The Annals of Statistics
2022, Vol. 50, No. 4, 1933–1959
https://doi.org/10.1214/21-AOS2151

ON UNIVERSALLY CONSISTENT AND FULLY DISTRIBUTION-FREE RANK TESTS OF VECTOR INDEPENDENCE

BY HONGJIAN SHI[1,a], MARC HALLIN[2,c], MATHIAS DRTON[1,b] AND FANG HAN[3,d]

[1]*Department of Mathematics, Technical University of Munich,* [a]*hongjian.shi@tum.de,* [b]*mathias.drton@tum.de*
[2]*ECARES and Department of Mathematics, Université Libre de Bruxelles,* [c]*mhallin@ulb.ac.be*
[3]*Department of Statistics, University of Washington,* [d]*fanghan@uw.edu*

Rank correlations have found many innovative applications in the last decade. In particular, suitable rank correlations have been used for consistent tests of independence between pairs of random variables. Using ranks is especially appealing for continuous data as tests become distribution-free. However, the traditional concept of ranks relies on ordering data and is, thus, tied to univariate observations. As a result, it has long remained unclear how one may construct distribution-free yet consistent tests of independence between random vectors. This is the problem addressed in this paper, in which we lay out a general framework for designing dependence measures that give tests of multivariate independence that are not only consistent and distribution-free but which we also prove to be statistically efficient. Our framework leverages the recently introduced concept of center-outward ranks and signs, a multivariate generalization of traditional ranks, and adopts a common standard form for dependence measures that encompasses many popular examples. In a unified study, we derive a general asymptotic representation of center-outward rank-based test statistics under independence, extending to the multivariate setting the classical Hájek asymptotic representation results. This representation permits direct calculation of limiting null distributions and facilitates a local power analysis that provides strong support for the center-outward approach by establishing, for the first time, the nontrivial power of center-outward rank-based tests over root-n neighborhoods within the class of quadratic mean differentiable alternatives.

1. Introduction.

Quantifying the dependence between two variables and testing for their independence are among the oldest and most fundamental problems of statistical inference. The (marginal) distributions of the two variables under study, in that context, typically play the role of nuisances, and the need for a nonparametric approach naturally leads, when they are univariate, to distribution-free methods based on their ranks. This paper is dealing with the multivariate extension of that approach.

1.1. *Measuring vector dependence and testing independence.*

Consider two absolutely continuous random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, with values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively. The problems of measuring the dependence between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ and testing their independence when $d_1 = d_2 = 1$ (call this the univariate case) have a long history that goes back more than a century (Pearson (1895), Spearman (1904)). The same problem when d_1 and d_2 are possibly unequal and larger than one (the multivariate case) is of equal practical interest but considerably more challenging. Following early attempts (Wilks (1935)), a large literature has emerged, with renewed interest in recent years.

Received May 2021; revised October 2021.
MSC2020 subject classifications. 62G10.
Key words and phrases. Multivariate ranks and signs, center-outward ranks and signs, multivariate dependence measure, independence test, Hájek representation, Le Cam's third lemma.

When the marginal distributions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are unspecified and $d_1 = d_2 = 1$, rank correlations provide a natural and appealing nonparametric approach to testing for independence, as initiated in the work of Spearman (1904) and Kendall (1938); cf. Chapter III.6 in Hájek and Šidák (1967). On one hand, ranks yield distribution-free tests because, under the null hypothesis of independence, their distributions do not depend on the unspecified marginal distributions (Han, Chen and Liu (2017), Drton, Han and Shi (2020)). On the other hand, they can be designed (Hoeffding (1948), Blum, Kiefer and Rosenblatt (1961), Bergsma and Dassios (2014), Yanagimoto (1970)) to consistently estimate dependence measures that vanish if and only if independence holds, and so detect any type of dependence—something Spearman and Kendall's rank correlations cannot.

New subtleties arise, however, when attempting to extend the rank-based approach to the multivariate case. While d_k ranks can be constructed separately for each coordinate of $\boldsymbol{X}_k$, $k = 1, 2$, their joint distribution depends on the distribution of the underlying $\boldsymbol{X}_k$, preventing distribution-freeness of the $(d_1 + d_2)$-tuple of ranks. As a consequence, the existing tests of multivariate independence based on componentwise ranks (e.g., Puri, Sen and Gokhale (1970)) are not distribution-free, which has both computational implications (e.g., through a need for permutation analysis) and statistical implications (as we shall detail soon).

1.2. *Desirable properties.* In this paper, we develop a general framework for multivariate analogues of popular rank-based measures of dependence for the univariate case. Our objective is to achieve the following five desirable properties.

(1) *Full distribution-freeness*. Many statistical tests exploit asymptotic distribution-freeness for computationally efficient distributional approximations yielding pointwise asymptotic control of their size. This is the case, for instance, with Hallin and Paindaveine (2002a, 2002b, 2002c, 2008) due to estimation of a scatter matrix, or with Taskinen, Kankainen and Oja (2003, 2004), Taskinen, Oja and Randles (2005). Pointwise asymptotics yield, for any given significance level $\alpha \in (0, 1)$, a sequence of tests $\phi_\alpha^{(n)}$ indexed by the sample size n such that $\lim_{n\to\infty} \mathrm{E}_\mathrm{P}[\phi_\alpha^{(n)}] = \alpha$ for every distribution P from a class $\mathcal{P}$ of null distributions. Generally, however, the size fails to be controlled in a uniform sense, that is, it does not hold that $\lim_{n\to\infty} \sup_{\mathrm{P}\in\mathcal{P}} \mathrm{E}_\mathrm{P}[\phi_\alpha^{(n)}] \le \alpha$, which may explain poor finite-sample properties (see, e.g., Le Cam and Yang (2000), Leeb and Pötscher (2008), Belloni, Chernozhukov and Hansen (2014)). While uniform inferential validity is impossible to achieve for some problems, for example, when testing for conditional independence (Shah and Peters (2020), Azadkia and Chatterjee (2021)), we shall see that it is achievable for testing (unconditional) multivariate independence. Indeed, for fully distribution-free tests, as obtained from our rank-based approach, pointwise validity automatically implies uniform validity.

(2) *Transformation invariance*. A dependence measure μ is said to be invariant under orthogonal transformations, shifts, and global rescaling if

$$\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu(\boldsymbol{v}_1 + a_1\mathbf{O}_1\boldsymbol{X}_1, \boldsymbol{v}_2 + a_2\mathbf{O}_1\boldsymbol{X}_2)$$

for any scalars $a_k > 0$, vectors $\boldsymbol{v}_k \in \mathbb{R}^{d_k}$, and orthogonal $d_k \times d_k$ matrices $\mathbf{O}_k$, $k = 1, 2$. This invariance, here simply termed "transformation invariance", is a natural requirement in cases where the components of $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ do not have specific meanings and observations could have been recorded in another coordinate system. Such invariance is of considerable interest in multivariate statistics (see, e.g., Gieser and Randles (1997), Taskinen, Kankainen and Oja (2003), Taskinen, Oja and Randles (2005), Oja, Paindaveine and Taskinen (2016)).

(3) *Consistency*. Weihs, Drton and Meinshausen (2018) call a dependence measure μ *I-consistent* within a family of distributions $\mathcal{P}$ if independence between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with joint distribution in $\mathcal{P}$ implies $\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$. If $\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$ implies independence of $\boldsymbol{X}_1$

and $\boldsymbol{X}_2$ (i.e., dependence of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ implies $\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) \neq 0$), then μ is *D-consistent* within $\mathcal{P}$. Note that the measures considered in this paper do not necessarily take maximal value 1 if and only if one random vector is a measurable function of the other. While any reasonable dependence measure should be I-consistent, prominent examples (Pearson's correlation, Spearman's ρ, Kendall's τ) fail to be D-consistent. If a dependence measure μ is I- and D-consistent, then the consistency of tests based on an estimator $\mu^{(n)}$ of μ is guaranteed by the (strong or weak) consistency of that estimator. Dependence measures that are both I- and D-consistent (within a large nonparametric family) serve an important purpose as they are able to capture nonlinear dependences. Well-known I- and D-consistent measures for the univariate case include Hoeffding's D (Hoeffding (1948)), Blum–Kiefer–Rosenblatt's R (Blum, Kiefer and Rosenblatt (1961)), and Bergsma–Dassios–Yanagimoto's τ^* (Bergsma and Dassios (2014), Yanagimoto (1970), Drton, Han and Shi (2020)). Multivariate extensions have been proposed, for example, in Gretton et al. (2005a), Székely, Rizzo and Bakirov (2007), Heller, Gorfine and Heller (2012), Heller, Heller and Gorfine (2013), Heller and Heller (2016), Zhu et al. (2017), Weihs, Drton and Meinshausen (2018), Kim, Balakrishnan and Wasserman (2020a), Deb and Sen (2022), Shi, Drton and Han (2022a), Berrett, Kontoyiannis and Samworth (2021).

(4) *Statistical efficiency*. Once its size is controlled, the performance of a test may be evaluated through its power against local alternatives. For the proposed tests, our focus is on quadratic mean differentiable alternatives (Lehmann and Romano (2005), Section 12.2), which form a popular class for conducting local power analyses; for related recent examples see Bhattacharya ((2019), Section 3) and Cao and Bickel ((2020), Section 4.4). Our results then show the nontrivial local power of our tests in $n^{-1/2}$ neighborhoods within this class.

(5) *Computational efficiency*. Statistical properties aside, modern applications require the evaluation of a dependence measure and the corresponding test to be as computationally efficient as possible. We thus prioritize measures leading to low computational complexity.

The main challenge, with this list of five properties, lies in combining the full distribution-freeness from property (1) with properties (2)–(5). The solution, as we shall see, involves an adequate multivariate extension of the univariate concepts of ranks and signs.

1.3. *Contribution of this paper.* This paper proposes a class of dependence measures and tests that achieve the five properties from Section 1.2 by leveraging the recently introduced multivariate center-outward ranks and signs (Chernozhukov et al. (2017), Hallin (2017)); see Hallin et al. (2021) for a complete account. In contrast to earlier related concepts such as componentwise ranks (Puri and Sen (1971)), spatial ranks (Oja (2010), Han and Liu (2018)), depth-based ranks (Liu and Singh (1993), Zuo and He (2006)), and pseudo-Mahalanobis ranks and signs (Hallin and Paindaveine (2002a)), the new concept yields statistics that enjoy full distribution-freeness (in finite samples and, thus, asymptotically) as soon as the underlying probability measure is Lebesgue-absolutely continuous. This allows for a general multivariate strategy, in which the observations are replaced by functions of their center-outward ranks and signs when forming dependence measures and corresponding test statistics. This is also the idea put forth in Shi, Drton and Han (2022a) and, in a slightly different way, in Deb and Sen (2022), where the focus is on distance covariance between center-outward ranks and signs.

Methodologically, we are generalizing this approach in two important ways. First, we introduce a class of *generalized symmetric covariances* (GSCs) along with their center-outward rank versions, of which the distance covariance concepts from Deb and Sen (2022) and Shi, Drton and Han (2022a) are but particular cases. Second, we show how considerable additional flexibility and power results from incorporating score functions in the definition. Our simulations in Section 5.4 exemplify the benefits of this "score-based" approach.

From a theoretical point of view, we offer a new approach to asymptotic theory for the proposed rank-based statistics. Indeed, handling this general class with the methods of Shi, Drton and Han (2022a) or Deb and Sen (2022) would be highly nontrivial. Moreover, these methods would not provide any insights into local power—an issue receiving much attention also in other contexts (Hallin, La Vecchia and Liu (2022), Beirlant et al. (2020), Hallin, Mordant and Segers (2021), Hallin, Hlubinka and Hudecová (2022)). We thus develop a completely different method, based on a general asymptotic representation result applicable to all center-outward rank-based GSCs under the null hypothesis of independence and contiguous alternatives of dependence. Our result (Theorem 5.1; see also Proposition 3.1 in Shi et al. (2021)) is a multivariate extension of Hájek's classical asymptotic representation for univariate linear rank statistics (Hájek and Šidák (1967)) and also simplifies the derivation of limiting null distributions. Combined with a nontrivial use of Le Cam's third lemma in a context of non-Gaussian limits, our approach allows for the first local power results in the area; the statistical efficiency of the tests of Deb and Sen (2022) and Shi, Drton and Han (2022a) follows as a special case. In Proposition 4.2, we establish the strong consistency of our rank-based tests against any fixed alternative under a regularity condition on the score function. Thanks to a recent result by Deb, Bhattacharya and Sen (2021), that assumption can be relaxed: our tests, thus, enjoy *universal consistency* against fixed dependence alternatives.

Outline of the paper. The paper begins with a review of important dependence measures from the literature (Section 2). Generalizing the idea of symmetric rank covariances put forth in Weihs, Drton and Meinshausen (2018), we show that a single formula unifies them all; we term the concept *generalized symmetric covariance* (GSC). As further background, Section 3 introduces the notion of center-outward ranks and signs. Section 4 presents our streamlined approach of defining multivariate dependence measures, along with sample counterparts, and highlights some of their basic properties. Section 5 treats tests of independence and develops a theory of asymptotic representation for center-outward rank-based GSCs (Section 5.1) as well as the local power analysis of the corresponding tests against classes of quadratic mean differentiable alternatives (Section 5.2). Specific alternatives are exemplified in Section 5.3, and benefits of choosing standard score functions (such as normal scores) are illustrated in the numerical study in Section 5.4. All proofs are deferred to the Supplementary Material (Shi et al. (2022)).

Notation. For integer $m \geq 1$, put $[\![m]\!] := \{1, 2, \dots, m\}$, and let $\mathfrak{S}_m$ be the symmetric group, that is, the group of all permutations of $[\![m]\!]$. We write $\mathrm{sgn}(\sigma)$ for the sign of $\sigma \in \mathfrak{S}_m$. In the sequel, the subgroup

$$H_*^m := \langle (1\ 4), (2\ 3) \rangle = \{(1), (1\ 4), (2\ 3), (1\ 4)(2\ 3)\} \subset \mathfrak{S}_m \tag{1.1}$$

will play an important role. Here, we have made use of the cycle notation (omitting 1-cycles) so that, for example, (1) denotes the identity permutation and

$$(1\ 4) \equiv \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \cdots & m \\ 4 & 2 & 3 & 1 & 5 & 6 & \cdots & m \end{pmatrix},$$

$$(1\ 4)(2\ 3) \equiv \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \cdots & m \\ 4 & 3 & 2 & 1 & 5 & 6 & \cdots & m \end{pmatrix},$$

where the right-hand sides are in classical two-line notation listing $\sigma(i)$ below i, $i \in [\![m]\!]$.

A set with distinct elements $x_1, \dots, x_n$ is written either as $\{x_1, \dots, x_n\}$ or $\{x_i\}_{i=1}^n$. The corresponding sequence is denoted by $[x_1, \dots, x_n]$ or $[x_i]_{i=1}^n$. An arrangement of $\{x_i\}_{i=1}^n$ is a sequence $[x_{\sigma(i)}]_{i=1}^n$, where $\sigma \in \mathfrak{S}_n$. An r-arrangement is a sequence $[x_{\sigma(i)}]_{i=1}^r$ for $r \in [\![n]\!]$. Write I_r^n for the family of all $(n)_r := n!/(n-r)!$ possible r-arrangements of $[\![n]\!]$.

The set of nonnegative reals is denoted $\mathbb{R}_{\geq 0}$, and $\mathbf{0}_d$ stands for the origin in $\mathbb{R}^d$. For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$, we write $\boldsymbol{u} \preceq \boldsymbol{v}$ if $u_\ell \leq v_\ell$ for all $\ell \in [\![d]\!]$, and $\boldsymbol{u} \npreceq \boldsymbol{v}$ otherwise. Let $\mathsf{Arc}(\boldsymbol{u}, \boldsymbol{v}) := (2\pi)^{-1} \arccos\{\boldsymbol{u}^\top \boldsymbol{v}/(\|\boldsymbol{u}\|\|\boldsymbol{v}\|)\}$ if $\boldsymbol{u}, \boldsymbol{v} \neq \mathbf{0}_d$; $\mathsf{Arc}(\boldsymbol{u}, \boldsymbol{v}) := 0$ otherwise. Here, $\|\cdot\|$ stands for the Euclidean norm. For vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$, we use $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ as a shorthand for $(\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_k^\top)^\top$. We write $\mathbf{I}_d$ for the $d \times d$ identity matrix. For a function $f : \mathcal{X} \to \mathbb{R}$, we define $\|f\|_\infty := \max_{x \in \mathcal{X}} |f(x)|$. The symbols $\lfloor \cdot \rfloor$ and $\mathbb{1}(\cdot)$ stand for the floor and indicator functions.

The cumulative distribution function and the probability distribution of a real-valued random variable/vector $\boldsymbol{Z}$ are denoted as $F_{\boldsymbol{Z}}(\cdot)$ and $\mathrm{P}_{\boldsymbol{Z}}$, respectively. The class of probability measures on $\mathbb{R}^d$ that are absolutely continuous (with respect to the Lebesgue measure) is denoted as $\mathcal{P}_d^{\mathrm{ac}}$. We use $\rightsquigarrow$ and $\xrightarrow{\text{a.s.}}$ to denote convergence in distribution and almost sure convergence, respectively. For any symmetric kernel $h(\cdot)$ on $(\mathbb{R}^d)^m$, any integer $\ell \in [\![m]\!]$, and any probability measure $\mathrm{P}_{\boldsymbol{Z}}$, we write $h_\ell(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell; \mathrm{P}_{\boldsymbol{Z}})$ for $\mathrm{E}h(\boldsymbol{z}_1 \ldots, \boldsymbol{z}_\ell, \boldsymbol{Z}_{\ell+1}, \ldots, \boldsymbol{Z}_m)$ where $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ are m independent copies of $\boldsymbol{Z} \sim \mathrm{P}_{\boldsymbol{Z}}$, and $\mathrm{E}h := \mathrm{E}h(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m)$. The product measure of two distributions P_1 and P_2 is denoted $\mathrm{P}_1 \otimes \mathrm{P}_2$.

2. Generalized symmetric covariances. Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be two random vectors with values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively, and assume throughout this paper that they are both absolutely continuous with respect to the Lebesgue measure. Weihs, Drton and Meinshausen ((2018), Definition 3) introduced a general approach to defining rank-based measures of dependence via signed sums of indicator functions that are acted upon by subgroups of the symmetric group. In this section, we highlight that their resulting family of *symmetric rank covariances* can be extended to cover a much wider range of dependence measures including, in particular, the celebrated *distance covariance* (Székely, Rizzo and Bakirov (2007)). This enables us to handle a broad family of dependence measures in the following common standard form.

DEFINITION 2.1 (Generalized symmetric covariance). A measure of dependence μ is said to be an mth order *generalized symmetric covariance (GSC)* if there exist two kernel functions $f_1 : (\mathbb{R}^{d_1})^m \to \mathbb{R}_{\geq 0}$ and $f_2 : (\mathbb{R}^{d_2})^m \to \mathbb{R}_{\geq 0}$, and a subgroup $H \subseteq \mathfrak{S}_m$ containing an equal number of even and odd permutations such that

$$\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu_{f_1, f_2, H}(\boldsymbol{X}_1, \boldsymbol{X}_2) := \mathrm{E}\big[k_{f_1, f_2, H}\big((\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), \ldots, (\boldsymbol{X}_{1m}, \boldsymbol{X}_{2m})\big)\big].$$

Here $(\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), \ldots, (\boldsymbol{X}_{1m}, \boldsymbol{X}_{2m})$ are m independent copies of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$, and the dependence kernel function $k_{f_1, f_2, H}(\cdot)$ is defined as

$$\begin{aligned} & k_{f_1, f_2, H}\big((\boldsymbol{x}_{11}, \boldsymbol{x}_{21}), \ldots, (\boldsymbol{x}_{1m}, \boldsymbol{x}_{2m})\big) \\ & := \Big\{\sum_{\sigma \in H} \operatorname{sgn}(\sigma) f_1(\boldsymbol{x}_{1\sigma(1)}, \ldots, \boldsymbol{x}_{1\sigma(m)})\Big\}\Big\{\sum_{\sigma \in H} \operatorname{sgn}(\sigma) f_2(\boldsymbol{x}_{2\sigma(1)}, \ldots, \boldsymbol{x}_{2\sigma(m)})\Big\}. \end{aligned} \tag{2.1}$$

As the group H is required to have equal numbers of even and odd permutations, the order of a GSC satisfies $m \geq 2$. This requirement also justifies the term "generalized covariance" through the following property; compare Weihs, Drton and Meinshausen ((2018), Proposition 2).

PROPOSITION 2.1. *All GSCs are I-consistent. More precisely, the GSC* $\mu_{f_1, f_2, H}(\boldsymbol{X}_1, \boldsymbol{X}_2)$ *is I-consistent in the family of distributions such that* $\mathrm{E}[f_k] := \mathrm{E}[f_k(\boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{km})] < \infty$, $k = 1, 2$, *where* $\boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{km}$ *are* m *independent copies of* $\boldsymbol{X}_k$.

The concept of GSC unifies a surprisingly large number of well-known dependence measures. We consider here five noteworthy examples, namely, the distance covariance of Székely, Rizzo and Bakirov (2007) and Székely and Rizzo (2013), the multivariate version of Hoeffding's D based on marginal ordering (Weihs, Drton and Meinshausen (2018), Section 2.2, page 549), and the projection-averaging extensions of Hoeffding's D (Zhu et al. (2017)), of Blum–Kiefer–Rosenblatt's R (Kim, Balakrishnan and Wasserman (2020b), Proposition D.5), and of Bergsma–Dassios–Yanagimoto's τ^* (Kim, Balakrishnan and Wasserman (2020a), Theorem. 7.2). Only one type of subgroup, namely, $H_*^m := \langle (1\ 4), (2\ 3)\rangle \subseteq \mathfrak{S}_m$ for $m \geq 4$ is needed; recall (1.1). For simplicity, we write $\boldsymbol{w} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_m) \mapsto f_k(\boldsymbol{w})$ for the kernel functions of an mth order multivariate GSC for which the dimension of $\boldsymbol{w}_\ell$, $\ell = 1, \dots, m$, is d_k, hence may differ for $k = 1$ and $k = 2$. Not all components of $\boldsymbol{w}$ need to have an impact on $f_k(\boldsymbol{w})$. For instance, the kernels of distance covariance, a 4th order GSC, map $\boldsymbol{w} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_4)$ to $\mathbb{R}_{\geq 0}$ but depend neither on $\boldsymbol{w}_3$ nor $\boldsymbol{w}_4$.

EXAMPLE 2.1 (Examples of multivariate GSCs).

(a) Distance covariance is a 4th order GSC with $H = H_*^4$ and

$$f_k^{\mathrm{dCov}}(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}_1 - \boldsymbol{w}_2\| \quad \text{on } (\mathbb{R}^{d_k})^4,\ k = 1, 2.$$

Indeed, with $c_d := \pi^{(1+d)/2}/\Gamma((1+d)/2)$, we have

$$\begin{aligned}
&\mu_{f_1^{\mathrm{dCov}}, f_2^{\mathrm{dCov}}, H_*^4}(\boldsymbol{X}_1, \boldsymbol{X}_2) \\
&\quad = \frac{1}{4}\mathrm{E}\big[\big(\|\boldsymbol{X}_{11} - \boldsymbol{X}_{12}\| - \|\boldsymbol{X}_{11} - \boldsymbol{X}_{13}\| - \|\boldsymbol{X}_{14} - \boldsymbol{X}_{12}\| + \|\boldsymbol{X}_{14} - \boldsymbol{X}_{13}\|\big) \\
&\qquad \times \big(\|\boldsymbol{X}_{21} - \boldsymbol{X}_{22}\| - \|\boldsymbol{X}_{21} - \boldsymbol{X}_{23}\| - \|\boldsymbol{X}_{24} - \boldsymbol{X}_{22}\| + \|\boldsymbol{X}_{24} - \boldsymbol{X}_{23}\|\big)\big] \\
&\quad = \frac{1}{c_{d_1} c_{d_2}} \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} \frac{|\varphi_{(\boldsymbol{X}_1, \boldsymbol{X}_2)}(\boldsymbol{t}_1, \boldsymbol{t}_2) - \varphi_{\boldsymbol{X}_1}(\boldsymbol{t}_1)\varphi_{\boldsymbol{X}_2}(\boldsymbol{t}_2)|^2}{\|\boldsymbol{t}_1\|^{d_1+1}\|\boldsymbol{t}_2\|^{d_2+1}} \,\mathrm{d}\boldsymbol{t}_1 \,\mathrm{d}\boldsymbol{t}_2.
\end{aligned} \tag{2.2}$$

Identity (2.2) was established in Székely, Rizzo and Bakirov ((2007), Remark 3), Székely and Rizzo ((2009), Theorem 8), and Bergsma and Dassios ((2014), Section 3.4);

(b) Hoeffding's multivariate marginal ordering D is a 5th order GSC with $H = H_*^5$ and

$$f_k^M(\boldsymbol{w}) = \frac{1}{2}\mathbb{1}(\boldsymbol{w}_1, \boldsymbol{w}_2 \preceq \boldsymbol{w}_5) \quad \text{on } (\mathbb{R}^{d_k})^5,\ k = 1, 2,$$

since, by Weihs, Drton and Meinshausen ((2018), Proposition 1),

$$\mu_{f_1^M, f_2^M, H_*^5}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} \{F_{(\boldsymbol{X}_1, \boldsymbol{X}_2)}(\boldsymbol{u}_1, \boldsymbol{u}_2) - F_{\boldsymbol{X}_1}(\boldsymbol{u}_1)F_{\boldsymbol{X}_2}(\boldsymbol{u}_2)\}^2 \,\mathrm{d}F_{(\boldsymbol{X}_1, \boldsymbol{X}_2)}(\boldsymbol{u}_1, \boldsymbol{u}_2);$$

(c) Hoeffding's multivariate projection-averaging D is a 5th order GSC with $H = H_*^5$ and

$$f_k^D(\boldsymbol{w}) = \frac{1}{2}\mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_5, \boldsymbol{w}_2 - \boldsymbol{w}_5) \quad \text{on } (\mathbb{R}^{d_k})^5,\ k = 1, 2.$$

Indeed, by Zhu et al. ((2017), Equation (3)), we have

$$\begin{aligned}
\mu_{f_1^D, f_2^D, H_*^5}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \int_{\mathcal{S}_{d_1-1} \times \mathcal{S}_{d_2-1}} \int_{\mathbb{R}^2} \{&F_{(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_2)}(u_1, u_2) \\
&- F_{\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1}(u_1)F_{\boldsymbol{\alpha}_2^\top \boldsymbol{X}_2}(u_2)\}^2 \,\mathrm{d}F_{(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_2)}(u_1, u_2) \,\mathrm{d}\lambda_{d_1}(\boldsymbol{\alpha}_1) \,\mathrm{d}\lambda_{d_2}(\boldsymbol{\alpha}_2),
\end{aligned}$$

with λ_d the uniform measure on the unit sphere $\mathcal{S}_{d-1}$;

(d) Blum–Kiefer–Rosenblatt's multivariate projection-averaging R is a 6th order GSC with $H = H_*^6$ and

$$f_1^R(\boldsymbol{w}) = \frac{1}{2}\mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_5, \boldsymbol{w}_2 - \boldsymbol{w}_5) \quad \text{on } (\mathbb{R}^{d_1})^6,$$

$$f_2^R(\boldsymbol{w}) = \frac{1}{2}\mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_6, \boldsymbol{w}_2 - \boldsymbol{w}_6) \quad \text{on } (\mathbb{R}^{d_2})^6;$$

this follows from Kim, Balakrishnan and Wasserman ((2020b), Proposition D.5), who showed

$$\begin{aligned}\mu_{f_1^R, f_2^R, H_*^6}&(\boldsymbol{X}_1, \boldsymbol{X}_2)\\ &= \int_{\mathcal{S}_{d_1-1}\times\mathcal{S}_{d_2-1}} \int_{\mathbb{R}^2} \{F_{(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_2)}(u_1, u_2)\\ &\quad - F_{\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1}(u_1) F_{\boldsymbol{\alpha}_2^\top \boldsymbol{X}_2}(u_2)\}^2 \,\mathrm{d}F_{\boldsymbol{\alpha}_1^\top \boldsymbol{X}_1}(u_1)\,\mathrm{d}F_{\boldsymbol{\alpha}_2^\top \boldsymbol{X}_2}(u_2)\,\mathrm{d}\lambda_{d_1}(\boldsymbol{\alpha}_1)\,\mathrm{d}\lambda_{d_2}(\boldsymbol{\alpha}_2);\end{aligned}$$

(e) Bergsma–Dassios–Yanagimoto's multivariate projection-averaging τ^* is a 4th order GSC with $H = H_*^4$ and

$$f_k^{\tau^*}(\boldsymbol{w}) = \mathsf{Arc}(\boldsymbol{w}_1 - \boldsymbol{w}_2, \boldsymbol{w}_2 - \boldsymbol{w}_3) + \mathsf{Arc}(\boldsymbol{w}_2 - \boldsymbol{w}_1, \boldsymbol{w}_1 - \boldsymbol{w}_4) \quad \text{on } (\mathbb{R}^{d_k})^4,\ k = 1, 2,$$

since, by Kim, Balakrishnan and Wasserman ((2020a), Theorem 7.2), we have

$$\begin{aligned}\mu_{f_1^{\tau^*}, f_2^{\tau^*}, H_*^4}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \int_{\mathcal{S}_{d_1-1}\times\mathcal{S}_{d_2-1}} &\mathrm{E}\{a_{\mathrm{sign}}(\boldsymbol{\alpha}_1^\top \boldsymbol{X}_{11}, \boldsymbol{\alpha}_1^\top \boldsymbol{X}_{12}, \boldsymbol{\alpha}_1^\top \boldsymbol{X}_{13}, \boldsymbol{\alpha}_1^\top \boldsymbol{X}_{14})\\ &\times a_{\mathrm{sign}}(\boldsymbol{\alpha}_2^\top \boldsymbol{X}_{21}, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_{22}, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_{23}, \boldsymbol{\alpha}_2^\top \boldsymbol{X}_{24})\}\,\mathrm{d}\lambda_{d_1}(\boldsymbol{\alpha}_1)\,\mathrm{d}\lambda_{d_2}(\boldsymbol{\alpha}_2),\end{aligned}$$

with $a_{\mathrm{sign}}(w_1, w_2, w_3, w_4) := \mathrm{sign}(|w_1 - w_2| - |w_1 - w_3| - |w_4 - w_2| + |w_4 - w_3|)$.

REMARK 2.1. Sejdinovic et al. (2013) recognize distance covariance as an example of an HSIC-type statistic (Gretton et al. (2005a, 2005b, 2005c), Fukumizu, Bach and Gretton (2007)). The HSIC-type statistics are all 4th order multivariate GSCs, and we note that our results for distance covariance readily extend to other HSIC-type statistics.

REMARK 2.2. In the univariate case, the GSCs from Example 2.1(b)–(e) reduce to the D of Hoeffding (1948), R of Blum, Kiefer and Rosenblatt (1961), and τ^* of Bergsma and Dassios (2014), respectively. As shown by Drton, Han and Shi (2020), the latter is connected to the work of Yanagimoto (1970). In Appendix B.1, we simplify the kernels for the univariate case, and show that the GSC framework also covers the τ of Kendall (1938).

All the multivariate dependence measures we have introduced are D-consistent, albeit with some variations in the families of distributions for which this holds; see, for example, the discussions in Examples 2.1–2.3 of Drton, Han and Shi (2020). As these dependence measures all involve the group H_*^m, we highlight the following fact.

LEMMA 2.1. *A GSC $\mu = \mu_{f_1, f_2, H_*^m}$ with $m \geq 4$ is D-consistent in a family $\mathcal{P}$ if and only if the pair (f_1, f_2) is D-consistent in $\mathcal{P}$—namely, if and only if*

$$\begin{aligned}\mathrm{E}\Bigg[\prod_{k=1}^{2}\{&f_k(\boldsymbol{X}_{k1}, \boldsymbol{X}_{k2}, \boldsymbol{X}_{k3}, \boldsymbol{X}_{k4}, \boldsymbol{X}_{k5}, \ldots, \boldsymbol{X}_{km}) - f_k(\boldsymbol{X}_{k1}, \boldsymbol{X}_{k3}, \boldsymbol{X}_{k2}, \boldsymbol{X}_{k4}, \boldsymbol{X}_{k5}, \ldots, \boldsymbol{X}_{km})\\ &- f_k(\boldsymbol{X}_{k4}, \boldsymbol{X}_{k2}, \boldsymbol{X}_{k3}, \boldsymbol{X}_{k1}, \boldsymbol{X}_{k5}, \ldots, \boldsymbol{X}_{km}) + f_k(\boldsymbol{X}_{k4}, \boldsymbol{X}_{k3}, \boldsymbol{X}_{k2}, \boldsymbol{X}_{k1}, \boldsymbol{X}_{k5}, \ldots, \boldsymbol{X}_{km})\}\Bigg]\end{aligned}$$

is finite, nonnegative, and equal to 0 *only if $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent.*

THEOREM 2.1. *All the multivariate GSCs in Example 2.1 are D-consistent within the family* $\{\mathrm{P} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2} \mid \mathrm{E}_{\mathrm{P}}[f_k(\boldsymbol{X}_{k1}, \dots, \boldsymbol{X}_{km})] < \infty,\ k = 1, 2\}$ (*with* f_k, $k = 1, 2$ *denoting their respective kernels*).

The invariance/equivariance properties of GSCs depend on those of their kernels. We say that a kernel function $f : (\mathbb{R}^d)^m \to \mathbb{R}$ is *orthogonally invariant* if, for any orthogonal matrix $\mathbf{O} \in \mathbb{R}^{d\times d}$ and any $\boldsymbol{w}_1, \dots, \boldsymbol{w}_m \in (\mathbb{R}^d)^m$, $f(\boldsymbol{w}_1, \dots, \boldsymbol{w}_m) = f(\mathbf{O}\boldsymbol{w}_1, \dots, \mathbf{O}\boldsymbol{w}_m)$.

LEMMA 2.2. *If* f_1 *and* f_2 *both are orthogonally invariant, then any GSC of the form* $\mu = \mu_{f_1,f_2,H}$ *is orthogonally invariant, that is,* $\mu(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu(\mathbf{O}_1\boldsymbol{X}_1, \mathbf{O}_2\boldsymbol{X}_2)$ *for any pair of random vectors* $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ *and orthogonal matrices* $\mathbf{O}_1 \in \mathbb{R}^{d_1\times d_1}$ *and* $\mathbf{O}_2 \in \mathbb{R}^{d_2\times d_2}$.

PROPOSITION 2.2. *The kernels (a), (c)–(e) in Example 2.1, hence the corresponding GSCs, are orthogonally invariant.*

Turning from theoretical dependence measures to their empirical counterparts, it is clear that any GSC admits a natural unbiased estimator in the form of a U-statistic, which we call the *sample generalized symmetric covariance* (SGSC).

DEFINITION 2.2 (Sample generalized symmetric covariance). The sample generalized symmetric covariance of $\mu = \mu_{f_1,f_2,H}$ is $\widehat{\mu}^{(n)} = \widehat{\mu}^{(n)}([(\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})]_{i=1}^n; f_1, f_2, H)$, of the form

$$\widehat{\mu}^{(n)} = \binom{n}{m}^{-1} \sum_{i_1<i_2<\cdots<i_m} \overline{k}_{f_1,f_2,H}\big((\boldsymbol{x}_{1i_1}, \boldsymbol{x}_{2i_1}), \dots, (\boldsymbol{x}_{1i_m}, \boldsymbol{x}_{2i_m})\big),$$

where $\overline{k}_{f_2,f_2,H}$ is the "symmetrized" version of $k_{f_2,f_2,H}$:

$$\overline{k}_{f_1,f_2,H}\big([(\boldsymbol{x}_{1\ell}, \boldsymbol{x}_{2\ell})]_{\ell=1}^m\big) := \frac{1}{m!} \sum_{\sigma\in\mathfrak{S}_m} k_{f_1,f_2,H}\big([(\boldsymbol{x}_{1\sigma(\ell)}, \boldsymbol{x}_{2\sigma(\ell)})]_{\ell=1}^m\big).$$

If the kernels f_1 and f_2 are orthogonally invariant, then it also holds that all SGSCs of the form $\widehat{\mu}^{(n)}(\cdot\,; f_1, f_2, H)$ are orthogonally invariant, in the sense of remaining unaffected when the input $[(\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})]_{i=1}^n$ is transformed into $[(\mathbf{O}_1\boldsymbol{x}_{1i}, \mathbf{O}_2\boldsymbol{x}_{2i})]_{i=1}^n$ where $\mathbf{O}_1 \in \mathbb{R}^{d_1\times d_1}$ and $\mathbf{O}_2 \in \mathbb{R}^{d_2\times d_2}$ are arbitrary orthogonal matrices. Proposition 2.2 thus also implies the orthogonal invariance of SGSCs associated with kernels (a) and (c)–(e) in Example 2.1.

The SGSCs associated with the examples listed in Example 2.1, unfortunately, all fail to satisfy the crucial property of distribution-freeness. However, as we will show in Section 4, distribution-freeness, along with transformation invariance, can be obtained by computing SGSCs from (functions of) the center-outward ranks and signs of the observations.

3. Center-outward ranks and signs. This section briefly introduces the concepts of center-outward ranks and signs to be used in the sequel. The main purpose is to fix notation and terminology; for a comprehensive coverage, we refer to Hallin et al. (2021).

We are concerned with defining multivariate ranks for a sample of d-dimensional observations drawn from a distribution in the class $\mathcal{P}^{\mathrm{ac}}_d$ of absolutely continuous probability measures on $\mathbb{R}^d$ with $d \ge 2$. Let $\mathbb{S}_d$ and $\mathcal{S}_{d-1}$ denote the open unit ball and the unit sphere in $\mathbb{R}^d$, respectively. Denote by U_d the spherical uniform measure on $\mathbb{S}_d$, that is, the product of the uniform measures on $[0, 1)$ (for the distance to the origin) and on $\mathcal{S}_{d-1}$ (for the direction). The push-forward of a measure Q by a measurable transformation T is denoted as $T\sharp\mathrm{Q}$.

DEFINITION 3.1 (Center-outward distribution function). The *center-outward distribution function* of a probability measure $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ is the P-a.s. unique function $\mathbf{F}_\pm$ that (i) maps $\mathbb{R}^d$ to the open unit ball $\mathbb{S}_d$, (ii) is the gradient of a convex function on $\mathbb{R}^d$, and (iii) pushes P forward to U_d (i.e., such that $\mathbf{F}_\pm \sharp \mathrm{P} = \mathrm{U}_d$).

The center-outward distribution function $\mathbf{F}_\pm$ of P entirely characterizes P provided that $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$; cf. Hallin et al. ((2021), Proposition 2.1(iii)). Also, $\mathbf{F}_\pm$ is invariant under shift, global rescaling, and orthogonal transformations. We refer the readers to Appendix B.2 for details about these elementary properties of center-outward distribution functions.

The sample counterpart $\mathbf{F}_\pm^{(n)}$ of $\mathbf{F}_\pm$ is based on an n-tuple of data points $z_1, \dots, z_n \in \mathbb{R}^d$. The key idea is to construct n grid points in the unit ball $\mathbb{S}_d$ such that the corresponding discrete uniform distribution converges weakly to U_d as $n \to \infty$. For $d \geq 2$, the construction proposed in Hallin ((2017), Section 4.2) starts by factorizing n into

$$n = n_R n_S + n_0, \qquad n_R, n_S \in \mathbb{Z}_{>0}, \qquad 0 \leq n_0 < \min\{n_R, n_S\},$$

where in asymptotic scenarios n_R and $n_S \to \infty$, hence $n_0/n \to 0$, as $n \to \infty$. Next consider the intersection points between

- the n_R hyperspheres centered at $\mathbf{0}_d$, with radii $r/(n_R+1)$, $r \in [\![n_R]\!]$, and
- n_S rays given by distinct unit vectors $\{\boldsymbol{s}_s^{(n_S)}\}_{s \in [\![n_S]\!]}$ that divide the unit circle into arcs of equal length $2\pi/n_S$ for $d = 2$, and are distributed as regularly as possible on the unit sphere $\mathcal{S}_{d-1}$ for $d \geq 3$; asymptotic statements merely require that the discrete uniform distribution over $\{\boldsymbol{s}_s^{(n_S)}\}_{s=1}^{n_S}$ converges weakly to the uniform distribution on $\mathcal{S}_{d-1}$ as $n_S \to \infty$.

Letting $\boldsymbol{n} := (n_R, n_S, n_0)$, the grid $\mathfrak{G}_{\boldsymbol{n}}^d$ is defined as the set of $n_R n_S$ points $\{\frac{r}{n_R+1}\boldsymbol{s}_s^{(n_S)}\}$ with $r \in [\![n_R]\!]$ and $s \in [\![n_S]\!]$ as described above along with the origin $\mathbf{0}$ in case $n_0 = 1$ or, whenever $n_0 > 1$, the n_0 points $\{\frac{1}{2(n_R+1)}\boldsymbol{s}_s^{(n_S)}\}$, $s \in \mathcal{S}$ where $\mathcal{S}$ is chosen as a random sample of size n_0 without replacement from $[\![n_S]\!]$. For $d = 1$, letting $n_S = 2$, $n_R = \lfloor n/n_S \rfloor$, $n_0 = n - n_R n_S = 0$ or 1, $\mathfrak{G}_{\boldsymbol{n}}^d$ reduces to the points $\{\pm r/(n_R+1) : r \in [\![n_R]\!]\}$, along with the origin 0 in case $n_0 = 1$.

The empirical version $\mathbf{F}_\pm^{(n)}$ of $\mathbf{F}_\pm$ is then defined as the optimal coupling between the observed data points and the grid $\mathfrak{G}_{\boldsymbol{n}}^d$.

DEFINITION 3.2 (Center-outward ranks and signs). Let $z_1, \dots, z_n$ be distinct data points in $\mathbb{R}^d$. Let $\mathcal{T}$ be the collection of all bijective mappings between the set $\{z_i\}_{i=1}^n$ and the grid $\mathfrak{G}_{\boldsymbol{n}}^d = \{\boldsymbol{u}_i\}_{i=1}^n$. The *sample center-outward distribution function* is defined as

$$\mathbf{F}_\pm^{(n)} := \operatorname*{argmin}_{T \in \mathcal{T}} \sum_{i=1}^n \|z_i - T(z_i)\|^2, \tag{3.1}$$

and $(n_R+1)\|\mathbf{F}_\pm^{(n)}(z_i)\|$ and $\mathbf{F}_\pm^{(n)}(z_i)/\|\mathbf{F}_\pm^{(n)}(z_i)\|$ are called the *center-outward rank* and *center-outward sign* of z_i, respectively.

REMARK 3.1. The particular way that the grid $\mathfrak{G}_{\boldsymbol{n}}^d$ is constructed here produces center-outward ranks and signs that enjoy all the properties—uniform distributions and mutual independence—that are expected from ranks and signs (see Section B.2 of the online Appendix (Shi et al. (2022)). These properties, however, are not required for the finite-sample validity and asymptotic properties of the rank-based tests we are pursuing in the subsequent sections. Any sequence of grids $\mathfrak{G}_{\boldsymbol{n}}^d$, whether stochastic (defined over a different probability space than the observations) or deterministic, is fine provided that the corresponding empirical distribution converges to the spherical uniform U_d. In addition, for the reasons developed, for example, in Hallin (2022), we deliberately only consider the spherical uniform U_d. In practice, the uniform distribution over the unit cube $[0,1]^d$ could be considered as well, yielding similar tests enjoying similar properties, with proofs following along similar lines.

The next proposition describes the Glivenko–Cantelli property of empirical center-outward distribution functions, a result we shall heavily rely on.

PROPOSITION 3.1 (Hallin (2017), Proposition 5.1, del Barrio et al. (2018), Theorem 3.1, and Hallin et al. (2021), Proposition 2.3). *Consider the following classes of distributions*:

- *the class $\mathcal{P}_d^+$ of distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ with nonvanishing probability density, namely, with Lebesgue density f such that, for all $D > 0$ there exist constants $\lambda_{D;f} < \Lambda_{D;f} \in (0,\infty)$ such that $\lambda_{D;f} \le f(z) \le \Lambda_{D;f}$ for all $\|z\| \le D$;*
- *the class $\mathcal{P}_d^\#$ of all distributions $\mathrm{P} \in \mathcal{P}_d^{\mathrm{ac}}$ such that, denoting by $\mathbf{F}_\pm^{(n)}$ the sample distribution function computed from an n-tuple $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ of independent copies of $\mathbf{Z} \sim \mathrm{P}$,*

$$\max_{1\le i\le n} \|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i) - \mathbf{F}_\pm(\mathbf{Z}_i)\| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n_R \text{ and } n_S \to \infty. \tag{3.2}$$

It holds that $\mathcal{P}_d^+ \subsetneq \mathcal{P}_d^\# \subsetneq \mathcal{P}_d^{\mathrm{ac}}$.

More properties about the population and empirical center-outward distribution functions can be found in Appendix B.2 and Chernozhukov et al. (2017), Figalli (2018), del Barrio, González-Sanz and Hallin (2020), Hallin et al. (2021), Ghosal and Sen (2022), and references there in.

4. Rank-based dependence measures. We are now ready to present our proposed family of dependence measures based on the notions of GSCs and center-outward ranks and signs. Throughout, $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is a pair of random vectors with $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$, and $(\boldsymbol{X}_{11}, \boldsymbol{X}_{21}), (\boldsymbol{X}_{12}, \boldsymbol{X}_{22}), \dots, (\boldsymbol{X}_{1n}, \boldsymbol{X}_{2n})$ is an n-tuple of independent copies of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$. Let $\mathbf{F}_{k,\pm}$ denote the center-outward distribution function of $\boldsymbol{X}_k$, and write $\mathbf{F}_{k,\pm}^{(n)}$ for the sample center-outward distribution function corresponding to $\{\boldsymbol{X}_{ki}\}_{i=1}^n$, $k = 1, 2$.

Our ideas build on Shi, Drton and Han (2022a) and, in slightly different form, also on Deb and Sen (2022), where the authors introduce a multivariate dependence measure by applying distance covariance to $\mathbf{F}_{1,\pm}(\boldsymbol{X}_1)$ and $\mathbf{F}_{2,\pm}(\boldsymbol{X}_2)$, with a sample counterpart involving $\mathbf{F}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i})$ and $\mathbf{F}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i})$, $i \in [\![n]\!]$. Our generalization of this particular dependence measure involves *score functions* and requires further notation. The score functions are continuous functions $J_1, J_2 : [0,1) \to \mathbb{R}_{\ge 0}$. Classical examples include the *normal* or *van der Waerden score function* $J_{\mathrm{vdW}}(u) := (F_{\chi_d^2}^{-1}(u))^{1/2}$ (with $F_{\chi_d^2}$ the χ_d^2 distribution function), the *Wilcoxon score function* $J_{\mathrm{W}}(u) := u$, and the *sign test score function* $J_{\mathrm{sign}}(u) := 1$. For $k = 1, 2$, let $\mathbf{J}_k(\boldsymbol{u}) := J_k(\|\boldsymbol{u}\|)\boldsymbol{u}/\|\boldsymbol{u}\|$ if $\boldsymbol{u} \in \mathbb{S}_{d_k} \backslash \{\mathbf{0}_{d_k}\}$ and $\mathbf{0}_{d_k}$ if $\boldsymbol{u} = \mathbf{0}_{d_k}$. Define the population and sample *scored center-outward distribution functions* as $\mathbf{G}_{k,\pm}(\cdot) := \mathbf{J}_k(\mathbf{F}_{k,\pm}(\cdot))$ and $\mathbf{G}_{k,\pm}^{(n)}(\cdot) := \mathbf{J}_k(\mathbf{F}_{k,\pm}^{(n)}(\cdot))$, respectively.

DEFINITION 4.1 (Rank-based dependence measures). Let J_1, J_2 be two score functions. The *(scored) rank-based version* of a dependence measure μ is obtained by applying μ to the pair $(\mathbf{G}_{1,\pm}(\boldsymbol{X}_1), \mathbf{G}_{2,\pm}(\boldsymbol{X}_2))$. For a GSC $\mu = \mu_{f_1,f_2,H}$, the rank-based version is denoted

$$\mu_\pm(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu_{\pm;J_1,J_2,f_1,f_2,H}(\boldsymbol{X}_1, \boldsymbol{X}_2) := \mu_{f_1,f_2,H}\big(\mathbf{G}_{1,\pm}(\boldsymbol{X}_1), \mathbf{G}_{2,\pm}(\boldsymbol{X}_2)\big) \tag{4.1}$$

and termed a *rank-based GSC* for short. The associated *rank-based SGSC* is

$$W_{\mu}^{(n)} = W_{J_1,J_2,\mu_{f_1,f_2,H}}^{(n)} := \widehat{\mu}^{(n)}\big([(\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i}), \mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i}))]_{i=1}^{n}; f_1, f_2, H\big). \tag{4.2}$$

REMARK 4.1. There is no immediate reason why a rank-based GSC should itself by a GSC in the sense of Definition 2.1. In this context, an observation of Bergsma (2006), Bergsma (2011) is of interest. For distance covariance in the univariate case (equivalent to 4κ in his notation), Lemma 10 in Bergsma (2006) implies that

$$\frac{1}{16}\mu_{f_1^{\mathrm{dCov}}, f_2^{\mathrm{dCov}}, H_*^4}\big(\mathbf{G}_{X_1,\pm}(X_1), \mathbf{G}_{X_2,\pm}(X_2)\big) = \int (F_{(X_1,X_2)} - F_{X_1}F_{X_2})^2 \,\mathrm{d}F_{X_1}\,\mathrm{d}F_{X_2}.$$

In other words, for $d_1 = d_2 = 1$ and $J_1(u) = J_2(u) = u$, the rank-based distance covariance coincides with R of Blum, Kiefer and Rosenblatt (1961) up to a scalar multiple. Recall that R is a GSC, but of higher order than distance covariance; see Example B.1(c) in Appendix B.1.

Plugging the center-outward ranks and signs into the multivariate dependence measures from Section 2 in combination with various score functions, one immediately obtains a large variety of rank-based GSCs and SGSCs, as we exemplify below. In particular, the choice $f_1 = f_1^{\mathrm{dCov}}$, $f_2 = f_2^{\mathrm{dCov}}$, $J_1(u) = J_2(u) = u$, and $H = H_*^4$ recovers the multivariate rank-based distance covariance from Shi, Drton and Han (2022a).

EXAMPLE 4.1. Some rank-based SGSCs.

(a) Rank-based distance covariance

$$W_{\mathrm{dCov}}^{(n)} := \binom{n}{4}^{-1} \sum_{i_1<\cdots<i_4} h_{\mathrm{dCov}}\big((\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_1}), \mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_1})), \ldots, (\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_4}), \mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_4}))\big)$$

with $h_{\mathrm{dCov}} := \overline{k}_{f_1^{\mathrm{dCov}}, f_2^{\mathrm{dCov}}, H_*^4}$ as given in Example 2.1(a). We have by definition that

$$\begin{aligned}W_{\mathrm{dCov}}^{(n)} = \binom{n}{4}^{-1} \sum_{i_1\neq\cdots\neq i_4} \frac{1}{4\cdot 4!}\big[\{&\|\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_1}) - \mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_2})\| - \|\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_1}) - \mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_3})\| \\ &- \|\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_4}) - \mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_2})\| + \|\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_4}) - \mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i_3})\|\} \\ &\times \{\|\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_1}) - \mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_2})\| - \|\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_1}) - \mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_3})\| \\ &- \|\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_4}) - \mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_2})\| + \|\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_4}) - \mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i_3})\|\}\big].\end{aligned}$$

(b) Similarly, Hoeffding's rank-based multivariate marginal ordering D (giving $W_M^{(n)}$), Hoeffding's rank-based multivariate projection-averaging D ($W_D^{(n)}$), Blum–Kiefer–Rosenblatt's rank-based multivariate projection-averaging R ($W_R^{(n)}$), and Bergsma–Dassios–Yanagimoto's rank-based multivariate projection-averaging τ^* ($W_{\tau^*}^{(n)}$) can be defined with kernels $h_M := \overline{k}_{f_1^M, f_2^M, H_*^5}$, $h_D := \overline{k}_{f_1^D, f_2^D, H_*^5}$, $h_R := \overline{k}_{f_1^R, f_2^R, H_*^6}$, and $h_{\tau^*} := \overline{k}_{f_1^{\tau^*}, f_2^{\tau^*}, H_*^4}$ as given in Example 2.1, respectively.

Having proposed a general class of dependence measures, we now examine, for each rank-based GSC, the five desirable properties listed in Section 1.2. To this end, we first introduce two regularity conditions on the score functions.

DEFINITION 4.2. A score function $J : [0, 1) \to \mathbb{R}_{\geq 0}$ is called *weakly regular* if it is continuous over $[0, 1)$ and nondegenerate: $\int_0^1 J^2(u)\,\mathrm{d}u > 0$. If, moreover, J is Lipschitz-continuous, strictly monotone, and satisfies $J(0) = 0$, it is called *strongly regular*.

PROPOSITION 4.1. *The normal and sign test score functions are weakly but not strongly regular; the Wilcoxon score function is strongly regular.*

PROPOSITION 4.2. *Suppose the considered pair* $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ *has marginal distributions* $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ *and* $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$. *Consider any rank-based GSC* $\mu_{\pm} := \mu_{\pm;J_1,J_2,f_1,f_2,H}$ *and its rank-based SGSC* $\underset{\sim}{W}_{\mu}^{(n)} := \underset{\sim}{W}_{J_1,J_2,\mu_{f_1,f_2,H}}^{(n)}$ *as defined in* (4.1) *and* (4.2). *Further, let* $\mu_{*\pm} := \mu_{\pm;J_1,J_2,f_1,f_2,H_*^m}$ *be an instance using the group from* (1.1). *Then*:

(i) (*Exact distribution-freeness*) *Under independence of* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$, *the distribution of* $\underset{\sim}{W}_{\mu}^{(n)}$ *does not depend on* $\mathrm{P}_{\boldsymbol{X}_1}$ *nor* $\mathrm{P}_{\boldsymbol{X}_2}$.

(ii) (*Transformation invariance*) *If the kernels* f_1 *and* f_2 *are orthogonally invariant, it holds for any orthogonal matrix* $\mathbf{O}_k \in \mathbb{R}^{d_k \times d_k}$, *any vector* $\boldsymbol{v}_k \in \mathbb{R}^{d_k}$, *and any scalar* $a_k \in \mathbb{R}_{>0}$ *that* $\mu_{\pm}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mu_{\pm}(\boldsymbol{v}_1 + a_1\mathbf{O}_1\boldsymbol{X}_1, \boldsymbol{v}_2 + a_2\mathbf{O}_2\boldsymbol{X}_2)$.

(iii) (*I- and D-Consistency*)

(a) $\mu_{\pm}$ *is I-consistent in the family*

$$\{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \mid \mathrm{P}_{\boldsymbol{X}_k} \in \mathcal{P}_{d_k}^{\mathrm{ac}} \text{ and } \mathrm{E}[f_k([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki})]_{i=1}^m)] < \infty \text{ for } k = 1, 2\}.$$

(b) *If the pair of kernels is D-consistent in the class*

$$\{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}} \mid \mathrm{E}[f_k(\boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{km})] < \infty \text{ for } k = 1, 2\}$$

(*cf. Lemma* 2.1), *then* $\mu_{*\pm}$ *is D-consistent in the family*

$$\mathcal{P}_{d_1,d_2,\infty}^{\mathrm{ac}} := \{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1+d_2}^{\mathrm{ac}} \mid \mathrm{E}[f_k([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki})]_{i=1}^m)] < \infty \text{ for } k = 1, 2\} \tag{4.3}$$

provided that the score functions J_1 *and* J_2 *are strictly monotone*;

(iv) (*Strong consistency*) *If* $f_k([\mathbf{G}_{k,\pm}^{(n)}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m)$ *and* $f_k([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m)$ *are almost surely bounded, that is, if there exists a constant* C (*depending on* f_k, J_k, *and* $\mathrm{P}_{\boldsymbol{X}_k}$) *such that for any* n *and* $k = 1, 2$,

$$\mathrm{P}(|f_k([\mathbf{G}_{k,\pm}^{(n)}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m)| \le C) = 1 = \mathrm{P}(|f_k([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m)| \le C),$$

and

$$(n)_m^{-1} \sum_{[i_1,\ldots,i_m] \in I_m^n} |f_k([\mathbf{G}_{k,\pm}^{(n)}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m) - f_k([\mathbf{G}_{k,\pm}(\boldsymbol{X}_{ki_\ell})]_{\ell=1}^m)| \xrightarrow{\text{a.s.}} 0, \tag{4.4}$$

then

$$\underset{\sim}{W}_{\mu}^{(n)} = \underset{\sim}{W}_{J_1,J_2,\mu_{f_1,f_2,H}}^{(n)} \xrightarrow{\text{a.s.}} \mu_{\pm}(\boldsymbol{X}_1, \boldsymbol{X}_2). \tag{4.5}$$

THEOREM 4.1 (Examples). *As long as* $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\#}$, $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\#}$, *and* J_1, J_2 *are strongly regular, all the kernel functions in Example* 2.1(a)–(e) *satisfy Condition* (4.4).

REMARK 4.2. Unfortunately, Theorem 4.1 does not imply that the rank-based SGSCs with normal score functions satisfy (4.5) although, in view of Proposition 4.2(iii), their population counterparts are both I- and D-consistent within a fairly large nonparametric family of distributions. A weaker version (replacing a.s. convergence by convergence in probability) of (4.5) holds in the univariate case with $d_1 = d_2 = 1$ by Feuerverger ((1993), Section 6). Consistency for normal scores, however, follows from a recent and yet unpublished result of Deb, Bhattacharya and Sen ((2021), Proposition 4.3), which was not available to us at the time this paper was written and which is obtained via a completely different technique.

We conclude this section with a discussion of computational issues. Two steps, in the evaluation of multivariate rank-based SGSCs, are potentially costly: (i) calculating the center-outward ranks and signs in (3.1), and (ii) computing a GSC $\widehat{\mu}^{(n)}(\cdot)$ with n inputs. The optimal matching problem (3.1) yielding $[\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i})]_{i=1}^n$ and $[\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i})]_{i=1}^n$ can be solved in $O(n^{5/2}\log(nN))$ time if the costs $\|\boldsymbol{z}_i-\boldsymbol{u}_j\|^2$, $i, j \in [\![n]\!]$ are integers bounded by N (Gabow and Tarjan (1989)); in dimension $d=2$, this can improved to $O(n^{3/2+\delta}\log(N))$ time for some arbitrarily small constant $\delta>0$ (Sharathkumar and Agarwal (2012)). The problem can also be solved approximately in $O(n^{3/2}\Omega(n,\epsilon,\Delta))$ time if $d \geq 3$, where

$$\Omega(n,\epsilon,\Delta) := \epsilon^{-1}\tau(n,\epsilon)\log^4(n/\epsilon)\log(\Delta)$$

depends on n, ϵ (the accuracy of the approximation) and $\Delta := \max c_{ij}/\min c_{ij}$, with $\tau(n,\epsilon)$ a small term (Agarwal and Sharathkumar (2014)). Further details are deferred to Appendix B.3.

Once $[\mathbf{G}_{1,\pm}^{(n)}(\boldsymbol{X}_{1i})]_{i=1}^n$ and $[\mathbf{G}_{2,\pm}^{(n)}(\boldsymbol{X}_{2i})]_{i=1}^n$ are obtained, a naïve evaluation of $\underset{\sim}{W}^{(n)}$, on the other hand, requires $O(n^m)$ operations. Great speedups are possible, however, in particular cases such as the rank-based SGSCs from Example 4.1. A detailed summary is provided in Proposition B.4 of the Appendix. The total computational complexity of the five statistics in Example 4.1 is given in the last three rows of Table 1.

5. Local power of rank-based tests of independence. Besides quantifying the dependence between two groups of random variables, the rank-based GSCs from Section 4 allow for constructing tests of the null hypothesis

$$H_0 : \boldsymbol{X}_1 \text{ and } \boldsymbol{X}_2 \text{ are mutually independent},$$

based on a sample $(\boldsymbol{X}_{11},\boldsymbol{X}_{21}),\ldots,(\boldsymbol{X}_{1n},\boldsymbol{X}_{2n})$ of n independent copies of $(\boldsymbol{X}_1,\boldsymbol{X}_2)$. Shi, Drton and Han (2022a), and, in a slightly different manner, Deb and Sen (2022), studied the particular case of a test based on the Wilcoxon version of the rank-based distance covariance $\underset{\sim}{W}_{\mathrm{dCov}}^{(n)}$. Among other results, they derive the limiting null distribution of $\underset{\sim}{W}_{\mathrm{dCov}}^{(n)}$, using combinatorial limit theorems and "brute-force" calculation of permutation statistics. Although this led to a fairly general combinatorial noncentral limit theorem (Shi, Drton and Han (2022a), Theorems 4.1 and 4.2), the derivation is not intuitive and difficult to generalize. In contrast, in this paper, we take a new and more powerful approach to the asymptotic analysis of rank-based SGSCs, which resolves the following three main issues:

(i) Intuitively, the asymptotic behavior of rank-based dependence measures follows from that of their Hájek *asymptotic representations*, which are oracle versions in which the observations are transformed using the unknown actual center-outward distribution function $\mathbf{F}_\pm$ rather than its sample version $\mathbf{F}_\pm^{(n)}$. Here, we show the correctness of this intuition by proving asymptotic equivalence between rank-based SGSCs and their oracle versions.

(ii) Previous work does not perform any power analysis for the new rank-based tests. Here, we fill this gap by proving that these tests have nontrivial power in the context of the class of quadratic mean differentiable alternatives (Lehmann and Romano (2005), Definition 12.2.1).

(iii) Finally, our rank-based tests allow for the incorporation of score functions, which may improve their performance.

This novel approach rests on a generalization of the classical Hájek representation method (Hájek and Šidák (1967)) to the multivariate setting of center-outward ranks and signs, which simplifies the derivation of asymptotic null distributions and, via a nontrivial use of Le Cam's third lemma for nonnormal limits, enables our local power analysis.

TABLE 1
Properties of the center-outward GSCs in Example 4.1 with weakly regular score functions J_k

	$\underset{\sim}{W}^{(n)}_{\mu}$	$\underset{\sim}{W}^{(n)}_{\mathrm{dCov}}$	$\underset{\sim}{W}^{(n)}_{M}$	$\underset{\sim}{W}^{(n)}_{D}$	$\underset{\sim}{W}^{(n)}_{R}$	$\underset{\sim}{W}^{(n)}_{\tau^*}$
(1)	Distribution-freeness	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1} \otimes \mathcal{P}^{\mathrm{ac}}_{d_2}$ (a)	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1} \otimes \mathcal{P}^{\mathrm{ac}}_{d_2}$	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1} \otimes \mathcal{P}^{\mathrm{ac}}_{d_2}$	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1} \otimes \mathcal{P}^{\mathrm{ac}}_{d_2}$	$\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1} \otimes \mathcal{P}^{\mathrm{ac}}_{d_2}$
(2)	Transformation invariance	Orthogonal transf., shifts, and global scales	Shifts and global scales	Orthogonal transf., shifts, and global scales	Orthogonal transf., shifts, and global scales	Orthogonal transf., shifts, and global scales
(3)	D-consistency	J_k strictly monotone and integrable, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2}$ (b)	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2}$	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2}$	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2}$	J_k strictly monotone, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2}$
(3′)	Consistency of test	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\#}_{d_1,d_2}$ (c)	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\#}_{d_1,d_2}$	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\#}_{d_1,d_2}$	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\#}_{d_1,d_2}$	J_k strongly regular, $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\#}_{d_1,d_2}$
(4)	Efficiency	J_k square-integrable	J_k weakly regular (as assumed)	J_k weakly regular (as assumed)	J_k weakly regular (as assumed)	J_k weakly regular (as assumed)
(5)	Exact $d_1 \vee d_2 = 2$	$O(n^2)$	$O(n^{3/2+\delta} \log N)$ (d)	$O(n^3)$	$O(n^4)$	$O(n^4)$
	Exact $d_1 \vee d_2 = 3$	$O(n^{5/2} \log(nN))$ (d)	$O(n^{5/2} \log(nN))$	$O(n^3)$	$O(n^4)$	$O(n^4)$
	Fast approximation	$O(n^{3/2}\Omega \vee nK \log n)$ (d)	$O(n^{3/2}\Omega)$	$O(n^{3/2}\Omega \vee nK \log n)$	$O(n^{3/2}\Omega \vee nK \log n)$	$O(n^{3/2}\Omega \vee nK \log n)$

(a) $\mathcal{P}^{\mathrm{ac}}_{d_1} \otimes \mathcal{P}^{\mathrm{ac}}_{d_2}$ is the family of all $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}$ such that $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ independent, $\mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}^{\mathrm{ac}}_{d_1}$ and $\mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}^{\mathrm{ac}}_{d_2}$.

(b) $\mathcal{P}^{\mathrm{ac}}_{d_1+d_2}$ is the family of all absolutely continuous distributions on $\mathbb{R}^{d_1+d_2}$.

(c) $\mathcal{P}^{\#}_{d_1,d_2} := \{\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2} \mid \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}^{\#}_{d_1}, \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}^{\#}_{d_2}\}$.

(d) Here we assume without loss of generality that c_{ij}, $i, j \in [\![n]\!]$ are all integers and bounded by integer N, δ is some arbitrarily small constant, Ω is defined as $\epsilon^{-1}\tau(n,\epsilon)\log^4(n/\epsilon)\log(\max c_{ij}/\min c_{ij})$, and K is sufficiently large; as usual, $q_1 \vee q_2$ stands for the minimum of two quantities q_1 and q_2. Also refer to Propositions B.3 and B.4 in Section B of the Appendix.

5.1. *Asymptotic representation.* In order to develop our multivariate asymptotic representation, we first introduce formally the oracle counterpart to the rank-based SGSC $\underset{\sim}{W}_{\mu}^{(n)}$.

DEFINITION 5.1 (Oracle rank-based SGSCs). The oracle version of the rank-based SGSC $\underset{\sim}{W}_{J_1,J_2,\mu_{f_1,f_2,H}}^{(n)}$ associated with the GSC $\mu = \mu_{f_1,f_2,H}$ is

$$W_{\mu}^{(n)} = W_{J_1,J_2,\mu_{f_1,f_2,H}}^{(n)} := \widehat{\mu}^{(n)}\big(\big[\big(\mathbf{G}_{1,\pm}(\boldsymbol{X}_{1i}), \mathbf{G}_{2,\pm}(\boldsymbol{X}_{2i})\big)\big]_{i=1}^{n}; f_1, f_2, H\big).$$

Note that the oracle $W_{\mu}^{(n)}$ cannot be computed from the observations as it involves the population scored center-outward distribution functions $\mathbf{G}_{1,\pm}$ and $\mathbf{G}_{2,\pm}$. However, the limiting null distribution of $W^{(n)}$, unlike that of $\underset{\sim}{W}^{(n)}$, follows from standard theory for degenerate U-statistics (Serfling (1980), Chapter 5.5.2). This point can be summarized as follows.

PROPOSITION 5.1. *Let $\mu = \mu_{f_1,f_2,H_*^m}$ be a GSC with $m \geq 4$. Let the kernels f_1, f_2 and the score functions J_1, J_2 satisfy*

$$0 < \mathrm{Var}\big(g_k(\boldsymbol{W}_{k1}, \boldsymbol{W}_{k2})\big) < \infty, \quad k = 1, 2, \tag{5.1}$$

where $\boldsymbol{W}_{ki} := \mathbf{J}_k(\boldsymbol{U}_{ki})$ with $(\boldsymbol{U}_{1i}, \boldsymbol{U}_{2i})$, $i \in [\![m]\!]$ independent and distributed according to the product of spherical uniform distributions $\mathrm{U}_{d_1} \otimes \mathrm{U}_{d_2}$,

$$g_k(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}) := \mathrm{E}\big[2 f_{k,H_*^m}(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}, \boldsymbol{W}_{k3}, \boldsymbol{W}_{k4}, \ldots, \boldsymbol{W}_{km})\big], \tag{5.2}$$

and $f_{k,H_*^m} := \sum_{\sigma \in H_*^m} \mathrm{sgn}(\sigma) f_k(\boldsymbol{x}_{k\sigma(1)}, \ldots, \boldsymbol{x}_{k\sigma(m)})$, $k = 1, 2$. Then, under the null hypothesis H_0 that $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ are independent,

$$nW_{\mu}^{(n)} = nW_{J_1,J_2,\mu_{f_1,f_2,H_*^m}}^{(n)} \rightsquigarrow \sum_{v=1}^{\infty} \lambda_{\mu,v}(\xi_v^2 - 1),$$

where $[\lambda_{\mu,v}]_{v=1}^{\infty}$ are the nonzero eigenvalues of the integral equation

$$\mathrm{E}\big[g_1(\boldsymbol{w}_{11}, \boldsymbol{W}_{12}) g_2(\boldsymbol{w}_{21}, \boldsymbol{W}_{22}) \psi(\boldsymbol{W}_{12}, \boldsymbol{W}_{22})\big] = \lambda \psi(\boldsymbol{w}_{11}, \boldsymbol{w}_{21}). \tag{5.3}$$

and $[\xi_v]_{v=1}^{\infty}$ are independent standard Gaussian random variables.

The tests we are considering reject for large values of test statistics that estimate a nonnegative (I- and D-)consistent dependence measure. In all these tests

(5.4) all eigenvalues of the integral equation (5.3) are nonnegative.

However, it should be noted that, in view of the following multivariate representation result, a valid test of H_0 can be implemented also when (5.4) does not hold.

THEOREM 5.1 (Multivariate Hájek representation). *Let f_1, f_2 be kernel functions of order $m \geq 4$, and let J_1, J_2 be weakly regular score functions. Writing $\mathrm{U}_{d_k}^{(n)}$ for the discrete uniform distribution over the grid $\mathfrak{G}_{\boldsymbol{n}}^{d_k}$, let $\boldsymbol{W}_{ki}^{(n)} := \mathbf{J}_k(\boldsymbol{U}_{ki}^{(n)})$ where $(\boldsymbol{U}_{1i}^{(n)}, \boldsymbol{U}_{2i}^{(n)})$ for $i \in [\![m]\!]$ are independent with distribution $\mathrm{U}_{d_1}^{(n)} \otimes \mathrm{U}_{d_2}^{(n)}$. Define g_k, $k = 1, 2$, as in (5.2), and*

$$g_k^{(n)}(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}) := \mathrm{E}\big[2 f_{k,H_*^m}\big(\boldsymbol{w}_{k1}, \boldsymbol{w}_{k2}, \boldsymbol{W}_{k3}^{(n)}, \boldsymbol{W}_{k4}^{(n)}, \ldots, \boldsymbol{W}_{km}^{(n)}\big)\big], \quad k = 1, 2. \tag{5.5}$$

Assume that

$$f_k \text{ and } g_k \text{ are Lipschitz-continuous}, \qquad g_k^{(n)} \text{ converges uniformly to } g_k,$$
$$\sup_{i_1,\ldots,i_m \in [\![m]\!]} \mathrm{E}\big[f_k\big([\boldsymbol{W}_{ki_\ell}]_{\ell=1}^{m}\big)^2\big] < \infty \quad \text{and} \quad \int_0^1 J_k^2(u)\,\mathrm{d}u < \infty, \quad k = 1, 2. \tag{5.6}$$

Then, under the hypothesis H_0 that $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ are independent, the rank-based SGSC $\underset{\sim}{W}_{\mu}^{(n)} = \underset{\sim}{W}_{J_1,J_2,\mu}^{(n)}$ associated to the GSC $\mu = \mu_{f_1,f_2,H_*^m}$ is asymptotically equivalent to its oracle version $W_{\mu}^{(n)}$, that is, $\underset{\sim}{W}_{\mu}^{(n)} - W_{\mu}^{(n)} = o_{\mathrm{P}}(n^{-1})$ as $n_R, n_S \to \infty$.

THEOREM 5.2. *The conclusion of Theorem 5.1 still holds with (5.6) replaced by*

(5.7) $\quad f_k$ *is uniformly bounded, and almost everywhere continuous,* $\quad k = 1, 2.$

PROPOSITION 5.2 (Examples). *If $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ is independent of $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ and J_1, J_2 are weakly regular, then the kernel functions from Example 2.1(b)–(e) satisfy (5.1), (5.4), and (5.7). If, moreover, J_1, J_2 are square-integrable (viz., $\int_0^1 J_k^2(u)\,\mathrm{d}u < \infty$ for $k = 1, 2$), then (5.1), (5.4), and (5.6) hold also for the kernels in Example 2.1(a).*

COROLLARY 5.1 (Limiting null distribution). *Suppose the conditions in Proposition 5.1 and Theorem 5.1 hold. Then, for $\mu = \mu_{f_1,f_2,H_*^m}$ with $m \geq 4$, under the hypothesis H_0 that $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$ are independent,*

$$n\underset{\sim}{W}_{\mu}^{(n)} = n\underset{\sim}{W}_{J_1,J_2,\mu_{f_1,f_2,H_*^m}}^{(n)} \rightsquigarrow \sum_{v=1}^{\infty} \lambda_{\mu,v}(\xi_v^2 - 1) \tag{5.8}$$

with $[\lambda_{\mu,v}]_{v=1}^{\infty}$ and $[\xi_v]_{v=1}^{\infty}$ as defined in Proposition 5.1.

REMARK 5.1. Corollary 5.1 gives no rate, that is, no Berry–Esséen type bound for the convergence in (5.8). Indeed, deriving such bounds in the present context is quite challenging. Results for the univariate case with $d_1 = d_2 = 1$ were established for simpler statistics such as Spearman's ρ and Kendall's τ by Koroljuk and Borovskich ((1994), Chapter 6.2) and, more recently, by Pinelis and Molzon (2016). Extending these results to the multivariate measure-transportation-based ranks considered here is highly nontrivial and requires properties of empirical transports that have not yet been obtained. This pertains, in particular, to working out the rate of convergence in the Glivenko–Cantelli result for the center-outward distribution function given in (3.2); an open problem in the recent survey by Hallin ((2022), Section 5).

For any significance level $\alpha \in (0, 1)$, define the quantile

$$q_{\mu,1-\alpha} := \inf\left\{x \in \mathbb{R} : \mathrm{P}\left(\sum_{v=1}^{\infty} \lambda_{\mu,v}(\xi_v^2 - 1) \leq x\right) \geq 1 - \alpha\right\}, \tag{5.9}$$

where $[\lambda_{\mu,v}]_{v=1}^{\infty}$ and $[\xi_v]_{v=1}^{\infty}$ are as in Proposition 5.1. Let $\underset{\sim}{W}_{\mu}^{(n)}$ be as in Theorem 5.1, and define the test

$$\mathsf{T}_{\mu,\alpha}^{(n)} := \mathbb{1}\big(n\underset{\sim}{W}_{\mu}^{(n)} > q_{\mu,1-\alpha}\big).$$

The next proposition summarizes the asymptotic validity and properties of this test.

PROPOSITION 5.3 (Uniform validity and consistency). *Let J_1, J_2 be weakly regular score functions, and let $\mu = \mu_{f_1,f_2,H_*^m}$ be a GSC with $m \geq 4$ such that Conditions (5.1) and one of (5.6) and (5.7) hold. Then:*

(i) $\lim_{n\to\infty} \mathrm{P}(\mathsf{T}_{\mu,\alpha}^{(n)} = 1) = \alpha$ *for any $\mathrm{P} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}$, that is, for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ independent with $\boldsymbol{X}_1 \sim \mathrm{P}_{\boldsymbol{X}_1} \in \mathcal{P}_{d_1}^{\mathrm{ac}}$ and $\boldsymbol{X}_2 \sim \mathrm{P}_{\boldsymbol{X}_2} \in \mathcal{P}_{d_2}^{\mathrm{ac}}$.*

(ii) *It follows from Proposition 4.2(i) that* $\lim_{n\to\infty} \sup_{\mathrm{P} \in \mathcal{P}_{d_1}^{\mathrm{ac}} \otimes \mathcal{P}_{d_2}^{\mathrm{ac}}} \mathrm{P}(\mathsf{T}_{\mu,\alpha}^{(n)} = 1) = \alpha$.

(iii) *If, moreover, the pair of kernels (f_1, f_2) is D-consistent, J_1, J_2 are strictly monotone, and (4.5) holds,* $\lim_{n\to\infty} \mathrm{P}(\mathsf{T}_{\mu,\alpha}^{(n)} = 1) = 1$ *for any fixed alternative $\mathrm{P}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)} \in \mathcal{P}_{d_1,d_2,\infty}^{\mathrm{ac}}$ as defined in (4.3).*

5.2. *Local power analysis.* In this section, we conduct local power analyses of the proposed tests for quadratic mean differentiable classes of alternatives (Lehmann and Romano (2005), Definition 12.2.1), for which we establish nontrivial power in $n^{-1/2}$ neighborhoods. We begin with a model $\{q_{\boldsymbol{X}}(\boldsymbol{x};\delta)\}_{|\delta|<\delta^*}$ with $\delta^*>0$, under which $\boldsymbol{X}=(\boldsymbol{X}_1,\boldsymbol{X}_2)$ has Lebesgue-density $q_{\boldsymbol{X}}(\boldsymbol{x};\delta)=q_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}((\boldsymbol{x}_1,\boldsymbol{x}_2);\delta)$, with $q_{\boldsymbol{X}_1}(\boldsymbol{x}_1;\delta)$ and $q_{\boldsymbol{X}_2}(\boldsymbol{x}_2;\delta)$ being the marginal densities. We then make the following assumptions.

ASSUMPTION 5.1.

(i) Dependence of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$: $q_{\boldsymbol{X}}(\boldsymbol{x};\delta)=q_{\boldsymbol{X}_1}(\boldsymbol{x}_1;\delta)q_{\boldsymbol{X}_2}(\boldsymbol{x}_2;\delta)$ holds if and only if $\delta=0$.

(ii) The family $\{q_\delta(\boldsymbol{x})\}_{|\delta|<\delta^*}$ is quadratic mean differentiable at $\delta=0$ with score function $\dot{\ell}(\cdot;0)$, that is,

$$\int\left(\sqrt{q_{\boldsymbol{X}}(\boldsymbol{x};\delta)}-\sqrt{q_{\boldsymbol{X}}(\boldsymbol{x};0)}-\frac{1}{2}\delta\dot{\ell}(\boldsymbol{x};0)\sqrt{q_{\boldsymbol{X}}(\boldsymbol{x};0)}\right)^2\mathrm{d}\boldsymbol{x}=o(\delta^2)\quad\text{as }\delta\to 0.$$

(iii) The Fisher information is positive, that is, $\mathcal{I}_{\boldsymbol{X}}(0):=\int\{\dot{\ell}(\boldsymbol{x};0)\}^2q_{\boldsymbol{X}}(\boldsymbol{x},0)\,\mathrm{d}\boldsymbol{x}>0$; of note, Assumption 5.1(ii) implies that $\mathcal{I}_{\boldsymbol{X}}(0)<\infty$ and $\int\dot{\ell}(\boldsymbol{x};0)q_{\boldsymbol{X}}(\boldsymbol{x},0)\,\mathrm{d}\boldsymbol{x}=0$.

(iv) The score function $\dot{\ell}(\boldsymbol{x};0)$ is not additively separable, that is, there do not exist functions h_1 and h_2 such that $\dot{\ell}(\boldsymbol{x};0)=h_1(\boldsymbol{x}_1)+h_2(\boldsymbol{x}_2)$.

REMARK 5.2. For the sake of simplicity, we have restricted ourselves to one-parameter classes. Analogous results hold for families indexed by a multivariate parameter $\boldsymbol{\delta}$.

For a local power analysis, we consider a sequence of local alternatives obtained as

$$H_1^{(n)}(\delta_0):\delta=\delta^{(n)}\quad\text{where }\delta^{(n)}:=n^{-1/2}\delta_0 \tag{5.10}$$

with some constant $\delta_0\neq 0$. In this local model, testing the null hypothesis of independence reduces to testing $H_0:\delta_0=0$ versus $H_1:\delta_0\neq 0$.

THEOREM 5.3 (Power analysis). *Consider a GSC $\mu=\mu_{f_1,f_2,H_*^m}$ with $m\geq 4$ and kernel functions f_1, f_2 picked from Example 2.1. Assume that J_1, J_2 are weakly regular score functions that satisfy the assumptions of Proposition 5.2. Then if Assumption 5.1 holds, for any $\beta>0$, there exists a constant $C_\beta>0$ depending only on β such that, as long as $|\delta_0|>C_\beta$, $\lim_{n\to\infty}\mathrm{P}\{\mathsf{T}_{\mu,\alpha}^{(n)}=1|H_1^{(n)}(\delta_0)\}\geq 1-\beta$.*

Following the arguments from the proof of Theorem 5.3, one should be able to obtain similar local power results for the original (non-rank-based) tests associated with the kernels listed in Example 2.1. However, to the best of our knowledge, this analysis has not been performed in the literature, except for $d_1=d_2=1$ where results can be found, for example, in Dhar, Dassios and Bergsma (2016) and Shi, Drton and Han (2022b). We also emphasize that, although Theorem 5.3 only considers the specific cases listed also in Example 4.1, the proof technique applies more generally. We refrain, however, from stating a more general version of Theorem 5.3 as this would require a number of tedious technical conditions.

As a by-product of Theorem 5.3, the following corollary gives the asymptotic distribution of the test statistic under the local alternative.

COROLLARY 5.2 (Limiting local alternative distribution). *Suppose all the conditions in Theorem 5.3 hold. Then, under the local alternative hypothesis $H_1^{(n)}(\delta_0)$,*

$$n\underset{\sim}{W}_\mu^{(n)}\rightsquigarrow\sum_{v=1}^{\infty}\lambda_{\mu,v}\big((\xi_v+\delta_0\gamma_{\mu,v})^2-1\big),$$

where $[\lambda_{\mu,v}]_{v=1}^{\infty}$ *and* $[\xi_v]_{v=1}^{\infty}$ *are as defined in Proposition* 5.1,

$$\gamma_{\mu,v} := \mathrm{Cov}\big[\psi_{\mu,v}\big(\mathbf{G}^*_{1,\pm}(\boldsymbol{X}^*_1), \mathbf{G}^*_{2,\pm}(\boldsymbol{X}^*_2)\big), \dot{\ell}\big((\boldsymbol{X}^*_1, \boldsymbol{X}^*_2); 0\big)\big],$$

in which $\psi_{\mu,v}$ *is the eigenfunction associated with the eigenvalue* $\lambda_{\mu,v}$ *of the integral equation* (5.3), $\boldsymbol{X}^* = (\boldsymbol{X}^*_1, \boldsymbol{X}^*_2)$ *has the same distribution as* $\boldsymbol{X}$ *with* $\delta = 0$, *and* $\mathbf{G}^*_{1,\pm}$ *and* $\mathbf{G}^*_{2,\pm}$ *denote the respective population scored center-outward distribution functions of* $\boldsymbol{X}^*_1$ *and* $\boldsymbol{X}^*_2$.

Notice that $[\gamma_{\mu,v}]_{v=1}^{\infty}$ cannot be all zero, which is ensured by Assumption 5.1(iv), and thus the limiting local alternative distribution will differ from the limiting null distribution.

Combined with the following result, Theorem 5.3 yields nontrivial power of the proposed tests in $n^{-1/2}$ neighborhoods of $\delta = 0$.

THEOREM 5.4. *Let Assumption* 5.1 *hold. Then, for any* $\beta > 0$ *such that* $\alpha + \beta < 1$, *there exists an absolute constant* $c_\beta > 0$ *such that, as long as* $|\delta_0| \le c_\beta$,

$$\inf_{\overline{\mathsf{T}}^{(n)}_\alpha \in \mathcal{T}^{(n)}_\alpha} \mathrm{P}\big\{\overline{\mathsf{T}}^{(n)}_\alpha = 0 \,|\, H^{(n)}_1(\delta_0)\big\} \ge 1 - \alpha - \beta$$

for all sufficiently large n. *Here the infimum is taken over the class* $\mathcal{T}^{(n)}_\alpha$ *of all size-*α *tests.*

Table 1 summarizes our results for the rank-based SGSCs from Example 4.1 by giving an overview of the five properties listed in the Introduction. It also indicates consistency of the tests. In all cases, it is assumed that the score functions involved are weakly regular.

5.3. *Examples in the quadratic mean differentiable class.* This section presents two specific examples in the quadratic mean differentiable class that satisfy Assumption 5.1. First, we consider parametrized families that extend the bivariate *Konijn alternatives* (Konijn (1956)). These alternatives are classical in the context of testing for multivariate independence and have also been considered by Gieser (1993), Gieser and Randles (1997), Taskinen, Kankainen and Oja (2003), Taskinen, Kankainen and Oja (2004), Taskinen, Oja and Randles (2005), and Hallin and Paindaveine (2008).

Konijn families are constructed as follows. Let $\boldsymbol{X}^*_1 \sim \mathrm{P}_{\boldsymbol{X}^*_1} \in \mathcal{P}^{\mathrm{ac}}_{d_1}$ and $\boldsymbol{X}^*_2 \sim \mathrm{P}_{\boldsymbol{X}^*_2} \in \mathcal{P}^{\mathrm{ac}}_{d_2}$ be two (without loss of generality) mean zero (unobserved) independent random vectors with densities q_1 and q_2, respectively. Let $\mathrm{P}_{\boldsymbol{X}^*} \in \mathcal{P}^{\mathrm{ac}}_{d_1+d_2}$ be their joint distribution, $q_{\boldsymbol{X}^*}(\boldsymbol{x}) = q_{\boldsymbol{X}^*}((\boldsymbol{x}_1, \boldsymbol{x}_2)) = q_1(\boldsymbol{x}_1)q_2(\boldsymbol{x}_2)$ be their joint density. Define, for $\delta \in \mathbb{R}$,

$$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix} := \begin{pmatrix}\mathbf{I}_{d_1} & \delta\,\mathbf{M}_1\\ \delta\,\mathbf{M}_2 & \mathbf{I}_{d_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{X}^*_1\\ \boldsymbol{X}^*_2\end{pmatrix} = \mathbf{A}_\delta\begin{pmatrix}\boldsymbol{X}^*_1\\ \boldsymbol{X}^*_2\end{pmatrix} = \mathbf{A}_\delta \boldsymbol{X}^*, \tag{5.11}$$

where $\mathbf{M}_1 \in \mathbb{R}^{d_1\times d_2}$ and $\mathbf{M}_2 \in \mathbb{R}^{d_2\times d_1}$ are two deterministic matrices. For $\delta = 0$, the matrix $\mathbf{A}_\delta$ is the identity and, thus, invertible. By continuity, $\mathbf{A}_\delta$ is also invertible for δ in a sufficiently small neighborhood Θ of 0. For $\delta \in \Theta$, the density of $\boldsymbol{X}$ can be expressed as $q_{\boldsymbol{X}}(\boldsymbol{x}; \delta) = |\det(\mathbf{A}_\delta)|^{-1} q_{\boldsymbol{X}^*}(\mathbf{A}_\delta^{-1}\boldsymbol{x})$, which is differentiable with respect to δ. The following additional assumptions will be made on the generating scheme (5.11).

ASSUMPTION 5.2.

(i) The distributions of $\boldsymbol{X}$ have a common support for all $\delta \in \Theta$. Without loss of generality, we assume $\boldsymbol{\mathcal{X}} := \{\boldsymbol{x} : q_{\boldsymbol{X}}(\boldsymbol{x}; \delta) > 0\}$ does not depend on δ.

(ii) The map $\boldsymbol{x} \mapsto \sqrt{q_{\boldsymbol{X}^*}(\boldsymbol{x})}$ is continuously differentiable.

(iii) The Fisher information $\mathcal{I}_{\boldsymbol{X}}(0) := \int \{\dot{\ell}(\boldsymbol{x}; 0)\}^2 q_{\boldsymbol{X}}(\boldsymbol{x}; 0)\,\mathrm{d}\boldsymbol{x}$ of $\boldsymbol{X}$ relative to δ at $\delta = 0$ is strictly positive and finite.

EXAMPLE 5.1.

(i) Suppose $\boldsymbol{X}_1^*$ and $\boldsymbol{X}_2^*$ are elliptical with centers $\mathbf{0}_{d_1}$ and $\mathbf{0}_{d_2}$ and covariances $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively, that is, $q_k(\boldsymbol{x}_k) \propto \phi_k(\boldsymbol{x}_k^\top \boldsymbol{\Sigma}_k^{-1}\boldsymbol{x}_k)$, $k = 1, 2$, where ϕ_k is such that $\mathrm{Var}(\boldsymbol{X}_k^*) = \boldsymbol{\Sigma}_k$ and $\mathrm{E}[\|\boldsymbol{Z}_k^*\|^2 \rho_k(\|\boldsymbol{Z}_k^*\|^2)^2] < \infty$, $k = 1, 2$ where $\boldsymbol{Z}_k^*$ has density function proportional to $\phi_k(\|z_k\|^2)$ and $\rho_k(t) := \phi_k'(t)/\phi_k(t)$. Then Assumption 5.2 is satisfied for any $\mathbf{M}_1$, $\mathbf{M}_2$ such that $\boldsymbol{\Sigma}_1\mathbf{M}_2^\top + \mathbf{M}_1\boldsymbol{\Sigma}_2 \neq \mathbf{0}$.

(ii) As a specific example of (i), if $\boldsymbol{X}_1^*$ and $\boldsymbol{X}_2^*$ are centered multivariate normal or follow centered multivariate t-distributions with degrees of freedom strictly greater than two, then Assumption 5.2 is satisfied for any $\mathbf{M}_1$, $\mathbf{M}_2$ such that $\boldsymbol{\Sigma}_1\mathbf{M}_2^\top + \mathbf{M}_1\boldsymbol{\Sigma}_2 \neq \mathbf{0}$.

Next, consider the following mixture model extending the alternatives treated in Dhar, Dassios and Bergsma ((2016), Section 3). Let q_1 and q_2 be fixed (Lebesgue-)density functions for $\boldsymbol{X}_1$ an $\boldsymbol{X}_2$, respectively. The joint density of $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ under independence is q_1q_2. Letting $q^* \neq q_1q_2$ denote a fixed joint density, mixture alternatives indexed by $\delta \in [0, 1]$ are defined as $q_{\boldsymbol{X}}(\boldsymbol{x}; \delta) := (1-\delta)q_1q_2 + \delta q^*$.

ASSUMPTION 5.3. It is assumed that:

(i) $(1+\delta^*)q_1q_2 - \delta^* q^*$ is a bonafide joint density for some $\delta^* > 0$;
(ii) q^* and q_1q_2 are mutually absolutely continuous;
(iii) the function $\delta \mapsto \sqrt{q_{\boldsymbol{X}}(\boldsymbol{x}; \delta)}$ is continuously differentiable in some neighborhood of 0;
(iv) the Fisher information $\mathcal{I}_{\boldsymbol{X}}(\delta) := \int (q^* - q_1q_2)^2/\{(1-\delta)q_1q_2 + \delta q^*\}\,\mathrm{d}\boldsymbol{x}$ of $\boldsymbol{X}$ relative to δ is finite, strictly positive, and continuous at $\delta = 0$;
(v) $\dot{\ell}(\boldsymbol{x}; 0) = q^*(\boldsymbol{x})/\{q_1(\boldsymbol{x}_1)q_2(\boldsymbol{x}_2)\} - 1$ is not additively separable.

EXAMPLE 5.2. If $q_k(\boldsymbol{x}_k) = 1$ for $\boldsymbol{x}_k \in [0, 1]^{d_k}$, $k = 1, 2$, and $q^*(\boldsymbol{x}) \not\equiv 1$ is continuous and supported on $[0, 1]^{d_1+d_2}$, then Assumption 5.3 holds.

PROPOSITION 5.4. *Assumption 5.1 is satisfied by the Konijn alternatives under Assumption 5.2, and by the mixture alternatives under Assumption 5.3.*

5.4. *Numerical experiments.* Extensive simulations of Shi, Drton and Han (2022a) give evidence for the superiority, under non-Gaussian densities, of the Wilcoxon versions of our tests over the original distance covariance tests. That superiority is more substantial when non-Wilcoxon scores, such as the Gaussian ones, are considered (Figure 4). In view of these results, there is little point in pursuing simulations with non-Gaussian densities, and we instead focus on Gaussian cases (Figures 1–3) to study the impact on finite-sample performance of the dimensions d_1 and d_2, sample size n, and within- and between-sample correlations.

EXAMPLE 5.3. The data are a sample of n independent copies of the multivariate normal vector $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ in $\mathbb{R}^{d_1+d_2}$, with mean zero and covariance matrix $\boldsymbol{\Sigma}$, where

$$\Sigma_{ij} = \Sigma_{ji} = \begin{cases} 1, & i = j, \\ \tau, & i = 1,\ j = 2, \\ \rho, & i = 1,\ j = d_1 + 1, \\ 0, & \text{otherwise}. \end{cases}$$

Here τ characterizes the within-group correlation and we consider (a) $\tau = 0$, (b) $\tau = 0.5$, and (c) $\tau = 0.9$. Independence holds if and only if ρ, a between-group correlation, is zero.

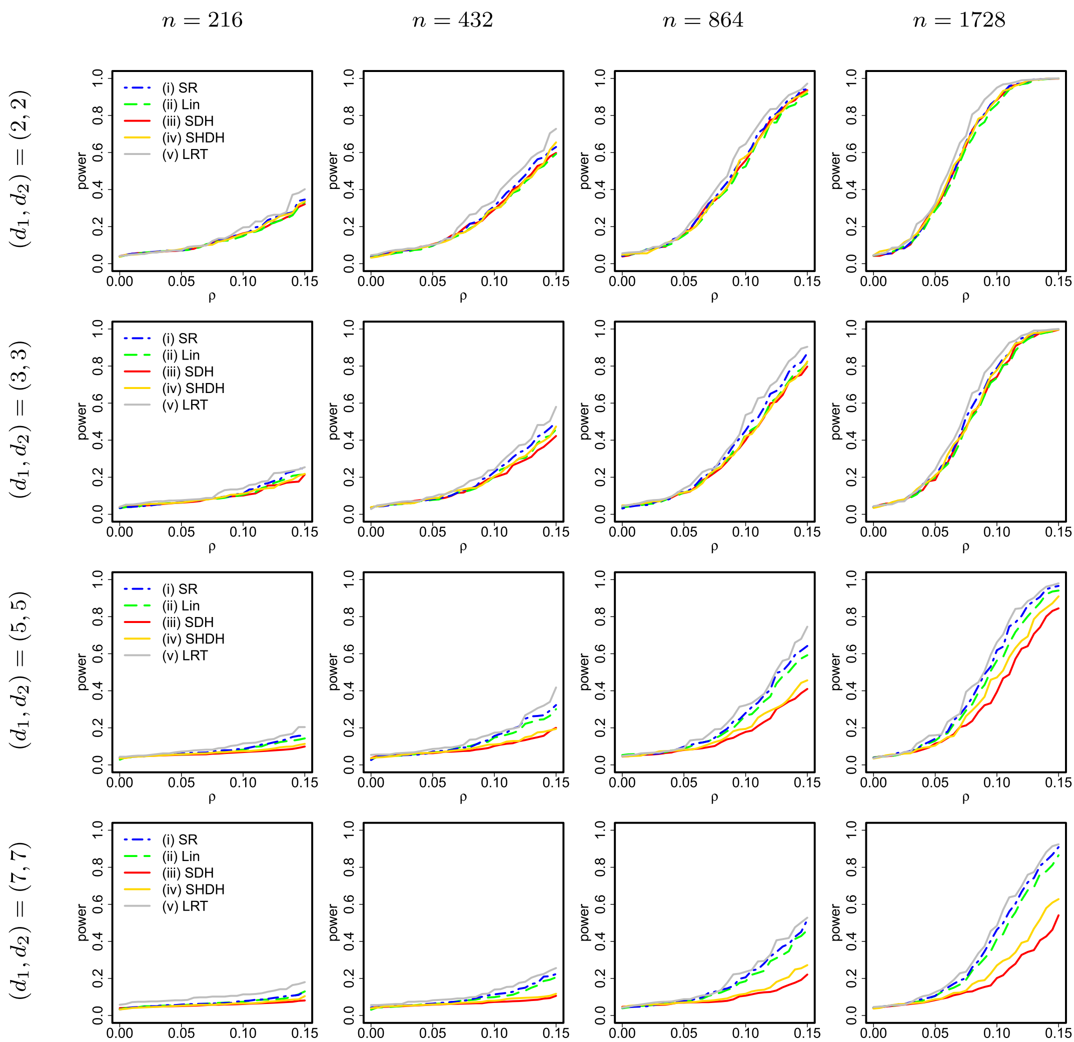

FIG. 1. *Empirical powers of the five competing tests in Example* 5.3(*a*) ($\tau = 0$, *no within-group correlation*). *The* y-*axis represents rejection frequencies based on* 1000 *replicates*, *the* x-*axis represents* ρ (*the between-group correlation*), *and the blue*, *green*, *red*, *gold*, *and grey lines represent the performance of* (*i*) *Szekely and Rizzo's original distance covariance test*, (*ii*) *Lin's marginal rank version of the distance covariance test*, (*iii*) *Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test*, (*iv*) *the center-outward normal-score version of the distance covariance test*, *and* (*v*) *the likelihood ratio test*, *respectively*.

EXAMPLE 5.4. The data are n independent copies of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ with $X_{1i} = Q_{t(1)}(\Phi(X'_{1i}))$ and $X_{2j} = Q_{t(1)}(\Phi(X'_{2j}))$ for $i \in [\![d_1]\!]$ and $j \in [\![d_2]\!]$; here $Q_{t(1)}$ denotes the quantile function of the standard Cauchy distribution and $(\boldsymbol{X}'_1, \boldsymbol{X}'_2)$ is generated according to Example 5.3(b).

We compare the empirical performance of the following five tests:

(i) permutation test using the original distance covariance (Székely and Rizzo (2013));
(ii) permutation test applying original distance covariance to marginal ranks (Lin (2017));
(iii) center-outward rank-based distance covariance test with Wilcoxon scores and critical values from the asymptotic distribution (Shi, Drton and Han (2022a));
(iv) new center-outward rank-based distance covariance test with normal scores and critical values from the asymptotic distribution;

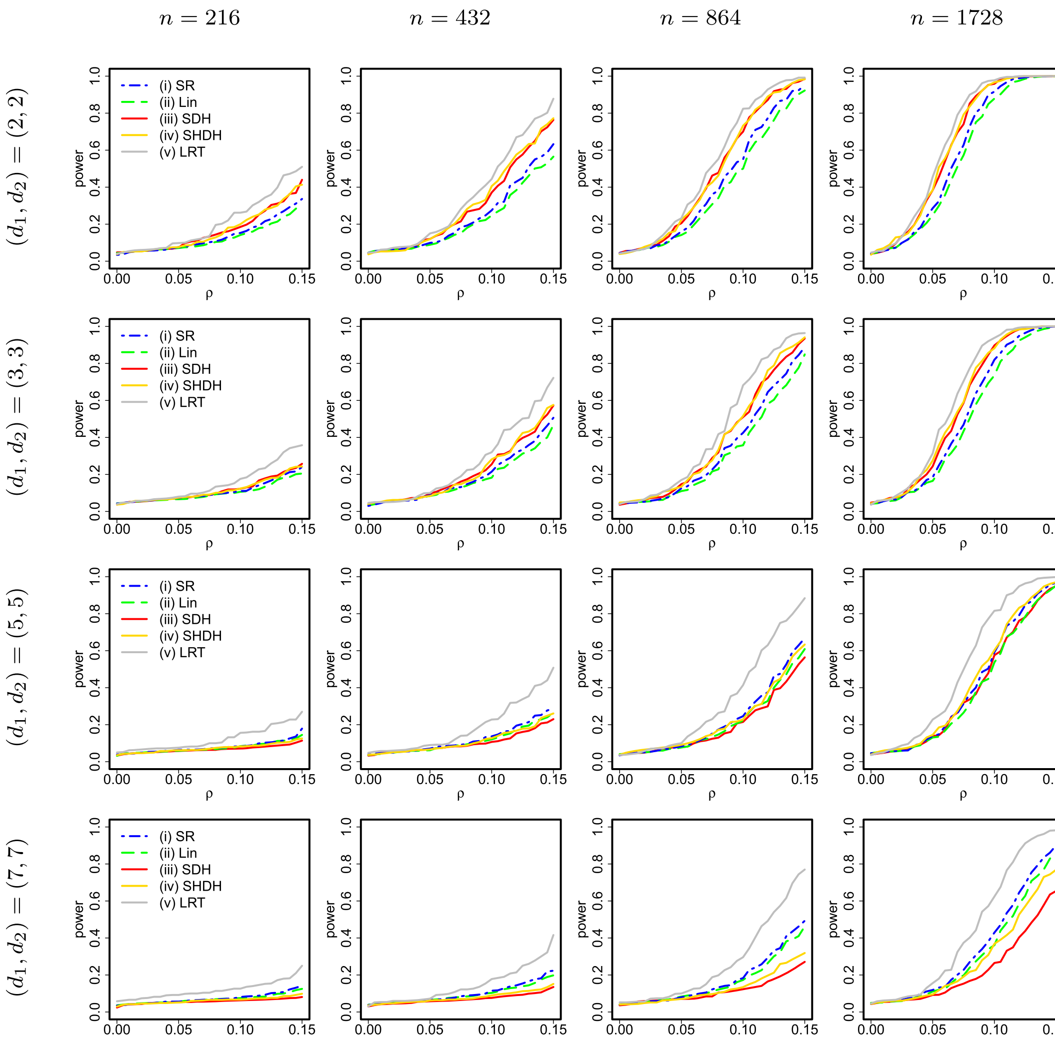

FIG. 2. *Empirical powers of the five competing tests in Example* 5.3(*b*) ($\tau = 0.5$, *moderate within-group correlation*). *The y-axis represents rejection frequencies based on* 1000 *replicates*, *the x-axis represents ρ* (*the between-group correlation*), *and the blue*, *green*, *red*, *gold*, *and grey lines represent the performance of* (*i*) *Szekely and Rizzo's original distance covariance test*, (*ii*) *Lin's marginal rank version of the distance covariance test*, (*iii*) *Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test*, (*iv*) *the center-outward normal-score version of the distance covariance test*, *and* (*v*) *the likelihood ratio test*, *respectively.*

(v) likelihood ratio test in the Gaussian model (Anderson (2003), Chapters 9.3.3 and 8.4.4).

The parametric test (v) is tailored for Gaussian densities and plays the role of a benchmark. Unsurprisingly, in the Gaussian experiments in Figures 1–3, it uniformly outperforms tests (i)-(iv). See Figure 4 for its unsatisfactory performance for non-Gaussian densities.

Figures 1–4 report empirical powers (rejection frequencies) of these five tests, based on 1000 simulations with nominal significance level 0.05, dimensions $d_1 = d_2 \in \{2, 3, 5, 7\}$, and sample size $n \in \{216, 432, 864, 1728\}$. The parameter ρ in the true covariance matrix takes values $\rho \in \{0, 0.005, \dots, 0.15\}$. The critical values for tests (i) and (ii) were computed on the basis of n random permutations. For tests (iii) and (iv), to determine the critical values from the asymptotic distribution given in Corollary 5.1, we numerically compute the eigen-

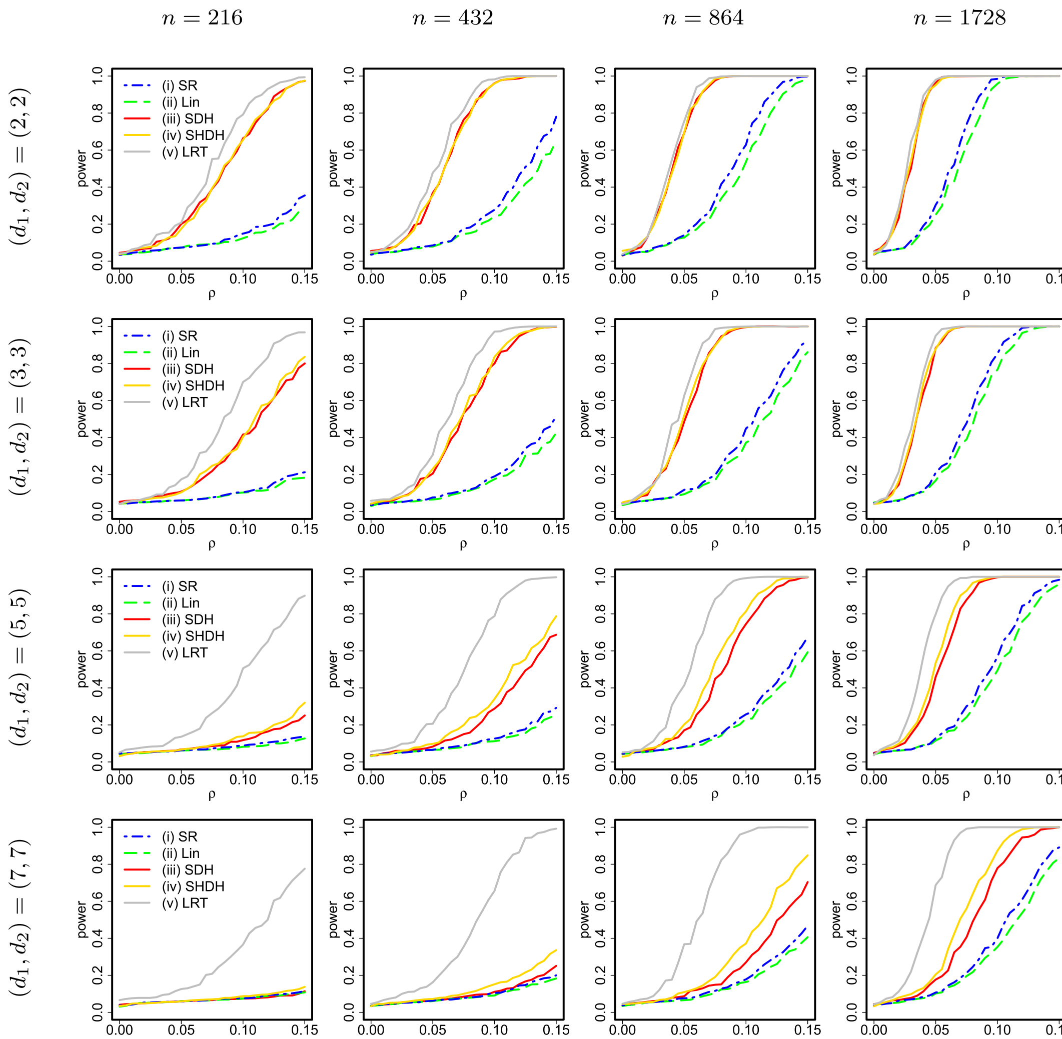

FIG. 3. *Empirical powers of the five competing tests in Example* 5.3(*c*) ($\tau = 0.9$, *high within-group correlation*). *The* y*-axis represents rejection frequencies based on* 1000 *replicates*, *the* x*-axis represents* ρ (*the between-group correlation*), *and the blue*, *green*, *red*, *gold*, *and grey lines represent the performance of* (*i*) *Szekely and Rizzo's original distance covariance test*, (*ii*) *Lin's marginal rank version of the distance covariance test*, (*iii*) *Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test*, (*iv*) *the center-outward normal-score version of the distance covariance test*, *and* (*v*) *the likelihood ratio test*, *respectively*.

values by adopting the same strategy as in Shi, Drton and Han ((2022a), Section 5.2); see also Lyons ((2013), page 3291).

It is evident from Figure 4 that, in non-Gaussian experiments, the potential benefits of rank-based tests are huge, particularly so when Gaussian scores are adopted (note the very severe bias of the Gaussian likelihood ratio test as d increases). In Gaussian experiments, the performance of the normal score–based test (iv) is uniformly better than that of its Wilcoxon score counterpart (iii); that superiority increases with the dimension and decreases with the within-group dependence τ. The superiority of both center-outward rank-based tests (iii) and (iv) over the traditional distance covariance one and its marginal rank version is quite significant for high values of the within-group correlation τ.

The way the normal-score rank-based test (and also the Wilcoxon-score one) outperforms the original distance covariance test may come as a surprise. However, the original distance

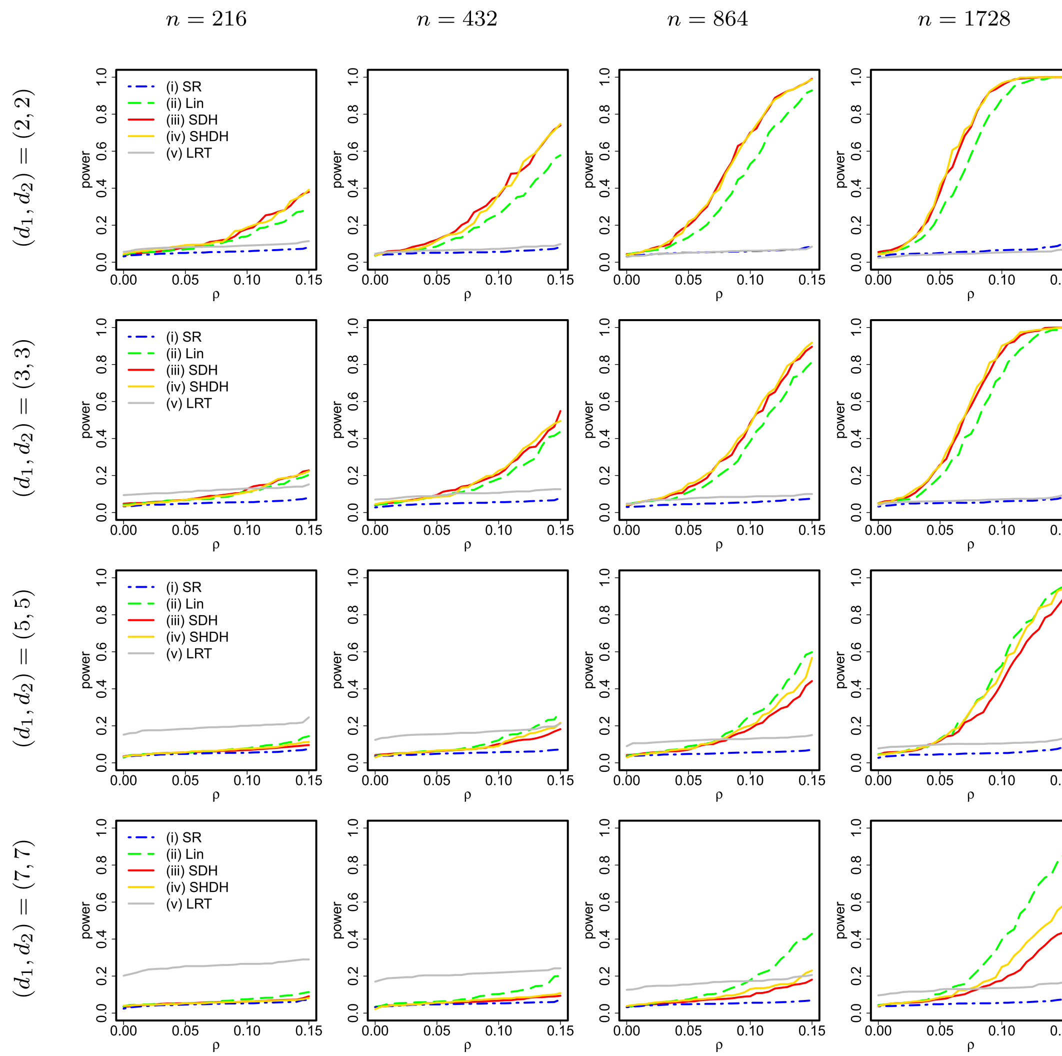

FIG. 4. *Empirical powers of the five competing tests in Example* 5.4. *The* y*-axis represents rejection frequencies based on* 1000 *replicates*, *the* x*-axis represents* ρ (*the between-group correlation*), *and the blue*, *green*, *red*, *gold*, *and grey lines represent the performance of* (*i*) *Szekely and Rizzo's original distance covariance test*, (*ii*) *Lin's marginal rank version of the distance covariance test*, (*iii*) *Shi–Drton–Han's center-outward Wilcoxon version of the distance covariance test*, (*iv*) *the center-outward normal-score version of the distance covariance test*, *and* (*v*) *the likelihood ratio test*, *respectively.*

covariance does not yield a Gaussian parametric test but rather a nonparametric test for which there is no reason to expect superiority over its rank-based versions in Gaussian settings. In a different context, we have long been used to the celebrated Chernoff–Savage phenomenon that normal-score rank statistics may (uniformly) outperform their pseudo-Gaussian counterparts (Chernoff and Savage (1958)). This is best known in the context of two-sample location problems; see, however, Hallin (1994), Hallin and Paindaveine (2008), and Deb, Bhattacharya and Sen (2021) for Chernoff–Savage results for linear time series (traditional univariate ranks and correlogram-based pseudo-Gaussian procedures) and vector independence (Mahalanobis ranks and signs under elliptical symmetry and Wilks' test as the pseudo-Gaussian procedure; measure-transportation-based ranks under elliptical symmetry or independent component assumptions). Although the present context is different, their superiority is another example in which restricting to rank-based methods brings distribution-freeness at no substantial cost in terms of efficiency/power.

6. Conclusion. This paper provides a general framework for specifying dependence measures that leverage the new concept of center-outward ranks and signs. The associated independence tests have the strong appeal of being fully distribution-free. Via the use of a flexible class of generalized symmetric covariances and the incorporation of score functions, our framework allows one to construct a variety of consistent dependence measures. This, as our numerical experiments demonstrate, can lead to significant gains in power.

The theory we develop facilitates the derivation of asymptotic distributions yielding easily computable approximate critical values. The key result is an asymptotic representation that also allows us to establish, for the first time, a nontrivial local power result for tests of vector independence based on center-outward ranks and signs.

Acknowledgments. We acknowledge with thanks the stimulating comments by the Editor, Ming Yuan, the anonymous Associate Editor, and three anonymous referees, which led to a significant improvement of this manuscript.

Funding. The research reported in this manuscript is supported by the United States NSF Grants DMS-1712536 and SES-2019363 and European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 883818).

SUPPLEMENTARY MATERIAL

Supplement to "On universally consistent and fully distribution-free rank tests of vector independence" (DOI: 10.1214/21-AOS2151SUPP; .pdf). This supplement contains all the technical proofs.

REFERENCES

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* 555–564. ACM, New York. MR3238983 https://doi.org/10.1145/2591796.2591844

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

AZADKIA, M. and CHATTERJEE, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49** 3070–3102. MR4352523 https://doi.org/10.1214/21-aos2073

BEIRLANT, J., BUITENDAG, S., DEL BARRIO, E., HALLIN, M. and KAMPER, F. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance Math. Econom.* **95** 79–100. MR4154446 https://doi.org/10.1016/j.insmatheco.2020.08.005

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BERGSMA, W. (2006). A new correlation coefficient, its orthogonal decomposition and associated tests of independence. Preprint. Available at arXiv:math/0604627v1.

BERGSMA, W. (2011). Nonparametric testing of conditional independence by means of the partial copula. Preprint. Available at arXiv:1101.4607v1.

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. MR3178526 https://doi.org/10.3150/13-BEJ514

BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2021). Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.* **49** 2457–2490. MR4338371 https://doi.org/10.1214/20-aos2041

BHATTACHARYA, B. B. (2019). A general asymptotic framework for distribution-free graph-based two-sample tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 575–602. MR3961499

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

CAO, S. and BICKEL, P. J. (2020). Correlations with tailored extremal properties. Preprint. Available at arXiv:2008.10177v2.

CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Stat.* **29** 972–994. MR0100322 https://doi.org/10.1214/aoms/1177706436

CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 https://doi.org/10.1214/16-AOS1450

DEB, N., BHATTACHARYA, B. B. and SEN, B. (2021). Efficiency lower bounds for distribution-free Hotelling-type two-sample tests based on optimal transport. Preprint. Available at arXiv:2104.01986v1.

DEB, N. and SEN, B. (2022). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2021.1923508

DEL BARRIO, E., GONZÁLEZ-SANZ, A. and HALLIN, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.* **180** 104671, 13 pp. MR4147635 https://doi.org/10.1016/j.jmva.2020.104671

DEL BARRIO, E., CUESTA-ALBERTOS, J. A., HALLIN, M. and MATRÁN, C. (2018). Smooth cyclically monotone interpolation and empirical center-outward distribution functions. Preprint. Available at arXiv:1806.01238v1.

DHAR, S. S., DASSIOS, A. and BERGSMA, W. (2016). A study of the power and robustness of a new test for independence against contiguous alternatives. *Electron. J. Stat.* **10** 330–351. MR3466185 https://doi.org/10.1214/16-EJS1107

DRTON, M., HAN, F. and SHI, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.* **48** 3206–3227. MR4185806 https://doi.org/10.1214/19-AOS1926

FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433. MR4205649

FIGALLI, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.* **177** 413–421. MR3886582 https://doi.org/10.1016/j.na.2018.05.008

FUKUMIZU, K., BACH, F. R. and GRETTON, A. (2007). Statistical consistency of kernel canonical correlation analysis. *J. Mach. Learn. Res.* **8** 361–383. MR2320675

GABOW, H. N. and TARJAN, R. E. (1989). Faster scaling algorithms for network problems. *SIAM J. Comput.* **18** 1013–1036. MR1015271 https://doi.org/10.1137/0218069

GHOSAL, P. and SEN, B. (2022). Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *Ann. Statist.* **50** 1012–1037. MR4404927 https://doi.org/10.1214/21-aos2136

GIESER, P. W. (1993). A new nonparametric test for independence between two sets of variates. PhD thesis, University of Florida. Available at https://ufdc.ufl.edu/AA00003658/00001 and https://www.proquest.com/docview/304041219. MR2691505

GIESER, P. W. and RANDLES, R. H. (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc.* **92** 561–567. MR1467849 https://doi.org/10.2307/2965704

GRETTON, A., SMOLA, A., BOUSQUET, O., HERBRICH, R., BELITSKI, A., AUGATH, M., MURAYAMA, Y., PAULS, J., SCHÖLKOPF, B. et al. (2005a). Kernel constrained covariance for dependence measurement. In *AISTATS05* (R. G. Cowell and Z. Ghahramani, eds.) 112–119. Society for Artificial Intelligence and Statistics.

GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005b). Measuring statistical dependence with Hilbert–Schmidt norms. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909 https://doi.org/10.1007/11564089_7

GRETTON, A., HERBRICH, R., SMOLA, A., BOUSQUET, O. and SCHÖLKOPF, B. (2005c). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6** 2075–2129. MR2249882

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York–London; Academia [Publishing House of the Czechoslovak Academy of Sciences], Prague. MR0229351

HALLIN, M. (1994). On the Pitman non-admissibility of correlogram-based methods. *J. Time Series Anal.* **15** 607–611. MR1312324 https://doi.org/10.1111/j.1467-9892.1994.tb00215.x

HALLIN, M. (2017). On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$: A measure transportation approach. Preprint. Available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

HALLIN, M. (2022). Measure transportation and statistical decision theory. *Annu. Rev. Stat. Appl.* **9** 401–424. MR4394914 https://doi.org/10.1146/annurev-statistics-040220-105948

HALLIN, M., HLUBINKA, D. and HUDECOVÁ, Š. (2022). Efficient fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2021.2021921

HALLIN, M., LA VECCHIA, D. and LIU, H. (2022). Center-outward R-estimation for semiparametric VARMA models. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2020.1832501

HALLIN, M., MORDANT, G. and SEGERS, J. (2021). Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat.* **15** 1328–1371. MR4255302 https://doi.org/10.1214/21-ejs1816

HALLIN, M. and PAINDAVEINE, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170 https://doi.org/10.1214/aos/1031689019

HALLIN, M. and PAINDAVEINE, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* **8** 787–815. MR1963662

HALLIN, M. and PAINDAVEINE, D. (2002c). Multivariate signed ranks: Randles' interdirections or Tyler's angles? In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (*Neuchâtel*, 2002). *Stat. Ind. Technol.* 271–282. Birkhäuser, Basel. MR2001322

HALLIN, M. and PAINDAVEINE, D. (2008). Chernoff–Savage and Hodges–Lehmann results for Wilks' test of multivariate independence. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat.* (*IMS*) *Collect.* **1** 184–196. IMS, Beachwood, OH. MR2462206 https://doi.org/10.1214/193940307000000130

HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. and MATRÁN, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Ann. Statist.* **49** 1139–1165. MR4255122 https://doi.org/10.1214/20-aos1996

HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306 https://doi.org/10.1093/biomet/asx050

HAN, F. and LIU, H. (2018). ECA: High-dimensional elliptical component analysis in non-Gaussian distributions. *J. Amer. Statist. Assoc.* **113** 252–268. MR3803462 https://doi.org/10.1080/01621459.2016.1246366

HELLER, R., GORFINE, M. and HELLER, Y. (2012). A class of multivariate distribution-free tests of independence based on graphs. *J. Statist. Plann. Inference* **142** 3097–3106. MR2956796 https://doi.org/10.1016/j.jspi.2012.06.003

HELLER, R. and HELLER, Y. (2016). Multivariate tests of association based on univariate tests. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 208–216. Curran Associates, Red Hook.

HELLER, R., HELLER, Y. and GORFINE, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100** 503–510. MR3068450 https://doi.org/10.1093/biomet/ass070

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93.

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020a). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.* **48** 3417–3441. MR4185814 https://doi.org/10.1214/19-AOS1936

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020b). Supplement to "Robust multivariate nonparametric tests via projection averaging". *Ann. Statist.* **48**. MR4185814 https://doi.org/10.1214/19-AOS1936SUPP

KONIJN, H. S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Stat.* **27** 300–323. MR0079384 https://doi.org/10.1214/aoms/1177728260

KOROLJUK, V. S. and BOROVSKICH, Y. V. (1994). *Theory of U-Statistics. Mathematics and Its Applications* **273**. Kluwer Academic, Dordrecht. MR1472486 https://doi.org/10.1007/978-94-017-3515-5

LE CAM, L. and YANG, G. L. (2000). *Asymptotics in Statistics*: *Some Basic Concepts*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1784901 https://doi.org/10.1007/978-1-4612-1166-2

LEEB, H. and PÖTSCHER, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* **24** 338–376. MR2422862 https://doi.org/10.1017/S0266466608080158

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

LIN, J. (2017). Copula versions of RKHS-based and distance-based criteria. PhD thesis, Pennsylvania State University. Available at https://etda.libraries.psu.edu/catalog/14485jul268. MR4172371

LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260. MR1212489

LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

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 https://doi.org/10.1007/978-1-4419-0468-3

OJA, H., PAINDAVEINE, D. and TASKINEN, S. (2016). Affine-invariant rank tests for multivariate independence in independent component models. *Electron. J. Stat.* **10** 2372–2419. MR3544291 https://doi.org/10.1214/16-EJS1174

PEARSON, K. (1895). Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **58** 240–242.

PINELIS, I. and MOLZON, R. (2016). Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electron. J. Stat.* **10** 1001–1063. MR3486424 https://doi.org/10.1214/16-EJS1133

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York. MR0298844

PURI, M. L., SEN, P. K. and GOKHALE, D. V. (1970). On a class of rank order tests for independence in multivariate distributions. *Sankhyā Ser. A* **32** 271–298. MR0298847

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0595165

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

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* 306–317. ACM, New York. MR3205219

SHI, H., DRTON, M. and HAN, F. (2022a). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* **117** 395–410. MR4399094 https://doi.org/10.1080/01621459.2020.1782223

SHI, H., DRTON, M. and HAN, F. (2022b). On the power of Chatterjee's rank correlation. *Biometrika*. To appear. https://doi.org/10.1093/biomet/asab028

SHI, H., DRTON, M., HALLIN, M. and HAN, F. (2021). Center-outward sign-and rank-based quadrant, Spearman, and Kendall tests for multivariate independence. Preprint. Available at arXiv:2111.15567v1.

SHI, H., HALLIN, M., DRTON, M. and HAN, F. (2022). Supplement to "On universally consistent and fully distribution-free rank tests of vector independence." https://doi.org/10.1214/21-AOS2151SUPP

SPEARMAN, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychol.* **15** 72–101.

SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

TASKINEN, S., KANKAINEN, A. and OJA, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21. MR1965367 https://doi.org/10.1016/S0167-7152(02)00399-1

TASKINEN, S., KANKAINEN, A. and OJA, H. (2004). Rank scores tests of multivariate independence. In *Theory and Applications of Recent Robust Methods*. *Stat. Ind. Technol.* 329–341. Birkhäuser, Basel. MR2088309

TASKINEN, S., OJA, H. and RANDLES, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.* **100** 916–925. MR2201019 https://doi.org/10.1198/016214505000000097

WEIHS, L., DRTON, M. and MEINSHAUSEN, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562. MR3842884 https://doi.org/10.1093/biomet/asy021

WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica* **3** 309–326.

YANAGIMOTO, T. (1970). On measures of association and a related problem. *Ann. Inst. Statist. Math.* **22** 57–63. https://doi.org/10.1007/BF02506323

ZHU, L., XU, K., LI, R. and ZHONG, W. (2017). Projection correlation between two random vectors. *Biometrika* **104** 829–843. MR3737307 https://doi.org/10.1093/biomet/asx043

ZUO, Y. and HE, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.* **34** 2879–2896. MR2329471 https://doi.org/10.1214/009053606000000876