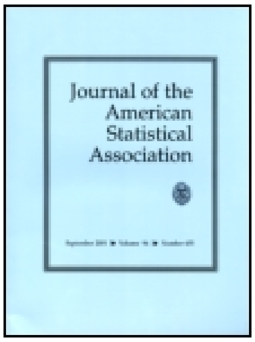

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://amstat.tandfonline.com/loi/uasa20

Distribution-Free Consistent Independence Tests via Center-Outward Ranks and Signs

Hongjian Shi, Mathias Drton & Fang Han

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2020, VOL. 00, NO. 0, 1–16: Theory and Methods
https://doi.org/10.1080/01621459.2020.1782223

Taylor & Francis
Taylor & Francis Group

Distribution-Free Consistent Independence Tests via Center-Outward Ranks and Signs

Hongjian Shi[a], Mathias Drton[b], and Fang Han[a]

[a]Department of Statistics, University of Washington, Seattle, WA; [b]Department of Mathematics, Technical University of Munich, Garching bei München, Germany

ABSTRACT

This article investigates the problem of testing independence of two random vectors of general dimensions. For this, we give for the first time a distribution-free consistent test. Our approach combines distance covariance with the center-outward ranks and signs developed by Marc Hallin and collaborators. In technical terms, the proposed test is consistent and distribution-free in the family of multivariate distributions with nonvanishing (Lebesgue) probability densities. Exploiting the (degenerate) *U*-statistic structure of the distance covariance and the combinatorial nature of Hallin's center-outward ranks and signs, we are able to derive the limiting null distribution of our test statistic. The resulting asymptotic approximation is accurate already for moderate sample sizes and makes the test implementable without requiring permutation. The limiting distribution is derived via a more general result that gives a new type of combinatorial noncentral limit theorem for double- and multiple-indexed permutation statistics. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received October 2019
Accepted June 2020

KEYWORDS
Center-outward ranks and signs; Combinatorial noncentral limit theorem; Degenerate *U*-statistics; Distance covariance; Independence test

1. Introduction

Let $\boldsymbol{X} \in \mathbb{R}^p$ and $\boldsymbol{Y} \in \mathbb{R}^q$ be two real random vectors defined on the same (otherwise unspecified) probability space. This article treats the problem of testing the null hypothesis

$$H_0 : \boldsymbol{X} \text{ and } \boldsymbol{Y} \text{ are independent,} \tag{1}$$

based on n independent copies $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \dots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ of $(\boldsymbol{X}, \boldsymbol{Y})$. Testing independence is a fundamental statistical problem that has received much attention in literature.

For the simplest instance, the bivariate case with $p = q = 1$, Hoeffding (1940), Hoeffding (1948), Blum, Kiefer, and Rosenblatt (1961), Yanagimoto (1970), Feuerverger (1993), Bergsma and Dassios (2014), among many others, have proposed tests that are consistent against all alternatives from slightly different but rather general classes of distributions. The tests are usually formulated using (univariate) ranks of the data, although recently more tests were proposed based on alternative summaries of the data, including (i) binning approaches based on a partition of the sample space (Heller, Heller, and Gorfine 2013; Heller et al. 2016; Ma and Mao 2019; Zhang 2019), (ii) mutual information (Kraskov, Stögbauer, and Grassberger 2004; Kinney and Atwal 2014; Berrett and Samworth 2019), and (iii) the maximal information coefficient (Reshef et al. 2011, 2016, 2018).

Testing independence of $\boldsymbol{X}$ and $\boldsymbol{Y}$ consistently when one or both of the dimensions p and q are larger than one is substantially more challenging, as noted in Feuerverger (1993, sec. 7). Solutions have not been discovered until much more recently. Two tracks were pursued. First, Székely, Rizzo, and Bakirov (2007) generalized Feuerverger's statistic to multivariate cases and proposed a new dependence measure termed "distance covariance." It has been shown that under the existence of finite marginal first moments, the distance covariance is zero if and only if H_0 holds. For further extensions, Lyons (2013) generalized distance covariance/correlation to general metric spaces, and Jakobsen (2017) considered the corresponding test of independence in metric spaces.

The second track to characterize nonlinear, nonmonotone dependence is based on the maximal correlation introduced in Hirschfeld (1935) and Gebelein (1941), reformulated and examined by Rényi (1959a, 1959b). Gretton, Smola, et al. (2005), Gretton, Bousquet, et al. (2005), and Gretton, Herbrich, et al. (2005) extended this idea to examine multivariate cases, resulting in the Hilbert–Schmidt independence criterion (HSIC), which is a consistent kernel-based measure of dependence in multivariate cases. Interestingly, Gretton et al. (2008) connected HSIC with a Gaussian kernel to the characteristic function-based statistic raised in Feuerverger (1993), and Sejdinovic et al. (2013) pointed out the equivalence between distance covariance in general metric spaces and the kernel-based independence criterion.

A notable feature of both distance- and kernel-based statistics is that their null distributions depend on the distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ even in the large-sample limit. This dependence arises already for $p = q = 1$ and is usually difficult to estimate. As a consequence, the tests are, unlike the rank tests of, for example, Hoeffding (1948) and Blum, Kiefer, and Rosenblatt (1961), no longer distribution-free and permutation analysis has to be conducted to implement them. To remedy this problem, Székely, Rizzo, and Bakirov (2007) proposed a nonparametric

CONTACT Mathias Drton mathias.drton@tum.de Department of Mathematics, Technical University of Munich, 85748 Garching bei München, Germany.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/JASA*.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

test based on distance correlation by applying a universal upper tail probability bound for all quadratic forms of centered Gaussian random variables that have their mean equal to one (Székely and Bakirov 2003). However, in practice this upper bound is usually too conservative for the approach to be a competitor to the computationally much more expensive permutation test (Gretton et al. 2008; Székely and Rizzo 2009). This triggers the following question: For general $p, q > 1$, does there exist an asymptotically accurate consistent test of H_0 that is distribution-free and hence directly implementable?

Rank-based tests constitute a natural approach to answering the above question. Indeed, in contrast to Székely and Rizzo (2009), Rémillard (2009) claimed that the methods based on marginal ranks are effective and as powerful as original ones when the sample size is moderately large and this idea has been explored in depth in Lin (2017). However, Bakirov, Rizzo, and Székely (2006) noted that the methods based on marginal ranks do not enjoy distribution-freeness except in dimension one, which is also recorded in, for example, Theorem 2.3.2 in Lin (2017). Using the idea of projection from Escanciano (2006), Zhu et al. (2017) generalized Hoeffding's D (Hoeffding 1948) to multivariate cases, and Kim, Balakrishnan, and Wasserman (2020) proposed the analogues of Blum–Kiefer–Rosenblatt's R (Blum, Kiefer, and Rosenblatt 1961) and Bergsma–Dassios–Yanagimoto's τ^* (Yanagimoto 1970; Bergsma and Dassios 2014; Drton, Han, and Shi 2020). Weihs, Drton, and Meinshausen (2018) proposed other multivariate extensions of Hoeffding's D, Blum–Kiefer–Rosenblatt's R, and Bergsma–Dassios–Yanagimoto's τ^*, and did numerical studies comparing them to distance covariance applied to marginal ranks. Alternatively, Heller, Heller, and Gorfine (2013) developed a consistent multivariate test based on ranked distance covariance by transferring the original problem to testing independence of an aggregated 2×2 contingency table. However, all the aforementioned tests are not distribution-free when p or q is larger than 1, and due to the difficulty of accounting for the dependence within $\boldsymbol{X}$ and $\boldsymbol{Y}$, permutation analysis is required for their implementation. On the other hand, Heller, Gorfine, and Heller (2012) and Heller and Heller (2016) introduced distribution-free graph-based and rank-based tests. However, it is unclear if the former is consistent, and the latter requires choosing two arbitrary reference points. The latter test is almost surely consistent in the sense that the choice of reference points needs to avoid an (unknown) measure zero set.

This article proposes a solution to the above question by combining Székely, Rizzo, and Bakirov's distance covariance with a recently defined concept of multivariate ranks due to Hallin (2017). Due to the lack of a canonical ordering on $\mathbb{R}^d$ for $d > 1$, fundamental concepts related to distribution functions in dimension $d = 1$, such as ranks and quantiles, do not admit a simple extension for $d \geq 2$ that maintains properties such as distribution-freeness. To overcome this limitation, several types of multivariate ranks have been introduced; see Hallin (2017, sec. 1.3) and, more recently, Ghosal and Sen (2019) for a literature review. None of them, however, is distribution-free except for pseudo-Mahalanobis ranks (Hallin and Paindaveine 2002a, 2002b), but these are restricted to the class of elliptically symmetric distributions (Fang, Kotz, and Ng 1990). Recently, Chernozhukov et al. (2017) introduced the concept of Monge–Kantorovich ranks and signs for all distributions with convex and compact supports, which is the first type of multivariate ranks that enjoys distribution-freeness for a rich class of distributions. Hallin (2017) generalized this definition by refraining from moment assumptions and making the solution more explicit. He also adopted the new terminology *center-outward ranks and signs*. Hallin et al. (2020) further showed that center-outward ranks and signs are not only distribution-free, but also essentially maximal ancillary, which can be interpreted as "maximal distribution-free" in view of Basu (1959). As shall be seen soon, the explicit nature of the solution is important as it allows for more delicate manipulations and ultimately allows us to form a test statistic of H_0 whose limiting null distribution can be determined. The limiting distribution furnishes an accurate approximation to the statistic's null distribution already for moderate sample sizes and allows us to avoid computationally more involved permutation analysis.

In detail, our proposed test is based on applying distance covariance to center-outward ranks and signs. We show that the test is consistent and distribution-free over the class of multivariate distributions with nonvanishing (Lebesgue) probability densities; see Section 2 for the precise definition of this class. The consistency is a consequence of a result of Figalli (2018). In light of the prior work of Székely, Rizzo, and Bakirov (2007), Hallin (2017), and Figalli (2018), our major new discovery is the form of the limiting null distribution of the test statistic, which is established with all parameters given explicitly. To this end, we study the weak convergence of U-statistics with a "degenerate" kernel and dependent (permutation) inputs, and derive a general combinatorial noncentral limit theorem (non-CLT) for double- and multiple-indexed permutation statistics. This theorem is new and of independent interest beyond our particular application of asymptotic calibration of the size of the independence test under H_0.

As we were completing this article, we became aware of an independent work by Deb and Sen (2019) who also proposed a rank-distance-covariance-based independence test. Their preprint was posted a few days before ours and presents, in particular, a result very similar to our Theorem 3.1. The derivations differ markedly, however. Deb and Sen's proof uses techniques based on characteristic functions, whereas we develop a general combinatorial non-CLT for double- and multiple-indexed permutation statistics that can be applied to the considered statistic as well as possible modifications. There are further differences in the precise setup of multivariate ranks: while we base ourselves directly on recent work by Hallin (2017) and by Figalli (2018), Deb and Sen (2019) considered transports to the unit cube rather than the unit ball (see Definition 2.2) and present weakened assumptions in the definition of the ranks.

The rest of the article is organized as follows. Section 2 introduces center-outward ranks and signs, and Section 3 specifies the proposed test. Section 4 gives the theoretical analysis, including the combinatorial non-CLT and a study of the proposed test. Computational aspects are discussed in Section 5, and numerical studies of the finite-sample behavior of our test and an analysis of stock market data are presented in Section 6. All proofs are relegated to the supplementary materials.

1.1. Notation

The sets of real and positive integer numbers are denoted $\mathbb{R}$ and $\mathbb{Z}_+$, respectively. For $n \in \mathbb{Z}_+$, we define $[\![n]\!] = \{1, 2, \ldots, n\}$. We write $\{x_1, \ldots, x_n\}$ and $\{x_i\}_{i=1}^n$ for the multiset consisting of (possibly duplicate) elements $x_1, \ldots, x_n$. We use $[x_1, \ldots, x_n]$ and $[x_i]_{i=1}^n$ to denote sequences. A permutation of a multiset $\mathcal{S} = \{x_1, \ldots, x_n\}$ is a sequence $[x_{\sigma(i)}]_{i=1}^n$, where σ is a bijection from $[\![n]\!]$ to itself. The family of all distinct permutations of a multiset $\mathcal{S}$ is denoted $\mathscr{P}(\mathcal{S})$. The Euclidean norm of $\boldsymbol{v} \in \mathbb{R}^d$ is written $\|\boldsymbol{v}\|$. We write $\mathbf{I}_d$ and $\mathbf{J}_d$ for the identity matrix and all-ones matrix in $\mathbb{R}^{d\times d}$, respectively. For a sequence of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$, we use $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d)$ as a shorthand of $(\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_d^\top)^\top$. For a function $f : \mathcal{X} \to \mathbb{R}$, we define $\|f\|_\infty := \max_{x\in\mathcal{X}} |f(x)|$. The greatest integer less than or equal to $x \in \mathbb{R}$ is denoted $\lfloor x \rfloor$. The symbol $\mathbb{1}(\cdot)$ stands for the indicator function. Throughout, c and C refer to positive absolute constants whose values may differ in different parts of the article. For any two real sequences $[a_n]_n$ and $[b_n]_n$, we write $a_n = O(b_n)$ if there exists $C > 0$ such that $|a_n| \le C|b_n|$ for all n large enough, and $a_n = o(b_n)$ if for any $c > 0$, $|a_n| \le c|b_n|$ holds for all n large enough. The symbols $\mathbb{S}_d$, $\overline{\mathbb{S}}_d$, and $\mathcal{S}_{d-1}$ stand for the open unit ball, closed unit ball, and unit sphere in $\mathbb{R}^d$, respectively. We use $\xrightarrow{\mathsf{d}}$ and $\xrightarrow{\mathsf{a.s.}}$ to denote convergence in distribution and almost surely. For any random vector $\boldsymbol{X}$, we use $P_{\boldsymbol{X}}$ to represent its probability measure.

2. Center-Outward Ranks and Signs

In this section, we introduce necessary background on center-outward ranks and signs. As in Hallin (2017), we will be focused on the family of absolutely continuous distributions on $\mathbb{R}^d$ that have a nonvanishing (Lebesgue) probability density (Definition 2.1). In what follows it is understood that the dimension d could be larger than 1 and that all considered probability measures are fixed, and not to be changed with the sample size n in particular.

Definition 2.1. Let P be an absolutely continuous probability measure on $\mathbb{R}^d$ with (Lebesgue) density f. Such P is said to be a *nonvanishing probability measure/distribution* if for all $D > 0$ there exist constants $\Lambda_{D;f} \ge \lambda_{D;f} > 0$ such that $\lambda_{D;f} \le f(\boldsymbol{x}) \le \Lambda_{D;f}$ for all $\|\boldsymbol{x}\| \le D$. We write $\mathcal{P}_d$ for the family of all nonvanishing probability measures/distributions on $\mathbb{R}^d$.

The considered generalization of ranks to higher dimensions rests on the following concept of a center-outward distribution function, whose existence and almost everywhere uniqueness within the family $\mathcal{P}_d$ is guaranteed by the main theorem in McCann (1995, p. 310).

Definition 2.2 (Hallin 2017, Definition 4.1). The *center-outward distribution function* $\mathbf{F}_\pm$ of a probability measure $P \in \mathcal{P}_d$ is the almost everywhere unique function that (i) is the gradient of a convex function on $\mathbb{R}^d$, (ii) maps $\mathbb{R}^d$ to the open unit ball $\mathbb{S}_d$, and (iii) pushes P forward to U_d, where U_d is the product of the uniform measure on $[0, 1)$ (for the radius) and the uniform measure on the unit sphere $\mathcal{S}_{d-1}$. To be explicit, property (iii) requires $U_d(B) = P(\mathbf{F}_\pm^{-1}(B))$ for any Borel set $B \subseteq \mathbb{S}_d$.

If $\boldsymbol{X} \sim P \in \mathcal{P}_d$ and we further have $E\|\boldsymbol{X}\|^2 < \infty$, then the center-outward distribution function $\mathbf{F}_\pm$ of P coincides with the L_2-optimal transport from P to U_d (Villani 2009, Theorem 9.4), i.e., it is the almost everywhere unique solution to the following optimization problem,

$$\inf_T \int_{\mathbb{R}^d} \left\| T(\boldsymbol{x}) - \boldsymbol{x} \right\|^2 dP \quad \text{subject to } T_\sharp P = U_d, \tag{2}$$

where $T_\sharp P$ denotes the push forward of P under map T. In other words, the optimization is done over all Borel-measurable maps from $\mathbb{R}^d$ to $\mathbb{R}^d$ pushing P forward to U_d. Assuming further that the Caffarelli's regularity conditions including compactness of support (Chernozhukov et al. 2017, Lemma 2.1) hold, $\mathbf{F}_\pm$ coincides with the Monge–Kantorovich vector rank transformation R_P proposed in Definition 2.1 in Chernozhukov et al. (2017). Lastly, it can be easily checked that when $d = 1$, $\mathbf{F}_\pm$ reduces to $2F - 1$, where F is the usual cumulative distribution function.

In dimension $d = 1$, the distribution function F determines the underlying probability distribution P. A natural question is then whether $\mathbf{F}_\pm$ similarly preserves all information about a distribution $P \in \mathcal{P}_d$ when $d > 1$. That this is indeed the case turns out to be highly nontrivial, and was not resolved until very recently. The following proposition shows that $\mathbf{F}_\pm$ is a homeomorphism from $\mathbb{R}^d$ to $\mathbb{S}_d$ except for a compact set with Lebesgue measure zero, indicating that all the information about the probability measure $P \in \mathcal{P}_d$ can be captured using $\mathbf{F}_\pm$. This proposition will play a key role in our later justification of the consistency of our proposed test (Theorem 3.2).

Proposition 2.1 (Figalli 2018, Theorem 1.1; Hallin 2017, Propositions 4.1 and 4.2). Let $P \in \mathcal{P}_d$, with center-outward distribution function $\mathbf{F}_\pm$. Then,

(i) $\mathbf{F}_\pm$ is a probability integral transformation of $\mathbb{R}^d$, that is, $\boldsymbol{X} \sim P$ iff $\mathbf{F}_\pm(\boldsymbol{X}) \sim U_d$;
(ii) The set $\mathbf{F}_\pm^{-1}(\mathbf{0})$ is compact and of Lebesgue measure zero. The restrictions of $\mathbf{F}_\pm$ and $\mathbf{F}_\pm^{-1}$ to $\mathbb{R}^d \backslash \mathbf{F}_\pm^{-1}(\mathbf{0})$ and $\mathbb{S}_d \backslash \{\mathbf{0}\}$ are homeomorphisms between $\mathbb{R}^d \backslash \mathbf{F}_\pm^{-1}(\mathbf{0})$ and $\mathbb{S}_d \backslash \{\mathbf{0}\}$. If $d = 1, 2$, then the set $\mathbf{F}_\pm^{-1}(\mathbf{0})$ is a singleton, and $\mathbf{F}_\pm$ and $\mathbf{F}_\pm^{-1}$ are homeomorphisms between $\mathbb{R}^d$ and $\mathbb{S}_d$.

We now move on to estimation of $\mathbf{F}_\pm$ based on n independent copies of $\boldsymbol{X} \sim P \in \mathcal{P}_d$. The considered estimator mimics the empirical version of the Monge–Kantorovich problem (2), and the key step is to "discretize" the unit ball $\mathbb{S}_d$ to n grid points. In the following, we sketch Hallin's approach to the construction of such a grid point set, with a focus on how to form the grid points when $d \ge 2$. To this end, let us first factorize n into the following form, whose existence is clear:

$$n = n_R n_S + n_0, \qquad n_R, n_S \in \mathbb{Z}_+,\ 0 \le n_0 < \min\{n_R, n_S\},$$
$$\text{with } n_R, n_S \to \infty \quad \text{as } n \to \infty. \tag{3}$$

Next, consider $n_R n_S$ intersection points between

- the n_R hyperspheres centered at $\mathbf{0}$ with radii $\frac{1}{n_R+1}, \ldots, \frac{n_R}{n_R+1}$, and
- n_S distinct unit vectors $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n_S}\}$.

The unit vectors in $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n_S}\}$ are selected such that the uniform discrete distribution on this set converges weakly to the uniform distribution on $\mathcal{S}_{d-1}$. For $d = 2$, we can choose unit vectors such that the unit circle is divided into n_S equal arcs. For $d \geq 3$, the requirement is satisfied almost surely when independently drawing n_S unit vectors from the uniform distribution on $\mathcal{S}_{d-1}$. Moreover, it is not difficult to give a deterministic construction that serves our purpose; see Section B in the supplementary materials.

Definition 2.3. When $d \geq 2$, the augmented grid $\mathcal{G}^d_{n_0,n_R,n_S}$ is the multiset consisting of n_0 copies of the origin $\mathbf{0}$ whenever $n_0 > 0$ and the intersection points $\{\frac{j}{n_R+1}\boldsymbol{r}_k : j \in [\![n_R]\!], k \in [\![n_S]\!]\}$. When $d = 1$, letting $n_S = 2$, $n_R = \lfloor n/n_S \rfloor$, and $n_0 = n - n_R n_S$, the augmented grid $\mathcal{G}^d_{n_0,n_R,n_S}$ is the multiset consisting of the origin 0 whenever $n_0 > 0$ and the points $\{\pm\frac{j}{n_R+1} : j \in [\![n_R]\!]\}$.

Proposition 2.2. As long as the uniform discrete distribution on $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n_S}\}$ converges weakly to the uniform distribution on $\mathcal{S}_{d-1}$, the uniform discrete distribution on the augmented grid $\mathcal{G}^d_{n_0,n_R,n_S}$, which assigns mass n_0/n to the origin and mass $1/n$ to every other grid point, weakly converges to U_d.

We are now ready to introduce Hallin's estimator, $\mathbf{F}^{(n)}_{\pm}$, of $\mathbf{F}_{\pm}$. It is defined via the optimal coupling between the observed data points and the augmented grid $\mathcal{G}^d_{n_0,n_R,n_S}$.

Definition 2.4 (Hallin 2017, Definition 4.2). Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be data points in $\mathbb{R}^d$. Let $\mathcal{T}$ be the collection of all bijective mappings between the multiset $\{\boldsymbol{x}_i\}_{i=1}^n$ and the augmented grid $\mathcal{G}^d_{n_0,n_R,n_S}$. The *empirical center-outward distribution function* is defined as

$$\mathbf{F}^{(n)}_{\pm} := \underset{T\in\mathcal{T}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \boldsymbol{x}_i - T(\boldsymbol{x}_i) \right\|^2, \tag{4}$$

the *center-outward rank* of $\boldsymbol{x}_i$ is defined as $(n_R + 1)\|\mathbf{F}^{(n)}_{\pm}(\boldsymbol{x}_i)\|$, and the *center-outward sign* of $\boldsymbol{x}_i$ is defined as $\mathbf{F}^{(n)}_{\pm}(\boldsymbol{x}_i)\|/\|\mathbf{F}^{(n)}_{\pm}(\boldsymbol{x}_i)\|$ if $\|\mathbf{F}^{(n)}_{\pm}(\boldsymbol{x}_i)\| \neq 0$, and $\mathbf{0}$ otherwise.

The following two propositions from Hallin (2017) give the Glivenko–Cantelli strong consistency and distribution-freeness of the empirical center-outward distribution function. Both shall play key roles for the limiting null distribution and asymptotic consistency of the test statistic that will be proposed in Section 3.

Proposition 2.3 (Glivenko–Cantelli, Hallin 2017, Proposition 5.1; del Barrio et al. 2018, Theorem 3.1). Let $P \in \mathcal{P}_d$ with center-outward distribution function $\mathbf{F}_{\pm}$, and let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be iid with distribution P with empirical center-outward distribution function $\mathbf{F}^{(n)}_{\pm}$. Then

$$\max_{1\leq i\leq n} \left\| \mathbf{F}^{(n)}_{\pm}(\boldsymbol{X}_i) - \mathbf{F}_{\pm}(\boldsymbol{X}_i) \right\| \xrightarrow{\text{a.s.}} 0 \tag{5}$$

when $n \to \infty$ and (3) holds.

Proposition 2.4 (Distribution-freeness, Hallin 2017, Proposition 6.1(ii); Hallin et al. 2020, Proposition 2.5(ii)). Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be iid with distribution $P \in \mathcal{P}_d$. Let $\mathbf{F}^{(n)}_{\pm}$ be their empirical center-outward distribution function. Then for any decomposition n_0, n_R, n_S of n, the random vector $[\mathbf{F}^{(n)}_{\pm}(\boldsymbol{X}_1), \ldots, \mathbf{F}^{(n)}_{\pm}(\boldsymbol{X}_n)]$ is uniformly distributed over $\mathscr{P}(\mathcal{G}^d_{n_0,n_R,n_S})$. The latter set is comprised of all permutations of the multiset $\mathcal{G}^d_{n_0,n_R,n_S}$; recall the notation introduced at the end of Section 1.

3. A Distribution-Free Test of Independence

This section introduces the proposed distribution-free test of H_0 in (1) built on center-outward ranks and signs. The main new methodological idea is simple: We propose to plug the calculated center-outward ranks and signs, instead of the original data, into the consistent test statistics presented in the introduction (Section 1). The distribution theory for the proposed test statistic, however, is nontrivial and requires new technical developments, which shall be detailed in Section 4.

To illustrate our idea, we will focus on one particular consistent test statistic in the sequel, namely, the distance covariance of Székely, Rizzo, and Bakirov (2007). Other choices including HSIC and more recent proposals like the ball covariance proposed in Pan et al. (2020) shall be discussed in Section 4 following the presentation of our general combinatorial non-CLT.

We begin with details on the distance covariance that are necessary to convey the main idea. We first introduce a representation of the associated measure of dependence.

Definition 3.1 (Distance covariance measure of dependence, Székely, Rizzo, and Bakirov 2007). Let $\boldsymbol{X} \in \mathbb{R}^p$ and $\boldsymbol{Y} \in \mathbb{R}^q$ be two random vectors with $E(\|\boldsymbol{X}\| + \|\boldsymbol{Y}\|) < \infty$, and let $(\boldsymbol{X}', \boldsymbol{Y}')$ be an independent copy of $(\boldsymbol{X}, \boldsymbol{Y})$. The distance covariance of $(\boldsymbol{X}, \boldsymbol{Y})$ is defined as

$$\mathrm{dCov}^2(\boldsymbol{X}, \boldsymbol{Y}) := E(d_{\boldsymbol{X}}(\boldsymbol{X}, \boldsymbol{X}') d_{\boldsymbol{Y}}(\boldsymbol{Y}, \boldsymbol{Y}')), \tag{6}$$

which is finite and uses the kernel function

$$\begin{aligned} d_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{x}') = d_{P_{\boldsymbol{X}}}(\boldsymbol{x}, \boldsymbol{x}') &:= \|\boldsymbol{x} - \boldsymbol{x}'\| - E\|\boldsymbol{x} - \boldsymbol{X}_2\| \\ &\quad - E\|\boldsymbol{X}_1 - \boldsymbol{x}'\| + E\|\boldsymbol{X}_1 - \boldsymbol{X}_2\|, \end{aligned} \tag{7}$$

and its analogue $d_{\boldsymbol{Y}}(\boldsymbol{y}, \boldsymbol{y}')$. Here $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent and both follow the distribution $P_{\boldsymbol{X}}$.

The finiteness of $\mathrm{dCov}^2(\boldsymbol{X}, \boldsymbol{Y})$ in (6) was proved by Lyons (2013, Proposition 2.3). It can be shown that under the same conditions as in Definition 3.1,

$$\mathrm{dCov}^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{4} E(s(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4) s(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y}_3, \boldsymbol{Y}_4)),$$

where $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_4, \boldsymbol{Y}_4)$ are independent copies of $(\boldsymbol{X}, \boldsymbol{Y})$ and

$$s(\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3, \boldsymbol{t}_4) := \|\boldsymbol{t}_1 - \boldsymbol{t}_2\| + \|\boldsymbol{t}_3 - \boldsymbol{t}_4\| - \|\boldsymbol{t}_1 - \boldsymbol{t}_3\| - \|\boldsymbol{t}_2 - \boldsymbol{t}_4\|;$$

see also Bergsma and Dassios (2014, sec. 3.4). Accordingly, we have an unbiased estimator of the distance covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$ as follows.

Definition 3.2 (Sample distance covariance, Székely and Rizzo 2013). Let $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ be independent copies of $(\boldsymbol{X}, \boldsymbol{Y})$ with $\boldsymbol{X} \in \mathbb{R}^p$, $\boldsymbol{Y} \in \mathbb{R}^q$, $E(\|\boldsymbol{X}\| + \|\boldsymbol{Y}\|) < \infty$. The sample distance covariance is defined as

$$\begin{aligned} \mathrm{dCov}^2_n\Big([\boldsymbol{X}_i]_{i=1}^n, [\boldsymbol{Y}_i]_{i=1}^n\Big) &= \binom{n}{4}^{-1} \\ &\sum_{1\leq i_1<\cdots<i_4\leq n} K\Big((\boldsymbol{X}_{i_1}, \boldsymbol{Y}_{i_1}), \ldots, (\boldsymbol{X}_{i_4}, \boldsymbol{Y}_{i_4})\Big), \end{aligned} \tag{8}$$

where

$$K\Big((\boldsymbol{x}_1,\boldsymbol{y}_1),\ldots,(\boldsymbol{x}_4,\boldsymbol{y}_4)\Big) := \frac{1}{4\cdot 4!} \sum_{[i_1,\ldots,i_4]\in\mathscr{P}([\![4]\!])} s(\boldsymbol{x}_{i_1},\boldsymbol{x}_{i_2},\boldsymbol{x}_{i_3},\boldsymbol{x}_{i_4})s(\boldsymbol{y}_{i_1},\boldsymbol{y}_{i_2},\boldsymbol{y}_{i_3},\boldsymbol{y}_{i_4}), \quad (9)$$

and recall $s(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{t}_3,\boldsymbol{t}_4) := \|\boldsymbol{t}_1-\boldsymbol{t}_2\| + \|\boldsymbol{t}_3-\boldsymbol{t}_4\| - \|\boldsymbol{t}_1-\boldsymbol{t}_3\| - \|\boldsymbol{t}_2-\boldsymbol{t}_4\|$.

The following is a direct consequence of Lemma 1 in Yao, Zhang, and Shao (2018b).

Proposition 3.1. Definition 1 in Székely and Rizzo (2013), Equation (3.2) in Székely and Rizzo (2014), Definition 5.3 (U-statistic) in Jakobsen (2017), and Definition 3.2 are equivalent.

We are now ready to describe our distribution-free test of independence, which combines distance covariance with center-outward ranks and signs.

Definition 3.3 (The proposed distribution-free test statistic). Let $(\boldsymbol{X}_1,\boldsymbol{Y}_1),\ldots,(\boldsymbol{X}_n,\boldsymbol{Y}_n)$ be independent copies of $(\boldsymbol{X},\boldsymbol{Y})$ with $P_{\boldsymbol{X}} \in \mathcal{P}_p$ and $P_{\boldsymbol{Y}} \in \mathcal{P}_q$. Let $\mathbf{F}^{(n)}_{\boldsymbol{X},\pm}$ and $\mathbf{F}^{(n)}_{\boldsymbol{Y},\pm}$ be the empirical center-outward distribution functions for $\{\boldsymbol{X}_i\}_{i=1}^n$ and $\{\boldsymbol{Y}_i\}_{i=1}^n$. We define the test statistic

$$\widehat{M}_n := n\cdot \mathrm{dCov}_n^2\Big([\mathbf{F}^{(n)}_{\boldsymbol{X},\pm}(\boldsymbol{X}_i)]_{i=1}^n, [\mathbf{F}^{(n)}_{\boldsymbol{Y},\pm}(\boldsymbol{Y}_i)]_{i=1}^n\Big). \quad (10)$$

By Proposition 2.4, the statistic $\widehat{M}_n$ is distribution-free under the independence hypothesis H_0 in (1). Hence, an exact critical value for rejection of H_0 can be approximated via Monte Carlo simulation. Numerically less demanding, one could instead adopt the critical value based on the limiting null distribution of $\widehat{M}_n$ derived from the following theorem.

Theorem 3.1 (Limiting null distribution). Let $(\boldsymbol{X}_1,\boldsymbol{Y}_1),\ldots,(\boldsymbol{X}_n,\boldsymbol{Y}_n)$ be independent copies of $(\boldsymbol{X},\boldsymbol{Y})$ with $P_{\boldsymbol{X}} \in \mathcal{P}_p$ and $P_{\boldsymbol{Y}} \in \mathcal{P}_q$, and $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent. Then we have

$$\widehat{M}_n \xrightarrow{\mathrm{d}} \sum_{k=1}^{\infty} \lambda_k(\xi_k^2-1), \quad (11)$$

as $n\to\infty$ and (3) holds, where λ_k, $k\in\mathbb{Z}_+$, are the nonzero eigenvalues of the integral equation

$$E\Big(d_{\boldsymbol{U}}(\boldsymbol{u},\boldsymbol{U})d_{\boldsymbol{V}}(\boldsymbol{v},\boldsymbol{V})\phi(\boldsymbol{U},\boldsymbol{V})\Big) = \lambda\phi(\boldsymbol{u},\boldsymbol{v}), \quad (12)$$

in which $d_{\boldsymbol{U}}(\boldsymbol{u},\boldsymbol{u}')$ and $d_{\boldsymbol{V}}(\boldsymbol{v},\boldsymbol{v}')$ are defined as in (7), $\boldsymbol{U}\sim U_p$ and $\boldsymbol{V}\sim U_q$ are independent, and $[\xi_k]_{k=1}^\infty$ is a sequence of independent standard Gaussian random variables.

Remark 3.1. In Section 4, we will prove Theorem 3.1 rigorously. Intuitively, it is helpful to first consider the following "oracle" test statistic $\widetilde{M}_n$:

$$\widetilde{M}_n := n\cdot \mathrm{dCov}_n^2\Big([\mathbf{F}_{\boldsymbol{X},\pm}(\boldsymbol{X}_i)]_{i=1}^n, [\mathbf{F}_{\boldsymbol{Y},\pm}(\boldsymbol{Y}_i)]_{i=1}^n\Big),$$

where $\mathbf{F}_{\boldsymbol{X},\pm}$ and $\mathbf{F}_{\boldsymbol{Y},\pm}$ denote the center-outward distribution functions of $P_{\boldsymbol{X}}$ and $P_{\boldsymbol{Y}}$, respectively. The infeasibility stems from the use of the (population) center-outward distribution functions. One can easily verify using the asymptotic theory of degenerate U-statistics that under the null

$$\widetilde{M}_n \xrightarrow{\mathrm{d}} \sum_{k=1}^{\infty} \lambda_k(\xi_k^2-1),$$

where $[\lambda_k]_{k=1}^\infty$ and $[\xi_k]_{k=1}^\infty$ are defined as in Theorem 3.1. Somewhat surprising to us, the limiting null distribution of $\widehat{M}_n$ is exactly the same as that of $\widetilde{M}_n$.

Therefore, for any prespecified significance level $\alpha\in(0,1)$, our proposed test is hence

$$\mathsf{T}_\alpha := \mathbb{1}\Big(\widehat{M}_n > Q_{1-\alpha}\Big),$$
$$Q_{1-\alpha} := \inf\Big\{x\in\mathbb{R} : P\Big(\sum_{k=1}^{\infty}\lambda_k(\xi_k^2-1)\le x\Big)\ge 1-\alpha\Big\}. \quad (13)$$

Consequently, by Theorem 3.1,

$$P(\mathsf{T}_\alpha = 1 \mid H_0) = \alpha + o(1). \quad (14)$$

It should be highlighted that, thanks to distribution-freeness, given any fixed dimensions p and q, the asymptotically small term in (14) is independent of the underlying distributions, and converges to zero uniformly over all the underlying distributions with $P_{\boldsymbol{X}} \in \mathcal{P}_p$, $P_{\boldsymbol{Y}} \in \mathcal{P}_q$, and $\boldsymbol{X}$ independent of $\boldsymbol{Y}$. The values of λ_k's, and hence also the critical value $Q_{1-\alpha}$ itself, are distribution-free and only depend on the dimensions p and q. The critical value may thus be calculated using numerical methods for each pair of p and q. Details will be described in Section 5.2. Table C.1 in the supplementary materials further records the critical values at significance levels $\alpha = 0.1, 0.05, 0.01$ for $(p,q) = (1,1),(1,2),\ldots,(10,10)$ with accuracy 2×10^{-3}.

Due to (i) the near-homeomorphism property of the center-outward distribution function shown in Proposition 2.1; (ii) the strong Glivenko–Cantelli consistency of empirical center-outward distribution functions shown in Proposition 2.3; and (iii) the fact that the distance covariance measure of dependence is zero if and only if H_0 holds under finiteness of marginal first moments (Lyons 2013, Theorem 3.11), it holds that $\widehat{M}_n$ is asymptotically consistent and the corresponding test T_α is consistent. This fact is summarized in the following theorem.

Theorem 3.2 (Consistency). Let $(\boldsymbol{X}_1,\boldsymbol{Y}_1),\ldots,(\boldsymbol{X}_n,\boldsymbol{Y}_n)$ be independent copies of $(\boldsymbol{X},\boldsymbol{Y})$, where $P_{\boldsymbol{X}} \in \mathcal{P}_p$ with center-outward distribution function $\mathbf{F}_{\boldsymbol{X},\pm}$, and $P_{\boldsymbol{Y}} \in \mathcal{P}_q$ with center-outward distribution function $\mathbf{F}_{\boldsymbol{Y},\pm}$. We then have, as long as $n\to\infty$ and (3) holds,

$$\widehat{M}_n/n \xrightarrow{\mathrm{a.s.}} \mathrm{dCov}^2\Big(\mathbf{F}_{\boldsymbol{X},\pm}(\boldsymbol{X}), \mathbf{F}_{\boldsymbol{Y},\pm}(\boldsymbol{Y})\Big), \quad (15)$$

where $\mathrm{dCov}^2(\mathbf{F}_{\boldsymbol{X},\pm}(\boldsymbol{X}), \mathbf{F}_{\boldsymbol{Y},\pm}(\boldsymbol{Y})) \ge 0$ with equality if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent. In addition, under any fixed alternative H_1, we obtain $\widehat{M}_n \xrightarrow{\mathrm{a.s.}} \infty$ if $n\to\infty$ and (3) holds, and thus

$$P(\mathsf{T}_\alpha = 1 \mid H_1) = 1 - o(1). \quad (16)$$

We conclude this section with one more remark that discusses an interesting connection between the proposed test and a famous dependence measure, Blum–Kiefer–Rosenblatt's R dependence measure (Blum, Kiefer, and Rosenblatt 1961), when $p = q = 1$.

Remark 3.2. In the univariate case ($p = q = 1$), the statistic $\widehat{M}_n/n$ is actually (up to a constant) a consistent estimator of Blum–Kiefer–Rosenblatt's R measure of dependence (Blum, Kiefer, and Rosenblatt 1961). In detail, Theorem 3.2 has shown that $\widehat{M}_n/n \xrightarrow{\text{a.s.}} \mathrm{dCov}^2(\mathbf{F}_{\mathbf{X},\pm}(\boldsymbol{X}), \mathbf{F}_{\mathbf{Y},\pm}(\boldsymbol{Y}))$. When X and Y are both absolutely continuous, Bergsma (2006, Lemma 10) showed that

$$\frac{1}{4}\mathrm{dCov}^2(X, Y) = \int \{F_{(X,Y)}(x, y) - F_X(x)F_Y(y)\}^2 dxdy,$$

where $F_Z(\cdot)$ denotes the cumulative distribution function of Z. This implies that

$$\frac{1}{16}\mathrm{dCov}^2(\mathbf{F}_{X,\pm}(X), \mathbf{F}_{Y,\pm}(Y)) = \int \{F_{(X,Y)}(x, y) - F_X(x)F_Y(y)\}^2 dF_X(x)dF_Y(y).$$

The right-hand side is Blum–Kiefer–Rosenblatt's R and $\widehat{M}_n/(16n)$ converges to it almost surely.

4. Theoretical Analysis

This section provides the theoretical justification for the test in (13). By Proposition 2.4, both $[\mathbf{F}^{(n)}_{\mathbf{X},\pm}(\boldsymbol{X}_i)]_{i=1}^n$ and $[\mathbf{F}^{(n)}_{\mathbf{Y},\pm}(\boldsymbol{Y}_i)]_{i=1}^n$ are generated from uniform permutation measures. In view of Definition 3.3, it is hence clear that under H_0 the test statistic $\widehat{M}_n$ is a summation over the product space of two uniform permutation measures, which belongs to the family of permutation statistics.

The study of permutation statistics can be traced back at least to Wald and Wolfowitz (1944), who proved an asymptotic normality result for single-indexed permutation statistics of the form $\sum_{i=1}^n x_i^{(n)} y_{\pi_i}^{(n)}$. Here $\boldsymbol{x}^{(n)}$ and $\boldsymbol{y}^{(n)}$ are vectors that are possibly varying with n, and $\boldsymbol{\pi}$ is uniformly distributed on $\mathscr{P}([\![n]\!])$. Later, Noether (1949), Hoeffding (1951), Motoo (1957), and Hájek (1961), among many others, generalized Wald and Wolfowitz's results in different ways, and Bolthausen (1984) gave a sharp Berry–Esseen bound for such permutation statistics using Stein's method.

Double-indexed permutation statistics, of the form $\sum_{i\neq j} A_{ij}^{(n)} B_{\pi_i\pi_j}^{(n)}$ with $\mathbf{A}^{(n)}$ and $\mathbf{B}^{(n)}$ as matrices possibly varying with n, are more difficult to tackle. They were first investigated by Daniels (1944), who gave sufficient conditions for asymptotic normality. Later, various weakened conditions were introduced in, for example, Bloemena (1964, chap. 4.1), Jogdeo (1968), Abe (1969), Cliff and Ord (1973, chap. 2.4), Shapiro and Hubert (1979), Barbour and Eagleson (1986), Pham, Möcks, and Sroka (1989), and the Berry–Esseen bound was established in Zhao et al. (1997), Barbour and Chen (2005), and Reinert and Röllin (2009).

Despite this vast literature, there is a notable absence of results on permutation statistics which, as its degenerate U-statistics "cousins", may weakly converge to a nonnormal distribution. Our analysis of $\widehat{M}_n$, however, hinges on such a combinatorial non-CLT. In the following, we present two general theorems that fill the gap.

Before stating the two theorems, we introduce some notions needed. For each $i = 1, 2$, let $\boldsymbol{Z}_i$ be a random vector taking values in Ω_i, a compact subset of $\mathbb{R}^{p_i}$. We consider triangular arrays $\{\boldsymbol{z}_{i;j}^{(n)}, n \in \mathbb{Z}_+, j \in [\![n]\!]\}$, for $i = 1, 2$, such that the random variables with uniform discrete distributions on the respective multisets $\{\boldsymbol{z}_{i;j}^{(n)}, j \in [\![n]\!]\}$, denoted by $\boldsymbol{Z}_i^{(n)}$, weakly converge to $\boldsymbol{Z}_i$ as $n \to \infty$. We further introduce an independent copy of $\boldsymbol{Z}_i$, denoted $\boldsymbol{Z}_i'$, and independent copies of the $\boldsymbol{Z}_i^{(n)}$, denoted $\boldsymbol{Z}_i^{(n)\prime}$. Finally, for $i = 1, 2$ and $n \in \mathbb{Z}_+$, let $g_i^{(n)}, g_i : \Omega_i \times \Omega_i \to \mathbb{R}$ be real-valued functions, the former of which may change with n.

Our first theorem is then focused on double-indexed permutation-statistics of the form

$$\widehat{D}^{(n)} = \binom{n}{2}^{-1} \sum_{1\le j_1<j_2\le n} g_1^{(n)}\Big(\boldsymbol{z}_{1;j_1}^{(n)}, \boldsymbol{z}_{1;j_2}^{(n)}\Big) g_2^{(n)}\Big(\boldsymbol{z}_{2;\pi_{j_1}}^{(n)}, \boldsymbol{z}_{2;\pi_{j_2}}^{(n)}\Big), \tag{17}$$

where $\boldsymbol{\pi}$ is uniformly distributed on $\mathscr{P}([\![n]\!])$.

Theorem 4.1. Assume that for each $i = 1, 2$, the functions $g_i^{(n)}$, $n \in \mathbb{Z}_+$, and g_i satisfy the following conditions:

(i) each $g_i^{(n)}$ is symmetric, that is, $g_i^{(n)}(\boldsymbol{z}, \boldsymbol{z}') = g_i^{(n)}(\boldsymbol{z}', \boldsymbol{z})$ for all $\boldsymbol{z}, \boldsymbol{z}' \in \Omega_i$;
(ii) the family $g_i^{(n)}$, $n \in \mathbb{Z}_+$, is equicontinuous;
(iii) each $g_i^{(n)}$ is nonnegative definite, that is,

$$\sum_{j_1,j_2=1}^{\ell} c_{j_1}c_{j_2}g_i^{(n)}(\boldsymbol{z}_{j_1}, \boldsymbol{z}_{j_2}) \ge 0$$

for all $c_1, \ldots, c_\ell \in \mathbb{R}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_\ell \in \Omega_i, \ell \in \mathbb{Z}_+$;
(iv) each $g_i^{(n)}$ has $E(g_i^{(n)}(\boldsymbol{z}, \boldsymbol{Z}_i^{(n)})) = 0$;
(v) each $g_i^{(n)}$ has $E(g_i^{(n)}(\boldsymbol{Z}_i^{(n)}, \boldsymbol{Z}_i^{(n)\prime})^2) \in (0, +\infty)$;
(vi) as $n \to \infty$, the functions $g_i^{(n)}$ converge uniformly on Ω_i to g_i, with $E(g_i(\boldsymbol{Z}_i, \boldsymbol{Z}_i')^2) \in (0, +\infty)$.

It then holds that

$$n\widehat{D}^{(n)} \xrightarrow{\text{d}} \sum_{k_1,k_2=1}^{\infty} \lambda_{1,k_1}\lambda_{2,k_2}(\xi_{k_1,k_2}^2 - 1)$$

as $n \to \infty$, where ξ_{k_1,k_2}, $k_1, k_2 \in \mathbb{Z}_+$, are iid standard Gaussian, and the $\lambda_{i,k} \ge 0$, $k \in \mathbb{Z}_+$, are eigenvalues of the Hilbert–Schmidt integral operator given by g_i, that is, for each i the $\lambda_{i,k}$'s solve the integral equations

$$E(g_i(\boldsymbol{z}_i, \boldsymbol{Z}_i)e_{i,k}(\boldsymbol{Z}_i)) = \lambda_{i,k}e_{i,k}(\boldsymbol{z}_i)$$

for a system of orthonormal eigenfunctions $e_{i,k}$.

Theorem 4.1 provides the essential component of our analysis for $\widehat{M}_n$. However, $\widehat{M}_n$ is a permutation statistic that is not double- but quadruple-indexed. To cover this case, we have to extend Theorem 4.1 to multiple-indexed permutation statistics, the study of which is much more sparse (see, e.g., Raič 2015 for some recent progresses). Further notation is needed.

For all $j \in \mathbb{Z}_+$, let $\boldsymbol{w}_j = (\boldsymbol{z}_{1;j}, \boldsymbol{z}_{2;j})$ be a vector with $\boldsymbol{z}_{i;j} \in \Omega_i$, for $i = 1, 2$. Let $h : (\Omega_1 \times \Omega_2)^m \to \mathbb{R}$ be a *symmetric* kernel of order m, that is, $h(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m) = h(\boldsymbol{w}_{\sigma_1}, \ldots, \boldsymbol{w}_{\sigma_m})$ for all permutations $\sigma \in \mathscr{P}(\llbracket m \rrbracket)$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m \in \Omega_1 \times \Omega_2$. For any integer $\ell \in \llbracket m \rrbracket$, and any measure $P_{\boldsymbol{W}}$, we let

$$h_\ell(\boldsymbol{w}_1 \ldots, \boldsymbol{w}_\ell; P_{\boldsymbol{W}}) := E(h(\boldsymbol{w}_1 \ldots, \boldsymbol{w}_\ell, \boldsymbol{W}_{\ell+1}, \ldots, \boldsymbol{W}_m)),$$

where $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_m$ are m independent random vectors with distribution $P_{\boldsymbol{W}}$.

The next theorem treats a multiple-indexed permutation-statistic of order m defined as

$$\widehat{\Pi}^{(n)} = \binom{n}{m}^{-1} \sum_{1 \le j_1 < \cdots < j_m \le n} h\Big((\boldsymbol{z}^{(n)}_{1;j_1}, \boldsymbol{z}^{(n)}_{2;\pi_{j_1}}), \ldots, (\boldsymbol{z}^{(n)}_{1;j_m}, \boldsymbol{z}^{(n)}_{2;\pi_{j_m}})\Big), \tag{18}$$

where $\boldsymbol{\pi}$ is uniformly distributed on $\mathscr{P}(\llbracket n \rrbracket)$, and the triangular arrays $\{\boldsymbol{z}^{(n)}_{i;j}, n \in \mathbb{Z}_+, j \in \llbracket n \rrbracket\}$, $i = 1, 2$ are as introduced before the statement of Theorem 4.1.

Theorem 4.2. Let $\boldsymbol{Z}_i$ and $\boldsymbol{Z}^{(n)}_i$, $i = 1, 2$, be defined as for Theorem 4.1. Assume the kernel h has the following three properties:

(I) h is continuous with $\|h\|_\infty < \infty$;
(II) $h_1\Big(\boldsymbol{w}_1; P_{\boldsymbol{Z}^{(n)}_1} \times P_{\boldsymbol{Z}^{(n)}_2}\Big) = 0$;
(III) one has

$$\binom{m}{2} \cdot h_2\Big(\boldsymbol{w}_1, \boldsymbol{w}_2; P_{\boldsymbol{Z}^{(n)}_1} \times P_{\boldsymbol{Z}^{(n)}_2}\Big) = g^{(n)}_1(\boldsymbol{z}_{1;1}, \boldsymbol{z}_{1;2}) g^{(n)}_2(\boldsymbol{z}_{2;1}, \boldsymbol{z}_{2;2}),$$

$$\text{and} \quad \binom{m}{2} \cdot h_2\Big(\boldsymbol{w}_1, \boldsymbol{w}_2; P_{\boldsymbol{Z}_1} \times P_{\boldsymbol{Z}_2}\Big) = g_1(\boldsymbol{z}_{1;1}, \boldsymbol{z}_{1;2}) g_2(\boldsymbol{z}_{2;1}, \boldsymbol{z}_{2;2}),$$

where for each $i = 1, 2$, $g^{(n)}_i$, $n \in \mathbb{Z}_+$, and g_i satisfy Assumptions (i)–(vi) from Theorem 4.1.

We then have

$$n\widehat{\Pi}^{(n)} \xrightarrow{\mathrm{d}} \sum_{k_1, k_2 = 1}^{\infty} \lambda_{1,k_1} \lambda_{2,k_2} (\xi^2_{k_1,k_2} - 1)$$

as $n \to \infty$, where $\lambda_{i,k}$ and ξ_{k_1,k_2} are defined as in Theorem 4.1.

With the aid of Theorem 4.2, we are now ready to prove Theorem 3.1, which presents the limiting null distribution of $\widehat{M}_n$. In our context, $p_1 = p$, $p_2 = q$, $m = 4$, and h is the kernel K defined in (9). The multisets $\{\boldsymbol{z}^{(n)}_{1;j}, j \in \llbracket n \rrbracket\}$ and $\{\boldsymbol{z}^{(n)}_{2;j}, j \in \llbracket n \rrbracket\}$ are taken to be $\{\boldsymbol{u}^{(n)}_j, j \in \llbracket n \rrbracket\} := \mathcal{G}^p_{n_0,n_R,n_S}$ and $\{\boldsymbol{v}^{(n)}_j, j \in \llbracket n \rrbracket\} := \mathcal{G}^q_{n_0,n_R,n_S}$, respectively. Accordingly, $\boldsymbol{Z}^{(n)}_1$ follows the uniform discrete distribution over $\mathcal{G}^p_{n_0,n_R,n_S}$, denoted by $\boldsymbol{U}^{(n)}$, and $\boldsymbol{Z}^{(n)}_2$ has a uniform discrete distribution over $\mathcal{G}^q_{n_0,n_R,n_S}$, denoted by $\boldsymbol{V}^{(n)}$. The functions $g^{(n)}_1, g_1, g^{(n)}_2$, and g_2 can be chosen as $-d_{\boldsymbol{U}^{(n)}}, -d_{\boldsymbol{U}}, -d_{\boldsymbol{V}^{(n)}}$, and $-d_{\boldsymbol{V}}$, defined in the manner of (7), respectively.

We now verify properties (I)–(III). Write $\boldsymbol{w} = (\boldsymbol{u}, \boldsymbol{v})$ and $\boldsymbol{w}' = (\boldsymbol{u}', \boldsymbol{v}')$. Notice that the kernel K is symmetric and continuous on $\overline{\mathbb{S}}_p \times \overline{\mathbb{S}}_q$. We have

$$K_1\Big(\boldsymbol{w}; P_{\boldsymbol{U}^{(n)}} \times P_{\boldsymbol{V}^{(n)}}\Big) = 0, \quad 6K_2\Big(\boldsymbol{w}, \boldsymbol{w}'; P_{\boldsymbol{U}^{(n)}} \times P_{\boldsymbol{V}^{(n)}}\Big) = \Big(-d_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}, \boldsymbol{u}')\Big)\Big(-d_{\boldsymbol{V}^{(n)}}(\boldsymbol{v}, \boldsymbol{v}')\Big),$$

$$\text{and} \quad 6K_2\Big(\boldsymbol{w}, \boldsymbol{w}'; P_{\boldsymbol{U}} \times P_{\boldsymbol{V}}\Big) = \Big(-d_{\boldsymbol{U}}(\boldsymbol{u}, \boldsymbol{u}')\Big)\Big(-d_{\boldsymbol{V}}(\boldsymbol{v}, \boldsymbol{v}')\Big),$$

by Yao, Zhang, and Shao (2018a, sec. 1.1). Moreover, the $-d_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}, \boldsymbol{u}')$ is symmetric, nonnegative definite (Lyons 2013, p. 3291), and equicontinuous since

$$|-d_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}, \boldsymbol{u}') - (-d_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}''', \boldsymbol{u}''))| \le 2\|\boldsymbol{u} - \boldsymbol{u}'''\| + 2\|\boldsymbol{u}' - \boldsymbol{u}''\|.$$

One can verify that $E[-d_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}, \boldsymbol{U}^{(n)})] = 0$, and $-d_{\boldsymbol{U}^{(n)}}(\boldsymbol{u}, \boldsymbol{u}')$ converges uniformly to $-d_{\boldsymbol{U}}(\boldsymbol{u}, \boldsymbol{u}')$ by combining the pointwise convergence using the portmanteau lemma (van der Vaart

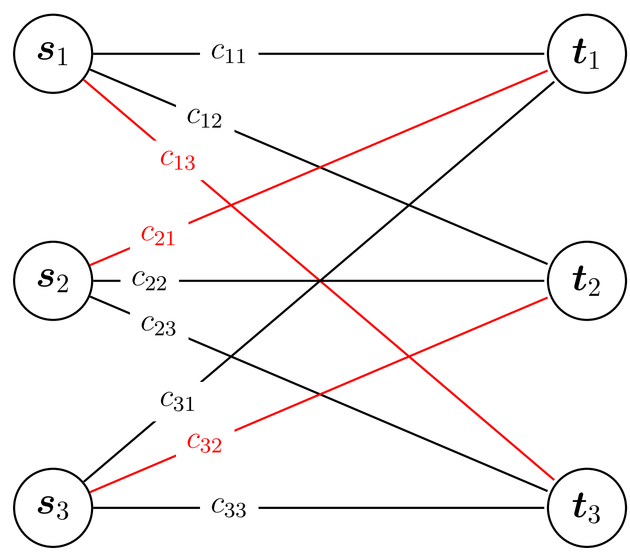

Figure 1. Bipartite graph formulation of a linear sum assignment problem (LSAP).

Table 1. Empirical sizes of the proposed test using theoretical (noted as Hallin(t)) and simulation-based (noted as Hallin(s)) rejection threshold, test via distance covariance with marginal ranks (noted as rdCov), and test via distance covariance (noted as dCov) in Example 6.1(a).

(p, q)	n	Hallin(t)	Hallin(s)	rdCov	dCov
(2, 2)	216	0.057	0.046	0.043	0.045
(2, 2)	432	0.054	0.051	0.048	0.050
(2, 2)	864	0.050	0.057	0.050	0.048
(2, 2)	1728	0.045	0.046	0.061	0.057
(3, 3)	216	0.052	0.056	0.058	0.053
(3, 3)	432	0.055	0.045	0.045	0.043
(3, 3)	864	0.052	0.046	0.053	0.048
(3, 3)	1728	0.056	0.050	0.043	0.050
(5, 5)	216	0.060	0.055	0.040	0.048
(5, 5)	432	0.047	0.046	0.048	0.043
(5, 5)	864	0.051	0.048	0.040	0.048
(5, 5)	1728	0.048	0.044	0.053	0.039
(7, 7)	216	0.053	0.048	0.053	0.056
(7, 7)	432	0.051	0.038	0.054	0.053
(7, 7)	864	0.047	0.055	0.048	0.046
(7, 7)	1728	0.046	0.047	0.048	0.052

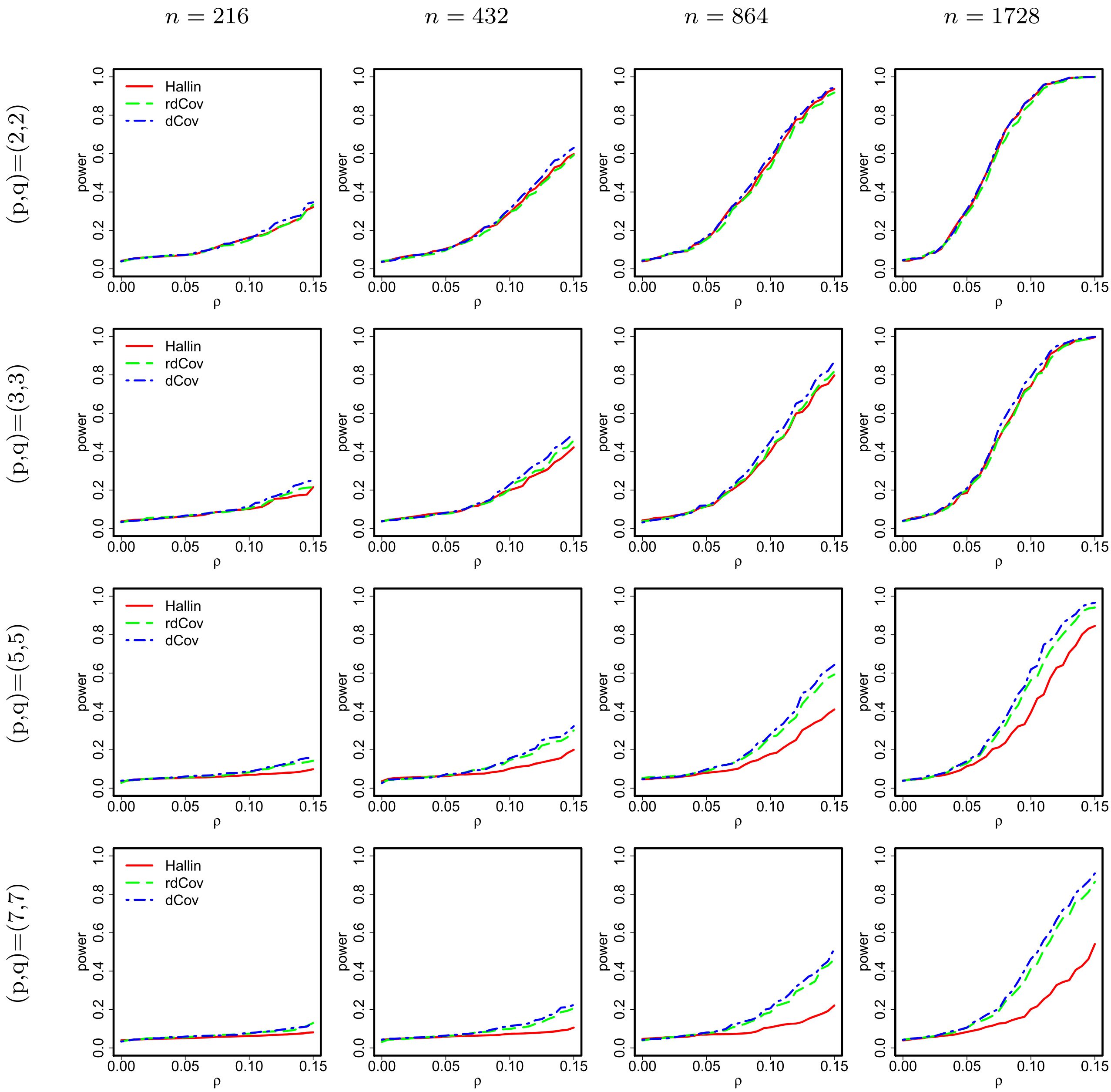

Figure 2. Empirical powers of the three competing tests in Example 6.1(a). The *y*-axis represents the power based on 1000 replicates and the *x*-axis represents the level of a desired signal.

1998, Lemma 2.2) and the equicontinuity of $-d_{\mathbf{U}^{(n)}}(\boldsymbol{u}, \boldsymbol{u}')$ (Rudin 1976, Exercise 7.16). The similar results hold for $-d_{\mathbf{V}^{(n)}}(\boldsymbol{v}, \boldsymbol{v}')$ and $-d_{\mathbf{V}}(\boldsymbol{v}, \boldsymbol{v}')$. Lastly, under H_0, $[\mathbf{F}^{(n)}_{\mathbf{X},\pm}(\boldsymbol{X}_i)]_{i=1}^n$ and $[\mathbf{F}^{(n)}_{\mathbf{Y},\pm}(\boldsymbol{Y}_i)]_{i=1}^n$ are independent with margins uniformly distributed on $\mathscr{P}(\mathcal{G}^p_{n_0,n_R,n_S})$ and $\mathscr{P}(\mathcal{G}^q_{n_0,n_R,n_S})$, respectively. Hence our statistic is distributed of the form (18).

In summary, Theorem 4.2 can be applied to the statistic $\widehat{M}_n$ and we have accordingly proven Theorem 3.1 rigorously. Furthermore, although our focus is on the combination of center-outward ranks and signs with the distance covariance statistic, the general form of our combinatorial non-CLTs (Theorems 4.1 and 4.2) also yields the limiting null distributions for test statistics based on plugging center-outward ranks and signs into HSIC-type or ball-covariance statistics (Gretton, Bousquet, et al. 2005; Gretton, Herbrich, et al. 2005; Gretton, Smola, et al. 2005; Pan et al. 2020). We omit the details for these analogies.

5. Computational Aspects

In this section, we describe the practical implementation of our test. To perform the proposed test, for any given n, we fix a factorization such that

$$n = n_R n_S + n_0, \quad n_R, n_S \in \mathbb{Z}_+,\ 0 \le n_0 < \min\{n_R, n_S\},$$
$$\text{with } n_R, n_S \to \infty \quad \text{as } n \to \infty.$$

First, we need to compute $[\mathbf{F}^{(n)}_{\mathbf{X},\pm}(\boldsymbol{X}_i)]_{i=1}^n$ and $[\mathbf{F}^{(n)}_{\mathbf{Y},\pm}(\boldsymbol{Y}_i)]_{i=1}^n$ as defined in (4). This is an assignment problem and will be discussed in Section 5.1. After obtaining $[\mathbf{F}^{(n)}_{\mathbf{X},\pm}(\boldsymbol{X}_i)]_{i=1}^n$ and $[\mathbf{F}^{(n)}_{\mathbf{Y},\pm}(\boldsymbol{Y}_i)]_{i=1}^n$, the test statistic $\widehat{M}_n$ in (10) can be computed using Equation (3.3) in Huo and Székely (2016) in $O(n^2)$ time. Second, we have to calculate the critical value $Q_{1-\alpha}$ defined in (13). This value can be estimated numerically, as detailed in Section 5.2. We have also provided the critical values at significance

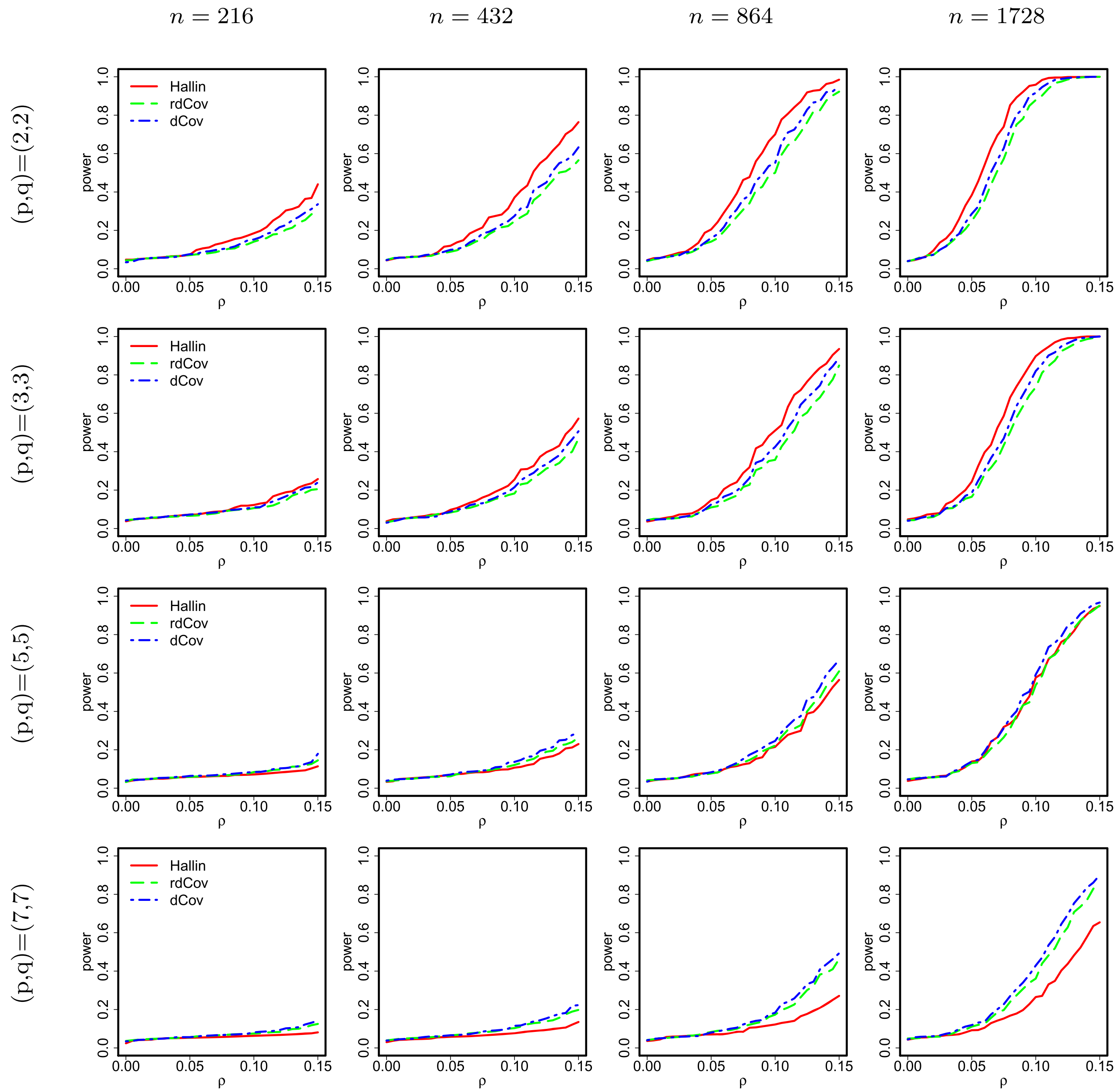

Figure 3. Empirical powers of the three competing tests in Example 6.1(b). The y-axis represents the power based on 1000 replicates and the x-axis represents the level of a desired signal.

levels $\alpha = 0.1, 0.05, 0.01$ for $(p, q) = (1, 1), (1, 2), \dots, (10, 10)$ with accuracy $2 \cdot 10^{-3}$ in Table C.1 in the supplementary materials.

As shall be shown soon, the total computation complexity of our proposed test is $O(n^{5/2} \log(n))$ in various cases. To contrast, to implement the distance covariance based test for instance, one has a time complexity $O(Rn^2)$, with R representing the number of permutations. For many choices of R, our test will have a clear computational advantage.

5.1. Assignment Problems

Problem (4) amounts to a linear sum assignment problem (LSAP), a fundamental problem in linear programming and combinatorial optimization. We define LSAP through graph theory. Consider a weighted (complete) bipartite graph $(S, T; E)$ with $S := \{\boldsymbol{s}_i\}_{i=1}^n$, $T := \{\boldsymbol{t}_j\}_{j=1}^n$, $\boldsymbol{s}_i, \boldsymbol{t}_j \in \mathbb{R}^d$, where in Problem (4), $S = \{\boldsymbol{x}_i\}_{i=1}^n$ and $T = \mathcal{G}^d_{n_0,n_R,n_S}$. The edge between $\boldsymbol{s}_i$ and $\boldsymbol{t}_j$, denoted by $(\boldsymbol{s}_i, \boldsymbol{t}_j)$, has a nonnegative weight $c_{ij} := \|\boldsymbol{s}_i - \boldsymbol{t}_j\|^2$, $i, j \in [\![n]\!]$. We want to find an *optimal matching*, that is, a subset of edges such that each vertex is an endpoint of exactly one edge in this subset with a minimum sum of weights of its edges; see Figure 1 for an illustration of $n = 3$, where edges in the optimal matching are marked in red.

We introduce some terms to state the theorem below. A *perfect matching* is a subset of edges such that each vertex is incident to exactly one edge. The *total weight* of a perfect matching is the sum of weights of the edges in this matching. A perfect matching is called $(1 + \epsilon)$-approximate for $\epsilon > 0$ if its total weight is no larger than $(1 + \epsilon)$ times the total weight of the optimal matching.

Theorem 5.1 (Gabow and Tarjan 1989; Sharathkumar and Agarwal 2012; Agarwal and Sharathkumar 2014). Assume that

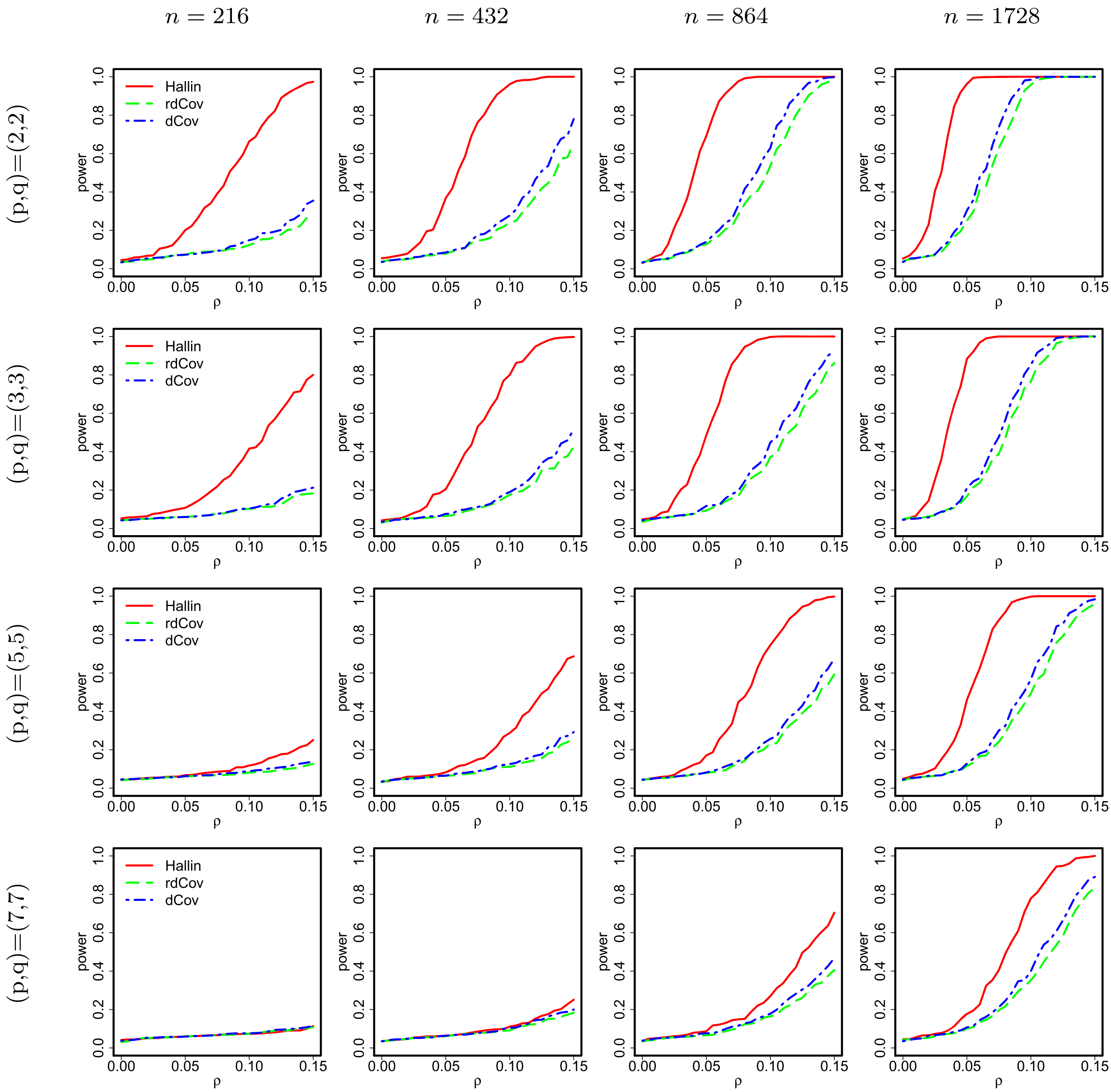

Figure 4. Empirical powers of the three competing tests in Example 6.1(c). The *y*-axis represents the power based on 1000 replicates and the *x*-axis represents the level of a desired signal.

points $\boldsymbol{s}_i, \boldsymbol{t}_j \in \mathbb{R}^d$, $i, j \in [\![n]\!]$, have bounded integer coordinates, and that the squared distances $\|\boldsymbol{s}_i - \boldsymbol{t}_j\|^2$, $i, j \in [\![n]\!]$ are all bounded by some integer N. Then there exists an algorithm to find the optimal matching in $O(n^{5/2}\log(nN))$ time. Furthermore,

(i) if $d = 2$, there exists an exact algorithm for computing the optimal matching in $O(n^{3/2+\delta}\log(N))$ time for any arbitrarily small constant $\delta > 0$;
(ii) if $d \geq 3$, there is an algorithm to compute a $(1+\epsilon)$-approximate perfect matching in $O(\epsilon^{-1}n^{3/2}\tau(n,\epsilon)\log^4(n/\epsilon)\log(\max c_{ij}/\min c_{ij}))$ time, where $\tau(n,\epsilon)$ depending on n, ϵ is small.

In the supplementary materials, we will describe the algorithm developed by Gabow and Tarjan (1989) under the basic settings. It is essentially the combination of the Hungarian method (Kuhn 1955, 1956; Munkres 1957) and the algorithm of Hopcroft and Karp (1973). We will ignore the details of the faster exact algorithm for $d = 2$ by Sharathkumar and Agarwal (2012) and the approximate algorithm for $d \geq 3$ by Agarwal and Sharathkumar (2014); both algorithms improve the Gabow–Tarjan algorithm by exploiting the geometric structure of the weight matrix.

5.2. Eigenvalues and Quadratic Forms in Normal Variables

In Theorem 3.1, λ_k, $k \in \mathbb{Z}_+$, are nonzero eigenvalues (counted with multiplicity) of the integral equation

$$E(d_U(\boldsymbol{u}, \boldsymbol{U})d_V(\boldsymbol{v}, \boldsymbol{V})\phi(\boldsymbol{U}, \boldsymbol{V})) = \lambda\phi(\boldsymbol{u}, \boldsymbol{v}).$$

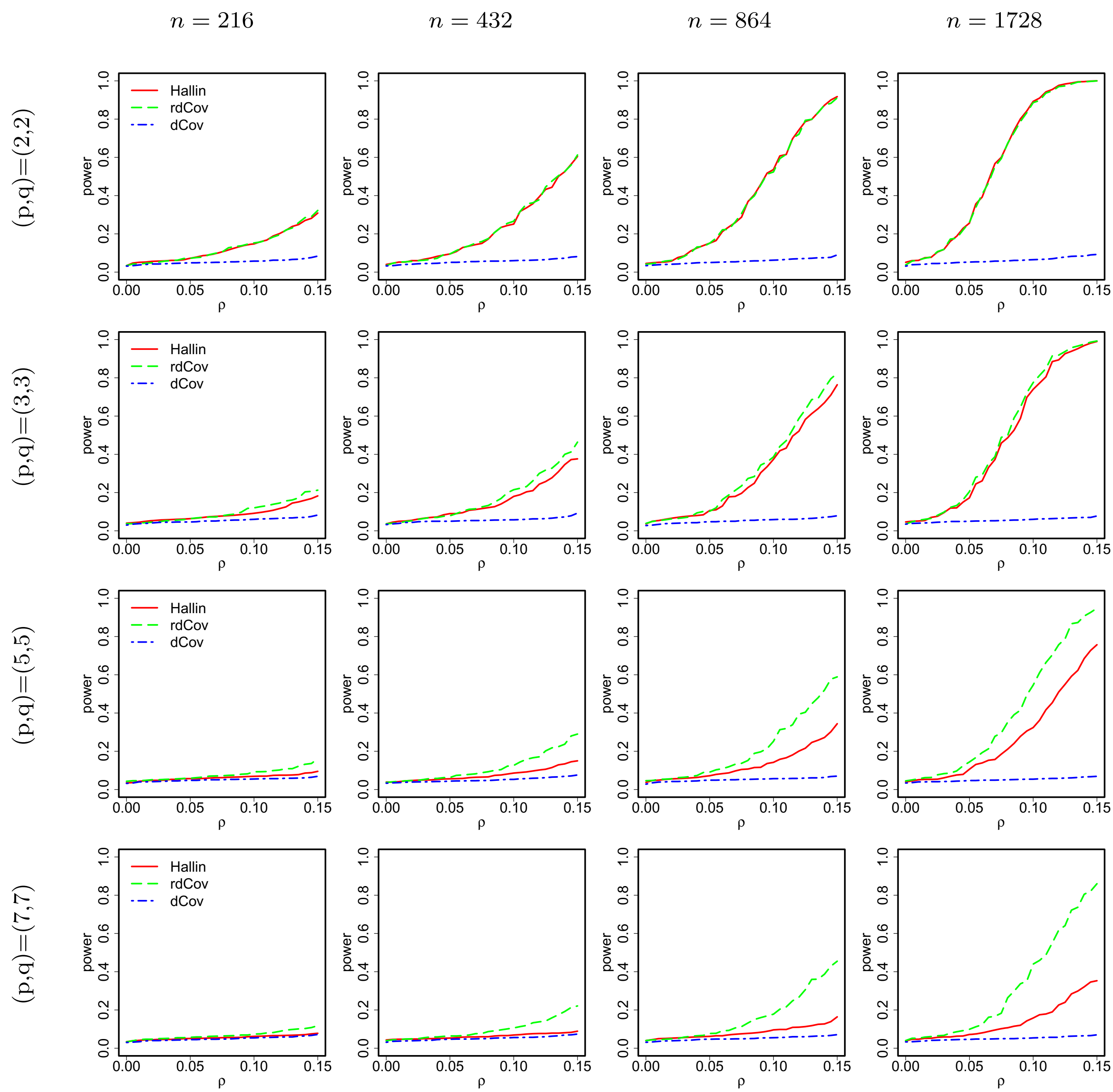

Figure 5. Empirical powers of the three competing tests in Example 6.2(a). The *y*-axis represents the power based on 1000 replicates and the *x*-axis represents the level of a desired signal.

Under the independence hypothesis H_0, the eigenvalues λ_k, $k \in \mathbb{Z}_+$, are given by all the products $\lambda_{1,j_1}\lambda_{2,j_2}$, $j_1, j_2 \in \mathbb{Z}_+$, where $\lambda_{1,j}$, $j \in \mathbb{Z}_+$, and $\lambda_{2,j}$, $j \in \mathbb{Z}_+$, are the nonzero eigenvalues of the integral equations

$$E(d_{\boldsymbol{U}}(\boldsymbol{u}, \boldsymbol{U})\phi_1(\boldsymbol{U})) = \lambda_1\phi_1(\boldsymbol{u}) \quad \text{and}$$
$$E(d_{\boldsymbol{V}}(\boldsymbol{v}, \boldsymbol{V})\phi_2(\boldsymbol{V})) = \lambda_2\phi_2(\boldsymbol{v}),$$

respectively (Nandy, Weihs, and Drton 2016, Lemma 4.2). The nonzero eigenvalues of integral equation $E(d_{\boldsymbol{U}}(\boldsymbol{u}, \boldsymbol{U})\phi_1(\boldsymbol{U})) = \lambda_1\phi_1(\boldsymbol{u})$ with $\boldsymbol{U} \sim U_p$ are given by

$$-4/(\pi^2 j^2), \quad \text{for all } j \in \mathbb{Z}_+ \text{ when } p = 1.$$

We are not aware of any closed form formulas for the eigenvalues when $p \geq 2$. However, in practice, the nonzero eigenvalues $\{\lambda_{1,j}\}_{j=1}^{\infty}$ can be numerically estimated by the nonzero eigenvalues of the matrix

$$(\mathbf{I}_M - \mathbf{J}_M/M)\mathbf{D}^{(M)}(\mathbf{I}_M - \mathbf{J}_M/M)/M,$$

denoted by $\lambda_{1,j}^{(M)}$, $j \in [\![M-1]\!]$, where $M := M_R M_S$, $\mathbf{D}^{(M)} = [D_{jj'}^{(M)}]$, $D_{jj'}^{(M)} = \|\boldsymbol{u}_j^{(M)} - \boldsymbol{u}_{j'}^{(M)}\|$ and $\boldsymbol{u}_j^{(M)}$, $j \in [\![M]\!]$, are points in the grid $\mathcal{G}^p_{0,M_R,M_S}$. Here $\lambda_{1,j}^{(M)}$, $j \in [\![M-1]\!]$ are all negative (Lyons 2013, p. 3291). For $p = 1$, we take $\lambda_{1,j}^{(M)} = -4/(\pi^2 j^2)$. We can obtain eigenvalues $\lambda_{2,j}^{(M)}$, $j \in [\![M-1]\!]$ based on the grid $\mathcal{G}^q_{0,M_R,M_S}$ similarly. Then we sort the positive products $\lambda_{1,j_1}^{(M)}\lambda_{2,j_2}^{(M)}$, $j_1, j_2 \in [\![M-1]\!]$ into a descendingly ordered sequence $[\lambda_k^{(M)}]_{k=1}^{(M-1)^2}$, and have the following theorem.

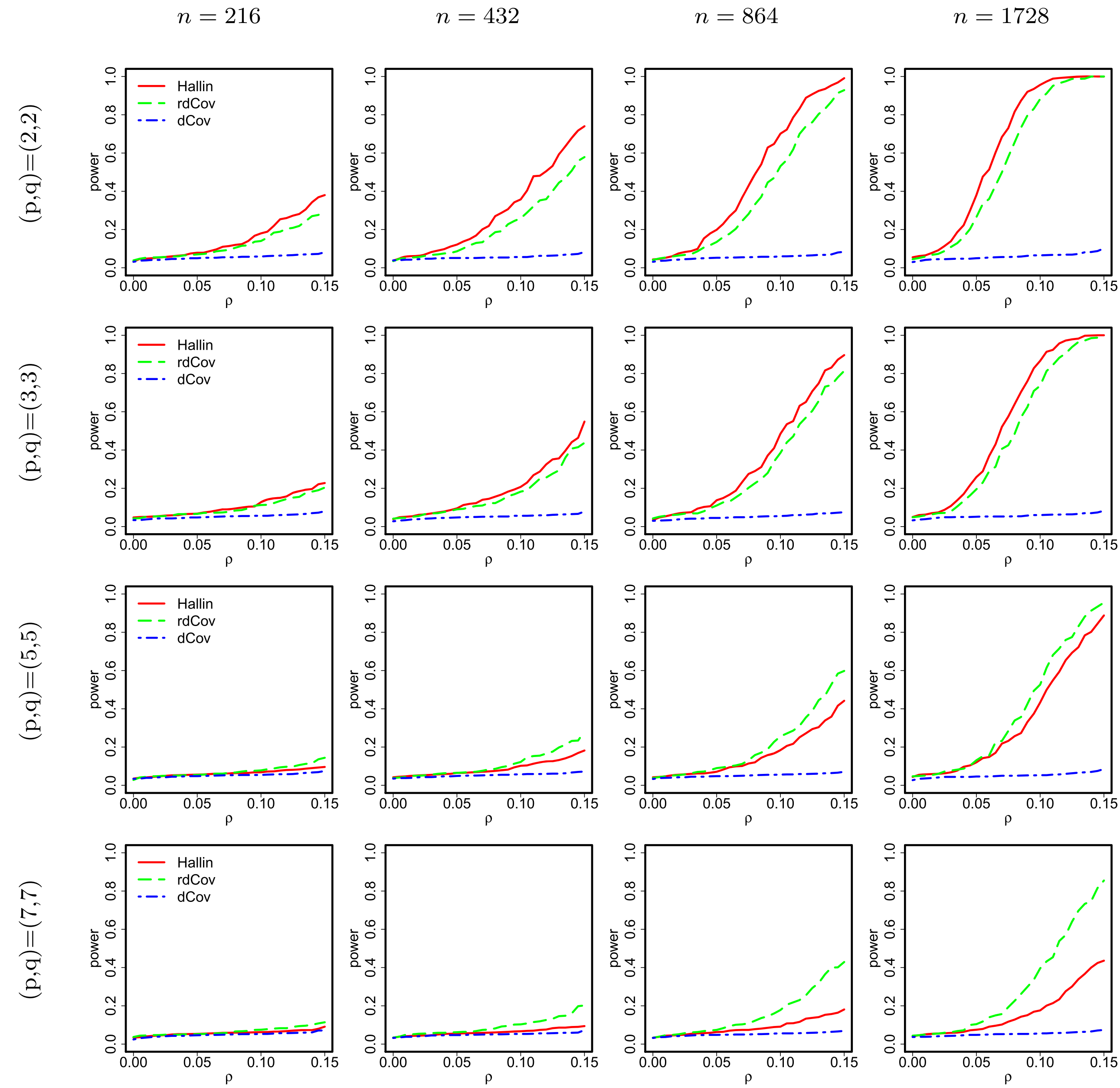

Figure 6. Empirical powers of the three competing tests in Example 6.2(b). The *y*-axis represents the power based on 1000 replicates and the *x*-axis represents the level of a desired signal.

Theorem 5.2. Let $[\lambda_k]_{k=1}^{\infty}$ and $[\lambda_k^{(M)}]_{k=1}^{(M-1)^2}$ be eigenvalues as defined in Theorem 3.1 and above, respectively. Let $[\xi_k]_{k=1}^{\infty}$ be a sequence of independent standard Gaussian random variables. Then it holds for any prespecified significance level $\alpha \in (0, 1)$ that

$$Q_{1-\alpha}^{(M)} \to Q_{1-\alpha}$$

as $M_R \to \infty$ and $M_S \to \infty$, where $Q_{1-\alpha}^{(M)}$ and $Q_{1-\alpha}$ are the $(1-\alpha)$ quantiles of

$$\sum_{k=1}^{(M-1)^2} \lambda_k^{(M)}(\xi_k^2 - 1) \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda_k(\xi_k^2 - 1),$$

respectively.

Consequently, we can approximate the $(1-\alpha)$ quantile of quadratic form $\sum_{k=1}^{\infty} \lambda_k(\xi_k^2 - 1)$ by estimating that of quadratic form $\sum_{k=1}^{(M-1)^2} \lambda_k^{(M)}(\xi_k^2 - 1)$ for a sufficiently large M. The latter is done by solving the inverse of the cumulative distribution function of quadratic form $\sum_{k=1}^{(M-1)^2} \lambda_k^{(M)}(\xi_k^2 - 1)$, which can be numerically evaluated using Farebrother's (1984) algorithm or Imhof's (1961) method.

6. Numerical Studies

This section compares the performances of our tests using (i) the theoretical rejection threshold $Q_{1-\alpha}$ defined in (13) and computed using the approximation in Section 5.2, and (ii) a Monte Carlo simulation-based rejection threshold to the existing tests of independence that use (iii) distance covariance with

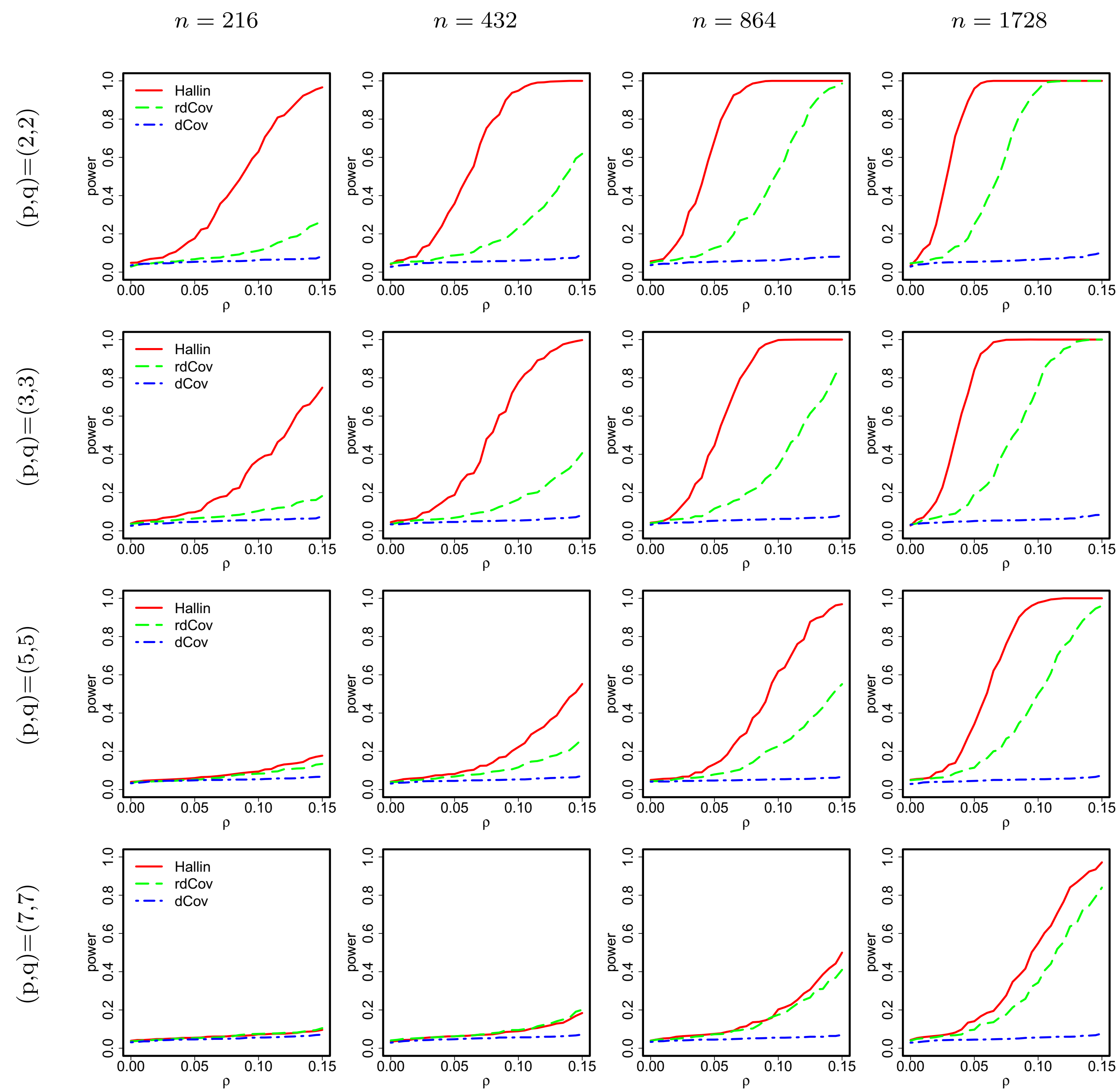

Figure 7. Empirical powers of the three competing tests in Example 6.2(c). The y-axis represents the power based on 1000 replicates and the x-axis represents the level of a desired signal.

marginal ranks (Lin 2017), and (iv) distance covariance (Székely and Rizzo 2013).

The test via distance covariance with marginal ranks proceeds as follows. Write $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,p})$ for $i \in [\![n]\!]$. Let $r_{i,k}$ be the rank of $x_{i,k}$ among $x_{1,k}, x_{2,k}, \dots, x_{n,k}$ for each $k \in [\![p]\!]$. The marginal rank (vector) of $\boldsymbol{x}_i$ is defined as $(r_{i,1}, \dots, r_{i,p})$. The marginal rank (vector) of $\boldsymbol{y}_i$ is defined similarly. Then we run the permutation-based distance covariance test on the marginal ranks instead of the original data.

6.1. Simulation Results

We first conduct Monte Carlo simulation experiments on the finite-sample performance of the proposed test from Section 3. We evaluate the empirical sizes and powers of the four competing tests stated above for both Gaussian and non-Gaussian distributions. The values reported below are based on 1000 simulations at the nominal significance level of 0.05, with sample size $n \in \{216, 432, 864, 1728\}$, dimensions $p = q \in \{2, 3, 5, 7\}$, and correlation $\rho \in \{0, 0.005, 0.01, \dots, 0.15\}$. More simulation studies on even higher dimensions of $p = q = 10$ and 30 are presented in the supplementary materials, Section C. For tests (iii) and (iv), we resample n times in the permutation procedure.

Example 6.1. The data are independently drawn from $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{R}^{p+q}$, which follows a multivariate normal distribution with mean zero and covariance matrix $\mathbf{I}_{p+q} + \tau \mathbf{L}_{p+q;1,2} + \rho \mathbf{L}_{p+q;1,p+1}$ (where $\mathbf{L}_{d;i,j} := \boldsymbol{e}_{d;i}\boldsymbol{e}_{d;j}^\top + \boldsymbol{e}_{d;j}\boldsymbol{e}_{d;i}^\top$ and $\boldsymbol{e}_{d;i} \in \mathbb{R}^d$ is the ith standard basis vector in d-dimensional space, that is, all entries are zero except for the one at the ith position) with (a) $\tau = 0$; (b) $\tau = 0.5$; and (c) $\tau = 0.9$.

Table 2. *p*-values based on the proposed test as well as two competing tests for the dataset of US stock closing prices between 2003 and 2012.

		(DD, DOW)	(DD, FCX)	(DD, MON)	(DOW, FCX)	(DOW, MON)	(FCX, MON)
Hallin	(T, VZ)	0.001	0.005	0.002	0.004	0.001	0.065
rdCov	(T, VZ)	0.002	0.013	0.005	0.009	0.002	0.070
dCov	(T, VZ)	0.002	0.018	0.003	0.012	0.002	0.101

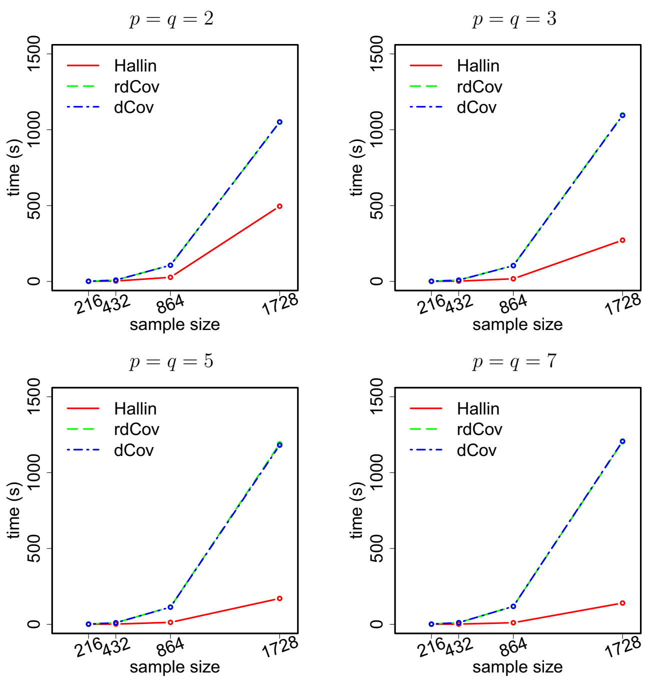

Figure 8. A comparison of computation time in Example 6.1(a) for the three tests. The *y*-axis represents the averaged computation elapsed time (in seconds) of 1000 replicates of a single experiment and the *x*-axis represents the sample size. To compute the optimal matching, we used the algorithm in Gabow and Tarjan (1989).

Example 6.2. The data are independently drawn from $(\boldsymbol{X}, \boldsymbol{Y})$, which is given by $X_i = Q_{t(1)}(\Phi(X_i^*))$, $i \in [\![p]\!]$ and $Y_j = Q_{t(1)}(\Phi(Y_j^*))$, $j \in [\![q]\!]$, where $Q_{t(1)}$ stands for the quantile function for Student's t-distribution with 1 degree of freedom (Cauchy distribution), and $(\boldsymbol{X}^*, \boldsymbol{Y}^*)$ are generated as in Example 6.1.

In these two examples, the independence hypothesis holds when $\rho = 0$. We first report the empirical sizes of all four considered tests, presented in Table 1. It can be observed that the proposed tests with either rejection threshold as well as their two competitors control the size effectively.

The empirical powers for Examples 6.1 and 6.2 are summarized in Figures 2–7. For the proposed test, we present results only for the theoretical rejection threshold as the results for the simulation-based threshold are similar and hence omitted.

Several facts are noteworthy. First, when the sample size is large and the dimension is relatively small, throughout all settings the performance of the proposed test is not much worse than the two competing ones. It should be highlighted that our method achieves this performance with smaller computational time, as shown in Figure 8 and also confirmed in our theoretical analysis of computational cost. Second, the proposed test beats the other two when the within-group correlation is high, i.e., as τ becomes larger from the setting (a) to (c), even when the dimension is high. Third, for heavy-tailed distributions, the tests via distance covariance with center-outward ranks and signs and marginal ranks perform better than the original distance covariance test. Finally, compared to its competitors, the proposed test appears to be more sensitive to dimension. This is as expected.

6.2. *Real Stock Market Data Analysis*

We analyze the monthly log returns of daily closing prices for stocks that are constantly in the Standard & Poor 100 (S&P 100) index during the time period 2003–2012. The data are from Yahoo! Finance (*finance.yahoo.com*), and the stocks are classified into 10 sectors by Global Industry Classification Standard (GICS). Stock market data tend to be heavy-tailed with many outliers, and monthly log returns may reasonably be modeled as independent and identically distributed random variables. The time period we analyzed includes some well-known turbulent stretches like the 2007–2008 financial crisis, which, however, could be either explained using heavy-tailed (e.g., elliptical or stable) distribution models or captured as outliers.

In this section we limit our scope and focus on detecting between-group dependence between two sectors in S&P 100 that contain a rather small number of stocks: (1) Telecommunication, including stocks "AT&T Inc. [T]" and "Verizon Communications [VZ]"; and (2) Materials, including stocks "Du Pont (E.I.) [DD]," "Dow Chemical [DOW]," "Freeport-McMoran Cp & Gld [FCX]," and "Monsanto Co. [MON]." We then consider detection of possible dependence between the Telecommunication sector and any two stocks in the Materials sector.

To this end, we apply the three considered tests to the monthly log returns of (T, VZ) coupled with either (DD, DOW), or (DD, FCX), or (DD, MON), or (DOW, FCX), or (DOW, MON), or (FCX, MON). The p-values for these three tests are reported in Table 2. There, one observes that using the proposed test yields uniformly the strongest evidence to conclude the existence of dependence between (T, VZ) and any two stocks in the Materials sector.

Supplementary Materials

The supplementary materials include a pdf file containing all the technical proofs and additional numerical results and R scripts used to perform the numerical studies.

Acknowledgments

We thank the co-editor Hongyu Zhao, the anonymous associate editor, and two anonymous referees for their very detailed and constructive comments and suggestions, which have helped greatly to improve the quality of the article.

Funding

This work was partially supported by the U.S. National Science Foundation under grant nos. DMS 1712535 and DMS 1712536.

References

Abe, O. (1969), "A Central Limit Theorem for the Number of Edges in the Random Intersection of Two Graphs," *The Annals of Mathematical Statistics*, 40, 144–151. [6]

Agarwal, P. K., and Sharathkumar, R. (2014), "Approximation Algorithms for Bipartite Matching With Metric and Geometric Costs," in *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing*, ACM, New York, NY, pp. 555–564. [9,10]

Bakirov, N. K., Rizzo, M. L., and Székely, G. J. (2006), "A Multivariate Nonparametric Test of Independence," *Journal of Multivariate Analysis*, 97, 1742–1756. [2]

Barbour, A. D., and Chen, L. H. Y. (2005), "The Permutation Distribution of Matrix Correlation Statistics," in *Stein's Method and Applications*, Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore (Vol. 5), Singapore: Singapore University Press, pp. 223–245. [6]

Barbour, A. D., and Eagleson, G. K. (1986), "Random Association of Symmetric Arrays," *Stochastic Analysis and Applications*, 4, 239–281. [6]

Basu, D. (1959), "The Family of Ancillary Statistics," *Sankhyā*, 21, 247–256. [2]

Bergsma, W. (2006), "A New Correlation Coefficient, Its Orthogonal Decomposition and Associated Tests of Independence," arXiv no. math/0604627v1. [6]

Bergsma, W., and Dassios, A. (2014), "A Consistent Test of Independence Based on a Sign Covariance Related to Kendall's Tau," *Bernoulli*, 20, 1006–1028. [1,2,4]

Berrett, T. B., and Samworth, R. J. (2019), "Nonparametric Independence Testing via Mutual Information," *Biometrika*, 106, 547–566. [1]

Bloemena, A. R. (1964), "Sampling From a Graph," in *Mathematical Centre Tracts* (No. 2), ed. W. R. van Zwet, Amsterdam: Mathematisch Centrum. [6]

Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961), "Distribution Free Tests of Independence Based on the Sample Distribution Function," *The Annals of Mathematical Statistics*, 32, 485–498. [1,2,6]

Bolthausen, E. (1984), "An Estimate of the Remainder in a Combinatorial Central Limit Theorem," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 66, 379–386. [6]

Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017), "Monge–Kantorovich Depth, Quantiles, Ranks and Signs," *The Annals of Statistics*, 45, 223–256. [2,3]

Cliff, A. D., and Ord, J. K. (1973), *Spatial Autocorrelation*, Monographs in Spatial and Environmental Systems Analysis (Vol. 5), London: Pion. [6]

Daniels, H. E. (1944), "The Relation Between Measures of Correlation in the Universe of Sample Permutations," *Biometrika*, 33, 129–135. [6]

Deb, N., and Sen, B. (2019), "Multivariate Rank-Based Distribution-Free Nonparametric Testing Using Measure Transportation," arXiv no. 1909.08733. [2]

del Barrio, E., Cuesta-Albertos, J. A., Hallin, M., and Matrán, C. (2018), "Smooth Cyclically Monotone Interpolation and Empirical Center-Outward Distribution Functions," arXiv no. 1806.01238v1. [4]

Drton, M., Han, F., and Shi, H. (2020), "High Dimensional Consistent Independence Testing With Maxima of Rank Correlations," *The Annals of Statistics* (in press). [2]

Escanciano, J. C. (2006), "A Consistent Diagnostic Test for Regression Models Using Projections," *Econometric Theory*, 22, 1030–1051. [2]

Fang, K. T., Kotz, S., and Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions*, Monographs on Statistics and Applied Probability (Vol. 36), London: Chapman and Hall, Ltd. [2]

Farebrother, R. W. (1984), "Algorithm AS 204: The Distribution of a Positive Linear Combination of χ^2 Random Variables," *Journal of the Royal Statistical Society*, Series C, 33, 332–339. [12]

Feuerverger, A. (1993), "A Consistent Test for Bivariate Dependence," *International Statistical Review*, 61, 419–433. [1]

Figalli, A. (2018), "On the Continuity of Center-Outward Distribution and Quantile Functions," *Nonlinear Analysis*, 177, 413–421. [2,3]

Gabow, H. N., and Tarjan, R. E. (1989), "Faster Scaling Algorithms for Network Problems," *SIAM Journal on Computing*, 18, 1013–1036. [9,10,14]

Gebelein, H. (1941), "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung," *Zeitschrift für Angewandte Mathematik und Mechanik*, 21, 364–379. [1]

Ghosal, P., and Sen, B. (2019), "Multivariate Ranks and Quantiles Using Optimal Transportation and Applications to Goodness-of-Fit Testing," arXiv no. 1905.05340. [2]

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005), "Measuring Statistical Dependence With Hilbert–Schmidt Norms," in *Algorithmic Learning Theory*, Lecture Notes in Computer Science (Vol. 3734), eds. S. Jain, H. U. Simon, and E. Tomita, Berlin: Springer-Verlag Berlin Heidelberg, pp. 63–77. [1,8]

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* (Vol. 20), eds. J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, Curran Associates, Inc., Red Hook, NY, pp. 984–991. [1,2]

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005), "Kernel Methods for Measuring Independence," *Journal of Machine Learning Research*, 6, 2075–2129. [1,8]

Gretton, A., Smola, A., Bousquet, O., Herbrich, R., Belitski, A., Augath, M., Murayama, Y., Pauls, J., Schölkopf, B., and Logothetis, N. (2005), "Kernel Constrained Covariance for Dependence Measurement," in *AISTATS05*, eds. R. G. Cowell and Z. Ghahramani, Society for Artificial Intelligence and Statistics, pp. 112–119. [1,8]

Hájek, J. (1961), "Some Extensions of the Wald–Wolfowitz–Noether Theorem," *The Annals of Mathematical Statistics*, 32, 506–523. [6]

Hallin, M. (2017), "On Distribution and Quantile Functions, Ranks and Signs in $\mathbb{R}^d$: A Measure Transportation Approach," available at *https://ideas.repec.org/p/eca/wpaper/2013-258262.html*. [2,3,4]

Hallin, M., del Barrio, E., Cuesta-Albertos, J. A., and Matrán, C. (2020), "Distribution and Quantile Functions, Ranks, and Signs in $\mathbb{R}^d$: A Measure Transportation Approach," *The Annals of Statistics* (in press). [2,4]

Hallin, M., and Paindaveine, D. (2002a), "Multivariate Signed Ranks: Randles' Interdirections or Tyler's Angles?," in *Statistical Data Analysis Based on the L_1-Norm and Related Methods (Neuchâtel, 2002)*, Statistics for Industry and Technology, ed. Y. Dodge, Basel: Birkhäuser, pp. 271–282. [2]

——— (2002b), "Optimal Tests for Multivariate Location Based on Interdirections and Pseudo-Mahalanobis Ranks," *The Annals of Statistics*, 30, 1103–1133. [2]

Heller, R., Gorfine, M., and Heller, Y. (2012), "A Class of Multivariate Distribution-Free Tests of Independence Based on Graphs," *Journal of Statistical Planning and Inference*, 142, 3097–3106. [2]

Heller, R., and Heller, Y. (2016), "Multivariate Tests of Association Based on Univariate Tests," in *Advances in Neural Information Processing Systems* (Vol. 29), eds. D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Curran Associates, Inc., pp. 208–216. [2]

Heller, R., Heller, Y., and Gorfine, M. (2013), "A Consistent Multivariate Test of Association Based on Ranks of Distances," *Biometrika*, 100, 503–510. [1,2]

Heller, R., Heller, Y., Kaufman, S., Brill, B., and Gorfine, M. (2016), "Consistent Distribution-Free K-Sample and Independence Tests for Univariate Random Variables," *Journal of Machine Learning Research*, 17, 1–54. [1]

Hirschfeld, H. O. (1935), "A Connection Between Correlation and Contingency," *Mathematical Proceedings of the Cambridge Philosophical Society*, 31, 520–524. [1]

Hoeffding, W. (1940), "Maszstabinvariante Korrelationstheorie," *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5, 181–233. [1]

——— (1948), "A Non-Parametric Test of Independence," *The Annals of Mathematical Statistics*, 19, 546–557. [1,2]

——— (1951), "A Combinatorial Central Limit Theorem," *The Annals of Mathematical Statistics*, 22, 558–566. [6]

Hopcroft, J. E., and Karp, R. M. (1973), "An $n^{5/2}$ Algorithm for Maximum Matchings in Bipartite Graphs," *SIAM Journal on Computing*, 2, 225–231. [10]

Huo, X., and Székely, G. J. (2016), "Fast Computing for Distance Covariance," *Technometrics*, 58, 435–447. [8]

Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426. [12]

Jakobsen, M. E. (2017), "Distance Covariance in Metric Spaces: Non-Parametric Independence Testing in Metric Spaces," Master's thesis, University of Copenhagen, arXiv no. 1706.03490. [1,5]

Jogdeo, K. (1968), "Asymptotic Normality in Nonparametric Methods," *The Annals of Mathematical Statistics*, 39, 905–922. [6]

Kim, I., Balakrishnan, S., and Wasserman, L. (2020), "Robust Multivariate Nonparametric Tests via Projection-Averaging," *The Annals of Statistics* (in press). [2]

Kinney, J. B., and Atwal, G. S. (2014), "Equitability, Mutual Information, and the Maximal Information Coefficient," *Proceedings of the National Academy of Sciences of the United States of America*, 111, 3354–3359. [1]

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004), "Estimating Mutual Information," *Physical Review E*, 69, 066138. [1]

Kuhn, H. W. (1955), "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, 2, 83–97. [10]

——— (1956), "Variants of the Hungarian Method for Assignment Problems," *Naval Research Logistics Quarterly*, 3, 253–258. [10]

Lin, J. (2017), "Copula Versions of RKHS-Based and Distance-Based Criteria," Ph.D. thesis, Pennsylvania State University, available at *https://etda.libraries.psu.edu/catalog/14485jul268*. [2,13]

Lyons, R. (2013), "Distance Covariance in Metric Spaces," *The Annals of Probability*, 41, 3284–3305. [1,4,5,7,11]

Ma, L., and Mao, J. (2019), "Fisher Exact Scanning for Dependency," *Journal of the American Statistical Association*, 114, 245–258. [1]

McCann, R. J. (1995), "Existence and Uniqueness of Monotone Measure-Preserving Maps," *Duke Mathematical Journal*, 80, 309–323. [3]

Motoo, M. (1957), "On the Hoeffding's Combinatorial Central Limit Theorem," *Annals of the Institute of Statistical Mathematics*, 8, 145–154. [6]

Munkres, J. (1957), "Algorithms for the Assignment and Transportation Problems," *Journal of the Society for Industrial and Applied Mathematics*, 5, 32–38. [10]

Nandy, P., Weihs, L., and Drton, M. (2016), "Large-Sample Theory for the Bergsma–Dassios Sign Covariance," *Electronic Journal of Statistics*, 10, 2287–2311. [11]

Noether, G. E. (1949), "On a Theorem of Wald and Wolfowitz," *The Annals of Mathematical Statistics*, 20, 455–458. [6]

Pan, W., Wang, X., Zhang, H., Zhu, H., and Zhu, J. (2020), "Ball Covariance: A Generic Measure of Dependence in Banach Space," *Journal of the American Statistical Association*, 115, 307–317. [4,8]

Pham, D. T., Möcks, J., and Sroka, L. (1989), "Asymptotic Normality of Double-Indexed Linear Permutation Statistics," *Annals of the Institute of Statistical Mathematics*, 41, 415–427. [6]

Raič, M. (2015), "Multivariate Normal Approximation: Permutation Statistics, Local Dependence and Beyond," available at *https://ims.nus.edu.sg/oldwww/Programs/015wstein/files/martin.pdf*. [7]

Reinert, G., and Röllin, A. (2009), "Multivariate Normal Approximation With Stein's Method of Exchangeable Pairs Under a General Linearity Condition," *The Annals of Probability*, 37, 2150–2173. [6]

Rémillard, B. (2009), "Discussion of: Brownian Distance Covariance," *The Annals of Applied Statistics*, 3, 1295–1298. [2]

Rényi, A. (1959a), "New Version of the Probabilistic Generalization of the Large Sieve," *Acta Mathematica Academiae Scientiarum Hungarica*, 10, 217–226. [1]

——— (1959b), "On Measures of Dependence," *Acta Mathematica Academiae Scientiarum Hungarica*, 10, 441–451. [1]

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C. (2011), "Detecting Novel Associations in Large Data Sets," *Science*, 334, 1518–1524. [1]

Reshef, D. N., Reshef, Y. A., Sabeti, P. C., and Mitzenmacher, M. (2018), "An Empirical Study of the Maximal and Total Information Coefficients and Leading Measures of Dependence," *The Annals of Applied Statistics*, 12, 123–155. [1]

Reshef, Y. A., Reshef, D. N., Finucane, H. K., Sabeti, P. C., and Mitzenmacher, M. (2016), "Measuring Dependence Powerfully and Equitably," *Journal of Machine Learning Research*, 17, 1–63. [1]

Rudin, W. (1976), *Principles of Mathematical Analysis*, International Series in Pure and Applied Mathematics (3rd ed.), New York: McGraw-Hill Book Co. [8]

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013), "Equivalence of Distance-Based and RKHS-Based Statistics in Hypothesis Testing," *The Annals of Statistics*, 41, 2263–2291. [1]

Shapiro, C. P., and Hubert, L. (1979), "Asymptotic Normality of Permutation Statistics Derived From Weighted Sums of Bivariate Functions," *The Annals of Statistics*, 7, 788–794. [6]

Sharathkumar, R., and Agarwal, P. K. (2012), "Algorithms for the Transportation Problem in Geometric Settings," in *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM, New York, NY, pp. 306–317. [9,10]

Székely, G. J., and Bakirov, N. K. (2003), "Extremal Probabilities for Gaussian Quadratic Forms," *Probability Theory and Related Fields*, 126, 184–202. [2]

Székely, G. J., and Rizzo, M. L. (2009), "Brownian Distance Covariance," *The Annals of Applied Statistics*, 3, 1236–1265. [2]

——— (2013), "The Distance Correlation *t*-Test of Independence in High Dimension," *Journal of Multivariate Analysis*, 117, 193–213. [4,5,13]

——— (2014), "Partial Distance Correlation With Methods for Dissimilarities," *The Annals of Statistics*, 42, 2382–2412. [5]

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances," *The Annals of Statistics*, 35, 2769–2794. [1,2,4]

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics (Vol. 3), Cambridge, UK: Cambridge University Press. [8]

Villani, C. (2009), *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] (Vol. 338), Berlin: Springer-Verlag Berlin Heidelberg. [3]

Wald, A., and Wolfowitz, J. (1944), "Statistical Tests Based on Permutations of the Observations," *The Annals of Mathematical Statistics*, 15, 358–372. [6]

Weihs, L., Drton, M., and Meinshausen, N. (2018), "Symmetric Rank Covariances: A Generalized Framework for Nonparametric Measures of Dependence," *Biometrika*, 105, 547–562. [2]

Yanagimoto, T. (1970), "On Measures of Association and a Related Problem," *Annals of the Institute of Statistical Mathematics*, 22, 57–63. [1,2]

Yao, S., Zhang, X., and Shao, X. (2018a), "Supplementary Material for 'Testing mutual Independence in High Dimension via Distance Covariance'," *Journal of the Royal Statistical Society*, Series B, 80, 455–480. [7]

——— (2018b), "Testing Mutual Independence in High Dimension via Distance Covariance," *Journal of the Royal Statistical Society*, Series B, 80, 455–480. [5]

Zhang, K. (2019), "BET on Independence," *Journal of the American Statistical Association*, 114, 1620–1637. [1]

Zhao, L., Bai, Z., Chao, C.-C., and Liang, W.-Q. (1997), "Error Bound in a Central Limit Theorem of Double-Indexed Permutation Statistics," *The Annals of Statistics*, 25, 2210–2227. [6]

Zhu, L., Xu, K., Li, R., and Zhong, W. (2017), "Projection Correlation Between Two Random Vectors," *Biometrika*, 104, 829–843. [2]