The Annals of Statistics
2011, Vol. 39, No. 5, 2448–2476
DOI: 10.1214/11-AOS906

SEMIPARAMETRICALLY EFFICIENT INFERENCE BASED ON SIGNED RANKS IN SYMMETRIC INDEPENDENT COMPONENT MODELS

BY PAULIINA ILMONEN[1] AND DAVY PAINDAVEINE[2]

Aalto University, University of Tampere, Université Libre de Bruxelles, and Université Pierre et Marie Curie

We consider semiparametric location-scatter models for which the p-variate observation is obtained as $X = \Lambda Z + \mu$, where μ is a p-vector, Λ is a full-rank $p \times p$ matrix and the (unobserved) random p-vector Z has marginals that are centered and mutually independent but are otherwise unspecified. As in blind source separation and independent component analysis (ICA), the parameter of interest throughout the paper is Λ. On the basis of n i.i.d. copies of X, we develop, under a symmetry assumption on Z, *signed-rank* one-sample testing and estimation procedures for Λ. We exploit the uniform local and asymptotic normality (ULAN) of the model to define signed-rank procedures that are semiparametrically efficient under correctly specified densities. Yet, as is usual in rank-based inference, the proposed procedures remain valid (correct asymptotic size under the null, for hypothesis testing, and root-n consistency, for point estimation) under a very broad range of densities. We derive the asymptotic properties of the proposed procedures and investigate their finite-sample behavior through simulations.

1. Introduction. In multivariate statistics, concepts of location and scatter are usually defined through affine transformations of a noise vector. To be more specific, assume that the observation X is obtained through

$$X = \Lambda Z + \mu, \tag{1.1}$$

where μ is a p-vector, Λ is a full-rank $p \times p$ matrix and Z is some *standardized* random vector. The exact nature of the resulting *location* parameter μ, *mixing matrix* parameter Λ, and *scatter* parameter $\Sigma = \Lambda\Lambda'$ crucially depends on the standardization adopted.

The most classical assumption on Z specifies that Z is standard p-normal. Then μ and Σ simply coincide with the mean vector $\mathrm{E}[X]$ and variance–covariance matrix $\mathrm{Var}[X]$ of X, respectively. In robust statistics, it is often rather assumed that Z

Received January 2011; revised June 2011.
[1]Supported by the Academy of Finland.
[2]Supported by an A.R.C. contract of the Communauté Française de Belgique. Davy Paindaveine is also member of ECORE, the association between CORE and ECARES.
MSC2010 subject classifications. Primary 62G05, 62G10; secondary 62G20, 62H99.
Key words and phrases. Independent component analysis, local asymptotic normality, rank-based inference, semiparametric efficiency, signed ranks.

is spherically symmetric about the origin of $\mathbb{R}^p$—in the sense that the distribution of OZ does not depend on the orthogonal $p \times p$ matrix O. The resulting model in (1.1) is then called the *elliptical* model. If Z has finite second-order moments, then $\mu = \mathrm{E}[X]$ and $\Sigma = c\,\mathrm{Var}[X]$ for some $c > 0$, but (1.1) allows to define μ and Σ in the absence of any moment assumption.

This paper focuses on an alternative standardization of Z, for which Z has mutually independent marginals with common median zero. The resulting model in (1.1)—the *independent component* (*IC*) *model*, say—is more flexible than the elliptical model, even if one restricts, as we will do, to vectors Z with symmetrically distributed marginals. The IC model indeed allows for heterogeneous marginal distributions for X, whereas, in contrast, marginals in the elliptical model all share—up to location and scale—the same distribution, hence also the same tail weight. This severely affects the relevance of elliptical models for practical applications, particularly so for moderate to large dimensions, since it is then very unlikely that all variables share, for example, the same tail weight.

The IC model provides the most standard setup for *independent component analysis* (*ICA*), in which the mixing matrix Λ is to be estimated on the basis of n independent copies $X_1, \dots, X_n$ of X, the objective being to recover (up to a translation) the original unobservable independent signals $Z_1, \dots, Z_n$ by premultiplying the X_i's with the resulting $\hat{\Lambda}^{-1}$. It is well known in ICA, however, that Λ is severely unidentified: for any $p \times p$ permutation matrix P and any full-rank diagonal matrix D, one can always write

$$X = [\Lambda PD][(PD)^{-1}Z] + \mu = \tilde{\Lambda}\tilde{Z} + \mu, \tag{1.2}$$

where $\tilde{Z}$ still has independent marginals with median zero. Provided that Z has at most one Gaussian marginal, two matrices Λ_1 and Λ_2 may lead to the same distribution for X in (1.1) if and only if they are equivalent (we will write $\Lambda_1 \sim \Lambda_2$) in the sense that $\Lambda_2 = \Lambda_1 PD$ for some matrices P and D as in (1.2); see, for example, [25]. In other words, under the assumption that Z has at most one Gaussian marginal, permutations (P), sign changes and scale transformations (D) of the independent components are the only sources of unidentifiability for Λ.

This paper considers inference on the mixing matrix Λ. More precisely, because of the identifiability issues above, we rather consider a normalized version L of Λ, where L is a well-defined representative of the class of mixing matrices that are equivalent to Λ. This parameter L is actually the parameter of interest in ICA: an estimate of L will indeed allow one to recover the independent signals $Z_1, \dots, Z_n$ equally well as an estimate of any other Λ with $\Lambda \sim L$. Interestingly, the situation is extremely similar when considering inference on Σ in the elliptical model. There, Σ is only identified up to a positive scalar factor, and it is often enough to focus on inference about the well-defined *shape* parameter $V = \Sigma/(\det \Sigma)^{1/p}$ (e.g., in PCA, principal directions, proportions of explained variance, etc. can be computed from V). Just as L is a normalized version of Λ in the IC model, V is a normalized version of Σ in the elliptical model, and in both classes of models, the normalized parameters actually are the natural parameters of interest in many inference problems. The similarities further extend to the semiparametric nature of both models: just as the density $g_{\|\cdot\|}$ of $\|Z\|$ in the elliptical model, the pdf g_r of the various independent components Z_r, $r = 1, \ldots, p$, in the IC model, can hardly be assumed to be known in practice.

These strong similarities motivate the approach we adopt in this paper: we plan to conduct inference on L (hypothesis testing and point estimation) in the IC model by adopting the methodology that proved extremely successful in [7, 8] for inference on V in the elliptical model. This methodology combines semiparametrically efficient inference and *invariance arguments*. In the IC model, the fixed-(μ, Λ) nonparametric submodels (indexed by $g_1, \ldots, g_p$) indeed enjoy a strong invariance structure that is parallel to the one of the corresponding elliptical submodels (indexed by $g_{\|\cdot\|}$). As in [7, 8], we exploit this invariance structure through a general result from [11] that allows one to derive invariant versions of efficient central sequences, on the basis of which one can define semiparametrically efficient (at fixed target densities $g_r = f_r$, $r = 1, \ldots, p$) invariant procedures. As the maximal invariant associated with the invariance structure considered turns out to be the vector of marginal signed ranks of the residuals, the proposed procedures are of a signed-rank nature and do not require to estimate densities. While they achieve semiparametric efficiency under correctly specified densities, they remain valid (correct asymptotic size under the null, for hypothesis testing, and root-n consistency, for point estimation) under misspecified densities.

We will consider the problem of estimating L and that of testing the null $\mathcal{H}_0 : L = L_0$ against the alternative $\mathcal{H}_1 : L \neq L_0$, for some fixed L_0. While point estimation is undoubtedly of primary importance for applications (e.g., in blind source separation), one might question the practical relevance of the testing problem considered, especially when L_0 is not the p-dimensional identity matrix. Solving this generic testing problem, however, is the main step in developing tests for any linear hypothesis on L, and we will explicitly describe the resulting tests in the sequel. An extensive study of these tests is beyond the scope of the present paper, though; we refer to [19] for an extension of our tests to the particular case of testing the (linear) hypothesis that L is block-diagonal, a problem that is obviously important in practice (nonrejection of the null would indeed allow practitioners to proceed with two separate, lower-dimensional, analyses). Testing linear hypotheses on L includes many other testing problems of high practical relevance, such as testing that a given column of L is equal to some fixed p-vector, and testing that a given entry of L is zero—the practical importance of these two testing problems, in relation, for example, with functional magnetic resonance imaging (fMRI), is discussed in [22].

The paper is organized as follows. In Section 2, we fix the notation and describe the model (Section 2.1), state the corresponding *uniformly locally and asymptotically normal* (*ULAN*) property that allows us to determine semiparametric efficiency bounds (Section 2.2) and then introduce, in relation with invariance arguments, *rank-based* efficient central sequences (Section 2.3). In Sections 3 and 4, we develop the resulting rank tests and estimators for the mixing matrix L, respectively. Our estimators actually require the delicate estimation of $2p(p-1)$ "cross-information coefficients," an issue we solve in Section 4.2 by generalizing the method recently developed in [5]. In Section 5, simulations are conducted both to compare the proposed estimators with some competitors and to investigate the validity of asymptotic results—simulation results for hypothesis testing are provided in the supplementary article [16]. Finally, the Appendix states some technical results (Appendix A) and reports proofs (Appendix B).

2. The model, the ULAN property and invariance arguments.

2.1. *The model.* As we already explained, the IC model above suffers from severe identifiability issues for Λ. To solve this, we map each Λ onto a unique representative $L = \Pi(\Lambda)$ of the collection of mixing matrices $\tilde{\Lambda}$ that satisfy $\tilde{\Lambda} \sim \Lambda$ (the equivalence class of Λ for $\sim$). We propose the mapping

$$\Lambda \mapsto \Pi(\Lambda) = \Lambda D_1^+ P D_2,$$

where D_1^+ is the positive definite diagonal matrix that makes each column of ΛD_1^+ have Euclidean norm one, P is the permutation matrix for which the matrix $B = (b_{ij}) = \Lambda D_1^+ P$ satisfies $|b_{ii}| > |b_{ij}|$ for all $i < j$ and D_2 is the diagonal matrix such that all diagonal entries of $\Pi(\Lambda) = \Lambda D_1^+ P D_2$ are equal to one.

If one restricts to the collection $\mathcal{M}_p$ of mixing matrices Λ for which no ties occur in the permutation step above, it can easily be shown that, for any $\Lambda_1, \Lambda_2 \in \mathcal{M}_p$, we have that $\Lambda_1 \sim \Lambda_2$ iff $\Pi(\Lambda_1) = \Pi(\Lambda_2)$, so that this mechanism succeeds in identifying a unique representative in each class of equivalence (this is ensured with the double scaling scheme above, which may seem a bit complicated at first). Besides, Π is then a continuously differentiable mapping from $\mathcal{M}_p$ onto $\mathcal{M}_{1p} := \Pi(\mathcal{M}_p)$. While ties may always be taken care of in some way (e.g., by basing the ordering on subsequent rows of the matrix B), they may prevent the mapping Π to be continuous, which would cause severe problems and would prevent us from using the Delta method in the sequel. It is clear, however, that the restriction to $\mathcal{M}_p$ only gets rid of a few particular mixing matrices, and will not have any implications in practice.

The parametrization of the IC model we consider is then associated with

$$X = LZ + \mu, \tag{2.1}$$

where $\mu \in \mathbb{R}^p$, $L \in \mathcal{M}_{1p}$ and Z has independent marginals with common median zero. Throughout, we further assume that Z admits a density with respect to the Lebesgue measure on $\mathbb{R}^p$, and that it has p symmetrically distributed marginals, among which at most one is Gaussian (as explained in the Introduction, this limitation on the number of Gaussian components is needed for L to be identifiable). We will denote by $\mathcal{F}$ the resulting collection of densities for Z. Of course, any $g \in \mathcal{F}$ naturally factorizes into $g(z) = \prod_{r=1}^{p} g_r(z_r)$, where g_r is the symmetric density of Z_r.

The hypothesis under which n mutually independent observations X_i, $i = 1, \dots, n$, are obtained from (2.1), where Z has density $g \in \mathcal{F}$, will be denoted as $\mathrm{P}^{(n)}_{\vartheta,g}$, with $\vartheta = (\mu', (\operatorname{vecd}^\circ L)')' \in \Theta = \mathbb{R}^p \times \operatorname{vecd}^\circ(\mathcal{M}_{1p})$, or alternatively, as $\mathrm{P}^{(n)}_{\mu,L,g}$; for any $p \times p$ matrix A, we write $\operatorname{vecd}^\circ A$ for the $p(p-1)$-vector obtained by removing the p diagonal entries of A from its usual vectorized form $\operatorname{vec} A$ (diagonal entries of L are all equal to one, hence should not be included in the parameter).

The resulting semiparametric model is then

$$\mathcal{P}^{(n)} := \bigcup_{g \in \mathcal{F}} \mathcal{P}^{(n)}_g := \bigcup_{g \in \mathcal{F}} \bigcup_{\vartheta \in \Theta} \{\mathrm{P}^{(n)}_{\vartheta,g}\}. \tag{2.2}$$

Performing semiparametrically efficient inference on ϑ, at a fixed $f \in \mathcal{F}$, typically requires that the corresponding parametric submodel $\mathcal{P}^{(n)}_f$ satisfies the *uniformly locally and asymptotically normal* (*ULAN*) property.

2.2. *The ULAN property.* As always, the ULAN property requires technical regularity conditions on f. In the present context, we need that each corresponding univariate pdf f_r, $r = 1, \dots, p$, is absolutely continuous (with derivative f_r', say) and satisfies

$$\sigma^2_{f_r} := \int_{-\infty}^{\infty} y^2 f_r(y)\, dy < \infty, \qquad \mathcal{I}_{f_r} := \int_{-\infty}^{\infty} \varphi^2_{f_r}(y) f_r(y)\, dy < \infty$$

and

$$\mathcal{J}_{f_r} := \int_{-\infty}^{\infty} y^2 \varphi^2_{f_r}(y) f_r(y)\, dy < \infty,$$

where we let $\varphi_{f_r} := -f_r'/f_r$. In the sequel, we denote by $\mathcal{F}_{\mathrm{ulan}}$ the collection of pdfs $f \in \mathcal{F}$ meeting these conditions.

For any $f \in \mathcal{F}_{\mathrm{ulan}}$, let $\gamma_{rs}(f) := \mathcal{I}_{f_r} \sigma^2_{f_s}$, define the optimal p-variate location score function $\varphi_f : \mathbb{R}^p \to \mathbb{R}^p$ through $z = (z_1, \dots, z_p)' \mapsto \varphi_f(z) = (\varphi_{f_1}(z_1), \dots, \varphi_{f_p}(z_p))'$, and denote by $\mathcal{I}_f$ the diagonal matrix with diagonal entries $\mathcal{I}_{f_r}$, $r = 1, \dots, p$. Further write I_ℓ for the ℓ-dimensional identity matrix and define

$$C := \sum_{r=1}^{p} \sum_{s=1}^{p-1} (e_r e_r' \otimes u_s e'_{s+\delta_{[s \geq r]}}),$$

where $\otimes$ is the usual Kronecker product, e_r and u_r stand for the rth vectors of the canonical basis of $\mathbb{R}^p$ and $\mathbb{R}^{p-1}$, respectively, and $\delta_{[s \geq r]}$ is equal to one if $s \geq r$ and to zero otherwise. The following ULAN result then easily follows from Proposition 2.1 in [19] by using a simple chain rule argument.

PROPOSITION 2.1. *Fix $f \in \mathcal{F}_{\mathrm{ulan}}$. Then the collection of probability distributions $\mathcal{P}_f^{(n)}$ is ULAN, with central sequence*

$$\Delta_{\vartheta,f} = \begin{pmatrix} \Delta_{\vartheta,f;1} \\ \Delta_{\vartheta,f;2} \end{pmatrix} = \begin{pmatrix} n^{-1/2}(L^{-1})' \sum_{i=1}^{n} \varphi_f(Z_i) \\ n^{-1/2} C(I_p \otimes L^{-1})' \sum_{i=1}^{n} \operatorname{vec}\bigl(\varphi_f(Z_i) Z_i' - I_p\bigr) \end{pmatrix}, \tag{2.3}$$

where $Z_i = Z_i(\vartheta) = L^{-1}(X_i - \mu)$, and full-rank information matrix

$$\Gamma_{L,f} = \begin{pmatrix} \Gamma_{L,f;1} & 0 \\ 0 & \Gamma_{L,f;2} \end{pmatrix},$$

where $\Gamma_{L,f;1} := (L^{-1})' \mathcal{I}_f L^{-1}$ and

$$\begin{aligned} \Gamma_{L,f;2} := C(I_p \otimes L^{-1})' \Biggl[& \sum_{r=1}^{p} (\mathcal{J}_{f_r} - 1)(e_r e_r' \otimes e_r e_r') \\ & + \sum_{r,s=1, r\neq s}^{p} \bigl(\gamma_{sr}(f)(e_r e_r' \otimes e_s e_s') + (e_r e_s' \otimes e_s e_r')\bigr) \Biggr] \\ & \times (I_p \otimes L^{-1}) C'. \end{aligned}$$

More precisely, for any $\vartheta_n = \vartheta + O(n^{-1/2})$ (with $\vartheta = (\mu', (\operatorname{vecd}^\circ L)')'$) and any bounded sequence (τ_n) in $\mathbb{R}^{p^2}$, we have that, under $\mathrm{P}^{(n)}_{\vartheta_n,f}$ as $n \to \infty$,

$$\log\bigl(d\mathrm{P}^{(n)}_{\vartheta_n + n^{-1/2}\tau_n, f} / d\mathrm{P}^{(n)}_{\vartheta_n,f}\bigr) = \tau_n' \Delta_{\vartheta_n,f} - \tfrac{1}{2} \tau_n' \Gamma_{L,f} \tau_n + o_{\mathrm{P}}(1),$$

and $\Delta_{\vartheta_n,f}$ converges in distribution to a p^2-variate normal distribution with mean zero and covariance matrix $\Gamma_{L,f}$.

Semiparametrically efficient (at f) inference procedures on L then may be based on the so-called *efficient central sequence* $\Delta^*_{\vartheta,f;2}$ resulting from $\Delta_{\vartheta,f;2}$ by performing adequate tangent space projections; see [3]. Under $\mathrm{P}^{(n)}_{\vartheta,f}$, $\Delta^*_{\vartheta,f;2}$ is still asymptotically normal with mean zero, but now with covariance matrix $\Gamma^*_{L,f;2}$ (the *efficient information matrix*). This matrix $\Gamma^*_{L,f;2}$ settles the semiparametric efficiency bound at f when performing inference on L. For instance, an estimator $\hat{L}$ is semiparametrically efficient at f if

$$\sqrt{n}\, \operatorname{vecd}^\circ(\hat{L} - L) \xrightarrow{\mathcal{L}} \mathcal{N}_{p(p-1)}(0, (\Gamma^*_{L,f;2})^{-1}). \tag{2.4}$$

The performance of semiparametrically efficient tests on L can similarly be characterized in terms of $\Gamma^*_{L,f;2}$: a test of $\mathcal{H}_0 : L = L_0$ is semiparametrically efficient at f (at asymptotic level α) if its asymptotic powers under local alternatives of the form $\mathcal{H}_1^{(n)}: L = L_0 + n^{-1/2} H$, where H is an arbitrary $p \times p$ matrix with zero diagonal entries, are given by

$$1 - \Psi_{p(p-1)}\big(\chi^2_{p(p-1),1-\alpha}; (\operatorname{vecd}^\circ H)' \Gamma^*_{L_0, f;2} (\operatorname{vecd}^\circ H)\big), \tag{2.5}$$

where $\chi^2_{p(p-1),1-\alpha}$ stands for the α-upper quantile of the $\chi^2_{p(p-1)}$ distribution, and $\Psi_{p(p-1)}(\cdot;\delta)$ denotes the cumulative distribution function of the noncentral $\chi^2_{p(p-1)}$ distribution with noncentrality parameter δ.

2.3. *Invariance arguments.* Instead of the classical tangent space projection approach to compute $\Delta^*_{\vartheta,f;2}$ (as in [6]), we adopt an approach—due to [11]—that rather exploits the invariance structure of the model considered. This will provide a version of the efficient central sequence (parallel to central sequences, efficient central sequences are defined up to $o_{\mathrm{P}}(1)$'s only) that is based on *signed ranks*. Here, signed ranks are defined as $S_i(\vartheta) = (S_{i1}(\vartheta), \ldots, S_{ip}(\vartheta))'$ and $R_i^+(\vartheta) = (R_{i1}^+(\vartheta), \ldots, R_{ip}^+(\vartheta))'$, where $S_{ir}(\vartheta)$ is the sign of $Z_{ir}(\vartheta) = (L^{-1}(X_i - \mu))_r$ and $R_{ir}^+(\vartheta)$ is the rank of $|Z_{ir}(\vartheta)|$ among $|Z_{1r}(\vartheta)|, \ldots, |Z_{nr}(\vartheta)|$. This signed-rank efficient central sequence—$\underline{\Delta}^*_{\vartheta,f;2}$, say—is given in Theorem 2.1 below (the asymptotic behavior of $\underline{\Delta}^*_{\vartheta,f;2}$ will be studied in Appendix A).

To be able to state Theorem 2.1, we need to introduce the following notation. Let $z \mapsto F_+(z) = (F_{+1}(z_1), \ldots, F_{+r}(z_p))'$, with $F_{+r}(t) := \mathrm{P}^{(n)}_{\vartheta,f}[|Z_r(\vartheta)| < t] = 2(\int_{-\infty}^t f_r(s)\,ds) - 1$, $t \geq 0$. Based on this, define $\underline{\Delta}^*_{\vartheta,f;2} := C(I_p \otimes L^{-1})' \operatorname{vec} \underline{T}_{\vartheta,f}$, with

$$\underline{T}_{\vartheta,f} := \operatorname{odiag}\Bigg[\frac{1}{\sqrt{n}} \sum_{i=1}^n \bigg(S_i(\vartheta) \odot \varphi_f\bigg(F_+^{-1}\bigg(\frac{R_i^+(\vartheta)}{n+1}\bigg)\bigg)\bigg) \times \bigg(S_i(\vartheta) \odot F_+^{-1}\bigg(\frac{R_i^+(\vartheta)}{n+1}\bigg)\bigg)'\Bigg],$$

where $\odot$ is the Hadamard (i.e., entrywise) product of two vectors, and where $\operatorname{odiag}(A)$ denotes the matrix obtained from A by replacing all diagonal entries with zeros. Finally, let $\underline{\mathcal{F}}_{\text{ulan}}$ be the collection of pdfs $f \in \mathcal{F}_{\text{ulan}}$ for which each φ_{f_r}, $r = 1, \ldots, p$, is continuous and can be written as the difference of two monotone increasing functions. We then have the following result (see Appendix B for a proof).

THEOREM 2.1. *Fix $\vartheta = (\mu', (\operatorname{vecd}^\circ L)')' \in \Theta$ and $f \in \underline{\mathcal{F}}_{\text{ulan}}$. Then,* (i) *denoting by $\mathrm{E}^{(n)}_{\vartheta,f}$ expectation under $\mathrm{P}^{(n)}_{\vartheta,f}$,*

$$\begin{aligned}\underline{\Delta}^*_{\vartheta,f;2} &:= C(I_p \otimes L^{-1})' \operatorname{vec} \underline{T}_{\vartheta,f} \\ &= \mathrm{E}^{(n)}_{\vartheta,f}[\Delta_{\vartheta,f;2} | S_1(\vartheta), \ldots, S_n(\vartheta), R_1^+(\vartheta), \ldots, R_n^+(\vartheta)] + o_{L^2}(1)\end{aligned}$$

as $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\vartheta,f}$; (ii) *the signed-rank quantity* $\underline{\Delta}^*_{\vartheta,f;2}$ *is a version of the efficient central sequence at* f [*i.e.*, $\underline{\Delta}^*_{\vartheta,f;2} = \Delta^*_{\vartheta,f;2} + o_{L^2}(1)$ *as* $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\vartheta,f}$].

Would the (nonparametric) fixed-ϑ submodels $\mathcal{P}^{(n)}_{\vartheta} := \bigcup_{g\in\mathcal{F}}\{\mathrm{P}^{(n)}_{\vartheta,g}\}$ of the semiparametric model $\bigcup_{\theta\in\Theta}\bigcup_{g\in\mathcal{F}}\{\mathrm{P}^{(n)}_{\theta,g}\}$ in (2.2) be invariant under a group of transformations $\mathcal{G}^{\vartheta}$ that generates $\mathcal{P}^{(n)}_{\vartheta}$, then the main result of [11] would show that the expectation of the original central sequence $\Delta_{\vartheta,f;2}$ conditional upon the corresponding maximal invariant—$\mathcal{I}^{(n)}_{\max}(\vartheta)$, say—is a version of the efficient central sequence $\Delta^*_{\vartheta,f;2}$ at f: as $n \to \infty$, under $\mathrm{P}^{(n)}_{\vartheta,f}$,

$$\Delta^*_{\vartheta,f;2} = \mathrm{E}^{(n)}_{\vartheta,f}\big[\Delta_{\vartheta,f;2} | \mathcal{I}^{(n)}_{\max}(\vartheta)\big] + o_{L^2}(1). \tag{2.6}$$

Such an invariance structure actually exists and the relevant group $\mathcal{G}^{\vartheta}$ collects all transformations

$$\begin{aligned} g^{\vartheta}_h : \mathbb{R}^p \times \cdots \times \mathbb{R}^p &\to \mathbb{R}^p \times \cdots \times \mathbb{R}^p, \\ (x_1, \ldots, x_n) &\mapsto \big(Lh(z_1(\vartheta)) + \mu, \ldots, Lh(z_n(\vartheta)) + \mu\big), \end{aligned}$$

with $z_i(\vartheta) := L^{-1}(x_i - \mu)$ and $h((z_1, \ldots, z_p)') = (h_1(z_1), \ldots, h_p(z_p))'$, where each h_r, $r = 1, \ldots, p$, is continuous, odd, monotone increasing and fixes $+\infty$. It is easy to check that $\mathcal{P}^{(n)}_{\vartheta}$ is invariant under (and is generated by) $\mathcal{G}^{\vartheta}$, and that the corresponding maximal invariant is the vector of signed ranks

$$\mathcal{I}^{(n)}_{\max}(\vartheta) = (S_1(\vartheta), \ldots, S_n(\vartheta), R^+_1(\vartheta), \ldots, R^+_n(\vartheta)); \tag{2.7}$$

Theorem 2.1(ii) then follows from (2.6) and Theorem 2.1(i).

Inference procedures based on $\underline{\Delta}^*_{\vartheta,f;2}$, unlike those (from [6]) based on the efficient central sequence $\Delta^*_{\vartheta,f;2}$ obtained through tangent space projections, are measurable with respect to signed ranks, hence enjoy all nice properties usually associated with rank methods: robustness, ease of computation, validity without density estimation (and, for hypothesis testing, even distribution-freeness), etc.

3. Hypothesis testing. We now consider the problem of testing the null hypothesis $\mathcal{H}_0 : L = L_0$ against the alternative $\mathcal{H}_1 : L \neq L_0$, with unspecified underlying density g. Beyond their intrinsic interest, the resulting tests will play an important role in the construction of the R-estimators of Section 4 below, and they pave the way to testing linear hypotheses on L.

The objective here is to define a test that is semiparametrically efficient at some target density f, yet that remains valid—in the sense that it meets asymptotically the level constraint—under a very broad class of densities g. As we will show, this objective is achieved by the signed-rank test—$\underline{\phi}_f$, say—that rejects $\mathcal{H}_0$ at asymptotic level $\alpha \in (0,1)$ whenever

$$\underline{Q}_f := (\underline{\Delta}^*_{\hat{\vartheta}_0,f;2})'(\Gamma^*_{L_0,f;2})^{-1}\underline{\Delta}^*_{\hat{\vartheta}_0,f;2} > \chi^2_{p(p-1),1-\alpha}, \tag{3.1}$$

where $\Gamma^*_{L,f;2}$ was introduced on Page 2453 (an explicit expression is given below) and where $\hat{\vartheta}_0 = (\hat{\mu}', (\mathrm{vecd}^\circ L_0)')'$ is based on a sequence of estimators $\hat{\mu}$ that is locally asymptotically discrete (see Appendix A for a precise definition) and root-n consistent under the null.

Possible choices for $\hat{\mu}$ include (discretized versions of) the sample mean $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ or the transformation-retransformation componentwise median $\hat{\mu}_{\mathrm{Med}} := L_0 \,\mathrm{Med}[L_0^{-1}X_1, \dots, L_0^{-1}X_n]$, where $\mathrm{Med}[\cdot]$ returns the vector of univariate medians. We favor the sign estimator $\hat{\mu}_{\mathrm{Med}}$, since it is very much in line with the signed-rank tests $\underline{\phi}_f$ and enjoys good robustness properties. However, we stress that Theorem 3.1 below, which states the asymptotic properties of the proposed signed-rank tests, implies that the choice of $\hat{\mu}$ does not affect the asymptotic properties of $\underline{\phi}_f$, at any $g \in \mathcal{F}_{\mathrm{ulan}}$.

In order to state this theorem, we need to define

$$\begin{aligned}\Gamma^*_{L,f,g;2} &:= C(I_p \otimes L^{-1})' G_{f,g}(I_p \otimes L^{-1})C' \\ &:= C(I_p \otimes L^{-1})' \\ &\quad \times \left[\sum_{r,s=1, r\neq s}^{p} \big(\gamma_{sr}(f,g)(e_r e_r' \otimes e_s e_s') + \rho_{rs}(f,g)(e_r e_s' \otimes e_s e_r')\big)\right] \\ &\quad \times (I_p \otimes L^{-1})C',\end{aligned} \tag{3.2}$$

where we let

$$\gamma_{rs}(f,g) := \int_0^1 \varphi_{f_r}(F_r^{-1}(u))\varphi_{g_r}(G_r^{-1}(u))\,du \times \int_0^1 F_s^{-1}(u)G_s^{-1}(u)\,du \tag{3.3}$$

and

$$\rho_{rs}(f,g) := \int_0^1 F_r^{-1}(u)\varphi_{g_r}(G_r^{-1}(u))\,du \times \int_0^1 \varphi_{f_s}(F_s^{-1}(u))G_s^{-1}(u)\,du. \tag{3.4}$$

We also let $\Gamma^*_{L,f;2} := \Gamma^*_{L,f,f;2}$ and $G_f := G_{f,f}$, that involve $\gamma_{rs}(f,f) = \gamma_{rs}(f)$ (see Section 2.2) and $\rho_{rs}(f,f) = 1$. We then have the following result (see Appendix B for a proof).

THEOREM 3.1. *Fix $f \in \underline{\mathcal{F}}_{\mathrm{ulan}}$. Then,* (i) *under $\mathrm{P}^{(n)}_{\vartheta_0,g}$ and under $\mathrm{P}^{(n)}_{\vartheta_0+n^{-1/2}\tau,g}$, with $\vartheta_0 = (\mu', (\mathrm{vecd}^\circ L_0)')'$, $\tau = (\tau_1', \tau_2')' \in \mathbb{R}^p \times \mathbb{R}^{p(p-1)}$ and $g \in \mathcal{F}_{\mathrm{ulan}}$,*

$$\underline{Q}_f \xrightarrow{\mathcal{L}} \chi^2_{p(p-1)} \quad \text{and} \quad \underline{Q}_f \xrightarrow{\mathcal{L}} \chi^2_{p(p-1)}(\tau_2'(\Gamma^*_{L_0,f,g;2})'(\Gamma^*_{L_0,f;2})^{-1}\Gamma^*_{L_0,f,g;2}\tau_2),$$

respectively, as $n \to \infty$. (ii) *The sequence of tests $\underline{\phi}_f^{(n)}$ has asymptotic level α under $\bigcup_{\mu\in\mathbb{R}^p}\bigcup_{g\in\mathcal{F}_{\mathrm{ulan}}}\{\mathrm{P}^{(n)}_{\mu,L_0,g}\}$.* (iii) *The sequence of tests $\underline{\phi}_f^{(n)}$ is semiparamet-*

rically efficient, still at asymptotic level α, *when testing* $\mathcal{H}_0 : L = L_0$ *against* $H_1^f : L \neq L_0$ *with noise density* f (*i.e., when testing* $\bigcup_{\mu\in\mathbb{R}^p}\bigcup_{g\in\mathcal{F}_{\text{ulan}}}\{\mathrm{P}^{(n)}_{\mu,L_0,g}\}$ *against* $\bigcup_{\mu\in\mathbb{R}^p}\bigcup_{L\in\mathcal{M}_{1p}\setminus\{L_0\}}\{\mathrm{P}^{(n)}_{\mu,L,f}\}$).

The test $\underline{\phi}_f$ achieves semiparametric efficiency at f [Theorem 3.1(iii)], and also at any f_σ, with $f_\sigma(z) := \prod_{r=1}^p \sigma_r^{-1} f_r(z_r/\sigma_r)$, where $\sigma_r > 0$ for all r—it can indeed be checked that $\underline{\phi}_{f_\sigma} = \underline{\phi}_f$. Most importantly, Theorem 3.1 shows also that $\underline{\phi}_f$ remains valid under any $g \in \mathcal{F}_{\text{ulan}}$. By proceeding as in Lemma 4.2 of [19], this can even be extended to any $g \in \mathcal{F}$, which allows us to avoid any finite moment condition.

This is to be compared to the semiparametric approach of Chen and Bickel [6]—these authors focus on point estimation, but their methodology also leads to tests that enjoy the same properties as their estimators. Their procedures achieve uniform (in g) semiparametric efficiency, while our methods achieve semiparametric efficiency at the target density f only—more precisely, at any corresponding f_σ. However, it turns out that the performances of our procedures do not depend much on the target density f, so that our procedures are close to achieving uniform (in g) semiparametric efficiency; see the simulations in the supplemental article [16]. As any uniformly semiparametrically efficient procedures (see [1]), Chen and Bickel's procedures require estimating g, hence choosing various smoothing parameters. In contrast, our procedures, by construction, are invariant (here, signed-rank) ones. As such, they do not require us to estimate densities, and they are robust, easy to compute, etc.

One might still object that the choice of f is quite arbitrary. This choice should be based on the practitioner's prior belief on the underlying densities. If he/she has no such prior belief, a kernel estimate $\hat{f}$ of f could be used. The resulting test $\underline{\phi}_{\hat{f}}$ would then enjoy the same properties as any $\underline{\phi}_f$ in terms of validity, since kernel density estimators, in the symmetric case considered, typically are measurable with respect to the order statistics of the $|Z_{ir}(\hat{\vartheta}_0)|$'s, that, asymptotically, are stochastically independent of the signed ranks $S_{ir}(\hat{\vartheta}_0)$, $R^+_{ir}(\hat{\vartheta}_0)$ used in $\underline{\phi}_f$; see [11] for details. The test $\underline{\phi}_{\hat{f}}$ would further achieve uniform semiparametric efficiency.

Further results on the proposed tests are given in the supplemental article [16]. More precisely, a simple explicit expression of the test statistics, local asymptotic powers of the corresponding tests, and simulation results can be found there.

We finish this section by describing the extension of our signed-rank tests to the problem of testing a fixed (arbitrary) linear hypothesis on L, which includes many instances of high practical relevance (we mentioned a few in the Introduction). Denoting by $\mathcal{V}(\Omega)$ the vector space that is spanned by the columns of the $p(p-1)\times\ell$ matrix Ω (which is assumed to have full rank ℓ), we consider the testing problem

$$\begin{cases}\mathcal{H}_0(L_0,\Omega) : (\operatorname{vecd}^\circ L) \in (\operatorname{vecd}^\circ L_0) + \mathcal{V}(\Omega)\\ \mathcal{H}_1(L_0,\Omega) : (\operatorname{vecd}^\circ L) \notin (\operatorname{vecd}^\circ L_0) + \mathcal{V}(\Omega),\end{cases} \tag{3.5}$$

for some fixed $L_0 \in \mathcal{M}_{1p}$. If one forgets about the tacitly assumed constraint that $L \in \mathcal{M}_{1p}$ in (3.5), the null hypothesis above imposes a set of linear constraints on L. This clearly includes all testing problems mentioned in the Introduction: testing that a given column of L is equal to a fixed vector, testing that a given (off-diagonal) entry of L is zero and testing block-diagonality of L.

Inspired by the tests from [18] (Section 10.9), the analog of our signed-rank test $\underline{\phi}_f$ above then rejects $\mathcal{H}_0(L_0, \Omega)$ for large values of

$$\underline{Q}_f(L_0, \Omega) := (\underline{\Delta}^*_{\hat{\vartheta},f;2})' P_\Omega \underline{\Delta}^*_{\hat{\vartheta},f;2}$$

with $P_\Omega := (\Gamma^*_{\hat{L},f;2})^- - \Omega(\Omega'\Gamma^*_{\hat{L},f;2}\Omega)^-\Omega'$, where B^- denotes the Moore–Penrose pseudoinverse of B, and where $\hat{\vartheta} = (\hat{\mu}', (\mathrm{vecd}^\circ \hat{L})')'$ is an estimator of ϑ that is locally and asymptotically discrete, root-n consistent under the null, and *constrained*—in the sense that $\hat{L}$ satisfies the linear constraints in $\mathcal{H}_0(L_0, \Omega)$.

It can be shown that this signed-rank test achieves semiparametric optimality at f (the relevant optimality concept here is *most stringency*; see, e.g., [19] for a discussion) and remains valid under any $g \in \underline{\mathcal{F}}_{\mathrm{ulan}}$. Its null asymptotic distribution is still chi-square, now with $r := \mathrm{Trace}[P_\Omega \Gamma^*_{L,f;2}]$ degrees of freedom (this directly follows from Theorem 9.2.1 in [24] and Theorem A.1); at asymptotic level α, the resulting asymptotic critical value (that actually does not depend on the true value L) therefore is $\chi^2_{r;1-\alpha}$. Just as for the tests $\underline{\phi}_f$, it is still possible to compute asymptotic powers under sequences of local alternatives. It is clear, however, that a thorough study of the properties of the tests above, for a general linear hypothesis, is beyond the scope of the present paper, hence is left for future research. In the important particular case of testing block-diagonality of L, a complete investigation of the signed-rank tests can be found in [19].

4. Point estimation.

We turn to the problem of estimating L, which is of primary importance for applications. Denoting by $\underline{Q}_f = \underline{Q}_f(L_0)$ the signed-rank test statistic for $\mathcal{H}_0 : L = L_0$ in (3.1), a natural signed-rank estimator of L is obtained by "inverting the corresponding test,"

$$\underline{\hat{L}}_{f;\arg\min} = \arg\min_{L \in \mathcal{M}_{1p}} \underline{Q}_f(L).$$

This estimator, however, is not satisfactory: as any signed-rank quantity, the objective function $L \mapsto \underline{Q}_f(L)$ is piecewise constant, hence discontinuous and nonconvex, which makes it very difficult to derive the asymptotic properties of $\underline{\hat{L}}_{f;\arg\min}$. It is also virtually impossible to compute $\underline{\hat{L}}_{f;\arg\min}$ in practice, since this lack of smoothness and convexity essentially forces computing the estimator by simply running over a grid of possible values of the $p(p-1)$-dimensional parameter L—a strategy that cannot provide a reasonable approximation of $\underline{\hat{L}}_{f;\arg\min}$, even for moderate values of p. Finally, there is no way to estimate the asymptotic covariance matrix of $\underline{\hat{L}}_{f;\arg\min}$, which rules out the possibility to derive confidence zones for L, hence drastically restricts the practical relevance of this estimator.

In order to avoid the aforementioned drawbacks, we propose adopting a one-step approach that was first used in [7] for the problem of estimating the shape of an elliptical distribution or in [9] in a more general context. The resulting one-step signed-rank estimators—in the sequel, we simply speak of *one-step rank estimators* or *one-step R-estimators*—can easily be computed in practice, their asymptotic properties can be derived explicitly, and their asymptotic covariance matrix can be estimated consistently.

4.1. *One-step R-estimators of L.* To initiate the one-step procedure, a preliminary estimator is needed. In the present context, we will assume that a root-n consistent and locally asymptotically discrete estimator $\tilde{\vartheta} = (\tilde{\mu}', (\operatorname{vecd}^\circ \tilde{L})')'$ is available. As we will show, the asymptotic properties of the proposed one-step R-estimators will not be affected by the choice of $\tilde{\vartheta}$. Practical choices will be provided in Section 5.

Describing our one-step R-estimators requires:

ASSUMPTION (A). For all $r \neq s \in \{1, \ldots, p\}$, we dispose of sequences of estimators $\hat{\gamma}_{rs}(f)$ and $\hat{\rho}_{rs}(f)$ that: (i) are locally asymptotically discrete and that (ii), for any $g \in \mathcal{F}_{\mathrm{ulan}}$, satisfy $\hat{\gamma}_{rs}(f) = \gamma_{rs}(f, g) + o_{\mathrm{P}}(1)$ and $\hat{\rho}_{rs}(f) = \rho_{rs}(f, g) + o_{\mathrm{P}}(1)$ as $n \to \infty$, under $\bigcup_{\vartheta \in \Theta}\{\mathrm{P}^{(n)}_{\vartheta,g}\}$.

Sequences of estimators fulfilling this assumption will be provided in Section 4.2 below. At this point, just note that plugging in (3.2) the estimators from Assumption (A) and the preliminary estimator $\tilde{L}$, defines a statistic—$\hat{\Gamma}^*_{\tilde{L},f;2}$, say—that consistently estimates $\Gamma^*_{L,f,g;2}$ under $\bigcup_{\vartheta \in \Theta}\{\mathrm{P}^{(n)}_{\vartheta,g}\}$.

For any target density f, we propose the one-step R-estimator $\underline{\hat{L}}_f$, with values in $\mathcal{M}_{1p}$, defined by

$$\operatorname{vecd}^\circ \underline{\hat{L}}_f := (\operatorname{vecd}^\circ \tilde{L}) + n^{-1/2}(\hat{\Gamma}^*_{\tilde{L},f;2})^{-1}\underline{\Delta}^*_{\tilde{\vartheta},f;2}. \tag{4.1}$$

The following result states the asymptotic properties of this estimator (see Appendix B for a proof).

THEOREM 4.1. *Let Assumption (A) hold, and fix $f \in \underline{\mathcal{F}}_{\mathrm{ulan}}$. Then* (i) *under $\mathrm{P}^{(n)}_{\vartheta,g}$, with $\vartheta = (\mu', (\operatorname{vecd}^\circ L)')' \in \Theta$ and $g \in \mathcal{F}_{\mathrm{ulan}}$, we have that*

$$\sqrt{n}\operatorname{vec}(\underline{\hat{L}}_f - L) = C'(\Gamma^*_{L,f,g;2})^{-1}\underline{\Delta}^*_{\vartheta,f;2} + o_{\mathrm{P}}(1) \tag{4.2}$$

$$= C'(\Gamma^*_{L,f,g;2})^{-1}\Delta^*_{\vartheta,f,g;2} + o_{\mathrm{P}}(1) \tag{4.3}$$

$$\xrightarrow{\mathcal{L}} \mathcal{N}_{p(p-1)}(0, C'(\Gamma^*_{L,f,g;2})^{-1}\Gamma^*_{L,f;2}(\Gamma^*_{L,f,g;2})^{-1\prime}C) \tag{4.4}$$

as $n \to \infty$, *where* $\Delta^*_{\vartheta,f,g;2}$ *is defined in Theorem* A.1 (*see Appendix* A). (ii) *The estimator* $\underline{\hat{L}}_f$ *is semiparametrically efficient at* f.

The result in (4.2) justifies calling $\underline{\hat{L}}_f$ an R-estimator since it shows that $n^{1/2}(\underline{\hat{L}}_f - L)$ is asymptotically equivalent to a random matrix that is measurable with respect to the signed ranks $S_i(\vartheta)$, $R_i^+(\vartheta)$ in (2.7). The asymptotic equivalence in (4.3) gives a Bahadur-type representation result for $\underline{\hat{L}}_f$ with summands that are independent and identically distributed, hence leads trivially to the asymptotic normality result in (4.4). Recalling that $\hat{\Gamma}^*_{\tilde{L},f;2}$ consistently estimates $\Gamma^*_{L,f,g;2}$ under $\bigcup_{\vartheta\in\Theta}\{\mathrm{P}^{(n)}_{\vartheta,g}\}$, it is clear that asymptotic (signed-rank) confidence zones for L may easily be obtained from this asymptotic normality result.

For $r \neq s \in \{1,\dots,p\}$, define $\hat{\alpha}_{rs}(f)$ and $\hat{\beta}_{rs}(f)$ as the statistics obtained by plugging the estimators $\hat{\gamma}_{rs}(f)$ and $\hat{\rho}_{rs}(f)$ from Assumption (A) in

$$\begin{cases} \alpha_{rs}(f,g) := \dfrac{\gamma_{rs}(f,g)}{\gamma_{rs}(f,g)\gamma_{sr}(f,g) - \rho_{rs}(f,g)\rho_{sr}(f,g)} \\ \beta_{rs}(f,g) := \dfrac{-\rho_{rs}(f,g)}{\gamma_{rs}(f,g)\gamma_{sr}(f,g) - \rho_{rs}(f,g)\rho_{sr}(f,g)}, \end{cases} \tag{4.5}$$

and let $\hat{\alpha}_{rr}(f) := 0 =: \hat{\beta}_{rr}(f)$, $r = 1,\dots,p$. The estimator $\underline{\hat{L}}_f$ then admits the following explicit expression (see Appendix B for a proof).

THEOREM 4.2. *Let Assumption* (A) *hold, and fix* $f \in \underline{\mathcal{F}}_{\mathrm{ulan}}$. *Let* $\hat{N}_f := (\hat{\mathcal{A}}'_f \odot \underline{T}_{\tilde{\vartheta},f}) + (\hat{\mathcal{B}}'_f \odot \underline{T}'_{\tilde{\vartheta},f})$, *where we let* $\hat{\mathcal{A}}_f := (\hat{\alpha}_{rs}(f))$ *and* $\hat{\mathcal{B}}_f := (\hat{\beta}_{rs}(f))$. *Then the estimator* $\underline{\hat{L}}_f$ *rewrites*

$$\underline{\hat{L}}_f = \tilde{L} + \frac{1}{\sqrt{n}}\tilde{L}[\hat{N}_f - \mathrm{diag}(\tilde{L}\hat{N}_f)], \tag{4.6}$$

where $\mathrm{diag}(A) = A - \mathrm{odiag}(A)$ *stands for the diagonal matrix with the same diagonal entries as* A.

It is straightforward to check that the role of the term $-\frac{1}{\sqrt{n}}\tilde{L}\,\mathrm{diag}(\tilde{L}\hat{N}_f)$ in the one-step correction $\frac{1}{\sqrt{n}}\tilde{L}[\hat{N}_f - \mathrm{diag}(\tilde{L}\hat{N}_f)]$ of $\tilde{L}$ is merely to ensure that the diagonal entries of $\underline{\hat{L}}_f$ remain equal to one, hence that $\underline{\hat{L}}_f$ takes values in $\mathcal{M}_{1p}$ (for n large enough).

As shown above, the estimator $\underline{\hat{L}}_f$ enjoys very nice properties: its asymptotic behavior is completely characterized, it is semiparametrically efficient under correctly specified densities, yet remains root-n consistent and asymptotically normal under a broad range of densities g, its asymptotic covariance matrix can easily be estimated consistently, etc.

However, $\underline{\hat{L}}_f$ requires estimates $\hat{\gamma}_{rs}(f)$ and $\hat{\rho}_{rs}(f)$ that fulfill Assumption (A). We now provide such estimates.

4.2. *Estimation of cross-information coefficients.* Of course, it is always possible to estimate consistently the cross-information coefficients $\gamma_{rs}(f, g)$ and $\rho_{rs}(f, g)$ by replacing g in (3.3) and (3.4) with appropriate window or kernel density estimates—this can be achieved since the residuals $Z_{ir}(\tilde{\vartheta})$, $i = 1, \ldots, n$ typically are asymptotically i.i.d. with density g_r. Rank-based methods, however, intend to eliminate—through invariance arguments—the nuisance g without estimating it, so that density estimation methods simply are antinomic to the spirit of rank-based methods.

Therefore, we rather propose a solution that is based on ranks and avoids estimating the underlying nuisance g. The method, that relies on the asymptotic linearity—under g—of an appropriate rank-based statistic $\underline{S}_{\vartheta,f}$, was first used in [7], where there is only one cross-information coefficient $J(f, g)$ to be estimated. There, it is crucial that $J(f, g)$ is involved as a scalar factor in the asymptotic covariance matrix, under g, between the rank-based efficient central sequence $\underline{\Delta}^*_{\vartheta,f}$ and the parametric central sequence $\Delta_{\vartheta,g}$. In [5], the method was extended to allow for the estimation of a cross-information coefficient that appears as a scalar factor in the linear term of the asymptotic linearity, under g, of a (possibly vector-valued) rank-based statistic $\underline{S}_{\vartheta,f}$.

In all cases, thus, this method was only used to estimate a *single* cross-information coefficient that appears as a *scalar factor* in some structural—typically, cross-information—matrix. In this respect, our problem, which requires us to estimate $2p(p-1)$ cross-information quantities appearing in various entries of the cross-information matrix $\Gamma^*_{L,f,g;2}$, is much more complex. Yet, as we now show, it allows for a solution relying on the same basic idea of exploiting the asymptotic linearity, under g, of an appropriate f-score rank-based statistic.

Based on the preliminary estimator $\tilde{\vartheta} := (\tilde{\mu}', (\mathrm{vecd}^\circ\, \tilde{L})')'$ at hand, define $\tilde{\vartheta}^{\gamma_{rs}}_\lambda := (\tilde{\mu}', (\mathrm{vecd}^\circ\, \tilde{L}^{\gamma_{rs}}_\lambda)')'$, $\lambda \geq 0$, with

$$\tilde{L}^{\gamma_{rs}}_\lambda := \tilde{L} + n^{-1/2}\lambda(\underline{T}_{\tilde{\vartheta},f})_{rs}\tilde{L}\big(e_r e_s' - \mathrm{diag}(\tilde{L}e_r e_s')\big),$$

and $\tilde{\vartheta}^{\rho_{rs}}_\lambda := (\tilde{\mu}', (\mathrm{vecd}^\circ\, \tilde{L}^{\rho_{rs}}_\lambda)')'$, $\lambda \geq 0$, with

$$\tilde{L}^{\rho_{rs}}_\lambda := \tilde{L} + n^{-1/2}\lambda(\underline{T}_{\tilde{\vartheta},f})_{sr}\tilde{L}\big(e_r e_s' - \mathrm{diag}(\tilde{L}e_r e_s')\big);$$

note that, at $\lambda = 0$, $\tilde{\vartheta}^{\gamma_{rs}}_\lambda = \tilde{\vartheta}^{\rho_{rs}}_\lambda = \tilde{\vartheta}$. We then have the following result that is crucial for the construction of the estimators $\hat{\gamma}_{rs}(f)$ and $\hat{\rho}_{rs}(f)$; see Appendix B for a proof.

LEMMA 4.1. *Fix $\vartheta \in \Theta$, $f \in \underline{\mathcal{F}}_{\mathrm{ulan}}$, $g \in \mathcal{F}_{\mathrm{ulan}}$ and $r \neq s \in \{1, \ldots, p\}$. Then $h^{\gamma_{rs}}(\lambda) := (\underline{T}_{\tilde{\vartheta},f})_{rs}(\underline{T}_{\tilde{\vartheta}^{\gamma_{rs}}_\lambda,f})_{rs} = (1 - \lambda\gamma_{rs}(f, g))((\underline{T}_{\tilde{\vartheta},f})_{rs})^2 + o_{\mathrm{P}}(1)$ and $h^{\rho_{rs}}(\lambda) := (\underline{T}_{\tilde{\vartheta},f})_{sr}(\underline{T}_{\tilde{\vartheta}^{\rho_{rs}}_\lambda,f})_{sr} = (1 - \lambda\rho_{rs}(f, g))((\underline{T}_{\tilde{\vartheta},f})_{sr})^2 + o_{\mathrm{P}}(1)$ as $n \to \infty$, under $\mathrm{P}^{(n)}_{\vartheta,g}$.*

The mappings $\lambda \mapsto h^{\gamma_{rs}}(\lambda)$ and $\lambda \mapsto h^{\rho_{rs}}(\lambda)$ assume a positive value in $\lambda = 0$, and, as shown by Lemma 4.1, are—up to $o_{\mathrm{P}}(1)$'s as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta,g}$—monotone decreasing functions that become negative at $\lambda = (\gamma_{rs}(f,g))^{-1}$ and $\lambda = (\rho_{rs}(f,g))^{-1}$, respectively. Restricting to a grid of values of the form $\lambda_j = j/c$ for some large discretization constant c (which is needed to achieve the required discreteness), this naturally leads—via linear interpolation—to the estimators $\hat{\gamma}_{rs}(f)$ and $\hat{\rho}_{rs}(f)$ defined through

$$
\begin{aligned}
(\hat{\gamma}_{rs}(f))^{-1} := \lambda_{\gamma_{rs}} &:= \lambda^-_{\gamma_{rs}} + \frac{(\lambda^+_{\gamma_{rs}} - \lambda^-_{\gamma_{rs}}) h^{\gamma_{rs}}(\lambda^-_{\gamma_{rs}})}{h^{\gamma_{rs}}(\lambda^-_{\gamma_{rs}}) - h^{\gamma_{rs}}(\lambda^+_{\gamma_{rs}})} \\
&= \lambda^-_{\gamma_{rs}} + \frac{c^{-1} h^{\gamma_{rs}}(\lambda^-_{\gamma_{rs}})}{h^{\gamma_{rs}}(\lambda^-_{\gamma_{rs}}) - h^{\gamma_{rs}}(\lambda^+_{\gamma_{rs}})}
\end{aligned}
\tag{4.7}
$$

with $\lambda^-_{\gamma_{rs}} := \inf\{j \in \mathbb{N} : h^{\gamma_{rs}}(\lambda_{j+1}) < 0\}$ and $\lambda^+_{\gamma_{rs}} := \lambda^-_{\gamma_{rs}} + \frac{1}{c}$, and

$$
(\hat{\rho}_{rs}(f))^{-1} := \lambda_{\rho_{rs}} := \lambda^-_{\rho_{rs}} + \frac{c^{-1} h^{\rho_{rs}}(\lambda^-_{\rho_{rs}})}{h^{\rho_{rs}}(\lambda^-_{\rho_{rs}}) - h^{\rho_{rs}}(\lambda^+_{\rho_{rs}})}
\tag{4.8}
$$

with $\lambda^-_{\rho_{rs}} := \inf\{j \in \mathbb{N} : h^{\rho_{rs}}(\lambda_{j+1}) < 0\}$ and $\lambda^+_{\rho_{rs}} := \lambda^-_{\rho_{rs}} + \frac{1}{c}$. We have the following result (see the supplemental article [16] for a proof).

THEOREM 4.3. *Fix $\vartheta \in \Theta$, $f \in \underline{\mathcal{F}}_{\mathrm{ulan}}$, and $g \in \mathcal{F}_{\mathrm{ulan}}$. Assume that $\tilde{\vartheta}$ is such that, for all $\varepsilon > 0$, there exist $\delta_\varepsilon > 0$ and an integer N_ε such that*

$$
\mathrm{P}^{(n)}_{\vartheta,g}[(\underline{T}_{\tilde{\vartheta},f})_{rs} \geq \delta_\varepsilon] \geq 1 - \varepsilon
\tag{4.9}
$$

for all $n \geq N_\varepsilon$, $r \neq s \in \{1,\ldots,p\}$. Then, for any such r, s, $\hat{\gamma}_{rs}(f) = \gamma_{rs}(f,g) + o_{\mathrm{P}}(1)$ and $\hat{\rho}_{rs}(f) = \rho_{rs}(f,g) + o_{\mathrm{P}}(1)$, as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta,g}$, hence $\hat{\gamma}_{rs}(f)$ and $\hat{\rho}_{rs}(f)$ satisfy Assumption (A).

We point out that the assumption in (4.9) is extremely mild, as it only requires that there is no couple (r,s), $r \neq s$, for which $(\underline{T}_{\tilde{\vartheta},f})_{rs}$ asymptotically has an atom in zero. It therefore rules out preliminary estimators $\tilde{L}$ defined through the (rank-based) f-likelihood equation $(\underline{T}_{\vartheta,f})_{rs} = 0$.

5. Simulations. Here we report simulation results for point estimation only—simulation results for hypothesis testing can be found in the supplemental article [16]. Our aim is to both compare the proposed estimators with some competitors and to investigate the validity of asymptotic results.

We used the following competitors: (i) FastICA from [12, 13], which is by far the most commonly used estimate in practice; we used here its deflation based version with the standard nonlinearity function pow3. (ii) FOBI from [4], which is one of the earliest solutions to the ICA problem and is often used as a benchmark estimate. (iii) The estimate based on two scatter matrices from [20]; here the two scatter matrices used are the regular empirical covariance matrix (COV) and the van der Waerden rank-based estimator (HOP) from [7] (actually, HOP is not a scatter matrix but rather a shape matrix, which is allowed in [20]). Root-n consistency of the resulting estimates $\hat{L}_{\mathrm{FICA}}$, $\hat{L}_{\mathrm{FOBI}}$ and $\hat{L}_{\mathrm{COV\_HOP}}$ of L requires finite sixth-, eighth- and fourth-order moments, respectively, and follows from [14, 15] and [21].

We focused on the bivariate case $p = 2$, and we generated, for three different setups indexed by $d \in \{1, 2, 3\}$, $M = 2{,}000$ independent random samples $Z_i^{(d,m)} = (Z_{i1}^{(d,m)}, Z_{i2}^{(d,m)})'$, $i = 1, \dots, n$, of size $n = 4{,}000$. Denoting by $g^{(d)}(z) = g_1^{(d)}(z_1)g_2^{(d)}(z_2)$ the common pdf of $Z_i^{(d,m)}$, $i = 1, \dots, n, m = 1, \dots, M$, the marginal densities $g_1^{(d)}$ and $g_2^{(d)}$ were chosen as follows:

(i) In Setup $d = 1$, $g_1^{(d)}$ is the pdf of the standard normal distribution ($\mathcal{N}$), and $g_2^{(d)}$ is the pdf of the Student distribution with 5 degrees of freedom (t_5);

(ii) In Setup $d = 2$, $g_1^{(d)}$ is the pdf of the logistic distribution with scale parameter one (log), and $g_2^{(d)}$ is t_5;

(iii) In Setup $d = 3$, $g_1^{(d)}$ is t_8 and $g_2^{(d)}$ is t_5.

We chose to use $L = I_2$ and $\mu = (0, 0)'$, so that the observations are given by $X_i^{(d,m)} = LZ_i^{(d,m)} + \mu = Z_i^{(d,m)}$ (other values of L and μ led to extremely similar results).

For each sample, we computed the competing estimates $\hat{L}_{\mathrm{FICA}}$, $\hat{L}_{\mathrm{FOBI}}$ and $\hat{L}_{\mathrm{COV\_HOP}}$ defined above. Each of these were also used as a preliminary estimator $\tilde{L}$ in the construction of three R-estimators: $\underline{\hat{L}}_{f^{(j)}}$, $j = 1, 2, 3$, with $f^{(j)} = g^{(j)}$ for all j. In the resulting nine R-estimators, we used the location estimate $\hat{\mu} = \tilde{L}\,\mathrm{Med}[\tilde{L}^{-1}X_1, \dots, \tilde{L}^{-1}X_n]$, based on the preliminary estimate $\tilde{L}$ used to initiate the one-step procedure.

Figure 1 reports, for each setup d, a boxplot of the M squared errors

$$\|\hat{L}(X_1^{(d,m)}, \dots, X_n^{(d,m)}) - L\|^2 = \sum_{\substack{r,s=1 \\ r \neq s}}^{p} (\hat{L}_{rs}(X_1^{(d,m)}, \dots, X_n^{(d,m)}) - L_{rs})^2 \tag{5.1}$$

for each of the twelve estimators $\hat{L}$ considered (the nine R-estimators and their three competitors).

The results show that, in each setup, all R-estimators dramatically improve over their competitors. The behavior of the R-estimators does not much depend on the preliminary estimator $\tilde{L}$ used. Optimality of $\underline{\hat{L}}_{f^{(d)}}$ in Setup d is confirmed. Most importantly, as stated for hypothesis testing at the end of Section 3, the performances of the R-estimators do not depend much on the target density $f^{(j)}$ adopted,

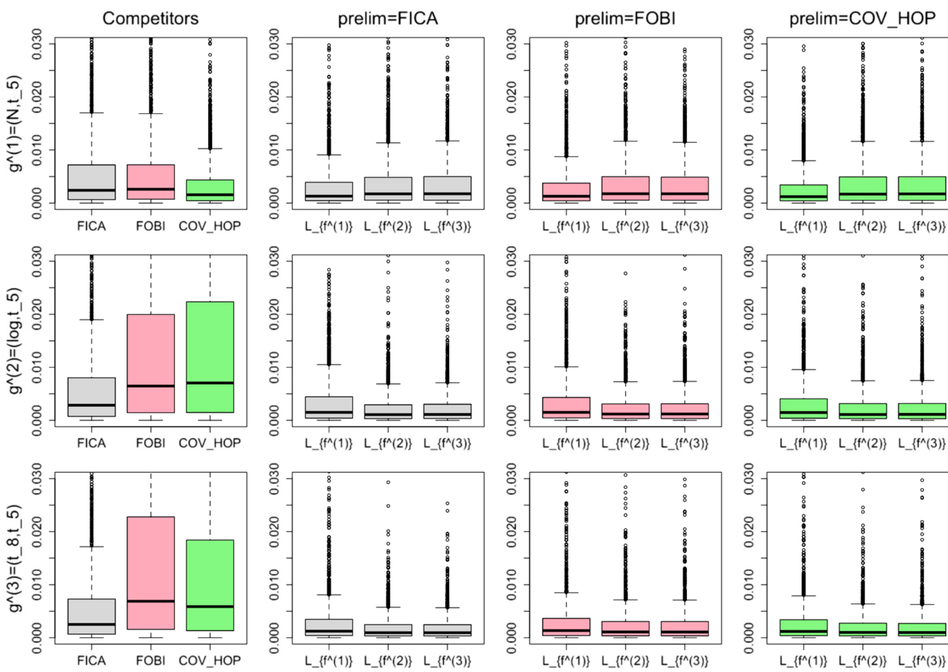

FIG. 1. *Boxplots of the squared errors* $\|\hat{L} - L\|^2$ [*see* (5.1)] *obtained in* $M = 2{,}000$ *replications from setups* $d = 1, 2, 3$ (*associated with underlying distributions* $g^{(d)}$, $d = 1, 2, 3$) *for the competitors* $\hat{L}_{\mathrm{FICA}}$, $\hat{L}_{\mathrm{FOBI}}$ *and* $\hat{L}_{\mathrm{COV\_HOP}}$, *and the nine* R-*estimators* $\hat{L}_f$ *resulting from all combinations of a target density* $f^{(j)} = g^{(j)}$, $j = 1, 2, 3$, *and one of the three preliminary estimators* $\hat{L}_{\mathrm{FICA}}$, $\hat{L}_{\mathrm{FOBI}}$ *and* $\hat{L}_{\mathrm{COV\_HOP}}$; *see Section* 5 *for details. The sample size is* $n = 4{,}000$.

so that one should not worry much about the choice of the target density in practice. Quite surprisingly, R-estimators behave remarkably well even when based on preliminary estimators that, due to heavy tails, fail to be root-n consistent.

In order to investigate small-sample behavior of the estimates, we reran the exact same simulation with sample size $n = 800$; in ICA, where most applications involve sample sizes that are not in hundreds, but much larger, this sample size can indeed be considered small. Results are reported in Figure 2. They indicate that, in Setups 2 and 3, R-estimators still improve significantly over their competitors, and particularly over $\hat{L}_{\mathrm{FOBI}}$ and $\hat{L}_{\mathrm{COV\_HOP}}$. In Setup 1, there seem to be no improvement. Compared to results for $n = 4{,}000$, the behavior of one-step R-estimators here depends more on the preliminary estimator used. Performances of R-estimators again do not depend crucially on the target density, and optimality under correctly specified densities is preserved in most cases.

As a conclusion, for practical sample sizes, the proposed R-estimators outperform the standard competitors considered, and their behavior is very well in line with our asymptotic results.

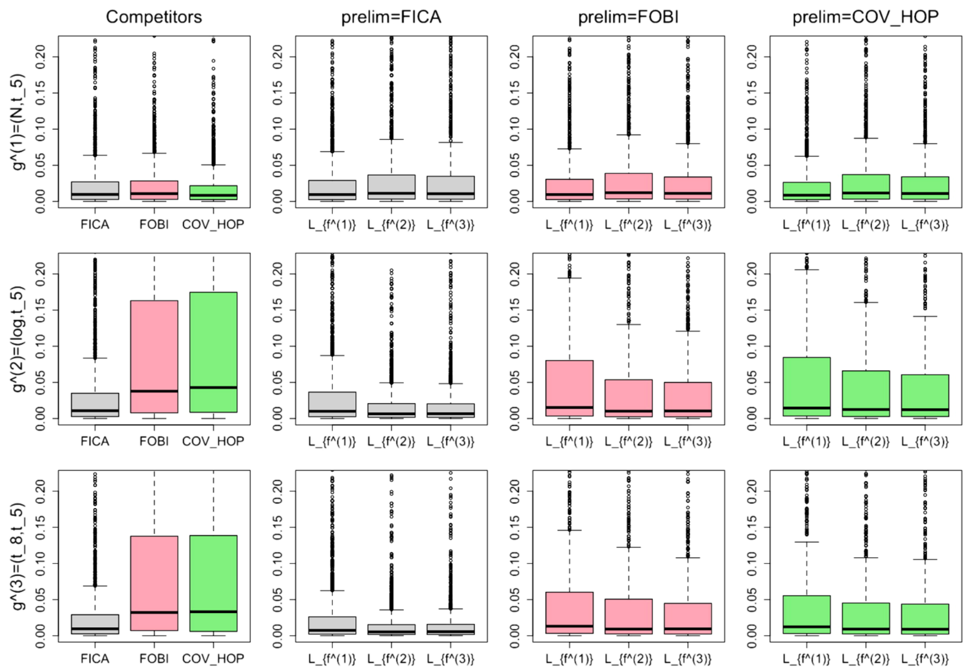

FIG. 2. *The same boxplots as in Figure 1, but based on sample size* $n = 800$.

Finally, we illustrate the proposed method for estimating cross-information coefficients. We consider again the first 50 replications of our simulation with $n = 4{,}000$, and focus on Setup 1 ($g = g^{(1)}$) and the target density $f = f^{(3)}$ ($\neq g^{(1)}$). The cross-information coefficients to be estimated then are $\gamma_{12}(f, g) \approx 1.478$, $\gamma_{21}(f, g) \approx 0.862$, $\rho_{12}(f, g) \approx 1.149$ and $\rho_{21}(f, g) \approx 0.887$. The upper left picture in Figure 3 shows 150 graphs of the mapping $\lambda \mapsto h^{\gamma_{12}}(\lambda)$ (based on $f = f^{(3)}$), among which the 50 pink curves are based on $\tilde{L} = \hat{L}_{\mathrm{FICA}}$, the 50 green curves are based on $\tilde{L} = \hat{L}_{\mathrm{FOBI}}$, and the 50 blue ones are based on $\tilde{L} = \hat{L}_{\mathrm{COV\_HOP}}$. The upper right, bottom left and bottom right pictures of the same figure provide the corresponding graphs for the mappings $\lambda \mapsto h^{\gamma_{21}}(\lambda)$, $\lambda \mapsto h^{\rho_{12}}(\lambda)$, and $\lambda \mapsto h^{\rho_{21}}(\lambda)$, respectively. The value at which each graph crosses the λ-axis is the resulting estimate of the inverse of the associated cross-information coefficient. To be able to evaluate the results, we plotted, in each picture, a vertical black line at the corresponding theoretical value, namely at $1/\gamma_{12}(f, g)$, $1/\gamma_{21}(f, g)$, $1/\rho_{12}(f, g)$ and $1/\rho_{21}(f, g)$. Clearly, the results are excellent, and there does not seem to be much dependence on the preliminary estimator $\tilde{L}$ used.

APPENDIX A: RANK-BASED EFFICIENT CENTRAL SEQUENCES

In this first Appendix, we study the asymptotic behavior of the rank-based efficient central sequences $\underline{\Delta}^{*}_{\vartheta, f;2}$. The main result is the following (see Appendix B for a proof).

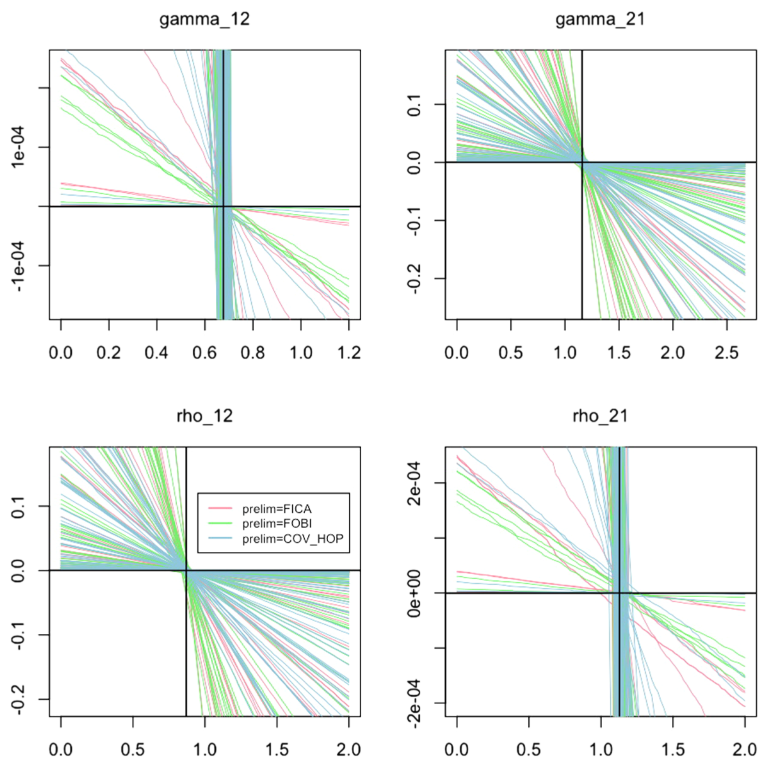

FIG. 3. *Top left*: 150 *graphs of the mapping* $\lambda \mapsto h^{\gamma_{12}}(\lambda)$ *based on* $f = f^{(3)}$, *associated with the first* 50 *replications from Setup* 1 ($g = g^{(1)}$) *in Figure* 1 (*sample size is* $n = 4{,}000$): *the* 50 *curves in pink, green, and blue are based on the preliminary estimators* $\hat{L}_{\mathrm{FICA}}$, $\hat{L}_{\mathrm{FOBI}}$ *and* $\hat{L}_{\mathrm{COV\_HOP}}$, *respectively. Top right, bottom left, and bottom right*: *the corresponding plots for the mappings* $\lambda \mapsto h^{\gamma_{21}}(\lambda)$, $\lambda \mapsto h^{\rho_{12}}(\lambda)$ *and* $\lambda \mapsto h^{\rho_{21}}(\lambda)$, *respectively.*

THEOREM A.1. *Fix* $\vartheta = (\mu', (\mathrm{vecd}^\circ L)')' \in \Theta$ *and* $f \in \underline{\mathcal{F}}_{\mathrm{ulan}}$. *Then*, (i) *for any* $g \in \mathcal{F}$,

$$\underline{\Delta}^*_{\vartheta,f;2} = \Delta^*_{\vartheta,f,g;2} + o_{L^2}(1)$$

as $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\vartheta,g}$, *where* $\Delta^*_{\vartheta,f,g;2} := C(I_p \otimes L^{-1})' \,\mathrm{vec}[\mathrm{odiag}(\frac{1}{\sqrt{n}} \sum_{i=1}^n (S_i \odot \varphi_f(F_+^{-1}(G_+(|Z_i|))))(S_i \odot F_+^{-1}(G_+(|Z_i|)))')]$. (ii) *Under* $\mathrm{P}^{(n)}_{\vartheta+n^{-1/2}\tau,g}$, *with* $\tau = (\tau_1', \tau_2')' \in \mathbb{R}^p \times \mathbb{R}^{p(p-1)}$ *and* $g \in \mathcal{F}_{\mathrm{ulan}}$,

$$\underline{\Delta}^*_{\vartheta,f;2} \xrightarrow{\mathcal{L}} \mathcal{N}_{p(p-1)}(\Gamma^*_{L,f,g;2}\tau_2, \Gamma^*_{L,f;2})$$

as $n \to \infty$ (*for* $\tau = 0$, *the result only requires that* $g \in \mathcal{F}$). (iii) *Still with* $\tau = (\tau_1', \tau_2')' \in \mathbb{R}^p \times \mathbb{R}^{p(p-1)}$ *and* $g \in \mathcal{F}_{\mathrm{ulan}}$, $\underline{\Delta}^*_{\vartheta+n^{-1/2}\tau,f;2} - \underline{\Delta}^*_{\vartheta,f;2} = -\Gamma^*_{L,f,g;2}\tau_2 + o_{\mathrm{P}}(1)$ *as* $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\vartheta,g}$.

Both for hypothesis testing and point estimation, we had to replace in $\underline{\Delta}^*_{\vartheta,f;2}$ the parameter ϑ with some estimator ($\check{\vartheta}^{(n)}$, say). The asymptotic behavior of the resulting (so-called *aligned*) rank-based efficient central sequence $\underline{\Delta}^*_{\check{\vartheta}^{(n)},f;2}$ is given in the following result.

COROLLARY A.1. *Fix* $\vartheta = (\mu', (\mathrm{vecd}^\circ L)')' \in \Theta$, $f \in \underline{\mathcal{F}}_{\mathrm{ulan}}$, *and* $g \in \mathcal{F}_{\mathrm{ulan}}$. *Let* $\check{\vartheta} = \check{\vartheta}^{(n)} = (\check{\mu}', (\mathrm{vecd}^\circ \check{L})')'$ *be a locally asymptotically discrete sequence of random vectors satisfying* $n^{1/2}(\check{\vartheta} - \vartheta) = O_{\mathrm{P}}(1)$ *as* $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\vartheta,g}$. *Then* $\underline{\Delta}^*_{\check{\vartheta},f;2} - \underline{\Delta}^*_{\vartheta,f;2} = -\Gamma^*_{L,f,g;2} n^{1/2}\,\mathrm{vecd}^\circ(\check{L} - L) + o_{\mathrm{P}}(1)$, *still as* $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\vartheta,g}$.

Since the sequence of estimators $\check{\vartheta}^{(n)}$ is assumed to be locally asymptotically discrete [which means that the number of possible values of $\check{\vartheta}^{(n)}$ in balls with $O(n^{-1/2})$ radius centered at ϑ is bounded as $n \to \infty$], this result is a direct consequence of Theorem A.1(iii) and Lemma 4.4 from [17]. Local asymptotic discreteness is a concept that goes back to Le Cam and is quite standard in one-step estimation; see, for example, [2] or [17].

Of course, a sequence of estimators $\check{\vartheta}^{(n)}$ can always be discretized by replacing each component $(\check{\vartheta}^{(n)})_\ell$ with

$$(\check{\vartheta}^{(n)}_{\#})_\ell := (cn^{1/2})^{-1}\,\mathrm{sign}\big((\check{\vartheta}^{(n)})_\ell\big)\big\lceil cn^{1/2}\big|(\check{\vartheta}^{(n)})_\ell\big|\big\rceil, \qquad \ell = 1, \dots, p^2,$$

for some arbitrary constant $c > 0$. In practice, however, one can safely forget about such discretizations: irrespective of the accuracy of the computer used, the discretization constant c can always be chosen large enough to make discretization be irrelevant at the fixed sample size n_0 at hand—hence also at any $n > n_0$.

APPENDIX B: PROOFS

B.1. Proofs of Theorems 2.1 and A.1. The proofs of this section make use of the Hájek projection theorem for linear signed-rank statistics (see, e.g., [23], Chapter 3), which states that, if $Y_i = \mathrm{Sign}(Y_i)|Y_i|$, $i = 1, \dots, n$, are i.i.d. with (absolutely continuous) cdf G and if $K : (0, 1) \to \mathbb{R}$ is a continuous and square-integrable score function that can be written as the difference of two monotone increasing functions, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{Sign}(Y_i) K(G_+(|Y_i|))$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{Sign}(Y_i) K\left(\frac{R_i^+}{n+1}\right) + o_{L^2}(1) \tag{B.1}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{Sign}(Y_i) \mathrm{E}[K(G_+(|Y_i|)) | R_i^+] + o_{L^2}(1) \tag{B.2}$$

as $n \to \infty$, where G_+ stands for the common cdf of the $|Y_i|$'s and R_i^+ denotes the rank of $|Y_i|$ among $|Y_1|, \dots, |Y_n|$. The quantities in (B.1) and (B.2) are linear signed-rank quantities that are said to be based on *approximate* and *exact* scores, respectively.

In the rest of this section, we fix $\vartheta \in \Theta$, $f \in \underline{\mathcal{F}}_{\text{ulan}}$, and $g \in \mathcal{F}$. We write throughout Z_i, S_i, and R_i^+, for $Z_i(\vartheta)$, $S_i(\vartheta)$, and $R_i^+(\vartheta)$, respectively. We also write E_h instead of $\mathrm{E}^{(n)}_{\vartheta,h}$, with $h = f, g$. We then start with the proof of Theorem A.1(i).

Proof of Theorem A.1(i). Fix $r \neq s \in \{1, \dots, p\}$ and two score functions $K_a, K_b : (0,1) \to \mathbb{R}$ with the same properties as K above. Then, by using (i) $\mathrm{E}_g[S_{ir}] = 0$, (ii) the independence (under $\mathrm{P}^{(n)}_{\vartheta,g}$) between the S_{ir}'s and the $(R_{ir}, |Z_{ir}|)$'s, and (iii) the independence between the Z_{ir}'s and the Z_{is}'s, we obtain

$$\begin{aligned}
&\mathrm{E}_g\Bigg[\Bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^n S_{ir}S_{is}\Bigg(K_a(G_{+r}(|Z_{ir}|))K_b(G_{+s}(|Z_{is}|)) \\
&\qquad\qquad\qquad - K_a\bigg(\frac{R_{ir}^+}{n+1}\bigg)K_b\bigg(\frac{R_{is}^+}{n+1}\bigg)\Bigg)\Bigg)^2\Bigg] \\
&\quad = \frac{1}{n}\sum_{i=1}^n \mathrm{E}_g\Bigg[\Bigg(K_a(G_{+r}(|Z_{ir}|))K_b(G_{+s}(|Z_{is}|)) \\
&\qquad\qquad\qquad - K_a\bigg(\frac{R_{ir}^+}{n+1}\bigg)K_b\bigg(\frac{R_{is}^+}{n+1}\bigg)\Bigg)^2\Bigg] \\
&\quad \le 2\mathrm{E}_g\Bigg[\Bigg(K_a(G_{+r}(|Z_{ir}|)) - K_a\bigg(\frac{R_{ir}^+}{n+1}\bigg)\Bigg)^2\Bigg]\mathrm{E}_g[K_b^2(G_{+s}(|Z_{is}|))] \\
&\qquad + 2\mathrm{E}_g\bigg[K_a^2\bigg(\frac{R_{ir}^+}{n+1}\bigg)\bigg]\mathrm{E}_g\Bigg[\Bigg(K_b(G_{+s}(|Z_{is}|)) - K_b\bigg(\frac{R_{is}^+}{n+1}\bigg)\Bigg)^2\Bigg].
\end{aligned}$$

Consequently, the square integrability of K_a, K_b, and the convergence to zero of both $\mathrm{E}_g[(K_a(G_{+r}(|Z_{ir}|)) - K_a(\frac{R_{ir}^+}{n+1}))^2]$ and $\mathrm{E}_g[(K_b(G_{+r}(|Z_{is}|)) - K_b(\frac{R_{is}^+}{n+1}))^2]$ [which directly follows from (B.1)] entail

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n S_{ir}S_{is}K_a(G_{+r}(|Z_{ir}|))K_b(G_{+s}(|Z_{is}|)) \\
&\qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^n S_{ir}S_{is}K_a\bigg(\frac{R_{ir}^+}{n+1}\bigg)K_b\bigg(\frac{R_{is}^+}{n+1}\bigg) + o_{L^2}(1)
\end{aligned}$$

as $n \to \infty$, under $\mathrm{P}^{(n)}_{\vartheta,g}$. Theorem A.1(i) follows by taking $K_a = \varphi_{f_r} \circ F_{+r}^{-1}$ and $K_b = F_{+s}^{-1}$. $\square$

We go on with the proof of Theorem 2.1, for which it is important to note that, by proceeding as in the proof of Theorem A.1(i) but with (B.2) instead of (B.1), we further obtain that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{ir}S_{is}K_a(G_{+r}(|Z_{ir}|))K_b(G_{+s}(|Z_{is}|))\\
&\qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{ir}S_{is}K_a\left(\frac{R_{ir}^{+}}{n+1}\right)K_b\left(\frac{R_{is}^{+}}{n+1}\right) + o_{L^2}(1)\\
&\qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{ir}S_{is}\mathrm{E}_g[K_a(G_{+r}(|Z_{ir}|))|R_{+ir}]\\
&\qquad\qquad\qquad \times \mathrm{E}_g[K_b(G_{+s}(|Z_{is}|))|R_{+is}] + o_{L^2}(1),
\end{aligned}\tag{B.3}$$

still as $n \to \infty$ under $\mathrm{P}_{\vartheta,g}^{(n)}$.

PROOF OF THEOREM 2.1. It is sufficient to prove Theorem 2.1(i) only, since, as already mentioned at the end of Section 2.3, Theorem 2.1(ii) follows from (2.6) and Theorem 2.1(i). That is, we have to show that, for any $r, s \in \{1, \dots, p\}$,

$$\begin{aligned}
&\mathrm{E}_f\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\varphi_f(Z_i)Z_i' - I_p\big)_{rs}|S_1, \dots, S_n, R_1^{+}, \dots, R_n^{+}\right]\\
&\qquad = (\underline{T}_{\vartheta,f})_{rs} + o_{L^2}(1)
\end{aligned}\tag{B.4}$$

as $n \to \infty$, under $\mathrm{P}_{\vartheta,f}^{(n)}$. Now, the left-hand side of (B.4) rewrites

$$\begin{aligned}
&\mathrm{E}_f\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\varphi_f(Z_i)Z_i' - I_p\big)_{rs}|S_1, \dots, S_n, R_1^{+}, \dots, R_n^{+}\right]\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{E}_f[S_{ir}S_{is}\varphi_f(|Z_{ir}|)|Z_{is}| - \delta_{rs}|S_1, \dots, S_n, R_1^{+}, \dots, R_n^{+}]\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(S_{ir}S_{is}\mathrm{E}_f[\varphi_f(|Z_{ir}|)|Z_{is}||R_{1r}^{+}, \dots, R_{nr}^{+}, R_{1s}^{+}, \dots, R_{ns}^{+}] - \delta_{rs}\big).
\end{aligned}\tag{B.5}$$

For $r \neq s$, this yields

$$\begin{aligned}
&\mathrm{E}_f\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\varphi_f(Z_i)Z_i' - I_p\big)_{rs}|S_1, \dots, S_n, R_1^{+}, \dots, R_n^{+}\right]\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{ir}S_{is}\mathrm{E}_f[\varphi_f(|Z_{ir}|)|R_{1r}^{+}, \dots, R_{nr}^{+}]\mathrm{E}_f[|Z_{is}||R_{1s}^{+}, \dots, R_{ns}^{+}]
\end{aligned}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_{ir} S_{is} \varphi_{f_r}\biggl(F_{+r}^{-1}\biggl(\frac{R_{ir}^{+}}{n+1}\biggr)\biggr) F_{+r}^{-1}\biggl(\frac{R_{is}^{+}}{n+1}\biggr) + o_{L^2}(1)$$
$$= (\underline{T}_{\vartheta,f})_{rs} + o_{L^2}(1)$$

as $n \to \infty$, under $\mathrm{P}_{\vartheta,f}^{(n)}$, where we have used (B.3), still with $K_a = \varphi_{f_r} \circ F_{+r}^{-1}$ and $K_b = F_{+s}^{-1}$, but this time at $g = f$. This establishes (B.4) for $r \neq s$. As for $r = s$, (B.5) now entails [writing $K_{ab}(u) := \varphi_f(F_{+r}^{-1}(u)) \times F_{+r}^{-1}(u)$ for all u]

$$\mathrm{E}_f\biggl[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\varphi_f(Z_i) Z_i' - I_p)_{rs} | S_1, \ldots, S_n, R_1^{+}, \ldots, R_n^{+}\biggr]$$
$$= \biggl(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{E}_f[\varphi_f(|Z_{ir}|)|Z_{ir}| | R_{1r}^{+}, \ldots, R_{nr}^{+}]\biggr) - \sqrt{n}$$
$$= \mathrm{E}_f\biggl[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} K_{ab}(F_{+r}(|Z_{ir}|)) | R_{1r}^{+}, \ldots, R_{nr}^{+}\biggr] - \sqrt{n}$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} K_{ab}\biggl(\frac{R_i^{+}}{n+1}\biggr) - \sqrt{n} + o_{L^2}(1) \tag{B.6}$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} K_{ab}\biggl(\frac{i}{n+1}\biggr) - \sqrt{n} + o_{L^2}(1)$$
$$= \sqrt{n} \int_0^1 K_{ab}(u)\, du - \sqrt{n} + o_{L^2}(1) \tag{B.7}$$
$$= o_{L^2}(1), \tag{B.8}$$

still as $n \to \infty$, under $\mathrm{P}_{\vartheta,f}^{(n)}$, where (B.6), (B.7) and (B.8) follow from the Hájek projection theorem for linear *rank* (not *signed-rank*) statistics (see, e.g., [23], Chapter 2), the square-integrability of $K_{ab}(\cdot)$ (see the proof of Proposition 3.2(i) in [10]), and integration by parts, respectively. This further proves (B.4) for $r = s$, hence also the result. □

PROOF OF THEOREM A.1(ii) AND (iii). (ii) In view of Theorem A.1(i), it is sufficient to show that both asymptotic normality results hold for $\Delta_{\vartheta,f,g;2}^{*}$. The result under $\mathrm{P}_{\vartheta,g}^{(n)}$ then straightforwardly follows from the multivariate CLT. As for the result under local alternatives [which, just as the result in part (iii), requires that $g \in \mathcal{F}_{\mathrm{ulan}}$], it is obtained as usual, by establishing the joint normality under $\mathrm{P}_{\vartheta,g}^{(n)}$ of $\log(d\mathrm{P}_{\vartheta+n^{-1/2}\tau,g}^{(n)}/d\mathrm{P}_{\vartheta,g}^{(n)})$ and $\Delta_{\vartheta,f,g;2}^{*}$, then applying Le Cam's third lemma; the required joint normality follows from a routine application of the classical

Cramér–Wold device. (iii) The proof, that is long and tedious, is also a quite trivial adaptation of the proof of Proposition A.1 in [7]. We therefore omit it. □

B.2. Proof of Theorem 3.1. (i) Applying Corollary A.1, with $\check{\vartheta} := \hat{\vartheta}_0 = (\hat{\mu}', (\mathrm{vecd}^\circ L_0)')'$ and $\vartheta := \vartheta_0 = (\mu', (\mathrm{vecd}^\circ L_0)')'$, entails that $\underline{\Delta}^*_{\hat{\vartheta}_0,f;2} = \underline{\Delta}^*_{\vartheta_0,f;2} + o_{\mathrm{P}}(1)$ as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta_0,g}$. Consequently, we have that

$$\underline{Q}_f = (\mathrm{vec}\,\underline{\Delta}^*_{\vartheta_0,f;2})'(\Gamma^*_{L_0,f;2})^{-1}(\mathrm{vec}\,\underline{\Delta}^*_{\vartheta_0,f;2}) + o_{\mathrm{P}}(1), \tag{B.9}$$

still as $n \to \infty$, under $\mathrm{P}^{(n)}_{\vartheta_0,g}$—hence also under $\mathrm{P}^{(n)}_{\vartheta_0+n^{-1/2}\tau,g}$ (from contiguity). The result then follows from Theorem A.1(ii). (ii) It directly follows from (i) that, under the sequence of local alternatives $\mathrm{P}^{(n)}_{\vartheta_0+n^{-1/2}\tau,f}$, $\underline{\phi}^{(n)}_f$ has asymptotic power $1 - \Psi_{p(p-1)}(\chi^2_{p(p-1),1-\alpha}; \tau_2'\Gamma^*_{L_0,f;2}\tau_2)$. This establishes the result, since these local powers coincide with the semiparametrically optimal (at f) powers in (2.5).

B.3. Proofs of Lemma 4.1, Theorems 4.1 and 4.2.

PROOF OF THEOREM 4.1. (i) Fix $\vartheta \in \Theta$ and $g \in \mathcal{F}_{\mathrm{ulan}}$. From (4.1), the fact that $\hat{\Gamma}^*_{\tilde{L},f;2} - \Gamma^*_{L,f,g;2} = o_{\mathrm{P}}(1)$ as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta,g}$, and Corollary A.1, we obtain

$$\begin{aligned}
\sqrt{n}\,\mathrm{vecd}^\circ(\hat{\underline{L}}_f - L) &= \sqrt{n}\,\mathrm{vecd}^\circ(\tilde{L} - L) + (\hat{\Gamma}^*_{\tilde{L},f;2})^{-1}\underline{\Delta}^*_{\tilde{\vartheta},f;2}\\
&= \sqrt{n}\,\mathrm{vecd}^\circ(\tilde{L} - L) + (\Gamma^*_{L,f,g;2})^{-1}\underline{\Delta}^*_{\tilde{\vartheta},f;2} + o_{\mathrm{P}}(1)\\
&= (\Gamma^*_{L,f,g;2})^{-1}\underline{\Delta}^*_{\vartheta,f;2} + o_{\mathrm{P}}(1) \qquad \text{(B.10)}
\end{aligned}$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta,g}$. Consequently, Theorem A.1(i) and (ii) entails that, still as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta,g}$,

$$\begin{aligned}
&\sqrt{n}\,\mathrm{vecd}^\circ(\hat{\underline{L}}_f - L)\\
&\quad = (\Gamma^*_{L,f,g;2})^{-1}\Delta^*_{\vartheta,f,g;2} + o_{\mathrm{P}}(1) \qquad \text{(B.11)}\\
&\quad \xrightarrow{\mathcal{L}} \mathcal{N}_{p(p-1)}(0, (\Gamma^*_{L,f,g;2})^{-1}\Gamma^*_{L,f;2}(\Gamma^*_{L,f,g;2})^{-1\prime}). \qquad \text{(B.12)}
\end{aligned}$$

Now, by using the fact that $C'(\mathrm{vecd}^\circ H) = (\mathrm{vec}\, H)$ for any $p \times p$ matrix H with only zero diagonal entries, we have that $\sqrt{n}\,\mathrm{vec}(\hat{\underline{L}}_f - L) = \sqrt{n}C'\,\mathrm{vecd}^\circ(\hat{\underline{L}}_f - L)$, so that (4.2), (4.3) and (4.4) follow from (B.10), (B.11) and (B.12), respectively.

(ii) The asymptotic covariance matrix of $\sqrt{n}\,\mathrm{vecd}^\circ(\hat{\underline{L}}_f - L)$, under $\mathrm{P}^{(n)}_{\vartheta,f}$, reduces to $(\Gamma^*_{L,f;2})^{-1}$ [let $g = f$ in (B.12)], which establishes the result. □

To prove Theorem 4.2, we will need the following result.

LEMMA B.1. *Fix* $\vartheta = (\mu', (\mathrm{vecd}^\circ L)')' \in \Theta$ *and* $f, g \in \mathcal{F}_{\mathrm{ulan}}$. *Then*

$$(I_p \otimes L^{-1})C'(\Gamma^*_{L,f,g;2})^{-1}C(I_p \otimes L^{-1})'$$
$$= \sum_{r,s=1, r\neq s}^{p} \{\alpha_{rs}(f,g)\big(e_r e_r' \otimes (L_{rs}^2 e_r e_r' + e_s e_s' - L_{rs} e_r e_s' - L_{rs} e_s e_r')\big)$$
$$+ \beta_{rs}(f,g)\big(e_r e_s' \otimes (L_{rs}L_{sr} e_r e_s' - L_{rs} e_r e_r' - L_{sr} e_s e_s' + e_s e_r')\big)\},$$

where L_{rs} *denotes the entry* (r,s) *of* L.

PROOF OF THEOREM 4.2. By using again the fact that $C'(\mathrm{vecd}^\circ H) = (\mathrm{vec}\, H)$ for any $p \times p$ matrix H with only zero diagonal entries, and then Lemma B.1, we obtain

$$\begin{aligned}
&\mathrm{vec}(\underline{\hat{L}}_f - \tilde{L})\\
&\quad = C' \,\mathrm{vecd}^\circ(\underline{\hat{L}}_f - \tilde{L})\\
&\quad = \frac{1}{\sqrt{n}} C'(\hat{\Gamma}^*_{\tilde{L},f;2})^{-1} C (I_p \otimes \tilde{L}^{-1})' \,\mathrm{vec}\, \underline{T}_{\tilde{\vartheta},f}\\
&\quad = \frac{1}{\sqrt{n}} (I_p \otimes \tilde{L})\\
&\qquad \times \Bigg[\sum_{r,s=1, r\neq s}^{p} \{\hat{\alpha}_{rs}(f)\big(e_r e_r' \otimes (\tilde{L}_{rs}^2 e_r e_r' + e_s e_s' - \tilde{L}_{rs} e_r e_s' - \tilde{L}_{rs} e_s e_r')\big)\\
&\qquad\qquad + \hat{\beta}_{rs}(f)\big(e_r e_s' \otimes (\tilde{L}_{rs}\tilde{L}_{sr} e_r e_s' - \tilde{L}_{rs} e_r e_r'\\
&\qquad\qquad\qquad - \tilde{L}_{sr} e_s e_s' + e_s e_r')\big)\}\Bigg]\\
&\qquad \times \mathrm{vec}\, \underline{T}_{\tilde{\vartheta},f}.
\end{aligned}$$

Since all diagonal entries of $\underline{T}_{\tilde{\vartheta},f}$ are zeros, we have that

$$\begin{aligned}
&\mathrm{vec}(\underline{\hat{L}}_f - \tilde{L})\\
&\quad = \frac{1}{\sqrt{n}} (I_p \otimes \tilde{L})\\
&\qquad \times \Bigg[\sum_{r,s=1, r\neq s}^{p} \{\hat{\alpha}_{rs}(f)\big(e_r e_r' \otimes (e_s e_s' - \tilde{L}_{rs} e_r e_s')\big)\\
&\qquad\qquad + \hat{\beta}_{rs}(f)\big(e_r e_s' \otimes (e_s e_r' - \tilde{L}_{rs} e_r e_r')\big)\}\Bigg] \mathrm{vec}\, \underline{T}_{\tilde{\vartheta},f}.
\end{aligned} \tag{B.13}$$

The identity $(C' \otimes A)(\mathrm{vec}\, B) = \mathrm{vec}(ABC)$ then yields

$$\mathrm{vec}(\underline{\hat{L}}_f - \tilde{L}) = \frac{1}{\sqrt{n}}(I_p \otimes \tilde{L})\,\mathrm{vec}\left[\sum_{r,s=1, r\neq s}^{p} (\hat{N}_f)_{sr}(e_s e_r' - \tilde{L}_{rs} e_r e_r')\right].$$

Hence, we have

$$\begin{aligned}
\underline{\hat{L}}_f - \tilde{L} &= \frac{1}{\sqrt{n}}\tilde{L} \sum_{r,s=1, r\neq s}^{p} (\hat{N}_f)_{sr}(e_s e_r' - \tilde{L}_{rs} e_r e_r') \\
&= \frac{1}{\sqrt{n}}\tilde{L} \sum_{r,s=1}^{p} (\hat{N}_f)_{sr}(e_s e_r' - \tilde{L}_{rs} e_r e_r') \\
&= \frac{1}{\sqrt{n}}\tilde{L}\left(N_f - \sum_{r,s=1}^{p} \tilde{L}_{rs}(\hat{N}_f)_{sr} e_r e_r'\right) \\
&= \frac{1}{\sqrt{n}}\tilde{L}\left(\hat{N}_f - \sum_{r=1}^{p} (\tilde{L} N_f)_{rr} e_r e_r'\right) \\
&= \frac{1}{\sqrt{n}}\tilde{L}\big(\hat{N}_f - \mathrm{diag}(\tilde{L} N_f)\big),
\end{aligned}$$

which proves the result. □

PROOF OF LEMMA 4.1. In this proof, all stochastic convergences are as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta,g}$. First note that, if $\check{\vartheta} := (\check{\mu}', (\mathrm{vecd}^\circ\, \check{L})')'$ is an arbitrary locally asymptotically discrete root-n consistent estimator for $\vartheta = (\mu', (\mathrm{vecd}^\circ\, L)')'$, we then have that

$$\begin{aligned}
\mathrm{vec}(\underline{T}_{\check{\vartheta},f} - \underline{T}_{\vartheta,f}) &= -G_{f,g}(I_p \otimes \check{L}^{-1})C'\sqrt{n}\,\mathrm{vecd}^\circ(\check{L} - L) \\
&\quad + o_{\mathrm{P}}(1)
\end{aligned} \tag{B.14}$$

(compare with Corollary A.1). Incidentally, note that (B.14) implies that $\mathrm{vec}\, \underline{T}_{\check{\vartheta},f}$ is $O_{\mathrm{P}}(1)$ [by proceeding exactly as in the proof of Theorem A.1(i) and (ii), we can indeed show that, under $\mathrm{P}^{(n)}_{\vartheta,g}$, $\mathrm{vec}\, \underline{T}_{\vartheta,f}$ is asymptotically multinormal, hence stochastically bounded].

Now, from (B.14), we obtain

$$\begin{aligned}
&\mathrm{vec}(\underline{T}_{\tilde{\vartheta}_\lambda^{\gamma_{rs}},f} - \underline{T}_{\tilde{\vartheta},f}) \\
&\quad = -G_{f,g}(I_p \otimes \tilde{L}^{-1})C'\sqrt{n}\,\mathrm{vecd}^\circ(\tilde{L}_\lambda^{\gamma_{rs}} - \tilde{L}) + o_{\mathrm{P}}(1) \\
&\quad = -\lambda(\underline{T}_{\tilde{\vartheta},f})_{rs} G_{f,g}(I_p \otimes \tilde{L}^{-1})C'\,\mathrm{vecd}^\circ\big(\tilde{L}e_r e_s' - \tilde{L}\,\mathrm{diag}(\tilde{L}e_r e_s')\big) + o_{\mathrm{P}}(1),
\end{aligned}$$

which, by using the fact that $C'(\mathrm{vecd}^\circ H) = (\mathrm{vec}\, H)$ for any $p \times p$ matrix H with only zero diagonal entries, leads to

$$\begin{aligned}
&\mathrm{vec}(\underline{T}_{\tilde{\vartheta}_\lambda^{\gamma_{rs}},f} - \underline{T}_{\tilde{\vartheta},f}) \\
&\quad = -\lambda(\underline{T}_{\tilde{\vartheta},f})_{rs} G_{f,g}(I_p \otimes \tilde{L}^{-1}) \,\mathrm{vec}\big(\tilde{L}e_r e_s' - \tilde{L}\,\mathrm{diag}(\tilde{L}e_r e_s')\big) + o_{\mathrm{P}}(1) \\
&\quad = -\lambda(\underline{T}_{\tilde{\vartheta},f})_{rs} G_{f,g} \,\mathrm{vec}\big(e_r e_s' - \mathrm{diag}(\tilde{L}e_r e_s')\big) + o_{\mathrm{P}}(1).
\end{aligned}$$

This yields

$$\begin{aligned}
&\mathrm{vec}(\underline{T}_{\tilde{\vartheta}_\lambda^{\gamma_{rs}},f} - \underline{T}_{\tilde{\vartheta},f}) \\
&\quad = -\lambda(\underline{T}_{\tilde{\vartheta},f})_{rs} G_{f,g} \,\mathrm{vec}(e_r e_s') + o_{\mathrm{P}}(1) \\
&\quad = -\lambda(\underline{T}_{\tilde{\vartheta},f})_{rs}\big(\gamma_{rs}(f,g)\,\mathrm{vec}(e_r e_s') + \rho_{rs}(f,g)\,\mathrm{vec}(e_s e_r')\big) \\
&\qquad + o_{\mathrm{P}}(1).
\end{aligned}$$

Premultiplying by $(\underline{T}_{\tilde{\vartheta},f})_{rs}(e_s \otimes e_r)'$, we then obtain

$$(\underline{T}_{\tilde{\vartheta},f})_{rs}(\underline{T}_{\tilde{\vartheta}_\lambda^{\gamma_{rs}},f})_{rs} - ((\underline{T}_{\tilde{\vartheta},f})_{rs})^2 = -\lambda((\underline{T}_{\tilde{\vartheta},f})_{rs})^2 \gamma_{rs}(f,g) + o_{\mathrm{P}}(1)$$

[recall indeed that $\underline{T}_{\tilde{\vartheta},f} = O_{\mathrm{P}}(1)$], which establishes the γ-part of the lemma. The proof of the ρ-part follows along the exact same lines, but for the fact that the premultiplication is by $(\underline{T}_{\tilde{\vartheta},f})_{sr}(e_r \otimes e_s)'$. $\square$

Acknowledgments. We would like to express our gratitude to the Co-Editor, Professor Peter Bühlmann, an Associate Editor and one referee. Their careful reading of a previous version of the paper and their comments and suggestions led to a considerable improvement of the present paper. We are also grateful to Klaus Nordhausen for sending to us the R code for FastICA authored by Abhijit Mandal.

SUPPLEMENTARY MATERIAL

Further results on tests and a proof of Theorem 4.3 (DOI: 10.1214/11-AOS906SUPP; .pdf). This supplement provides a simple explicit expression for the proposed test statistics, derives local asymptotic powers of the corresponding tests, and presents simulation results for hypothesis testing. It also gives a proof of Theorem 4.3.

REFERENCES

[1] AMARI, S. (2002). Independent component analysis and method of estimating functions. *IEICE Trans. Fundamentals Electronics, Communications and Computer Sciences* **E85-A** 540–547.

[2] BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671. MR0663424

[3] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Statistical Inference for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore.

[4] CARDOSO, J. F. (1989). Source separation using higher moments. In *Proceedings of IEEE International Conference on Acoustics*, *Speech and Signal Processing*, *Glasgow* 2109–2112.

[5] CASSART, D., HALLIN, M. and PAINDAVEINE, D. (2010). On the estimation of cross-information quantities in R-estimation. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis*: *A Festschrift in Honor of Professor Jana Jurečková* (J. Antoch, M. Hušková and P. K. Sen, eds.) 35–45. IMS, Beachwood, OH.

[6] CHEN, A. and BICKEL, P. J. (2006). Efficient independent component analysis. *Ann*. *Statist*. **34** 2825–2855. MR2329469

[7] HALLIN, M., OJA, H. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *Ann*. *Statist*. **34** 2757–2789. MR2329466

[8] 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

[9] HALLIN, M. and PAINDAVEINE, D. (2008). Semiparametrically efficient one-step R-estimation. Unpublished manuscript. Univ. Libre de Bruxelles.

[10] HALLIN, M., VERMANDELE, C. and WERKER, B. (2006). Serial and nonserial sign-and-rank statistics: Asymptotic representation and asymptotic normality. *Ann*. *Statist*. **34** 254–289. MR2275242

[11] HALLIN, M. and WERKER, B. J. M. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165. MR1963675

[12] HYVÄRINEN, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans*. *Neural Networks* **10** 626–634.

[13] HYVÄRINEN, A. and OJA, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Comput*. **9** 1483–1492.

[14] ILMONEN, P., NEVALAINEN, J. and OJA, H. (2010). Characteristics of multivariate distributions and the invariant coordinate system. *Statist*. *Probab. Lett*. **80** 1844–1853. MR2734250

[15] ILMONEN, P., NORDHAUSEN, K., OJA, H. and OLLILA, E. (2011). Independent component (IC) functionals and a new performance index. Unpublished manuscript. Univ. Tampere.

[16] ILMONEN, P. and PAINDAVEINE, D. (2011). Supplement to "Semiparametrically efficient inference based on signed ranks in symmetric independent component models." DOI:10.1214/11-AOS906SUPP.

[17] KREISS, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann*. *Statist*. **15** 112–133. MR0885727

[18] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411

[19] OJA, H., PAINDAVEINE, D. and TASKINEN, S. (2011). Parametric and nonparametric tests for multivariate independence in IC models. Unpublished manuscript. Univ. Libre de Bruxelles.

[20] OJA, H., SIRKIÄ, S. and ERIKSSON, J. (2006). Scatter matrices and independent component analysis. *Austrian J*. *Statist*. **35** 175–189.

[21] OLLILA, E. (2010). The deflation-based FastICA estimator: Statistical analysis revisited. *IEEE Trans*. *Signal Process*. **58** 1527–1541. MR2758026

[22] OLLILA, E. and KIM, H. J. (2011). On testing hypotheses of mixing vectors in the ICA model using FastICA. In *Proceedings of IEEE International Symposium on Biomedical Imaging* (*ISBI*'), *Chicago*, *IL* **11** 325–328.

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

[24] RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley, New York. MR0338013

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

AALTO UNIVERSITY SCHOOL OF ECONOMICS
QUANTITATIVE METHODS IN ECONOMICS
FI-00076 AALTO
FINLAND
E-MAIL: Pauliina.Ilmonen@gmail.com

E.C.A.R.E.S.
AND DÉPARTEMENT DE MATHÉMATIQUE
UNIVERSITÉ LIBRE DE BRUXELLES
50, AVENUE F.D. ROOSEVELT, CP114/04
B-1050 BRUSSELS
BELGIUM
E-MAIL: dpaindav@ulb.ac.be
URL: http://homepages.ulb.ac.be/~dpaindav