Signed-rank Tests for Location in the Symmetric Independent Component Model

Klaus Nordhausen [a,*] Hannu Oja [a] Davy Paindaveine [b]

[a] *Tampere School of Public Health, University of Tampere, 33014 University of Tampere, FINLAND*

[b] *E.C.A.R.E.S., Institut de Recherche en Statistique, and Département de Mathématique, Université Libre de Bruxelles, Campus de la Plaine CP 210, 1050 Bruxelles, BELGIUM*

Abstract

The so-called independent component (IC) model states that the observed p-vector X is generated via $X = \Lambda Z + \mu$, where μ is a p-vector, Λ is a full-rank matrix, and the centered random vector Z has independent marginals. We consider the problem of testing the null hypothesis $\mathcal{H}_0 : \mu = 0$ on the basis of i.i.d. observations $X_1, \ldots, X_n$ generated by the symmetric version of the IC model above (for which all ICs have a symmetric distribution about the origin). In the spirit of Hallin & Paindaveine (2002a), we develop nonparametric (signed-rank) tests, which are valid without any moment assumption and are, for adequately chosen scores, locally and asymptotically optimal (in the Le Cam sense) at given densities. Our tests are measurable with respect to the marginal signed ranks computed in the collection of null residuals $\hat{\Lambda}^{-1} X_i$, where $\hat{\Lambda}$ is a suitable estimate of Λ. Provided that $\hat{\Lambda}$ is affine-equivariant, the proposed tests, unlike the standard marginal signed-rank tests developed in Puri & Sen (1971) or any of their obvious generalizations, are affine-invariant. Local powers and asymptotic relative efficiencies (AREs) with respect to Hotelling's T^2 test are derived. Quite remarkably, when Gaussian scores are used, these AREs are always greater than or equal to one, with equality in the multinormal model only. Finite-sample efficiencies and robustness properties are investigated through a Monte-Carlo study.

Key words: One-sample location problem, Rank tests, Independent component models, Elliptical symmetry, Hotelling's test, Local asymptotic normality

* Corresponding author.
Email address: klaus.nordhausen@uta.fi (Klaus Nordhausen).

1 Introduction

Let $X_1, \ldots, X_n$ be a sample of p-variate random vectors generated by the location-scatter model

$$X_i = \Lambda Z_i + \mu, \quad i = 1, \ldots, n,$$

where the p-vector μ is the location center, the full-rank $p \times p$ matrix Λ is called the *mixing matrix*, and the Z_i's are i.i.d. *standardized* p-variate random vectors. We consider the multivariate one-sample location problem, that is, we wish to test $\mathcal{H}_0 : \mu = 0$ versus $\mathcal{H}_1 : \mu \neq 0$ (any other null value μ_0 can be tested by replacing X_i with $X_i - \mu_0$). Of course, different standardizations of the Z_i's lead to different location-scatter models—and to different definitions of μ and Λ. Such models include

- *The multinormal model:* Z_i has a standard multinormal distribution. This is a parametric model with mean vector μ and covariance matrix $\Sigma = \Lambda\Lambda'$.
- *The elliptic model:* Z_i has a spherical distribution around the origin ($OZ_i \stackrel{\mathcal{D}}{=} Z_i$ for any orthogonal $p \times p$ matrix O; throughout, $\stackrel{\mathcal{D}}{=}$ stands for equality in distribution) with $\text{Med}[\|Z_i\|^2] = \chi^2_{p,.5}$, where $\text{Med}[\cdot]$ denotes the population median and $\chi^2_{\ell,\alpha}$ denotes the α quantile of the χ^2_ℓ distribution. This is a semiparametric model with symmetry center μ and scatter matrix $\Sigma = \Lambda\Lambda'$ (in the multinormal submodel, Σ is the covariance matrix).
- *The symmetric independent component (IC) model:* the components of Z_i are independent and symmetric ($-Z_i^{(r)} \stackrel{\mathcal{D}}{=} Z_i^{(r)}$) with $\text{Med}[(Z_i^{(r)})^2] = \chi^2_{1,.5}$, $r = 1, \ldots, p$. This is a semiparametric model with symmetry center μ and mixing matrix Λ (again, in the multinormal submodel, $\Sigma = \Lambda\Lambda'$ is the covariance matrix). This model is used in the so-called independent component analysis (ICA), where the problem is to estimate Λ.
- *The symmetric nonparametric model:* Z_i has a distribution symmetric around the origin ($-Z_i \stackrel{\mathcal{D}}{=} Z_i$). Then, neither Λ nor Σ are uniquely defined.

Note that the semiparametric/nonparametric models above do not require any moment assumption, and that μ, irrespective of the model adopted, is properly identified as the center of symmetry of X_i. The assumption of symmetry is common in the one-sample location case. It is for example quite natural in the classical matched pairs design for the comparison of two treatments: if for pair i, $i = 1, \ldots, n$, the response variable is $X_{1i} = Y_i + \varepsilon_{1i} + \mu_1$ for treatment 1 and $X_{2i} = Y_i + \varepsilon_{2i} + \mu_2$ for treatment 2, with mutually independent Y_i, ε_{1i}, and ε_{2i} ($\stackrel{\mathcal{D}}{=} \varepsilon_{1i}$), then the difference used in the analysis, namely $X_i = X_{2i} - X_{1i}$, is symmetric about $\mu = \mu_2 - \mu_1$. The literature proposes a vast list of multivariate one-sample location tests. Some of the tests do not require symmetry; note however that only in the symmetric case the different tests are for the same population quantity. The tests include

- *The Hotelling's T^2 test*, which is equivalent to the Gaussian likelihood ratio test (and actually is uniformly most powerful affine-invariant at the multinormal), is asymptotically valid (i.e., asymptotically meets the nominal level constraint) under any distribution with finite variances. However, its power is poor away from the multinormal (particularly so under heavy tails), and it is also very sensitive to outlying observations.
- *The optimal signed-rank scores tests by Hallin and Paindaveine (2002a,b)* are based on standardized spatial signs (or Randles' interdirections; see Randles (1989) for the corresponding sign test) and the ranks of Mahalanobis distances between the data points and the origin. They do not require any moment assumption and are optimal (in the Le Cam sense) at correctly specified (elliptical) densities. They are affine-invariant, robust, and highly efficient under a broad range of densities (AREs of their Gaussian-score version with respect to Hotelling's test are uniformly larger than or equal to one in the elliptic model). Later Oja & Paindaveine (2005) showed that interdirections together with the so-called lift-interdirections allow for building hyperplane-based versions of these tests. All these tests however strictly require ellipticity.
- *The signed-rank scores tests by Puri & Sen (1971)* combine marginal signed-rank scores tests in the widest symmetric nonparametric model. Unfortunately, these tests are not affine-invariant and may be poorly efficient for dependent margins. Invariant tests are obtained if the data points are first transformed to invariant coordinates; see Chakraborty & Chaudhuri (1999) and Nordhausen et al. (2006).
- *The spatial sign and signed-rank tests* (for a review, see Möttönen & Oja, 1995), which are based on spatial signs and signed ranks, are also valid in the symmetric nonparametric model. They improve over the Puri and Sen tests in terms of efficiency, but not in terms of affine-invariance. Again, affine-invariance can be achieved if the data is first transformed by using any scatter matrix (the spatial sign test based on Tyler (1987)'s scatter matrix is strictly distribution-free in the elliptic model and even in the wider directional elliptic model; see Randles (2000)).
- *The sign and signed-rank tests by Hettmansperger et al. (1994, 1997)* are based on multivariate Oja signs and ranks. They can be used in all models above, are asymptotically equivalent to spatial sign and signed-rank tests in the spherical case, and are affine-invariant. However, at the elliptic model, their efficiency (as well as that of the spatial sign and signed-rank tests) may be poor when compared with the Hallin and Paindaveine tests.

Only the Hallin & Paindaveine (2002a,b) and Oja & Paindaveine (2005) tests combine robustness and affine-invariance with a locally optimal—and uniformly excellent—power behavior. The required ellipticity assumption, however, may not be appropriate in practice. This model assumption is often easily discarded just by a visual inspection of bivariate scatter plots or marginal density plots; equidensity contours should be elliptical, and the marginal densities should be similar in shape. The IC model which serves as an alternative extension of the multivariate normal model cannot be ruled out as easily in practice. Of course, more statistical tools should be developed for the important model choice problem.

This paper introduces signed-rank tests which enjoy the nice properties of the Hallin & Paindaveine (2002a) ones (absence of moment assumption, robustness, affine-invariance, Le Cam optimality at prespecified densities, uniform dominance over Hotelling for Gaussian scores, etc.), but are valid in the *symmetric IC model*. The proposed tests are marginal signed-rank tests (with optimal scores) applied to the residuals $\hat{\Lambda}^{-1}X_i$, $i = 1, \ldots, n$, where $\hat{\Lambda}$ is a suitable (see Section 3) estimate of the mixing matrix Λ. Although they are based on marginal signed-rank statistics, our tests, unlike the marginal Puri and Sen signed-rank tests or any of their obvious generalizations, are affine-invariant.

The outline of the paper is as follows. Section 2 defines more carefully the IC models under consideration. Section 3 introduces the proposed tests and studies their asymptotic null behavior. In Section 4, we explain how to choose score functions to achieve Le Cam optimality at prespecified densities, derive the local powers of our tests under contiguous alternatives, and compute their AREs with respect to Hotelling's T^2 test. Section 5 discusses the practical implementation of our tests and presents a simulation that investigates their finite-sample efficiencies and robustness properties. Finally, the appendix collects proofs of technical results.

2 IC models and identifiability

In the absolutely continuous case, the IC model will be indexed by the location vector μ, mixing matrix Λ, and the pdf g of the standardized vectors. The location vector μ is a p-vector and Λ belongs to the collection $\mathcal{M}_p$ of invertible $p \times p$ matrices. As for g, it throughout belongs to the collection $\mathcal{F}$ of densities of absolutely continuous p-vectors $Z = (Z^{(1)}, \ldots, Z^{(p)})'$ whose marginals are (i) mutually independent, (ii) symmetric about the origin (i.e., $-Z^{(r)} \stackrel{\mathcal{D}}{=} Z^{(r)}$ for all r), and (iii) standardized so that $\text{Med}[(Z^{(r)})^2] = \chi^2_{1,.5}$ for all $r = 1, \ldots, p$. Any $g \in \mathcal{F}$ of course decomposes into $z = (z^{(1)}, \ldots, z^{(p)})' \mapsto g(z) =: \prod_{r=1}^{p} g_r(z^{(r)})$.

Denote then by $P^n_{\mu,\Lambda,g}$, $g \in \mathcal{F}$, the hypothesis under which the p-variate observations $X_1, \ldots, X_n$ are generated by the model $X_i = \Lambda Z_i + \mu$, $i = 1, \ldots, n$, where $Z_i = (Z_i^{(1)}, \ldots, Z_i^{(p)})'$, $i = 1, \ldots, n$ are i.i.d. with pdf g. Clearly, the likelihood, under $P^n_{\mu,\Lambda,g}$, is given by $L^n_{\mu,\Lambda,g} = |\det \Lambda|^{-n} \prod_{i=1}^{n} (\prod_{r=1}^{p} g_r(e_r'\Lambda^{-1}(X_i - \mu)))$, where e_r is the vector with a one in position r and zeros elsewhere.

In the symmetric IC model above, the location parameter μ is the unique center of symmetry of the common distribution of the X_i's and therefore is a well-defined parameter. In sharp contrast, the parameters Λ and g are not identifiable: letting P be any $p \times p$ permutation matrix and S be any $p \times p$ diagonal matrix with diagonal entries in $\{-1, 1\}$, one can write $X_i = (\Lambda PS)(SP^{-1}Z_i) + \mu =: \tilde{\Lambda}\tilde{Z}_i + \mu$, where $\tilde{Z}_i$ still satisfies (i), (ii) and (iii) above. If $\tilde{g}$ is the density of $\tilde{Z}_i$, then $P^n_{\mu,\Lambda,g} = P^n_{\mu,\tilde{\Lambda},\tilde{g}}$. This indeterminacy can be avoided by requiring, for instance, that marginal densities are given in a specified (e.g., kurtosis) order and that the entry having largest absolute value in each column of Λ is positive.

In the independent component analysis (ICA) one wishes to find an estimate of *any* Λ such that $\Lambda^{-1}X_i$ has independent components. If $\Lambda^{-1}X_i$ has independent components then so has $DSP\Lambda^{-1}X_i$, where D is any diagonal matrix with positive diagonal elements. This same identifiability problem is well recognized in the ICA literature, and it has been proven (see, e.g., Theis (2004) for a simple proof) that these three sources of non-identifiability are the only ones, *provided that not more than one IC is Gaussian*, an assumption that is therefore made throughout in the ICA literature. Note that the third source of non-identifiability D is avoided in our model building by fixing the scales of the marginals of Z_i in (iii) above. In the classical ICA the estimation of Λ is the main goal, whereas in our problem it is only a primary device to yield the components used for the testing. The sign-change or permutation of the components will not be a problem in our test construction. We naturally also would like to deal with distributions where there are more than one Gaussian IC. In particular, we do not want to rule out the multinormal case, for which all ICs are Gaussian! Quite fortunately, the resulting lack of identifiability will not affect the behavior of our tests (we discuss this further in Section 5).

3 The proposed tests

Define the (null) residual associated with observation X_i and value Λ of the mixing matrix as $Z_i(\Lambda) := \Lambda^{-1}X_i$. The signed ranks of these residuals are the quantities $S_i(\Lambda)R_i(\Lambda)$, with $S_i(\Lambda) := (S_i^{(1)}(\Lambda), \ldots, S_i^{(p)}(\Lambda))'$ and $R_i(\Lambda) := (R_i^{(1)}(\Lambda), \ldots, R_i^{(p)}(\Lambda))'$, $i = 1, \ldots, n$, where $S_i^{(r)}(\Lambda) := I_{[Z_i^{(r)}(\Lambda)>0]} - I_{[Z_i^{(r)}(\Lambda)<0]}$ is the sign of $Z_i^{(r)}(\Lambda)$ and $R_i^{(r)}(\Lambda)$ is the rank of $|Z_i^{(r)}(\Lambda)|$ among $|Z_1^{(r)}(\Lambda)|, \ldots, |Z_n^{(r)}(\Lambda)|$. Let $K^{(r)} : (0, 1) \to \mathbb{R}$, $r = 1, \ldots, p$ be score functions and consider the corresponding p-variate score function K defined by $u = (u^{(1)}, \ldots, u^{(p)})' \mapsto K(u) := (K^{(1)}(u^{(1)}), \ldots, K^{(p)}(u^{(p)}))'$. We throughout assume that the $K^{(r)}$'s are (i) continuous, (ii) satisfy $\int_0^1 (K^{(r)}(u))^{2+\delta}\, du < \infty$ for some $\delta > 0$, and (iii) can be expressed as the difference of two monotone increasing functions. These assumptions are required for Hájek's classical projection result for linear signed-rank statistics; see, e.g., Puri & Sen (1985), Chapter 3 (actually, Hájek's result requires square-integrability rather than the reinforcement of square integrability in (ii); we will need the latter however to control the unspecification of Λ; see the proof of Lemma 3.3 below).

The (K-score version of the) test statistic we propose is then

$$Q_K(\Lambda) := (T_K(\Lambda))'\Gamma_K^{-1}\, T_K(\Lambda),$$

where $T_K(\Lambda) := n^{-1/2}\sum_{i=1}^n T_{K;i}(\Lambda) := n^{-1/2}\sum_{i=1}^n [S_i(\Lambda)\odot K(\frac{R_i(\Lambda)}{n+1})]$ and $\Gamma_K :=$ diag$(\mathrm{E}[(K^{(1)}(U))^2], \ldots, \mathrm{E}[(K^{(p)}(U))^2])$; throughout, $\odot$ denotes the Hadamard (i.e., entrywise) product and U stands for a random variable that is uniformly distributed over $(0,1)$.

The asymptotic behavior of $Q_K(\Lambda)$ can be investigated quite easily by using the *representation* result in Lemma 3.1 below. In order to state this result, we define $z = (z^{(1)}, \ldots, z^{(p)})' \mapsto G_+(z) := (G_+^{(1)}(z^{(1)}), \ldots, G_+^{(p)}(z^{(p)}))'$, where $G_+^{(r)}$ stands for the cdf of $|Z_1^{(r)}(\Lambda)|$ under $P^n_{0,\Lambda,g}$. Symmetry of g_r yields $G_+^{(r)}(t) = 2G^{(r)}(t) - 1$, where $t \mapsto G^{(r)}(t) = \int_{-\infty}^t g_r(s)\, ds$ is the cdf of $Z_1^{(r)}(\Lambda)$ under $P^n_{0,\Lambda,g}$.

Lemma 3.1 *Define* $T_{K;g}(\Lambda) := n^{-1/2}\sum_{i=1}^n T_{K;g;i}(\Lambda) := n^{-1/2}\sum_{i=1}^n [S_i(\Lambda)\odot K(G_+(|Z_i(\Lambda)|))]$, *where* $|Z_i(\Lambda)| := (|Z_i^{(1)}(\Lambda)|, \ldots, |Z_i^{(p)}(\Lambda)|)'$. *Then, for any* $\Lambda \in \mathcal{M}_p$ *and* $g \in \mathcal{F}$, $\mathrm{E}[\|T_K(\Lambda) - T_{K;g}(\Lambda)\|^2] = o(1)$ *as* $n \to \infty$, *under* $P^n_{0,\Lambda,g}$.

Lemma 3.1 implies that under the null—hence also under sequences of contiguous alternatives (see Section 4.2 for the form of those alternatives)—$T_K(\Lambda)$ is asymptotically equivalent to $T_{K;g}(\Lambda)$, where g is the "true" underlying noise density. Since $T_{K;g}(\Lambda)$ is a sum of i.i.d. terms, the asymptotic null distribution of $T_K(\Lambda)$ then follows from the multivariate CLT.

Lemma 3.2 *For any* $\Lambda \in \mathcal{M}_p$, $T_K(\Lambda)$, *under* $\cup_{g\in\mathcal{F}}\{P^n_{0,\Lambda,g}\}$, *is asymptotically multinormal with mean zero and covariance matrix* Γ_K.

It readily follows from Lemma 3.2 that $Q_K(\Lambda)$, under $\cup_{g\in\mathcal{F}}\{P^n_{0,\Lambda,g}\}$, is asymptotically chi-square with p degrees of freedom. The resulting test therefore consists in rejecting the null at asymptotic level α iff $Q_K(\Lambda) > \chi^2_{p,1-\alpha}$.

Of course, as already mentioned, Λ in practice is unspecified and should be replaced with some suitable estimate $\hat{\Lambda}$. The choice of this estimate is discussed in Section 5, but we will throughout assume that $\hat{\Lambda}$ is (i) root-n consistent, (ii) invariant under permutations of the observations, and (iii) invariant under individual reflections of the observations with respect to the origin (i.e., $\hat{\Lambda}(s_1X_1, \ldots, s_nX_n) = \hat{\Lambda}(X_1, \ldots, X_n)$ for all $s_1, \ldots, s_n \in \{-1, 1\}$). The replacement of Λ with $\hat{\Lambda}$ in $Q_K(\Lambda)$ yields the genuine test statistic $\hat{Q}_K := Q_K(\hat{\Lambda})$. The following result establishes that this replacement has no effect on the asymptotic null behavior of the test (see the appendix for a proof).

Lemma 3.3 *For any $\Lambda \in \mathcal{M}_p$, $T_K(\hat{\Lambda}) = T_K(\Lambda) + o_P(1)$ (hence also $\hat{Q}_K = Q_K(\Lambda) + o_P(1)$) as $n \to \infty$, under $\cup_{g\in\mathcal{F}}\{P^n_{0,\Lambda,g}\}$.*

The following theorem, which is the main result of this section, is then a direct corollary of Lemmas 3.2 and 3.3.

Theorem 3.1 *Under $\cup_{\Lambda\in\mathcal{M}_p} \cup_{g\in\mathcal{F}} \{P^n_{0,\Lambda,g}\}$, $\hat{Q}_K$ is asymptotically χ^2_p, so that, still under $\cup_{\Lambda\in\mathcal{M}_p} \cup_{g\in\mathcal{F}} \{P^n_{0,\Lambda,g}\}$, the test ϕ_K that rejects the null as soon as $\hat{Q}_K > \chi^2_{p,1-\alpha}$ has asymptotic level α.*

The behavior of our tests under local alternatives will be studied in Section 4.

Let us finish this section with some particular cases of the proposed test statistics $\hat{Q}_K$. To this end, write $\hat{S}_i$ and $\hat{R}_i$ for the empirical signs $S_i(\hat{\Lambda})$ and ranks $R_i(\hat{\Lambda})$, respectively. Then (i) sign test statistics are obtained with constant score functions ($K^{(r)}(u) = 1$ for all r, say). The resulting test statistics are

$$\hat{Q}_S = \hat{T}'_S\hat{T}_S = \frac{1}{n}\sum_{i,j=1}^{n} \hat{S}'_i\hat{S}_j = \frac{1}{n}\sum_{i,j=1}^{n}\sum_{r=1}^{p} \hat{S}^{(r)}_i\hat{S}^{(r)}_j. \tag{1}$$

(ii) Wilcoxon-type test statistics, associated with linear score functions ($K^{(r)}(u) = u$ for all r, say), take the form

$$\hat{Q}_W = 3\,\hat{T}'_W\hat{T}_W = \frac{3}{n(n+1)^2}\sum_{i,j=1}^{n}\sum_{r=1}^{p} \hat{S}^{(r)}_i\hat{S}^{(r)}_j\hat{R}^{(r)}_i\hat{R}^{(r)}_j. \tag{2}$$

(iii) Gaussian (or *van der Waerden*) scores are obtained with $K^{(r)}(u) = \Phi^{-1}_+(u) = \Phi^{-1}((u+1)/2)$, where Φ is the cdf of the standard normal distribution. The corresponding test statistics are

$$\hat{Q}_{\mathrm{vdW}} = \hat{T}'_{\mathrm{vdW}}\hat{T}_{\mathrm{vdW}} = \frac{1}{n}\sum_{i,j=1}^{n}\sum_{r=1}^{p} \hat{S}^{(r)}_i\hat{S}^{(r)}_j\Phi^{-1}_+\Big(\frac{\hat{R}^{(r)}_i}{n+1}\Big)\Phi^{-1}_+\Big(\frac{\hat{R}^{(r)}_j}{n+1}\Big). \tag{3}$$

As we show in the next section, this van der Waerden test is optimal in the Le Cam sense (more precisely, locally and asymptotically maximin) at the multinormal submodel.

4 Optimality, local powers, and AREs

In this section, we exploit Le Cam's theory of asymptotic experiments in order to define versions of our tests that achieve Le Cam optimality under correctly specified noise densities. We also study the behavior of our tests under sequences of local alternatives and compare their asymptotic performances with those of Hotelling's T^2 test in terms of asymptotic relative efficiencies (AREs).

The main technical result here is the locally and asymptotically normal (LAN) structure of the IC model with respect to μ, for fixed values of Λ and g. Such LAN property requires more stringent assumptions on g. Define accordingly $\mathcal{F}_{\mathrm{LAN}}$ as the collection of noise densities $g \in \mathcal{F}$ that (i) are absolutely continuous and (ii) have finite Fisher information for location, i.e., $\mathcal{I}_{g_r} := \int_{-\infty}^{\infty} (\varphi_{g_r}(z))^2 g_r(z)\, dz < \infty$ for all r, where, denoting by g_r' the a.e.-derivative of g_r, we let $\varphi_{g_r} := -g_r'/g_r$. For $g \in \mathcal{F}_{\mathrm{LAN}}$, 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 LAN result, which is an immediate corollary of the more general result established in Oja et al. (2008).

Proposition 4.1 *For any $\Lambda \in \mathcal{M}_p$ and $g \in \mathcal{F}_{\mathrm{LAN}}$, the family of distributions $\mathcal{P}^n_{\Lambda,g} := \{P^n_{\mu,\Lambda,g}, \mu \in \mathbb{R}^p\}$ is LAN. More precisely, for any p-vector μ and any bounded sequence of p-vectors (τ_n), we have that (letting $S_i(\mu,\Lambda)$ stand for the sign of $Z_i(\mu,\Lambda) := \Lambda^{-1}(X_i - \mu)$) (i) under $P^n_{\mu,\Lambda,g}$, as $n \to \infty$,*

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

with central sequence $\Delta^{(n)}_{\mu,\Lambda,g} := n^{-1/2}(\Lambda^{-1})' \sum_{i=1}^n \varphi_g(Z_i(\mu,\Lambda)) = n^{-1/2}(\Lambda^{-1})' \sum_{i=1}^n [S_i(\mu,\Lambda) \odot \varphi_g(|Z_i(\mu,\Lambda)|)]$ *and* information matrix $\Gamma_{\Lambda,g} := (\Lambda^{-1})'\mathcal{I}_g\Lambda^{-1} := (\Lambda^{-1})'\mathrm{diag}(\mathcal{I}_{g_1}, \ldots, \mathcal{I}_{g_p})\Lambda^{-1}$, *and that (ii) still under $P^n_{\mu,\Lambda,g}$, $\Delta^{(n)}_{\mu,\Lambda,g}$ is asymptotically multinormal with mean zero and covariance matrix $\Gamma_{\Lambda,g}$.*

Fix now some noise density $f \in \mathcal{F}_{\mathrm{LAN}}$. Le Cam's theory of asymptotic experiments (see, e.g., Chapter 11 of Le Cam (1986)) implies that an f-optimal (actually, locally and asymptotically maximin at f) test for $\mathcal{H}_0 : \mu = 0$ versus $\mathcal{H}_1 : \mu \neq 0$, under fixed $\Lambda \in \mathcal{M}_k$, consists, at asymptotic level α, in rejecting the null as soon as

$$Q_f(\Lambda) := \left(\Delta^{(n)}_{0,\Lambda,f}\right)' \Gamma^{-1}_{\Lambda,f}\, \Delta^{(n)}_{0,\Lambda,f} > \chi^2_{p,1-\alpha}.$$

Letting K_f be the p-variate score function defined by $K^{(r)} := \varphi_{f_r} \circ F^{-1}_{+r}$, $r = 1, \ldots, p$ (with the same notation as in Section 3), one straightforwardly checks that $Q_f(\Lambda) = (T_{K_f;f}(\Lambda))'\, \Gamma^{-1}_{K_f}\, T_{K_f;f}(\Lambda)$, which, by Lemmas 3.1 and 3.3 (provided that the score function K_f satisfies the assumptions of Section 3), is asymptotically equivalent to $\hat{Q}_{K_f}$ under $P^n_{0,\Lambda,f}$. Therefore, denoting by $\mathcal{F}^{\mathrm{opt}}_{\mathrm{LAN}}$ the collection of densities $f \in \mathcal{F}_{\mathrm{LAN}}$ for which the K_{f_r}'s (i) are continuous, (ii) satisfy $\int_0^1 (K_{f_r}(u))^{2+\delta}\, du < \infty$ for some $\delta > 0$, and (iii) can be expressed as the difference of two monotone increasing functions, we have proved the following.

Theorem 4.1 *For any $f \in \mathcal{F}^{\mathrm{opt}}_{\mathrm{LAN}}$, the test ϕ_{K_f} that rejects the null as soon as $\hat{Q}_{K_f} > \chi^2_{p,1-\alpha}$ (i) has asymptotic level α under $\cup_{\Lambda \in \mathcal{M}_p} \cup_{g \in \mathcal{F}} \{P^n_{0,\Lambda,g}\}$ and (ii)* *is locally and asymptotically maximin, at asymptotic level* α*, for* $\cup_{\Lambda\in\mathcal{M}_p}\cup_{g\in\mathcal{F}}\{P^n_{0,\Lambda,g}\}$ *against alternatives of the form* $\cup_{\mu\neq 0}\cup_{\Lambda\in\mathcal{M}_p}\{P^n_{\mu,\Lambda,f}\}$*.*

This justifies the claim (see the end of the previous section) stating that the van der Waerden version of the proposed signed-rank tests is optimal at the multinormal model. More generally, Theorem 4.1 indicates how to achieve Le Cam optimality at a fixed (smooth) noise density f.

4.2 Local powers and asymptotic relative efficiencies

Local powers of our signed-rank tests ϕ_K under local alternatives of the form $P^n_{n^{-1/2}\tau,\Lambda,g}$, $g\in\mathcal{F}_{\rm LAN}$ can be straightforwardly computed from the following result (the proof is given in the appendix).

Theorem 4.2 *Fix* $g\in\mathcal{F}_{\rm LAN}$ *and define* $I_{K,g}:=\mathrm{diag}(I_{K^{(1)},g_1},\ldots,I_{K^{(p)},g_p})$*, with* $I_{K^{(r)},g_r}:=\mathrm{E}[K^{(r)}(U)\,\varphi_{g_r}((G_+^{(r)})^{-1}(U))]$*, where* U *is uniformly distributed over* $(0,1)$*. Then,* $\hat{Q}_K$ *is asymptotically* $\chi^2_p(\tau'(\Lambda^{-1})'I_{K,g}\Gamma_K^{-1}I_{K,g}\Lambda^{-1}\tau)$ *under* $P^n_{n^{-1/2}\tau,\Lambda,g}$*, where* $\chi^2_\ell(c)$ *stands for the noncentral chi-square distribution with* ℓ *degrees of freedom and noncentrality parameter* c*.*

This also allows for computing asymptotic relative efficiencies (AREs) with respect to our benchmark competitor, namely Hotelling's T^2 test. In the following result (see the appendix for a proof), we determine these AREs at any g belonging to the collection $\mathcal{F}^2_{\rm LAN}$ of noise densities in $\mathcal{F}_{\rm LAN}$ with finite variances. We want to stress however that our signed-rank tests ϕ_K, unlike Hotelling's test, remain valid without such moment assumption, so that, when the underlying density does not admit a finite variance, the ARE of any ϕ_K with respect to Hotelling's test actually can be considered as being infinite.

Theorem 4.3 *Fix* $g\in\mathcal{F}^2_{\rm LAN}$*. Then the asymptotic relative efficiency of* ϕ_K *with respect to Hotelling's* T^2 *test, when testing* $\mathcal{H}_0:\mu=0$ *against* $\mathcal{H}_1(\tau):\mu=n^{-1/2}\tau$*, under mixing matrix* $\Lambda\in\mathcal{M}_p$ *and noise density* g*, is given by*

$$\mathrm{ARE}_{\Lambda,\tau,g}[\phi_K,T^2]=\frac{\tau'(\Lambda^{-1})'I_{K,g}\Gamma_K^{-1}I_{K,g}\Lambda^{-1}\tau}{\tau'(\Lambda^{-1})'\Sigma_g^{-1}\Lambda^{-1}\tau},\qquad(4)$$

where $\Sigma_g:=\mathrm{diag}(\sigma^2_{g_1},\ldots,\sigma^2_{g_p})$*, with* $\sigma^2_{g_r}:=\int_{-\infty}^{z}z^2g_r(z)\,dz$*.*

For $p=1$, ϕ_K (resp., T^2) boils down to the standard univariate location signed-rank test $\phi_K^{\rm univ}$ based on the score function K (resp., to the one-sample Student test St), and the ARE in (4) reduces to the well-known result

$$\mathrm{ARE}^{\rm univ}_{\Lambda,\tau,g}[\phi_K^{\rm univ},St]=\frac{\sigma_g^2\,I^2_{K,g}}{\mathrm{E}[K^2(U)]},\qquad(5)$$

which does not depend on τ, nor on Λ. For $p \geq 2$, however, the ARE in (4) depends on τ and Λ. Letting $v = (v^{(1)}, \ldots, v^{(p)})' := \frac{\Sigma_g^{-1/2}\Lambda^{-1}\tau}{\|\Sigma_g^{-1/2}\Lambda^{-1}\tau\|}$, we can write

$$\mathrm{ARE}_{\Lambda,\tau,g}[\phi_K, T^2] = \sum_{r=1}^{p} (v^{(r)})^2 \frac{\sigma_r^2 I^2_{K^{(r)},g_r}}{\mathrm{E}[(K^{(r)}(U))^2]} = \sum_{r=1}^{p} (v^{(r)})^2 \, \mathrm{ARE}^{\mathrm{univ}}_{\Lambda=1,\tau^{(r)},g_r}[\phi^{\mathrm{univ}}_{K^{(r)}}, St], \tag{6}$$

which shows that $\mathrm{ARE}_{\Lambda,\tau,g}[\phi_K, T^2]$ can be seen as *a weighted mean of the corresponding univariate* AREs (those of the univariate signed-rank tests with respect to Student's). The weights depend on the shift τ through the "standardized" shift $\Lambda^{-1}\tau$; if the latter is in the direction of the rth coordinate axis, then $\mathrm{ARE}_{\Lambda,\tau,g}[\phi_K, T^2] = \mathrm{ARE}^{\mathrm{univ}}_{\Lambda=1,\tau^{(r)},g_r}[\phi^{\mathrm{univ}}_{K^{(r)}}, St]$. In all cases, irrespective of τ and Λ, $\mathrm{ARE}_{\Lambda,\tau,g}[\phi_K, T^2]$ always lies between the smallest and the largest "univariate" AREs in $\{\mathrm{ARE}^{\mathrm{univ}}_{\Lambda=1,\tau^{(r)},g_r}[\phi^{\mathrm{univ}}_{K^{(r)}}, T^2], r = 1, \ldots, p\}$.

This explains that it is sufficient to give numerical values for these univariate AREs. Such values are provided in Table 1, for various scores (sign, Wilcoxon, and van der Waerden scores, as well as scores achieving optimality at fixed t distributions) and various underlying densities (t, Gaussian, and power-exponential densities with lighter-than-normal tails). Power-exponential densities refer to densities of the form $g_\eta(r) = c_\eta \exp(-a_\eta r^{2\eta})$, where c_η is a normalization constant, $\eta > 0$ determines the tail weight, and $a_\eta > 0$ standardizes g_η in the same way as the marginal densities in $\mathcal{F}$ (see Section 2).

		underlying density						
		t_3	t_6	t_{12}	$\mathcal{N}$	e_2	e_3	e_5
score	S	1.621	0.879	0.733	0.637	0.411	0.370	0.347
	W	1.900	1.164	1.033	0.955	0.873	0.881	0.907
	vdW	1.639	1.093	1.020	1.000	1.129	1.286	1.533
	t_{12}	1.816	1.151	1.040	0.981	0.973	1.024	1.102
	t_6	1.926	1.167	1.026	0.936	0.820	0.800	0.779
	t_3	2.000	1.124	0.944	0.820	0.569	0.479	0.385

Table 1
AREs of various univariate signed-rank tests (with sign, Wilcoxon, and van der Waerden scores, as well as scores achieving optimality under t_{12}, t_6, and t_3 densities) with respect to Student's test, under t (with 3, 6, 12 degrees of freedom), Gaussian, and power-exponential densities (with tail parameter $\eta = 2, 3, 5$).

All numerical values for the van der Waerden signed-rank test ϕ_{vdW} in Table 1 are larger than one, except in the normal case, where it is equal to one. This is an empirical illustration of the Chernoff and Savage (1958) result showing that $\mathrm{ARE}^{\mathrm{univ}}_{\Lambda=1,\tau,g}[\phi^{\mathrm{univ}}_{vdW}, St] \geq 1$ for all τ and g (with equality iff g is Gaussian). Hence, (6) entails that, in the IC model under consideration, the AREs of our

p-variate van der Waerden test ϕ_{vdW}, with respect to Hotelling's, are always larger than or equal to one, with equality in the multinormal model only.

Coming back to the general expressions of our AREs in (4) and (6), it is clear (in view of (5)) that, in order to maximize the local powers/AREs above with respect to the score function K, one should maximize the cross-information quantities $I_{K^{(r)},g_r}$, $r = 1, \ldots, p$. The Cauchy-Schwarz inequality shows that $I_{K^{(r)},g_r}$ is maximal at $K^{(r)} = \varphi_{g_r} \circ (G_+^{(r)})^{-1}$, which confirms the rule for determining optimal score functions that was derived in Section 4.1.

5 Practical implementation and simulations

In this section, we first focus on the main issue for the practical implementation of our tests, namely the estimation of the mixing matrix Λ. Several approaches are possible, but the approach presented in Oja et al. (2006) is chosen here. Then finite-sample efficiencies and robustness properties of our tests are investigated through Monte-Carlo studies.

Computations were done using the statistical software package R 2.6.0 (R Development Core Team, 2007). Note that the proposed method for estimating Λ is implemented in the R-package ICS (Nordhausen et al., 2007a), whereas the tests proposed in this paper are implemented in the R-package ICSNP (Nordhausen et al., 2007b). Both packages are available on the CRAN website.

5.1 Estimation of Λ

An interesting way to obtain a root-n consistent estimate of Λ is to use two different root-n consistent scatter matrix estimates as in Oja et al. (2006).

Let X be a p-variate random vector and denote its cdf by F_X. A *scatter matrix functional* S (with respect to the null value of the location center, namely the origin) is a $p \times p$ matrix-valued functional such that $S(F_X)$ is positive definite, symmetric, and affine-equivariant in the sense that $S(F_{AX}) = AS(F_X)A'$, $\forall A \in \mathcal{M}_p$. Examples of scatter matrices are the covariance matrix $S_{\text{cov}}(F_X) := E[XX']$, the scatter matrix based on fourth-order moments $S_{\text{kurt}}(F_X) := E\left[(X'(S_{\text{cov}}(F_X))^{-1}X)XX'\right]$, and Tyler (1987)'s scatter matrix S_{Tyl} defined implicitly by $S_{\text{Tyl}}(F_X) = E\left[(X'(S_{\text{Tyl}}(F_X))^{-1}X)^{-1}XX'\right]$.

As we now show, the mixing matrix Λ can be estimated by using a couple of different scatter matrices (S_1, S_2). Recall that our tests require a root-n consistent estimate of Λ *under the null*, that is, under $\mathcal{P}_0^n := \{P^n_{0,\Lambda,g}, \Lambda \in \mathcal{M}_p, g \in \mathcal{F}\}$.

However, since Λ is not identifiable in $\mathcal{P}_0^n$ (see Section 2), estimation of Λ is an ill-posed problem. We therefore restrict to a submodel by using a couple of scatter matrices (S_1, S_2) as follows.

Define the model $\mathcal{P}_0^n(S_1, S_2)$ as the collection of probability distributions of $(X_1, \ldots, X_n)$ generated by $X_i = \Lambda Z_i$, $i = 1, \ldots, n$, where $Z_i = (Z_i^{(1)}, \ldots, Z_i^{(p)})'$, $i = 1, \ldots, n$ are i.i.d. from a distribution F_Z for which $S_1(F_Z) = I$ and $S_2(F_Z) = \Omega$, where $\Omega = (\Omega_{ij})$ is diagonal with $\Omega_{11} > \Omega_{22} > \ldots > \Omega_{pp} (> 0)$. Theorem 5.3 of Tyler et al. (2008) and our assumption that Z has independent and symmetric marginals imply that $S_\ell(F_Z)$, $\ell = 1, 2$ are diagonal matrices, so that this submodel actually only imposes that the quantities Ω_{rr}, $r = 1, \ldots, p$ are pairwise different. Before discussing the severity of this restriction, we note that $\mathcal{P}_0^n(S_1, S_2)$ takes care of the permutation (and scale) indetermination by merely assuming that the ICs are first standardized in terms of their "S_1-scales" and then ordered according to their "(S_1, S_2)-kurtoses". As for the signs of the ICs, they can be fixed by requiring, e.g., that the entry having largest absolute value in each column of Λ is positive (and similarly with $\hat{\Lambda}$); see Section 2.

Most importantly, the affine-equivariance of S_1 and S_2 then implies that

$$(S_2(F_X))^{-1} S_1(F_X)(\Lambda^{-1})' = (\Lambda^{-1})'\Omega^{-1} \tag{7}$$

(where X stands for a p-variate random vector with the same distribution as X_i, $i = 1, \ldots, n$), that is, Λ^{-1} and Ω^{-1} list the eigenvectors and eigenvalues of $(S_2(F_X))^{-1} S_1(F_X)$, respectively. Replacing $S_1(F_X)$ and $S_2(F_X)$ with their natural estimates $\hat{S}_1$ and $\hat{S}_2$ in (7) yields estimates $\hat{\Lambda}$ and $\hat{\Omega}$. Clearly, if $\hat{S}_1$ and $\hat{S}_2$ are root-n consistent, then $\hat{\Lambda}$ is root-n consistent as well. Since our tests are based on statistics that are invariant under heterogeneous rescaling and reordering of the ICs, their versions based on such a $\hat{\Lambda}$ will remain valid (i.e., will meet the asymptotic level constraint) independently of the particular signs, scales, and order of the ICs fixed above in $\mathcal{P}_0^n(S_1, S_2)$. Note that their optimality properties, however, require to order the scores K_{f_r}, $r = 1, \ldots, p$ according to the corresponding "(S_1, S_2)-kurtoses".

As we have seen above, the only restriction imposed by $\mathcal{P}_0^n(S_1, S_2)$ is that the "(S_1, S_2)-kurtoses" of the ICs are pairwise different, so that the ordering of the ICs is well-defined. Note that this rules out cases for which two (or more) ICs would be identically distributed. More precisely, consider the case for which exactly $k (\geq 2)$ ICs are equally distributed and the distributions of the remaining $p - k$ ICs are pairwise different. Then the estimator $\hat{\Lambda}$ above allows for recovering the $p-k$ ICs with different distributions, but estimates the remaining k ones up to some random rotation. Note however that if those k ICs are Gaussian, the components of $\hat{\Lambda}^{-1}X$—conditional on this random rotation—converge in distribution to Z (since—possibly rotated—uncorrelated Gaussian variables with a common scale are independent), so that the asymptotic null distribution of our test statistics is still χ^2_p (also unconditionally, since this conditional asymptotic distribution does not depend on the value of the random rotation). As a conclusion, while our tests, when based on such $\hat{\Lambda}$, would fail being valid when several ICs share the same distribution, they are valid in the case where the only equally distributed ICs are Gaussian, which includes the important multinormal case.

If however one thinks that ruling out equally distributed non-Gaussian ICs is too much of a restriction, then he/she can still use a root-n consistent estimator of Λ that does not require this assumption. See for example Hyvärinen et al. (2001) for an overview.

5.2 Finite-sample performances

We conducted a simulation study in the trivariate case ($p = 3$) in order to evaluate the finite-sample performances of our signed-rank tests.

We started by generating i.i.d. centered random vectors $Z_i = (Z_i^{(1)}, Z_i^{(2)}, Z_i^{(3)})'$, $i = 1, \ldots, n$ (we used $n = 50$ and $n = 200$) with marginals that are standardized so that $\text{Med}[(Z_1^{(r)})^2] = 1$, $r = 1, 2, 3$. We considered four settings with the following marginal distributions for $Z_1^{(1)}$, $Z_1^{(2)}$, and $Z_1^{(3)}$:

Setting I: t_9, Gaussian, and power-exponential with $\eta = 2$ (see Section 4.2) distributions
Setting II: t_3, t_6, and Gaussian distributions
Setting III: t_1, t_6, and Gaussian distributions
Setting IV: three Gaussian distributions (the multinormal case).

Denoting by I_ℓ the ℓ-dimensional identity matrix, samples were then obtained from the IC models $X_i = \Lambda Z_i + \mu$, $i = 1, \ldots, n$, with mixing matrix $\Lambda = I_3$ (this is without loss of generality, since all tests involved in this study are affine-invariant) and location values $\mu = 0$ (null case) and $\mu = n^{-1/2}\tau_\ell e_r$, $\ell = 1, 2, 3, 4$, $r = 1, 2, 3$, (cases in the alternative), where $\tau_1 = 2.147$, $\tau_2 = 3.145$, $\tau_3 = 3.966$, and $\tau_4 = 4.895$ were chosen so that the asymptotic powers of Hotelling's T^2 test, in Setting IV, are equal to .2, .4, .6, and .8, respectively.

First, we studied the sensitivity of our tests with respect to the choice of the estimator $\hat{\Lambda}$ in Setting I. To this end, we considered three estimators in the class of estimators introduced in Section 5.1:

(1) The estimator $\hat{\Lambda}_1$ is based on $S_1 = S_{\text{cov}}$ and $S_2 = S_{\text{kurt}}$; root-n consistency of $\hat{\Lambda}_1$ requires finite eighth-order moments.

(2) The estimate $\hat{\Lambda}_2$ is based on $S_1 = S_{\text{Tyl}}$ and $S_2 = S_{\text{Düm}}$, where $S_{\text{Düm}}$ stands

for Dümbgen (1998)'s scatter matrix (which is the symmetrized version of S_{Tyl}); although $\hat{\Lambda}_2$ is root-n consistent without any moment assumption, it does not fulfill the assumptions of Section 3, since $S_{\mathrm{Düm}}$ (hence also $\hat{\Lambda}_2$) is not invariant under individual sign changes of observations.

(3) Finally, defining $S_{\mathrm{rank}} = E[\Psi_p^{-1}(F_{\|S_{\mathrm{Tyl}}^{-1/2}X\|}(\|S_{\mathrm{Tyl}}^{-1/2}X\|))\frac{XX'}{X'S_{\mathrm{Tyl}}^{-1}X}]$, where Ψ_p denotes the distribution function of a χ^2_p random variable, the estimate $\hat{\Lambda}_3$, based on $S_1 = S_{\mathrm{Tyl}}$ and $S_2 = S_{\mathrm{rank}}$ fulfills all the assumptions of Section 3 and is root-n consistent without any moment conditions.

For the sake of comparison, we also considered the unrealistic case for which Λ is known. For brevity reasons we refrain from showing the results and only point out that the behavior of our tests does not depend much on the choice of the estimator for Λ. Actually even knowing the true value of Λ did not show to be of any clear advantage. However, it is crucial that the estimator $\hat{\Lambda}$ that is used is root-n consistent, which, in Setting I, is the case of $\hat{\Lambda}_i$, $i = 1, 2, 3$. In Settings I, II and III, the "(S_1, S_2)-kurtoses" from (1), (2) and (3) order the marginal distributions in the same way.

Second, we compared, in Settings I to IV, several versions of our tests with Hotelling's T^2 test. We considered the following signed-rank tests: the sign test based on $\hat{Q}_S$ in (1), the Wilcoxon test based on $\hat{Q}_W$ in (2), and the van der Waerden test based on $\hat{Q}_{\mathrm{vdW}}$ in (3). In each setting, we also included the corresponding optimal signed-rank test (based on $\hat{Q}_{K_f}$ in Section 4.1); we denote by $\hat{Q}^{\mathrm{I}}_{\mathrm{opt}}$, $\hat{Q}^{\mathrm{II}}_{\mathrm{opt}}$, and $\hat{Q}^{\mathrm{III}}_{\mathrm{opt}}$ the statistics of these setting-dependent tests (the optimal test in Setting IV is the van der Waerden test based on $\hat{Q}_{\mathrm{vdW}}$). Of course, these optimal tests use the unspecified underlying density, which is unrealistic, but this is done in order to check how much is gained, in each setting, by using optimal scores. Since the properties of the proposed tests are not very sensitive to the choice of $\hat{\Lambda}$, each signed-rank test was based on the estimator $\hat{\Lambda}_3$ (only the latter satisfies our assumptions on $\hat{\Lambda}$ in all settings). All tests were performed at asymptotic level 5%.

Figures 1 to 4 report rejection proportions (based on 5,000 replications) and asymptotic powers of the above tests in Settings I to IV, respectively. We should stress that preliminary simulations showed that, under the null in Setting I, the van der Waerden test and the test based on $\hat{Q}^{\mathrm{I}}_{\mathrm{opt}}$, when based on their asymptotic chi-square critical values, are conservative and significantly biased at small sample size $n = 50$. In order to remedy this, we rather used critical values based on the estimation of the (distribution-free) quantile of the test statistic under $\mu = 0$ and under known value $\Lambda = I_3$ of the mixing parameter. These estimations, just as the asymptotic chi-square quantile, are consistent approximations of the corresponding exact quantiles under the null, and were obtained, for the van der Waerden test and the test based on $\hat{Q}^{\mathrm{I}}_{\mathrm{opt}}$, as the empirical 0.05-upper quantiles $q_{.95}$ of the corresponding signed-rank test statistics

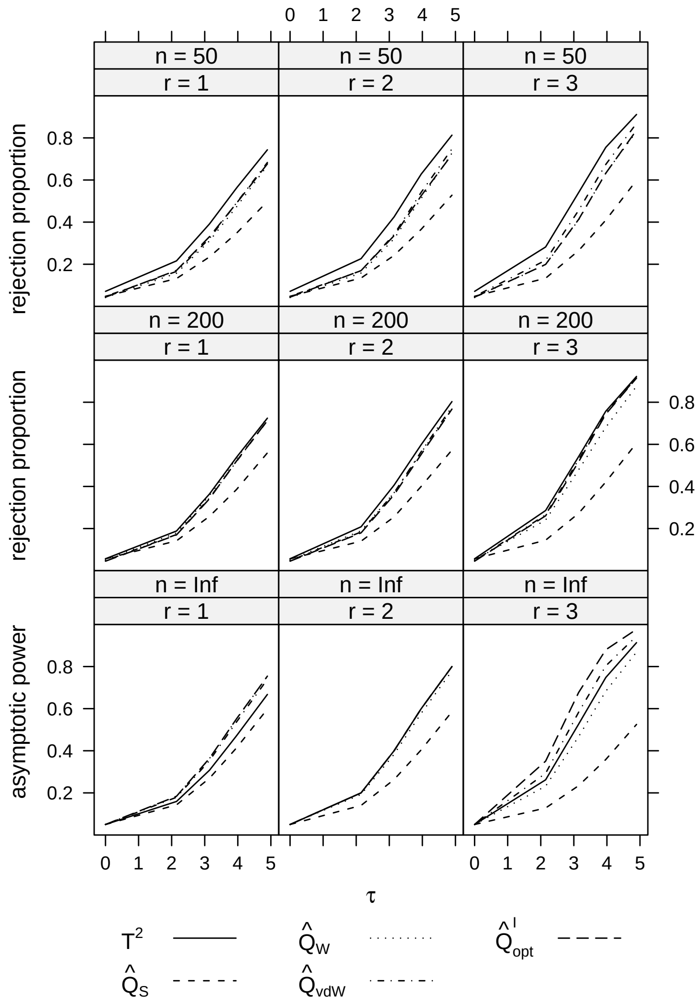

Fig. 1. Rejection proportions (for $n = 50$ and $n = 200$, based on $5,000$ replications) and asymptotic powers, in Setting I, of Hotelling's T^2 test and of the $\hat{\Lambda}_3$-based versions of the sign, Wilcoxon, van der Waerden, and Setting I-optimal signed-rank tests. The integer r indicates in which coordinate the shift occurs.

in a collection of $10,000$ simulated (standard) multinormal samples, yielding $q^{\text{vdW}}_{.95} = 7.239$ and $q^{\text{opt,I}}_{.95} = 6.859$, respectively. These bias-corrected critical values both are smaller than the asymptotic chi-square one $\chi^2_{3,.95} = 7.815$, so that the resulting tests are uniformly less conservative than the original ones. Note that these critical values were always applied when any of those tests were used with $n = 50$ since in practice one does not know the underlying distribution. In all other cases (i.e., for all other tests at $n = 50$, and for all tests at $n = 200$), the asymptotic chi-square critical value $\chi^2_{3,.95}$ was used.

Based on the simulation studies we therefore recommend that for small sample sizes one should calculate the p-value based on simulations or just use a conditionally distribution-free test version. This is not a problem with the

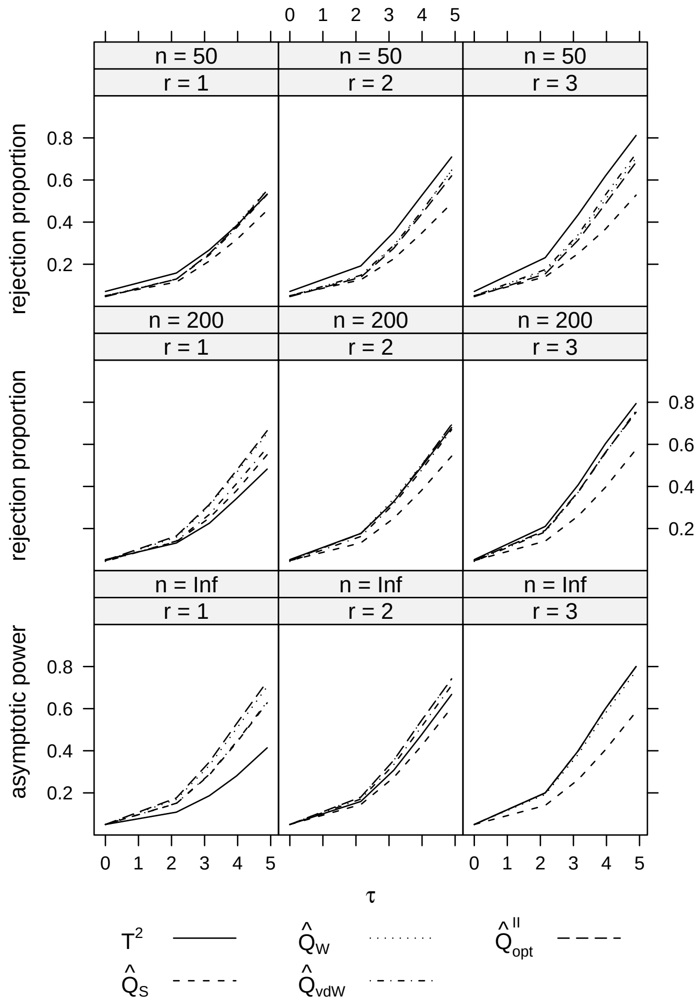

Fig. 2. Rejection proportions (for $n = 50$ and $n = 200$, based on $5,000$ replications) and asymptotic powers, in Setting II, of Hotelling's T^2 test and of the $\hat{\Lambda}_3$-based versions of the sign, Wilcoxon, van der Waerden, and Setting II-optimal signed-rank tests. The integer r indicates in which coordinate the shift occurs.

current speed of computers, and all three approaches have been implemented in the package ICSNP. Our simulations show that alternative ways to calculate p-values are needed especially when one of the score functions K_{f_r} used is associated with a light-tailed density f_r.

A glance at the rejection proportions under the null in Figures 1 to 4 shows that all signed-rank tests appear to satisfy the 5% probability level constraint. In particular, for $n = 50$, the bias-corrected versions of the tests based on $\hat{Q}_{\mathrm{vdW}}$ and on $\hat{Q}^{\mathrm{I}}_{\mathrm{opt}}$ are reasonably unbiased, whereas the asymptotic χ^2_3 approximation seems to work fine in all other cases. Note that Hotelling's T^2 test satisfies the 5% probability level constraint also in Setting III, which was unexpected since one of the marginals (the t_1 distributed one) has infinite second-order

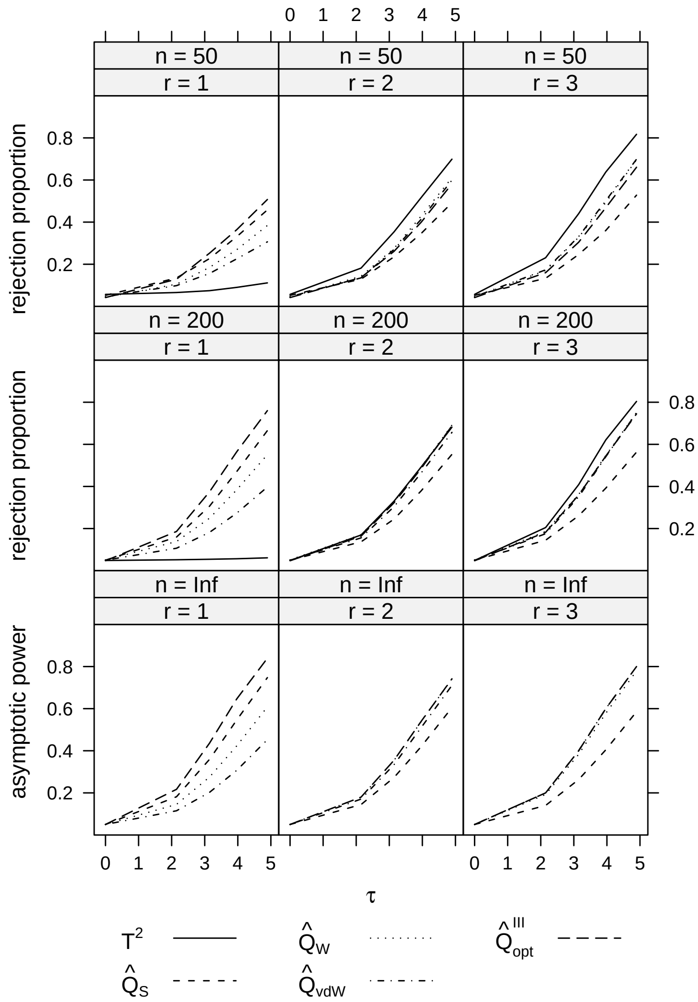

Fig. 3. Rejection proportions (for $n = 50$ and $n = 200$, based on $5,000$ replications) and asymptotic powers, in Setting III, of Hotelling's T^2 test and of the $\hat{\Lambda}_3$-based versions of the sign, Wilcoxon, van der Waerden, and Setting III-optimal signed-rank tests. The integer r indicates in which coordinate the shift occurs.

moments whereas in all other settings Hotelling's T^2 seems to reject too often.

As for the power properties, the proposed signed-rank tests behave uniformly well in all settings, unlike Hotelling's test, which, for instance, basically never detects the shift in the t_1 component of Setting III (still, it should be noticed that, in the same setting, Hotelling's test works pretty well if the shift is another component; we will explain this unexpected behavior of Hotelling's test in Section 5.3 below). In Setting II (see Figure 2), Hotelling's test competes reasonably well with our tests for small sample sizes, when the shift occurs in a heavy-tailed component. For larger sample sizes, however, our tests outperform Hotelling's and, except for $\hat{Q}_S$, behave essentially as Hotelling's test when the

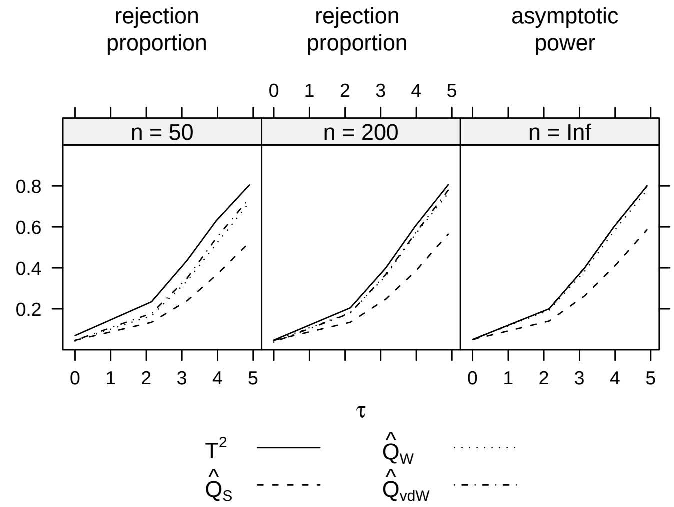

Fig. 4. Rejection proportions (for $n = 50$ and $n = 200$, based on $5,000$ replications) and asymptotic powers, in Setting IV, of Hotelling's T^2 test and of the $\hat{\Lambda}_3$-based versions of the sign, Wilcoxon, and van der Waerden (which is optimal in Setting IV) signed-rank tests. Without loss of generality (since the underlying distribution is spherically symmetric), the shift occurs in the first coordinate only.

shift occurs in the Gaussian component (this is totally in line with the ARE values in Table 1). Note that when a light-tailed component is present as in Setting I (see Figure 1), our tests perform as expected. Furthermore the proposed tests also work well in the multinormal model (Figure 4), although $\hat{\Lambda}_3$ is there only a random rotation; see the comments at the end of Section 5.1.

As a conclusion, our optimal tests exhibit very good finite-sample performances in IC models, both in terms of level and power.

5.3 Robustness evaluation

In this section, we investigate the robustness properties of the proposed signed-rank tests (in the bivariate case) by studying their power functions under contamination, and by comparing the results with Hotelling's test.

Starting with bivariate i.i.d. random vectors $Z_i = (Z_i^{(1)}, Z_i^{(2)})'$, $i = 1, \ldots, n$ (we used $n = 50$ in this section) with centered t_3 and Gaussian marginals in the first and second components, respectively (still standardized so that $\text{Med}[(Z_1^{(r)})^2] = 1$, $r = 1, 2$), we generated bivariate observations according to $X_i = \Lambda Z_i + \frac{\tau}{\sqrt{n}}(0, 1)'$, $i = 1, \ldots, n$, where $\Lambda = I_2$ and where $\tau = 3.301$ is so that the asymptotic power of Hotelling's test (at asymptotic level $\alpha = .05$) is .5. For any fixed $\delta = (\delta^{(1)}, \delta^{(2)})' \in \mathbb{R}^2$, denote then by $\mathbf{X}(\delta)$ the sample of size n obtained by replacing the first observation X_1 with $X_1 + \delta$.

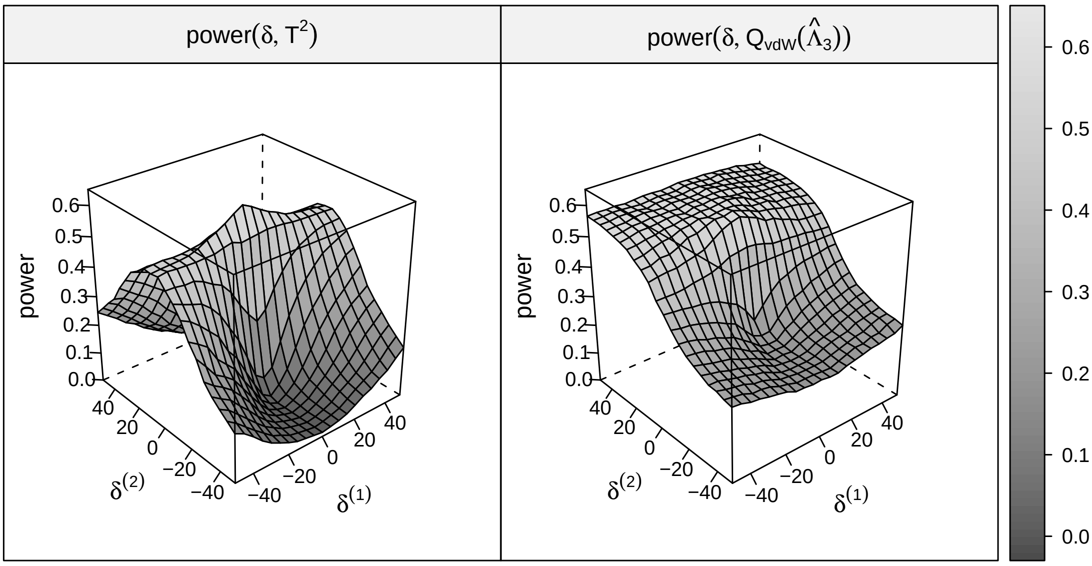

Fig. 5. Estimates of the power functions $power(\delta, T^2)$, $power(\delta, Q_{\rm vdW}(\Lambda))$, $power(\delta, Q_{\rm vdW}(\hat{\Lambda}_1))$, and $power(\delta, Q_{\rm vdW}(\hat{\Lambda}_3))$. The sample size is $n = 50$ and the estimation is based on $1,000$ replications.

Clearly, the value of a test statistic computed on $\mathbf{X}(\delta)$—hence, also the power of the corresponding test—depends on δ. For any test ϕ rejecting $\mathcal{H}_0 : \mu = 0$ at asymptotic level α whenever $Q > \chi^2_{2,1-\alpha}$, we define the *power function* of ϕ as $\delta \mapsto power(\delta, Q) := P[Q(\mathbf{X}(\delta)) > \chi^2_{2,1-\alpha}]$. Of course, this function can be estimated by generating a large number of independent samples $\mathbf{X}(\delta)$ and by computing rejection frequencies.

We estimated the power functions over $\delta = (\pm 5i, \pm 5j)'$, with $i, j = 0, \ldots, 10$, of Hotelling's T^2 test and of two versions of the van der Waerden signed-rank tests based on (3): the first one (resp., the second one) is based on $\hat{\Lambda}_1$ (resp., on $\hat{\Lambda}_3$), where $\hat{\Lambda}_i$, $i = 1, 3$ are as in Section 5.2. To be in line with what we did there, all van der Waerden tests were based on an estimate (under the null) of the exact (at $n = 50$) distribution-free 95%-quantile of the known-Λ van der Waerden test statistic. In this bivariate case, this estimated quantile, based on $10,000$ independent values of this statistic, took the value 5.354 ($< 5.991 = \chi^2_{2,.95}$).

Figure 5 presents the estimated power functions (based on 1,000 replications) of Hotelling's T^2 test and of the $\hat{\Lambda}_3$-based van der Waerden test. Results for the $\hat{\Lambda}_1$-based version of the latter are not shown since they are very similar to those of the $\hat{\Lambda}_3$-based one (which is actually surprising since one would guess that the lack of robustness of $\hat{\Lambda}_1$ would severely affect the test).

Quite unexpectedly, for $\delta^{(2)} = 0$, the power of Hotelling's test does not suffer under the value of $\delta^{(1)}$. It is even so that compared with the non-contaminated case $\delta = 0$, for which the power functional of Hotelling has the value .516, the functional shows higher power for $|\delta^{(1)}| < 10$ and $0 < \delta^{(2)} \leq 10$. However, if $|\delta^{(2)}|$ is large, the power drops quickly, especially so when there is no or little contamination in $\delta^{(1)}$. The power can then drop even below the size value of .05; e.g., at $\delta = (0, -20)'$, it is only .012.

The puzzling robustness of Hotelling's test with respect to an outlying observation in the first variate can be explained as follows. Let $\mathbf{X} = (X_1 \; X_2 \; \cdots \; X_n)$ be a sample of i.i.d. p-variate observations (whose common distribution admits finite second-order moments) and partition it into

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} X_{11} \; X_{21} \; \cdots \; X_{n1} \\ X_{12} \; X_{22} \; \cdots \; X_{n2} \end{pmatrix},$$

where the X_{i1}'s are random variables and the X_{i2}'s are $(p-1)$-random vectors. Now, by using (14) in Rencher (1993), it can be shown that, if one replaces $X_1 = (X_{11}, X_{12}')'$ with $(X_{11}+\delta, X_{12}')'$ and lets $\delta \to \infty$, then, under the assumption (as in the setting above) that the X_{i2}'s are i.i.d. with mean $\tau/\sqrt{n}$ and covariance matrix Σ_{22}, $\lim_{\delta\to\infty} T^2(\mathbf{X}) = T^2(\mathbf{X}_2)+1+o_P(1) \xrightarrow{\mathcal{L}} \chi^2_{p-1}(\tau'\Sigma_{22}^{-1}\tau)+1$, as $n \to \infty$, where $\xrightarrow{\mathcal{L}}$ denotes convergence in law. This is to be compared with the asymptotic $\chi^2_p(\tau'\Sigma_{22}^{-1}\tau)$ distribution of $T^2(\mathbf{X})$ under the assumption that the $X_i = (X_{i1}', X_{i2}')'$'s are i.i.d. with mean $(0, \tau')'$ and with an arbitrary covariance matrix such that $\text{Var}[X_{i2}] = \Sigma_{22}$. For small dimensions p, obtaining (by contaminating a single observation) a $\chi^2_{p-1}(\tau'\Sigma_{22}^{-1}\tau)+1$ distribution rather than the expected $\chi^2_p(\tau'\Sigma_{22}^{-1}\tau)$ one can bias the results considerably.

Hence, one can say that an outlier in one variate (i) destroys all information about that variate and (ii) biases the result for the "remaining data". This also explains the unexpected behavior of Hotelling's test in Setting III of Section 5.2: the t_1-distributed variate can be seen as a completely contaminated variate which therefore basically contains no information; still, Hotelling's test can detect shifts in the remaining variates.

Figure 5 shows that on the other hand the test based on $Q_{\text{vdW}}(\hat{\Lambda}_3)$ proves much more robust than Hotelling's and is hardly affected by the value of δ_1. Note that if the contamination δ_2 is negative (resp., positive), the power of this test slightly goes down (resp., up) as δ_1 goes through the Z_{i1} data cloud. This slight decrease (resp., increase) of the power function can be explained by the fact that, for any negative (resp., positive) value of δ_2, the contaminated observation—with the scale used in our setting—immediately gets the smallest (resp., largest) rank assigned. The range of the $Q_{\text{vdW}}(\hat{\Lambda}_3)$-power function in Figure 5 goes from .193 to .582, which is comparable with those associated with $Q_{\text{vdW}}(\Lambda)$ (from .263 to .576) and with $Q_{\text{vdW}}(\hat{\Lambda}_1)$ (from .237 to .580).

Acknowledgement

The research work of Klaus Nordhausen and Hannu Oja was partially supported by research grants from the Academy of Finland and by the Tampere Graduate School in Information Science and Engineering. The research work of Davy Paindaveine was supported by a Mandat d'Impulsion Scientifique of the Fonds National de la Recherche Scientifique, Communauté française de Belgique. The authors would like to thank the editor, the associate editor, and two anonymous referees for their valuable comments that helped us improve significantly the presentation and the quality of the paper.

A Proofs of Lemmas 3.1, 3.2, and 3.3

In this section, we will write, $T_K^{(r)}(\Lambda)$ (resp., $T_{K;g}^{(r)}(\Lambda)$) for the rth component of $T_K(\Lambda)$ (resp., of $T_{K;g}(\Lambda)$), $r = 1, \ldots, p$.

Proof of Lemma 3.1. Fix $\Lambda \in \mathcal{M}_p$, $g \in \mathcal{F}$, and $r \in \{1, \ldots, p\}$. Then, under $P^n_{0,\Lambda,g}$, the vector of signs $(S_1^{(r)}(\Lambda), \ldots, S_n^{(r)}(\Lambda))$ collects i.i.d. random variables with $P^n_{0,\Lambda,g}[S_i^{(r)}(\Lambda) = 1] = P^n_{0,\Lambda,g}[S_i^{(1)}(\Lambda) = -1] = 1/2$, (ii) the vector of ranks $(R_1^{(r)}(\Lambda), \ldots, R_n^{(r)}(\Lambda))$ is uniformly distributed over the set of all permutations of $\{1, 2, \ldots, n\}$, and (iii) the vector of signs is independent of the vector of ranks. Consequently, Hájek's classical projection result for signed rank linear statistics (see, e.g., Puri and Sen 1985, Chapter 3) yields that $\mathrm{E}[(T_K^{(r)}(\Lambda) - T_{K;g}^{(r)}(\Lambda))^2] = \mathrm{E}[(n^{-1/2} \sum_{i=1}^n S_i^{(r)}(\Lambda)[K^{(r)}(\frac{R_i^{(r)}(\Lambda)}{n+1}) - K^{(r)}(G_+^{(r)}(|Z_i^{(r)}(\Lambda)|))])^2]$ is $o(1)$ under $P^n_{0,\Lambda,g}$, as $n \to \infty$, which establishes the result. □

Note that this also shows that $\mathrm{E}[(K^{(r)}(R_1^{(r)}(\Lambda)/(n+1)) - K^{(r)}(G_+^{(r)}(|Z_1^{(r)}(\Lambda)|)))^2]$ $= \mathrm{E}[(n^{-1/2} \sum_{i=1}^n S_i^{(r)}(\Lambda)[K^{(r)}(R_i^{(r)}(\Lambda)/(n+1)) - K^{(r)}(G_+^{(r)}(|Z_i^{(r)}(\Lambda)|))])^2]$ is $o(1)$ as $n \to \infty$, under $P^n_{0,\Lambda,g}$.

Proof of Lemma 3.2. Fix $\Lambda \in \mathcal{M}_p$ and $g \in \mathcal{F}$. For any $r = 1, \ldots, p$, the CLT shows that $T_{K;g}^{(r)}(\Lambda)$ is, under $P^n_{0,\Lambda,g}$, asymptotically normal with mean zero and variance $\mathrm{E}[(K^{(r)}(U))^2]$. Therefore, the mutual independence (still under $P^n_{0,\Lambda,g}$) of $T_{K;g}^{(r)}(\Lambda)$, $r = 1, \ldots, p$ entails that $T_{K;g}(\Lambda)$ is asymptotically multinormal with mean zero and covariance matrix Γ_K. The result then follows from Lemma 3.1. □

It remains to prove Lemma 3.3. We do so by showing that, for any $\Lambda \in \mathcal{M}_p$, $g \in \mathcal{F}$, and $r \in \{1, \ldots, p\}$,

$$E[(T_K^{(r)}(\hat{\Lambda}) - T_K^{(r)}(\Lambda))^2] = o(1) \tag{A.1}$$

as $n \to \infty$, under $P^n_{0,\Lambda,g}$. In the rest of this section, we therefore fix such Λ, g, and r. All expectations and stochastic convergences will then be under $P^n_{0,\Lambda,g}$, and we will write $Z_i^{(r)}$, $S_i^{(r)}$, and $R_i^{(r)}$ for $Z_i^{(r)}(\Lambda)$, $S_i^{(r)}(\Lambda)$, and $R_i^{(r)}(\Lambda)$, respectively. Finally, we will denote the empirical counterparts of these quantities (based on $\hat{\Lambda}$) by $\hat{Z}_i^{(r)}$, $\hat{S}_i^{(r)}$, and $\hat{R}_i^{(r)}$.

We will need the following preliminary result.

Lemma A.1 *As $n \to \infty$, (i) $\hat{Z}_1^{(r)} - Z_1^{(r)} = o_P(1)$, (ii) $\mathrm{E}[(K^{(r)}(\hat{R}_1^{(r)}/(n+1)) - K^{(r)}(G_+^{(r)}(|Z_1^{(r)}(\Lambda)|)))^2] = o(1)$ and (iii) $\mathrm{E}[|\hat{S}_1^{(r)} - S_1^{(r)}|^a] = o(1)$ for any $a > 0$.*

Proof of Lemma A.1. (i) Denoting by $\|A\|_{\mathcal{L}}$ the sup norm of the array A, we have $|\hat{Z}_1^{(r)} - Z_1^{(r)}| \leq \|\hat{Z}_1 - Z_1\| \leq \|\hat{\Lambda}^{-1} - \Lambda^{-1}\|_{\mathcal{L}}\|X_1\|$. The claim therefore follows from the root-n consistency of $\hat{\Lambda}$.

(ii) Applying Lemma 2 of Peters & Randles (1990), with $\alpha = (\mathrm{vec}\,\Lambda)$ and $g(X, \alpha) = |e_r'[\Lambda^{-1}X]|$ yields that $(\hat{R}_1^{(r)}/(n+1)) - G_+^{(r)}(|Z_1^{(r)}|)$ is $o(1)$ as $n \to \infty$ (note that Conditions (a)-(b) of that lemma are fulfilled: (a) is our root-n consistency assumption on $\hat{\Lambda}$, whereas (b) can be checked exactly along the same lines as in Peters & Randles (1990), once it is noticed that $||e_r'[(\Lambda + n^{-1/2}L)^{-1}X]| - |e_r'[\Lambda^{-1}X]|| \leq \|[(\Lambda + n^{-1/2}L)^{-1} - \Lambda^{-1}]X\|$, for any fixed $p \times p$ matrix L).

Now, the continuity of $K^{(r)}$ entails that

$$K^{(r)}\Big(\frac{\hat{R}_1^{(r)}}{n+1}\Big) - K^{(r)}(G_+^{(r)}(|Z_1^{(r)}|)) \tag{A.2}$$

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 (A.2) is uniformly integrable. The second term in (A.2) is of course uniformly integrable since the integrable random variable $K_r(G_+^{(r)}(|Z_1^{(1r)}|))$ does not depend on n. As for the first term in (A.2), recall that $K^{(r)}(R_1^{(r)}/(n+1)) - K^{(r)}(G_+^{(r)}(|Z_1^{(r)}|)) = o_{L^2}(1)$ as $n \to \infty$ (see the remark after the proof of Lemma 3.1), which implies that $K^{(r)}(\frac{R_1^{(r)}}{n+1})$ is uniformly integrable. Finally, the latter uniform integrability and the invariance of $\hat{\Lambda}$ under permutations of the observations in turn imply that $K^{(r)}(\frac{\hat{R}_1^{(r)}}{n+1})$ is uniformly integrable. We conclude that (A.2) is indeed uniformly integrable, and the result follows.

(iii) Since $\hat{S}_1^{(r)} - S_1^{(r)} = (|\hat{Z}_1^{(r)}|^{-1} - |Z_1^{(r)}|^{-1})\hat{Z}_1^{(r)} + |Z_1^{(r)}|^{-1}(\hat{Z}_1^{(r)} - Z_1^{(r)})$, we have $|\hat{S}_1^{(r)} - S_1^{(r)}| \leq 2|\hat{Z}_1^{(r)} - Z_1^{(r)}|/|Z_1^{(r)}| =: Y_1^{(r)}$. Now, fix some $\delta > 0$. Then, for all $\eta > 0$, $P[Y_1^{(r)} > \delta] \leq P[Y_1^{(r)}\, I_{[|Z_1^{(r)}|<\eta]} > \delta/2] + P[Y_1^{(r)}\, I_{[|Z_1^{(r)}|\geq\eta]} > \delta/2]$ $\leq P[|Z_1^{(r)}| < \eta] + P[Y_1^{(r)}\, I_{[|Z_1^{(r)}|\geq\eta]} > \delta/2] =: p_1^{(n)} + p_2^{(n)}$, say. For all $\varepsilon > 0$, there exists $\eta = \eta(\varepsilon)$ such that $p_1^{(n)} < \varepsilon/2$. As for $p_2^{(n)}$, note that $Y_1^{(r)} I_{[|Z_1^{(r)}| \geq \eta]} \leq (2/\eta)|\hat{Z}_1^{(r)} - Z_1^{(r)}|$, so that Part (i) of the lemma entails that $p_2^{(n)} < \varepsilon/2$ for large n. We conclude that $|\hat{S}_1^{(r)} - S_1^{(r)}| \leq Y_1^{(r)}$ converges to zero in probability, which establishes the result (since $|\hat{S}_1^{(r)} - S_1^{(r)}|$ is bounded). □

Proof of Lemma 3.3. We have to prove (A.1). Since the proof of Lemma 3.1 establishes that $E[(T_K^{(r)}(\Lambda) - T_{K;g}^{(r)}(\Lambda))^2] = o(1)$ as $n \to \infty$, it is sufficient to show that $E[(T_K^{(r)}(\hat{\Lambda}) - T_{K;g}^{(r)}(\Lambda))^2] = o(1)$ as $n \to \infty$. To do so, write $T_K^{(r)}(\hat{\Lambda}) - T_{K;g}^{(r)}(\Lambda) = H_1 + H_2$, with $H_1 := n^{-1/2}\sum_{i=1}^n \hat{S}_i^{(r)}\,(K^{(r)}(\hat{R}_i^{(r)}/(n+1)) - K^{(r)}(G_+^{(r)}(|Z_i^{(r)}|)))$ and $H_2 := n^{-1/2}\sum_{i=1}^n (\hat{S}_i^{(r)} - S_i^{(r)})\,K^{(r)}(G_+^{(r)}(|Z_i^{(r)}|))$. Then, by using the invariance of $\hat{\Lambda}$ under individual reflections of the observations about the origin, we obtain

$$\begin{aligned}\mathrm{E}[(H_1)^2] &= \frac{1}{n}\sum_{i=1}^n \mathrm{E}[(\hat{S}_i^{(r)})^2(K^{(r)}(\hat{R}_i^{(r)}/(n+1)) - K^{(r)}(G_+^{(r)}(|Z_i^{(r)}|)))^2]\\ &= \mathrm{E}[(K^{(r)}(\hat{R}_1^{(r)}/(n+1)) - K^{(r)}(G_+^{(r)}(|Z_1^{(r)}|)))^2]\end{aligned}$$

and, by using Holder's inequality,

$$\mathrm{E}[(H_2)^2] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[(\hat{S}_i^{(r)} - S_i^{(r)})^2\,(K^{(1)})^2(G_+^{(r)}(|Z_i^{(r)}|))] =$$

$$\mathrm{E}[(\hat{S}_1^{(r)} - S_1^{(r)})^2\,(K^{(r)}(G_+^{(r)}(|Z_1^{(r)}|)))^2] \leq (\mathrm{E}[(\hat{S}_1^{(r)} - S_1^{(r)})^{\frac{2d_\delta}{\delta}}])^{\frac{\delta}{d_\delta}}(\mathrm{E}[(K^{(r)}(U))^{d_\delta}])^{\frac{2}{d_\delta}},$$

where $d_\delta := 2 + \delta$, U is uniformly distributed over $(0,1)$, and $\delta > 0$ is the real number involved in our assumptions on $K^{(r)}$ (see the beginning of Section 3). By applying Lemma A.1(ii)-(iii), we then conclude that $E[(T_K^{(r)}(\hat{\Lambda}) - T_{K;g}^{(r)}(\Lambda))^2] \leq 2(E[(H_1)^2] + E[(H_2)^2])$ is $o(1)$ as $n \to \infty$. □

B Proofs of Theorems 4.2 and 4.3

Proof of Theorem 4.2. Fix $\Lambda \in \mathcal{M}_p$ and $g \in \mathcal{F}_{\mathrm{LAN}}$. Applying successively Lemmas 3.3 and 3.1 yields that, as $n \to \infty$, under $P^n_{0,\Lambda,g}$,

$$\hat{Q}_K = (T_{K;g}(\Lambda))'\Gamma_K^{-1}T_{K;g}(\Lambda) + o_P(1). \qquad \text{(B.1)}$$

Recall that $T_{K;g}(\Lambda)$, under $P^n_{0,\Lambda,g}$, is asymptotically multinormal with mean zero and covariance matrix Γ_K; see the proof of Lemma 3.2. Now, it is easy to see that, under $P^n_{0,\Lambda,g}$, $T_{K;g}(\Lambda)$ and the local log-likelihood $\log(dP^n_{n^{-1/2}\tau,\Lambda,g}/dP^n_{0,\Lambda,g})$ asymptotically are jointly multinormal with covariance $I_{K,g}\Lambda^{-1}\tau$. Le Cam's third Lemma thus yields that $T_{K;g}(\Lambda)$, under $P^n_{n^{-1/2}\tau,\Lambda,g}$, is asymptotically multinormal with mean $I_{K,g}\Lambda^{-1}\tau$ and covariance matrix Γ_K. The result then follows from the fact contiguity implies (B.1) holds also under $P^n_{n^{-1/2}\tau,\Lambda,g}$. $\square$

Proof of Theorem 4.3. Fix $\Lambda \in \mathcal{M}_p$ and $g \in \mathcal{F}^2_{\mathrm{LAN}}$. In this proof, all expectations, variances, and covariances are under $P^n_{0,\Lambda,g}$.

Since $\mathrm{Var}[X_1] = \Lambda\Sigma_g\Lambda'$ (where Σ_g is defined in the statement of the theorem), we have that $S := \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' = \Lambda\Sigma_g\Lambda' + o_P(1)$ as $n \to \infty$, under $P^n_{0,\Lambda,g}$. Consequently, letting $Z_i := Z_i(\Lambda) = \Lambda^{-1}X_i$ and $\bar{Z} := \frac{1}{n}\sum_{i=1}^n Z_i$, Hotelling's test statistic T^2 satisfies $T^2 = n\bar{X}'S^{-1}\bar{X} + o_P(1) = (\sqrt{n}\bar{Z})'\Sigma_g^{-1}(\sqrt{n}\bar{Z}) + o_P(1)$ as $n \to \infty$, under $P^n_{0,\Lambda,g}$, hence also under $P^n_{n^{-1/2}\tau,\Lambda,g}$ (from contiguity). Clearly, $\sqrt{n}\bar{Z}$ is asymptotically multinormal with mean zero and covariance matrix Σ_g under $P^n_{0,\Lambda,g}$. Proceeding as in the previous proof, one then shows that $\sqrt{n}\bar{Z}$ and the local log-likelihood $\log(dP^n_{n^{-1/2}\tau,\Lambda,g}/dP^n_{0,\Lambda,g})$ asymptotically are jointly multinormal under $P^n_{0,\Lambda,g}$, with asymptotic covariance $\Lambda^{-1}\tau$. Le Cam's third Lemma thus implies that $\sqrt{n}\bar{Z}$, under $P^n_{n^{-1/2}\tau,\Lambda,g}$, is asymptotically multinormal with mean $\Lambda^{-1}\tau$ and covariance matrix Σ_g. Therefore, T^2 is asymptotically $\chi^2_p(\tau'(\Lambda^{-1})'\Sigma_g^{-1}\Lambda^{-1}\tau)$ under $P^n_{n^{-1/2}\tau,\Lambda,g}$.

This establishes the result since the AREs of ϕ_K with respect to Hotelling's T^2 test are obtained by computing the ratios of the noncentrality parameters in their respective asymptotic distributions under local alternatives. $\square$

References

Chakraborty, B. & Chaudhuri, P., On affine invariant sign and rank tests in one sample and two sample multivariate problems. In Ghosh, S., Multivariate, design and sample survey, Marcel-Dekker, (New York 1999), 499-414.

Chernoff, H., & Savage, I. R., Asymptotic normality and efficiency of certain nonparametric tests, 1958, Annals of Institute of Mathematical Statistics, 29, 972–994.

Dümbgen, L., On Tyler's M-functional of scatter in high dimensions, 1998, Annals of Institute of Statistical Mathematics, 50, 1269–1292.

Hallin, M., & D. Paindaveine, Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks, 2002, Annals of Statistics, 30, 1103–1133.

Hallin, M., & D. Paindaveine, Multivariate signed ranks : Randles' interdirections or Tyler's angles?, 2002, In Y. Dodge, Ed., Statistical data analysis based on the L1-norm and related methods, Birkh?user, Basel, 271-282.

Hettmansperger, T.P.J., Nyblom, J., & Oja, H., Affine invariant multivariate one-sample sign test, 1994, Journal of the Royal Statistical Society, B 56, 221–234.

Hettmansperger, T.P.J., Möttönen, J., & Oja, H., Affine invariant multivari-

ate one-sample signed-rank test, 1997, Journal of the American Statistical Association, 92, 1591–1600.
Hyvärinen A., Karhunen J. & Oja E., Independent component analysis, Wiley & Sons (New York, 2001).
Le Cam, L.M. , Asymptotic Methods in Statistical Decision Theory, Springer-Verlag (New York, 1986).
Möttönen, J. & Oja, H., Multivariate spatial sign and rank methods, 1995, Journal of Nonparametric Statistics, 5, 201-213.
Nordhausen, K., Oja, H. & Tyler, D.E., On the efficiency of invariant multivariate sign and rank test. In Liski, E.P., Isotalo, J., Niemelä, J., Puntanen, S. & Styan, G.P.H., Festschrift for Tarmo Pukkila on his 60th birthday, University of Tampere, (Tampere, 2006), 217–231.
Nordhausen, K., Oja, H. & Tyler, D.E., ICS: Tools for Exploring Multivariate Data via ICS/ICA, 2007, R package version 1.1-0.
Nordhausen, K., Sirkiä, S., Oja, H. & Tyler, D.E., ICSNP: Tools for Multivariate Nonparametrics, 2007, R package version 1.0-1.
Oja, H. & Paindaveine, D., Optimal signed-rank tests based on hyperplanes, 2005, Journal of Statistical Planning and Inference 135, 300-323.
Oja, H., Sirkiä, S., & Eriksson, J., Scatter matrices and independent component analysis, 2006, Austrian Journal of Statistics, 35, 175-189.
Oja, H., Paindaveine, D., & Taskinen, S., Parametric and nonparametric tests for multivariate independence in IC models. Manuscript in preparation.
Peters, D. & Randles, R. H., A multivariate signed-rank test for the one-sample location problem, 1990, Journal of the American Statistical Association, 85, 552-557.
Puri, M. L. & Sen, P. K., Nonparametric Methods in Multivariate Analysis, Wiley & Sons (New York, 1971).
Puri, M. L. & Sen, P. K., Nonparametric Methods in General Linear Models, J. Wiley (New York, 1985).
R Development Core Team, R: A language and environment for statistical computing, ISBN 3-900051-07-0, (Vienna, 2007).
Randles, R. H., A distribution-free multivariate sign test based on interdirections, 1989 Journal of the American Statistical Association, 84, 1045-1050.
Randles, R. H., A simpler, affine-invariant, multivariate, distribution-free sign test, 2000, Journal of the American Statistical Association, 95, 1263–1268.
Rencher, A. C., The contribution of individual variables to Hotelling's T^2, Wilks' Λ, and R^2, 1993, Biometrics, 49, 479–489.
Theis, F. J., A new concept for separability problems in blind source separation, 2004, Neural Computation, 16, 1827–1850.
Tyler, D. E., A distribution-free M-estimator of multivariate scatter, 1987, Annals of Statistics, 15, 234–251.
Tyler, D. E., Critchley, F., Dümbgen, L., & Oja, H., Invariant coordinate selection, 2008. Conditionally accepted.