposed test under a sparse alternative.

We introduce three assumptions that will be used later. Assumption **(A1)** specifies the sparsity of $\mathbf{U} = \mathbf{U}_1 = \mathbf{U}_2$. Assumption **(A2)** specifies the scaling of q, n. Assumption **(A3)** is a technical condition that we impose for obtaining the limiting distribution of M_n. In Section 3.2, we will show that Assumption **(A3)** can be further relaxed under a semiparametric Gaussian copula model.

In detail, for a fixed constant $\alpha_0 > 0$, we define

$$\mathrm{supp}_j(\alpha_0) := \{1 \le i \le q : |u_{1,ij}| \ge (\log q)^{-1-\alpha_0} \text{ or } |u_{2,ij}| \ge (\log q)^{-1-\alpha_0}\}.$$

$\mathrm{supp}_j(\alpha_0)$ is the set of indices i such that either the ith variable of $\boldsymbol{X}$ is highly correlated ($|u_{a,ij}| > (\log q)^{-1-\alpha_0}$) with the jth variable of $\boldsymbol{X}$, or the ith variable of $\boldsymbol{Y}$ is highly correlated with the jth variable of $\boldsymbol{Y}$. We then introduce Assumption **(A1)** as follows.

(A1) We assume that there exits a subset $\Gamma \subset \{1, 2, \ldots, q\}$ with $|\Gamma| = o(q)$ and a constant $\alpha_0 > 0$ such that for all $\gamma > 0$, we have

$$\max_{1 \le j \le q, j \notin \Gamma} |\mathrm{supp}_j(\alpha_0)| = o(q^\gamma).$$

Before stating Assumption **(A2)**, we need some additional notations. Set

$$\Psi_{ij}(\boldsymbol{X}_{\ell_1}, \ldots, \boldsymbol{X}_{\ell_m}) := \Phi_{ij}(\boldsymbol{X}_{\ell_1}, \ldots, \boldsymbol{X}_{\ell_m}) - u_{1,ij}.$$

For $\ell = 1, \ldots, n_1$, we also denote $g_{ij}(\boldsymbol{X}_\ell)$ and $h_{ij}(\boldsymbol{X}_\ell)$ as

$$\begin{aligned} g_{ij}(\boldsymbol{X}_\ell) &:= \mathbb{E}[\Phi_{ij}(\boldsymbol{X}_{\ell_1}, \ldots, \boldsymbol{X}_{\ell_m}) | \boldsymbol{X}_\ell], \\ h_{ij}(\boldsymbol{X}_\ell) &:= \mathbb{E}[\Psi_{ij}(\boldsymbol{X}_{\ell_1}, \ldots, \boldsymbol{X}_{\ell_m}) | \boldsymbol{X}_\ell], \end{aligned} \tag{2.9}$$

where $\{\ell_1, \ldots, \ell_m\}$ is an arbitrary subset of $\{1, \ldots, n_1\}$ with distinct elements and contains ℓ. $g_{ij}(\boldsymbol{Y}_\ell)$ and $h_{ij}(\boldsymbol{Y}_\ell)$ are similarly defined for $\ell = 1, \ldots, n_2$. We then denote $\zeta_{1,ij}$ to be the variance of $g_{ij}(\boldsymbol{X}_\ell)$, that is,

$$\zeta_{1,ij} := \mathbb{E}[\mathbb{E}[\Psi_{ij}(\boldsymbol{X}_{\ell_1}, \ldots, \boldsymbol{X}_{\ell_m}) | \boldsymbol{X}_\ell]^2] = \mathrm{Var}(g_{ij}(\boldsymbol{X}_\ell)). \tag{2.10}$$

Similarly, we define $\zeta_{2,ij} := \mathrm{Var}(g_{ij}(\boldsymbol{Y}_\ell))$.

With these introduced notations, we are now ready to state Assumption **(A2)**.

(A2) We assume $n_1 \asymp n_2 \asymp n$ and $\log q = O(n^{1/3-\epsilon})$ for an arbitrary $0 < \epsilon < 1/3$. We also assume $\zeta_{a,ij} > r_a > 0$ for $a = 1, 2$, where r_1 and r_2 are constants which are irrelevant to i and j.

The condition that $\zeta_{a,ij} > r_a > 0$ is mild. It is used to exclude the degenerate cases of U-statistics and has been widely used for analyzing U-statistics.

To describe Assumption **(A3)**, we write

$$S = \{(i, j) : 1 \le i, j \le q\} \quad \text{and} \quad S_0 = \{(i, j) : 1 \le i \le q, i \in \operatorname{supp}_j(\alpha_0)\}. \tag{2.11}$$

By the definition of S_0, for any $(i, j) \in S \setminus S_0$, we have $|u_{a,ij}| \le (\log q)^{-1-\alpha_0}$. Moreover, we use $u_{1,ijk\ell}$ and $u_{2,ijk\ell}$ to denote $\mathbb{E}[g_{ij}(\boldsymbol{X}_\ell)g_{k\ell}(\boldsymbol{X}_\ell)]$ and $\mathbb{E}[g_{ij}(\boldsymbol{Y}_\ell)g_{k\ell}(\boldsymbol{Y}_\ell)]$.

(A3) Assume $u_{a,ijk\ell} = O((\log q)^{-1-\alpha_0})$ for any $(i, j) \ne (k, \ell) \in S \setminus S_0$ and $a = 1, 2$.

Under fully nonparametric models, we note that $u_{a,ijk\ell}$ is estimable (Klüppelberg and Kuhn [25]). Thus it is possible to verify Assumption **(A3)** in applications. When we test the equality of two Kendall's tau correlation matrices $\mathbf{U}_1^\tau$ and $\mathbf{U}_2^\tau$, under a semiparametric Gaussian copula model Assumption **(A3)** can be replaced by a simplified condition which is easier to be verified. More details are provided in Section 3.2.

Under the above assumptions, our main theoretical result quantifies the limiting distribution of the extreme value statistic M_n.

Theorem 2.2. *Assuming* **(A1)**, **(A2)**, **(A3)** *hold*, *under* $\mathbf{H}_0$ *of* (1.1), *we have*

$$\mathbb{P}\bigl(M_n - 4\log q + \log(\log q) \le x\bigr) \to \exp\biggl(-\frac{1}{\sqrt{8\pi}}\exp\biggl(-\frac{x}{2}\biggr)\biggr), \tag{2.12}$$

for any $x \in \mathbb{R}$, *as* $n, q \to \infty$. *Furthermore*, (2.12) *holds uniformly for all random vectors* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *satisfying Assumptions* **(A1)**, **(A2)** *and* **(A3)**.

Proof. We list a sketch of this proof. The detailed proof is in Appendix A. The proof proceeds in three steps.

Step (i) (Sketch). We set $\widehat{\sigma}^2(\widehat{u}_{a,ij})$ as the Jackknife variance estimator of $\widehat{u}_{a,ij}$ and $\sigma^2(\widehat{u}_{a,ij})$ as the true variance of $\widehat{u}_{a,ij}$. We then analyze the estimation error of Jackknife variance estimator by providing an upper bound of $|n_a\widehat{\sigma}^2(\widehat{u}_{a,ij}) - m^2\zeta_{a,ij}|$, where $\zeta_{a,ij}$ is defined in (2.10). The central limit theorem for U-statistics (Lemma D.3 in Supplement D of Supplementary Material (Zhou et al. [46])) implies that $m^2\zeta_{a,ij}$ is the limit of $n_a\sigma^2(\widehat{u}_{a,ij})$ as n_a goes to infinity. This motivates us to define

$$M_{ij} := \frac{(\widehat{u}_{1,ij} - \widehat{u}_{2,ij})^2}{\widehat{\sigma}^2(\widehat{u}_{1,ij}) + \widehat{\sigma}^2(\widehat{u}_{2,ij})} \quad \text{and} \quad \widetilde{M}_{ij} := \frac{(\widehat{u}_{1,ij} - \widehat{u}_{2,ij})^2}{m^2\zeta_{1,ij}/n_1 + m^2\zeta_{2,ij}/n_2}. \tag{2.13}$$

In $\widetilde{M}_{ij}$, we use $m^2\zeta_{a,ij}/n_1$ to replace $\widehat{\sigma}^2(\widehat{u}_{a,ij})$ of M_{ij}.

By using the obtained upper bound of $|n_a\widehat{\sigma}^2(\widehat{u}_{a,ij}) - m^2\zeta_{a,ij}|$, we prove $\max_{1\le i,j\le q} M_{ij}$ and $\max_{1\le i,j\le q} \widetilde{M}_{ij}$ have the same limiting distribution, that is, it suffices to prove that

$$\lim_{n,q\to\infty} \mathbb{P}\big(\widetilde{M}_n - 4\log q + \log(\log q) \le x\big) = \exp\big(-\exp(-x/2)/\sqrt{8\pi}\big), \tag{2.14}$$

where $\widetilde{M}_n := \max_{1\le i,j\le q} \widetilde{M}_{ij}$.

Step (ii) (Sketch). We use the Hoeffding decomposition (Lemma D.4 in Supplement D of Supplementary Material (Zhou et al. [46])) to decompose the U-statistic $\widetilde{u}_{a,ij} := \widehat{u}_{a,ij} - u_{a,ij}$. By the definition of $\widetilde{u}_{a,ij}$, we have $\mathbb{E}[\widetilde{u}_{a,ij}] = 0$. By the Hoeffding decomposition, we decompose $\widetilde{u}_{a,ij}$ into two pieces. One is the sum of independent and identically distributed (i.i.d.) random variables and the other is the residual term. In detail, decompose $\widetilde{u}_{a,ij}$ as

$$\begin{aligned} \widetilde{u}_{1,ij} &= \frac{m}{n_1}\sum_{\alpha=1}^{n_1} h_{ij}(\boldsymbol{X}_\alpha) + \binom{n_1}{m}^{-1}\Delta_{n_1,ij}, \\ \widetilde{u}_{2,ij} &= \frac{m}{n_2}\sum_{\alpha=1}^{n_2} h_{ij}(\boldsymbol{Y}_\alpha) + \binom{n_2}{m}^{-1}\Delta_{n_2,ij}, \end{aligned} \tag{2.15}$$

where we set

$$\Delta_{n_1,ij} = \sum_{1\le \ell_1<\ell_2<\cdots<\ell_m\le n_1}\left(\Phi_{ij}(\boldsymbol{X}_{\ell_1},\dots,\boldsymbol{X}_{\ell_m}) - u_{1,ij} - \sum_{k=1}^{m} h_{ij}(\boldsymbol{X}_{\ell_k})\right),$$

$$\Delta_{n_2,ij} = \sum_{1\le \ell_1<\ell_2<\cdots<\ell_m\le n_2}\left(\Phi_{ij}(\boldsymbol{Y}_{\ell_1},\dots,\boldsymbol{Y}_{\ell_m}) - u_{2,ij} - \sum_{k=1}^{m} h_{ij}(\boldsymbol{Y}_{\ell_k})\right).$$

Apparently, $m\sum_{\alpha=1}^{n_1} h_{ij}(\boldsymbol{X}_\alpha)/n_1$ and $m\sum_{\alpha=1}^{n_2} h_{ij}(\boldsymbol{Y}_\alpha)/n_2$ are terms for the sum of i.i.d. random variables and $\binom{n_a}{m}^{-1}\Delta_{n_a,ij}$ is the residual term. We then use $m\sum_{\alpha=1}^{n_1} h_{ij}(\boldsymbol{X}_\alpha)/n_1$ and $m\sum_{\alpha=1}^{n_2} h_{ij}(\boldsymbol{Y}_\alpha)/n_2$ as the approximations of $\widetilde{u}_{1,ij}$ and $\widetilde{u}_{2,ij}$ and define

$$T_{ij} := \frac{\sum_{\alpha=1}^{n_1} h_{ij}(\boldsymbol{X}_\alpha)/n_1 - \sum_{\alpha=1}^{n_2} h_{ij}(\boldsymbol{Y}_\alpha)/n_2}{\sqrt{\zeta_{1,ij}/n_1 + \zeta_{2,ij}/n_2}} \quad \text{and} \quad T_n := \max_{1\le i,j\le q}(T_{ij})^2. \tag{2.16}$$

We then prove that the small residual term $\binom{n_a}{m}^{-1}\Delta_{n_a,ij}$ is negligible for our theorem, that is, to obtain Theorem 2.2, it suffices to prove that as $n, q\to\infty$, we have

$$\mathbb{P}\big(T_n - 4\log q + \log(\log q) \le x\big) \to \exp\big(-\exp(-x/2)/\sqrt{8\pi}\big). \tag{2.17}$$

Step (iii) (Sketch). In the last step, we derive the limiting distribution of T_n to prove (2.17). T_n is the maximum of $(T_{ij})^2$ over $\{1\le i, j\le q\}$ and T_{ij} is not independent of each other. Therefore, we cannot straightforwardly exploit the extreme value theorem under the independent setting to obtain the limiting distribution of T_n. To solve this problem, we exploit the normal approximation to get the extreme value distribution of $\{(T_{ij})^2\}_{1\le i,j\le q}$ under the setting that T_{ij} can be dependent of each other. The detailed proof of this theorem is in Appendix A. $\square$

Theorem 2.2 justifies the size of the proposed test T_α in (2.6). It shows that under $\mathbf{H}_0$ of (1.1), $M_n - 4\log q + \log(\log q)$ converges weakly to an extreme value Type I distribution with the distribution function $F(t) = \exp(-\exp(t/2)/\sqrt{8\pi})$.

Remark 2.3. Theorem 2.2 provides a unified framework for testing the equality of two large U-statistic based matrices, which include ranked-based correlation matrices as special examples. Our test method exploits the Jackknife strategy and extreme value statistics, and it works under a fully nonparametric model. Technically, for proving Theorem 2.2, we develop a set of tools for analyzing the Jackknife variance estimator defined in (2.5), which is technically nontrivial and is of independent interest for analyzing U-statistics in more general settings.

Next, we analyze the power of T_α. To this end, we first introduce an alternative hypothesis characterized by the following set of matrix pairs

$$\mathbb{A}(C) = \left\{(\mathbf{U}_1, \mathbf{U}_2) : \max_{1\le i,j\le q} \frac{|u_{1,ij} - u_{2,ij}|}{\sqrt{m^2\zeta_{1,ij}/n_1 + m^2\zeta_{2,ij}/n_2}} \ge C\sqrt{\log q}\right\},$$

where $C > 0$ is a constant. The setting that only one entry of $\mathbf{U}_1$ and $\mathbf{U}_2$ differentiates large enough will make $(\mathbf{U}_1, \mathbf{U}_2) \in \mathbb{A}(C)$ for some constant C. The next theorem shows that the null hypothesis is asymptotically distinguishable from $\mathbb{A}(4)$ by T_α, that is, we can use T_α to reject $\mathbf{H}_0$ in (1.1) with an overwhelming probability if $(\mathbf{U}_1, \mathbf{U}_2) \in \mathbb{A}(4)$.

Theorem 2.4 (Power of the Test T_α). *If* **(A2)** *is satisfied, as $n, q \to \infty$ we have*

$$\inf_{(\mathbf{U}_1,\mathbf{U}_2)\in\mathbb{A}(4)} \mathbb{P}(\mathrm{T}_\alpha = 1) \to 1. \tag{2.18}$$

Remark 2.5. From the above theorem, for big enough $C > 0$, only one entry of $\mathbf{U}_1 - \mathbf{U}_2$ has a magnitude more than $C\sqrt{\log q/n}$ is enough for the test T_α to correctly reject $\mathbf{H}_0$ of (1.1). We don't impose Assumptions (**A1**) and (**A3**) to obtain such results.

2.3. Testing rows or columns of two U-statistic based matrices

In some applications, instead of testing the equality of two full matrices, we are interested in testing the equality of a particular row or column of the given matrix pair. This requires us to test the hypothesis in (2.7). For simplicity, we only present the result for row comparison here. The application to the column comparison is straightforward.

To test the hypothesis in (2.7), we define the test statistic as

$$M_{n,i} = \max_{1\le j\le q} M_{ij}.$$

The following theorem derives the limiting distribution of $M_{n,i}$ under the null hypothesis.

Theorem 2.6. *If the null hypothesis* $\mathbf{H}_{0,i}$ *in* (2.7) *and conditions in Theorem* 2.2 *hold*, *we have*

$$\mathbb{P}(M_{n,i} - 2\log q + \log\log q \le x) \to \exp\left(-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right), \tag{2.19}$$

for any given $x \in \mathbb{R}$, *as* $n, q \to \infty$.

The above theorem can be proved in a similar way to Theorem 2.2.

Remark 2.7. For analyzing the power of $T_{\alpha,i}$, we define the following set of vector pairs,

$$\mathbb{A}_{i\star}(C) = \left\{(\mathbf{u}_{1,i\star}, \mathbf{u}_{2,i\star}) : \max_{1\le j\le q} \frac{|u_{1,ij} - u_{2,ij}|}{\sqrt{m^2\zeta_{1,ij}/n_1 + m^2\zeta_{2,ij}/n_2}} \ge C\sqrt{\log q}\right\}.$$

This allows us to yield a similar result to Theorem 2.4.

3. Applications to testing large Kendall's tau correlation matrix

In this section, we focus on testing the equality of two Kendall's tau matrices $\mathbf{U}_1^\tau$ and $\mathbf{U}_2^\tau$. This section contains two parts. In the first part, we assume the samples are from a fully nonparametric model. Under this model, in addition to the general Jackknife-based approach outlined in the previous section, we introduce two additional methods for testing (1.2) and analyze their theoretical properties (e.g., size and power). In the second part, we assume the samples are generated from a Gaussian copula model, under which we can relax Assumption **(A3)** to a much simplified form.

Kendall's tau provides a way to describe the nonlinear relationship between two random variables. As it is rank-based, it is especially suitable to analyze data from heavy-tailed or corrupted distributions. In this section, we aim to test the equality of two Kendall's tau matrices. More specifically, we set

$$\Phi_{ij}(\boldsymbol{X}_k, \boldsymbol{X}_\ell) := \operatorname{sign}(X_{ki} - X_{\ell i})\operatorname{sign}(X_{kj} - X_{\ell j}),$$
$$\Phi_{ij}(\boldsymbol{Y}_k, \boldsymbol{Y}_\ell) := \operatorname{sign}(Y_{ki} - Y_{\ell i})\operatorname{sign}(Y_{kj} - Y_{\ell j}),$$

and $q = d$. We aim to test whether $\mathbf{U}_1^\tau = \mathbf{U}_2^\tau$.

3.1. Methods and theory under fully nonparametric models

Section 3.1 contains two parts. The first part introduces two additional test procedures tailored for testing the equality of Kendall's tau matrices $\mathbf{U}_1^\tau$ and $\mathbf{U}_2^\tau$. The second part presents the theoretical properties of all the three tests. In addition, we further prove the rate-optimality of the proposed tests. Our technical contributions include providing an upper bound of the traditional plug-in variance estimation error, which enables us to establish the explicit rate of convergence of the plug-in variance estimator. We also prove an upper bound of the variance difference between two Kendall's tau correlation coefficients. These bounds allow us to derive the limiting distribution of the additional test statistic. The construction of these bounds requires the nontrivial usage of special structures of variance estimators and is of independent interest themselves. Moreover, for proving our test methods' optimality for the Kendall's tau matrix comparison, we construct a collection of least favourable multivariate normal distributions with regard to the test hypothesis. This novel construction technique is developed for correlation matrix comparison and is one of our technical contributions.

Recall that $\mathbf{U}_a^\tau$ and $\widehat{\mathbf{U}}_a^\tau$, defined in (2.3), are symmetric and we have $\mathrm{Diag}(\mathbf{U}_a^\tau) = \mathrm{Diag}(\widehat{\mathbf{U}}_a^\tau) = \mathbf{I}_d$. Therefore, we don't need to compare the main diagonals of $\mathbf{U}_a^\tau$. Hence, we reset $S = \{(i, j) : 1 \le i < j \le d\}$ for testing the equality of large Kendall's tau correlation matrices. The Jackknife-based statistic M_n in Section 2.1 then becomes

$$M_n^{\tau,\mathrm{jack}} := \max_{(i,j)\in S} \frac{(\widehat{\tau}_{1,ij} - \widehat{\tau}_{2,ij})^2}{\widehat{\sigma}^2(\widehat{\tau}_{1,ij}) + \widehat{\sigma}^2(\widehat{\tau}_{2,ij})}. \tag{3.1}$$

Here, we still use $\widehat{\sigma}^2(\cdot)$ to denote the Jackknife variance estimator. Accordingly, we obtain $\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$:

$$\mathrm{T}_\alpha^{\tau,\mathrm{jack}} := \mathbb{1}\big\{M_n^{\tau,\mathrm{jack}} \ge G^-(\alpha) + 4\log d - \log(\log d)\big\}.$$

3.1.1. *Three procedures to compare Kendall's tau matrices*

In this section, we present two additional methods for comparing two Kendall's tau matrices. We start with the introduction of a plug-in method, which directly estimates the variances of $\{\widehat{\tau}_{a,ij}\}_{a=1,2}$ and plugs them into the test statistic. For this, recall that the Kendall's tau sample correlation between two random variable U and V is set as

$$\widehat{\tau} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathrm{sign}(U_i - U_j)\,\mathrm{sign}(V_i - V_j),$$

where $U_1, \ldots, U_n$ and $V_1, \ldots, V_n$ are n random samples from U and V. Let Π_c be the probability of the event that among two members drawn from the sample without replacement, they are concordant with each other. In other words, we have

$$\Pi_c = \mathbb{P}\big((U_2 - U_1)(V_2 - V_1) > 0\big). \tag{3.2}$$

Kruskal [26] prove that the variance of $\widehat{\tau}$ can be written as

$$\frac{8}{n(n-1)}\Pi_c(1 - \Pi_c) + 16\frac{1}{n}\frac{n-2}{n-1}\big(\Pi_{cc} - \Pi_c^2\big), \tag{3.3}$$

where Π_{cc} is the probability of the event that among three members drawn from the sample without replacement, the second and third are concordant with the first. In other words, we have

$$\Pi_{cc} = \mathbb{P}\big(\big[(U_2 - U_1)(V_2 - V_1) > 0\big] \cap \big[(U_3 - U_1)(V_3 - V_1) > 0\big]\big). \tag{3.4}$$

As $n \to \infty$, the quantity in (3.3) multiplied by n has the limit $16(\Pi_{cc} - \Pi_c^2)$. Motivated by this result, we propose the following plug-in variance estimator

$$\widehat{\sigma}^2_{\text{plug}}(\widehat{\tau}) = \frac{16}{n}\big(\widehat{\Pi}_{cc} - \widehat{\Pi}_c^2\big), \tag{3.5}$$

as an alternative to the Jackknife based one for estimating the variance of $\widehat{\tau}$. Here $\widehat{\Pi}_{cc}$ and $\widehat{\Pi}_c$ are the corresponding U-statistics to estimate Π_{cc} and Π_c.[4] We replace $\widehat{\sigma}^2(\cdot)$ in $M_n^{\tau,\text{jack}}$ with $\widehat{\sigma}^2_{\text{plug}}(\cdot)$ to construct $M_n^{\tau,\text{plug}}$:

$$M_n^{\tau,\text{plug}} := \max_{1\le i<j\le d} \frac{(\widehat{\tau}_{1,ij} - \widehat{\tau}_{2,ij})^2}{\widehat{\sigma}^2_{\text{plug}}(\widehat{\tau}_{1,ij}) + \widehat{\sigma}^2_{\text{plug}}(\widehat{\tau}_{2,ij})}. \tag{3.6}$$

Accordingly, we construct the plug-in type test $\mathrm{T}_\alpha^{\tau,\text{plug}}$ as follows:

$$\mathrm{T}_\alpha^{\tau,\text{plug}} := \mathbb{1}\big\{M_n^{\tau,\text{plug}} \ge G^-(\alpha) + 4\log d - \log(\log d)\big\}.$$

In Section 3.1.2, we will provide the theoretical justification for this plug-in procedure.

Both the theoretical and numerical results indicate that the variance estimation error is also a key factor influencing the test statistics' powers. Up to now, we consider two kinds of variance estimation procedures (Jackknife based and plug-in based) for testing the equality of two Kendall's tau matrices. To exploit the sparsity of $\mathbf{U}^\tau$, we next propose to use the exact variance under the uncorrelated condition ($\tau = 0$). We name this procedure as "pseudo method". It calculates the variance of $\widehat{\tau}_{a,ij}$ by assuming $\tau_{a,ij} = 0$. We set $\widetilde{\sigma}^2_{1,\text{ps}}$ and $\widetilde{\sigma}^2_{2,\text{ps}}$ as the variances of $\sqrt{n_1}\widehat{\tau}_1$ and $\sqrt{n_2}\widehat{\tau}_2$, under $\tau_1 = 0$ and $\tau_2 = 0$. We also set

$$\sigma_{a,\text{ps}} := \lim_{n_a\to\infty} \widetilde{\sigma}_{a,\text{ps}} \qquad \text{for } a = 1, 2. \tag{3.7}$$

The test statistic becomes

$$M_n^{\tau,\text{ps}} := \max_{1\le i<j\le d} \frac{(\widehat{\tau}_{1,ij} - \widehat{\tau}_{2,ij})^2}{\sigma^2_{1,\text{ps}}/n_1 + \sigma^2_{2,\text{ps}}/n_2}. \tag{3.8}$$

Similarly, we construct the test $\mathrm{T}_\alpha^{\tau,\text{ps}}$:

$$\mathrm{T}_\alpha^{\tau,\text{ps}} := \mathbb{1}\big\{M_n^{\tau,\text{ps}} \ge G^-(\alpha) + 4\log d - \log(\log d)\big\}. \tag{3.9}$$

For example, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are generated from continuous Gaussian copula model, we have

$$\widetilde{\sigma}^2_{1,\text{ps}} = \frac{2(2n_1+5)}{9(n_1-1)}, \qquad \widetilde{\sigma}^2_{2,\text{ps}} = \frac{2(2n_2+5)}{9(n_2-1)} \quad \text{and} \quad \sigma^2_{1,\text{ps}} = \sigma^2_{2,\text{ps}} = \frac{4}{9}.$$

[4] By the definition of Π_{cc} in (3.4), we should build a U-statistic with an asymmetric kernel to estimate it.

Remark 3.1. As long as $|\widetilde{\sigma}^2_{a,\mathrm{ps}} - \sigma^2_{a,\mathrm{ps}}| = o((\log d)^{-1-\epsilon})$ with an arbitrary $\epsilon > 0$, we can show that replacing $\sigma^2_{a,\mathrm{ps}}$ with $\widetilde{\sigma}^2_{a,\mathrm{ps}}$ still gives a valid test. Details are provided in the proof of Theorem 3.3.

3.1.2. *Theoretical properties of three testing procedures*

We now present the theoretical properties (size, power, and optimality) of the three tests introduced in the former sections. More specifically, we prove their validity under the null hypothesis and conduct power analysis similarly to Theorems 2.2 and 2.4. Furthermore, we show that these tests are rate optimal against the sparse alternative.

In the beginning, the following theorem gives the limiting distribution for plug-in and Jackknife based test statistics.

Theorem 3.2. *Assuming* **(A1)**, **(A2)** *and* **(A3)** *hold*, *under* $\mathbf{H}_0^\tau$ *of* (1.2), *we have*

$$\mathbb{P}\big(M_n^{\tau,\mathrm{jack}} - 4\log d + \log(\log d) \le x\big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right), \tag{3.10}$$

$$\mathbb{P}\big(M_n^{\tau,\mathrm{plug}} - 4\log d + \log(\log d) \le x\big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right), \tag{3.11}$$

for any $x \in \mathbb{R}$, *as* $n, d \to \infty$. *Furthermore*, *the results hold uniformly for all* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *satisfying* **(A1)**, **(A2)** *and* **(A3)**.

The following theorem gives the limiting distribution of the pseudo method. It holds under an additional meta-elliptical (defined in Supplement E of Supplementary Material(Zhou et al. [46])) distributional assumption on the data.

Theorem 3.3. *We assume that* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *belong to the meta-elliptical distribution* (*Fang*, *Fang and Kotz* [14]).[5] *If Assumptions* **(A1)**, **(A2)** *and* **(A3)** *hold*, *under* $\mathbf{H}_0^\tau$ *in* (1.2), *we have*

$$\mathbb{P}\big(M_n^{\tau,\mathrm{ps}} - 4\log d + \log(\log d) \le x\big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right), \tag{3.12}$$

for any $x \in \mathbb{R}$, *as* $n, d \to \infty$. *Furthermore*, *the result holds uniformly for all* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *satisfying* **(A1)**, **(A2)**, **(A3)**.

We now analyze the powers of $\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$, $\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$ and $\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$. Similarly to Theorem 2.4, we define

$$\mathbb{U}(C) = \left\{\big(\mathbf{U}_1^\tau, \mathbf{U}_2^\tau\big) : \max_{1\le i<j\le d} \frac{|\tau_{1,ij} - \tau_{2,ij}|}{\sqrt{4\zeta_{1,ij}/n_1 + 4\zeta_{2,ij}/n_2}} \ge C\sqrt{\log d}\right\},$$

[5]Detailed introduction of the meta-elliptical distribution family is provided in Supplement E of Supplementary Material (Zhou et al. [46]).

$$\mathbb{V}(C)=\left\{\left(\mathbf{U}_1^\tau,\mathbf{U}_2^\tau\right):\max_{1\le i<j\le d}\frac{|\tau_{1,ij}-\tau_{2,ij}|}{\sqrt{\sigma_{1,\mathrm{ps}}^2/n_1+\sigma_{2,\mathrm{ps}}^2/n_2}}\ge C\sqrt{\log d}\right\}.$$

They are Kendall's tau versions of $\mathbb{A}(C)$ in Theorem 2.4.

Theorem 3.4. (*Power Analysis*) *Assuming* **(A2)** *holds*, *we have*

$$\inf_{(\mathbf{U}_1^\tau,\mathbf{U}_2^\tau)\in\mathbb{U}(4)}\mathbb{P}\left(\mathrm{T}_\alpha^{\tau,\mathrm{jack}}=1\right)\to 1, \tag{3.13}$$

$$\inf_{(\mathbf{U}_1^\tau,\mathbf{U}_2^\tau)\in\mathbb{U}(4)}\mathbb{P}\left(\mathrm{T}_\alpha^{\tau,\mathrm{plug}}=1\right)\to 1, \tag{3.14}$$

as $n,d\to\infty$. *If* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *belong to the meta-elliptical family and* **(A1)**, **(A2)** *are satisfied*, *as* $n,d\to\infty$, *we have*

$$\inf_{(\mathbf{U}_1^\tau,\mathbf{U}_2^\tau)\in\mathbb{V}(4)}\mathbb{P}\left(\mathrm{T}_\alpha^{\tau,\mathrm{ps}}=1\right)\to 1. \tag{3.15}$$

Theorem 3.4 implies that just one entry of $\mathbf{U}_1^\tau-\mathbf{U}_2^\tau$ has a magnitude no smaller than $C\sqrt{\log d/n}$ is enough for the introduced tests to correctly reject $\mathbf{H}_0^\tau$.

Next, we show that all the three proposed methods are rate optimal by matching the obtained rates of convergence to a lower bound for correlation matrix comparison. We adopt the general framework used in Baraud [7] to obtain the lower bound for testing the equality of correlation matrices. The core of the proof is the construction of collections of least favourable multivariate normal distributions with regard to the test hypothesis. Our work is related to Cai, Liu and Xia [10] which prove the lower bound for testing the equality of covariance matrices. However, their construction technique is developed for covariance matrices but not the correlation matrices. Specifically, they only perturb the diagonal elements of the covariance, which does not affect the resulting correlation matrices. To test correlation matrices, we need to develop a novel construction by perturbing the off-diagonal elements of the correlation matrices. Details are provided in the proof of Theorem 3.5.

Theorem 3.5. *Let* $\alpha,\beta>0$ *and* $\alpha+\beta<1$. *Assuming that* $\log d/n=o(1)$, *there exits a sufficiently small positive number* c_0, *such that for any distribution family that contains Gaussian as a subfamily, and all large enough* n *and* d, *we have*

$$\inf_{(\mathbf{U}_1^\tau,\mathbf{U}_2^\tau)\in\mathbb{U}(c_0)}\sup_{T_\alpha\in\mathcal{T}_\alpha}\mathbb{P}(T_\alpha=1)\le 1-\beta, \tag{3.16}$$

where $\mathcal{T}_\alpha$ *represents all level* α *tests for testing the equality of two correlation matrices.*

Cai, Liu and Xia [10] give a similar result for testing the equality of two covariance matrices. They show that the rate $C\sqrt{\log d/n}$ is optimal for comparing covariance matrices under conditions that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have sub-Gaussian-type or polynomial-type tails. In comparison, the lower bound result in Theorem 3.5 illustrates that our proposed methods are rate optimal under the fully nonparametric model. In particular, we don't impose assumptions on the marginal distributions.

3.2. Methods and theory under semiparametric Gaussian copula models

In this section, we assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-dimensional random vectors from the Gaussian copula with latent correlation matrices $\mathbf{\Sigma}_a = (\sigma_{a,ij})$, $a = 1, 2$ and $\mathrm{Diag}(\mathbf{\Sigma}_a) = \mathbf{I}_d$.[6] Under the Gaussian copula model, the technical assumption **(A3)** in Section 2.2 can be replaced by a much simplified condition. Specifically, for $r \in (0, 1)$, we define

$$\Omega(r) := \left\{1 \le i \le d : |\tau_{1,ij}| > r \,\mathrm{or}\, |\tau_{2,ij}| > r \text{ for some } j \ne i\right\}. \tag{3.17}$$

We describe the technical assumption **(A4)** as follows:

(A4) For some $r < 1$ and a sequence of numbers $\Omega_{d,r} = o(d)$, we have $|\Omega(r)| \le \Omega_{d,r}$.

After introducing Assumption **(A4)**, we then discuss its relationship with Assumptions **(A1)** and **(A3)**. For **(A1)**, although it has similar form to **(A4)**, they are essentially different. Assumption **(A1)** is related to the largest eigenvalues of $\mathbf{U}_\gamma^\tau$. In fact, bounded $\lambda_{\max}(\mathbf{U}_\gamma^\tau)$ implies $\max_{1\le j\le d} \mathrm{supp}_j(\alpha_0) \le C(\log d)^{2+2\alpha_0}$. On the contrary, Assumption **(A4)** is related to $\lambda_{\min}(\mathbf{U}_\gamma^\tau)$. For example, if the correlation between two Gaussian random variables goes to 1, the corresponding correlation matrix will be asymptotically degenerated with the least eigenvalue infinite small. In proof, we first use Assumption **(A4)** to select largest sub-matrix of $\mathbf{U}_\gamma^\tau$ so that all its entries' absolute values are less than r. We then use Assumption **(A1)** to exclude the influence of entries with $|\tau_{\gamma,ij}| \ge (\log d)^{-1-\alpha_0}$ on the asymptotic results.

Assumptions **(A4)** and **(A3)** are highly related. However, Assumption **(A3)** cannot be straightforwardly implied by Assumptions **(A1)**, **(A2)** and **(A4)**. In fact, the relationship between **(A3)** and **(A4)** is complicated. To see the exact relationship, we need some additional definitions.

First, we have $S = \{(i, j) : 1 \le i < j \le d\}$. We then define

$$C_0 = \left\{(i, j) : i \in \Omega(r) \cup \Gamma\right\} \cup \left\{(i, j) : j \in \Omega(r) \cup \Gamma\right\} \quad \text{and} \quad B_0 = S_0 \cup C_0,$$

where Γ is defined in in Assumption **(A1)** and S_0 is defined in (2.11). Furthermore, we denote A to be the biggest subset of $S \setminus B_0$, such that any two pairs $(i, j) \ne (k, \ell) \in A$ must satisfy a condition ($\star$). More detailed description of condition ($\star$) will be provided in the proof of Theorem 3.6. Essentially, it specifies that, for any $(i, j) \ne (k, \ell) \in S \setminus B_0$, there exits an $i_1 \in \{i, j, k, \ell\}$ such that for any $j_1 \in \{i, j, k, \ell\} \setminus i_1$, we have $|\tau_{a,i_1 j_1}| = O((\log d)^{-1-\alpha_0})$. We also define $\tau_{a,ijk\ell}$ as the Kendall's tau version of $u_{a,ijk\ell}$ in Assumption **(A3)**.

Under Assumptions **(A1)**, **(A2)** and **(A4)**, we can prove that for any $(i, j) \ne (k, \ell) \in A$, we have $|\tau_{a,ijk\ell}| = O((\log d)^{-1-\alpha_0})$, which is essentially Assumption **(A3)** with $u_{a,ijk\ell}$ replaced by $\tau_{a,ijk\ell}$. The only difference is that these conditions hold on A but instead of $S \setminus S_0$ as in Assumption **(A3)**. Theorem 3.6 below specifics that Assumptions **(A1)**, **(A2)** and **(A4)** can be used to replace Assumptions **(A1)**, **(A2)** and **(A3)** when we test the equality of Kendall's tau correlation matrices under the Gaussian copula model.

[6]Detailed definition of the Gaussian copula is put in Supplement E of Supplementary Material (Zhou et al. [46]).

Theorem 3.6. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be Gaussian copula random vectors with latent correlation matrices $\boldsymbol{\Sigma}_a$, $a = 1$ or 2 and $\mathrm{Diag}(\boldsymbol{\Sigma}_a) = \mathbf{I}_d$. We assume that the smallest eigenvalue of any 4 by 4 principal sub-matrix of $\boldsymbol{\Sigma}_a$ is uniformly bounded away from 0. Assuming* **(A1)**, **(A2)** *and* **(A4)** *hold, under $\mathbf{H}_0^\tau$ of* (1.2), *we have*

$$\mathbb{P}\big(M_n^{\tau,\mathrm{jack}} - 4\log d + \log(\log d) \le x\big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right),$$
$$\mathbb{P}\big(M_n^{\tau,\mathrm{plug}} - 4\log d + \log(\log d) \le x\big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right),$$
$$\mathbb{P}\big(M_n^{\tau,\mathrm{ps}} - 4\log d + \log(\log d) \le x\big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right),$$

for any $x \in \mathbb{R}$, as $n, d \to \infty$. Furthermore, these limiting results hold uniformly for all $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfying **(A1)**, **(A2)** *and* **(A4)**.

Proof. Recall that $M_n^{\tau,\mathrm{jack}}$, $M_n^{\tau,\mathrm{plug}}$ and $M_n^{\tau,\mathrm{ps}}$ in (3.1), (3.6) and (3.8) are defined by taking maximum over S. The main idea is to show that it is sufficient to use a version of these quantities taking the maximum over the smaller set A as defined before. The proof is technical and left to Supplement A.6 of Supplementary Material (Zhou et al. [46]). □

Remark 3.7. In Supplement E of Supplementary Material (Zhou et al. [46]), we show that $\mathbf{U}_a^\tau$ and $\boldsymbol{\Sigma}_a$ are related in terms of $\sigma_{a,ij} = \sin(\tau_{a,ij}\pi/2)$. Hence, testing (1.2) is equivalent to testing

$$\mathbf{H}_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \quad \text{v.s.} \quad \mathbf{H}_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2,$$

under the Gaussian copula model.

Remark 3.8. To test the row or column of Kendall's tau matrices, if any of the conditions of Theorems 3.2, 3.3 and 3.6 hold, we get the same limiting result as in (2.19).

4. Experiments

In this section, we demonstrate numerical performances of proposed methods on simulated and real data sets. In particular, we compare proposed methods with the state-of-the-art method in the literature.

4.1. Numerical simulations

We compare proposed methods with the sample covariance based method (denoted by $\mathrm{T}_\alpha^{\mathrm{CLX}}$) in Cai, Liu and Xia [10]. To test our methods under various covariance structures, we introduce the following matrices.

- (Block matrix $\mathbf{\Sigma}^*$) Let $\mathbf{R}^* = (r^*_{ij}) \in \mathbb{R}^{d\times d}$ with $r^*_{ij} = 0.6$ for $5(k-1)+1 \le i \ne j \le 5k$ and $k = 1, \dots, \lfloor d/5 \rfloor$. For other entries in $\mathbf{R}^*$, we set $r^*_{ii} = 1$ and $r^*_{ij} = 0$ when $i \ne j$. Let $\mathbf{D}$ as a diagonal matrix with each nonzero entry following independent uniform distribution on the interval $(0.5, 1.5)$. We then set $\mathbf{\Sigma}^* = \mathbf{D}\mathbf{R}^*\mathbf{D}$.
- (Tridiagonal matrix $\mathbf{\Sigma}'$) Let $\mathbf{R}' = (r'_{ij}) \in \mathbb{R}^{d\times d}$ be a tridiagonal matrix with 1 on the main diagonal and 0.5 on the first diagonal. We then set $\mathbf{\Sigma}' = \mathbf{D}\mathbf{R}'\mathbf{D}$.
- (Multidiagonal matrix $\mathbf{\Sigma}^\star$) Let $\mathbf{R}^\star = (r^\star_{ij}) \in \mathbb{R}^{d\times d}$ with $r_{ij} = 0.8^{|i-j|}$ and $\mathbf{\Sigma}^\star = \mathbf{D}\mathbf{R}^\star\mathbf{D}$.

Under the null hypothesis, we sample $n_1 + n_2$ data points from the following 3 models with $\mathbf{\Sigma} = \mathbf{\Sigma}^*$, $\mathbf{\Sigma}'$, and $\mathbf{\Sigma}^\star$.

- **Model 1** (Normal distribution) In this model, under the null hypothesis we generate $n_1 + n_2$ random vectors from $N(\mathbf{0}, \mathbf{\Sigma})$.
- **Model 2** (Multivariate t distribution) We sample from $\boldsymbol{\mu} + \boldsymbol{Z}/\sqrt{W/\nu}$ with $W \sim \chi^2(\nu)$ and $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbf{\Sigma})$, where W and $\boldsymbol{Z}$ are independent. Under the null hypothesis, we generate $n_1 + n_2$ data points with $\boldsymbol{\mu} = \mathbf{0}$ and $\nu = 3$.
- **Model 3** (Marginal Cauchy distribution) Generate $n_1 + n_2$ random vectors from $N(\mathbf{0}, \mathbf{\Sigma})$. We then use a monotone function to transform each coordinate to follow the Cauchy distribution Cauchy(μ, s) whose density function is $s/\pi(s^2 + (x-\mu)^2)$. In the simulation, we set $\mu = 0$ and $s = 1$.

Under above models, the two populations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same covariance matrices. We use them to show that our proposed methods can control the size correctly under the null hypothesis. For the power analysis, we introduce a random symmetric matrix $\mathbf{\Delta} = (\delta_{k\ell}) \in \mathbb{R}^{d\times d}$ with exactly 8 nonzero entries. Among the 8 entries, 4 entries are randomly selected from the upper triangle of $\mathbf{\Delta}$, with a magnitude generated from the uniform distribution on $(0, \zeta\sigma^2_{\max})$, where $\sigma^2_{\max}$ is the maximal value of $\mathbf{\Sigma}$'s main diagonal. Other 4 entries are determined by symmetry. We then set $\widetilde{\mathbf{\Sigma}}_1 = \mathbf{\Sigma} + \delta\mathbf{I}$ and $\widetilde{\mathbf{\Sigma}}_2 = \mathbf{\Sigma} + \mathbf{\Delta} + \delta\mathbf{I}$ with $\delta = |\min\{\lambda_{\min}(\mathbf{\Sigma} + \mathbf{\Delta}), \lambda_{\min}(\mathbf{\Sigma})\}| + 0.05$. In place of $\mathbf{\Sigma}$, we use the matrices $\widetilde{\mathbf{\Sigma}}_1$ and $\widetilde{\mathbf{\Sigma}}_2$ to generate samples for $\boldsymbol{X}$ and $\boldsymbol{Y}$ under the alternative hypothesis.

We set $n_1 = n_2 = n$ with $n = 200, 500$ and $d = 50, 100, 200, 300, 500, 700, 1000$. The nominal significance level α is 0.05. Table 1 presents empirical sizes. We see that $\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$ always attains the desired size even for extremely large d. When d is significantly larger than n, both $\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$ and $\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$ suffer from the size distortion. When d approximates n, $\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$ is still valid but $\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$ fails. These size distortions decrease as n increases. Although the theoretical limiting results are similar for all the proposed methods, the simulation results show that the estimation errors of variance heavily affect the proposed tests' finite sample performances, and $\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$ benefits a lot from avoiding estimating the variance directly. Moreover, for heavy tail distributions such as multivariate t and Cauchy distributions, we also see that $\mathrm{T}_\alpha^{\mathrm{CLX}}$ from Cai, Liu and Xia [10] becomes too conservative.

By examining the empirical powers in Table 2, for distributions with heavy tails or strong tail dependence, $\mathrm{T}_\alpha^{\mathrm{CLX}}$'s power decreases dramatically, making $\mathrm{T}_\alpha^{\mathrm{CLX}}$ inappropriate for such applications. These finite sample results also suggest that among three proposed methods $\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$ is most aggressive and $\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$ is most conservative.

These finite sample (with n around several hundreds) results suggest that $\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$ is useful only when d is smaller than n. With d approximates n, we recommend to use $\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$ because it has averagely higher power. When d is significantly larger than n, $\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$ is recommended because of its good size control.

4.2. Real data example

In this section, we use proposed methods to analyze the dependence structure of brain activity. We use the resting-state functional magnetic resonance imaging (fMRI) data of normal children and diseased children with the disease attention deficit hyperactivity disorder (ADHD). Functional neuroimaging studies have revealed abnormalities in various brain regions of ADHD patients (Lou, Henriksen and Bruhn [30], Giedd et al. [15], Shafritz et al. [39], Yufeng et al. [44], Zou et al. [48]). As a marker of brain activity, amplitude of low-frequency fluctuation (ALFF) is a powerful tool to investigate this disorder. ALFF is the total power within the frequency range between 0.01 and 0.1 Hz of the fMRI time series. Generally speaking, it captures average slow fluctuations of brain activity. For the detailed definition of ALFF, we refer to Yufeng et al. [44]. Existing literature suggests the existence of significant differences in mean values of ALFF between the normal and diseased children (Zou et al. [48]). By using our methods, we aim to test the dependence structure of ALFF between brain regions. Considering the nonlinear relationship and robustness, we use Kendall's tau matrix to measure the dependence structure.

We then introduce our data processing procedure. We use the standard methodology of software C-PAC[7] to correct body motion, brain heterogeneity, and many other kinds of measure errors. We then calculate voxel-wise ALFF of each person's fMRI images. As the voxel number is very large ($61 \times 73 \times 61$ for 3 mm brain template), to limit the number of testing parameters, a common approach is to extract signals from specified regions of interest (ROIs) based on the anatomical structure of brain. In our experiments, we combine two kinds of brain areas including Brodmann (BA) and automated anatomical labeling (AAL) on the gray matter to build new 227 brain regions. In each brain region, we average obtained voxel-wise ALFF to get data points with the dimension $d = 227$.

After the introduction of data processing, we then describe the data set in detail. The resting-state fMRI data for ADHD is available on the Internet.[8] Considering the dimension ($d = 227$) of data points, we use samples from Peking University and Kennedy Krieger Institute of Johns Hopkins University to build a sample with 119 ADHD patients and 200 control members.

In the application, we test both mean vectors and Kendall's tau matrices between the diseased and normal groups and show the results in Table 3. In the context of high-dimensional mean tests, we use three existing methods: $\mathrm{T}_\alpha^{\mathrm{Bai}}$ in Bai and Yin [5], $\mathrm{T}_\alpha^{\mathrm{Sri}}$ in Srivastava and Du [41], and $\mathrm{T}_\alpha^{\mathrm{Cai}}$ in Cai, Liu and Xia [11]. Except for the known mean differences, the results for Kendall's tau matrices also suggest that the dependence structure of brain activities for ADHD patients are also very different from normal children, which is worth investigating for related researchers.

[7] See the website http://fcp-indi.github.io/docs/user/.
[8] See the website http://fcon_1000.projects.nitrc.org/indi/adhd200/.

Table 1. Empirical sizes of **Model 1**, **2** and **3** under $\alpha = 0.05$ based on 2000 repetitions

		$\boldsymbol{\Sigma}^*$							$\boldsymbol{\Sigma}'$							$\boldsymbol{\Sigma}^\star$						
n	d	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000
		Model 1																				
500	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.05	0.05	0.07	0.07	0.07	0.08	0.09	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.05	0.05	0.05	0.06	0.06	0.08	0.07
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.05	0.05	0.06	0.06	0.06	0.06	0.07	0.05	0.05	0.05	0.05	0.06	0.06	0.08	0.04	0.05	0.05	0.05	0.05	0.06	0.06
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.04	0.04	0.05	0.04	0.05	0.05
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.04	0.04	0.04	0.05	0.04	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.04	0.05	0.03	0.03	0.05	0.05	0.05	0.06	0.06
		Model 2																				
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.05	0.06	0.07	0.07	0.08	0.08	0.09	0.05	0.05	0.05	0.06	0.06	0.08	0.08
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.05	0.05	0.05	0.06	0.06	0.06	0.07	0.05	0.05	0.06	0.06	0.07	0.07	0.08	0.05	0.05	0.05	0.05	0.06	0.06	0.06
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.04	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.03	0.04	0.04	0.04	0.05	0.05	0.05
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Model 3																				
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.05	0.05	0.06	0.07	0.08	0.08	0.09	0.05	0.05	0.05	0.06	0.06	0.08	0.08
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.05	0.05	0.05	0.06	0.06	0.06	0.07	0.05	0.05	0.05	0.06	0.07	0.07	0.08	0.04	0.04	0.04	0.05	0.05	0.06	0.06
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.03	0.03	0.04	0.04	0.04	0.05	0.05
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.00	0.01	0.02	0.05	0.08	0.16	0.22	0.00	0.01	0.02	0.05	0.10	0.16	0.25	0.00	0.01	0.02	0.04	0.08	0.15	0.23

Table 1. (Continued)

		$\boldsymbol{\Sigma}^*$							$\boldsymbol{\Sigma}'$							$\boldsymbol{\Sigma}^\star$						
n	d	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000
											Model 1											
200	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.06	0.08	0.09	0.12	0.14	0.16	0.19	0.06	0.08	0.11	0.13	0.15	0.16	0.19	0.05	0.06	0.08	0.09	0.12	0.15	0.14
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.05	0.05	0.06	0.08	0.10	0.10	0.12	0.05	0.06	0.07	0.09	0.10	0.11	0.12	0.03	0.05	0.05	0.07	0.08	0.10	0.10
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.05	0.06	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.04
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.04	0.04	0.04	0.04	0.05	0.05	0.06	0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.04	0.04	0.05	0.05	0.06	0.05	0.06
											Model 2											
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.06	0.08	0.09	0.12	0.14	0.16	0.19	0.06	0.08	0.11	0.13	0.15	0.16	0.20	0.05	0.06	0.08	0.09	0.12	0.15	0.15
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.05	0.05	0.06	0.08	0.10	0.10	0.12	0.05	0.06	0.07	0.09	0.10	0.11	0.12	0.03	0.05	0.05	0.07	0.08	0.10	0.10
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.05	0.06	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.05
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
											Model 3											
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.06	0.08	0.09	0.12	0.14	0.16	0.19	0.06	0.08	0.11	0.13	0.15	0.16	0.20	0.05	0.06	0.08	0.09	0.12	0.15	0.15
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.05	0.05	0.06	0.08	0.10	0.10	0.12	0.05	0.06	0.07	0.09	0.10	0.11	0.12	0.03	0.05	0.05	0.07	0.08	0.10	0.10
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.05	0.06	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.05
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.00	0.01	0.02	0.06	0.12	0.23	0.37	0.00	0.00	0.02	0.07	0.14	0.26	0.37	0.00	0.01	0.01	0.08	0.15	0.28	0.39

Table 2. Empirical powers of **Model 1**, **2** and **3** under $\alpha = 0.05$ based on 2000 repetitions. We set $\zeta = 0.2$ for $n = 500$ and $\zeta = 0.3$ for $n = 200$

		$\mathbf{\Sigma}^*$							$\mathbf{\Sigma}'$							$\mathbf{\Sigma}^\star$						
n	d	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000
		Model 1																				
500	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.85	0.81	0.76	0.76	0.76	0.74	0.71	0.86	0.82	0.78	0.76	0.72	0.72	0.71	0.86	0.81	0.75	0.74	0.71	0.70	0.67
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.84	0.80	0.75	0.75	0.75	0.73	0.70	0.86	0.81	0.77	0.76	0.72	0.71	0.70	0.87	0.81	0.75	0.73	0.70	0.69	0.64
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.83	0.79	0.74	0.73	0.73	0.71	0.68	0.85	0.81	0.75	0.74	0.70	0.68	0.67	0.85	0.79	0.72	0.70	0.67	067	0.61
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.83	0.78	0.74	0.74	0.72	0.72	0.67	0.86	0.80	0.77	0.75	0.71	0.68	0.66	0.80	0.77	0.71	0.70	0.66	0.66	0.60
		Model 2																				
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.78	0.72	0.70	0.68	0.66	0.65	0.62	0.77	0.72	0.68	0.66	0.63	0.63	0.60	0.79	0.71	0.63	0.62	0.58	0.58	0.55
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.77	0.72	0.69	0.68	0.67	0.66	0.61	0.76	0.72	0.67	0.66	0.62	0.61	0.59	0.79	0.70	0.64	0.62	0.57	0.57	0.54
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.76	0.70	0.67	0.66	0.65	0.64	0.61	0.75	0.69	0.66	0.63	0.60	0.60	0.58	0.75	0.67	0.58	0.57	0.53	0.52	0.52
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.12	0.09	0.07	0.06	0.03	0.02	0.02	0.09	0.06	0.05	0.04	0.02	0.02	0.02	0.07	0.04	0.03	0.02	0.01	0.01	0.01
		Model 3																				
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.84	0.81	0.78	0.77	0.74	0.72	0.71	0.85	0.81	0.80	0.78	0.72	0.72	0.71	0.87	0.82	0.74	0.75	0.70	0.70	0.68
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.84	0.80	0.78	0.76	0.73	0.71	0.70	0.85	0.81	0.78	0.77	0.71	0.70	0.68	0.87	0.82	0.77	0.74	0.69	0.69	0.65
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.82	0.80	0.75	0.75	0.72	0.69	0.68	0.84	0.81	0.77	0.75	0.70	0.69	0.67	0.85	0.79	0.73	0.72	0.66	0.66	0.61
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.00	0.01	0.02	0.04	0.08	0.16	0.23	0.00	0.00	0.04	0.05	0.10	0.15	0.24	0.01	0.01	0.03	0.04	0.09	0.16	0.25

Table 2. (Continued)

		$\mathbf{\Sigma}^*$							$\mathbf{\Sigma}'$							$\mathbf{\Sigma}^\star$						
n	d	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000	50	100	200	300	500	700	1000
											Model 1											
200	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.86	0.81	0.79	0.78	0.76	0.74	0.73	0.80	0.76	0.74	0.71	0.70	0.68	0.68	0.81	0.72	0.66	0.64	0.61	0.58	0.56
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.85	0.80	0.79	0.73	0.73	0.71	0.69	0.79	0.74	0.71	0.67	0.66	0.64	0.63	0.79	0.70	0.62	0.60	0.56	0.50	0.52
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.82	0.77	0.75	0.68	0.68	0.64	0.60	0.77	0.71	0.66	0.63	0.61	0.57	0.55	0.75	0.64	0.56	0.54	0.47	0.43	0.42
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.76	0.72	0.68	0.62	0.60	0.53	0.50	0.73	0.68	0.62	0.58	0.51	0.50	0.47	0.63	0.58	0.50	0.46	0.39	0.32	0.31
											Model 2											
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.79	0.74	0.72	0.67	0.62	0.60	0.59	0.73	0.66	0.65	0.61	0.58	0.56	0.56	0.70	0.60	0.54	0.48	0.46	0.45	0.45
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.78	0.71	0.70	0.64	0.59	0.55	0.55	0.72	0.64	0.61	0.58	0.55	0.52	0.51	0.69	0.58	0.50	0.46	0.44	0.40	0.39
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.73	0.64	0.63	0.58	0.53	0.50	0.50	0.69	0.58	0.57	0.53	0.49	0.46	0.45	0.62	0.51	0.42	0.37	0.35	0.34	0.34
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.07	0.04	0.02	0.01	0.01	0.01	0.00	0.07	0.04	0.02	0.01	0.00	0.00	0.00	0.03	0.02	0.01	0.00	0.00	0.00	0.00
											Model 3											
	$\mathrm{T}_\alpha^{\tau,\mathrm{plug}}$	0.85	0.82	0.80	0.78	0.76	0.74	0.73	0.80	0.76	0.74	0.71	0.70	0.68	0.68	0.81	0.73	0.66	0.64	0.61	0.58	0.56
	$\mathrm{T}_\alpha^{\tau,\mathrm{jack}}$	0.85	0.80	0.79	0.75	0.73	0.71	0.69	0.79	0.74	0.72	0.68	0.66	0.64	0.63	0.79	0.71	0.62	0.60	0.56	0.51	0.52
	$\mathrm{T}_\alpha^{\tau,\mathrm{ps}}$	0.82	0.78	0.75	0.70	0.68	0.64	0.61	0.77	0.71	0.66	0.61	0.57	0.57	0.55	0.75	0.65	0.56	0.54	0.47	0.44	0.42
	$\mathrm{T}_\alpha^{\mathrm{CLX}}$	0.00	0.00	0.03	0.07	0.16	0.22	0.37	0.00	0.01	0.02	0.06	0.15	0.26	0.37	0.00	0.00	0.00	0.06	0.16	0.26	0.37

Table 3. Region based two-sample tests of ALFF between ADHD patients and control members

	Mean vectcor			Kendall's tau matrix		
	$\mathrm{T}_{\alpha}^{\mathrm{Bai}}$	$\mathrm{T}_{\alpha}^{\mathrm{Sri}}$	$\mathrm{T}_{\alpha}^{\mathrm{Cai}}$	$\mathrm{T}_{\alpha}^{\mathrm{plug}}$	$\mathrm{T}_{\alpha}^{\mathrm{jack}}$	$\mathrm{T}_{\alpha}^{\mathrm{CLX}}$
Test statistics	2.8358	2.5978	22.7085	25.892	24.371	23.572
P-values	0.0023	0.0047	0.0006	0.0105	0.0223	0.0330

5. Summary and discussion

This paper considers the problem of testing the equality of high-dimensional U-statistic based matrices. We provide a lower bound for testing the equality of correlation matrices and prove the proposed methods' optimality. Based on thorough numerical comparisons, $\mathrm{T}_{\alpha}^{\mathrm{plug}}$ performs well only when d is significantly smaller than n. When d is very large, we recommend to use $\mathrm{T}_{\alpha}^{\mathrm{ps}}$ for correctly controlling the size. In addition, $\mathrm{T}_{\alpha}^{\mathrm{ps}}$ performs quite well for distributions with heavy tails or strong tail dependence. Therefore, $\mathrm{T}_{\alpha}^{\mathrm{ps}}$ is potentially more useful for financial applications in which heavy-tailness is a common phenomenon. There are many possible future directions of this work. For example, instead of two-sample problems, it is interesting to generalize the idea to k-sample testing problems ($k > 2$). This may require a nontrivial extension of theoretical analysis.

For testing Kendall's tau matrices, we show that the variance estimation error is a key factor influencing a test procedure's power. In fact, the test $\mathrm{T}_{\alpha}^{\mathrm{ps}}$, which exploits the exact value of variance under the uncorrelated condition ($\tau = 0$), achieves a better finite-sample performance especially when d is very large. We can generalize such idea to many other applications. We also provide an upper bound of the Jackknife variance estimation error in the proof of Theorem 2.2. This result is also useful for other properties of U-statistics.

Next, we discuss the imposed assumptions. We note that the sparsity assumption **(A1)** plays a key role for obtaining the limiting extreme value distribution. It is not clear on whether this assumption is necessary, but it is satisfied in many high-dimensional applications. When **(A1)** is not satisfied, it is possible to exploit the bootstrap method to construct a test statistic. This is left as for future investigation. Regarding **(A2)**, we note that Cai, Liu and Xia [10] assume a stronger scaling assumption: $\log(d) = o(n^{1/5})$. We strengthen this scaling by assuming $\log(d) = O(n^{1/3-\epsilon})$ for an arbitrary $\epsilon > 0$. This is from the fact that U-statistics studied in this paper are assumed to have bounded kernels.

In the simulation studies, we use $\mathrm{T}_{\alpha}^{\mathrm{CLX}}$ as a comparison benchmark. In Supplement F of Supplementary Material (Zhou et al. [46]), we provide another heuristic test (denoted by $\mathrm{T}_{\alpha}^{\mathrm{R}}$) for testing the equality of Pearson's correlation matrices. The performances of $\mathrm{T}_{\alpha}^{\mathrm{CLX}}$ and $\mathrm{T}_{\alpha}^{\mathrm{R}}$ are similar for off diagonal disturbances.

Appendix A: The proof of main theorem

This appendix contains the proof of main theorem, that is, Theorem 2.2. In the sequel, we use C, C_1, C_2, $\dots$, to denote constants that do not depend on n, d, q and they can vary from place to place.

Proof. As explained in the sketch of proof, our analysis proceeds in three steps.

Step (i). In this step, we prove that it is sufficient to establish (2.14) for proving the theorem. For this, we need to sharply characterize the estimation error of the Jackknife variance estimator of U-statistics. For this, we introduce the following lemma.

Lemma A.1. *Let $\widehat{\sigma}^2(\widehat{u}_{a,ij})$ be the Jackknife estimator of $\widehat{u}_{a,ij}$ and $\sigma^2(\widehat{u}_{a,ij})$ be the variance of $\widehat{u}_{a,ij}$. Recalling the definition of h_{ij} and $\zeta_{a,ij}$ in (2.9) and (2.10), $\zeta_{1,ij}$ and $\zeta_{2,ij}$ are the variances of $h_{ij}(\boldsymbol{X}_\ell)$ and $h_{ij}(\boldsymbol{Y}_\ell)$. We have that $m^2\zeta_{a,ij}$ is the limit of $n_a\sigma^2(\widehat{u}_{a,ij})$ as n_a goes to infinity. We also have that $m^2\zeta_{a,ij}$ is the limit of $n_a\widehat{\sigma}^2(\widehat{u}_{a,ij})$ as n_a goes to infinity. Moreover, under Assumption* (**A2**)*, as $n, q \to \infty$ we have*

$$\mathbb{P}\left(\max_{1\le i,j\le q}\left|n_a\widehat{\sigma}^2(\widehat{u}_{a,ij}) - m^2\zeta_{a,ij}\right| \ge C\frac{\varepsilon_n}{\log q}\right) = o(1), \tag{A.1}$$

where $\varepsilon_n = o(1)$ and $a = 1, 2$.

The detailed proof of Lemma A.1 is in Supplement B.1 of Supplementary Material (Zhou et al. [46]). This Lemma presents an upper bound of Jackknife variance estimation error, which enables us to obtain the convergence rate of Jackknife variance estimator. To prove this lemma, we decompose $\widehat{\sigma}^2(\widehat{u}_{a,ij})$ into different pieces and bound each piece separately. The details of this decomposition are in Supplements B.1 and C.1. Both the result and the proof of Lemma A.1 are nontrivial and are of independent technical interest.

Lemma A.1 implies that both of the following two events

$$\mathcal{E}_1 := \left\{\max_{1\le i,j\le q}\left|n_1\widehat{\sigma}^2(\widehat{u}_{1,ij}) - m^2\zeta_{1,ij}\right| < C\frac{\varepsilon_n}{\log q}\right\},$$

$$\mathcal{E}_2 := \left\{\max_{1\le i,j\le q}\left|n_2\widehat{\sigma}^2(\widehat{u}_{2,ij}) - m^2\zeta_{2,ij}\right| < C\frac{\varepsilon_n}{\log q}\right\},$$

happen with probability going to one as $n, q \to \infty$. Under $\mathcal{E}_1$ and $\mathcal{E}_2$, by $\zeta_{a,ij} \ge r_a > 0$ (Assumption (**A2**)), we have

$$\left|n_1\widehat{\sigma}^2(u_{1,ij})/\left(m^2\zeta_{1,ij}\right) - 1\right| < C\varepsilon_n/\log q \quad \text{and} \quad \left|n_2\widehat{\sigma}^2(u_{2,ij})/\left(m^2\zeta_{2,ij}\right) - 1\right| < C\varepsilon_n/\log q.$$

We set $M_n := \max_{1\le i,j\le q} M_{ij}$ and $\widetilde{M}_n := \max_{1\le i,j\le q}\widetilde{M}_{ij}$. By the definition of M_n and $\widetilde{M}_n$, we calculate the relative difference of M_{ij} and $\widetilde{M}_{ij}$ as

$$\left|\frac{M_{ij} - \widetilde{M}_{ij}}{\widetilde{M}_{ij}}\right| \le \left|\frac{\widehat{\sigma}^2(\widehat{u}_{1,ij}) - m^2\zeta_{1,ij}/n_1}{\widehat{\sigma}^2(\widehat{u}_{1,ij})}\right| + \left|\frac{\widehat{\sigma}^2(\widehat{u}_{2,ij}) - m^2\zeta_{2,ij}/n_2}{\widehat{\sigma}^2(\widehat{u}_{2,ij})}\right| \le C\frac{\varepsilon_n}{\log q}. \tag{A.2}$$

Therefore, we have $|M_{ij} - \widetilde{M}_{ij}| \le C\varepsilon_n \widetilde{M}_{ij}/\log q$, which implies that

$$|M_n - \widetilde{M}_n| \le \max_{1\le i,j\le n_1} |M_{ij} - \widetilde{M}_{ij}| \le C\widetilde{M}_n \varepsilon_n/\log q. \tag{A.3}$$

Combining $\widetilde{M}_n/\log q = O_p(1)$ and $\varepsilon_n = o(1)$, to prove Theorem 2.2 it suffices to show that as $n, q \to \infty$, (2.14) holds for any $x \in \mathbb{R}$.

Step (ii). In this step, we use the Hoeffding decomposition (Lemma D.4 in Supplementary Material (Zhou et al. [46])) to decompose U-statistics. We then prove the residual term $\Delta_{n_a,ij}/\binom{n_a}{m}$ is negligible, i.e., to prove the theorem it is sufficient to prove (2.17) as $n, q \to \infty$.

For notational simplicity, we set

$$\widetilde{N}_{ij} := (\widehat{u}_{1,ij} - \widehat{u}_{2,ij})/\sqrt{m^2\zeta_{1,ij}/n_1 + m^2\zeta_{2,ij}/n_2}. \tag{A.4}$$

Recall that in (2.13) and (2.16) we define $\widetilde{M}_{ij}$ and T_{ij} as

$$\begin{aligned} \widetilde{M}_{ij} &:= \frac{(\widehat{u}_{1,ij} - \widehat{u}_{2,ij})^2}{m^2\zeta_{1,ij}/n_1 + m^2\zeta_{2,ij}/n_2}, \\ T_{ij} &:= \frac{\sum_{\alpha=1}^{n_1} h_{ij}(\boldsymbol{X}_\alpha)/n_1 - \sum_{\alpha=1}^{n_2} h_{ij}(\boldsymbol{Y}_\alpha)/n_2}{\sqrt{\zeta_{1,ij}/n_1 + \zeta_{2,ij}/n_2}}. \end{aligned} \tag{A.5}$$

By the definition of $\widetilde{M}_{ij}$, we have $\widetilde{M}_{ij} = (\widetilde{N}_{ij})^2$. Combining the definition of T_{ij} and (2.15), we have

$$\widetilde{N}_{ij} = T_{ij} + \frac{\binom{n_1}{m}^{-1}\Delta_{n_1,ij} - \binom{n_2}{m}^{-1}\Delta_{n_2,ij}}{\sqrt{m^2\zeta_{1,ij}/n_1 + m^2\zeta_{2,ij}/n_2}}. \tag{A.6}$$

We then introduce the following lemma to analyze the difference of $\widetilde{N}_{ij}$ and T_{ij}.

Lemma A.2. *As $n, q \to \infty$, we have*

$$\left| \max_{1\le i,j\le q} (\widetilde{N}_{ij})^2 - \max_{1\le i,j\le q} (T_{ij})^2 \right| = o_p(1). \tag{A.7}$$

The detailed proof of Lemma A.2 is in Supplement B.2. This lemma illustrates that $\max_{1\le,i,j\le q} \widetilde{N}_{ij}$ and $T_n := \max_{1\le i,j\le q} T_{ij}$ have the same limiting distribution. Hence, to prove Theorem 2.2 it suffices to show (2.17) as $n, q \to \infty$.

Step (iii). In this step, we aim to prove (2.17). In (2.17), T_n is the maximum of T_{ij} over $S := \{(i, j) : 1 \le i, j \le q\}$ and these T_{ij}'s are not independent of each other. Therefore, we cannot straightforwardly exploit the extreme value theorem under the independent setting to obtain the limiting distribution of T_n. To solve this problem, we construct normal approximation to obtain the extreme value distribution of $(T_{ij})_{1\le i,j\le q}$ under the setting that T_{ij} can be dependent of each other. The construction of such normal approximation requires most correlations of different T_{ij} to be small. Correlations between different T_{ij}'s are related to the correlations of entries of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Assumption **(A1)** specifies sufficient conditions on the correlations of entries of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

To obtain more insight of Assumption **(A1)**, we introduce the following notations. We use S_0 to denote pairs of (i, j) such that X_i and X_j are highly correlated ($|u_{1,ij}| > (\log q)^{-1-\alpha_0}$) or Y_i and Y_j are highly correlated ($|u_{2,ij}| > (\log q)^{-1-\alpha_0}$). Recalling the formal definition of S_0 in (2.11), Assumption **(A1)** implies that the number of highly correlated ($|u_{a,ij}| > (\log q)^{-1-\alpha_0}$) entries of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is small. More specifically, Assumption **(A1)** assumes $|S_0| = o(q^2)$.

We can prove that correlations between T_{ij}'s on $S \setminus S_0$ are all small. We then use the Bofferroni inequality (Lemma 1 of Cai, Liu and Xia [10]) and normal approximation to obtain the limiting distribution of $\max_{(i,j)\in S\setminus S_0}(T_{ij})^2$ so as to prove (2.17).

We then present the detailed proof of (2.17). First, we prove that it suffices to take the maximum of T_{ij} over $S \setminus S_0$ but instead of over S as in (2.17). By setting $y_q = x + 4\log q - \log(\log q)$, we have

$$\left|\mathbb{P}\Big(\max_{(i,j)\in S}(T_{ij})^2 \geq y_q\Big) - \mathbb{P}\Big(\max_{(i,j)\in S\setminus S_0}(T_{ij})^2 \geq y_q\Big)\right| \leq \mathbb{P}\Big(\max_{(i,j)\in S_0}(T_{ij})^2 \geq y_q\Big). \tag{A.8}$$

The next lemma implies that, as $n,\ q \to \infty$, we have $\mathbb{P}(\max_{(i,j)\in S_0}(T_{ij})^2 \geq y_q) \to 0$.

Lemma A.3. *Under Assumptions* **(A1)** *and* **(A2)**, *as $n, q \to \infty$, we have*

$$\mathbb{P}\Big(\max_{(i,j)\in S_0}(T_{ij})^2 \geq y_q\Big) \to 0.$$

The detailed proof of Lemma A.3 is in Supplement B.3 of Supplementary Material (Zhou et al. [46]). By Lemma A.3, we have $\mathbb{P}(\max_{(i,j)\in S_0}(T_{ij})^2 \geq y_q) \to 0$ as $n, q \to \infty$. Moreover, by (A.8), we have that $\mathbb{P}(\max_{(i,j)\in S}(T_{ij})^2 \geq y_q)$ and $\mathbb{P}(\max_{(i,j)\in S\setminus S_0}(T_{ij})^2 \geq y_q)$ have the same limit value as $n, q \to \infty$. Therefore, to obtain (2.17), it suffices to prove

$$\mathbb{P}\Big(\max_{(i,j)\in S\setminus S_0}(T_{ij})^2 - 4\log q + \log(\log q) \leq x\Big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right), \tag{A.9}$$

as $n, q \to \infty$. The problem is then reduced to prove (A.9).

For simplicity, by rearranging the two-dimensional indices $\{(i, j) : (i, j) \in S \setminus S_0\}$ in any order, we set them as $\{(i_k, j_k) : 1 \leq k \leq h\}$ with $h = |S \setminus S_0|$. If we denote $T_k := T_{i_k j_k}$, (A.9) becomes

$$\mathbb{P}\Big(\max_{1\leq k\leq h}(T_k)^2 - 4\log q + \log(\log q) \leq x\Big) \to \exp\left(-\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right). \tag{A.10}$$

Secondly, we exploit normal approximation to obtain the limiting distribution of $\max_{1\leq k\leq h}(T_k)^2$. This normal approximation is useful for getting the extreme value distribution of weakly dependent data. By excluding all the pairs in S_0, correlations between T_k's are all small. Therefore, we can use this normal approximation to get the limiting distribution of $\max_{1\leq k\leq h}(T_k)^2$. In detail, we first use the Boferroni inequality to obtain both lower and upper bounds of $\mathbb{P}(\max_{1\leq k\leq h}(T_k)^2 \geq y_q)$. The obtained lower and upper bounds can then be shown to have the same limiting distribution, which is the extreme value distribution with the cumulative distribution function of $\exp(-(8\pi)^{-1/2}\exp(-x/2))$.

To describe the procedure of normal approximation, we need some additional notations. We introduce

$$\begin{cases} \widehat{Z}_{\beta,ij} = n_2 h_{ij}(\boldsymbol{X}_\beta)/n_1 & \text{for } 1 \le \beta \le n_1, \\ \widehat{Z}_{\beta,ij} = -h_{ij}(\boldsymbol{Y}_{\beta-n_1}) & \text{for } n_1+1 \le \beta \le n_1+n_2, \end{cases} \tag{A.11}$$

where h_{ij} is defined in (2.9). Moreover, by the definition of T_{ij} in (2.16), we have

$$T_k := T_{i_k j_k} = \sum_{\beta=1}^{n_1+n_2} \widehat{Z}_{\beta,i_k j_k} \Big/ \sqrt{n_2^2 \zeta_{1,i_k j_k}/n_1 + n_2 \zeta_{2,i_k j_k}}. \tag{A.12}$$

After introducing these notations, we explain how to use normal approximation to get the extreme value distribution of $\max_{1 \le k \le h}(T_k)^2$. First, by the Boferroni inequality (Lemma 1 of Cai, Liu and Xia [10]), for any integer M with $0 < M < [h/2]$, we have

$$\begin{aligned} \sum_{\ell=1}^{2M}(-1)^{\ell-1} \sum_{1 \le k_1 < \cdots < k_\ell \le h} \mathbb{P}\Bigg(\bigcap_{j=1}^{\ell} E_{k_j}\Bigg) &\le \mathbb{P}\Big(\max_{1 \le k \le h}(T_k)^2 \ge y_q\Big) \\ &\le \sum_{\ell=1}^{2M-1}(-1)^{\ell-1} \sum_{1 \le k_1 < \cdots < k_\ell \le h} \mathbb{P}\Bigg(\bigcap_{j=1}^{\ell} E_{k_j}\Bigg), \end{aligned} \tag{A.13}$$

where we set $E_{k_j} = \{(T_{k_j})^2 \ge y_q\}$. In next step, to simplify $\mathbb{P}(\bigcap_{j=1}^{\ell} E_{k_j})$, we define

$$\widetilde{Z}_{\beta k} = \widehat{Z}_{\beta,i_k j_k}/(n_2 \zeta_{1,i_k j_k}/n_1 + \zeta_{2,i_k j_k})^{1/2} \quad \text{and} \quad \boldsymbol{W}_\beta = (\widetilde{Z}_{\beta k_1}, \ldots, \widetilde{Z}_{\beta k_\ell})^T, \tag{A.14}$$

for $1 \le k \le h$ and $1 \le \beta \le n_1 + n_2$. Therefore, we have $T_{k_j} = (n_2)^{-1} \sum_{\beta=1}^{n_1+n_2} \widetilde{Z}_{\beta k_j}$. Define $\|\mathbf{v}\|_{\min} = \min_{1 \le i \le \ell} |v_i|$ for vector $\mathbf{v} \in \mathbb{R}^\ell$. With these notations, we rewrite $\mathbb{P}(\bigcap_{j=1}^{\ell} E_{k_j})$ as

$$\mathbb{P}\Bigg(\bigcap_{j=1}^{\ell} E_{k_j}\Bigg) = \mathbb{P}\Bigg(\Bigg\| n_2^{-1/2} \sum_{\beta=1}^{n_1+n_2} \boldsymbol{W}_\beta \Bigg\|_{\min} \ge y_q^{1/2}\Bigg).$$

Second, we use a normal vector $\boldsymbol{N}_\ell$ to approximate $n_2^{-1/2} \sum_{\beta=1}^{n_1+n_2} \boldsymbol{W}_\beta$. In detail, we set $\boldsymbol{N}_\ell$ as a normal vector with the same mean vector and the same covariance matrix as $n_2^{-1/2} \sum_{\beta=1}^{n_1+n_2} \boldsymbol{W}_\beta$. More specifically, we have

$$\begin{aligned} &\boldsymbol{N}_\ell := (N_{k_1}, \ldots, N_{k_\ell})^T \\ &\quad \text{with } \mathbb{E}[\boldsymbol{N}_\ell] = 0, \operatorname{Var}(\boldsymbol{N}_\ell) = n_1 \operatorname{Var}(\boldsymbol{W}_1)/n_2 + \operatorname{Var}(\boldsymbol{W}_{n_1+1}). \end{aligned} \tag{A.15}$$

The following lemma uses $\boldsymbol{N}_\ell$ to rewrite the the upper and lower bounds in (A.13).

Lemma A.4. *Under Assumption* **(A2)**, *as* $n, q \to \infty$, *we have*

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le k\le h}(T_k)^2 \ge y_q\Big)\\
&\quad \le \sum_{\ell=1}^{2M-1}(-1)^{\ell-1}\sum_{1\le k_1<\cdots<k_\ell\le h}\mathbb{P}\big(\|\boldsymbol{N}_\ell\|_{\min}\ge y_q^{1/2}-\epsilon_n(\log q)^{-1/2}\big) \qquad \text{(A.16)}\\
&\qquad + o(1),\\
&\mathbb{P}\Big(\max_{1\le k\le h}(T_k)^2 \ge y_q\Big)\\
&\quad \ge \sum_{\ell=1}^{2M}(-1)^{\ell-1}\sum_{1\le k_1<\cdots<k_\ell\le h}\mathbb{P}\big(\|\boldsymbol{N}_\ell\|_{\min}\ge y_q^{1/2}+\epsilon_n(\log q)^{-1/2}\big) \qquad \text{(A.17)}\\
&\qquad - o(1).
\end{aligned}$$

The detailed proof of Lemma A.4 is in Supplement B.4 of Supplementary Material (Zhou et al. [46]). At last, to complete the proof, we need to prove that the right-hand sides of (A.16) and (A.17) have the same limit value $1-\exp(-(\sqrt{8\pi})^{-1}\exp(-x/2))$ as $n, q \to \infty$. To calculate the limit value, we need the following lemma.

Lemma A.5. *Under Assumption* **(A3)**, *for any integer* $\ell \ge 1$ *and* $x \in \mathbb{R}$, *we have*

$$\sum_{1\le k_1<\cdots<k_\ell\le h}\mathbb{P}\big(\|\boldsymbol{N}_\ell\|_{\min}\ge y_q^{1/2}\pm\epsilon_n(\log q)^{-1/2}\big) = \frac{1}{\ell!}\left(\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right)^{\ell}\big(1+o(1)\big). \qquad \text{(A.18)}$$

The detailed proof of Lemma A.5 is in Supplement B.5. By plugging (A.18) into (A.16) and (A.17), we construct the following inequities:

$$\limsup_{n,q\to\infty}\mathbb{P}\Big(\max_{1\le k\le h}(T_k)^2\ge y_q\Big) \le \sum_{\ell=1}^{2M-1}(-1)^{\ell-1}\frac{1}{\ell!}\left(\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right)^{\ell},$$

$$\liminf_{n,q\to\infty}\mathbb{P}\Big(\max_{1\le k\le h}(T_k)^2\ge y_q\Big) \ge \sum_{\ell=1}^{2M}(-1)^{\ell-1}\frac{1}{\ell!}\left(\frac{1}{\sqrt{8\pi}}\exp\left(-\frac{x}{2}\right)\right)^{\ell},$$

for any positive integer M. Letting $M \to \infty$, we prove (A.10). Therefore, we finish the proof of Theorem 2.2. □

Acknowledgements

The authors thank the Editor, Associate Editor, and two referees for their careful reading and helpful comments. Cheng acknowledges support by a fellowship from CSC (China Scholarship Council). Fang's research is supported by a Google research fellowship and NSF DMS-1712536. Xin-Sheng's research is supported by the National Natural Science Foundation of China (Grant No. 11571080). Han Liu's research is supported by an NSF CAREER Award DMS1454377, NSF IIS1408910, NSF IIS1332109, NIH R01MH102339, NIH R01GM083084, and NIH R01HG06841.

Supplementary Material

Technical Proofs and More Simulation for "An extreme-value approach for testing the equality of large U-statistic based correlation matrices" (DOI: 10.3150/18-BEJ1027SUPP; .pdf). We provide additional proof and simulation in Supplementary Material (Zhou et al. [46]). The Supplementary Material consists of 6 parts: Supplements A–F. Among them, Supplements A–D prove the theorems that are not proven in Appendix A. Supplement E introduces some useful definitions. Supplement F presents more simulation results.

References

[1] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons]. MR1990662

[2] Aslam, S. and Rocke, D.M. (2005). A robust testing procedure for the equality of covariance matrices. *Comput*. *Statist*. *Data Anal*. **49** 863–874. MR2141423

[3] Bai, Z., Jiang, D., Yao, J.-F. and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann*. *Statist*. **37** 3822–3840. MR2572444

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

[5] Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann*. *Probab*. **21** 1275–1294. MR1235416

[6] Bao, Z., Lin, L.-C., Pan, G. and Zhou, W. (2015). Spectral statistics of large dimensional Spearman's rank correlation matrix and its application. *Ann*. *Statist*. **43** 2588–2623. MR3405605

[7] Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[8] Bickel, P.J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann*. *Statist*. **36** 199–227. MR2387969

[9] Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J*. *Amer*. *Statist*. *Assoc*. **106** 672–684. MR2847949

[10] Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J*. *Amer*. *Statist*. *Assoc*. **108** 265–277. MR3174618

[11] Cai, T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **76** 349–372. MR3164870

[12] El Maache, H. and Lepage, Y. (2003). Spearman's rho and Kendall's tau for multivariate data sets. In *Mathematical Statistics and Applications*: *Festschrift for Constance van Eeden*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **42** 113–130. IMS, Beachwood, OH. MR2138289

[13] Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance* **8** 329–384.

[14] Fang, H.-B., Fang, K.-T. and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal*. **82** 1–16. MR1918612

[15] Giedd, J.N., Blumenthal, J., Molloy, E. and Castellanos, F.X. (2001). Brain imaging of attention deficit/hyperactivity disorder. *Ann. N.Y. Acad. Sci*. **931** 33–49.

[16] Han, F., Chen, S. and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306

[17] Han, F., Xu, S. and Zhou, W.-X. (2018). On Gaussian comparison inequality and its application to spectral analysis of large random matrices. *Bernoulli* **24** 1787–1833. MR3757515

[18] Han, F., Zhao, T. and Liu, H. (2013). CODA: High dimensional copula discriminant analysis. *J. Mach. Learn. Res*. **14** 629–671. MR3033343

[19] Ho, J.W.K., Stefani, M., dos Remedios, C.G. and Charleston, M.A. (2008). Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics* **24** i390–i398.

[20] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat*. **19** 293–325. MR0026294

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

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

[23] Jiang, D., Jiang, T. and Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference* **142** 2241–2256. MR2911842

[24] Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93.

[25] Klüppelberg, C. and Kuhn, G. (2009). Copula structure analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **71** 737–753. MR2749917

[26] Kruskal, W.H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc*. **53** 814–861. MR0100941

[27] Li, J. and Chen, S.X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist*. **40** 908–940. MR2985938

[28] Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *Ann. Statist*. **40** 2293–2326. MR3059084

[29] Lopez-Paz, D., Hennig, P. and Schölkopf, B. (2013). The randomized dependence coefficient. In *Advances in Neural Information Processing Systems* 1–9.

[30] Lou, H., Henriksen, L. and Bruhn, P. (1990). Focal cerebral dysfunction in developmental learning disabilities. *Lancet* **335** 8–11.

[31] Mai, Q. and Zou, H. (2015). Sparse semiparametric discriminant analysis. *J. Multivariate Anal*. **135** 175–188. MR3306434

[32] Markowitz, H.M. (1991). Foundations of portfolio theory. *J. Finance* **46** 469–477.

[33] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons, Inc. Wiley Series in Probability and Mathematical Statistics. MR0652932

[34] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist*. **1** 700–709. MR0339405

[35] O'Brien, P.C. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics* **48** 819–827.

[36] Ravikumar, P., Wainwright, M.J., Raskutti, G. and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat*. **5** 935–980. MR2836766

[37] Roy, S.N. (1957). *Some Aspects of Multivariate Analysis*. Calcutta: Wiley, New York; Indian Statistical Institute. MR0092296

[38] Schott, J.R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput*. *Statist*. *Data Anal*. **51** 6535–6542. MR2408613

[39] Shafritz, K.M., Marchione, K.E., Gore, J.C., Shaywitz, S.E. and Shaywitz, B.A. (2004). The effects of methylphenidate on neural systems of attention in attention deficit hyperactivity disorder. *Amer. J. Psychiatry* **161** 1990–1997.

[40] Spearman, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychology* **15** 72–101.

[41] Srivastava, M.S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal*. **99** 386–402. MR2396970

[42] Srivastava, M.S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *J. Multivariate Anal*. **101** 1319–1329. MR2609494

[43] Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *Ann*. *Statist*. **35** 2769–2794. MR2382665

[44] Yufeng, Z., Yong, H., Chaozhe, Z., Qingjiu, C., Manqiu, S., Meng, L., Lixia, T., Tianzi, J. and Yufeng, W. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development* **29** 83–91.

[45] Zhao, T., Roeder, K. and Liu, H. (2014). Positive semidefinite rank-based correlation matrix estimation with application to semiparametric graph estimation. *J. Comput*. *Graph*. *Statist*. **23** 895–922. MR3270703

[46] Zhou, C., Han, F., Zhang, X.-S. and Liu, H. (2019). Supplement to "An extreme-value approach for testing the equality of large U-statistic based correlation matrices." DOI:10.3150/18-BEJ1027SUPP.

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

[48] Zou, Q., Zhu, C., Yang, Y., Zuo, X., Long, X., Cao, Q., Wang, Y. and Zang, Y. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *J. Neurosci*. *Methods* **172** 137–141.

Received January 2017 and revised February 2018