Author(s): Oja, Hannu; Paindaveine, Davy; Taskinen, Sara

Title: Affine-invariant rank tests for multivariate independence in independent component models

Year: 2016

Version: Publisher's PDF

Please cite the original version:

Oja, H., Paindaveine, D., & Taskinen, S. (2016). Affine-invariant rank tests for multivariate independence in independent component models. Electronic Journal of Statistics, 10 (2), 2372-2419. doi:10.1214/16-EJS1174

Electronic Journal of Statistics
Vol. 10 (2016) 2372–2419
ISSN: 1935-7524
DOI: 10.1214/16-EJS1174

Affine-invariant rank tests for multivariate independence in independent component models

Hannu Oja

Department of Mathematics and Statistics
University of Turku, FI-20014 Turku, Finland
e-mail: hannu.oja@utu.fi
url: http://users.utu.fi/hfvoja

Davy Paindaveine

Department of Mathematics and ECARES, Université libre de Bruxelles
Avenue F.D. Roosevelt, 50, CP114/04,1050 Brussels, Belgium
e-mail: dpaindav@ulb.ac.be
url: http://homepages.ulb.ac.be/~dpaindav

and

Sara Taskinen

Department of Mathematics and Statistics
University of Jyvaskyla, P.O.Box 35, FI-40014 Jyvaskyla, Finland
e-mail: sara.l.taskinen@jyu.fi
url: http://users.jyu.fi/~slahola

Abstract: We consider the problem of testing for multivariate independence in independent component (IC) models. Under a symmetry assumption, we develop parametric and nonparametric (signed-rank) tests. Unlike in independent component analysis (ICA), we allow for the singular cases involving more than one Gaussian independent component. The proposed rank tests are based on componentwise signed ranks, à la Puri and Sen. Unlike the Puri and Sen tests, however, our tests (i) are affine-invariant and (ii) are, for adequately chosen scores, locally and asymptotically optimal (in the Le Cam sense) at prespecified densities. Asymptotic local powers and asymptotic relative efficiencies with respect to Wilks' LRT are derived. Finite-sample properties are investigated through a Monte-Carlo study.

MSC 2010 subject classifications: Primary 62G10, 62H15; secondary 62G35.
Keywords and phrases: Distribution-free tests, independent component models, rank tests, singular information matrices, tests for multivariate independence, uniform local asymptotic normality.

Received January 2016.

Contents

1. Introduction

In many sampling and experimental designs, multiple measurements are obtained on each observational unit, resulting in multivariate observation vectors. It is often of interest to explore whether two or several subvectors are interrelated. This typically requires a test of independence between two vectors. To be more specific, let $X_1, \ldots, X_n$ be a sample of i.i.d. observations from a p-variate distribution with cumulative distribution function (cdf) F and write $X_i = (X_i^{(1)\prime}, X_i^{(2)\prime})'$, $i = 1, \ldots, n$, where $X_i^{(1)}$ and $X_i^{(2)}$ are p_1-variate and p_2-variate subvectors, respectively ($p = p_1 + p_2$). We wish to test the null hypothesis

$$\mathcal{H}_0 : \quad X_i^{(1)} \quad \text{and} \quad X_i^{(2)} \text{ are independent.}$$

Denoting by $F^{(\ell)}$ the cdf of $X_i^{(\ell)}$ ($\ell = 1, 2$), this null hypothesis equivalently states that $F(x) = F^{(1)}(x^{(1)})F^{(2)}(x^{(2)})$ for all $x = (x^{(1)\prime}, x^{(2)\prime})'$.

If the observations $X_1, \ldots, X_n$ come from a multinormal distribution with mean μ and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{1.1}$$

the null hypothesis of independence says that the $p_1 \times p_2$ matrix Σ_{12} is zero. The likelihood ratio test for $\mathcal{H}_0 : \Sigma_{12} = 0$ is due to Wilks (1935) and rejects the null at asymptotic level α if and only if

$$W := -n \log \left[\frac{\det S}{(\det S_{11})(\det S_{22})} \right] > \chi^2_{p_1 p_2; 1-\alpha}, \tag{1.2}$$

where the sample covariance matrix S is partitioned as Σ in (1.1) and where $\chi^2_{d;1-\alpha}$ stands for the α-upper quantile of the χ^2_d distribution. Another classical test of independence based on the sample covariance matrix is the Pillai (1955) trace test, which rejects the null (still at asymptotic level α) if and only if

$$P := n \operatorname{tr}\big(S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}\big) = n \, (\operatorname{vec} S_{12})'(S_{22} \otimes S_{11})^{-1} (\operatorname{vec} S_{12}) > \chi^2_{p_1 p_2; 1-\alpha} \tag{1.3}$$

(as usual, $\otimes$ is the Kronecker product and $(\operatorname{vec} A)$ stands for the vector stacking the columns of A on top of each other). If the subvectors are standardized so that

$$\hat{Z}_i^{(1)} := S_{11}^{-1/2} X_i^{(1)} \quad \text{and} \quad \hat{Z}_i^{(2)} := S_{22}^{-1/2} X_i^{(2)}, \quad i = 1, \ldots, n \tag{1.4}$$

(in the sequel, $A^{1/2}$, for a symmetric and positive definite matrix A, stands for the symmetric and positive definite root of A), then we simply have $P = n\|S_{12}(Z)\|^2$, where $S_{12}(Z)$ stands for the statistic S_{12} computed from $\hat{Z}_i = (\hat{Z}_i^{(1)\prime}, \hat{Z}_i^{(2)\prime})'$, $i = 1, \ldots, n$ (throughout, $\|A\| = (\operatorname{tr}(AA'))^{1/2}$ is the Frobenius norm of the matrix A).

It is important to note that Wilks' test and Pillai's test may be used for testing independence even when the normality assumption is not met. This is due to the fact that the covariance matrix functional $\Sigma = \Sigma(F)$ has the so-called *independence property*: if $X_i^{(1)}$ and $X_i^{(2)}$ are independent, then $\Sigma_{12}(F) = 0$. Under $\mathcal{H}_0$ with finite second-order moments, W and P are asymptotically equivalent (that is, $W = P + o_P(1)$ as $n \to \infty$; see (5.1)), so that they admit the same asymptotic null ($\chi^2_{p_1 p_2}$) distribution and share the same asymptotic powers under any sequence of contiguous alternatives; see Section 5.

Note also that both W and P are invariant under the group of block-affine transformations $x \to Ax + b$, associated with any p-vector b and any invertible matrix A of the form $A = \operatorname{diag}(A_{11}, A_{22})$, where $A_{\ell\ell}$ is $p_\ell \times p_\ell$ ($\ell = 1, 2$); throughout, $\operatorname{diag}(B_1, B_2, \ldots, B_m)$ stands for the block-diagonal matrix with diagonal blocks $B_1, B_2, \ldots, B_m$. This *block-affine-invariance* property—in the sequel, we will simply write *affine-invariance*—is natural in cases where the components of $X_i^{(1)}$ and $X_i^{(2)}$ do not have any fixed specified meaning or label, so that the observations (subvectors) could have been taken in another coordinate system as

well. Affine-invariance also ensures distribution-freeness of W and P (under the null) with respect to the variance-covariance structures of both $X_i^{(1)}$ and $X_i^{(2)}$.

Aiming at invariance with respect to componentwise monotone increasing transformations as well (hence, at distribution-freeness and validity under broad conditions, excluding any moment assumption), Puri and Sen (1971) proposed a class of nonparametric tests based on componentwise rankings and componentwise score functions $K_1, \ldots, K_p$ defined over $(0,1)$ (the latter are normalized so that $E[K_r(U)] = 0$ and $E[K_r^2(U)] = 1$ for all r, where U is uniformly distributed over $(0,1)$). As Wilks' statistic in (1.2), the proposed test statistic is

$$W_K := -n \log \left[\frac{\det \tilde{S}}{(\det \tilde{S}_{11})(\det \tilde{S}_{22})} \right],$$

where the $p \times p$ rank-based covariance matrix $\tilde{S}$ (still partitioned as Σ in (1.1) above) has entry (r,s) given by

$$\frac{1}{n} \sum_{i=1}^{n} K_r \left(\frac{R_{ir}}{n+1} \right) K_s \left(\frac{R_{is}}{n+1} \right),$$

where R_{ir} denotes the rank of X_{ir} among $X_{1r}, \ldots, X_{nr}$ (here, X_{ir} stands for the rth component of X_i). The classical sign covariance matrix and Spearman's rho matrix are obtained as special cases (through sign and Wilcoxon score functions, respectively). At the null of independence, W_K, under general assumptions, is asymptotically $\chi^2_{p_1 p_2}$. The resulting tests extend to any dimension the univariate ($p_1 = p_2 = 1$) quadrant test of Blomqvist (1950) (sign score function) and the popular univariate test due to Spearman (1904) (Wilcoxon score function). The test statistic W_K is invariant under monotone transformations of the components of the X_i's, but *is not affine-invariant* (in the sense described above). Most importantly, using componentwise ranks of the standardized subvectors

$$\hat{Z}_i^{(1)} := S_{11}^{-1/2} X_i^{(1)} \quad \text{and} \quad \hat{Z}_i^{(2)} := S_{22}^{-1/2} X_i^{(2)}, \quad i = 1, \ldots, n,$$

instead of the original ones still would *not* provide affine-invariant test statistics.

Affine-invariant nonparametric tests for independence have been developed as well. Gieser and Randles (1997) proposed a simple nonparametric test that generalizes the quadrant test and is based on the Randles (1989) concept of interdirection counts. Taskinen, Kankainen and Oja (2003) proposed a related but more practical affine-invariant extension of the quadrant test based on spatial signs. Later, Taskinen, Oja and Randles (2005) developed invariant tests which are multivariate extensions of the univariate tests due to Kendall (1938) and Spearman (1904). Their tests are based on interdirection counts, spatial signs and spatial ranks, and provide intuitive, practical and robust alternatives to multivariate normal theory methods. To make the test statistics affine-invariant, both subvectors are standardized before spatial signs and ranks are formed. Taskinen, Kankainen and Oja (2004) developed rank score tests based on the spatial signs and the ranks of the lengths of standardized marginal vectors. All these affine-invariant nonparametric tests avoid any moment assumption.

At first sight, the multivariate nonparametric tests listed in the previous paragraph improve on the classical Wilks/Pillai tests (in terms of moment assumptions) and on the Puri and Sen ones (in terms of affine-invariance). However, a closer inspection reveals that they are not invariant (not even asymptotically so) under componentwise monotone increasing transformations, and that their asymptotic properties were derived for *elliptically symmetric* subvectors only. Actually, they are designed for the model

$$\begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \begin{pmatrix} Z_i^{(1)} \\ Z_i^{(2)} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad i = 1, \ldots, n, \tag{1.5}$$

where the independent random vectors $Z_i^{(1)}$ and $Z_i^{(2)}$ have a spherically symmetric distribution (i.e., are such $OZ_i^{(\ell)} \stackrel{\mathcal{D}}{=} Z_i^{(\ell)}$ for any $p_\ell \times p_\ell$ orthogonal matrix O, $\ell = 1, 2$, where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution). The null hypothesis then naturally is $\mathcal{H}_0 : \Lambda_{12} = \Lambda_{21}' = 0$. Note that, in this model, parametric submodels are obtained by fixing a couple of univariate densities only, namely those of $\|Z_i^{(1)}\|$ and $\|Z_i^{(2)}\|$, and that, under the null, the marginals of $X_i^{(1)}$ (as those of $X_i^{(2)}$) do all share the same density type, that is, have—up to location and scale—a common distribution, which is of course a very severe restriction. Finally, note that $X_i^{(1)}$ and $X_i^{(2)}$ in general do not have an elliptical distribution under the alternative.

In this paper, we consider a more flexible and natural model, namely the *independent component (IC) model*, where the p components of $Z_i = (Z_i^{(1)\prime}, Z_i^{(2)\prime})'$ are assumed to be symmetric mutually independent random variables. This model is widely used by engineers in blind source separation problems, and is related to *independent component analysis* (ICA). After some possible permutation of the components of Z_i, the null of independence still is $\mathcal{H}_0 : \Lambda_{12} = \Lambda_{21}' = 0$ in this IC model. But we argue that the latter (i) is more natural (since alternatives to independence belong to the IC model, whereas dependent marginal in the "elliptical" model in (1.5) are not elliptical) and (ii) is also richer (since parametric IC submodels are obtained by fixing p univariate densities, namely those of the marginals of $Z_i^{(1)}$ and $Z_i^{(2)}$, which allows, e.g., for heterogeneous tail weights across marginals). Note that, although the central symmetry assumption on Z_i may seem strong at first sight, it is of course much weaker than the elliptical symmetry assumption required by the affine-invariant nonparametric tests above.

In those IC models, we adopt the same methodology as in Ilmonen and Paindaveine (2011)[1] and define classes of affine-invariant parametric and nonparametric tests of multivariate independence. The nonparametric procedures are based on the componentwise signed ranks of the estimated (in the null model)

[1] As far as hypothesis testing is concerned, Ilmonen and Paindaveine (2011) restricts to the problem of testing the null that the "mixing matrix" (the $p \times p$ matrix in (1.5)) is equal to a given value, which, unlike the problem of testing for multivariate independence we consider in the present work, is mainly of academic interest only.

independent components. Similarly as the affine-invariant nonparametric procedures designed for the elliptical model, the tests we propose do not require any moment assumption. Our tests however have two important advantages over their affine-invariant nonparametric competitors: (i) as explained above, they are defined in a model of dependence that is much more satisfactory than the elliptical one, and (ii) they allow for local and asymptotical optimality (in the Le Cam sense) at prespecified densities (provided that they are based on adequate score functions K_r).

An important issue in the paper will be the singularity arising in IC models when the assumption that *at most one independent component is Gaussian*, is violated. In ICA (that is, in a point estimation context), this standard assumption is made throughout since it essentially guarantees identifiability of Λ; see Section 2 or Ilmonen and Paindaveine (2011). However, in the problem of testing for multivariate independence considered here, this assumption is much too strong, as it would, e.g., rule out the multinormal case. We therefore *must* study carefully the resulting possible singularity of IC models, which, as we will see, has a deep impact on the asymptotic distributions of our optimal tests. To the best of our knowledge, the problem of investigating the nature of this singularity has never been touched in the literature.

The paper is organized as follows. Section 2 describes the IC model under consideration and states its *uniform local asymptotic normality* (ULAN) property. Section 3 exploits this ULAN structure to define optimal *parametric* tests of independence in IC models. Nonparametric (signed-rank) versions of these tests are proposed in Section 4. The properties of the classical Gaussian tests (the Wilks and Pillai tests) in IC models are investigated in Section 5.1, whereas Section 5.2 derives the asymptotic relative efficiencies of our nonparametric tests with respect to those Gaussian competitors. Section 6 discusses the practical implementation of the proposed tests and Section 7 investigates their finite-sample properties through a Monte-Carlo study. Finally, the Appendix collects technical proofs.

2. IC models, ULAN, and multivariate independence

In this section, we define the IC models in which we will test for multivariate independence, and state the ULAN property on which the construction of the proposed optimal tests will be based. We then introduce the problem of testing for multivariate independence in such models.

2.1. IC models and ULAN

Denote by $\mathcal{G}^p$ a subset of the collection of invertible $p \times p$ real matrices Λ obtained by fixing the order and "signs" of columns in some prespecified way, in the sense that, if $\Lambda \in \mathcal{G}^p$, then the only matrix ΛPS that also belongs to $\mathcal{G}^p$ is Λ itself, where P and S are any permutation and sign-change matrices, respectively (i.e., matrices respectively obtained by permuting the columns of the p-dimensional identity matrix I_p or by changing signs of the entries of the

same). For instance, one can let the sign of the entry with largest absolute values in each column be positive, and then order columns in such a way that those largest absolute values form an increasing sequence (in case of ties, one can then base the ordering/signs on the basis of the second largest absolute values, etc.) Now, further denote by $\mathcal{F}$ the collection of probability density functions (pdf's) g (with respect to the Lebesgue measure on $\mathbb{R}^p$) of absolutely continuous random vectors $Z = (Z_1, \ldots, Z_p)'$ whose marginals are (i) mutually independent, (ii) symmetric about the origin ($-Z_r \stackrel{\mathcal{D}}{=} Z_r$ for all r), and (iii) standardized so that $\mathrm{Med}[Z_r^2] = \chi^2_{1;.5}$ for all r. For $g \in \mathcal{F}$, we will often decompose g into $g(z) =: \prod_{r=1}^p g_r(e_r' z)$ in the sequel, where e_r denotes the rth vector of the canonical basis of $\mathbb{R}^p$.

We then throughout assume that the following *independent component (IC) model* holds.

Assumption (A). For some $\mu \in \mathbb{R}^p$, $\Lambda \in \mathcal{G}^p$, and $g \in \mathcal{F}$, the p-variate observations $X_1, \ldots, X_n$ are generated by $X_i = \Lambda Z_i + \mu$, $i = 1, \ldots, n$, where $Z_i = (Z_{i1}, \ldots, Z_{ip})'$, $i = 1, \ldots, n$ are i.i.d. with pdf g.

In the sequel, we denote the corresponding hypothesis by $P^n_{\mu,\Lambda,g}$ or $P^n_{\vartheta,g}$ (where $\vartheta = (\mu', (\mathrm{vec}\,\Lambda)')' \in \Theta = \mathbb{R}^p \times \mathrm{vec}\,\mathcal{G}^p$), and the marginals of Z_i are called the *independent components* (ICs). The location μ is of course a well-defined parameter since it is the unique center of symmetry of the common distribution of the X_i's. Also, it follows from Theis (2004) that, in cases where at most one IC is Gaussian (cases we do not want to restrict to in the sequel; see the comments at the end of this section for more details), the parameters Λ and g are identifiable. Note indeed that, since $\Lambda \in \mathcal{G}^p$ and $g \in \mathcal{F}$, the order, scales, and signs of the ICs are fixed.

For the sake of illustration, we will later consider Gaussian and t-distributed ICs. We say the rth IC is Gaussian (resp., is t_ν, $\nu > 0$) if and only if $g_r(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ (resp., $g_r(z) = c_\nu \sigma_\nu^{-1} (1 + \nu^{-1} \sigma_\nu^{-2} z^2)^{-(\nu+1)/2}$, where c_ν is a normalization constant). Here, σ_ν is such that $\mathrm{Med}[Z_r^2] = \chi^2_{1;.5}$ if Z_r has pdf g_r (note that $\lim_{\nu\to\infty} \sigma_\nu = 1$). When compared to variance-based standardizations, the median-based one we use throughout has the advantage of avoiding any moment assumption (other such standardizations might be adopted, though; see, e.g., Chen and Bickel (2006)).

As the density g remains unspecified in practice, the semiparametric IC model $\mathcal{P}^n = \{P^n_{\vartheta,g} : \vartheta \in \Theta, g \in \mathcal{F}\}$ is to be considered. Proposition 2.1 below, which is crucial for the construction of optimal tests in Section 3, states that most fixed-g parametric submodels $\mathcal{P}^n_g = \{P^n_{\vartheta,g} : \vartheta \in \Theta\}$ are ULAN. More precisely, ULAN requires that the noise density g belongs to the collection $\mathcal{F}_{\mathrm{ULAN}}$ of densities in $\mathcal{F}$ that (i) are absolutely continuous with respect to the Lebesgue measure (in the sequel, we let $\varphi_{g_r} := -\dot{g}_r/g_r$, where $\dot{g}_r$ stands for the a.e.-derivative of g_r) and (ii) have finite second-order moments ($\sigma^2_{g_r} := \int_{-\infty}^{\infty} z^2 g_r(z) dz < \infty$ for all r), finite Fisher information for location ($\mathcal{I}_{g_r} := \int_{-\infty}^{\infty} \varphi^2_{g_r}(z) g_r(z)\, dz < \infty$ for all r), and finite Fisher information for scale ($\mathcal{J}_{g_r} := \int_{-\infty}^{\infty} z^2 \varphi^2_{g_r}(z) g_r(z)\, dz < \infty$ for all r).

For $g \in \mathcal{F}_{\mathrm{ULAN}}$, let $\gamma_{r,s}(g) := \mathcal{I}_{g_r}\sigma^2_{g_s}$, $\mathcal{I}_g := \mathrm{diag}(\mathcal{I}_{g_1},\ldots,\mathcal{I}_{g_p})$, and define the p-variate optimal location score function φ_g by $z = (z_1,\ldots,z_p)' \mapsto \varphi_g(z) = (\varphi_{g_1}(z_1),\ldots,\varphi_{g_p}(z_p))'$. We then have the following ULAN result.

Proposition 2.1. For any $g \in \mathcal{F}_{\mathrm{ULAN}}$, the family of probability distributions $\mathcal{P}^n_g = \{P^n_{\vartheta,g} : \vartheta \in \Theta\}$ is ULAN. More precisely, for any $\vartheta_n = \vartheta + O(n^{-1/2})$, any bounded sequence (τ_n) in $\mathbb{R}^{p(p+1)}$, and any $g \in \mathcal{F}_{\mathrm{ULAN}}$, we have that (i) under $P^n_{\vartheta_n,g}$, as $n \to \infty$,

$$\log\left(dP^n_{\vartheta_n+n^{-1/2}\tau_n,g}/dP^n_{\vartheta_n,g}\right) = \tau_n'\,\Delta^{(n)}_{\vartheta_n,g} - \frac{1}{2}\tau_n'\Gamma_{\vartheta,g}\tau_n + o_P(1),$$

where, letting $Z_i(\vartheta) := \Lambda^{-1}(X_i - \mu)$,

$$\Delta^{(n)}_{\vartheta,g} = \begin{pmatrix} \Delta^{(n)}_{\vartheta,g;1} \\ \Delta^{(n)}_{\vartheta,g;2} \end{pmatrix} = \begin{pmatrix} n^{-1/2}\big(\Lambda^{-1}\big)'\sum_{i=1}^n \varphi_g\big(Z_i(\vartheta)\big) \\ n^{-1/2}\big(I_p\otimes\Lambda^{-1}\big)'\sum_{i=1}^n \mathrm{vec}\left(\varphi_g\big(Z_i(\vartheta)\big)Z_i'(\vartheta) - I_p\right) \end{pmatrix},$$

and that (ii) the *central sequence* $\Delta^{(n)}_{\vartheta_n,g}$, still under $P^n_{\vartheta_n,g}$, is asymptotically normal with mean zero and covariance matrix $\Gamma_{\vartheta,g} = \mathrm{diag}(\Gamma_{\vartheta,g;1},\Gamma_{\vartheta,g;2})$, with $\Gamma_{\vartheta,g;1} = \big(\Lambda^{-1}\big)'\mathcal{I}_g\Lambda^{-1}$ and $\Gamma_{\vartheta,g;2} = \big(I_p\otimes\Lambda^{-1}\big)'\big\{\sum_{r=1}^p(\mathcal{J}_{g_r}-1)\,(e_re_r'\otimes e_re_r') + \sum_{r,s=1,r\neq s}^p\big(\gamma_{s,r}(g)(e_re_r'\otimes e_se_s') + (e_re_s'\otimes e_se_r')\big)\big\}\big(I_p\otimes\Lambda^{-1}\big)$.

Clearly, $\Gamma_{\vartheta,g;1}$ has full rank for all $\vartheta\in\Theta$ and for all $g\in\mathcal{F}_{\mathrm{ULAN}}$ with $\mathcal{I}_{g_r}>0$ for all r. Now,

$$\det\Gamma_{\vartheta,g;2} = (\det\Lambda)^{-2p}\left(\prod_{r=1}^p(\mathcal{J}_{g_r}-1)\right)\left(\prod_{r=1}^{p-1}\prod_{s=r+1}^p(\mathcal{I}_{g_r}\mathcal{I}_{g_s}\sigma^2_{g_r}\sigma^2_{g_s}-1)\right).$$

If $g\in\mathcal{F}_{\mathrm{ULAN}}$, $\mathcal{J}_{g_r}>1$ for all r (see, e.g., Hallin and Paindaveine (2006)). As for the other factors, $\mathcal{I}_{g_r}\sigma^2_{g_r}\geq 1$, where the equality holds if and only if g_r is Gaussian. Hence, *the information matrix $\Gamma_{\vartheta,g}$ is nonsingular if and only if at most one IC is Gaussian.* Actually, it can be shown that if exactly q ICs are Gaussian, then the rank of $\Gamma_{\vartheta,g}$ is $p(p+1)-q(q-1)/2$ (an explicit expression of the Moore-Penrose pseudo-inverse of $\Gamma_{\vartheta,g}$ is readily obtained from (A.3) in the Appendix).

This possible singularity of IC models is well-known. In standard ICA, i.e., in a point estimation context, it is actually assumed that at most one IC is Gaussian, which guarantees identifiability (up to postmultiplication by some permutation, sign-change, and scale matrices) of the mixing matrix Λ to be estimated; see, e.g., Theis (2004). The ULAN result above sheds some light on the nature of this singularity in an asymptotic sense, in terms of Fisher information matrices. In the problem of testing for multivariate independence however (see Section 2.2 below), the parameter Λ need not be fully identified, and having several Gaussian ICs does not hurt (for instance, testing for multivariate independence at the multinormal model, where all ICs are Gaussian, clearly makes sense). Still, a careful treatment of the possible singularity of $\Gamma_{\vartheta,g}$ will be required when studying the asymptotic properties of the proposed tests.

2.2. Multivariate independence in IC models

Assume that the true model is the semiparametric IC model $\mathcal{P}^n$ above. As in the Introduction, we partition the observations into $X_i = (X_i^{(1)\prime}, X_i^{(2)\prime})'$, where $X_i^{(\ell)}$ is a p_ℓ-variate vector ($\ell = 1, 2$), and partition accordingly Z_i, μ, and Λ into $(Z_i^{(1)\prime}, Z_i^{(2)\prime})'$, $(\mu_1', \mu_2')'$, and

$$\begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix},$$

respectively. We consider the problem of testing $\mathcal{H}_0^{\mathrm{IC}} : \Lambda_{12} = \Lambda_{21}' = 0$ (under which the subvectors $X_i^{(1)}$ and $X_i^{(2)}$ are independent) against the alternative that at least one entry in Λ_{12} or Λ_{21} is non-zero. Clearly, the location vector μ, the marginal null mixing matrices $\Lambda_{\ell\ell}$ ($\ell = 1, 2$), and the noise density g are nuisance parameters.

Denoting by $\mathcal{M}(\Omega)$ the vector space that is spanned by the columns of the full-rank $(p + p^2) \times (p + p_1^2 + p_2^2)$ matrix

$$\Omega := \mathrm{diag}\big(\Omega_1, \Omega_2\big) := \mathrm{diag}\Big(I_p, \mathrm{diag}\Big(I_{p_1} \otimes (I_{p_1} \vdots 0_{p_1 \times p_2})', I_{p_2} \otimes (0_{p_2 \times p_1} \vdots I_{p_2})'\Big)\Big), \tag{2.1}$$

the null can be written as $\mathcal{H}_0^{\mathrm{IC}} : \vartheta \in \mathcal{M}(\Omega)(\cap\Theta)$, hence imposes a set of linear constraints on ϑ. This plays an important role in the sequel, as the form of optimal (in the Le Cam sense) tests for linear constraints on the parameters of ULAN models is well-known; see Section 3.1.

Clearly, ϑ is not specified under the null. As we will see in Section 3.1, our parametric tests will be based on a sequence of estimators $\hat{\vartheta}^{(n)} = (\hat{\mu}^{(n)\prime}, (\mathrm{vec}\,\hat{\Lambda}^{(n)})')'$ satisfying Assumption (B) below—the nonparametric ones actually require slightly different estimators, satisfying Assumption (B′); see Section 4.2.

Assumption (B). The sequence $(\hat{\vartheta}^{(n)} = (\mu^{(n)\prime}, (\mathrm{vec}\,\Lambda^{(n)})')', n \in \mathbb{N})$ is

(i) *constrained:* $P_{\vartheta,g}^n\big[\hat{\vartheta}^{(n)} \in \mathcal{M}(\Omega)\big] = 1$ for all n, $\vartheta \in \mathcal{M}(\Omega)$, and $g \in \mathcal{F}$;

(ii) *root-n consistent:* for all $\vartheta \in \mathcal{M}(\Omega)$, $n^{1/2}(\hat{\vartheta}^{(n)} - \vartheta) = O_P(1)$, as $n \to \infty$, under $\bigcup_{g\in\mathcal{F}} P_{\vartheta,g}^n$;

(iii) *affine-equivariant:* for any invertible matrix $A = \mathrm{diag}(A_{11}, A_{22})$, where $A_{\ell\ell}$ is a $p_\ell \times p_\ell$ matrix ($\ell = 1, 2$), and any p-vector b,

$$\hat{\mu}^{(n)}(AX_1 + b, \ldots, AX_n + b) = A\hat{\mu}^{(n)}(X_1, \ldots, X_n) + b$$

and

$$\hat{\Lambda}^{(n)}(AX_1 + b, \ldots, AX_n + b) = (A\hat{\Lambda}^{(n)}(X_1, \ldots, X_n))^\circ,$$

where, for any invertible $p \times p$ matrix B, B° stands for the unique matrix that can be obtained in $\mathcal{G}_p$ by postmultiplying B with permutation and/or sign-change matrices.

(iv) *locally asymptotically discrete:* for all $\vartheta \in \mathcal{M}(\Omega)$, and all $c > 0$, there exists an $M = M(c) > 0$ such that the number of possible values of $\hat{\vartheta}_n$ in balls of the form $\{t \in \mathbb{R}^{p(p+1)} : n^{1/2}\|t - \vartheta\| \leq c\}$ is bounded by M, uniformly as $n \to \infty$.

Assumption (B) is extremely mild; (B(i)) stresses that $\hat{\vartheta}^{(n)}$ should be obtained by fitting the *null* model (typically by running two *separate* ICAs on the subvectors $X_i^{(1)}$ and $X_i^{(2)}$, $i = 1, \ldots, n$). The rate required in (B(ii)) is the regular one in ICA (see, e.g., Chen and Bickel (2006)). The (natural) affine-equivariance of $\hat{\mu}^{(n)}$ and $\hat{\Lambda}^{(n)}$ in Assumption (B(iii)) will guarantee the affine-invariance of the proposed tests. Finally, Assumption (B(iv))—which is needed for the parametric versions of our tests only (compare with Assumption (B′) in Section 4.2)—is a purely technical requirement, with little practical implications (for fixed sample size, any estimator indeed can be considered part of a locally asymptotically discrete sequence). Most importantly, as far as the asymptotic properties of our tests are concerned, it turns out that no best choice of $\hat{\vartheta}^{(n)}$ does exist in the class of estimators satisfying Assumption (B) (or (B′)); we will indeed show that the asymptotic behavior of our tests is not affected by this choice (see however Section 6.1 for a discussion of finite-sample issues and a recommended practical solution).

We end this section with the following important remark: the problem of testing the null $\mathcal{H}_0$, under which $X_i^{(1)}$ and $X_i^{(2)}$ are independent, only *imperfectly* translates, in the semiparametric IC model $\mathcal{P}^n$, into that of testing $\mathcal{H}_0^{\mathrm{IC}} : \Lambda_{12} = \Lambda_{21}' = 0$. Indeed, whereas $\mathcal{H}_0^{\mathrm{IC}} \subset \mathcal{H}_0$ always holds, $\mathcal{H}_0 \subset \mathcal{H}_0^{\mathrm{IC}}$ may fail to hold at some noise densities $g \in \mathcal{F}$; an extreme example is the following: if $Z_i = (Z_i^{(1)\prime}, Z_i^{(2)\prime})'$ is standard multinormal, $X_i = OZ_i + \mu$ has independent subvectors $X_i^{(1)}$ and $X_i^{(2)}$ for *any* orthogonal matrix O and any p-vector μ. More generally, the number of Gaussian ICs in each subvector plays a crucial role in the non-equivalence between $\mathcal{H}_0$ and $\mathcal{H}_0^{\mathrm{IC}}$, which is thus clearly related to the possible singularity of IC models.

Although the primary objective of this paper is to test the null $\mathcal{H}_0$ of multivariate independence, the main focus in the next sections will be on the null $\mathcal{H}_0^{\mathrm{IC}}$, mainly because, as already mentioned, the form of optimal tests for the latter is known in this ULAN setup. This implies, however, that the tests we will derive for $\mathcal{H}_0^{\mathrm{IC}}$ will have to be reevaluated when considered as tests for $\mathcal{H}_0$. As we will see, our tests will also appear as excellent procedures for the null $\mathcal{H}_0$ of interest.

3. Optimal parametric tests

In this section, we build optimal tests for $\mathcal{H}_0^{\mathrm{IC}}$ under the assumption that the underlying noise density g is known to be some fixed $f \in \mathcal{F}_{\mathrm{ULAN}}$ (this highly unrealistic assumption will be relaxed later). We start with a definition of the optimality concept that is considered in the paper and a reference to Le Cam (1986) explaining how to define optimal tests in the present context. We then

provide an explicit expression for the optimal test statistics and derive their asymptotic distributions, both under the null and under sequences of contiguous alternatives.

3.1. Local and asymptotic most stringency in ULAN models

Let $\mathcal{C}_\alpha$ be the collection of α-level tests for some generic testing problem $\mathcal{H}_0$ versus $\mathcal{H}_1$. We say that the test $\phi_\star$ is *most stringent* in $\mathcal{C}_\alpha$ if and only if (i) $\phi_\star \in \mathcal{C}_\alpha$ and

$$\text{(ii)} \qquad \sup_{P\in\mathcal{H}_1} r_{\phi_\star}(P) \le \sup_{P\in\mathcal{H}_1} r_{\phi}(P), \quad \forall \phi \in \mathcal{C}_\alpha,$$

where $r_{\phi_0}(P) := \sup_{\phi\in\mathcal{C}_\alpha} \mathrm{E}_P[\phi] - \mathrm{E}_P[\phi_0]$, $P \in \mathcal{H}_1$ is the so-called *regret* of ϕ_0 at P (of course we regard here ϕ, ϕ_0, and $\phi_\star$ as *test functions* taking values in $[0,1]$). In other words, a test is most stringent at level α if and only if it minimizes the maximum (in $P \in \mathcal{H}_1$) lack of power (at P) with respect to the maximal power that can be achieved (at P) by an α-level test.

Lucien Le Cam showed that when a ULAN result such as that of Proposition 2.1 holds at $g = f$, a *locally and asymptotically most stringent* test (see Le Cam (1986), Section 10.9, for a precise definition of what is meant here by "locally and asymptotically") for a linear null hypothesis of the form $\mathcal{H}_0 : \vartheta \in \mathcal{M}(\Omega)$ in the fixed-f model $\mathcal{P}^n_f = \{P^n_{\vartheta,f} : \vartheta \in \Theta\}$ rejects the null for large values of

$$Q_f := Q^{(n)}_{\hat\vartheta,f} := \big(\Delta^{(n)}_{\hat\vartheta,f}\big)' \Big(\Gamma^-_{\hat\vartheta,f} - \Omega\big(\Omega'\Gamma_{\hat\vartheta,f}\Omega\big)^-\Omega'\Big)\,\Delta^{(n)}_{\hat\vartheta,f}, \tag{3.1}$$

where $\hat\vartheta$ is an estimator satisfying Assumption (B(i)–(ii), (iv)) and where B^- denotes the Moore-Penrose pseudoinverse of B (that is, the unique matrix C such that $BCB = B$, $CBC = C$, $(BC)' = BC$, and $(CB)' = CB$).

3.2. Optimal parametric tests for multivariate independence

It follows from the previous sections that a locally and asymptotically most stringent test for $\mathcal{H}^{\mathrm{IC}}_0$ in the *parametric IC model* $\mathcal{P}^n_f = \{P^n_{\vartheta,f} : \vartheta \in \Theta\}$ rejects the null for large values of the test statistic in (3.1), where $\Delta^{(n)}_{\vartheta,f}$ and $\Gamma_{\vartheta,f}$ are respectively the central sequence and information matrix in Proposition 2.1 and where Ω is the matrix in (2.1).

We now show that Q_f can be rewritten under a simple explicit form, which makes clear why Q_f might detect some possible dependence between $X^{(1)}_i$ and $X^{(2)}_i$. First note that, since both Ω and $\Gamma_{\vartheta,f}$ are block-diagonal and since $\Omega_1 = I_p$,

$$Q_f = \Big(\Delta^{(n)}_{\hat\vartheta,f;2}\Big)' \Big(\Gamma^-_{\hat\vartheta,f;2} - \Omega_2\big(\Omega_2'\Gamma_{\hat\vartheta,f;2}\Omega_2\big)^-\Omega_2'\Big)\,\Delta^{(n)}_{\hat\vartheta,f;2}. \tag{3.2}$$

For obvious reasons, the following result is crucial to obtain explicit expressions for Q_f.

Lemma 3.1. Fix $\vartheta = (\mu', (\operatorname{vec}\Lambda)')' \in \mathcal{M}(\Omega) \cap \Theta$ and $f \in \mathcal{F}_{\mathrm{ULAN}}$, and consider the summation operator

$$\sum_{r<>s} b_{rs} := \sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2}(b_{rs}+b_{sr}), \quad \mathcal{S}_1 := \{1,\ldots,p_1\},\ \mathcal{S}_2 := \{p_1+1,\ldots,p\}.$$

Then

$$\Gamma^{-}_{\vartheta,f;2} - \Omega_2\big(\Omega_2'\Gamma_{\vartheta,f;2}\Omega_2\big)^{-}\Omega_2' = \big(I_p\otimes\Lambda\big)M_f\big(I_p\otimes\Lambda\big)',$$

with $M_f := \sum_{r<>s}\big(\alpha_{r,s}(f)\,(e_re_r'\otimes e_se_s') + \beta_{r,s}(f)(e_re_s'\otimes e_se_r')\big)$, where we let

$$\alpha_{r,s}(f) := \frac{\gamma_{r,s}(f)}{(1+\gamma^2_{r,s}(f))^2} = \frac{1}{4} \quad \text{and} \quad \beta_{r,s}(f) := \frac{\gamma^2_{r,s}(f)}{(1+\gamma^2_{r,s}(f))^2} = \frac{1}{4}$$

if both f_r and f_s are Gaussian, and

$$\alpha_{r,s}(f) := \frac{\gamma_{r,s}(f)}{\gamma_{r,s}(f)\gamma_{s,r}(f)-1} \quad \text{and} \quad \beta_{r,s}(f) := \frac{-1}{\gamma_{r,s}(f)\gamma_{s,r}(f)-1} \tag{3.3}$$

otherwise.

If f_r and f_s are Gaussian (with variance one, since $f\in\mathcal{F}$), $I_{f_r} = 1/\sigma^2_{f_r} = 1 = 1/\sigma^2_{f_s} = I_{f_s}$, which explains that $\alpha_{r,s}(f)$ and $\beta_{r,s}(f)$ then cannot be defined through (3.3). In all other cases, $\gamma_{r,s}(f)\gamma_{s,r}(f) < 1$ (see the comments after Proposition 2.1), so that the quantities in (3.3) are thus well-defined. Applying Lemma 3.1 to (3.2) and writing $\hat{Z}_i$ for $Z_i(\hat{\vartheta})$ straightforwardly yields

$$Q_f = \frac{1}{n}\sum_{i,j=1}^{n}\Big[\operatorname{vec}\Big(\varphi_f\big(\hat{Z}_i\big)\,\hat{Z}_i' - I_p\Big)\Big]' M_f\Big[\operatorname{vec}\Big(\varphi_f\big(\hat{Z}_j\big)\,\hat{Z}_j' - I_p\Big)\Big].$$

Now, since the rows (hence also the columns) $(\ell-1)p+1,\ldots,(\ell-1)p+p_1$ ($\ell\in\mathcal{S}_1$) and $(\ell-1)p+p_1+1,\ldots,(\ell-1)p+p$ ($\ell\in\mathcal{S}_2$) of the symmetric matrix M_f contain zeros only, we obtain that

$$Q_f = T'_{\hat{\vartheta},f}\,M_f\,T_{\hat{\vartheta},f}, \quad \text{with} \quad T_{\vartheta,f} := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\operatorname{vec}_d\Big(\varphi_f\big(Z_i(\vartheta)\big)\,Z_i'(\vartheta)\Big), \tag{3.4}$$

where the vec_d operator is defined by

$$\operatorname{vec}_d\begin{pmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix} := \Big[I_{p^2} - \Omega_2\Omega_2'\Big]\operatorname{vec}\begin{pmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix} = \operatorname{vec}\begin{pmatrix}0 & \Lambda_{12}\\ \Lambda_{21} & 0\end{pmatrix}.$$

As M_f is symmetric and positive semidefinite, Q_f can be interpreted as a squared norm of the *generalized* cross-covariances $n^{-1/2}T_{\hat{\vartheta},f}$ between the estimated residuals $Z_i^{(1)}(\hat{\vartheta})$ and $Z_i^{(2)}(\hat{\vartheta})$. The word "generalized" stresses that the residuals $Z_i(\hat{\vartheta})$ are weighted by φ_f, which allows for achieving (local and asymptotic) optimality at f. Note that, at the multinormal, those generalized

cross-covariances boil down to the standard ones (see (3.6) below). This intuitive interpretation for Q_f makes clear why $\mathcal{H}_0^{\mathrm{IC}}$ should indeed be rejected for large values of Q_f.

Finally, denoting by $G_{\vartheta,f}$ the $p \times p$ matrix defined by $T_{\vartheta,f} = (\mathrm{vec}\, G_{\vartheta,f})$, the structure of M_f implies that

$$\begin{aligned} Q_f &= \sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2} Q_{f;r,s} \\ &:= \sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2} \begin{pmatrix} (G_{\hat{\vartheta},f})_{sr} \\ (G_{\hat{\vartheta},f})_{rs} \end{pmatrix}' \begin{pmatrix} \alpha_{r,s}(f) & \beta_{r,s}(f) \\ \beta_{r,s}(f) & \alpha_{s,r}(f) \end{pmatrix} \begin{pmatrix} (G_{\hat{\vartheta},f})_{sr} \\ (G_{\hat{\vartheta},f})_{rs} \end{pmatrix}. \end{aligned} \tag{3.5}$$

This decomposition of Q_f into a sum of (asymptotically independent) quadratic forms in $((G_{\hat{\vartheta},f})_{sr}, (G_{\hat{\vartheta},f})_{rs})'$ is explained by the asymptotic independence (under $P^n_{\vartheta,f}$) of the pairs $((G_{\vartheta,f})_{sr}, (G_{\vartheta,f})_{rs})'$ (see Proposition 2.1); incidently, note that these pairs would not be asymptotically independent in asymmetric IC models, which would lead to much more complicated test statistics. If both f_r and f_s are Gaussian, then $(G_{\vartheta,f})_{sr} = (G_{\vartheta,f})_{rs}$ and $\alpha_{r,s}(f) = \alpha_{s,r}(f) = \beta_{r,s}(f) = 1/4$, so that

$$Q_{f;r,s} = ((G_{\hat{\vartheta},f})_{rs})^2,$$

which, as we will show, is asymptotically χ^2_1 under the null. In all other cases, $Q_{f;r,s}$ is asymptotically χ^2_2 under the null. This explains that the number of degrees of freedom in the asymptotic null distribution of Q_f is $d(f) := 2p_1p_2 - q_1(f)q_2(f)$, where $q_\ell(f)$ is the number of Gaussian densities in $\{f_r : r \in S_\ell\}$ $(\ell = 1, 2)$.

More precisely, we have the following result, which summarizes the asymptotic properties of the test based on Q_f (below, $\chi^2_d(\delta)$ denotes the noncentral chi-square distribution with d degrees of freedom and noncentrality parameter δ).

Theorem 3.1. Let Assumptions (A) and (B) hold at $g = f \in \mathcal{F}_{\mathrm{ULAN}}$, and denote by $\phi_f = \phi_f^{(n)}$ the test that rejects $\mathcal{H}_0^{\mathrm{IC}}$ as soon as $Q_f > \chi^2_{d(f);1-\alpha}$. Then,

(i) under $\cup_{\vartheta\in\mathcal{M}(\Omega)}\{P^n_{\vartheta,f}\}$, Q_f is asymptotically $\chi^2_{d(f)}$;
(ii) under $P^n_{\vartheta+n^{-1/2}\tau,f}$, with $\vartheta \in \mathcal{M}(\Omega)$ and $\tau \notin \mathcal{M}(\Omega)$ $(\tau = (\tau_1', \tau_2')' \in \mathbb{R}^p \times \mathbb{R}^{p^2})$, Q_f is asymptotically $\chi^2_{d(f)}(\tau_2'(I_p\otimes\Lambda^{-1})'H_f(I_p\otimes\Lambda^{-1})\tau_2)$, with $H_f := \sum_{r<>s}\big(\gamma_{s,r}(f)(e_re_r'\otimes e_se_s') + (e_re_s'\otimes e_se_r')\big)$;
(iii) the sequence of tests $\phi_f^{(n)}$ is locally and asymptotically most stringent, at asymptotic level α, for $\cup_{\vartheta\in\mathcal{M}(\Omega)}\{P^n_{\vartheta,f}\}$ against $\cup_{\vartheta\notin\mathcal{M}(\Omega)}\{P^n_{\vartheta,f}\}$.

It is easy to check that, if $\hat{\vartheta}$ is affine-equivariant (in the sense of Assumption (B(iii)), then Q_f is affine-invariant.

3.3. Two particular cases

Consider the case for which all ICs are Gaussian, that is, the multinormal case ($f = \phi$, say). Then $q_\ell(\phi) = p_\ell$ ($\ell = 1, 2$), $\alpha_{r,s}(\phi) = \beta_{r,s}(\phi) = 1/4$ for all r, s, $\varphi_\phi(z) = z$, and

$$T_{\vartheta,\phi} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{vec}_d\Big(Z_i(\vartheta)\, Z_i'(\vartheta)\Big), \tag{3.6}$$

so that the corresponding test rejects $\mathcal{H}_0^{\mathrm{IC}}$ at asymptotic level α as soon as

$$Q_\phi = T_{\hat\vartheta,\phi}'\, M_\phi\, T_{\hat\vartheta,\phi} = n \left\| \frac{1}{n} \sum_{i=1}^{n} \hat Z_i^{(1)}\, \hat Z_i^{(2)\prime} \right\|^2 > \chi^2_{p_1 p_2; 1-\alpha}. \tag{3.7}$$

This test is valid at the multinormal only. Under finite second-order moments, however, it can be robustified via some standard "studentization", which actually yields—as we show in Section 5—a test that is asymptotically equivalent, under the null (hence also under sequences of contiguous alternatives), to the classical Wilks and Pillai tests of multivariate independence described in the Introduction.

As a second particular case, we consider the test designed to be optimal when the rth IC is t_{ν_r}, $r = 1, \ldots, p$. Denote by ϕ_ν the corresponding p-variate pdf. Assuming that $\nu_r > 2$ for all r (which ensures that $\phi_\nu \in \mathcal{F}_{\mathrm{ULAN}}$), one easily checks that (with obvious notation) $\varphi_{\phi_{\nu,r}}(z) = (\nu_r + 1)z/(\sigma^2_{\nu_r}\nu_r + z^2)$, $\sigma^2_{\phi_{\nu,r}} = \sigma^2_{\nu_r}\nu_r/(\nu_r - 2)$, and $\mathcal{I}_{\phi_{\nu,r}} = (\nu_r + 1)/(\sigma^2_{\nu_r}(\nu_r + 3))$, so that the resulting optimal test statistic is given by

$$\begin{aligned} Q_{\phi_\nu} = \frac{1}{6n} \sum_{r\in\mathcal{S}_1} \sum_{s\in\mathcal{S}_2} \frac{\kappa(\nu_r)\kappa(\nu_s)}{\nu_r^2 + \nu_s^2 + \nu_r + \nu_s - 6} \Bigg\{ & \frac{\nu_r(\nu_r+1)}{(\nu_r-2)(\nu_s+3)} \left[\sum_{i=1}^{n} \frac{\check Z_{ir}\check Z_{is}}{\nu_r + \check Z_{ir}^2}\right]^2 \\ & + \frac{\nu_s(\nu_s+1)}{(\nu_s-2)(\nu_r+3)} \left[\sum_{i=1}^{n} \frac{\check Z_{ir}\check Z_{is}}{\nu_s + \check Z_{is}^2}\right]^2 - 2 \left[\sum_{i=1}^{n} \frac{\check Z_{ir}\check Z_{is}}{\nu_r + \check Z_{ir}^2}\right]\left[\sum_{i=1}^{n} \frac{\check Z_{ir}\check Z_{is}}{\nu_s + \check Z_{is}^2}\right] \Bigg\}, \end{aligned} \tag{3.8}$$

where we let $\check Z_{ir} := \hat Z_{ir}/\sigma_{\nu_r}$ and $\kappa(\nu) := (\nu - 2)(\nu + 1)(\nu + 3)$. Theorem 3.1 shows that, *for all fixed* $\nu_1, \ldots, \nu_p > 2$, the asymptotic null distribution of Q_{ϕ_ν} (at ϕ_ν) is $\chi^2_{2p_1p_2}$. Quite interestingly, if one puts $\nu_1 = \ldots = \nu_p = \nu$ in (3.8) and then lets ν go to infinity, the resulting (a.e.-)limit Q_{ϕ_∞} differs from Q_ϕ in (3.7). Actually, writing $v^{\odot 3} := v \odot v \odot v$, where $\odot$ denotes the Hadamard (i.e., componentwise) product, it can be shown that

$$\begin{aligned} Q_{\phi_\infty} &= Q_\phi + Q_{\mathrm{new}} \\ &:= n \left\| \frac{1}{n} \sum_{i=1}^{n} \hat Z_i^{(1)}\, \hat Z_i^{(2)\prime} \right\|^2 + \frac{n}{12} \left\| \frac{1}{n} \sum_{i=1}^{n} \Big(\hat Z_i^{(1)\odot 3}\, \hat Z_i^{(2)\prime} - \hat Z_i^{(1)} \big(\hat Z_i^{(2)\odot 3}\big)' \Big) \right\|^2 \end{aligned} \tag{3.9}$$

and that the asymptotic null distribution of Q_{ϕ_∞} at the multinormal is still $\chi^2_{2p_1p_2}$ (there, Q_ϕ and Q_{new} asymptotically are independent $\chi^2_{p_1p_2}$ random variables). In Section 5.2, we provide some insight about the relevance of the extra term Q_{new} in Q_{ϕ_∞} and on the reason why this term can safely be dropped without affecting optimality at the multinormal.

4. Optimal signed-rank tests

The main drawback of the parametric tests ϕ_f above is their lack of robustness, as they in general do not meet the asymptotic α-level constraint if the noise density is misspecified. In this section, we robustify those tests by defining asymptotically distribution-free (*signed-rank-based*) counterparts, and investigate the properties of the resulting nonparametric tests.

4.1. Signed ranks and invariance

The signed ranks of the residuals $Z_i(\vartheta) = \Lambda^{-1}(X_i - \mu)$, $i = 1, \ldots, n$ are the quantities $S_i(\vartheta) = (S_{i1}(\vartheta), \ldots, S_{ip}(\vartheta))'$ and $R_i(\vartheta) = (R_{i1}(\vartheta), \ldots, R_{ip}(\vartheta))'$, $i = 1, \ldots, n$, where $S_{ir}(\vartheta) := I_{[Z_{ir}(\vartheta)>0]} - I_{[Z_{ir}(\vartheta)<0]}$ is the sign of $Z_{ir}(\vartheta)$ and where $R_{ir}(\vartheta)$ denotes the rank of $|Z_{ir}(\vartheta)|$ among $|Z_{1r}(\vartheta)|, \ldots, |Z_{nr}(\vartheta)|$ (here, I_A stands for the indicator function of set A). When no ambiguity is possible, we will not stress the dependence in ϑ.

Restricting to signed-rank tests (i.e., to tests that are measurable with respect to the signed ranks of the residuals) is justified by standard invariance arguments, which, in IC models, take the following form. Denote by $\mathcal{H}$ the collection of transformations h of $\mathbb{R}^p$ defined by $h((z_1, \ldots, z_p)') = (h_1(z_1), \ldots, h_p(z_p))'$, where the functions h_r, $r = 1, \ldots, p$ are continuous, odd, and monotone increasing functions that fix $(\chi^2_{1;.5})^{1/2}$ and $+\infty$. For each $\vartheta = (\mu', (\text{vec}\,\Lambda)')' \in \Theta$, consider then the group of componentwise monotone increasing transformations (of $(\mathbb{R}^p)^n$) $\mathcal{G}^\vartheta = \{g^\vartheta_h : h \in \mathcal{H}\}$, defined by

$$g^\vartheta_h(X_1 \ldots, X_n) = (\Lambda h(Z_1(\vartheta)) + \mu, \ldots, \Lambda h(Z_n(\vartheta)) + \mu).$$

It is easy to check that the corresponding maximal invariant is the collection of signed ranks $(S_i(\vartheta), R_i(\vartheta)$, $i = 1, \ldots, n)$.

Now, the null submodel with value ϑ of the parameter—that is, the family $\mathcal{P}^n_\vartheta := \{P^n_{\vartheta,g} : g \in \mathcal{F}\}$, $\vartheta \in \mathcal{M}(\Omega)$—is invariant under $\mathcal{G}^\vartheta$. The invariance principle therefore suggests restricting to tests that are measurable with respect to the corresponding maximal invariants, i.e., restricting to signed-rank tests. Since, moreover, $\mathcal{P}^n_\vartheta$ is generated by $\mathcal{G}^\vartheta$ for each ϑ, signed-rank tests are strictly distribution-free with respect to the noise density g (actually, since ϑ is to be estimated, only *asymptotic* invariance—hence, also *asymptotic* distribution-freeness—will be achieved).

In the discussion above, distribution-freeness is with respect to $g \in \mathcal{F}$. Note that distribution-freeness with respect to $\vartheta (\in \mathcal{M}(\Omega))$ will also follow from invariance arguments; the relevant group of transformations, in this case, is the group of affine transformations defined in the Introduction (see the top of page 2374).

4.2. Componentwise signed-rank statistics

The proposed signed-rank statistics will involve score functions that must satisfy the following assumption.

ASSUMPTION (C). The score functions $K, L : (0,1)^p \to \mathbb{R}^p$ are of the form

$$K((u_1,\dots,u_p)') = (K_1(u_1),\dots,K_p(u_p))',$$

$$L((u_1,\dots,u_p)') = (L_1(u_1),\dots,L_p(u_p))',$$

where $K_r, L_r : (0,1) \to \mathbb{R}$, $r = 1,\dots,p$ (i) are normalized in such a way that $\mathrm{E}[K_r(U)L_r(U)] = 1$, where U is uniformly distributed over $(0,1)$, and (ii) can be expressed as the difference of two continuous monotonically increasing functions ($K = K_{r+} - K_{r+}$ and $L = L_{r+} - L_{r+}$, say) with $\int_0^1 (K_{r\pm}(u))^{2+\delta}\,du < \infty$ and $\int_0^1 (L_{r\pm}(u))^{2+\delta}\,du < \infty$ for some $\delta > 0$.

Assumption (C) is needed for Hájek's classical projection result for linear signed-rank statistics, which actually only requires square-integrability of the scores (see, e.g., Chapter 3 of Puri and Sen (1985)). As we will see in the proof of Lemma 4.2 below, controlling the replacement of ϑ with an estimate however requires the reinforcement of square-integrability into Assumption (C(ii)).

The (K,L)-score version of our tests will be based on the signed-rank statistic

$$\tilde{T}_{\vartheta,K,L} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{vec}_d\Big(\Big(S_i(\vartheta) \odot K\Big(\frac{R_i(\vartheta)}{n+1}\Big)\Big)\Big(S_i(\vartheta) \odot L\Big(\frac{R_i(\vartheta)}{n+1}\Big)\Big)'\Big),$$

which clearly collects signed-rank cross-covariances between the $Z_i^{(1)}(\vartheta)$'s and the $Z_i^{(2)}(\vartheta)$'s. Now, let G_+ be the function defined by $G_+((z_1,\dots,z_p)') = (G_{1+}(z_1),\dots,G_{p+}(z_p))'$, where G_{r+} stands for the cdf of $|Z_{1r}(\vartheta)|$ under $P^n_{\vartheta,g}$. Since g_r is symmetric, we have that $G_{r+}(z) = 2G_r(z) - 1$ (hence, that $G_{r+}^{-1}(u) = G_r^{-1}(\frac{u+1}{2})$), where G_r is the cdf of $Z_{1r}(\vartheta)$ under $P^n_{\vartheta,g}$. We then have the following result.

Lemma 4.1. Let Assumptions (A) and (C) hold. Then, for any $\vartheta \in \Theta$ and $g \in \mathcal{F}$, $\mathrm{E}[\|\tilde{T}_{\vartheta,K,L} - T_{\vartheta,K,L;g}\|^2] = o(1)$ as $n \to \infty$, under $P^n_{\vartheta,g}$, where, writing $|Z_i(\vartheta)| := (|Z_{i1}(\vartheta)|,\dots,|Z_{ip}(\vartheta)|)'$, we let $T_{\vartheta,K,L;g} := n^{-1/2}\sum_{i=1}^n \mathrm{vec}_d\big(\big(S_i \odot K(G_+(|Z_i(\vartheta)|))\big)\big(S_i \odot L(G_+(|Z_i(\vartheta)|))\big)'\big)$.

Again, to obtain proper test statistics, appropriate estimators need be substituted for nuisance parameters. Actually, we will need to consider specific centerings for each set of scores, or, more precisely, statistics of the form

$$\tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{vec}_d\Big(\Big(S_i(\hat{\vartheta}_K) \odot K\Big(\frac{R_i(\hat{\vartheta}_K)}{n+1}\Big)\Big)\Big(S_i(\hat{\vartheta}_L) \odot L\Big(\frac{R_i(\hat{\vartheta}_L)}{n+1}\Big)\Big)'\Big),$$

where the estimators $\hat{\vartheta}_K := \hat{\vartheta}_K^{(n)}$ and $\hat{\vartheta}_L := \hat{\vartheta}_L^{(n)}$ fulfill

ASSUMPTION (B′). The sequences $(\hat{\vartheta}_K^{(n)} = (\hat{\mu}_K^{(n)\prime}, (\mathrm{vec}\,\hat{\Lambda}^{(n)})')', n \in \mathbb{N})$ and $(\hat{\vartheta}_L^{(n)} = (\hat{\mu}_L^{(n)\prime}, (\mathrm{vec}\,\hat{\Lambda}^{(n)})')', n \in \mathbb{N})$ satisfy Assumptions (B(i)–(iii)) and are such that

(i) for all r, $\sum_{i=1}^n S_{ir}(\hat{\vartheta}_K^{(n)}) K_r\big(\frac{R_{ir}(\hat{\vartheta}_K^{(n)})}{n+1}\big)$ and $\sum_{i=1}^n S_{ir}(\hat{\vartheta}_L^{(n)}) L_r\big(\frac{R_{ir}(\hat{\vartheta}_L^{(n)})}{n+1}\big)$ are $o(n)$ in quadratic mean as $n \to \infty$, under $\bigcup_{\vartheta\in\mathcal{M}(\Omega)} \bigcup_{g\in\mathcal{F}} \{P_{\vartheta,g}^n\}$, and

(ii) $\hat{\mu}_K^{(n)}$, $\hat{\mu}_L^{(n)}$, and $\hat{\Lambda}^{(n)}$ are invariant under permutations of the observations.

Assumption (B′(ii)) is extremely mild, but Assumption (B′(i)) may appear quite peculiar. While estimators satisfying Assumptions (B′) are described in Section 6, we point out that the latter assumption could actually be replaced with the same Assumption (B) as for parametric tests, but at the expense of second-order moment assumptions—the replacement of ϑ with a single estimator $\hat{\vartheta}^{(n)} = (\hat{\mu}^{(n)\prime}, (\mathrm{vec}\,\hat{\Lambda}^{(n)})')'$ could indeed then be controlled through appropriate *asymptotic linearity results*, in the same way as in Lemma A.4(ii) (see the Appendix) for the parametric tests, but this would require ULAN, hence finite second-order moments. Since we want to avoid any moment assumption, we rather adopt Assumption (B′), but the considerations above imply that, if finite moments are not an issue, any estimator $\hat{\vartheta}^{(n)}$ satisfying Assumption (B) can then be used in our signed-rank tests.

Jointly with Lemma 4.1, the following result then provides the key for defining distribution-free counterparts to the parametric test ϕ_f introduced in Section 3.

Lemma 4.2. Let Assumptions (A), (B′), and (C) hold. Then for all $\vartheta \in \mathcal{M}(\Omega)$, $\mathrm{E}[\|\tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} - \tilde{T}_{\vartheta,K,L}\|^2] = o(1)$ as $n \to \infty$, under $\cup_{g\in\mathcal{F}}\{P_{\vartheta,g}^n\}$.

4.3. Definition of the proposed tests

It directly follows from the *representation result* in Lemma 4.1 above that, for any $f \in \mathcal{F}_{\mathrm{ULAN}}$, the signed-rank statistic $\tilde{T}_{\vartheta,f} := \tilde{T}_{\vartheta,K_f,L_f}$, with $K = K_f := \varphi_f \circ F_+^{-1}$ and $L = L_f := F_+^{-1}$, is asymptotically equivalent in probability (under $P_{\vartheta,f}^n$) to $T_{\vartheta,f}$ $(= T_{\vartheta,K_f,L_f;f})$. Hence, the test rejecting $\mathcal{H}_0^{\mathrm{IC}}$ at asymptotic level α as soon as

$$\tilde{Q}_{\vartheta,f} = \tilde{T}_{\vartheta,f}'\, M_f\, \tilde{T}_{\vartheta,f} > \chi^2_{d(f);1-\alpha},$$

will inherit, under noise density f, the optimality properties of ϕ_f. However, unlike ϕ_f, this signed-rank test is distribution-free under $\mathcal{H}_0^{\mathrm{IC}}$, hence has asymptotic level α under any noise density $g \in \mathcal{F}$. Of course, the actual test ($\tilde{\phi}_f$, say) is based on $\tilde{Q}_f := \tilde{T}_f'\, M_f\, \tilde{T}_f$, with $\tilde{T}_f := \tilde{T}_{\hat{\vartheta}_{K_f},\hat{\vartheta}_{L_f},K_f,L_f}$, but we will show (see Theorem 4.1) that this estimation of ϑ (i) does not affect optimality at f and (ii) actually only weakens (strict) distribution-freeness into asymptotic distribution-freeness (which is sufficient to ensure asymptotic level α under any noise density $g \in \mathcal{F}$).

More generally, the (K, L)-score version of the proposed signed-rank tests is the test $\tilde{\phi}_{K,L}$ that rejects $\mathcal{H}_0^{\mathrm{IC}}$ for large values of

$$\tilde{Q}_{K,L} := \tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L}'\, M_{K,L}\, \tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L},$$

with $M_{K,L} := \sum_{r<>s} \big(\alpha_{r,s}(K,L)(e_r e_r' \otimes e_s e_s') + \beta_{r,s}(K,L)(e_r e_s' \otimes e_s e_r')\big)$, where, defining $\gamma_{r,s}(K,L) = \mathrm{E}[K_r^2(U)] \times \mathrm{E}[L_s^2(U)]$, we let

$$\alpha_{r,s}(K,L) := \frac{\gamma_{r,s}(K,L)}{\gamma_{r,s}(K,L)\gamma_{s,r}(K,L) - 1}, \quad \beta_{r,s}(K,L) := \frac{-1}{\gamma_{r,s}(K,L)\gamma_{s,r}(K,L) - 1}$$

if $\gamma_{r,s}(K,L)\gamma_{s,r}(K,L) > 1$, and

$$\alpha_{r,s}(K,L) := \frac{\gamma_{r,s}(K,L)}{(1+\gamma_{r,s}^2(K,L))^2} \quad \text{and} \quad \beta_{r,s}(K,L) := \frac{\gamma_{r,s}^2(K,L)}{(1+\gamma_{r,s}^2(K,L))^2}$$

if $\gamma_{r,s}(K,L)\gamma_{s,r}(K,L) = 1$ (which, in view of Assumption (C(i)), is the case if and only if $L_r(u) = K_r(u)/\mathrm{E}[K_r^2(U)]$ and $L_s(u) = K_s(u)/\mathrm{E}[K_s^2(U)]$ for all $u \in (0,1)$). Since $\tilde{\phi}_{K_f,L_f} = \phi_f$, the tests $\tilde{\phi}_{K,L}$ extend the f-score ones defined above.

The nonparametric tests $\tilde{\phi}_{K,L}$ are to be interpreted in the same way as the parametric ones from Section 3: they reject the null of multivariate independence when the norm of some cross-covariance matrix — in this case, the *signed-rank* one $n^{-1/2}\tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L}$ — is too large.

4.4. Some particular cases

Before stating the asymptotic properties of $\tilde{\phi}_{K,L}$, we first consider some particular cases. We start with the important particular case for which there exist λ_r, $r = 1,\ldots,p$ such that $L_r(u) = \lambda_r K_r(u)$ for all $u \in (0,1)$ (note that Assumption (C(i)) then implies that $\lambda_r = 1/\mathrm{E}[K_r^2(U)]$). One can then take $\hat{\vartheta}_K^{(n)} = \hat{\vartheta}_L^{(n)}$, and it can be checked that

$$\tilde{Q}_{K,L} = n\|\tilde{C}_K\|^2, \tag{4.1}$$

where (with obvious notation)

$$\tilde{C}_K = \frac{1}{n}\sum_{i=1}^n \Big(\hat{S}_i^{(1)} \odot \bar{K}^{(1)}\Big(\frac{\hat{R}_i^{(1)}}{n+1}\Big)\Big)\Big(\hat{S}_i^{(2)} \odot \bar{K}^{(2)}\Big(\frac{\hat{R}_i^{(2)}}{n+1}\Big)\Big)', \tag{4.2}$$

where the p-variate score function $\bar{K}$ is obtained from K by replacing K_r with $\bar{K}_r = K_r/(\mathrm{E}[K_r^2(U)])^{1/2}$, $r = 1,\ldots,p$; this shows that $\tilde{Q}_{K,L}$ then has the simple structure of the Puri and Sen (1971) test statistics. However, we point out that the tests based on $\tilde{Q}_{K,L}$, unlike the Puri and Sen ones, are affine-invariant.

Three classical score functions and corresponding tests are of this type. (i) *Sign* tests are obtained for constant score functions ($\bar{K}_r(u) = 1$ for all r). The resulting test statistic is

$$\tilde{Q}_S := n\|\tilde{C}_S\|^2, \text{ with } \tilde{C}_S := \frac{1}{n}\sum_{i=1}^n \hat{S}_i^{(1)}\hat{S}_i^{(2)\prime}. \tag{4.3}$$

(ii) *Wilcoxon*-type tests are associated with linear score functions ($\bar{K}_r(u) = \sqrt{3}u$ for all r) and reject $\mathcal{H}_0^{\mathrm{IC}}$ for large values of

$$\tilde{Q}_W := n\|\tilde{C}_W\|^2, \text{ with } \tilde{C}_W := \frac{3}{n(n+1)^2}\sum_{i=1}^{n}\big(\hat{S}_i^{(1)} \odot \hat{R}_i^{(1)}\big)\big(\hat{S}_i^{(2)} \odot \hat{R}_i^{(2)}\big)'. \quad (4.4)$$

(iii) Gaussian scores—$\bar{K}_r(u) = \Phi_+^{-1}(u) = \Phi^{-1}(\frac{u+1}{2})$ for all r, where Φ stands for the cdf of the standard normal distribution—yield the *van der Waerden* test statistic

$$\tilde{Q}_{\mathrm{vdW}} := n\|\tilde{C}_{\mathrm{vdW}}\|^2, \quad (4.5)$$
$$\text{with } \tilde{C}_{\mathrm{vdW}} := \frac{1}{n}\sum_{i=1}^{n}\Big(\hat{S}_i^{(1)} \odot \Phi_+^{-1}\Big(\frac{\hat{R}_i^{(1)}}{n+1}\Big)\Big)\Big(\hat{S}_i^{(2)} \odot \Phi_+^{-1}\Big(\frac{\hat{R}_i^{(2)}}{n+1}\Big)\Big)'.$$

It is easy to check that $\tilde{Q}_{\mathrm{vdW}}$ coincides with the signed-rank test statistic $\tilde{Q}_\phi$ achieving optimality at the multinormal (that is, that based on the score functions $K_f = \varphi_f \circ F_+^{-1}$ and $L_f = F_+^{-1}$, with $f = \phi$). The resulting van der Waerden test is therefore the distribution-free counterpart to the Gaussian test based on (3.7).

As shown in Theorem 4.1 below, $\tilde{Q}_{\mathrm{vdW}} = \tilde{Q}_\phi$—as $\tilde{Q}_S$ and $\tilde{Q}_W$—is asymptotically $\chi^2_{p_1p_2}$ under the null. *For any other noise density f, however, the nonparametric test statistic achieving optimality at f, namely $\tilde{Q}_f$, gives rise to a larger number of degrees of freedom (and to a more complicated structure than that of (4.1)–(4.2)).* As an example, we consider the nonparametric counterpart $\tilde{Q}_{\phi_\nu}$ to Q_{ϕ_ν} in (3.8), that is, the signed-rank test statistic designed to achieve optimality when the rth IC is t_{ν_r}, $r = 1, \ldots, p$. Letting

$$\mathcal{T}_{ir}^K := \hat{S}_{ir}(\vartheta^{(n)}_{K_{\phi_\nu}})\sqrt{F^{-1}_{1,\nu_r}\Big(\frac{\hat{R}_{ir}(\vartheta^{(n)}_{K_{\phi_\nu}})}{n+1}\Big)} \text{ and } \mathcal{T}_{ir}^L := \hat{S}_{ir}(\vartheta^{(n)}_{L_{\phi_\nu}})\sqrt{F^{-1}_{1,\nu_r}\Big(\frac{\hat{R}_{ir}(\vartheta^{(n)}_{L_{\phi_\nu}})}{n+1}\Big)},$$

where $F_{1,\nu}$ stands for the cdf of the Fisher-Snedecor distribution with 1 and ν degrees of freedom, it is easy to show that

$$\begin{aligned}\tilde{Q}_{\phi_\nu} = \frac{1}{6n}\sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2}\frac{\kappa(\nu_r)\kappa(\nu_s)}{\nu_r^2+\nu_s^2+\nu_r+\nu_s-6}\Bigg\{&\frac{\nu_r(\nu_r+1)}{(\nu_r-2)(\nu_s+3)}\Bigg[\sum_{i=1}^{n}\frac{\mathcal{T}_{ir}^K\mathcal{T}_{is}^L}{\nu_r+(\mathcal{T}_{ir}^K)^2}\Bigg]^2\\ &+\frac{\nu_s(\nu_s+1)}{(\nu_s-2)(\nu_r+3)}\Bigg[\sum_{i=1}^{n}\frac{\mathcal{T}_{ir}^L\mathcal{T}_{is}^K}{\nu_s+(\mathcal{T}_{is}^K)^2}\Bigg]^2\\ &-2\Bigg[\sum_{i=1}^{n}\frac{\mathcal{T}_{ir}^K\mathcal{T}_{is}^L}{\nu_r+(\mathcal{T}_{ir}^K)^2}\Bigg]\Bigg[\sum_{i=1}^{n}\frac{\mathcal{T}_{ir}^L\mathcal{T}_{is}^K}{\nu_s+(\mathcal{T}_{is}^K)^2}\Bigg]\Bigg\},\end{aligned} \quad (4.6)$$

which (see again Theorem 4.1 below) is asymptotically $\chi^2_{2p_1p_2}$ under $\mathcal{H}_0^{\mathrm{IC}}$ (irrespective of the underlying noise density $g \in \mathcal{F}$).

4.5. Asymptotic properties of the proposed signed-rank tests

Lemma 4.1 implies that, at noise density g, $\tilde{T}_{\vartheta,K,L}$ and $T_{\vartheta,K,L;g}$ have the same asymptotic behavior under the null, hence also under sequences of contiguous alternatives. Unlike $\tilde{T}_{\vartheta,K,L}$, the random variable $T_{\vartheta,K,L;g}$ is a sum of i.i.d. terms, hence can be studied easily. Defining

$$H_{K,L} := \sum_{r<>s} \Big(\gamma_{s,r}(K,L)(e_r e_r' \otimes e_s e_s') + (e_r e_s' \otimes e_s e_r')\Big)$$

and

$$H_{K,L;g} := \sum_{r<>s} \Big(\delta_{s,r}(K,L;g)(e_r e_r' \otimes e_s e_s') + \delta_{r,s}(L,K;g)(e_r e_s' \otimes e_s e_r')\Big), \quad (4.7)$$

where we let $\delta_{r,s}(K,L;g) = \mathrm{E}[K_r(U)\varphi_{g_r}(G_{r+}^{-1}(U))] \times \mathrm{E}[L_s(U)G_{s+}^{-1}(U)]$, standard arguments (mainly Le Cam's third lemma) then yield the following lemma (see the Appendix for the proof).

Lemma 4.3. Let Assumptions (A) and (C) hold. Then, for any $\vartheta \in \Theta$, $\tilde{T}_{\vartheta,K,L}$ is asymptotically normal with mean zero and mean $H_{K,L;g}(I_p \otimes \Lambda^{-1})\tau_2$ under $P_{\vartheta,g}^n$ ($g \in \mathcal{F}$) and under $P_{\vartheta+n^{-1/2}\tau,g}^n$ ($\tau = (\tau_1', \tau_2')' \in \mathbb{R}^p \times \mathbb{R}^{p^2}$, $g \in \mathcal{F}_{\mathrm{ULAN}}$), respectively, and covariance matrix $H_{K,L}$ under both.

Now, let $d(K,L) := 2p_1p_2 - q_1(K,L)q_2(K,L)$, where $q_\ell(K,L)$, $\ell = 1,2$ is the number of indices $r \in \mathcal{S}_\ell$ (see Lemma 3.1 for the definition of $\mathcal{S}_\ell$) such that $L_r = K_r/\mathrm{E}[K_r^2(U)]$. We can then state the main theorem of this paper.

Theorem 4.1. Let Assumptions (A), (B′), and (C) hold, and denote by $\tilde{\phi}_{K,L} = \tilde{\phi}_{K,L}^{(n)}$ the test that rejects $\mathcal{H}_0^{\mathrm{IC}}$ as soon as $\tilde{Q}_{K,L} > \chi^2_{d(K,L);1-\alpha}$. Then,

(i) under $\cup_{\vartheta\in\mathcal{M}(\Omega)} \cup_{g\in\mathcal{F}} \{P_{\vartheta,g}^n\}$, $\tilde{Q}_{K,L}$ is asymptotically $\chi^2_{d(K,L)}$;

(ii) under $P_{\vartheta+n^{-1/2}\tau,g}^n$, with $\vartheta \in \mathcal{M}(\Omega)$, $\tau = (\tau_1', \tau_2')' \notin \mathcal{M}(\Omega)$, and $g \in \mathcal{F}_{\mathrm{ULAN}}$, $\tilde{Q}_{K,L}$ is asymptotically $\chi^2_{d(K,L)}(\tau_2'(I_p\otimes\Lambda^{-1})'H_{K,L;g}'M_{K,L}H_{K,L;g}(I_p\otimes\Lambda^{-1})\tau_2)$;

(iii) for any $f \in \mathcal{F}_{\mathrm{ULAN}}$, the sequence of tests $\tilde{\phi}_f^{(n)} = \tilde{\phi}_{K_f,L_f}^{(n)}$ is locally and asymptotically most stringent, at asymptotic level α, for $\cup_{\vartheta\in\mathcal{M}(\Omega)} \cup_{g\in\mathcal{F}} \{P_{\vartheta,g}^n\}$ against $\cup_{\vartheta\notin\mathcal{M}(\Omega)}\{P_{\vartheta,f}^n\}$.

Three comments are in order. First, Part (i) of the result confirms that the asymptotic null distribution of the proposed nonparametric test statistics only depends on the adopted score functions K and L; the resulting asymptotic distribution-freeness in particular is not affected by the (typically unknown) number of underlying Gaussian marginals. Second, since $H_{K_f,L_f;f} = H_{K_f,L_f} = H_f$ actually is the Moore-Penrose pseudoinverse of $M_{K_f,L_f} = M_f$ (see Lemma A.6 in the Appendix), the local asymptotic powers of $\tilde{\phi}_f$ under noise density f coincide with those of ϕ_f in Theorem 3.1 (as expected, since both tests share the same local and asymptotic optimality properties at f). Third, we stress once more that, unlike the Puri and Sen (1971) tests, our signed-rank tests—when based on affine-equivariant estimators in the sense of Assumption (B(iii))—are affine-invariant.

5. Wilks' and Pillai's tests in IC models and AREs

As mentioned in the Introduction, the most classical parametric tests for multivariate independence are the Wilks (1935) and Pillai (1955) tests in (1.2) and (1.3), respectively. In this section, we first investigate the asymptotic properties of these tests in IC models, and then evaluate the performances of our signed-rank tests by computing their asymptotic relative efficiencies (AREs) with respect to those classical benchmarks.

5.1. Wilks' and Pillai's tests in IC models

Writing $S = I_p + (S - I_p)$ and $S_{\ell\ell} = I_p + (S_{\ell\ell} - I_p)$, $\ell = 1, 2$, in (1.2) and performing a Taylor expansion, it can be shown that (we use the same notation as in the Introduction)

$$W = P + o_P(1) = n \left(\operatorname{vec} S_{12}\right)' \left(\Sigma_{22} \otimes \Sigma_{11}\right)^{-1} \left(\operatorname{vec} S_{12}\right) + o_P(1) \tag{5.1}$$

as $n \to \infty$, under any null distribution with finite second-order moments, where $\Sigma = \operatorname{diag}(\Sigma_{11}, \Sigma_{22})$ stands for the common population covariance matrix of the X_i's. For any g in the collection $\mathcal{F}_2$ of noise densities in $\mathcal{F}$ with finite second-order moments, define $S_g := \operatorname{diag}(\sigma^2_{g_1}, \ldots, \sigma^2_{g_p})$. In the IC model under consideration, Σ, at $P^n_{\vartheta,g}$, with $\vartheta \in \mathcal{M}(\Omega)$ and $g \in \mathcal{F}_2$, is given by $\Sigma_{\vartheta,g} = \operatorname{diag}(\Sigma_{\vartheta,g;11}, \Sigma_{\vartheta,g;22}) = \Lambda S_g \Lambda' = \operatorname{diag}(\Lambda_{11} S^{(1)}_g \Lambda'_{11}, \Lambda_{22} S^{(2)}_g \Lambda'_{22})$ (with obvious notation). Direct computations yield that, under $P^n_{\vartheta,g}$, still with $\vartheta \in \mathcal{M}(\Omega)$ and $g \in \mathcal{F}_2$, hence also under sequences of contiguous alternatives,

$$\begin{aligned} W &= n \left\| (S^{(2)}_g \otimes S^{(1)}_g)^{-1/2} \frac{1}{n} \sum_{i=1}^n \operatorname{vec}\Big(Z^{(1)}_i(\vartheta)\, Z^{(2)\prime}_i(\vartheta) \Big) \right\|^2 + o_P(1) \\ &= n \left\| (S^{(1)}_g)^{-1/2} \Big(\frac{1}{n} \sum_{i=1}^n Z^{(1)}_i(\vartheta)\, Z^{(2)\prime}_i(\vartheta) \Big) (S^{(2)}_g)^{-1/2} \right\|^2 + o_P(1) \qquad (5.2) \\ &= T'_{\vartheta,\phi,g}\, M_\phi\, T_{\vartheta,\phi,g} + o_P(1) \qquad (5.3) \end{aligned}$$

as $n \to \infty$, where we let $T_{\vartheta,\phi,g} := n^{-1/2} \sum_{i=1}^n \operatorname{vec}_d\big(S_g^{-1/2} Z_i(\vartheta)\, Z'_i(\vartheta)\, S_g^{-1/2} \big)$. The fact that W is equal to (5.2) now makes clear why Wilks' test can be regarded as a robustified version of the parametric Gaussian test in (3.7). The following result summarizes the asymptotic properties of $T_{\vartheta,\phi,g}$.

Lemma 5.1. Let Assumption (A) hold, and define $H_{\phi;g} := \sum_{r<>s} \big(\frac{\sigma_{g_r}}{\sigma_{g_s}} (e_r e'_r \otimes e_s e'_s) + \frac{\sigma_{g_s}}{\sigma_{g_r}} (e_r e'_s \otimes e_s e'_r) \big)$ for any $g \in \mathcal{F}_2$. Then, for any $\vartheta \in \Theta$, $T_{\vartheta,\phi,g}$ is asymptotically normal with mean zero and mean $H_{\phi;g}(I_p \otimes \Lambda^{-1})\tau_2$ under $P^n_{\vartheta,g}$ ($g \in \mathcal{F}_2$) and under $P^n_{\vartheta + n^{-1/2}\tau, g}$ ($\tau = (\tau'_1, \tau'_2)' \in \mathbb{R}^p \times \mathbb{R}^{p^2}$, $g \in \mathcal{F}_{\mathrm{ULAN}}$), respectively, and covariance matrix H_ϕ under both.

The asymptotic properties of Wilks' test (hence, also of Pillai's; see (5.1)) in IC models easily follow from (5.3) and Lemma 5.1.

Theorem 5.1. Let Assumption (A) hold, and denote by $\phi_{\text{Wilks}} = \phi^{(n)}_{\text{Wilks}}$ the test that rejects $\mathcal{H}_0^{\text{IC}}$ as soon as $W > \chi^2_{p_1p_2;1-\alpha}$. Then,

(i) under $\cup_{\vartheta\in\mathcal{M}(\Omega)} \cup_{g\in\mathcal{F}_2} \{P^n_{\vartheta,g}\}$, W is asymptotically $\chi^2_{p_1p_2}$;
(ii) under $P^n_{\vartheta+n^{-1/2}\tau,g}$, with $\vartheta \in \mathcal{M}(\Omega)$, $\tau = (\tau_1', \tau_2')' \notin \mathcal{M}(\Omega)$, and $g \in \mathcal{F}_{\text{ULAN}}$, W is asymptotically $\chi^2_{p_1p_2}(\tau_2'(I_p \otimes \Lambda^{-1})' H'_{\phi;g} M_\phi H_{\phi;g}(I_p \otimes \Lambda^{-1})\tau_2)$;
(iii) the sequence of tests $\phi^{(n)}_{\text{Wilks}}$ is locally and asymptotically most stringent, at asymptotic level α, for $\cup_{\vartheta\in\mathcal{M}(\Omega)} \cup_{g\in\mathcal{F}_2} \{P^n_{\vartheta,g}\}$ against $\cup_{\vartheta\notin\mathcal{M}(\Omega)} \{P^n_{\vartheta,\phi}\}$.

This also shows that Wilks' test does not require finite fourth-order moments (as it is often stated), but second-order ones only; this follows from (5.1) and the fact that, unlike $\sqrt{n}\,\text{vec}(S-\Sigma)$, $\sqrt{n}(\text{vec}\, S_{12})$ is asymptotically normal under the null as soon as the common distribution of the X_i's has a finite covariance matrix.

5.2. Asymptotic relative efficiencies

We here compare the performances of the proposed signed-rank tests $\tilde{\phi}_{K,L}$ with Wilks' (equivalently, with Pillai's) through asymptotic relative efficiencies (AREs). If the score functions K, L are such that $d(K,L) = p_1p_2$, these AREs are simply obtained from Theorems 4.1 and 5.1 by computing the ratios of the noncentrality parameters in the asymptotic non-null distributions of $\tilde{\phi}_{K,L}$ and ϕ_{Wilks}. If $d(K,L) > p_1p_2$, however, the degrees of freedom in the limiting distributions of $\tilde{\phi}_{K,L}$ and ϕ_{Wilks} do not match and a direct use of the ratio of the noncentrality parameters is no longer valid. We then use the extension of the concept of Pitman ARE to cases where the limiting distributions of the competing tests are of different types; see Nyblom and Mäkeläinen (1983) and Möttönen, Hüsler and Oja (2003). In such a case, the resulting relative efficiency may depend on the common asymptotic level α and power β of the tests. The general result is the following.

Proposition 5.1. Let Assumptions (A), (B′), and (C) hold, and fix $\vartheta \in \mathcal{M}(\Omega)$, $\tau = (\tau_1', \tau_2')' \notin \mathcal{M}(\Omega)$, and $g \in \mathcal{F}_{\text{ULAN}}$. Then, when testing $P^n_{\vartheta,g}$ against $P^n_{\vartheta+n^{-1/2}\tau,g}$, the ARE of $\tilde{\phi}_{K,L}$ with respect to ϕ_{Wilks} is given by

$$\begin{aligned}&\text{ARE}_{\vartheta,\tau,g}\big[\tilde{\phi}_{K,L}/\phi_{\text{Wilks}}\big]\\ &\quad= \frac{c^{\alpha,\beta}_{p_1p_2}}{c^{\alpha,\beta}_{d(K,L)}} \frac{\tau_2'(I_p \otimes \Lambda^{-1})' H'_{K,L;g} M_{K,L} H_{K,L;g}(I_p \otimes \Lambda^{-1})\tau_2}{\tau_2'(I_p \otimes \Lambda^{-1})' H'_{\phi;g} M_\phi H_{\phi;g}(I_p \otimes \Lambda^{-1})\tau_2},\end{aligned} \tag{5.4}$$

where α (resp., β) is the common asymptotic size (resp., power) of the tests and where $c^{\alpha,\beta}_k$ is the value of the noncentrality parameter δ such that the β upper-quantile of the $\chi^2_k(\delta)$ distribution is $\chi^2_{k;1-\alpha}$.

To present some numerical values of these AREs, we consider the generic bivariate case ($p_1 = p_2 = 1$) for which $\vartheta = (\mu', (\text{vec}\,\Lambda)')'$ and $\tau = (t', (\text{vec}\, D)')'$,

TABLE 1

AREs with respect to Wilks' test, under type 1 alternatives (see (5.5)) with various error densities of the form $z = (z_1, z_2)' \mapsto g(z) = g_1(z_1)g_2(z_2)$, of the sign ($\tilde{\phi}_S$), Wilcoxon ($\tilde{\phi}_W$), van der Waerden ($\tilde{\phi}_{\text{vdW}}$) signed-rank tests, and of the signed-rank tests ($\tilde{\phi}_{t_3}$, $\tilde{\phi}_{t_5}$, $\tilde{\phi}_{t_8}$) achieving optimality when both ICs are t_ν, with $\nu = 3, 5$, and 8, respectively.

		g_2			
	g_1	t_3	t_5	t_8	$\mathcal{N}$
$\tilde{\phi}_S$	t_3	0.657	0.608	0.595	0.581
	t_5		0.519	0.495	0.468
	t_8			0.467	0.438
	$\mathcal{N}$				0.405
$\tilde{\phi}_W$	t_3	1.299	1.211	1.186	1.161
	t_5		1.070	1.032	0.996
	t_8			0.992	0.953
	$\mathcal{N}$				0.912
$\tilde{\phi}_{\text{vdW}}$	t_3	1.357	1.245	1.216	1.199
	t_5		1.106	1.072	1.053
	t_8			1.038	1.019
	$\mathcal{N}$				1.000
$\tilde{\phi}_{t_3}$	t_3	1.222	1.043	0.993	0.953
	t_5		0.872	0.828	0.798
	t_8			0.787	0.761
	$\mathcal{N}$				0.737
$\tilde{\phi}_{t_5}$	t_3	1.163	1.051	1.020	0.997
	t_5		0.916	0.882	0.859
	t_8			0.848	0.825
	$\mathcal{N}$				0.801
$\tilde{\phi}_{t_8}$	t_3	1.130	1.035	1.010	0.993
	t_5		0.913	0.883	0.863
	t_8			0.852	0.831
	$\mathcal{N}$				0.809

with arbitrary $\mu, t \in \mathbb{R}^2$, $\Lambda = I_2$, and either

$$D = \begin{pmatrix} 0 & \sigma_{g_1}/\sigma_{g_2} \\ \sigma_{g_2}/\sigma_{g_1} & 0 \end{pmatrix} \qquad \text{(type 1 alternatives)} \tag{5.5}$$

or

$$D = \begin{pmatrix} 0 & \sigma_{g_1}/\sigma_{g_2} \\ -\sigma_{g_2}/\sigma_{g_1} & 0 \end{pmatrix} \qquad \text{(type 2 alternatives);} \tag{5.6}$$

the rescaling in D ensures that the resulting AREs do not depend on the variance of the ICs. Table 1 provides, for type 1 alternative sequences, numerical values of the AREs, with respect to Wilks' test and under various distributions of the ICs, for the following six signed-rank tests: (i) the sign test $\tilde{\phi}_S$ based on (4.3), (ii) the Wilcoxon test $\tilde{\phi}_W$ based on (4.4), (iii) the van der Waerden test $\tilde{\phi}_{\text{vdW}}$ based on (4.5), and (iv)-(vi) the tests $\tilde{\phi}_{t_\nu}$, $\nu = 3, 5, 8$ based on (4.6) with $\nu_1 = \nu_2 (=: \nu)$, which achieve optimality when both ICs are t_ν. These last three tests do not have the same limiting null distribution as Wilks' (under the null, their statistics are asymptotically χ^2_2, whereas Wilks' is asymptotically χ^2_1); their AREs with respect to Wilks' test were then evaluated from Proposition 5.1 with $\alpha = 5\%$ and $\beta = 80\%$.

Table 1 shows that, while the sign test is poorly efficient, the Wilcoxon and van der Waerden tests behave uniformly very well. This is actually an illustration of the following general result.

Proposition 5.2. Let Assumptions (A), (B′), and (C) hold. Consider a signed-rank test $\tilde{\phi}_{K,L}$ with $d(K,L) = p_1 p_2$ (equivalently, with $L_r = K_r/\mathrm{E}[K_r^2(U)]$ for all $r = 1, \ldots, p$), and fix $\vartheta =: (\mu', (\mathrm{vec}\,\Lambda)')'$, $\tau =: (t', (\mathrm{vec}\,D)')'$, and $g \in \mathcal{F}_{\mathrm{ULAN}}$. Then,

(i) letting $B := S_g^{-1/2}\Lambda^{-1}DS_g^{1/2}$ and $a_{r,s}(\bar{K};g) := \frac{\sigma_{g_r}}{\sigma_{g_s}}\,\delta_{r,s}(\bar{K},\bar{K};g)$, we have

$$\mathrm{ARE}_{\vartheta,\tau,g}\big[\tilde{\phi}_{K,L}/\phi_{\mathrm{Wilks}}\big] = \frac{\sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2}\big(a_{r,s}(\bar{K};g)B_{rs} + a_{s,r}(\bar{K};g)B_{sr}\big)^2}{\sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2}(B_{rs}+B_{sr})^2}, \tag{5.7}$$

where $\bar{K}$, $\delta_{r,s}$, and S_g were defined in Sections 4.4, 4.5, and 5.1, respectively.

(ii) the Wilcoxon test $\tilde{\phi}_W$ and the van der Waerden test $\tilde{\phi}_{\mathrm{vdW}}$ satisfy

$$\inf_{\vartheta,\tau,g}\mathrm{ARE}_{\vartheta,\tau,g}\big[\tilde{\phi}_W/\phi_{\mathrm{Wilks}}\big] = \frac{81\pi^8}{2^{20}} \approx .733$$

and

$$\inf_{\vartheta,\tau,g}\mathrm{ARE}_{\vartheta,\tau,g}\big[\tilde{\phi}_{\mathrm{vdW}}/\phi_{\mathrm{Wilks}}\big] = 1,$$

where the infima are taken over all $\vartheta, \tau \in \mathbb{R}^{p+p^2}$ and $g \in \mathcal{F}_{\mathrm{ULAN}}$ for which the corresponding B is symmetric. Moreover, for $\tilde{\phi}_{\mathrm{vdW}}$, the lower bound is reached if and only if all ICs are Gaussian.

Part (ii) of this result establishes the very good uniform efficiency properties of our Wilcoxon and—overall—of our van der Waerden signed-rank tests, under *type 1* alternatives (this restriction is associated with the symmetry of B in Proposition 5.2(ii)). Such uniform efficiency results, for location problems, were first derived in Hodges and Lehmann (1956) and Chernoff and Savage (1958), for Wilcoxon scores and van der Waerden scores, respectively. As for Part (i), it provides, for arbitrary sequences of alternatives and any dimensions p_1 and p_2, a very simple expression for the AREs of *our Puri and Sen type signed-rank tests* (the ones based on (4.1)–(4.2)) with respect to Wilks'.

Unfortunately, such a simple expression does not exist for the other proposed signed-rank tests, namely those for which $d(K,L) > p_1 p_2$. To give some insight on the AREs of the latter with respect to Wilks' test, we consider the AREs, under identically distributed ICs (with common pdf g_1, say), of the signed-rank test ($\tilde{\phi}_{f_1}$, say) designed to achieve optimality when both ICs share some (non-Gaussian) pdf f_1 (we may safely exclude the case for which f_1 is Gaussian since the resulting test, namely $\tilde{\phi}_{\mathrm{vdW}}$, is then of the Puri and Sen type). Lengthy yet straightforward calculations yield that, under the bivariate type j (j=1,2)

alternatives in (5.5)–(5.6), we have

$$\mathrm{ARE}^{(j)}_{\vartheta,\tau,g_1}\left[\tilde{\phi}_{f_1}/\phi_{\mathrm{Wilks}}\right] = \frac{c_1^{\alpha,\beta}}{c_2^{\alpha,\beta}} \frac{2(\mathcal{I}_{f_1,g_1}\sigma^2_{f_1,g_1} + s_j\mathcal{L}_{f_1;g_1}\mathcal{L}_{g_1;f_1})^2/(\mathcal{I}_{f_1}\sigma^2_{f_1} + s_j)}{(1+s_j)^2}, \tag{5.8}$$

with $s_1 := 1$, $s_2 := -1$, $\mathcal{I}_{f_1,g_1} := \mathrm{E}[\varphi_{f_1}(F^{-1}_{1+}(U))\varphi_{g_1}(G^{-1}_{1+}(U))]$, $\sigma^2_{f_1,g_1} := \mathrm{E}[F^{-1}_{1+}(U)G^{-1}_{1+}(U)]$, and $\mathcal{L}_{f_1;g_1} := \mathrm{E}[\varphi_{f_1}(F^{-1}_{1+}(U))G^{-1}_{1+}(U)]$ (in (5.8), we stick to the same numerator/denominator structure as in (5.4), and we allow for zero noncentrality parameters in the denominator, with obvious interpretation). In particular, at $g_1 = f_1$, we simply have

$$\mathrm{ARE}^{(j)}_{\vartheta,\tau,f_1}\left[\tilde{\phi}_{f_1}/\phi_{\mathrm{Wilks}}\right] = \frac{c_1^{\alpha,\beta}}{c_2^{\alpha,\beta}} \frac{2(\mathcal{I}_{f_1}\sigma^2_{f_1} + s_j)}{(1+s_j)^2}, \quad j = 1,2, \tag{5.9}$$

which, when f_1 is the pdf of the t_ν distribution (with $\nu > 2$, so that ULAN holds), gives

$$\mathrm{ARE}^{(j)}_{\vartheta,\tau,t_\nu}\left[\tilde{\phi}_{t_\nu}/\phi_{\mathrm{Wilks}}\right] = \frac{c_1^{\alpha,\beta}}{c_2^{\alpha,\beta}} \frac{2(1+s_j) + \frac{12}{(\nu-2)(\nu+3)}}{(1+s_j)^2}, \quad j = 1,2. \tag{5.10}$$

For type 1 alternatives, it is clear that, since $c_1^{\alpha,\beta} < c_2^{\alpha,\beta}$ for all α, β, the AREs in (5.10) are strictly smaller than one for large ν (e.g., if $\alpha = .05$ and $\beta = .80$, these AREs are .916 and .852 for $\nu = 5$ and $\nu = 8$, respectively; see Table 1), so that Wilks' test asymptotically dominates, when both ICs are t_ν, the signed-rank test that is optimal under such conditions. This of course is puzzling at first sight. However, our concept of optimality, namely *most stringency* (see Section 3.1), clearly does not imply that the optimal tests are most powerful under *all* alternatives, but only that their lack of power with respect to the best test for any fixed alternative is minimal. What occurs in the AREs (5.8)–(5.10) is totally in line with most stringency: our optimal tests $\tilde{\phi}_{f_1}$ (f_1 non-Gaussian) pay a price in terms of efficiency along type 1 alternatives (which allows for the superiority of Wilks' there) in order to gain some power along type 2 alternatives, where the local asymptotic powers of Wilks' are equal to the nominal level α. The AREs of the optimal tests $\tilde{\phi}_{f_1}$ (f_1 non-Gaussian) with respect to Wilks' under (non-Gaussian) Type 2 alternatives may then be considered as being infinite; see (5.8)–(5.10) again.

Note that $\mathrm{Cov}[X_1, X_2] = 0$ under type 2 alternatives, which explains why such alternatives are more difficult to detect than those of type 1. At the multinormal, X_1 and X_2 are then independent, and type 2 "alternatives" actually belong to the null; hence, optimal tests at the multinormal model can concentrate on being most powerful along type 1 alternatives. That is exactly what Wilks' test and our van der Waerden test $\tilde{\phi}_{\mathrm{vdW}}$ do. Away from the multinormal, however, tests with more degrees of freedom are needed to discriminate between the null and type 2 alternatives. This also explains why, in Q_{ϕ_∞} (see (3.9)), the term Q_{new}, which is the limit (as $\nu \to \infty$) of a quadratic form allowing to detect type 2 alternatives, may be dropped without affecting optimality at the multinormal model.

6. Practical implementation

The practical implementation of our parametric tests (resp., of our nonparametric tests) crucially relies on the existence of an estimator $\hat{\vartheta}^{(n)} = (\hat{\mu}^{(n)\prime}, (\operatorname{vec}\hat{\Lambda}^{(n)})')'$ satisfying Assumption (B) (resp., the existence of a couple of estimators $\hat{\vartheta}_K^{(n)} = (\hat{\mu}_K^{(n)\prime}, (\operatorname{vec}\hat{\Lambda}^{(n)})')'$ and $\hat{\vartheta}_L^{(n)} = (\hat{\mu}_L^{(n)\prime}, (\operatorname{vec}\hat{\Lambda}^{(n)})')'$ satisfying Assumption (B′)). Also, our nonparametric tests turn out to be strongly affected, when $d(K,L) > p_1p_2$, by the slow convergence of our test statistics to their limiting distributions. This section discusses these issues and provides practical solutions.

6.1. Estimation of ϑ

As stated in Theorems 3.1 and 4.1, the asymptotic properties of our tests do not depend on the choice of $\hat{\vartheta}$, $\hat{\vartheta}_K$, and $\hat{\vartheta}_L$ (we drop superscripts $^{(n)}$ in this section); still, their finite-sample properties may be affected by this choice. Here, we suggest using a particular class of practical estimates that are robust and easy to implement.

To describe conveniently these estimates, we define $\mathcal{G}_k^{\text{unit}}$ as the collection of matrices $\Lambda \in \mathcal{G}_k$ for which each column has Euclidean norm one, and consider the k-dimensional IC model

$$X = \Lambda Z + \mu, \tag{6.1}$$

where $\mu \in \mathbb{R}^k$, $\Lambda \in \mathcal{G}_k^{\text{unit}}$, and $Z(\overset{\mathcal{D}}{=} -Z)$ has independent marginals (ICs). Note that the ICs are not standardized in (6.1), but that the requirement $\Lambda \in \mathcal{G}_k^{\text{unit}}$ (rather than just $\Lambda \in \mathcal{G}_k$) plays the same role in the mutual identification of Λ and (the distribution of) Z as the standardization of Z in the IC models from Section 2.1.

Consider then two k-variate scatter matrix functionals S_a and S_b (recall that, if F_X is the cdf of a k-variate random vector X, a *scatter matrix functional* S is a $k\times k$ matrix-valued functional such that $S(F_X)$ is positive definite, symmetric, and affine-equivariant in the sense that $S(F_{AX+b}) = AS(F_X)A'$ for any $k \times k$ invertible matrix A and any k-vector b). If X comes from the IC model (6.1), the affine-equivariance of S_a and S_b entails that (see also (7) in Nordhausen, Oja and Paindaveine (2009))

$$S_b(F_X)(S_a(F_X))^{-1}\Lambda = \Lambda D, \tag{6.2}$$

where $D := S_b(F_Z)(S_a(F_Z))^{-1}$. Since D is diagonal (this follows from the independence and the symmetry of the marginals of Z; see Theorem 7 in Tyler et al. (2009)), the columns of Λ are made of eigenvectors of $S_b(F_X)(S_a(F_X))^{-1}$; here the order, signs, and norms of these eigenvectors are fixed by the requirement $\Lambda \in \mathcal{G}_k^{\text{unit}}$.

Of course, the resulting estimator $\hat{\Lambda}$ is obtained by replacing S_a and S_b with root-n consistent estimates $\hat{S}_a$ and $\hat{S}_b$ in (6.2). Actually, one should not bother too much about signs and norms of the columns of $\hat{\Lambda}$ since our tests are invariant under reflections (about zero) and rescaling of the estimated ICs (which

explains that we may safely consider here the IC model in (6.1) rather than the one in Section 2.1). However, if they involve score functions (K_r, L_r, say, to adopt the notation used in our nonparametric tests) that are not homogeneous across components (i.e., that depend on r), then our tests are *not* invariant under permutations of the estimated ICs, and it is crucial to order the score functions K_r, L_r so that the ordering matches that of the underlying ICs (note that failing to achieve this matching would actually affect only local powers and optimality properties of our tests, but not their asymptotic validity). Typically, if the score functions are the ones allowing to achieve optimality at $f \in \mathcal{F}$ (i.e., are given by K_f, L_f), then one way to achieve this matching is to reorder the score functions (equivalently, the marginal densities f_r of f) so that the diagonal entries of $S_b(F_f)S_a^{-1}(F_f)$ (where F_f denotes the cdf associated with f) have the same vector of ranks as of the diagonal entries of the matrix $\hat{D}_{\text{ordered}}$ defined through $\hat{S}_b\hat{S}_a^{-1}\hat{\Lambda} = \hat{\Lambda}\,\hat{D}_{\text{ordered}}$ (with $\hat{\Lambda} \in \mathcal{G}_k^{\text{unit}}$).

Now, if $2 \le q < k$ ICs are identically distributed, the corresponding q eigenvalues of $S_b(F_X)(S_a(F_X))^{-1}$ do coincide, which implies that only the subspace of those q ICs can be recovered through (6.2) but not the individual ICs themselves; in such cases, our tests are not valid, unless those q ICs are Gaussian—see Nordhausen, Oja and Paindaveine (2009) for a discussion. However, if one feels that excluding (non-Gaussian) identically distributed ICs is too much of an assumption, then it is always possible to resort to another estimator of Λ in the literature; but we feel that the estimators we propose in this section are not only easy to compute in practice, but are also very well in line with the nonparametric sign-and-rank spirit of our tests (see the practical estimators we propose below).

Focusing on the problem of testing for independence, we consider the null model (Assumption (B(i)) indeed states that $\hat{\vartheta}$ (or $\hat{\vartheta}_K/\hat{\vartheta}_L$) should be obtained by fitting the null model)

$$\begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} = \begin{pmatrix} \Lambda_{11}Z_i^{(1)} \\ \Lambda_{22}Z_i^{(2)} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad i = 1, \ldots, n, \tag{6.3}$$

where the marginals of the i.i.d. noise inputs $Z_i = (Z_i^{(1)\prime}, Z_i^{(2)\prime})'$, $i = 1, \ldots, n$ are independent and symmetric about the origin (again, adopting the requirement $\Lambda \in \mathcal{G}_p^{\text{unit}}$ or rather standardizing noise inputs does not affect the behavior of our tests, so that one may here adopt the IC model he/she is most comfortable with). By using, for each $\ell = 1, 2$, two different p_ℓ-variate scatter matrix estimates $\hat{S}_a^{(\ell)}$ and $\hat{S}_b^{(\ell)}$, we can as above define (separately) estimators $\hat{\Lambda}_{\ell\ell}$ based on $X_i^{(\ell)}$, $i = 1, \ldots, n$. The estimator we then propose for $\vartheta = (\mu', (\operatorname{vec}\Lambda)')'$ is $\hat{\vartheta} = (\hat{\mu}', (\operatorname{vec}\hat{\Lambda})')'$, with

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} = \begin{pmatrix} \hat{\Lambda}_{11}\hat{T}^{(1)}(\hat{\Lambda}_{11}^{-1}X_1^{(1)}, \ldots, \hat{\Lambda}_{11}^{-1}X_n^{(1)}) \\ \hat{\Lambda}_{22}\hat{T}^{(2)}(\hat{\Lambda}_{22}^{-1}X_1^{(2)}, \ldots, \hat{\Lambda}_{22}^{-1}X_n^{(2)}) \end{pmatrix} \text{ and } \hat{\Lambda} = \operatorname{diag}(\hat{\Lambda}_{11}, \hat{\Lambda}_{22}), \tag{6.4}$$

where $\hat{T}^{(\ell)}$, $\ell = 1, 2$, is some p_ℓ-variate location estimator obtained by computing in a *componentwise* fashion a univariate (equivariant) location estimator $\hat{T}$.

Provided that $\hat{S}_a^{(\ell)}$, $\hat{S}_b^{(\ell)}$, and $\hat{T}^{(\ell)}$ ($\ell = 1, 2$) are root-n consistent, this estimator $\hat{\vartheta}$ then clearly fulfills Assumptions (B(i)–(iii)); note that, in relation with (B(iii)), the transformed data set

$$\begin{pmatrix} \hat{Z}_i^{(1)} \\ \hat{Z}_i^{(2)} \end{pmatrix} = \begin{pmatrix} \hat{\Lambda}_{11}^{-1}(X_i^{(1)} - \hat{\mu}_1) \\ \hat{\Lambda}_{22}^{-1}(X_i^{(2)} - \hat{\mu}_2) \end{pmatrix}, \quad i = 1, \ldots, n$$

is then *invariant* under the transformation $z \to Az + b$ for any $p \times p$ invertible matrix $A = \operatorname{diag}(A_{11}, A_{22})$ and any p-vector b. After appropriate discretization, the estimator $\hat{\vartheta}$ therefore satisfies Assumption (B), hence can be used in the parametric tests of Section 3.

Our nonparametric tests of Section 4 however require a couple of estimators $\hat{\vartheta}_K = (\hat{\mu}_K', (\operatorname{vec}\hat{\Lambda})')'$ and $\hat{\vartheta}_L = (\hat{\mu}_L', (\operatorname{vec}\hat{\Lambda})')'$ satisfying Assumption (B′). While one can use the same estimator $\hat{\Lambda}$ as above, we need to define appropriate location estimates $\hat{\mu}_K$ and $\hat{\mu}_L$. For $\hat{\mu}_K$ (one can define $\hat{\mu}_L$ accordingly), we propose adopting the location estimator obtained when using K-score location R-estimators in (6.4) above. More precisely, we suggest using

$$\hat{\mu}_K := \begin{pmatrix} \hat{\mu}_{K,1} \\ \hat{\mu}_{K,2} \end{pmatrix} := \begin{pmatrix} \hat{\Lambda}_{11} & 0 \\ 0 & \hat{\Lambda}_{22} \end{pmatrix} \hat{\xi}_K, \tag{6.5}$$

where the rth component of $\hat{\xi}_K$ is defined as an arbitrary "zero" of the step function

$$t \mapsto h_r^K(t) := \sum_{i=1}^{n} \hat{S}_{ir}(t)\hat{K}_{ir}(t),$$

where the signed ranks $\hat{S}_{ir}(t)\hat{K}_{ir}(t)$ are those of $y_{ir} - t$, $i = 1, \ldots, n$, with $y_{ir} := [e_r'(\hat{\Lambda}_{11}^{-1}X_i^{(1)})]$ for $r \le p_1$ and $y_{ir} := [e_{r-p_1}'(\hat{\Lambda}_{22}^{-1}X_i^{(2)})]$ for $r > p_1$. By "zero", we here mean an arbitrary value t_0 for which $h_r^K(t_0^-) \times h_r^K(t_0^+) \le 0$; in order to define $\hat{\mu}_K$ unambiguously, we could always choose the such zero that is closest to the sample median of the y_{ir}'s, $i = 1, \ldots, n$. Beyond being robust and root-n consistent without any moment condition, the resulting estimators $\hat{\mu}_K$ and $\hat{\mu}_L$ also satisfy Assumption (B′(i)–(ii)) (see the end of Section A.2 for a proof), hence can be used in our nonparametric tests.

Finally, we point out that, to avoid any moment assumption in the construction above, our choices for $\hat{S}_a^{(\ell)}$ and $\hat{S}_b^{(\ell)}$ are the Tyler (1987) and Dümbgen (1998) scatter matrix estimates, respectively. These statistics are *weakly* affine-equivariant only, in the sense that, for invertible matrices A and vectors b with appropriate dimensions, $\hat{S}(AX_1 + b, \ldots, AX_n + b)$ is proportional (but in general not equal) to $A\hat{S}(X_1, \ldots, X_n)A'$; however, it is easy to check that this weak affine-equivariance is sufficient to guarantee the (standard) affine-equivariance of the resuting estimators of ϑ. For Tyler's estimate, we need a simultaneous location estimate; a natural choice is the Hettmansperger and Randles (2002)

estimate. These estimates are root-n consistent under very weak assumptions, which do not involve any moment condition.

6.2. An alternative critical value

Our signed-rank tests in Section 4 are based on the fact that the statistics

$$\tilde{Q}_{K,L} = \tilde{Q}^{(n)}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} := \tilde{T}'_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L}\, M_{K,L}\, \tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L}$$

are asymptotically $\chi^2_{d(K,L)}$ under the null. For some of these statistics however (typically, this happens when $d(K,L) > p_1p_2$), this convergence is very slow and the test rejecting the null at asymptotic level α when $\tilde{Q}_{K,L} > \chi^2_{d(K,L);1-\alpha}$ is strongly conservative (hence, biased) for small to moderate sample sizes. Consequently, we recommend using the following alternative critical values, which allow for implementing bias-corrected versions of our nonparametric tests.

If $\vartheta(\in\mathcal{M}(\Omega))$ was known, one could use the test statistic $\tilde{Q}^{(n)}_{\vartheta,\vartheta,K,L}$, which is (strictly) distribution-free under the null; see Section 4.1. The exact critical values of $\tilde{Q}^{(n)}_{\vartheta,\vartheta,K,L}$ can easily be estimated by using simulations; distribution-freeness indeed allows for simply generating M independent samples $S_1,\ldots,S_M$ of n i.i.d. *multinormal* observations with location $\mu_0 = 0$ and covariance matrix $\Lambda_0\Lambda_0' = I_p$, and to use empirical quantiles of the resulting M values of $\tilde{Q}^{(n)}_{\vartheta_0,\vartheta_0,K,L}$, $\vartheta_0 = (\mu_0', (\mathrm{vec}\,\Lambda_0)')' = (0', (\mathrm{vec}\, I_p)')'$ as estimates for the exact quantiles of $\tilde{Q}^{(n)}_{\vartheta,\vartheta,K,L}$

Now, since $\tilde{Q}_{K,L} = \tilde{Q}^{(n)}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L}$ and $\tilde{Q}^{(n)}_{\vartheta,\vartheta,K,L}$ are asymptotically equivalent under the null (this is a direct consequence of Lemma 4.2), the exact critical values of $\tilde{Q}^{(n)}_{\vartheta,\vartheta,K,L}$—hence, also their estimated versions above—may be regarded as approximate ones for $\tilde{Q}_{K,L}$. As we show through a Monte-Carlo experiment in the next section, this approach works very well in practice.

7. Simulations

In this section, we report the results of Monte-Carlo experiments that were conducted to study the small-sample performances of the proposed signed-rank tests and to see how they compete with Wilks'. It was also of interest to investigate how well the finite-sample behaviors of the various procedures were in accordance with the asymptotic results of the previous sections, and we therefore compared non-null rejection frequencies with the corresponding asymptotic powers.

We first considered the bivariate case $p_1 = p_2 = 1$ and fixed the location μ and the null value of the mixing matrix Λ to 0 and I_2, respectively (this is without loss of generality since all tests here are affine-invariant). For various distributions of the ICs (see below) and some selected values of δ (including the null value $\delta = 0$), we generated $N = 5,000$ independent samples of sizes $n = 100, 200$, and 500 from the bivariate IC models

$$\begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} = \left(\Lambda + n^{-1/2}\delta\, D\right) \begin{pmatrix} Z_i^{(1)} \\ Z_i^{(2)} \end{pmatrix} + \mu, \quad i = 1, \ldots, n, \tag{7.1}$$

where D is as in (5.5) or (5.6) (type 1 and type 2 alternatives, respectively). For each sample, we performed Wilks' test and the following six signed-rank tests, all at asymptotic nominal level 5%: the sign test $\tilde{\phi}_S$, the Wilcoxon test $\tilde{\phi}_W$, the van der Waerden test $\tilde{\phi}_{\mathrm{vdW}}$ (achieving Le Cam optimality in the multinormal case), and the tests $\tilde{\phi}_{t_{\nu_1} t_{\nu_2}}$, with $(\nu_1, \nu_2) = (3,3)$, $(3,5)$, and $(5,5)$, based on the (bivariate version of the) statistics in (4.6) (achieving Le Cam optimality when the ICs are t_{ν_1} and t_{ν_2}). For the last three tests (for which $d(K,L) = 2$, the convergence to the null asymptotic χ^2_2 distribution appeared to be quite slow, and we therefore used the alternative critical values of Section 6.2 (evaluated, for each such test and each sample size n, on the basis of $M = 10,000$ independent samples of n i.i.d. bivariate standard normal observations). For all other tests, the critical values were simply based on the asymptotic χ^2_1 approximation of the null distributions.

For type 1 alternatives, rejection frequencies are reported, as functions of δ, in Figure 1 (where both $Z_i^{(1)}$ and $Z_i^{(2)}$ are Gaussian), Figure 2 (where $Z_i^{(1)}$ and $Z_i^{(2)}$ are t_3 and t_5, respectively), and Figure 4(a) (where both $Z_i^{(1)}$ and $Z_i^{(2)}$ are t_1). When second-order moments are finite (that is, in the designs considered in Figures 1–2), we also present the corresponding asymptotic powers—computed from Theorems 4.1(ii) and 5.1(ii). The results show that, when both ICs are Gaussian, Wilks' test slightly dominates the asymptotically optimal van der Waerden test (and of course the other signed-rank tests) at sample size 100, but that this dominance fades out as the sample size increases. When the ICs are t_3 and t_5, the signed-rank tests (excluding the sign test, as expected) are a bit more powerful than Wilks' test, which is in accordance with the AREs of Section 5.2. In each case, finite-sample powers seem to converge quickly to the asymptotic ones. When both ICs are t_1, it is seen that Wilks' test (which requires finite second-order moments) exhibits poor performances, while the signed-rank tests behave as expected: in particular, the "closer" (t_{ν_1}, t_{ν_2}) to the underlying couple of IC distributions (t_1, t_1), the better the performances of the asymptotically χ^2_2 tests $\tilde{\phi}_{t_{\nu_1} t_{\nu_2}}$.

We then consider type 2 alternatives, for which rejection frequencies are reported in Figure 3 (where $Z_i^{(1)}$ and $Z_i^{(2)}$ are t_3 and t_5, respectively) and Figure 4(b) (where both $Z_i^{(1)}$ and $Z_i^{(2)}$ are t_1). In the setup of Figure 3, all tests based on a statistic that is asymptotically χ^2_1 under the null should, according to our asymptotic results, exhibit poor asymptotic powers (in particular, asymptotic powers of Wilks' test should coincide with the nominal level 5%). Quite surprisingly, Wilks' seems to gain some power at the sample sizes considered (we have checked, however, that this unexpected behavior of Wilks' test disappears for larger sample sizes). The finite-sample powers of the signed-rank tests $\tilde{\phi}_{t_{\nu_1} t_{\nu_2}}$, which here are by far the most powerful ones, converge to the limiting ones, although relatively slowly. For type 2 alternatives with t_1 ICs (Figure 4(b)), Wilks' test, which is extremely conservative, is again strongly dominated by the signed-rank tests $\tilde{\phi}_{t_{\nu_1} t_{\nu_2}}$.

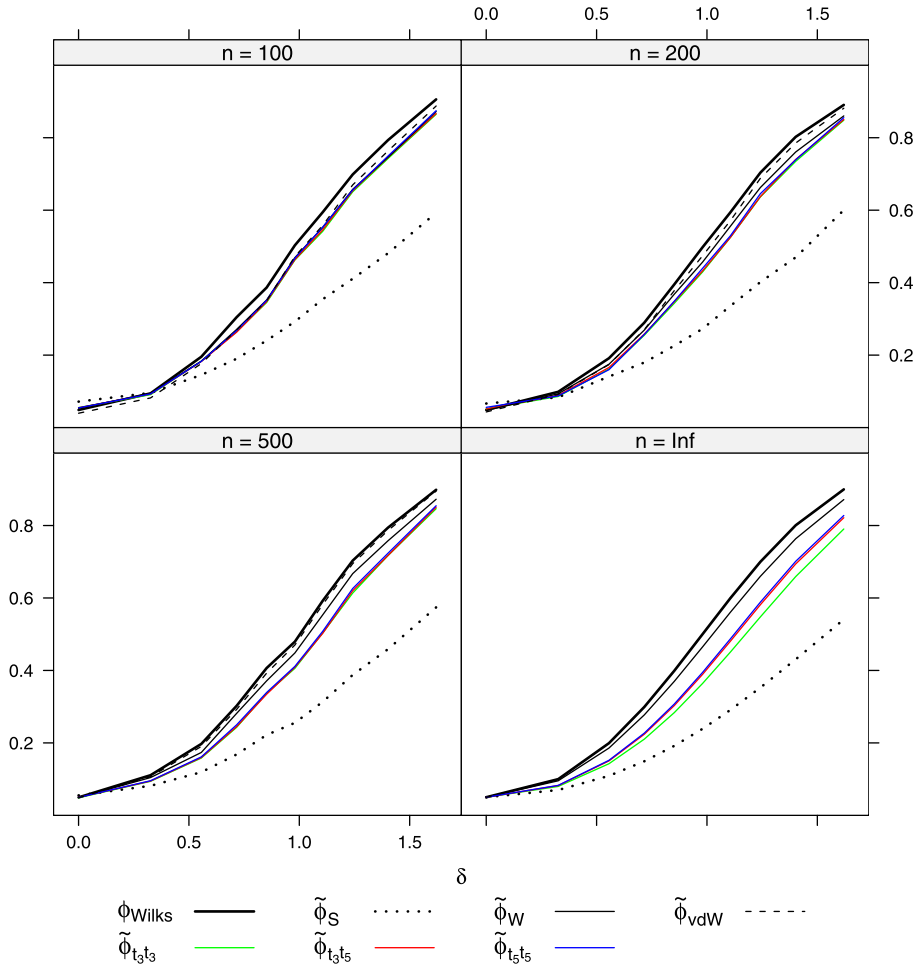

FIG 1. *Rejection frequencies and asymptotic powers, under $p_1 = p_2 = 1$ type 1 alternatives with Gaussian ICs, of Wilks' test (ϕ_{Wilks}), the sign test ($\tilde{\phi}_S$), the Wilcoxon signed-rank test ($\tilde{\phi}_W$), the van der Waerden signed-rank test ($\tilde{\phi}_{\text{vdW}}$), and various signed-rank tests based on (t_{ν_1}, t_{ν_2})-score functions ($\tilde{\phi}_{t_{\nu_1} t_{\nu_2}}$); see Section 7 for details.*

To explore cases where $Z_i^{(1)}$ and $Z_i^{(2)}$ are multivariate, we generated observations through

$$\begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} = \left(\Lambda + n^{-1/2}\delta\, D\right) \begin{pmatrix} Z_i^{(1)} \\ Z_i^{(2)} \end{pmatrix} + \mu, \quad i = 1, \ldots, n,$$

where $\mu = 0 \in \mathbb{R}^{p_1+p_2} = \mathbb{R}^{2p_1}$, $\Lambda = I_{p_1+p_2}$,

$$D = \begin{pmatrix} 0 & \sigma_{g_1}/\sigma_{g_2} \\ \sigma_{g_2}/\sigma_{g_1} & 0 \end{pmatrix} \otimes I_{p_1} \qquad \text{(type 1 alternatives)}$$

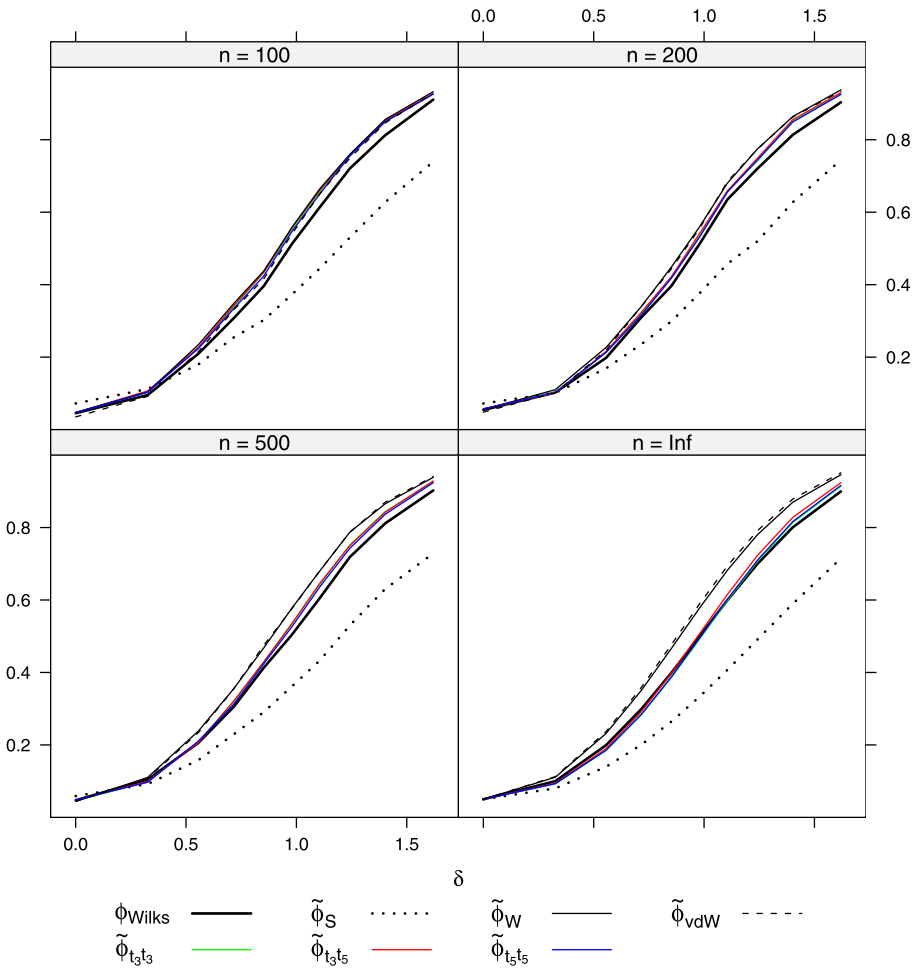

FIG 2. *Finite-sample and exact limiting powers, under $p_1 = p_2 = 1$ type 1 alternatives with t_3 and t_5 ICs, of Wilks' test (ϕ_{Wilks}), the sign test ($\tilde{\phi}_S$), the Wilcoxon signed-rank test ($\tilde{\phi}_W$), the van der Waerden signed-rank test ($\tilde{\phi}_{\text{vdW}}$), and various signed-rank tests based on (t_{ν_1}, t_{ν_2})-score functions ($\tilde{\phi}_{t_{\nu_1}t_{\nu_2}}$); see Section 7 for details.*

or

$$D = \begin{pmatrix} 0 & \sigma_{g_1}/\sigma_{g_2} \\ -\sigma_{g_2}/\sigma_{g_1} & 0 \end{pmatrix} \otimes I_{p_1} \qquad \text{(type 2 alternatives)},$$

and where the first (resp., second) half of the mutually independent marginals of both $Z_i^{(1)}$ and $Z_i^{(2)}$ are t_3 distributed (resp., t_5 distributed) — which of course implies that we restrict to even values of $p_1 (= p_2)$. In all cases below, the results are based on $5,000$ independent replications and the same tests as for $p_1 = p_2 = 1$ were implemented at asymptotic level $\alpha = 5\%$ in

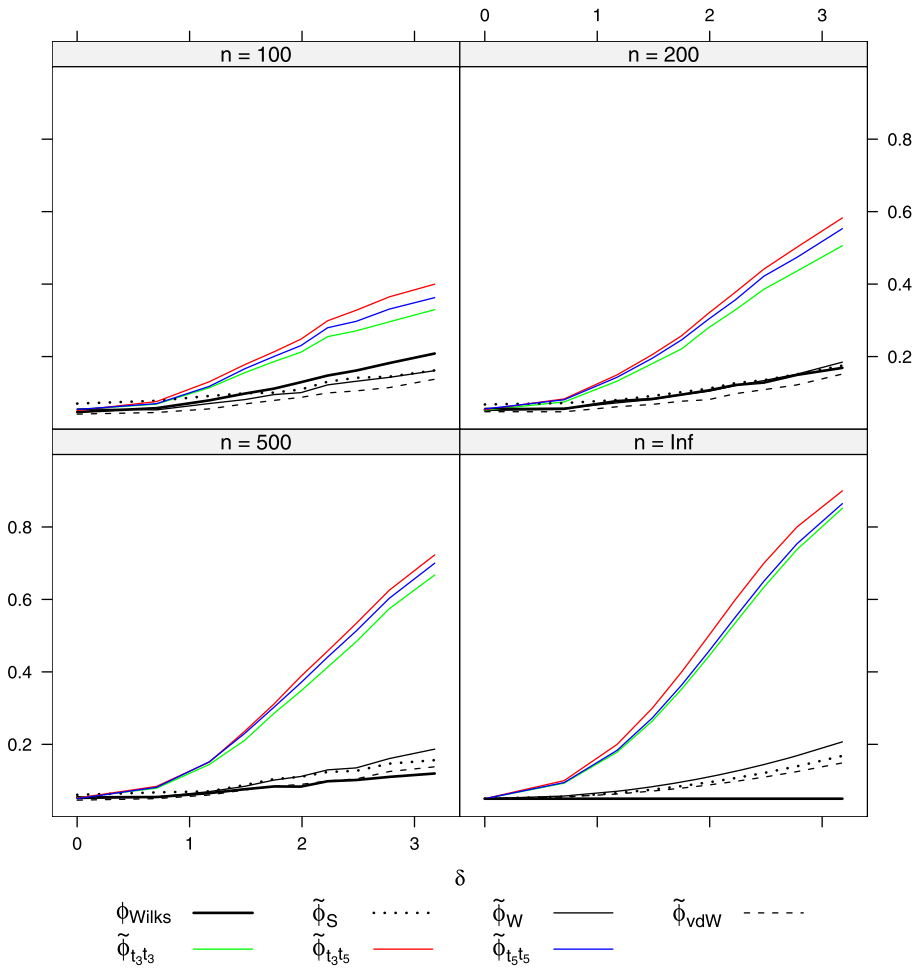

FIG 3. *Rejection frequencies and asymptotic powers, under $p_1 = p_2 = 1$ type 2 alternatives with t_3 and t_5 ICs, of Wilks' test (ϕ_{Wilks}), the sign test ($\tilde{\phi}_S$), the Wilcoxon signed-rank test ($\tilde{\phi}_W$), the van der Waerden signed-rank test ($\tilde{\phi}_{\text{vdW}}$), and various signed-rank tests based on (t_{ν_1}, t_{ν_2})-score functions ($\tilde{\phi}_{t_{\nu_1}t_{\nu_2}}$); see Section 7 for details.*

exactly the same fashion as above (here, the tests $\tilde{\phi}_{t_{\nu_1}t_{\nu_2}}$ are to be replaced with tests of the form $\tilde{\phi}_{t_{\nu_1}t_{\nu_2}\ldots t_{\nu_{p_1+p_2}}}$, achieving optimality when the marginals of $Z_i = (Z_i^{(1)\prime}, Z_i^{(2)\prime})'$ are $t_{\nu_1}, t_{\nu_2}, \ldots, t_{\nu_{p_1+p_2}}$, respectively). Obviously, unlike in the simulations considered for $p_1 = p_2 = 1$, a setup with $\min(p_1, p_2) > 1$ requires to run two separate ICAs to compute the estimated ICs.

For $p_1 = p_2 = 2$, the resulting rejection frequencies (and asymptotic powers) are reported in Figure 5 and Figure 6, for type 1 and type 2 alternatives, respectively. These figures show that our tests behave similarly as for $p_1 = p_2 = 1$ and

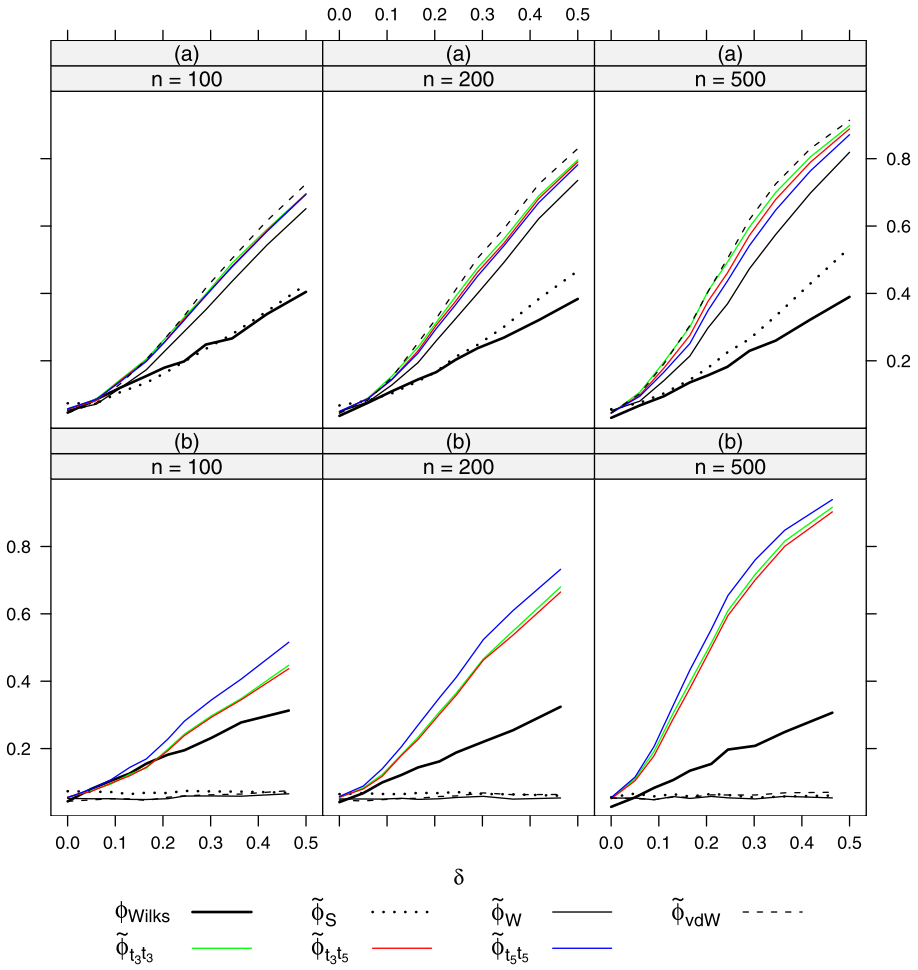

FIG 4. *Rejection frequencies, under $p_1 = p_2 = 1$ (a) type 1 or (b) type 2 alternatives, with t_1 ICs in both cases, of Wilks' test (ϕ_{Wilks}), the sign test ($\tilde{\phi}_S$), the Wilcoxon signed-rank test ($\tilde{\phi}_W$), the van der Waerden signed-rank test ($\tilde{\phi}_{\text{vdW}}$), and various signed-rank tests based on (t_{ν_1}, t_{ν_2})-score functions ($\tilde{\phi}_{t_{\nu_1}t_{\nu_2}}$); see Section 7 for details.*

that they are not affected by the estimation of nuisance parameters through the ICA procedures proposed in Section 6 (at least when based, as it was the case here, on the Tyler (1987) and Dümbgen (1998) scatter matrix estimates and on the location estimates $\hat{\mu}_K$ and $\hat{\mu}_L$ from (6.5)). Figure 7 provides the corresponding results for type 1 alternatives in the higher-dimensional setup $p_1 = p_2 = 10$. Clearly, at the same sample sizes as in the low-dimensional cases above, the tests still behave in an excellent agreement with the fixed-p asymptotic theory. In contrast, simulations results not reported here indicate that, for high-dimensional type 2 alternatives, such an agreement shows only for larger sample sizes. Note

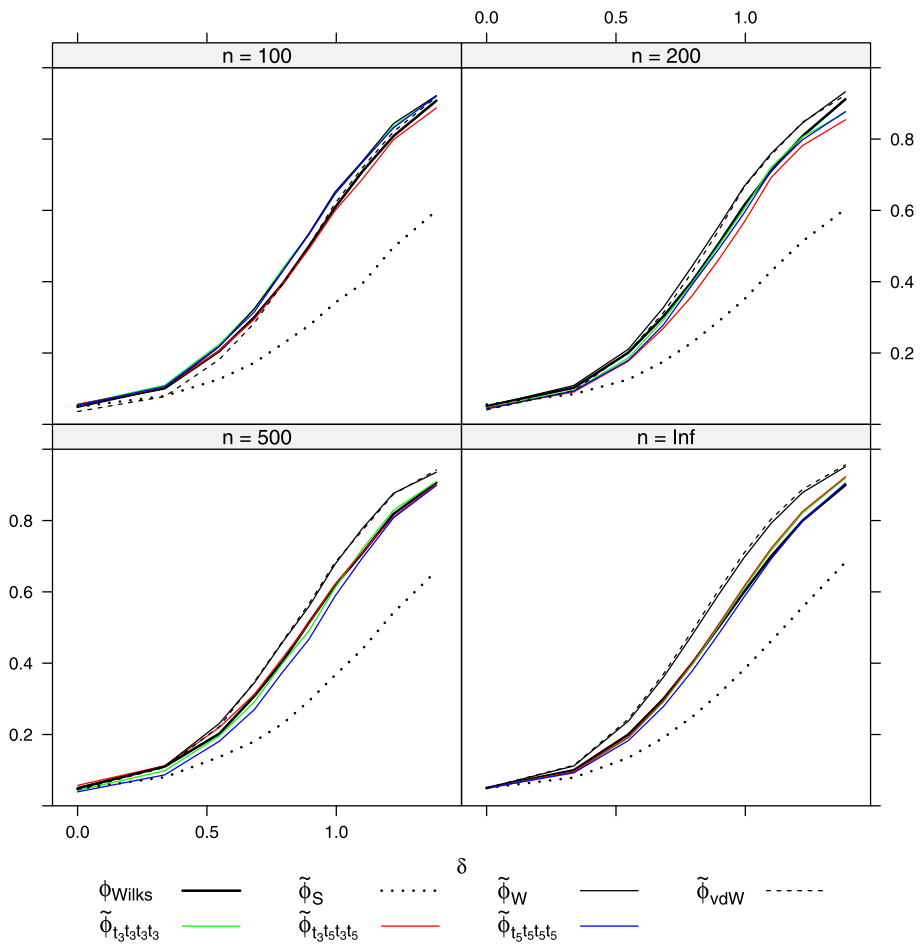

FIG 5. *Rejection frequencies and asymptotic powers, under $p_1 = p_2 = 2$ type 1 alternatives with both $Z_i^{(1)}$ and $Z_i^{(2)}$ having a t_3 marginal and a t_5 marginal, of Wilks' test (ϕ_{Wilks}), the sign test ($\tilde{\phi}_S$), the Wilcoxon signed-rank test ($\tilde{\phi}_W$), the van der Waerden signed-rank test ($\tilde{\phi}_{\text{vdW}}$), and various signed-rank tests based on $(t_{\nu_1}, t_{\nu_2}, t_{\nu_3} t_{\nu_4})$-score functions ($\tilde{\phi}_{t_{\nu_1} t_{\nu_2} t_{\nu_3} t_{\nu_4}}$); see Section 7 for details.*

that our tests still remain of high practical value even for such high-dimensional type 2 alternatives, since many ICA applications, e.g., in signal processing, typically offer large sample sizes.

To comment on the computational efforts required by the proposed nonparametric procedures, we ran (for sample sizes $n = 100$, $n = 200$ and $n = 500$) 100 replications of the null distributional setups considered in Figures 1, 5 and 7 (this relates to $p_1 = p_2 = 1$, $p_1 = p_2 = 2$ and $p_1 = p_2 = 10$, respectively). In Figure 8, we show the computing times (in seconds), averaged over these 100 replications, of the sign test statistic, the Wilcoxon test statistic, the van der

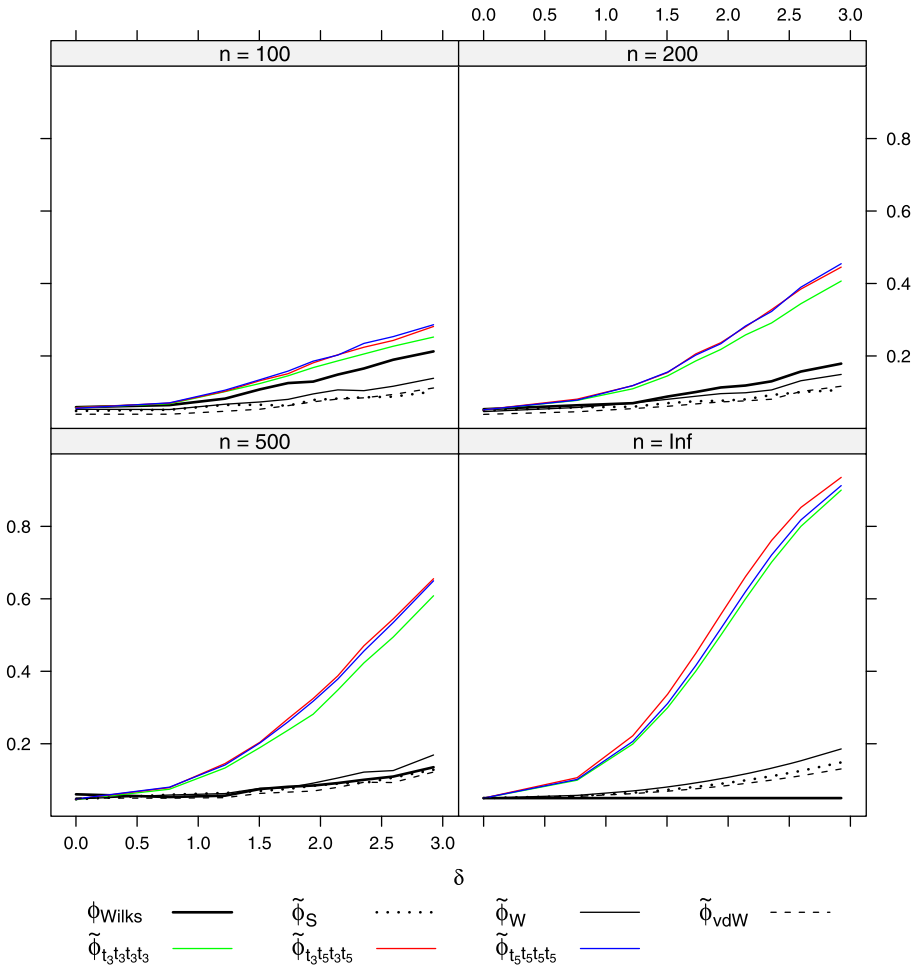

FIG 6. *Rejection frequencies and asymptotic powers, under* $p_1 = p_2 = 2$ *type 2 alternatives with both* $Z_i^{(1)}$ *and* $Z_i^{(2)}$ *having a* t_3 *marginal and a* t_5 *marginal, of Wilks' test* (ϕ_{Wilks})*, the sign test* ($\tilde{\phi}_S$)*, the Wilcoxon signed-rank test* ($\tilde{\phi}_W$)*, the van der Waerden signed-rank test* ($\tilde{\phi}_{\text{vdW}}$)*, and various signed-rank tests based on* $(t_{\nu_1}, t_{\nu_2}, t_{\nu_3} t_{\nu_4})$*-score functions* ($\tilde{\phi}_{t_{\nu_1} t_{\nu_2} t_{\nu_3} t_{\nu_4}}$)*; see Section 7 for details.*

Waerden test statistic, and the signed-rank test statistics based on t_3-scores only, t_5-scores only, and on mixed t_3-t_5 scores (the three corresponding t-score tests are those used in Figures 1, 5 and 7). Whenever $p_1 = p_2 > 1$, the implementation of all these test statistics requires evaluating the Tyler and Dümbgen scatter matrices, and the corresponding (still, averaged) computing times are then also shown in Figure 8. The results reveal that, as expected, the computational effort increases both with the dimension and with the sample size. The more complex test statistics, relying on t-scores, are more costly than the sign,

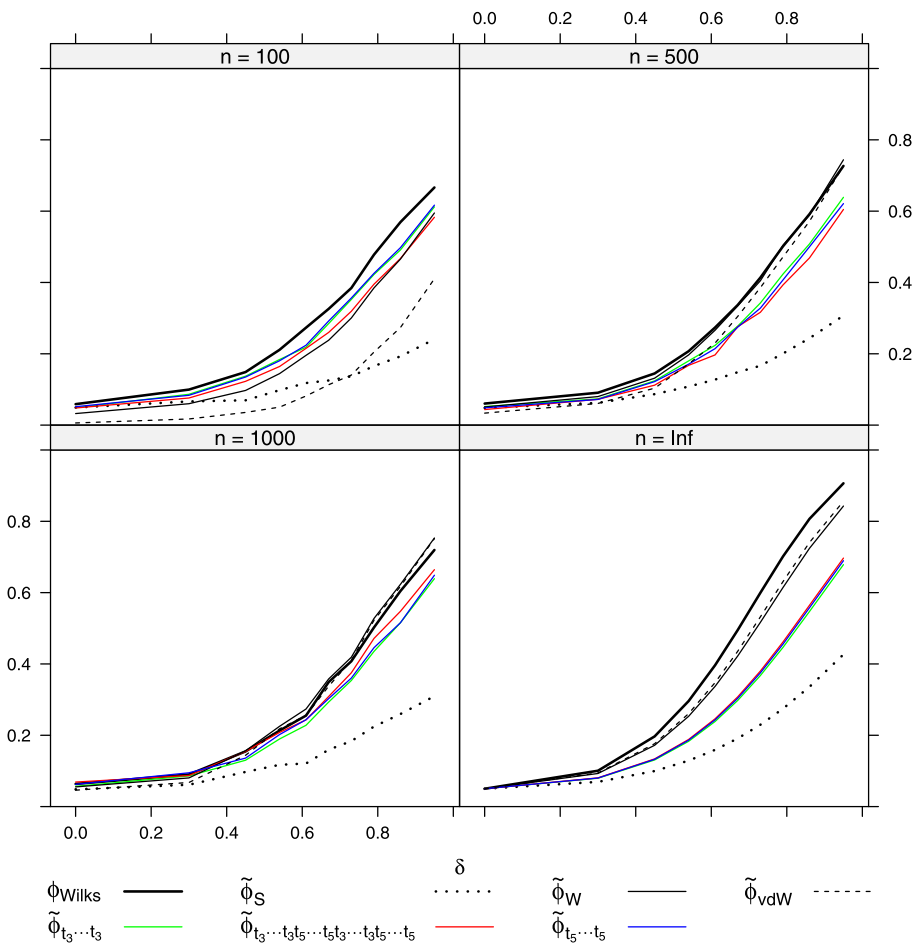

FIG 7. *Rejection frequencies and asymptotic powers, under* $p_1 = p_2 = 10$ *type 1 alternatives with all* $Z_i^{(1)}$ *and* $Z_i^{(2)}$ *respectively having their first five (resp., last five) marginals that are* t_3 *distributed (resp.,* t_5 *distributed), of Wilks' test (*ϕ_{Wilks}*), the sign test (*$\tilde{\phi}_S$*), the Wilcoxon signed-rank test (*$\tilde{\phi}_W$*), the van der Waerden signed-rank test (*$\tilde{\phi}_{\text{vdW}}$*), and various signed-rank tests based on* $(t_{\nu_1}, \ldots, t_{\nu_{10}})$*-score functions (*$\tilde{\phi}_{t_{\nu_1} \ldots t_{\nu_{10}}}$*); see Section 7 for details (the test denoted as* $\tilde{\phi}_{t_3 \ldots t_3 t_5 \ldots t_5 t_3 \ldots t_3 t_5 \ldots t_5}$ *is the one that is optimal in the distributional setup considered).*

Wilcoxon and van der Waerden ones. Clearly, as the sample size increases, the time to compute the Dümbgen scatter matrix is increasingly large compared to the time required to compute the full test statistic (this can be corrected for by using a less robust scatter matrix). Most importantly, for all dimensions and sample sizes considered, computing times remain quite small for every nonparametric test statistic, with a maximum value of less than 2.5 seconds for sample size $n = 500$ in dimension $p_1 = p_2 = 10$.

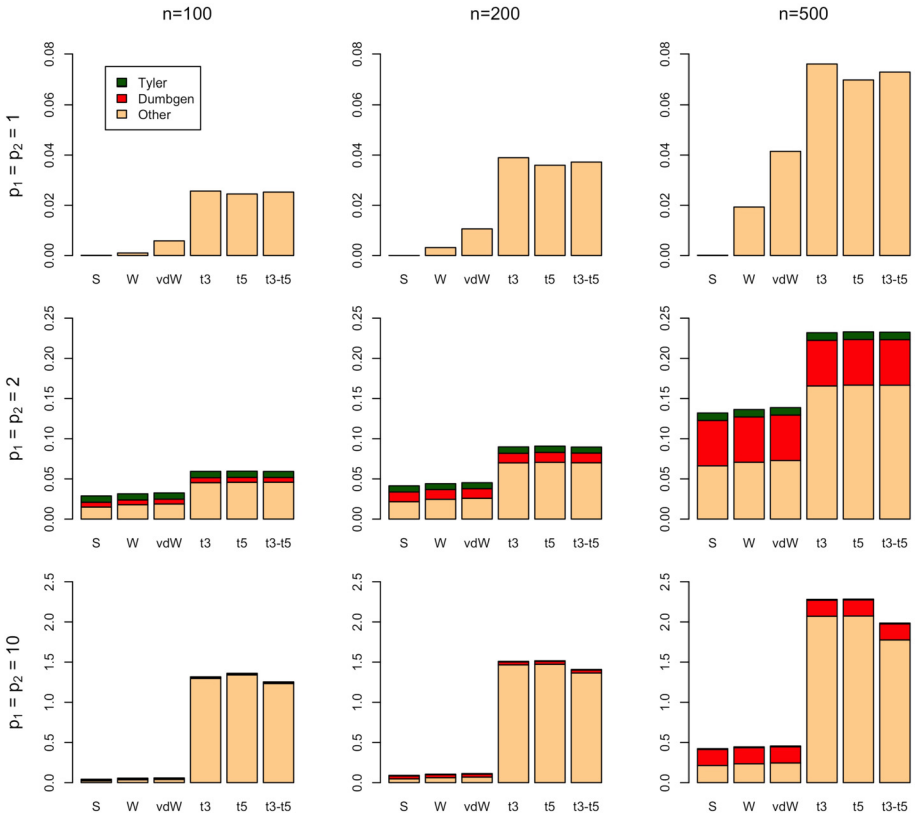

FIG 8. *Computing times (in seconds), averaged over 100 replications, of the sign test statistic* (S)*, the Wilcoxon test statistic* (W)*, the van der Waerden test statistic* (vdW)*, and various signed-rank test statistics based on* t_3*-scores only* (t_3)*,* t_5*-scores only* (t_5)*, and on mixed* t_3*-*t_5 *scores* (t_3*-*t_5)*. The distributional setups are those considered under the null in Figure 1, Figure 5 and Figure 7, for* $p_1 = p_2 = 1$*,* $p_1 = p_2 = 2$ *and* $p_1 = p_2 = 10$*, respectively. Whenever* $p_1 = p_2 > 1$*, the total computing time is partitioned into the (still, averaged) times used to compute the Tyler scatter matrix (green), the Dümbgen scatter matrix (red), and the remaining time used to compute the test statistic.*

As a conclusion, this Monte-Carlo study shows that (i) all tests succeed in meeting the asymptotic 5% level constraint at all sample sizes; (ii) the non-null rejection frequencies of the proposed signed-rank tests are compatible with the corresponding asymptotic powers and AREs derived in the previous sections; most importantly, while they compete reasonably well with the other tests under type 1 alternatives, the optimal signed-rank tests (that are based on a higher number of degrees of freedom) outperform the other tests under (non-Gaussian) type 2 alternatives. Again, this is totally in line with the optimality concept (namely, most stringency) considered. Finally, the computational efforts required remains quite low, which makes the proposed tests applicable in practice.

Appendix A: Appendix

A.1. Proofs of Proposition 2.1 and Lemma 3.1

Clearly, the likelihood, under $P^n_{\vartheta,g}$, is given by $L^n_{\vartheta,g} = \prod_{i=1}^n g_\vartheta(X_i)$, where $g_\vartheta(x) := |\det \Lambda|^{-1} \prod_{r=1}^p g_r\big(e_r'\Lambda^{-1}(x-\mu)\big)$ (here, e_r still denotes the rth vector of the canonical basis of $\mathbb{R}^p$). The proof of Proposition 2.1 is then based on Lemma 1 from Swensen (1985), or, more precisely, on its extension by Garel and Hallin (1995), which, in the present context, takes the following form.

Lemma A.1. Define $\xi_i^{(n)} := (g^{1/2}_{\vartheta+n^{-1/2}\tau_n}(X_i)/g^{1/2}_\vartheta(X_i)) - 1$ and assume that there exist a triangular array of random variables $\zeta_i^{(n)}$, $i = 1,\ldots,n$ and a bounded sequence (d_n) satisfying the following seven conditions, where all convergences are as $n \to \infty$ and all expectations and stochastic convergences are under $P^n_{\vartheta,g}$: (i) $\mathrm{E}[\sum_{i=1}^n(\xi_i^{(n)} - \zeta_i^{(n)})^2] = o(1)$; (ii) $\sup_n \mathrm{E}[\sum_{i=1}^n(\zeta_i^{(n)})^2] < \infty$; (iii) $\max_{1\le i\le n}|\zeta_i^{(n)}| = o_P(1)$; (iv) $\sum_{i=1}^n(\zeta_i^{(n)})^2 - (1/4)d_n^2 = o_P(1)$; (v) $\sum_{i=1}^n \mathrm{E}[(\zeta_i^{(n)})^2 I_{[|\zeta_i^{(n)}|>1/2]}] = o(1)$; (vi) $\mathrm{E}[\zeta_i^{(n)}] = 0$ for all i,n; (vii) $\sum_{i=1}^n \mathrm{E}[(\xi_i^{(n)})^2 + 2\xi_i^{(n)}] = o(1)$. Then, under $P^n_{\vartheta,g}$, with $g \in \mathcal{F}_{\mathrm{ULAN}}$,

$$\log(dP^n_{\vartheta_n+n^{-1/2}\tau_n,g}/dP^n_{\vartheta_n,g}) = 2\sum_{i=1}^n \zeta_i^{(n)} - (1/2)d_n^2 + o_P(1) \tag{A.1}$$

and $[\log(dP^n_{\vartheta_n+n^{-1/2}\tau_n,g}/dP^n_{\vartheta_n,g}) + (1/2)d_n^2]/d_n$ is asymptotically standard normal.

The proof of Lemma A.1 in turn crucially relies on the quadratic mean differentiability of the mapping $\vartheta \mapsto g^{1/2}_\vartheta(\cdot)$ stated in Lemma A.2 below (we omit the tedious proof, which can be done by combining arguments from Lind and Roussas (1972) and from the proof of Lemma A.1 in Hallin and Paindaveine (2006)).

Lemma A.2. Let $g \in \mathcal{F}_{\mathrm{ULAN}}$ and define

$$\begin{aligned}
D_\mu g^{1/2}_\vartheta(x) &:= \frac{1}{2}\, g^{1/2}_\vartheta(x)(\Lambda^{-1})'\varphi_g(\Lambda^{-1}(x-\mu)),\\
D_\Lambda g^{1/2}_\vartheta(x) &:= \frac{1}{2}\, g^{1/2}_\vartheta(x)\left(I_p\otimes\Lambda^{-1}\right)' \mathrm{vec}\,[\varphi_g(\Lambda^{-1}(x-\mu))\,(\Lambda^{-1}(x-\mu))' - I_p],\\
D_\vartheta g^{1/2}_\vartheta(x) &= ((D_\mu g^{1/2}_\vartheta(x))', (D_\Lambda g^{1/2}_\vartheta(x))')'.
\end{aligned}$$

Then, (i) $\int \{g^{1/2}_{\mu+m,\Lambda}(x) - g^{1/2}_{\mu,\Lambda}(x) - m'D_\mu g^{1/2}_\vartheta(x)\}^2 dx = o(\|m\|^2)$, as $m \to 0$, (ii) $\int \{g^{1/2}_{\mu,\Lambda+L}(x) - g^{1/2}_{\mu,\Lambda}(x) - (\mathrm{vec}\,L)'D_\Lambda g^{1/2}_\vartheta(x)\}^2 dx = o(\|L\|^2)$, as $(\mathrm{vec}\,L) \to 0$, and (iii) $\int \{g^{1/2}_{\vartheta+t}(x) - g^{1/2}_\vartheta(x) - t'D_\vartheta g^{1/2}_\vartheta(x)\}^2 dx = o(\|t\|^2)$, as $t = (m', (\mathrm{vec}\,L)')' \to 0$.

Proof of Proposition 2.1. Letting $T_i^{(n)} := Dg^{1/2}_\vartheta(X_i)/g^{1/2}_\vartheta(X_i)$, $i = 1,\ldots,n$, it is easy to check that $\mathrm{E}[T_i^{(n)}] = \int g^{1/2}_\vartheta(x)\, Dg^{1/2}_\vartheta(x)\, dx = 0$ and that $\mathrm{E}[T_i^{(n)}T_i^{(n)\prime}] =$

$\int Dg_\vartheta^{1/2}(x)(Dg_\vartheta^{1/2}(x))'\,dx = (1/4)\Gamma_{\vartheta,g}$. We then show that Conditions (i)-(vii) in Lemma A.1 are satisfied with $\zeta_i^{(n)} := n^{-1/2}\,\tau_n' T_i^{(n)}$, $i = 1,\ldots,n$ and $d_n = \Gamma_{\vartheta,g}^{1/2}\tau_n$.

(i) Lemma A.2(iii) directly entails that $\mathrm{E}[\sum_{i=1}^n(\xi_i^{(n)} - \zeta_i^{(n)})^2] = n\int\{g_{\vartheta+n^{-1/2}\tau_n}^{1/2}(x) - g_\vartheta^{1/2}(x) - n^{-1/2}\tau_n' Dg_\vartheta^{1/2}(x)\}^2\,dx$ is $o(1)$. (ii) Since $\mathrm{E}[(\zeta_i^{(n)})^2] = n^{-1}\tau_n'\Gamma_{\vartheta,g}\tau_n$ for all i, Condition (ii) follows from the boundedness of (τ_n). (iii) Letting $\varepsilon > 0$ be fixed and defining $C_\tau := \sup_n \|\tau_n\|$, we have

$$P[\max_{1\le i\le n}|\zeta_i^{(n)}| \le \varepsilon] = (P[(\zeta_i^{(n)})^2 \le \varepsilon^2])^n \ge (P[C_\tau^2\|T_i^{(n)}\|^2 \le \varepsilon^2 n])^n \\ \ge \Big(1 - \frac{1}{\varepsilon^2 n}\mathrm{E}\Big[C_\tau^2\|T_i^{(n)}\|^2 I_{[C_\tau^2\|T_i^{(n)}\|^2 > \varepsilon^2 n]}\Big]\Big)^n \quad \text{(A.2)}$$

which converges to one (since $T_i^{(n)}$ has a finite covariance matrix under the considered sequence of probability measures, we indeed have that the second expectation in (A.2) is $o(1)$). (iv) The WLLN directly entails that $\sum_{i=1}^n(\zeta_i^{(n)})^2 - (1/4)d_n^2 = \tau_n'[n^{-1}\sum_{i=1}^n T_i^{(n)}T_i^{(n)\prime} - (1/4)\Gamma_{\vartheta,g}]\tau_n = o(1)$. (v) denoting by $\|A\|_{\mathcal{L}} := \sup\{\|Ax\| : \|x\| = 1\}$ the sup norm of the matrix A, we have

$$\sum_{i=1}^n \mathrm{E}[(\zeta_i^{(n)})^2 I_{[|\zeta_i^{(n)}|>1/2]}] \le C_\tau^2\|\mathrm{E}[T_1^{(n)}T_1^{(n)\prime} I_{[|T_1^{(n)}|>n^{1/2}/(2C_\tau)]}]\|_{\mathcal{L}},$$

which is $o(1)$ since $\mathrm{E}[T_1^{(n)}T_1^{(n)\prime}]$ is finite. (vi) $\mathrm{E}[\zeta_i^{(n)}] = n^{-1/2}\,\tau_n'\mathrm{E}[T_i^{(n)}] = 0$. (vii) Using the definition of $\xi_i^{(n)}$, we obtain $\mathrm{E}[(\xi_i^{(n)})^2 + 2\xi_i^{(n)}] + 1 = \mathrm{E}[(\xi_i^{(n)} + 1)^2] = \int g_{\vartheta+n^{-1/2}\tau_n}(x)\,dx = 1$.

Consequently, Lemma A.1 applies and establishes the result since the linear term in the right-hand side of (A.1) is $2\sum_{i=1}^n \zeta_i^{(n)} = \tau_n'(2n^{-1/2}\sum_{i=1}^n T_i^{(n)})$, hence coincide with $\tau_n'\Delta_{\vartheta,g}^{(n)}$ (see Lemma A.2) and the expression of the central sequence $\Delta_{\vartheta,g}^{(n)}$ in Proposition 2.1). □

Proof of Lemma 3.1. Lengthy but straightforward calculations allow to check that

$$\Gamma_{\vartheta,f;2}^{-} = (I_p\otimes\Lambda)\Bigg\{\sum_{r=1}^p \frac{1}{\mathcal{J}_{f_r}-1}(e_re_r'\otimes e_re_r') + \\ \sum_{\substack{r\ne s\\ r,s=1}}^p\Big(\alpha_{r,s}(f)\,(e_re_r'\otimes e_se_s') + \beta_{r,s}(f)(e_re_s'\otimes e_se_r')\Big)\Bigg\}(I_p\otimes\Lambda)' \quad \text{(A.3)}$$

is the Moore-Penrose pseudoinverse of $\Gamma_{\vartheta,f;2}$. It can easily be checked that (with $\vartheta\in\mathcal{M}(\Omega)$)

$$\Omega_2'\Gamma_{\vartheta,f;2}\Omega_2 = \mathrm{diag}\Big(\Gamma_{\vartheta,f;2}^{(1)}, \Gamma_{\vartheta,f;2}^{(2)}\Big),$$

where $\Gamma_{\vartheta,f}^{(\ell)} = \mathrm{diag}(\Gamma_{\vartheta,f;1}^{(\ell)}, \Gamma_{\vartheta,f;2}^{(\ell)})$ denotes the $(p_\ell(p_\ell+1)) \times (p_\ell(p_\ell+1))$ information matrix in the marginal IC model $X_i^{(\ell)} = \Lambda_{\ell\ell}Z_i^{(\ell)} + \mu_\ell$, $\ell = 1,2$. Since

$\Gamma^{(\ell)}_{\vartheta,f;2}$, $\ell = 1, 2$ are information matrices with the same structure as $\Gamma_{\vartheta,f;2}$, one can use (A.3) to obtain the Moore-Penrose pseudoinverses $\big(\Gamma^{(\ell)}_{\vartheta,f;2}\big)^-$, $\ell = 1, 2$. Simplifying the resulting explicit expression for

$$\Gamma^-_{\vartheta,f;2} - \Omega_2\big(\Omega_2'\Gamma_{\vartheta,f;2}\Omega_2\big)^-\Omega_2' = \Gamma^-_{\vartheta,f;2} - \Omega_2\left[\mathrm{diag}\big(\big(\Gamma^{(1)}_{\vartheta,f;2}\big)^-, \big(\Gamma^{(2)}_{\vartheta,f;2}\big)^-\big)\right]\Omega_2'.$$

then provides the result. □

A.2. Proofs of Lemmas 4.1 and 4.2

In this section, we fix some $\vartheta \in \Theta$ and $g \in \mathcal{F}$ (in the proof of Lemma 4.2, the fixed value of ϑ is in $\mathcal{M}(\Omega)$). All expectations and stochastic convergences are then under $P^n_{\vartheta,g}$. We write throughout Z_{ir}, S_{ir}, R_{ir}, $\hat{Z}^K_{ir}$, $\hat{S}^K_{ir}$, and $\hat{R}^K_{ir}$, for $Z_{ir}(\vartheta)$, $S_{ir}(\vartheta)$, $R_{ir}(\vartheta)$, $Z_{ir}(\hat{\vartheta}_K)$, $S_{ir}(\hat{\vartheta}_K)$, and $R_{ir}(\hat{\vartheta}_K)$, respectively. We also let $\hat{K}_{ir} := K_r(\hat{R}^K_{ir}/(n+1))$, $\tilde{K}_{ir} := K_r(R_{ir}/(n+1))$, and $K_{ir} := K_r(G_{+r}(|Z_{ir}|))$. Of course, Z^L_{ir}, $\hat{S}^L_{ir}$, $\hat{R}^L_{ir}$, $\hat{L}_{ir}$, $\tilde{L}_{ir}$, and L_{ir} will denote the corresponding quantities based on $\hat{\vartheta}_L$ and the score function L_r. Eventually, we denote by $\tilde{G}_{\vartheta,K,L}$, $\tilde{G}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L}$, and $G_{\vartheta,K,L;g}$, the $p\times p$ matrices defined by $\tilde{T}_{\vartheta,K,L} =: (\mathrm{vec}\,\tilde{G}_{\vartheta,K,L})$, $\tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} =: (\mathrm{vec}\,\tilde{G}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L})$, $T_{\vartheta,K,L;g} =: (\mathrm{vec}\,G_{K,L;g})$, respectively.

Proof of Lemma 4.1. Since

$$\mathrm{E}[\|\tilde{T}_{\vartheta,K,L} - T_{\vartheta,K,L;g}\|^2] = \sum_{r<>s} E[((\tilde{G}_{\vartheta,K,L} - G_{\vartheta,K,L;g})_{rs})^2], \tag{A.4}$$

it is sufficient to show that each term in this sum is $o(1)$ as $n \to \infty$. Then using (i) $E[S_{ir}] = 0$, (ii) the independence (under $P^n_{\vartheta,g}$) between the S_{ir}'s and the $(R_{ir}, |Z_{ir}|)$'s, and (iii) the independence, for fixed $r \neq s$ (note that r are s are never equal in (A.4)), between the Z_{ir}'s and the Z_{is}'s, we obtain $E[((\tilde{G}_{\vartheta,K,L} - G_{\vartheta,K,L;g})_{rs})^2] = E[(n^{-1/2}\sum_i S_{ir}S_{is}(\tilde{K}_{ir}\tilde{L}_{is} - K_{ir}L_{is}))^2] = n^{-1}\sum_i E[(\tilde{K}_{ir}\tilde{L}_{is} - K_{ir}L_{is})^2]$. Hence, $E[((\tilde{G}_{\vartheta,K,L} - G_{\vartheta,K,L;g})_{rs})^2] = E[(\tilde{K}_{1r}\tilde{L}_{1s} - K_{1r}L_{1s})^2] \le 2E[(\tilde{K}_{1r} - K_{1r})^2\tilde{L}^2_{1s}] + 2\,E[K^2_{1r}(\tilde{L}_{1s} - L_{1s})^2] \le 2(E[L^2_s(U)] + o(1))E[(\tilde{K}_{1r} - K_{1r})^2] + 2\,E[K^2_r(U)]E[(\tilde{L}_{1s} - L_{1s})^2]$, as $n \to \infty$.

This proves the result since the Hájek projection theorem for linear signed-rank statistics (see Puri and Sen (1985), Chapter 3) yields $E[(\tilde{K}_{1r} - K_{1r})^2] = n^{-1}\sum_i E[(\tilde{K}_{ir} - K_{ir})^2] = n^{-1}\sum_{i,j} E[S_{ir}S_{jr}(\tilde{K}_{ir} - K_{ir})(\tilde{K}_{jr} - K_{jr})] = E[((n^{-1/2}\sum_i S_{ir}\tilde{K}_{ir}) - (n^{-1/2}\sum_i S_{ir}K_{ir}))^2]$ is $o(1)$ as $n \to \infty$ (of course, one similarly obtains that $E[(\tilde{L}_{1s} - L_{1s})^2] = o(1)$ as $n \to \infty$). □

It remains to prove Lemma 4.2. In view of Lemma 4.1, it is sufficient to show that $\mathrm{E}[\|\tilde{T}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} - T_{\vartheta,K,L;g}\|^2] = o(1)$ as $n \to \infty$, or equivalently, that

$$E[((\tilde{G}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} - G_{\vartheta,K,L;g})_{rs})^2] = o(1) \tag{A.5}$$

as $n \to \infty$, for any $(r, s) \in (\mathcal{S}_1 \times \mathcal{S}_2) \cup (\mathcal{S}_2 \times \mathcal{S}_1)$. In the rest of this section, we therefore fix $(r, s) \in \mathcal{S}_1 \times \mathcal{S}_2$ (the case $(r, s) \in \mathcal{S}_2 \times \mathcal{S}_1$ of course follows along the

same lines) and proceed to prove (A.5). We will need the following preliminary result.

Lemma A.3. As $n \to \infty$, (i) $\hat{Z}^K_{1r} - Z_{1r} = o_P(1) = \hat{Z}^L_{1r} - Z_{1r}$ under $P^n_{\vartheta,g}$ (with the ϑ and g fixed at the beginning of this section), (ii) $\mathrm{E}[(\hat{K}_{1r} - K_{1r})^2] = o(1) = \mathrm{E}[(\hat{L}_{1r} - L_{1r})^2]$, (iii) $\mathrm{E}[|\hat{S}^K_{1r} - S_{1r}|^a] = o(1) = \mathrm{E}[|\hat{S}^L_{1r} - S_{1r}|^a]$ for any $a > 0$, and (iv) $\mathrm{E}[(\hat{S}^K_{1r}\hat{K}_{1r} - S_{1r}K_{1r})^2] = o(1) = \mathrm{E}[(\hat{S}^L_{1r}\hat{L}_{1r} - S_{1r}L_{1r})^2]$.

Proof of Lemma A.3. Clearly, it is sufficient to prove the lemma for the K-based quantities (the proof for the L-based ones is entirely similar).

(i) Since $|\hat{Z}^K_{1r} - Z_{1r}| \le \|\hat{Z}^K_1 - Z_1\| \le \|\hat{\Lambda}^{-1} - \Lambda^{-1}\|_{\mathcal{L}}\|X_1 - \mu\| + \|\hat{\Lambda}^{-1}\|_{\mathcal{L}}\|\hat{\mu}_K - \mu\|$, the claim for $\hat{Z}^K_{1r} - Z_{1r}$ follows from the root-n consistency of $\hat{\vartheta}_K$.

(ii) Applying Lemma 2 in page 555 of Peters and Randles (1990), with $\alpha = \vartheta$ and $g(X, \alpha) = |e_r'[\Lambda^{-1}(X - \mu)]|$ yields that $\frac{\hat{R}^K_{1r}}{n+1} - G_{r+}(|Z_{1r}|)$ is $o_P(1)$ as $n \to \infty$ (note that Conditions (a)-(b) of that lemma are fulfilled: (a) is our Assumption (B(ii)), whereas (b) can be checked exactly along the same lines as in Peters and Randles (1990), once it is noticed that $||e_r'[(\Lambda + n^{-1/2}H)^{-1}(X - \mu_0)]| - |e_r'[\Lambda^{-1}(X - \mu_0)]|| \le \|[(\Lambda + n^{-1/2}H)^{-1} - \Lambda^{-1}](X - \mu_0)\|$, for any fixed $p \times p$ matrix H).

Now, the continuity of K_r entails that $\hat{K}_{1r} - K_{1r}$ is $o_P(1)$ as $n \to \infty$. To prove that this convergence also holds in quadratic mean (which is precisely Part (ii) of the lemma), it is sufficient to show that $\hat{K}_{1r} - K_{1r}$ is uniformly integrable. Clearly, K_{1r} is uniformly integrable since the integrable random variable $K_{1r} = K_r(G_{r+}(|Z_{1r}|))$ does not depend on n. As for $\hat{K}_{1r}$, recall that $\tilde{K}_{1r} - K_{1r} = o_{L^2}(1)$ as $n \to \infty$ (see the proof of Lemma 4.1), which implies that $\tilde{K}_{1r}$ is uniformly integrable. Finally, the latter uniform integrability and the invariance of $\hat{\vartheta}_K$ under permutations of the observations in turn imply that $\hat{K}_{1r}$ is also uniformly integrable. We conclude that $\hat{K}_{1r} - K_{1r}$ is indeed uniformly integrable, and the result follows.

(iii) Since $\hat{S}^K_{1r} - S_{1r} = (|\hat{Z}^K_{1r}|^{-1} - |Z_{1r}|^{-1})\hat{Z}^K_{1r} + |Z_{1r}|^{-1}(\hat{Z}^K_{1r} - Z_{1r})$, we have $|\hat{S}^K_{1r} - S_{1r}| \le 2|\hat{Z}^K_{1r} - Z_{1r}|/|Z_{1r}| =: Y_{1r}$. Now, fix some $\delta > 0$. Then, for all $\eta > 0$,

$$\begin{aligned} P[Y_{1r} > \delta] &\le P[Y_{1r} I_{[|Z_{1r}|<\eta]} > \delta/2] + P[Y_{1r} I_{[|Z_{1r}|\ge\eta]} > \delta/2] \\ &\le P[|Z_{1r}| < \eta] + P[Y_{1r} I_{[|Z_{1r}|\ge\eta]} > \delta/2] \\ &=: p_1^{(n)} + p_2^{(n)}, \end{aligned}$$

say. For all $\varepsilon > 0$, there exists $\eta = \eta(\varepsilon)$ such that $p_1^{(n)} < \varepsilon/2$. As for $p_2^{(n)}$, note that $Y_{1r} I_{[|Z_{1r}|\ge\eta]} \le (2/\eta)|\hat{Z}^K_{1r} - Z_{1r}|$, so that Part (i) of the Lemma entails that $p_2^{(n)} < \varepsilon/2$ for large n. We conclude that $|\hat{S}^K_{1r} - S_{1r}| \le Y_{1r}$ converges to zero in probability, which establishes the result (since $|\hat{S}^K_{1r} - S_{1r}|$ is bounded).

(iv) By using Hölder's inequality, we obtain

$$\begin{aligned} \mathrm{E}[(\hat{S}^K_{1r}\hat{K}_{1r} - S^K_{1r}K_{1r})^2] &\le 2\mathrm{E}\big[(\hat{S}^K_{1r})^2(\hat{K}_{1r} - K_{1r})^2\big] + 2\mathrm{E}\big[(\hat{S}^K_{1r} - S_{1r})^2 K^2_{1r}\big] \\ &\le 2\mathrm{E}\big[(\hat{K}_{1r} - K_{1r})^2\big] + 2\big(\mathrm{E}[(\hat{S}^K_{1r} - S_{1r})^{\frac{2(2+\delta)}{\delta}}]\big)^{\frac{\delta}{2+\delta}}\big(\mathrm{E}[K^{2+\delta}_{1r}]\big)^{\frac{2}{2+\delta}}, \end{aligned}$$

where $\delta > 0$ is as in Assumption (C(ii)). The latter and Parts (ii)-(iii) of this Lemma then yield the result. □

Proof of Lemma 4.2. We have to prove (A.5). To do so, note that, writing $\hat{u}_i := \hat{S}^K_{ir}\hat{K}_{ir}$, $\hat{v}_i := \hat{S}^L_{is}\hat{L}_{is}$, $u_i := S_{ir}K_{ir}$, and $v_i := S_{is}L_{is}$, and using the exchangeability of the observations, we have

$$\begin{aligned} &E[((\tilde{G}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} - G_{\vartheta,K,L;g})_{rs})^2] \\ &\quad = \mathrm{E}[(n^{-1/2}\textstyle\sum_{i=1}^n (\hat{u}_i\hat{v}_i - u_iv_i))^2] \\ &\quad = \mathrm{E}[(\hat{u}_1\hat{v}_1 - u_1v_1)^2] + (n-1)\mathrm{E}[(\hat{u}_1\hat{v}_1 - u_1v_1)(\hat{u}_2\hat{v}_2 - u_2v_2)] \\ &\quad =: T_1^{(n)} + T_2^{(n)}, \end{aligned}$$

say. Now, by using the independence between $(\hat{u}_1, u_1)$ and $(\hat{v}_1, v_1)$ under the null, we obtain

$$\begin{aligned} T_1^{(n)} &\le 2\mathrm{E}\big[(\hat{u}_1 - u_1)^2\hat{v}_1^2\big] + 2\mathrm{E}\big[u_1^2(\hat{v}_1 - v_1)^2\big] \\ &= 2\mathrm{E}\big[(\hat{u}_1 - u_1)^2\big]\mathrm{E}\big[\hat{v}_1^2\big] + 2\mathrm{E}\big[u_1^2\big]\mathrm{E}\big[(\hat{v}_1 - v_1)^2\big], \end{aligned}$$

which is $o(1)$ as $n \to \infty$ by Lemma A.3(iv). As for $T_2^{(n)}$, we have

$$\begin{aligned} T_2^{(n)} &= (n-1)\mathrm{E}[\hat{u}_1\hat{v}_1\hat{u}_2\hat{v}_2] - 2(n-1)\mathrm{E}[u_1v_1\hat{u}_2\hat{v}_2] + (n-1)\mathrm{E}[u_1v_1u_2v_2] \\ &= (n-1)\mathrm{E}[\hat{u}_1\hat{u}_2]\mathrm{E}[\hat{v}_1\hat{v}_2] - 2(n-1)\mathrm{E}[u_1\hat{u}_2]\mathrm{E}[v_1\hat{v}_2] + 0, \end{aligned} \tag{A.6}$$

where we used again the independence (under the null) between $(u_1, \hat{u}_1, \hat{u}_2)$ and $(v_1, \hat{v}_1, \hat{v}_2)$, along with the fact that $\mathrm{E}[u_1] = 0$. Defining $W_u := \sum_{i=1}^n \hat{u}_i$ and writing $\dot{u}_1$ for either $\hat{u}_1$ or u_1, we have

$$\begin{aligned} |(n-1)^{1/2}\,\mathrm{E}[\dot{u}_1\hat{u}_2]| &\le (n-1)^{-1/2}|\mathrm{E}[\dot{u}_1\big(\textstyle\sum_{i=2}^n \hat{u}_i\big)]| \\ &\le (n-1)^{-1/2}\,\mathrm{E}[|\dot{u}_1(W_u - \hat{u}_1)|] \\ &\le (n-1)^{-1/2}\big((\mathrm{E}[\dot{u}_1^2]\,\mathrm{E}[W_u^2])^{1/2} + (\mathrm{E}[\dot{u}_1^2]\,\mathrm{E}[\hat{u}_1^2])^{1/2}\big), \end{aligned}$$

which is $o(1)$ as $n \to \infty$ by using Lemma A.3(iv) and Assumption (B′(i)). Similarly, $(n-1)^{1/2}\mathrm{E}[\hat{v}_1\hat{v}_2] = o(1) = (n-1)^{1/2}\mathrm{E}[v_1\hat{v}_2]$ as $n \to \infty$, so that (A.6) implies that $T_2^{(n)}$—hence also $E[((\tilde{G}_{\hat{\vartheta}_K,\hat{\vartheta}_L,K,L} - G_{\vartheta,K,L;g})_{rs})^2]$—is $o(1)$ as $n \to \infty$. □

Finally, we prove that the nonparametric location estimate $\hat{\mu}_K$ (hence, also its companion estimate $\hat{\mu}_L$) from (6.5) satisfies Assumption (B′).

Proof that $\hat{\mu}_K$ satisfies Assumption (B′). We only check that $\hat{\mu}_K$ satisfies Assumption (B′(i)), since the other conditions are trivially met. To this end, fix $r \in \mathcal{S}_1$ (the proof is entirely similar for $r \in \mathcal{S}_2$). Using the same notation as in (6.5), we stress that the piecewise constant function h_r^K may only have jumps at the Walsh averages $(y_{ir} + y_{jr})/2$, $1 \le i \le j \le n$, with sizes $2|K_r(\frac{i}{n+1}) -$

$K_r(\frac{i-1}{n+1})|$, $i = 2, \dots, n$, or $2|K_r(\frac{1}{n+1})|$; see Bauer (1972). This implies that $\big|\sum_{i=1}^n \hat{S}_{ir}^K \hat{K}_{ir}\big| \le 2\max(\max_{i=2,\dots,n} \big|K_r(\frac{i}{n+1}) - K_r(\frac{i-1}{n+1})\big|, |K_r(\frac{1}{n+1})|)$, hence that

$$
\begin{aligned}
&n^{-1}\big|\textstyle\sum_{i=1}^n \hat{S}_{ir}^K \hat{K}_{ir}\big|^2 \\
&\quad \le 4\max\big(\tfrac{1}{n}\textstyle\sum_{i=2}^n \big(K_r\big(\tfrac{i}{n+1}\big) - K_r\big(\tfrac{i-1}{n+1}\big)\big)^2, \tfrac{1}{n} K_r^2\big(\tfrac{1}{n+1}\big)\big) \\
&\quad \le 4\max\big(\tfrac{1}{n}\textstyle\sum_{i=2}^n \big(K_r^2\big(\tfrac{i}{n+1}\big) + K_r^2\big(\tfrac{i-1}{n+1}\big) - 2K_r\big(\tfrac{i}{n+1}\big)K_r\big(\tfrac{i-1}{n+1}\big)\big), \tfrac{1}{n} K_r^2\big(\tfrac{1}{n+1}\big)\big) \\
&\quad =: 4\max(a_n + b_n - 2c_n, d_n),
\end{aligned}
$$

almost surely. The finiteness of $\int_0^1 K_r^2(u)\,du$ clearly implies that d_n is $o(1)$ as $n \to \infty$. Now, if one assumes that K_r is monotone increasing, we have that $K_r^2(\frac{i-1}{n+1}) \le K_r(\frac{i}{n+1})K_r(\frac{i-1}{n+1}) \le K_r^2(\frac{i}{n+1})$, hence that c_n, as a_n and b_n, converges to $\int_0^1 K^2(u)\,du$ as $n \to \infty$, which yields the result in the monotone case considered. In the general case, the result follows similarly by writing $K_r = K_{r+} - K_{r-}$, where $K_{r\pm}$ are monotone increasing functions; see Assumption (C(ii)). □

A.3. Proofs of Lemmas 4.3 and 5.1

Proof of Lemma 4.3. Under $P_{\vartheta,g}^n$, the multivariate CLT yields that $T_{\vartheta,K,L;g}$ is asymptotically normal with mean zero and covariance matrix $H_{K,L}$. Under $P_{\vartheta+n^{-1/2}\tau,g}^n$, the asymptotic normality of $T_{\vartheta,K,L;g}$ with mean $H_{K,L;g}(I_p \otimes \Lambda^{-1})\tau_2$ and covariance matrix $H_{K,L}$ follows as usual, by (i) establishing the joint normality (under $P_{\vartheta,g}^n$) of $T_{\vartheta,K,L;g}$ and $\log(dP_{\vartheta+n^{-1/2}\tau,g}^n/dP_{\vartheta,g}^n)$, then (ii) applying Le Cam's third Lemma (the required joint normality follows from a routine application of the classical Cramér-Wold device).

This establishes the result since Lemma 4.1 shows that $\tilde{T}_{\vartheta,K,L} = T_{\vartheta,K,L;g} + o_P(1)$ as $n \to \infty$, under $P_{\vartheta,g}^n$ (hence, by contiguity, also under $P_{\vartheta+n^{-1/2}\tau,g}^n$). □

Proof of Lemma 5.1. Along the same lines as in the proof of Lemma 4.3, the result under $P_{\vartheta,g}^n$ easily follows from the multivariate CLT, and the one under $P_{\vartheta+n^{-1/2}\tau,g}^n$ can be obtained by establishing the joint normality (under $P_{\vartheta,g}^n$) of $T_{\vartheta,\phi,g}$ and $\log(dP_{\vartheta+n^{-1/2}\tau,g}^n/dP_{\vartheta,g}^n)$, then applying Le Cam's third Lemma. □

We end this section with

Lemma A.4. Let Assumptions (A) and (B) hold at $g = f \in \mathcal{F}_{\mathrm{ULAN}}$. Then (i) $T_{\vartheta,f}$ is asymptotically normal with mean zero and mean $H_f(I_p \otimes \Lambda^{-1})\tau_2$ under $P_{\vartheta,f}^n$ and under $P_{\vartheta+n^{-1/2}\tau,f}^n$ ($\tau = (\tau_1', \tau_2')' \in \mathbb{R}^p \times \mathbb{R}^{p^2}$), respectively, and covariance matrix H_f under both; (ii) writing $\hat{\vartheta} = (\hat{\mu}', (\mathrm{vec}\,\hat{\Lambda})')'$, $T_{\hat{\vartheta},f} - T_{\vartheta,f} + n^{1/2}H_f(I_p \otimes \Lambda^{-1})\mathrm{vec}(\hat{\Lambda} - \Lambda)$ is $o_P(1)$ as $n \to \infty$, under $P_{\vartheta,f}^n$.

Proof of Lemma A.4. Part (i) of the result readily follows from Lemma 4.3 by using Lemma 4.1 and the identities $T_{\vartheta,K_f,L_f;f} = T_{\vartheta,f}$ and $H_{K_f,L_f;f} = H_{K_f,L_f} = H_f$. As for Part (ii), it is obtained by first showing that $T_{\vartheta+n^{1/2}\tau,f} - T_{\vartheta,f} + H_f(I_p \otimes \Lambda^{-1})\tau_2$ is $o_P(1)$ as $n \to \infty$, under $P^n_{\vartheta,f}$ (which is a direct consequence of the equivalent *asymptotic linearity* formulation of ULAN at $f \in \mathcal{F}_{\text{ULAN}}$), then by using Lemma 4.4 of Kreiss (1987). □

A.4. Proofs of Theorems 3.1, 4.1, and 5.1

Those proofs are based on both following lemmas. The first one is actually Theorem 9.2.1 in Rao and Mitra (1971), and the second one follows from standard algebra.

Lemma A.5. Let Y be a Gaussian k-vector with mean μ and covariance matrix Σ (that may be singular). Let A be a $k \times k$ matrix. Then $Y'AY \sim \chi^2_d(\delta)$ for some d and δ if and only if (i) $\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$, (ii) $\Sigma A \mu \in \mathcal{M}(\Sigma A \Sigma)$, and (iii) $\mu' A' \Sigma A \mu = \mu' A \mu$, in which case, $d = \text{tr}(A\Sigma)$ and $\delta = \mu' A' \Sigma A \Sigma A \mu$.

Lemma A.6. (i) $M_{K,L}$ is the Moore-Penrose pseudoinverse of $H_{K,L}$ (in particular, $M_{K,L} H_{K,L} M_{K,L} = M_{K,L}$), and (ii) $\text{tr}(M_{K,L} H_{K,L}) = d(K,L)$.

Proof of Theorem 3.1. In this proof, whenever we refer to Lemma A.6, it is for $K = K_f = \varphi_f \circ F_+^{-1}$ and $L = L_f = F_+^{-1}$ (note that $H_{K_f,L_f} = H_f$, $M_{K_f,L_f} = M_f$, and $d(K_f, L_f) = d(f)$). Now, since $\hat{\vartheta}$ satisfies Assumption (B(i)) and $\vartheta \in \mathcal{M}(\Omega)$, we have that $M_f H_f (I_p \otimes \Lambda^{-1}) \text{vec}(\hat{\Lambda} - \Lambda) = 0$. Hence, Lemma A.4(ii) yields that

$$Q_f = T'_{\vartheta,f}\, M_f\, T_{\vartheta,f} + o_P(1), \tag{A.7}$$

as $n \to \infty$, under $P^n_{\vartheta,f}$, with $\vartheta \in \mathcal{M}(\Omega)$. Part (i) of the theorem then follows trivially from Lemma A.5, Lemma A.4(i), and Lemma A.6.

As for Part (ii), contiguity implies that (A.7) also holds under local alternatives of the form $P^n_{\vartheta+n^{-1/2}\tau,f}$, with $\vartheta \in \mathcal{M}(\Omega)$ and $\tau \notin \mathcal{M}(\Omega)$. Hence, Lemma A.5 (Conditions (i)-(iii) readily follow from Lemma A.6(i)) yields that Q_f is asymptotically $\chi^2_{d(f)}(\delta)$ under $P^n_{\vartheta+n^{-1/2}\tau,f}$, with $\delta = (H_f(I_p \otimes \Lambda^{-1})\tau_2)' M_f H_f M_f H_f M_f (H_f(I_p \otimes \Lambda^{-1})\tau_2) = \tau_2'(I_p \otimes \Lambda^{-1})' H_f M_f H_f (I_p \otimes \Lambda^{-1})\tau_2$ (where we used again Lemma A.6(i)).

Eventually, Part (iii) follows from the general form (in (3.1)) of locally and asymptotically most stringent tests (see Le Cam (1986), Section 11.9). □

Proof of Theorem 4.1. Applying Lemma 4.2 then Lemma 4.1 yields that, as $n \to \infty$, under $P^n_{\vartheta,g}$, with $\vartheta \in \mathcal{M}(\Omega)$,

$$\begin{aligned} \tilde{Q}_{K,L} &= \tilde{T}'_{\vartheta,K,L}\, M_{K,L} \tilde{T}_{\vartheta,K,L} + o_P(1) \\ &= T'_{\vartheta,K,L;g}\, M_{K,L}\, T_{\vartheta,K,L;g} + o_P(1). \end{aligned} \tag{A.8}$$

Part (i) of the theorem then follows from Lemmas A.5, 4.3, and A.6.

Now, by contiguity, (A.8) also holds under local alternatives of the form $P^n_{\vartheta+n^{-1/2}\tau,g}$, with $\vartheta \in \mathcal{M}(\Omega)$ and $\tau \notin \mathcal{M}(\Omega)$. In order to use Lemma A.5 again, we further need checking that Conditions (ii) and (iii) in that lemma hold: Condition (ii) states that $H_{K,L}M_{K,L}H_{K,L;g}(I_p \otimes \Lambda^{-1})\tau_2 = H_{K,L}M_{K,L}H_{K,L}v$ for some vector v, which is a direct consequence of Lemma A.6(i); as for Condition (iii), it trivially follows from Lemma A.6(i). Hence, Lemma A.5 shows that $\tilde{Q}_{K,L}$ is asymptotically $\chi^2_{d(K,L)}(\delta)$ under $P^n_{\vartheta+n^{-1/2}\tau,g}$, where, by using again Lemma A.6(i),

$$\begin{aligned}\delta &= (H_{K,L;g}(I_p \otimes \Lambda^{-1})\tau_2)' M_{K,L}H_{K,L}M_{K,L}H_{K,L}M_{K,L}(H_{K,L;g}(I_p \otimes \Lambda^{-1})\tau_2) \\ &= \tau_2'(I_p \otimes \Lambda^{-1})' H'_{K,L;g} M_{K,L}H_{K,L;g}(I_p \otimes \Lambda^{-1})\tau_2,\end{aligned}$$

which establishes Part (ii) of the theorem.

As for Part (iii), it is a direct consequence of the asymptotic equivalence, under $\cup_{\vartheta\in\mathcal{M}(\Omega)}\{P^n_{\vartheta,f}\}$ (hence, also under sequences of contiguous alternatives), between $\tilde{Q}_f = \tilde{Q}_{K_f,L_f}$ and the optimal parametric test statistic Q_f (this equivalence follows from (A.7), (A.8), and the identities $T_{\vartheta,K_f,L_f;f} = T_{\vartheta,f}$ and $M_{K_f,L_f} = M_f$). □

Proof of Theorem 5.1. This theorem follows from (5.3) and Lemma A.5, along the same lines as the proof of Theorem 4.1. The required Gaussian version of Lemma A.6, which is obtained for $K = K_\phi$ and $L = L_\phi$, states that (i) $M_\phi H_\phi M_\phi = M_\phi$ and (ii) $\mathrm{tr}(M_\phi H_\phi) = p_1 p_2$. □

A.5. Proof of Proposition 5.2

It only remains to prove Proposition 5.2.

Proof of Proposition 5.2. Part (i) follows—after painful computations—by evaluating (5.4) in the particular case for which $L_r = K_r/\mathrm{E}[K_r^2(U)]$ for all $r = 1,\ldots,p$. Let us therefore concentrate on Part (ii). By using $(x+y)^2 \geq 4xy$ in (5.7), we see that alternatives for which B is symmetric satisfy

$$\mathrm{ARE}_{\vartheta,\tau,g}\big[\tilde{\phi}_{K,L}/\phi_{\mathrm{Wilks}}\big] \geq \frac{\sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2} B_{rs}^2\, a_{r,s}(\bar{K};g)a_{s,r}(\bar{K};g)}{\sum_{r\in\mathcal{S}_1}\sum_{s\in\mathcal{S}_2} B_{rs}^2}, \tag{A.9}$$

which shows that the ARE in (5.7) is bounded from below by a weighted mean of the quantities $a_{r,s}(\bar{K};g)a_{s,r}(\bar{K};g) = \delta_{r,r}(\bar{K},\bar{K};g)\delta_{s,s}(\bar{K},\bar{K};g)$, $r \in \mathcal{S}_1$, $s \in \mathcal{S}_2$. The result for the Wilcoxon test then follows from Hallin and Tribel (2000), which establishes that $\inf_{g\in\mathcal{F}_{\mathrm{ULAN}}} \delta_{r,r}(\bar{K},\bar{K};g) \geq \frac{9\pi^4}{2^{10}}$ for the linear scores $u \mapsto \bar{K}_r(u) = \sqrt{3}u$. As for the van der Waerden test, the result similarly follows from the fact that, for the Gaussian scores $u \mapsto \bar{K}_r(u) = \Phi^{-1}(\frac{u+1}{2})$, one has $\delta_{r,r}(\bar{K},\bar{K};g) \geq 1$ for all $g \in \mathcal{F}_{\mathrm{ULAN}}$, where the equality holds if and only if g_r is Gaussian; see Hallin (1994). □

Acknowledgments

Hannu Oja's and Sara Taskinen's research work was supported by a grant from the Academy of Finland. Davy Paindaveine's was supported by an A.R.C. contract from the Communauté Française de Belgique and by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). The authors are grateful to the reviewers for their careful reading of the original version of this manuscript and for their insightful comments.

References

Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *J. Amer. Statist. Assoc.* **67** 687–690.

Blomqvist, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.* **21** 593–600. MR0039190

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York. MR0856411

Chen, A. and Bickel, P. J. (2006). Efficient independent component analysis. *Ann. Statist.* **34** 2825–2855. MR2329469

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.* **29** 972–994. MR0100322

Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.* **50** 471–491. MR1664575

Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Statist. Math.* **47** 551–579. MR1364260

Gieser, P. W. and Randles, R. H. (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc.* **92** 561–567. MR1467849

Hallin, M. (1994). On the Pitman-nonadmissibility of correlogram-based methods. *J. Time Series Anal.* **15** 607–612. MR1312324

Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape I: Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465

Hettmansperger, T. P. and Randles, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860. MR1946515

Hodges, J. L. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27** 324–335. MR0079383

Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39** 2448–2476. MR2906874

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

Lind, B. and Roussas, G. (1972). A remark on quadratic mean differentiability. *Ann. Math. Statist.* **43** 1030–1034. MR0307329

MÖTTÖNEN, J., HÜSLER, J. and OJA, H. (2003). Multivariate nonparametric tests in randomized complete block design. *J. Multivariate Anal.* **85** 106–129. MR1978180

NORDHAUSEN, K., OJA, H. and PAINDAVEINE, D. (2009). Signed-rank tests for location in the symmetric independent component model. *J. Multivariate Analysis* **100** 821–834. MR2498716

NYBLOM, J. and MÄKELÄINEN, T. (1983). Comparisons of tests for the presence of random walk coefficients in a simple linear model. *J. Amer. Statist. Assoc.* **78** 856–864. MR0727571

PETERS, D. and RANDLES, R. H. (1990). A multivariate signed-rank test for the one-sample location problem. *J. Amer. Statist. Assoc.* **85** 552–557. MR1141757

PILLAI, K. C. S. (1955). Some new test criteria in multivariate analysis. *Ann. Math. Statist.* **26** 117–121. MR0067429

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

PURI, M. L. and SEN, P. K. (1985). *Nonparametric Methods in General Linear Models*. J. Wiley, New York. MR0794309

RANDLES, R. H. (1989). A distribution-free multivariate test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045–1050. MR1134492

RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverses of Matrices and its Applications*. J. Wiley, New York. MR0338013

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

SWENSEN, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16** 54–70. MR0778489

TASKINEN, S., KANKAINEN, A. and OJA, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21. MR1965367

TASKINEN, S., KANKAINEN, A. and OJA, H. (2004). Rank scores tests of multivariate independence. In *Theory and Applications of recent Robust Methods* A. S. M. HUBERT, G. PISON and S. VAN AELST, eds. 329–342. Birkhauser, Basel. MR2088309

TASKINEN, S., OJA, H. and RANDLES, R. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.* **100** 916–925. MR2201019

THEIS, F. J. (2004). A new concept for separability problems in blind source separation. *Neural Comput.* **16** 1827–1850.

TYLER, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251. MR0885734

TYLER, D. E., CRITCHLEY, F., DÜMBGEN, L. and OJA, H. (2009). Invariant co-ordinate selection. *J. Roy. Statist. Soc. Ser. B* **71** 549–592. MR2749907

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