SUPPLEMENTARY MATERIAL

Further results on tests and a proof of Theorem 4.3
(doi: ???http://lib.stat.cmu.edu/aos/???/???; .pdf). This supplement provides a simple explicit expression for the proposed test statistics (Section 1), derives local asymptotic powers of the corresponding tests (Section 2), and presents simulation results for hypothesis testing (Section 3). It also gives a proof of Theorem 4.3 (Section 4).

Below, (M-3.1), Page M.9, Section M.3, Lemma M-4.1, etc. refer to Expression (3.1), Page 9, Section 3, Lemma 4.1, etc. from [2]. Unless otherwise stated, other cross-references relate to this supplement itself.

1. Explicit expressions of the proposed test statistics. The following result provides a simple and explicit expression of the signed-rank test statistic $\underset{\sim}{Q}_f$ in (M-3.1).

THEOREM 1.1. *Fix $f \in \underset{\sim}{\mathcal{F}}_{\mathrm{ulan}}$. Then the test statistic $\underset{\sim}{Q}_f$ rewrites*

$$\begin{aligned} \underset{\sim}{Q}_f &= (\mathrm{vec}\, \underset{\sim}{T}_{\hat{\vartheta}_0,f})' M_f (\mathrm{vec}\, \underset{\sim}{T}_{\hat{\vartheta}_0,f}) \\ &= \sum_{r,s=1, r\neq s}^{p} \big(\alpha_{rs}(f) (\underset{\sim}{T}_{\hat{\vartheta}_0,f})^2_{sr} + \beta_{rs}(f) (\underset{\sim}{T}_{\hat{\vartheta}_0,f})_{rs} (\underset{\sim}{T}_{\hat{\vartheta}_0,f})_{sr} \big), \qquad (1.1) \end{aligned}$$

where we let $\alpha_{rs}(f) = \alpha_{rs}(f,f)$, and $\beta_{rs}(f) = \beta_{rs}(f,f)$ (see (M-4.5)) and $M_f := \sum_{r,s=1,r\neq s}^{p} (\alpha_{rs}(f)(e_r e_r' \otimes e_s e_s') + \beta_{rs}(f)(e_r e_s' \otimes e_s e_r')$.

PROOF. Applying to $\underset{\sim}{Q}_f$ Lemma M-B.1 with $g = f$, we obtain

$$\begin{aligned} \underset{\sim}{Q}_f &= (\mathrm{vec}\, \underset{\sim}{T}_{\hat{\vartheta}_0,f})' \\ &\times \Bigg[\sum_{r,s=1, r\neq s}^{p} \Big\{ \alpha_{rs}(f) \big(e_r e_r' \otimes \big(L^2_{0rs} e_r e_r' + e_s e_s' - L_{0rs} e_r e_s' - L_{0rs} e_s e_r' \big)\big) \\ &+ \beta_{rs}(f) \big(e_r e_s' \otimes \big(L_{0rs} L_{0sr} e_r e_s' - L_{0rs} e_r e_r' - L_{0sr} e_s e_s' + e_s e_r' \big)\big) \Big\} \Bigg] (\mathrm{vec}\, \underset{\sim}{T}_{\hat{\vartheta}_0,f}), \end{aligned}$$

which, as all diagonal entries of $\underset{\sim}{T}_{\hat{\vartheta}_0,f}$ are equal to zero, indeed yields $\underset{\sim}{Q}_f = (\mathrm{vec}\, \underset{\sim}{T}_{\hat{\vartheta}_0,f})' M_f (\mathrm{vec}\, \underset{\sim}{T}_{\hat{\vartheta}_0,f})$. The equality (1.1) then easily follows from the identity $(C' \otimes A)(\mathrm{vec}\, B) = \mathrm{vec}(ABC)$. ∎

2. Local asymptotic powers. Theorem M.3.1 allows to compute the asymptotic powers of $\underline{\phi}_f$ under sequences of local alternatives of the form $\mathrm{P}^{(n)}_{\mu,L_0+n^{-1/2}H,g}$, where H is an arbitrary $p \times p$ matrix with zero diagonal entries (only such a H provides a perturbed mixing matrix $L_0 + n^{-1/2}H$ that belongs—for n large enough—to the parameter space $\mathcal{M}_{1p}$). The corresponding asymptotic powers are given by

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

where $\Psi_{p(p-1)}(\,\cdot\,;\delta)$ and $\chi^2_{p(p-1),1-\alpha}$ were defined in Page M.7. By using the fact that $C'(\mathrm{vecd}^\circ H) = (\mathrm{vec}\, H)$ and then applying Lemma M-B.1, the noncentrality parameter above, after painful yet straightforward computations, simplifies to

$$\sum_{r,s=1,r\neq s}^{p} \big(\,\xi_{rs}(f,g)\,((L_0^{-1}H)_{sr})^2 + \eta_{rs}(f,g)\,(L_0^{-1}H)_{rs}(L_0^{-1}H)_{sr}\big), \tag{2.1}$$

with

$$\xi_{rs}(f,g) = \frac{\gamma_{rs}(f)\gamma^2_{sr}(f,g) + \rho^2_{rs}(f,g)\gamma_{sr}(f) - 2\rho_{rs}(f,g)\gamma_{sr}(f,g)}{\gamma_{rs}(f)\gamma_{sr}(f) - 1}$$

and

$$\eta_{rs}(f,g) = \frac{\rho_{sr}(f,g)(\gamma_{rs}(f)\gamma_{sr}(f,g) - \rho_{rs}(f,g))}{\gamma_{rs}(f)\gamma_{sr}(f) - 1} + \frac{\gamma_{rs}(f,g)(\gamma_{sr}(f)\rho_{rs}(f,g) - \gamma_{sr}(f,g))}{\gamma_{rs}(f)\gamma_{sr}(f) - 1}.$$

At $g = f$, this reduces to $\sum_{r,s=1,r\neq s}^{p}(\gamma_{sr}(f)\,((L_0^{-1}H)_{sr})^2 + (L_0^{-1}H)_{rs}(L_0^{-1}H)_{sr})$. In the simulations of the next section, we will compare the ranking of finite-sample rejection frequencies associated with various tests $\underline{\phi}_f$ with the corresponding theoretical ranking derived from (2.1).

3. Simulations for hypothesis testing. We considered the trivariate case $p = 3$ and concentrated on the particular case for which the null value of L is $L_0 = I_3$. For three trivariate densities of the form $z \mapsto g(z) = g^{(d)}(z) = \prod_{r=1}^{3} g_r^{(d)}(z_r)$, $d \in \{1, 2, 3\}$, we generated $M = 5,000$ independent random samples $Z_i^{(d,m)} = (Z_{i1}^{(d,m)}, Z_{i2}^{(d,m)}, Z_{i3}^{(d,m)})'$, $i = 1, \ldots, n$, $m = 1, \ldots, M$, of size $n = 500$. The pdfs $g^{(d)}$ have the following marginals:

(i) In Setup $d = 1$, $g_1^{(d)}$, $g_2^{(d)}$ and $g_3^{(d)}$ are the pdfs of the standard normal distribution ($\mathcal{N}$), the Student distribution with 6 degrees of freedom (t_6), and the beta distribution with parameters 3 and 3 ($\beta_{3,3}$), respectively;

(ii) In Setup $d = 2$, $g_1^{(d)}$ is t_6, $g_2^{(d)}$ is $\beta_{3,3}$, and $g_3^{(d)}$ is the pdf of the double-exponential distribution with scale parameter one (d-exp);

(iii) In Setup $d = 3$, $g_1^{(d)}$ is t_6, $g_2^{(d)}$ is d-exp, and $g_3^{(d)}$ is the pdf of the logistic distribution with scale parameter one (log).

We then generated samples of n observations $X_1, \ldots, X_n$ according to

$$X_i^{(d,m)} = (L_0 + a\,\kappa^{(d)} H)\, Z_i^{(d,m)} + \mu, \tag{3.1}$$

with $a = 0, 1, 2, 3, 4$,

$$\begin{pmatrix} \kappa^{(1)} \\ \kappa^{(2)} \\ \kappa^{(3)} \end{pmatrix} = \begin{pmatrix} .002 \\ .007 \\ .0025 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 2 & 0 \end{pmatrix}, \quad \text{and} \quad \mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Clearly, these samples correspond to the null hypothesis for $a = 0$ and to increasingly severe alternatives for $a = 1, 2, 3, 4$. The quantities $\kappa^{(d)}$ were chosen in such a way that the rejection frequencies obtained for $a = 4$ were approximately .95 for all d. All samples were subjected, at asymptotic level $\alpha = 5\%$, to the signed-rank tests $\underline{\phi}_{f^{(j)}}$, $j = 1, 2, 3, 4$, where $f^{(j)} = g^{(j)}$ for $j = 1, 2, 3$, and where $f^{(4)}$ uses a t_3 pdf for each marginal density. The first three tests therefore achieve asymptotic optimality in Setups 1 to 3, respectively. In all tests, the location estimate $\hat{\mu}$ used is the componentwise median defined in Page M-9.

Rejection frequencies are plotted against a in the first column of Figure 1. These rejection frequencies indicate that, when based on their asymptotic chi-square critical values, the signed-rank tests are conservative and significantly biased at the sample size considered. In order to remedy this, we also implemented versions of each of the signed-rank procedures based on estimations of the (distribution-free) quantile of the test statistic under known parameter values μ and L_0. These estimations, just as the asymptotic chi-square quantile, are consistent approximations of the corresponding exact quantiles under the null, and were obtained, for each of the four tests above, as the empirical 0.05-upper quantiles $q_{.95}^{(n)}$ of each signed-rank test statistic in a collection of 10^6 simulated multinormal samples, yielding $q_{.95}^{(n)} = 10.34$, 11.56, 10.88, and 9.74, respectively. These bias-corrected critical values are all smaller than the asymptotic chi-square one $\chi^2_{6;.95} = 12.60$, so that the resulting tests are uniformly less conservative than the original ones. The resulting rejection frequencies are plotted in the second column of Figure 1, where it is readily seen that all tests now are roughly unbiased.

At the sample size $n = 500$, the asymptotic properties derived in Section M.3 do not show so clearly in the simulation results, not only because the signed-rank tests are biased, but also because the test $\underline{\phi}_{f^{(d)}}$ does not seem to be the most powerful one in Setup d. To question correctness of our asymptotic results, we reran the same simulation as above, but now with $n = 10,000$ and with $(\kappa^{(1)}, \kappa^{(2)}, \kappa^{(3)})'$ divided by $\sqrt{10,000/500}$. The resulting simulated critical values are given by $q^{(n)}_{.95} = 11.59$, 12.38, 11.83, and 11.46, respectively, and are all much closer to the asymptotic one $\chi^2_{6;.95} = 12.60$, so that the signed-rank tests, in their asymptotic versions, may only suffer a small bias for this large sample size. Consequently, it is justified to restrict to these asymptotic versions. The corresponding rejection frequencies are plotted in the last column of Figure 1 and confirm, under any $g^{(d)}$, $d = 1, 2, 3$, both the optimality of $\underline{\phi}_{f^{(d)}}$ and—more generally—the whole ranking of the local asymptotic powers of $\underline{\phi}_{f^{(j)}}$, $j = 1, 2, 3, 4$, which can be obtained from (2.1).

Finally, we point out that, for each fixed sample size, setup, and type of critical values considered, the various signed-rank tests exhibit very similar performances. This implies that, just as for point estimation, one should not worry too much about the choice of the target density f in hypothesis testing.

4. Proof of Theorem M-4.3. The proof follows the same scheme as that of Proposition 2.1 in [1]. We report the proof here for the sake of completeness.

Proof. We fix $\vartheta \in \Theta$, $f \in \underline{\mathcal{F}}_{\text{ulan}}$, $g \in \mathcal{F}_{\text{ulan}}$, and $r \neq s \in \{1, \ldots, p\}$, and concentrate on establishing that $\hat{\gamma}_{rs}(f) = \gamma_{rs}(f, g) + o_{\mathrm{P}}(1)$, as $n \to \infty$ under $\mathrm{P}^{(n)}_{\vartheta,g}$ (the proof of the ρ-result is entirely similar). In the sequel, we stress the dependence in n of the various statistics with superscripts ${}^{(n)}$.

Let us first show that, under $\mathrm{P}^{(n)}_{\vartheta,g}$, $\lambda^{(n)-}_{\gamma_{rs}}$, hence also $\lambda^{(n)+}_{\gamma_{rs}}$, is $O_{\mathrm{P}}(1)$ as $n \to \infty$. Assume therefore it is not: then, there exist $\epsilon > 0$ and a sequence $n_i \nearrow \infty$ such that, for all $\ell \in \mathbb{R}$ and i, $\mathrm{P}^{(n_i)}_{\vartheta,g}[\lambda^{(n_i)-}_{\gamma_{rs}} > \ell] > \epsilon$. This implies, for arbitrarily large ℓ, that $\mathrm{P}^{(n_i)}_{\vartheta,g}[h^{(n_i)\gamma_{rs}}(\ell) > 0] > \epsilon$, hence, in view of Lemma M-4.1,

$$\mathrm{P}^{(n_i)}_{\vartheta,g}\big[(1 - \ell\gamma_{rs}(f, g))\, h^{(n_i)\gamma_{rs}}(0) + \zeta^{(n_i)} > 0\big] > \epsilon$$

for all i, where $\zeta^{(n)}$, $n \in \mathbb{N}$ is some $o_{\mathrm{P}}(1)$ sequence. For $\ell > (\gamma_{rs}(f, g))^{-1}$,

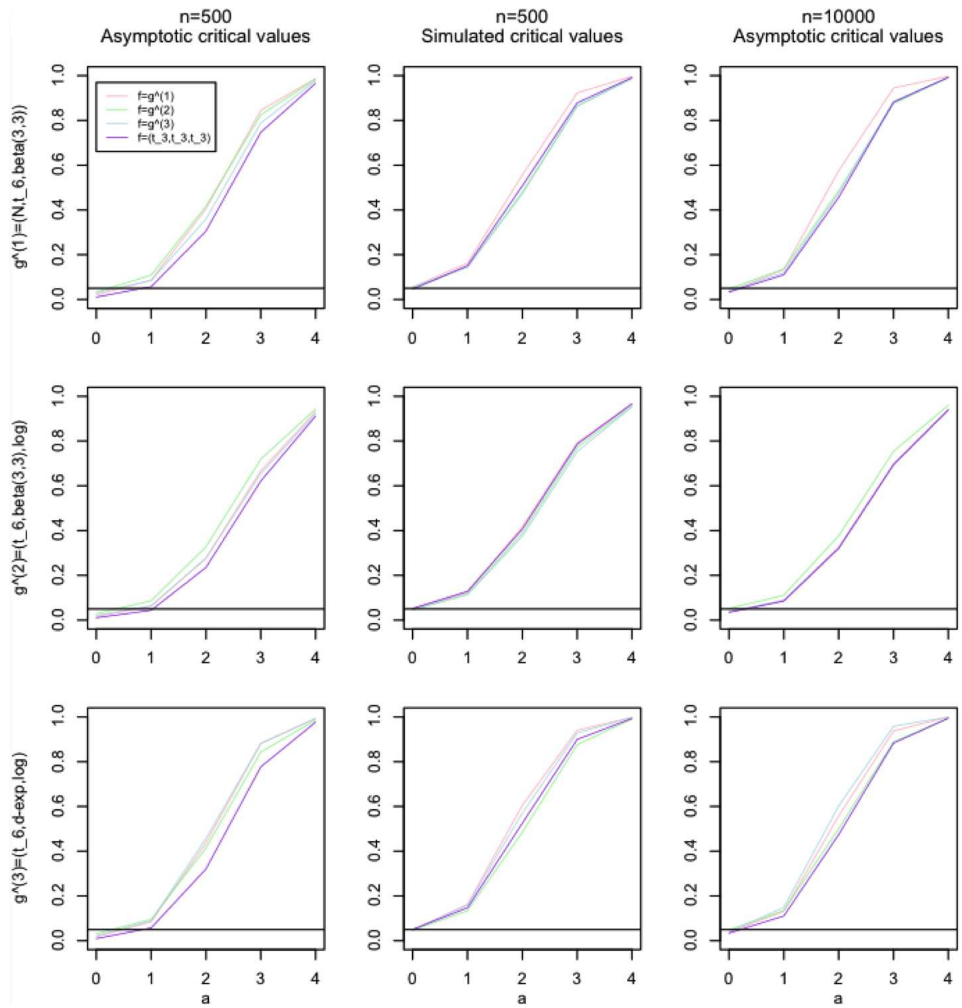

FIG 1. *Rejection frequencies (out of $M = 5,000$ replications), under the null ($a = 0$) and increasingly severe alternatives ($a = 1, 2, 3, 4$), of the signed-rank tests $\underline{\phi}_{f^{(j)}}$, $j = 1, 2, 3, 4$; see Section 3 for details. The sample size is $n = 500$ in both first columns and $n = 10,000$ in the third one. In the first and third columns, tests are based on their asymptotic null distribution, whereas the second column uses simulated critical values, obtained from 10^6 standard multinormal samples.*

this entails, for all i,

$$\mathrm{P}^{(n_i)}_{\vartheta,g}\big[0 < h^{(n_i)\gamma_{rs}}(0) < (\ell\gamma_{rs}(f,g)-1)^{-1}|\zeta^{(n_i)}|\big] > \epsilon,$$

which contradicts (M-4.9). It follows that $\lambda^{(n)-}_{\gamma_{rs}}$ is $O_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\vartheta,g}$.

By using again (M-4.9), there exist, for all $\eta > 0$, a positive real number δ_η and an integer N_η such that

$$\mathrm{P}^{(n)}_{\vartheta,g}\big[h^{(n)\gamma_{rs}}(0) \geq \delta_\eta\big] \geq 1 - \frac{\eta}{2}$$

for all $n \geq N_\eta$. Since $\lambda^{(n)-}_{\gamma_{rs}}$ and $\lambda^{(n)+}_{\gamma_{rs}}$ are $O_{\mathrm{P}}(1)$, Lemma M-4.1 implies that, for all $\eta > 0$ and $\varepsilon > 0$, there exists an integer $N_{\varepsilon,\delta} \geq N_\eta$ such that, for all $n \geq N_{\varepsilon,\delta}$ (with $\lambda^{(n)\pm}_{\gamma_{rs}}$ standing for either $\lambda^{(n)-}_{\gamma_{rs}}$ or $\lambda^{(n)+}_{\gamma_{rs}}$),

$$\mathrm{P}^{(n)}_{\vartheta,g}\big[(1-\lambda^{(n)\pm}_{\gamma_{rs}}\gamma_{rs}(f,g))h^{(n)\gamma_{rs}}(0) \in [h^{(n)\gamma_{rs}}(\lambda^{(n)\pm}_{\gamma_{rs}}) \pm \varepsilon]\big] \geq 1 - \frac{\eta}{2}.$$

It follows that for all $\eta > 0$, $\varepsilon > 0$ and $n \geq N_{\varepsilon,\delta}$, letting $\delta = \delta_\eta$,

$$\begin{aligned}\mathrm{P}^{(n)}_{\vartheta,g}\big[A^{(n)}_{\varepsilon,\delta}\big] := \mathrm{P}^{(n)}_{\vartheta,g}\Big[&(1-\lambda^{(n)\pm}_{\gamma_{rs}}\gamma_{rs}(f,g))h^{(n)\gamma_{rs}}(0) \in [h^{(n)\gamma_{rs}}(\lambda^{(n)\pm}_{\gamma_{rs}}) \pm \varepsilon] \\ &\text{and } h^{(n)\gamma_{rs}}(0) \geq \delta\Big] \\ \geq 1 - \eta.&\end{aligned}$$

Next, denote by $\hat{D}^{(n)}$, $D^{(n)}$, and $D^{(n)}_{\pm}$ the graphs of the mappings

$$\begin{aligned}\lambda &\mapsto h^{(n)\gamma_{rs}}(\lambda^{(n)-}_{\gamma_{rs}}) - c(\lambda - \lambda^{(n)-}_{\gamma_{rs}})(h^{(n)\gamma_{rs}}(\lambda^{(n)-}_{\gamma_{rs}}) - h^{(n)\gamma_{rs}}(\lambda^{(n)+}_{\gamma_{rs}})) \\ \lambda &\mapsto (1-\lambda\gamma_{rs}(f,g))\, h^{(n)\gamma_{rs}}(0),\end{aligned}$$

and

$$\lambda \mapsto (1-\lambda\gamma_{rs}(f,g))\, h^{(n)\gamma_{rs}}(0) \pm \epsilon,$$

respectively. These graphs take the form of four random straight lines, intersecting the horizontal axis at $\lambda^{(n)}_{\gamma_{rs}}$ (our estimator of $(\gamma_{rs}(f,g))^{-1}$), $\lambda_0 := (\gamma_{rs}(f,g))^{-1}$, $\lambda_0^{(n)+}$ and $\lambda_0^{(n)-}$, respectively. Since $D^{(n)}_{\pm}$ and $D^{(n)}$ are parallel, with a negative slope, we have that $\lambda_0^{(n)-} \leq \lambda_0 \leq \lambda_0^{(n)+}$. Under $A^{(n)}_{\varepsilon,\delta}$, that common slope has absolute value at least $\delta\gamma_{rs}(f,g)$, which implies that $\lambda_0^{(n)+} - \lambda_0^{(n)-} \leq \frac{2\varepsilon}{\delta\gamma_{rs}(f,g)}$. Still under $A^{(n)}_{\varepsilon,\delta}$, for λ values between $\lambda^{(n)-}_{\gamma_{rs}}$

and $\lambda_{\gamma_{rs}}^{(n)+}$, $\hat{D}^{(n)}$ is lying between $D_{-}^{(n)}$ and $D_{+}^{(n)}$, which entails $\lambda_0^{(n)-} \leq \lambda_{\gamma_{rs}}^{(n)} \leq \lambda_0^{(n)+}$.

Summing up, for all $\eta > 0$ and $\varepsilon > 0$, there exist $\delta = \delta_\eta > 0$, and $N = N_{\varepsilon\gamma_{rs}(f,g)\delta/2,\delta}$ such that, for any $n \geq N$, with $\mathrm{P}_{\vartheta,g}^{(n)}$ probability larger than $1 - \eta$, $|\lambda_{\gamma_{rs}}^{(n)} - \lambda_0| \leq \lambda_0^{(n)+} - \lambda_0^{(n)-} \leq \varepsilon$. □

References.

[1] Cassart, D., Hallin, M., and Paindaveine, D. (2010). On the estimation of cross-information quantities in R-estimation. In J. Antoch, M. Hušková and P.K. Sen, Editors: *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková*, I.M.S. Monographs-Lecture Notes, 35–45.

[2] Ilmonen, P., and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models.