

## TESTING INDEPENDENCE IN HIGH DIMENSIONS WITH SUMS OF RANK CORRELATIONS

BY DENNIS LEUNG AND MATHIAS DRTON<sup>1</sup>

*Chinese University of Hong Kong and University of Washington*

We treat the problem of testing independence between  $m$  continuous variables when  $m$  can be larger than the available sample size  $n$ . We consider three types of test statistics that are constructed as sums or sums of squares of pairwise rank correlations. In the asymptotic regime where both  $m$  and  $n$  tend to infinity, a martingale central limit theorem is applied to show that the null distributions of these statistics converge to Gaussian limits, which are valid with no specific distributional or moment assumptions on the data. Using the framework of U-statistics, our result covers a variety of rank correlations including Kendall's tau and a dominating term of Spearman's rank correlation coefficient ( $\rho$ ), but also degenerate U-statistics such as Hoeffding's  $D$ , or the  $\tau^*$  of Bergsma and Dassios [*Bernoulli* **20** (2014) 1006–1028]. As in the classical theory for U-statistics, the test statistics need to be scaled differently when the rank correlations used to construct them are degenerate U-statistics. The power of the considered tests is explored in rate-optimality theory under a Gaussian equicorrelation alternative as well as in numerical experiments for specific cases of more general alternatives.

**1. Introduction.** This paper is concerned with nonparametric tests of independence between the coordinates of a continuous random vector  $\mathbf{X} = (X^{(1)}, \dots, X^{(m)})$ . Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be an i.i.d. sample, with each  $\mathbf{X}_i = (X_i^{(1)}, \dots, X_i^{(m)})$  following the same distribution as  $\mathbf{X}$ . We then wish to test the null hypothesis

$$(1.1) \quad H_0 : X^{(1)}, \dots, X^{(m)} \text{ are independent.}$$

The natural approach is to form a test statistic that measures the dependence among the variables  $X^{(1)}, \dots, X^{(m)}$  based on the sample, and reject  $H_0$  when its value is too large, where the critical value of rejection is calibrated by the asymptotic distribution of the test statistic under the null. Our focus is on the use of rank correlations in problems where the dimension  $m$  can be larger than the sample size  $n$ . Specifically, our testing procedures will be studied under the asymptotic regime where  $m = m(n)$  grows as a function of  $n$  such that  $m$  also tends to infinity. This regime is denoted by  $m, n \rightarrow \infty$  throughout our paper.

---

Received October 2015; revised December 2016.

<sup>1</sup>Supported in part by NSF Grant DMS-15-61814.

*MSC2010 subject classifications.* 60K35.

*Key words and phrases.* High-dimensional statistics, independence, U-statistics, minimax optimality, rank correlations.

There is a vast literature on the problem of testing independence. If  $\mathbf{X}$  is normal, then under the traditional asymptotic setup in which  $n$  goes to  $\infty$  while  $m$  is fixed, the likelihood ratio test (LRT) statistic converges to a chi-square distribution when  $H_0$  is true [Anderson \(2003\)](#). This test is known to be unimplementable for  $m > n$  due to the singularity of the sample covariance matrix, but recent work of [Jiang and Qi \(2015\)](#), Corollary 1, shows asymptotic normality for the LRT statistic under the regime where  $m, n \rightarrow \infty$  while  $n > m + 4$ . When  $m$  can actually be larger than  $n$ , one line of work uses the *maximum* of many pairwise dependency measures to test for (1.1). For  $p = 1, \dots, m$ , let  $\mathbf{X}^{(p)} = (X_1^{(p)}, \dots, X_n^{(p)})$  be the sample of observations for the  $p$ th variable. For  $1 \leq p \neq q \leq m$ , let  $r^{(pq)}$  denote the sample Pearson (product-moment) correlation of  $\mathbf{X}^{(p)}$  and  $\mathbf{X}^{(q)}$ . [Jiang \(2004\)](#) proved that, under suitable centering and scaling, the null distribution of the statistic

$$(1.2) \quad \max_{1 \leq p < q \leq m} (r^{(pq)})^2$$

converges to an extreme value distribution of type 1 when  $m/n$  converges to a constant  $\gamma \in (0, \infty)$  as  $m, n \rightarrow \infty$ . We will abbreviate such convergence as  $m/n \rightarrow \gamma \in (0, \infty)$ . He assumed higher-order moment conditions that were weakened in subsequent work [[Zhou \(2007\)](#), [Liu, Lin and Shao \(2008\)](#), [Li, Liu and Rosalsky \(2010\)](#), [Li, Qi and Rosalsky \(2012\)](#)]. [Cai and Jiang \(2011\)](#) derived a similar asymptotic distribution for the statistic from (1.2), allowing for subexponential growth in the dimension  $m$ . Further weakening distributional assumptions, the recent work of [Han and Liu \(2014\)](#) treated maxima of rank correlations, that is, the sample Pearson correlation in (1.2) is replaced by a rank correlation measure such as Kendall's tau. This maximum was shown to have a similar extreme value type null distribution. Statistics such as (1.2) are of obvious appeal when strong dependence is expected between some variables.

This paper, however, aligns with a different approach that is appealing when moderate dependence is expected between many variables. In this approach, tests are based on estimates of the *sum* of many pairwise dependency signals. Let  $\Sigma = (\sigma^{(pq)})$  and  $R = (\rho^{(pq)})$  be, respectively, the population covariance and Pearson correlation matrix of the random vector  $\mathbf{X}$ . Under a Gaussian assumption for  $\mathbf{X}$ , [Schott \(2005\)](#) proposed the use of the “plug-in” estimate

$$(1.3) \quad S_r := \sum_{1 \leq p < q \leq m} (r^{(pq)})^2$$

for the overall dependency signal  $\sum_{p < q} (\rho^{(pq)})^2$ . Subsequent work of [Chen and Shao \(2012\)](#) obtained a Berry–Esseen bound for this statistic's weak convergence to normality under  $H_0$  as  $m, n \rightarrow \infty$ . The statistic  $S_r$  is in fact Rao's score statistic for the multivariate normal setting; see the [Appendix](#). [Mao \(2014\)](#) suggested a related statistic, namely, the sum of  $f(r^{(pq)})$  for  $f(x) = x^2/(1-x^2)$ , and again the null distribution is shown to be asymptotically normal. For the two related problems of testing the equality and the proportionality of  $\Sigma$  to the identity matrix,

similar statistics have been studied [Ledoit and Wolf (2002), Nagao (1973), John (1972)]. Motivated by this approach, we construct our first class of test statistics by plugging in rank correlations to obtain nonparametric tests for (1.1). We illustrate it here for Kendall’s tau. For  $1 \leq p \neq q \leq m$ , let

$$(1.4) \quad \tau^{(pq)} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \operatorname{sgn}(X_i^{(p)} - X_j^{(p)}) \operatorname{sgn}(X_i^{(q)} - X_j^{(q)})$$

be the sample Kendall’s tau correlation coefficient for  $\mathbf{X}^{(p)}$  and  $\mathbf{X}^{(q)}$ . A natural test is then to reject  $H_0$  for large values of the statistic

$$(1.5) \quad S_\tau := \sum_{1 \leq p < q \leq m} (\tau^{(pq)})^2.$$

As an estimator of the dependency signal

$$(1.6) \quad \sum_{1 \leq p < q \leq m} (\mathbb{E}[\tau^{(pq)}])^2,$$

the “plug-in” statistic  $S_\tau$  from (1.5) is biased and thus needs to be recentered to obtain a mean zero asymptotic null distribution under our considered regime  $m, n \rightarrow \infty$ . Alternatively, we may instead attempt to form an unbiased estimator of (1.6) to serve as a test statistic. As shown in Section 3, such an unbiased estimator is given by

$$(1.7) \quad T_\tau := \frac{1}{4! \binom{n}{4}} \sum \operatorname{sgn}(X_{i_{\pi(1)}}^{(p)} - X_{i_{\pi(2)}}^{(p)}) \operatorname{sgn}(X_{i_{\pi(3)}}^{(p)} - X_{i_{\pi(4)}}^{(p)}) \\ \times \operatorname{sgn}(X_{i_{\pi(1)}}^{(q)} - X_{i_{\pi(2)}}^{(q)}) \operatorname{sgn}(X_{i_{\pi(3)}}^{(q)} - X_{i_{\pi(4)}}^{(q)}),$$

where the summation is over all variable pairs  $1 \leq p < q \leq m$ , ordered 4-tuples of indices  $1 \leq i_1 < i_2 < i_3 < i_4 \leq n$ , and permutations  $\pi$  on four elements. This type of statistics is motivated by the work of Chen, Zhang and Zhong (2010) and Cai and Ma (2013), who tested the equality of  $\Sigma$  to the identity based on unbiased estimates of the squared Frobenius norm  $\|\Sigma - I_m\|_F^2$ , where  $I_m$  is the  $m$ -by- $m$  identity matrix. Under a Gaussian assumption for  $\mathbf{X}$ , Cai and Ma (2013) showed their test to be asymptotically minimax rate optimal.

As a last variant, when testing for *positive* associations, it may be of interest to consider the statistic

$$(1.8) \quad Z_\tau := \sum_{1 \leq p < q \leq m} \tau^{(pq)},$$

which sums all pairwise sample correlations for a “one-sided” test. As we explain below, such a statistic also provides a “two-sided” test for  $H_0$  when rank correlations such as the  $\tau^*$  of Bergsma and Dassios (2014) are used. In Section 4, we

show that all the statistics introduced above are asymptotically normal under suitable recentering and rescaling.

Kendall’s tau is an example of a U-statistic whose values depend on the data only via ranks [van der Vaart \(1998\)](#), Example 12.5. Indeed, the values of (1.4), (1.7) and (1.8) remain unchanged if each observation  $X_i^{(p)}$  is replaced with its rank  $R_i^{(p)}$ . To be specific,  $R_i^{(p)}$  is the rank of  $X_i^{(p)}$  among  $X_1^{(p)}, \dots, X_n^{(p)}$ . Other examples of measures of association that are both U-statistics and rank correlations are the  $D$  of [Hoeffding \(1948a\)](#) and the aforementioned  $\tau^*$  of [Bergsma and Dassios \(2014\)](#). We note that for a pair of continuous random variables both of these statistics lead to consistent tests of independence, that is, their expectations are zero if and only if the two random variables are independent. Another classical example is Spearman’s rho, which is not a U-statistic but can be approximated by a rank-based U-statistic.

The above examples of U-statistics are reviewed in Section 2, which also introduces a general framework of rank-based U-statistics that we adopt for a unified theory. In Section 3, we construct our classes of test statistics for the null hypothesis  $H_0$  from (1.1). Their asymptotic null distributions when  $m, n \rightarrow \infty$  are derived in Section 4. Our arguments make use of a central limit theorem for martingale arrays and U-statistic theory. We emphasize that all our statistics admit a normal limit after appropriate rescaling, but just as in the classical theory for U-statistics, the scaling factors have a different order when degenerate U-statistics are considered. In Section 5, we explore aspects of power of our tests from a minimax point of view. Simulation experiments are presented in Section 6, which also discusses computational considerations in the implementation of the tests. Throughout, for our null distributional theory, we make no distributional or moment assumption on  $(X^{(1)}, \dots, X^{(m)})$  other than that it is a continuous random vector. This assumption is needed to avoid ties in observations and ranks. We conclude with a brief discussion in Section 7.

1.1. *Notation.* For  $p \in \{1, \dots, m\}$ , we let  $\mathbf{R}^{(p)} := (R_1^{(p)}, \dots, R_n^{(p)})$  be the vector of ranks of  $\mathbf{X}^{(p)} = (X_1^{(p)}, \dots, X_n^{(p)})$ . The symmetric group of order  $l$  is denoted by  $\mathfrak{S}_l$ . Depending on the context, its elements are treated either as permutation functions or as ordered tuples from the set  $\{1, \dots, l\}$ . For  $k \leq n$ ,  $\mathcal{P}(n, k)$  denotes the set of  $k$ -tuples  $\mathbf{i} = (i_1, \dots, i_k)$  with  $1 \leq i_1 < \dots < i_k \leq n$ , and we will also identify the tuple  $\mathbf{i}$  with its set of elements  $\{i_1, \dots, i_k\}$ . Hence, for any two elements  $\mathbf{i} \in \mathcal{P}(n, k_1)$  and  $\mathbf{j} \in \mathcal{P}(n, k_2)$  with  $1 \leq k_1, k_2 \leq n$ , the operations  $\mathbf{i} \cup \mathbf{j}$ ,  $\mathbf{i} \cap \mathbf{j}$ , and  $\mathbf{i} \setminus \mathbf{j}$  give the tuples with increasing components that, as sets, equal the union, intersection and difference of  $\mathbf{i}$  and  $\mathbf{j}$ , respectively. For  $\mathbf{i} \in \mathcal{P}(n, k)$ , we let  $\mathbf{X}_{\mathbf{i}}^{(p)} := (X_{i_1}^{(p)}, \dots, X_{i_k}^{(p)})$ , and define the rank vector

$$\mathbf{R}_{\mathbf{i}}^{(p)} := (R_{i_1}^{(p)}, \dots, R_{i_k}^{(p)}),$$

where  $R_{i_c}^{(p)}$  is the rank of  $X_{i_c}^{(p)}$  among  $X_{i_1}^{(p)}, \dots, X_{i_k}^{(p)}$ .

Let  $p \neq q$  index two distinct variables. Then  $\mathbf{X}_c^{(pq)}$  and  $\mathbf{R}_c^{(pq)}$  denotes the pairs  $(X_c^{(p)}, X_c^{(q)})$  and  $(R_c^{(p)}, R_c^{(q)})$ , respectively, for  $c = 1, \dots, n$ . Similarly, given  $\mathbf{i} = (i_1, \dots, i_k) \in \mathcal{P}(n, k)$ , we let  $\mathbf{X}_{\mathbf{i},c}^{(pq)} := (X_{i_c}^{(p)}, X_{i_c}^{(q)})$  and  $\mathbf{R}_{\mathbf{i},c}^{(pq)} := (R_{i,c}^{(p)}, R_{i,c}^{(q)})$  for  $c \in \{1, \dots, k\}$ . We then define the  $k$ -tuples that are observation and rank vectors of pairs

$$\mathbf{R}_{\mathbf{i}}^{(pq)} := (\mathbf{R}_{\mathbf{i},1}^{(pq)}, \dots, \mathbf{R}_{\mathbf{i},k}^{(pq)}) \quad \text{and} \quad \mathbf{X}_{\mathbf{i}}^{(pq)} := (\mathbf{X}_{\mathbf{i},1}^{(pq)}, \dots, \mathbf{X}_{\mathbf{i},k}^{(pq)}).$$

When taking expectations under the null hypothesis  $H_0$ , we write  $\mathbb{E}_0[\cdot]$ , whereas  $\mathbb{E}[\cdot]$  is the general expectation operator, possibly under alternative hypotheses. Similarly, we write  $P_0[\cdot]$ ,  $P[\cdot]$ ,  $\text{Var}_0[\cdot]$ ,  $\text{Var}[\cdot]$ ,  $\text{Cov}_0[\cdot]$  and  $\text{Cov}[\cdot]$  for the probability, variance and covariance operator under  $H_0$  and possibly alternatives, respectively. Finally,  $\|\cdot\|_\infty$  and  $\|\cdot\|_2$  are the max norm and Euclidean norm for vectors, respectively, and the Froebenius norm of a matrix is denoted by  $\|\cdot\|_F$ . For two sequences  $(a_n)$  and  $(b_n)$ , the symbol  $a_n \asymp b_n$  is used to indicate the existence of constants  $c, C > 0$  such that  $c|a_n| \leq |b_n| \leq C|a_n|$  for all indices  $n$ .

**2. Rank correlations as U-statistics.** This section lays out a rank-based U-statistic framework that encompasses all rank correlations we will use when constructing specific test statistics for  $H_0$  in Section 3. Let

$$h : (\mathbb{R}^2)^k \longrightarrow \mathbb{R}$$

be a symmetric function of  $k \geq 2$  arguments in  $\mathbb{R}^2$ , that is, for all choices of  $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}) \in \mathbb{R}^2$ ,  $i = 1, \dots, k$ , and any permutation  $\pi \in \mathfrak{S}_k$ , it holds that  $h(\mathbf{x}_1, \dots, \mathbf{x}_k) = h(\mathbf{x}_{\pi(1)}, \dots, \mathbf{x}_{\pi(k)})$ . For any pair of distinct variable indices  $p, q \in \{1, \dots, m\}$ , the function  $h$  yields a *U-statistic*

$$(2.1) \quad U_h^{(pq)} = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h(\mathbf{X}_{\mathbf{i},1}^{(pq)}, \dots, \mathbf{X}_{\mathbf{i},k}^{(pq)}) = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h(\mathbf{X}_{\mathbf{i}}^{(pq)}).$$

In this context,  $h$  is termed the *kernel* of the U-statistics and is said to be of *degree*  $k$ .

Subsequently, we always assume that the kernel  $h$  and the induced U-statistics from (2.1) are *rank-based*, that is, the kernel has the property that  $h(\mathbf{x}_1, \dots, \mathbf{x}_k) = h(\mathbf{r}_1, \dots, \mathbf{r}_k)$  for all arguments  $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^2$ . Here, for each argument  $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}) \in \mathbb{R}^2$ , we let  $\mathbf{r}_i = (r_i^{(1)}, r_i^{(2)})$  with  $r_i^{(j)}$  being the rank of  $x_i^{(j)}$  among  $x_1^{(j)}, \dots, x_k^{(j)}$  for  $j = 1, 2$ . If  $U_h^{(pq)}$  from (2.1) is rank-based, then

$$(2.2) \quad U_h^{(pq)} = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \dots, \mathbf{R}_{\mathbf{i},k}^{(pq)}) = \frac{1}{\binom{n}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h(\mathbf{R}_{\mathbf{i}}^{(pq)}).$$

We note that all  $k$ -tuples  $(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \dots, \mathbf{R}_{\mathbf{i},k}^{(pq)})$  are uniquely determined by  $(\mathbf{R}_1^{(pq)}, \dots, \mathbf{R}_n^{(pq)})$ .

The following lemma lists elementary properties of  $U_h^{(pq)}$  under  $H_0$ . It relies on the fact that under  $H_0$  the distribution of  $h(\mathbf{R}_i^{(pq)})$  does not depend on the choice of  $i$ ,  $p$  and  $q$  because the rank vectors  $\mathbf{R}^{(1)}, \dots, \mathbf{R}^{(m)}$  are i.i.d. according to a uniform distribution on the symmetric group  $\mathfrak{S}_n$ ; recall that we assume the original observations to be continuous random vectors such that ties among the ranks have probability zero. A proof of the lemma is given in the Supplementary Material [Leung and Drton (2018)].

LEMMA 2.1. *Suppose  $g(\cdot)$  is a real-valued function defined on  $(\mathbb{R}^2)^n$ , and for  $1 \leq p \neq q \leq m$ ,*

$$g^{(pq)} := g(\mathbf{R}_1^{(pq)}, \dots, \mathbf{R}_n^{(pq)})$$

*is symmetric in the  $n$  arguments  $\mathbf{R}_1^{(pq)}, \dots, \mathbf{R}_n^{(pq)}$ . The random variables  $g^{(pq)}$  satisfy the following properties under  $H_0$ :*

- (i) *If  $p \neq q$ , then  $g^{(pq)}$  has the same distribution as  $g^{(12)}$ .*
- (ii) *If  $p \neq q$ , then  $g^{(pq)}$  is independent of  $\mathbf{X}^{(p)}$  (and also independent of  $\mathbf{X}^{(q)}$ ).*
- (iii) *For any fixed  $1 \leq l \leq m$ , the  $m - 1$  random variables  $g^{(pl)}$ ,  $p \neq l$ , are mutually independent.*
- (iv) *If  $p \neq q, r \neq s$  and  $\{p, q\} \neq \{r, s\}$ , then  $g^{(pq)}$  and  $g^{(rs)}$  are independent.*

In this paper, we assume all kernel functions  $h$  to be *bounded*. Since  $h$  can be recentered if needed, without loss of generality, we will further assume that  $\mathbb{E}_0[h(\mathbf{R}_i^{(pq)})] = 0$ , a property exhibited by all the examples below.

EXAMPLE 2.1 (Kendall’s tau). If we take  $h$  in (2.2) to be the kernel of degree  $k = 2$  given by

$$h_\tau(\mathbf{r}_1, \mathbf{r}_2) = \text{sgn}((r_1^{(1)} - r_2^{(1)})(r_1^{(2)} - r_2^{(2)})),$$

then  $\tau^{(pq)} := U_{h_\tau}^{(pq)}$  is Kendall’s tau, which measures the association of  $\mathbf{X}^{(p)}$  and  $\mathbf{X}^{(q)}$  by counting concordant versus discordant pairs of points.

EXAMPLE 2.2 (Spearman’s rho). Let

$$(2.3) \quad \rho_s^{(pq)} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i^{(p)} - R_i^{(q)})^2$$

be the Spearman’s rank correlation coefficient (rho) between  $\mathbf{X}^{(p)}$  and  $\mathbf{X}^{(q)}$ . Define  $\hat{\rho}_s^{(pq)} := U_{h_{\hat{\rho}_s}}^{(pq)}$ , where  $h_{\hat{\rho}_s}$  is the kernel function of degree 3 given by

$$(2.4) \quad h_{\hat{\rho}_s}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \frac{1}{2} \sum_{\pi \in \mathfrak{S}_3} \text{sgn}(r_{\pi_1}^{(1)} - r_{\pi_2}^{(1)}) \text{sgn}(r_{\pi_1}^{(2)} - r_{\pi_3}^{(2)}).$$

Hoeffding (1948b), page 318, showed that

$$(2.5) \quad \rho_s^{(pq)} = \frac{n-2}{n+1} \hat{\rho}_s^{(pq)} + \frac{3}{n+1} \tau^{(pq)}.$$

Hence, the dominating term  $\hat{\rho}_s$  of Spearman’s rho is a U-statistic.

EXAMPLE 2.3 (Hoeffding’s  $D$  statistic). Let

$$h_D(\mathbf{r}_1, \dots, \mathbf{r}_5) = \frac{1}{5!} \sum_{\pi \in \mathfrak{S}_5} \frac{\phi(r_{\pi_1}^{(1)}, \dots, r_{\pi_5}^{(1)}) \phi(r_{\pi_1}^{(2)}, \dots, r_{\pi_5}^{(2)})}{4},$$

where

$$\phi(r_1, \dots, r_5) = (I(r_1 \geq r_2) - I(r_1 \geq r_3))(I(r_1 \geq r_4) - I(r_1 \geq r_5))$$

and  $I(\cdot)$  is the indicator function. Hoeffding (1948a) suggested the statistic  $D^{(pq)} := U_{h_D}^{(pq)}$  to measure association between the vectors  $\mathbf{X}^{(p)}$  and  $\mathbf{X}^{(q)}$ . When the joint distribution of  $(X^{(p)}, X^{(q)})$  has continuous joint and marginal densities, the expectation

$$\mathbb{E}[h_D(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \dots, \mathbf{R}_{\mathbf{i},5}^{(pq)})]$$

is zero if and only if  $X^{(p)}$  and  $X^{(q)}$  are independent [Hoeffding (1948a), Theorem 3.1].

EXAMPLE 2.4 (Bergsma and Dassios’  $t^*$ ). In a recent paper, Bergsma and Dassios (2014) introduced  $t^{*(pq)} := U_{h_{t^*}}^{(pq)}$ , a U-statistic of degree 4 with the kernel

$$h_{t^*}(\mathbf{r}_1, \dots, \mathbf{r}_4) = \frac{1}{4!} \sum_{\pi \in \mathfrak{S}_4} \phi(r_{\pi_1}^{(1)}, \dots, r_{\pi_4}^{(1)}) \phi(r_{\pi_1}^{(2)}, \dots, r_{\pi_4}^{(2)}),$$

where now

$$\begin{aligned} \phi(r_1, \dots, r_4) &= I(r_1, r_3 < r_2, r_4) + I(r_1, r_3 > r_2, r_4) \\ &\quad - I(r_1, r_2 < r_3, r_4) - I(r_1, r_2 > r_3, r_4). \end{aligned}$$

According to Theorem 1 in Bergsma and Dassios (2014),  $t^*$  is an improvement over Hoeffding’s  $D$  in the sense that the vanishing of the expectation  $\mathbb{E}[h_{t^*}(\mathbf{R}_{\mathbf{i},1}^{(pq)}, \dots, \mathbf{R}_{\mathbf{i},4}^{(pq)})]$  characterizes the independence of  $X^{(p)}$  and  $X^{(q)}$  under the weaker assumption that  $(X^{(p)}, X^{(q)})$  has a bivariate distribution that is discrete or (absolutely) continuous, or a mixture of both. In fact, Bergsma and Dassios (2014) conjectured that even this assumption is not necessary.

Returning to our general setup, the variance and also the large-sample behavior of the statistic  $U_h^{(pq)}$  is determined by the covariance quantities

$$(2.6) \quad \zeta_c^h := \text{Cov}[h(\mathbf{R}_i^{(pq)})h(\mathbf{R}_j^{(pq)})], \quad c = 0, \dots, k,$$

where  $\mathbf{i}, \mathbf{j} \in \mathcal{P}(n, k)$  are such that  $|\mathbf{i} \cap \mathbf{j}| = c$ . When  $H_0$  is true,

$$(2.7) \quad \zeta_c^h = \mathbb{E}_0[h(\mathbf{R}_i^{(pq)})h(\mathbf{R}_j^{(pq)})]$$

as we are assuming that  $\mathbb{E}_0[h(\mathbf{R}_i^{(pq)})] = 0$ . Furthermore, the value of  $\zeta_c^h$  does not depend on the choice of  $(\mathbf{i}, p, q)$  under  $H_0$ . In the sequel, it will be clear from the context whether  $\zeta_c^h$  is defined under  $H_0$  or an alternative hypothesis.

It is well known that  $0 = \zeta_0^h \leq \zeta_1^h, \dots, \leq \zeta_k^h$ , and the kernel  $h$  is said to have order of degeneracy  $d$  if  $\zeta_0^h = \zeta_1^h = \dots = \zeta_{d-1}^h = 0$  and  $\zeta_d^h > 0$  [Serfling (1980), Chapter 5]. If  $d \geq 2$ , the kernel and the U-statistic it defines are referred to as degenerate. For any  $c = 1, \dots, k$ , it holds under  $H_0$  that

$$(2.8) \quad \zeta_c^h = 0 \iff \mathbb{E}_0[h(\mathbf{R}_i^{(pq)}) | \mathbf{X}_{i'}^{(pq)}] = 0 \quad \text{almost surely,}$$

where  $i' \subset i$  may be any subset with  $|i'| = c$ . In particular, for the kernels  $h_D$  and  $h_{t^*}$ , the right-hand side of (2.8) holds with  $c \leq 1$ .

As in the classical theory of U-statistics,  $\zeta_d^h$  will play a role in our asymptotic results for the test statistics we construct from rank-based U-statistics, for which the kernels have order of degeneracy  $d = 1$  or  $d = 2$  under  $H_0$ . However, when  $d = 2$ , an additional quantity is needed to describe our asymptotic results. For a symmetric kernel  $h : (\mathbb{R}^2)^k \rightarrow \mathbb{R}$  with order of degeneracy  $d = 2$  under  $H_0$ , we define

$$(2.9) \quad \eta^h := \mathbb{E}_0[h(\mathbf{R}_{i^1}^{(pq)})h(\mathbf{R}_{i^2}^{(pq)})h(\mathbf{R}_{i^3}^{(pq)})h(\mathbf{R}_{i^4}^{(pq)})],$$

where  $\mathbf{i}^1, \dots, \mathbf{i}^4 \in \mathcal{P}(n, k)$  are any four tuples such that:

- (i)  $|\bigcup_{\omega=1}^4 \mathbf{i}^\omega| = 4k - 4$ ,
- (ii)  $|\mathbf{i}^1 \cap \mathbf{i}^2| = |\mathbf{i}^2 \cap \mathbf{i}^3| = |\mathbf{i}^3 \cap \mathbf{i}^4| = |\mathbf{i}^4 \cap \mathbf{i}^1| = 1$  and
- (iii) no index  $i \in \bigcup_{\omega=1}^4 \mathbf{i}^\omega$  is an element of more than two of the sets  $\mathbf{i}^1, \dots, \mathbf{i}^4$ .

For our purpose, we only need to define  $\eta^h$  under  $H_0$ , and it is also easy to see that the choice of  $p, q, \mathbf{i}^\omega, \omega = 1, \dots, 4$ , does not matter in its definition. Table 1 collects the order of degeneracy  $d$  under  $H_0$ , and the quantities  $\zeta_d^h$  and  $\eta^h$  for the kernels in Examples 2.1–2.4. The latter are found in Hoeffding (1948b, 1948a), and by our own calculations.

Finally, it is easy to check that all the kernels in Example 2.1–2.4 satisfy the following property that will be assumed for our null asymptotic results.



TABLE 1  
 Degree  $k$ , order of degeneracy  $d$ , covariance  $\zeta_d^h$  and fourth moment  $\eta^h$  for the kernel functions in Examples 2.1–2.4 when independence holds

Kernel	$h_\tau$	$h_{\hat{\rho}_s}$	$h_D$	$h_{t^*}$
$k$	2	3	5	4
$d$	1	1	2	2
$\zeta_d^h$	1/9	1/9	1/810,000	1/225
$\eta^h$	–	–	$(7/864,000)^2$	$(2/525)^2$

ASSUMPTION 2.2. Let  $h : (\mathbb{R}^2)^k \rightarrow \mathbb{R}$  be a symmetric kernel with order of degeneracy  $d \geq 1$  under  $H_0$ . Then given  $\mathbf{i} = (i_1, \dots, i_k) \in \mathcal{P}(n, k)$  and  $1 \leq p \neq q \leq m$ ,

$$\mathbb{E}_0[h(\mathbf{R}_{\mathbf{i}}^{(pq)}) \mid \mathbf{X}_{\mathbf{j}}^{(p)}, \mathbf{X}_{\mathbf{j}' }^{(q)}] = 0$$

for all  $\mathbf{j}, \mathbf{j}' \subset \mathbf{i}$  such that  $\min(|\mathbf{j}|, |\mathbf{j}'|) < d$ .

**3. Test statistics.** We now proceed to construct test statistics for the independence hypothesis  $H_0$  from (1.1). Building on the pairwise rank correlations from Section 2, we introduce general classes of statistics and derive their respective asymptotic null distributions when  $m, n \rightarrow \infty$ .

3.1. *Sum of squared sample rank correlations.* Let  $U_h^{(pq)}$  be a rank-based U-statistic as defined in (2.2), with mean zero when  $X^{(p)}$  and  $X^{(q)}$  are independent. Suppose further that large absolute values of  $U_h^{(pq)}$  indicate strong association (positive or negative) between  $\mathbf{X}^{(p)}$  and  $\mathbf{X}^{(q)}$ . It is then natural to reject  $H_0$  for large values of the centered quantity

$$(3.1) \quad S_h := \sum_{1 \leq p < q \leq m} (U_h^{(pq)})^2 - \binom{m}{2} \mu_h.$$

Here,  $\mu_h := \mathbb{E}_0[(U_h^{(pq)})^2]$ . Note that, as indicated by our notation, this expectation does not depend on the choice of  $p$  and  $q$  by Lemma 2.1(i). The following lemma specifies  $\mu_h$  and gives a result on other moments of  $U_h^{(pq)}$  that will be used later.

LEMMA 3.1. Let  $n \geq 2k \geq 2$ , and suppose that  $U_h^{(pq)}$  from (2.2) has a kernel  $h$  with order of degeneracy  $d$  under  $H_0$ . Then the following three facts hold under  $H_0$ :

(i)

$$\mu_h = \binom{n}{k}^{-1} \sum_{c=1}^k \binom{k}{c} \binom{n-k}{k-c} \zeta_c = \binom{k}{d}^2 \frac{d! \zeta_d}{n^d} + O(n^{-d-1}).$$

(ii) For any  $r > 2$ ,

$$\mathbb{E}_0[(U_h^{(pq)})^r] = O(n^{-\lfloor (rd+1)/2 \rfloor}),$$

where  $\lfloor \cdot \rfloor$  denotes the floor function.

(iii)

$$\mathbb{E}_0[(U_h^{(pq)})^4] = \begin{cases} \frac{3k^4(\zeta_1^h)^2}{n^2} + O(n^{-3}) & \text{if } d = 1, \\ \binom{k}{2}^4 \frac{12}{n^4} ((\zeta_2^h)^2 + 4\eta^h) + O(n^{-5}) & \text{if } d = 2. \end{cases}$$

For Lemma 3.1(i) and (ii), see Lemma 5.2.1A and 5.2.2B in Serfling (1980). The last claim about the leading term of the fourth moment is proven in the Supplementary Material [Leung and Drton (2018)]. Let  $\mu_\tau, \mu_{\hat{\rho}_s}, \mu_D$  and  $\mu_{t^*}$  be the values of  $\mu_h$  when  $h$  is equal to  $h_\tau, h_{\hat{\rho}_s}, h_D$  and  $h_{t^*}$ , respectively. Then

$$\begin{aligned} \mu_\tau &= \frac{2(2n+5)}{9n(n-1)}, & \mu_{\hat{\rho}_s} &= \frac{(n^2-3)}{n(n-1)(n-2)}, \\ \mu_D &= \frac{2(n^2+5n-32)}{9n(n-1)(n-3)(n-4)}, & \mu_{t^*} &= \frac{8}{75} \frac{3n^2+5n-18}{n(n-1)(n-2)(n-3)}. \end{aligned}$$

The first three quantities can be found in Hoeffding (1948b, 1948a). The stated value of  $\mu_{t^*}$  is based on our own calculations.

3.2. *Unbiased estimator of the sum of squared population correlations.* The kernel function  $h$  is central to the role of  $U_h^{(pq)}$  as a measure of association between the vectors of observations  $\mathbf{X}^{(p)}$  and  $\mathbf{X}^{(q)}$ . At the population level, the association (positive or negative) is captured by the expectation of  $U_h^{(pq)}$ , which is also equal to

$$(3.2) \quad \theta_h^{(pq)} := \mathbb{E}[h(\mathbf{R}_j^{(pq)})],$$

where  $\mathbf{j}$  may be any element in  $\mathcal{P}(n, k)$ . Hence,

$$(3.3) \quad \sum_{1 \leq p < q \leq m} (\theta_h^{(pq)})^2$$

is a population measure of overall dependency in the joint distribution of  $X^{(1)}, \dots, X^{(m)}$ . As an alternative approach to Section 3.1, we now construct an

unbiased estimator of (3.3), targeting more directly the problem of global (in-)dependence.

Recall that given  $\mathbf{i} \in \mathcal{P}(n, 2k)$  and  $\mathbf{j} \in \mathcal{P}(n, k)$  such that  $\mathbf{j} \subset \mathbf{i}$  as sets,  $\mathbf{i} \setminus \mathbf{j}$  is the  $k$ -tuple in  $\mathcal{P}(n, k)$  that is given by their set difference. The function

$$(3.4) \quad h^W(\mathbf{R}_i^{(pq)}) := \binom{2k}{k}^{-1} \sum_{\substack{\mathbf{j} \subset \mathbf{i} \\ \mathbf{j} \in \mathcal{P}(n, k)}} h(\mathbf{R}_j^{(pq)})h(\mathbf{R}_{i \setminus j}^{(pq)}),$$

defined on  $(\mathbb{R}^2)^{2k}$ , is symmetric in its  $2k$  arguments  $\mathbf{R}_{i,1}^{(pq)}, \dots, \mathbf{R}_{i,2k}^{(pq)}$ , due to the symmetry of  $h$  and the summation over all possible tuples  $\mathbf{j} \in \mathcal{P}(n, k)$  contained in  $\mathbf{i}$  on the right-hand side of (3.4). Moreover,  $h^W$  is an unbiased estimator of the square of the expectation in (3.2), since each summand on the right-hand side of (3.4) is a product of two independent unbiased estimators of  $\theta_h^{(pq)}$ . Therefore, defining the U-statistic

$$(3.5) \quad W_h^{(pq)} = W_h^{(pq)}(\mathbf{R}_1^{(pq)}, \dots, \mathbf{R}_n^{(pq)}) = \binom{n}{2k}^{-1} \sum_{\mathbf{i} \in \mathcal{P}(n, 2k)} h^W(\mathbf{R}_i^{(pq)}),$$

we have that the sum

$$(3.6) \quad T_h := \sum_{1 \leq p < q \leq m} W_h^{(pq)}$$

is an unbiased estimator of (3.3). The statistic  $T_h$  is a U-statistic itself and serves as a natural test statistic for  $H_0$ . Large values of  $T_h$  indicate departures from  $H_0$ . When  $h = h_\tau$ , that is, the case of Kendall’s tau,  $T_h$  equals the statistic displayed in (1.7) in the Introduction.

Clearly,  $W_h^{(pq)}$  is a rank-based U-statistic with the kernel  $h^W$  of degree  $2k$ . The following lemma summarizes the degeneracy properties of  $h^W$  under  $H_0$ .

LEMMA 3.2. *Suppose  $h : (\mathbb{R}^2)^k \rightarrow \mathbb{R}$  is a symmetric kernel function of degree  $k$  with order of degeneracy  $d \in \{1, 2\}$  under  $H_0$ . So,  $\zeta_d^h > 0$ . Then, under  $H_0$ , the induced symmetric kernel function  $h^W$  defined in (3.4) has order of degeneracy  $2d$  and*

$$\begin{aligned} \zeta_{2d}^{h^W} &= \mathbb{E}_0[h^W(\mathbf{R}_i^{(pq)})h^W(\mathbf{R}_j^{(pq)})] \\ &= \begin{cases} 4 \binom{2k-2}{k-1}^2 \binom{2k}{k}^{-2} (\zeta_d^h)^2 & \text{if } d = 1, \\ 12 \binom{2k-4}{k-2}^2 \binom{2k}{k}^{-2} \{(\zeta_d^h)^2 + 2\eta^h\} & \text{if } d = 2, \end{cases} \end{aligned}$$

where  $\mathbf{i}, \mathbf{j} \in \mathcal{P}(n, 2k)$  and  $|\mathbf{i} \cap \mathbf{j}| = 2d$ .

The proof of the lemma is deferred to the Supplementary Material [Leung and Drton (2018)].

3.3. *Sum of sample rank correlations.* For testing  $H_0$ , it is also interesting to consider the simple sum

$$(3.7) \quad Z_h := \sum_{1 \leq p < q \leq m} U_h^{(pq)},$$

which unbiasedly estimates the signal

$$(3.8) \quad \sum_{1 \leq p < q \leq m} \theta_h^{(pq)};$$

compare with (3.3). When the kernel  $h$  is  $h_{\hat{\rho}_s}$  or  $h_\tau$ , without the squaring as in (3.3), (3.8) may not be an effective measure for the overall dependency of  $X^{(1)}, \dots, X^{(m)}$  since any pairwise signal  $\theta_h^{(pq)}$  can be either negative or positive depending on the direction of association [Kruskal (1958)]. Hence, the rejection of  $H_0$  for large value of  $Z_h$  is only good for testing against the “one-sided” alternative

$$\sum_{1 \leq p < q \leq m} \theta_h^{(pq)} > 0, \quad \theta_h^{(pq)} \geq 0 \text{ for all } p < q.$$

However, when  $h = h_{t^*}$  or  $h = h_D$ , (3.8) is an effective measure of the overall dependency of  $X^{(1)}, \dots, X^{(m)}$ , since any pairwise signal  $\theta_h^{(pq)}$  is non-negative and equals zero if and only if  $X^{(p)}$  and  $X^{(q)}$  are independent under the weak assumptions in the work of Hoeffding (1948a) and Bergsma and Dassios (2014). In this case, large values of  $Z_h$  detect dependency among  $X^{(1)}, \dots, X^{(m)}$ , without any restrictions to the direction of the pairwise associations.

**4. Asymptotic null distributions.** We are now ready to state our results on the asymptotic distributions for the test statistics introduced in Section 3. As mentioned in Section 2, we focus on rank-based U-statistics with a kernel  $h$  satisfying Assumption 2.2 and order of degeneracy  $d \in \{1, 2\}$  under  $H_0$ .

**THEOREM 4.1.** *Suppose the null hypothesis  $H_0$  from (1.1) is true. Let  $h$  be a symmetric bounded kernel function of degree  $k$  satisfying Assumption 2.2, and consider the asymptotic regime  $m, n \rightarrow \infty$ . If  $d = 1$ , after suitable rescaling,  $S_h, T_h$  and  $Z_h$  are asymptotically normal, namely,*

$$\frac{nS_h}{k^2m\zeta_1^h}, \quad \frac{nT_h}{k^2m\zeta_1^h}, \quad \frac{\sqrt{2n}Z_h}{km\sqrt{\zeta_1^h}} \implies \mathcal{N}(0, 1).$$

If  $d = 2$ , then

$$\frac{n^2 \binom{k}{2}^{-2} S_h}{2m\sqrt{(\zeta_2^h)^2 + 6\eta^h}}, \quad \frac{n^2 \binom{k}{2}^{-2} T_h}{2m\sqrt{(\zeta_2^h)^2 + 2\eta^h}}, \quad \frac{n \binom{k}{2}^{-1} Z_h}{m\sqrt{\zeta_2^h}} \implies \mathcal{N}(0, 1).$$

The theorem covers in particular the rank correlations from Examples 2.1–2.4. A critical value for an approximate  $\alpha$ -size test can thus be calibrated based on normal quantiles. As in the classical theory for U-statistics, the rescaling factors for the nondegenerate and degenerate cases differ in order; for instance, we have to multiply  $S_h$  with a factor of order  $O(n/m)$  when  $h$  has order of degeneracy  $d = 1$ , and with a factor of order  $O(n^2/m)$  when  $h$  has order of degeneracy  $d = 2$ . The ingredients needed to compute the rescaling factors were given in Table 1. In slight abbreviation, we write  $S_\tau, S_{\hat{\rho}_s}, S_D$  and  $S_{t^*}$  for the four versions of the statistic  $S_h$  from (3.1) with the different kernels reviewed in Section 2, and analogously,  $T_\tau, T_{\hat{\rho}_s}, T_D, T_{t^*}$  and  $Z_\tau, Z_{\hat{\rho}_s}, Z_D, Z_{t^*}$  for the versions of  $T_h$  and  $Z_h$  from (3.6) and (3.7). This matches the notation used in (1.5), (1.7) and (1.8).

We remark that while the classical Spearman’s rho is not a U-statistic one may of course consider the centered test statistic

$$(4.1) \quad S_{\rho_s} := \sum_{1 \leq p < q \leq m} (\rho_s^{(pq)})^2 - \binom{m}{2} \mu_{\rho_s},$$

where  $\mu_{\rho_s} := \mathbb{E}_0[(\rho_s^{(pq)})^2] = 1/(n - 1)$ ; see Hoeffding (1948b), page 321. The convergence of  $\frac{n}{m} S_{\hat{\rho}_s}$  to a standard normal distribution, as suggested by Theorem 4.1 and Table 1, implies the following distributional convergence for  $S_{\rho_s}$ . Its proof, given in the Supplementary Material [Leung and Drton (2018)], makes use of the decomposition from (2.5). The same result has been obtained by Zhou (2007) and Wang, Zou and Wang (2013) via different methods.

COROLLARY 4.2. *Under  $H_0, \frac{n}{m} S_{\rho_s} \implies N(0, 1)$  as  $m, n \rightarrow \infty$ .*

Our proof of Theorem 4.1 is based on a central limit theorem for martingale arrays Hall and Heyde (1980), Corollary 3.1, that was also applied by Schott (2005). We outline the approach here, postponing computations verifying the conditions of the martingale CLT to the Supplementary Material [Leung and Drton (2018)].

PROOF OF THEOREM 4.1. Fix a sample size  $n$ . For  $q = 1, \dots, m$ , let  $\mathcal{F}_{nq}$  be the  $\sigma$ -algebra generated by  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(q)}$  (or for our purposes, equivalently,  $\mathbf{R}^{(1)}, \dots, \mathbf{R}^{(q)}$ ) under  $H_0$ . For convenience, we will use the shorthand  $\bar{U}_h^{(pq)} := (U_h^{(pq)})^2 - \mu_h$  for  $1 \leq p < q \leq m$ . Let

$$(4.2) \quad D_{nq}^S := \sum_{p=1}^{q-1} \bar{U}_h^{(pq)}, \quad D_{nq}^T := \sum_{p=1}^{q-1} W_h^{(pq)}, \quad \text{and} \quad D_{nq}^Z := \sum_{p=1}^{q-1} U_h^{(pq)}$$

and set  $D_{n1}^S = D_{n1}^T = D_{n1}^Z = 0$ . Writing  $S_{nq} = \sum_{l=1}^q D_{nl}^S, T_{nq} = \sum_{l=1}^q D_{nl}^T$  and  $Z_{nq} = \sum_{l=1}^q D_{nl}^Z$ , we have that  $S_h = S_{nm}, T_h = T_{nm}$  and  $Z_h = Z_{nm}$ .

We claim that, for each  $n$ , the sequences

$$(4.3) \quad \{S_{nq}, \mathcal{F}_{nq}, 1 \leq q \leq m\}, \quad \{T_{nq}, \mathcal{F}_{nq}, 1 \leq q \leq m\} \quad \text{and} \\ \{Z_{nq}, \mathcal{F}_{nq}, 1 \leq q \leq m\}$$

are martingales, that is,  $\mathbb{E}_0[S_{nq} | \mathcal{F}_{n,q-1}] = S_{n,q-1}$ ,  $\mathbb{E}_0[T_{nq} | \mathcal{F}_{n,q-1}] = T_{n,q-1}$  and  $\mathbb{E}_0[Z_{nq} | \mathcal{F}_{n,q-1}] = Z_{n,q-1}$  for  $q = 2, \dots, m$ . Given the way  $S_{nq}$ ,  $T_{nq}$  and  $Z_{nq}$  are defined as sums, it suffices to show that

$$(4.4) \quad \mathbb{E}_0[\bar{U}_h^{(pq)} | \mathcal{F}_{n,q-1}] = \mathbb{E}_0[W_h^{(pq)} | \mathcal{F}_{n,q-1}] = \mathbb{E}_0[U_h^{(pq)} | \mathcal{F}_{n,q-1}] = 0$$

for all  $1 \leq p < q \leq m$ . Since  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}$  are independent under  $H_0$ , conditioning on  $\mathcal{F}_{n,q-1}$  is the same as conditioning on  $\mathbf{X}^{(p)}$  alone in (4.4). As  $\bar{U}_h^{(pq)}$ ,  $W_h^{(pq)}$  and  $U_h^{(pq)}$  are all symmetric functions of the  $n$  arguments  $\mathbf{R}_1^{(pq)}, \dots, \mathbf{R}_n^{(pq)}$ , (4.4) follows from Lemma 2.1(i) and (ii).

By the boundedness of our kernel  $h$ , each of the martingales in (4.3) is trivially square-integrable. As such, the central limit theorem for martingale arrays from Corollary 3.1 in Hall and Heyde (1980) implies the assertion of Theorem 4.1 if we can show that the squares of the martingale differences  $D_{nl}^S$ ,  $D_{nl}^T$  and  $D_{nl}^Z$  each satisfy the following two conditions. The first condition requires that as  $m, n \rightarrow \infty$ ,

$$(4.5) \quad \frac{n^2}{m^2} \sum_{l=2}^m \mathbb{E}_0[(D_{nl}^S)^2 | \mathcal{F}_{n,l-1}], \quad \frac{n^2}{m^2} \sum_{l=2}^m \mathbb{E}_0[(D_{nl}^T)^2 | \mathcal{F}_{n,l-1}] \xrightarrow{p} k^4 (\zeta_1^h)^2,$$

$$(4.6) \quad \frac{n}{m^2} \sum_{l=2}^m \mathbb{E}_0[(D_{nl}^Z)^2 | \mathcal{F}_{n,l-1}] \xrightarrow{p} 2^{-1} k^2 \zeta_1^h,$$

for  $d = 1$ , and

$$(4.7) \quad \frac{n^4}{m^2} \sum_{l=2}^m \mathbb{E}_0[(D_{nl}^S)^2 | \mathcal{F}_{n,l-1}] \xrightarrow{p} 4 \binom{k}{2}^4 \{(\zeta_2^h)^2 + 6\eta^h\},$$

$$(4.8) \quad \frac{n^4}{m^2} \sum_{l=2}^m \mathbb{E}_0[(D_{nl}^T)^2 | \mathcal{F}_{n,l-1}] \xrightarrow{p} 4 \binom{k}{2}^4 \{(\zeta_2^h)^2 + 2\eta^h\},$$

$$(4.9) \quad \frac{n^2}{m^2} \sum_{l=2}^m \mathbb{E}_0[(D_{nl}^Z)^2 | \mathcal{F}_{n,l-1}] \xrightarrow{p} \binom{k}{2}^2 \zeta_2^h,$$

for  $d = 2$ , where the convergence symbol stands for convergence in probability. The second condition is a Lindeberg condition. In the Supplementary Material [Leung and Drton (2018)], we show that, in fact, (4.5)–(4.9) also hold in the stronger sense of  $L^2$  (or quadratic mean). It also contains the proof of a Lyapunov condition that implies the Lindeberg condition, which completes the proof of Theorem 4.1.  $\square$

**5. Aspects of power.** In order to investigate the power of our tests we adopt an asymptotic minimax perspective. While our null distributional results in Section 3 are valid under the more general asymptotic regime  $m, n \rightarrow \infty$ , we treat here the particular regime  $\frac{m}{n} \rightarrow \gamma \in (0, \infty)$ . Recall the definition in (3.2), and let  $\Theta = (\theta_h^{(pq)})_{1 \leq p < q \leq m}$  be the  $\binom{m}{2}$ -vector comprising all these pairwise measures of association. In our exploration of power, it is at times convenient to have U-statistics with a kernel  $h$  of degree 2. For instance, we apply results for U-statistics of degree 2 from Chen (2016). Consequently, our power analysis focuses on the two classes of statistics  $S_h$  and  $T_h$  for the kernel  $h = h_\tau$  of Kendall’s tau. To indicate this restriction, we write  $\theta_\tau^{(pq)} := \mathbb{E}[h_\tau(\mathbf{R}_i^{(pq)})]$  for  $\mathbf{i} \in \mathcal{P}(n, 2)$  and  $\Theta_\tau = (\theta_\tau^{(pq)})_{1 \leq p < q \leq m}$ .

Let  $\mathcal{D}_m$  be a family of continuous joint distributions on  $\mathbb{R}^m$  containing all  $m$ -variate Gaussian distributions, to be considered as joint distributions for  $(X^{(1)}, \dots, X^{(m)})$ . For a given significance level  $\alpha \in (0, 1)$ , we study which sequences of lower bounds  $\varepsilon_n$  on the dependency signal  $\|\Theta_\tau\|_2$  allow tests to uniformly achieve a fixed power  $\beta > \alpha$  over the set of alternative distributions

$$(5.1) \quad \mathcal{D}_m(\|\Theta_\tau\|_2 \geq \varepsilon_n) := \{D \in \mathcal{D}_m : \|\Theta_\tau\|_2 \geq \varepsilon_n\}.$$

As usual, we take a test  $\phi$  to be a function mapping the data into the unit interval  $[0, 1]$ . Given a test statistic  $S = S(\mathbf{X}_1, \dots, \mathbf{X}_n)$ , we write  $\phi_\alpha(S)$  for the test that rejects for large values of  $S$  and has (asymptotic) size  $\alpha$ .

The statistics  $S_\tau$  and  $T_\tau$  estimate the squared Euclidean norm of the signal  $\|\Theta_\tau\|_2^2$ . They are thus natural when the interest is in detecting the alternatives in (5.1). The following theorem gives a rough lower bound on the signal size  $\|\Theta_\tau\|_2$  that is needed for detectability.

**THEOREM 5.1.** *Let  $0 < \alpha < \beta < 1$ . Under the asymptotic regime  $m/n \rightarrow \gamma \in (0, \infty)$ , there exist constants  $C_i = C_i(\alpha, \beta, \gamma) > 0$  for  $i = 1, 2$ , such that:*

(i)

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 \geq \varepsilon_n)} \mathbb{E}[\phi_\alpha(S_\tau)] > \beta \quad \text{for } \varepsilon_n = C_1\sqrt{n}, \quad \text{and}$$

(ii)

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 \geq \varepsilon_n)} \mathbb{E}[\phi_\alpha(T_\tau)] > \beta \quad \text{for } \varepsilon_n = C_2\sqrt{n}.$$

Our proof of Theorem 5.1 uses rather general concentration bounds and it should be possible to sharpen the analysis to show asymptotic power for  $\phi_\alpha(S_\tau)$  and  $\phi_\alpha(T_\tau)$  under smaller signal strength. Indeed, we conjecture that a test based on  $T_\tau$  can asymptotically attain uniform power  $\beta$  when the signal size  $\|\Theta_\tau\|_2$  is of constant order  $O(1)$ . This conjecture is partially supported by Theorem 5.2 below.

5.1. *Rate-optimality under equicorrelation.* When the joint distribution of  $X^{(1)}, \dots, X^{(m)}$  is a regular Gaussian distribution, then  $H_0$  is equivalent to  $R - I_m = 0$ , where  $I_m$  is the  $m$ -by- $m$  identity matrix; recall that  $R$  is the population Pearson correlation matrix. For any  $\varepsilon > 0$ , define the alternative

$$(5.2) \quad \mathcal{N}_m(\|R - I_m\|_F \geq \varepsilon)$$

as the family of regular  $m$ -variate Gaussian distributions whose correlation matrix  $R$  satisfies  $\|R - I_m\|_F \geq \varepsilon$ . Fix any  $\alpha, \beta \in (0, 1)$  with  $\alpha < \beta$ . A result of Remark 1(a) from Cai and Ma (2013) implies that in the regime  $m/n \rightarrow \gamma$ , there exists a sufficiently small constant  $c = c(\alpha, \beta, \gamma) > 0$  such that

$$\limsup_{n \rightarrow \infty} \inf_{\mathcal{N}_m(\|R - I_m\|_F \geq c)} \mathbb{E}[\phi] < \beta$$

for any  $\alpha$ -level test  $\phi$ . In other words, asymptotically, no  $\alpha$ -level test can uniformly achieve the desired power against the alternative (5.2) when the signal size  $\|R - I_m\|_F$  is allowed to be as small as  $c$ . It follows immediately that in our nonparametric setup there also exists a constant  $\tilde{c} = \tilde{c}(\alpha, \beta, \gamma) > 0$  such that

$$\limsup_{n \rightarrow \infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 > \tilde{c})} \mathbb{E}[\phi] < \beta$$

for any  $\alpha$ -level test  $\phi$ . This is true because the nonparametric class  $\mathcal{D}_m$  contains all  $m$ -variate Gaussian distributions, and because  $\theta_\tau^{(pq)} \asymp \rho^{(pq)}$  when  $X^{(p)}$  and  $X^{(q)}$  are jointly Gaussian. The latter fact follows from  $\rho^{(pq)} = \sin(\frac{\pi}{2}\theta_\tau^{(pq)})$  for nondegenerate elliptical distributions; see Lindskog, McNeil and Schmock (2003).

Given the observation just made, an  $\alpha$ -level test  $\phi$  that satisfies

$$(5.3) \quad \liminf_{n \rightarrow \infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_2 \geq \tilde{C})} \mathbb{E}[\phi] > \beta$$

for a large enough constant  $\tilde{C} = \tilde{C}(\alpha, \beta, \gamma) > 0$  would be rate-optimal. If the signal  $\|\Theta_\tau\|_2$  is large, being an unbiased estimator of  $\|\Theta_\tau\|_2^2$  our statistic  $T_\tau$  always centers around the same large value regardless of the true underlying distribution of  $\mathbf{X}$ . It is hence natural to conjecture that the optimality condition (5.3) is satisfied by the test  $\phi_\alpha(T_\tau)$ , for a reasonable class of elliptical distributions  $\mathcal{D}_m$  that extends beyond the Gaussians. Our next result supports the conjecture.

Let  $\mathcal{N}_m^{\text{equi}}(\|\Theta_\tau\|_2 \geq \tilde{C})$  be the set of  $m$ -variate Gaussian distributions that have all pairwise (Pearson and thus also Kendall) correlations equal to a common value such that  $\|\Theta_\tau\|_2 \geq \tilde{C}$ . If  $\theta_\tau^{(pq)} = \theta$  for all  $1 \leq p \neq q \leq m$ , then  $\|\Theta_\tau\|_2^2 = \theta^2 \binom{m}{2}$ .

**THEOREM 5.2.** *As  $\frac{m}{n} \rightarrow \gamma$ , there exists a constant  $\tilde{C} = \tilde{C}(\alpha, \beta, \gamma) > 0$  such that*

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{N}_m^{\text{equi}}(\|\Theta_\tau\|_2 \geq \tilde{C})} \mathbb{E}[\phi_\alpha(T_\tau)] > \beta.$$

The theorem is proved in the Supplementary Material [Leung and Drton (2018)]. Our simulation experiments on power in Section 6 corroborate the conjecture made above.



5.2. *Comparison with the “max” statistic.* The work of Han and Liu (2014) considered testing the independence hypothesis  $H_0$  from (1.1) using maxima of rank correlations and, in particular, the statistic

$$(5.4) \quad S_\tau^{\max} := \max_{1 \leq p < q \leq m} |\tau^{(pq)}|$$

that is based on Kendall’s tau. Han and Liu (2014) derived the asymptotic null distribution under the regime  $\log m = o(n^{1/3})$ . Let  $\phi_\alpha(S_\tau^{\max})$  be the level  $\alpha$  test that rejects for large values of  $S_\tau^{\max}$ . Naturally, this test is powerful against alternatives belonging to the set

$$(5.5) \quad \mathcal{D}_m(\|\Theta_\tau\|_\infty \geq \varepsilon_n) := \{D \in \mathcal{D}_m : \|\Theta_\tau\|_\infty \geq \varepsilon_n\},$$

which is characterized by the max norm of  $\Theta_\tau$ . Indeed, when  $\log m = o(n^{1/3})$ , for a given significance level  $\alpha$  and targeted power  $\beta \in (\alpha, 1)$ , it was shown that there exists a constant  $c_1 = c_1(\alpha, \beta)$  such that

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_\infty \geq c_1 \sqrt{(\log m)/n})} \mathbb{E}[\phi_\alpha(S_\tau^{\max})] > \beta.$$

Han and Liu (2014) also showed rate-optimality of this test, that is, there exists a constant  $c_2 = c_2(\alpha, \beta) < c_1$  such that for any  $\alpha$ -level test  $\phi$ ,

$$(5.6) \quad \limsup_{n \rightarrow \infty} \inf_{\mathcal{D}_m(\|\Theta_\tau\|_\infty \geq c_2 \sqrt{(\log m)/n})} \mathbb{E}[\phi] < \beta.$$

Note that in the regime  $m/n \rightarrow \gamma$  that we consider in this section we have  $\log m = o(n^{1/3})$ .

While a test based on  $S_\tau^{\max}$  is rate-optimal in detecting alternatives of the form (5.5) characterized by the max norm signal, it is—as intuition suggests—not powerful in detecting alternatives with small but nonzero dependence among many pairs of random variables. The latter scenario is best described via the Euclidean norm as in (5.1). This is demonstrated by the following theorem about equicorrelation alternatives; recall the positive result in Theorem 5.2.

**THEOREM 5.3.** *As  $\frac{m}{n} \rightarrow \gamma$ , there does not exist any constant  $C = C(\alpha, \beta, \gamma) > 0$  such that*

$$\liminf_{n \rightarrow \infty} \inf_{\mathcal{N}_m^{\text{equi}}(\|\Theta_\tau\|_2 \geq C)} \mathbb{E}[\phi_\alpha(S_\tau^{\max})] > \beta.$$

The proof of the theorem is deferred to the Supplementary Material [Leung and Drton (2018)]. It relies on a comparison lemma of Chernozhukov, Chetverikov and Kato (2013) and a recent result on Gaussian approximation for high-dimensional U-statistics in Chen (2016). Theorem 5.3 says that a signal size  $\|\Theta_\tau\|_2$  of constant order is not enough to guarantee a preset asymptotic power for a test based on  $S_\tau^{\max}$  under the regime  $\frac{m}{n} \rightarrow \gamma$ . We demonstrate this in our simulations in the next section.

**6. Implementation and simulation experiments.** We now compare several tests of the independence hypothesis  $H_0$  based on specific versions of the statistics introduced in this paper. Our simulations first explore the size of the tests when critical values are set using asymptotic normal approximations. We then compare their power. Before turning to the simulations, however, we discuss the computation of the test statistics.

6.1. *Implementation.* In order to compute the statistics  $S_h$  from (3.1) and  $Z_h$  from (3.7) for  $m$  variables, one has to make  $\binom{m}{2}$  evaluations of the U-statistics  $U_h^{(pq)}$ . In general, for a U-statistic of degree  $k$ , a naïve calculation following the definition in (2.2) requires  $O(n^k)$  operations. Fortunately, more efficient algorithms are available for the specific examples covered here. For instance, Spearman’s  $\rho_s^{(pq)}$  from Example 2.2 can be computed in  $O(n \log n)$  operations. The same is true for Kendall’s  $\tau^{(pq)}$  from Example 2.1 [Christensen (2005)]. Similarly, Weihs, Drton and Leung (2016) showed how to compute the Bergsma–Dassios sign covariance  $t^{*(pq)}$  in  $O(n^2 \log n)$  operations despite the fact that its kernel has degree  $k = 4$ , as reviewed in Example 2.4. An improvement to  $O(n^2)$  was given by Heller and Heller (2016). Finally, Hoeffding (1948a) gives formulas for efficient computation of his statistic  $D$  in Section 5 of his paper.

The situation with the class of statistics  $T_h$  from (3.6) is more complicated. Since a kernel  $h$  of degree  $k$  gives rise to an induced kernel  $h^W$  of degree  $2k$ , the number of operations equals  $O(n^{2k})$  if we compute  $W_h^{(pq)}$  by naïvely following its definition. This would lead to a total of  $\binom{m}{2} O(n^{2k})$  operations to find all  $W_h^{(pq)}$ ,  $1 \leq p < q \leq m$ . A more efficient way to compute each  $W_h^{(pq)}$  in  $O(n^k)$  time proceeds as follows. Using (3.4) and (3.5), we see that

$$(6.1) \quad W_h^{(pq)} = \frac{1}{\binom{n}{k} \binom{n-k}{k}} \sum_{\mathbf{i} \in \mathcal{P}(n,k)} h_{\mathbf{i}} \bar{h}_{\mathbf{i}},$$

where for each  $\mathbf{i} \in \mathcal{P}(n, k)$ , and suppressing the dependence on the pair  $(p, q)$ , we define

$$h_{\mathbf{i}} := h(\mathbf{R}_{\mathbf{i}}^{(pq)}) \quad \text{and} \quad \bar{h}_{\mathbf{i}} := \sum_{\mathbf{j} \in \mathcal{P}(n,k); \mathbf{j} \cap \mathbf{i} = \emptyset} h_{\mathbf{j}}.$$

Hence, it suffices to calculate:

- (i)  $h_{\mathbf{i}}$  for all  $\mathbf{i} \in \mathcal{P}(n, k)$ ,
- (ii)  $\bar{h}_{\mathbf{i}}$  for all  $\mathbf{i} \in \mathcal{P}(n, k)$  and
- (iii) the summation in (6.1),

in that order. Evidently, step (i) involves  $O(n^k)$  operations. By the inclusion-exclusion principle,

$$(6.2) \quad \bar{h}_{\mathbf{i}} = \sum_{\mathbf{j} \in \mathcal{P}(n,k)} h_{\mathbf{j}} + \sum_{1 \leq \ell \leq k} (-1)^\ell \sum_{\substack{\mathbf{j}' \in \mathcal{P}(n,\ell): \\ \mathbf{j}' \subset \mathbf{i}}} h_{\mathbf{j}'},$$

where  $h_{\mathbf{j}'} := \sum_{\mathbf{j} \subset \mathcal{P}(n,k); \mathbf{j}' \subset \mathbf{j}} h_{\mathbf{j}}$  for each  $1 \leq \ell < k$  and  $\mathbf{j}' \subset \mathcal{P}(n, \ell)$ . Note that there are  $O(n^\ell)$  many  $\mathbf{j}' \in \mathcal{P}(n, \ell)$ , and each  $h_{\mathbf{j}'}$  is a sum of  $O(n^{k-\ell})$  many terms. Finding  $h_{\mathbf{j}'}$  for all  $\mathbf{j}' \in \mathcal{P}(n, \ell)$  and  $1 \leq \ell < k$  thus requires  $O(n^k)$  operations, and with these as ingredients, by (6.2), one can compute each  $\bar{h}_i$  in  $O(1)$  operations if  $\sum_{\mathbf{j} \in \mathcal{P}(n,k)} h_{\mathbf{j}}$  is already known. But the quantity  $\sum_{\mathbf{j} \in \mathcal{P}(n,k)} h_{\mathbf{j}}$  only has to be computed once, with another  $O(n^k)$  computations. Consequently, step (ii) involves  $O(n^k)$  operations, and so does the final summation in step (iii).

6.2. *Simulations.* We first consider the sizes of tests based on our statistics  $S_\tau$ ,  $S_{\rho_s}$ ,  $S_{t^*}$ ,  $T_\tau$ ,  $T_{\hat{\rho}_s}$  and  $Z_{t^*}$  that we introduced in Section 4. For comparison, we also consider the sum of squared Pearson correlations  $S_r$  from Schott (2005); recall (1.3). Each test compares a rescaled test statistic to the limiting standard normal distribution from Theorem 4.1 and Corollary 4.2. Targeting a size of 0.05, the null hypothesis  $H_0$  is rejected if the value of the rescaled statistic exceeds the 95th percentile of the standard normal distribution. Table 2 gives Monte-Carlo estimates of finite-sample sizes for different combinations of  $n$  and  $m$ . The data underlying the table are i.i.d. noncentral  $t$  with  $\nu = 3$  degrees of freedom and noncentrality parameter  $\mu = 2$ . For each combination of  $m$  and  $n$ , the sizes of the tests are calculated from 5000 independently generated data sets. As expected, the tests that use rank-based statistics all have their sizes get closer to the nominal 0.05 when  $m$  and  $n$  increase, but the test based on  $S_r$  is not valid as it rejects too often. Recall that Schott’s limit theorem is derived under a Gaussian assumption. For certain new nonparametric tests introduced in this paper, the test sizes are not very satisfactory when  $n$  is small, but they all get close to the nominal 0.05 level once  $n$  becomes 128, indicating that the asymptotics described by Theorem 4.1 kicks in. Surprisingly, the test given by  $S_\rho$  has good size even for very small  $n$ . It would be of interest to explore more refined results, such as a Berry–Esseen bound or an Edgeworth expansion for the normal convergences of Theorem 4.1 in future research.

Next, we consider the power of the tests, as studied in Section 5. For different combinations of  $(m, n)$ , we generate data as  $n$  independent draws from three different  $m$ -variate elliptical distributions. These are:

- (i) the  $m$ -variate normal distribution:  $N_m(0, \Sigma)$ ,
- (ii) the  $m$ -variate  $t$  distribution:  $t_{\nu=20,m}(\mu = 2 \cdot \mathbf{1}_m, \Sigma)$ , and
- (iii) the  $m$ -variate power exponential distribution:  $PE(\mu = 0, \Sigma, \nu = 20)$ .

Here,  $\mathbf{1}_m$  is the  $m$ -vector with all entries equal to 1, and the parametrizations of these distributions are in accordance with Oja (2010), pages 8–10. For each distribution, the scatter matrix  $\Sigma = (\sigma_{ij})$  is taken to be a matrix with 1’s on the diagonal and equal values for the off-diagonal entries, which are set to obtain the signal strengths  $\|\Theta_\tau\|_2^2 = 0.1, 0.3, \text{ and } 0.7$  based on Kendall’s  $\tau$ . We refer again to Lindskog, McNeil and Schmock (2003) for the relationship between  $\Sigma$  and  $\|\Theta_\tau\|_2^2$ .

TABLE 2  
*Simulated size of tests when  $X^{(1)}, \dots, X^{(m)}$  are i.i.d.  $t_{3,2}$  data. For each combination of  $(m, n)$  and each test, the sizes are computed from 5000 independently generated datasets*

Statistics	$n$	$m$							
		4	8	16	32	64	128	256	512
$S_\tau$	16	0.060	0.065	0.062	0.067	0.071	0.063	0.071	0.072
$S_\tau$		0.069	0.079	0.080	0.090	0.094	0.093	0.086	0.089
$T_\tau$		0.088	0.096	0.102	0.113	0.120	0.110	0.113	0.114
$S_{\rho_S}$		0.046	0.050	0.052	0.057	0.059	0.053	0.053	0.055
$T_{\hat{\rho}_S}$		0.079	0.093	0.099	0.107	0.111	0.107	0.104	0.109
$S_{\tau^*}$		0.079	0.098	0.115	0.112	0.123	0.122	0.111	0.121
$Z_{\tau^*}$		0.079	0.092	0.098	0.098	0.111	0.104	0.096	0.099
$S_\tau$	32	0.066	0.078	0.076	0.081	0.076	0.089	0.079	0.086
$S_\tau$		0.059	0.069	0.067	0.077	0.073	0.071	0.070	0.077
$T_\tau$		0.064	0.078	0.075	0.087	0.081	0.082	0.080	0.086
$S_{\rho_S}$		0.047	0.054	0.052	0.061	0.056	0.053	0.056	0.058
$T_{\hat{\rho}_S}$		0.062	0.075	0.072	0.082	0.080	0.079	0.072	0.083
$S_{\tau^*}$		0.056	0.081	0.085	0.090	0.088	0.078	0.087	0.085
$Z_{\tau^*}$		0.062	0.069	0.067	0.081	0.077	0.077	0.079	0.078
$S_\tau$	64	0.073	0.083	0.095	0.095	0.102	0.097	0.096	0.091
$S_\tau$		0.057	0.061	0.062	0.065	0.058	0.058	0.065	0.059
$T_\tau$		0.058	0.064	0.066	0.069	0.061	0.064	0.067	0.062
$S_{\rho_S}$		0.048	0.053	0.055	0.055	0.050	0.052	0.057	0.048
$T_{\hat{\rho}_S}$		0.057	0.061	0.065	0.067	0.060	0.064	0.059	0.062
$S_{\tau^*}$		0.045	0.074	0.064	0.070	0.068	0.070	0.069	0.063
$Z_{\tau^*}$		0.054	0.061	0.058	0.064	0.065	0.062	0.063	0.064
$S_\tau$	128	0.072	0.089	0.107	0.112	0.101	0.109	0.110	0.115
$S_\tau$		0.047	0.061	0.053	0.061	0.052	0.056	0.053	0.055
$T_\tau$		0.049	0.063	0.053	0.064	0.054	0.060	0.054	0.058
$S_{\rho_S}$		0.043	0.059	0.049	0.056	0.048	0.052	0.048	0.051
$T_{\hat{\rho}_S}$		0.048	0.062	0.052	0.060	0.055	0.057	0.058	0.054
$S_{\tau^*}$		0.041	0.066	0.070	0.071	0.060	0.058	0.052	0.058
$Z_{\tau^*}$		0.050	0.055	0.058	0.062	0.053	0.056	0.055	0.055

The power, computed based on 500 repetitions of experiments, for tests based on  $S_\tau$ ,  $T_\tau$ , and the statistic  $S_\tau^{\max}$  of Han and Liu (2014) are compared in Table 3. As expected,  $S_\tau^{\max}$  is not well adapted for detecting the alternatives we generated. For each  $(m, n)$  combination and a given value of  $\|\Theta_\tau\|_2^2$ , the power of the test based on  $T_\tau$  is similar across different data-generating distributions. In contrast,  $S_\tau$  tends to yield more power for t-distributed data, and less power for data with power exponential distribution. The stability of the power rendered by  $T_\tau$  points to our conjecture in Section 5 on the minimax optimality of  $T_\tau$  over a wider class of distributions.

When the data are generated from multivariate normal distributions, Table 3 includes a comparison to three further tests. First, Schott's  $S_r$  from (1.3) yields a valid (asymptotic) test in this case. As seen in Table 3, the three statistics,  $S_\tau$ ,

TABLE 3

*Simulated power of tests when data are generated from the multivariate normal (MVN), multivariate  $t$  (MVT) and multivariate power exponential (MVPE) distributions with three different values for the dependency signal  $\|\Theta_\tau\|_2^2$ . All pairwise (population) Kendall's tau correlations  $\theta_\tau^{(pq)}$ ,  $1 \leq p < q \leq m$  are equal to the same value  $\theta$  so that  $\|\Theta_\tau\|_2^2 = \binom{m}{2}\theta^2$ . For each combination of  $(m, n)$  and each test, the power is calculated from 500 independently generated datasets*

Statistic	$n$	$\ \Theta_\tau\ _2^2 = 0.1$			$\ \Theta_\tau\ _2^2 = 0.3$			$\ \Theta_\tau\ _2^2 = 0.7$		
		$m$			$m$			$m$		
		64	128	256	64	128	256	64	128	256
MVN										
$S_\tau$	64	0.094	0.054	0.070	0.182	0.108	0.092	0.424	0.218	0.114
$T_\tau$		0.100	0.068	0.078	0.194	0.110	0.090	0.426	0.228	0.134
$S_\tau^{\max}$		0.046	0.046	0.020	0.040	0.058	0.046	0.056	0.054	0.058
$S_r$		0.070	0.058	0.070	0.178	0.114	0.080	0.448	0.222	0.110
Cai & Ma		0.076	0.076	0.060	0.190	0.116	0.086	0.456	0.278	0.130
$S_\tau$	128	0.130	0.086	0.056	0.342	0.164	0.080	0.794	0.444	0.176
$T_\tau$		0.132	0.088	0.058	0.352	0.174	0.084	0.806	0.446	0.186
$S_\tau^{\max}$		0.062	0.064	0.052	0.046	0.058	0.060	0.094	0.058	0.060
$S_r$		0.142	0.072	0.066	0.378	0.172	0.084	0.832	0.514	0.198
LRT		0.094	–	–	0.204	–	–	0.396	–	–
Cai & Ma		0.134	0.064	0.068	0.386	0.172	0.096	0.834	0.520	0.204
$S_\tau$	256	0.256	0.108	0.096	0.780	0.358	0.198	0.992	0.838	0.476
$T_\tau$		0.262	0.114	0.094	0.782	0.364	0.200	0.992	0.830	0.470
$S_\tau^{\max}$		0.048	0.050	0.046	0.064	0.056	0.058	0.124	0.082	0.052
$S_r$		0.282	0.126	0.094	0.816	0.420	0.224	1.000	0.880	0.502
LRT		0.166	0.086	–	0.450	0.152	–	0.876	0.370	–
Cai & Ma		0.282	0.124	0.110	0.812	0.422	0.234	1.000	0.882	0.494
MVT										
$S_\tau$	64	0.506	0.866	0.998	0.628	0.896	0.998	0.802	0.926	0.998
$T_\tau$		0.130	0.080	0.078	0.232	0.128	0.096	0.488	0.234	0.114
$S_\tau^{\max}$		0.080	0.066	0.060	0.086	0.074	0.060	0.110	0.074	0.068
$S_\tau$	128	0.554	0.912	0.998	0.806	0.948	1.000	0.962	0.990	1.000
$T_\tau$		0.130	0.102	0.094	0.384	0.210	0.114	0.796	0.494	0.244
$S_\tau^{\max}$		0.064	0.060	0.054	0.080	0.064	0.066	0.114	0.074	0.076
$S_\tau$	256	0.694	0.924	1.000	0.972	0.992	1.000	1.000	1.000	1.000
$T_\tau$		0.268	0.130	0.084	0.740	0.348	0.188	0.998	0.832	0.456
$S_\tau^{\max}$		0.076	0.062	0.072	0.110	0.066	0.076	0.186	0.102	0.078

TABLE 3  
(Continued)

Statistic	$n$	$\ \Theta_\tau\ _2^2 = 0.1$			$\ \Theta_\tau\ _2^2 = 0.3$			$\ \Theta_\tau\ _2^2 = 0.7$		
		$m$			$m$			$m$		
		64	128	256	64	128	256	64	128	256
					MVPE					
$S_\tau$	64	0.052	0.042	0.022	0.128	0.056	0.044	0.358	0.122	0.060
$T_\tau$		0.114	0.076	0.076	0.222	0.110	0.082	0.462	0.216	0.134
$S_\tau^{\max}$		0.056	0.050	0.032	0.046	0.050	0.034	0.062	0.054	0.036
$S_\tau$	128	0.074	0.038	0.028	0.274	0.094	0.036	0.744	0.314	0.112
$T_\tau$		0.128	0.084	0.056	0.398	0.174	0.096	0.836	0.454	0.214
$S_\tau^{\max}$		0.038	0.054	0.050	0.050	0.056	0.044	0.084	0.060	0.046
$S_\tau$	256	0.134	0.066	0.050	0.638	0.256	0.102	0.992	0.794	0.306
$T_\tau$		0.232	0.152	0.100	0.768	0.370	0.184	0.998	0.862	0.450
$S_\tau^{\max}$		0.052	0.036	0.060	0.074	0.040	0.060	0.120	0.064	0.062

$T_\tau$  and  $S_r$  give comparable power for different combinations of  $(m, n)$  and signal strength  $\|\Theta_\tau\|_2^2$ . Second, we tried the likelihood ratio test (LRT) with critical rejection region calibrated based on Corollary 1 in Jiang and Qi (2015) whenever it is implementable, that is, when  $m < n$  in the table. It is generally less powerful than our new tests and  $S_r$  in detecting the alternatives we consider. Lastly, we experimented with the statistic proposed in Cai and Ma (2013), which again demonstrates similar power. The test of Cai and Ma (2013) is minimax rate optimal in detecting the Frobenius norm signal  $\|\Sigma - I_m\|_2$ , but only for testing the different hypothesis  $\tilde{H}_0 : \Sigma = I_m$  and under a Gaussian assumption on  $\mathbf{X}$ . Under Gaussianity, our hypothesis of independence  $H_0$  from (1.1) is of course equivalent to the  $R = I_m$  instead. Despite this mismatch, the comparable power of the test of Cai and Ma (2013) indicates that the three statistics  $S_\tau$ ,  $T_\tau$  and  $S_r$  are all powerful in detecting the signal  $\|R - I_m\|_2 \asymp \|\Theta_\tau\|_2$ ; recall that our experiment has  $\Sigma$  with 1's on the diagonal so that  $\Sigma = R$ . Lastly, we speculate that  $S_r$  is minimax optimal in detecting the signal  $\|R - I_m\|_2$  for the null hypothesis  $H_0$  under a Gaussian assumption on  $\mathbf{X}$ , although to our knowledge this has not yet been demonstrated theoretically in the literature; see also the last section of Cai and Ma (2013) for other related open problems.

To provide further evidence for the conjectures we have made, we repeated the above simulation study in a case without equicorrelation. Specifically, we generated data from elliptical distributions with scatter matrices  $\Sigma$  that are pentadiagonal. The precise setup has  $\Sigma$  with 1's on the diagonal, equal values for the entries  $\sigma_{ij}$ ,  $1 \leq |i - j| \leq 2$ , and zeros elsewhere. The results are reported in the Supplementary Material [Leung and Drton (2018)] and lead to similar conclusions as Table 3.

TABLE 4

*Simulated power when contaminating 5% of data generated from  $N_m(0, \Sigma_{\text{band2}})$ , where  $\Sigma_{\text{band2}} = (\sigma_{ij})$  has diagonal entries  $\sigma_{ii} = 1$  and off-diagonal entry  $\sigma_{ij} = 0.1$  if  $1 \leq |i - j| \leq 2$  and  $\sigma_{ij} = 0$  if  $|i - j| \geq 3$ . For each combination of  $(m, n)$  and each test, the power is calculated from 500 independently generated datasets*

Statistics	$n$	$m$					
		4	8	16	32	64	128
$S_r$	16	0.058	0.058	0.038	0.072	0.086	0.092
$S_\tau$		0.074	0.090	0.094	0.096	0.116	0.120
$T_\tau$		0.094	0.108	0.122	0.108	0.144	0.146
$S_{\rho_s}$		0.034	0.068	0.056	0.070	0.076	0.074
$T_{\hat{\rho}_s}$		0.088	0.096	0.118	0.116	0.136	0.152
$S_{t^*}$		0.078	0.114	0.114	0.130	0.150	0.162
$Z_{t^*}$		0.100	0.112	0.118	0.096	0.112	0.138
$S_r$		32	0.072	0.100	0.078	0.110	0.106
$S_\tau$	0.086		0.112	0.114	0.130	0.136	0.126
$T_\tau$	0.090		0.130	0.128	0.132	0.150	0.138
$S_{\rho_s}$	0.072		0.098	0.086	0.110	0.106	0.096
$T_{\hat{\rho}_s}$	0.084		0.126	0.114	0.138	0.136	0.128
$S_{t^*}$	0.068		0.114	0.130	0.122	0.148	0.112
$Z_{t^*}$	0.088		0.120	0.130	0.118	0.146	0.116
$S_r$	64		0.110	0.156	0.128	0.158	0.172
$S_\tau$		0.134	0.164	0.176	0.216	0.222	0.204
$T_\tau$		0.138	0.176	0.182	0.220	0.240	0.202
$S_{\rho_s}$		0.114	0.166	0.152	0.190	0.190	0.192
$T_{\hat{\rho}_s}$		0.134	0.176	0.180	0.204	0.228	0.200
$S_{t^*}$		0.110	0.168	0.148	0.184	0.184	0.168
$Z_{t^*}$		0.130	0.170	0.174	0.192	0.184	0.190
$S_r$		128	0.224	0.290	0.332	0.342	0.384
$S_\tau$	0.306		0.390	0.408	0.436	0.454	0.484
$T_\tau$	0.308		0.392	0.418	0.440	0.462	0.484
$S_{\rho_s}$	0.296		0.376	0.392	0.418	0.444	0.470
$T_{\hat{\rho}_s}$	0.302		0.398	0.414	0.434	0.452	0.424
$S_{t^*}$	0.198		0.292	0.338	0.356	0.370	0.412
$Z_{t^*}$	0.274		0.336	0.402	0.388	0.412	0.414

Finally, in Table 4, we report Monte Carlo estimates of power in a setting of data contamination and without restricting solely to Kendall’s tau. We generate data as  $n$  independent random vectors  $\mathbf{X}_1, \dots, \mathbf{X}_n$  whose  $m$  coordinates are dependent. Each  $\mathbf{X}_i$  is multivariate normal, with mean vector zero and pentadiagonal covariance matrix. Precisely,  $\mathbf{X}_i \sim N_m(0, \Sigma_{\text{band2}})$ , where  $\Sigma_{\text{band2}} = (\sigma_{ij})$  has diagonal entries  $\sigma_{ii} = 1$  and entry  $\sigma_{ij} = 0.1$  if  $1 \leq |i - j| \leq 2$  and  $\sigma_{ij} = 0$  if  $|i - j| \geq 3$ . For each combination of  $(n, m)$ , we randomly select 5% of the  $nm$  values of the

data matrix to be contaminated. Each selected value is replaced by an independent draw from  $N(2.5, 0.2)$  multiplied with a random sign. Such outliers tend to decrease observed correlations, but the rank correlations are affected less than Pearson correlations. The empirical power of these tests is computed based on 500 repetitions of experiments. As the results in Table 4 show, Schott's  $S_r$  tends to give smaller power than the other statistics. At the larger sample sizes, when the test have approximately nominal size (recall Table 2), the "Kendall statistics"  $S_\tau$  and  $T_\tau$  show rather similar power, and the same happens for  $S_{\rho_s}$  and  $T_{\hat{\rho}_s}$ . For the Bergsma–Dassios statistics, there is some evidence that  $Z_{t^*}$  has greater power than  $S_{t^*}$  in this setting.

6.3. *Comparison of the statistics.* When data are approximately Gaussian, the statistic  $S_r$  of Schott (2005) yields a powerful test. Since the computation of a Pearson correlation is linear in the sample size  $n$ , it is inexpensive to compute, and its distribution is well approximated by a normal limit at surprisingly small sample sizes [see Table 1 in the original paper of Schott (2005)]. However, as one would expect, our simulations show that the size of the test may be far from nominal in non-Gaussian settings.

The Kendall and Spearman "sum of squares"  $S_\tau$  and  $S_{\rho_s}$  are attractive alternatives that are nearly as efficient to compute as  $S_r$ . The use of rank correlations guards against effects of non-Gaussianity all the while leading to rather little loss in power when data are indeed Gaussian. Compared to  $S_{\rho_s}$ ,  $S_\tau$  requires somewhat larger samples for the normal approximation to the null distribution to be useful.

The statistics  $T_h$  similarly guard against non-Gaussianity but are computationally more costly to use. However, as we explored in the case of  $T_\tau$ , their unbiasedness as an estimator of the signal strength leads to power that is similarly large across very different alternatives. We consider this an attractive feature and conjecture that these statistics are minimax optimal, at least for a wide class of elliptical distributions.

Another interesting assessment of independence is obtained by using the statistics  $Z_{t^*}$  and  $Z_D$  based on Bergsma and Dassios' sign covariance  $t^*$  and Hoeffding's  $D$ , respectively. Both  $t^*$  and  $D$  have the intriguing property of providing a consistent assessment of pairwise independence. For continuous observation, their expectations are zero if and only if the considered pair of random variables is independent. In the case of  $t^*$ , this also holds for discrete variables. Under independence,  $t^*$  and  $D$  are degenerate U-statistics [Nandy, Weihs and Drton (2016)]. The computational cost of their use in  $Z_{t^*}$  and  $Z_D$  is comparable to that of  $T_\tau$ . However, determining the signal strength relevant for  $Z_{t^*}$  is more complicated than for  $T_\tau$ . We are not aware of any literature that would offer a simple relationship between the expectation of  $t^*$  or  $D$  and the scatter matrix of an elliptical distribution.



**7. Discussion.** This paper treats nonparametric tests of independence using pairwise rank correlations or, more precisely, rank correlations that are also U-statistics. As reviewed in Section 2, the motivating examples are Kendall’s tau and Spearman’s rho but also Hoeffding’s  $D$  and Bergsma and Dassios’ sign covariance  $t^*$ . The latter two correlations allow for consistent assessment of pairwise independence but form degenerate U-statistics. With a view toward alternatives in which dependence is “spread out over many coordinates”, we proposed statistics that are formed as sums of many pairwise dependency signals as explained in Section 3. In a high-dimensional regime in which both the number of variables  $m$  and the sample size  $n$  tend to infinity, we derived normal limits for the null distributions of these statistics (Section 4). Our general framework gives results for U-statistic degeneracy of order up to two. Finally, we explored aspects of power theoretically and in simulations (Sections 5 and 6).

Under the null hypothesis of independence, the  $m$  rank vectors are independent, each following a uniform distribution on the symmetric group  $\mathfrak{S}_n$ . In small to moderate size problems, we may thus implement exact tests using Monte Carlo simulation to compute critical values. However, for large-scale problems and/or when using the computationally more involved  $t^*$  or  $D$ , the asymptotic normal distributions we derived furnish accurate approximations and allow for great computational savings.

Our study of power has focused on the case of Kendall’s tau. In a minimax paradigm and for Gaussian equicorrelation alternatives, we showed rate-optimality for the test based on  $T_\tau$ , the unbiased estimator of the signal strength defined via (3.6) with kernel  $h = h_\tau$ . It would be an interesting problem for future work to prove such rate-optimality more broadly, for more general alternatives as well as other kernels. In particular, for the kernel associated to Kendall’s tau, we conjectured in Section 5.1 that rate-optimality holds for alternatives from a wide class of elliptical distributions.

APPENDIX: MOTIVATION OF SCHOTT’S STATISTIC AS A RAO SCORE

We show that, up to a rescaling by the squared sample size, the statistic  $S_r$  from (1.3) is Rao’s score statistic in the multivariate normal setting. Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be i.i.d.  $m$ -variate normal random vectors with mean vector  $\boldsymbol{\mu}$  and precision matrix  $K = (\kappa^{(pq)})$ . Let  $\bar{\mathbf{X}} := \frac{1}{n} \sum_i \mathbf{X}_i$  be the sample mean vector, and let  $W = (w^{(pq)}) := \frac{1}{n} \sum_i (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$  be the sample covariance matrix. The score test considers the gradient  $\nabla l_n$  of the multivariate normal log-likelihood function

$$l_n(\boldsymbol{\mu}, K) = \frac{1}{2}n \log |K| - \frac{1}{2} \sum_i (\mathbf{X}_i - \boldsymbol{\mu})^T K (\mathbf{X}_i - \boldsymbol{\mu})$$

at the maximum likelihood estimate  $(\hat{\boldsymbol{\mu}}_0, \hat{K}_0)$  under the null hypothesis  $H_0$  from (1.1). Specifically, the score test rejects  $H_0$  for large values of

(A.1)  $\quad \nabla l_n(\hat{\boldsymbol{\mu}}_0, \hat{K}_0)^T I(\hat{\boldsymbol{\mu}}_0, \hat{K}_0)^{-1} \nabla l_n(\hat{\boldsymbol{\mu}}_0, \hat{K}_0),$

where  $I(\boldsymbol{\mu}, K)$  is the Fisher-information matrix, and the estimates are  $\hat{\boldsymbol{\mu}}_0 = \bar{\mathbf{X}}$  and  $\hat{K}_0 = \text{diag}(w^{(11)}, \dots, w^{(mm)})^{-1}$ .

Routine calculations show that

$$(A.2) \quad \left. \frac{\partial l_n}{\partial \boldsymbol{\mu}} \right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}_0, K=\hat{K}_0} = 0, \quad \left. \frac{\partial l_n}{\partial \kappa^{(pq)}} \right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}_0, K=\hat{K}_0} = \begin{cases} 0 & \text{if } p = q, \\ -nw^{(pq)} & \text{if } p < q. \end{cases}$$

Moreover, for  $p < q$  and  $p' < q'$ ,

$$(A.3) \quad \begin{aligned} & [I(\hat{\boldsymbol{\mu}}_0, \hat{K}_0)]_{\kappa^{(pq)}, \kappa^{(p'q')}} \\ &= \mathbb{E}_{\hat{\boldsymbol{\mu}}_0, \hat{K}_0} [(X^{(p)} - \mu^{(p)})(X^{(q)} - \mu^{(q)})(X^{(p')} - \mu^{(p')})(X^{(q')} - \mu^{(q')})] \\ &= \begin{cases} ([\hat{K}_0]_{pp}[\hat{K}_0]_{qq})^{-1} & \text{if } (p, q) = (p', q'), \\ 0 & \text{if } (p, q) \neq (p', q'), \end{cases} \end{aligned}$$

where  $\mathbb{E}_{\hat{\boldsymbol{\mu}}_0, \hat{K}_0}$  means taking expectation under a multivariate normal distribution with mean  $\hat{\boldsymbol{\mu}}_0$  and precision matrix  $\hat{K}_0$ . In light of (A.2) and (A.3), one obtains that the statistic from (A.1) is equal to  $n^2$  times Schott's statistic  $S_r$  from (1.3).

**Acknowledgments.** The authors would like to thank the Editor and the referees for their very helpful comments which lead to many improvements in this paper.

## SUPPLEMENTARY MATERIAL

**Supplement to “Testing independence in high dimensions with sums of rank correlations”** (DOI: [10.1214/17-AOS1550SUPP](https://doi.org/10.1214/17-AOS1550SUPP); .pdf). Due to space limitations, the proofs of the results in this paper are deferred to the supplementary document [Leung and Drton (2018)].

## REFERENCES

- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. [MR1990662](#)
- BERGSMAN, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. [MR3178526](#)
- 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** 1496–1525. [MR2850210](#)
- CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. [MR3160557](#)
- CHEN, X. (2016). Gaussian approximation for the sup-norm of high-dimensional matrix-variate U-statistics and its applications. ArXiv preprint. Available at [arXiv:1602.00199](https://arxiv.org/abs/1602.00199).
- CHEN, Y. and SHAO, Q.-M. (2012). Berry–Esseen inequality for unbounded exchangeable pairs. In *Probability Approximations and Beyond. Lecture Notes in Statistics* **205** 13–30. Springer, New York. [MR3289373](#)

- CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. [MR2724863](#)
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. [MR3161448](#)
- CHRISTENSEN, D. (2005). Fast algorithms for the calculation of Kendall's  $\tau$ . *Comput. Statist.* **20** 51–62. [MR2162534](#)
- HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. [MR0624435](#)
- HAN, F. and LIU, H. (2014). Distribution-free tests of independence with applications to testing more structures. Available at [arXiv:1410.4179](#).
- HELLER, Y. and HELLER, R. (2016). Computing the Bergsma Dassios sign-covariance. Available at [arXiv:1605.08732](#).
- HOEFFDING, W. (1948a). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. [MR0029139](#)
- HOEFFDING, W. (1948b). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. [MR0026294](#)
- JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. [MR2052906](#)
- JIANG, T. and QI, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* **42** 988–1009. [MR3426306](#)
- JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. [MR0312619](#)
- KRUSKAL, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **53** 814–861. [MR0100941](#)
- LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. [MR1926169](#)
- LEUNG, D. and DRTON, M. (2018). Supplement to “Testing independence in high dimensions with sums of rank correlations.” DOI:[10.1214/17-AOS1550SUPP](#).
- LI, D., LIU, W.-D. and ROSALSKY, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148** 5–35. [MR2653220](#)
- LI, D., QI, Y. and ROSALSKY, A. (2012). On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivariate Anal.* **111** 256–270. [MR2944420](#)
- LINDSKOG, F., MCNEIL, A. and SCHMOCK, U. (2003). Credit risk. In *Kendall's Tau for Elliptical Distributions* 149–156. Physica-Verlag, Heidelberg.
- LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. [MR2474539](#)
- MAO, G. (2014). A new test of independence for high-dimensional data. *Statist. Probab. Lett.* **93** 14–18. [MR3244549](#)
- NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. [MR0339405](#)
- NANDY, P., WEIHS, L. and DRTON, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electron. J. Stat.* **10** 2287–2311. [MR3541972](#)
- OJA, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks. Lecture Notes in Statistics* **199**. Springer, New York. [MR2598854](#)
- SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. [MR2234197](#)
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [MR0595165](#)

- VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- WANG, G., ZOU, C. and WANG, Z. (2013). A necessary test for complete independence in high dimensions using rank-correlations. *J. Multivariate Anal.* **121** 224–232. [MR3090478](#)
- WEIHS, L., DRTON, M. and LEUNG, D. (2016). Efficient computation of the Bergsma–Dassios sign covariance. *Comput. Statist.* **31** 315–328. [MR3481807](#)
- ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345–5363. [MR2327033](#)

DEPARTMENT OF STATISTICS  
CHINESE UNIVERSITY OF HONG KONG  
SHATIN N.T., HONG KONG  
CHINA  
E-MAIL: [dmhleung@uw.edu](mailto:dmhleung@uw.edu)

DEPARTMENT OF STATISTICS  
UNIVERSITY OF WASHINGTON  
SEATTLE, WASHINGTON 98195  
USA  
E-MAIL: [md5@uw.edu](mailto:md5@uw.edu)