Inference on the Shape of Elliptical Distributions Based on the MCD

Davy PAINDAVEINE
Germain VAN BEVER
Inference on the Shape of Elliptical Distributions based on the MCD

Davy Paindaveine\textsuperscript{a,1}, Germain Van Bever\textsuperscript{b}

\textsuperscript{a}Université libre de Bruxelles, ECARES and Département de Mathématique, Avenue F. D. Roosevelt, 50, CP 114/04, B-1050 Brussels, Belgium

\textsuperscript{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. 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. Our results are illustrated through a Monte-Carlo study.

\textit{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 \( \hat{\theta}_\gamma \) and scatter \( \hat{\Sigma}_\gamma \), for any \( \gamma \in (0,1] \), are defined as the sample average and covariance matrix computed from “the”\textsuperscript{2} subsample leading to a covariance matrix with smallest determinant over the collection of all possible subsamples of size larger than or equal to \( \lfloor n\gamma \rfloor \) (it was shown in Cator & Lopuhaä (2012) that this optimal subsample is always of size \( \lfloor n\gamma \rfloor \)).

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 \( \gamma \), MCD estimators indeed show the highest breakdown points that can be achieved in the class of affine-equivariant estimators; see Lopuhaä & Rousseeuw (1991) and Agulló et al. (2008). Another

\textsuperscript{1}Corresponding author.

\textsuperscript{2}Uniqueness is actually not guaranteed.
advantage 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 \( \hat{\theta}_y \) and \( \hat{\Sigma}_y \), as well as asymptotic normality (at the standard root-\( n \) rate) of \( \hat{\Sigma}_y \). Croux & Haesbroeck (1999) computed the influence function of \( \hat{\Sigma}_y \), and, assuming the validity of the usual von Mises expansion linking estimators and their influence functions, deduced the asymptotic covariance matrix of \( \sqrt{n} \hat{\Sigma}_y \) 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 asymptotic normality for \( \hat{\Sigma}_y \) (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 \( \hat{\Sigma}_y \), 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 \( \Sigma \) into \( \sigma^2 V \), where \( \sigma^2 = (\det \Sigma)^{1/k} \) is a scale parameter and \( V = \Sigma / (\det \Sigma)^{1/k} \) is a shape parameter, it is often so that the parameter of interest is \( V \) (with dimension \( K := k(k+1)/2 - 1 \)), while \( \sigma^2 \) plays the role of a nuisance. In principal component analysis, for instance, principal directions may be interchangeably computed from \( \Sigma \) or from \( V \), 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 \( \sigma^2 = (\text{tr} \Sigma)/k \) 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. 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 \( \hat{\Sigma}_y \). 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. A Monte-Carlo study is conducted in Section 7 in order to confirm our asymptotic results. Finally, the Appendix collects technical proofs.

2. Elliptical densities and MCD

Let \( S_k \) be the collection of \( k \times k \) symmetric positive definite matrices, and \( \mathcal{F} := \{ f : \mathbb{R}^+ \to \mathbb{R}^+ \text{ with } \mu_{k-1,f} < \infty \} \), where \( \mu_{k-1,f} = \int_0^\infty r^{k-1} f(r) \, dr \). The random \( k \)-vector \( \mathbf{X} \) is said to be elliptically symmetric with location \( \mathbf{\theta} (\in \mathbb{R}^k) \), scatter \( \mathbf{\Sigma} (\in S_k) \), and radial density \( f \in \mathcal{F} \) (this will be denoted as \( \mathbf{X} \sim \text{Ell}_k(\mathbf{\theta}, \mathbf{\Sigma}, f) \)) if it is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^k \), with density

\[
 f_{\mathbf{X}} : \mathbb{R}^k \to \mathbb{R},
 \mathbf{x} \mapsto (\mu_{k-1,f} \omega_{k-1})^{-1} f \left( \sqrt{(\mathbf{x} - \mathbf{\theta}) \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{\theta})} \right),
\]

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

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

If \( \mathbf{X} \) has finite second-order moments (equivalently, if \( \mu_{k+1,f} < \infty \)), the covariance matrix of \( \mathbf{X} \) is proportional to \( \mathbf{\Sigma} \). 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) \( f(r) = f^\nu_0(r) := (1 + a_k r^2 / \nu)^{-(k+\nu)/2} \), and the power-exponential distributions, with radial densities of the form \( f(r) = f^\eta_0(r) := \exp(-b_k r^{2\eta}) \), \( \eta > 0 \) (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 \( \text{Ell}_k(\mathbf{\theta}, \mathbf{\Sigma}, f) \) involving a monotone decreasing \( f \).

**Assumption (B).** The observations \( \mathbf{X}_i, i = 1, \ldots, n \) are i.i.d. with a common distribution \( \text{Ell}_k(\mathbf{\theta}, \mathbf{\Sigma}, f) \) 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 continuous (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 \( r = r_k, (f)^b \) et \( \alpha \), then \( \alpha \) is the corresponding shape matrix.

\[
\begin{align*}
\Pi_{\gamma, \theta, \Sigma} & := d_{\theta, \Sigma}[d_{\theta, \Sigma} \leq r_{\gamma}],
D^{(t)}_{\gamma} := D^{(t)}_{k, \gamma}(f) := E[\Pi_{\gamma, \theta, \Sigma}^{(t)}] \int_0^{r_{\gamma}} r^t \tilde{f}_k(r) \, dr,
\alpha_{\gamma} := \alpha_{k, \gamma}(f) := \sqrt{\frac{D^{(2)}_{\gamma}}{k \gamma}},
\end{align*}
\]

and

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

where the last equality follows by integrating by parts. Note that, under Assumption (A), \( \beta_{\gamma} \) is positive and increases monotonically in \( \gamma \).

Under ellipticity, the MCD estimator of scatter \( \hat{\Sigma}_n \) is not consistent for \( \Sigma \), but rather for \( \alpha_{\gamma}^2 \Sigma \); see Proposition 2.1 below. Our derivations will rely on the following Bahadur representation result for \( \hat{\Sigma}_n \) which follows directly from Corollary 4.1 of Cator & Lopuhaä (2010) by using the affine-equivariance of \( \hat{\Sigma} \) 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

\[
\sqrt{n} (\hat{\Sigma}_n - \alpha_{\gamma}^2 \Sigma) = \frac{\alpha_{\gamma}^2}{\beta_{\gamma} \sqrt{n}} \Sigma^{1/2} \sum_{i=1}^n (\Pi_{i; \gamma, \theta, \Sigma} - k \gamma \alpha_{\gamma}^2 \Sigma) \left( U_{i; \theta, \Sigma} U^\prime_{i; \theta, \Sigma} - \frac{1}{k} I_k \right) \Sigma^{1/2}
+ \frac{1}{k \gamma \sqrt{n}} \sum_{i=1}^n (\Pi_{i; \gamma, \theta, \Sigma} - k \gamma \alpha_{\gamma}^2 \Sigma) - \frac{r_{\gamma}^2}{k \gamma \sqrt{n}} \sum_{i=1}^n (\Pi_{i; \gamma, \theta, \Sigma} - \gamma) \Sigma + o_p(1),
\]

as \( n \to \infty \), where \( I_k \) 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 are 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 \( \Sigma \) up to a positive scalar factor only. In other words, the parameter of interest, in such problems, is the corresponding \( S \)-shape matrix

\[
V_S := \Sigma / S(\Sigma)
\]
(while the scale parameter $\sigma^2_S := S(\Sigma)$ plays the role of a nuisance), where the scale functional $S : S_k \to \mathbb{R}_0^+$ (i) is homogeneous (for all $\lambda > 0$, $S(\lambda \Sigma) = \lambda S(\Sigma)$), (ii) is differentiable, with $\frac{\partial S}{\partial \Sigma_{ii}}(\Sigma) \neq 0$ for all $\Sigma \in S_k$, and (iii) satisfies $S(I_k) = 1$; see Paindaveine (2008) for comments on the requirements (i)-(iii). The collection of $k \times k$ $S$-shape matrices will be denoted by $V_k^S$.

Classical scale functionals include

(a) $S(\Sigma) = \Sigma_{11}$ (Randles (2000) and Hettmansperger & Randles (2002)),

(b) $S(\Sigma) = (\text{tr} \Sigma)/k$ (Tyler (1987), Dümbgen (1998), Visuri et al. (2003), and Taskinen et al. (2010)),

(c) $S(\Sigma) = |\Sigma|^{1/k}$ (Tatsuoka & Tyler (2000), Dümbgen & Tyler (2005), and Taskinen et al. (2006)), and

(d) $S(\Sigma) = k/(\text{tr} \Sigma^{-1})$ (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 $V_S$ and scale $\sigma^2_S$. 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 $\text{vech} \mathbf{A}$ denote the $k^2$-dimensional vector resulting from stacking the columns of $\mathbf{A}$ on top of each other. Write $\text{vech} \mathbf{A}$ for the $(K+1)$-vector (recall that $K = (k+1)/2 - 1$) obtained by stacking the upper-triangular elements of $\mathbf{A}$; $\text{vech} \mathbf{A}$ will denote the $K$-vector obtained by depriving $\text{vech} \mathbf{A}$ of its first component. Write $\mathbf{A} \otimes \mathbf{2}$ for the Kronecker product $\mathbf{A} \otimes \mathbf{2}$. Denoting by $e_k$ the $k$th vector of the canonical basis of $\mathbb{R}^k$, let $\mathbf{K}_k := \sum_{i,j=1}^k(e_i e'_j) \otimes (e_j e'_i)$ be the $k^2 \times k^2$ commutation matrix, and put $\mathbf{J}_k := \sum_{i,j=1}^k(e_i e'_j) \otimes (e_i e'_j) = (\text{vec} \mathbf{I}_k)(\text{vec} \mathbf{I}_k)'$. Finally, define $\mathbf{N}_k$ as the $k^2 \times K$ matrix such that $\mathbf{N}_k(\text{vec} \mathbf{A}) = (\text{vech} \mathbf{A})$ for any $k \times k$ matrix $\mathbf{A}$.

The algebra of $S$-shape matrices then requires introducing the following quantities. For any $\Sigma \in S_k$ and any $S$ as above, let $\mathbf{D}_S^\Sigma := (\mathbf{C}_S^\Sigma + (\mathbf{C}_S^\Sigma)' )/2$, where $\mathbf{C}_S^\Sigma := \mathbf{C}_{S,k}^\Sigma$ is the upper-triangular $k \times k$ matrix such that $\text{vech} \mathbf{C}_S^\Sigma = \nabla S(\text{vech} \Sigma)$ (whenever we consider gradients of $S$, we look at $\Sigma \mapsto S(\Sigma)$ as a function of $\text{vech} \Sigma$). Define $\mathbf{M}_S^\Sigma := \mathbf{M}_{S,k}^\Sigma$ as the $K \times k^2$ matrix such that $(\mathbf{M}_S^\Sigma)'(\text{vech} \mathbf{v}) = \mathbf{v}$ for any symmetric $k \times k$ matrix $\mathbf{v}$ satisfying $(\nabla (\text{vech} \Sigma))'(\text{vech} \mathbf{v}) = 0$ (equivalently, $(\text{vec} \mathbf{D}_S^\Sigma)'(\text{vec} \mathbf{v}) = 0$, or $\text{tr} [\mathbf{D}_S^\Sigma \mathbf{v}] = 0$). Finally, for any $\Sigma$ and $\mathbf{V} \in V_k^S$, define $\mathbf{e}_k^\mathbf{V} := \text{tr}([\mathbf{D}_S^\Sigma \mathbf{V}]^2)$. For $\Sigma = \Sigma_{11}$, $S(\Sigma) = (\text{tr} \Sigma)/k$, $S(\Sigma) = |\Sigma|^{1/k}$, and $S(\Sigma) = k/(\text{tr} \Sigma^{-1})$, one has $\mathbf{D}_S(\Sigma) = e_1 e'_1$ (with $(1,0, \ldots ,0)' \in \mathbb{R}^k$), $\mathbf{D}_S^2 = 1/k \mathbf{I}_k$, $\mathbf{D}_S^3 = 1/k |\Sigma|^{1/2} \Sigma^{-1}$, and $\mathbf{D}_S^{1/2} = k\Sigma^{-1/2}/(\text{tr} \Sigma^{-1})^{1/2}$. Hence $\mathbf{e}_k^\mathbf{V} = 1$, $\mathbf{e}_k^\mathbf{V} = 1/k \text{tr} [\mathbf{V}^2]$, $\mathbf{e}_k^\mathbf{V} = 1/k$, and $\mathbf{e}_k^\mathbf{V} = 1/k \text{tr} [\mathbf{V}^2]$, 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$_S$ estimator of scatter as $\hat{\Sigma}_S$, the corresponding MCD estimator for $S$-shape is naturally defined as $\hat{V}_{S,\gamma} := \hat{\Sigma}_S/S(\hat{\Sigma}_S)$. The affine-equivariance of $\hat{\Sigma}_S$ implies that, for
any $k \times k$ invertible matrix $A$ and any $k$-vector $b$,
\[
\hat{V}_{S,\gamma}(AX_1 + b, \ldots, AX_n + b) = \frac{A\hat{V}_{S,\gamma}(X_1, \ldots, X_n)A'}{S(A\hat{V}_{S,\gamma}(X_1, \ldots, X_n)A')},
\]
which is the natural affine-equivariance property for $S$-shape matrices.

We are primarily interested in the asymptotic properties of $\hat{V}_{S,\gamma}$. 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 $\hat{V}_{S,\gamma}$, we let
\[
c_{k,\gamma} := \frac{k(k+2)\beta_\gamma^2}{D_\gamma^{(4)}}
\]
and
\[
Q_k^{V_S} := (I_{k^2} + K_k)(\hat{V}_S^{\otimes 2}) - 2(\hat{V}_S^{\otimes 2})(\text{vec } D_S^{V_S})(\text{vec } V_S)'
- 2(\text{vec } V_S)(\text{vec } D_S^{V_S})'(\hat{V}_S^{\otimes 2}) + 2\varepsilon_k^{V_S}(\text{vec } V_S)(\text{vec } V_S)'.
\]

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

**Theorem 4.1.** Let Assumption (A) hold. Then (i) we have that
\[
\sqrt{n}\text{vec}(\hat{V}_{S,\gamma} - V_S) = \frac{1}{\beta\sqrt{n}} \left[ I_{k^2} - (\text{vec } V_S)(\text{vec } D_S^{V_S})' \right]
\times \left( \hat{V}_S^{\otimes 2} \right)^{1/2} \sum_{i=1}^{n} \hat{U}_i^{[2]} \text{vec}(U_i:\theta, V_S U_i':\theta, V_S - \frac{1}{k}I_k) + o_P(1)
\]
as $n \to \infty$; hence, (ii) $\sqrt{n}\text{vec}(\hat{V}_{S,\gamma} - V_S)$ is asymptotically normal with mean zero and covariance matrix $c_{k,\gamma}^{-1}Q_k^{V_S}$.

Building confidence zones for $\hat{V}_{S,\gamma}$ from Theorem 4.1 requires to estimate consistently the quantity $c_{k,\gamma}$ (the continuous mapping theorem indeed trivially implies that $Q_k^{V_S}$ may simply be estimated consistently by $Q_k^{V_S,\gamma}$). 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 moments, then $\mu_{k+3, f} := (\mu_{k-1, f})^{-1} \int_0^\infty r^3 \hat{f}_k(r) dr$ is finite. This implies that $r^3 \hat{f}_k(r)$ must go to zero as $\gamma \to 1$, which yields that, still as $\gamma \to 1$,
\[
c_{k,\gamma}^{-1} = \left( 1 - \frac{r^3 \hat{f}_k(r)}{(k+2)D_\gamma^{(2)}} \right)^{-2} \times \frac{kD_\gamma^{(4)}}{(k+2)(D_\gamma^{(2)})^2}
=: \left( 1 - \frac{r^3 \hat{f}_k(r)}{(k+2)D_\gamma^{(2)}} \right)^{-2} (1 + \kappa_\gamma) \to 1 + \kappa := \frac{kD_1^{(4)}}{(k+2)(D_1^{(2)})^2};
\]
the quantity $\kappa = \kappa_k(f)$ is the usual kurtosis coefficient for $k$-dimensional elliptical distributions with radial density $f$; see, e.g., Muirhead & Wateraux (1980) or Tyler (1982). The coefficient $\kappa_\gamma$
may be interpreted as a truncated elliptical kurtosis coefficient (where truncation is governed by
the population MCD, ellipsoid). Writing the asymptotic covariance matrix in terms of $\kappa$, also
makes clear 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 $\hat{V}_{S,7}$.

**Theorem 4.2.** The influence function of $\hat{V}_{S,7}$, under location $\theta$, scale $\sigma^2_S$, shape $V_S$, and radial density $f$, is given by

$$\mathbf{x} \mapsto \text{IF} (\mathbf{x}, \hat{V}_{S,7}; \theta, \sigma^2_S, V_S, f) := \frac{1}{\beta_n \sigma^2_S} d_{\theta V_S}^2 \mathbb{I}[d_{\theta V_S} \leq r_7]$$

$$\times V_S^{1/2} \left( u_{\theta V_S} u_{\theta V_S}^T V_S + [u_{\theta V_S} V_S^{1/2} D_S V_S^{1/2} u_{\theta V_S}] I_k \right) V_S^{1/2},$$

where $d_{\theta V_S} := ((\mathbf{x} - \theta) V_S^{-1} (\mathbf{x} - \theta))^{1/2}$ and $u_{\theta V_S} := V_S^{-1/2} (\mathbf{x} - \theta)/d_{\theta V_S}$.

As expected, the support of the influence function of $\hat{V}_{S,7}$ is the hyper-ellipsoid $\{ \mathbf{x} \in \mathbb{R}^k : d_{\theta V_S} \leq r_7 \}$, hence coincides with the support of the influence function of $\hat{\Sigma}_{S,7}$; see Croux & Haesbroeck (1999). Note also that the influence function of $\hat{V}_{S,7}$ depends on $f$ (hence, on the distribution of $d_{\theta \Sigma}$) and on $\gamma$ only through the scalar factor $1/\beta_n$, whereas the influence function of $\hat{\Sigma}_{S,7}$ depends on $\gamma$ and $f$ in a much more complicated way (implying, e.g., that the influence functions of $\hat{\Sigma}_{S,7}$ at elliptical $t$-distributions and at the multinormal are not proportional to each other). Of course, the smaller $\gamma$, the smaller the support of $\hat{V}_{S,7}$’s influence function, but also the larger the influence function itself within this support (recall that $\beta_n$ is monotonically increasing in $\gamma$).

As an illustration, Figure 1 plots, for $S(\Sigma) = (\det \Sigma)^{1/k}$, the influence functions of $(\hat{V}_{S,7})_{22}$ (first column) and $(\hat{V}_{S,7})_{12}$ (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 $(\hat{V}_{S,7})_{12}$ does not depend on the scale functional $S$, and that, in the spherical setup considered, the scale functionals $S(\Sigma) = (\det \Sigma)^{1/k}$, $S(\Sigma) = (\text{tr} \Sigma)/k$, and $S(\Sigma) = k/(\text{tr} \Sigma^{-1})$ lead to the same influence function for $(\hat{V}_{S,7})_{22}$, and that the influence function of $(\hat{V}_{S,7})_{22}$ for $S(\Sigma) = \Sigma_{11}$ is equal to twice the common influence function obtained for the three other scale functionals.

4.2. **MCD-test for shape**

In this section, we construct a Wald-type test, based on the MCD shape estimator $\hat{V}_{S,7}$ above for the problem

$$\begin{cases}
H_0 : V_S = V_S^0 \\
H_1 : V_S \neq V_S^0,
\end{cases}$$

(7)

where $V_S^0 \in V_S$ is fixed. The important case for which $V_S^0 = I_k$ 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 $\sqrt{n} \vec{V}_{S,7} - V_S$ is not invertible. This non-invertibility is explained by the fact that only $K$ of the $k^2$ entries of $\vec{V}(V_S)$ are functionally independent (which follows from symmetry of $V_S$ and the normalization constraint $S(V_S) = 1$).
Figure 1: Plots of the Influence functions, for the scale functional \( S(\Sigma) = (\det \Sigma)^{1/k} \), of \((\hat{V}_{S,\gamma})_{22}\) (first column) and \((\hat{V}_{S,\gamma})_{12}\) (second column) at the bivariate standard normal distribution. The first row (resp., second row) corresponds to \( \gamma = 0.5 \) (resp., \( \gamma = 0.75 \)).
To solve this issue, one can rather base a Wald test on the random $K$-vector $\sqrt{n} \text{vech}(\hat{V}_{S,\gamma} - V_S)$, which, in view of Theorem 4.1(ii), is asymptotically normal with mean zero and covariance matrix $c_{k,\gamma}^{-1} \mathbf{N}_k Q_k^S \mathbf{N}_k'^t$. 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

$$\hat{Q}_{S,\gamma} = n \hat{c}_{k,\gamma} \left[ \text{vech}(\hat{V}_{S,\gamma} - V_S^0)'(N_k Q_k^S \mathbf{N}_k)'^{-1} \text{vech}(\hat{V}_{S,\gamma} - V_S^0) \right],$$

(8)

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

We actually propose rather using the simpler test statistic

$$Q_{S,\gamma} = \frac{n \hat{c}_{k,\gamma}}{2} \left( \text{tr}\left[ ((V_S^0)^{-1}\hat{V}_{S,\gamma})^2 \right] - \frac{1}{k} \text{tr}^2 \left[ (V_S^0)^{-1}\hat{V}_{S,\gamma} \right] \right),$$

(9)

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

**Lemma 4.1.** The matrix $N_k Q_k^S N_k'^t$ has full rank $K$, and its inverse is given by $(N_k Q_k^S N_k'^t)^{-1} = \frac{1}{4} M_k^S (V_S^{(2)})^{-1/2} [I_k^2 + K_k - \frac{2}{k} J_k] (V_S^{(2)})^{-1/2} (M_k^S)'$.

**Theorem 4.3.** Let Assumption (A) hold. Then, (i) under $H_0 : V_S = V_S^0$, $Q_{S,\gamma} = \hat{Q}_{S,\gamma} + o_P(1)$, as $n \to \infty$; (ii) under $H_0 : V_S = V_S^0$, $Q_{S,\gamma}$ is asymptotically $\chi^2_k$; (ii) under sequences of local alternatives $H_1^{(n)} : V_S^{(n)} = V_S^0 + n^{-1/2} \mathbf{v}$, with $\text{tr}[D_S^0 \mathbf{v}] = 0$, $Q_{S,\gamma}$ is asymptotically non-central $\chi^2_k$, with non-centrality parameter

$$\frac{c_{k,\gamma}}{2} \left( \text{tr}[((V_S^0)^{-1}\mathbf{v})^2] - \frac{1}{k} \text{tr}^2 [(V_S^0)^{-1}\mathbf{v}] \right),$$

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

The condition $\text{tr}[D_S^0 \mathbf{v}] = 0$ in the local alternatives $V_S^{(n)} = V_S^0 + n^{-1/2} \mathbf{v}$ above ensures that, at the first order as $n \to \infty$, $S(V_S^{(n)}) = 1$, hence that $V_S^{(n)}$ 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 $V_S$ 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 $V_S^{(n)} = V_S^0 + n^{-1/2} \mathbf{v}$ indeed remains an $S$-shape matrix (for large $n$). The intuition is similar for “non-linear” scale functionals (such as the determinant-based one), where the constraint $S(V_S^{(n)}) = 1$, however, can only be achieved at the first order.
The null hypothesis \( H_0 : V_S = V_S^0 \) is not invariant under the group of affine transformations, but it is invariant under the subgroup of affine transformations of the form
\[
(X_1, \ldots, X_n) \mapsto ((V_S^0)^{1/2}O(V_S^0)^{-1/2}X_1 + b, \ldots, (V_S^0)^{1/2}O(V_S^0)^{-1/2}X_n + b),
\]
where \( O \) is an arbitrary orthogonal \( k \times k \) matrix and \( b \) 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, building confidence zones for \( V_{S,\gamma} \) on the basis of Theorem 4.1 or implementing the corresponding test for \( H_0 : V_S = V_S^0 \) requires consistent estimators of the quantity \( c_{k,\gamma} \) in (5). Such estimators may be obtained as follows.

Unlike \( \tilde{f}_k \), the mapping \( \tilde{f}_{k,\text{shape}} \), defined through \( r \mapsto \tilde{f}_{k,\text{shape}}(r) = \sigma^{-1}\tilde{f}_k(r/\sigma) \), does not depend on \( \sigma \), which follows from the fact that \( \tilde{f}_k \) (resp., \( \tilde{f}_{k,\text{shape}} \)) is the pdf of \( d_{\theta,\Sigma} \) (resp., \( d_{\theta,V_S} \)). Similarly, \( s_{\gamma} := r_\gamma \sigma \), unlike \( r_\gamma \), does not depend on \( \sigma \), since \( s_{\gamma} \) (resp., \( r_{\gamma} \)) is the order-\( \gamma \) quantile of \( d_{\theta,V_S} \) (resp., \( d_{\theta,\Sigma} \)). Consequently, the quantity \( c_{k,\gamma} \), that, by using the identity
\[
D_{\gamma}^{(1)} = E[d_{\theta,\Sigma}^{\ell} I[d_{\theta,\Sigma} \leq r_{\gamma}]] = \sigma^{-d} E[d_{\theta,V}^{\ell} I[d_{\theta,V} \leq s_{\gamma}]],
\]
rewrites
\[
c_{k,\gamma} = \frac{k(k+2)\beta_{\gamma}^2}{D_{\gamma}^{(4)}} = \frac{((k+2)D_{\gamma}^{(2)} - r_{\gamma}^3\tilde{f}_k(r_{\gamma}))^2}{k(k+2)D_{\gamma}^{(4)}}
= \frac{((k+2)E[d_{\theta,V}^{\ell} I[d_{\theta,V} \leq s_{\gamma}]] - s_{\gamma}^3\tilde{f}_{k,\text{shape}}(s_{\gamma}))^2}{k(k+2)E[d_{\theta,V}^{\ell} I[d_{\theta,V} \leq s_{\gamma}]]},
\]
(10) does not depend on \( \sigma \), hence may be estimated without estimating that parameter. Since the MCD_{\gamma}-estimators of location and S-shape \( \hat{\theta}_\cdot, \) and \( \hat{V}_{S,\gamma} \) are consistent for \( \theta \) and \( V_S \), respectively, (10) leads to the estimator
\[
\hat{c}_{k,\gamma} := \frac{((k+2)\frac{1}{n} \sum_{i=1}^{n} d_{\hat{\theta},\gamma}^{\ell} I[d_{\hat{\theta},\gamma} \leq \hat{s}_{\gamma}]) - \hat{s}_{\gamma}^3\hat{f}_{k,\text{shape}}(\hat{s}_{\gamma})}{k(k+2)\frac{1}{n} \sum_{i=1}^{n} d_{\hat{\theta},\gamma}^{\ell} I[d_{\hat{\theta},\gamma} \leq \hat{s}_{\gamma}]},
\]
(11)
where \( \hat{s}_{\gamma} \), quite naturally, is taken as the order-\( \gamma \) quantile of the \( d_{\hat{\theta},\gamma} \)'s, and where
\[
\hat{f}_{k,\text{shape}}(\hat{s}_{\gamma}) := \frac{1}{nh_n} \sum_{i=1}^{n} K \left( \frac{\hat{s}_{\gamma} - d_{\hat{\theta},\gamma}}{h_n} \right)
\]
is a kernel-based estimator for \( \tilde{f}_{k,\text{shape}}(s_{\gamma}) \).

The estimators \( \hat{c}_{k,\gamma} \) are consistent, but the convergence may be slow, especially for small values of \( \gamma \). To illustrate this, we generated \( M = 5,000 \) independent random samples of sizes \( n = 50, 400, \) and \( 10,000 \) from the bivariate standard normal distribution, and evaluated in each sample the estimators \( \hat{c}_{k,\gamma}, \gamma = 0.5, 0.6, 0.7, 0.8, 0.9 \) as well as the estimator \( \hat{c}_k \) of corresponding covariance-based quantity (see Section 6). Boxplots of the resulting estimates are reported in Figure 2, that indicates that, for small \( \gamma \)-values, the estimators \( \hat{c}_{k,\gamma} \) are severely biased, even for \( n = 400 \). Consistency, of course, ensures that this improves for much larger sample sizes (\( n = 10,000 \)). In contrast, the estimators \( \hat{c}_{k,\gamma} \) behave much better for large \( \gamma \)-values, even for moderate sample sizes.
Figure 2: Boxplots of the estimators $\hat{c}_{k,\gamma}$ in (11), for $\gamma = 0.5, 0.6, 0.7, 0.8, 0.9$, and of the corresponding covariance-based estimators $\hat{c}_k$ (see Section 6), computed from 5,000 independent bivariate standard normal samples of size $n = 50, 400$ and $10,000$. The corresponding population quantities ($c_{k,\gamma}$ or, in the lower right panel, $c_k$) are throughout reported in orange.
6. Covariance-based procedures and AREs

The goal of this section is to derive the asymptotic relative efficiencies of the MCD_, procedures of Section 4 with respect to their competitors based on the empirical covariance matrix \( \hat{\Sigma} := \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})' \). Although \( \Sigma = \Sigma_\gamma \) for \( \gamma = 1 \), the asymptotic properties of these covariance-based procedures cannot be obtained by taking \( \gamma = 1 \) in Theorem 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 \( \Sigma \), however, can be obtained quite trivially. Of course, unlike for the MCD, scatter estimator, finite fourth-order moments here are needed.

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

\[
\sqrt{n} \text{vec} \left( \hat{\Sigma} - \frac{D^{(2)}}{k} \Sigma \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Sigma_{i,\theta} \left( \mathbf{U}_{i,\theta,\Sigma} \mathbf{U}_{i,\theta,\Sigma} - \frac{1}{k} \mathbf{I}_k \right) \Sigma_{i,\theta}^{1/2} + \frac{1}{k} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (d_{i,\theta,\Sigma} - D^{(2)})^{1/2} + o_p(1),
\]

as \( n \to \infty \), where \( D^{(2)} = D_1^{(2)} = \int_0^{\infty} r^2 f_k(r) \, dr \).

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 \( \hat{V}_S = \hat{\Sigma}/S(\hat{\Sigma}) \).

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

\[
\sqrt{n} \text{vec}(\hat{V}_S - V_S) = -\frac{k}{D^{(2)}} \frac{1}{\sqrt{n}} \left[ \mathbf{I}_{k^2} - (\text{vec} \, V_S)(\text{vec} \, D_S^{Y_S})' \right] \cdot (V_S^{\otimes 2})^{1/2} \sum_{i=1}^{n} d_{i,\theta,\Sigma} \text{vec} \left( \mathbf{U}_{i,\theta,\Sigma} \mathbf{U}_{i,\theta,\Sigma} - \frac{1}{k} \mathbf{I}_k \right) + o_p(1)
\]

as \( n \to \infty \); hence, (ii) \( \sqrt{n} \text{vec}(\hat{V}_S - V_S) \) is asymptotically normal with mean zero and covariance matrix \( c_k^{-1} Q_k^{Y_S} \), where \( c_k = 1/(1 + \kappa) \) involves the kurtosis coefficient defined in Page 6.

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

\[
\text{ARE}_f[\hat{V}_{S,\gamma}/\hat{V}_S] = c_{k,\gamma}/c_k.
\]

Such AREs are unambiguously defined since the asymptotic covariance matrices in Theorems 4.1 and 6.1 are of the form \( \lambda f Q_\gamma \); for a common matrix \( Q_\gamma \), 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 \( \lambda_{1,f} Q_1 + \lambda_{2,f} Q_2 \); see, e.g., Tyler (1982, 1983). Some plots of the AREs in (12) 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, $\phi_S$ say, rejects the null at asymptotic level $\alpha$ whenever

$$Q_S = \frac{n \hat{c}_k}{2} \left( \text{tr} \left[ \left( (V_S^0)^{-1} V_S \right)^2 \right] - \frac{1}{k} \text{tr}^2 \left[ (V_S^0)^{-1} V_S \right] \right) > \chi^2_{K,1-\alpha},$$

where $\hat{c}_k := 1/(1+\hat{\kappa})$ involves a consistent estimator for the kurtosis coefficient $\kappa$. For instance, one may use $\hat{\kappa} := \left[ k \left( \frac{1}{n} \sum_{i=1}^n d_i^2 \langle X, \Sigma \rangle \right) \right] / \left[ (k+2) \left( \frac{1}{n} \sum_{i=1}^n d_i^2 \langle X, \Sigma \rangle \right)^2 \right] - 1$. 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 $\hat{c}_k$, 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 $H_0 : V_S = V_S^0$, $Q_S$ is asymptotically $\chi^2_K$; (ii) under sequences of local alternatives $H_i^{(n)} : V_S^{(n)} = V_S^0 + n^{-1/2}v$, with $\text{tr} [D^2_{V_S^0} v] = 0$, $Q_S$ is asymptotically non-central $\chi^2_K$, with non-centrality parameter

$$\frac{c_k}{2} \left( \text{tr} \left[ \left( (V_S^0)^{-1} v \right)^2 \right] - \frac{1}{k} \text{tr}^2 \left[ (V_S^0)^{-1} v \right] \right),$$

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, test for shape $\phi_{\gamma,S}$ with respect to its covariance-based competitor $\phi_S$ is given by

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

which coincides with the ARE obtained in (12) 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 procedures remain valid (in the sense that $V_{S,\gamma}$ remains root-$n$ consistent and that $\phi_{\gamma,S}$ still meets the asymptotic $\alpha$-level constraint).

Figure 3 provides several plots (as functions of $\gamma$ or of the number of degrees of freedom $\nu$ of the underlying standard elliptical $t_v$ distribution) of the AREs in (12)-(13), under $k$-variate standard normal and $t_v$ densities; see Figure 3 for details. It is seen that the AREs decrease with the tail weight $\nu$ : 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 $k$. 

13
Figure 3: Plots of asymptotic relative efficiencies (AREs) of MCD, shape estimators and tests with respect to their covariance-based competitors, under $k$-variate standard normal and $t_\nu$ densities.
7. A Monte-Carlo study

In this section, we illustrate the finite-sample behaviors of the MCD\(_\gamma\) shape estimators and tests 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.

We start with hypothesis testing, where we focused on the problem of testing for sphericity, i.e., on the null hypothesis \(H_0 : V_S = I_k\). Throughout, we adopted the determinant-based scale functional \(S(\Sigma) = (\det \Sigma)^{1/k}\). We generated collections of \(M = 2,000\) independent random samples of sizes \(n = 50, 400,\) and \(10,000\), from a bivariate Gaussian distribution with mean \(\theta = 0\), scale \(\sigma_S^2 = 1\), and shape matrices

\[
V_S(m; \xi) = \frac{I_2 + \frac{m}{\xi \sqrt{n}} \begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix}}{\left(\det \left[ I_2 + \frac{m}{\xi \sqrt{n}} \begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix} \right] \right)^{1/2}}, \quad m = 0, 1, 2, \ldots, 6, \tag{14}
\]

with \(\xi = 1.2\). We also generated collections of \(M = 2,000\) independent random samples with the same sizes from a bivariate \(t_5\) distribution with mean zero, \(S\)-scale one, and shape matrices \(V_S(m; \xi)\), still for \(m = 0, 1, 2, \ldots, 6\), but here with \(\xi = 1\); these heterogeneous \(\xi\)-values were chosen so that the most severe alternatives—associated with the shape matrices \(V_S(6; \xi)\)—lead to roughly similar rejection frequencies in the multinormal and \(t_5\) cases.

For each such sample, we performed, at asymptotic level \(\alpha = 5\%\), the MCD\(_\gamma\) tests of sphericity for \(\gamma = 0.5, 0.75, 0.9\) and \(0.95\), as well as their covariance-based competitor from Section 6. Figure 4 plots the corresponding rejection frequencies as functions of \(m\). This figure also reports the corresponding exact asymptotic powers, that are readily obtained from Theorems 4.3 and 6.2. For sample size \(n = 50\), we also report there, for the sake of comparison, the rejection frequencies of the “pseudo-tests” based on the true underlying value of the coefficients \(c_{k, \gamma}\) in (5) (for MCD tests) or \(c_k\) in (5) (for their covariance-based competitor). The genuine MCD\(_\gamma\) tests were based on the estimators \(\hat{c}_{k, \gamma}\) in (11); for the kernel density estimation involved, we 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 remarkably confirms our asymptotic results in Theorems 4.3 and 6.2 as the empirical rejection frequencies for \(n = 10,000\) 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\(_\gamma\) tests, for low \(\gamma\)-values, dominate the covariance-based one under \(t_5\)). For small to moderate sample sizes, low \(\gamma\)-values lead to (for \(n = 50\) and \(\gamma = 0.5\), severely) liberal tests. This is only due to the corresponding poor estimation of \(c_{k, \gamma}\) (see Section 5); note indeed that the pseudo-tests using the true value of \(c_{k, \gamma}\) have a null size that is extremely close to the nominal level \(\alpha = 5\%\).
Figure 4: Rejection frequencies (second, third and fourth columns, for \( n = 50, 400 \) and 10,000, respectively) and asymptotic powers (rightmost column) of the MCD tests of sphericity, for \( \gamma = 0.5, 0.75, 0.9, \) and 0.95, and of their covariance-based competitor, under bivariate normal and \( t_5 \) densities. Rejection frequencies, for \( n = 50, \) of the pseudo-tests using the exact values of \( c_k, \) and \( c_k^* \) are also provided (first column). We refer to Section 7 for details.
For the sake of completeness, we also performed 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) Gaussian 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 \( \mathbf{V}_S \). For each of these shape estimators \( \hat{\mathbf{V}} = (\hat{V}_{ij}) \), Figure 5 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 \( \gamma \) decreases. Under heavy tails, however, MCD\( _\gamma \) estimators, for large values of \( \gamma \), 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. Of course, in terms of robustness, MCD\( _\gamma \)-based procedures strongly dominate their covariance-based competitors.

**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.

**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 \( n \to \infty \),

\[
\sqrt{n} \text{vec}(\mathbf{V}_{S,\gamma} - \mathbf{V}_S) = \frac{1}{S(\alpha^2 \Sigma)} \left[ \mathbf{I}_{k^2} - (\text{vec} \mathbf{V}_S)(\text{vec} \mathbf{D}_S^{Y_s})' \right] \sqrt{n} \text{vec}(\Sigma - \alpha^2 \Sigma) + o_p(1).
\]

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

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

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

(ii) Since

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

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

\[
\frac{D^{(4)}_{\theta, \Sigma, \mathbf{f}}}{\beta^2} \left[ \mathbf{I}_{k^2} - (\text{vec} \mathbf{V}_S)(\text{vec} \mathbf{D}_S^{Y_s})' \right](\mathbf{V}_S^\otimes 2)^{1/2} \mathbf{A}_k (\mathbf{V}_S^\otimes 2)^{1/2} \left[ \mathbf{I}_{k^2} - (\text{vec} \mathbf{V}_S)(\text{vec} \mathbf{D}_S^{Y_s})' \right]'.
\]
Figure 5: Boxplots, obtained from 2,000 independent random multinormal 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 $\hat{V}_{S,\gamma}$, $\gamma = 0.5$, 0.75, 0.9 and 0.95, and their covariance-based competitors $\hat{V}_{S}$; see Section 7 for details.
By using (A.1), \( K_k(A \otimes B) = (A \otimes B)K_k \), and \( K_k(\text{vec } A) = \text{vec } (A^0) \), this covariance matrix rewrites

\[
\frac{D^{(i)}}{k(k + 2)^{3/2}} \left[ I_{k^2} - (\text{vec } V_S)(\text{vec } D_S^{V_S})' \right] \left( V_S^{(2)} \right)^{1/2} \left[ I_{k^2} + K_k \right] \left( V_S^{(2)} \right)^{1/2} \left[ I_{k^2} - (\text{vec } V_S)(\text{vec } D_S^{V_S})' \right]'
\]

\[
= c_{k, \gamma}^{-1} \left[ I_{k^2} - (\text{vec } V_S)(\text{vec } D_S^{V_S})' \right] \left( V_S^{(2)} \right) \left[ I_{k^2} - (\text{vec } V_S)(\text{vec } D_S^{V_S})' \right]'
\]

\[
= c_{k, \gamma}^{-1} \left[ I_{k^2} - (\text{vec } V_S)(\text{vec } D_S^{V_S})' \right] \left( I_{k^2} + K_k \right) \left( V_S^{(2)} \right) - 2 \left( V_S^{(2)} \right) \left( \text{vec } D_S^{V_S} \right)(\text{vec } V_S)' .
\]

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

**Proof of Lemma 4.1.** The result follows by noting that \( Q_k^{V_S} = Q_k^{V_S} \), where \( Q_k^{V_S} \) 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 vec(ABC) = (C' \otimes A)(vec B) and \( K_k(\text{vec } A) = \text{vec } (A') \), and Lemma 5.1 from Hallin & Paindaveine (2006a), we obtain that, under the null as \( n \to \infty \),

\[
\hat{Q}_{S, \gamma} = \frac{n}{4c_{k, \gamma}} \left[ \text{vec}(\hat{V}_{S, \gamma} - V_S^0) \right]' \left( V_S^{(2)} \right)^{-1/2} \left[ I_{k^2} + K_k \right] \left( V_S^{(2)} \right)^{-1/2} \text{vec}(\hat{V}_{S, \gamma} - V_S^0) + o_p(1)
\]

\[
= \frac{n}{2c_{k, \gamma}} \left[ \text{vec}(V_S^0)^{-1/2} \hat{V}_{S, \gamma}(V_S^0)^{-1/2} - I_k \right]' \left[ I_{k^2} + \frac{1}{k} J_k \right] \left[ \text{vec}(V_S^0)^{-1/2} \hat{V}_{S, \gamma}(V_S^0)^{-1/2} - I_k \right] + o_p(1).
\]

From the identities \( \left[ I_{k^2} - \frac{1}{k} J_k \right](\text{vec } I_k) = 0 \) and \( (\text{vec } A)'(\text{vec } B) = \text{tr}[A'B] \), we then obtain that, still under the null as \( n \to \infty \),

\[
\hat{Q}_{S, \gamma} = \frac{n}{2c_{k, \gamma}} \left( \text{tr} \left[ ((V_S^0)^{-1/2} \hat{V}_{S, \gamma}(V_S^0)^{-1/2})^2 \right] - \frac{1}{k} \text{tr}^2 \left[ (V_S^0)^{-1/2} \hat{V}_{S, \gamma}(V_S^0)^{-1/2} \right] \right) + o_p(1),
\]

which establishes the result.

(ii) This readily follows from Part (i) of the result, the consistency of \( c_{k, \gamma} \), and the fact that \( \sqrt{n} \text{vech}(\hat{V}_{S, \gamma} - V_S) \) is asymptotically normal with mean zero and (full rank \( K \); see Lemma 4.1) covariance \( c_{k, \gamma}^{-1} N_k Q_k^{V_S} N_k' \).

(iii) Under Assumption (A'), the fixed-\( f \) parametric model described by \( P_f := \{ P^{(n)}_{\theta, \sigma^2_S, \text{vech } V_S; f} \} \) (where \( P^{(n)}_{\theta, \sigma^2_S, \text{vech } V_S; f} \) denotes the probability measure of \( n \) i.i.d. \( k \)-variate elliptical observations with location \( \theta \), scale \( \sigma^2_S \), shape \( V_S \), and radial density \( f \)) is uniformly locally and asymptotically normal with a central sequence of the form \( \Delta_f = ((\Delta_f^\theta)' , \Delta_f^{\sigma^2} , (\Delta_f^{V_S})')' \), where

\[
\Delta_f^{V_S} := \frac{1}{2\sqrt{n}} M_k V_S^{(2)} \left( V_S^{(2)} \right)^{-1/2} \sum_{i=1}^{n} \text{vec} \left( \frac{d_i \theta, V_S}{\sigma_S} \varphi_f \left( \frac{d_i \theta, V_S}{\sigma_S} \right) U_{i, \theta, V_S} U_{i, \theta, V_S}' - \frac{1}{k} I_k \right);
\]
see, e.g., Paindaveine (2008). It can then be shown that, under $\theta$, $\sigma^2$, $V_S^0$, and $f$, the joint asymptotic distribution of $S^{(n)} := \sqrt{n} \, \text{vech}(\hat{V}_{S,\gamma} - V_S^0)$ and

$$T^{(n)} := \log(\text{d}P^{(n)}_{\theta, \sigma^2, \text{vech}(V_{S,\gamma}^0 + n^{1/2}v)}; f / \text{d}P^{(n)}_{\theta, \sigma^2, \text{vech}(V_S^0); f})$$

is asymptotically multinormal, with an asymptotic covariance between $S^{(n)}$ and $T^{(n)}$ that is given by $w = \lim_{n \to \infty} E_{\theta, \sigma^2, V_S^0; f} [S^{(n)}(\Delta_f^S)^\prime]^\prime (\text{vech } v)$. By first using Theorem 4.1(i) and $N_k(\text{vec } A) = (\text{vech } A)$, then by simplifying $w$ along the same lines as in the previous proofs, we obtain

$$w = \frac{1}{k(k + 2)\beta^\gamma} E_{\theta, \sigma^2, V_S^0; f} \left[ \frac{d^2_{i, \theta, \text{vech}(V_S^0)} f}{\sigma_S^2} \left[ \frac{d_{i, \theta, \text{vech}(V_S^0)}}{\sigma_S} \leq r_\gamma \right] \times \frac{d_{i, \theta, \text{vech}(V_S^0)}}{\sigma_S} \varphi(f(\frac{d_{i, \theta, \text{vech}(V_S^0)}}{\sigma_S})) \right] (\text{vech } v),$$

which, in view of (3), yields $w = (\text{vech } v)$. Le Cam’s third lemma then yields that $S^{(n)}$ is asymptotically normal, under $P^{(n)}_{\theta, \sigma^2, \text{vech}(V_S^0 + n^{1/2}v); f}$, with mean $(\text{vech } v)$ and the same covariance matrix $c^{-1}_{k, \gamma} N_k Q_k^{V_S^0} N_k^\prime$ as under the null. Hence, still under $P^{(n)}_{\theta, \sigma^2, \text{vech}(V_S^0 + n^{1/2}v); f}$

$$Q_{S, \gamma} = Q_{S, \gamma} + o_P(1) = c_{k, \gamma}(S^{(n)})^\prime (N_k Q_k^{V_S^0} N_k^\prime)^{-1} S^{(n)} + o_P(1)$$

(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 $\chi_k^2$ with non-centrality parameter

$$\hat{c}_{k, \gamma}(\text{vech } v)^\prime (N_k Q_k^{V_S^0} N_k^\prime)^{-1} (\text{vech } v),$$

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

\textbf{Proof of Proposition 6.1.} Decomposing as usual $\Sigma$ into $\Sigma_{\theta} + (\bar{X} - \theta)(\bar{X} - \theta)^\prime$, with $\Sigma_{\theta} := \frac{1}{n} \sum_{i=1}^n (X_i - \theta)(X_i - \theta)^\prime$, we obtain that, as $n \to \infty$,

$$\sqrt{n} \left( \frac{\Sigma - D^{(2)}_k}{\Sigma} \right) = \sqrt{n} \left( \Sigma_{\theta} - \frac{D^{(2)}_k}{\Sigma} \right) + o_P(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( d_{i, \theta, \Sigma}^{1/2} U_{i, \theta, \Sigma} U_{i, \theta, \Sigma}^\prime \Sigma_{i, \theta, \Sigma}^{1/2} - \frac{D^{(2)}_k}{\Sigma} \right) + o_P(1),$$

which establishes the result.

\textbf{References}


