High dimensional independence testing with maxima of rank correlations

Mathias Drton[*], Fang Han[†], Hongjian Shi[‡]

Abstract

Testing mutual independence for high dimensional observations is a fundamental statistical challenge. Popular tests based on linear and simple rank correlations are known to be incapable of detecting non-linear, non-monotone relationships, calling for methods that can account for such dependences. To address this challenge, we propose a family of tests that are constructed using maxima of pairwise rank correlations that permit consistent assessment of pairwise independence. Built upon a newly developed Cramér-type moderate deviation theorem for degenerate U-statistics, our results cover a variety of rank correlations including Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^*. The proposed tests are distribution-free, implementable without the need for permutation, and are shown to be rate-optimal against sparse alternatives under the Gaussian copula model. As a by-product of the study, we reveal an identity between the aforementioned three rank correlation statistics, and hence make a step towards proving a conjecture of Bergsma and Dassios.

Keywords: Degenerate U-statistics, extreme value distribution, independence test, maximum-type test, rank statistics, rate-optimality.

1 Introduction

Let $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ be a random vector taking values in $\mathbb{R}^p$ and having all univariate marginal distributions continuous. This paper is concerned with testing the null hypothesis

$$H_0 : X_1, \ldots, X_p \text{ are mutually independent}, \tag{1.1}$$

based on n independent realizations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of $\boldsymbol{X}$. Testing H_0 is a core problem in multivariate statistics that has attracted the attention of statisticians for decades; see e.g. the exposition in Anderson (2003, Chap. 9) or Muirhead (1982, Chap. 11). Traditional methods such as the likelihood ratio test, Roy's largest root test (Roy, 1957), and Nagao's L_2-type test (Nagao, 1973) target the case where the dimension p is small and perform poorly when p is comparable to or even larger than n. A line of recent work seeks to address this issue and develops tests that are suitable for modern

[*]Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen Ø, Denmark, and Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: `md5@uw.edu`

[†]Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: `fanghan@uw.edu`

[‡]Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: `hongshi@uw.edu`

applications involving data with large dimension p. This high dimensional regime is in the focus of our work, which develops distribution theory based on asymptotic regimes where $p = p_n$ increases to infinity with the sample size n.

Many tests of independence in high dimensions have been proposed recently. For example, Bai et al. (2009) and Jiang and Yang (2013) derived corrected likelihood ratio tests for Gaussian data. Using covariance/correlation statistics such as Pearson's r, Spearman's ρ, and Kendall's τ, Bao et al. (2012), Gao et al. (2017), Han et al. (2018), and Bao (2018) proposed revised versions of Roy's largest root test. Schott (2005) and Leung and Drton (2018) derived corrected Nagao's L_2-type tests. Finally, Jiang (2004), Zhou (2007), and Han et al. (2017) proposed tests using the magnitude of the largest pairwise correlation statistics. Subsequently we shall refer to tests of this latter type as maximum-type tests.

The aforementioned approaches are largely built on linear and simple rank correlations. These, however, are incapable of detecting more complicated non-linear, non-monotone dependences as Hoeffding (1948) noted in his seminal paper. Recent work thus proposed the use of consistent rank (Bergsma and Dassios, 2014a), kernel-based (Gretton et al., 2008; Pfister et al., 2018), and distance covariance/correlation statistics (Székely et al., 2007). However, much less is known about high dimensional tests of H_0 that use these more involved statistics. Notable exceptions include Leung and Drton (2018) and Yao et al. (2018). There, the authors combined Nagao's L_2-type methods with rank and distance covariance statistics that in a tour de force are shown to weakly converge to a Gaussian limit under the null. In addition, Yao et al. (2018) proved that an infeasible version of their test is rate-optimal against a Gaussian dense alternative (Gaussian distribution with equal correlation), while still little is known about optimality of Leung and Drton's.

In this paper, we derive maximum-type tests that are counterparts of Leung–Drton and Yao–Zhang–Shao L_2-type ones. As noted in Han et al. (2017), Leung and Drton (2018), and Yao et al. (2018), maximum-type tests will be more sensitive to strong but sparse dependence. Our tests are formed using statistics based on pairwise rank correlation measures such as Hoeffding's D (Hoeffding, 1948), Blum–Kiefer–Rosenblatt's R (Blum et al., 1961), and Bergsma–Dassios–Yanagimoto's τ^* (Bergsma and Dassios, 2014a; Yanagimoto, 1970). Assuming the pair of random variables X_i and X_j to have an absolutely continuous distribution, these measures all satisfy the following three desirable properties summarized in Weihs et al. (2018):

I-consistency. If X_i and X_j are independent, the correlation measure is zero.

D-consistency. If X_i and X_j are dependent, the correlation measure is nonzero.

Monotonic invariance. The correlation measure is invariant to monotone transformations.

As we shall review in Section 2, the aforementioned correlation statistics can all be treated as instances of degenerate U-statistics, with important special properties.

The contributions of our work are threefold. First, we prove that all the maximum-type test statistics we propose in Section 3 have a null distribution that converges to a (non-standard) Gumbel distribution under high dimensional asymptotics. No distributional assumption is required for this result and the parameters for the Gumbel limit can be given explicitly. This allows one to avoid permutation analysis in problems of larger scale. Second, we conduct a power analysis and give explicit conditions on a sparse local alternative under which our proposed tests have power tending to one. Third, we show that the maximum-type tests based on Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^* are all rate-optimal in the class of Gaussian (copula) distributions with sparse and strong dependence as characterized in the power analysis. These results are developed in Section 4. The theoretical advantages of our tests are highlighted in simulation studies (Section 5). We note that, as an interesting by-product of the study, we give an identity among the above three statistics that helps make a step towards proving Bergsma–Dassios's conjecture about general D-consistency of τ^*. We end the paper with a discussion (Section 6). All proofs are deferred to a supplement.

Notation. The sets of real, integer, and positive integer numbers are denoted $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{Z}^+$, respectively. The cardinality of a set $\mathcal{A}$ is written $\#\mathcal{A}$. For $m \in \mathbb{Z}^+$, we define $[m] = \{1, 2, \ldots, m\}$ and write $\mathcal{P}_m$ for the set of all $m!$ permutations of $[m]$. Let $\boldsymbol{v} = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$, $\mathbf{M} = [\mathbf{M}_{jk}] \in \mathbb{R}^{p\times p}$, and I, J be two subsets of $[p]$. Then $\boldsymbol{v}_I$ is the sub-vector of $\boldsymbol{v}$ with entries indexed by I, and both $\mathbf{M}_{I,J}$ and $\mathbf{M}[I,J]$ are used to refer to the sub-matrix of $\mathbf{M}$ with rows indexed by I and columns indexed by J. The matrix $\mathrm{diag}(\mathbf{M}) \in \mathbb{R}^{p\times p}$ is the diagonal matrix whose diagonal is the same as that of $\mathbf{M}$. We write $\mathbf{I}_p$ for the identity matrix in $\mathbb{R}^{p\times p}$. For a function $f : \mathcal{X} \to \mathbb{R}$, we define $\|f\|_\infty := \max_{x\in\mathcal{X}} |f(x)|$. For any function $f : \mathbb{R} \to \mathbb{R}$, we write $f(\boldsymbol{v}) = (f(v_1), \ldots, f(v_p))^\top$. The greatest integer less than or equal to $x \in \mathbb{R}$ is denoted $\lfloor x \rfloor$. The symbol $\mathbb{1}(\cdot)$ is used for indicator functions. For any two real sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$, $a_n = O(b_n)$, or equivalently $b_n \gtrsim a_n$, if there exists $C > 0$ such that $|a_n| \le C|b_n|$ for any large enough n. We write $a_n \asymp b_n$ if both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold. We write $a_n = o(b_n)$ if for any $c > 0$, $|a_n| \le c|b_n|$ holds for any large enough n. Throughout, c and C will refer to positive absolute constants whose values may differ in different parts of the paper.

2 Rank correlations and degenerate U-statistics

This section introduces the pairwise rank correlations that will later be aggregated in a maximum-type test of the independence hypothesis in (1.1). We present these correlations in a general U-statistic framework. In the sequel, unless otherwise stated, the random vector $\boldsymbol{X}$ is assumed to have continuous margins, that is, its marginal distributions are continuous, though not necessarily absolutely continuous.

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. copies of $\boldsymbol{X}$, with $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,p})^\top$. Let $j \ne k \in [p]$, and let $h : (\mathbb{R}^2)^m \to \mathbb{R}$ be a kernel of order m. The kernel h defines a *U-statistic* of order m:

$$\widehat{U}_{jk} = \binom{n}{m}^{-1} \sum_{1\le i_1<i_2<\cdots<i_m\le n} h\Big\{\binom{X_{i_1,j}}{X_{i_1,k}}, \ldots, \binom{X_{i_m,j}}{X_{i_m,k}}\Big\}. \tag{2.1}$$

For our purposes h may always be assumed to be *symmetric*, i.e., $h(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m) = h(\boldsymbol{z}_{\sigma(1)}, \ldots, \boldsymbol{z}_{\sigma(m)})$ for all permutations $\sigma \in \mathcal{P}_m$ and $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \in \mathbb{R}^2$. Letting $\boldsymbol{z}_i = (z_{i,1}, z_{i,2})^\top$, if both vectors $(z_{1,1}, \ldots, z_{m,1})$ and $(z_{1,2}, \ldots, z_{m,2})$ are free of ties, i.e., have marginal distinct entries, then we have well-defined vectors of ranks $(r_{1,1}, \ldots, r_{m,1})$ and $(r_{1,2}, \ldots, r_{m,2})$, and we define $\boldsymbol{r}_i = (r_{i,1}, r_{i,2})^\top$ for

$1 \leq i \leq n$. Now a kernel is *rank-based* if

$$h(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m) = h(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m)$$

for all $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \in \mathbb{R}^2$ with $(z_{1,1}, \ldots, z_{m,1})$ and $(z_{1,2}, \ldots, z_{m,2})$ free of ties. In this case, we also say that the "correlation" statistic $\widehat{U}_{jk}$ as well as the corresponding "correlation measure" $\mathbb{E}\widehat{U}_{jk}$ is rank-based.

Rank-based statistics have many appealing properties with regard to independence. The following three will be of particular importance for us. Proofs can be found, e.g., in Chapter 31 in Kendall and Stuart (1979), Lemma C4 in the supplement of Han et al. (2017), and Lemma 2.1 in Leung and Drton (2018). We also note that, in finite samples, the statistics $\{\widehat{U}_{jk}; j < k\}$ are generally not mutually independent.

Proposition 2.1. *Under the null hypothesis in (1.1) and assuming continuous margins, we have:*

(i) The statistics $\{\widehat{U}_{jk}, j \neq k\}$ are all identically distributed and are distribution-free, i.e., the distribution of $\widehat{U}_{jk}$ does not depend on the marginal distributions of $X_1, \ldots, X_p$;

(ii) Fix any $j \in [p]$, then the statistics $\{\widehat{U}_{jk}, k \neq j\}$, are mutually independent;

(iii) For any $j \neq k \in [p]$, the statistic $\widehat{U}_{jk}$ is independent of $\{\widehat{U}_{j'k'}; j', k' \notin \{j, k\}, j' \neq k'\}$.

Our focus will be on those rank-based correlation statistics and the corresponding measures that are induced by the kernel $h(\cdot)$ and are both I- and D-consistent. The kernels of these measures satisfy important additional properties that we will assume in our general treatment. Further concepts concerning U-statistics are needed to state this assumption. For any kernel h, any number $\ell \in [m]$, and any measure $\mathbb{P}_{\boldsymbol{Z}}$, we write

$$h_\ell(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell; \mathbb{P}_{\boldsymbol{Z}}) = \mathbb{E}h(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell, \boldsymbol{Z}_{\ell+1}, \ldots, \boldsymbol{Z}_m)$$

and

$$h^{(\ell)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell; \mathbb{P}_{\boldsymbol{Z}}) = h_\ell(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell; \mathbb{P}_{\boldsymbol{Z}}) - \mathbb{E}h - \sum_{k=1}^{\ell-1} \sum_{1 \leq i_1 < \cdots < i_k \leq \ell} h^{(k)}(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_k}; \mathbb{P}_{\boldsymbol{Z}}), \tag{2.2}$$

where $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ are m independent random vectors with distribution $\mathbb{P}_{\boldsymbol{Z}}$. The kernel as well as the corresponding U-statistic is *degenerate* under $\mathbb{P}_{\boldsymbol{Z}}$ if $h_1(\cdot)$ has variance zero. We use the term *completely degenerate* to indicate that the variances of $h_1(\cdot), \ldots, h_{m-1}(\cdot)$ are all zero. Finally, let $\mathbb{P}_0$ be the uniform distribution on $[0, 1]$, and write $\mathbb{P}_0 \otimes \mathbb{P}_0$ for its product measure, the uniform distribution on $[0, 1]^2$. Note that by Proposition 2.1(i), the study of $\widehat{U}_{jk}$ under independent continuous margins X_j and X_k can be reduced to the case with $(X_j, X_k)^\top \sim \mathbb{P}_0 \otimes \mathbb{P}_0$.

Assumption 2.1. *The kernel h is rank-based, symmetric, and has the following three properties:*

(i) h is bounded.

(ii) h is mean-zero and degenerate under independent continuous margins, i.e., $h_1(\boldsymbol{z}_1; \mathbb{P}_0 \otimes \mathbb{P}_0)$ is almost surely zero.

(iii) $h_2(\boldsymbol{z}_1, \boldsymbol{z}_2; \mathbb{P}_0 \otimes \mathbb{P}_0)$ has uniformly bounded eigenfunctions, that is, it admits the expansion

$$h_2(\boldsymbol{z}_1, \boldsymbol{z}_2; \mathbb{P}_0 \otimes \mathbb{P}_0) = \sum_{v=1}^{\infty} \lambda_v \phi_v(\boldsymbol{z}_1) \phi_v(\boldsymbol{z}_2),$$

where $\{\lambda_v\}$ *and* $\{\phi_v\}$ *are the eigenvalues and eigenfunctions satisfying the integral equation*

$$\mathbb{E}h_2(\boldsymbol{z}_1, \boldsymbol{Z}_2)\phi(\boldsymbol{Z}_2) = \lambda\phi(\boldsymbol{z}_1) \quad \textit{for all } \boldsymbol{z}_1 \in \mathbb{R}^2,$$

with $\boldsymbol{Z}_2 \sim \mathbb{P}_0 \otimes \mathbb{P}_0$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\Lambda := \sum_{v=1}^{\infty} \lambda_v \in (0, \infty)$, *and* $\sup_v \|\phi_v\|_\infty < \infty$.

The first boundedness property is satisfied for the commonly used rank correlations, including Kendall's τ, Spearman's ρ, and many others. The latter two properties are much more specific, but exhibited by rank correlation measures for which consistency properties are known; we discuss the main examples below. Note also that the assumption $\Lambda > 0$ implies $\lambda_1 > 0$, so that $h_2(\cdot)$ is not a constant function.

Example 2.1 (Hoeffding's D)**.** From the symmetric kernel

$$\begin{aligned} &h_D(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_5) \\ := &\frac{1}{16} \sum_{(i_1,\ldots,i_5)\in\mathcal{P}_5} \Big[\Big\{\mathbb{1}(z_{i_1,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_2,1} \leq z_{i_5,1})\Big\}\Big\{\mathbb{1}(z_{i_3,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_4,1} \leq z_{i_5,1})\Big\}\Big] \\ &\qquad\qquad \Big[\Big\{\mathbb{1}(z_{i_1,2} \leq z_{i_5,2}) - \mathbb{1}(z_{i_2,2} \leq z_{i_5,2})\Big\}\Big\{\mathbb{1}(z_{i_3,2} \leq z_{i_5,2}) - \mathbb{1}(z_{i_4,2} \leq z_{i_5,2})\Big\}\Big], \end{aligned}$$

we recover Hoeffding's D statistic, which is a rank-based U-statistic of order 5 and gives rise to the Hoeffding's D correlation measure $\mathbb{E}h_D$. The kernel $h_D(\cdot)$ satisfies the first two properties in Assumption 2.1 in view of the results in Hoeffding (1948). To verify the last property, we note that under the measure $\mathbb{P}_0 \otimes \mathbb{P}_0$, $h_{D,2}(\cdot)$ is known to have eigenvalues

$$\lambda_{i,j;D} = 3/(\pi^4 i^2 j^2), \quad i, j \in \mathbb{Z}^+;$$

see, e.g., Theorem 4.4 in Nandy et al. (2016) or Proposition 7 in Weihs et al. (2018). The corresponding eigenfunctions are

$$\phi_{i,j;D}\{(z_{1,1}, z_{1,2})^\top\} = 2\cos(\pi i z_{1,1})\cos(\pi j z_{1,2}), \quad i, j \in \mathbb{Z}^+.$$

The eigenvalues are positive and sum to $\Lambda_D := \sum_{i,j} \lambda_{i,j;D} = 1/12$, and $\sup_{i,j}\|\phi_{i,j;D}\|_\infty \leq 2$. In addition, it has been proven that, once the pair is absolutely continuous in $\mathbb{R}^2$, the correlation measure $\mathbb{E}h_D = 0$ if and only if the pair is independent (Hoeffding, 1948; Yanagimoto, 1970). This property, however, generally does not hold for discrete data or data generated from a bivariate distribution that is continuous but not absolutely continuous; see Remark 1 in Yanagimoto (1970) for a counterexample.

Example 2.2 (Blum–Kiefer–Rosenblatt's R)**.** The symmetric kernel

$$\begin{aligned} &h_R(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_6) \\ := &\frac{1}{32} \sum_{(i_1,\ldots,i_6)\in\mathcal{P}_6} \Big[\Big\{\mathbb{1}(z_{i_1,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_2,1} \leq z_{i_5,1})\Big\}\Big\{\mathbb{1}(z_{i_3,1} \leq z_{i_5,1}) - \mathbb{1}(z_{i_4,1} \leq z_{i_5,1})\Big\}\Big] \\ &\qquad\qquad \Big[\Big\{\mathbb{1}(z_{i_1,2} \leq z_{i_6,2}) - \mathbb{1}(z_{i_2,2} \leq z_{i_6,2})\Big\}\Big\{\mathbb{1}(z_{i_3,2} \leq z_{i_6,2}) - \mathbb{1}(z_{i_4,2} \leq z_{i_6,2})\Big\}\Big] \end{aligned}$$

yields Blum–Kiefer–Rosenblatt's R statistic (Blum et al., 1961), which is a rank-based U-statistic of order 6. One can verify the three properties in Assumption 2.1 similarly to Hoeffding's D by using that $h_{R,2} = 2h_{D,2}$. In addition, for any pair of random variables, we have the correlation measure $\mathbb{E}h_R = 0$ if and only if the pair is independent; cf. page 490 of Blum et al. (1961).

Example 2.3 (Bergsma–Dassios–Yanagimoto's τ^*)**.** Bergsma and Dassios (2014a) introduced a rank correlation statistic as a U-statistic of order 4 with the symmetric kernel

$$\begin{aligned}h_{\tau^*}(\boldsymbol{z}_1,\dots,\boldsymbol{z}_4) := \frac{1}{16}\sum_{(i_1,\dots,i_4)\in\mathcal{P}_4} &\Big\{ \mathbb{1}(z_{i_1,1},z_{i_3,1}<z_{i_2,1},z_{i_4,1}) + \mathbb{1}(z_{i_2,1},z_{i_4,1}<z_{i_1,1},z_{i_3,1}) \\ &-\mathbb{1}(z_{i_1,1},z_{i_4,1}<z_{i_2,1},z_{i_3,1}) - \mathbb{1}(z_{i_2,1},z_{i_3,1}<z_{i_1,1},z_{i_4,1})\Big\} \\ &\Big\{ \mathbb{1}(z_{i_1,2},z_{i_3,2}<z_{i_2,2},z_{i_4,2}) + \mathbb{1}(z_{i_2,2},z_{i_4,2}<z_{i_1,2},z_{i_3,2}) \\ &-\mathbb{1}(z_{i_1,2},z_{i_4,2}<z_{i_2,2},z_{i_3,2}) - \mathbb{1}(z_{i_2,2},z_{i_3,2}<z_{i_1,2},z_{i_4,2})\Big\}.\end{aligned}$$

Here, $\mathbb{1}(y_1,y_2<y_3,y_4) := \mathbb{1}(y_1<y_3)\mathbb{1}(y_1<y_4)\mathbb{1}(y_2<y_3)\mathbb{1}(y_2<y_4)$. It holds that $h_{\tau^*,2} = 3h_{D,2}$ and all properties in Assumption 2.1 also hold for $h_{\tau^*}(\cdot)$. Theorem 1 in Bergsma and Dassios (2014a) shows that for a pair of random variables whose distribution is discrete, absolutely continuous, or a mixture of both, the correlation measure $\mathbb{E}h_{\tau^*}=0$ if and only if the variables are independent. It has been conjectured that this fact is true for any distribution on $\mathbb{R}^2$.

The Bergsma-Dassios-Yanagimoto's correlation measure $\tau^* := \mathbb{E}h_{\tau^*}$ deserves more discussion. Hoeffding (1948) stated a problem about the relationship between equiprobable rankings and independence that was solved by Yanagimoto (1970). In the proof of his Proposition 9, Yanagimoto (1970) presented a correlation measure that is proportional to τ^* of Bergsma–Dassios if the margins are absolutely continuous (cf. Proposition 6.1 in Section 6). Accordingly, we term the correlation "Bergsma–Dassios–Yanagimoto's τ^*". Yanagimoto's key insight is a relation that gives rise to an interesting identity between Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^* statistics. This identity appears to be unknown in the literature. In detail, if $\boldsymbol{z}_1,\dots,\boldsymbol{z}_6\in\mathbb{R}^2$ have no tie among their first and their second entries, then

$$\begin{aligned}&3\cdot\binom{6}{5}^{-1}\sum_{1\le i_1<\cdots<i_5\le 6} h_D(\boldsymbol{z}_{i_1},\dots,\boldsymbol{z}_{i_5}) + 2h_R(\boldsymbol{z}_1,\dots,\boldsymbol{z}_6) \\ &= 5\cdot\binom{6}{4}^{-1}\sum_{1\le i_1<\cdots<i_4\le 6} h_{\tau^*}(\boldsymbol{z}_{i_1},\dots,\boldsymbol{z}_{i_4}). \qquad (2.3)\end{aligned}$$

Equation (2.3) can be easily verified by calculating all 6! entrywise permutations of $\{1,2,\dots,6\}$. Using the identity, we can make a step towards proving the conjecture raised in Bergsma and Dassios (2014a), that is, for an arbitrary random pair $(Z_1,Z_2)^\top\in\mathbb{R}^2$, do we have $\mathbb{E}h_{\tau^*}\ge 0$ with equality if and only if Z_1 and Z_2 are independent?

Theorem 2.1. *For any random vector $\boldsymbol{Z}=(Z_1,Z_2)^\top\in\mathbb{R}^2$ with continuous marginal distributions, we have $\mathbb{E}h_{\tau^*}\ge 0$ and the equality holds if and only if Z_1 is independent of Z_2.*

Theorem 2.1 extends the results in Theorem 1 in Bergsma and Dassios (2014a) to random vectors with continuous margins but a bivariate joint distribution that need not be (absolutely) continuous. The proof hinges on the identity (2.3), the fact that random vectors of continuous margins almost surely have no ties among the values of each coordinate, and the following proposition, which is a simple consequence of the definitions of Hoeffding's D and Blum–Kiefer–Rosenblatt's R; see Hoeffding (1948, Sec. 3) and Blum et al. (1961, p. 490).

Proposition 2.2. *For any random vector $\boldsymbol{Z} = (Z_1, Z_2)^\top \in \mathbb{R}^2$, we have $\mathbb{E}h_D \geq 0$ and $\mathbb{E}h_R \geq 0$.*

The identity (2.3) now gives that $\mathbb{E}h_{\tau^*} \geq 0$ and that $\mathbb{E}h_{\tau^*} = 0$ if and only if $\mathbb{E}h_D = \mathbb{E}h_R = 0$, which in turn implies independence of the considered pair of random variables.

Similarly, a monotonicity property of $\mathbb{E}h_D$ and $\mathbb{E}h_R$ proved by Yanagimoto (1970, Sec. 2) extends to $\mathbb{E}h_{\tau^*}$. We state the Gaussian version of this property.

Theorem 2.2. *If $\boldsymbol{Z} = (Z_1, Z_2)^\top \in \mathbb{R}^2$ is bivariate Gaussian with (Pearson) correlation ρ, then $\mathbb{E}h_D$ and $\mathbb{E}h_R$ and, thus, also $\mathbb{E}h_{\tau^*}$ are increasing functions of $|\rho|$.*

Example 2.4. Unfortunately, identity (2.3) may be false when ties exist. Indeed, if we take $\boldsymbol{z}_i = (\lfloor (i+2)/3 \rfloor, i)^\top$ for $i \in [6]$, then

$$\binom{6}{5}^{-1} \sum_{1 \leq i_1 < \cdots < i_5 \leq 6} h_D(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_5}) = 1/2, \quad h_R(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_6) = 3/2,$$

$$\text{and} \quad \binom{6}{4}^{-1} \sum_{1 \leq i_1 < \cdots < i_4 \leq 6} h_{\tau^*}(\boldsymbol{z}_{i_1}, \ldots, \boldsymbol{z}_{i_4}) = 3/5.$$

Remark 2.1. In view of Example 2.4, the proof of Theorem 2.1 cannot be directly extended to pairs consisting of both discrete and continuous random variables, and the question if Bergsma–Dassios's conjecture is correct remains open in that regard. However, Theorem 2.1 does imply that $\mathbb{E}h_{\tau^*} > 0$ when $\boldsymbol{Z}$ is uniformly distributed on the unit-circle in $\mathbb{R}^2$, giving an affirmative answer to a comment raised in Weihs et al. (2018, p. 551); see also Proposition 6.2 in the discussion section. Moreover, by the Lebesgue decomposition theorem, in order to prove Bergsma–Dassios's conjecture it suffices to prove the case where the pair follows a mixture of discrete and singular measures.

3 Maximum-type tests of mutual independence

We now turn to tests of the mutual independence hypothesis H_0 in (1.1). As in Han et al. (2017), we propose maximum-type tests. However, in contrast to Han et al. (2017), we suggest the use of consistent and rank-based correlations with the practical choices being these from Examples 2.1–2.3. As these measures are all nonnegative, it is appropriate to consider a one-sided test in which we aggregate pairwise U-statistics $\widehat{U}_{jk}$ in (2.1) into the test statistic

$$\widehat{M}_n := (n-1) \max_{j<k} \widehat{U}_{jk}.$$

We then reject H_0 if $\widehat{M}_n$ is larger than a certain threshold. Note that we tacitly assumed $\widehat{U}_{jk} = \widehat{U}_{kj}$ when maximizing over $j < k$; this symmetry holds for any reasonable correlation statistic.

By Proposition 2.1(i), the statistic $\widehat{M}_n$ is distribution-free. An exact critical value for rejection of H_0 could thus be approximated by Monte Carlo simulation. However, as we will show, extreme-value theory yields asymptotic critical values that avoid any extra computation all the while giving good finite-sample control of the test's size. When presenting this theory, we write $X \stackrel{\mathsf{d}}{=} Y$ if two random variables X and Y have the same distribution, and we use $\stackrel{\mathsf{d}}{\longrightarrow}$ to denote "weak convergence".

If, under H_0, the studied statistic $(n-1)\widehat{U}_{jk}$ weakly converged to a chi-square distribution with one degree of freedom, as in Theorems 1 and 2 of Han et al. (2017), then extreme-value theory combined with Proposition 2.1 would imply that a suitably standardized version of $\widehat{M}_n$ would weakly converge to a type-I Gumbel distribution with distribution function $\exp\{-(8\pi)^{-1/2}\exp(-y/2)\}$. However, the degeneracy stated in Assumption 2.1(ii) rules out this possibility. Classical theory yields that instead of a single chi-square variable, we encounter convergence to much more involved infinite weighted series (Serfling, 1980, Chap. 5.5.2).

Proposition 3.1. *Let $\boldsymbol{X}$ have continuous margins, and let $j \neq k$. If $h(\cdot)$ satisfies Assumption 2.1, then under H_0,*

$$\binom{m}{2}^{-1}(n-1)\widehat{U}_{jk} \xrightarrow{\mathsf{d}} \sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1),$$

where $\{\xi_v, v=1,2,\ldots\}$ are i.i.d. standard Gaussian random variables.

Our intuition for the asymptotic forms of the maxima now comes from the following fact.

Proposition 3.2. *Let $Y_1,\ldots,Y_d$ be $d=p(p-1)/2$ i.i.d. copies of $\zeta \overset{\mathsf{d}}{=} \sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)$. Then, as $p\to\infty$,*

$$\max_{j\in[d]}\frac{Y_j}{\lambda_1} - 4\log p - (\mu_1-2)\log\log p + \frac{\Lambda}{\lambda_1} \xrightarrow{\mathsf{d}} G.$$

Here G follows a Gumbel distribution with distribution function

$$\exp\Big\{-\frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp(-y/2)\Big\},$$

where μ_1 is the multiplicity of the largest eigenvalue λ_1 in the sequence $\{\lambda_1,\lambda_2,\cdots\}$, $\kappa := \prod_{v=\mu_1+1}^{\infty}(1-\lambda_v/\lambda_1)^{-1/2}$, and $\Gamma(z) := \int_0^{\infty}x^{z-1}e^{-x}dx$ is the gamma function.

Obviously, when setting $\lambda_1=1, \lambda_2=\lambda_3=\cdots=0$ in Proposition 3.2, we recover the Gumbel distribution derived by Han et al. (2017). Based on Propositions 3.1 and 3.2, for any pre-specified significance level $\alpha\in(0,1)$, our proposed test is

$$\mathsf{T}_\alpha := \mathbb{1}\Big\{\frac{n-1}{\lambda_1\binom{m}{2}}\max_{j<k}\widehat{U}_{jk} - 4\log p - (\mu_1-2)\log\log p + \frac{\Lambda}{\lambda_1} > Q_\alpha\Big\}, \tag{3.1}$$

where

$$Q_\alpha := \log\frac{2^{\mu_1-4}\kappa^2}{\{\Gamma(\mu_1/2)\}^2} - 2\log\log(1-\alpha)^{-1}$$

is the $1-\alpha$ quantile of the Gumbel distribution $\exp\{-\exp(-y/2)\cdot 2^{\mu_1/2-2}\kappa/\Gamma(\mu_1/2)\}$.

Example 3.1 ("Extreme D")**.** Hoeffding's D statistic introduced in Example 2.1 is

$$\widehat{D}_{jk} := \binom{n}{5}^{-1}\sum_{i_1<\cdots<i_5}h_D(\boldsymbol{X}_{i_1,\{j,k\}},\ldots,\boldsymbol{X}_{i_5,\{j,k\}}).$$

According to (3.1), the corresponding test is

$$\mathsf{T}_{D,\alpha} := \mathbb{1}\Big\{\frac{\pi^4(n-1)}{30}\max_{j<k}\widehat{D}_{jk} - 4\log p + \log\log p + \frac{\pi^4}{36} > Q_{D,\alpha}\Big\},$$

where $Q_{D,\alpha} := \log\{\kappa^2/(8\pi)\} - 2\log\log(1-\alpha)^{-1}$ and $\kappa_D := \{2\prod_{n=2}^{\infty}(\pi/n)/\sin(\pi/n)\}^{1/2} \approx 2.466$.

Example 3.2 ("Extreme R")**.** Blum–Kiefer–Rosenblatt's R statistic from Example 2.2 is

$$\widehat{R}_{jk} := \binom{n}{6}^{-1} \sum_{i_1<\cdots<i_6} h_R(\boldsymbol{X}_{i_1,\{j,k\}}, \ldots, \boldsymbol{X}_{i_6,\{j,k\}}).$$

According to (3.1), the corresponding test is

$$\mathsf{T}_{R,\alpha} := \mathbb{1}\Big\{\frac{\pi^4(n-1)}{90}\max_{j<k}\widehat{R}_{jk} - 4\log p + \log\log p + \frac{\pi^4}{36} > Q_{R,\alpha}\Big\},$$

where $Q_{R,\alpha} := Q_{D,\alpha}$.

Example 3.3 ("Extreme τ^*")**.** Bergsma–Dassios–Yanagimoto's τ^* statistic from Example 2.3 is

$$\widehat{\tau}^*_{jk} := \binom{n}{4}^{-1} \sum_{i_1<\cdots<i_4} h_{\tau^*}(\boldsymbol{X}_{i_1,\{j,k\}}, \ldots, \boldsymbol{X}_{i_4,\{j,k\}}).$$

According to (3.1), it yields the test

$$\mathsf{T}_{\tau^*,\alpha} := \mathbb{1}\Big\{\frac{\pi^4(n-1)}{54}\max_{j<k}\widehat{\tau}^*_{jk} - 4\log p + \log\log p + \frac{\pi^4}{36} > Q_{\tau^*,\alpha}\Big\},$$

where $Q_{\tau^*,\alpha} := Q_{D,\alpha}$.

Note that, by the definitions of the kernels and the identity (2.3), as long as there is no tie in the data, for any $j,k \in [p]$,

$$\widehat{D}_{jj} = \widehat{R}_{jj} = \widehat{\tau}^*_{jj} = 1 \quad \text{and} \quad 3\widehat{D}_{jk} + 2\widehat{R}_{jk} = 5\widehat{\tau}^*_{jk}. \tag{3.2}$$

4 Theoretical analysis

This section provides theoretical justifications of the tests proposed in Section 3. The section is split into two parts. The first part rigorously justifies the proposed asymptotic critical values. The second part gives a power analysis and shows optimality properties.

4.1 Size control

In this section, we derive the limiting distribution of the statistic $\widehat{M}_n$ under H_0. The below Cramér-type moderate deviation theorem for degenerate U-statistics under a general probability measure is the foundation of our theory. There has been a large literature on deriving the moderate deviation theorem for non-degenerate U-statistics (see, for example, Shao and Zhou (2016) for some recent developments) as well as Berry–Esseen-type bounds for degenerate U-statistics (see Bentkus and Götze (1997) and Götze and Zaitsev (2014) among many). However, to our knowledge, the literature does not provide a comparable moderate deviation theorem for degenerate U-statistics.

Theorem 4.1 (Cramér-type moderate deviation for degenerate U-statistics)**.** *Let $Z_1, \ldots, Z_n$ be (not necessarily continuous) i.i.d. random variables with distribution $\mathbb{P}_Z$. Consider the U-statistic*

$$\widehat{U}_n = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} h(Z_{i_1}, \ldots, Z_{i_m}),$$

where the kernel $h(\cdot)$ is symmetric and such that (i) $\|h\|_\infty < \infty$, (ii) $h_1(z_1; \mathbb{P}_Z) = 0$ almost surely, and (iii) $h_2(z_1, z_2; \mathbb{P}_Z)$ admits the eigenfunction expansion,

$$h_2(z_1, z_2; \mathbb{P}_Z) = \sum_{v=1}^{\infty} \lambda_v \phi_v(z_1) \phi_v(z_2),$$

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\Lambda := \sum_{v=1}^{\infty} \lambda_v \in (0, \infty)$, and $\sup_v \|\phi_v\|_\infty < \infty$. Then, uniformly over $x \in (0, o(n^\theta))$, we have

$$\frac{\mathbb{P}\Big\{ \binom{m}{2}^{-1} (n-1) \widehat{U}_n > x \Big\}}{\mathbb{P}\Big\{ \sum_{v=1}^{\infty} \lambda_v (\xi_v^2 - 1) > x \Big\}} = 1 + o(1),$$

where $\{\xi_v, v = 1, 2, \ldots\}$ are i.i.d. standard Gaussian,

$$\theta < \sup \Big\{ q \in [0, 1/3) : \sum\nolimits_{v > \lfloor n^{(1-3q)/5} \rfloor} \lambda_v = O(n^{-q}) \Big\} \tag{4.1}$$

if infinitely many of eigenvalues λ_v are nonzero, and $\theta = 1/3$ otherwise.

In Theorem 4.1, when there are only finitely many nonzero eigenvalues, the range $(0, o(n^{1/3}))$, which is the standard one for Cramér moderate deviation, is recovered. When there is infinite number of nonzero eigenvalues, it is still unclear if the range $(0, o(n^\theta))$ is the best possible one. It is certainly an interesting question to investigate the optimal range for degenerate U-statistics in the future. With the aid of Theorem 4.1 and combining it with Proposition 3.2, we have now been able to show that, under H_0, even if p is exponentially larger than the sample size n, our maximum-type test statistic still weakly converges to the Gumbel distribution specified in Proposition 3.2. Hence, the proposed test T_α in (3.1) can effectively control the size.

Theorem 4.2 (Limiting null distribution)**.** *Assume $X_1, \ldots, X_p$ are continuous and the independence hypothesis H_0 holds. Let $\widehat{U}_{jk}$, $j < k$, have a common kernel h that satisfies Assumption 2.1. Define the parameter θ as in (4.1). Then if $p = p_n$ goes to infinity with n such that $\log p = o(n^\theta)$, it holds for all $y \in \mathbb{R}$ that*

$$\begin{aligned} &\mathbb{P}\Big\{ \frac{n-1}{\lambda_1 \binom{m}{2}} \max_{j<k} \widehat{U}_{jk} - 4 \log p - (\mu_1 - 2) \log \log p + \frac{\Lambda}{\lambda_1} \leq y \Big\} \\ =\ &\exp\Big\{ - \frac{2^{\mu_1/2 - 2} \kappa}{\Gamma(\mu_1/2)} \exp\Big(- \frac{y}{2} \Big) \Big\} + o(1). \end{aligned}$$

Consequently,

$$\mathbb{P}(\mathsf{T}_\alpha = 1 \,|\, H_0) = \alpha + o(1).$$

We emphasize that our theory holds without any moment assumption on $\boldsymbol{X}$. This property of being distribution-free is essentially shared by all rank-based correlation measures, but is clearly not satisfied by other measures like linear or distance covariance as was illustrated, for example, by Jiang (2004) and Yao et al. (2018).

As a simple consequence of Theorem 4.2, the following corollary shows that the tests in Examples 3.1–3.3 have asymptotically correct sizes, with θ being explicitly calculated.

Corollary 4.1. *Let $X_1, \ldots, X_p$ be continuous. Let p go to infinity with n in such a way that $\log p = o(n^{1/8-\delta})$ for some arbitrarily small pre-specified constant $\delta > 0$. Then*

$$\mathbb{P}(\mathsf{T}_{D,\alpha} = 1 \mid H_0) = \alpha + o(1), \quad \mathbb{P}(\mathsf{T}_{R,\alpha} = 1 \mid H_0) = \alpha + o(1), \quad \textit{and} \quad \mathbb{P}(\mathsf{T}_{\tau^*,\alpha} = 1 \mid H_0) = \alpha + o(1).$$

4.2 Power analysis and rate-optimality

We now investigate the power of the proposed tests from an asymptotic minimax perspective. The key ingredient is the choice of a suitable distribution family as an alternative to the null hypothesis in (1.1). Recall the definition of $h^{(1)}(\cdot)$ in (2.2). For any kernel function $h(\cdot)$ and constants $\gamma > 0$ and $q \in \mathbb{Z}^+$, define a general q-dimensional (not necessarily continuous) distribution family as follows:

$$\mathcal{D}(\gamma, q; h) := \Big\{\mathcal{L}(\boldsymbol{X}) : \boldsymbol{X} \in \mathbb{R}^q, \mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathbb{P}_{jk})\} \leq \gamma \mathbb{E}_{jk} h \ \text{ for all } j \neq k \in [q]\Big\},$$

where $\mathcal{L}(\boldsymbol{X})$ is the distribution (law) of $\boldsymbol{X}$, and $\mathbb{P}_{jk}$, $\mathbb{E}_{jk}(\cdot)$, and $\mathrm{Var}_{jk}(\cdot)$ stand for the probability measure, expectation, and variance operator on the bivariate distribution of $(X_j, X_k)^\top$, respectively.

The family $\mathcal{D}(\gamma, q; h)$ intrinsically characterizes the slope of the function $\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathbb{P}_{jk})\}$ with regard to the dependence between X_j and X_k characterized by the "correlation measure" $\mathbb{E}_{jk} h$. Furthermore, once X_j is independent of X_k, $\mathbb{P}_{jk}$ reduces to $\mathbb{P}_j \otimes \mathbb{P}_k$, the product measure of the two marginal measures $\mathbb{P}_j, \mathbb{P}_k$. If $\mathbb{P}_j$ and $\mathbb{P}_k$ are further continuous, the according variance is

$$\mathrm{Var}_{jk}\{h^{(1)}(\cdot; \mathbb{P}_j \otimes \mathbb{P}_k)\} = 0 = \mathbb{E}_{jk} h, \tag{4.2}$$

provided that $h(\cdot)$ satisfies the properties in Assumption 2.1. Therefore, in view of (4.2), it is reasonable to expect that $\mathcal{D}(\gamma, q; h)$, for large enough γ, contains many interesting continuous multivariate distributions. Indeed, the Gaussian family belongs to $\mathcal{D}(\gamma, q; h)$ for all the kernels $h(\cdot)$ considered in Examples 2.1 to 2.3, provided γ is large enough.

Lemma 4.1. *There exists an absolute constant $\gamma > 0$ such that for all $q \in \mathbb{Z}^+$, any q-dimensional Gaussian distribution is in $\mathcal{D}(\gamma, q; h_D)$, $\mathcal{D}(\gamma, q; h_R)$, and $\mathcal{D}(\gamma, q; h_{\tau^*})$.*

Next we introduce a class of matrices indexed by a positive constant C as

$$\mathcal{U}_p(C) := \Big\{\mathbf{M} \in \mathbb{R}^{p \times p} : \max_{j<k} \Big\{M_{jk}\Big\} \geq C(\log p/n)\Big\}.$$

Such matrices will define a "sparse local alternative" as considered also in Section 4.1 in Han et al. (2017). Note, however, that in our case the scale is at the order of $\log p/n$ as opposed to $(\log p/n)^{1/2}$ in Han et al. (2017). This is due to our statistics being degenerate under independence. Hence, the variance of $h^{(1)}(\cdot)$ is zero under the null, while nonzero for these statistics investigated in Han et al. (2017).

The following theorem now describes "local alternatives" under which the power of our general test T_α tends to one as n, and with it p, goes to infinity.

Theorem 4.3 (Power analysis, general)**.** *Given any $\gamma > 0$ and a kernel $h(\cdot)$ satisfying Assumption 2.1, there exists some sufficiently large C_γ depending on γ such that*

$$\liminf_{n,p\to\infty} \inf_{\mathbf{U} \in \mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{U}}(\mathsf{T}_\alpha = 1) = 1,$$

where, for each specified (n, p), the infimum is taken over all distributions in $\mathcal{D}(\gamma, p; h)$ that have the matrix of population dependence coefficients $\mathbf{U} = [U_{jk}]$ in $\mathcal{U}_p(C_\gamma)$. Here, $U_{jk} := \mathbb{E}\widehat{U}_{jk}$.

The proof of Theorem 4.3 only uses the Hoeffding decomposition for U-statistics, Bernstein's inequality for the sample mean part, and Arcones and Giné's inequality for the degenerate U-statistics parts (Arcones and Giné, 1993). Consequently, we do not have to assume any continuity of $\boldsymbol{X}$. The theorem immediately yields the following corollary, characterizing the local alternatives under which the three rank-based tests from Examples 3.1–3.3 have power tending to 1.

Corollary 4.2 (Power analysis, examples)**.** *Given any $\gamma > 0$, we have, for some sufficiently large C_γ depending on γ,*

$$\liminf_{n,p\to\infty} \inf_{\mathbf{D}\in\mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{D}}(\mathsf{T}_{D,\alpha} = 1) = 1, \quad \liminf_{n,p\to\infty} \inf_{\mathbf{R}\in\mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{R}}(\mathsf{T}_{R,\alpha} = 1) = 1,$$
$$\liminf_{n,p\to\infty} \inf_{\mathbf{T}^*\in\mathcal{U}_p(C_\gamma)} \mathbb{P}_{\mathbf{T}^*}(\mathsf{T}_{\tau^*,\alpha} = 1) = 1,$$

where, for each specified (n, p), the infima are taken over all distributions in $\mathcal{D}(\gamma, p; h_D)$, $\mathcal{D}(\gamma, p; h_R)$, and $\mathcal{D}(\gamma, p; h_{\tau^*})$ with population dependence coefficient matrices $\mathbf{D} = [D_{jk}]$, $\mathbf{R} = [R_{jk}]$, and $\mathbf{T}^* = [\tau^*_{jk}]$ for $D_{jk} := \mathbb{E}\widehat{D}_{jk}$, $R_{jk} := \mathbb{E}\widehat{R}_{jk}$, and $\tau^*_{jk} := \mathbb{E}\widehat{\tau}^*_{jk}$, respectively.

We now turn to optimality of the proposed tests. There have been long debates on the power of consistent rank-based tests compared to those based on linear and simple rank correlation measures. As a matter of fact, Blum et al. (1961) have given interesting comments on this topic, stating that the required sample size for the bivariate independence test based on $h_R(\cdot)$ is of the same order as that in common parametric cases, hinting that even under a particular parametric model these nonparametric consistent tests of independence can be as rate-efficient as tests that specifically target the considered model. Recently, Yao et al. (2018) made a first step towards a minimax optimality result for consistent tests of independence. Their result shows an infeasible version of a test based on distance covariance to be rate-optimal against a Gaussian dense alternative. However, it remained an open question if there exists a feasible (consistent) test of mutual independence in high dimensions that is rate-optimal against certain alternatives. Below we are able to give an affirmative answer.

We shall focus on the proposed tests in Examples 3.1–3.3 and show their rate-optimality in the Gaussian model. To this end, we define a new alternative class of matrices

$$\mathcal{V}(C) := \Big\{\mathbf{M} \in \mathbb{R}^{p\times p} : \mathbf{M} \succeq 0, \mathrm{diag}(\mathbf{M}) = \mathbf{I}_p, \mathbf{M} = \mathbf{M}^\top, \max_{j\neq k} |M_{jk}| \geq C(\log p/n)^{1/2}\Big\},$$

where $\mathbf{M} \succeq 0$ denotes positive semi-definiteness. We then have the following theorem as a consequence of Corollary 4.2. It concerns the proposed tests' power under a Gaussian model with some nonzero pairwise correlations but for which these are decaying to zero as the sample size increases. Since the test statistics are all rank-based and thus invariant to monotone marginal transformations, extension of the following result to the corresponding Gaussian copula family is straightforward.

Theorem 4.4 (Power analysis, Gaussian)**.** *For a sufficiently large absolute constant $C_0 > 0$, we* *have*

$$\inf_{\mathbf{\Sigma}\in\mathcal{V}(C_0)} \mathbb{P}_{\mathbf{\Sigma}}(\mathsf{T}_{D,\alpha}=1)=1-o(1), \qquad \inf_{\mathbf{\Sigma}\in\mathcal{V}(C_0)} \mathbb{P}_{\mathbf{\Sigma}}(\mathsf{T}_{R,\alpha}=1)=1-o(1),$$

$$\text{and} \quad \inf_{\mathbf{\Sigma}\in\mathcal{V}(C_0)} \mathbb{P}_{\mathbf{\Sigma}}(\mathsf{T}_{\tau^*,\alpha}=1)=1-o(1),$$

where infima are over centered Gaussian distributions with (Pearson) covariance matrix $\mathbf{\Sigma}=[\Sigma_{jk}]$.

The proof of Theorem 4.4 is given in the supplement. It relies on Lemma 4.1 and the fact that $D_{jk}, R_{jk}, \tau^*_{jk} \asymp \Sigma^2_{jk}$ as $\Sigma_{jk}\to 0$. Combined with the following result from Han et al. (2017), Theorem 4.4 yields minimax rate-optimality of the tests in Examples 3.1–3.3 against the sparse Gaussian alternative.

Theorem 4.5 (Rate optimality, Theorem 5 in Han et al., 2017)**.** *There exists an absolute constant* $c_0>0$ *such that for any number* $\beta>0$ *satisfying* $\alpha+\beta<1$*, in any asymptotic regime with* $p\to\infty$ *as* $n\to\infty$ *but* $\log p/n=o(1)$*, it holds for all sufficiently large* n *and* p *that*

$$\inf_{\overline{\mathsf{T}}_\alpha\in\mathcal{T}_\alpha}\ \sup_{\mathbf{\Sigma}\in\mathcal{V}(c_0)} \mathbb{P}(\overline{\mathsf{T}}_\alpha=0)\geq 1-\alpha-\beta.$$

Here the infimum is taken over all size-α *tests, and the supremum is taken over all centered Gaussian distributions with (Pearson) covariance matrix* $\mathbf{\Sigma}$.

5 Simulation studies

In this section we study the finite-sample performance of the three tests (Extreme D, Extreme R, and Extreme τ^*) from Section 3 via Monte Carlo simulations. We compare their performances to the following eight existing tests proposed in the literature. The first five tests are rank-based and hence distribution-free, while the other three tests are distribution-dependent.

LD_τ: the L_2-type test based on Kendall's τ (Leung and Drton, 2018);

LD_ρ: the L_2-type test based on Spearman's ρ (Leung and Drton, 2018);

LD_{τ^*}: the L_2-type test based on Bergsma–Dassios–Yanagimoto's τ^* (Leung and Drton, 2018);

HCL_τ: the maximum-type test based on Kendall's τ (Han et al., 2017);

HCL_ρ: the maximum-type test based on Spearman's ρ (Han et al., 2017);

YZS: the L_2-type test based on the distance covariance statistic (Yao et al., 2018);

SC: the L_2-type test based on Pearson's r (Schott, 2005);

CJ: the maximum-type test based on Pearson's r (Cai and Jiang, 2011);

5.1 Computational aspects

Throughout this section $\{\boldsymbol{z}_i=(z_{i,1},z_{i,2})^\top\}_{i\in[n]}$ is a bivariate sample that contains no tie. We first discuss how to compute the U-statistics $\widehat{D}$, $\widehat{R}$, and $\widehat{\tau}^*$ for Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^*, respectively. As we review below, efficient algorithms are available for $\widehat{D}$ and $\widehat{\tau}^*$. The value of $\widehat{R}$ may then be found using the relation in (3.2).

Hoeffding (1948) himself observed that $\widehat{D}$ can be computed in $O(n \log n)$ time via the following formula

$$\frac{\widehat{D}}{30} = \frac{P - 2(n-2)Q + (n-2)(n-3)S}{n(n-1)(n-2)(n-3)(n-4)}.$$

Here

$$P := \sum_{i=1}^{n} (r_i - 1)(r_i - 2)(s_i - 1)(s_i - 2), \quad Q := \sum_{i=1}^{n} (r_i - 2)(s_i - 1)c_i, \quad S := \sum_{i=1}^{n} c_i(c_i - 1),$$

and r_i and s_i are the ranks of $z_{i,1}$ among $\{z_{1,1}, \dots, z_{n,1}\}$ and $z_{i,2}$ among $\{z_{1,2}, \dots, z_{n,2}\}$, respectively. Moreover, c_i is the number of pairs $\boldsymbol{z}_{i'}$ for which $z_{i',1} < z_{i,1}$ and $z_{i',2} < z_{i,2}$.

Weihs et al. (2016) and Heller and Heller (2016) proposed algorithms for efficient computation of the Bergsma–Dassios–Yanagimoto statistic $\widehat{\tau}^*$. Without loss of generality, let $z_{1,1} < \cdots < z_{n,1}$, i.e., $r_i = i$. Weihs et al. (2016) proved that $2\widehat{\tau}^*/3 = N_c/\binom{n}{4} - 1/3$ with

$$N_c = \sum_{3 \le \ell < \ell' \le n} \binom{\mathbf{B}_<[\ell, \ell']}{2} + \binom{\mathbf{B}_>[\ell, \ell']}{2},$$

where for all $\ell < \ell'$,

$$\begin{aligned} &\mathbf{B}_<[\ell, \ell'] := \#\{i : i \in [\ell - 1], z_{i,2} < \min(z_{\ell,2}, z_{\ell',2})\} \\ \text{and} \quad &\mathbf{B}_>[\ell, \ell'] := \#\{i : i \in [\ell - 1], z_{i,2} > \max(z_{\ell,2}, z_{\ell',2})\}. \end{aligned}$$

Weihs et al. (2016) went on to give an algorithm to compute these counts, and thus $\widehat{\tau}^*$, in $O(n^2 \log n)$ time with little memory use. Heller and Heller (2016) showed that the computation time can be further lowered to $O(n^2)$ via calculation of the following matrix based on the empirical distribution of the ranks r_i and s_i,

$$\mathbf{B}[r, s] := \sum_{i=1}^{n} \mathbb{1}(r_i \le r, s_i \le s), \quad 0 \le r, s \le n.$$

Here, $\mathbf{B}[r, 0] := 0$ and $\mathbf{B}[0, s] := 0$. We may then find $\mathbf{B}_<[\ell, \ell'] = \mathbf{B}[\ell - 1, \min(s_\ell, s_{\ell'}) - 1]$ and $\mathbf{B}_>[\ell, \ell'] = \ell - \mathbf{B}[\ell, \max(s_\ell, s_{\ell'})]$ for all $\ell < \ell'$; recall that s_i is the rank of $z_{i,2}$ in $\{z_{1,2}, \dots, z_{n,2}\}$. As a consequence, formula (3.2) now also yields an $O(n^2)$ algorithm for $\widehat{R}$.

Regarding other competing statistics, note that Pearson's r and Spearman's ρ can be naively computed in time $O(n)$ and $O(n \log n)$, respectively. Knight (1966) proposed an efficient algorithm for computing Kendall's τ that has time complexity $O(n \log n)$. Finally, the algorithm of Huo and Székely (2016) computes the distance covariance statistic in $O(n \log n)$ time.

5.2 Simulation results

We evaluate the size and power of the eleven competing tests introduced above for both Gaussian and non-Gaussian distributions. The values reported below are based on $5,000$ simulations at the nominal significance level of 0.05, with sample size $n \in \{100, 200\}$ and dimension $p \in \{50, 100, 200, 400, 800\}$. All data sets are generated as an i.i.d. sample from the distribution specified for the p-dimensional random vector $\boldsymbol{X}$.

We investigate the sizes of the tests in four settings, where $\boldsymbol{X} = (X_1, \dots, X_p)^\top$ has mutually independent entries.

Example 5.1.

(a) $\boldsymbol{X} \sim N_p(0, \mathbf{I}_p)$ (standard Gaussian).
(b) $\boldsymbol{X} = \boldsymbol{W}^{1/3}$ with $\boldsymbol{W} \sim N_p(0, \mathbf{I}_p)$ (light-tailed Gaussian copula).
(c) $\boldsymbol{X} = \boldsymbol{W}^3$ with $\boldsymbol{W} \sim N_p(0, \mathbf{I}_p)$ (heavy-tailed Gaussian copula).
(d) $X_1, \ldots, X_p$ are i.i.d. with a t-distribution with 3 degrees of freedom.

The simulated sizes of the eight distribution-free tests are reported in Table 1. Those of the three distribution-dependent tests are given in Table 2. As expected, the tests derived from Gaussianity (SC, CJ) fail to control the size for heavy-tailed distributions. In contrast, the other tests control the size effectively in most circumstances. A slight size inflation is observed for Extreme D at small sample size, which can be addressed using Monte Carlo approximation to set the critical value.

In order to study the power properties of the different statistics, we consider three sets of examples. In the first set (Example 5.2), the signal is rather dense. The other two sets of examples focus on sparse settings.

Example 5.2.

(a) The data are generated from the distribution of the random vector

$$\boldsymbol{X} = (\boldsymbol{\omega}^\top, \sin(2\pi\boldsymbol{\omega})^\top, \cos(2\pi\boldsymbol{\omega})^\top, \sin(4\pi\boldsymbol{\omega})^\top, \cos(4\pi\boldsymbol{\omega})^\top)^\top \in \mathbb{R}^p$$

where $\boldsymbol{\omega} \sim N_{p/5}(0, \mathbf{I}_{p/5})$.

(b) The data are generated from the distribution of the random vector

$$\boldsymbol{X} = (\boldsymbol{\omega}^\top, \log(\boldsymbol{\omega}^2)^\top)^\top \in \mathbb{R}^p$$

where $\boldsymbol{\omega} \sim N_{p/2}(0, I_{p/2})$.

Example 5.3.

(a) The data are generated as $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, where

$$\boldsymbol{X}_1 = (\boldsymbol{\omega}^\top, \sin(2\pi\boldsymbol{\omega})^\top, \cos(2\pi\boldsymbol{\omega})^\top, \sin(4\pi\boldsymbol{\omega})^\top, \cos(4\pi\boldsymbol{\omega})^\top)^\top \in \mathbb{R}^{10}$$

with $\boldsymbol{\omega} \sim N_2(0, \mathbf{I}_2)$, and where $\boldsymbol{X}_2 \sim N_{p-10}(0, \mathbf{I}_{p-10})$ is independent of $\boldsymbol{X}_1$.

(b) The data are generated as $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, where

$$\boldsymbol{X}_1 = (\boldsymbol{\omega}^\top, \log(\boldsymbol{\omega}^2)^\top)^\top \in \mathbb{R}^{10}$$

with $\boldsymbol{\omega} \sim N_5(0, \mathbf{I}_5)$, and where $\boldsymbol{X}_2 \sim N_{p-10}(0, \mathbf{I}_{p-10})$ is independent of $\boldsymbol{X}_1$.

Example 5.4.

(a) The data are drawn as $\boldsymbol{X} \sim N_p(0, \mathbf{R}^*)$ with $\mathbf{R}^*$ generated as follows: Consider a random matrix $\boldsymbol{\Delta}$ with all but eight random nonzero entries. We select the locations of four nonzero entries randomly from the upper triangle of $\boldsymbol{\Delta}$, each with a magnitude randomly drawn from the uniform distribution in $[0,1]$. The other four nonzero entries in the lower triangle are determined to make $\mathbf{\Delta}$ symmetric. Finally,

$$\mathbf{R}^* = (1+\delta)\mathbf{I}_p + \mathbf{\Delta},$$

where $\delta = \{-\lambda_{\min}(\mathbf{I}_p + \mathbf{\Delta}) + 0.05\} \cdot \mathbb{1}\{\lambda_{\min}(\mathbf{I}_p + \mathbf{\Delta}) \le 0\}$ and $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of the input.

(b) The data are drawn as $\boldsymbol{X} = \sin(2\pi \boldsymbol{Z}^{1/3}/3)$, where $\boldsymbol{Z} \sim N_p(0, \mathbf{R}^*)$ with $\mathbf{R}^*$ as in (a).

(c) The data are drawn as $\boldsymbol{X} = \sin(\pi \boldsymbol{Z}^3/4)$, where $\boldsymbol{Z} \sim N_p(0, \mathbf{R}^*)$ with $\mathbf{R}^*$ as in (a).

The powers for Examples 5.2–5.4 are reported in Tables 3–5. Several observations stand out. First, throughout all examples, we found that the proposed tests have the highest powers on average, followed by LD_{τ^*} that performs well except for Example 5.4. Among the three proposed tests, the power of Extreme D is highest on average, followed by Extreme τ^*. Recall, however, that D can be subject to slight size inflation. Second, Table 3 shows that, under the dense alternative case, the powers of the proposed tests, YZS, and LD_{τ^*} are all perfect, while the performance of the other methods is considerably worse. Third, comparing the results in Examples 5.2 and 5.3 shows that, as more independent components are added, the power of YZS significantly decreases. This is as expected and indicates that YZS is less powerful in detection of sparse dependences. In addition, both HCL_τ and HCL_ρ perform unsatisfactorily in both settings, indicating that they are powerless in detecting the considered non-linear, non-monotone dependences, an observation that was also made in Yao et al. (2018). Fourth, Tables 4 and 5 jointly confirm the intuition that, for sparse alternatives, the proposed maximum-type tests dominate L_2-type ones including both YZS and LD_{τ^*}, especially when p is large. In addition, we note that, under the setting of Example 5.4, the performances of HCL_τ and HCL_ρ are the second best to the proposed consistent rank-based tests, indicating that there exist cases in which simple rank correlation measures like Kendall's τ and Spearman's ρ can still detect aspects of non-linear non-monotone dependences.

We end this section with a discussion of the simulated-based approach. In view of Proposition 2.1, the distributions of rank-based test statistics are invariant to the generating distribution, and hence we may use simulations to approximate the exact distribution of

$$S := \frac{n-1}{\lambda_1\binom{m}{2}} \max_{j<k} \widehat{U}_{jk} - 4\log p - (\mu_1 - 2)\log\log p + \frac{\Lambda}{\lambda_1}.$$

In detail, we pick a large integer M to be the number of independent replications. For each $t \in [M]$, compute $S^{(t)}$ as the value of S for an $n \times p$ data matrix drawn as having i.i.d. Uniform(0,1) entries. Let $\widehat{F}_{n,p;M}(y) = \frac{1}{M}\sum_{t=1}^{M} \mathbb{1}\{S^{(t)} \le y\}$, $y \in \mathbb{R}$, be the resulting empirical distribution function. For a specified significance level $\alpha \in (0,1)$, we may now use the simulated quantile $\widehat{Q}_{\alpha,n,p;M} := \inf\{y \in \mathbb{R} : \widehat{F}_{n,p;M}(y) \ge 1-\alpha\}$ to form the test

$$\mathsf{T}_\alpha^{\text{exact}} := \mathbb{1}\Big\{\frac{n-1}{\lambda_1\binom{m}{2}} \max_{j<k} \widehat{U}_{jk} - 4\log p - (\mu_1 - 2)\log\log p + \frac{\Lambda}{\lambda_1} > \widehat{Q}_{\alpha,n,p;M}\Big\}. \tag{5.1}$$

The test becomes exact in the large M limit, by the Dvoretzky–Kiefer–Wolfowitz inequality for discontinuous distribution functions (e.g. Kosorok, 2008, Theorem 11.6).

Table 6 gives the sizes and powers of the proposed tests with simulation-based critical values ($M = 5{,}000$). The table shows results only for Examples 5.1 and 5.4 as the simulated powers under

Examples 5.2 and 5.3 were all perfectly one. It can be observed that all sizes are now well controlled, with powers of the proposed tests only slightly different from the ones without using simulation.

Table 1: Empirical sizes of the eight distribution-free tests on Example 5.1

n	p	Extreme D	Extreme R	Extreme τ^*	LD$_\tau$	LD$_\rho$	LD$_{\tau^*}$	HCL$_\tau$	HCL$_\rho$
100	50	0.070	0.042	0.047	0.054	0.048	0.056	0.037	0.028
	100	0.073	0.035	0.042	0.055	0.047	0.066	0.034	0.021
	200	0.076	0.028	0.036	0.058	0.050	0.059	0.028	0.015
	400	0.084	0.025	0.035	0.054	0.045	0.065	0.025	0.012
	800	0.088	0.021	0.032	0.055	0.049	0.062	0.023	0.008
200	50	0.054	0.042	0.044	0.048	0.044	0.051	0.037	0.034
	100	0.057	0.042	0.044	0.052	0.047	0.052	0.038	0.032
	200	0.059	0.038	0.042	0.052	0.050	0.055	0.037	0.032
	400	0.064	0.040	0.045	0.051	0.048	0.053	0.038	0.027
	800	0.065	0.034	0.040	0.051	0.047	0.055	0.034	0.024

Table 2: Empirical sizes of the three distribution-dependent tests on Example 5.1

n	p	YZS	SC	CJ	YZS	SC	CJ	YZS	SC	CJ	YZS	SC	CJ
		Results for Case (a)			Results for Case (b)			Results for Case (c)			Results for Case (d)		
100	50	0.048	0.051	0.029	0.052	0.052	0.036	0.055	0.210	0.974	0.055	0.081	0.479
	100	0.054	0.052	0.018	0.048	0.047	0.032	0.052	0.206	1.000	0.053	0.083	0.781
	200	0.059	0.051	0.013	0.055	0.055	0.024	0.052	0.207	1.000	0.058	0.089	0.974
	400	0.049	0.049	0.011	0.053	0.051	0.022	0.052	0.210	1.000	0.055	0.089	1.000
	800	0.050	0.045	0.005	0.050	0.048	0.018	0.055	0.222	1.000	0.051	0.092	1.000
200	50	0.050	0.044	0.032	0.050	0.052	0.040	0.054	0.194	0.955	0.050	0.086	0.527
	100	0.049	0.049	0.029	0.049	0.051	0.036	0.048	0.190	1.000	0.052	0.089	0.850
	200	0.053	0.049	0.030	0.052	0.053	0.035	0.055	0.193	1.000	0.050	0.085	0.996
	400	0.051	0.049	0.022	0.050	0.048	0.035	0.050	0.193	1.000	0.050	0.091	1.000
	800	0.050	0.053	0.018	0.051	0.053	0.033	0.052	0.188	1.000	0.049	0.088	1.000

6 Further discussion of τ^*

We conclude our work with some further comments on the correlation measure τ^*, which has attracted a lot of attention following the seminal work of Bergsma and Dassios (2014a). For an absolutely continuous bivariate random vector, we have $\tau^* \geq 0$ and equality holds if and only if the pair is independent. Our Theorem 2.1 generalizes this result and has a simple proof based on a relation among U-statistic kernels. The relation is obtained from insights of Yanagimoto (1970). We now provide a second proof of Theorem 2.1 that connects the correlation measures raised by Bergsma and Dassios (2014a) and the one in the proof of Proposition 9 in Yanagimoto (1970). We believe the resulting alternative representation of the population τ^* is of independent interest, e.g., from the point of view of multivariate extensions of τ^* as considered by Weihs et al. (2018).

Table 3: Empirical powers of the eleven competing tests on Example 5.2

n	p	Extreme D	Extreme R	Extreme τ^*	LD$_\tau$	LD$_\rho$	LD$_{\tau^*}$	HCL$_\tau$	HCL$_\rho$	YZS	SC	CJ
					Results for Example 5.2(a)							
100	50	1.000	1.000	1.000	0.086	0.052	1.000	0.286	0.053	1.000	0.052	0.025
	100	1.000	1.000	1.000	0.086	0.052	1.000	0.326	0.053	1.000	0.062	0.025
	200	1.000	1.000	1.000	0.078	0.053	1.000	0.372	0.039	1.000	0.053	0.018
	400	1.000	1.000	1.000	0.075	0.054	1.000	0.417	0.030	1.000	0.058	0.015
	800	1.000	1.000	1.000	0.075	0.051	1.000	0.462	0.024	1.000	0.056	0.012
200	50	1.000	1.000	1.000	0.082	0.052	1.000	0.315	0.071	1.000	0.057	0.036
	100	1.000	1.000	1.000	0.084	0.054	1.000	0.361	0.069	1.000	0.057	0.036
	200	1.000	1.000	1.000	0.078	0.052	1.000	0.399	0.056	1.000	0.051	0.031
	400	1.000	1.000	1.000	0.071	0.050	1.000	0.449	0.054	1.000	0.052	0.027
	800	1.000	1.000	1.000	0.073	0.053	1.000	0.517	0.045	1.000	0.058	0.024
					Results for Example 5.2(b)							
100	50	1.000	1.000	1.000	0.196	0.072	1.000	0.342	0.063	1.000	0.054	0.046
	100	1.000	1.000	1.000	0.189	0.073	1.000	0.386	0.056	1.000	0.049	0.050
	200	1.000	1.000	1.000	0.180	0.069	1.000	0.430	0.045	1.000	0.050	0.055
	400	1.000	1.000	1.000	0.166	0.063	1.000	0.485	0.038	1.000	0.053	0.071
	800	1.000	1.000	1.000	0.184	0.074	1.000	0.546	0.028	1.000	0.053	0.088
200	50	1.000	1.000	1.000	0.192	0.072	1.000	0.369	0.085	1.000	0.051	0.049
	100	1.000	1.000	1.000	0.181	0.073	1.000	0.411	0.075	1.000	0.054	0.050
	200	1.000	1.000	1.000	0.185	0.069	1.000	0.462	0.065	1.000	0.050	0.064
	400	1.000	1.000	1.000	0.171	0.062	1.000	0.522	0.057	1.000	0.049	0.075
	800	1.000	1.000	1.000	0.167	0.070	1.000	0.596	0.047	1.000	0.051	0.093

Table 4: Empirical powers of the eleven competing tests on Example 5.3

n	p	Extreme D	Extreme R	Extreme τ^*	LD$_\tau$	LD$_\rho$	LD$_{\tau^*}$	HCL$_\tau$	HCL$_\rho$	YZS	SC	CJ
						Results for Example 5.3(a)						
100	50	1.000	1.000	1.000	0.058	0.049	1.000	0.089	0.033	0.442	0.047	0.024
	100	1.000	1.000	1.000	0.055	0.045	1.000	0.070	0.025	0.156	0.049	0.018
	200	1.000	1.000	1.000	0.052	0.046	1.000	0.049	0.017	0.071	0.048	0.011
	400	1.000	1.000	1.000	0.058	0.049	0.973	0.043	0.014	0.057	0.050	0.011
	800	1.000	0.827	1.000	0.061	0.052	0.520	0.029	0.009	0.054	0.050	0.007
200	50	1.000	1.000	1.000	0.053	0.045	1.000	0.099	0.038	0.955	0.053	0.033
	100	1.000	1.000	1.000	0.055	0.051	1.000	0.080	0.038	0.435	0.050	0.032
	200	1.000	1.000	1.000	0.048	0.045	1.000	0.060	0.028	0.142	0.045	0.023
	400	1.000	1.000	1.000	0.052	0.047	1.000	0.049	0.023	0.078	0.048	0.023
	800	1.000	1.000	1.000	0.057	0.052	1.000	0.044	0.020	0.053	0.050	0.021
						Results for Example 5.3(b)						
100	50	1.000	1.000	1.000	0.065	0.049	1.000	0.106	0.037	0.984	0.049	0.026
	100	1.000	1.000	1.000	0.054	0.046	1.000	0.078	0.026	0.660	0.046	0.020
	200	1.000	1.000	1.000	0.059	0.052	1.000	0.055	0.018	0.266	0.051	0.014
	400	1.000	1.000	1.000	0.059	0.052	0.996	0.039	0.014	0.107	0.046	0.010
	800	1.000	0.897	1.000	0.059	0.051	0.642	0.030	0.007	0.067	0.052	0.005
200	50	1.000	1.000	1.000	0.062	0.053	1.000	0.120	0.042	1.000	0.050	0.033
	100	1.000	1.000	1.000	0.053	0.047	1.000	0.087	0.040	0.996	0.045	0.036
	200	1.000	1.000	1.000	0.051	0.047	1.000	0.061	0.030	0.729	0.045	0.023
	400	1.000	1.000	1.000	0.053	0.050	1.000	0.050	0.023	0.272	0.053	0.023
	800	1.000	1.000	1.000	0.047	0.044	1.000	0.042	0.021	0.102	0.046	0.016

Table 5: Empirical powers of the eleven competing tests on Example 5.4

n	p	Extreme D	Extreme R	Extreme τ^*	LD_τ	LD_ρ	LD_{τ^*}	HCL_τ	HCL_ρ	YZS	SC	CJ
						Results for Example 5.4(a)						
100	50	0.967	0.962	0.964	0.705	0.586	0.946	0.970	0.966	0.555	0.624	0.973
	100	0.959	0.952	0.954	0.392	0.259	0.914	0.960	0.956	0.252	0.283	0.962
	200	0.950	0.938	0.942	0.161	0.107	0.840	0.950	0.943	0.109	0.115	0.950
	400	0.936	0.924	0.928	0.089	0.064	0.727	0.938	0.931	0.064	0.073	0.941
	800	0.931	0.911	0.918	0.061	0.049	0.539	0.929	0.916	0.051	0.051	0.931
200	50	0.991	0.991	0.991	0.912	0.891	0.988	0.993	0.992	0.871	0.906	0.993
	100	0.984	0.985	0.985	0.728	0.627	0.974	0.988	0.987	0.579	0.650	0.989
	200	0.984	0.983	0.983	0.408	0.278	0.954	0.987	0.985	0.255	0.299	0.988
	400	0.986	0.983	0.983	0.166	0.110	0.917	0.986	0.985	0.111	0.115	0.989
	800	0.980	0.976	0.978	0.073	0.060	0.839	0.983	0.980	0.058	0.063	0.986
						Results for Example 5.4(b)						
100	50	0.759	0.642	0.687	0.244	0.167	0.623	0.623	0.553	0.277	0.260	0.786
	100	0.747	0.624	0.670	0.131	0.091	0.555	0.607	0.540	0.131	0.125	0.758
	200	0.720	0.583	0.635	0.082	0.062	0.444	0.578	0.502	0.080	0.075	0.714
	400	0.702	0.557	0.615	0.065	0.054	0.333	0.549	0.471	0.060	0.061	0.678
	800	0.679	0.512	0.577	0.057	0.048	0.218	0.517	0.431	0.052	0.051	0.638
200	50	0.897	0.843	0.866	0.423	0.343	0.825	0.810	0.767	0.577	0.550	0.928
	100	0.880	0.819	0.846	0.248	0.170	0.753	0.784	0.732	0.287	0.273	0.912
	200	0.855	0.789	0.818	0.128	0.088	0.670	0.757	0.714	0.129	0.128	0.891
	400	0.849	0.768	0.799	0.074	0.059	0.571	0.743	0.689	0.065	0.064	0.875
	800	0.820	0.738	0.772	0.051	0.045	0.450	0.713	0.654	0.053	0.051	0.852
						Results for Example 5.4(c)						
100	50	0.654	0.579	0.608	0.209	0.137	0.541	0.582	0.513	0.111	0.106	0.365
	100	0.656	0.566	0.599	0.109	0.072	0.464	0.580	0.502	0.071	0.064	0.344
	200	0.635	0.527	0.571	0.069	0.055	0.364	0.539	0.455	0.056	0.051	0.311
	400	0.617	0.496	0.546	0.068	0.059	0.256	0.516	0.421	0.053	0.058	0.277
	800	0.597	0.455	0.507	0.055	0.049	0.164	0.487	0.370	0.055	0.049	0.238
200	50	0.824	0.789	0.803	0.396	0.302	0.750	0.785	0.753	0.238	0.211	0.606
	100	0.812	0.773	0.788	0.219	0.143	0.681	0.768	0.732	0.113	0.100	0.570
	200	0.792	0.752	0.767	0.101	0.072	0.596	0.750	0.711	0.063	0.059	0.543
	400	0.776	0.728	0.744	0.070	0.054	0.499	0.730	0.689	0.058	0.057	0.513
	800	0.755	0.699	0.723	0.052	0.048	0.360	0.699	0.646	0.044	0.051	0.473

Table 6: Empirical sizes and powers of simulation-based rejection threshold on Examples 5.1–5.4 (The powers under Examples 5.2 and 5.3 are all perfectly 1.000 and hence omitted)

n	p	Extreme D	Extreme R	Extreme τ^*	Extreme D	Extreme R	Extreme τ^*
		Results for Example 5.1(a)			Results for Example 5.4(a)		
100	50	0.053	0.053	0.053	0.964	0.964	0.965
	100	0.051	0.051	0.050	0.955	0.954	0.955
	200	0.045	0.045	0.044	0.943	0.944	0.945
	400	0.045	0.046	0.046	0.930	0.931	0.932
	800	0.054	0.051	0.051	0.921	0.921	0.923
200	50	0.050	0.053	0.051	0.991	0.991	0.991
	100	0.048	0.048	0.047	0.984	0.985	0.985
	200	0.046	0.045	0.044	0.983	0.984	0.984
	400	0.051	0.058	0.055	0.983	0.984	0.984
	800	0.042	0.044	0.046	0.978	0.978	0.979
		Results for Example 5.4(b)			Results for Example 5.4(c)		
100	50	0.746	0.651	0.694	0.639	0.591	0.611
	100	0.731	0.636	0.676	0.638	0.581	0.607
	200	0.698	0.602	0.643	0.609	0.549	0.580
	400	0.674	0.577	0.624	0.592	0.524	0.557
	800	0.651	0.548	0.594	0.567	0.490	0.526
200	50	0.896	0.853	0.872	0.822	0.800	0.810
	100	0.874	0.824	0.847	0.803	0.775	0.787
	200	0.852	0.794	0.820	0.785	0.757	0.769
	400	0.842	0.778	0.805	0.766	0.738	0.751
	800	0.809	0.746	0.776	0.741	0.708	0.727

Proposition 6.1. *For any pair of absolutely continuous random variables $(X,Y)^\top \in \mathbb{R}^2$ with joint distribution function $F(x,y)$ and marginal distribution functions $F_1(x), F_2(y)$, we have*

$$
\begin{aligned}
&\frac{1}{18}\,\mathbb{E}h_{\tau^*}\\
&\overset{(i)}{=} \int F^2 d(F+F_1F_2) - \int F^2 d(FF_1) - 2\int FF_1 d(FF_2) + \int FF_1 d(F^2) + \frac{1}{18}\\
&\overset{(ii)}{=} \int F^2 dF - 2\int FF_1F_2 dF + 2\int F^2 dF_1 dF_2 - \frac{1}{9}\\
&\overset{(iii)}{=} \int (F-F_1F_2)^2 dF + 2\int (F-F_1F_2)^2 dF_1 dF_2\\
&= \frac{1}{30}\,\mathbb{E}h_D + \frac{1}{45}\,\mathbb{E}h_R,
\end{aligned}
$$

where the term on the righthand side of the identity (ii) is Yagagimoto's correlation measure.

Like previous results, Proposition 6.1 is proven in the appendix and the proof is different from the conditioning argument used in Section 2 of Bergsma and Dassios (2014b).

Finally, we revisit a comment of Weihs et al. (2018) who note that based on existing literature "it is not guaranteed that $\mathbb{E}h_{\tau^*} > 0$ when $(X,Y)^\top$ is generated uniformly on the unit circle in $\mathbb{R}^2$." We are able to calculate the values of D and R for this example and, thus, can deduce the value of τ^*.

Proposition 6.2. *For $(X,Y)^\top$ following the uniform distribution on the unit circle in $\mathbb{R}^2$, we have $\mathbb{E}h_D = \mathbb{E}h_R = \mathbb{E}h_{\tau^*} = 1/16$.*

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis (3rd ed.)*. John Wiley and Sons, Inc., Hoboken, NJ.

Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, 21(3):1494–1542.

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25.

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

Bao, Z. (2018+). Tracy-Widom limit for Kendall's tau. *Ann. Statist.*, in press.

Bao, Z., Pan, G., and Zhou, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.*, 17:no. 88, 1–32.

Bentkus, V. and Götze, F. (1997). Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Probab. Theory Related Fields*, 109(3):367–416.

Bergsma, W. and Dassios, A. (2014a). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, 20(2):1006–1028.

Bergsma, W. and Dassios, A. (2014b). Supplement to "A consistent test of independence based on a sign covariance related to Kendall's tau". *Bernoulli*.

Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32(2):485–498.

Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.*, 39(3):1496–1525.

de la Peña, V. H. and Giné, E. (1999). *Decoupling: From Dependence to Independence*. Springer-Verlag, New York.

Gao, J., Han, X., Pan, G., and Yang, Y. (2017). High dimensional correlation matrices: the central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):677–693.

Götze, F. and Zaitsev, A. Y. (2014). Explicit rates of approximation in the CLT for quadratic forms. *Ann. Probab.*, 42(1):354–397.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems 20*, pages 984–991. Curran Associates, Inc., Red Hook, NY.

Han, F., Chen, S., and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika*, 104(4):813–828.

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(3):1787–1833.

Hashorva, E., Korshunov, D., and Piterbarg, V. I. (2015). Asymptotic expansion of Gaussian chaos via probabilistic approach. *Extremes*, 18(3):315–347.

Heller, Y. and Heller, R. (2016). Computing the Bergsma Dassios sign-covariance. *arXiv preprint arXiv:1605.08732*.

Hoeffding, W. (1948). A non-parametric test of independence. *Ann. Math. Statistics*, 19(4):546–557.

Huo, X. and Székely, G. J. (2016). Fast computing for distance covariance. *Technometrics*, 58(4):435–447.

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

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

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics. Vol. 2: Inference and Relationship (4th ed.).* Charles Griffin & Co. Ltd., London.

Knight, W. R. (1966). A computer method for calculating Kendall's tau with ungrouped data. *J. Am. Stat. Assoc.*, 61(314):436–439.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.*, 46(1):280–307.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley and Sons, Inc., New York.

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

Nandy, P., Weihs, L., and Drton, M. (2016). Large-sample theory for the Bergsma-Dassios sign covariance. *Electron. J. Stat.*, 10(2):2287–2311.

Nelsen, R. B. (2006). *An Introduction to Copulas (2nd ed.).* Springer, New York.

Pfister, N., Bühlmann, P., Schölkopf, B., and Peters, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80(1):5–31.

Roy, S. N. (1957). *Some Aspects of Multivariate Analysis.* John Wiley and Sons Inc., New York; Indian Statistical Institute, Calcutta.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, 92(4):951–956.

Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.*, 9(4):879–885.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley and Sons, Inc., New York.

Shao, Q.-M. and Zhou, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli*, 22(4):2029–2079.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35(6):2769–2794.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge University Press, Cambridge.

Weihs, L., Drton, M., and Leung, D. (2016). Efficient computation of the Bergsma-Dassios sign covariance. *Comput. Statist.*, 31(1):315–328.

Weihs, L., Drton, M., and Meinshausen, N. (2018). Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika*, 105(3):547–562.

Yanagimoto, T. (1970). On measures of association and a related problem. *Ann. Inst. Stat. Math.*, 22(1):57–63.

Yao, S., Zhang, X., and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80(3):455–480.

Zaĭtsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probab. Theory Related Fields*, 74(4):535–566.

Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.*, 359(11):5345–5363.

Zolotarev, V. M. (1962). Concerning a certain probability problem. *Theory Probab. Appl.*, 6(2):201–204.

A Proofs

We first introduce more notation. For any vector $\boldsymbol{v} \in \mathbb{R}^p$, we denote $\|\boldsymbol{v}\|$ as its Euclidean norm. We define the L^∞ norm of a random variable as $\|X\|_\infty = \inf\{t \geq 0 : |X| \leq t \text{ a.s.}\}$, the ψ_2 (sub-gaussian) norm as $\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}\exp(X^2/t^2) \leq 2\}$, and the ψ_1 (sub-exponential) norm as $\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}\exp(|X|/t) \leq 2\}$. For any measure $\mathbb{P}_Z$ and kernel h, we let $H_n^{(\ell)}(\cdot;\mathbb{P}_Z)$ be the U-statistic based on the completely degenerate kernel $h^{(\ell)}(\cdot;\mathbb{P}_Z)$ from (2.2):

$$H_n^{(\ell)}(\cdot;\mathbb{P}_Z) := \binom{n}{\ell}^{-1} \sum_{1\leq i_1<i_2<\cdots<i_\ell\leq n} h^{(\ell)}\Big(Z_{i_1},\ldots,Z_{i_\ell};\mathbb{P}_Z\Big). \tag{A.1}$$

A.1 Proofs for Section 3 of the main paper

Proof of Proposition 3.2. Since $Y_1,\ldots,Y_d$ are i.i.d. realizations of ζ, we have

$$\mathbb{P}\Big(\max_{j\in[d]} Y_j \leq x\Big) = \{\mathbb{P}(\zeta \leq x)\}^d = \{F_\zeta(x)\}^d = \{1-\overline{F}_\zeta(x)\}^d, \tag{A.2}$$

where

$$\overline{F}_\zeta(x) := \mathbb{P}(\zeta > x) = \frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1+o(1)\} \tag{A.3}$$

for $x > -\Lambda$ as $x \to \infty$ by Equation (6) in Zolotarev (1962). Take $x = 4\lambda_1\log p + \lambda_1(\mu_1-2)\log\log p - \Lambda + \lambda_1 y$. Noticing that $x \to \infty$ as $p \to \infty$ and recalling $d = p(p-1)/2$, we obtain

$$\begin{aligned}
d\cdot\overline{F}_\zeta(x) &= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1+o(1)\}\\
&= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}(2\log p)^{\mu_1/2-1}\exp\Big\{-2\log p - \Big(\frac{\mu_1}{2}-1\Big)\log\log p - \frac{y}{2}\Big\}\{1+o(1)\}\\
&= \frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp\Big(-\frac{y}{2}\Big)\{1+o(1)\}.
\end{aligned} \tag{A.4}$$

Combing (A.2) and (A.4), we deduce that

$$\mathbb{P}(\max_{j\in[d]} Y_j \leq x) = \{1-\overline{F}_\zeta(x)\}^d \to \exp\Big\{-\lim_{d\to\infty} d\cdot\overline{F}_\zeta(x)\Big\} = \exp\Big\{-\frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp\Big(-\frac{y}{2}\Big)\Big\},$$

which concludes the proof of the lemma. $\square$

A.2 Proofs for Section 4 of the main paper

A.2.1 Proof of Theorem 4.1

Proof of Theorem 4.1. It is easily derived from the results in Chapter 5.5.2 of Serfling (1980) that $(n-1)\widehat{U}_n$ weakly converges to $\binom{m}{2}\sum_{v=1}^\infty \lambda_v(\xi_v^2-1)$, which implies, for any bounded x,

$$\frac{\mathbb{P}\Big\{\binom{m}{2}^{-1}(n-1)\widehat{U}_n > x\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x\Big\}} = 1 + o(1).$$

For the proof we may thus only consider, without loss of generality, those x's that goes to infinity and θ's that are positive. We proceed in two steps, proving first the case $m = 2$ and then generalizing to $m \geq 2$. For notational convenience we introduce the constants $b_1 := \|h\|_\infty < \infty$ and $b_2 := \sup_v \|\phi_v\|_\infty < \infty$.

Step I. Suppose $m = 2$. We start with the scenario that there are infinitely many nonzero eigenvalues. For a large enough integer K to be specified later, we define the "truncated" kernel of $h_2(z_1, z_2; \mathbb{P}_Z)$ as $h_{2,K}(z_1, z_2; \mathbb{P}_Z) = \sum_{v=1}^K \lambda_v \phi_v(z_1)\phi_v(z_2)$, with corresponding U-statistic

$$\widehat{U}_{K,n} := \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} h_{2,K}(Z_i, Z_j; \mathbb{P}_Z).$$

For simpler presentation, define $Y_{v,i} = \phi_v(Z_i)$ for all $v = 1, 2, \ldots$ and $i \in [n]$. In view of the expansions of $h_{2,K}(\cdot)$ and $h_2(\cdot)$, $\widehat{U}_{K,n}$ and $\widehat{U}_n$ can be written as

$$\widehat{U}_{K,n} = \frac{1}{n-1}\Big\{ \sum_{v=1}^K \lambda_v \Big(n^{-1/2} \sum_{i=1}^n Y_{v,i} \Big)^2 - \sum_{v=1}^K \lambda_v \Big(\frac{\sum_{i=1}^n Y_{v,i}^2}{n} \Big) \Big\}$$
$$\text{and} \quad \widehat{U}_n = \frac{1}{n-1}\Big\{ \sum_{v=1}^\infty \lambda_v \Big(n^{-1/2} \sum_{i=1}^n Y_{v,i} \Big)^2 - \sum_{v=1}^\infty \lambda_v \Big(\frac{\sum_{i=1}^n Y_{v,i}^2}{n} \Big) \Big\}.$$

We now quantify the approximation accuracy of $\widehat{U}_{K,n}$ to $\widehat{U}_n$. Using Slutsky's argument, we obtain

$$\begin{aligned}
&\mathbb{P}\Big\{ (n-1)\widehat{U}_n \geq x \Big\} = \mathbb{P}\Big\{ \sum_{v=1}^\infty \lambda_v \Big(n^{-1/2} \sum_{i=1}^n Y_{v,i} \Big)^2 - \sum_{v=1}^\infty \lambda_v \Big(\frac{\sum_{i=1}^n Y_{v,i}^2}{n} \Big) \geq x \Big\} \\
&\leq \mathbb{P}\Big\{ \sum_{v=1}^K \lambda_v \Big(n^{-1/2} \sum_{i=1}^n Y_{v,i} \Big)^2 - \sum_{v=1}^K \lambda_v \Big(\frac{\sum_{i=1}^n Y_{v,i}^2}{n} \Big) \geq x - \epsilon_1 \Big\} + \mathbb{P}\Big\{ \Big| (n-1)(\widehat{U}_n - \widehat{U}_{K,n}) \Big| \geq \epsilon_1 \Big\} \\
&\leq \mathbb{P}\Big\{ \sum_{v=1}^K \lambda_v \Big(n^{-1/2} \sum_{i=1}^n Y_{v,i} \Big)^2 - \sum_{v=1}^K \lambda_v \geq x - \epsilon_1 - \epsilon_2 \Big\} + \mathbb{P}\Big\{ \Big| (n-1)(\widehat{U}_n - \widehat{U}_{K,n}) \Big| \geq \epsilon_1 \Big\} \\
&\quad + \mathbb{P}\Big\{ \Big| \sum_{v=1}^K \lambda_v \frac{\sum_{i=1}^n (Y_{v,i}^2 - 1)}{n} \Big| \geq \epsilon_2 \Big\},
\end{aligned} \tag{A.5}$$

where ϵ_1, ϵ_2 are constants to be specified later.

The first term on the right-hand side of (A.5) may be controlled using Zaĭtsev's multivariate moderate deviation theorem. For this, we require a dimension-free bound on $\sum_{v=1}^K u_v \lambda_v^{1/2} (n^{-1/2} Y_{v,i})$ for any $u \in \mathbb{R}^K$ satisfying $\|u\| = 1$. Indeed, we have

$$\Big\| \sum_{v=1}^K u_v \lambda_v^{1/2} \frac{Y_{v,i}}{n^{1/2}} \Big\|_\infty \leq \sum_{v=1}^K |u_v| \lambda_v^{1/2} \frac{\|Y_{v,i}\|_\infty}{n^{1/2}} \leq \Big(\sum_{v=1}^K u_v^2 \Big)^{1/2} \Big(\sum_{v=1}^K \lambda_v \Big)^{1/2} n^{-1/2} b_2 \leq n^{-1/2} \Lambda^{1/2} b_2.$$

Thus all assumptions in Theorem 1.1 in Zaĭtsev (1987) are satisfied with the τ in his Equation (1.5) chosen to be $n^{-1/2}\Lambda^{1/2}b_2$. We obtain the following bound:

$$
\begin{aligned}
&\mathbb{P}\Big\{\sum_{v=1}^{K}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2-\sum_{v=1}^{K}\lambda_v\geq x-\epsilon_1-\epsilon_2\Big\}\\
&=\mathbb{P}\Big[\Big\{\sum_{v=1}^{K}\Big(\lambda_v^{1/2}\sum_{i=1}^{n}n^{-1/2}Y_{v,i}\Big)^2\Big\}^{1/2}\geq\Big(x-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)^{1/2}\Big]\\
&\leq\mathbb{P}\Big[\Big\{\sum_{v=1}^{K}(\lambda_v^{1/2}\xi_v)^2\Big\}^{1/2}\geq\Big(x-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)^{1/2}-\epsilon_3\Big]+c_1K^{5/2}\exp\Big\{-\frac{\epsilon_3}{c_2K^{5/2}(n^{-1/2}\Lambda^{1/2}b_2)}\Big\}\\
&=\mathbb{P}\Big[\sum_{v=1}^{K}\lambda_v\xi_v^2\geq\Big\{\Big(x-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)^{1/2}-\epsilon_3\Big\}^2\Big]+c_1K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2K^{5/2}}\Big),
\end{aligned}
\tag{A.6}
$$

where ϵ_3 is a constant to be specified later. Combining (A.5) and (A.6), we find using Slutsky's argument once again that

$$
\begin{aligned}
&\mathbb{P}\Big\{\sum_{v=1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2-\sum_{v=1}^{\infty}\lambda_v\Big(\frac{\sum_{i=1}^{n}Y_{v,i}^2}{n}\Big)\geq x\Big\}\\
&\leq\mathbb{P}\Big[\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)\geq\Big\{\Big(x-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)^{1/2}-\epsilon_3\Big\}^2-\sum_{v=1}^{K}\lambda_v-\epsilon_4\Big]\\
&\quad+\mathbb{P}\Big\{\Big|(n-1)(\widehat{U}_n-\widehat{U}_{K,n})\Big|\geq\epsilon_1\Big\}+\mathbb{P}\Big\{\Big|\sum_{v=1}^{K}\lambda_v\frac{\sum_{i=1}^{n}(Y_{v,i}^2-1)}{n}\Big|\geq\epsilon_2\Big\}\\
&\quad+c_1K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2K^{5/2}}\Big)+\mathbb{P}\Big\{\Big|\sum_{v=K+1}^{\infty}\lambda_v(\xi_v^2-1)\Big|\geq\epsilon_4\Big\},
\end{aligned}
\tag{A.7}
$$

where ϵ_4 is another constant to be specified later. In the following, we separately study the five terms on the right-hand side of (A.7), starting from the first term.

Let $\epsilon:=x-[\{(x-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v)^{1/2}-\epsilon_3\}^2-\sum_{v=1}^{K}\lambda_v-\epsilon_4]$. Then

$$
\epsilon=\epsilon_1+\epsilon_2+2\epsilon_3\Big(x-\epsilon_1-\epsilon_2+\sum_{v=1}^{K}\lambda_v\Big)^{1/2}-\epsilon_3^2+\epsilon_4
$$

and

$$
\mathbb{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)\geq x-\epsilon\Big\}\leq\mathbb{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)\geq x\Big\}+\epsilon\cdot\max_{x'\in[x-\epsilon,x]}p_\zeta(x')
\tag{A.8}
$$

where $p_\zeta(x)$ is the density of the random variable $\zeta:=\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)$.

We turn to the second term in (A.7). Proposition 2.6.1 and Example 2.5.8 in Vershynin (2018) yield that

$$
\Big\|n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big\|_{\psi_2}^2\leq 8n^{-1}\sum_{i=1}^{n}\Big\|Y_{v,i}\Big\|_{\psi_2}^2\leq 8(\log 2)^{-1}b_2^2\leq 12b_2^2.
$$

Applying the triangle inequality and Lemma 2.7.6 in Vershynin (2018), we deduce that

$$\Big\|\sum_{v=K+1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2\Big\|_{\psi_1}\le\sum_{v=K+1}^{\infty}\lambda_v\Big\|\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2\Big\|_{\psi_1}$$
$$=\sum_{v=K+1}^{\infty}\lambda_v\Big\|n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big\|_{\psi_2}^2\le 12b_2^2\sum_{v=K+1}^{\infty}\lambda_v.$$

Using Proposition 2.7.1 in Vershynin (2018), this is seen to further imply that, for any $\epsilon_1'>0$,

$$\mathbb{P}\Big\{\sum_{v=K+1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2\ge\epsilon_1'\Big\}\le 2\exp\Big(-\frac{\epsilon_1'}{12b_2^2\sum_{v=K+1}^{\infty}\lambda_v}\Big).$$

Noting that

$$\Big|\sum_{v=K+1}^{\infty}\lambda_v\Big(n^{-1}\sum_{i=1}^{n}Y_{v,i}^2\Big)\Big|\le b_2^2\sum_{v=K+1}^{\infty}\lambda_v,$$

we obtain, for any $\epsilon_1>b_2^2\sum_{v=K+1}^{\infty}\lambda_v$,

$$\begin{aligned}&\mathbb{P}\Big\{\Big|(n-1)(\widehat{U}_n-\widehat{U}_{K,n})\Big|\ge\epsilon_1\Big\}\\
&\le\mathbb{P}\Big\{\Big|\sum_{v=K+1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2\Big|+\Big|\sum_{v=K+1}^{\infty}\lambda_v\Big(n^{-1}\sum_{i=1}^{n}Y_{v,i}^2\Big)\Big|\ge\epsilon_1\Big\}\\
&\le\mathbb{P}\Big\{\Big|\sum_{v=K+1}^{\infty}\lambda_v\Big(n^{-1/2}\sum_{i=1}^{n}Y_{v,i}\Big)^2\Big|\ge\epsilon_1-b_2^2\sum_{v=K+1}^{\infty}\lambda_v\Big\}\le 2e^{1/12}\exp\Big(-\frac{\epsilon_1}{12b_2^2\sum_{v=K+1}^{\infty}\lambda_v}\Big).\end{aligned}\tag{A.9}$$

We next study the third term in (A.7). Again, Proposition 2.6.1 and Example 2.5.8 in Vershynin (2018) give

$$\Big\|n^{-1}\sum_{i=1}^{n}(Y_{v,i}^2-1)\Big\|_{\psi_2}^2\le 8n^{-2}\sum_{i=1}^{n}\Big\|Y_{v,1}^2-1\Big\|_{\psi_2}^2\le 12n^{-1}(b_2^2+1)^2,$$

which further yields

$$\Big\|\sum_{v=1}^{K}\lambda_v\sum_{i=1}^{n}\frac{Y_{v,i}^2-1}{n}\Big\|_{\psi_2}\le\sum_{v=1}^{K}\lambda_v\Big\|n^{-1}\sum_{i=1}^{n}(Y_{v,i}^2-1)\Big\|_{\psi_2}\le 12^{1/2}n^{-1/2}\Lambda(b_2^2+1).$$

Using Proposition 2.5.2 in Vershynin (2018), we have, for any $\epsilon_2>0$,

$$\mathbb{P}\Big(\Big|\sum_{v=1}^{K}\lambda_v\sum_{i=1}^{n}\frac{Y_{v,i}^2-1}{n}\Big|\ge\epsilon_2\Big)\le 2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\}.\tag{A.10}$$

The fourth term in (A.7) is explicit, and it remains to bound the fifth and last term. Since ξ_v is a sub-gaussian random variable, ξ_v^2-1 is sub-exponential. One readily verifies $\|\xi_v^2-1\|_{\psi_1}\le 4$, and accordingly

$$\Big\|\sum_{v=K+1}^{\infty}\lambda_v(\xi_v^2-1)\Big\|_{\psi_1}\le\sum_{v=K+1}^{\infty}\lambda_v\|\xi_v^2-1\|_{\psi_1}\le 4\sum_{v=K+1}^{\infty}\lambda_v.$$

By Proposition 2.7.1 in Vershynin (2018), this further implies that, for any $\epsilon_4 > 0$,

$$\mathbb{P}\Big\{\Big|\sum_{v=K+1}^{\infty}\lambda_v(\xi_v^2-1)\Big|\geq\epsilon_4\Big\}\leq 2\exp\Big(-\frac{\epsilon_4}{4\sum_{v=K+1}^{\infty}\lambda_v}\Big). \tag{A.11}$$

We now specify the integer K to be $\lfloor n^{(1-3\theta)/5}\rfloor$. By the definition of θ, there exists a positive absolute constant C_λ such that $\sum_{v=K+1}^{\infty}\lambda_v\leq C_\lambda n^{-\theta}$ for all sufficiently large n. Combining this fact and inequalities (A.7)–(A.11), we obtain

$$\begin{aligned}
&\frac{\mathbb{P}\Big\{(n-1)\widehat{U}_n>x\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)>x\Big\}}-1\\
&\leq\{\overline{F}_\zeta(x)\}^{-1}\Big[\epsilon\cdot\max_{x'\in[x-\epsilon,x]}p_\zeta(x')+2e^{1/12}\exp\Big(-\frac{\epsilon_1}{12b_2^2\sum_{v=K+1}^{\infty}\lambda_v}\Big)+2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\}\\
&\quad+c_1K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2K^{5/2}}\Big)+2\exp\Big(-\frac{\epsilon_4}{4\sum_{v=K+1}^{\infty}\lambda_v}\Big)\Big]\\
&\leq\{\overline{F}_\zeta(x)\}^{-1}\Big[\epsilon\cdot\max_{x'\in[x-\epsilon,x]}p_\zeta(x')+2e^{1/12}\exp\Big(-\frac{\epsilon_1}{12b_2^2C_\lambda n^{-\theta}}\Big)+2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\}\\
&\quad+c_1n^{(1-3\theta)/2}\exp\Big\{-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2n^{(1-3\theta)/2}}\Big\}+2\exp\Big(-\frac{\epsilon_4}{4C_\lambda n^{-\theta}}\Big)\Big],
\end{aligned} \tag{A.12}$$

which we shall prove to be $o(1)$. The starting point for proving this are Equations (5) and (6) in Zolotarev (1962), which yield that the density $p_\zeta(x)$ and the survival function $\overline{F}_\zeta(x)=\mathbb{P}(\zeta>x)$ of $\zeta=\sum_{v=1}^{\infty}\lambda_v(\xi_v^2-1)$ satisfy

$$p_\zeta(x)=\frac{\kappa}{2\lambda_1\cdot\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1+o(1)\}$$

$$\text{and}\quad\overline{F}_\zeta(x):=\frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\{1+o(1)\}$$

for $x>-\Lambda$ tending to infinity. Here μ_1 is the multiplicity of the largest eigenvalue λ_1 and $\kappa:=\prod_{v=\mu_1+1}^{\infty}(1-\lambda_v/\lambda_1)^{-1/2}$.

Consider the first term in (A.12). We claim that there exists an absolute constant $x_0>0$ such that, for all $0<\epsilon\leq\lambda_1/2$,

$$\sup_{x\geq x_0}\Big|\{\overline{F}_\zeta(x)\}^{-1}\cdot\max_{x'\in[x-\epsilon,x]}p_\zeta(x')\Big|\leq\frac{2}{\lambda_1}. \tag{A.13}$$

Indeed, we have $p_\zeta(x)/\overline{F}_\zeta(x)=(2\lambda_1)^{-1}\{1+o(1)\}$, and thus there exists $x_0'>-\Lambda$ such that $p_\zeta(x)/\overline{F}_\zeta(x)\leq\lambda_1^{-1}$ for $x\geq x_0'$. Then for any $0<\epsilon\leq\lambda_1/2$,

$$\frac{\max_{x'\in[x-\epsilon,x]}p_\zeta(x')}{\overline{F}_\zeta(x)}=\frac{p_\zeta(x-\epsilon')}{\overline{F}_\zeta(x)}\leq\frac{p_\zeta(x-\epsilon')}{\overline{F}_\zeta(x-\epsilon')-\epsilon'\cdot p_\zeta(x-\epsilon')}=\frac{1}{\overline{F}_\zeta(x-\epsilon')/p_\zeta(x-\epsilon')-\epsilon'}\leq\frac{2}{\lambda_1}$$

for $x-\epsilon\geq x_0'$, where $\epsilon'\in[0,\epsilon]$ is chosen such that $p_\zeta(x-\epsilon')=\max_{x'\in[x-\epsilon,x]}p_\zeta(x')$. Now (A.13) holds when taking $x_0=x_0'+\lambda_1/2$.

From (A.13), we conclude that the first term in (A.12) tends to 0 as $\epsilon\to 0$ uniformly over

$x \geq x_0$. Choosing

$$\epsilon_1 = 12b_2^2 C_\lambda n^{-\theta}\Big(\frac{x+\Lambda}{2\lambda_1} + n^{\theta/2}\Big), \quad \epsilon_2 = n^{-1/3}, \quad \epsilon_3 = n^{-\theta/2}, \quad \epsilon_4 = 4C_\lambda n^{-\theta}\Big(\frac{x+\Lambda}{2\lambda_1} + n^{\theta/2}\Big),$$

we deduce that the first term in (A.12) is $o(1)$ as $n \to \infty$ by

$$\epsilon \asymp \epsilon_1 + \epsilon_2 + 2\epsilon_3 x^{1/2} - \epsilon_3^2 + \epsilon_4 = \epsilon_1 + \epsilon_2 + 2(x/n^\theta)^{1/2} - \epsilon_3^2 + \epsilon_4 = o(1).$$

Recall that we consider $x \in (0, o(n^\theta))$. One can further verify that the other four terms in (A.12) are also $o(1)$ as $n \to \infty$:

$$\begin{aligned}
&\Big\{\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\Big\}^{-1}\exp\Big(-\frac{\epsilon_1}{12b_2^2 C_\lambda n^{-a}}\Big) \lesssim n^{\theta/2}\exp(-n^{\theta/2}) = o(1),\\
&\Big\{\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\Big\}^{-1}\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} \lesssim n^{\theta/2}\exp(-C' n^{1/3}) = o(1),\\
&\Big\{\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\Big\}^{-1} n^{(1-3\theta)/2}\exp\Big\{-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2 n^{(1-3\theta)/2}}\Big\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\lesssim n^{1/2-\theta}\exp(-C'' n^\theta) = o(1),\\
&\Big\{\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\Big\}^{-1}\exp\Big(-\frac{\epsilon_4}{4C_\lambda n^{-\theta}}\Big) \lesssim n^{\theta/2}\exp(-n^{\theta/2}) = o(1).
\end{aligned}$$

Here C' and C'' are some absolute positive constants.

If there are only finitely many nonzero eigenvalues, a simple modification to (A.12) gives

$$\begin{aligned}
&\frac{\mathbb{P}\Big\{(n-1)\widehat{U}_n > x\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2 - 1) > x\Big\}} - 1\\
&\leq \frac{1}{\overline{F}_\zeta(x)}\Big[\epsilon \cdot \max_{x' \in [x-\epsilon, x]} p_\zeta(x') + 2\exp\Big\{-\frac{n\epsilon_2^2}{48\Lambda^2(b_2^2+1)^2}\Big\} + c_1 K^{5/2}\exp\Big(-\frac{n^{1/2}\epsilon_3}{c_2\Lambda^{1/2}b_2 K^{5/2}}\Big)\Big], \quad \text{(A.14)}
\end{aligned}$$

where K is the number of nonzero eigenvalues. Choosing $\epsilon_2 = n^{-1/3}$, $\epsilon_3 = n^{-1/6}$, and using $\epsilon \asymp \epsilon_2 + 2\epsilon_3 x^{1/2} - \epsilon_3^2 = o(1)$, we obtain that the right-hand side of (A.14) is $o(1)$.

We thus proved that

$$\frac{\mathbb{P}\Big\{(n-1)\widehat{U}_n > x\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2 - 1) > x\Big\}} - 1 \leq o(1)$$

and that this holds uniformly over $x \in (0, o(n^\theta))$. It can be shown similarly that the left hand side in the display is lower-bounded by an $o(1)$ term uniformly over $x \in (0, o(n^\theta))$, which concludes the proof of the case $m = 2$.

Step II. We use the Hoeffding decomposition and the exponential inequality for bounded completely degenerate U-statistics of Arcones and Giné (1993) to prove the general case $m \geq 2$. Write

$$\binom{m}{2}^{-1}(n-1)\widehat{U}_n = (n-1)H_n^{(2)}(\cdot;\mathbb{P}_Z) + \sum_{\ell=3}^m \binom{m}{2}^{-1}\binom{m}{\ell}(n-1)H_n^{(\ell)}(\cdot;\mathbb{P}_Z).$$

Using Slutsky's argument, we have

$$\frac{\mathbb{P}\Big\{\binom{m}{2}^{-1}(n-1)\widehat{U}_n > x\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x\Big\}}$$
$$\leq \frac{\mathbb{P}\Big\{(n-1)H_n^{(2)}(\cdot;\mathbb{P}_Z) > x-\epsilon_5\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x\Big\}} + \sum_{\ell=3}^m \frac{\mathbb{P}\Big\{\binom{m}{2}^{-1}\binom{m}{\ell}(n-1)\cdot|H_n^{(\ell)}(\cdot;\mathbb{P}_Z)| \geq \epsilon_{5,\ell}\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x\Big\}} \tag{A.15}$$

where $\{\epsilon_{5,\ell}, \ell = 3,\ldots,m\}$ are constants to be specified later and $\epsilon_5 := \sum_{\ell=3}^m \epsilon_{5,\ell}$.

We analyze the first term and the remaining terms on the right-hand side of (A.15) separately. To bound the latter, we employ Theorem 4.1.12 in de la Peña and Giné (1999), which states that there exist absolute positive constants C'_ℓ and C''_ℓ such that for all $\epsilon'_5 > 0$,

$$\mathbb{P}(n^{\ell/2}|H_n^{(\ell)}(\cdot;\mathbb{P}_Z)| \geq \epsilon'_5) \leq C'_\ell \exp\{-C''_\ell(\epsilon'_5/\|h^{(\ell)}(\cdot;\mathbb{P}_Z)\|_\infty)^{2/\ell}\}, \tag{A.16}$$

where $\|h^{(\ell)}(\cdot;\mathbb{P}_Z)\|_\infty \leq 2^\ell b_1$ by the alternative definition of $h^{(\ell)}(z_1,\ldots,z_\ell;\mathbb{P}_Z)$ as below:

$$h^{(\ell)}(z_1,\ldots,z_\ell;\mathbb{P}_Z)$$
$$= h_\ell(z_1,\ldots,z_\ell;\mathbb{P}_Z) + \sum_{k=1}^{\ell-1}(-1)^{\ell-k}\sum_{1\leq i_1<\cdots<i_k\leq\ell} h_k(z_{i_1},\ldots,z_{i_k};\mathbb{P}_Z) + (-1)^\ell\mathbb{E}h.$$

Plugging (A.16) into each term in the sum on the right of (A.15) implies, for $n \geq 2$,

$$\sum_{\ell=3}^m \frac{\mathbb{P}\Big\{\binom{m}{2}^{-1}\binom{m}{\ell}(n-1)\cdot|H_n^{(\ell)}(\cdot;\mathbb{P}_Z)| \geq \epsilon_{5,\ell}\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x\Big\}}$$
$$\leq \sum_{\ell=3}^m C'_\ell \exp\Big[-C''_\ell\Big\{n^{\ell/2-1}\binom{m}{2}\binom{m}{\ell}^{-1}\epsilon_{5,\ell}\Big/\|h^{(\ell)}(\cdot;\mathbb{P}_Z)\|_\infty\Big\}^{2/\ell}\Big]\{\overline{F}_\zeta(x)\}^{-1}$$
$$\lesssim \sum_{\ell=3}^m \exp\Big[-C''_\ell n^\theta\Big\{\binom{m}{2}\binom{m}{\ell}^{-1}\epsilon_{5,\ell}\Big/(2^\ell b_1)\Big\}^{2/\ell}\Big]\Big\{\Big(\frac{x+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{x+\Lambda}{2\lambda_1}\Big)\Big\}^{-1}, \tag{A.17}$$

where the last step is due to the fact that $a \leq 1/3 \leq 1-2/\ell$ for $\ell \geq 3$. Taking

$$\epsilon_{5,\ell} = b_1\binom{m}{2}^{-1}\binom{m}{\ell}\Big\{\frac{4}{C''_\ell n^\theta}\Big(\frac{x+\Lambda}{2\lambda_1}+n^{\theta/2}\Big)\Big\}^{\ell/2},$$

the sum on the right-hand side of (A.17) is seen to be $o(1)$. It remains to control the first term in (A.15). We start by writing the term as

$$\frac{\mathbb{P}\Big\{(n-1)H_n^{(2)} > x-\epsilon_5\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x\Big\}}$$
$$= \frac{\mathbb{P}\Big\{(n-1)H_n^{(2)} > x-\epsilon_5\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x-\epsilon_5\Big\}}\frac{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x-\epsilon_5\Big\}}{\mathbb{P}\Big\{\sum_{v=1}^\infty \lambda_v(\xi_v^2-1) > x\Big\}}. \tag{A.18}$$

The first factor in (A.18) is $1+o(1)$ by the proof in Step I. For the second term in (A.18), we have

$$1 \le \frac{\overline{F}_\zeta(x-\epsilon_5)}{\overline{F}_\zeta(x)} \le 1 + \frac{\epsilon_5 \cdot \max_{x' \in [x-\epsilon_5, x]} p_\zeta(x')}{\overline{F}_\zeta(x)} \le 1 + \frac{2\epsilon_5}{\lambda_1}$$

for $x > x_0$ and $\epsilon_5 \le \lambda_1/2$ by (A.13). Since $\epsilon_5 = \sum_{\ell=3}^m \epsilon_{5,\ell} = o(1)$, we have the second term in (A.18) is $1+o(1)$ as well. Therefore, we obtain the right-hand side of (A.15) is $1+o(1)$ uniformly over $x \in (0, o(n^\theta))$. Consequently,

$$\frac{\mathbb{P}\Big\{ \binom{m}{2}^{-1}(n-1)\widehat{U}_n > x \Big\}}{\mathbb{P}\Big\{ \sum_{v=1}^\infty \lambda_v(\xi_v^2 - 1) > x \Big\}} - 1 \le o(1)$$

uniformly over $x \in (0, o(n^\theta))$. Again a similar derivation yields a corresponding lower bound of order $o(1)$, completing the proof of the general case $m \ge 2$. □

A.2.2 Proof of Theorem 4.2

Proof of Theorem 4.2. Since the marginal distributions are assumed continuous, we may assume, without loss of generality, that they are uniform distributions on $[0,1]$. Theorem 4.1 can then directly apply to the studied kernel $h(\cdot)$ in view of Assumption 2.1.

The main tool in this proof is Theorem 1 in Arratia et al. (1989). Specifically, we use the version presented in Lemma C2 in Han et al. (2017). We let $I := \{(j,k) : 1 \le j < k \le p\}$, and for all $u := (j,k) \in I$, we define $B_u = \{(\ell, v) \in I : \{\ell, v\} \cap \{j,k\} \neq \varnothing\}$ and

$$\eta_u := \eta_{jk} := \binom{m}{2}^{-1}(n-1)\widehat{U}_{jk}.$$

Then the theorem yields that

$$\Big|\mathbb{P}\Big(\max_{u \in I} \eta_u \le t\Big) - \exp(-L_n)\Big| \le A_1 + A_2 + A_3, \tag{A.19}$$

where $L_n = \sum_{u \in I} \mathbb{P}(\eta_u > t)$,

$$A_1 = \sum_{u \in I}\sum_{\beta \in B_u} \mathbb{P}(\eta_u > t)\mathbb{P}(\eta_\beta > t), \qquad A_2 = \sum_{u \in I}\sum_{\beta \in B_u \setminus \{u\}} \mathbb{P}(\eta_u > t, \eta_\beta > t),$$

$$\text{and} \quad A_3 = \sum_{u \in I} \mathbb{E}|\mathbb{P}\{\eta_u > t \mid \sigma(\eta_\beta : \beta \notin B_u)\} - \mathbb{P}(\eta_u > t)|.$$

We now choose an appropriate value of t such that L_n tends to a constant independent of p as $n \to \infty$. Let

$$t = 4\lambda_1 \log p + \lambda_1(\mu_1 - 2)\log\log p - \Lambda + \lambda_1 y \asymp 4\lambda_1 \log p = o(n^\theta).$$

By Theorem 4.1,

$$L_n = \frac{p(p-1)}{2}\mathbb{P}(\eta_{12} > t) = \frac{p(p-1)}{2}\overline{F}_\zeta(t)\{1+o(1)\}. \tag{A.20}$$

Using Example 5 in Hashorva et al. (2015), we have for any $t > -\Lambda$,

$$\overline{F}_\zeta(t) = \frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{t+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{t+\Lambda}{2\lambda_1}\Big)[1 + O\{(\log p)^{-1}\}]. \tag{A.21}$$

Combining (A.20) and (A.21) yields

$$
\begin{aligned}
L_n &= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}\Big(\frac{t+\Lambda}{2\lambda_1}\Big)^{\mu_1/2-1}\exp\Big(-\frac{t+\Lambda}{2\lambda_1}\Big)\{1+o(1)\}\\
&= \frac{p(p-1)}{2}\frac{\kappa}{\Gamma(\mu_1/2)}(2\log p)^{\mu_1/2-1}\exp\Big\{-2\log p-\Big(\frac{\mu_1}{2}-1\Big)\log\log p-\frac{y}{2}\Big\}\{1+o(1)\}\\
&= \frac{2^{\mu_1/2-2}\kappa}{\Gamma(\mu_1/2)}\exp\Big(-\frac{y}{2}\Big)\{1+o(1)\}, \qquad\qquad (A.22)
\end{aligned}
$$

where $\kappa := \prod_{v=\mu_1+1}(1-\lambda_v/\lambda_1)^{-1/2}$.

Next we bound A_1, A_2, and A_3 separately. We have

$$
A_1 = \tfrac{1}{2}p(p-1)(2p-3)\{\mathbb{P}(\eta_{12}>t)\}^2.
$$

Moreover, since Hoeffding's D is a rank-based U-statistic, Proposition 2.1(ii) yields that η_u is independent of η_β for all $u\in I, \beta\in B_u\backslash\{u\}$. Hence,

$$
A_2 = \sum_{u\in I}\sum_{\beta\in B_u\backslash\{u\}}\mathbb{P}(\eta_u>t)\mathbb{P}(\eta_\beta>t) = p(p-1)(p-2)\{\mathbb{P}(\eta_{12}>t)\}^2.
$$

Again, by Proposition 2.1(iii), we have $A_3=0$. Accordingly,

$$
A_1+A_2+A_3 \le 2p(p-1)^2\{\mathbb{P}(\eta_{12}>t)\}^2 = \frac{2(2L_n)^2}{p} = O\Big(\frac{1}{p}\Big). \qquad (A.23)
$$

Let $L=\exp(-y/2)\cdot 2^{\mu_1/2-2}\kappa/\Gamma(\mu_1/2)$. Combining (A.19)–(A.23) yields

$$
\begin{aligned}
&\Big|\mathbb{P}\Big\{\binom{m}{2}^{-1}(n-1)\max_{j<k}\widehat{U}_{jk}-4\lambda_1\log p-\lambda_1(\mu_1-2)\log\log p+\Lambda\le\lambda_1 y\Big\}-\exp(-L)\Big|\\
&\le \Big|\mathbb{P}\Big(\max_{u\in I}\eta_\alpha\le t\Big)-\exp(-L_n)\Big| + \Big|\exp(-L_n)-\exp(-L)\Big| = o(1).
\end{aligned}
$$

This completes the proof. □

A.2.3 Proof of Corollary 4.1

Proof of Corollary 4.1. We only give the proof for Hoeffding's D here. The proofs for the other two tests are very similar and hence omitted. As in the proof of Theorem 4.2, we may assume the margins to be uniformly distributed on $[0,1]$ without loss of generality. To employ Theorem 4.2, we only need to compute θ. We claim that

$$
\sum_{v=K+1}^{\infty}\lambda_v \asymp \frac{(\log K)^2}{K}. \qquad (A.24)
$$

If this claim is true, then by the definition of θ, one obtains $\theta=1/8-\delta$, where δ is an arbitrarily small pre-specified positive absolute constant.

We now prove (A.24). Notice that the K largest eigenvalues are corresponding to the K smallest products ij, $i,j\in\mathbb{Z}^+$. We begin by assuming that there exists an integer M such that the number of pairs (i,j) satisfying $ij \le M$ is exactly K:

$$2\sum_{i=1}^{\lfloor M^{1/2}\rfloor} \lfloor M/i\rfloor - \lfloor M^{1/2}\rfloor^2 = K. \tag{A.25}$$

To analyze $\sum_{v=K+1}^{\infty}\lambda_v$, we first quantify M. An upper bound on $\sum_{i=1}^{\lfloor M^{1/2}\rfloor}\lfloor M/i\rfloor$ is

$$\sum_{i=1}^{\lfloor M^{1/2}\rfloor}\Big\lfloor\frac{M}{i}\Big\rfloor \le \sum_{i=1}^{\lfloor M^{1/2}\rfloor}\frac{M}{i} = M\sum_{i=1}^{\lfloor M^{1/2}\rfloor}\frac{1}{i} \le M\Big(\log\lfloor M^{1/2}\rfloor + 1\Big) \le M\Big(\frac{1}{2}\log M + 1\Big),$$

and a lower bound is

$$\begin{aligned}\sum_{i=1}^{\lfloor M^{1/2}\rfloor}\Big\lfloor\frac{M}{i}\Big\rfloor &\ge \sum_{i=1}^{\lfloor M^{1/2}\rfloor}\Big(\frac{M}{i}-1\Big) = M\sum_{i=1}^{\lfloor M^{1/2}\rfloor}\frac{1}{i} - \lfloor M^{1/2}\rfloor\\ &\ge M\log\lfloor M^{1/2}\rfloor - \lfloor M^{1/2}\rfloor \ge M\log(M^{1/2}-1) - M^{1/2}.\end{aligned}$$

Thus we have $M\log M \asymp K$, which implies that $M \asymp K/\log K$. Then we obtain

$$\begin{aligned}\sum_{v=K+1}^{\infty}\lambda_v &\asymp \sum_{i=1}^{\lfloor M^{1/2}\rfloor}\sum_{j=\lfloor M/i\rfloor+1}^{\infty}\frac{1}{i^2j^2} + \sum_{j=1}^{\lfloor M^{1/2}\rfloor}\sum_{i=\lfloor M/j\rfloor+1}^{\infty}\frac{1}{i^2j^2} + \sum_{i=\lfloor M^{1/2}\rfloor+1}^{\infty}\sum_{j=\lfloor M^{1/2}\rfloor+1}^{\infty}\frac{1}{i^2j^2}\\ &\asymp \sum_{i=1}^{\lfloor M^{1/2}\rfloor}\frac{1}{i^2(M/i)} + \sum_{j=1}^{\lfloor M^{1/2}\rfloor}\frac{1}{(M/j)j^2} + \frac{1}{(M^{1/2})(M^{1/2})}\\ &\asymp 2\Big\{\frac{\log(M^{1/2})}{M}\Big\} + \frac{1}{M} \asymp \frac{(\log K)^2}{K}.\end{aligned}$$

If there is no integer M such that (A.25) holds, then we pick the largest integer M_1 and the smallest integer M_2 such that

$$2\sum_{i=1}^{\lfloor M_1^{1/2}\rfloor}\Big\lfloor\frac{M_1}{i}\Big\rfloor - \lfloor M_1^{1/2}\rfloor^2 < K < 2\sum_{i=1}^{\lfloor M_2^{1/2}\rfloor}\Big\lfloor\frac{M_2}{i}\Big\rfloor - \lfloor M_2^{1/2}\rfloor^2,$$

and let K_1 and K_2 denote the left-hand side and the right-hand side, respectively. One can verify that $K_1 > K/2$ and $K_2 < 2K$ for sufficiently large K. Then we have

$$\begin{aligned}&\sum_{v=K+1}^{\infty}\lambda_v \le \sum_{v=K_1+1}^{\infty}\lambda_v \asymp \frac{(\log K_1)^2}{K_1} \le \frac{2(\log K)^2}{K}\\ \text{and}\quad &\sum_{v=K+1}^{\infty}\lambda_v \ge \sum_{v=K_2+1}^{\infty}\lambda_v \asymp \frac{(\log K_2)^2}{K_2} \ge \frac{(\log K)^2}{2K}.\end{aligned}$$

Therefore, the asymptotic result for $\sum_{v=K+1}^{\infty}\lambda_v$ given in (A.24) still holds. $\square$

A.2.4 Proof of Lemma 4.1

Proof of Lemma 4.1. Again we only prove the claim for Hoeffding's D; Blum–Kiefer–Rosenblatt's R and Bergsma–Dassios–Yanagimoto's τ^* can be treated similarly. We first establish the fact that $D_{jk} \asymp \Sigma_{jk}^2$ as $\Sigma_{jk} \to 0$. Let $\{(X_{ij}, X_{ik})^\top : i \in [5]\}$ be a collection of independent and identically distributed random vectors that follow a bivariate normal distribution with mean $(0,0)^\top$ and covariance matrix

$$\begin{bmatrix} 1 & \Sigma_{jk} \\ \Sigma_{jk} & 1 \end{bmatrix}.$$

We have

$$D_{jk} = \mathbb{E}_{jk} h_D = \int h_D(x_{1j}, x_{1k}, \ldots, x_{5j}, x_{5k}) \phi(x_{1j}, x_{1k}, \ldots, x_{5j}, x_{5k}; \Sigma_{jk}) \prod_{i=1}^5 dx_{ij} \prod_{i=1}^5 dx_{ik},$$

where

$$\phi(x_{1j}, x_{1k}, \ldots, x_{5j}, x_{5k}; \Sigma_{jk}) = \prod_{i=1}^5 \phi(x_{ij}, x_{ik}; \Sigma_{jk}),$$

and

$$\phi(x_{ij}, x_{ik}; \Sigma_{jk}) = \frac{1}{2\pi(1-\Sigma_{jk}^2)^{1/2}} \exp\Big\{ - \frac{x_{ij}^2 + x_{ik}^2 - 2\Sigma_{jk} x_{ij} x_{ik}}{2(1-\Sigma_{jk}^2)} \Big\}$$

is the joint density of $(X_{ij}, X_{ik})^\top$. Notice that D_{jk} is smooth with respect to Σ_{jk}:

$$\frac{\partial^s D_{jk}}{\partial \Sigma_{jk}^s} = \int h_D(x_{1j}, x_{1k}, \ldots, x_{5j}, x_{5k}) \frac{\partial^s \phi(x_{1j}, x_{1k}, \ldots, x_{5j}, x_{5k}; \Sigma_{jk})}{\partial \Sigma_{jk}^s} \prod_{i=1}^5 dx_{ij} \prod_{i=1}^5 dx_{ik}.$$

In order to prove $D_{jk} \asymp \Sigma_{jk}^2$, it suffices to establish that $D_{jk} = 0$ when $\Sigma_{jk} = 0$, the first derivative of D_{jk} with respect to Σ_{jk} is 0 at $\Sigma_{jk} = 0$, and the second derivative of D_{jk} with respect to Σ_{jk} is $5/\pi^2$ at $\Sigma_{jk} = 0$, which can be confirmed by a lengthy but straightforward computation.

Now we turn to our claim. Recall that $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathbb{P}_{jk})\} = 0$ when $\Sigma_{jk} = 0$. We will show that the first-order term in the Taylor series of $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathbb{P}_{jk})\}$ with respect to Σ_{jk} is also 0. Suppose, for contradiction, the first-order coefficient (denoted by a_1) in the Taylor series of $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathbb{P}_{jk})\}$ is not 0, then for Σ_{jk} in a sufficiently small neighborhood of 0, $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathbb{P}_{jk})\} < 0$ for $\Sigma_{jk} < 0$ if $a_1 > 0$, and for $\Sigma_{jk} > 0$ if $a_1 < 0$, which contradicts the definition of $\mathrm{Var}_{jk}\{h_D^{(1)}(\cdot; \mathbb{P}_{jk})\}$. This together with $\mathbb{E}_{jk} h_D \asymp \Sigma_{jk}^2$ completes the proof. □

A.2.5 Proof of Theorem 4.3

Proof of Theorem 4.3. It is clear that we only have to consider $\max_{j<k} U_{jk} = C_\gamma (\log p / n)$ for some sufficiently large C_γ. The main idea here is to bound $\max_{j<k} \widehat{U}_{jk} - \max_{j<k} U_{jk}$ with high probability. To do this, we first construct a concentration inequality for $|\widehat{U}_{jk} - U_{jk}|$. The Hoeffding decomposition of the difference is

$$\widehat{U}_{jk} - U_{jk} = \frac{m}{n} \sum_{i=1}^n h^{(1)}(\boldsymbol{X}_{i,\{j,k\}}; \mathbb{P}_{jk}) + \sum_{\ell=2}^k \binom{m}{\ell} H_n^{(\ell)}(\cdot; \mathbb{P}_{jk}). \tag{A.26}$$

For controlling the first term in (A.26), recall that $\|h\|_\infty \le b_1 < \infty$, and then $h^{(1)}(\cdot; \mathbb{P}_{jk}) = h_1(\cdot; \mathbb{P}_{jk}) - \mathbb{E}h$ is bounded by $2b_1$ almost surely and $\mathbb{E}h^{(1)}(\cdot; \mathbb{P}_{jk}) = 0$. We then apply Bernstein's inequality, giving

$$\mathbb{P}\Big\{\frac{m}{n}\Big|\sum_{i=1}^{n} h^{(1)}(\cdot;\mathbb{P}_{jk})\Big| > t_1\Big\} \le \exp\Big(-\frac{n(t_1/m)^2}{2[\mathrm{Var}_{jk}\{h^{(1)}(\cdot;\mathbb{P}_{jk})\} + 2b_1(t_1/m)/3]}\Big). \tag{A.27}$$

By the definition of the distribution family $\mathcal{D}(\gamma, p; h)$, we have

$$\mathrm{Var}_{jk}\{h^{(1)}(\cdot;\mathbb{P}_{jk})\} \le \gamma\mathbb{E}_{jk}h = \gamma U_{jk} \le \gamma C_\gamma(\log p/n).$$

Plugging this into (A.27) and taking $t_1 = C_1(\log p/n)$, where C_1 is a constant to be specified later, yields

$$\begin{aligned}
&\mathbb{P}\Big\{\frac{m}{n}\Big|\sum_{i=1}^{n} h^{(1)}(\cdot;\mathbb{P}_{jk})\Big| > C_1\frac{\log p}{n}\Big\} \\
&\le 2\exp\Big\{-\frac{C_1^2\log p}{2(m^2\gamma C_\gamma + 2mb_1C_1/3)}\Big\} = 2\Big(\frac{1}{p}\Big)^{C_1^2/(2m^2\gamma C_\gamma + 4mb_1C_1/3)}.
\end{aligned} \tag{A.28}$$

We then handle the remaining term. By Theorem 4.1.12 in de la Peña and Giné (1999), there exist absolute constants $C_\ell', C_\ell'' > 0$ such that for all $t \in (0,1]$, $2 \le \ell \le m$,

$$\begin{aligned}
&\mathbb{P}(|H_n^{(\ell)}(\cdot;\mathbb{P}_{jk})| \ge t) \le C_\ell'\exp\Big\{-C_\ell'' n\Big(\frac{t}{\|h^{(\ell)}(\cdot;\mathbb{P}_{jk})\|_\infty}\Big)^{2/\ell}\Big\} \\
&\le C_\ell'\exp\Big\{-C_\ell'' n\Big(\frac{t}{2^\ell b_1}\Big)^{2/\ell}\Big\} \le C_\ell'\exp\Big(-\frac{C_\ell'' nt}{4b_1^{2/\ell}}\Big),
\end{aligned}$$

which further implies that

$$\begin{aligned}
\mathbb{P}\Big\{\Big|\sum_{\ell=2}^{k}\binom{m}{\ell}H_n^{(\ell)}(\cdot;\mathbb{P}_{jk})\Big| \ge t_2\Big\} &\le \sum_{\ell=2}^{m}\mathbb{P}\Big\{\binom{m}{\ell}|H_n^{(\ell)}(\cdot;\mathbb{P}_{jk})| \ge t_2\frac{4b_1^{2/\ell}\binom{m}{\ell}/C_\ell''}{\sum_{\ell=2}^m 4b_1^{2/\ell}\binom{m}{\ell}/C_\ell''}\Big\} \\
&\le \Big(\sum_{\ell=2}^{m}C_\ell'\Big)\exp\Big\{-\frac{nt_2}{\sum_{\ell=2}^m 4b_1^{2/\ell}\binom{m}{\ell}/C_\ell''}\Big\}.
\end{aligned}$$

Taking $t_2 = C_2(\log p/n)$, where C_2 is another constant to be specified later, we have

$$\mathbb{P}\Big\{\Big|\sum_{\ell=2}^{k}\binom{m}{\ell}H_n^{(\ell)}(\cdot;\mathbb{P}_{jk})\Big| \ge C_2(\log p/n)\Big\} \le \Big(\sum_{\ell=2}^{m}C_\ell'\Big)\Big(\frac{1}{p}\Big)^{C_2/\{\sum_{\ell=2}^m 4b_1^{2/\ell}\binom{m}{\ell}/C_\ell''\}}. \tag{A.29}$$

Putting (A.28) and (A.29) together, and choosing

$$C_1 = 2mb_1 + m(4b_1^2 + 6\gamma C_\gamma)^{1/2} \quad \text{and} \quad C_2 = 12\sum_{\ell=2}^{m}\frac{b_1^{2/\ell}\binom{m}{\ell}}{C_\ell''},$$

we deduce

$$\mathbb{P}\Big[|\widehat{U}_{jk} - U_{jk}| \ge \Big\{2mb_1 + m(4b_1^2 + 6\gamma C_\gamma)^{1/2} + 12\sum_{\ell=2}^{m}\frac{b_1^{2/\ell}\binom{m}{\ell}}{C_\ell''}\Big\}\frac{\log p}{n}\Big] \le \Big(2 + \sum_{\ell=2}^{m}C_\ell'\Big)\frac{1}{p^3}.$$

Then using Slutsky's argument gives

$$\mathbb{P}\Big[\max_{j<k}|\widehat{U}_{jk} - U_{jk}| \ge \Big\{2mb_1 + m(4b_1^2 + 6\gamma C_\gamma)^{1/2} + 12\sum_{\ell=2}^{m}\frac{b_1^{2/\ell}\binom{m}{\ell}}{C_\ell''}\Big\}\frac{\log p}{n}\Big] \le \frac{2 + \sum_{\ell=2}^m C_\ell'}{2}\cdot\frac{1}{p},$$

which implies that, with probability at least $1-(1+\sum_{\ell=2}^m C'_\ell/2)p^{-1}$,

$$\max_{j<k}|\widehat{U}_{jk}-U_{jk}| \le \Big\{2mb_1+m(4b_1^2+6\gamma C_\gamma)^{1/2}+12\sum_{\ell=2}^m \frac{b_1^{2/\ell}\binom{m}{\ell}}{C''_\ell}\Big\}\frac{\log p}{n}.$$

Hence for $n\ge 2$, we have with probability no smaller than $1-(1+\sum_{\ell=2}^m C'_\ell/2)p^{-1}$,

$$\max_{j<k}\widehat{U}_{jk} \ge \max_{j<k}U_{jk}-\max_{j<k}|\widehat{U}_{jk}-U_{jk}|$$
$$\ge \Big\{C_\gamma-2mb_1-m(4b_1^2+6\gamma C_\gamma)^{1/2}-12\sum_{\ell=2}^m \frac{b_1^{2/\ell}\binom{m}{\ell}}{C''_\ell}\Big\}\frac{\log p}{n} \ge \frac{5\lambda_1\binom{m}{2}\log p}{n-1},$$

where the last inequality is satisfied by choosing C_γ large enough. Accordingly, for any given Q_α, the probability that

$$\frac{n-1}{\lambda_1\binom{m}{2}}\max_{j<k}\widehat{U}_{jk} \ge 5\log p > 4\log p+(\mu_1-2)\log\log p-\frac{\Lambda}{\lambda_1}+Q_\alpha$$

tends to 1 as p goes to infinity. The proof is thus completed. □

A.2.6 Proof of Theorem 4.4

Proof of Theorem 4.4. In view of Corollary 4.2, the results follow from Lemma 4.1 and the fact that $D_{jk}, R_{jk}, \tau^*_{jk} \asymp \Sigma_{jk}^2$ as $\Sigma_{jk}\to 0$, which has been shown in the proof of Lemma 4.1. □

A.3 Proofs for Section 6 of the main paper

A.3.1 Proof of Proposition 6.1

Proof of Proposition 6.1. We prove identities (i)–(iii) sequentially. Let Ψ_1,Ψ_2,Ψ_3 denote the expressions on the right-hand side of identities (i), (ii), (iii), respectively.

Identity (i). Let $\{(X_i,Y_i)^\top\}_{i\in[4]}$ be four independent realizations of $(X,Y)^\top$. For Bergsma–Dassios–Yanagimoto's τ^*, we have, by Equation (6) in Bergsma and Dassios (2014a),

$$\begin{aligned}\frac{1}{18}\mathbb{E}h_{\tau^*} = {}& \frac{1}{3}\mathbb{P}\{\max(X_1,X_2)<\min(X_3,X_4),\ \max(Y_1,Y_2)<\min(Y_3,Y_4)\}\\ &+\frac{1}{3}\mathbb{P}\{\max(X_1,X_2)<\min(X_3,X_4),\ \max(Y_3,Y_4)<\min(Y_1,Y_2)\}\\ &-\frac{2}{3}\mathbb{P}\{\max(X_1,X_2)<\min(X_3,X_4),\ \max(Y_1,Y_3)<\min(Y_2,Y_4)\}. \end{aligned}\tag{A.30}$$

We study the three terms in (A.30) separately, starting from the first term. Using Fubini's theorem, we get

$$\begin{aligned}&\mathbb{P}\{\max(X_1,X_2)<\min(X_3,X_4),\ \max(Y_1,Y_2\}<\min(Y_3,Y_4)\}\\ ={}&\int \mathbb{P}\{\max(X_1,X_2)<x,\ \max(Y_1,Y_2)<y\}d\mathbb{P}\{\min(X_3,X_4)\le x,\ \min(Y_3,Y_4)\le y\}\\ ={}&\int F(x,y)^2 d\mathbb{P}\{\min(X_3,X_4)\le x,\ \min(Y_3,Y_4)\le y\}, \end{aligned}\tag{A.31}$$

where

$$\mathbb{P}\{\min(X_3, X_4) \le x,\ \ \min(Y_3, Y_4) \le y\} = \mathbb{P}\{(A \cup B) \cap (C \cup D)\} = \mathbb{P}(I \cup II \cup III \cup IV)$$

and $A := \{X_3 \le x\}$, $B := \{X_4 \le X\}$, $C := \{Y_3 \le y\}$, $D := \{Y_4 \le y\}$,

$$I := A \cap C = \{X_3 \le x, Y_3 \le y\}, \qquad II := A \cap D = \{X_3 \le x, Y_4 \le y\},$$
$$III := B \cap C = \{X_4 \le x, Y_3 \le y\}, \qquad IV := B \cap D = \{X_4 \le x, Y_4 \le y\}.$$

From the inclusion–exclusion principle, we obtain

$$\begin{aligned}
&\mathbb{P}\{\min(X_3, X_4) \le x,\ \ \min(Y_3, Y_4) \le y\} \\
&= \mathbb{P}(I) + \mathbb{P}(II) + \mathbb{P}(III) + \mathbb{P}(IV) \\
&\quad - \mathbb{P}(I \cap II) - \mathbb{P}(I \cap III) - \mathbb{P}(I \cap IV) - \mathbb{P}(II \cap III) - \mathbb{P}(II \cap IV) - \mathbb{P}(III \cap IV) \\
&\quad + \mathbb{P}(I \cap II \cap III) + \mathbb{P}(I \cap II \cap IV) + \mathbb{P}(I \cap III \cap IV) + \mathbb{P}(II \cap III \cap IV) \\
&\quad - \mathbb{P}(I \cap II \cap III \cap IV) \\
&= F + F_1F_2 + F_1F_2 + F - FF_2 - FF_1 - F^2 - F^2 - FF_1 - FF_2 + F^2 + F^2 + F^2 + F^2 - F^2 \\
&= 2F + 2F_1F_2 - 2FF_1 - 2FF_2 + F^2.
\end{aligned} \tag{A.32}$$

Plugging (A.32) into (A.31) implies that

$$\begin{aligned}
&\mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_2\} < \min(Y_3, Y_4)\} \\
&= \int F^2 d(2F + 2F_1F_2 - 2FF_1 - 2FF_2 + F^2).
\end{aligned} \tag{A.33}$$

The second term in (A.30) can be written as

$$\begin{aligned}
&\mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_3, Y_4) < \min(Y_1, Y_2)\} \\
&= \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4)\} \\
&\quad - \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \min(Y_1, Y_2) \le \max(Y_3, Y_4)\} \\
&= \int \mathbb{P}\{\max(X_1, X_2) < x\} d\mathbb{P}\{\min(X_3, X_4) \le x\} \\
&\quad - \mathbb{P}\{\max(X_1, X_2) < x,\ \ \min(Y_1, Y_2) \le y\} d\mathbb{P}\{\min(X_3, X_4) \le x,\ \ \max(Y_3, Y_4) \le y\},
\end{aligned} \tag{A.34}$$

where we have

$$\mathbb{P}\{\min(X_3, X_4) \le x\} = \mathbb{P}(A \cup B) = 2F_1 - F_1^2, \tag{A.35}$$

and

$$\begin{aligned}
&\mathbb{P}\{\max(X_1, X_2) < x,\ \ \min(Y_1, Y_2) \le y\} \\
&= \mathbb{P}\{\max(X_1, X_2) < x,\ \ Y_1 \le y\} + \mathbb{P}\{\max(X_1, X_2) < x,\ \ Y_2 \le y\} \\
&\quad - \mathbb{P}\{\max(X_1, X_2) < x,\ \ \max(Y_1, Y_2) \le y\} \\
&= 2FF_1 - F^2,
\end{aligned} \tag{A.36}$$

and

$$
\begin{aligned}
&\mathbb{P}(\min\{X_3, X_4\} \le x,\ \ \max\{Y_3, Y_4\} \le y) \\
&= \mathbb{P}[\{(X_3 \le x) \cup (X_4 \le x)\} \cap \{(Y_3 \le y) \cap (Y_4 \le y)\}] \\
&= \mathbb{P}[\{A \cap (C \cap D)\} \cup \{B \cap (C \cap D)\}] \\
&= 2FF_2 - F^2.
\end{aligned} \tag{A.37}
$$

Plugging (A.35)–(A.37) into (A.34) yields

$$
\begin{aligned}
&\mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_3, Y_4) < \min(Y_1, Y_2)\} \\
&= \int F_1^2 d(2FF_1 - F^2) - \int (2FF_1 - F^2) d(2FF_2 - F^2).
\end{aligned} \tag{A.38}
$$

Next we handle the third term in (A.30). We have by symmetry that

$$
\begin{aligned}
&\mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_3) < \min(Y_2, Y_4)\} \\
&= \mathbb{P}\{\max(X_1, X_2) < \min(X_4, X_3),\ \ \max(Y_1, Y_4) < \min(Y_2, Y_3)\} \\
&= \mathbb{P}\{\max(X_2, X_1) < \min(X_3, X_4),\ \ \max(Y_2, Y_3) < \min(Y_1, Y_4)\} \\
&= \mathbb{P}\{\max(X_2, X_1) < \min(X_4, X_3),\ \ \max(Y_2, Y_4) < \min(Y_1, Y_3)\},
\end{aligned} \tag{A.39}
$$

and notice that

$$
\begin{aligned}
&\mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4) \\
&= \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_2) < \min(Y_3, Y_4)\} \\
&\quad + \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_3, Y_4) < \min(Y_1, Y_2)\} \\
&\quad + \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_3) < \min(Y_2, Y_4)\} \\
&\quad + \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_4) < \min(Y_2, Y_3)\} \\
&\quad + \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_2, Y_3) < \min(Y_1, Y_4)\} \\
&\quad + \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_2, Y_4) < \min(Y_1, Y_3)\}
\end{aligned} \tag{A.40}
$$

assuming marginal continuity of $(X, Y)^\top$. Combining (A.39) and (A.40) gives

$$
\begin{aligned}
&\mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_3) < \min(Y_2, Y_4)\} \\
&= \frac{1}{4}\Big[\mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4) \\
&\quad - \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_1, Y_2) < \min(Y_3, Y_4)\} \\
&\quad - \mathbb{P}\{\max(X_1, X_2) < \min(X_3, X_4),\ \ \max(Y_3, Y_4) < \min(Y_1, Y_2)\}\Big] \\
&= \frac{1}{4}\Big\{\int (2FF_1 - F^2) d(2FF_2 - F^2) - \int F^2 d(2F + 2F_1F_2 - 2FF_1 - 2FF_2 + F^2)\Big\}.
\end{aligned} \tag{A.41}
$$

The identity (i) follows by plugging (A.33), (A.38), and (A.41) into (A.30).

Identity (ii)**.** Next we prove that $\Psi_1 - \Psi_2 = 0$. A straightforward computation gives

$$
\begin{aligned}
&\Psi_1 - \Psi_2 \\
&= -\int F^2 d(F_1F_2) - \int F^2 d(FF_1) - 2\int FF_1 d(FF_2) + \int FF_1 d(F^2) + 2\int FF_1F_2 dF + \frac{1}{6} \\
&= -\iint F^2 \frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}dxdy - \iint F^2 \frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}dxdy + \iint F^2 F_1 \frac{\partial^2 F}{\partial x \partial y}dxdy \\
&\quad - 2\iint FF_1 \frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}dxdy + 2\iint FF_1 \frac{\partial F}{\partial x}\frac{\partial F}{\partial y}dxdy + \frac{1}{6}.
\end{aligned} \tag{A.42}
$$

To further simplify (A.42), notice that

$$
\iint \Big(F^2 \frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y} + 2FF_1\frac{\partial F}{\partial x}\frac{\partial F}{\partial y} + F^2F_1\frac{\partial^2 F}{\partial x\partial y}\Big)dxdy = \iint \frac{\partial^2(F^3F_1/3)}{\partial x \partial y}dxdy = \frac{1}{3}. \tag{A.43}
$$

Adding (A.42) and (A.43) together yields

$$
\begin{aligned}
\Psi_1 - \Psi_2 &= -\Big(\iint F^2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}dxdy + 2\iint FF_1\frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}dxdy\Big) - 2\iint F^2 \frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}dxdy + \frac{1}{2} \\
&= -\int \frac{\partial F_2}{\partial y}\int \frac{\partial (F^2F_1)}{\partial x}dxdy - 2\int \frac{\partial F_1}{\partial x}\int F^2 \frac{\partial F}{\partial y}dydx + \frac{1}{2} \\
&= -\int \frac{\partial F_2}{\partial y}F_2^2 dy - 2\int \frac{\partial F_1}{\partial x}\frac{F_1^3}{3}dx + \frac{1}{2} \\
&= -\frac{1}{3} - 2\Big(\frac{1}{12}\Big) + \frac{1}{2} = 0,
\end{aligned}
$$

which completes the proof of identity (ii).

Identity (iii)**.** This identity was discovered by Yanagimoto (1970). To see this, it suffices to show that

$$
\Psi_3 - \Psi_2 = \int F_1^2F_2^2 dF - 4\int FF_1F_2 dF_1 dF_2 + 2\int F_1^2F_2^2 dF_1 dF_2 + \frac{1}{9} = 0.
$$

We start from the identity

$$
\begin{aligned}
1 = \iint \frac{\partial^2(FF_1^2F_2^2)}{\partial x\partial y}dxdy = &\iint F_1^2F_2^2\frac{\partial^2 F}{\partial x \partial y}dxdy + \iint 4FF_1F_2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}dxdy \\
&+ \iint 2F_1^2F_2\frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}dxdy + \iint 2F_1F_2^2\frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}dxdy.
\end{aligned} \tag{A.44}
$$

We also note that

$$
\begin{aligned}
&\iint 2FF_1F_2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}dxdy + \iint F_1^2F_2\frac{\partial F}{\partial x}\frac{\partial F_2}{\partial y}dxdy \\
&= \int F_2\frac{\partial F_2}{\partial y}\int \frac{\partial (FF_1^2)}{\partial x}dxdy \\
&= \int F_2^2\frac{\partial F_2}{\partial y}dy = \frac{1}{3},
\end{aligned} \tag{A.45}
$$

and

$$\begin{aligned}
&\iint 2FF_1F_2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}dxdy + \iint F_1F_2^2\frac{\partial F_1}{\partial x}\frac{\partial F}{\partial y}dxdy \\
&= \int F_1\frac{\partial F_1}{\partial x}\int\frac{\partial(FF_2^2)}{\partial y}dydx \\
&= \int F_1^2\frac{\partial F_1}{\partial x}dx = \frac{1}{3},
\end{aligned} \tag{A.46}$$

and

$$\iint 2F_1^2F_2^2\frac{\partial F_1}{\partial x}\frac{\partial F_2}{\partial y}dxdy = 2\int F_1^2\frac{\partial F_1}{\partial x}dx\int F_2^2\frac{\partial F_2}{\partial y}dy = 2\Big(\frac{1}{3}\Big)\Big(\frac{1}{3}\Big) = \frac{2}{9}. \tag{A.47}$$

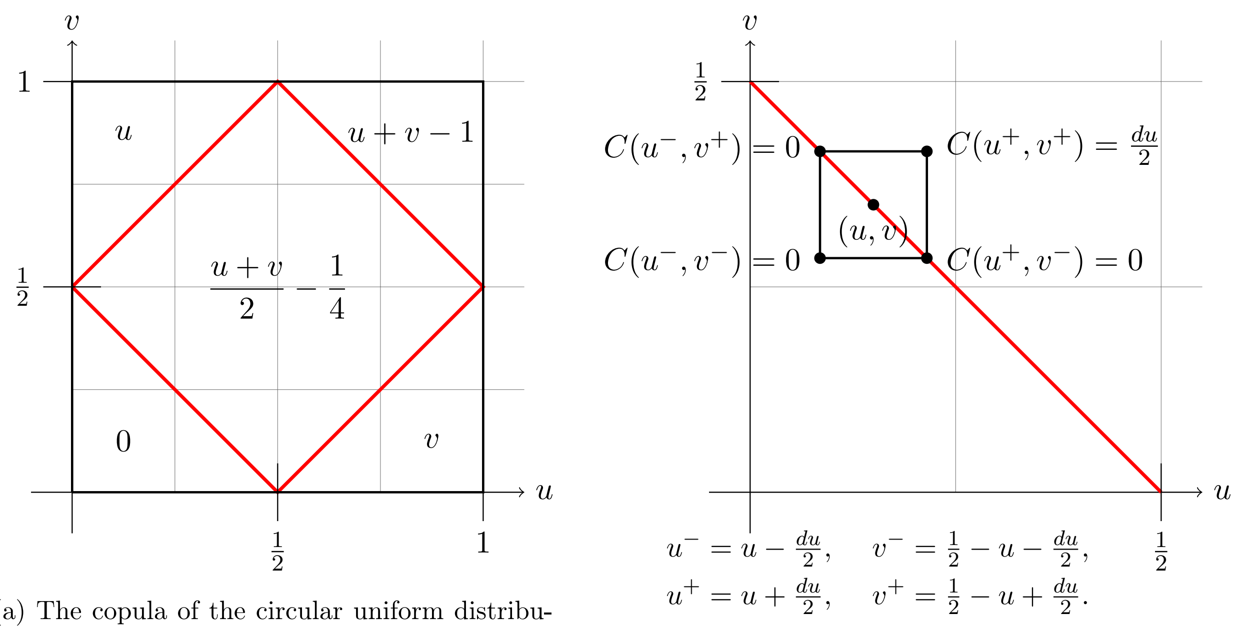
Combining (A.44)–(A.47) concludes the claim. □

A.3.2 Proof of Proposition 6.2

Proof of Proposition 6.2. The copula of $(X,Y)^\top$ is given by Nelsen (2006, p. 56):

$$C(u,v) = \begin{cases} \min(u,v), & \text{if } |u-v| \geq \frac{1}{2}, \\ \max(0,u+v-1), & \text{if } |u+v-1| \geq \frac{1}{2}, \\ \frac{u+v}{2}-\frac{1}{4}, & \text{otherwise.} \end{cases}$$

We summarize the copula in Figure 1a.



(a) The copula of the circular uniform distribution with its support marked in red.

(b) Integral on part of the support.

Figure 1: The copula of the circular uniform distribution.

Since both X and Y are continuous, by the arguments in Schweizer and Wolff (1981), we obtain

$$\begin{aligned} \mathbb{E}h_D &= 30\int\{F(x,y) - F_1(x)F_2(y)\}^2 dF(x,y) \\ &= 30\int\{C(u,v) - uv\}^2 dC(u,v) \\ \text{and} \quad \mathbb{E}h_R &= 90\int\{F(x,y) - F_1(x)F_2(y)\}^2 dF_1(x)dF_2(y) \\ &= 90\int\{C(u,v) - uv\}^2 dudv. \end{aligned}$$

We first compute $\mathbb{E}h_D$. Notice that $\partial^2 C(u,v)/\partial u \partial v = 0$ in $[0,1]\times[0,1]$ except for the support of $C(u,v)$ (marked in red in Figure 1a). Therefore, we only need to compute the integral on the support consisting of four line segments. In Figure 1b, we illustrate how to find $dC(u,v)$ on the line segment from $(0,1/2)$ to $(1/2,0)$ (denoted by $\mathcal{C}_1$). We have

$$dC(u,v) = C(u^+,v^+) - C(u^+,v^-) - C(u^-,v^+) + C(u^-,v^-) = \frac{du}{2},$$

and thus the integral on the line segment $\mathcal{C}_1$ is given by

$$30\int_{\mathcal{C}_1}\{C(u,v) - uv\}^2 dC(u,v) = 30\int\left\{0 - u\Big(\frac{1}{2} - u\Big)\right\}^2\frac{du}{2} = \frac{1}{64}.$$

We can evaluate the integral on the other three line segments (denoted by $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$, respectively) similarly, and we find

$$\mathbb{E}h_D = 30\int_{\mathcal{C}_1+\mathcal{C}_2+\mathcal{C}_3+\mathcal{C}_4}\{C(u,v) - uv\}^2 dC(u,v) = \frac{1}{16}.$$

The computation of $\mathbb{E}h_R = 90\int\{C(u,v) - uv\}^2 dudv = 1/16$ is standard, and we omit details. Finally, using the identity (2.3), we deduce that $\mathbb{E}h_{\tau^*} = 1/16$. $\square$