ECARES

Tyler Shape Depth

Davy Paindaveine
SBS-EM, ECARES, Université libre de Bruxelles

Germain Van Bever
SBS-EM, ECARES, Université libre de Bruxelles

July 2017

Tyler Shape Depth

Davy Paindaveine[1]   and   Germain Van Bever

ECARES and Département de Mathématique, Université libre de Bruxelles, Avenue Roosevelt, 50, ECARES, CP114/04, B-1050, Brussels, Belgium

Abstract

In many problems from multivariate analysis (principal component analysis, testing for sphericity, etc.), the parameter of interest is a *shape* matrix, that is, a normalised version of the corresponding scatter or dispersion matrix. In this paper, we propose a depth concept for shape matrices which is of a *sign* nature, in the sense that it involves data points only through their directions from the center of the distribution. We use the terminology *Tyler shape depth* since the resulting estimator of shape — namely, the deepest shape matrix — is the depth-based counterpart of the celebrated M-estimator of shape from Tyler (1987). We investigate the invariance, quasi-concavity and continuity properties of Tyler shape depth, as well as the topological and boundedness properties of the corresponding depth regions. We study existence of a deepest shape matrix and prove Fisher consistency in the elliptical case. We derive a Glivenko-Cantelli-type result and establish the almost sure consistency of the deepest shape matrix estimator. We also consider depth-based tests for shape and investigate their finite-sample performances through simulations. Finally, we illustrate the practical relevance of the proposed depth concept on a real data example.

Keywords: Elliptical distribution; Robustness; Shape matrix; Statistical depth; Test for sphericity.

---

1   Introduction

Location depths measure the centrality of an arbitrary k-vector θ with respect to a probability measure $P = P^X$ over $\mathbb{R}^k$. Letting $\mathcal{S}^{k-1} = \{x \in \mathbb{R}^k : \|x\|^2 = x^T x = 1\}$ be the unit sphere in $\mathbb{R}^k$, the most famous instance is the (Tukey, 1975) *halfspace depth*

$$D(\theta, P) = \inf_{u \in \mathcal{S}^{k-1}} P[u^T(X - \theta) \geq 0], \tag{1.1}$$

the lower bound of the probability mass of any halfspace whose boundary hyperplane contains θ. The halfspace depth regions $\{\theta \in \mathbb{R}^k : D(\theta, P) \geq \alpha\}$ form a family of nested convex subsets of $\mathbb{R}^k$. The innermost region $M_P = \{\theta \in \mathbb{R}^k : D(\theta, P) = \max_{\xi \in \mathbb{R}^k} D(\xi, P)\}$ (the maximum always exists; see Rousseeuw and Ruts, 1999) extends the univariate median interval to the multivariate case. Whenever a unique representative of M_P is needed, one often considers the *Tukey median* θ_P defined as the barycentre of M_P (from convexity, θ_P has maximal depth). The use of θ_P as a robust alternative to

---

[1]Corresponding author: dpaindav@ulb.ac.be

the expectation $E(X)$ is only one out of the numerous applications of halfspace depth. Many inference problems can indeed be tackled in a robust and nonparametric way by using the center-outward order resulting from depth; see, e.g., Liu et al. (1999). Halfspace depth is also important through its links with multivariate quantiles; see, e.g., Hallin et al. (2010) and the references therein.

In this paper, the focus is not on location parameters as above, but rather on some specific multivariate dispersion parameters, namely on *shape matrices.* We now describe shape in a context where its definition makes a large consensus, that is, in the elliptical setup. Denoting as $\mathcal{P}_k$ the collection of $k \times k$ symmetric positive definite matrices and writing $A^{1/2}$, with $A \in \mathcal{P}_k$, for the unique square root of A in $\mathcal{P}_k$, we will say that $P = P^X$ is elliptical with location $\theta(\in \mathbb{R}^k)$, *shape* $V(\in \mathcal{P}_{k,\mathrm{tr}} = \{V \in \mathcal{P}_k : \ \mathrm{tr}(V) = k\})$ and generating variate R if X has the same distribution as $\theta + RV^{1/2}U$, where U is uniformly distributed over $\mathcal{S}^{k-1}$ and is independent of the nonnegative random variable R. The shape V is identifiable as soon as $P[\{\theta\}] < 1$. Shape is the parameter of interest in many multivariate statistics problems, including principal component analysis (PCA), canonical correlation analysis (CCA), testing for sphericity, etc. In PCA, for instance, since V is proportional to the covariance matrix Σ of X (when it exists), both V and Σ provide the same principal directions and the same proportions of explained variance. Moreover, under infinite second-order moments, shape remains well-defined (unlike Σ) and still fixes the elliptical geometry of the distribution that allows to conduct PCA.

The paramount importance of shape in multivariate statistics explains the huge literature dedicated to M-estimation for shape/scatter parameters; see, among many others, Tyler (1987), Kent and Tyler (1988, 1991), Dudley et al. (2009), or the survey paper Dümbgen et al. (2015). It also makes it desirable to extend the concept outside the elliptical setup. Letting $U_{\theta,V}$ be the multivariate sign defined as $V^{-1/2}(X-\theta)/\|V^{-1/2}(X-\theta)\|$ if $X \neq \theta$ and as 0 otherwise (throughout, $A^{-1/2}$ is the inverse of $A^{1/2}$), Tyler (1987) defined the shape of $P = P^X$ as the matrix $V(\in \mathcal{P}_{k,\mathrm{tr}})$ satisfying

$$E\big(W_{\theta,V}\big) = 0, \qquad \text{with} \ \ W_{\theta,V} = \mathrm{vec}\big(U_{\theta,V}U_{\theta,V}^T - \tfrac{1}{k}I_k\big), \tag{1.2}$$

where vec stacks the columns of a matrix on top of each other and where I_k denotes the k-dimensional identity matrix. If P does not charge any hyperplane containing θ, then (1.2) admits a unique solution $V(\in \mathcal{P}_{k,\mathrm{tr}})$ that agrees with the true shape if P is elliptical with location θ; see Tyler (1987), Kent and Tyler (1988) or Dümbgen (1998). In essence, (1.2) identifies the shape V making the origin equal to the expectation of $W_{\theta,V}$, that is, making the origin most central in an L_2-sense with respect to the distribution $P^{W_{\theta,V}}$ of $W_{\theta,V}$. The present work finds its source in the idea that, alternatively, one may define the shape of P as the matrix $V(\in \mathcal{P}_{k,\mathrm{tr}})$ making the origin most central in the, L_1, halfspace depth sense, that is, as the value of V maximising $D(0, P^{W_{\theta,V}})$. This leads to defining shape as the value of V maximising the following shape depth.

Definition 1.1 (Tyler shape depth). *Let $P = P^X$ be a probability measure over $\mathbb{R}^k$ and fix $V \in \mathcal{P}_{k,\mathrm{tr}}$. (i) For any $\theta \in \mathbb{R}^k$, the fixed-θ shape depth of V with respect to P is $D_\theta(V,P) = D(0, P^{W_{\theta,V}}) = \inf_{u \in \mathcal{S}^{k^2-1}} P\big[u^T W_{\theta,V} \geq 0\big]$. (ii) The shape depth of V with respect to P is $D(V,P) = D_{\theta_P}(V,P)$, where θ_P is the Tukey median of P.*

Whenever the argument of $D(\cdot, P)$ is a vector (resp., a shape matrix), then the notation refers to halfspace depth (resp., to Tyler shape depth). Of course, the terminology *Tyler shape depth* refers to the use of the quantity $W_{\theta,V}$ from (1.2) to define the proposed shape depth. The definition of Tyler shape depth in the unspecified location case calls for some comments. For an unspecified location, it is natural to consider the location θ and shape V satisfying both

$$E(U_{\theta,V}) = 0 \quad \text{and} \quad E(W_{\theta,V}) = 0; \tag{1.3}$$

see Tyler (1987) and Hettmansperger and Randles (2002). In the elliptical setup, the resulting location and shape still agree with those defined above for an elliptical distribution. Here, the corresponding L_1-approach leads to considering the values of θ and V maximising $D(0, P^{U_{\theta,V}}) + \lambda D(0, P^{W_{\theta,V}})$ for some $\lambda > 0$. Interestingly, the solution does not depend on λ; the properties of halfspace depth indeed ensure that, for any V, the mapping $\theta \mapsto D(0, P^{U_{\theta,V}})$ is maximised at θ_P. Hence, for any λ, the L_1-approach identifies the location θ_P and the shape V maximising $D(0, P^{W_{\theta_P,V}})$. This justifies the "plug-in approach" adopted in Definition 1.1 for the unspecified location case. Interestingly, to the best of our knowledge, there is no formal proof that, under appropriate smoothness conditions on the underlying probability measure, there exists a solution (θ, V) of (1.3); this is why, as far as theory is concerned, the plug-in approach is adopted in the M-estimation framework; see Tyler (1987) and Hettmansperger and Randles (2002). The depth construction above has the advantage that the plug-in approach and the joint location-scatter one provide the *same* depth-based functionals.

The above concept of shape depth raises many questions: does a deepest shape matrix exist for any P? If it exists, does it coincide, under ellipticity assumptions, with the shape defined in the elliptical case? What are the properties of Tyler shape depth and of the corresponding depth regions $R_\theta(\alpha, P) = \{V \in \mathcal{P}_{k,\mathrm{tr}} : D_\theta(V, P) \geq \alpha\}$ and $R(\alpha, P) = R_{\theta_P}(\alpha, P)$? Can one perform inference based on the sample version of these concepts? Is the resulting ordering of shape matrices of interest for applications? We answer these questions in this paper. Depth for a generic parameter has been discussed in Mizera (2002). To the best of our knowledge, however, depth for covariance or scatter matrices has only been considered in Zhang (2002), Chen et al. (2015) and Paindaveine and Van Bever (2017a), and only the latter work considers depth for shape matrices. As we will explain in the sequel, Tyler shape depth presents important advantages over both the shape depth resulting from the general concept of Mizera (2002) and the one from Paindaveine and Van Bever (2017a).

2 Main properties

In this section, we study the main properties of the fixed-θ Tyler shape depth and of the corresponding depth regions. Topological statements for subsets of $\mathcal{P}_{k,\mathrm{tr}}$ and for functions defined on $\mathcal{P}_{k,\mathrm{tr}}$ will refer to the topology whose open sets are generated by balls of the form $B(V_0, r) = \{V \in \mathcal{P}_{k,\mathrm{tr}} : d(V, V_0) < r\}$, where d is the geodesic distance that is usually defined on $\mathcal{P}_k$: with the usual logarithmic mapping on $\mathcal{P}_k$, the distance d

is defined through $d(V_a, V_b) = \|\log(V_a^{-1/2} V_b V_a^{-1/2})\|_F$, where $\|A\|_F = \{\text{tr}(AA^T)\}^{1/2}$ is the Frobenius norm of A; see, e.g., Bhatia (2007). We start with the following continuity result.

Theorem 2.1. *Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, (i) $V \mapsto D_\theta(V, P)$ is upper semicontinuous on $\mathcal{P}_{k,\text{tr}}$; (ii) the depth region $R_\theta(\alpha, P)$ is closed for any $\alpha \geq 0$; (iii) if P is absolutely continuous with respect to the Lebesgue measure, then $V \mapsto D_\theta(V, P)$ is also lower semicontinuous, hence continuous, on $\mathcal{P}_{k,\text{tr}}$.*

We will say that a subset R of $\mathcal{P}_{k,\text{tr}}$ is bounded if and only if $R \subset B(I_k, r)$ for some $r > 0$ (the fact that the distance d actually satisfies the triangle inequality allows to restrict to balls centered at I_k). Defining $t_{\theta,P} = \sup_{u \in \mathcal{S}^{k-1}} P[u^T(X - \theta) = 0]$, we say that P is smooth at θ if $t_{\theta,P} = 0$, that is, if P does not charge any hyperplane containing θ. We then have the following boundedness result.

Theorem 2.2. *Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, the depth region $R_\theta(\alpha, P)$ is bounded and compact for any $\alpha > t_{\theta,P}$.*

The main reason to work with the geodesic distance rather than the Frobenius one $d_F(V_1, V_2) = \|V_2 - V_1\|_F$ is that, unlike $(\mathcal{P}_{k,\text{tr}}, d_F)$, the metric space $(\mathcal{P}_{k,\text{tr}}, d)$ is complete; see, e.g., Proposition 10 in Bhatia and Holbrook (2006). This is what allows to establish compacity in Theorem 2.2, which in turn is the main ingredient for the following result stating the existence of a deepest shape matrix.

Theorem 2.3. *Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. (i) If $R_\theta(t_{\theta,P}, P)$ is non-empty, then there exists a shape $V_* \in \mathcal{P}_{k,\text{tr}}$ maximising $D_\theta(V, P)$. In particular, (ii) if P is smooth at θ, then such a deepest shape V_* exists.*

The deepest shape matrix V_* is a natural candidate for *the (fixed-θ) shape matrix $V_{\theta,P}$* of P. While the previous result guarantees existence in particular for absolutely continuous probability measures, unicity is not guaranteed in general. Parallel to what is done for the Tukey median θ_P, we then define the (fixed-θ) shape matrix of P as the barycentre of the deepest shape region of P, that is, as the shape matrix $V_{\theta,P}$ satisfying

$$\text{vec } V_{\theta,P} = \int_{\text{vec } R_\theta(\alpha_*, P)} v \, dv \Big/ \int_{\text{vec } R_\theta(\alpha_*, P)} dv, \tag{2.1}$$

where $\alpha_* = \max_V D_\theta(V, P)$. Two remarks are in order. First, the integrals in (2.1) exist and are finite since $\text{vec } \mathcal{P}_{k,\text{tr}}$ is a bounded subset of $\mathbb{R}^{k^2}$ ($V_{ij}^2 \leq V_{ii} V_{jj} \leq k^2$ for any $V = (V_{ij}) \in \mathcal{P}_{k,\text{tr}}$). Second, the shape $V_{\theta,P}$ has maximal depth, which follows from the following convexity result.

Theorem 2.4. *Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, (i) $V \mapsto D_\theta(V, P)$ is quasi-concave on $\mathcal{P}_{k,\text{tr}}$, in the sense that $D_\theta((1 - t)V_a + tV_b, P) \geq \min(D_\theta(V_a, P), D_\theta(V_b, P))$ for any $V_a, V_b \in \mathcal{P}_{k,\text{tr}}$ and any $t \in [0, 1]$; (ii) the depth region $R_\theta(\alpha, P)$ is convex for any $\alpha \geq 0$.*

4

This defines the (fixed-θ) shape of an arbitrary probability measure P under the extremely mild condition that $R_\theta(t_{\theta,P}, P)$ is non-empty. Of course, it is important that, under ellipticity, this agrees with the elliptical definition of shape provided in the introduction. The following Fisher consistency result confirms this is the case.

Theorem 2.5. *Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and shape V_0. Then, $D_{\theta_0}(V_0, P) \geq D_{\theta_0}(V, P)$ for any $V \in \mathcal{P}_{k,\mathrm{tr}}$, and, provided that $P[\{\theta_0\}] < 1$, the equality holds if and only if $V = V_0$. Letting $Y_k \sim \mathrm{Beta}(1/2, (k-1)/2)$, the maximal depth is $D_{\theta_0}(V_0, P) = (1 - P[\{\theta_0\}])P[Y_k > 1/k]$.*

The only role of the constraint $P[\{\theta_0\}] < 1$ in this result is to guarantee identifiability of V_0. Note that Lemma 2 in Paindaveine and Van Bever (2017b) implies that the maximal depth in Theorem 2.5 is monotone decreasing in k as soon as $P[\{\theta_0\}]$ does not depend on k, in which case the maximal depth converges as k goes to infinity. Since Y_k has the same distribution as $Z_1^2/(\sum_{\ell=1}^{k} Z_\ell^2)$, where $Z = (Z_1, \ldots, Z_k)^T$ is standard normal, the limit is then equal to $P[Z_1^2 > 1] \approx 0.317$. The proof of Theorem 2.5 requires the following affine-invariance/equivariance result.

Theorem 2.6. *Let $P = P^X$ be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, for any shape matrix V, any invertible $k \times k$ matrix A and any k-vector b,*

$$D_{A\theta+b}(V_A, P^{AX+b}) = D_\theta(V, P^X) \quad and \quad R_{A\theta+b}(\alpha, P^{AX+b}) = \{V_A : V \in R_\theta(\alpha, P)\},$$

where $V_A = kAVA^T/\mathrm{tr}(AVA^T)$ is the shape matrix proportional to AVA^T.

This result, which is of independent interest, shows that the fixed-θ shape depth and the corresponding regions behave well under affine transformations, hence in particular under changes of the measurement units. In the location setup, the corresponding affine-invariance/equivariance property is one of the classical requirements for depth; see Property (P1) in Zuo and Serfling (2000).

Tyler shape depth is a sign concept in the sense that it depends on the underlying random vector X only through its multivariate sign $U_{\theta,V}$. In the elliptical case, it follows that, as soon as the distribution does not charge the center of the distribution, this depth is distribution-free in the sense that it does not depend on the distribution of the underlying generating variate R, hence on the fact that the elliptic distribution is Gaussian, t, etc. More precisely, we have the following result, that plays an important role for inference based on Tyler shape depth; see Section 4.

Theorem 2.7. *Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and shape V_0. Then, (i) for some $h : \mathcal{P}_{k,\mathrm{tr}} \to [0, 1]$ that does not depend on V nor on P,*

$$D_{\theta_0}(V, P) = (1 - P[\{\theta_0\}]) \, h\left(\frac{k(V_0^{-1/2} V V_0^{-1/2})}{\mathrm{tr}(V_0^{-1}V)}\right);$$

(ii) for $k = 2$,

$$D_{\theta_0}(V, P) = (1 - P[\{\theta_0\}]) \, P\left[Y_2 \geq \frac{1}{2} + \frac{1}{2}\left[1 - \det\left\{\frac{2V_0^{-1}V}{\mathrm{tr}(V_0^{-1}V)}\right\}\right]^{1/2}\right], \tag{2.2}$$

with $Y_2 \sim \mathrm{Beta}(1/2, 1/2)$.

This result shows that, while depth in the elliptical case depends on P through V_0 and $P[\{\theta_0\}]$, its dependence on $P[\{\theta_0\}]$ does not have any impact on the induced ranking of shape matrices. The explicit bivariate elliptical depth in (2.2) is compatible with all results of this section. In particular, it is easy to check that, provided $P[\{\theta_0\}] < 1$, (2.2) is uniquely maximised at $V = V_0$ and that the corresponding maximal depth is the one provided in Theorem 2.5. Boundedness of all depth regions $R(\alpha, P)$, $\alpha > 0$, can also be seen from (2.2): if $d(V, I_2)$ converges to infinity, then $d(V, V_0)$ also does, which implies that the smallest eigenvalue of $V_0^{-1}V$, hence also the depth in (2.2), converges to zero. This strengthens the general result in Theorem 2.2, that was only ensuring that the regions $R(\alpha, P)$, $\alpha > P[\{\theta_0\}]$, are bounded (note indeed that $t_{\theta_0, P} = P[\{\theta_0\}]$ for an elliptical probability measure).

3 Consistency results

Whenever k-variate observations $X_1, \ldots, X_n$ are available, the sample (fixed-θ) depth of a shape matrix V may simply be defined as $D_\theta(V, P_n)$, where P_n denotes the empirical probability measure associated with $X_1, \ldots, X_n$. In this section, we state a Glivenko-Cantelli-type result for this sample depth and investigate consistency of max-depth shape estimators. The Glivenko-Cantelli result is the following.

Theorem 3.1. *Let P be a probability measure over $\mathbb{R}^k$ and let P_n denote the empirical probability measure associated with a random sample of size n from P. Then, for any $\theta \in \mathbb{R}^k$, $\sup_{V \in \mathcal{P}_{k, \mathrm{tr}}} |D_\theta(V, P_n) - D_\theta(V, P)| \to 0$ almost surely as $n \to \infty$.*

We illustrate this result in the bivariate elliptical case associated with Theorem 2.7(ii). Figure 1 provides, for three different bivariate normal probability measures P, contour plots of

$$(V_{11}, V_{12}) \mapsto D_\theta(V, P), \quad \text{with } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12} & 2 - V_{11} \end{pmatrix}$$

as well as the empirical contour plots obtained from a random sample of size $n = 800$ drawn from the corresponding distributions. Clearly, the results support the consistency in Theorem 3.1.

In the previous section, the shape $V_{\theta, P}$ of the probability measure P was defined as the barycentre of the collection of P-deepest shape matrices. In the empirical case, a natural estimator is of course the corresponding shape matrix V_{θ, P_n} computed from the empirical probability measure P_n associated with the sample at hand. Since empirical probability measures are not smooth at θ, existence of deepest shape matrices in the sample case does not rely on Theorem 2.3; existence, however, merely follows from the fact that the only possible values of $D_\theta(V, P_n)$ are of the form ℓ/n, $\ell = 0, 1, \ldots, n$.

Theorem 3.2. *Let P be a probability measure over $\mathbb{R}^k$ and let P_n denote the empirical probability measure associated with a random sample of size n from P. Fix $\theta \in \mathbb{R}^k$ and assume that $R_\theta(t_{\theta, P}, P)$ is non-empty. Then, $d(V_{\theta, P_n}, V_{\theta, P}) \to 0$ almost surely as $n \to \infty$.*

6

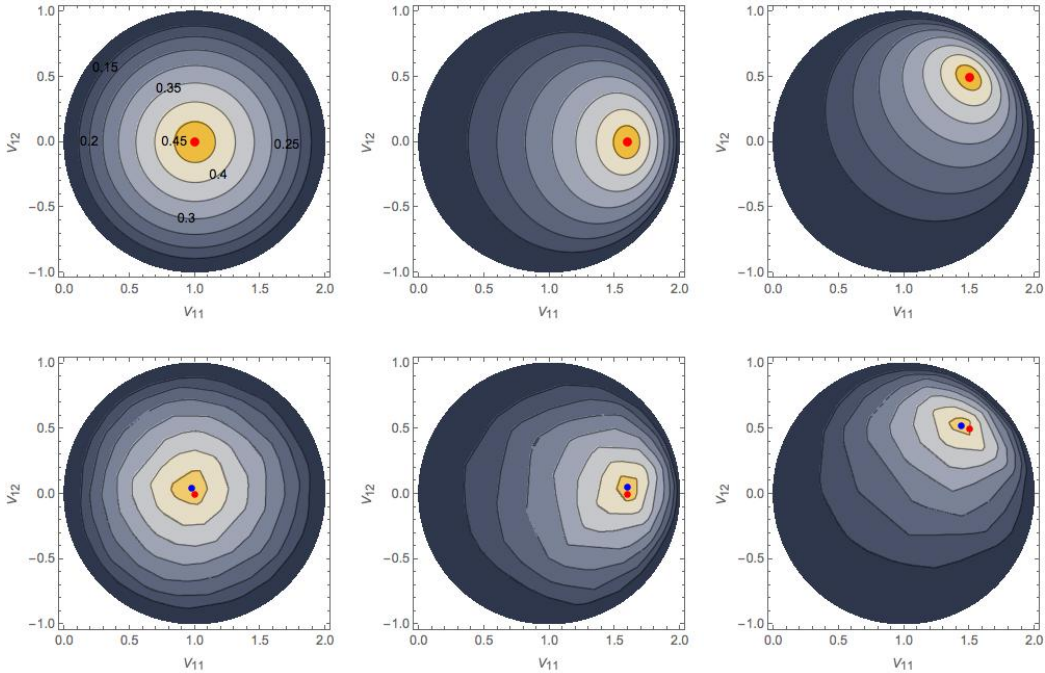Figure 1: (First row:) Contour plots of $(V_{11}, V_{12}) \mapsto D_\theta(V, P)$, for $V = \left(\begin{smallmatrix} V_{11} & V_{12} \\ V_{12} & 2-V_{11} \end{smallmatrix} \right)$, where P is bivariate normal with location 0 and shape $V_A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ (left), $V_B \propto \left(\begin{smallmatrix} 4 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ (center) and $V_C \propto c \left(\begin{smallmatrix} 3 & 1 \\ 1 & 1 \end{smallmatrix} \right)$ (right). (Second row:) Contour plots for $(V_{11}, V_{12}) \mapsto D_0(V, P_n)$, where P_n is the empirical probability measure associated with a random sample of size $n = 800$ from the centered bivariate normal with shape V_A (left), V_B (center), and V_C (right). The "true" shapes $V_{0,P}$ (resp., sample deepest shapes V_{0,P_n}) are marked in red (resp., in blue).

The only role of the assumption that $R_\theta(t_{\theta,P}, P)$ is non-empty is to guarantee the existence of the (fixed-θ) shape matrix $V_{\theta,P}$ of P. Again, this assumption is fulfilled in particular if P is smooth at θ. Figure 1 also supports Theorem 3.2 since, in each sample considered, the sample deepest shape is close to its population counterpart.

4 Depth-based tests for shape

We turn to hypothesis testing and focus on one-sample shape testing in the elliptical model. More specifically, based on a random sample $X_1, \ldots, X_n$ from a k-variate elliptical distribution with known location θ and unknown shape V, we want to test $\mathcal{H}_0 : V = V_0$ against $\mathcal{H}_1 : V \neq V_0$ at level $\alpha \in (0,1)$, where $V_0 \in \mathcal{P}_{k,\mathrm{tr}}$ is fixed ($V_0 = I_k$ provides the problem of testing sphericity against elliptical alternatives). From Theorem 2.5, a natural depth-based test, ϕ_D say, rejects the null whenever $T_{\theta,n} = D_\theta(V_0, P_n) < t_{\alpha,n}$,

where P_n is the empirical distribution associated with the random sample at hand and where $t_{\alpha,n}$ denotes the null α-quantile of $T_{\theta,n}$. Under the mild assumption that P does not charge the center of the distribution, $T_{\theta,n}$ is distribution-free under the null, which allows to approximate $t_{\alpha,n}$ arbitrarily well through simulations : from $5,000$ mutually independent standard normal samples for each sample size, we obtained the approximations $\hat{t}_{.05,200} = 0.395$ and $\hat{t}_{.05,500} = 0.434$ for $k = 2$.

We performed two simulations in the bivariate case ($k = 2$). The first one considers the problem of testing sphericity about the origin ($V_0 = I_2$, $\theta = 0$) and compares the finite-sample powers of ϕ_D with those of some competitors. For each value of $\ell = 0, 1, \ldots, 6$ we generated $M = 3,000$ independent random samples X_i, $i = 1, \ldots, n$ of size $n = 500$ from the normal with location $\theta = 0$ and shape

$$V_{\ell,\xi} = I_2 + \ell\xi\begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix}$$

and from the corresponding elliptical Cauchy. The value $\ell = 0$ corresponds to the null, whereas $\ell = 1, \ldots, 6$ provide increasingly severe alternatives. We took $\xi = 0.035$ and 0.045 for the normal and Cauchy samples in order to obtain roughly the same rejection frequencies in both cases.

For each sample, we carried out five (fixed-θ) tests at nominal level 5%: (i) the test ϕ_D rejecting the null if $T_{\theta,500} > \hat{t}_{\alpha,500} = 0.434$; (ii) the Gaussian test from John (1972)—more precisely, its extension to elliptical distributions with finite fourth-order moments from Hallin and Paindaveine (2006b); (iii)-(iv) the sign test and van der Waerden signed-rank test from the same paper; (v) the test based on the MCD_γ shape estimator from Paindaveine and Van Bever (2014) with $\gamma = 0.8$ (to achieve a good balance between efficiency and robustness). The tests (ii)-(v) were performed based on their asymptotic null distribution. The resulting rejection frequencies in Figure 2 reveal that the depth-based test ϕ_D performs very similarly to (although it may be slightly dominated by) the sign test in (ii), which is in line with the sign nature of ϕ_D. Consequently, ϕ_D performs very well under heavy tails, where it beats all other tests. As expected, the MCD test shows low empirical powers and the Gaussian test collapses under heavy tails.

The second simulation compares the five tests above in terms of robustness when testing $\mathcal{H}_0 : V = V_0$, with $V_0 = \text{diag}(2, 1/2)$ and specified location $\theta = 0$. We focused on "level robustness" (He et al., 1990) under various contaminations. We considered mixture distributions $P^{X(\eta)} = (1 - \eta)P^X + \eta P^Y$, with $\eta = 0$ (no contamination), 0.025, 0.05, 0.1, 0.2, 0.25 or 0.3 (increasingly severe contaminations). Here, X is a bivariate, normal or elliptical Cauchy, null random vector. The bivariate random vector Y determines the contamination pattern and was chosen as follows: (i) *non-uniform directional contamination*: Y has the same distribution as the vector obtained by rotating X about the origin by an angle $\pi/4$ radiant; (ii) *uniform directional contamination*: Y has the same elliptical distribution as X but for the fact that its shape is $V = I_2$; (iii) *radial and uniform directional contamination*: Y is obtained by multiplying by four the vector Y in (ii). The uncontaminated distribution P^X puts more mass along the horizontal axis. In (i), the contamination is directional and typically shows along the main bisector, whereas the contamination in (ii) is uniformly distributed over the unit
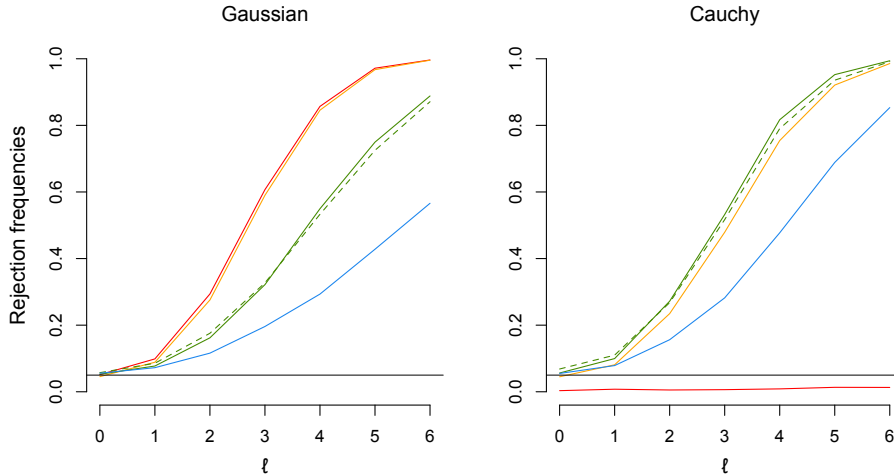
Figure 2: Rejection frequencies, under bivariate normal (left) and elliptical Cauchy (right) densities, of five tests of sphericity: the Gaussian test (red), the van der Waerden signed-rank test (orange), the sign test (green), the depth-based test (dashed green), and the MCD-based test (blue). Results are based on 3,000 replications and the sample size is $n = 500$. See Section 4 for details.

circle. As for (iii), the contamination combines the directional feature of (ii) with a radial outlyingness.

For each combination of a distribution type (normal or Cauchy), of a contamination pattern ((i)-(iii)), and of a contamination level η ($\eta = 0$, 0.025, 0.05, 0.1, 0.2, 0.25 or 0.3), we generated 3,000 independent random samples $X_{(\eta)i}$, $i = 1, \ldots, n$ of size $n = 200$. The resulting rejection frequencies are plotted in Figure 3 and reveal the very good robustness of the depth-based test ϕ_D. In particular, ϕ_D always dominates its sign-based competitor. The MCD test seems to dominate ϕ_D in some configurations but, as shown in the first simulation, exhibits very low finite-sample powers. Finally, radial outliers strongly affect the Gaussian and van der Waerden tests.

5   Comparison with other shape depths

In this section, we shortly comment on how Tyler shape depth compares with other shape depths. We start with the shape depth that would result from the generic parametric depth concept proposed in Mizera (2002). Consider a random k-vector X with a distribution $P = P_{\vartheta_0}$ from the parametric family $\mathcal{P} = \left\{ P_\vartheta : \vartheta \in \Theta \subset \mathbb{R}^\ell \right\}$ and let $\vartheta \mapsto F_\vartheta(X)$ be a measure of fit of the parameter value ϑ for an observation X. The Mizera (2002) *tangent depth* of ϑ with respect to $P = P^X$ is then $TD(\vartheta, P) = D(0, P^{\nabla_\vartheta F_\vartheta(X)})$, where $P^{\nabla_\vartheta F_\vartheta(X)}$ stands for the distribution of $\nabla_\vartheta F_\vartheta(X)$ under P. As advocated, e.g., in Mizera and Müller (2004) and Müller (2005), a likelihood-guided approach consists in taking $F_\vartheta(X) = \log L_\vartheta(X)$, where $L_\vartheta(X)$ is the likelihood of X under P_ϑ. For location
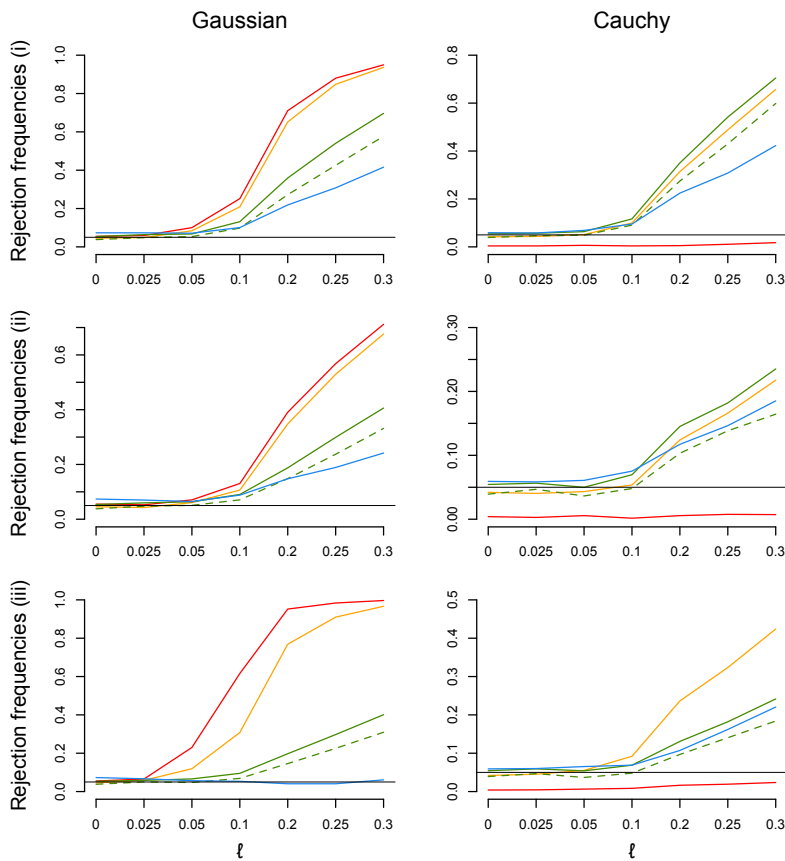
Figure 3: Null rejection frequencies, as a function of the contamination level η, of the same five tests (using the same colours) as in Figure 2, under bivariate normal (left) and elliptical Cauchy (right) densities. The labels (i)-(iii) refer to the three contaminations patterns considered; see Section 4 for details. Results are based on 3,000 replications and the sample size is $n = 200$.

and shape parameters, it is natural to consider the elliptical likelihood

$$x \mapsto L_{\theta,V,g}(x) = \frac{c_{k,g}}{(\det V)^{1/2}}\, g\big(\{(x-\theta)^T V^{-1}(x-\theta)\}^{1/2}\big), \tag{5.1}$$

where $\theta \in \mathbb{R}^k$, $V \in \mathcal{P}_{k,\mathrm{tr}}$, $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is a smooth monotone decreasing function and $c_{k,g}$ is a normalising constant. Irrespective of V and g, the resulting tangent location depth $TD(\theta, P) = D(0, P^{\nabla_\theta \log L_{\theta,V,g}(X)})$ coincides with the halfspace depth of θ with respect to P.

Describing the corresponding tangent shape depth requires the following notation. Let $\mathrm{vech}(A)$ be the vector stacking the upper-diagonal entries of A on top of each other and write $\mathrm{vech}_0(A)$ for the d_k-vector ($d_k = k(k+1)/2-1$) obtained by depriving $\mathrm{vech}(A)$ of its first component. Further, let B_k be the $d_k \times k^2$ matrix such that $B_k^T(\mathrm{vech}_0 A) = \mathrm{vec}\, A$ for any $k \times k$ symmetric matrix A with trace zero. Writing ∇_V for the gradient

10

with respect to $\text{vech}_0(V)$, the tangent shape depth of V with respect to $P = P^X$ is then

$$TD_{\theta,g}(V, P) = D(0, P^{\nabla_V \log L_{\theta,V,g}(X)}),$$

where the score (see Hallin and Paindaveine, 2006a or Paindaveine, 2008)

$$\nabla_V \log L_{\theta,V,g}(X) = \tfrac{1}{2} B_k (V^{\otimes 2})^{-1/2} \text{vec}\big(\varphi_g\big(d_{\theta,V}\big) d_{\theta,V} U_{\theta,V} U_{\theta,V}^T - \tfrac{1}{k} I_k\big)$$

involves the Mahalanobis distance $d_{\theta,V} = \{(X-\theta)^T V^{-1}(X-\theta)\}^{1/2}$. Tangent shape depth is much less satisfactory than its location counterpart : first, it depends on g in (5.1); for instance, a Gaussian likelihood and a t_ν likelihood will provide different tangent shape depths. Second, more importantly, the tangent deepest shape is not Fisher-consistent under ellipticity, irrespective of g, as the following example shows. Let X be a bivariate normal random vector with location $\theta_0 = 0$ and shape $V_0 = \text{diag}(3/4, 5/4)$. Then it can be showed (a proof is available on request from the authors) that, even when considering the tangent shape depth associated with the "true" location θ and function g (i.e., the one obtained with $\theta = \theta_0$ and $g(r) = g_0(r) = \exp(-r^2/2)$ in (5.1)), one has $TD_{\theta_0,g_0}(V_0, P^X) < TD_{\theta_0,g_0}(I_2, P^X)$, which implies that the true shape V_0 does not maximise $V \mapsto TD_{\theta_0,g_0}(V, P^X)$. This clearly disqualifies tangent depth for shape parameters.

We turn to a comparison with the *halfspace shape depth* from Paindaveine and Van Bever (2017a). In its specified-θ version, this depth is obtained as $HD_\theta(V, P) = \sup_{\sigma^2 > 0} HD_\theta^{\text{sc}}(\sigma^2 V, P)$, where $HD_\theta^{\text{sc}}(\cdot, P)$ is a companion concept of halfspace depth for (unnormalised) scatter matrices. The halfspace shape depth $HD_\theta(V, P)$ therefore requires a (delicate) maximisation in σ^2 of the halfspace scatter depth, that itself requires a projection-pursuit optimisation in $\mathbb{R}^k$. In comparison, Tyler shape depth has the advantage to be intrinsically a depth for shape matrices. A possible drawback of Tyler shape depth, however, is that it in principle requires to evaluate halfspace (location) depth in $\mathbb{R}^{k^2}$, which is computationally prohibitive even for relatively small dimensions k. Interestingly, the next result shows that Tyler shape depth only requires evaluating halfspace depth in $\mathbb{R}^{d_k}$ (recall that we let $d_k = k(k+1)/2 - 1$ above).

Theorem 5.1. *Let $P = P^X$ be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then,* $D_\theta(V, P) = D(0, P^{\tilde{W}_{\theta,V}}) = \inf_{u \in \mathcal{S}^{d_k-1}} P\big[u^T \tilde{W}_{\theta,V} \geq 0\big]$, *with* $\tilde{W}_{\theta,V} = \text{vech}_0\big(U_{\theta,V} U_{\theta,V}^T - \tfrac{1}{k} I_k\big)$.

From a computational point of view, thus, Tyler shape depth competes well for $k = 2$ and 3 with its halfspace counterpart; the latter, however, would dominate Tyler shape depth in this respect for larger dimensions k. Most importantly, a strong advantage of Tyler shape depth over its halfspace competitor is its distribution-freeness (Theorem 2.7), which results from the sign nature of the concept. As illustrated in Section 4, distribution-freeness allows to perform inference on Tyler shape depth in the elliptical framework. This cannot be done with the halfspace shape depth that turns out to crucially depend on the type of elliptical distribution at hand. We defer to the next section a comparison of both shape depths on a real data example.

11

6 The unspecified location case and a real-data example

The previous sections focused on the fixed-θ shape depth $D_\theta(V, P)$. Most of the result extend, with only minor modifications (if any), to the unspecified-location shape depth $D(V, P) = D_{\theta_P}(V, P)$ from Definition 1.1. Theorems 2.1 to 2.4 hold for any fixed θ and their unspecified-θ versions are simply obtained by substituting θ_P for θ throughout. In particular, the existence of an unspecified-location deepest shape matrix is guaranteed if P is smooth at θ_P, or, more generally, if $R(t_{\theta_P, P}, P)$ is non-empty. The same construction allows to identify a unique representative V_P of the collection of deepest shape matrices. Theorem 2.5 and Theorem 2.7 also readily extend to the unspecified-location case since $\theta_P = \theta_0$ for any elliptical probability measure P with location θ_0. In particular, if P is elliptical with shape V_0, then the unspecified-θ shape depth $D(V, P)$ is uniquely maximised at $V = V_0$ (as soon as the distribution is not degenerate at a single point). In view of the affine equivariance of θ_P (i.e., $\theta_{P^{AX+B}} = A\theta_{P^X} + b$), the affine-invariance/equivariance properties

$$D\big(V_A, P^{AX+b}\big) = D(V, P^X) \quad \text{and} \quad R(\alpha, P^{AX+b}) = \big\{ V_A : V \in R(\alpha, P) \big\}$$

directly follow from Theorem 2.6 (see this result for the definition of V_A). As a matter of fact, only the consistency results in Theorems 3.1-3.2 require stronger assumptions (absolute continuity) in the unspecified-location case compared to the specified-location one. More precisely, we have

Theorem 6.1. *Let P be a probability measure over $\mathbb{R}^k$ that is absolutely continuous with respect to the Lebesgue measure and let P_n denote the empirical probability measure associated with a random sample of size n from P. Then, (i) $\sup_{V \in \mathcal{P}_{k,\mathrm{tr}}} |D(V, P_n) - D(V, P)| \to 0$ almost surely as $n \to \infty$; (ii) $d(V_{P_n}, V_P) \to 0$ almost surely as $n \to \infty$.*

We illustrate the use of the unspecified-location Tyler shape depth on a real data example. For each trading day between February 1st, 2015 and February 1st, 2017, we collected every five minutes the Nasdaq Composite and S&P500 stock indices and computed their returns, that is, the differences between two consecutive index values. The returns on a given day form a bivariate dataset of usually 78 observations. The exact number of observations per day varies due to some missing values and days with less than 70 bivariate returns were discarded. The resulting dataset comprises $n = 38489$ observations distributed over $D = 478$ trading days.

The analysis conducted here studies the joint behavior of the bivariate returns. Its goal is to determine which trading days present an atypical pattern, different from the "global" behavior of the volatility. Of course, an important source of atypicality is associated with the overall scale of the bivariate returns that alternate between periods of high and low volatility. Such deviations, however, can easily be detected by comparing, e.g., the trace of any scatter measure on intraday data with that on the whole dataset. Therefore, we rather focus on detecting atypicality in the shape of the joint volatility.

In other words, we aim at detecting days for which the ratios of the marginal volatilities or the correlation between the returns much deviate from their global behavior.

To this end, let $\hat{V}_{\text{full}}$ denote the unspecified-location Tyler shape M-estimate computed from the full collection of n returns, that is, (the shape part in) the solution of the empirical version of (1.3); see Tyler (1987) or Hettmansperger and Randles (2002). For each day $d = 1, \ldots, D$, we evaluated the depth $D(\hat{V}_{\text{full}}, P_d)$ of the global shape estimate with respect to the empirical distribution P_d of the bivariate returns on day d. The reason why we base this measure of (a)typicality on $\hat{V}_{\text{full}}$ is twofold. First, $\hat{V}_{\text{full}}$ is a robust shape estimate which takes into account outliers that occurs naturally in the dataset (returns at the beginning of each trading period are notoriously more volatile and should be downweighted in the shape estimation procedure). Second, $\hat{V}_{\text{full}}$ is very deep in the global series of returns (denoting as P_{full} the empirical distribution of the full collection of bivariate returns, we have $D(\hat{V}_{\text{full}}, P_{\text{full}}) = 0.4965$), hence is an excellent proxy for the (global) deepest shape matrix, whose computation seems to be a difficult task.

The left panel of Figure 4 presents the depth values $D(\hat{V}_{\text{full}}, P_d)$ as a function of $d = 1, \ldots, D$. Vertical lines mark major events affecting the shape of the volatility, while the two greyed rectangles cover two periods during which the markets notoriously knew some atypical returns. The first period follows the devaluation of the Yuan on August 11th, 2015 which saw rapid changes in the stock markets, including large devaluations on the "Black Monday" of August 24th (marked in orange). The second period covers the beginning of 2016, which was hit by slump in oil prices, making stocks relying on oil very volatile compared to non oil-based ones. This resulted in atypical shape behavior during the fortnight spanning January 22 - February 9 (this last day is marked in blue, and is known to have the sharpest loss for the S&P500 index). The other events are (3) the decision of the European Central Bank on March 10th, 2016 (in green) to extend quantitative easing thereby slashing interest rates (known to have had a significant positive impact on both Nasdaq and S&P500, but more pronounced for the latter), (4) the positive impact on the financial stocks following Fed officials' comments on the possibility of rate hike made on May 27, 2016 (in red), (5) the slump in the S&P500 Futures prices on August 15th, 2016 (in purple), and (6) the aftermath of Donald Trump's election at the US presidency on November 9th (in teal). All events are associated with days known to have atypical volatilities and are seen to have a low shape depth value. Some other major financial events – such as OPEC refusing to reduce oil production in early 2015 or the aftermath of the Brexit vote on June 24, 2016 (midway between events (4) and (5)) – had an effect on the overall size of the bivariate returns but not on their shape, which explains that the corresponding days are not flagged as overly atypical by Tyler shape depth.

For the sake of comparison, we also computed the halfspace shape depth $HD(\hat{V}_{\text{full}}, P_d)$ (see Paindaveine and Van Bever, 2017a) of the global estimate for each day d. The right panel of Figure 4 provides the plot of $D(\hat{V}_{\text{full}}, P_d)$ versus $HD(\hat{V}_{\text{full}}, P_d)$ for $d = 1, \ldots, D$. The plot shows a clear positive association (correlation between the two variables is 0.6413). The following two remarks are, however, in order. First, halfspace shape depth values seem to have a higher concentration than Tyler's. This is due to the fact that the

former maximises a concept of scatter depth in scale and has, possibly, more leeway to find scatter estimates better suited to the data. Indeed, a decrease in volatility in one of the marginals might be balanced by considering a scatter with a smaller scale and hence keep a large depth value. A byproduct of this is the fact that, when evaluating halfspace shape depth, the (difficult) maximisation step in scale seems to be crucial in correctly computing the depth ranking of the data (small deviations can indeed cause changes in this ranking). More importantly, while events (1) to (3) receive low depth with respect to both concepts, only Tyler shape depth succeeds in flagging days associated with events (4) to (6) as "outlying".

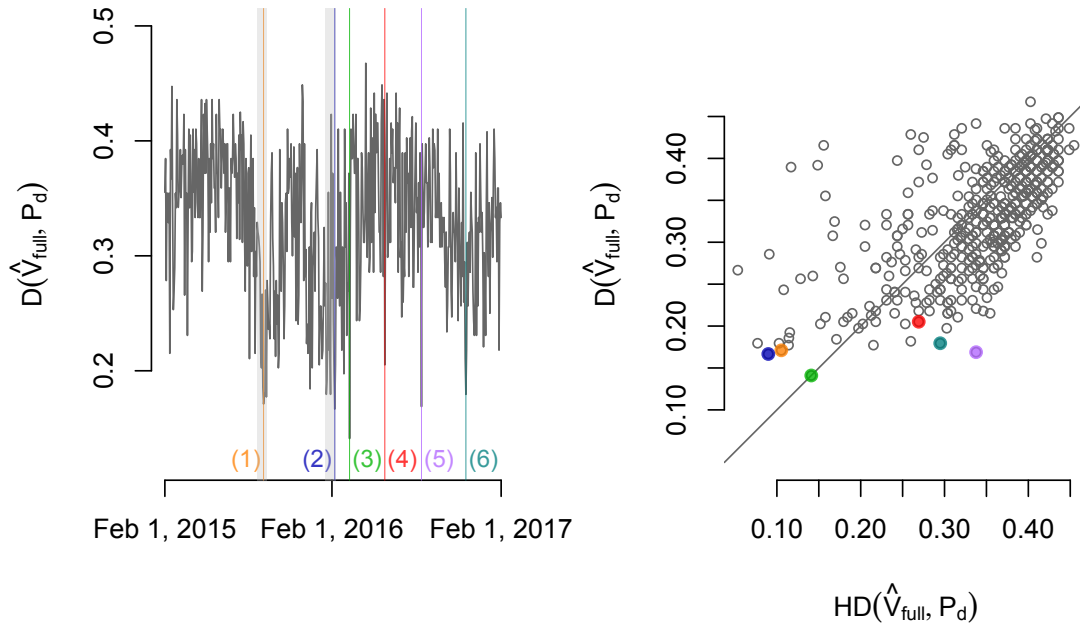Figure 4: (Left:) Plot of $D(\hat{V}_{\text{full}}, P_d)$ as a function of d. Events (1) to (6) are described in Section 6. (Right:) Plot of $D(\hat{V}_{\text{full}}, P_d)$ vs $HD(\hat{V}_{\text{full}}, P_d)$ for each trading day d. Events from the left panel are highlighted using the same colour.

Acknowledgement

A   Appendix

Many of the subsequent results require the following lemma.

Lemma A.1. *Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Write $C_{\theta,V}^M = \{x \in \mathbb{R}^k \setminus \{\theta\} : (u_{\theta,V}^x)^T M u_{\theta,V}^x \geq \frac{1}{k}\mathrm{tr}(M)\}$ and $\tilde{C}_{\theta,V}^M = \{x \in \mathbb{R}^k : (u_{\theta,V}^x)^T M u_{\theta,V}^x \geq \frac{1}{k}\mathrm{tr}(M)\}$, where $u_{\theta,V}^x$ is defined as $V^{-1/2}(x-\theta)/\|V^{-1/2}(x-\theta)\|$ if $x \neq \theta$ and as 0 otherwise. Then, for any $V \in \mathcal{P}_{k,\mathrm{tr}}$ and any $r \in \mathbb{R}$,*

$$D_\theta(V,P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[\tilde{C}_{\theta,V}^M\big] = \inf_{M \in \mathcal{M}_{k,F}^{\mathrm{all}}} P\big[\tilde{C}_{\theta,V}^M\big] = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[C_{\theta,V}^M\big] = \inf_{M \in \mathcal{M}_k^r} P\big[C_{\theta,V}^M\big],$$

where $\mathcal{M}_k^{\mathrm{all}}$ (resp., $\mathcal{M}_k^r$) collects the $k \times k$ symmetric matrices with arbitrary trace (resp., with trace r) and where $\mathcal{M}_{k,F}^{\mathrm{all}}$ is the collection of matrices in $\mathcal{M}_k^{\mathrm{all}}$ with Frobenius norm one.

Proof. It directly follows from the definition of Tyler shape depth that

$$D_\theta(V,P) = \inf_{v \in \mathbb{R}^{k^2}} P\big[\big\{x \in \mathbb{R}^k : v^T \mathrm{vec}\,(u_{\theta,V}^x (u_{\theta,V}^x)^T - \tfrac{1}{k}I_k) \geq 0\big\}\big].$$

When v runs over $\mathbb{R}^{k^2}$, the matrix M satisfying $v = \mathrm{vec}(M^T)$ runs over the collection $\mathcal{N}_k$ of $k \times k$ matrices. Since $(u_{\theta,V}^x)^T M u_{\theta,V}^x = (u_{\theta,V}^x)^T \{(M + M^T)/2\} u_{\theta,V}^x$ for any $M \in \mathcal{N}_k$, this yields

$$
\begin{aligned}
D_\theta(V,P) &= \inf_{M \in \mathcal{N}_k} P\big[\big\{x \in \mathbb{R}^{k^2} : \mathrm{tr}\big[M\{u_{\theta,V}^x (u_{\theta,V}^x)^T - \tfrac{1}{k}I_k\}\big] \geq 0\big\}\big] \\
&= \inf_{M \in \mathcal{N}_k} P\big[\tilde{C}_{\theta,V}^M\big] = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[\tilde{C}_{\theta,V}^M\big]. \tag{A.1}
\end{aligned}
$$

Letting $\mathbb{I}[A]$ be equal to one if condition A holds and to zero otherwise, this provides

$$D_\theta(V,P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} \Big(P\big[C_{\theta,V}^M\big] + P[\{\theta\}]\mathbb{I}\big[\mathrm{tr}(M) \leq 0\big]\Big) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[C_{\theta,V}^M\big], \tag{A.2}$$

where we have used the fact that $P\big[C_{\theta,V}^M\big]$ is unchanged when M is replaced with $M + \lambda I_k$ for any $\lambda \in \mathbb{R}$. The same invariance property explains that the infimum over $\mathcal{M}_k^{\mathrm{all}}$ in (A.2) may be replaced with an infimum over $\mathcal{M}_k^r$ for any r. Finally, the result for $\mathcal{M}_{k,F}^{\mathrm{all}}$ follows from (A.1) by noting that $\tilde{C}_{\theta,V}^{\lambda M} = \tilde{C}_{\theta,V}^M$ for any $\lambda > 0$ and that $M = 0$ cannot provide the infimum in (A.1). The proof is complete. $\square$

Proof of Theorem 2.1. (i) Fix $M \in \mathcal{M}_k^{\mathrm{all}}$ and consider $\tilde{C}^M = \tilde{C}_{0,I_k}^M$, where $\tilde{C}_{\theta,V}^M$ was defined in Lemma A.1. Since $\tilde{C}^M$ is closed, the mapping $P \mapsto P[\tilde{C}^M]$ is upper semicontinuous for weak convergence. Now, Slutzky's lemma entails that, as $d(V,V_0) \to 0$, the measure defined by $B \mapsto P[\theta + V^{1/2}B]$ converges weakly to the one defined by $B \mapsto P[\theta + V_0^{1/2}B]$. Therefore, $V \mapsto P[\theta + V^{1/2}\tilde{C}^M] = P[\tilde{C}_{\theta,V}^M]$ is upper semicontinuous at V_0. From Lemma A.1, we then obtain that

$$V \mapsto D_\theta(V,P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P[\tilde{C}_{\theta,V}^M],$$

15

is upper semicontinuous (as the infimum of a collection of upper semicontinuous functions). (ii) The result follows from the fact that the depth region $R_\theta(\alpha, P)$ is the inverse image of $[\alpha, \infty)$ by the upper semicontinuous function $V \mapsto D_\theta(V, P)$. (iii) Fix a sequence (V_n) in $\mathcal{P}_{k,\mathrm{tr}}$ such that $d(V_n, V_0) \to 0$. In view of Lemma A.1 again, we can, for any n, pick $M_n (\in \mathcal{M}_{k,F}^{\mathrm{all}})$ such that $P[\tilde{C}_{\theta,V_n}^{M_n}] \leq D_\theta(V_n, P) + \frac{1}{n}$. Compactness of $\mathcal{M}_{k,F}^{\mathrm{all}}$ ensures that we can extract a subsequence (M_{n_ℓ}) of (M_n) that converges to $M_0 (\in \mathcal{M}_{k,F}^{\mathrm{all}})$. Writing $\mathbb{I}[B]$ for the indicator function of the set B, the dominated convergence theorem then yields that

$$P[\tilde{C}_{\theta,V_{n_\ell}}^{M_{n_\ell}}] - P[\tilde{C}_{\theta,V_0}^{M_0}] = \int_{\mathbb{R}^k} (\mathbb{I}[\tilde{C}_{\theta,V_{n_\ell}}^{M_{n_\ell}}] - \mathbb{I}[\tilde{C}_{\theta,V_0}^{M_0}]) \, dP \to 0$$

as $\ell \to \infty$ (the absolute continuity assumption on P guarantees that $\mathbb{I}[\tilde{C}_{\theta,V_{n_\ell}}^{M_{n_\ell}}] - \mathbb{I}[\tilde{C}_{\theta,V_0}^{M_0}] \to 0$ P-almost everywhere). Consequently,

$$\liminf_{n \to \infty} D_\theta(V_n, P) = \liminf_{n \to \infty} P[\tilde{C}_{\theta,V_n}^M] = \liminf_{\ell \to \infty} P[\tilde{C}_{\theta,V_{n_\ell}}^{M_{n_\ell}}] = P[\tilde{C}_{\theta,V_0}^{M_0}] \geq D_\theta(V_0, P).$$

We conclude that, if P is absolutely continuous with respect to the Lebesgue measure, then $V \to D_\theta(V, P)$ is also lower semicontinuous, hence continuous. $\qquad\square$

The proof of Theorem 2.2 requires the following result.

Lemma A.2. *Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Write $u_\theta^x = (x - \theta)/\|x - \theta\|$ if $x \neq \theta$ and 0 otherwise. For any $c \geq 0$, further let $t_{\theta,P}(c) = \sup_{v \in \mathcal{S}^{k-1}} P[|v^T u_\theta^X| \leq c]$, so that $t_{\theta,P} = t_{\theta,P}(0) = \sup_{v \in \mathcal{S}^{k-1}} P[v^T(X - \theta) = 0]$. Then, $t_{\theta,P}(c) \to t_{\theta,P}$ as $c \to 0$.*

Proof of Lemma A.2. Since $t_{\theta,P}(c)$ is increasing in c over $[0, \infty)$ and is larger than or equal to $t_{\theta,P}$ for any positive c, we have that $\tilde{t}_{\theta,P} = \lim_{c \to 0} t_{\theta,P}(c)$ exists and is such that $\tilde{t}_{\theta,P} \geq t_{\theta,P}$. Now, fix a decreasing sequence (c_n) converging to 0 and consider an arbitrary sequence (v_n) such that

$$P[|v_n^T u_\theta^X| \leq c_n] \geq t_{\theta,P}(c_n) - (1/n).$$

Since $\mathcal{S}^{k-1}$ is compact, we can consider a subsequence (v_{n_ℓ}) that converges to $v_0 (\in \mathcal{S}^{k-1})$; without loss of generality, we can of course assume that this subsequence is such that $(v_0^T v_{n_\ell})$ is an increasing sequence. Let then $C_\ell = \{v \in \mathcal{S}^{k-1} : v_0^T v \geq v_0^T v_{n_\ell}\}$. Clearly, C_ℓ is a decreasing sequence of sets with $\cap_\ell C_\ell = \{v_0\}$, so that

$$\lim_{\ell \to \infty} P[u_\theta^X \in \cup_{v \in C_\ell} \{|v^T y| \leq c_{n_\ell}\}] = P[u_\theta^X \in \{|v_0^T y| \leq 0\}] = P[|v_0^T u_\theta^X| = 0].$$

Now, for any ℓ, we have $P[u_\theta^X \in \cup_{v \in C_\ell} \{|v^T y| \leq c_{n_\ell}\}] \geq P[|v_{n_\ell}^T u_\theta^X| \leq c_{n_\ell}] \geq t_{\theta,P}(c_{n_\ell}) - (1/n_\ell)$, which implies that $t_{\theta,P} \geq P[|v_0^T u_\theta^X| = 0] \geq \tilde{t}_{\theta,P}$. $\qquad\square$

Proof of Theorem 2.2. Fix $V \in \mathcal{P}_{k,\mathrm{tr}}$ and denote as $\lambda_1(V)$ (resp., $\lambda_k(V)$) the largest (resp., smallest) eigenvalue of V (possible ties are unimportant below). Letting $v_1(V)$ and $v_k(V)$ be arbitrary corresponding unit eigenvectors, Lemma A.1 provides (with $M_V = v_1(V)v_1^T(V) \in \mathcal{M}_k^{\mathrm{all}}$)

$$
\begin{aligned}
D_\theta(V, P) &\leq P\Big[(u_{\theta,V}^X)^T M_V u_{\theta,V}^X \geq \tfrac{1}{k}\mathrm{tr}(M_V), X \neq \theta\Big] \\
&= P\Big[k\lambda_1^{-1}(V)\{v_1^T(V)(X - \theta)\}^2 \geq \|V^{-1/2}(X - \theta)\|^2, X \neq \theta\Big] \\
&\leq P\Big[\|V^{-1/2}(X - \theta)\|^2 \leq k\|X - \theta\|^2, X \neq \theta\Big] = P\big[\|V^{-1/2}u_{\theta,I_k}^X\|^2 \leq k, u_\theta^X \neq 0\big],
\end{aligned}
$$

where we used the inequality $\lambda_1(V) \geq 1$ (which follows from the constraint $\mathrm{tr}(V) = k$) and where u_θ^s is defined in Lemma A.2. Therefore,

$$
D_\theta(V, P) \leq P\big[\lambda_1(V^{-1})(v_1^T(V)u_\theta^X)^2 \leq k\big] \leq t_{\theta,P}\big((k\lambda_k(V))^{1/2}\big). \tag{A.3}
$$

Now, ad absurdum, take $\varepsilon > 0$ such that $R_\theta(t_{\theta,P} + \varepsilon, P)$ is unbounded. This implies that there exists a sequence (V_n) in $\mathcal{P}_{k,\mathrm{tr}}$ satisfying $D_\theta(V_n, P) \geq t_{\theta,P} + \varepsilon$ for any n and for which $d(V_n, I_k) \to \infty$. Since $\lambda_1(V_n) < \mathrm{tr}[V_n] = k$, we must have that $\lambda_k(V_n) \to 0$. Lemma A.2 and (A.3) then imply that $D_\theta(V_n, P) < t_{\theta,P} + \varepsilon$ for n large enough, a contradiction. Consequently, $R_\theta(\alpha, P)$ is bounded for any $\alpha > t_{\theta,P}$.

Now, Lemma C.1 in Paindaveine and Van Bever (2017a) readily implies that a bounded subset of $\mathcal{P}_{k,\mathrm{tr}}$ is also totally bounded, in the sense that, for any $\varepsilon > 0$, it can be covered by finitely many balls of the form $B(V, \varepsilon) = \{\tilde{V} \in \mathcal{P}_{k,\mathrm{tr}} : d(\tilde{V}, V) < \varepsilon\}$. Part (i) of the result and Theorem 2.1(ii) thus entail that, for any $\alpha > t_{\theta,P}$, the region $R_\theta(\alpha, P)$ is closed and totally bounded. The result then follows from the completeness of the metric space $(\mathcal{P}_{k,\mathrm{tr}}, d)$. $\qquad\square$

Proof of Theorem 2.3. Let $\alpha_* = \sup_{V \in \mathcal{P}_{k,\mathrm{tr}}} D_\theta(V, P)$. By assumption, $R_\theta(t_{\theta,P}, P)$ is non-empty. Thus, $\alpha_* \geq t_{\theta,P}$ and the result holds if $\alpha_* = t_{\theta,P}$. We may therefore assume that $\alpha_* > t_{\theta,P}$. For any n, pick then V_n in $R_\theta((\alpha_* - 1/n)_+, P)$, where $u_+ = \max(u, 0)$. Fix $\varepsilon \in (0, \alpha_* - t_{\theta,P})$. For n large enough, all terms of the sequence (V_n) belong to the compact set $R_\theta(\alpha_* - \varepsilon, P)$; see Theorem 2.2. Thus, there exists a subsequence (V_{n_k}) that converges in $R_\theta(\alpha_* - \varepsilon, P)$, to V_* say. For any $\varepsilon' \in (0, \varepsilon)$, all (V_{n_k}) eventually belong to the closed set $R_\theta(\alpha_* - \varepsilon', P)$, so that $V_* \in R_\theta(\alpha_* - \varepsilon', P)$. Therefore, $\alpha_* - \varepsilon' \leq D_\theta(V_*, P) \leq \alpha_*$ for any such ε', which establishes the result. $\qquad\square$

The proof of Theorem 2.4 requires the following preliminary result.

Lemma A.3. *For any $y \in \mathbb{R}^k$ and any $k \times k$ symmetric matrix M, the mapping $V \mapsto \mathrm{tr}(MV)y^T V^{-1}y$ is quasi-convex, that is, for any $V_a, V_b \in \mathcal{P}_{k,\mathrm{tr}}$ and any $t \in [0,1]$, $\mathrm{tr}(MV_t)y^T V_t^{-1}y \leq \max\{\mathrm{tr}(MV_a)y^T V_a^{-1}y, \mathrm{tr}(MV_b)y^T V_b^{-1}y\}$, with $V_t = (1-t)V_a + tV_b$.*

Proof. We treat two cases separately. (i) Assume first that $\mathrm{tr}(MV_a)\mathrm{tr}(MV_b) > 0$. Write

$$
\frac{V_t}{\mathrm{tr}(MV_t)} = (1 - s_t)\frac{V_a}{\mathrm{tr}(MV_a)} + s_t\frac{V_b}{\mathrm{tr}(MV_b)}, \quad \text{with } s_t = \frac{t\,\mathrm{tr}(MV_b)}{(1-t)\mathrm{tr}(MV_a) + t\,\mathrm{tr}(MV_b)}.
$$

17

Since $s_t \in [0, 1]$, the (weighted) harmonic-arithmetic matrix inequality (see, e.g., Lemma 2.1 (vii) in Lawson and Lim, 2013) then shows that, for any $y \in \mathbb{R}^k$,

$$y^T \left\{ \frac{V_t}{\text{tr}(MV_t)} \right\}^{-1} y \;\leq\; y^T \left[(1 - s_t) \left\{ \frac{V_a}{\text{tr}(MV_a)} \right\}^{-1} + s_t \left\{ \frac{V_b}{\text{tr}(MV_b)} \right\}^{-1} \right] y$$

$$\leq\; \max \left[y^T \left\{ \frac{V_a}{\text{tr}(MV_a)} \right\}^{-1} y, y^T \left\{ \frac{V_b}{\text{tr}(MV_b)} \right\}^{-1} y \right],$$

as was to be showed. (ii) Assume then that $\text{tr}(MV_a)\text{tr}(MV_b) \leq 0$. Without loss of generality, assume that $\text{tr}(MV_a) \leq 0$ and $\text{tr}(MV_b) \geq 0$. If $\text{tr}(MV_a) = \text{tr}(MV_b) = 0$, then $\text{tr}(MV_t) = 0$ for any t and the result trivially holds. Hence, we may assume that $\text{tr}(MV_a) \neq 0$ or $\text{tr}(MV_b) \neq 0$, which implies that $\text{tr}[MV_{t_0}] = 0$ for a unique $t_0 \in [0, 1]$. From continuity, pick then $\delta \in (0, 1 - t_0)$ such that, for any $t \in [t_0, t_0 + \delta)$,

$$\text{tr}(MV_t)y^T V_t^{-1} y \;\leq\; \text{tr}(MV_b)y^T V_b^{-1} y$$

$$\leq\; \max\{\text{tr}(MV_a)y^T V_a^{-1} y, \text{tr}(MV_b)y^T V_b^{-1} y\}.$$

By applying Part (i) of the proof with $V_{t_0+\delta}$ and V_b, we obtain that, for any $t \in [t_0+\delta, 1]$,

$$\text{tr}(MV_t)y^T V_t^{-1} y \;\leq\; \max\{\text{tr}(MV_{t_0+\delta})y^T V_{t_0+\delta}^{-1} y, \text{tr}(MV_b)y^T V_b^{-1} y\}$$

$$\leq\; \max\{\text{tr}(MV_a)y^T V_a^{-1} y, \text{tr}(MV_b)y^T V_b^{-1} y\}.$$

Since $\text{tr}(MV_t)y^T V_t^{-1} y \leq 0 \leq \max\{\text{tr}(MV_a)y^T V_a^{-1} y, \text{tr}(MV_b)y^T V_b^{-1} y\}$ for any $t \in [0, t_0]$, the result follows. $\qquad\square$

Proof of Theorem 2.4. (i) Write $V_t = (1 - t)V_a + tV_b$, where $V_a, V_b \in \mathcal{P}_{k,\text{tr}}$ and $t \in [0, 1]$ are fixed. First note that, letting $d_\theta^2(V) = (X - \theta)^T V^{-1}(X - \theta)$, Lemma A.1 yields

$$D_\theta(V, P) \;=\; \inf_{M \in \mathcal{M}_k^{\text{all}}} P\big[(X - \theta)^T V^{-1/2} M V^{-1/2}(X - \theta) \geq \tfrac{1}{k}\text{tr}(M)d_\theta^2(V), X \neq \theta\big]$$

$$=\; \inf_{M \in \mathcal{M}_k^{\text{all}}} P\big[(X - \theta)^T M (X - \theta) \geq \tfrac{1}{k}\text{tr}(MV)d_\theta^2(V), X \neq \theta\big]. \qquad (A.4)$$

Writing again $V_t = (1 - t)V_a + tV_b$, Lemma A.3 thus yields that, for any $M \in \mathcal{M}_k^{\text{all}}$,

$$P\big[(X - \theta)^T M (X - \theta) \geq \tfrac{1}{k}\text{tr}(MV_t)d_\theta^2(V_t), X \neq \theta\big]$$

$$\geq P\big[(X - \theta)^T M (X - \theta) \geq \tfrac{1}{k}\max\{\text{tr}(MV_a)d_\theta^2(V_a), \text{tr}(MV_b)d_\theta^2(V_b)\}, X \neq \theta\big]$$

$$= \min \Big(P\big[(X - \theta)^T M (X - \theta) \geq \tfrac{1}{k}\text{tr}(MV_a)d_\theta^2(V_a), X \neq \theta\big],$$

$$P\big[(X - \theta)^T M (X - \theta) \geq \tfrac{1}{k}\text{tr}(MV_b)d_\theta^2(V_b), X \neq \theta\big] \Big)$$

$$\geq \min(D_\theta(V_a, P), D_\theta(V_b, P)).$$

The result then follows from (A.4). (ii) If $V_a, V_b \in R_\theta(\alpha, P)$, then Part (i) of the result entails that $D_\theta((1-t)V_a + tV_b, P) \geq \min\{D_\theta(V_a, P), D_\theta(V_b, P)\} \geq \alpha$, so that $(1-t)V_a + tV_b \in R_\theta(\alpha, P)$. $\qquad\square$

The proof of Theorem 2.5 requires both following lemmas.

Lemma A.4. *Let P be elliptical over $\mathbb{R}^k$ with location 0 and shape I_k. Then, $D_0(I_k, P) = (1 - P[\{0\}])P[U_1^2 > 1/k]$, where $U = (U_1, \ldots, U_k)^T$ is uniformly distributed over the unit sphere $\mathcal{S}^{k-1}$.*

Lemma A.5. *Let P be elliptical over $\mathbb{R}^k$ with location 0 and shape I_k. Then, for any $V \in \mathcal{P}_{k,\mathrm{tr}} \setminus \{I_k\}$, $D_0(V, P) < (1 - P[\{0\}])P[U_1^2 > 1/k]$, where $U = (U_1, \ldots, U_k)^T$ is uniformly distributed over $\mathcal{S}^{k-1}$.*

Proof of Lemma A.4. In the spherical setup considered, we have that, for any $M \in \mathcal{M}_k^{\mathrm{all}}$,

$$P\left[\frac{X^T M X}{\|X\|^2} \geq \frac{1}{k}\operatorname{tr}(M),\, X \neq 0\right] = P\left[U^T M U \geq \tfrac{1}{k}\operatorname{tr}(M)\right]P[X \neq 0],$$

where $U = (U_1, \ldots, U_k)^T$ is uniform over $\mathcal{S}^{k-1}$. Lemma A.1 then entails that

$$D_0(I_k, P) = (1 - P[\{0\}])\inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\left[U^T M U \geq \tfrac{1}{k}\operatorname{tr}(M)\right].$$

Decomposing M into $O\Lambda O^T$, where O is a $k \times k$ orthogonal matrix and where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ is a diagonal matrix, this yields

$$D_0(I_k, P) = (1 - P[\{0\}])\inf_{\lambda \in \mathbb{R}^k} P\left[\sum_{\ell=1}^k \lambda_\ell U_\ell^2 \geq \frac{1}{k}\sum_{\ell=1}^k \lambda_\ell\right]$$

$$= (1 - P[\{0\}])\inf_{\lambda \in \mathbb{R}^k} P\left[\sum_{\ell=1}^k \lambda_\ell\left(U_\ell^2 - \frac{1}{k}\right) \geq 0\right] =: (1 - P[\{0\}])\inf_{\lambda \in \mathbb{R}^k} p(\lambda).$$

By using successively the facts that $p(0) = 1$ and $p(\lambda) = p(\lambda/\|\lambda\|)$ for any $\lambda \in \mathbb{R}^k \setminus \{0\}$, we obtain

$$D_0(I_k, P) = (1 - P[\{0\}])\inf_{\lambda \in \mathbb{R}^k \setminus \{0\}} p(\lambda) = (1 - P[\{0\}])\inf_{\lambda \in \mathcal{S}^{k-1}} p(\lambda). \tag{A.5}$$

The result then follows from Theorem 2 from Paindaveine and Van Bever (2017b), that states that the last infimum in (A.5) is equal to $P[U_1^2 > 1/k]$. $\square$

Proof of Lemma A.5. Fix $V \in \mathcal{P}_{k,\mathrm{tr}}$ and let X be a random k-vector with $P = P^X$. Write $V = O\Lambda O^T$, where O is a $k \times k$ orthogonal matrix and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ is a diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$. Affine invariance (see Theorem 2.6) entails that

$$D_0(V, P^X) = D_0\left(O^T V O, P^{O^T X}\right) = D_0\left(\Lambda, P^X\right).$$

Denoting by e_1 the first vector of the canonical basis of $\mathbb{R}^{k^2}$, we then have

$$D_0(V, P^X) = D_0\left(\Lambda, P^X\right) \leq P\left[e_1^T \operatorname{vec}\left(U_{0,\Lambda} U_{0,\Lambda}^T - \tfrac{1}{k}I_k\right) \geq 0\right] = P\left[((U_{0,\Lambda})_1)^2 - \tfrac{1}{k} \geq 0\right]$$

$$= P\left[((U_{0,\Lambda})_1)^2 - \tfrac{1}{k} \geq 0,\, X \neq 0\right] = P\left[\lambda_1^{-1} X_1^2 \geq \tfrac{1}{k}\sum_{\ell=1}^k \lambda_\ell^{-1} X_\ell^2,\, X \neq 0\right] \tag{A.6}$$

$$\leq P\left[X_1^2 \geq \tfrac{1}{k}\sum_{\ell=1}^k X_\ell^2,\, X \neq 0\right] = P\left[X_1^2/\|X\|^2 \geq \tfrac{1}{k},\, X \neq 0\right] = P[X \neq 0]P\left[U_1^2 \geq \tfrac{1}{k}\right],$$

19

where $U = (U_1, \ldots, U_k)^T$ is uniform over $\mathcal{S}^{k-1}$. To have $D_0(V, P^X) = P[X \neq 0]P\left[U_1^2 \geq \frac{1}{k}\right]$, the inequality in (A.6) needs to be an equality, which requires that $\lambda_\ell = \lambda_1$ for all ℓ, hence that $V = I_k$. $\qquad\square$

We can now prove Theorem 2.5.

Proof of Theorem 2.5. Lemmas A.4-A.5 establish the result in the spherical case associated with $\theta_0 = 0$ and $V_0 = I_k$. For general values of θ_0 and V_0, note that $Y = V_0^{-1/2}(X - \theta_0)$ is elliptical with location 0, shape I_k, and satisfies $P[Y = 0] = P[X = \theta_0]$. By affine invariance,

$$D_{\theta_0}(V, P^X) = D_0\left(\frac{kV_0^{-1/2}VV_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2})}, P^Y\right)$$

$$\leq D_0(I_k, P^Y) = D_0\left(\frac{kV_0^{-1/2}V_0V_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}V_0V_0^{-1/2})}, P^Y\right) = D_{\theta_0}(V_0, P^X),$$

with equality if and only if $kV_0^{-1/2}VV_0^{-1/2}/\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2}) = I_k$, that is, if and only if $V = V_0$. $\qquad\square$

Proof of Theorem 2.6. In the proof of Theorem 2.4, we showed that

$$D_\theta(V, P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\left[(X-\theta)^T V^{-1/2}MV^{-1/2}(X-\theta) \geq \tfrac{1}{k}\mathrm{tr}(M)(X-\theta)^T V^{-1}(X-\theta), X \neq \theta\right].$$

Using the fact that $V_A^{1/2} = k^{1/2}AV^{1/2}O/\{\mathrm{tr}(AVA^T)\}^{1/2}$ for some $k \times k$ orthogonal matrix O (recall that $V_A^{1/2}$ stands for the symmetric positive definite square root of V_A), this readily yields

$$D_{A\theta+b}(V_A, P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[(X - \theta)^T V^{-1/2}OMO^T V^{-1/2}(X - \theta)$$

$$\geq \tfrac{1}{k}\mathrm{tr}(OMO^T)(X - \theta)^T V^{-1}(X - \theta), X \neq \theta\big] = D_\theta(V, P),$$

as was to be showed. The affine-equivariance property of the depth regions readily follows. $\qquad\square$

The proof of Theorem 2.7 requires the following lemma (the proof is straightforward, hence is omitted).

Lemma A.6. *For any v_1, v_2 such that $v_1^2 + v_2^2 < 1$, we have*

$$\frac{(1 - v_1^2)^{1/2} - |v_2|}{(1 - v_1^2)^{1/2} + |v_2|} \leq \frac{(1 - v_1^2)^{1/2} + |v_2|}{(1 - v_1^2)^{1/2} - |v_2|} \leq \frac{1 + (v_1^2 + v_2^2)^{1/2}}{1 - (v_1^2 + v_2^2)^{1/2}}.$$

Proof of Theorem 2.7. (i) If $P = P^X$ is elliptical with location θ_0 and shape V_0, then $V_0^{-1/2}(X - \theta_0)$ is equal in distribution to RU, where U is uniformly distributed over the unit sphere $\mathcal{S}^{k-1}$ and is independent of the nonnegative random variable R. Theorem 2.6 then yields

$$D_{\theta_0}(V, P^X) = D_0\left(\frac{kV_0^{-1/2}VV_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2})}, P^{RU}\right). \tag{A.7}$$

Now, for any $\tilde{V} \in \mathcal{P}_{k,\mathrm{tr}}$, Lemma A.1 entails that

$$D_0(\tilde{V}, P^{RU}) = \inf_{M \in \mathcal{M}_k^0} P\left[U^T \tilde{V}^{-1/2} M \tilde{V}^{-1/2} U \geq 0, R > 0\right] \tag{A.8}$$

$$= P[R > 0] \inf_{M \in \mathcal{M}_k^0} P\left[U^T \tilde{V}^{-1/2} M \tilde{V}^{-1/2} U \geq 0\right] = P[R > 0] D_0(\tilde{V}, P^U).$$

Combining with (A.7), we obtain

$$D_{\theta_0}(V, P^X) = (1 - P^X[\{\theta_0\}]) D_0\left(\frac{kV_0^{-1/2}VV_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2})}, P^U\right),$$

which establishes Part (i) of the result. (ii) Assume that $P = P^X$ is bivariate standard normal and fix $V \in \mathcal{P}_{2,\mathrm{tr}}$. We aim at evaluating

$$D_0(V, P^X) = \inf_{M \in \mathcal{M}_k^0} P\left[X^T V^{-1/2} M V^{-1/2} X \geq 0\right]; \tag{A.9}$$

see (A.8). To do so, it will be convenient to parametrise V and the matrix M from as

$$V = \begin{pmatrix} 1 + v_1 & v_2 \\ v_2 & 1 - v_1 \end{pmatrix} \quad \text{and} \quad M = m_1 \begin{pmatrix} 1 & m_2 \\ m_2 & -1 \end{pmatrix},$$

with $v_1^2 + v_2^2 < 1$ and $m_1 \neq 0$ (note indeed that $m_1 = 0$ makes the probability in (A.9) equal to one, which cannot be the infimum). Decomposing $V^{-1/2} M V^{-1/2}$ into $O\Lambda O^T$, where O is a 2×2 orthogonal matrix and where $\Lambda = \mathrm{diag}(\lambda_1(V^{-1}M), \lambda_2(V^{-1}M))$, with $\lambda_1(V^{-1}M) \geq \lambda_2(V^{-1}M)$, involves the eigenvalues of $V^{-1}M$ (equivalently, of $V^{-1/2} M V^{-1/2}$), we have

$$D_0(V, P) = \inf_{(m_1, m_2) \in \mathbb{R}_0 \times \mathbb{R}} P\left[\lambda_1(V^{-1}M)X_1^2 + \lambda_2(V^{-1}M)X_2^2 \geq 0\right], \tag{A.10}$$

where $X = (X_1, X_2)^T$ is still bivariate standard normal. Since $\lambda_1(-V^{-1}M) = -\lambda_2(V^{-1}M)$ for any $M \in \mathcal{M}_k^0$ (we will show below that $\lambda_2(V^{-1}M) < 0 < \lambda_1(V^{-1}M)$ for any $M \in \mathcal{M}_k^0$), we have

$$D_0(V, P) = \min\left(\inf_{(m_1, m_2) \in \mathbb{R}_0^+ \times \mathbb{R}} P\left[\lambda_1(V^{-1}M)X_1^2 + \lambda_2(V^{-1}M)X_2^2 \geq 0\right], \tag{A.11}\right.$$

$$\left.\inf_{(m_1, m_2) \in \mathbb{R}_0^+ \times \mathbb{R}} P\left[\lambda_1(V^{-1}M)X_1^2 + \lambda_2(V^{-1}M)X_2^2 \leq 0\right]\right),$$

21

which allows us to restrict to positive values of m_1. A direct computation shows that, for $m_1 > 0$,

$$\lambda_1(V^{-1}M) = \frac{m_1}{\det V}\left[-(v_1 + m_2 v_2) + \{(v_1 + m_2 v_2)^2 + (1 + m_2^2)\det V\}^{1/2}\right] > 0$$

and

$$\lambda_2(V^{-1}M) = \frac{m_1}{\det V}\left[-(v_1 + m_2 v_2) - \{(v_1 + m_2 v_2)^2 + (1 + m_2^2)\det V\}^{1/2}\right] < 0.$$

Since $f(m_2) = -\lambda_2(V^{-1}M))/\lambda_1(V^{-1}M)$ does not depend on m_1, (A.11) leads to

$$
\begin{aligned}
D_0(V, P) &= \min\left(P\left[\frac{X_1^2}{X_2^2} \geq \sup_{m_2 \in \mathbb{R}} f(m_2)\right], P\left[\frac{X_1^2}{X_2^2} \leq \inf_{m_2 \in \mathbb{R}} f(m_2)\right]\right) \\
&= P\left[\frac{X_1^2}{X_2^2} \geq \max\left(\sup_{m_2 \in \mathbb{R}} f(m_2)\,,\, 1 \Big/ \inf_{m_2 \in \mathbb{R}} f(m_2)\right)\right]. \qquad \text{(A.12)}
\end{aligned}
$$

It is easy to check that f is differentiable over $\mathbb{R}$ with a derivative of the form $c_{v_1,v_2}(m_2)(v_2 - v_1 m_2)$, where $c_{v_1,v_2}(m_2) > 0$ for any m_2, and that

$$f(\pm\infty) = \lim_{m_2 \to \pm\infty} f(m_2) = \frac{\sqrt{1 - v_1^2} \pm v_2}{\sqrt{1 - v_1^2} \mp v_2}.$$

We treat the cases $v_1 = 0$ and $v_1 \neq 0$ separately.

(a) Assume that $v_1 = 0$. If $v_2 = 0$, then $V = I_2$ and Theorem 2.5 establishes the result. If $v_2 \neq 0$, then f has no critical point and

$$\sup_{m_2 \in \mathbb{R}} f(m_2) = \max\left(f(-\infty), f(\infty)\right) = \frac{1 + |v_2|}{1 - |v_2|}$$

and

$$\inf_{m_2 \in \mathbb{R}} f(m_2) = \min\left(f(-\infty), f(\infty)\right) = \frac{1 - |v_2|}{1 + |v_2|},$$

so that (A.12) yields

$$D_0(V, P) = P\left[\frac{X_1^2}{X_2^2} \geq \frac{1 + |v_2|}{1 - |v_2|}\right] = P\left[\frac{X_1^2}{X_2^2} \geq \frac{1 + (1 - \det V)^{1/2}}{1 - (1 - \det V)^{1/2}}\right] = P\left[Y_2 \geq \frac{1}{2} + \frac{1}{2}\{1 - \det(V)\}^{1/2}\right],$$

where we have used the fact that if Z has a $F(1,1)$ Fisher-Snedecor distribution, then $Z/(1 + Z)$ has a Beta$(1/2, 1/2)$ distribution.

(b) Assume now that $v_1 \neq 0$. Then the only critical point of f is $m_2^{\mathrm{crit}} = v_2/v_1$, so that, irrespective of the fact that this critical point is a local minimum/maximum of f,

$$\sup_{m_2 \in \mathbb{R}} f(m_2) = \max\left(f(-\infty), f(\infty), f(m_2^{\mathrm{crit}})\right) = \max\left(\frac{(1 - v_1^2)^{1/2} + |v_2|}{(1 - v_1^2)^{1/2} - |v_2|}, \frac{\mathrm{Sign}(v_1) + (v_1^2 + v_2^2)^{1/2}}{\mathrm{Sign}(v_1) - (v_1^2 + v_2^2)^{1/2}}\right)$$

and

$$\inf_{m_2 \in \mathbb{R}} f(m_2) = \min\left(f(-\infty), f(\infty), f(m_2^{\mathrm{crit}})\right) = \min\left(\frac{(1-v_1^2)^{1/2} - |v_2|}{(1-v_1^2)^{1/2} + |v_2|}, \frac{\mathrm{Sign}(v_1) + (v_1^2 + v_2^2)^{1/2}}{\mathrm{Sign}(v_1) - (v_1^2 + v_2^2)^{1/2}}\right).$$

Lemma A.6 yields

$$\sup_{m_2 \in \mathbb{R}} f(m_2) = \frac{(1-v_1^2)^{1/2} + |v_2|}{(1-v_1^2)^{1/2} - |v_2|}\, \mathbb{I}[v_1 < 0] + \frac{1 + (v_1^2 + v_2^2)^{1/2}}{1 - (v_1^2 + v_2^2)^{1/2}}\, \mathbb{I}[v_1 > 0]$$

and

$$\inf_{m_2 \in \mathbb{R}} f(m_2) = \frac{-1 + (v_1^2 + v_2^2)^{1/2}}{-1 - (v_1^2 + v_2^2)^{1/2}}\, \mathbb{I}[v_1 < 0] + \frac{(1-v_1^2)^{1/2} - |v_2|}{(1-v_1^2)^{1/2} + |v_2|}\, \mathbb{I}[v_1 > 0],$$

hence also

$$\max\left(\sup_{m_2 \in \mathbb{R}} f(m_2)\,,\, 1 \Big/ \inf_{m_2 \in \mathbb{R}} f(m_2)\right) = \frac{1 + (v_1^2 + v_2^2)^{1/2}}{1 - (v_1^2 + v_2^2)^{1/2}} = \frac{1 + (1 - \det(V))^{1/2}}{1 - (1 - \det(V))^{1/2}}.$$

Therefore, (A.12) finally provides

$$D_0(V, P) = P\left[\frac{X_1^2}{X_2^2} \geq \frac{1 + (1 - \det V)^{1/2}}{1 - (1 - \det V)^{1/2}}\right] = P\left[Y_2 \geq \frac{1}{2} + \frac{1}{2}\{1 - \det(V)\}^{1/2} x\right].$$

This proves the result for the case where P is bivariate standard normal. The general result then follows from Part (i) of the Theorem. $\square$

Proof of Theorem 3.1. Let P and Q be two probability measures over $\mathbb{R}^k$ and fix $V \in \mathcal{P}_{k,\mathrm{tr}}$. Fix $\varepsilon > 0$ and assume, without loss of generality, that $D_\theta(V, P) \leq D_\theta(V, Q)$. Lemma A.1 entails that there exists $M_0 \in \mathcal{M}_k^0$ such that $P[C_{\theta,V}^{M_0}] \leq D_\theta(V, P) + \varepsilon$, where we still use the notation $C_{\theta,V}^M = \{x \in \mathbb{R}^k \setminus \{\theta\} : (u_{\theta,V}^x)^T M u_{\theta,V}^x \geq \frac{1}{k}\mathrm{tr}(M)\}$. Consequently, using Lemma A.1 again,

$$|D_\theta(V, Q) - D_\theta(V, P)| = D_\theta(V, Q) - D_\theta(V, P)$$

$$\leq Q[C_{\theta,V}^{M_0}] - P[C_{\theta,V}^{M_0}] + \varepsilon \leq \sup_{C \in \mathcal{C}_\theta} |Q[C] - P[C]| + \varepsilon,$$

with $\mathcal{C}_\theta = \{C_{\theta,V}^M : M \in \mathcal{M}_k^0, V \in \mathcal{P}_{k,\mathrm{tr}}\}$. Since this holds for any $\varepsilon > 0$ and for any $V \in \mathcal{P}_{k,\mathrm{tr}}$, we have

$$\sup_{V \in \mathcal{P}_{k,\mathrm{tr}}} |D_\theta(V, Q) - D_\theta(V, P)| \leq \sup_{C \in \mathcal{C}_\theta} |Q[C] - P[C]|.$$

It thus only remains to show that $\mathcal{C}_\theta$ is a Vapnik-Chervonenkis class. To do so, note that $C_{\theta,V}^M = \{x \in \mathbb{R}^k \setminus \{\theta\} : (x - \theta)^T V^{-1/2} M V^{-1/2}(x - \theta) \geq 0\}$, so that $\mathcal{C}_\theta \subset \{D_{\theta,A} \cap (\mathbb{R}^k \setminus \{\theta\}) : A \in \mathcal{M}_k^{\mathrm{all}}\}$, with $D_{\theta,A} = \{x \in \mathbb{R}^k : (x - \theta)^T A(x - \theta) \geq 0\}$. Theorem 4.6 from Dudley (2014) implies that $\{D_{\theta,A} : A \in \mathcal{M}_k^{\mathrm{all}}\}$ is a Vapnik-Chervonenkis class $\mathcal{D}_\theta$. It then follows from Lemma 2.6.17(ii) in Van der Vaart and Wellner (1996) that $\{D_{\theta,A} \cap (\mathbb{R}^k \setminus \{\theta\}) : A \in \mathcal{M}_k^{\mathrm{all}}\}$, hence also $\mathcal{C}_\theta$, is a Vapnik-Chervonenkis class. $\square$

Proof of Theorem 3.2. Recall from (2.1) that $V_{\theta,P}$ is defined as the barycentre of $R_\theta(\alpha_*, P)$, with $\alpha_* = \max_V D_\theta(V, P)$. The mapping $V \mapsto D_\theta(V, P)$ is upper semicontinuous (Theorem 2.1) and constant over $R_\theta(\alpha_*, P)$. Clearly, it is easy to define a mapping $V \mapsto \tilde{D}_\theta(V, P)$ that is upper semicontinuous, agrees with $V \mapsto D_\theta(V, P)$ in the complement of $R_\theta(\alpha_*, P)$, and for which $V_{\theta,P}$ is the unique maximiser. By using Theorem 3.1, the result then follows from Theorem 2.12 and Lemma 14.3 in Kosorok (2008). $\qquad\square$

Proof of Theorem 5.1. Write $L_{\theta,V} = U_{\theta,V} U_{\theta,V}^T - \frac{1}{k} I_k$. Since $(L_{\theta,V})_{11} = -\sum_{\ell=2}^{k} (L_{\theta,V})_{\ell\ell}$, there exists a $(d_k + 1) \times d_k$ full-rank matrix H_0 such that $\mathrm{vech}(L_{\theta,V}) = H_0 \,\mathrm{vech}_0(L_{\theta,V})$. Therefore, there exists a $k^2 \times d_k$ full-rank matrix H (one can take $H = D H_0$, where D is the usual duplication matrix) such that $W_{\theta,V} = \mathrm{vec}(L_{\theta,V}) = H \tilde{W}_{\theta,V}$. It follows that

$$D_\theta(V, P) = D(0, P^{W_{\theta,V}}) = \inf_{u \in \mathbb{R}^{k^2}} P\big[u^T W_{\theta,V} \geq 0\big]$$

$$= \inf_{u \in \mathbb{R}^{k^2}} P\big[(H^T u)^T \tilde{W}_{\theta,V} \geq 0\big] = \inf_{v \in \mathbb{R}^{d_k}} P\big[v^T \tilde{W}_{\theta,V} \geq 0\big] = D(0, P^{\tilde{W}_{\theta,V}}),$$

where we used the fact that H^T has full column rank. $\qquad\square$

The proof of Theorem 6.1(i) is long and technical, but follows along the same lines as the proof of Theorem 2.2 in Paindaveine and Van Bever (2017a). As for the proof of Theorem 6.1(ii), it is strictly the same as that of Theorem 3.2. We therefore omit the proof of Theorem 6.1.

References

Bhatia, R. (2007) *Positive definite matrices.* Princeton Series in Applied Mathematics. Princeton, NJ: Princeton University Press.

Bhatia, R. and Holbrook, J. (2006) Riemannian geometry and matrix geometric means. *Linear Algebra Appl.*, **413**, 594–618.

Chen, M., Gao, C. and Ren, Z. (2015) Robust covariance matrix estimation via matrix depth. arXiv:1506.00691v3.

Dudley, R. M. (2014) *Uniform Central Limit Theorems.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2nd edition edn.

Dudley, R. M., Sidenko, S. and Wang, Z. (2009) Differentiability of t-functionals of location and scatter. *Ann. Statist.*, **37**, 939–960.

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

Dümbgen, L., Pauly, M. and Schweizer, T. (2015) m-functionals of multivariate scatter. *Statist. Surv.*, **9**, 32–105.

Hallin, M. and Paindaveine, D. (2006a) Parametric and semiparametric inference for shape: the role of the scale functional. *Statist. Decisions*, **24**, 327–350.

— (2006b) Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.*, **34**, 2707–2756.

Hallin, M., Paindaveine, D. and Siman, M. (2010) Multivariate quantiles and multiple-output regression quantiles: From l_1 optimization to halfspace depth (with discussion). *Ann. Statist.*, **38**, 635–669.

He, X., Simpson, D. and Portnoy, S. (1990) Breakdown robustness of tests. *J. Amer. Statist. Assoc.*, **85**, 446–452.

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

John, S. (1972) The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, **59**, 169–173.

Kent, J. and Tyler, D. E. (1988) Maximum likelihood estimation for the wrapped cauchy distribution. *J. Appl. Statist.*, **15**, 247–254.

— (1991) Redescending m-estimates of multivariate location and scatter. *Ann. Statist.*, **19**, 2102–2119.

Kosorok, M. R. (2008) *Introduction to Empirical Processes and Semiparametric Inference.* Springer Series in Statistics. New York: Springer.

Lawson, J. and Lim, Y. (2013) Weighted means and karcher equations of positive operators. *Proc. Natl. Acad. Sci. USA*, **110**, 15626–15632.

Liu, R. Y., Parelius, J. M. and Singh, K. (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion). *Ann. Statist.*, **27**, 783–858.

Mizera, I. (2002) On depth and deep points: a calculus. *Ann. Statist.*, **30**, 1681–1736.

Mizera, I. and Müller, C. H. (2004) Location-scale depth (with discussion). *J. Amer. Statist. Assoc.*, **99**, 949–989.

Müller, C. H. (2005) Depth estimators and tests based on the likelihood principle with application to regression. *J. Multivariate Anal.*, **95**, 153–181.

Paindaveine, D. (2008) A canonical definition of shape. *Statist. Probab. Lett.*, **78**, 2240–2247.

Paindaveine, D. and Van Bever, G. (2014) Inference on the shape of elliptical distributions based on the mcd. *J. Multivariate Anal.*, **129**, 125–144.

25

— (2017a) Halfspace depth for scatter, concentration and shape matrices. *arXiv:1704.06160*

— (2017b) On the maximal halfspace depth of permutation-invariant distributions on the simplex. *arXiv:1705.04974*

Rousseeuw, P. J. and Ruts, I. (1999) The depth function of a population distribution. *Metrika*, **49**, 213–244.

Tukey, J. W. (1975) Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, 523–531. Canad. Math. Congress, Montreal, Que.

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

Van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York: Springer.

Zhang, J. (2002) Some extensions of Tukey's depth function. *J. Multivariate Anal.*, **82**, 134–165.

Zuo, Y. and Serfling, R. (2000) General notions of statistical depth function. *Ann. Statist.*, **28**, 461–482.