Efficient Fully Distribution-Free Center-Outward Rank Tests for Multiple-Output Regression and MANOVA

Marc Hallin
ECARES and Département de Mathématique
Université libre de Bruxelles, Belgium

Daniel Hlubinka
Faculty of Mathematics and Physics
Charles University, Prague, Czech Republic

Sarka Hudecova
Faculty of Mathematics and Physics
Charles University, Prague, Czech Republic

July 2021

ECARES working paper 2021-13

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Efficient Fully Distribution-Free Center-Outward Rank Tests for Multiple-Output Regression and MANOVA

Marc Hallin
ECARES and Département de Mathématique
Université libre de Bruxelles, Brussels, Belgium

Daniel Hlubinka and Šárka Hudecová
Faculty of Mathematics and Physics
Charles University, Prague, Czech Republic

Abstract

Extending rank-based inference to a multivariate setting such as multiple-output regression or MANOVA with unspecified d-dimensional error density has remained an open problem for more than half a century. None of the many solutions proposed so far is enjoying the combination of distribution-freeness and efficiency that makes rank-based inference a successful tool in the univariate setting. A concept of *center-outward* multivariate ranks and signs based on measure transportation ideas has been introduced recently. Center-outward ranks and signs are not only distribution-free but achieve in dimension $d > 1$ the (essential) maximal ancillarity property of traditional univariate ranks. In the present case, we show that fully distribution-free testing procedures based on center-outward ranks can achieve parametric efficiency. We establish the Hájek representation and asymptotic normality results required in the construction of such tests in multiple-output regression and MANOVA models. Simulations and an empirical study demonstrate the excellent performance of the proposed procedures.

Keywords: Distribution-free tests; Multivariate ranks; Multivariate signs; Hájek representation.

1 Introduction

Linear models—regression (single- and multiple-output), Analysis of Variance (ANOVA and MANOVA)—are probably the most popular and most useful of all statistical models; they are found in the table of contents of all statistical textbooks and statistical softwares, and are part of daily statistical practice in all domains of application. The pseudo-Gaussian approach—Gaussian quasi maximum likelihood estimation and pseudo-Gaussian F tests—is largely dominant in that context on the ground that pseudo-Gaussian methods remain asymptotically valid under a broad class of non-Gaussian densities satisfying mild moment conditions. One should beware of excessive confidence in such asymptotics, though.

1.1 Pseudo-Gaussian tests

Let us concentrate on hypothesis testing. The problem with pseudo-Gaussian tests under unspecified noise density is twofold:

(a) although pseudo-Gaussian tests are asymptotically valid under a broad range of non-Gaussian densities, that asymptotic validity is far from uniform: actually, in a semiparametric model with parameter θ where the underlying noise has unspecified density f in some broad class $\mathcal{F}$ of densities, a sequence $\varphi^{(n)}$ of tests of the null hypothesis $\theta = \theta_0$ has asymptotic level α iff $\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \mathrm{E}_{\theta_0,f}[\varphi^{(n)}] \leq \alpha$, whereas pseudo-Gaussian tests $\varphi_{\mathcal{G}}^{(n)}$ only satisfy the pointwise condition $\lim_{n\to\infty} \mathrm{E}_{\theta_0,f}[\varphi_{\mathcal{G}}^{(n)}] \leq \alpha$ for all $f \in \mathcal{F}$;

(b) still for fixed n, the performance of pseudo-Gaussian tests may rapidly deteriorate away from the Gaussian.

Appendix A.1 illustrates these pitfalls in the case of Hotelling's bivariate two-sample test.

1.2 Rank-based tests

A natural way to restore uniform asymptotics, thereby solving the validity problem in (a) consists in resorting to distribution-free tests, and this is how rank tests enter the picture. Rank-based testing methods have been quite successful in testing problems for single-ouput regression and linear models such as ANOVA (see the classical monographs by Hájek and Šidák (1967), Randles and Wolfe (1979) or Puri and Sen (1985)) and univariate linear time series (Hallin et al. (1985), Koul and Saleh (1993), Hallin and Puri (1994)). Being distribution-free, rank tests remain valid over the full class of absolutely continuous distributions. In linear models (this includes testing for single-output regression slopes, testing for treatment effects in analysis of variance, testing against location shifts in two-sample problems) and ARMA time series, they do reach parametric or semiparametric efficiency bounds at given reference densities, thus reconciling the conflicting objectives of robustness and efficiency. The celebrated Chernoff-Savage result (Chernoff and Savage (1958) and, for time-series, Hallin (1994)) moreover indicates that, far from losing power with respect to their pseudo-Gaussian counterparts, rank tests make the latter non-admissible under any non-Gaussian density f.

Extending these attractive features to a multivariate (multiple-output) context, of course, is highly desirable and the problem of defining multivariate concepts of ranks has been a long-standing open problem, for which many solutions have been proposed in the literature. Puri and Sen (1971) for a variety of problems in multivariate analysis (including multiple-output regression and MANOVA) and Hallin et al. (1989) for VARMA time series models construct tests based on it componentwise ranks which, however, fail to be distribution-free. Building upon an ingenious multivariate extension of the L_1 definition of quantiles, Oja (1999, 2010) defines the so-called *spatial ranks*; the resulting tests are neither distribution-free nor efficient. Tests based on the ranks of various concepts of statistical depth also have been proposed (Liu (1992), Liu and Singh (1993), He and Wang (1997), Zuo and He (2006)). While distribution-free, these ranks are failing to exploit any directional information, and hence typically do not allow for any type of asymptotic efficiency. As for the tests based on the *Mahalanobis ranks and signs* proposed by Hallin and Paindaveine (2002a,b, 2004, 2005), they do achieve, within the class of linear models and linear time series with elliptical densities, parametric or semiparametric efficiency at correctly specified elliptical reference densities; their distribution-freeness, hence their validity, unfortunately, is limited to the class of elliptical distributions.

Inspired by measure transportation ideas, a new concept of ranks and signs for multivariate observations has been introduced recently under the name of *Monge-Kantorovich ranks and signs* in Chernozhukov et al. (2017), under the name of *center-outward ranks and signs* in Hallin (2017) and Hallin et al. (2021a), along with the related population concepts of *center-outward distribution and quantile functions*. Unlike earlier concepts, these ranks and signs extend to dimension $d > 1$ the *essential maximal ancillarity* property (see Section 2.4 and Appendices D1 and D.2 of Hallin et al. (2021a)) of univariate ranks; the corresponding empirical center-outward distribution functions, moreover, satisfy a Glivenko-Cantelli result.

Center-outward ranks and signs have been successfully applied (Boeckel et al. (2018), Deb and Sen (2019), Ghosal and Sen (2019), Shi et al. (2021, 2020)) in the construction of distribution-free tests of independence between random vectors and multivariate goodness-of-fit; applications to the study of tail behavior and extremes can be found in De Valk and Segers (2018); Beirlant et al. (2020) are using the related center-outward empirical quantiles in the analysis of multivariate risk; Hallin et al. (2021b, 2020b) are proposing center-outward tests and R-estimators for VAR and VARMA time series models with unspecified innovation densities. The present paper goes one step further in the direction of a toolkit of distribution-free tests for multiple-output multivariate analysis by deriving a Hájek-type asymptotic representation result for linear center-outward rank statistics. Asymptotic normality follows as a corollary, from which center-outward rank tests are constructed for multiple-output regression models (including, as special cases, MANOVA and two-sample location models). Those tests are fully distribution-free, hence valid, over the entire family of absolutely continuous distributions; for adequate choice of the scores, parametric efficiency is attained at chosen densities. Since this paper was written (Hallin et al., 2020a), some further results (among them, partial Chernoff-Savage and Hodges-Lehmann properties) on the two-sample problem have been obtained by Deb et al. (2021).

1.3 A motivating example

The importance of center-outward rank tests in daily statistical practice is illustrated with the following real-life motivating example. The Wisconsin Diagnostic Breast Cancer (WDBC) data[1], first analyzed in Street et al. (1993) in a classification context, contains records on $n = 569$ patients from two groups—benign or malignant tumor diagnosis.

For each patient, several features were recorded from the digitized image of a fine needle aspirate of the breast mass, resulting in $d = 30$ variables, labeled V1–V30. The two groups of patients are well separated: the two-sample Hotelling test in dimension $d = 30$ very significantly rejects the null hypothesis of equal locations (R delivers a p-value 0.000, meaning that the actual p-value is less than 10^{-22}!). However, for some smaller subsets of variables (dimension three or four), the Hotelling test remains non-significant; the subset consisting of V12 (mean of fractal dimension), V14 (standard error of texture), V21 (standard error of symmetry), and V22 (standard error of fractal dimension) is an example. Figure 1 shows bivariate scatterplots and histograms for these four variables, revealing skewness in

[1]The dataset is available at Machine Learning Repository Dua and Graff (2017).

all univariate marginals and deviations from elliptical symmetry in bivariate marginals.

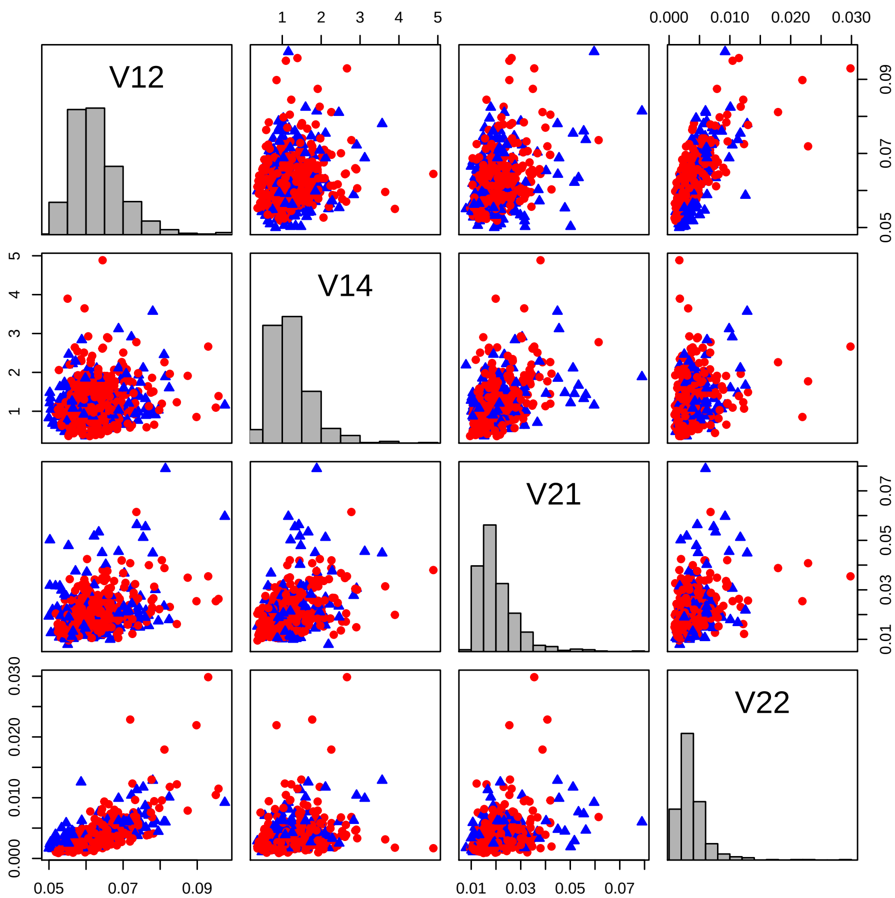

Figure 1: Wisconsin Diagnostic Breast Cancer (WDBC) data: bivariate scatterplots and univariate histograms for mean fractal dimension (V12), standard error of texture (V14), standard error of symmetry (V21), and standard error of fractal dimension (V22) in 212 malignant patients (triangles) and 357 benign patients (circles).

The Hotelling and Wilcoxon[2] center-outward rank tests have been performed for the corresponding four-dimensional dataset and all its three-dimensional marginals[3]; p-values are shown in Table 1. With p-value 0.0090, the Wilcoxon test in dimension 4 is significant at 5% and 1% levels, while Hotelling (p-value 0.0595) is not. Turning to dimension 3, Wilcoxon is always significant at 5% level (at 1% level in all cases but one), while Hotelling never rejects on 1%. The most spectacular case is that of the subset $\{V12, V14, V21\}$ where Wilcoxon and Hotelling yield p-values 0.0327 and 0.9899., respectively. Such discrepancies most likely originate in the skewness, non-ellipticity and/or the heavy tails of

[2]See Section 5.3.1 for a precise description.

[3]The center-outward ranks were computed for a 569-point random grid with $n_R = 20$, $n_S = 28$, $n_0 = 9$; the $n_S = 28$-points over the sphere were generated (seed `1111` in R Core Team (2021)) as in Section A.7.2.

Variables	(12,14,21, 22)	(12,14,21)	(12,14,22)	(12,21,22)	(14,21,22)
Hotelling	0.0595	0.9899	0.0299	0.0346	0.2136
c-o Wilcoxon	0.0090	0.0327	0.0007	0.0000	0.0018

Table 1: Wisconsin Diagnostic Breast Cancer (WDBC) data: p-values of the two-sample location Hotelling and center-outward Wilcoxon rank tests for the 4-dimensional marginal WDBC data corresponding to the set of variables {V12,V14,VV21,V22} and its 3-dimensional subsets.

the observations; their impact in terms of diagnostic power may have crucial consequences.

1.4 Outline of the paper

The paper is organized as follows. Section 2 briefly describes the main tools to be used: center-outward distribution and quantile functions (Section 2.1) and their empirical counterparts, the center-outward ranks and signs (Section 2.2). The main properties of these concepts are summarized in Section 2.3 (Proposition 2.1); their invariance/equivariance properties are established in Proposition 2.2. Section 3 is entirely devoted to the key theoretical results of this paper, which extend and generalize the classical approach by Hájek and Šidák (1967): a Hájek-type asymptotic representation for multivariate center-outward linear rank statistics and the resulting asymptotic normality result. Section 4.1 describes the multiple-output regression model to be considered throughout, which contains, as particular cases, the two-sample location and MANOVA models, of obvious practical importance. Local asymptotic normality is established in Section 4.2 for this model under general error densities (Proposition 4.1) and, for the purpose of future comparisons, for the particular case of elliptical distributions (Proposition 4.2). The center-outward rank tests we are proposing are described in Section 5.2, along with (Corollary 5.2) their local asymptotic optimality properties. Due to their importance in applications, the particular cases of the hypotheses of equal locations in the two-sample problem and no treatment effect in MANOVA are considered in Section 5.3. Sections 6.1 and 6.2 propose some simple choices of score functions, extending the classical median-test-score (based on center-outward signs only), Wilcoxon, and van der Waerden (normal-score) tests. Section 6.3 discusses affine invariance issues. Section 7 is devoted to a Monte Carlo exploration, in dimension $d = 2$, of the finite-sample performance of our rank tests which appear to outperform their competitors in non-elliptical situations while performing equally well under ellipticity. Section 7.3 presents an archaeological MANOVA application in dimension $d = 4$; while traditional MANOVA methods cannot reject the hypothesis of no treatment effect, our fully distribution-free center-outward rank-based test rejects it quite significantly, which might lead to revising some of the conclusions (Phelps et al., 2016) on Middle-East economic exchanges between Egypt and Syro-Palestine in the Byzantine-Islamic transition period. All proofs are concentrated in an online appendix where we also provide simulations in dimension $d = 6$.

2 Center-outward ranks and signs in $\mathbb{R}^d$

2.1 Center-outward distribution functions

Throughout, denote by $\mathbf{Z}^{(n)}$ a triangular array $(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)})$, $n \in \mathbb{N}$ of i.i.d. d-dimensional random vectors with distribution P in the family $\mathcal{P}_d$ of absolutely continuous distributions on $\mathbb{R}^d$. The notation $\overline{\text{spt}}(\text{P})$ is used for the support of P, spt(P) for its interior. The open (resp. closed) unit ball and the unit hypersphere in $\mathbb{R}^d$ are denoted by $\mathbb{S}_d$ (resp. $\overline{\mathbb{S}}_d$) and $\mathcal{S}_{d-1}$, respectively; U_d stands for the spherical[4] uniform distribution over $\mathbb{S}_d$, μ_d for the Lebesgue measure over $\mathbb{R}^d$; $\mathbf{I}_d$ is the $d \times d$ unit matrix, $\mathbf{1}_A$ the indicator of the Borel set A.

The definition of the center-outward distribution function of P is particularly simple for P in the so-called class $\mathcal{P}_d^+$ of distributions *with nonvanishing densities*—namely, the class of all distributions with density $f := \text{dP}/\text{d}\mu_d$ such that, for all $D \in \mathbb{R}^+$, there exist constants $\lambda^-_{D;\text{P}}$ and $\lambda^+_{D;\text{P}}$ satisfying $0 < \lambda^-_{D;\text{P}} \leq f(\mathbf{z}) \leq \lambda^+_{D;\text{P}} < \infty$ for all $\mathbf{z}$ with $\|\mathbf{z}\| \leq D$ (so that spt(P) $= \mathbb{R}^d$ and P-a.s. is equivalent to μ_d-a.e.). The main result in McCann (1995) then implies the existence of an a.e. unique convex lower semi-continuous function $\varphi : \mathbb{R}^d \to \mathbb{R}$ with gradient $\nabla\varphi$ such that[5] $\nabla\varphi\#\text{P} = \text{U}_d$. Call $\mathbf{F}_\pm := \nabla\varphi$ the *center-outward distribution function* of P. It follows from Figalli (2018) that $\mathbf{F}_\pm$ defines a homeomorphism between the punctured unit ball $\mathbb{S}_d \setminus \{\mathbf{0}\}$ and its image $\mathbb{R}^d \setminus \mathbf{F}_\pm^{-1}(\mathbf{0})$: call $\mathbf{Q}_\pm : \mathbf{u} \mapsto \mathbf{Q}_\pm(\mathbf{u}) := \mathbf{F}_\pm^{-1}(\mathbf{u})$, $\mathbf{u} \neq \mathbf{0}$ the *center-outward quantile function.* Figalli also shows that, defining $\mathbf{Q}_\pm(\mathbf{0}) := \mathbf{F}_\pm^{-1}(\mathbf{0})$ yields a convex and compact subset with Lebesgue measure zero in $\mathbb{R}^d$, the *center-outward median set of* P.

[4]Namely, the spherical distribution with uniform (over $[0, 1]$) radial density—equivalently, the product of a uniform over the distances to the origin and a uniform over the unit sphere $\mathcal{S}_{d-1}$. For $d = 1$, it coincides with the Lebesgue uniform; for $d \geq 2$, it has unbounded density at the origin.

[5]We borrow from measure transportation the convenient notation $T\#\text{P}$ ($T\colon \mathbb{R}^d \to \mathbb{R}^d$ *pushes* P *forward to* $T\#\text{P}$) for the distribution under $\mathbf{Z} \sim \text{P}$ of $T(\mathbf{Z})$.

All the intuition and all the properties of center-outward distribution and quantile functions hold for $\mathrm{P} \in \mathcal{P}_d^+$; this special case is the one considered in Hallin (2017). A more general case is addressed in del Barrio et al. (2020) and Hallin et al. (2021a) where we refer to for details, but requires more technical definitions, which we are skipping here. Note, however, that while some statements below only hold under $\mathrm{P} \in \mathcal{P}_d^+$, many others (including validity), due to distribution-freeness, can be made under the very general condition $\mathrm{P} \in \mathcal{P}_d$.

2.2 Center-outward ranks and signs

Except for a few particular cases such as spherical distributions, the above definitions are not meant for an analytical derivation of $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$ which typically involves Monge-Ampère equations;[6] estimation is possible, though, via their empirical counterparts $\mathbf{F}_\pm^{(n)}$ and $\mathbf{Q}_\pm^{(n)}$, based on center-outward ranks and signs, which we now describe.

Associated with the n-tuple $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$, the *empirical center-outward distribution function* $\mathbf{F}_\pm^{(n)}$ is mapping $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ to a "regular" grid $\mathfrak{G}_n$ of the unit ball $\mathbb{S}_d$. That grid $\mathfrak{G}_n$ is obtained as follows:

(a) first factorize n into $n = n_R n_S + n_0$, with $0 \leq n_0 < \min(n_R, n_S)$;

(b) next consider a "regular array" $\mathfrak{S}_{n_S} := \{\mathbf{s}_1^{n_S}, \ldots, \mathbf{s}_{n_S}^{n_S}\}$ of n_S points on the sphere $\mathcal{S}_{d-1}$ (see the comment below);

(c) finally, the grid consists in the collection $\mathfrak{G}_n$ of the $n_R n_S$ points $\mathfrak{g}$ of the form

$$\big(r/(n_R+1)\big)\mathbf{s}_s^{n_S}, \quad r = 1, \ldots, n_R, \quad s = 1, \ldots, n_S,$$

along with (n_0 copies of) the origin in case $n_0 \neq 0$: a total number $n - (n_0 - 1)$ or n of distinct points, thus, according as $n_0 > 0$ or $n_0 = 0$.

By "regular" we mean "as uniform as possible", in the sense, for example, of the *low-discrepancy sequences* of the type considered in numerical integration and Monte-Carlo methods (see, e.g., Niederreiter (1992), Judd (1998), or Santner et al. (2003)). The only mathematical requirement needed for Proposition 2.1 below is the weak convergence, as $n \to \infty$, of the uniform discrete distribution over $\mathfrak{G}_n$ to the uniform distribution over $\mathbb{S}_d$; all sequences $\mathfrak{G}_n$ satisfying that requirement yield the same asymptotic results. A uniform i.i.d. sample of points over $\mathbb{S}_d$, for example, satisfies the requirement but fails to produce mutually independent ranks and signs; moreover, one easily can construct arrays that

[6] In particular, no closed forms of $\mathbf{F}_\pm$ and $\mathbf{Q}_\pm$ are known for non-spherical elliptical distributions.

are “more regular” than an i.i.d. one. For instance, one could see that n_S or $n_S - 1$ of the points $\mathbf{s}_s^{n_S}$ in $\mathfrak{G}_n$ are such that $-\mathbf{s}_s^{n_S}$ also belongs to $\mathfrak{G}_{n_S}$, so that $\|\sum_{s=1}^{n_S}\mathbf{s}_s^{n_S}\|$ is 0 or 1 according as n_S is even or odd. One also could consider factorizations of the form $n = n_R n_S + n_0$ with n_S even and $0 \le n_0 < \min(2n_R, n_S)$, then require $\mathfrak{G}_n$ to be symmetric with respect to the origin, automatically yielding $\sum_{s=1}^{n_S}\mathbf{s}_s^{n_S} = \mathbf{0}$.

The empirical counterpart $\mathbf{F}_{\pm}^{(n)}$ of $\mathbf{F}_{\pm}$ is defined as the (bijective, once the origin is given multiplicity n_0) mapping from $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ to the grid $\mathfrak{G}_n$ that minimizes the sum of squared Euclidean distances $\sum_{i=1}^{n}\big\|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) - \mathbf{Z}_i^{(n)}\big\|^2$. That mapping is unique with probability one; in practice, it is obtained via a simple optimal assignment (pairing) algorithm (a linear program; see Section 4 of Hallin (2017) for details).

Call *center-outward rank* of $\mathbf{Z}_i^{(n)}$ the integer (in $\{1, \ldots, n_R\}$ or $\{0, \ldots, n_R\}$ according as $n_0 = 0$ or not) $R_{i;\pm s}^{(n)} := (n_R + 1)\big\|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\big\|$ and *center-outward sign* of $\mathbf{Z}_i^{(n)}$ the unit vector $\mathbf{S}_{i;\pm}^{(n)} := \mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})/\big\|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\big\|$ for $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) \neq \mathbf{0}$; for $\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)}) = \mathbf{0}$, put $\mathbf{S}_{i;\pm s}^{(n)} = \mathbf{0}$.

Some desirable finite-sample properties, such as strict independence between the ranks and the signs, only hold for $n_0 = 0$ or 1, due to the fact that the mapping from the sample to the grid is no longer injective for $n_0 \ge 2$. This, which has no asymptotic consequences (since the number n_0 of tied values involved is $o(n)$ as $n \to \infty$), is easily taken care of by the following tie-breaking device:

(i) randomly select n_0 directions $\mathbf{s}_1^0, \ldots, \mathbf{s}_{n_0}^0$ in $\mathfrak{G}_{n_S}$, then

(ii) replace the n_0 copies of the origin with the new gridpoints $\frac{1}{2(n_R+1)}\mathbf{s}_1^0, \ldots, \frac{1}{2(n_R+1)}\mathbf{s}_{n_0}^0$.

The resulting grid (for simplicity, the same notation $\mathfrak{G}_n$ is used) no longer has multiple points, and the optimal pairing between the sample and the grid is bijective; the n_0 smallest ranks, however, take the non-integer value $1/2$. Again, this tie-breaking device has no influence on asymptotic results.

2.3 Main properties

This section summarizes the main properties of the concepts defined in Sections 2.1 and 2.2; further properties and a proof for Proposition 2.1 can be found in Hallin et al. (2021a).

Proposition 2.1. *Let* $\mathbf{F}_{\pm}$ *denote the center-outward distribution function of* $\mathrm{P} \in \mathcal{P}_d$. *Then,*

(i) $\mathbf{F}_{\pm}$ *is a probability integral transformation of* $\mathbb{R}^d$*: namely,* $\mathbf{Z} \sim \mathrm{P}$ *iff* $\mathbf{F}_{\pm}(\mathbf{Z}) \sim \mathrm{U}_d$*; by construction,* $\|\mathbf{F}_{\pm}(\mathbf{Z})\|$ *is uniform over the interval* $[0, 1]$, $\mathbf{F}_{\pm}(\mathbf{Z})/\|\mathbf{F}_{\pm}(\mathbf{Z})\|$ *uniform*

over the sphere $\mathcal{S}_{d-1}$, and they are mutually independent.

Let $\mathbf{Z}_i^{(n)}, \dots, \mathbf{Z}_i^{(n)}$ be i.i.d. with distribution $\mathrm{P} \in \mathcal{P}_d$ and center-outward distribution function $\mathbf{F}_\pm$. Then,

(ii) $\big(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1^{(n)}), \dots, \mathbf{F}_\pm^{(n)}(\mathbf{Z}_n^{(n)})\big)$ *is uniformly distributed over the $n!/n_0!$ permutations with repetitions of the gridpoints in $\mathfrak{G}_n$ with the origin counted as n_0 indistinguishable points (resp. the $n!$ permutations of $\mathfrak{G}_n$ if either $n_0 \leq 1$ or the tie-breaking device described in Section 2.2 is adopted);*

(iii) if either $n_0 = 0$ or the tie-breaking device described in Section 2.2 is adopted, the n-tuple of center-outward ranks $\big(R_{1;\pm}^{(n)}, \dots, R_{n;\pm}^{(n)}\big)$ and the n-tuple of center-outward signs $\big(\mathbf{S}_{1;\pm}^{(n)}, \dots, \mathbf{S}_{n;\pm}^{(n)}\big)$ are mutually independent;

(iv) if either $n_0 \leq 1$ or the tie-breaking device described in Section 2.2 is adopted, the n-tuple $\big(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1^{(n)}), \dots, \mathbf{F}_\pm^{(n)}(\mathbf{Z}_n^{(n)})\big)$ *is* essentially maximal ancillary.[7]

Assuming, moreover, that $\mathrm{P} \in \mathcal{P}_d^+$,

(v) (Glivenko-Cantelli) $\max_{1\leq i\leq n} \Big\| \mathbf{F}_\pm^{(n)}(\mathbf{Z}_i^{(n)}) - \mathbf{F}_\pm(\mathbf{Z}_i^{(n)}) \Big\| \to 0$ a.s. *as $n \to \infty$.*

Center-outward distribution functions, ranks, and signs also inherit, from the invariance features of Euclidean distances, elementary but quite remarkable invariance and equivariance properties under orthogonal transformations. Denote by $\mathbf{F}_\pm^{\mathbf{Z}}$ the center-outward distribution function of $\mathbf{Z}$ and by $\mathbf{F}_\pm^{\mathbf{Z};(n)}$ the empirical distribution function of a sample $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ associated with a grid $\mathfrak{G}_n$.

Proposition 2.2. *Let $\boldsymbol{\mu} \in \mathbb{R}^d$ and denote by $\mathbf{O}$ a $d \times d$ orthogonal matrix. Then,*

(i) $\mathbf{F}_\pm^{\boldsymbol{\mu}+\mathbf{OZ}}(\boldsymbol{\mu} + \mathbf{Oz}) = \mathbf{O}\mathbf{F}_\pm^{\mathbf{Z}}(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^d$;

(ii) denoting by $\mathbf{F}_\pm^{\boldsymbol{\mu}+\mathbf{OZ};(n)}$ the empirical distribution function of the sample $\boldsymbol{\mu}+\mathbf{OZ}_1, \dots, \boldsymbol{\mu}+\mathbf{OZ}_n$ associated with the grid $\mathbf{O}\mathfrak{G}_n$ (hence, by $\mathbf{F}_\pm^{\mathbf{Z};(n)}$ the empirical distribution function of the sample $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ associated with the grid $\mathfrak{G}_n$),

$$\mathbf{F}_\pm^{\boldsymbol{\mu}+\mathbf{OZ};(n)}(\boldsymbol{\mu} + \mathbf{OZ}_i) = \mathbf{O}\mathbf{F}_\pm^{\mathbf{Z};(n)}(\mathbf{Z}_i), \quad i = 1, \dots, n; \tag{2.1}$$

(iii) the center-outward ranks $R_{i;\pm}^{(n)}$ and the cosines $\mathbf{S}_{i;\pm}^{(n)\prime}\mathbf{S}_{j;\pm}^{(n)}$ computed from the sample $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ and the grid $\mathfrak{G}_n$ are the same as those computed from the sample $\boldsymbol{\mu} + \mathbf{OZ}_1, \dots, \boldsymbol{\mu} + \mathbf{OZ}_n$ and the grid $\mathbf{O}\mathfrak{G}_n$.

See Appendix A.2 for the proof.

[7]See Section 2.4 and Appendices D1 and D.2 of Hallin et al. (2021a) for a precise definition of this crucial property (which entails distribution-freeness) and a proof.

These orthogonal equivariance and invariance properties, however, do not extend to non-orthogonal affine transformations.

3 Hájek representation and asymptotic normality

As in Hájek and Šidák (1967), the rank-based statistics to be used in this context are quadratic forms in vectors of linear rank statistics—involving center-outward ranks and signs instead of ordinary ranks, though. Fundamental in Hájek's approach is an asymptotic representation result establishing the asymptotic equivalence between linear rank statistics and sums of independent variables. We start with a center-outward version of that result; asymptotic normality follows as a corollary.

3.1 Linear center-outward rank statistics

Linear rank statistics in this context depend on a *score function* $\mathbf{J} : \mathbb{S}_d \to \mathbb{R}^d$ and are indexed by triangular arrays $\{c_1^{(n)}, \ldots, c_n^{(n)}\}$ of real numbers (regression constants). On those score functions and regression constants we are making the following assumptions.

Assumption 3.1. *(i)* $\mathbf{J} : \mathbb{S}_d \to \mathbb{R}^d$ is continuous over $\mathbb{S}_d$;

(ii) for any sequence $\mathfrak{s}^{(n)} = \{\mathbf{s}_1^{(n)}, \ldots, \mathbf{s}_n^{(n)}\}$ of n-tuples in $\mathbb{S}_d$ such that the uniform discrete distribution over $\mathfrak{s}^{(n)}$ converges weakly to U_d as $n \to \infty$,

$$\lim_{n\to\infty} n^{-1}\mathrm{tr} \sum_{r=1}^{n} \mathbf{J}(\mathbf{s}_r^{(n)})\mathbf{J}'(\mathbf{s}_r^{(n)}) = \mathrm{tr} \int_{\mathbb{S}_d} \mathbf{J}(\mathbf{u})\mathbf{J}'(\mathbf{u})\, \mathrm{dU}_d \tag{3.2}$$

where $\int_{\mathbb{S}_d} \mathbf{J}(\mathbf{u})\mathbf{J}'(\mathbf{u})\, \mathrm{dU}_d < \infty$ has full rank.

As we shall see, a special role is played, in relation with spherical distributions, by score functions of the form

$$\mathbf{J}(\mathbf{u}) := J(\|\mathbf{u}\|)\frac{\mathbf{u}}{\|\mathbf{u}\|}\mathbf{1}_{[\|\mathbf{u}\|\neq 0]} \qquad \mathbf{u} \in \mathbb{S}_d \tag{3.3}$$

for some function $J : [0, 1) \to \mathbb{R}$. Assumption 3.1 then holds if *(i)* J is continuous and *(ii)*

$$0 < \lim_{n\to\infty} n^{-1} \sum_{r=1}^{n} J^2\big(r/(n+1)\big) = \int_0^1 J^2(u)\, \mathrm{d}u < \infty \tag{3.4}$$

(a sufficient condition for (3.4) is the traditional assumption that J has bounded variation, i.e. is the difference of two nondecreasing functions). Both (3.2) and (3.4) extend the conditions on univariate scores in Section V.1.6 of Hájek and Šidák (1967).

As for the regression constants, we assume that the classical *Noether conditions* hold.

Assumption 3.2. The $c_i^{(n)}$'s are not all equal (for given n) and satisfy

$$\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2 / \max_{1\leq i\leq n}(c_i^{(n)} - \bar{c}^{(n)})^2 \longrightarrow \infty \quad \text{as } n\to\infty \tag{3.5}$$

where $\bar{c}^{(n)} := n^{-1}\sum_{i=1}^{n} c_i^{(n)}$.

Associated with the score functions $\mathbf{J}$, consider the d-dimensional statistics

$$\underset{\sim}{\mathbf{T}}_a^{(n)} = \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1/2}\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})\mathbf{J}(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})), \tag{3.6}$$

$$\underset{\sim}{\mathbf{T}}_e^{(n)} := \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1/2}\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})\mathrm{E}\left[\mathbf{J}(\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)}))\Big|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\right],$$

and

$$\mathbf{T}^{(n)} := \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1/2}\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})\mathbf{J}(\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})).$$

Adopting Hájek's terminology, call $\underset{\sim}{\mathbf{T}}_a^{(n)}$ an *approximate-score* linear rank statistic and $\underset{\sim}{\mathbf{T}}_e^{(n)}$ an *exact-score* linear rank statistic. As we shall see, both $\underset{\sim}{\mathbf{T}}_a^{(n)}$ and $\underset{\sim}{\mathbf{T}}_e^{(n)}$ admit the same asymptotic representation $\mathbf{T}^{(n)}$, hence are asymptotically equivalent. For score functions of the form (3.3), we have

$$\underset{\sim}{\mathbf{T}}_a^{(n)} = \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1/2}\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})J\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\Big)\mathbf{S}_{i;\pm}^{(n)},$$

$$\begin{aligned}\underset{\sim}{\mathbf{T}}_e^{(n)} =& \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1/2}\\ &\times\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})\mathrm{E}\left[J(\|\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})\|)\frac{\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})}{\|\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})\|}\Big|\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i^{(n)})\right],\end{aligned}$$

and

$$\mathbf{T}^{(n)} = \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1/2}\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})J(\|\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})\|)\frac{\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})}{\|\mathbf{F}_{\pm}(\mathbf{Z}_i^{(n)})\|}.$$

3.2 Asymptotic representation and asymptotic normality

The following proposition is a center-outward multivariate counterpart of the asymptotic results in Section V.1.6 of Hájek and Šidák (1967). Throughout this section, we assume that $\mathbf{F}_{\pm}^{(n)}$ is computed from a triangular array $(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)})$, $n \in \mathbb{N}$ of i.i.d. d-dimensional random vectors with distribution $\mathrm{P} \in \mathcal{P}_d^+$ and center-outward distribution function $\mathbf{F}_{\pm}$.

Proposition 3.1 (Hájek representation). *Let Assumptions 3.1 and 3.2 hold and $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ be i.i.d. with distribution $\mathrm{P} \in \mathcal{P}_d^+$. Then,*

$$(i)\ \underset{\sim}{\mathbf{T}}_a^{(n)} - \mathbf{T}^{(n)} = o_{\text{q.m.}}(1), \quad (ii)\ \underset{\sim}{\mathbf{T}}_e^{(n)} - \mathbf{T}^{(n)} = o_{\text{q.m.}}(1), \quad \text{and} \quad (iii)\ \underset{\sim}{\mathbf{T}}_a^{(n)} - \underset{\sim}{\mathbf{T}}_e^{(n)} = o_{\text{q.m.}}(1)$$

as $n \to \infty$ in such a way that $n_R \to \infty$ and $n_S \to \infty$.[8]

See Appendix A.3 for the proof.

The asymptotic normality of $\underset{\sim}{\mathbf{T}}_a^{(n)}$ and $\underset{\sim}{\mathbf{T}}_e^{(n)}$ then follows from Proposition 3.1 and the asymptotic normality of $\mathbf{T}^{(n)}$, along with the distribution-freeness of $\underset{\sim}{\mathbf{T}}_a^{(n)}$ and $\underset{\sim}{\mathbf{T}}_e^{(n)}$.

Proposition 3.2 (Asymptotic normality). *Let Assumptions 3.1 and 3.2 hold and $\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)}$ be i.i.d. with distribution $\mathrm{P} \in \mathcal{P}_d$. Then, $\underset{\sim}{\mathbf{T}}_a^{(n)}$, $\underset{\sim}{\mathbf{T}}_e^{(n)}$, and $\mathbf{T}^{(n)}$ are asymptotically normal as $n \to \infty$ (in such a way that $n_R \to \infty$ and $n_S \to \infty$), with mean $\mathbf{0}$ and covariance $\int_{\mathbb{S}_d} \mathbf{J}(\mathbf{u})\mathbf{J}'(\mathbf{u})\, \mathrm{dU}_d$ reducing, for $\mathbf{J}$ of the form (3.3), to $d^{-1}\int_0^1 J^2(u)\, \mathrm{d}u\, \mathbf{I}_d$.*

See Appendix A.4 for the proof.

4 Multiple-output linear models

Based on the center-outward ranks and signs of Section 3, we now construct rank tests for the slopes of multiple-output linear models, extending to a multivariate setting the methods developed, e.g. in Puri and Sen (1985) for the single-output case.

4.1 The model

Consider the multiple-output linear (or multiple-output regression) model under which an observed $\mathbf{Y}^{(n)}$ satisfies

$$\mathbf{Y}^{(n)} = \mathbf{1}_n \boldsymbol{\beta}_0' + \mathbf{C}^{(n)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{(n)}, \tag{4.1}$$

where $\mathbf{1}_n := (1, \ldots, 1)'$,

$$\mathbf{Y}^{(n)} = \begin{pmatrix} Y_{11}^{(n)} & Y_{12}^{(n)} & \ldots & Y_{1d}^{(n)} \\ \vdots & \vdots & & \vdots \\ Y_{n1}^{(n)} & Y_{n2}^{(n)} & \ldots & Y_{nd}^{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1^{(n)\prime} \\ \vdots \\ \mathbf{Y}_n^{(n)\prime} \end{pmatrix}$$

[8] The notation $o_{\text{q.m.}}(1)$ stands for a sequence of random vectors tending to zero in quadratic mean (hence also in probability).

is an $n \times d$ matrix of n observed d-dimensional outputs,

$$\mathbf{C}^{(n)} = \begin{pmatrix} c_{11}^{(n)} & c_{12}^{(n)} & \dots & c_{1m}^{(n)} \\ \vdots & \vdots & & \vdots \\ c_{n1}^{(n)} & c_{n2}^{(n)} & \dots & c_{nm}^{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1^{(n)\prime} \\ \vdots \\ \mathbf{c}_n^{(n)\prime} \end{pmatrix}$$

an $n \times m$ matrix of (specified) deterministic covariates,

$$\boldsymbol{\beta}_0' = (\beta_{01}, \dots, \beta_{0d}) \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1d} \\ \vdots & \vdots & & \vdots \\ \beta_{m1} & \beta_{m2} & \dots & \beta_{md} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1' \\ \vdots \\ \boldsymbol{\beta}_m' \end{pmatrix}$$

a d-dimensional intercept and an $m \times d$ matrix of regression coefficients, and

$$\boldsymbol{\varepsilon}^{(n)} = \begin{pmatrix} \varepsilon_{11}^{(n)} & \varepsilon_{12}^{(n)} & \dots & \varepsilon_{1d}^{(n)} \\ \vdots & \vdots & & \vdots \\ \varepsilon_{n1}^{(n)} & \varepsilon_{n2}^{(n)} & \dots & \varepsilon_{nd}^{(n)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\varepsilon}_1^{(n)\prime} \\ \vdots \\ \boldsymbol{\varepsilon}_n^{(n)\prime} \end{pmatrix}$$

an $n \times d$ matrix of nonobserved i.i.d. d-dimensional errors $\boldsymbol{\varepsilon}_i^{(n)}$, $i = 1, \dots, n$ with density $f^{\boldsymbol{\varepsilon}}$. If $\boldsymbol{\beta}_0$ is to be identified, a location constraint has to be imposed on $f^{\boldsymbol{\varepsilon}}$. One could think of the classical constraint $\mathrm{E}\boldsymbol{\varepsilon}_i^{(n)} = \mathbf{0}$ (requiring the existence of a finite mean): $\boldsymbol{\beta}_0 + \boldsymbol{\beta}'\mathbf{c}_i^{(n)}$ then is to be interpreted as the expected value of $\mathbf{Y}_i^{(n)}$ for covariate values $\mathbf{c}_i^{(n)}$. In the context of this paper, however, a more natural location constraint (which moreover does not require any integrability condition) is $\mathbf{F}_{\pm}^{\boldsymbol{\varepsilon}}(\mathbf{0}) = \mathbf{0}$, where $\mathbf{F}_{\pm}^{\boldsymbol{\varepsilon}}$ stands for the center-outward distribution function of the $\boldsymbol{\varepsilon}_i^{(n)}$'s: $\mathbf{0}$ and $\boldsymbol{\beta}_0 + \boldsymbol{\beta}'\mathbf{c}_i^{(n)}$ then are *center-outward medians* for $\boldsymbol{\varepsilon}$ and $\mathbf{Y}_i^{(n)}$, respectively.

In most applications, however, one is interested mainly in the impact of the input covariates $\mathbf{c}_i^{(n)}$ on the output $\mathbf{Y}_i^{(n)}$: the matrix $\boldsymbol{\beta}$ is the parameter of interest, and $\boldsymbol{\beta}_0$ is a nuisance. There is no need, then, for identifying $\boldsymbol{\beta}_0$ nor qualifying $\boldsymbol{\beta}_0 + \boldsymbol{\beta}'\mathbf{c}_i^{(n)}$ as a mean or a center-outward median for $\mathbf{Y}_i^{(n)}$: $\boldsymbol{\beta}$ is to be interpreted as a matrix of treatment effects governing the shift $\boldsymbol{\delta}'\boldsymbol{\beta}$ in the distribution of the d-dimensional output produced by a variation $\boldsymbol{\delta}$ in the m-dimensional covariate. Center-outward ranks and signs being insensitive to shifts, there is even no need to specify, nor to estimate $\boldsymbol{\beta}_0$.

4.2 Local Asymptotic Normality (LAN)

The model (4.1) is easily seen to be locally asymptotically normal (LAN) under the following two classical assumptions.

Assumption 4.1. The square root $\mathbf{z} \mapsto (f^\varepsilon)^{1/2}(\mathbf{z})$ of the error density is *differentiable in quadratic mean*,[9] with quadratic mean gradient $\nabla (f^\varepsilon)^{1/2}$. Letting $\boldsymbol{\varphi}_{f^\varepsilon} := -2\nabla (f^\varepsilon)^{1/2}/(f^\varepsilon)^{1/2}$, assume moreover that the *information matrix* $\boldsymbol{\mathcal{I}}_{f^\varepsilon} := \mathrm{E}\left[\boldsymbol{\varphi}_{f^\varepsilon}(\boldsymbol{\varepsilon})\boldsymbol{\varphi}'_{f^\varepsilon}(\boldsymbol{\varepsilon})\right]$ has full rank d.

On the regression constants $\mathbf{C}^{(n)}$, we borrow from Hallin and Paindaveine (2005) the following assumptions; note that Part *(iii)* requires that each of the m triangular arrays of constants $c_{ij}^{(n)}$, $i \in \mathbb{N}$, $j = 1, \ldots, m$ satisfies Assumption 3.2.

Assumption 4.2. Let $\bar{\mathbf{c}}^{(n)} := n^{-1}\sum_{i=1}^n \mathbf{c}_i^{(n)}$, $\mathbf{V}_\mathbf{c}^{(n)} := n^{-1}\sum_{i=1}^n \left(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)}\right)\left(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)}\right)'$, and denote by $\mathbf{D}_\mathbf{c}^{(n)}$ the diagonal matrix with diagonal elements $\left(\mathbf{V}_\mathbf{c}^{(n)}\right)_{jj}$, $j = 1, \ldots, m$:

(i) $\left(\mathbf{V}_\mathbf{c}^{(n)}\right)_{jj} > 0$ for $j = 1, \ldots, m$;

(ii) defining $\mathbf{R}_\mathbf{c}^{(n)} := \mathbf{D}_\mathbf{c}^{(n)-1/2}\mathbf{V}_\mathbf{c}^{(n)}\mathbf{D}_\mathbf{c}^{(n)-1/2}$, the limit $\mathbf{R}_\mathbf{c} := \lim_{n\to\infty}\mathbf{R}_\mathbf{c}^{(n)}$ exists, is positive definite, and factorizes into $\mathbf{R}_\mathbf{c} = \left(\mathbf{K}_\mathbf{c}\mathbf{K}_\mathbf{c}'\right)^{-1}$ for some full-rank $m \times m$ matrix $\mathbf{K}_\mathbf{c}$;

(iii) letting $\bar{c}_j^{(n)} := n^{-1}\sum_{i=1}^n c_{ij}^{(n)}$, the following Noether conditions hold:

$$\lim_{n\to\infty}\sum_{i=1}^n \left(c_{ij}^{(n)} - \bar{c}_j^{(n)}\right)^2 \Big/ \max_{1\le i\le n}\left(c_{ij}^{(n)} - \bar{c}_j^{(n)}\right)^2 = \infty, \qquad j = 1, \ldots, m.$$

Letting $\mathbf{Z}_i^{(n)} = \mathbf{Z}_i^{(n)}(\boldsymbol{\beta}) := \mathbf{Y}_i^{(n)} - \mathbf{1}_n\boldsymbol{\beta}_0' - \boldsymbol{\beta}'\mathbf{c}_i^{(n)}$, the following result readily follows from, e.g., (Lehmann and Romano, 2005, Theorem 12.2.3). In order to simplify the notation, we throughout adopt the same contiguity rates as in Hallin and Paindaveine (2005). Namely, we consider local perturbations of the parameter $\boldsymbol{\beta}$ of the form $\boldsymbol{\beta} + \boldsymbol{\nu}(n)\boldsymbol{\tau}$ where $\boldsymbol{\tau}$ is an $m \times d$ matrix and $\boldsymbol{\nu}(n) := n^{-1/2}\mathbf{K}_\mathbf{c}^{(n)}$, with $\mathbf{K}_\mathbf{c}^{(n)} := \left(\mathbf{D}_\mathbf{c}^{(n)}\right)^{-1/2}\mathbf{K}_\mathbf{c}$. This is a notational convenience and has no impact on the form of locally asymptotically optimal test statistics.

Proposition 4.1. *Under Assumptions 4.1 and 4.2, the model (4.1) is LAN (with respect to $\boldsymbol{\beta}$), with central sequence $\boldsymbol{\Delta}_{\boldsymbol{\beta}_0;f^\varepsilon}^{(n)}(\boldsymbol{\beta}) := n^{1/2}\mathrm{vec}\boldsymbol{\Lambda}_{\boldsymbol{\beta}_0;f^\varepsilon}^{(n)}$ where*

$$\boldsymbol{\Lambda}_{\boldsymbol{\beta}_0;f^\varepsilon}^{(n)} := \frac{1}{n}\sum_{i=1}^n \mathbf{K}_\mathbf{c}^{(n)\prime}\left(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)}\right)\boldsymbol{\varphi}'_{f^\varepsilon}(\mathbf{Z}_i^{(n)}) \tag{4.2}$$

and Fisher information $\boldsymbol{\mathcal{I}}_{f^\varepsilon} \otimes \mathbf{I}_m$

LAN for the same linear model (4.1) has been established (in the broader context of regression with VARMA errors in Hallin and Paindaveine (2005)) under the assumption

[9] It follows from a result by Lind and Roussas (1972) independently rediscovered by Garel and Hallin (1995) that quadratic mean differentiability is equivalent to partial quadratic mean derivability with respect to all variables.

that the error density $f^{\boldsymbol{\varepsilon}}$ is *centered elliptical*,[10] that is, has the form

$$f^{\boldsymbol{\varepsilon}}(\mathbf{z}) = \kappa_{d,\mathfrak{f}}^{-1}\big(\det\boldsymbol{\Sigma}\big)^{-1/2}\mathfrak{f}\big((\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z})^{1/2}\big) \tag{4.3}$$

with $\kappa_{d,\mathfrak{f}} := \big(2\pi^{d/2}/\Gamma(d/2)\big)\int_0^\infty r^{d-1}\mathfrak{f}(r)\,\mathrm{d}r$ for some symmetric positive definite *shape matrix* $\boldsymbol{\Sigma}$ and some *radial density* $\mathfrak{f}$ (over $\mathbb{R}_0^+$) such that $\mathfrak{f}(z) > 0$ Lebesgue-a.e. in $\mathbb{R}_0^+$ and $\int_0^\infty r^{d-1}\mathfrak{f}(r)\,\mathrm{d}r < \infty$. When $\boldsymbol{\varepsilon}$ is elliptical with shape matrix $\boldsymbol{\Sigma}$ and radial density $\mathfrak{f}$, $\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}\|$ has density $\mathfrak{f}_d^\star(r) = (\mu_{d-1;\mathfrak{f}})^{-1}r^{d-1}\mathfrak{f}(r)I[r>0]$, where $\mu_{d-1;\mathfrak{f}} := \int_0^\infty r^{d-1}\mathfrak{f}(r)\mathrm{d}\,r$, and distribution function $F^\star_{d;\mathfrak{f}}$.

Assumption 4.1 then is equivalent to the mean square differentiability, with quadratic mean derivative $\big(\mathfrak{f}^{1/2}\big)'$, of $x \mapsto \mathfrak{f}^{1/2}(x)$, $x \in \mathbb{R}_0^+$ (a scalar); letting $\varphi_{\mathfrak{f}} := -2\big(\mathfrak{f}^{1/2}\big)'/\mathfrak{f}^{1/2}$, we automatically get $\mathcal{I}_{d;\mathfrak{f}} := \int_0^1\Big(\varphi_{\mathfrak{f}} \circ \big(F^\star_{d;\mathfrak{f}}\big)^{-1}(u)\Big)^2\mathrm{d}u < \infty$. Define the *sphericized residuals*

$$\mathbf{Z}_i^{(n)\,\mathrm{ell}} := \big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\big(\mathbf{Y}_i^{(n)} - \boldsymbol{\beta}_0 - \boldsymbol{\beta}'\mathbf{c}_i^{(n)}\big) = \big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\big(\mathbf{Z}_i^{(n)}\big), \quad i = 1,\ldots,n \tag{4.4}$$

where the matrix $\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{1/2}$ is the symmetric root of a consistent estimator $\widehat{\boldsymbol{\Sigma}}^{(n)}$ of some multiple $a\,\boldsymbol{\Sigma}$ of $\boldsymbol{\Sigma}$ ($a > 0$ an arbitrary constant) satisfying the following consistency assumption.

Assumption 4.3. Under (4.1), $\widehat{\boldsymbol{\Sigma}}^{(n)} - a\boldsymbol{\Sigma} = O_{\mathrm{P}}(n^{-1/2})$ as $n \to \infty$, for some $a > 0$; moreover, $\widehat{\boldsymbol{\Sigma}}^{(n)}$ is invariant under permutations and reflections (with respect to the origin) of the residuals $\mathbf{Z}_i^{(n)} = (\mathbf{Y}_i^{(n)} - \mathbf{1}_n\boldsymbol{\beta}_0' - \boldsymbol{\beta}'\mathbf{c}_i^{(n)})$'s, and equivariant under their affine transformations.

A traditional choice which, however, rules out heavy-tailed radial densities with infinite second-order moments, is the empirical covariance of the $\mathbf{Z}_i^{(n)}$'s. An alternative, satisfying Assumption 4.3 without any moment assumptions, is Tyler's estimator of scatter, see Theorem 4.1 in Tyler (1987) for strong consistency, Theorem 4.2 for asymptotic normality.

Under Assumption 4.3, which entails the affine invariance of $\mathbf{Z}_i^{(n)\,\mathrm{ell}}$, Proposition 4.1 takes the following form.

Proposition 4.2. *Under Assumptions 4.2 and 4.3, the model (4.1) with error density $f^{\boldsymbol{\varepsilon}}$ of the elliptical type (4.3) and quadratic mean differentiable $\mathfrak{f}^{1/2}$ is LAN (with respect to $\boldsymbol{\beta}$), with central sequence*

$$\boldsymbol{\Delta}^{(n)\,\mathrm{ell}}_{\widehat{\boldsymbol{\Sigma}}^{(n)},\boldsymbol{\beta}_0;\mathfrak{f}}(\boldsymbol{\beta}) := n^{1/2}\left(\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}\otimes\mathbf{I}_m\right)\mathrm{vec}\boldsymbol{\Lambda}^{(n)\,\mathrm{ell}}_{\widehat{\boldsymbol{\Sigma}}^{(n)},\boldsymbol{\beta}_0;\mathfrak{f}}(\boldsymbol{\beta}) = \boldsymbol{\Delta}^{(n)\,\mathrm{ell}}_{\boldsymbol{\Sigma},\boldsymbol{\beta}_0;\mathfrak{f}}(\boldsymbol{\beta}) + o_{\mathrm{P}}(1) \tag{4.5}$$

where

[10] For simplicity, we henceforth are dropping the word "centered."

$$\boldsymbol{\Lambda}^{(n)\,\mathrm{ell}}_{\widehat{\boldsymbol{\Sigma}}^{(n)},\boldsymbol{\beta}_0;\mathfrak{f}}(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^{n} \varphi_{\mathfrak{f}}\big(\|\mathbf{Z}_i^{(n)\,\mathrm{ell}}\|\big)\mathbf{K}_{\mathbf{c}}^{(n)\prime}\big(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)}\big)\left(\frac{\mathbf{Z}_i^{(n)\,\mathrm{ell}}}{\|\mathbf{Z}_i^{(n)\,\mathrm{ell}}\|}\right)^{\prime}, \tag{4.6}$$

yielding a Fisher information matrix $\frac{1}{d}\mathcal{I}_{d;\mathfrak{f}}\,\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_m$.

This LAN result, where the residuals are subjected to preliminary (empirical) sphericization via $\big(\widehat{\boldsymbol{\Sigma}}^{(n)}\big)^{-1/2}$, stresses the fact that elliptical families with given $\mathfrak{f}$ are parametrized *spherical families* (indexed by $\boldsymbol{\Sigma}$). Actually, since $\boldsymbol{\Delta}^{(n)\,\mathrm{ell}}_{\boldsymbol{\Sigma},\boldsymbol{\beta}_0;\mathfrak{f}}(\boldsymbol{\beta}) = \Big(\mathbf{I}_d\otimes\boldsymbol{\Sigma}^{-1/2}\Big)\,\boldsymbol{\Delta}^{(n)\,\mathrm{ell}}_{\mathbf{I}_d,\boldsymbol{\beta}_0;\mathfrak{f}}(\boldsymbol{\beta})$, the limiting Gaussian shift experiments associated with elliptical and spherical errors coincide (with the perturbation $\mathrm{vec}(\boldsymbol{\tau})$ of $\mathrm{vec}(\boldsymbol{\beta})$ in the elliptical case corresponding to a perturbation $\mathrm{vec}(\boldsymbol{\varsigma}) = \Big(\mathbf{I}_d\otimes\boldsymbol{\Sigma}^{-1/2}\Big)\,\mathrm{vec}(\boldsymbol{\tau})$ in the spherical case). That invariance under linear sphericization of local limiting Gaussian shifts, however, does not extend to the general case of Proposition 4.1.

5 Rank tests for multiple-output linear models

5.1 Elliptical (Mahalanobis) rank tests

Rank-based inference for elliptical multiple-output linear models was developed in Hallin and Paindaveine (2005). The ranks and the signs there are the *elliptical* or *Mahalanobis ranks and signs*—namely, the ranks $R_i^{(n)\,\mathrm{ell}}$ of the moduli $\|\mathbf{Z}_i^{(n)\,\mathrm{ell}}\|$ and the signs (directions) $\mathbf{S}_i^{(n)\,\mathrm{ell}} := \mathbf{Z}_i^{(n)\,\mathrm{ell}}/\|\mathbf{Z}_i^{(n)\,\mathrm{ell}}\|$, both computed, in agreement with the above remark on the spherical nature of elliptical families, after the empirical sphericization (4.4).

Consider the null hypothesis $H_0^{(n)}(\boldsymbol{\beta}^0)$ under which $\mathbf{Y}^{(n)}$ satisfies (4.1) with $\boldsymbol{\beta} = \boldsymbol{\beta}^0$, specified $\boldsymbol{\beta}_0$, elliptical $f^{\boldsymbol{\varepsilon}}$, and radial density $\mathfrak{f}$. Hallin and Paindaveine (2005) define

$$\underset{\sim}{\boldsymbol{\Lambda}}_J^{(n)\,\mathrm{ell}} := n^{-1}\sum_{i=1}^{n} J\Big(\frac{R_i^{(n)\,\mathrm{ell}}}{n+1}\Big)\,\mathbf{S}_i^{(n)\,\mathrm{ell}}\big(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)}\big)'\mathbf{K}_{\mathbf{c}}^{(n)}$$

for a score function $J : [0,1) \to \mathbb{R}$ and show that the test of $H_0^{(n)}(\boldsymbol{\beta}^0)$ can be based on

$$\underset{\sim}{Q}_J^{(n)\,\mathrm{ell}}(\boldsymbol{\beta}^0) = \frac{n\,d}{\int_0^1 J^2(u)du}\big(\mathrm{vec}\underset{\sim}{\boldsymbol{\Lambda}}_J^{(n)\,\mathrm{ell}}\big)'\big(\mathrm{vec}\underset{\sim}{\boldsymbol{\Lambda}}_J^{(n)\,\mathrm{ell}}\big),$$

which has asymptotically χ^2_{md} distribution under the null.

The validity of tests based on those elliptical ranks and signs, unfortunately, requires an elliptical $f^{\boldsymbol{\varepsilon}}$. A welcome relaxation of stricter Gaussianity assumptions, ellipticity remains an extremely strong symmetry requirement; it is made, essentially, for lack of anything better but is unlikely to hold in practice. If the assumption of ellipticity is to be waived, elliptical ranks and signs are losing their distribution-freeness for the benefit of the center-outward ranks and signs. And, since center-outward ranks and signs, in view of Proposition 2.2, are invariant under location shift, center-outward rank tests can address the (more realistic) unspecified intercept case without any additional estimation step.

5.2 Center-outward rank tests

Denote by $\mathbf{F}_{\pm}^{(n)}$ the empirical center-outward distribution associated with the observed n-tuple $(\mathbf{Z}_1^{(n)}, \ldots, \mathbf{Z}_n^{(n)})$ where $\mathbf{Z}_i^{(n)}$ now is defined as $\mathbf{Y}_i^{(n)} - \boldsymbol{\beta}'\mathbf{c}_i^{(n)}$, by $R_{i;\pm}^{(n)}$ and $\mathbf{S}_{i;\pm}^{(n)}$, respectively, the corresponding center-outward ranks and signs. In line with the form of the central sequence (4.2), consider

$$\underset{\sim}{\boldsymbol{\Lambda}}_{\mathbf{J}}^{(n)\pm} := n^{-1} \sum_{i=1}^{n} \mathbf{K}_{\mathbf{c}}^{(n)\prime}\big(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)}\big)\mathbf{J}'\left(\frac{R_{i;\pm}^{(n)}}{n_R + 1}\mathbf{S}_{i;\pm}^{(n)}\right). \tag{5.1}$$

It follows from the asymptotic representation result of Proposition 3.1 that, when the actual density is $f^{\boldsymbol{\varepsilon}}$, for the scores $\mathbf{J} = \boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}} \circ \mathbf{F}_{\pm}^{-1}$, with $\boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}}$ defined in Assumption 4.1

$$\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\beta}_0;f^{\boldsymbol{\varepsilon}}}^{(n)}(\boldsymbol{\beta}) := n^{1/2}\text{vec}\underset{\sim}{\boldsymbol{\Lambda}}_{\mathbf{J}}^{(n)\pm} = \boldsymbol{\Delta}_{\boldsymbol{\beta}_0;f^{\boldsymbol{\varepsilon}}}^{(n)}(\boldsymbol{\beta}) + o_{\mathrm{P}}(1) \tag{5.2}$$

and $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\beta}_0;f^{\boldsymbol{\varepsilon}}}^{(n)}(\boldsymbol{\beta})$ thus constitutes a version, based on the center-outward ranks and signs and hence distribution-free, of the central sequence $\boldsymbol{\Delta}_{\boldsymbol{\beta}_0;f^{\boldsymbol{\varepsilon}}}^{(n)}(\boldsymbol{\beta})$ in (4.2). The following asymptotic normality result then holds.

Proposition 5.1. *Let $\mathbf{Y}_i^{(n)}$ satisfy (4.1) and Assumptions 3.1 and 4.2 hold. Then,*

(i) $n^{1/2}\text{vec}\underset{\sim}{\boldsymbol{\Lambda}}_{\mathbf{J}}^{(n)\pm}$ *is asymptotically normal, with mean* $\mathbf{0}$ *and covariance* $\boldsymbol{\mathcal{I}}_{\mathbf{J}} \otimes \mathbf{I}_m$ *where* $\boldsymbol{\mathcal{I}}_{\mathbf{J}} := \int_{\mathcal{S}_d} \mathbf{J}(\mathbf{u})\mathbf{J}'(\mathbf{u})\mathrm{dU}_d$, *under the null hypothesis* $H_0^{(n)}(\boldsymbol{\beta}^0)$ *that* $\boldsymbol{\beta} = \boldsymbol{\beta}^0$ *while the intercept* $\boldsymbol{\beta}_0$ *and the distribution* $\mathrm{P} \in \mathcal{P}_d$ *of the* $\boldsymbol{\varepsilon}$*'s remain unspecified;*

(ii) the test rejecting $H_0^{(n)}(\boldsymbol{\beta}^0)$ *whenever the test statistic*

$$\underset{\sim}{\mathbf{Q}}_{\mathbf{J}}^{(n)\pm} := n\left(\text{vec}\underset{\sim}{\boldsymbol{\Lambda}}_{\mathbf{J}}^{(n)\pm}\right)' \boldsymbol{\mathcal{I}}_{\mathbf{J}}^{-1} \otimes \mathbf{I}_m \left(\text{vec}\underset{\sim}{\boldsymbol{\Lambda}}_{\mathbf{J}}^{(n)\pm}\right) \tag{5.3}$$

exceeds the $(1-\alpha)$ *quantile of a chi-square distribution with* md *degrees of freedom has asymptotic level* α *as* $n \to \infty$*;*[11]

[11]Since $\underset{\sim}{\mathbf{Q}}_{\mathbf{J}}^{(n)\pm}$ is distribution-free under the null hypothesis $H_0^{(n)}(\boldsymbol{\beta}^0)$, the finite-$n$ size of this test is uniform over $H_0^{(n)}(\boldsymbol{\beta}^0)$, hence uniformly close to α for n large enough. This is in sharp contrast with daily practice pseudo-Gaussian tests, which remain asymptotically valid under a broad range of distributions, albeit not uniformly so (see Section 1.1).

(iii) for $\mathbf{J} = \boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}} \circ \mathbf{F}_{\pm}^{-1}$ *where* $\mathbf{F}_{\pm}$ *denotes the center-outward distribution function associated with* $f^{\boldsymbol{\varepsilon}}$*, the covariance* $\boldsymbol{\mathcal{I}}_{\mathbf{J}} \otimes \mathbf{I}_m$ *coincides with* $\boldsymbol{\mathcal{I}}_{f^{\boldsymbol{\varepsilon}}} \otimes \mathbf{I}_m$ *and the test based on* $\underset{\sim}{\mathbf{Q}}{}_{\mathbf{J}_{f^{\boldsymbol{\varepsilon}}}}^{(n)\pm}$ *is, under error density* $f^{\boldsymbol{\varepsilon}}$*, locally asymptotically maximin, at asymptotic level* α*, for the null hypothesis* $H_0^{(n)}(\boldsymbol{\beta}^0)$*.*

Corollary 5.2. *(i) In the particular case of a* spherical score *of the form* (3.3)*, the test statistic* $\underset{\sim}{\mathbf{Q}}{}_{\mathbf{J}}^{(n)\pm}$ *simplifies into*

$$\underset{\sim}{\mathbf{Q}}{}_{J}^{(n)\pm} = \frac{nd}{\int_0^1 J^2(u)\mathrm{d}u} \left(\mathrm{vec}\underset{\sim}{\boldsymbol{\Lambda}}{}_J^{(n)\pm}\right)' \left(\mathrm{vec}\underset{\sim}{\boldsymbol{\Lambda}}{}_J^{(n)\pm}\right) \tag{5.4}$$

where $\underset{\sim}{\boldsymbol{\Lambda}}{}_J^{(n)\pm} := n^{-1}\sum_{i=1}^n J\left(\frac{R_{i;\pm}^{(n)}}{n_R+1}\right)\mathbf{K}_{\mathbf{c}}^{(n)\prime}(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})\mathbf{S}_{i;\pm}^{(n)\prime}$ *and* $n^{1/2}\mathrm{vec}\underset{\sim}{\boldsymbol{\Lambda}}{}_J^{(n)\pm}$ *is asymptotically normal with mean* $\mathbf{0}$ *and variance* $d^{-1}\int_0^1 J^2(u)\mathrm{d}u\,\mathbf{I}_{md}$*.*

(ii) The test statistic $\underset{\sim}{\mathbf{Q}}{}_{J}^{(n)\pm}$ *with spherical score* $J_{\mathfrak{f}} := \varphi_{\mathfrak{f}} \circ \left(F_{d;\mathfrak{f}}^{\star}\right)^{-1}$ *yields locally asymptotically optimal tests under the spherical density with radial density* $\mathfrak{f}$*.*

The results of this section provide asymptotic (centered normal and chi-square) distributions under the null. Deriving asymptotic (shifted normal and noncentral chi-square) distributions under local alternatives is a straightforward application of Le Cam's third lemma. The shifts and noncentrality parameters, however, take the form of integrals involving the score function $\mathbf{J} = \boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}} \circ \mathbf{F}_{\pm}^{-1}$ where $\mathbf{F}_{\pm}$ denotes the center-outward distribution function associated with the actual density $f^{\boldsymbol{\varepsilon}}$. Unless $f^{\boldsymbol{\varepsilon}}$ is spherical, these scores, accordingly, cannot be expressed under analytical form and these integrals thus are of little practical interest: rather than overloading the paper with cumbersome but useless formulas, we do not report them. The case of spherical densities $f^{\boldsymbol{\varepsilon}}$ is an exception, though: provided that the test statistic itself is based on spherical scores, the noncentrality parameters of its asymptotic non-null chi-square distributions coincide with those obtained under ellipticity in Hallin and Paindaveine (2005): see Hallin and Paindaveine (2002b) for numerical values, AREs, Chernoff-Savage and Hodges-Lehmann results.

5.3 Some particular cases

In this section, we provide explicit forms of the test statistic for the two-sample and MANOVA problems. Because of their simplicity and practical value (see Section 6.1), we concentrate on the case (5.4) of spherical scores, from which the general case (5.3) is easily deduced (essentially, by substituting $\mathbf{J}\left(\frac{R_{i;\pm}^{(n)}}{n_R+1}\mathbf{S}_{i;\pm}^{(n)}\right)$ for $J\left(\frac{R_{i;\pm}^{(n)}}{n_R+1}\right)\mathbf{S}_{i;\pm}^{(n)}$).

5.3.1 Center-outward rank tests for two-sample location

An important particular case is the two-sample location model, where $n = n_1 + n_2$ and (4.1) holds with covariates of the form $\mathbf{C}^{(n)} = (\mathbf{1}'_{n_1}, \mathbf{0}'_{n_2})'$ (with $\mathbf{1}_{n_1}$ an n_1-dimensional column vector of ones, $\mathbf{0}_{n_2}$ an n_2-dimensional column vector of zeros); the parameter $\boldsymbol{\beta} = (\beta_{11}, \ldots, \beta_{1d})'$ here is a d-dimensional row vector. The objective is to test the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}_d$ under which the distributions of $\mathbf{Y}_1^{(n)}, \ldots, \mathbf{Y}_{n_1}^{(n)}$ and $\mathbf{Y}_{n_1+1}^{(n)}, \ldots, \mathbf{Y}_n^{(n)}$ coincide. Elementary computation yields $\bar{c}^{(n)} = n_1/n$, $V_{\mathbf{c}}^{(n)} = n_1 n_2/n^2$, and $K_{\mathbf{c}} = 1$. If the regular grid $\mathfrak{G}_n$ is chosen such that $\| \sum_{s=1}^{n_S} \mathbf{s}_s^{n_S} \| = \mathbf{0}$ (which is always possible in view of Section 2.2), $\sum_{i=1}^{n} J\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\Big)\mathbf{S}_{i;\pm}^{(n)} = \mathbf{0}$ and the test statistic (5.4) takes the simple form

$$\underset{\sim}{Q}_J^{(n)\pm} = \left(nd/n_1 n_2 \int_0^1 J^2(u)\mathrm{d}u\right) \left\| \sum_{i=1}^{n_1} J\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\Big)\mathbf{S}_{i;\pm}^{(n)} \right\|^2 ; \tag{5.5}$$

else, a centering term $\frac{n_1}{n} \sum_{i=1}^{n} J\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\Big)\mathbf{S}_{i;\pm}^{(n)}$ is to be subtracted. Assumption 4.2 *(iii)* requires $\lim_{n\to\infty} n \min\{n_1, n_2\}/\max\{n_1, n_2\} = \infty$, which holds whenever

$$\lim_{n\to\infty} \min\{n_1, n_2\} = \infty. \tag{5.6}$$

Under Assumptions 3.1 and (5.6), with $\mathrm{P} \in \mathcal{P}_d$, $\underset{\sim}{Q}_J^{(n)\pm}$ is, under H_0, asymptotically χ^2 with d degrees of freedom and the null hypothesis can be rejected at asymptotic level α whenever $\underset{\sim}{Q}_J^{(n)\pm}$ exceeds the $(1-\alpha)$ quantile of a χ^2_d distribution.

5.3.2 Center-outward rank tests for MANOVA

Another important special case of model (4.1) is the multivariate K-sample location or MANOVA model. The observation here decomposes into K samples, with respective sizes $n_1, \ldots, n_K$ and $n = \sum_{k=1}^{K} n_k$. Precisely, $\mathbf{Y}^{(n)} =: \big(\mathbf{Y}^{(n;1)}, \ldots, \mathbf{Y}^{(n;k)}, \ldots, \mathbf{Y}^{(n;K)}\big)$ with

$$\mathbf{Y}^{(n;k)} = \begin{pmatrix} Y_{k;11}^{(n)} & Y_{k;12}^{(n)} & \ldots & Y_{k;1d}^{(n)} \\ \vdots & \vdots & & \vdots \\ Y_{k;n_k1}^{(n)} & Y_{k;n_k2}^{(n)} & \ldots & Y_{k;n_kd}^{(n)} \end{pmatrix}$$

and (4.1) holds with the matrix of covariates

$$\mathbf{C}^{(n)} = \begin{pmatrix} \mathbf{d}_{11}^{(n)} & \mathbf{d}_{12}^{(n)} & \ldots & \mathbf{d}_{1,K-1}^{(n)} \\ \mathbf{d}_{21}^{(n)} & \mathbf{d}_{22}^{(n)} & \ldots & \mathbf{d}_{2,K-1}^{(n)} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{d}_{K1}^{(n)} & \mathbf{d}_{K2}^{(n)} & \ldots & \mathbf{d}_{K,K-1}^{(n)}, \end{pmatrix}$$

where $\mathbf{d}_{ij}^{(n)} = \mathbf{1}_{n_i} I[i=j]$, $i = 1, \ldots, K$ and $j = 1, \ldots, K-1$. The null hypothesis is the hypothesis of no treatment effect $H_0 : \boldsymbol{\beta} = \mathbf{0}_{(K-1)\times d}$.

Letting $\mathbf{v}^{(n)} := (n_1/n, \ldots, n_{K-1}/n)'$, the matrix $\mathbf{V_c}^{(n)}$ in Assumption 4.2 takes the form $\mathbf{V_c}^{(n)} = \text{diag}\{\mathbf{v}^{(n)}\} - \mathbf{v}^{(n)}\mathbf{v}^{(n)\prime}$, where $\text{diag}\{\mathbf{v}^{(n)}\}$ stands for the diagonal matrix with diagonal entries $\mathbf{v}^{(n)}$. If the regular grid $\mathfrak{S}_n$ is chosen such that $\| \sum_{s=1}^{n_S} \mathbf{s}_s^{n_S} \| = 0$ and $(\mathbf{V}_\mathbf{c}^{(n)})^{-1/2}$ is substituted for its limit $\mathbf{K}_\mathbf{c}^{(n)}$, the test statistic (5.4) simplifies into

$$\underset{\sim}{Q}_J^{(n)\pm} = \frac{d}{\int_0^1 J^2(u)\mathrm{d}u} \sum_{k=1}^{K} \frac{1}{n_k} \left\| \sum_{i=n_1+\ldots+n_{k-1}+1}^{n_1+\ldots+n_k} J\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\Big) \mathbf{S}_{i;\pm}^{(n)} \right\|^2 .$$

Assumption 4.2*(iii)* is satisfied as soon as $\lim_{n\to\infty} \min\{n_1, \ldots, n_K\} \to \infty$. Assuming moreover that $0 < \liminf_{n\to\infty} n_k/n \leq \limsup_{n\to\infty} n_k/n < 1$ for $1 \leq k \leq K$, the limit matrix $\mathbf{R}_c$ is positive definite[12] and Assumption 4.2*(ii)* is satisfied as well. Then, under the null hypothesis of no treatment effect, $\underset{\sim}{Q}_J^{(n)\pm}$ is asymptotically chi-square with $(K-1)d$ degrees of freedom and the test rejecting H_0 whenever $\underset{\sim}{Q}_J^{(n)\pm}$ exceeds the corresponding $(1-\alpha)$ quantile has asymptotic level α irrespective of the actual error distribution $\mathrm{P} \in \mathcal{P}_d$. This test is a multivariate generalization of the well-known univariate rank test for K-sample equality of location (the univariate one-way ANOVA hypothesis of no treatment effect), see (Hájek and Šidák, 1967, p.170). Note that, for $K = 2$, $\underset{\sim}{Q}_J^{(n)\pm}$ coincides with the two-sample test statistic obtained in Section 5.3.1.

6 Choosing a score function

Section 5 allows us to construct, based on any $\mathbf{J}$ or J satisfying Assumption 3.1 (either with (3.2) or (3.4)), strictly distribution-free center-outward rank tests of the null hypothesis $H_0^{(n)}(\boldsymbol{\beta}^0)$ under which $\boldsymbol{\beta} = \boldsymbol{\beta}^0$ while the intercept $\boldsymbol{\beta}_0$ and the error distribution $\mathrm{P} \in \mathcal{P}_d$ remain unspecified. All these tests, however, depend on a score function to be selected by the practitioner. Some will favor simple scores of the spherical type (see Section 6.1); others may want to base their choice on efficiency considerations (see Section 6.2).

[12] This limit possibly can exist along subsequences, with asymptotic statements modified accordingly. For the sake of simplicity, we do not include this in subsequent results.

6.1 Standard score functions

Popular choices are the spherical sign test, Wilcoxon and van der Waerden scores. Let us describe them, in more details, in the particular case of the two-sample problem.

The two-sample sign test is based on the degenerate score $J_{\text{sign}}(r) := 1$ for $r \in [0, 1)$; using the fact that $\sum_{i=1}^{n} \mathbf{S}_{i;\pm}^{(n)} = \mathbf{0}$, one gets for (5.5), with the notation of Section 5.3.1, the very simple test statistic

$$\underset{\sim}{\mathbf{Q}}_{\text{sign}}^{(n)\pm} = \frac{nd}{n_1 n_2} \left\| \sum_{i=1}^{n_1} \mathbf{S}_{i;\pm}^{(n)} \right\|^2 .$$

The choice $J_{\text{Wilcoxon}}(r) := r$ similarly characterizes the Wilcoxon two-sample test: noting that $\sum_{i=1}^{n} R_{i;\pm}^{(n)} \mathbf{S}_{i;\pm}^{(n)} = \mathbf{0}$ holds if $\sum_{i=1}^{n} \mathbf{S}_{i;\pm}^{(n)} = \mathbf{0}$ and that $\int_0^1 r^2 \mathrm{d}u = 1/3$, this yields

$$\underset{\sim}{\mathbf{Q}}_{\text{Wilcoxon}}^{(n)\pm} = \frac{3nd}{n_1 n_2 (n_R + 1)} \left\| \sum_{i=1}^{n_1} R_{i;\pm}^{(n)} \mathbf{S}_{i;\pm}^{(n)} \right\|^2 .$$

As for the two-sample van der Waerden test, it is based on the Gaussian or van der Waerden scores $J_{\text{vdW}}(r) := \big(\Psi_d^{-1}(r)\big)^{1/2}$, where Ψ_d denotes the cumulative distribution function of a chi-square variable with d degrees of freedom. Clearly $\int_0^1 J_{\text{vdW}}^2(r)\mathrm{d}r = \int_0^\infty x \mathrm{d}\Psi_d(x) = d$ and, provided that $\sum_{i=1}^{n} \mathbf{S}_{i;\pm}^{(n)} = 0$, $\sum_{i=1}^{n} \Big(\Psi_d^{-1}\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\Big)\Big)^{1/2} \mathbf{S}_{i;\pm}^{(n)} = 0$. Hence, the van der Waerden center-outward rank test statistics takes the form

$$\underset{\sim}{\mathbf{Q}}_{\text{vdW}}^{(n)\pm} = \frac{n}{n_1 n_2} \left\| \sum_{i=1}^{n_1} \Big(\Psi_d^{-1}\Big(\frac{R_{i;\pm}^{(n)}}{n_R + 1}\Big)\Big)^{1/2} \mathbf{S}_{i;\pm}^{(n)} \right\|^2 .$$

6.2 Score functions and efficiency

The tests statistics in Section 6.1 offer the advantage of a structure paralleling the structure of the numerator of the classical Gaussian F test—basically substituting, in the latter, $\mathbf{S}_{i;\pm}^{(n)}$ (sign test scores), $R_{i;\pm}^{(n)} \mathbf{S}_{i;\pm}^{(n)}$ (Wilcoxon scores), or $\Big(\Psi_d^{-1}\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\Big)\Big)^{1/2} \mathbf{S}_{i;\pm}^{(n)}$ (van der Waerden scores) for the sphericized residuals (4.4)[13] and adopting the adequate standardization.

The choice of a score function also can be guided by efficiency considerations, selecting $\mathbf{J}$ in relation to some reference distribution under which efficiency is to be attained. This, in the univariate case, yields the normal (van der Waerden), Wilcoxon or sign test scores, achieving efficiency under Gaussian, logistic, or double exponential reference densities;

[13] The computation of which, moreover, requires the specification of $\boldsymbol{\beta}_0$ or its consistent estimation—something center-outward ranks and signs do not need in view of their shift-invariance.

as we shall see, $\underset{\sim}{\mathbf{Q}}^{(n)\pm}_{\text{sign}}$ and $\underset{\sim}{\mathbf{Q}}^{(n)\pm}_{\text{vdW}}$ similarly achieve efficiency at spherical exponential and Gaussian reference distributions.[14]

In the same spirit, one could contemplate the idea of achieving, based on center-outward rank tests, efficiency at some selected reference distribution $\mathrm{P}^{\boldsymbol{\varepsilon}}_0$ in $\mathcal{P}_d$ (with density $f^{\boldsymbol{\varepsilon}}_0$ and center-outward distribution function $\mathbf{F}^{\boldsymbol{\varepsilon}}_{0;\pm}$ satisfying the adequate regularity assumptions). Indeed, it follows from Proposition 5.1 that efficiency under $\mathrm{P}^{\boldsymbol{\varepsilon}}_0$ can be achieved by a test based on the test statistic $\underset{\sim}{\mathbf{Q}}^{(n)\pm}_{\mathbf{J}}$ given in (5.3) with score $\mathbf{J} = \boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}_0} \circ \big(\mathbf{F}^{\boldsymbol{\varepsilon}}_{0;\pm}\big)^{-1}$. This, however, raises two problems. First, in order for $\boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}_0}$ to be analytically computable, the distribution $\mathrm{P}^{\boldsymbol{\varepsilon}}_0$ has to be fully specified (up to location and a global scaling parameter), with closed-form density function $f^{\boldsymbol{\varepsilon}}_0$. Second, the corresponding score function $\mathbf{J} = \boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}_0} \circ \big(\mathbf{F}^{\boldsymbol{\varepsilon}}_{0;\pm}\big)^{-1}$ also involves the center-outward quantile function $(\mathbf{F}^{\boldsymbol{\varepsilon}}_{0;\pm})^{-1}$ for which, except for a few particular cases (spherical distributions), no explicit form is available in the literature. Once $\mathrm{P}^{\boldsymbol{\varepsilon}}_0$ is fully specified, in principle, it can be simulated, and an arbitrarily precise numerical evaluation of $(\mathbf{F}^{\boldsymbol{\varepsilon}}_{0;\pm})^{-1}$ can be obtained, to be plugged into $\mathbf{J}$. This may be computationally heavy, but increasingly efficient algorithms are available in the domain of numerical measure transportation: see, e.g., Mérigot (2011) or Peyré and Cuturi (2019).

Now, choosing a fully specified reference $\mathrm{P}^{\boldsymbol{\varepsilon}}_0$ may be embarrassing—this means, for instance, a skew-t distribution with specified degrees of freedom, shape matrix, and skewness parameter (without loss of generality, location can be taken as $\mathbf{0}$), a multinormal or elliptical distribution with specified radial density and specified (up to a positive global factor) covariance (again, the mean can be taken as $\mathbf{0}$), ... Fortunately, a full specification of $\mathrm{P}^{\boldsymbol{\varepsilon}}_0$ can be relaxed to the specification of a parametric family with parameter $\boldsymbol{\vartheta}$, say, such as the family $\mathcal{P}_{\text{skew } t}$ of all skew-t distributions with location $\mathbf{0}$ (parameters: a shape matrix and a d-tuple of skewness parameters) or the family $\mathcal{P}^{\text{ell}}_{\mathfrak{f}}$ of all elliptical distributions (4.3) with radial density $\mathfrak{f}$ (parameter: a scatter matrix). The unspecified parameter $\boldsymbol{\vartheta}$ of $\mathrm{P}^{\boldsymbol{\varepsilon}}_0$ indeed can be replaced, in the numerical evaluation of $\mathbf{F}^{\boldsymbol{\varepsilon}}_{0;\pm}$, with consistent estimated values provided that the estimator $\hat{\boldsymbol{\vartheta}}$ is measurable with respect to the *order statistic*[15] of the residuals $\mathbf{Z}^{(n)}_i$. Plugging these estimators into the score $\mathbf{J}$—this include

[14]Due to the fact that the density $f^{\star}_{d;\mathfrak{f}}$ of the modulus of a spherical logistic fails to be logistic for $d > 1$, the Wilcoxon test based on $\underset{\sim}{\mathbf{Q}}^{(n)\pm}_{\text{Wilcoxon}}$, however, does not enjoy efficiency under spherical logistic; this is also the case of the elliptical rank tests based on Wilcoxon scores in Hallin and Paindaveine (2002a,b, 2005).

[15]The order statistic of the n-tuple $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ of d-dimensional ($d > 1$) random vectors can be defined as any reordering $\mathbf{Z}_{(1)}, \ldots, \mathbf{Z}_{(n)}$ generating the σ-field of permutation-invariant Borel sets of $\sigma\big(\mathbf{Z}_1, \ldots, \mathbf{Z}_n\big)$; for instance, the one resulting from ordering the observations $\mathbf{Z}_i$ from smallest to largest first component.

the standardization factor and the numerical evaluation of $\mathbf{F}^{\varepsilon}_{0;\pm}$—yields data-driven (order-statistic-driven) scores $\mathbf{J}^{(n)}$.[16] Conditionally on the order statistic, the corresponding test statistic is still distribution-free and its (conditional) critical values yield unconditionally correct size. However, these critical values involve the order statistic: the resulting tests therefore no longer are ranks tests but permutation tests.[17] The theoretical properties, feasibility, and finite-sample performance of this data-driven approach should be explored and numerically assessed—this is, however, beyond the scope of this paper and we leave it for future research.

In view of this, no obvious non-spherical convenient candidate emerges as a reference density in dimension $d > 1$. The center-outward test statistic achieving optimality at the spherical distributions with radial density $\mathfrak{f}$ is $\underset{\sim}{\mathbf{Q}}^{(n)\pm}_{J_\mathfrak{f}}$ with $J_\mathfrak{f}$ as in part *(iii)* of Corollary 5.2.

6.3 Affine invariance and sphericization

Affine invariance (testing) or equivariance (estimation), in "classical multivariate analysis," is often considered an essential and inescapable property. Closer examination, however, reveals that this particular role of affine transformations is intimately related to the affine invariance of Gaussian and elliptical families of distributions. When Gaussian or elliptical assumptions are relaxed, affine transformations are losing this privileged role and the relevance of affine invariance/equivariance properties is much less obvious. We refer to Appendix A.6 for a more detailed discussion of that invariance issue.

7 Some numerical results

A Monte Carlo simulation study is conducted (Sections 7.1–7.2) in order to explore the finite-sample performance of our tests. Results are presented for two-sample location and MANOVA models, and limited to the Wilcoxon score function $J(r) = r$; other choices for J lead to very similar figures, which we therefore do not report.

[16] Similar data-driven scores have been proposed in the univariate case by Dodge and Jurečková (2000).

[17] A *permutation test* is a test enjoying *Neyman α-structure* with respect to the sufficient and complete order statistic.

7.1 Two-sample location, $d = 2$

Consider first the two-sample location problem in dimension $d = 2$. Two independent random samples of size $n_1 = n_2 = n/2$ were generated and the two test statistics $\underset{\sim}{Q}^{(n)\,\text{ell}}_{\text{Wilcoxon}}$ and $\underset{\sim}{Q}^{(n)\pm}_{\text{Wilcoxon}}$ (see Sections 5 and 5.3.1) were computed. The sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$ was used for the computation of $\underset{\sim}{\boldsymbol{\Lambda}}^{(n)\,\text{ell}}_{\text{Wilcoxon}}$ and the elliptical or Mahalanobis ranks and signs.

Rejection frequencies were computed for the following error densities:

(a) a centered bivariate normal distribution with unit variances and correlation $\rho = 1/4$;

(b) a centered bivariate t-distribution with the same scaling matrix as in *(a)* and ν degree of freedom, $\nu = 1$ (Cauchy) and $\nu = 3$;[18]

(c) a mixture, with weights $w_1 = 1/4$ and $w_2 = 3/4$, of two bivariate normal distributions with means $\boldsymbol{\mu}_1 = (3/4, 0)'$ and $\boldsymbol{\mu}_2 = (-1/4, 0)'$ and covariance matrices

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 2/3 \\ 2/3 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & -2/3 \\ -2/3 & 1 \end{pmatrix},$$

respectively;

(d) a mixture, with weights $w_1 = 1/4$ and $w_2 = 3/4$, of two bivariate t_1 (Cauchy) distributions centered at $\boldsymbol{\mu}_1 = (3/4, 0)'$ and $\boldsymbol{\mu}_2 = (-1/4, 0)'$, with the same scaling matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as in *(c)*;

(e) a "U-shaped" mixture, with weights $w_1 = 1/2$, $w_2 = 1/4$, and $w_3 = 1/4$, of three bivariate normal distributions, $\mathcal{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\mathcal{N}_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, and $\mathcal{N}_2(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$ where

$$\boldsymbol{\mu}_1 = (0, 0)', \quad \boldsymbol{\mu}_2 = (-3, 1)', \quad \boldsymbol{\mu}_3 = (3, 1)',$$

and

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1/8 \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1/2 & -1/3 \\ -1/3 & 1/2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_3 = \begin{pmatrix} 1/2 & 1/3 \\ 1/3 & 1/2 \end{pmatrix};$$

(f) an "S-shaped" mixture, with equal weights $w = 1/3$, of three bivariate normal distributions, $\mathcal{N}_2(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_4)$, $\mathcal{N}_2(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_5)$, and $\mathcal{N}_2(\boldsymbol{\mu}_6, \boldsymbol{\Sigma}_4)$ where

$$\boldsymbol{\mu}_4 = (-9/2, -1/2)', \quad \boldsymbol{\mu}_5 = (0, -1/2)', \quad \boldsymbol{\mu}_6 = (9/2, 1)',$$

and

$$\boldsymbol{\Sigma}_4 = \begin{pmatrix} 3/2 & -\sqrt{3/8} \\ -\sqrt{3/8} & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_5 = \begin{pmatrix} 3/2 & \sqrt{3/8} \\ \sqrt{3/8} & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_6 = \begin{pmatrix} 3/2 & -\sqrt{3/8} \\ -\sqrt{3/8} & 1 \end{pmatrix};$$

[18] The bivariate t-distribution with m degrees of freedom and scaling matrix $\mathbf{A}'\mathbf{A}$ is the one defined in Example 2.5 of Fang et al. (2017) as the distribution of a random vector $\boldsymbol{\xi} := \boldsymbol{\mu} + \mathbf{A}'\boldsymbol{\zeta}\sqrt{m}/\sqrt{s}$ where $\boldsymbol{\zeta} \sim \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$ and $s \sim \chi^2_m$, independent of $\boldsymbol{\zeta}$—not to be confused with the elliptical distribution with Student radial density $\mathfrak{f}$.

(g) a skew-t-distribution with ν degrees of freedom, $\nu = 1$ and 3, with skewness parameter $\boldsymbol{\alpha} = (5, -3)'$, scaling matrix $\boldsymbol{\Sigma}_7 = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$, and location $\boldsymbol{\xi} = \mathbf{0}$.

Mixture error densities naturally appear in the context of hidden heterogeneities due, for instance, to omitted covariates; as for asymmetries, they are likely to be the rule rather than the exception. Samples of size 200 from the Gaussian mixtures *(c)*, *(e)*, and *(f)* and the skew-t distribution with 3 degrees of freedom *(g)* are shown in Appendix A.7.1, Figure 9.

To investigate finite-sample performance, a first sample was generated from one of the distributions *(a)*–*(g)*, a second one from the same distribution shifted by the vector $(\delta, \delta)'$ for $\delta \in [0.00, 0.24]$. Three sample sizes $n_1 = n_2 = 50$, 200, and 450 (hence, $n = 100$, 400, and 900) were considered, yielding three groups of curves (from light gray to black, colors in the online version). The regular grids $\mathfrak{G}_n$ for computation of the center-outward ranks and signs are constructed with $n_S = n_R = 10$ for $n = 100$, $n_S = n_R = 20$ for $n = 400$, and $n_S = n_R = 30$ for $n = 900$. Each simulation was replicated $N = 1000$ times and the empirical size and power of the test were computed for $\alpha = 0.05$. The resulting rejection frequencies show the dependence of the power on δ; they are provided in Figures 2—4. For the sake of comparison, we also provide the power of Hotelling's classical two-sample test.

.

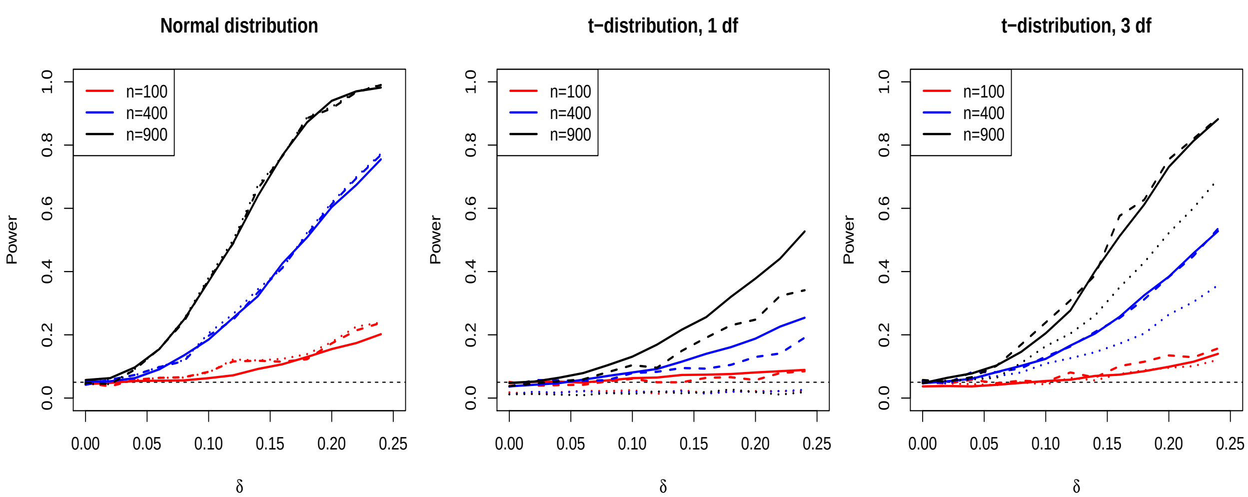

Figure 2: Empirical powers of two-sample location tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon elliptical rank statistic (dashed line), and Hotelling's two-sample test (dotted line), as functions of the shift δ under bivariate normal and elliptical Student (1 and 3 degrees of freedom) error densities; sample sizes $n_1 = n_2 = 50$, 200, and 450.

Figure 2 displays the empirical power curves for the elliptical distributions *(a)* and *(b)*. The results for the normal distribution are very similar for the three tests: rank-based tests (Wilcoxon scores), thus, are no less powerful than the optimal Hotelling test. As expected, Hotelling crashes under the t_1 distribution, while the Wilcoxon elliptical test, although based on the sample covariance matrix, performs surprisingly well (the robustness of ranks offsets infinite variance). The tests based on $\underset{\sim}{Q}^{(n)\,\mathrm{ell}}_{\mathrm{Wilcoxon}}$ and $\underset{\sim}{Q}^{(n)\pm}_{\mathrm{Wilcoxon}}$ both outperform Hotelling also for the t-distribution with 3 degrees of freedom. The conclusion is that center-outward rank tests perform equally well as elliptical rank tests under elliptical densities.

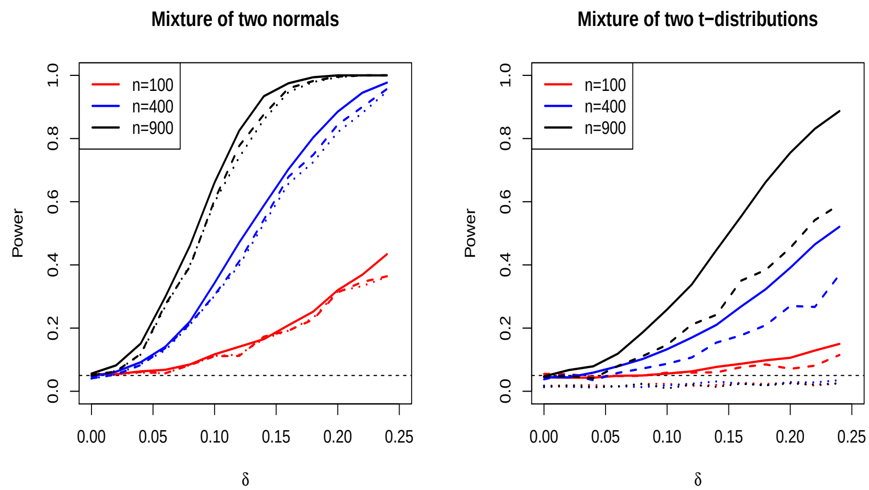

Figure 3: Empirical powers of two-sample location tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon elliptical rank test statistic (dashed line), and Hotelling's two-sample test (dotted line), as functions of the shift δ, for the mixtures of two normal (left panel) and two t_1 error densities (right panel), respectively; sample sizes $n_1 = n_2 = 50$, 200, and 450.

The remaining distributions *(c)*–*(g)* are non-elliptical ones. Results for the mixtures *(c)* and *(d)* are shown in Figure 3. For the mixture *(c)* of two normals, the results obtained for the three tests are still quite similar, but the center-outward rank test based on $\underset{\sim}{Q}^{(n)\pm}_{\mathrm{Wilcoxon}}$, in general, yields the largest power. For the mixture *(d)* of two t_1 (Cauchy) distributions, the Hotelling test fails miserably and the center-outward rank test very clearly outperforms the elliptical rank test for all sample sizes. Figure 4 provides the results for the mixtures *(e)*–*(f)* and the skew-t-distribution *(g)*, respectively. The power curve for the test statistic $\underset{\sim}{Q}^{(n)\pm}_{\mathrm{Wilcoxon}}$ computed from the linearly sphericized residuals (using the sample mean and the sample covariance matrix as estimators of location and scatter) is added as a dot-dashed line. In all these plots, the center-outward rank test statistic leads to the largest power. Note that the linear sphericization of the residuals, which makes the test affine-invariant, may noticeably deteriorate the power (see the discussion in Section 6.3 and Appendix A.6).

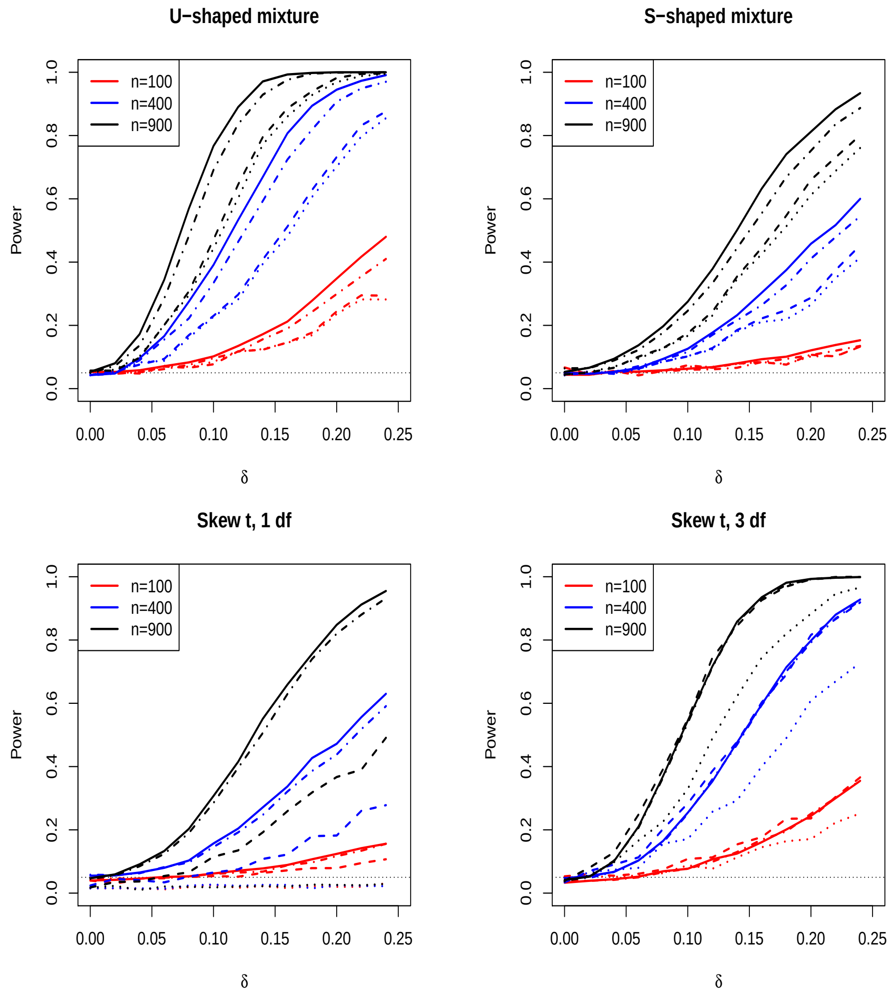

Figure 4: Empirical powers of two-sample location tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon center-outward rank statistic computed from linearly sphericized residuals (dot-dashed line), the Wilcoxon elliptical rank test statistic (dashed line), and Hotelling's two-sample test (dotted line), as functions of the shift δ for the "U-shaped" (upper left panel) and the "S-shaped" (upper right panel) mixtures of three normal error densities, and skew-t error densities with $\nu = 1.1$ (bottom left panel) and $\nu = 3$ (bottom right panel) degrees of freedom, respectively; sample sizes $n_1 = n_2 = 50$, 200, and 450.

7.2 One-way MANOVA, $d = 2$

The performance of center-outward rank tests is very briefly studied here for one-way MANOVA with $K = 3$ groups, still for $d = 2$. Two random samples were generated from the distribution *(a)* (Gaussian) or *(e)* (U-shaped mixture of three Gaussians), as described in Section 7.1, and the third sample was drawn from the same distribution shifted by the vector $(\delta, \delta)'$ for $\delta \in [0.00, 0.24]$. A balanced design with groups of size $n_1 = n_2 = n_3 = 75$ (hence $n = 225$) and $n_1 = n_2 = n_3 = 300$ (hence $n = 900$) was considered. For $n = 225$, the grid $\mathfrak{G}_n$ is constructed with $n_R = n_S = 15$; for $n = 900$, we set $n_R = n_S = 30$. As in

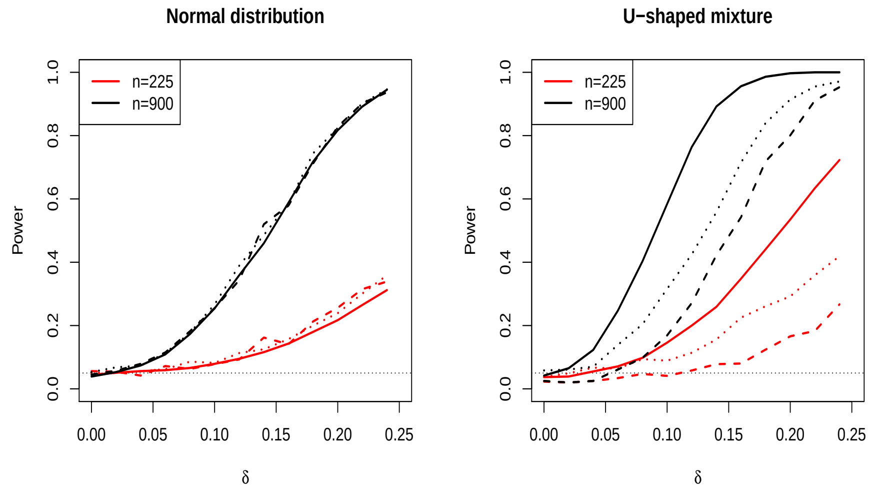

Figure 5: Empirical powers of MANOVA tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon elliptical rank test statistic (dashed line), and Pillai's classical test (dotted line), as functions of the shift δ, for the normal distribution (left panel) and the U-shaped mixture of three normals (right panel); the sample sizes are $n_1 = n_2 = n_3 = 75$ and 300.

Section 7.1, the results are presented for the Wilcoxon scores $J(r) = r$ only—other choices lead to very similar conclusions.

Rejection frequencies are plotted in Figure 5 for the center-outward rank test based on $\underset{\sim}{Q}^{(n)\pm}_{\text{Wilcoxon}}$ (solid line), the elliptical rank test statistic $\underset{\sim}{Q}^{(n)\,\text{ell}}_{\text{Wilcoxon}}$ (dashed line), and the Pillai trace test based on an approximate F-distribution (dotted line). Under normal density, the three tests perform very similarly. For the non-elliptical mixture distribution, however, the center-outward rank test achieves sizeably larger power than the other two. Further simulations yielding, in dimension $d = 6$, similar conclusions, are provided in Appendix A.7.

7.3 An empirical illustration

The practical value of the center-outward rank tests developed in the previous sections is illustrated with the following archeological application where classical methods fail to detect any treatment effect. The data consist of $n = 126$ measurements of MgO (Magnesium oxide), P_2O_5 (Phosphorus pentoxide), CoO (Cobalt monoxide), and Sb_2O_3 (Antimony trioxide) (dimension $d = 4$, thus) in natron glass vessels excavated from three Syro-Palestinian sites in present-day Israel: Apollonia ($n_1 = 54$ observations), Bet Eli'ezer ($n_2 = 17$ observations), and Egypt ($n_3 = 55$ observations); a fourth site only has two observations and was dropped from the analysis. This dataset has been originally analyzed by Phelps et al.

(2016) with the objective of detecting possible differences among the three sites. Bivariate plots of these four variables are shown in Figure 6, where one can observe that the marginal distributions of CoO, and Sb_2O_3 exhibit heavy tails and are very far from normal, and their joint distribution far from elliptically symmetric. A traditional (pseudo-Gaussian) test here is Pillai's trace test[19] reducing, in the two-sample case, to Hotelling's classical T-square test.

First, all the two-dimensional data subsets corresponding to the bivariate plots in Figure 6 were analyzed (six bivariate MANOVA models, thus). Pillai's test yields non-significant p-values for all combinations, see Table 2. But the center-outward tests we are proposing in this paper do detect significant differences between the three groups whenever the variable CoO is included in the analysis. Two versions[20] of the center-outward ranks and signs are considered in Table 2 below (c-o tests I and II, respectively).

Inspection of Table 2 reveals that, unlike Pillai's trace, the Wilcoxon center-outward rank tests (c-o I and II) reject the null hypothesis at significance level $\alpha = 0.05$. As for the Wilcoxon tests based on elliptical ranks (based on the sample covariance function), they yield highly non-significant p-values for all couples of variables; the corresponding results are not presented here. Next, the MANOVA comparison is conducted for the full 4-dimensional dataset. Pillai's test p-value is 0.1553: no difference detected among the three groups, thus, at level $\alpha \leq 0.15$. In sharp contrast, the Wilcoxon center-outward rank test (with $n_R = 7$ and $n_S = 18$) yields a p-value 10^{-15}, which is highly significant. The elliptical Wilcoxon rank test (based on the sample covariance matrix), on the other hand, with p-value 0.5827, also fails to detect anything at any level $\alpha \leq 0.5$.

This, according to archeological sources, might lead to revising some of the conclusions made by Phelps et al. (2016) on Middle-East economic exchanges between Egypt and Syro-Palestine in the Byzantine-Islamic transition period.

8 Conclusion and perspectives

Classical multivariate analysis methods, which are daily practice in a number of applied domains, remain deeply marked by Gaussian and elliptical assumptions. In particular,

[19] Alternatives are Wilks' Lambda, the Lawley-Hotelling Trace, and Roy's largest root tests. In the two-sample case, they all coincide; else, they are asymptotically equivalent.

[20] These two versions correspond to two choices of the grid $\mathfrak{G}_n$, with either $n_S = 7$ and $n_R = 18$ or $n_S = 18$ and $n_R = 7$—see Section 2.2 for an explanation.

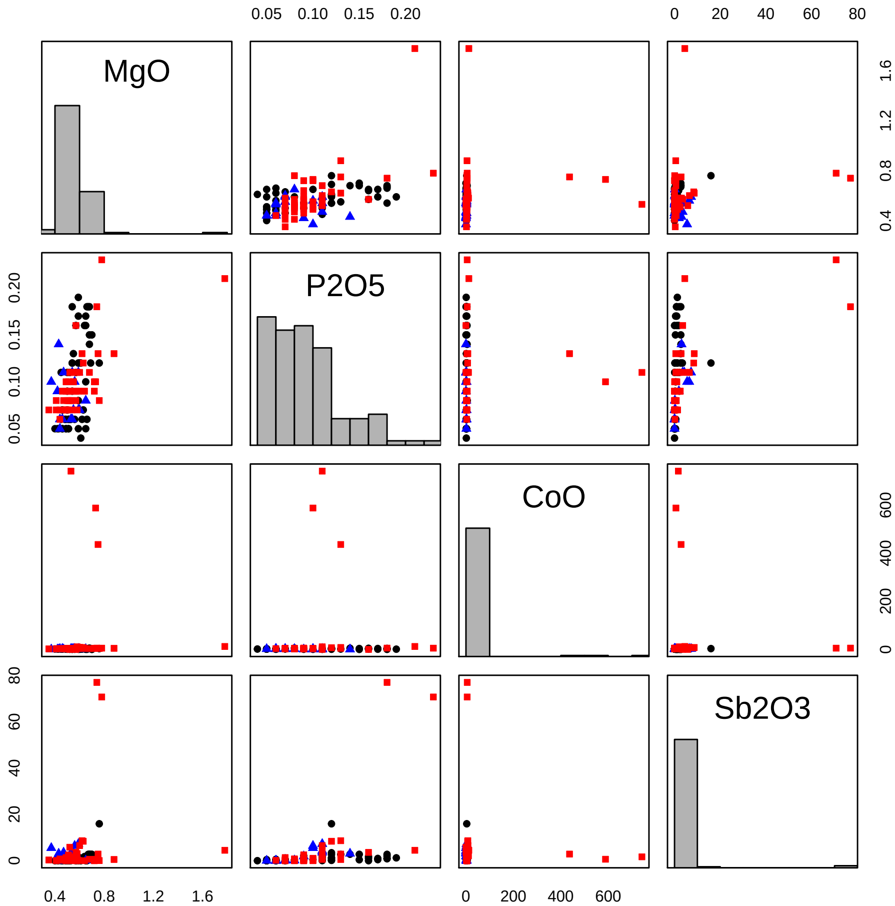

Figure 6: The content of MgO, P_2O_5, CoO, and Sb_2O_3 in natron glass vessels from Appolonia (circles), Bet Eli'ezer (triangles), and Egypt (squares).

no distribution-free approach is available so far for hypothesis testing in multiple-output regression models, which include the fundamental two-sample and MANOVA models—except for the elliptical or Mahalanobis rank tests developed in Hallin and Paindaveine (2005) which, however, require the strong assumption of elliptic symmetry, an assumption which is unlikely to hold in most applications. Based on the recent concept of center-outward ranks and signs, this paper proposes the first efficient fully distribution-free tests of the hypothesis of no treatment effect in that multiple-output context, thereby extending to the multivariate case the classical Hájek approach to univariate rank-based inference (Hájek and Šidák, 1967). Simulations and an empirical example demonstrate the excellent performance of the method. This lays the theoretical bases (asymptotic representation and asymptotic normality results for linear center-outward rank statistics) and theoretical guidelines (*Hájek projection* of LAN central sequences) for further developments.

		Pillai's test	c-o test I	c-o test II
MgO	P_2O_5	0.3547	0.3817	0.0946
MgO	CoO	0.1217	0.0000	0.0000
MgO	Sb_2O_3	0.2268	0.1865	0.3236
P_2O_5	CoO	0.1491	0.0000	0.0000
P_2O_5	Sb_2O_3	0.1957	0.0561	0.3004
CoO	Sb_2O_3	0.1453	0.0000	0.0000

Table 2: p-values for the bivariate MANOVA Pillai trace and Wilcoxon center-outward rank tests based on $n_R = 7$, $n_S = 18$ (c-o test I) and $n_R = 18$, $n_S = 7$ (c-o test II), respectively.

References

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families.* Cambridge University Press, New York.

Beirlant, J., Buitendag, S., del Barrio, E., and Hallin, M. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance, Math. & Econ.*, 95:79–100.

Boeckel, M., Spokoiny, V., and Suvorikova, A. L. (2018). Multivariate Brenier cumulative distribution functions and their application to non-parametric testing. *arXiv:1809.04090v1.*

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. of Math. Statist.*, 29(4):972 – 994.

Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks, and signs. *Ann. Statist.*, 45:223–256.

Cuesta-Albertos, J., Rüschendorf, L., and Tuero-Diaz, A. (1993). Optimal coupling of multivariate distributions and stochastic processes. *J. Multivariate Anal.*, 46:335–361.

De Valk, C. and Segers, J. (2018). Stability and tail limits of transport-based quantile contours. *arXiv:1811.12061.*

Deb, N., Bhattacharya, B. B., and Sen, B. (2021). Efficiency lower bounds for distribution-free Hotelling-type two-sample tests based on optimal transport. *arXiv:2104.01986.*

Deb, N. and Sen, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *JASA*, to appear.

del Barrio, E., González-Sanz, A., and Hallin, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *Journal of Multivariate Analysis*, 180(C):S0047259X20302529.

Dodge, Y. and Jurečková, J. (2000). *Adaptive Regression*. Springer, New York.

Dua, D. and Graff, C. (2017). UCI machine learning repository.

Dudley, R. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, California.

Fang, K., Kotz, S., and Ng, K. (2017). *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC.

Figalli, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlin. Anal.: Theory, Methods & Appl.*, 177:413–421.

Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics*, 47:551–579.

Ghosal, P. and Sen, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. *arXiv:1905.05340*.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Test*. Academic Press, New York.

Hallin, M. (1994). On the Pitman non-admissibility of correlogram-based methods. *Journal of Time Series Analysis*, 15(6):607–611.

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

Hallin, M., del Barrio, T., Cuesta-Albertos, J., and Matrán, C. (2021a). On distribution and quantile functions, ranks, and signs in $\mathbb{R}^d$: a measure transportation approach. *Ann. Statist.*, 49:1139–1165.

Hallin, M., Hlubinka, D., and Šárka Hudecová (2020a). Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. *arXiv:2007.15496*.

Hallin, M., Ingenbleek, J. F., and Puri, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.*, 13:1156–1181.

Hallin, M., Ingenbleek, J. F., and Puri, M. L. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *J. Multivariate Anal.*, 30:34–71.

Hallin, M., La Vecchia, D., and Liu, H. (2020b). Rank-based testing for semiparametric VAR models: a measure transportation approach. *arXiv:2011.06062*.

Hallin, M., La Vecchia, D., and Liu, H. (2021b). Center-outward R-estimation for semi-parametric VARMA models. *JASA*, to appear.

Hallin, M., Mordant, G., and Segers, J. (2020c). Multivariate goodness-of-fit tests based on Wasserstein distance. *arXiv:2003.06684*.

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

Hallin, M. and Paindaveine, D. (2002b). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.*, 30:1103–1133.

Hallin, M. and Paindaveine, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.*, 32(6):2642–2678.

Hallin, M. and Paindaveine, D. (2005). Affine-invariant aligned rank tests for multivariate general linear models with VARMA errors. *J. Multivariate Anal.*, 93:122–163.

Hallin, M. and Puri, M. L. (1994). Aligned rank tests for linear models with autocorrelated errors. *J. Multivariate Anal.*, 50:175–237.

He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.*, 25:495–504.

Judd, K. L. (1998). *Numerical Methods in Economics.* MIT Press, Cambridge, MA.

Koul, H. L. and Saleh, A. K. M. (1993). R-estimation of the parameters of autoregressive AR(p) models. *Ann. Statist.*, 21:534–551.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, N. Y.

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

Liu, R. Y. (1992). Data depth and multivariate rank tests. In Dodge, Y., editor, *L^1 Statistics and Related Methods*, pages 279–294. North-Holland, Amsterdam.

Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, 88:257–260.

McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–324.

Mérigot, Q. (2011). A multiscale approach to optimal transport. In *Computer Graphics Forum*, volume 30, pages 1583–1592. Wiley Online Library.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NSF Conference Series in Applied Mathematics. SIAM, Philadelphia, PA.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics and Probab. Letters*, 1:327–332.

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.*, 26:319–343.

Oja, H. (2010). *Multivariate Nonparametric Methods with R: an approach based on spatial signs and ranks*. Springer, New York.

Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Foundations and Trends® in Machine Learning*, 11(5–6):355–607.

Phelps, M., Freestone, I. C., Gorin-Rosen, Y., and Gratuze, B. (2016). Natron glass production and supply in the late antique and early medieval Near East: The effect of the Byzantine-Islamic transition. *Journal of Archaeological Science*, pages 57–71.

Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

Puri, M. L. and Sen, P. K. (1985). *Nonparametric Methods in General Linear Models*. Wiley, New York.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Randles, R. and Wolfe, D. (1979). *Introduction to the Theory of Nonparametric Statistics*. Wiley, New York.

Rockafellar, R. T. (1966). Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17:1497–510.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York.

Shi, H., Drton, M., and Han, F. (2021). Distribution-free consistent independence tests via center-outward ranks and signs. *JASA*, to appear.

Shi, H., Hallin, M., Drton, M., and Han, F. (2020). Rate-optimality of consistent distribution-free tests of independence based on center-outward ranks and signs. *arXiv:2007.02186*.

Shorack, G. R. (2000). *Probability for Statisticians*. Springer, New York.

Street, W. N., Wolberg, W. H., and Mangasarian, O. L. (1993). Nuclear feature extraction for breast tumor diagnosis. In Acharya, R. S. and Goldgof, D. B., editors, *Biomedical Image Processing and Biomedical Visualization*, pages 86–870. International Society for Optics and Photonics, SPIE.

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

Zuo, Y. and He, X. (2006). On limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.*, 34:2879–2896.

Appendix

A.1 The power of of pseudo-Gaussian tests away from the Gaussian

This section provides evidence of the deterioration (highligted in the Introduction) of the performance of pseudo-Gaussian tests away from Gaussian distributions.

Figure 7 illustrates the deterioration of pseudo-Gaussian tests performance in the case of Hotelling's bivariate two-sample test under skew-Gaussian distributions with location $\mathbf{0}$, correlation ρ, and skewness parameter $\boldsymbol{\lambda} = (a, a)^\top$. Recall that bivariate skew-normal distributions (see pp. 128–131 in Azzalini and Capitanio (2014) for details) are indexed by a triplet $(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$ where $\boldsymbol{\xi}$ is a location, $\boldsymbol{\Psi}$ a shape matrix, and $\boldsymbol{\lambda}$ a skewness parameter. Setting

$$\boldsymbol{\xi} = (0, 0)^\top, \qquad \boldsymbol{\Psi} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\lambda} = (a, a)^\top, \quad \rho \in (-1, 1), \quad a \in \mathbb{R},$$

refer to this distribution as $\mathsf{SN}_{\rho,a}$. For $a = 0$, $\mathsf{SN}_{\rho,0}$ coincides with the bivariate centered Gaussian distribution with covariance $\boldsymbol{\Psi}$.

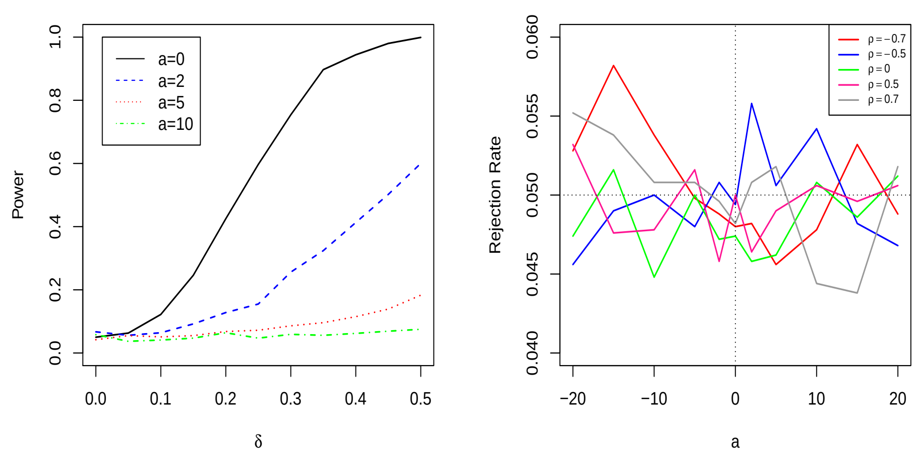

Figure 7: Rejection frequencies of Hotelling's two-sample test (nominal size $\alpha = 5\%$) for sample sizes $n_1 = n_2 = 100$ under the skew-normal distribution $\mathsf{SN}_{\rho,a}$. Left panel: the empirical power against shift $(\delta, \delta)^\top$ for $\rho = 0$ and various values of the skewness parameter a. Right panel: empirical sizes for various values of (ρ, a).

Although skew-Gaussian distributions satisfy the conditions for the asymptotic validity of Hotelling's pseudo-Gaussian test, the left-hand panel of Figure 7 shows the rapid deterioration of power as the skewness parameter a increases (correlation $\rho = 0$) while the right-hand panel highlights the erraticism of the size of the same Hotelling test across skewness and correlation parameter values despite a relatively large sample of size 200.

Skewness is not the only feature damaging Hotelling's power, though. The settings in Figure 8 are the same as in the left-hand panel of Figure 7, now for the following "increasingly non-Gaussian" distributions:

(a) a "U-shaped" mixture of three bivariate normal distributions, $\mathcal{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\mathcal{N}_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, and $\mathcal{N}_2(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$ with weights $1 - 2w$, w, and w, respectively,

$$\boldsymbol{\mu}_1 = (0,0)^\top, \quad \boldsymbol{\mu}_2 = (-3,1)^\top, \quad \boldsymbol{\mu}_3 = (3,1)^\top,$$

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{8} \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{2} \end{pmatrix}, \quad \boldsymbol{\Sigma}_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} \end{pmatrix},$$

and $w \in [0.00, 0.45]$;

(b) an "S-shaped" mixture of three bivariate normal distributions, $\mathcal{N}_2(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_4)$, $\mathcal{N}_2(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_5)$, and $\mathcal{N}(\boldsymbol{\mu}_6, \boldsymbol{\Sigma}_4)$ with weights w, $1 - 2w$, and w, respectively,

$$\boldsymbol{\mu}_4 = (-9/2, -1/2)^\top, \quad \boldsymbol{\mu}_5 = (0, -1/2)^\top, \quad \boldsymbol{\mu}_6 = (9/2, 1)^\top,$$

$$\boldsymbol{\Sigma}_4 = \begin{pmatrix} \sigma_1^2 & -\rho\sigma_1 \\ -\rho\sigma_1 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_5 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix}$$

where $\sigma_1^2 = 3/2$ and $\rho = 1/2$, and $w \in [0.00, 0.45]$;

(c) a bivariate distribution with standard Gaussian marginals and joint distribution modelled via the Clayton copula with parameter t, $t \in \{0, 1/2, 1, 2, 5, 10\}$;

(d) a bivariate t-distribution with ν degrees of freedom for $\nu \in \{\infty, 5, 3, 2, 1\}$ (where $\nu = \infty$ corresponds to the standard bivariate Gaussian).

The first sample ($n_1 = 100$) was drawn from one of the distributions (a)–(d), the second one ($n_2 = 100$) from the same distribution shifted by $(\delta, \delta)^\top$ for $\delta \in [0.00, 0.24]$. The rejection frequencies were computed from 1000 replications for $\alpha = 0.05$. Figure 8 illustrates how the rejection frequencies, for given δ, deteriorate as the underlying distributions move away from the Gaussian.

A.2 Proof of Proposition 2.2.

Starting with *(ii)*, note that, for any $(\mathbf{z}_1, \dots, \mathbf{z}_n) \in \mathbb{R}^{nd}$, $(\mathbf{u}_1, \dots, \mathbf{u}_n) \in \mathbb{R}^{nd}$, and $\boldsymbol{\mu} \in \mathbb{R}^d$, denoting by π a permutation of $\{1, \dots, n\}$,

$$\sum_{i=1}^{n} \|\boldsymbol{\mu} + \mathbf{z}_i - \mathbf{u}_{\pi(i)}\|^2 - \sum_{i=1}^{n} \|\mathbf{z}_i - \mathbf{u}_{\pi(i)}\|^2 = n\boldsymbol{\mu}'\boldsymbol{\mu} + 2\boldsymbol{\mu}' \sum_{i=1}^{n} \mathbf{z}_i - 2\boldsymbol{\mu}' \sum_{i=1}^{n} \mathbf{u}_i$$

does not depend on π; the optimal pairing between the $\boldsymbol{\mu} + \mathbf{z}_i$'s and the $\mathbf{u}_i$'s thus does not depend on $\boldsymbol{\mu}$, so that $\mathbf{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{Z};(n)}(\boldsymbol{\mu} + \mathbf{Z}_i) = \mathbf{F}_{\pm}^{\mathbf{Z};(n)}(\mathbf{Z}_i)$ for all i (with $\mathbf{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{Z};(n)}$ and $\mathbf{F}_{\pm}^{\mathbf{Z};(n)}$ con-

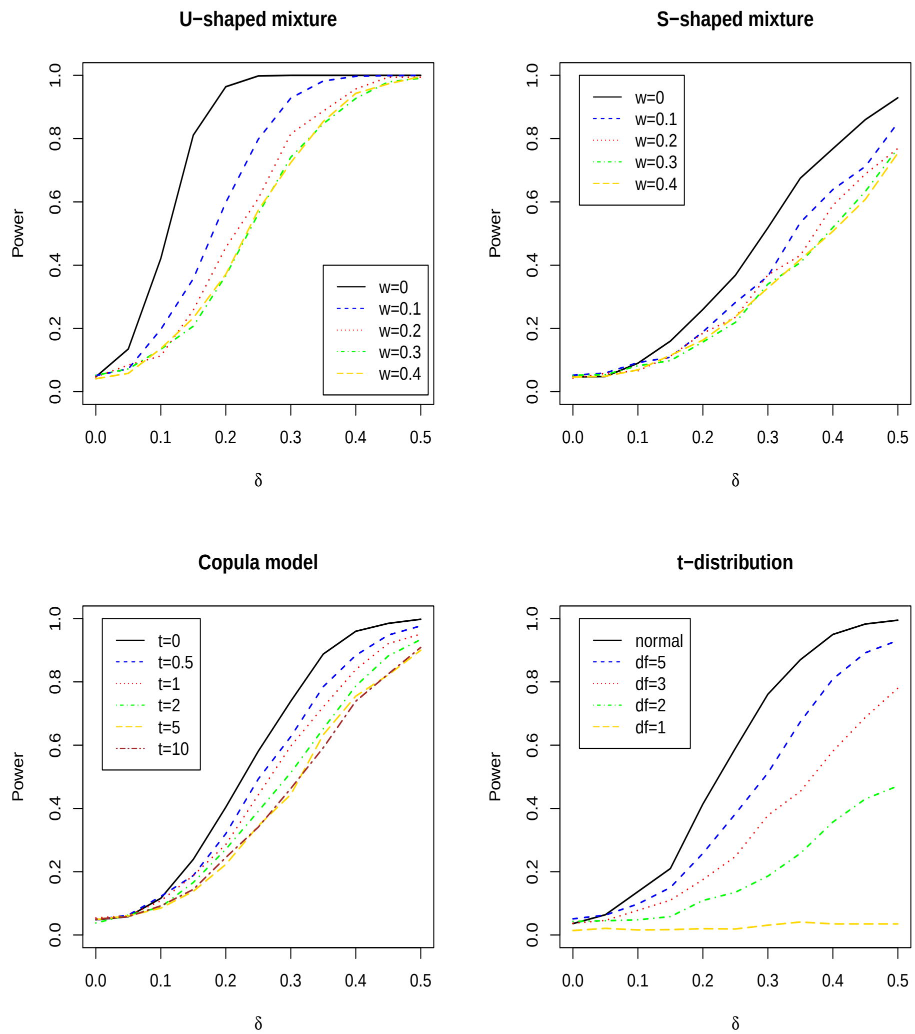

Figure 8: Rejection rates of the two-sample Hotelling test when the second sample is shifted by $(\delta, \delta)^\top$ for samples of sizes $n_1 = n_2 = 100$ from the U-shaped mixture (a)(top left panel), the S-shaped mixture (b) (top right panel), the Clayton copula (c) (bottom left panel), and the bivariate t distribution (d) (bottom right panel).

structed from the same grid $\mathfrak{G}_n$). As for $\mathbf{F}_{\pm}^{\mathbf{OZ};(n)}(\mathbf{OZ}_i)$ computed from $\mathbf{O}\mathfrak{G}_n$ and $\mathbf{OF}_{\pm}^{\mathbf{Z};(n)}(\mathbf{Z}_i)$ computed from $\mathfrak{G}_n$, they obviously coincide since the Euclidean distances on which they are based coincide. Part *(iii)* of the proposition is an immediate consequence.

Turning to *(i)*, note that $\mathbf{F}_{\pm}$, as the gradient of a convex function, enjoys (see, e.g., Rockafellar (1966)) *cyclical monotonicity*: for any finite collection of points $\mathbf{z}_1, \ldots, \mathbf{z}_k \in \mathbb{R}^{nd}$, it holds that

$$\langle \mathbf{F}_{\pm}(\mathbf{z}_1), \mathbf{z}_2 - \mathbf{z}_1 \rangle + \langle \mathbf{F}_{\pm}(\mathbf{z}_2), \mathbf{z}_3 - \mathbf{z}_2 \rangle + \ldots + \langle \mathbf{F}_{\pm}(\mathbf{z}_k), \mathbf{z}_1 - \mathbf{z}_k \rangle \leq 0.$$

Equivalently, considering the grid $\mathfrak{G}_k := \{\mathbf{F}_{\pm}(\mathbf{z}_1), \ldots, \mathbf{F}_{\pm}(\mathbf{z}_k)\}$, any k-tuple of the form $(\mathbf{z}_i, \mathbf{F}_{\pm}(\mathbf{z}_i))$, $i = 1, \ldots, k$ constitutes an optimal coupling minimizing

$$S_{\mathbf{z}}^{(k)} := \sum_{i=1}^{k} \|\mathbf{F}_{\pm}(\mathbf{z}_i) - \mathbf{z}_{\pi(i)}\|^2$$

over the $k!$ permutations π of $\{1, \ldots, k\}$: denoting by $\mathbf{F}_{\pm}^{\mathbf{z};(k)}$ the minimizer of $S_{\mathbf{z}}^{(k)}$, thus,

$$\mathbf{F}_{\pm}^{\mathbf{z};(k)}(\mathbf{z}_i) = \mathbf{F}_{\pm}(\mathbf{z}_i), \quad i = 1, \ldots, k \text{ for any } k. \tag{A1}$$

Now, for fixed k, *(ii)* applies, so that, similar to (2.1),

$$\mathbf{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{Oz};(k)}(\boldsymbol{\mu} + \mathbf{Oz}_i) = \mathbf{OF}_{\pm}^{\mathbf{z};(k)}(\mathbf{z}_i). \tag{A2}$$

In view of (A1) (for $\mathbf{F}_{\pm} = \mathbf{F}_{\pm}^{\mathbf{Z}}$), however,

$$\mathbf{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{Oz};(k)}(\boldsymbol{\mu} + \mathbf{Oz}_i) = \mathbf{F}_{\pm}^{\boldsymbol{\mu}+\mathbf{OZ}}(\boldsymbol{\mu} + \mathbf{Oz}_i) \quad \text{and} \quad \mathbf{F}_{\pm}^{\mathbf{z};(k)}(\mathbf{z}_i) = \mathbf{F}_{\pm}^{\mathbf{Z}}(\mathbf{z}_i). \tag{A3}$$

The result follows from piecing together (A2) and (A3). □

A.3 Proof of Proposition 3.1

We throughout write $\mathbf{Z}_i$ for $\mathbf{Z}_i^{(n)}$. First consider Part *(i)* of the proposition. We have

$$\underset{\sim}{\mathbf{T}}_a^{(n)} - \mathbf{T}^{(n)} = \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1/2} \sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})[\mathbf{J}(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i)) - \mathbf{J}(\mathbf{F}_{\pm}(\mathbf{Z}_i))].$$

Let $\mathbf{a}_i^{(n)} := \mathbf{a}(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i), \mathbf{F}_{\pm}(\mathbf{Z}_i)) := \mathbf{J}(\mathbf{F}_{\pm}^{(n)}(\mathbf{Z}_i)) - \mathbf{J}(\mathbf{F}_{\pm}(\mathbf{Z}_i))$. Then,

$$\|\underset{\sim}{\mathbf{T}}_a^{(n)} - \mathbf{T}^{(n)}\|^2 = \Big(\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2\Big)^{-1} \Bigg[\sum_{i=1}^{n}(c_i^{(n)} - \bar{c}^{(n)})^2 \|\mathbf{a}_i^{(n)}\|^2 + \sum_{i \neq j}(c_i^{(n)} - \bar{c}^{(n)})(c_j^{(n)} - \bar{c}^{(n)})\mathbf{a}_i^{(n)\prime}\mathbf{a}_j^{(n)}\Bigg].$$

Since $\mathrm{E}\|\mathbf{a}_i^{(n)}\|^2 = \mathrm{E}\|\mathbf{a}_1^{(n)}\|^2$ and $\mathrm{E}\mathbf{a}_i^{(n)\prime}\mathbf{a}_j^{(n)} = \mathrm{E}\mathbf{a}_1^{(n)\prime}\mathbf{a}_2^{(n)}$, we get

$$\mathrm{E}\big\|\underset{\sim}{\mathbf{T}}_a^{(n)} - \mathbf{T}^{(n)}\big\|^2 = \mathrm{E}\big\|\mathbf{a}_1^{(n)}\big\|^2 - \mathrm{E}\mathbf{a}_1^{(n)\prime}\mathbf{a}_2^{(n)} \leq 2\mathrm{E}\big\|\mathbf{a}_1^{(n)}\big\|^2.$$

Hence, it only remains to show that

$$\begin{aligned}\mathrm{E}\big\|\mathbf{a}_1^{(n)}\big\|^2 &= \mathrm{E}\left\|\mathbf{J}(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)) - \mathbf{J}(\mathbf{F}_\pm(\mathbf{Z}_1))\right\|^2 \\ &= \mathrm{E}\big\|\boldsymbol{\zeta}_n - \boldsymbol{\zeta}\big\|^2 \to 0\end{aligned}$$

with $\boldsymbol{\zeta}_n := \mathbf{J}(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1))$ and $\boldsymbol{\zeta} := \mathbf{J}(\mathbf{F}_\pm(\mathbf{Z}_1))$. It follows from the Glivenko-Cantelli theorem in Hallin et al. (2021a), the continuity of $\mathbf{J}$, and the continuity over $\mathbb{R}\setminus\{0\}$ of $\mathbf{x} \mapsto \mathbf{x}/\big\|\mathbf{x}\big\|$ that $\boldsymbol{\zeta}_n \to \boldsymbol{\zeta}$ a.s. Furthermore, Assumption 3.1 implies that

$$\mathrm{E}\big\|\boldsymbol{\zeta}_n\big\|^2 = \mathrm{tr}\,\mathrm{E}\mathbf{J}(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1))\mathbf{J}'(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)) \to \mathrm{tr}\,\mathrm{E}\mathbf{J}(\mathbf{F}_\pm(\mathbf{Z}_1))\mathbf{J}'(\mathbf{F}_\pm(\mathbf{Z}_1)) = \mathrm{E}\big\|\boldsymbol{\zeta}\big\|^2.$$

It follows (see, for instance, part *(iv)* of Theorem 5.7 in (Shorack, 2000, Chapter 3)) that $\mathrm{E}\big\|\boldsymbol{\zeta}_n - \boldsymbol{\zeta}\big\|^2 \to 0$. This concludes the proof for Part *(i)* of the proposition.

Turning to Part *(ii)*, put

$$\mathbf{b}_i^{(n)} := \mathbf{J}(\mathbf{F}_\pm(\mathbf{Z}_i)) - \mathrm{E}\left[\mathbf{J}(\mathbf{F}_\pm(\mathbf{Z}_i))\middle|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_i)\right],$$

and let us show that $\mathrm{E}\big\|\mathbf{b}_1^{(n)}\big\|^2 = o(1)$. Since $\|\boldsymbol{\zeta}_n - \boldsymbol{\zeta}\|$ tends a.s. to zero, it follows from the Egorov theorem (see, e.g., Theorem 7.5.1 in Dudley (1989)) that, for any $\varepsilon > 0$, there is a set $A \subset \Omega$ such that

$$\mathrm{P}(A) > 1 - \varepsilon \qquad \text{and} \qquad \sup_{\omega\in A}\|\boldsymbol{\zeta}_n(\omega) - \boldsymbol{\zeta}(\omega)\| \to 0.$$

Denoting by A^c the complement of A in Ω, we have

$$\begin{aligned}\mathrm{E}\big\|\mathbf{b}_1^{(n)}\big\|^2 &= \mathrm{E}\big\|\boldsymbol{\zeta} - \mathrm{E}[\boldsymbol{\zeta}|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)]\big\|^2 \\ &= \mathrm{E}\big\|\boldsymbol{\zeta}\mathbf{1}_A + \boldsymbol{\zeta}\mathbf{1}_{A^c} - \mathrm{E}[\boldsymbol{\zeta}\mathbf{1}_A + \boldsymbol{\zeta}\mathbf{1}_{A^c}|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)]\big\|^2 \\ &\leq 3\mathrm{E}\big\|\boldsymbol{\zeta}\mathbf{1}_A - \mathrm{E}[\boldsymbol{\zeta}\mathbf{1}_A|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)]\big\|^2 + 3\mathrm{E}\big\|\boldsymbol{\zeta}\mathbf{1}_{A^c}\big\|^2 \\ &\qquad\qquad + 3\mathrm{E}\left\|\mathrm{E}[\boldsymbol{\zeta}\mathbf{1}_{A^c}|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)]\right\|^2 \\ &=: 3\left(I_1^{(n)} + I_2 + I_3^{(n)}\right), \quad \text{say.}\end{aligned}$$

In view of the square-integrability of $\boldsymbol{\zeta}$, I_2 can be made arbitrarily small as $\varepsilon \to 0$. As for $I_3^{(n)}$, we have

$$I_3^{(n)} = \mathrm{E}\left\|\mathrm{E}[\boldsymbol{\zeta}\mathbf{1}_{A^c}|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)]\right\|^2 \leq \mathrm{E}\Big(\mathrm{E}\big[\|\boldsymbol{\zeta}\mathbf{1}_{A^c}\|^2|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)\big]\Big) = \mathrm{E}\big\|\boldsymbol{\zeta}\mathbf{1}_{A^c}\big\|^2$$

where $\mathrm{P}(A^c) \leq \varepsilon$, so that $I_3^{(n)}$ also is arbitrarily small as $\varepsilon \to 0$.

It remains to prove that $I_1^{(n)} \to 0$ as $n \to \infty$. Recall that $\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)$, with probability $(1 - n_0/n)$ tending to one, is a point of the regular grid $\mathfrak{G}_n$ of $n - (n_0 - 1)\mathbf{1}_{n_0>0}$ points in the unit ball used in the construction of $\mathbf{F}_\pm^{(n)}$. Moreover, for any $\mathfrak{g} \in \mathfrak{G}_n \setminus \{\mathbf{0}\}$, we have $\mathrm{P}[\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1) = \mathfrak{g}] = 1/n$. Define

$$B_\mathfrak{g}^{(n)} := \{\omega : \mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)(\omega) = \mathfrak{g}\}.$$

Clearly, $\{B_\mathfrak{g}^{(n)}, \mathfrak{g} \in \mathfrak{G}_n\}$ constitutes a disjoint partition of Ω and $\mathrm{P}(B_\mathfrak{g}^{(n)}) \to 0$ uniformly in $\mathfrak{g} \in \mathfrak{G}_n$ as $n \to \infty$. Then,

$$\begin{aligned} I_1^{(n)} &= \mathrm{E}\Big\| \sum_{\mathfrak{g}\in\mathfrak{G}_n} \big(\boldsymbol{\zeta}\mathbf{1}_A\mathbf{1}_{B_\mathfrak{g}^{(n)}} - \mathrm{E}[\boldsymbol{\zeta}\mathbf{1}_A|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)]\mathbf{1}_{B_\mathfrak{g}^{(n)}}\big)\Big\|^2 \\ &= \mathrm{E}\sum_{\mathfrak{g}\in\mathfrak{G}_n} \Big\|\boldsymbol{\zeta}\mathbf{1}_A\mathbf{1}_{B_\mathfrak{g}^{(n)}} - \mathrm{E}[\boldsymbol{\zeta}\mathbf{1}_A|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)]\mathbf{1}_{B_\mathfrak{g}^{(n)}}\Big\|^2 \end{aligned}$$

where the latter equality follows form the fact that $\mathbf{1}_{B_\mathfrak{g}^{(n)}}\mathbf{1}_{B_\mathfrak{h}^{(n)}} = \mathbf{1}_{(\mathfrak{g}=\mathfrak{h})}$. Since $B_\mathfrak{g}^{(n)}$ is an atom of $\sigma\big(\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)\big)$, the latter conditional expectation is a constant on $B_\mathfrak{g}^{(n)}$, namely

$$\mathrm{E}\big[\boldsymbol{\zeta}\mathbf{1}_A\big|\mathbf{F}_\pm^{(n)}(\mathbf{Z}_1)\big]\mathbf{1}_{B_\mathfrak{g}^{(n)}} = \frac{\mathbf{1}_{B_\mathfrak{g}^{(n)}}}{\mathrm{P}(B_\mathfrak{g}^{(n)})}\int_{\eta\in B_\mathfrak{g}^{(n)}} \boldsymbol{\zeta}(\eta)\mathbf{1}_A(\eta)\mathrm{dP}(\eta).$$

Hence,

$$\begin{aligned} I_1^{(n)} &= \sum_{\mathfrak{g}\in\mathfrak{G}_n}\int_\Omega \Big\|\mathbf{1}_{B_\mathfrak{g}^{(n)}}(\omega)\int_{\eta\in B_\mathfrak{g}^{(n)}} [\boldsymbol{\zeta}(\omega)\mathbf{1}_A(\omega) - \boldsymbol{\zeta}(\eta)\mathbf{1}_A(\eta)]\ \frac{\mathrm{dP}(\eta)}{\mathrm{P}(B_\mathfrak{g}^{(n)})}\Big\|^2 \mathrm{dP}(\omega) \\ &= \sum_{\mathfrak{g}\in\mathfrak{G}_n}\int_\Omega \Big\|\mathbf{1}_{B_\mathfrak{g}^{(n)}}(\omega)\int_{\eta\in B_\mathfrak{g}^{(n)}} \big[\big(\boldsymbol{\zeta}(\omega) - \boldsymbol{\zeta}_n(\omega)\big)\mathbf{1}_A(\omega) \\ &\qquad\qquad + \big(\boldsymbol{\zeta}_n(\eta) - \boldsymbol{\zeta}(\eta)\big)\mathbf{1}_A(\eta)\big]\ \frac{\mathrm{dP}(\eta)}{\mathrm{P}(B_\mathfrak{g}^{(n)})}\Big\|^2 \mathrm{dP}(\omega) \end{aligned}$$

since $\boldsymbol{\zeta}_n(\omega)\mathbf{1}_A(\omega)\mathbf{1}_{B_\mathfrak{g}^{(n)}}(\omega) = \mathbf{J}(\mathfrak{g}) = \boldsymbol{\zeta}_n(\eta)\mathbf{1}_A(\eta)\mathbf{1}_{B_\mathfrak{g}^{(n)}}(\eta)$ on $A \cap B_\mathfrak{g}^{(n)}$. Now, we are almost done. Since, for $\omega \in A$, we have the uniform convergence of $\|\boldsymbol{\zeta}_n(\omega) - \boldsymbol{\zeta}(\omega)\|$ to zero, we may bound the integrand uniformly. More precisely, for any $\tilde{\varepsilon} > 0$ there exists $n_{\tilde{\varepsilon}}$ such that $\|\boldsymbol{\zeta}(\omega) - \boldsymbol{\zeta}_n(\omega)\| < \tilde{\varepsilon}$ for all $n \geq n_{\tilde{\varepsilon}}$ and all $\omega \in A$, so that, from Jensen's inequality,

$$\begin{aligned} I_1^{(n)} &\leq \sum_{\mathfrak{g}\in\mathfrak{G}_n}\int_\Omega \mathbf{1}_{B_\mathfrak{g}^{(n)}}(\omega)\int_{B_\mathfrak{g}^{(n)}} \Big[2\big\|\boldsymbol{\zeta}(\omega) - \boldsymbol{\zeta}_n(\omega)\big\|^2\mathbf{1}_A(\omega) \\ &\qquad\qquad + 2\big\|\boldsymbol{\zeta}(\eta) - \boldsymbol{\zeta}_n(\eta)\big\|^2\mathbf{1}_A(\eta)\Big]\frac{\mathrm{dP}(\eta)}{\mathrm{P}(B_\mathfrak{g}^{(n)})}\mathrm{dP}(\omega) \\ &\leq \sum_{\mathfrak{g}\in\mathfrak{G}_n}\int_\Omega \mathbf{1}_{B_\mathfrak{g}^{(n)}}(\omega)\,4\tilde{\varepsilon}^2\frac{\mathrm{P}(B_\mathfrak{g}^{(n)})}{\mathrm{P}(B_\mathfrak{g}^{(n)})}\mathrm{dP}(\omega) = 4\tilde{\varepsilon}^2\mathrm{E}\sum_{\mathfrak{g}\in\mathfrak{G}_n}\mathbf{1}_{B_\mathfrak{g}^{(n)}} = 4\tilde{\varepsilon}^2. \end{aligned}$$

Part *(ii)* of the proposition follows. Part *(iii)* is an immediate consequence of Parts *(i)* and *(ii)*. □

A.4 Proof of Proposition 3.2

First assume that $\mathrm{P} \in \mathcal{P}_d^+$, with center-outward distribution function $\mathbf{F}_\pm$. In view of Proposition 3.1, establishing the result for $\mathbf{T}^{(n)}$ is sufficient.

Put $\mathbf{V} := \mathbf{F}_\pm(\mathbf{Z}_1^{(n)})$. Then $\mathbf{V} \overset{D}{=} U\mathbf{W}$, where U and $\mathbf{W}$ are mutually independent, U is uniform over $[0,1]$, and $\mathbf{W}$ is uniform over the unit sphere $\mathcal{S}_{d-1}$. Clearly,

$$\mathrm{E}\mathbf{T}^{(n)} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}\,(\mathbf{T}^{(n)}) = \mathrm{Var}\,\mathbf{J}(\mathbf{V}) = \int_{\mathbb{S}_d} \mathbf{J}(\mathbf{u})\mathbf{J}'(\mathbf{u})\,\mathrm{dU}_d$$

so that, for $\mathbf{J}$ of the form (3.3),

$$\mathrm{Var}\,(\mathbf{T}^{(n)}) = \mathrm{E}J^2(U)\mathrm{Var}\,\mathbf{W} = \frac{1}{d}\int_0^1 J^2(u)\,\mathrm{d}u\,\mathbf{I}_d$$

since $\mathrm{Var}\,\mathbf{W} = \frac{1}{d}\mathbf{I}_d$ (see, e.g. page 34 of Fang et al. (2017)). Now, $\mathbf{T}^{(n)}$ is a sum of independent variables, and the Noether condition (3.5) ensures that the Feller-Lindenberg condition holds. The desired asymptotic normality result (for $\underset{\sim}{\mathbf{T}}_a^{(n)}$ and $\underset{\sim}{\mathbf{T}}_e^{(n)}$, under $\mathrm{P} \in \mathcal{P}_d^+$) thus follows from the central limit theorem. Finally, consider the general case $\mathrm{P} \in \mathcal{P}_d$. Distribution-freeness implies that the finite-n distributions of $\underset{\sim}{\mathbf{T}}_a^{(n)}$ and $\underset{\sim}{\mathbf{T}}_e^{(n)}$ are the same under $\mathrm{P} \in \mathcal{P}_d$ as under $\mathrm{P}' \in \mathcal{P}_d^+$. Hence, their asymptotic distributions under $\mathrm{P} \in \mathcal{P}_d$ and $\mathrm{P}' \in \mathcal{P}_d^+$ also coincide. This completes the proof. □

A.5 Proof of Proposition 5.1

First assume that the error distribution P, with center-outward distribution function $\mathbf{F}_\pm$, is in $\mathcal{P}_d^+$. Noting that, for column vectors $\mathbf{a}$ and $\mathbf{b}$, we have $\mathrm{vec}\big(\mathbf{a}\mathbf{b}'\big) = \mathbf{b}\otimes\mathbf{a}$,

$$\begin{aligned} n^{1/2}\mathrm{vec}\underset{\sim}{\mathbf{\Lambda}}_{\mathbf{J}}^{(n)\pm} &= n^{-1/2}\sum_{i=1}^n \mathrm{vec}\Big[\mathbf{K}_{\mathbf{c}}^{(n)\prime}(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})\mathbf{J}'\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\mathbf{S}_{i;\pm}^{(n)}\Big)\Big] \\ &= n^{-1/2}\sum_{i=1}^n \mathbf{J}\Big(\frac{R_{i;\pm}^{(n)}}{n_R+1}\mathbf{S}_{i;\pm}^{(n)}\Big)\otimes\Big(\mathbf{K}_{\mathbf{c}}^{(n)\prime}(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})\Big). \end{aligned}$$

It follows from Proposition 3.1 that this latter statistic is asymptotically equivalent to

$$\mathbf{T}_J = n^{-1/2}\sum_{i=1}^n \mathbf{J}\Big(\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\Big)\otimes\Big(\mathbf{K}_{\mathbf{c}}^{(n)\prime}(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})\Big),$$

which is a sum of independent variables such that $\mathrm{E}\mathbf{T}_J = 0$, and

$$\begin{aligned}
\mathrm{Var}\,\mathbf{T}_J &= n^{-1}\sum_{i=1}^{n} \mathrm{Var}\,\Big[\mathbf{J}\Big(\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\Big) \otimes \Big(\mathbf{K}_{\mathbf{c}}^{(n)\prime}(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})\Big)\Big] \\
&= n^{-1}\sum_{i=1}^{n} \mathrm{E}\Big[\mathbf{J}\Big(\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\Big) \otimes \Big(\mathbf{K}_{\mathbf{c}}^{(n)\prime}(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})\Big) \\
&\qquad\qquad\qquad \times \mathbf{J}'\Big(\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\Big) \otimes \Big((\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})'\mathbf{K}_{\mathbf{c}}^{(n)}\Big] \\
&= n^{-1}\sum_{i=1}^{n} \mathrm{E}\Big[\mathbf{J}\Big(\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\Big)\mathbf{J}'\Big(\mathbf{F}_\pm(\mathbf{Z}_i^{(n)})\Big) \\
&\qquad\qquad\qquad \otimes \Big(\mathbf{K}_{\mathbf{c}}^{(n)\prime}(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})'\mathbf{K}_{\mathbf{c}}^{(n)}\Big)\Big] \\
&= \int_{\mathcal{S}_d} \mathbf{J}(\mathbf{u})\mathbf{J}'(\mathbf{u})\,\mathrm{dU}_d \otimes n^{-1}\mathbf{K}_{\mathbf{c}}^{(n)\prime}\sum_{i=1}^{n}\Big[(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})(\mathbf{c}_i^{(n)} - \bar{\mathbf{c}}^{(n)})'\Big]\mathbf{K}_{\mathbf{c}}^{(n)},
\end{aligned}$$

which tends to $\boldsymbol{\mathcal{I}}_{\mathbf{J}} \otimes \mathbf{I}_m$ as $n \to \infty$. The Lindeberg condition is satisfied, so that $\mathbf{T}_J$, hence also $n^{1/2}\mathrm{vec}\underset{\sim}{\boldsymbol{\Lambda}}_{\mathbf{J}}^{(n)\pm}$, has the announced asymptotic normal distribution.

Finally, consider the general case of an absolutely continuous $\mathrm{P} \in \mathcal{P}_d$: as in the proof of Proposition 3.2, distribution-freeness implies that the asymptotic distribution of $n^{1/2}\mathrm{vec}\underset{\sim}{\boldsymbol{\Lambda}}_{J}^{(n)\pm}$ is the same under $\mathrm{P} \in \mathcal{P}_d$ as under $\mathrm{P} \in \mathcal{P}_d^+$. This completes the proof of Part *(i)*. In view of (5.2); Parts *(ii)* and *(iii)* readily follow. □

A.6 Affine invariance and sphericization

Affine invariance (testing) or equivariance (estimation), in "classical multivariate analysis," is generally considered an essential and inescapable property. Closer examination, however, reveals that this particular role of affine transformations is intimately related to the affine invariance of Gaussian and elliptical families of distributions. When Gaussian or elliptical assumptions are relaxed, affine transformations are losing this privileged role and the relevance of affine invariance/equivariance properties is much less obvious.

When $\mathbf{Y}^{(n)}\mathbf{A}'$ (where $\mathbf{A}$ is an arbitrary full-rank $d \times d$ matrix), is observed instead of $\mathbf{Y}^{(n)}$, $\widehat{\boldsymbol{\Sigma}}^{(n)}$ is replaced with $\widehat{\boldsymbol{\Sigma}}_{\mathbf{A}}^{(n)} = \mathbf{A}\widehat{\boldsymbol{\Sigma}}^{(n)}\mathbf{A}'$, yielding sphericized residuals of the form $\mathbf{Z}_{\mathbf{A};i}^{(n)\,\mathrm{ell}} := \big(\mathbf{A}\widehat{\boldsymbol{\Sigma}}^{(n)}\mathbf{A}'\big)^{-1/2}\mathbf{A}\mathbf{Z}_i^{(n)}$ instead of $\mathbf{Z}_i^{(n)\,\mathrm{ell}}$. It follows from elementary calculation that $\mathbf{Z}_{\mathbf{A};i}^{(n)\,\mathrm{ell}} = \mathbf{P}\mathbf{Z}_i^{(n)\,\mathrm{ell}}$ with $\mathbf{P} = (\mathbf{A}\widehat{\boldsymbol{\Sigma}}^{(n)}\mathbf{A}')^{-1/2}\mathbf{A}(\widehat{\boldsymbol{\Sigma}}^{(n)})^{1/2}$ orthogonal. Strictly speaking, sphericized residuals, thus, are not affine-invariant. This possible discrepancy between sphericized residuals is due to the fact that square roots such as $(\widehat{\boldsymbol{\Sigma}}^{(n)})^{-1/2}$ are only defined up to an orthogonal transformation; choosing the symmetric root is a convenient choice, but does not yield $\mathbf{P} = \mathbf{I}_d$.[21] However, the moduli $\|\mathbf{Z}^{(n)\,\text{ell}}_{\mathbf{A};i}\|$ and $\|\mathbf{Z}^{(n)\,\text{ell}}_{i}\|$ coincide, irrespective of $\mathbf{P}$, and so do the cosines

$$\frac{\langle \mathbf{Z}^{(n)\,\text{ell}}_{\mathbf{A};i}, \mathbf{Z}^{(n)\,\text{ell}}_{\mathbf{A};j}\rangle}{\|\mathbf{Z}^{(n)\,\text{ell}}_{\mathbf{A};i}\|\|\mathbf{Z}^{(n)\,\text{ell}}_{\mathbf{A};j}\|} \quad \text{and} \quad \frac{\langle \mathbf{Z}^{(n)\,\text{ell}}_{i}, \mathbf{Z}^{(n)\,\text{ell}}_{j}\rangle}{\|\mathbf{Z}^{(n)\,\text{ell}}_{i}\|\|\mathbf{Z}^{(n)\,\text{ell}}_{j}\|}, \quad i, j = 1, \ldots, n.$$

The affine-invariance of typical elliptical-rank-based test statistics, which are quadratic forms involving those moduli and cosines, follows.

Being measurable with respect to the ranks of the moduli $\|\mathbf{Z}^{(n)\,\text{ell}}_{i}\|$ and the scalar products $\langle \mathbf{Z}^{(n)\,\text{ell}}_{i}, \mathbf{Z}^{(n)\,\text{ell}}_{j}\rangle / \|\mathbf{Z}^{(n)\,\text{ell}}_{i}\|\|\mathbf{Z}^{(n)\,\text{ell}}_{j}\|$, the elliptical rank statistics developed in Hallin and Paindaveine (2005) are affine-invariant; this is in full agreement with our previous remark that the limiting local Gaussian shifts in elliptical experiments are unaffected under affine transformations.

The center-outward distribution functions, ranks and signs cannot be expected to enjoy similar affine-invariance properties—actually, it has been proved (Proposition 3.14 in Cuesta-Albertos et al. (1993)) that they do not. If, however, affine invariance is considered an indispensable property, it is easily restored: choosing your favorite (consistent under ellipticity) estimator of scatter $\widehat{\mathbf{\Sigma}}^{(n)}$(which also requires, in case $\boldsymbol{\beta}_0$ is not specified, an estimator of location $\hat{\boldsymbol{\mu}}^{(n)}$), just compute the sphericized residuals $\mathbf{Z}^{(n)\,\text{ell}}_{i}$ in (4.4) prior to computing the center-outward ranks and signs and performing the tests: in view of Proposition 2.2, the resulting center-outward ranks and signs enjoy the same affine-invariance properties (invariance of the ranks and the cosines of signs) as the elliptical ones.

If the actual density $f^{\boldsymbol{\varepsilon}}$ is elliptical, this linear sphericization does not modify the local experiment, hence local asymptotic powers and, in case the scores themselves are spherical, efficiency properties, are preserved. If the actual density $f^{\boldsymbol{\varepsilon}}$ is not elliptical, however, such linear sphericization[22] has a nonlinear impact[23] on $f^{\boldsymbol{\varepsilon}}$-based central sequences: the corresponding Gaussian shift experiments are not preserved and, irrespective of the scores they are based on, the local asymptotic powers of center-outward rank tests are affected. Summing up, preliminary sphericization does restore affine-invariance of center-outward rank

[21] The Cholesky square root does: see Proposition 2 in Hallin et al. (2020c).

[22] Actually, only a "second-order sphericization," as the distribution of $\mathbf{Z}^{(n)\,\text{ell}}_{i}$ still fails to be spherical unless the error distribution itself was.

[23] The signs, indeed, now are sitting "inside" the function $\boldsymbol{\varphi}_{f^{\boldsymbol{\varepsilon}}}$.

tests while preserving their local powers under ellipticity, but distorts those local powers under non-elliptical error densities. Figure 4 provides examples where that distortion significantly deteriorates the power.

Whether affine-invariance is desirable or not is open to discussion. In "classical multivariate analysis," that is, under Gaussian or elliptical densities, linear sphericization preserves local experiments, making affine invariance a natural requirement. When considering more general error distributions P, linear transformations are losing their privileged status: they no longer sphericize the distribution P of a typical $\mathbf{Z} \sim \mathrm{P}$ and no longer preserve local experiments. Moreover, while all consistent-under-ellipticity estimators $\widehat{\mathbf{\Sigma}}^{(n)}$ and $\hat{\boldsymbol{\mu}}^{(n)}$ yield, under ellipticity, the same limiting location and scatter values, distinct estimators, under non-elliptical densities, will converge to distinct and sometimes hardly interpretable limits: $\hat{\boldsymbol{\mu}}^{(n)} = \overline{\mathbf{X}}^{(n)}$ (the arithmetic mean) and $\hat{\boldsymbol{\mu}}^{(n)} = \mathbf{X}^{(n)}_{\text{Oja}}$ (the *Oja median*, Oja (1983)), which asymptotically coincide under ellipticity, may yield completely distinct locations; what is the relevance of Tyler's scatter matrix in a distribution where sign curves are not straight lines and decorrelation of radii (through which center $\hat{\boldsymbol{\mu}}^{(n)}$?) makes little sense? etc. The distortion of local powers under non-elliptical error densities thus depends on the choice of $\widehat{\mathbf{\Sigma}}^{(n)}$ and $\hat{\boldsymbol{\mu}}^{(n)}$, which is hard to justify. While easily implementable, affine invariance/equivariance, in such a context, is thus a disputable requirement.

A.7 Further simulations

A.7.1 Scatterplots fom the Gaussian mixtures *(c)*, *(e)*, and *(f)* and the skew-t distribution with 3 degrees of freedom *(g)*

Figure 9 provides scatterplots (samples of size 200) from the bivariate Gaussian mixtures *(c)*, *(e)*, and *(f)* and from the skew-t distribution with 3 degrees of freedom *(g)* as described in Section 7.1.

A.7.2 Two-sample location, $d = 6$

A Monte Carlo simulation study was conducted for the two-sample location test in $\mathbb{R}^d$ for $d = 6$. The setting is similar to that for $d = 2$. Two independent random samples of size $n_1 = n_2 = n/2$ were generated and the test statistics $\underset{\sim}{Q}^{(n)\,\text{ell}}_{\text{Wilcoxon}}$ and $\underset{\sim}{Q}^{(n)\pm}_{\text{Wilcoxon}}$ were computed, along with Hotelling's test statistic . The sample covariance matrix $\widehat{\mathbf{\Sigma}}_n$ was used for the computation of $\underset{\sim}{Q}^{(n)\,\mathrm{ell}}_{\mathrm{Wilcoxon}}$. Rejection frequencies were computed for the following error distributions:

(a) the centered multivariate normal with identity variance matrix,

(b) the centered spherical Cauchy distribution (i.e., the elliptical t-distribution with 1 degree of freedom) and identity scatter matrix,

(c) the distribution with Clayton copula with parameter $\theta \in \{1/2, 2\}$ and exponential marginals (with mean 1). Note that these choices of θ correspond to Kendall's τ values 0.2 and 0.5, respectively.

The first sample was generated from one of the distributions in (a)–(c), the second one from the same distribution shifted by the vector $(\delta, \ldots, \delta)^\top$ with $\delta \in \{0.00, 0.05, 0.10, 0.20\}$. A random grid, simulated specifically for each joint random sample (of size n), was used for the computation of $\mathbf{F}^{(n)}_{\pm}$. For specified n_R and n_S such that $n_R \cdot n_S = n$, direc-

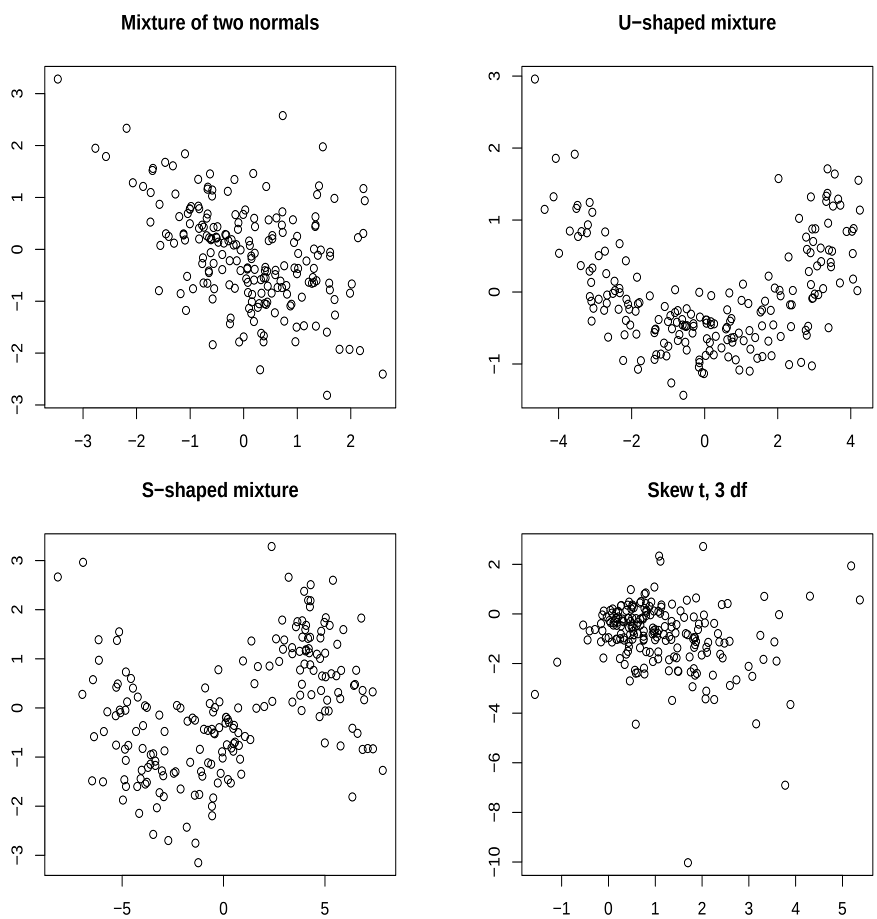

Figure 9: Sample plots of 200 observations drawn form the the Gaussian mixtures *(c)*, *(e)*, and *(f)* and from the skew-t-distribution with 3 degrees of freedom *(g)* described in Section 7.1.

tions $\mathbf{s}_1^r, \ldots, \mathbf{s}_{n_S}^r$, $r = 1, \ldots, n_R$ were generated uniformly over the sphere $\mathcal{S}_5$; the grid then consists of the points

$$\frac{r}{n_R + 1} \cdot \mathbf{s}_s^r, \quad r = 1, \ldots, n_R, \ s = 1, \ldots, n_S.$$

The results we are presenting below were computed for $n_R = 8, 10, 16, 20, 25$, corresponding to $n = 2n_1 = 200, 400, 800, 1200, 1600$ respectively.

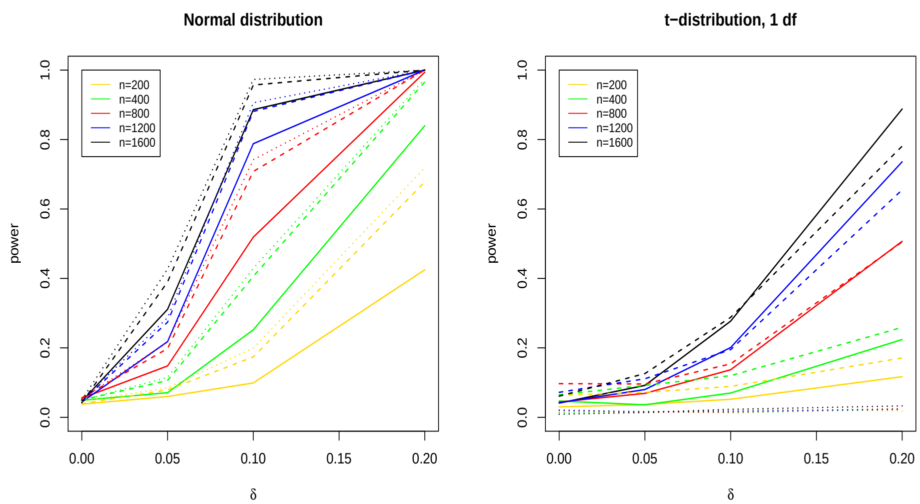

Figure 10: The empirical powers of two-sample location tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon elliptical rank statistic (dashed line), and the two-sample Hotelling test (dotted line) as functions of the shift δ for the 6-variate normal distribution (left panel) and Cauchy distribution (right panel). Sample sizes are $n_1 = n/2 = n_2$.

Each simulation was replicated $N = 1000$ times and the empirical power and size of the test were computed for $\alpha = 0.05$. The resulting rejection frequencies show the dependence of the power on the parameter δ and they are provided in Figures 10 and 11. The left panel of Figures 10 unsurprisingly reveals that if the underlying distribution is Gaussian, then the classical Hotelling test yields the largest power, closely followed by the elliptical Wilcoxon test. The loss of power resulting from using Wilcoxon rather than normal scores, thus, seems to increase with the dimension (for $d = 2$, it was hardly visible). Under the Cauchy distribution, the two rank tests yield comparable results, with an increasing advantage for the center-outward one as n and δ grow, while the Hotelling test fails miserably (the power lies below α for all δ).

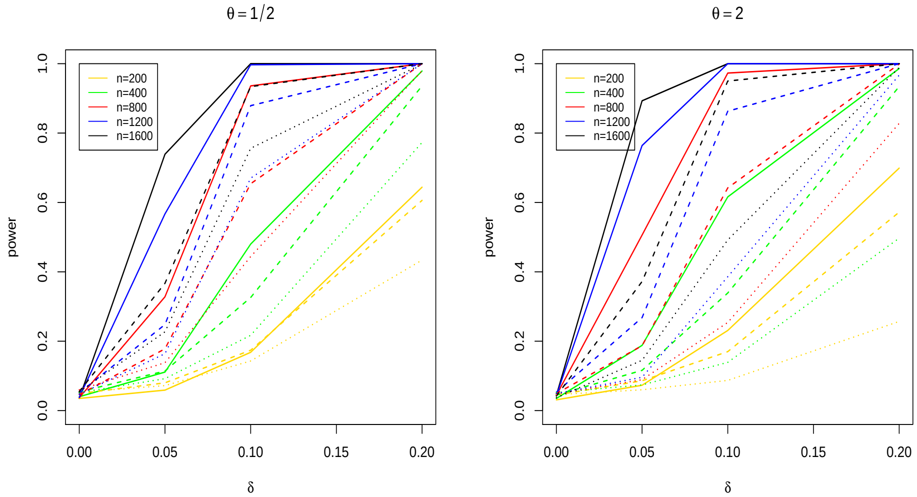

Figure 11: The empirical powers of two-sample location tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon elliptical rank statistic (dashed line), and the two-sample Hotelling test (dotted line) as functions of the shift δ for the distribution (c) with exponential marginals and joint Clayton copula with $\theta = 1/2$ (left panel) and $\theta = 2$ (right panel). Sample sizes are $n_1 = n/2 = n_2$.

Figure 11 indicates that, under non-elliptical distributions of the form (c), the center-outward rank test spectacularly outperforms its two competitors; the differences between the three tests are particularly large for $\theta = 2$ (stronger copula dependence).

A.7.3 One-way MANOVA, $d = 6$

A similar simulation study was conducted for MANOVA ($K = 3$ groups) in $\mathbb{R}^6$. As for $d = 2$ in Section 7.2, two random samples were generated from one of the distributions (a)–(c) listed in Section A.7.2 and the third sample was drawn from the same distribution shifted by a vector $(\delta, \ldots, \delta)^\top$ with $\delta \in \{0.00, 0.05, 0.10, 0.20\}$. A balanced design with $n_1 = n_2 = n_3 = n/3$ was considered for $n = 300, 750, 1500$. Figures 12 and 13 are reporting the rejection rates for $\alpha = 0.05$ of the Wilcoxon center-outward rank test (solid line), the Wilcoxon elliptical rank test (dashed line) and Pillai's classical trace test (dotted line). If the underlying distribution of the three samples is Gaussian, the elliptical and Pillai tests outperform the center-outward one—which, again, is hardly surprising since they are exploiting the information that the observations are elliptical or Gaussian. Under the Cauchy distribution, however, the elliptical test is over-rejecting, with rejection

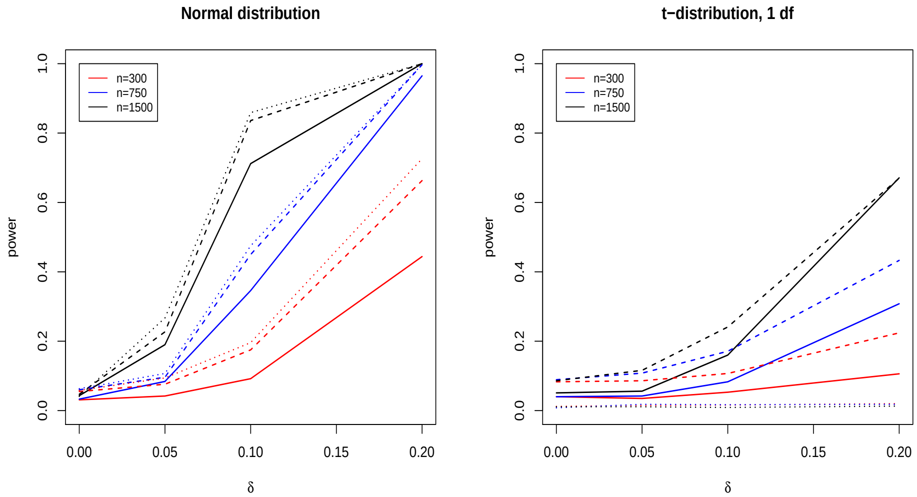

Figure 12: The empirical powers of MANOVA tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon elliptical rank statistic (dashed line), and Pillai's test (dotted line) as functions of the shift δ for the 6-variate normal distribution (left panel) and Cauchy distribution (right panel). Balanced samples of sizes $n_1 = n_2 = n_3 = n/3$ are considered.

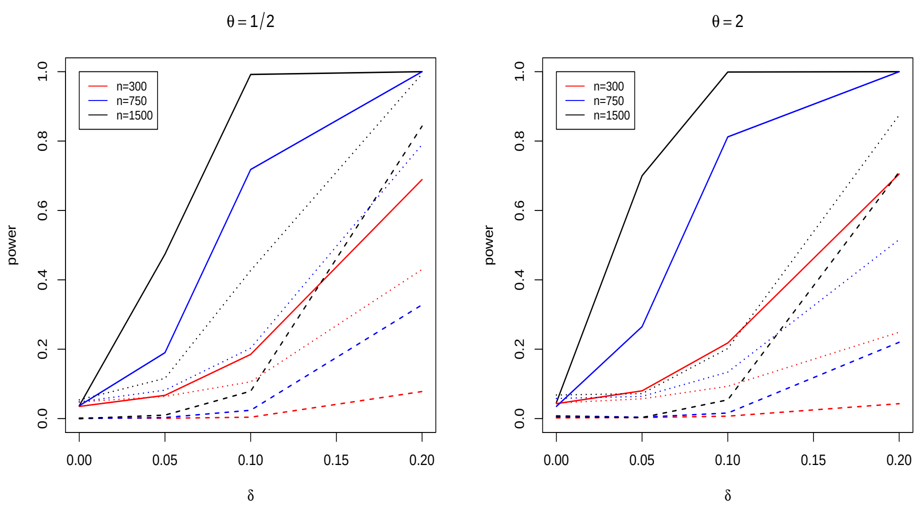

Figure 13: The empirical powers of MANOVA tests based on the Wilcoxon center-outward rank statistic (solid line), the Wilcoxon elliptical rank statistic (dashed line), and Pillai's test (dotted line) as functions of the shift δ for the normal distribution (left panel) and Cauchy distribution (right panel). Balanced samples of sizes $n_1 = n_2 = n_3 = n/3$ are considered.

frequency significantly larger than the nominal size $\alpha = 0.05$. As for Pillai's test , it fails completely. For the non-elliptical distributions with exponential marginals (Figure 13), the center-outward rank test clearly outperforms the other two. The elliptical test seems to be very conservative here.