Inference on the Shape of Elliptical Distributions based on the MCD

Davy Paindaveine[a,1], Germain Van Bever[b]

[a] *Université libre de Bruxelles, ECARES and Département de Mathématique, Avenue F. D. Roosevelt, 50, CP 114/04, B-1050 Brussels, Belgium*

[b] *Université libre de Bruxelles, ECARES and Département de Mathématique, Campus de la Plaine, Boulevard du Triomphe, CP 210, B-1050 Brussels, Belgium*

Abstract

The minimum covariance determinant (MCD) estimator of scatter is one of the most famous robust procedures for multivariate scatter. Despite the quite important research activity related to this estimator, culminating in the recent thorough asymptotic study of Cator & Lopuhaä (2010, 2012), no results have been obtained on the corresponding estimator of *shape*, which is the parameter of interest in many multivariate problems (including principal component analysis, canonical correlation analysis, testing for sphericity, etc.) In this paper, we therefore propose and study MCD-based inference procedures for shape, that inherit the good robustness properties of the MCD. The main emphasis is on asymptotic results, for point estimation (Bahadur representation and asymptotic normality results) as well as for hypothesis testing (asymptotic distributions under the null and under local alternatives). Influence functions of the MCD-estimators of shape are obtained as a corollary. Monte-Carlo studies illustrate our asymptotic results and assess the robustness of the proposed procedures.

Keywords: Bahadur representation results; elliptical distributions; MCD estimators; robustness; shape parameters; tests of sphericity

1. Introduction

The minimum covariance determinant (MCD) estimators of location and scatter, that were introduced in Rousseeuw (1985), are among the most famous estimators in robust statistics. Assuming that k-variate observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are available, the MCD estimators of location $\hat{\boldsymbol{\theta}}_\gamma$ and scatter $\hat{\boldsymbol{\Sigma}}_\gamma$, for any $\gamma \in (0, 1]$, are defined as the sample average and covariance matrix computed from "the"[2] subsample leading to a covariance matrix with smallest determinant over the collection of all possible subsamples of size larger than or equal to $\lceil n\gamma \rceil$ (it was shown in Cator & Lopuhaä (2012) that the smallest determinant is always obtained for a subsample of size $\lceil n\gamma \rceil$).

Despite their relatively poor efficiency under multinormality, MCD estimators have been quite successful. This is explained by their very good robustness properties: for appropriately chosen γ, MCD estimators indeed show the highest breakdown points that can be achieved in the class of

Email addresses: dpaindav@ulb.ac.be (Davy Paindaveine), gvbever@ulb.ac.be (Germain Van Bever)

[1] Corresponding author.

[2] Uniqueness is actually not guaranteed.

affine-equivariant estimators; see Lopuhaä & Rousseeuw (1991) and Agullò et al. (2008). Another advantadge over competing methods is that they can be computed very efficiently through the so-called FAST-MCD algorithm from Rousseeuw & Van Driessen (1999) (that is available in the R package *MASS*). This holds for relatively high dimensions, where Rousseeuw & Van Driessen (1999) could treat a dataset involving up to $n = 137,256$ observations with $k = 27$ variables.

Asymptotic results were slow to come. Within the framework of elliptical distributions, Butler et al. (1993) established strong consistency of $\hat{\boldsymbol{\theta}}_\gamma$ and $\hat{\boldsymbol{\Sigma}}_\gamma$, as well as asymptotic normality (at the standard root-n rate) of $\hat{\boldsymbol{\theta}}_\gamma$. Croux & Haesbroeck (1999) computed the influence function of $\hat{\boldsymbol{\Sigma}}_\gamma$, and, assuming the validity of the usual von Mises expansion linking estimators and their influence functions, deduced the asymptotic covariance matrix of $\sqrt{n}\,\hat{\boldsymbol{\Sigma}}_\gamma$ in the elliptical setup. Recently, Cator & Lopuhaä (2010, 2012) showed that this von Mises expansion indeed holds under very broad distributional assumptions, which provides as a corollary the first proof of aymptotic normality for $\hat{\boldsymbol{\Sigma}}_\gamma$ (and validates the asymptotic covariance computation of Croux & Haesbroeck (1999)); their results apply in particular in the context of elliptical densities.

It is argued in Cator & Lopuhaä (2010, 2012) that, beyond their initial purpose to estimate location and scatter, the MCD estimators, in particular $\hat{\boldsymbol{\Sigma}}_\gamma$, also serve as robust plug-ins in other multivariate statistical techniques. It is often the case, however, that these techniques do only require to know or to estimate the scatter matrix up to a positive scalar factor. In other words, factorizing the population scatter matrix $\boldsymbol{\Sigma}$ into $\sigma^2\mathbf{V}$, where $\sigma^2 = (\det \boldsymbol{\Sigma})^{1/k}$ is a *scale* parameter and $\mathbf{V} = \boldsymbol{\Sigma}/(\det \boldsymbol{\Sigma})^{1/k}$ is a *shape* parameter, it is often so that the parameter of interest is $\mathbf{V}$ (with dimension $K := k(k+1)/2 - 1$), while σ^2 plays the role of a nuisance. In principal component analysis, for instance, principal directions may be interchangeably computed from $\boldsymbol{\Sigma}$ or from $\mathbf{V}$, and both scatter and shape matrices will lead to the same proportions of explained variance. Other factorizations of scatter into scale $\times$ shape are possible, such as those based on $\sigma^2 = (\mathrm{tr}\,\boldsymbol{\Sigma})/k$ or on $\sigma^2 = \Sigma_{11}$ that lead to shape matrices with fixed trace k or upper-left entry equal to one, respectively.

There have been many recent works developing specific inference procedures for shape; see, among others, Hallin & Paindaveine (2006b), Hallin et al. (2006), Frahm (2009), and Taskinen et al. (2010). For many robust scatter estimators, the corresponding estimators of shape have been studied. In particular, a quite systematic investigation of the properties of robust estimators of shape has been performed in Frahm (2009), where M-, S-, and R-estimators of shape are considered.

To the best of our knowledge, however, MCD-estimators of shape have not been considered, which may seem surprising in view of (i) the importance of the MCD estimators of (location and) scatter in robust statistics and (ii) the continued research related to the MCD. The goal of this paper is therefore to provide, in the elliptical case, MCD estimators and tests for shape, that inherit the good robustness properties of the MCD. Emphasis is put on asymptotic results (Bahadur representation and asymptotic normality results, for point estimation, and asymptotic distribution under the null and under local alternatives, for hypothesis testing). Influence functions of the MCD-estimators of shape considered will also be obtained as a corollary. Rather than adopting a particular definition of shape (e.g., the determinant-based or trace-based definitions above), we throughout derive our results for a generic shape concept.

The outline of the paper is as follows. In Section 2, we first introduce the notation and assumption we will need on elliptical densities, and then state, in a form that is adapted to our purposes, the Cator & Lopuhaä (2010) Bahadur representation result for $\hat{\boldsymbol{\Sigma}}_\gamma$. In Section 3, we introduce and discuss the concept of shape based on a general "scale functional". In Section 4, we develop MCD-based inference procedures for shape; point estimation and hypothesis testing are considered in Sections 4.1 and 4.2, respectively. In Section 5, we describe how to estimate consistently the nuisance parameters involved in these procedures, which is required for their practical implementation. Section 6 derives the corresponding result for the procedures based on the empirical covariance matrix, which allows to obtain asymptotic relative efficiencies of the MCD shape procedures with respect to these covariance-based competitors. Monte-Carlo studies are conducted in Section 7 in order to confirm our asymptotic results and to assess the robustness properties of the proposed procedures. Finally, the Appendix collects technical proofs.

2. Elliptical densities and MCD

Let $\mathcal{S}_k$ be the collection of $k \times k$ symmetric and positive definite matrices, and let $\mathcal{F}$ be the collection of functions from $\mathbb{R}^+$ to $\mathbb{R}^+$ that satisfy the integrability condition $\mu_{k-1,f} < \infty$, where we wrote $\mu_{\ell,f} = \int_0^\infty r^\ell f(r)\,dr$. The random k-vector $\mathbf{X}$ is said to be elliptically symmetric with *location* $\boldsymbol{\theta}\,(\in \mathbb{R}^k)$, *scatter* $\boldsymbol{\Sigma}\,(\in \mathcal{S}_k)$, and *radial density* $f \in \mathcal{F}$ (this will be denoted as $\mathbf{X} \sim \mathrm{Ell}_k(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$) if it is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^k$, with density

$$
\begin{aligned}
f^{\mathbf{X}} &: \mathbb{R}^k \to \mathbb{R} \\
\mathbf{x} &\mapsto \frac{(\mu_{k-1,f}\,\omega_{k-1})^{-1}}{\sqrt{\det \boldsymbol{\Sigma}}}\, f\Big(\sqrt{(\mathbf{x}-\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\theta})}\,\Big),
\end{aligned}
\tag{1}
$$

where $\omega_{k-1} = 2\pi^{k/2}/\Gamma(k/2)$ is the $(k-1)$-measure of the unit sphere $\mathcal{S}^{k-1}$ in $\mathbb{R}^k$. The Mahalanobis distance $d_{\boldsymbol{\theta},\boldsymbol{\Sigma}} := \sqrt{(\mathbf{X}-\boldsymbol{\theta})'\,\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\theta})}$ has then density $r \mapsto \tilde{f}_k(r) = (\mu_{k-1,f})^{-1} r^{k-1} f(r)\mathbb{I}[r > 0]$, where $\mathbb{I}$ denotes the indicator function. Unlike this distance, the unit vector $\mathbf{U}_{\boldsymbol{\theta},\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\theta})/d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ is distribution-free, with a uniform distribution over $\mathcal{S}^{k-1}$, and is independent of $d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ (throughout, $\mathbf{A}^{1/2}$, for a symmetric and positive definite matrix $\mathbf{A}$, will stand for the symmetric and positive definite square root of $\mathbf{A}$). To make $\boldsymbol{\Sigma}$ and f identifiable without imposing any moment assumption, we will assume that $d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ has median one, i.e., that

$$
\int_0^1 \tilde{f}_k(r)\,dr = 1/2. \tag{2}
$$

If $\mathbf{X}$ has finite second-order moments (equivalently, if $\mu_{k+1,f} < \infty$), the covariance matrix of $\mathbf{X}$ is proportional to $\boldsymbol{\Sigma}$. Classical examples of elliptical distributions are the multinormal distributions, with radial density $f(r) = \phi(r) := \exp(-a_k r^2/2)$, the Student distributions, with radial densities (for $\nu > 0$ degrees of freedom) $f(r) = f^t_\nu(r) := (1 + a_{k,\nu} r^2/\nu)^{-(k+\nu)/2}$, and the power-exponential distributions, with radial densities of the form $f(r) = f^e_\eta(r) := \exp(-b_{k,\eta} r^{2\eta})$, $\eta > 0$ (the positive constants a_k, $a_{k,\nu}$, and $b_{k,\eta}$ are such that (2) is fulfilled).

For the sake of convenience, we are listing here the assumptions needed in the sequel.

Assumption (A). The observations $\mathbf{X}_i$, $i = 1, \ldots, n$ are i.i.d. with a common distribution $\mathrm{Ell}_k(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$ involving a monotone decreasing f.

Assumption (B). The observations $\mathbf{X}_i$, $i = 1, \ldots, n$ are i.i.d. with a common distribution $\mathrm{Ell}_k(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$ admitting finite fourth-order moments (i.e., involving a radial density f such that $\mu_{k+3,f} < \infty$).

Assumption (A′) (resp., (B′)). Reinforcement of Assumption (A) (resp., (B)), further imposing that f is absolutely continous (with a.e. derivative f', say) and $\int_0^\infty r^2 \varphi_f^2(r) \tilde{f}_k(r)\, dr < \infty$, where we wrote $\varphi_f = -f'/f$.

We also report here the various notations we will use in relation with elliptical distributions. Let $r_\gamma = r_{k,\gamma}(f)$ be the γ-quantile of $d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$, that satisfies $\int_0^{r_\gamma} \tilde{f}_k(r)\, dr = \gamma$ (note that our parametrization of elliptical densities implies that $r_{k,1/2}(f) = 1$ for any k and f). Writing $\mathbb{I}^{(\ell)}_{\gamma,\boldsymbol{\theta},\boldsymbol{\Sigma}} := d^\ell_{\boldsymbol{\theta},\boldsymbol{\Sigma}} \mathbb{I}[d_{\boldsymbol{\theta},\boldsymbol{\Sigma}} \leq r_\gamma]$, define then

$$D^{(\ell)}_\gamma := D^{(\ell)}_{k,\gamma}(f) := \mathrm{E}\big[\mathbb{I}^{(\ell)}_{\gamma,\boldsymbol{\theta},\boldsymbol{\Sigma}}\big] = \int_0^{r_\gamma} r^\ell \tilde{f}_k(r)\, dr, \qquad \alpha_\gamma := \alpha_{k,\gamma}(f) := \sqrt{\frac{D^{(2)}_\gamma}{k\gamma}},$$

and

$$\beta_\gamma := \beta_{k,\gamma}(f) := \frac{1}{k(k+2)} \int_0^{r_\gamma} r^3 \varphi_f(r) \tilde{f}_k(r)\, dr = \frac{(k+2)D^{(2)}_\gamma - r^3_\gamma \tilde{f}_k(r_\gamma)}{k(k+2)}, \tag{3}$$

where the last equality follows by integrating by parts. Note that, under Assumption (A), β_γ is positive and increases monotonically in γ.

Under ellipticity, the MCD estimator of scatter $\hat{\boldsymbol{\Sigma}}_\gamma$ is not consistent for $\boldsymbol{\Sigma}$, but rather for $\alpha^2_\gamma \boldsymbol{\Sigma}$; see Proposition 2.1 below. Our derivations will rely on the following Bahadur representation result for $\hat{\boldsymbol{\Sigma}}_\gamma$ which follows directly from Corollary 4.1 of Cator & Lopuhaä (2010) by using the affine-equivariance of $\hat{\boldsymbol{\Sigma}}$ and by rearranging the terms there (note that the radial function h in Cator & Lopuhaä (2010) is linked to the f introduced above through $h(r^2) = (\mu_{k-1,f}\omega_{k-1})^{-1} f(r)$).

Proposition 2.1. *Under Assumption (A), we have that*

$$\begin{aligned}\sqrt{n}\,(\hat{\boldsymbol{\Sigma}}_\gamma - \alpha^2_\gamma \boldsymbol{\Sigma}) = {} & \frac{\alpha^2_\gamma}{\beta_\gamma \sqrt{n}} \boldsymbol{\Sigma}^{1/2} \sum_{i=1}^n \mathbb{I}^{(2)}_{i;\gamma,\boldsymbol{\theta},\boldsymbol{\Sigma}} \Big(\mathbf{U}_{i;\boldsymbol{\theta},\boldsymbol{\Sigma}} \mathbf{U}'_{i;\boldsymbol{\theta},\boldsymbol{\Sigma}} - \frac{1}{k} \mathbf{I}_k \Big) \boldsymbol{\Sigma}^{1/2} \\ & + \frac{1}{k\gamma\sqrt{n}} \sum_{i=1}^n (\mathbb{I}^{(2)}_{i;\gamma,\boldsymbol{\theta},\boldsymbol{\Sigma}} - k\gamma\alpha^2_\gamma) \boldsymbol{\Sigma} - \frac{r^2_\gamma}{k\gamma\sqrt{n}} \sum_{i=1}^n (\mathbb{I}^{(0)}_{i;\gamma,\boldsymbol{\theta},\boldsymbol{\Sigma}} - \gamma) \boldsymbol{\Sigma} + o_{\mathrm{P}}(1),\end{aligned} \tag{4}$$

as $n \to \infty$, where $\mathbf{I}_k$ denotes the k-dimensional identity matrix.

As we will see, this formulation of the Bahadur result from Cator & Lopuhaä (2010) is suitable for our purposes. It will be convenient that each of the first three terms in the right-hand side of (4) has zero mean and bounded variance, hence is bounded in probability. This will indeed allow to apply the continuous mapping theorem in order to derive the asymptotic behavior of the corresponding shape estimators.

3. The shape parameter

As mentioned in the Introduction, many problems in multivariate analysis (principal component analysis, canonical correlation analysis, testing for sphericity, etc.) require to know or estimate the scatter $\boldsymbol{\Sigma}$ up to a positive scalar factor only. In other words, the parameter of interest, in such problems, is the corresponding S-shape matrix

$$\mathbf{V}_S := \boldsymbol{\Sigma}/S(\boldsymbol{\Sigma})$$

(while the scale parameter $\sigma_S^2 := S(\boldsymbol{\Sigma})$ plays the role of a nuisance), where the *scale* functional $S : \mathcal{S}_k \to \mathbb{R}_0^+$ (i) is *homogeneous* (for all $\lambda > 0$, $S(\lambda\boldsymbol{\Sigma}) = \lambda S(\boldsymbol{\Sigma})$), (ii) is differentiable, with $\frac{\partial S}{\partial \Sigma_{11}}(\boldsymbol{\Sigma}) \neq 0$ for all $\boldsymbol{\Sigma} \in \mathcal{S}_k$, and (iii) satisfies $S(\mathbf{I}_k) = 1$; see Paindaveine (2008) for comments on the requirements (i)-(iii). The collection of $k \times k$ S-shape matrices will be denoted by $\mathcal{V}_k^S$.

Classical scale functionals include

(a) $S(\boldsymbol{\Sigma}) = \Sigma_{11}$ (Randles (2000) and Hettmansperger & Randles (2002)),
(b) $S(\boldsymbol{\Sigma}) = (\mathrm{tr}\,\boldsymbol{\Sigma})/k$ (Tyler (1987), Dümbgen (1998),Visuri et al. (2003), and Taskinen et al. (2010)),
(c) $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$ (Tatsuoka & Tyler (2000), Dümbgen & Tyler (2005), and Taskinen et al. (2006)), and
(d) $S(\boldsymbol{\Sigma}) = k/(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})$ (Frahm (2009)).

The scale functional in (c) was shown to be "canonical" in Paindaveine (2008), in the sense that it is the only scale functional that provides parameter-orthogonality between shape $\mathbf{V}_S$ and scale σ_S^2 (parameter orthogonality here refers to block-diagonality of the corresponding information matrix; see, e.g., Cox & Reid (1987), Section 2.1). A directly related result is that this particular scale functional is the only one for which asymptotically normal shape and scale estimators are asymptotically independent; see Frahm (2009).

The following notation will be used throughout. For any $k \times k$ matrix $\mathbf{A}$, let $\mathrm{vec}\,\mathbf{A}$ denote the k^2-dimensional vector resulting from stacking the columns of $\mathbf{A}$ on top of each other. Write $\mathrm{vech}\,\mathbf{A}$ for the $(K+1)$-vector (recall that $K = k(k+1)/2 - 1$) obtained by stacking the upper-triangular elements of $\mathbf{A}$; $\overset{\circ}{\mathrm{vech}}\,\mathbf{A}$ will denote the K-vector obtained by depriving $\mathrm{vech}\,\mathbf{A}$ of its first component. Write $\mathbf{A}^{\otimes 2}$ for the Kronecker product $\mathbf{A} \otimes \mathbf{A}$. Denoting by $\mathbf{e}_\ell$ the ℓth vector of the canonical basis of $\mathbb{R}^k$, let $\mathbf{K}_k := \sum_{i,j=1}^k (\mathbf{e}_i\mathbf{e}_j') \otimes (\mathbf{e}_j\mathbf{e}_i')$ be the $k^2 \times k^2$ *commutation matrix*, and put $\mathbf{J}_k := \sum_{i,j=1}^k (\mathbf{e}_i\mathbf{e}_j') \otimes (\mathbf{e}_i\mathbf{e}_j') = (\mathrm{vec}\,\mathbf{I}_k)(\mathrm{vec}\,\mathbf{I}_k)'$. Finally, define $\mathbf{N}_k$ as the $K \times k^2$ matrix such that $\mathbf{N}_k(\mathrm{vec}\,\mathbf{A}) = (\overset{\circ}{\mathrm{vech}}\,\mathbf{A})$ for any $k \times k$ matrix $\mathbf{A}$.

The algebra of S-shape matrices then requires introducing the following quantities. For any $\boldsymbol{\Sigma} \in \mathcal{S}_k$ and any S as above, let $\mathbf{D}_S^{\boldsymbol{\Sigma}} := (\mathbf{C}_S^{\boldsymbol{\Sigma}} + (\mathbf{C}_S^{\boldsymbol{\Sigma}})')/2$, where $\mathbf{C}_S^{\boldsymbol{\Sigma}} := \mathbf{C}_{S,k}^{\boldsymbol{\Sigma}}$ is the upper-triangular $k \times k$ matrix such that $\mathrm{vech}\,\mathbf{C}_S^{\boldsymbol{\Sigma}} = \nabla S(\mathrm{vech}\,\boldsymbol{\Sigma})$; here, $\nabla S(\mathrm{vech}\,\boldsymbol{\Sigma})$ stands for the gradient $\mathrm{grad}_{\mathrm{vech}\,\boldsymbol{\Sigma}} S(\mathrm{vech}\,\boldsymbol{\Sigma})$. Define $\mathbf{M}_S^{\boldsymbol{\Sigma}} := \mathbf{M}_{S,k}^{\boldsymbol{\Sigma}}$ as the $K \times k^2$ matrix such that $(\mathbf{M}_S^{\boldsymbol{\Sigma}})'(\overset{\circ}{\mathrm{vech}}\,\mathbf{v}) = \mathrm{vec}\,\mathbf{v}$ for any symmetric $k \times k$ matrix $\mathbf{v}$ satisfying $(\nabla S(\mathrm{vech}\,\boldsymbol{\Sigma}))'(\mathrm{vech}\,\mathbf{v}) = 0$ (equivalently, $(\mathrm{vec}\,\mathbf{D}_S^{\boldsymbol{\Sigma}})'(\mathrm{vec}\,\mathbf{v}) = 0$, or $\mathrm{tr}\,[\mathbf{D}_S^{\boldsymbol{\Sigma}}\mathbf{v}] = 0$). Finally, for any S and $\mathbf{V} \in \mathcal{V}_k^S$, define $\mathcal{E}_k^{\mathbf{V}} := \mathrm{tr}[(\mathbf{D}_S^{\mathbf{V}}\mathbf{V})^2]$. For $S(\boldsymbol{\Sigma}) = \Sigma_{11}$, $S(\boldsymbol{\Sigma}) = (\mathrm{tr}\,\boldsymbol{\Sigma})/k$, $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$, and $S(\boldsymbol{\Sigma}) = k/(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})$, one has $\mathbf{D}_S^{\boldsymbol{\Sigma}} = \mathbf{e}_1\mathbf{e}_1'$, $\mathbf{D}_S^{\boldsymbol{\Sigma}} = \frac{1}{k}\mathbf{I}_k$, $\mathbf{D}_S^{\boldsymbol{\Sigma}} = \frac{1}{k}|\boldsymbol{\Sigma}|^{1/k}\boldsymbol{\Sigma}^{-1}$, and $\mathbf{D}_S^{\boldsymbol{\Sigma}} = k\boldsymbol{\Sigma}^{-2}/(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})^2$ — hence $\mathcal{E}_k^{\mathbf{V}} = 1$, $\mathcal{E}_k^{\mathbf{V}} = \frac{1}{k^2}\mathrm{tr}[\mathbf{V}^2]$, $\mathcal{E}_k^{\mathbf{V}} = \frac{1}{k}$, and $\mathcal{E}_k^{\mathbf{V}} = \frac{1}{k^2}\mathrm{tr}[\mathbf{V}^{-2}]$, respectively.

4. Inference on shape based on the MCD

In this section, we provide the main results of the paper. First, we determine the asymptotic behavior of the MCD estimator of S-shape (Section 4.1). Then we exploit this result to propose and study a test for the null hypothesis that the S-shape is equal to a given possible value (Section 4.2).

4.1. MCD-estimator of shape

Denoting again the MCD_γ estimator of scatter as $\hat{\mathbf{\Sigma}}_\gamma$, the corresponding MCD estimator for S-shape is naturally defined as $\hat{\mathbf{V}}_{S,\gamma} := \hat{\mathbf{\Sigma}}_\gamma / S(\hat{\mathbf{\Sigma}}_\gamma)$. The affine-equivariance of $\hat{\mathbf{\Sigma}}_\gamma$ implies that, for any $k \times k$ invertible matrix $\mathbf{A}$ and any k-vector $\mathbf{b}$,

$$\hat{\mathbf{V}}_{S,\gamma}(\mathbf{A}\mathbf{X}_1 + \mathbf{b}, \ldots, \mathbf{A}\mathbf{X}_n + \mathbf{b}) = \frac{\mathbf{A}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{X}_1, \ldots, \mathbf{X}_n)\mathbf{A}'}{S(\mathbf{A}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{X}_1, \ldots, \mathbf{X}_n)\mathbf{A}')},$$

which is the natural affine-equivariance property for S-shape matrices.

We are primarily interested in the asymptotic properties of $\hat{\mathbf{V}}_{S,\gamma}$. These can be derived from Proposition 2.1 by applying the Delta method. In order to state a Bahadur representation and asymptotic normality result for $\hat{\mathbf{V}}_{S,\gamma}$, we let

$$c_{k,\gamma} := \frac{k(k+2)\beta_\gamma^2}{D_\gamma^{(4)}} \tag{5}$$

and

$$\begin{aligned}\mathbf{Q}_k^{\mathbf{V}_S} :=& (\mathbf{I}_{k^2} + \mathbf{K}_k)\big(\mathbf{V}_S^{\otimes 2}\big) - 2\big(\mathbf{V}_S^{\otimes 2}\big)(\operatorname{vec}\mathbf{D}_S^{\mathbf{V}_S})(\operatorname{vec}\mathbf{V}_S)' \\ &- 2(\operatorname{vec}\mathbf{V}_S)(\operatorname{vec}\mathbf{D}_S^{\mathbf{V}_S})'\big(\mathbf{V}_S^{\otimes 2}\big) + 2\,\mathcal{E}_k^{\mathbf{V}_S}(\operatorname{vec}\mathbf{V}_S)(\operatorname{vec}\mathbf{V}_S)'.\end{aligned} \tag{6}$$

We then have the following result (see Appendix A for the proof).

Theorem 4.1. Let Assumption (A) hold. Then (i) we have that

$$\begin{aligned}\sqrt{n}\operatorname{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S) \;=\; & \frac{1}{\beta_\gamma\sqrt{n}}\Big[\mathbf{I}_{k^2} - (\operatorname{vec}\mathbf{V}_S)(\operatorname{vec}\mathbf{D}_S^{\mathbf{V}_S})'\Big] \\ & \times \big(\mathbf{V}_S^{\otimes 2}\big)^{1/2}\sum_{i=1}^n \mathbb{I}_{i;\gamma,\boldsymbol{\theta},\mathbf{\Sigma}}^{(2)}\operatorname{vec}\Big(\mathbf{U}_{i;\boldsymbol{\theta},\mathbf{V}_S}\mathbf{U}_{i;\boldsymbol{\theta},\mathbf{V}_S}' - \frac{1}{k}\mathbf{I}_k\Big) + o_{\mathrm{P}}(1)\end{aligned}$$

as $n \to \infty$; hence, (ii) $\sqrt{n}\operatorname{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S)$ is asymptotically normal with mean zero and covariance matrix $c_{k,\gamma}^{-1}\mathbf{Q}_k^{\mathbf{V}_S}$.

Building confidence zones for $\hat{\mathbf{V}}_{S,\gamma}$ from Theorem 4.1 requires to estimate consistently the quantity $c_{k,\gamma}$ (the continuous mapping theorem indeed trivially implies that $\mathbf{Q}_k^{\mathbf{V}_S}$ may simply be consistently estimated by $\mathbf{Q}_k^{\hat{\mathbf{V}}_{S,\gamma}}$). Estimation of $c_{k,\gamma}$ will be discussed in Section 5 below.

If Assumption (B) also holds, that is, if the elliptical distribution at hand has finite fourth-order (hence also third-order) moments, then $\int_0^\infty r^3 \tilde{f}_k(r)\, dr$ is finite. This implies that $r_\gamma^3 \tilde{f}_k(r_\gamma)$ must go to zero as $\gamma \to 1$, which yields that, still as $\gamma \to 1$,

$$\begin{aligned} c_{k,\gamma}^{-1} &= \Big(1 - \frac{r_\gamma^3 \tilde{f}_k(r_\gamma)}{(k+2)D_\gamma^{(2)}}\Big)^{-2} \times \frac{kD_\gamma^{(4)}}{(k+2)(D_\gamma^{(2)})^2} \\ &=: \Big(1 - \frac{r_\gamma^3 \tilde{f}_k(r_\gamma)}{(k+2)D_\gamma^{(2)}}\Big)^{-2} (1+\kappa_\gamma) \to 1 + \kappa := \frac{kD^{(4)}}{(k+2)(D^{(2)})^2}, \end{aligned}$$

where we let $D^{(\ell)} = D_1^{(\ell)} = \int_0^\infty r^\ell \tilde{f}_k(r)\, dr$. The quantity $\kappa = \kappa_k(f)$ is the usual *kurtosis* coefficient for k-dimensional elliptical distributions with radial density f; see, e.g., Muirhead & Waternaux (1980) or Tyler (1982). The coefficient κ_γ may be interpreted as a truncated elliptical kurtosis coefficient (where truncation is governed by the population MCD$_\gamma$ ellipsoid). Writing the asymptotic covariance matrix in terms of κ_γ also clarifies the link with the corresponding result for the usual empirical covariance matrix; see Theorem 6.1 below.

Theorem 4.1 straightforwardly provides the influence function of the MCD estimator $\hat{\mathbf{V}}_{S,\gamma}$.

Theorem 4.2. The influence function of $\hat{\mathbf{V}}_{S,\gamma}$, under location $\boldsymbol{\theta}$, scale σ_S^2, shape $\mathbf{V}_S$, and radial density f, is given by

$$\begin{aligned} \mathbf{x} \mapsto \mathrm{IF}(\mathbf{x}, \hat{\mathbf{V}}_{S,\gamma}; \boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f) := \frac{1}{\beta_\gamma \sigma_S^2}\, d^2_{\boldsymbol{\theta},\mathbf{V}_S} \mathbb{I}[d_{\boldsymbol{\theta},\mathbf{V}_S} \leq \sigma_S r_\gamma] \\ \times \mathbf{V}_S^{1/2} \Big(\mathbf{u}_{\boldsymbol{\theta},\mathbf{V}_S} \mathbf{u}'_{\boldsymbol{\theta},\mathbf{V}_S} - \big[\mathbf{u}'_{\boldsymbol{\theta},\mathbf{V}_S} \mathbf{V}_S^{1/2} \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S^{1/2} \mathbf{u}_{\boldsymbol{\theta},\mathbf{V}_S}\big] \mathbf{I}_k \Big) \mathbf{V}_S^{1/2}, \end{aligned}$$

where $d_{\boldsymbol{\theta},\mathbf{V}_S} := ((\mathbf{x}-\boldsymbol{\theta})'\mathbf{V}_S^{-1}(\mathbf{x}-\boldsymbol{\theta}))^{1/2}$ and $\mathbf{u}_{\boldsymbol{\theta},\mathbf{V}_S} := \mathbf{V}_S^{-1/2}(\mathbf{x}-\boldsymbol{\theta})/d_{\boldsymbol{\theta},\mathbf{V}_S}$.

As expected, the support of the influence function of $\hat{\mathbf{V}}_{S,\gamma}$ is the hyper-ellipsoid $\{\mathbf{x} \in \mathbb{R}^k : d_{\boldsymbol{\theta},\mathbf{V}_S} \leq \sigma_S r_\gamma\}$, hence coincides with the support of the influence function of $\hat{\boldsymbol{\Sigma}}_\gamma$; see Croux & Haesbroeck (1999). Note also that, in this support, the influence function of $\hat{\mathbf{V}}_{S,\gamma}$ takes a value that depends on f (hence, on the distribution of $d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$) and on γ only through the scalar factor $1/\beta_\gamma$, whereas the influence function of $\hat{\boldsymbol{\Sigma}}_\gamma$ depends on f and γ in a much more complicated way (implying, e.g., that the influence functions of $\hat{\boldsymbol{\Sigma}}_\gamma$ at elliptical t-distributions and at the multinormal are not proportional to each other). Of course, the smaller γ, the smaller the support of $\hat{\mathbf{V}}_{S,\gamma}$'s influence function, but also the larger the influence function itself within this support (recall that β_γ is monotonically increasing in γ).

As an illustration, Figure 1 plots, for $S(\boldsymbol{\Sigma}) = (\det \boldsymbol{\Sigma})^{1/k}$, the influence functions of $(\hat{\mathbf{V}}_{S,\gamma})_{22}$ (first column) and $(\hat{\mathbf{V}}_{S,\gamma})_{12}$ (second column) at the bivariate standard normal distribution; first row (resp., second row) corresponds to $\gamma = 0.5$ (resp., $\gamma = 0.75$). Note that the influence function of $(\hat{\mathbf{V}}_{S,\gamma})_{12}$ does not depend on the scale functional S. In the spherical setup considered, the scale functionals $S(\boldsymbol{\Sigma}) = (\det \boldsymbol{\Sigma})^{1/k}$, $S(\boldsymbol{\Sigma}) = (\mathrm{tr}\, \boldsymbol{\Sigma})/k$, and $S(\boldsymbol{\Sigma}) = k/(\mathrm{tr}\, \boldsymbol{\Sigma}^{-1})$ lead to the same influence function for $(\hat{\mathbf{V}}_{S,\gamma})_{22}$, and the influence function of $(\hat{\mathbf{V}}_{S,\gamma})_{22}$ for $S(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}_{11}$ is equal to twice the common influence function obtained for the three other scale functionals.

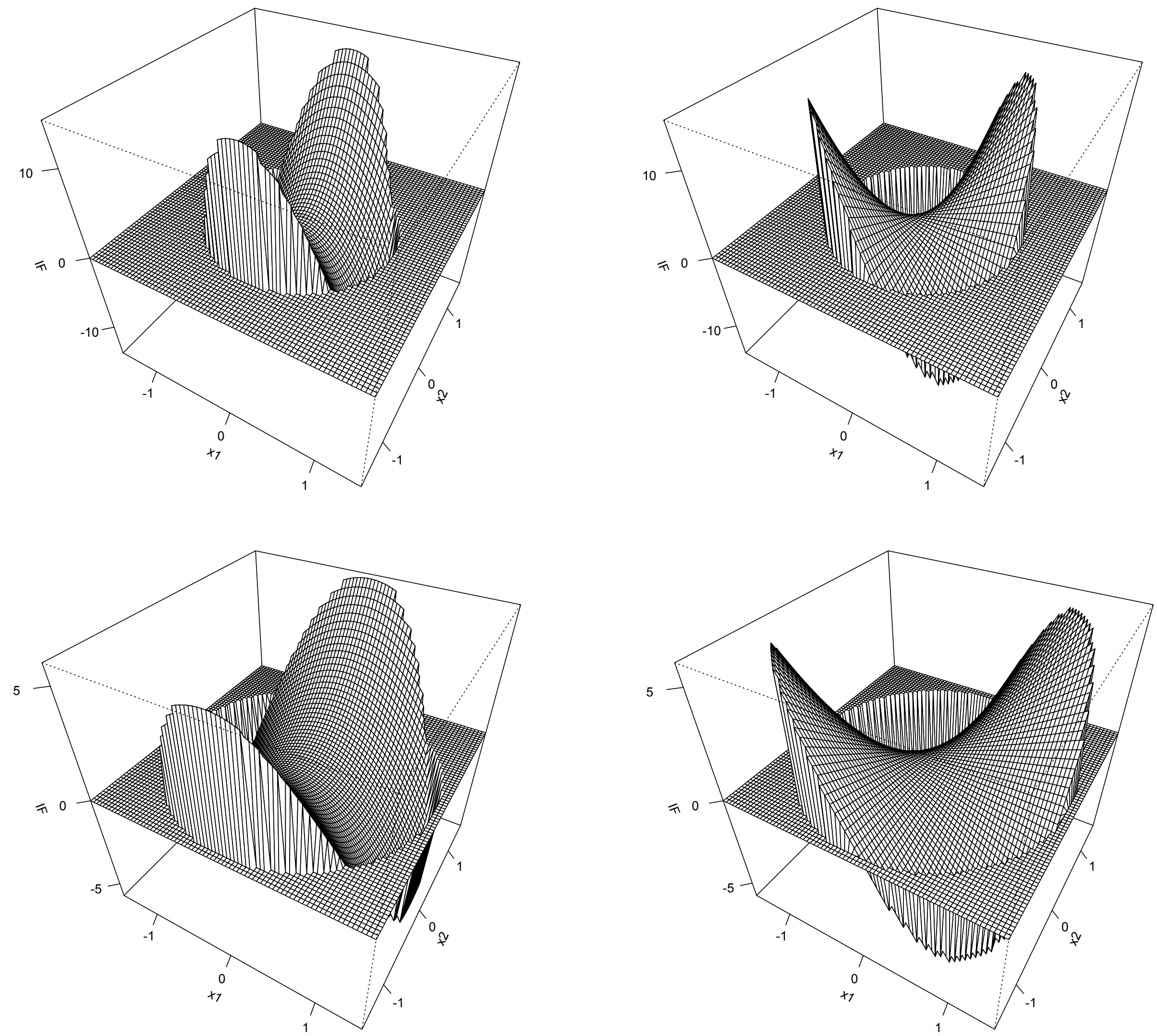

Figure 1: Plots of the influence functions, for the scale functional $S(\mathbf{\Sigma}) = (\det \mathbf{\Sigma})^{1/k}$, of $(\hat{\mathbf{V}}_{S,\gamma})_{22}$ (first column) and $(\hat{\mathbf{V}}_{S,\gamma})_{12}$ (second column) at the bivariate standard normal distribution. The first row (resp., second row) corresponds to $\gamma = 0.5$ (resp., $\gamma = 0.75$).

4.2. MCD-test for shape

In this section, we construct a Wald-type test, based on the MCD shape estimator $\hat{\mathbf{V}}_{S,\gamma}$ above for the problem

$$\begin{cases} \mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0 \\ \mathcal{H}_1 : \mathbf{V}_S \neq \mathbf{V}_S^0, \end{cases} \tag{7}$$

where $\mathbf{V}_S^0 \in \mathcal{V}_k^S$ is fixed. The important case for which $\mathbf{V}_S^0 = \mathbf{I}_k$ corresponds to testing the null of *sphericity*. A Wald test cannot be directly based on Theorem 4.1(ii) because the asymptotic covariance matrix of $\sqrt{n}\,\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S)$ is not invertible. This non-invertibility is explained by the fact that only K of the k^2 entries of $\mathrm{vec}(\mathbf{V}_S)$ are functionally independent (which follows from symmetry of $\mathbf{V}_S$ and the normalization constraint $S(\mathbf{V}_S) = 1$).

To solve this issue, one can rather base a Wald test on the random K-vector $\sqrt{n}\,\mathring{\mathrm{vech}}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S)$, which, in view of Theorem 4.1(ii), is asymptotically normal with mean zero and covariance matrix $c_{k,\gamma}^{-1}\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S}\mathbf{N}_k'$. As we learn from Lemma 4.1 below, this asymptotic covariance matrix is invertible, so that a MCD Wald test for (7) may be based on

$$\mathring{Q}_{S,\gamma} = n\hat{c}_{k,\gamma}\big[\mathring{\mathrm{vech}}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)\big]'\big(\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S^0}\mathbf{N}_k'\big)^{-1}\,\mathring{\mathrm{vech}}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0), \tag{8}$$

where $\hat{c}_{k,\gamma}$ is an arbitrary consistent estimator of $c_{k,\gamma}$; see Section 5 for possible estimators.

We actually propose rather using the simpler test statistic

$$Q_{S,\gamma} = \frac{n\hat{c}_{k,\gamma}}{2}\Big(\mathrm{tr}\Big[((\mathbf{V}_S^0)^{-1}\hat{\mathbf{V}}_{S,\gamma})^2\Big] - \frac{1}{k}\,\mathrm{tr}^2\Big[(\mathbf{V}_S^0)^{-1}\hat{\mathbf{V}}_{S,\gamma})\Big]\Big), \tag{9}$$

that, under the null (hence also under sequences of contiguous alternatives), is asymptotically equivalent to $\mathring{Q}_{S,\gamma}$ in probability; see Theorem 4.3(i). Denoting by $\hat{\lambda}_j$, $j = 1, \ldots, k$ the eigenvalues of $(\mathbf{V}_S^0)^{-1/2}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{V}_S^0)^{-1/2}$, note that $Q_{S,\gamma}$ is proportional to $\mathrm{Var}_{\hat{\lambda}} = \frac{1}{k}\sum_{j=1}^k\{\hat{\lambda}_j - (\frac{1}{k}\sum_{j=1}^k \hat{\lambda}_j)\}^2$, so that the larger $\mathrm{Var}_{\hat{\lambda}}$, the more $(\mathbf{V}_S^0)^{-1/2}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{V}_S^0)^{-1/2}$ is far from being proportional to $\mathbf{I}_k$, and the more severe the deviation from the null. The corresponding test, $\phi_{S,\gamma}$ say, then rejects the null at asymptotic level α whenever $Q_{S,\gamma} > \chi^2_{K,1-\alpha}$, where $\chi^2_{K,1-\alpha}$ stands for the upper α-quantile of the χ^2_K distribution. Theorem 4.3 below gives the asymptotic properties of this test; its proof requires the following preliminary result (see Appendix A for the proofs).

Lemma 4.1. *The matrix $\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S}\mathbf{N}_k'$ has full rank K, and its inverse is given by* $(\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S}\mathbf{N}_k')^{-1} = \frac{1}{4}\mathbf{M}_k^{\mathbf{V}_S}\big(\mathbf{V}_S^{\otimes 2}\big)^{-1/2}\big[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k}\mathbf{J}_k\big]\big(\mathbf{V}_S^{\otimes 2}\big)^{-1/2}\big(\mathbf{M}_k^{\mathbf{V}_S}\big)'$.

Theorem 4.3. Let Assumption (A) hold. Then, (i) under $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$, $Q_{S,\gamma} = \mathring{Q}_{S,\gamma} + o_{\mathrm{P}}(1)$, as $n \to \infty$; (ii) under $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$, $Q_{S,\gamma}$ is asymptotically χ^2_K; (iii) under sequences of local alternatives $\mathcal{H}_1^{(n)} : \mathbf{V}_S^{(n)} = \mathbf{V}_S^0 + n^{-1/2}\mathbf{v}$, with $\mathrm{tr}[\mathbf{D}_S^{\mathbf{V}_S^0}\mathbf{v}] = 0$, $Q_{S,\gamma}$ is asymptotically non-central χ^2_K, with non-centrality parameter

$$\frac{c_{k,\gamma}}{2}\left(\mathrm{tr}\big[((\mathbf{V}_S^0)^{-1}\mathbf{v})^2\big] - \frac{1}{k}\,\mathrm{tr}^2\big[(\mathbf{V}_S^0)^{-1}\mathbf{v}\big]\right),$$

provided, however, that Assumption (A) is reinforced into (A$'$).

The condition $\mathrm{tr}[\mathbf{D}_S^{\mathbf{V}_S^0}\mathbf{v}] = 0$ in the local alternatives $\mathcal{H}_1^{(n)} : \mathbf{V}_S^{(n)} = \mathbf{V}_S^0 + n^{-1/2}\mathbf{v}$ above ensures that, at the first order as $n \to \infty$, $S(\mathbf{V}_S^{(n)}) = 1$, hence that $\mathbf{V}_S^{(n)}$ remains an S-shape matrix; see (4.3) in Hallin & Paindaveine (2006a) for details. For "linear" scale functionals, this can easily be understood : if S normalizes $\mathbf{V}_S$ to have trace k (resp., upper-left entry equal to one), then $\mathbf{v}$ is constrained to have trace zero (resp., to have upper-left entry equal to zero), so that the perturbed value $\mathbf{V}_S^{(n)} = \mathbf{V}_S^0 + n^{-1/2}\mathbf{v}$ indeed remains an S-shape matrix (for n large enough). The intuition is similar for "non-linear" scale functionals (such as the determinant-based one), where the constraint $S(\mathbf{V}_S^{(n)}) = 1$, however, can only be achieved at the first order.

The null hypothesis $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$ is not invariant under the group of affine transformations, but it is invariant under the subgroup of affine transformations of the form

$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \mapsto ((\mathbf{V}_S^0)^{1/2}\mathbf{O}(\mathbf{V}_S^0)^{-1/2}\mathbf{X}_1 + \mathbf{b}, \ldots, (\mathbf{V}_S^0)^{1/2}\mathbf{O}(\mathbf{V}_S^0)^{-1/2}\mathbf{X}_n + \mathbf{b}), \tag{10}$$

where $\mathbf{O}$ is an arbitrary orthogonal $k \times k$ matrix and $\mathbf{b}$ is an arbitrary k-vector. Note that the test statistic $Q_{S,\gamma}$ in (9) is invariant under this group of transformations.

5. Estimation of nuisance parameters

As already mentioned, implementing the test $\phi_{S,\gamma}$ for $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$ requires to estimate consistently (at least under the null) the quantity $c_{k,\gamma}$ in (5). We now present two such estimators, establish their consistency, and compare their finite-sample performances through simulations.

To describe the first estimator, consider the mapping $r \mapsto \tilde{f}_{k;\mathrm{shape}}(r) = \sigma_S^{-1}\tilde{f}_k(r/\sigma_S)$. Note that this mapping — unlike $\tilde{f}_k$ — does not depend on σ_S, which follows from the fact that $\tilde{f}_{k;\mathrm{shape}}$ (resp., $\tilde{f}_k$) is the pdf of $d_{\boldsymbol{\theta},\mathbf{V}_S}$ (resp., $d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$). Similarly, $s_\gamma := \sigma_S r_\gamma$ — unlike r_γ itself — does not depend on σ_S, since s_γ (resp., r_γ) is the order-γ quantile of $d_{\boldsymbol{\theta},\mathbf{V}_S}$ (resp., $d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$). Consequently, the quantity $c_{k,\gamma}$, that, by using the identity

$$D_\gamma^{(\ell)} = \mathrm{E}[d_{\boldsymbol{\theta},\boldsymbol{\Sigma}}^\ell \mathbb{I}[d_{\boldsymbol{\theta},\boldsymbol{\Sigma}} \leq r_\gamma]] = \sigma_S^{-\ell}\mathrm{E}[d_{\boldsymbol{\theta},\mathbf{V}_S}^\ell \mathbb{I}[d_{\boldsymbol{\theta},\mathbf{V}_S} \leq s_\gamma]], \tag{11}$$

rewrites

$$c_{k,\gamma} = \frac{k(k+2)\beta_\gamma^2}{D_\gamma^{(4)}} = \frac{((k+2)D_\gamma^{(2)} - r_\gamma^3\tilde{f}_k(r_\gamma))^2}{k(k+2)D_\gamma^{(4)}} \tag{12}$$

$$= \frac{((k+2)\mathrm{E}[d_{\boldsymbol{\theta},\mathbf{V}_S}^2 \mathbb{I}[d_{\boldsymbol{\theta},\mathbf{V}_S} \leq s_\gamma]] - s_\gamma^3\tilde{f}_{k;\mathrm{shape}}(s_\gamma))^2}{k(k+2)\mathrm{E}[d_{\boldsymbol{\theta},\mathbf{V}_S}^4 \mathbb{I}[d_{\boldsymbol{\theta},\mathbf{V}_S} \leq s_\gamma]]}, \tag{13}$$

does not depend on σ_S, hence may be estimated without estimating this scale parameter. Since the MCD$_\gamma$-estimators of location and S-shape $\hat{\boldsymbol{\theta}}_\gamma$ and $\hat{\mathbf{V}}_{S,\gamma}$ are consistent for $\boldsymbol{\theta}$ and $\mathbf{V}_S$, respectively, (13) leads to the estimator

$$\hat{c}_{k,\gamma} := \frac{((k+2)\frac{1}{n}\sum_{i=1}^n d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}}^2 \mathbb{I}[d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}} \leq \hat{s}_\gamma] - \hat{s}_\gamma^3\hat{\tilde{f}}_{k;\mathrm{shape}}(\hat{s}_\gamma))^2}{k(k+2)\frac{1}{n}\sum_{i=1}^n d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}}^4 \mathbb{I}[d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}} \leq \hat{s}_\gamma]}, \tag{14}$$

where $\hat{s}_\gamma$, quite naturally, is taken as the sample γ-quantile of the $d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}}$'s, and where

$$\hat{\tilde{f}}_{k;\mathrm{shape}}(s) := \frac{1}{nh_n}\sum_{i=1}^n K\bigg(\frac{s-d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}}}{h_n}\bigg) \tag{15}$$

is a kernel density estimator for $\tilde{f}_{k;\mathrm{shape}}(s)$. We then have the following consistency result (see Appendix B for a proof).

Theorem 5.1. Let Assumption (A) hold. Assume further that (i) the bandwidth sequence (h_n) satisfies $h_n \to 0$ and $nh_n^4 \to \infty$ as $n \to \infty$, and that (ii) the kernel function K has a compact support and is differentiable, and that there exists $C > 0$ such that the derivative of K satisfies $|K'(s)| \leq C$ for all s. Then $\hat{c}_{k,\gamma}$ in (14) converges to $c_{k,\gamma}$ in probability as $n \to \infty$.

This result shows in particular that $\hat{c}_{k,\gamma}$ is a consistent estimator of $c_{k,\gamma}$ when the usual optimal bandwidth $h_n \propto n^{-1/5}$ is used. Note also that consistency holds not only under the null $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$ but under an arbitrary value of $\mathbf{V}_S$. Consequently, this estimator may be used both in the tests of Section 4.2 and to build confidence zones for $\mathbf{V}_S$, based on the asymptotic normality result for $\hat{\mathbf{V}}_{S,\gamma}$ in Theorem 4.1. When performing hypothesis testing, though, it is of course preferable to replace $\hat{c}_{k,\gamma}$ with its null counterpart — $\hat{c}^0_{k,\gamma}$, say — obtained by replacing the $d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}}$'s in (14)-(15) above with their null versions $d_{i;\hat{\boldsymbol{\theta}}_\gamma,\mathbf{V}_S^0}$, $i = 1,\ldots,n$; this estimator $\hat{c}^0_{k,\gamma}$ involves in particular the sample γ-quantile $\hat{s}^0_\gamma$ of the $d_{i;\hat{\boldsymbol{\theta}}_\gamma,\mathbf{V}_S^0}$'s. The proof of Theorem 5.1 still applies and shows that the resulting estimator is weakly consistent under the null $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$.

We then present a second estimator of $c_{k,\gamma}$, that was suggested to us by one of the Referees. This alternative estimator has the advantage to avoid density estimation. However, it consistently estimates $c_{k,\gamma}$ under the null $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$ only, hence cannot be used to obtain confidence zones for $\mathbf{V}_S$. The construction of this estimator exploits Theorem 4.1, that indeed suggests that, under $\mathcal{H}_0$, the quantity $\sigma_S^2\beta_\gamma$ can be consistently estimated by

$$\begin{aligned}\rho^{(n)} \;=\; & \frac{1}{n\|\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma}-\mathbf{V}_S^0)\|^2}\big(\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma}-\mathbf{V}_S^0)\big)'\Big[\mathbf{I}_{k^2}-(\mathrm{vec}\,\mathbf{V}_S^0)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S^0})'\Big] \\ & \times\big((\mathbf{V}_S^0)^{\otimes 2}\big)^{1/2}\sum_{i=1}^n d^2_{\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0}\mathbb{I}[d_{\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} \leq \hat{s}^0_{\gamma\#}]\mathrm{vec}\Big(\mathbf{U}_{i;\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0}\mathbf{U}'_{i;\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0}-\frac{1}{k}\mathbf{I}_k\Big),\end{aligned}$$

where $\hat{\boldsymbol{\theta}}_{\gamma\#} = ((\hat{\boldsymbol{\theta}}_{\gamma\#})_1,\ldots,(\hat{\boldsymbol{\theta}}_{\gamma\#})_k)'$ and $\hat{s}^0_{\gamma\#}$ are *discretized* versions of $\hat{\boldsymbol{\theta}}_\gamma = ((\hat{\boldsymbol{\theta}}_\gamma)_1,\ldots,(\hat{\boldsymbol{\theta}}_\gamma)_k)'$ and $\hat{s}^0_\gamma$. The discretized estimators are obtained as

$$(\hat{\boldsymbol{\theta}}_{\gamma\#})_\ell := \mathrm{sign}\big((\hat{\boldsymbol{\theta}}_\gamma)_\ell\big)\frac{\lceil a\sqrt{n}|(\hat{\boldsymbol{\theta}}_\gamma)_\ell|\rceil}{a\sqrt{n}}, \quad \ell = 1,\ldots,k, \quad \text{and} \quad \hat{s}^0_{\gamma\#} := \frac{\lceil a\sqrt{n}|\hat{s}^0_\gamma|\rceil}{a\sqrt{n}},$$

for some arbitrary constant $a > 0$. These discretized estimators are still root-n consistent, but now are also *locally and asymptotically discrete*; see, e.g., Kreiss (1987) or Ilmonen & Paindaveine (2011), and the comments therein. Since a can be chosen arbitrarily large, such discretization has no impact in real data applications, where n is fixed, so that one may in practice simply use the original estimators $\hat{\boldsymbol{\theta}}_\gamma$ and $\hat{s}^0_\gamma$.

Under the null, it is then natural to estimate

$$c_{k,\gamma} = \frac{k(k+2)\beta_\gamma^2}{D_\gamma^{(4)}} = \frac{k(k+2)(\sigma^2\beta_\gamma)^2}{\mathrm{E}[d^4_{\boldsymbol{\theta},\mathbf{V}_S^0}\mathbb{I}[d_{\boldsymbol{\theta},\mathbf{V}_S^0} \leq s_\gamma]]}$$

(see (11)-(12)) by

$$\vec{c}^{\,0}_{k,\gamma} = \frac{k(k+2)(\rho^{(n)})^2}{\frac{1}{n}\sum_{i=1}^n d^4_{i;\hat{\boldsymbol{\theta}}_\gamma,\mathbf{V}_S^0}\mathbb{I}[d_{i;\hat{\boldsymbol{\theta}}_\gamma,\mathbf{V}_S^0} \leq \hat{s}^0_\gamma]}. \tag{16}$$

Consistency is established in the following result (see Appendix B for a proof, which requires such discretization).

Theorem 5.2. Let Assumption (A′) hold. Then, under the null $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$, $\vec{c}^{\,0}_{k,\gamma}$ in (16) converges to $c_{k,\gamma}$ in probability as $n \to \infty$.

We conducted the following numerical experiment in order to compare the finite-sample performances of the universally consistent density-based estimator $\hat{c}_{k,\gamma}$, with those of its null version $\hat{c}^0_{k,\gamma}$, and of the null density-free estimator $\vec{c}^{\,0}_{k,\gamma}$. We generated $M = 5,000$ independent random samples of sizes $n = 50$, 400, and 10,000 from the bivariate standard normal distribution ($\boldsymbol{\theta} = \mathbf{0}$ and $\mathbf{V}_S = \mathbf{I}_k$). In each of these samples, we evaluated, for $\gamma = 0.5, 0.6, 0.7, 0.8, 0.9$, the estimators $\hat{c}_{k,\gamma}$, $\hat{c}^0_{k,\gamma}$, and $\vec{c}^{\,0}_{k,\gamma}$, where the last two are based on the true value $\mathbf{V}_S^0 = \mathbf{I}_k$. We also computed the universally consistent estimator $\hat{c}_k$ of the corresponding covariance-based quantity c_k, along with the null version $\hat{c}^0_k$ of this estimator (see Section 6).

Boxplots of the resulting estimates are reported in Figure 2. The results indicate that the universally consistent estimators $\hat{c}_{k,\gamma}$ are severely biased for small γ-values (unless, of course, the sample size is very large) but behave well for larger γ-values. As expected, the corresponding null estimators $\hat{c}^0_{k,\gamma}$, that are based on the true underlying shape $\mathbf{V}_S^0$, are more accurate, and show a much smaller bias. Finally, the density-based estimators $\hat{c}^0_{k,\gamma}$ strongly dominate their competitors $\vec{c}^{\,0}_{k,\gamma}$, particularly so for large γ-values.

6. Covariance-based procedures and AREs

The goal of this section is to derive the asymptotic relative efficiencies (AREs) of the MCD$_\gamma$ procedures of Section 4 with respect to their competitors based on the empirical covariance matrix $\hat{\boldsymbol{\Sigma}} := \frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$. Although $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_\gamma$ for $\gamma = 1$, the asymptotic properties of these covariance-based procedures cannot be obtained by taking $\gamma = 1$ in Theorems 4.1 and 4.3, since these results were derived from Proposition 2.1, that is not valid for $\gamma = 1$ (if $f(r) > 0$ for all r, then we indeed have $r_1 = \infty$).

A Bahadur representation result for $\hat{\boldsymbol{\Sigma}}$, however, can be obtained quite trivially. Of course, unlike for the MCD$_\gamma$ scatter estimator, finite fourth-order moments here are needed.

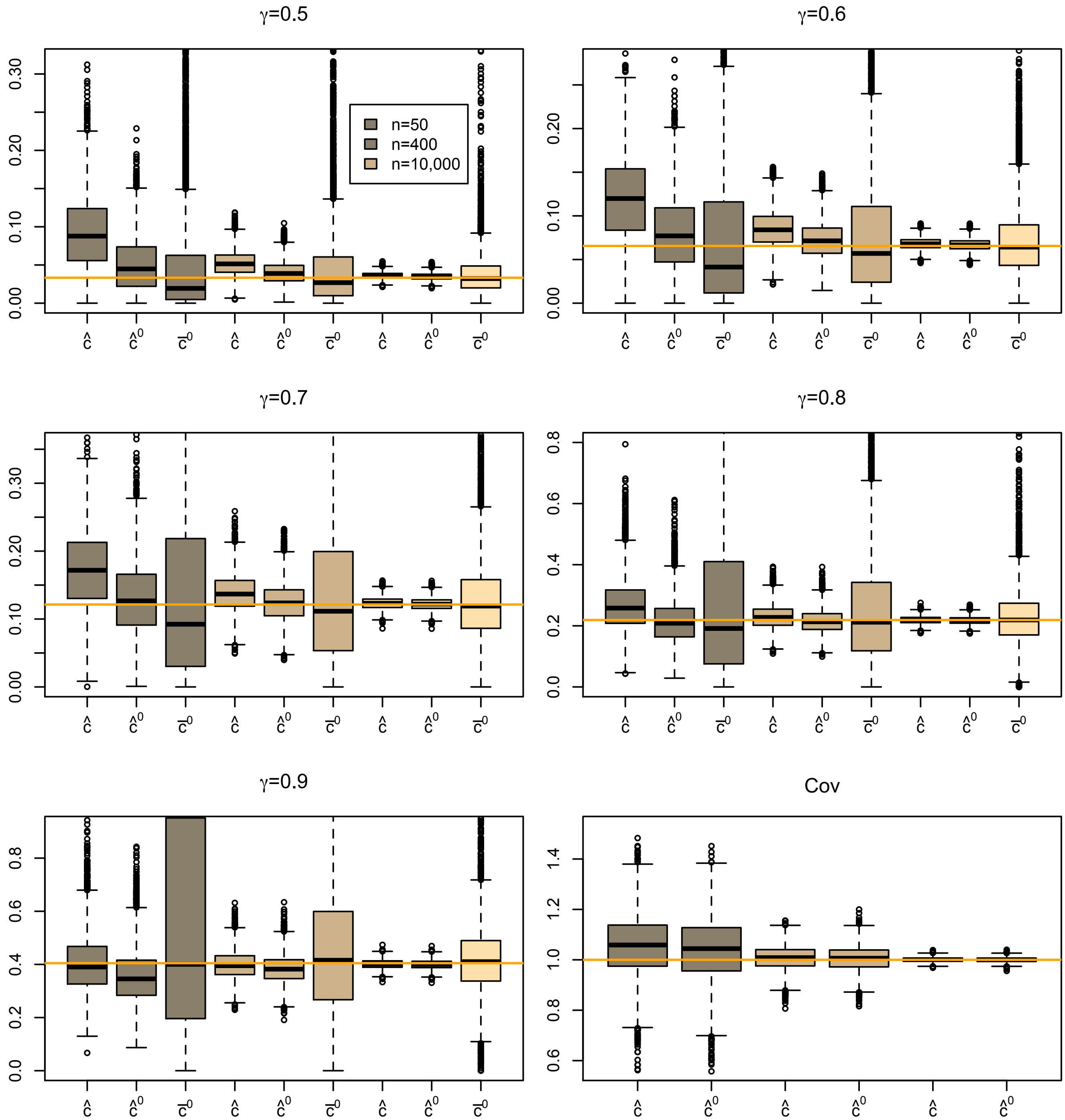

Figure 2: Boxplots, computed from 5,000 independent bivariate standard normal samples of size $n = 50$, 400 and $10,000$, of (i) the estimators $\hat{c}_{k,\gamma}$ in (14), (ii) their version $\hat{c}^0_{k,\gamma}$ based on the true value of $\mathbf{V}_S$, and of (iii) the estimators $\bar{c}^0_{k,\gamma}$ in (16), for $\gamma = 0.5, 0.6, 0.7, 0.8, 0.9$. The lower right panel reports the covariance-based estimators $\hat{c}_k$ and $\hat{c}^0_k$ (see Section 6). The corresponding population quantities ($c_{k,\gamma}$ or, in the lower right panel, c_k) are throughout reported in orange.

Proposition 6.1. Let Assumption (B) hold. Then we have that

$$\sqrt{n}\,\mathrm{vec}\Big(\hat{\mathbf{\Sigma}} - \frac{D^{(2)}}{k}\mathbf{\Sigma}\Big) = \frac{1}{\sqrt{n}}\big(\mathbf{\Sigma}^{\otimes 2}\big)^{1/2}\sum_{i=1}^{n} d^2_{i;\boldsymbol{\theta},\mathbf{\Sigma}}\,\mathrm{vec}\Big(\mathbf{U}_{i;\boldsymbol{\theta},\mathbf{\Sigma}}\mathbf{U}'_{i;\boldsymbol{\theta},\mathbf{\Sigma}} - \frac{1}{k}\mathbf{I}_k\Big)$$
$$+\frac{1}{k\sqrt{n}}\sum_{i=1}^{n}\big(d^2_{i;\boldsymbol{\theta},\mathbf{\Sigma}} - D^{(2)}\big)(\mathrm{vec}\,\mathbf{\Sigma}) + o_{\mathrm{P}}(1),$$

as $n \to \infty$, where $D^{(2)} = D^{(2)}_1 = \int_0^\infty r^2 \tilde{f}_k(r)\,dr$.

Proceeding along the exact same lines as in the proof of Theorem 4.1, we then obtain the asymptotic behavior of the covariance-based estimator of shape $\hat{\mathbf{V}}_S = \hat{\mathbf{\Sigma}}/S(\hat{\mathbf{\Sigma}})$.

Theorem 6.1. Let Assumption (B) hold. Then (i) we have that

$$\sqrt{n}\,\mathrm{vec}(\hat{\mathbf{V}}_S - \mathbf{V}_S) = \frac{k}{D^{(2)}\sqrt{n}}\Big[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})'\Big]$$
$$\times\big(\mathbf{V}_S^{\otimes 2}\big)^{1/2}\sum_{i=1}^{n} d^2_{i;\boldsymbol{\theta},\mathbf{\Sigma}}\,\mathrm{vec}\Big(\mathbf{U}_{i;\boldsymbol{\theta},\mathbf{V}}\mathbf{U}'_{i;\boldsymbol{\theta},\mathbf{V}} - \frac{1}{k}\mathbf{I}_k\Big) + o_{\mathrm{P}}(1)$$

as $n \to \infty$; hence, (ii) $\sqrt{n}\,\mathrm{vec}(\hat{\mathbf{V}}_S - \mathbf{V}_S)$ is asymptotically normal with mean zero and covariance matrix $c_k^{-1}\mathbf{Q}_k^{\mathbf{V}_S}$, where $c_k = 1/(1+\kappa)$ involves the kurtosis coefficient defined in Page 7.

It directly follows that the ARE, under radial density f, of the MCD estimator of shape $\hat{\mathbf{V}}_{S,\gamma}$ with respect to its covariance-based competitor $\hat{\mathbf{V}}_S$ is given by

$$\mathrm{ARE}_f[\hat{\mathbf{V}}_{S,\gamma}/\hat{\mathbf{V}}_S] = c_{k,\gamma}/c_k. \tag{17}$$

Such AREs are unambiguously defined since the asymptotic covariance matrices in Theorems 4.1 and 6.1 are of the form $\lambda_f\mathbf{Q}$, for a common matrix $\mathbf{Q}$, hence are proportional to each other. In contrast, AREs for (affine-equivariant) estimators of *scatter* would not be as easily defined, as such estimators have asymptotic covariance matrices (under radial density f) of the form $\lambda_{1,f}\mathbf{Q}_1 + \lambda_{2,f}\mathbf{Q}_2$; see, e.g., Tyler (1982, 1983). Some plots of the AREs in (17) will be provided below.

Turning to hypothesis testing, the exact similarity between Theorems 4.1 and 6.1 allows to readily deduce the form and asymptotic properties of the covariance-based tests for the problem (7). More precisely, the covariance-based test, ϕ_S say, rejects the null at asymptotic level α whenever

$$Q_S = \frac{n\hat{c}_k^0}{2}\Big(\mathrm{tr}\Big[((\mathbf{V}_S^0)^{-1}\hat{\mathbf{V}}_S)^2\Big] - \frac{1}{k}\,\mathrm{tr}^2\Big[(\mathbf{V}_S^0)^{-1}\hat{\mathbf{V}}_S\Big]\Big) > \chi^2_{K,1-\alpha},$$

with $\hat{c}_k^0 := 1/(1+\hat{\kappa}^0)$, where $\hat{\kappa}^0 := [k(\frac{1}{n}\sum_{i=1}^n d^4_{i;\bar{\mathbf{X}},\mathbf{V}_S^0})]/[(k+2)(\frac{1}{n}\sum_{i=1}^n d^2_{i;\bar{\mathbf{X}},\mathbf{V}_S^0})^2] - 1$ consistently estimates, under the null, the kurtosis coefficient κ. Of course, consistent estimation, for an arbitrary shape value, is achieved by considering $\hat{c}_k := 1/(1+\hat{\kappa})$, where $\hat{\kappa}$ is obtained by substituting $\hat{\mathbf{\Sigma}}$ for $\mathbf{V}_S^0$ in $\hat{\kappa}^0$. Finite-sample performances of these estimators of c_k were illustrated in the lower right panel of Figure 2.

This test coincides with the modified version defined in Hallin & Paindaveine (2006b) of the Gaussian test from John (1972). The modification, that consists in adding the factor $\hat{c}_k^0$, extends the validity of John's test to any elliptical distribution with finite fourth-order moments (John's test, originally, is only valid under elliptical distributions having the same kurtosis as in the multinormal case — i.e., $\kappa_k(f) = \kappa_k(\phi) = 0$). The following result summarizes the asymptotic properties of this test.

Theorem 6.2. Let Assumption (B) hold. Then, (i) under $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$, Q_S is asymptotically χ_K^2; (ii) under sequences of local alternatives $\mathcal{H}_1^{(n)} : \mathbf{V}_S^{(n)} = \mathbf{V}_S^0 + n^{-1/2}\mathbf{v}$, with $\mathrm{tr}[\mathbf{D}_S^{\mathbf{V}_S^0}\mathbf{v}] = 0$, Q_S is asymptotically non-central χ_K^2, with non-centrality parameter

$$\frac{c_k}{2}\left(\mathrm{tr}\big[((\mathbf{V}_S^0)^{-1}\mathbf{v})^2\big] - \frac{1}{k}\,\mathrm{tr}^2\big[(\mathbf{V}_S^0)^{-1}\mathbf{v}\big]\right),$$

provided, however, that Assumption (B) is reinforced into (B$'$).

Asymptotic relative efficiencies, as usual, are obtained as the ratios of the non-centrality parameters in the asymptotic non-null distributions of the corresponding tests. Therefore, the ARE, under radial density f, of the MCD$_\gamma$ test for shape $\phi_{S,\gamma}$ with respect to its covariance-based competitor ϕ_S is given by

$$\mathrm{ARE}_f[\phi_{S,\gamma}/\phi_S] = c_{k,\gamma}/c_k, \tag{18}$$

which coincides with the ARE obtained in (17) for point estimation. Both for hypothesis testing and point estimation, these AREs require that the underlying elliptical distribution has finite fourth-order moments ($\mu_{k+3,f} < \infty$). Note, however, that the AREs may be considered infinite when fourth-order moments themselves are infinite, since the covariance-based competitors then collapse, while the MCD$_\gamma$ procedures remain valid (in the sense that $\hat{\mathbf{V}}_{S,\gamma}$ remains root-n consistent and that $\phi_{S,\gamma}$ still meets the asymptotic α-level constraint).

Figure 3 provides several plots (as functions of γ or of the number of degrees of freedom ν of the underlying standard elliptical t_ν distribution) of the AREs in (17)-(18), under k-variate standard normal and t_ν densities. It is seen that the AREs decrease with the tail weight ν. At the multinormal, as expected, MCD-based shape procedures are poorly efficient, but they dominate their covariance-based competitors under heavy tails, particularly so for large dimensions k.

7. Monte-Carlo studies

In this section, we illustrate the finite-sample behaviors of the MCD$_\gamma$ inference procedures for shape from Section 4 and of their covariance-based competitors from Section 6. The goal is not so much to show how the former compare with the latter, but rather to confirm our asymptotic results and to learn how well these results approximate the finite-sample properties of the procedures considered. A robustness study will also be conducted.

We start with hypothesis testing, where we focused on the problem of testing for sphericity, i.e., on the null hypothesis $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{I}_k$. Throughout, we adopted the determinant-based scale functional $S(\mathbf{\Sigma}) = (\det \mathbf{\Sigma})^{1/k}$. We generated collections of $M = 2,000$ independent random samples

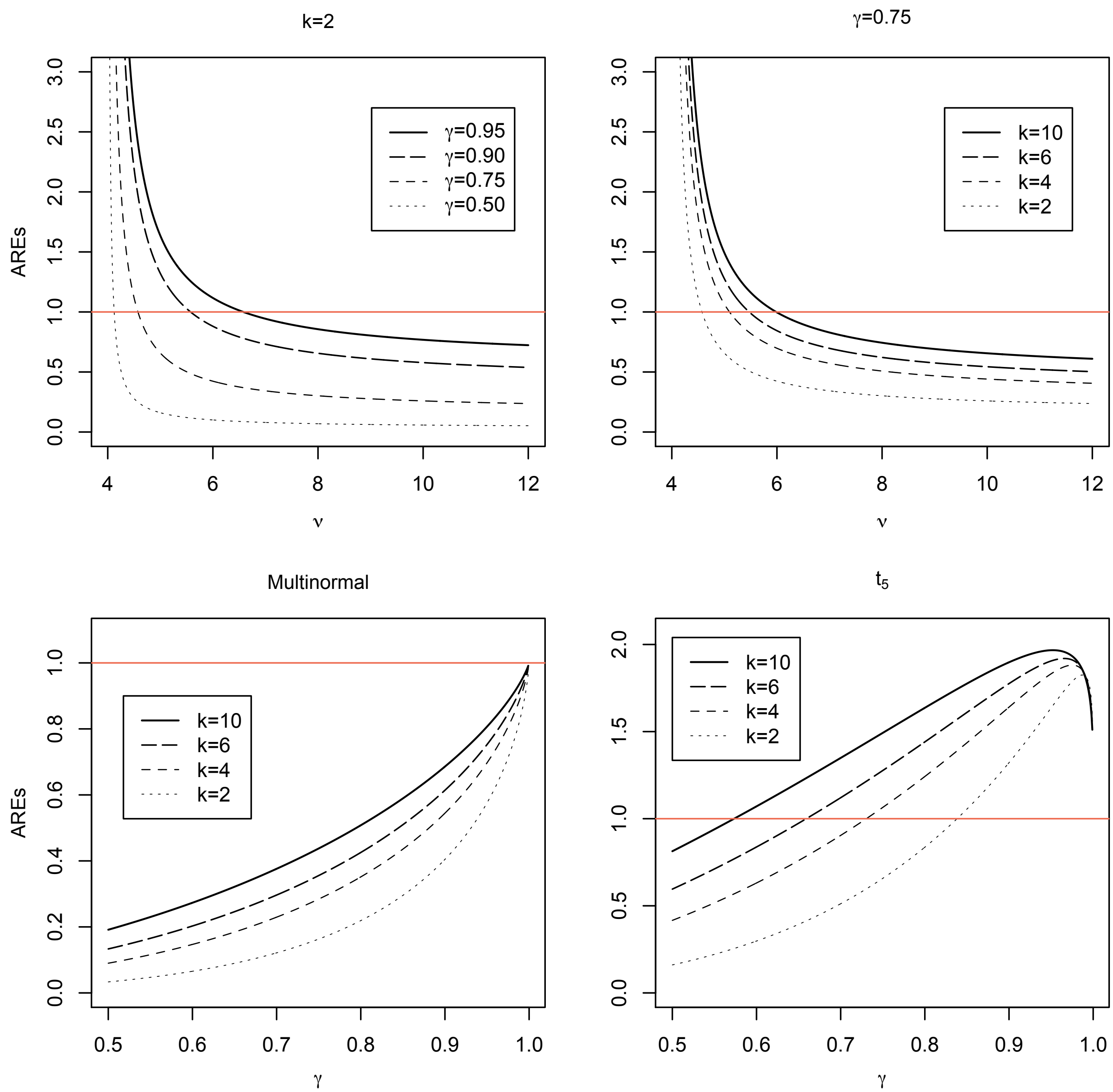

Figure 3: Plots of asymptotic relative efficiencies (AREs) of MCD_γ shape estimators and tests with respect to their covariance-based competitors, under k-variate standard normal and t_ν densities.

of sizes $n = 50$, 400, and 2,000, from a bivariate normal distribution with mean $\boldsymbol{\theta} = \mathbf{0}$, scale $\sigma_S = 1$, and shape

$$\mathbf{V}_S^{(n)}(m;\xi) = \frac{\mathbf{I}_2 + \frac{m}{\xi\sqrt{n}} \begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix}}{\left(\det\left[\mathbf{I}_2 + \frac{m}{\xi\sqrt{n}} \begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix}\right]\right)^{1/2}}, \qquad m = 0, 1, 2, \ldots, 6, \tag{19}$$

with $\xi = 1.2$. Figure 4 plots, for each sample size n above, a few equidensity contours of the bivariate normal distribution with shape $\mathbf{V}_S^{(n)}(6; 1.2)$, which corresponds to the most extreme alternative considered. We also generated collections of $M = 2,000$ independent random samples with the same sample sizes from a bivariate t_5 distribution with mean zero, S-scale one, and shape matrices $\mathbf{V}_S^{(n)}(m;\xi)$, still for $m = 0, 1, 2, \ldots, 6$, but here with $\xi = 1$; these heterogeneous ξ-values were chosen so that the most severe alternatives — associated with the shape matrices $\mathbf{V}_S^{(n)}(6;\xi)$ — lead to roughly similar rejection frequencies in the multinormal and t_5 cases.

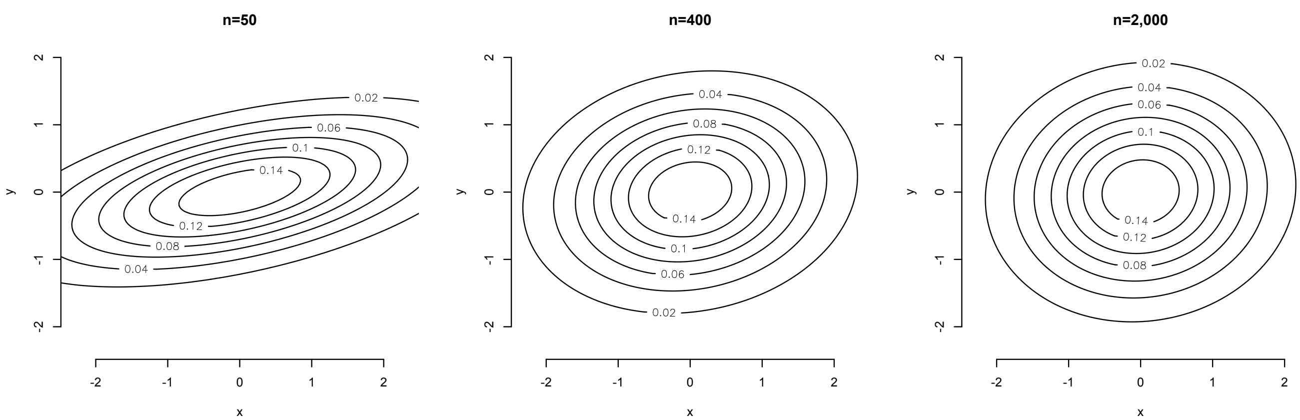

Figure 4: Some equidensity contours of the bivariate normal distribution with mean $\boldsymbol{\theta} = \mathbf{0}$, scale $\sigma_S = 1$, and shape $\mathbf{V}_S^{(n)}(6; 1.2)$ (see (19)), for $n = 50$, 400, and 2,000. These correspond to the most severe alternatives considered in the hypothesis testing simulation.

For each such sample, we performed, at asymptotic level $\alpha = 5\%$, the MCD$_\gamma$ tests of sphericity $\phi_{S,\gamma}$, for $\gamma = 0.5$, 0.75, 0.9 and 0.95, their covariance-based competitor ϕ_S from Section 6, as well as the sign test and van der Waerden signed-rank test from Hallin & Paindaveine (2006a). Figure 5 plots the corresponding rejection frequencies as functions of m. This figure also reports the corresponding asymptotic powers, that are readily obtained from Theorems 4.3(iii) and 6.2(ii) (and from Proposition 4.1 in Hallin & Paindaveine (2006a)). MCD$_\gamma$ tests were based on the null estimators $\hat{c}^0_{k,\gamma}$ from Section 5. The "covMcd" function from the "Robusbase" R-package was used to select the best subsample among nsamp=5000 subsamples. The MCD$_\gamma$ estimator of shape was then obtained as the shape matrix associated with the covariance matrix of this subsample. The kernel density estimation involved in the testing procedure used a Gaussian kernel and the automatic bandwidth selection in Equation (3.31) from Silverman (1986), as implemented in the "density()" R function.

This simulation exercise clearly confirms our asymptotic results in Theorems 4.3 and 6.2 as the empirical rejection frequencies for $n = 2,000$ very well match the corresponding asymptotic powers; all findings associated with the AREs derived in Section 6 therefore show at this large sample size (in particular, MCD_γ tests, for large γ-values, dominate the covariance-based one under t_5). For small sample size ($n = 50$), the lowest γ-value considered ($\gamma = 0.5$) leads to slightly liberal tests, which is due to the relatively poor estimation (see Figure 2) of $c_{k,\gamma}$ by $\hat{c}^0_{k,\gamma}$. Simulations based on other alternatives led to extremely similar conclusions.

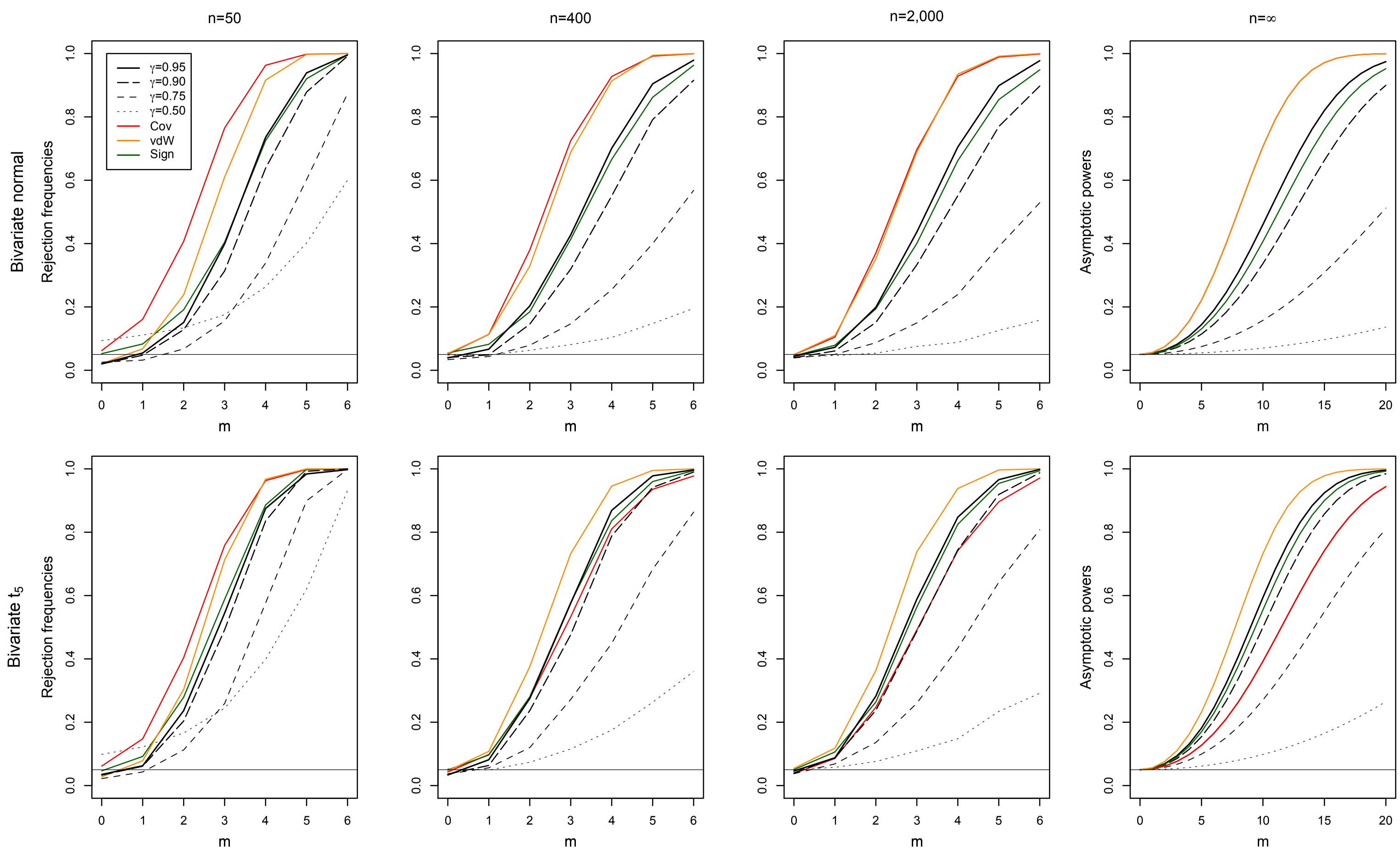

Figure 5: Rejection frequencies (first, second, and third columns, for $n = 50$, 400 and $2,000$, respectively) and asymptotic powers (rightmost column) of the MCD$_\gamma$ tests of sphericity $\phi_{S,\gamma}$, for $\gamma = 0.5$, 0.75, 0.9, and 0.95, their covariance-based competitor ϕ_S, the sign test and van der Waerden signed-rank test from Hallin & Paindaveine (2006a), under bivariate normal and t_5 densities. We refer to Section 7 for details.

We turn to simulations for point estimation. Parallel as above, we generated $M = 2,000$ independent random samples, of sizes $n = 400$ and $n = 10,000$, from the bivariate (without loss of generality, standard) normal and t_5 distributions. For each sample, we evaluated the MCD$_\gamma$ shape estimators $\hat{\mathbf{V}}_{S,\gamma}$, still for $\gamma = 0.5$, 0.75, 0.9 and 0.95, and their covariance-based competitor $\hat{\mathbf{V}}_S$. For the sake of comparison, we also computed the corresponding reweighted MCD$_\gamma$ estimators, obtained through the "covMcd" R function. For each shape estimators $\hat{\mathbf{V}} = (\hat{V}_{ij})$, Figure 6 provides the boxplots of the corresponding estimation errors for fixed diagonal and off-diagonal entries — more precisely, the boxplots of $(\hat{V}_{11} - 1)$ and $\hat{V}_{12}$ are reported there. The results confirm that, under multinormality, the covariance-based estimators dominate the MCD$_\gamma$ estimators, that become less and less accurate as γ decreases. Under heavy tails, however, MCD$_\gamma$ estimators, for large values of γ, are slightly more efficient than the covariance-based one, which is in line with the AREs in the lower-right panel of Figure 3. These finite-sample performances therefore thoroughly confirm our asymptotic (efficiency) results. Reweighted estimators dominate the original MCD estimators, but the difference is negligible for large γ.

Finally, we performed a simulation study in order to assess the robustness of MCD-based inference procedures for shape. As previously, we generated $M = 2,000$ independent random samples of size $n = 400$ from the bivariate standard normal and t_5 distributions. Contamination was then introduced by multiplying by four the first component of ψn observations in each sample; this was done for $\psi = 0, 0.05, 0.10, \ldots, 0.50$. Figure 7 shows the coverage frequencies of the asymptotic 95%-confidence intervals for $(\mathbf{V}_S)_{11}$ and for $(\mathbf{V}_S)_{12}$ based on $\hat{\mathbf{V}}_{S,\gamma}$, still for $\gamma = 0.5, 0.75, 0.9, 0.95$. These confidence intervals were obtained from Theorem 4.1, where the relevant asymptotic variance was estimated by plugging $\hat{\mathbf{V}}_{S,\gamma}$ and by using the estimator $\hat{c}_{k,\gamma}$ introduced in Section 5. As above, the raw MCD$_\gamma$ was computed through the "covMcd" R function, with nsamp= $5,000$ subsamples. For $(\mathbf{V}_S)_{11}$, robustness, as expected, increases as γ decreases. For $\gamma = 0.50$, high robustness is achieved despite the density-based estimator $\hat{c}_{k,\gamma}$ used in the procedure. Results are much more stable for $(\mathbf{V}_S)_{12}$ than for $(\mathbf{V}_S)_{11}$, which indicates that the increasingly poorer performances obtained for $(\mathbf{V}_S)_{11}$ as contamination increases, should not be attributed to the non-robustness of $\hat{c}_{k,\gamma}$.

Acknowledgments

Davy Paindaveine's research is supported by an A.R.C. contract from the Communauté Française de Belgique and by the IAP research network grant P7/06 of the Belgian government (Belgian Science Policy). Germain Van Bever thanks the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique, for its support via a Mandat d'Aspirant FNRS. The authors are grateful to two anonymous referees and an Associate Editor for their careful reading and insightful comments that led to substantial improvements of the manuscript. They also wish to thank Prof. Gentiane Haesbroeck for comments on the first version of this paper, and for her advices on MCD computation in R.

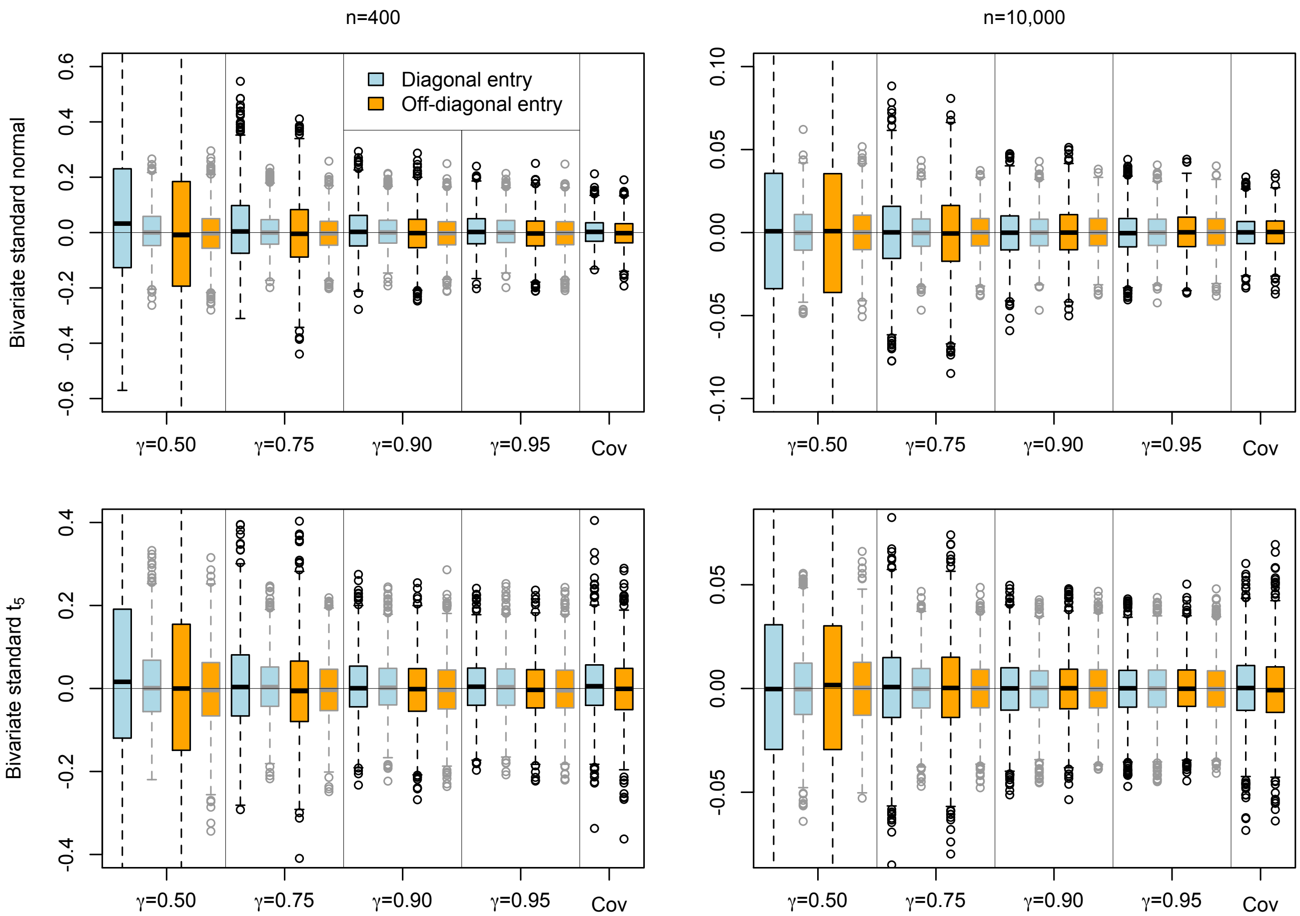

Figure 6: Boxplots, obtained from 2,000 independent bivariate normal or t_5 samples of size $n = 400$ or $n = 10,000$, of the diagonal estimation errors $\hat{V}_{11} - 1$ and of the off-diagonal ones $\hat{V}_{12}$, for the MCD$_\gamma$ shape estimators $\hat{\mathbf{V}}_{S,\gamma}$ (black borders) and their reweighted counterparts (grey borders), $\gamma = 0.5$, 0.75, 0.9 and 0.95, as well as for their covariance-based competitors $\hat{\mathbf{V}}_S$; see Section 7 for details.

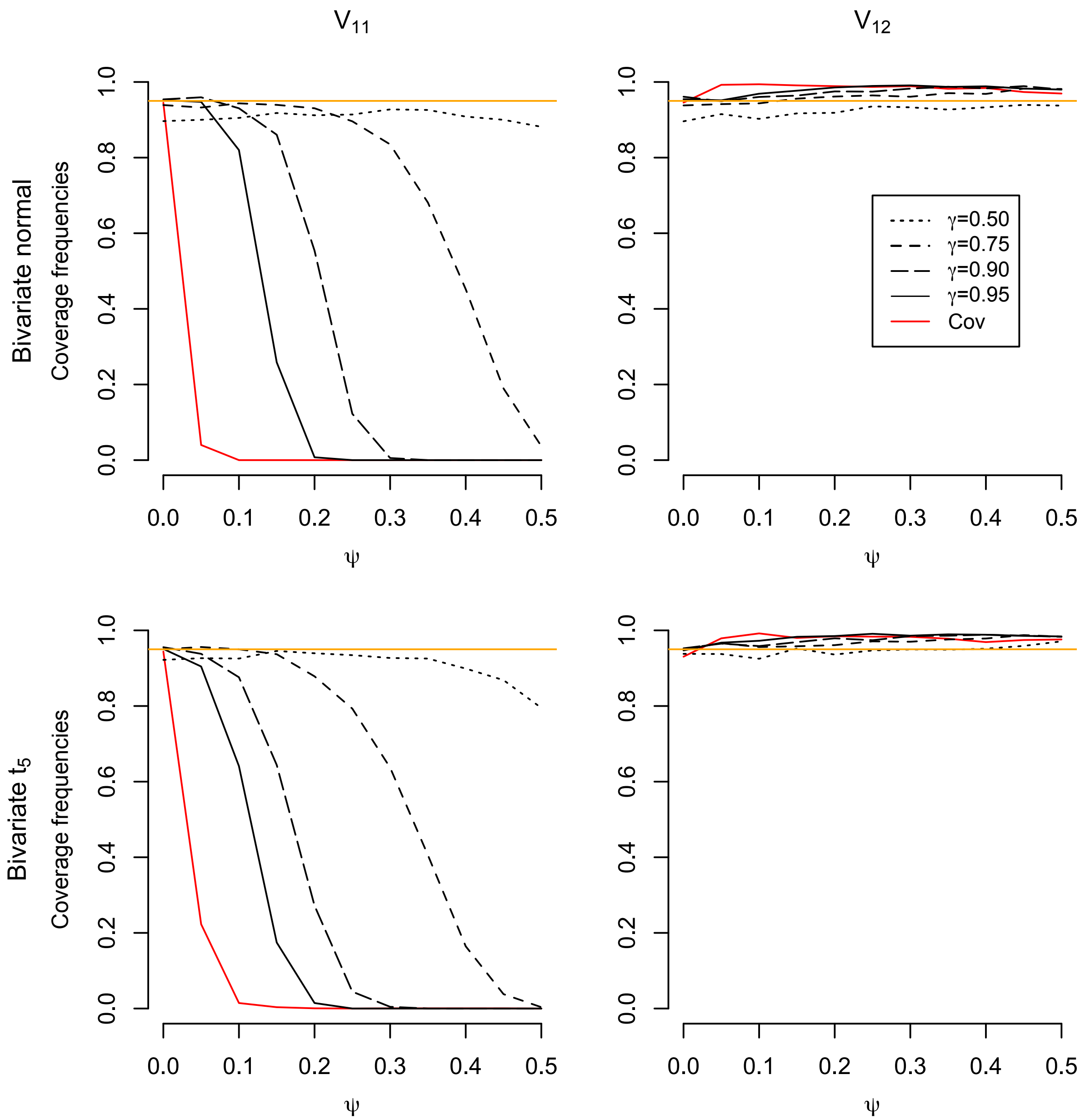

Figure 7: Coverage frequencies, as functions of the proportion ψ of contamination, of the asymptotic 95%-confidence intervals for $(\mathbf{V}_S)_{11}$ (left) and for $(\mathbf{V}_S)_{12}$ (right) based on $\hat{\mathbf{V}}_{S,\gamma}$, for $\gamma = 0.5, 0.75, 0.9, 0.95$, under bivariate normal (top) and t_5 (bottom) elliptical densities; see Section 7 for details. The target confidence level is shown in orange.

Appendix A.

In this appendix, we prove Theorems 4.1 and 4.3, Lemma 4.1, and Proposition 6.1.

Proof of Theorem 4.1. (i) The Delta method yields that, as $n \to \infty$,

$$\sqrt{n}\,\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S) = \frac{1}{S(\alpha_\gamma^2 \mathbf{\Sigma})} \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right] \sqrt{n}\,\mathrm{vec}(\hat{\mathbf{\Sigma}} - \alpha_\gamma^2 \mathbf{\Sigma}) + o_{\mathrm{P}}(1).$$

Since $\mathrm{tr}[\mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S] = S(\mathbf{V}_S) = 1$ (see Lemma 4.2(ii) in Paindaveine (2008)), this implies that

$$\left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right] (\mathrm{vec}\,\mathbf{V}_S) = (\mathrm{vec}\,\mathbf{V}_S) - \mathrm{tr}[\mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S](\mathrm{vec}\,\mathbf{V}_S) = \mathbf{0}. \tag{A.1}$$

The result then follows from the Bahadur representation result in Proposition 2.1, by using (A.1) and the identity $\mathrm{vec}\,(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,(\mathrm{vec}\,\mathbf{B})$.

(ii) Since

$$\mathrm{Var}_{\boldsymbol{\theta},\mathbf{\Sigma},f}\left[\mathrm{vec}\Big(\mathbf{U}_{i;\boldsymbol{\theta},\mathbf{V}_S} \mathbf{U}'_{i;\boldsymbol{\theta},\mathbf{V}_S} - \frac{1}{k}\mathbf{I}_k \Big) \right] = \frac{1}{k(k+2)} (\mathbf{I}_{k^2} + \mathbf{K}_k + \mathbf{J}_k) - \mathbf{J}_k =: \mathbf{A}_k,$$

we readily obtain that $\sqrt{n}\,\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S)$ is asymptotically normal with mean zero and covariance matrix

$$\frac{D_\gamma^{(4)}}{\beta_\gamma^2} \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right] \big(\mathbf{V}_S^{\otimes 2}\big)^{1/2} \mathbf{A}_k \big(\mathbf{V}_S^{\otimes 2}\big)^{1/2} \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right]'.$$

By using (A.1), $\mathbf{K}_k(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{A} \otimes \mathbf{B})\mathbf{K}_k$, and $\mathbf{K}_k(\mathrm{vec}\,\mathbf{A}) = \mathrm{vec}\,(\mathbf{A}')$, this covariance matrix rewrites

$$\frac{D_\gamma^{(4)}}{k(k+2)\beta_\gamma^2} \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right] \big(\mathbf{V}_S^{\otimes 2}\big)^{1/2} (\mathbf{I}_{k^2} + \mathbf{K}_k) \big(\mathbf{V}_S^{\otimes 2}\big)^{1/2} \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right]'$$

$$= c_{k,\gamma}^{-1} \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right] \big(\mathbf{V}_S^{\otimes 2}\big) (\mathbf{I}_{k^2} + \mathbf{K}_k) \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right]'$$

$$= c_{k,\gamma}^{-1} \left[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})' \right] \left[(\mathbf{I}_{k^2} + \mathbf{K}_k)\big(\mathbf{V}_S^{\otimes 2}\big) - 2\big(\mathbf{V}_S^{\otimes 2}\big)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})(\mathrm{vec}\,\mathbf{V}_S)' \right].$$

Performing this last product and using $(\mathrm{vec}\,\mathbf{A})'(\mathrm{vec}\,\mathbf{B}) = \mathrm{tr}[\mathbf{A}'\mathbf{B}]$ establishes the result. □

Proof of Lemma 4.1. The result follows by noting that $\mathbf{Q}_k^{\mathbf{V}_S} = \mathbf{Q}^{\mathbf{V}_S}_{k;1,2\mathcal{E}_k^{\mathbf{V}_S}}$, where $\mathbf{Q}^{\mathbf{V}_S}_{k;r,s}$ is defined in (5.3) from Hallin & Paindaveine (2006a), and by applying Lemma 5.2 from the same paper. □

Proof of Theorem 4.3. (i) Using first Lemma 4.1, then the identities $\mathrm{vec}\,(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,(\mathrm{vec}\,\mathbf{B})$ and $\mathbf{K}_k(\mathrm{vec}\,\mathbf{A}) = \mathrm{vec}\,(\mathbf{A}')$, and Lemma 5.1 from Hallin & Paindaveine (2006a), we obtain that, under the null as $n \to \infty$,

$$\begin{aligned}\mathring{Q}_{S,\gamma} &= \frac{n}{4\hat{c}_{k,\gamma}}\Big[\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)\Big]'\big(\mathbf{V}_S^{0\otimes 2}\big)^{-1/2}\Big[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k}\mathbf{J}_k\Big]\big(\mathbf{V}_S^{0\otimes 2}\big)^{-1/2}\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0) + o_{\mathrm{P}}(1) \\ &= \frac{n}{2\hat{c}_{k,\gamma}}\Big[\mathrm{vec}((\mathbf{V}_S^0)^{-1/2}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{V}_S^0)^{-1/2} - \mathbf{I}_k)\Big]'\Big[\mathbf{I}_{k^2} - \frac{1}{k}\mathbf{J}_k\Big]\Big[\mathrm{vec}((\mathbf{V}_S^0)^{-1/2}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{V}_S^0)^{-1/2} - \mathbf{I}_k)\Big] + o_{\mathrm{P}}(1).\end{aligned}$$

From the identities $\big[\mathbf{I}_{k^2} - \frac{1}{k}\mathbf{J}_k\big](\mathrm{vec}\,\mathbf{I}_k) = \mathbf{0}$ and $(\mathrm{vec}\,\mathbf{A})'(\mathrm{vec}\,\mathbf{B}) = \mathrm{tr}[\mathbf{A}'\mathbf{B}]$, we then obtain that, still under the null as $n \to \infty$,

$$\mathring{Q}_{S,\gamma} = \frac{n}{2\hat{c}_{k,\gamma}}\Big(\mathrm{tr}\Big[((\mathbf{V}_S^0)^{-1/2}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{V}_S^0)^{-1/2}))^2\Big] - \frac{1}{k}\,\mathrm{tr}^2\Big[(\mathbf{V}_S^0)^{-1/2}\hat{\mathbf{V}}_{S,\gamma}(\mathbf{V}_S^0)^{-1/2}\Big]\Big) + o_{\mathrm{P}}(1),$$

which establishes the result.

(ii) This readily follows from Part (i) of the result, the consistency of $\hat{c}_{k,\gamma}$, and the fact that $\sqrt{n}\,\mathrm{v\mathring{e}ch}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S)$ is asymptotically normal with mean zero and (full rank K; see Lemma 4.1) covariance $c_{k,\gamma}^{-1}\,\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S}\mathbf{N}_k'$.

(iii) Under Assumption (A′), the fixed-f parametric model described by $\mathcal{P}_f := \{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma_S^2,\mathrm{v\mathring{e}ch}\,\mathbf{V}_S;f}\}$ (where $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma_S^2,\mathrm{v\mathring{e}ch}\,\mathbf{V}_S;f}$ denotes the probability measure of n i.i.d. k-variate elliptical observations with location $\boldsymbol{\theta}$, scale σ_S^2, shape $\mathbf{V}_S$, and radial density f) is uniformly locally and asymptotically normal (ULAN) with a central sequence of the form $\boldsymbol{\Delta}_f = ((\boldsymbol{\Delta}_f^{\boldsymbol{\theta}})', \Delta_f^{\sigma_S^2}, (\boldsymbol{\Delta}_f^{\mathbf{V}_S})')'$, where

$$\boldsymbol{\Delta}_f^{\mathbf{V}_S} := \frac{1}{2\sqrt{n}}\,\mathbf{M}_k^{\mathbf{V}_S}\big(\mathbf{V}_S^{\otimes 2}\big)^{-1/2}\sum_{i=1}^n \mathrm{vec}\bigg(\frac{d_{i;\boldsymbol{\theta},\mathbf{V}_S}}{\sigma_S}\varphi_f\Big(\frac{d_{i;\boldsymbol{\theta},\mathbf{V}_S}}{\sigma_S}\Big)\,\mathbf{U}_{i;\boldsymbol{\theta},\mathbf{V}_S}\mathbf{U}'_{i;\boldsymbol{\theta},\mathbf{V}_S} - \frac{1}{k}\mathbf{I}_k\bigg);$$

see, e.g., Paindaveine (2008). This ULAN result in particular entails that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma_S^2,\mathrm{v\mathring{e}ch}\,\mathbf{V}_S;f}$,

$$\begin{aligned}T^{(n)} &:= \log\Big(d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\mathrm{v\mathring{e}ch}(\mathbf{V}_S^0+n^{1/2}\mathbf{v});f}/d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\mathrm{v\mathring{e}ch}\,\mathbf{V}_S^0;f}\Big) \\ &= (\mathrm{v\mathring{e}ch}\,\mathbf{v})'\boldsymbol{\Delta}_f^{\mathbf{V}_S} - \frac{1}{2}(\mathrm{v\mathring{e}ch}\,\mathbf{v})'\boldsymbol{\Gamma}_f^{\mathbf{V}_S}(\mathrm{v\mathring{e}ch}\,\mathbf{v}) + o_{\mathrm{P}}(1)\end{aligned}$$

as $n \to \infty$, where $\boldsymbol{\Gamma}_f^{\mathbf{V}_S}$ denotes the covariance matrix in the asymptotically normal distribution of $\boldsymbol{\Delta}_f^{\mathbf{V}_S}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma_S^2,\mathrm{v\mathring{e}ch}\,\mathbf{V}_S;f}$. Hence, the standard Cramér-Wold device shows that, still under the same, the joint asymptotic distribution of $\mathbf{S}^{(n)} := \sqrt{n}\,\mathrm{v\mathring{e}ch}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)$ and $T^{(n)}$ is asymptotically multinormal, with an asymptotic covariance between $\mathbf{S}^{(n)}$ and $T^{(n)}$ that is given by $\mathbf{w} = \lim_{n\to\infty} \mathrm{E}_{\boldsymbol{\theta},\sigma_S^2,\mathbf{V}_S^0;f}[\mathbf{S}^{(n)}(\boldsymbol{\Delta}_f^{\mathbf{V}_S})'](\mathrm{v\mathring{e}ch}\,\mathbf{v})$. By first using Theorem 4.1(i) and $\mathbf{N}_k(\mathrm{vec}\,\mathbf{A}) = (\mathrm{v\mathring{e}ch}\,\mathbf{A})$, then by simplifying $\mathbf{w}$ along the same lines as in the previous proofs, we obtain

$$\mathbf{w} = \frac{1}{k(k+2)\beta_\gamma}\,\mathrm{E}_{\boldsymbol{\theta},\sigma_S^2,\mathbf{V}_S^0;f}\bigg[\frac{d^2_{i;\boldsymbol{\theta},\mathbf{V}_S^0}}{\sigma_S^2}\mathbb{I}\Big[\frac{d_{i;\boldsymbol{\theta},\mathbf{V}_S^0}}{\sigma_S} \leq r_\gamma\Big] \times \frac{d_{i;\boldsymbol{\theta},\mathbf{V}_S^0}}{\sigma_S}\varphi_f\Big(\frac{d_{i;\boldsymbol{\theta},\mathbf{V}_S^0}}{\sigma_S}\Big)\bigg]\,(\mathrm{v\mathring{e}ch}\,\mathbf{v}),$$

which, in view of (3), yields $\mathbf{w} = (\overset{\circ}{\text{vech}}\ \mathbf{v})$. Le Cam's third lemma then yields that $\mathbf{S}^{(n)}$ is asymptotically normal, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\overset{\circ}{\text{vech}}(\mathbf{V}_S^0+n^{1/2}\mathbf{v});f}$, with mean $(\overset{\circ}{\text{vech}}\ \mathbf{v})$ and the same covariance matrix $c_{k,\gamma}^{-1}\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S^0}\mathbf{N}_k'$ as under the null. Hence, still under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\overset{\circ}{\text{vech}}(\mathbf{V}_S^0+n^{1/2}\mathbf{v});f}$,

$$Q_{S,\gamma} = \mathring{Q}_{S,\gamma} + o_{\mathrm{P}}(1) = c_{k,\gamma}(\mathbf{S}^{(n)})'\big(\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S}\mathbf{N}_k'\big)^{-1}\mathbf{S}^{(n)} + o_{\mathrm{P}}(1)$$

(contiguity implies that the first part of the theorem and the consistency of $\hat{c}_{k,\gamma}$ extend to the local alternatives considered) is asymptotically non-central χ^2_K with non-centrality parameter

$$\hat{c}_{k,\gamma}(\overset{\circ}{\text{vech}}\ \mathbf{v})'\big(\mathbf{N}_k\mathbf{Q}_k^{\mathbf{V}_S}\mathbf{N}_k'\big)^{-1}(\overset{\circ}{\text{vech}}\ \mathbf{v}),$$

which, after some computations, reduces to the non-centrality parameter in the statement of the theorem. □

Proof of Proposition 6.1. Decomposing as usual $\hat{\boldsymbol{\Sigma}}$ into $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} + (\bar{\mathbf{X}} - \boldsymbol{\theta})(\bar{\mathbf{X}} - \boldsymbol{\theta})'$, with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} := \frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i - \boldsymbol{\theta})(\mathbf{X}_i - \boldsymbol{\theta})'$, we obtain that, as $n \to \infty$,

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\Sigma}} - \frac{D^{(2)}}{k}\boldsymbol{\Sigma}\right) &= \sqrt{n}\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} - \frac{D^{(2)}}{k}\boldsymbol{\Sigma}\right) + o_{\mathrm{P}}(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(d^2_{i;\boldsymbol{\theta},\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{1/2}\mathbf{U}_{i;\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{U}'_{i;\boldsymbol{\theta},\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{1/2} - \frac{D^{(2)}}{k}\boldsymbol{\Sigma}\right) + o_{\mathrm{P}}(1),\end{aligned}$$

which establishes the result. □

Appendix B.

It remains to prove Theorem 5.1. The proof relies on several lemmas. We first introduce the following notation. Let $\hat{s}_\gamma^{(n)}$ and $s_\gamma^{(n)}$ be the sample γ-quantiles of $\hat{d}_i = d_{i;\hat{\boldsymbol{\theta}}_\gamma,\hat{\mathbf{V}}_{S,\gamma}}$, $i = 1, \ldots, n$, and $d_i = d_{i;\boldsymbol{\theta},\mathbf{V}_S}$, $i = 1, \ldots, n$, respectively. In the rest of the paper, all convergences, o_{P}'s, and O_{P}'s are as $n \to \infty$. Also, we will write $\max_i$ and $\sum_i$ for $\max_{i=1,\ldots,n}$ and $\sum_{i=1}^n$, respectively. Similarly, "for some i" will stand for "for some $i \in \{1, \ldots, n\}$".

Lemma B.1. *Let $(L^{(n)})$ be a sequence of random variables that is $O_{\mathrm{P}}(1)$. Then (i) $\max_i\big(|\hat{d}_i - d_i|\mathbb{I}[d_i \le L^{(n)}]\big) = O_{\mathrm{P}}(n^{-1/2})$ and (ii) $\max_i\big(|\hat{d}_i - d_i|\mathbb{I}[\hat{d}_i \le L^{(n)}]\big) = O_{\mathrm{P}}(n^{-1/2})$.*

Proof of Lemma B.1. (i) Using repeatedly the triangular inequality provides

$$\begin{aligned}|\hat{d}_i - d_i| &\le \|\hat{\mathbf{V}}_{S,\gamma}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\theta}}_\gamma) - \mathbf{V}_S^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})\|\\ &= \|(\hat{\mathbf{V}}_{S,\gamma}^{-1/2} - \mathbf{V}_S^{-1/2})(\mathbf{X}_i - \boldsymbol{\theta}) - \hat{\mathbf{V}}_{S,\gamma}^{-1/2}(\hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta})\|\\ &\le \|\hat{\mathbf{V}}_{S,\gamma}^{-1/2} - \mathbf{V}_S^{-1/2}\|_{\mathcal{L}}\|\mathbf{V}_S^{1/2}\|_{\mathcal{L}}\, d_i + \|\hat{\mathbf{V}}_{S,\gamma}^{-1/2}\|_{\mathcal{L}}\|\hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}\|,\end{aligned}$$

where $\|\mathbf{A}\|_{\mathcal{L}} = \sup\{\mathbf{A}\mathbf{u} : \mathbf{u} \in \mathcal{S}^{k-1}\}$ is the operator norm of $\mathbf{A}$. Hence, we obtain

$$\sqrt{n}\max_i\big(|\hat{d}_i - d_i|\,\mathbb{I}[d_i \leq L^{(n)}]\big) \leq L^{(n)}\|\sqrt{n}(\hat{\mathbf{V}}_{S,\gamma}^{-1/2} - \mathbf{V}_S^{-1/2})\|_{\mathcal{L}}\|\mathbf{V}_S^{1/2}\|_{\mathcal{L}} + \|\hat{\mathbf{V}}_{S,\gamma}^{-1/2}\|\|\sqrt{n}(\hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta})\|,$$

so that the result follows from the root-n consistency of $\hat{\boldsymbol{\theta}}_\gamma$ and $\hat{\mathbf{V}}_{S,\gamma}$. (ii) The proof is entirely similar but is based on the decomposition

$$\begin{aligned}
|\hat{d}_i - d_i| &\leq \|\hat{\mathbf{V}}_{S,\gamma}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\theta}}_\gamma) - \mathbf{V}_S^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})\| \\
&= \|(\hat{\mathbf{V}}_{S,\gamma}^{-1/2} - \mathbf{V}_S^{-1/2})(\mathbf{X}_i - \hat{\boldsymbol{\theta}}_\gamma) - \mathbf{V}_{S,\gamma}^{-1/2}(\hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta})\| \\
&\leq \|\hat{\mathbf{V}}_{S,\gamma}^{-1/2} - \mathbf{V}_S^{-1/2}\|_{\mathcal{L}}\|\hat{\mathbf{V}}_{S,\gamma}^{1/2}\|_{\mathcal{L}}\,\hat{d}_i + \|\mathbf{V}_{S,\gamma}^{-1/2}\|_{\mathcal{L}}\|\hat{\boldsymbol{\theta}}_\gamma - \boldsymbol{\theta}\|. \qquad \square
\end{aligned}$$

Lemma B.2. *Let $A^{(n)}$ and $B^{(n)}$ be two sequences of random variables that converge in probability to constants a and b, respectively, with $a < b$. Then $\mathrm{P}[d_i \leq A^{(n)}, \hat{d}_i \geq B^{(n)}$ for some $i]$ and $\mathrm{P}[\hat{d}_i \leq A^{(n)}, d_i \geq B^{(n)}$ for some $i]$ both converge to zero as $n \to \infty$.*

Proof of Lemma B.2. Fix $\delta \in (0, (b-a)/3)$. We then have

$$\begin{aligned}
&\mathrm{P}[d_i \leq A^{(n)}, \hat{d}_i \geq B^{(n)} \text{ for some } i] \\
&\leq \mathrm{P}[d_i \leq A^{(n)}, \hat{d}_i \geq B^{(n)} \text{ for some } i, A^{(n)} \leq a+\delta, B^{(n)} \geq b-\delta] + \mathrm{P}[A^{(n)} > a+\delta] + \mathrm{P}[B^{(n)} < b-\delta] \\
&\leq \mathrm{P}[d_i \leq a+\delta, \hat{d}_i \geq b-\delta \text{ for some } i] + \mathrm{P}[|A^{(n)} - a| > \delta] + \mathrm{P}[|B^{(n)} - b| > \delta] \\
&\leq \mathrm{P}\big[\max_i|\hat{d}_i - d_i|\,\mathbb{I}[d_i \leq a+\delta] > \delta\big] + \mathrm{P}[|A^{(n)} - a| > \delta] + \mathrm{P}[|B^{(n)} - b| > \delta],
\end{aligned}$$

so that the (i) follows from Lemma B.1(i). Interchanging d_i and $\hat{d}_i$, one concludes that $\mathrm{P}[\hat{d}_i \leq A^{(n)}, d_i \geq B^{(n)},$ for some $i]$ converges to zero as $n \to \infty$ (this time by using Lemma B.1(ii)). $\square$

Lemma B.3. *Recalling that s_γ is the γ-quantile of $d_{1;\boldsymbol{\theta},\mathbf{V}_S}$, we have that $\sqrt{n}(\hat{s}_\gamma^{(n)} - s_\gamma^{(n)})$ and $\sqrt{n}(\hat{s}_\gamma^{(n)} - s_\gamma)$ are $O_{\mathrm{P}}(1)$ as $n \to \infty$.*

Proof of Lemma B.3. Since the d_i's are i.i.d., the root-n consistency of sample quantiles trivially entails that $\sqrt{n}(s_\gamma^{(n)} - s_\gamma) = O_{\mathrm{P}}(1)$, so that it is sufficient to show that $\sqrt{n}(\hat{s}_\gamma^{(n)} - s_\gamma^{(n)}) = O_{\mathrm{P}}(1)$. To do so, fix $\varepsilon > 0$, and write

$$\begin{aligned}
\mathrm{P}\big[\sqrt{n}\,|\hat{s}_\gamma^{(n)} - s_\gamma^{(n)}| > \varepsilon\big] &\leq \mathrm{P}\big[\sqrt{n}\,|\hat{s}_\gamma^{(n)} - s_\gamma^{(n)}| > \varepsilon, s_\gamma^{(n)} \leq s_\gamma + 1\big] + \mathrm{P}\big[s_\gamma^{(n)} > s_\gamma + 1\big] \\
&\leq \mathrm{P}\big[\sqrt{n}\max_i|\hat{d}_i - d_i|\mathbb{I}[d_i \leq s_\gamma + 1] > \varepsilon\big] + \mathrm{P}\big[s_\gamma^{(n)} > s_\gamma + 1\big].
\end{aligned}$$

The result then follows from Lemma B.1(i) and the fact that $s_\gamma^{(n)} - s_\gamma$ is $o_{\mathrm{P}}(1)$. $\square$

Lemma B.4. *Let the assumptions of Theorem 5.1 hold. Then, for any $\alpha \in (0,1)$,*

$$\sup_{s \in [0, s_\alpha]} |\hat{f}^{(n)}(s) - f^{(n)}(s)| = o_{\mathrm{P}}(1),$$

where we let $\hat{f}^{(n)}(s) = (nh_n)^{-1}\sum_i K\big(\frac{s-\hat{d}_i}{h_n}\big)$ and $f^{(n)}(s) = (nh_n)^{-1}\sum_i K\big(\frac{s-d_i}{h_n}\big)$.

Proof of Lemma B.4. Pick an arbitrary $\alpha' \in (\alpha, 1)$, and write $|\hat{f}^{(n)}(s) - f^{(n)}(s)| \leq T_1^{(n)}(s) + T_2^{(n)}(s)$, with

$$T_1^{(n)}(s) = \frac{1}{nh_n}\sum_i \Big|K\Big(\frac{s-\hat{d}_i}{h_n}\Big) - K\Big(\frac{s-d_i}{h_n}\Big)\Big|\, \mathbb{I}[d_i \leq s_{\alpha'}]$$

and

$$T_2^{(n)}(s) = \frac{1}{nh_n}\sum_i \Big|K\Big(\frac{s-\hat{d}_i}{h_n}\Big) - K\Big(\frac{s-d_i}{h_n}\Big)\Big|\, \mathbb{I}[d_i > s_{\alpha'}].$$

From the mean value theorem and the boundedness of K', we obtain

$$\begin{aligned}\sup_{s\in[0,s_\alpha]} T_1^{(n)}(s) &\leq \frac{C}{nh_n}\sup_{s\in[0,s_\alpha]}\sum_i \Big|\big(\frac{s-\hat{d}_i}{h_n}\big) - \big(\frac{s-d_i}{h_n}\big)\Big|\, \mathbb{I}[d_i \leq s_{\alpha'}] \\ &= \frac{C}{nh_n^2}\sum_i |\hat{d}_i - d_i|\, \mathbb{I}[d_i \leq s_{\alpha'}] \leq \frac{C}{\sqrt{nh_n^4}}\Big(\sqrt{n}\max_i|\hat{d}_i - d_i|\, \mathbb{I}[d_i \leq s_{\alpha'}]\Big),\end{aligned}$$

which is $o_{\mathrm{P}}(1)$ (since $nh_n^4 \to \infty$ and the sequence in the brackets is $O_{\mathrm{P}}(1)$ in view of Lemma B.1(i)). Turning then to $T_2^{(n)}(s)$, pick $c > 0$ so that the support of K is a subset of $[-c, c]$. Fix $\delta \in (0, s_{\alpha'} - s_\alpha)$ and choose n_0 so that $s_\alpha + ch_n < s_{\alpha'} - \delta$ for all $n \geq n_0$. In the rest of the proof, we restrict (without loss of generality) to $n \geq n_0$. For any $s \in [0, s_\alpha]$, we then trivially have $s + ch_n \leq s_\alpha + ch_n < s_{\alpha'} - \delta$. Therefore, we have that, for all $i = 1, \ldots, n$,

$$K\Big(\frac{s-d_i}{h_n}\Big)\, \mathbb{I}[d_i > s_{\alpha'}] = 0,$$

almost surely, which implies that, still almost surely and any $s \in [0, s_\alpha]$,

$$T_2^{(n)}(s) = \frac{1}{nh_n}\sum_i \Big|K\Big(\frac{s-\hat{d}_i}{h_n}\Big)\Big|\, \mathbb{I}[d_i > s_{\alpha'}] = \frac{1}{nh_n}\sum_i \Big|K\Big(\frac{s-\hat{d}_i}{h_n}\Big)\Big|\, \mathbb{I}[d_i > s_{\alpha'}, \hat{d}_i \leq s_{\alpha'} - \delta]$$

(since $\hat{d}_i > s_{\alpha'} - \delta$ would entail $|s - \hat{d}_i| \geq ch_n$). Hence,

$$\mathrm{P}\Big[\Big|\sup_{s\in[0,s_\alpha]} T_2^{(n)}\Big| > \varepsilon\Big] \leq \mathrm{P}\Big[\Big|\sup_{s\in[0,s_\alpha]} T_2^{(n)}\Big| \neq 0\Big] \;\;\leq\;\; \mathrm{P}[d_i > s_{\alpha'}, \hat{d}_i \leq s_{\alpha'} - \delta, \text{ for some } i]$$

which, in view of Lemma B.2, converges to zero. We conclude that both $T_1^{(n)}(s)$ and $T_2^{(n)}(s)$, hence also $|\hat{f}^{(n)}(s) - f^{(n)}(s)|$, are $o_{\mathrm{P}}(1)$ uniformly in $s \in [0, s_\alpha]$. □

Lemma B.5. *For any integer ℓ, $\frac{1}{n}\sum_i \hat{d}_i^\ell\, \mathbb{I}[\hat{d}_i \leq \hat{s}_\gamma^{(n)}]$ converges in probability to $\mathrm{E}[d^\ell_{\boldsymbol{\theta},\mathbf{V}_S}\mathbb{I}[d_{\boldsymbol{\theta},\mathbf{V}_S} \leq s_\gamma]]$ as $n \to \infty$.*

Proof of Lemma B.5. The weak law of large numbers implies that it is sufficient to show that

$$S^{(n)} = \frac{1}{n}\sum_i \hat{d}_i^\ell\, \mathbb{I}[\hat{d}_i \leq \hat{s}_\gamma^{(n)}] - \frac{1}{n}\sum_i d_i^\ell\, \mathbb{I}[d_i \leq s_\gamma] = o_{\mathrm{P}}(1).$$

Decompose then $S^{(n)}$ into $S_1^{(n)} + S_2^{(n)}$, where

$$S_1^{(n)} = \frac{1}{n}\sum_i (\hat{d}_i^\ell - d_i^\ell)\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}] \quad \text{and} \quad S_2^{(n)} = \frac{1}{n}\sum_i d_i^\ell \Big(\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}] - \mathbb{I}[d_i \le s_\gamma]\Big).$$

Let us start with $S_1^{(n)}$ (which only needs to be considered if $\ell \ge 1$). We have

$$\begin{aligned}
|S_1^{(n)}| &\le \frac{1}{n}\sum_i |\hat{d}_i^\ell - d_i^\ell|\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i < \hat{s}_\gamma^{(n)} + 1] + \frac{1}{n}\sum_i |\hat{d}_i^\ell - d_i^\ell|\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i \ge \hat{s}_\gamma^{(n)} + 1] \\
&\le \max_i |\hat{d}_i^\ell - d_i^\ell|\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i < \hat{s}_\gamma^{(n)} + 1] + \frac{1}{n}\sum_i |\hat{d}_i^\ell - d_i^\ell|\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i \ge \hat{s}_\gamma^{(n)} + 1]. \\
&= S_{1a}^{(n)} + S_{1b}^{(n)},
\end{aligned}$$

say. Using the mean value theorem, then Lemma B.1, yields

$$S_{1a}^{(n)} \le \ell(\hat{s}_\gamma^{(n)} + 1)^{\ell-1} \max_i |\hat{d}_i - d_i|\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i < \hat{s}_\gamma^{(n)} + 1] = o_{\mathrm{P}}(1).$$

As for $S_{1b}^{(n)}$, we have that, for any $\varepsilon > 0$,

$$\mathrm{P}[S_{1b}^{(n)} > \varepsilon] \le \mathrm{P}[d_i \le \hat{s}_\gamma^{(n)}, d_i \ge \hat{s}_\gamma^{(n)} + 1, \text{ for some } i],$$

which converges to zero (Lemma B.2). We conclude that $S_1^{(n)}$ is $o_{\mathrm{P}}(1)$.

Turning to $S_2^{(n)}$,

$$\begin{aligned}
|S_2^{(n)}| &\le \frac{1}{n}\sum_i d_i^\ell \left|\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}] - \mathbb{I}[d_i \le s_\gamma]\right| \\
&= \frac{1}{n}\sum_i d_i^\ell\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i > s_\gamma] + \frac{1}{n}\sum_i d_i^\ell\,\mathbb{I}[\hat{d}_i > \hat{s}_\gamma^{(n)}, d_i \le s_\gamma] = S_{2a}^{(n)} + S_{2b}^{(n)},
\end{aligned}$$

say. For any $\eta > 0$, we may write

$$\begin{aligned}
&\mathrm{P}\big[\big|S_{2a}^{(n)}\big| > \varepsilon\big] \\
&\le \mathrm{P}\left[\frac{1}{n}\sum_i d_i^\ell\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i > s_\gamma + \eta] > \frac{\varepsilon}{2}\right] + \mathrm{P}\left[\frac{1}{n}\sum_i d_i^\ell\,\mathbb{I}[\hat{d}_i \le \hat{s}_\gamma^{(n)}, s_\gamma < d_i \le s_\gamma + \eta] > \frac{\varepsilon}{2}\right] \\
&\le \mathrm{P}\big[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i > s_\gamma + \eta \text{ for some } i\big] + \mathrm{P}\left[\frac{1}{n}\sum_i d_i^\ell\,\mathbb{I}[s_\gamma < d_i \le s_\gamma + \eta] > \frac{\varepsilon}{2}\right] \\
&\le \mathrm{P}\big[\hat{d}_i \le \hat{s}_\gamma^{(n)}, d_i > s_\gamma + \eta \text{ for some } i\big] + \frac{2}{\varepsilon}\,\mathrm{E}\big[d_1^\ell\,\mathbb{I}[s_\gamma < d_1 \le s_\gamma + \eta]\big], \qquad \text{(B.1)}
\end{aligned}$$

where the last inequality follows from Markov's inequality. The second term of (B.1) does not depend on n and can be made arbitrarily small by choosing η appropriately. Since Lemma B.2 implies that the first term of (B.1) converges to zero, this yields that $S_{2a}^{(n)}$ is $o_{\mathrm{P}}(1)$. The proof that $S_{2b}^{(n)}$ is also $o_{\mathrm{P}}(1)$ is extremely similar. We conclude that $S^{(n)}$ itself is $o_{\mathrm{P}}(1)$, which establishes the result. □

We can now prove Theorem 5.1.

Proof of Theorem 5.1. Fix $\alpha \in (\gamma, 1)$ and pick $\delta > 0$ such that $s_\gamma + \delta < s_\alpha$. Then (note that, using the notation of this Appendix, we have $\hat{\hat{f}}_{k;\text{shape}}(\hat{s}_\gamma) = \hat{f}^{(n)}(\hat{s}_\gamma)$),

$$\mathrm{P}\big[\big|\hat{\hat{f}}_{k;\text{shape}}(\hat{s}_\gamma^{(n)}) - f^{(n)}(s_\gamma)\big| > \varepsilon\big] \leq \mathrm{P}\Big[\big|\hat{f}^{(n)}(\hat{s}_\gamma^{(n)}) - f^{(n)}(\hat{s}_\gamma^{(n)})\big| > \frac{\varepsilon}{2}\Big] + \mathrm{P}\Big[\big|\hat{f}^{(n)}(s_\gamma) - f^{(n)}(s_\gamma)\big| > \frac{\varepsilon}{2}\Big]$$

$$\leq \mathrm{P}\Big[\big|\hat{f}^{(n)}(\hat{s}_\gamma^{(n)}) - f^{(n)}(\hat{s}_\gamma^{(n)})\big| > \frac{\varepsilon}{2}, |\hat{s}_\gamma^{(n)} - s_\gamma| \leq \delta|\Big] + \mathrm{P}\big[|\hat{s}_\gamma^{(n)} - s_\gamma| > \delta\big] + \mathrm{P}\Big[\big|\hat{f}^{(n)}(s_\gamma) - f^{(n)}(s_\gamma)\big| > \frac{\varepsilon}{2}\Big]$$

$$\leq 2\,\mathrm{P}\Big[\sup_{s\in[0,s_\alpha]} \big|\hat{f}^{(n)}(s) - f^{(n)}(s)\big| > \frac{\varepsilon}{2}\Big] + \mathrm{P}\big[|\hat{s}_\gamma^{(n)} - s_\gamma| > \delta\big],$$

which, by using Lemmas B.3-B.4, shows that $\hat{\hat{f}}_{k;\text{shape}}(\hat{s}_\gamma^{(n)}) - f^{(n)}(s_\gamma)$ is $o_\mathrm{P}(1)$. The weak consistency of the standard kernel density estimator $f^{(n)}(s)$ then entails that $\hat{\hat{f}}_{k;\text{shape}}(\hat{s}_\gamma^{(n)}) - f(s_\gamma)$ is $o_\mathrm{P}(1)$. Hence, the result follows from the continuous mapping theorem, Lemma B.3, and Lemma B.5. □

It remains to prove Theorem 5.2, which requires the following preliminary result.

Lemma B.6. *Let Assumption (A′) hold. Let $\hat{\boldsymbol{\theta}}_\#$ and $\hat{s}_{\gamma\#}$ be root-n consistent and locally asymptotically discrete estimators of $\boldsymbol{\theta}$ and s_γ, respectively. Then*

$$\sqrt{n}\,\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S) \;=\; \frac{1}{\sigma^2\beta_\gamma\sqrt{n}}\Big[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S})'\Big]$$
$$\times\big(\mathbf{V}_S^{\otimes 2}\big)^{1/2}\sum_{i=1}^n d^2_{\hat{\boldsymbol{\theta}}_\#,\mathbf{V}_S}\mathbb{I}[d_{\hat{\boldsymbol{\theta}}_\#,\mathbf{V}_S} \leq \hat{s}_{\gamma\#}]\mathrm{vec}\Big(\mathbf{U}_{i;\hat{\boldsymbol{\theta}}_\#,\mathbf{V}_S}\mathbf{U}'_{i;\hat{\boldsymbol{\theta}}_\#,\mathbf{V}_S} - \frac{1}{k}\,\mathbf{I}_k\Big) + o_\mathrm{P}(1)$$

as $n \to \infty$.

Proof of Lemma B.6. In Sections 2 and 3, we parametrized the family of elliptical distributions by $(\boldsymbol{\theta}, \sigma^2, \mathbf{V}_S)$ and the radial density f, where identifiability of σ^2 and f follows by imposing that $d_{\boldsymbol{\theta},\mathbf{V}_S} = ((\mathbf{X}-\boldsymbol{\theta})'\mathbf{V}_S^{-1}(\mathbf{X}-\boldsymbol{\theta}))^{1/2}$ has median σ. For any given $\gamma \in (0,1)$, one may equivalently adopt the parametrization in $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, s_\gamma, \mathbf{V}_S)$ and f associated with the densities

$$\begin{aligned} f^{\mathbf{X}} &: \mathbb{R}^k \to \mathbb{R} \\ \mathbf{x} &\mapsto \frac{(\mu_{k-1,f}\omega_{k-1})^{-1}}{s_\gamma^k\sqrt{\det \mathbf{V}_S}}\, f\Big(s_\gamma^{-1}\sqrt{(\mathbf{x}-\boldsymbol{\theta})'\mathbf{V}_S^{-1}(\mathbf{x}-\boldsymbol{\theta})}\,\Big), \end{aligned} \tag{B.2}$$

where the scale parameter s_γ is defined as the γ-quantile of $d_{\boldsymbol{\theta},\mathbf{V}_S}$. Proceeding as in Paindaveine (2008) (Section 4), it is seen that, for fixed f (satisfying some mild regularity conditions), the parametric family of f-elliptical distributions is ULAN with a central sequence that, in this alternative parametrization, takes the form $\Delta^{(n)}_{\boldsymbol{\vartheta},f} := ((\Delta^{(n)}_{\boldsymbol{\vartheta},f;1})', \Delta^{(n)}_{\boldsymbol{\vartheta},f;2}, (\Delta^{(n)}_{\boldsymbol{\vartheta},f;3})')'$, with

$$\Delta^{(n)}_{\boldsymbol{\vartheta},f;1} := \frac{1}{s_\gamma\sqrt{n}}\sum_{i=1}^n \varphi_f\left(\frac{d_i}{s_\gamma}\right)\mathbf{V}_S^{-1/2}\mathbf{U}_i,$$

$$\Delta^{(n)}_{\boldsymbol{\vartheta},f;2} := \frac{1}{2s_\gamma^2\sqrt{n}} \sum_{i=1}^n \left(\varphi_f\left(\frac{d_i}{s_\gamma}\right)\frac{d_i}{s_\gamma} - k \right), \tag{B.3}$$

and

$$\Delta^{(n)}_{\boldsymbol{\vartheta},f;3} := \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} \big(\mathbf{V}_S^{\otimes 2}\big)^{-1/2} \sum_{i=1}^n \mathrm{vec}\left(\varphi_f\left(\frac{d_i}{s_\gamma}\right)\frac{d_i}{s_\gamma}\,\mathbf{U}_i\mathbf{U}_i' - \mathbf{I}_k \right), \tag{B.4}$$

where $d_i = d_{i;\boldsymbol{\theta},\mathbf{V}_S}$ and $\mathbf{U}_i = \mathbf{U}_{i;\boldsymbol{\theta},\mathbf{V}_S}$. Using classical techniques in ULAN experiments then allows to show the asymptotic linearity result stating that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},s_\gamma,\mathbf{V}_S,f}$,

$$\begin{aligned}
&\mathbf{T}^{(n)}(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{\tau}, s_\gamma + n^{1/2}\zeta) - \mathbf{T}^{(n)}(\boldsymbol{\theta}, s_\gamma) \\
&\qquad = \mathrm{E}_{\boldsymbol{\theta},s_\gamma,\mathbf{V}_S,f}\left[\mathbf{T}^{(n)}(\boldsymbol{\theta}, s_\gamma) \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\theta},s_\gamma,\mathbf{V}_S,f;1} \\ \boldsymbol{\Delta}_{\boldsymbol{\theta},s_\gamma,\mathbf{V}_S,f;3} \end{pmatrix}' \right] \begin{pmatrix} \boldsymbol{\tau} \\ \zeta \end{pmatrix} + o_{\mathrm{P}}(1),
\end{aligned} \tag{B.5}$$

as $n \to \infty$; see Van der Vaart (2000), Proposition A.10, for a classical reference, or Hallin et al. (2013) for a most recent one. Applying this to

$$\mathbf{T}^{(n)}(\boldsymbol{\theta}, s_\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^n d^2_{i;\boldsymbol{\theta},\mathbf{V}_S} \mathbb{I}[d_{i;\boldsymbol{\theta},\mathbf{V}_S} \leq s_\gamma] \mathrm{vec}\Big(\mathbf{U}_{i;\boldsymbol{\theta},\mathbf{V}_S}\mathbf{U}'_{i;\boldsymbol{\theta},\mathbf{V}_S} - \frac{1}{k}\,\mathbf{I}_k \Big),$$

we readily obtain that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},s_\gamma,\mathbf{V}_S,f}$,

$$\mathbf{T}^{(n)}(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{\tau}, s_\gamma + n^{1/2}\zeta) - \mathbf{T}^{(n)}(\boldsymbol{\theta}, s_\gamma) = o_{\mathrm{P}}(1),$$

as $n \to \infty$, since the expectation in (B.5) is then equal to zero. Therefore, Lemma 4.4 from Kreiss (1987) entails that

$$\mathbf{T}^{(n)}(\hat{\boldsymbol{\theta}}_{\#}, \hat{s}_{\gamma\#}) - \mathbf{T}^{(n)}(\boldsymbol{\theta}, s_\gamma) = o_{\mathrm{P}}(1), \tag{B.6}$$

still as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V}_S,s_\gamma,f}$. The result then readily follows from (B.6) and Theorem 4.1. □

We can now establish consistency of the estimator in (16) under the null $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$.

Proof of Theorem 5.2. Note first that $\hat{\boldsymbol{\theta}}_\gamma$ is root-n consistent for $\boldsymbol{\theta}$ (see Cator & Lopuhaä (2010)) and that, from Lemma B.3, $\hat{s}^0_\gamma$ is root-n consistent for s_γ under the null $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$. As mentioned in Section 5, the discretization of $\hat{\boldsymbol{\theta}}_\gamma$ and $\hat{s}^0_\gamma$ into $\hat{\boldsymbol{\theta}}_{\gamma\#}$ and $\hat{s}^0_{\gamma\#}$ does not affect root-n consistency, and we may therefore apply Lemma B.6 with these discretized estimators. This yields

$$\begin{aligned}
n\big\|\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)\|^2 \;&=\; \frac{1}{\sigma_S^2\beta_\gamma} \big(\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)\big)' \Big[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S^0)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S^0})' \Big] \\
&\quad \times \big((\mathbf{V}_S^0)^{\otimes 2}\big)^{1/2} \sum_{i=1}^n d^2_{\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} \mathbb{I}[d_{\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} \leq \hat{s}^0_{\gamma\#}] \mathrm{vec}\Big(\mathbf{U}_{i;\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0}\mathbf{U}'_{i;\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} - \frac{1}{k}\,\mathbf{I}_k \Big) + o_{\mathrm{P}}(1),
\end{aligned}$$

as $n \to \infty$ under the null. Since $n\|\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)\|^2$ is $O_{\mathrm{P}}(1)$ under the null but not $o_{\mathrm{P}}(1)$, we have that

$$\rho^{(n)} = \frac{1}{n\|\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)\|^2} \big(\mathrm{vec}(\hat{\mathbf{V}}_{S,\gamma} - \mathbf{V}_S^0)\big)' \Big[\mathbf{I}_{k^2} - (\mathrm{vec}\,\mathbf{V}_S^0)(\mathrm{vec}\,\mathbf{D}_S^{\mathbf{V}_S^0})'\Big]$$
$$\times \big((\mathbf{V}_S^0)^{\otimes 2}\big)^{1/2} \sum_{i=1}^n d^2_{\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} \mathbb{I}[d_{\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} \le \hat{s}^0_{\gamma\#}] \mathrm{vec}\Big(\mathbf{U}_{i;\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} \mathbf{U}'_{i;\hat{\boldsymbol{\theta}}_{\gamma\#},\mathbf{V}_S^0} - \frac{1}{k}\mathbf{I}_k\Big)$$

is a consistent estimator of $\sigma_S^2 \beta_\gamma$ under the null. Since Lemma B.5 ensures that, still under the null, $\frac{1}{n}\sum_i \hat{d}^\ell_{i;\hat{\boldsymbol{\theta}}_\gamma,\mathbf{V}_S^0} \mathbb{I}[\hat{d}_{i;\hat{\boldsymbol{\theta}}_\gamma,\mathbf{V}_S^0} \le \hat{s}^0_\gamma]$ consistently estimates $\mathrm{E}[d^4_{\boldsymbol{\theta},\mathbf{V}_S} \mathbb{I}[d_{\boldsymbol{\theta},\mathbf{V}_S} \le s_\gamma]]$, the result then follows from the continuous mapping theorem. □

References

Agullò, J., Croux, C., & Van Aelst, S. (2008). The multivariate least-trimmed squares estimator. *J. Multivariate Anal.*, *99*, 311–338.

Butler, R., Davies, P., & Jhun, M. (1993). Asymptotics for the minimum covariance determinant estimator. *Ann. Statist.*, *3*, 1385–1400.

Cator, E. A., & Lopuhaä, H. P. (2010). Asymptotic expansion of the minimum covariance determinant estimators. *J. Multivariate Anal.*, *101*, 2372–2388.

Cator, E. A., & Lopuhaä, H. P. (2012). Central limit theorem and influence function for the MCD estimators at general multivariate distributions. *Bernoulli*, *18*, 520–551.

Cox, D. R., & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. R. Stat. Soc. Ser. B*, *49*, 1–39.

Croux, C., & Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *J. Multivariate Anal.*, *71*, 161–190.

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

Dümbgen, L., & Tyler, D. E. (2005). On the breakdown properties of some multivariate M-functionals. *Scand. J. Statist.*, *32*, 247–264.

Frahm, G. (2009). Asymptotic distributions of robust shape matrices and scales. *J. Multivariate Anal.*, *100*, 1329–1337.

Hallin, M., van den Akker, R., & Werker, B. J. (2013). On quadratric expansions of log-likelihoods and a general asymptotic linearity result. *ECARES working paper 2013-34*, .

Hallin, M., Oja, H., & Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal *R*-estimation of shape. *Ann. Statist.*, *34*, 2757–2789.

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

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

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

Ilmonen, P., & Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Annals of Statistics*, *39*, 2448–2476.

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

Kreiss, J. (1987). On adaptive estimation in stationary arma processes. *Ann. Statist.*, *15*, 112–133.

Lopuhaä, H. P., & Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.*, *19*, 229–248.

Muirhead, R., & Waternaux, C. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, *67*, 31–43.

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

Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.*, *95*, 1263–1268.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, & W. Wertz (Eds.), *Mathematical Statistics and Applications* (pp. 283–297). Dordrecht: Reidel volume B.

Rousseeuw, P. J., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*, 212–223.

Silverman, B. (1986). *Density estimation*. London: Chapman and Hall.

Taskinen, S., Croux, C., Kankainen, A., Ollila, E., & Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *J. Multivariate Anal.*, *97*, 359–384.

Taskinen, S., Sirkiä, S., & Oja, H. (2010). k-step shape estimators based on spatial signs and ranks. *J. Statist. Plann. Inference*, *140*, 3376–3388.

Tatsuoka, K. S., & Tyler, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *Ann. Statist.*, *28*, 1219–1243.

Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika*, *69*, 429–436.

Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, *70*, 411–420.

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

Van der Vaart, A. (2000). *Statistical Estimation in Large Parameter Spaces*. CWI tract 44. Amsterdam: CWI.

Visuri, S., Ollila, E., Koivunen, V., Möttönen, J., & Oja, H. (2003). Affine equivariant multivariate rank methods. *J. Statist. Plann. Inference*, *114*, 161–185. C.R. Rao 80th Birthday Felicitation Volume.