

Parametric and semiparametric inference for shape: the role of the scale functional

Marc Hallin, Davy Paindaveine

Received: June 27, 2005; Accepted: August 13, 2006

Summary: We are considering the problem of efficient inference on the shape matrix of an elliptic distribution with unspecified location and either (a) fully specified radial density, (b) radial density specified up to a scale parameter, or (c) completely unspecified radial density. Bickel in [1] has shown that efficiencies under (b) and (c), while being strictly less than under (a), coincide: the efficiency loss caused by an unspecified radial density thus is entirely due to the non-specification of its scale (scale here is not necessarily measured by standard error, as second-order moments may be infinite). Defining scale however requires the choice of a particular *scale functional*, a choice which has an impact on efficiency bounds. We provide a closed form expression for this efficiency loss, both in hypothesis testing and in point estimation, as a function of the standardized radial density and the scale functional. We show that this loss is maximum at arbitrarily light tails whereas, under arbitrarily heavy tails, it is arbitrarily close to zero: hence, under very heavy tails, ignoring the scale (ignoring the exact density) asymptotically does not harm inference on shape. However, the same loss is nil, irrespective of the standardized radial density, when the scale functional (in dimension k) is the k -th root of the scatter determinant.

1 Introduction

1.1 Scatter, scale, and shape

Denote by $P_{\theta, \Sigma, f_1}^{(n)}$ the distribution of the n -tuple of k -dimensional observations $\mathbf{X}^{(n)} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, where, letting $d(\mathbf{x}, \theta; \Sigma) := [(\mathbf{x} - \theta)' \Sigma^{-1} (\mathbf{x} - \theta)]^{1/2}$, the \mathbf{X}_i 's are i.i.d. with common *elliptical density*

$$\mathbf{x} \mapsto c_{k, f_1} |\Sigma|^{-1/2} f_1(d(\mathbf{x}, \theta; \Sigma)). \quad (1.1)$$

The *center of symmetry* θ is a k -dimensional real vector, the *scatter matrix* Σ belongs to the collection \mathcal{S}_k of all symmetric positive definite real $k \times k$ matrices, the *standardized radial density* $f_1 : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ satisfies $\mu_{k-1, f_1} := \int_0^\infty r^{k-1} f_1(r) dr < \infty$, and c_{k, f_1} is a normalization factor. To ensure identifiability of Σ and $c_{k, f_1} \times f_1$ without imposing

any moment conditions, we require the radial density f_1 to be *standardized* in such a way that, writing $d_i(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ for $d(\mathbf{X}_i, \boldsymbol{\theta}; \boldsymbol{\Sigma})$,

$$\text{Med}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1} [d_i(\boldsymbol{\theta}, \boldsymbol{\Sigma})] = 1, \tag{1.2}$$

where $\text{Med}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1} [d_i(\boldsymbol{\theta}, \boldsymbol{\Sigma})]$ denotes the median of $d_i(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ under $\mathbb{P}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1}^{(n)}$. Note however that f_1 , strictly speaking, is not a probability density (more precisely, it is not a density with respect to the Lebesgue measure over \mathbb{R}^+), and does not even integrate to one. Actually, under $\mathbb{P}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1}^{(n)}$, $d_i(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ has density $r \mapsto \tilde{f}_{1k}(r) := (\mu_{k-1, f_1})^{-1} r^{k-1} f_1(r)$, $r \in \mathbb{R}_0^+$. The class of all standardized radial densities will be denoted as \mathcal{F}_1 :

$$\mathcal{F}_1 := \left\{ f_1 : \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+ : \int_0^1 r^{k-1} f_1(r) dr = \int_1^\infty r^{k-1} f_1(r) dr < \infty \right\}.$$

Special instances of (1.1) are the k -variate multinormal distribution, with radial density $f_1(r) = \phi_1(r) := \exp(-a_k r^2/2)$, the k -variate Student distributions, with radial densities (for $\nu \in \mathbb{R}_0^+$ degrees of freedom) $f_1(r) = f_{1,\nu}^t(r) := (1 + a_{k,\nu} r^2/\nu)^{-(k+\nu)/2}$, and the k -variate power-exponential distributions, with radial densities of the form $f_1(r) = f_{1,\eta}^e(r) := \exp(-b_{k,\eta} r^{2\eta})$, $\eta \in \mathbb{R}_0^+$; the positive constants a_k , $a_{k,\nu}$, and $b_{k,\eta}$ are such that $f_1 \in \mathcal{F}_1$.

Now, let $S : \mathcal{S}_k \rightarrow \mathbb{R}_0^+$ be a *homogeneous* function (satisfying $S(\lambda \boldsymbol{\Sigma}) = \lambda S(\boldsymbol{\Sigma})$ for all $\lambda > 0$), and define the *scale* parameter $\sigma_S^2 := S(\boldsymbol{\Sigma}) > 0$ and the *shape* parameter $\mathbf{V}_S := \boldsymbol{\Sigma}/S(\boldsymbol{\Sigma}) \in \mathcal{V}_k^S$, where \mathcal{V}_k^S denotes the collection of matrices \mathbf{V} in \mathcal{S}_k satisfying $S(\mathbf{V}) = 1$. Clearly, $d_i(\boldsymbol{\theta}, \mathbf{V}_S)$ under $\mathbb{P}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1}^{(n)}$ has median σ_S , and we therefore refer to S as a *scale functional*.

Classical choices for S include $S(\boldsymbol{\Sigma}) = \Sigma_{11}$ ([6], [5], [9], and [16]), $S(\boldsymbol{\Sigma}) = (\text{tr } \boldsymbol{\Sigma})/k$ ([4], [13], and [20]), and $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$ ([3], [17], [18], and [19]). This leads to rewriting (1.1) as

$$\mathbf{x} \mapsto \frac{c_{k, f_1}}{\sigma_S^k |\mathbf{V}_S|^{1/2}} f_1 \left(\frac{1}{\sigma_S} d(\mathbf{x}, \boldsymbol{\theta}; \mathbf{V}_S) \right) = \frac{c_{k, f}}{|\mathbf{V}_S|^{1/2}} f(d(\mathbf{x}, \boldsymbol{\theta}; \mathbf{V}_S)), \tag{1.3}$$

and $\mathbb{P}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_1}^{(n)}$ as $\mathbb{P}_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^{(n)}$ or $\mathbb{P}_{\boldsymbol{\theta}, \mathbf{V}_S, f}^{(n)}$, where $r \mapsto f(r) := f_\sigma(r) := \sigma^{-1} f_1(r/\sigma)$ ranges over the class

$$\mathcal{F} := \left\{ f : \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+ : \int_0^\infty r^{k-1} f(r) dr < \infty \right\}.$$

Irrespective of the choice of S , this shape matrix \mathbf{V}_S is a normalized version of the scatter matrix $\boldsymbol{\Sigma}$ —hence, under finite second-order moments, a normalized version of the ordinary covariance matrix. Therefore, it is a parameter of crucial interest in most standard multivariate analysis problems. Many problems in principal component analysis (PCA) and canonical correlation analysis (CCA), and the problem of testing for sphericity, among others, depend on shape rather than on the scatter matrix: see, for instance, [2], [6], [18]. Inference on shape in this context appears as an essential issue.

Intuitively, the choice of a particular scale functional should not affect inference about shape: irrespectively of S , \mathbf{V}_S indeed is just a representative of the class of all matrices that are proportional to Σ —a class that does not depend on S . This intuition is confirmed by the fact (see Sections 4.2 and 5.4) that semiparametric efficiency bounds for \mathbf{V}_S do not depend on S . However, a very subtle relation exists between parametric efficiencies for \mathbf{V}_S , the underlying radial density f , and the scale functional S . The subject of this paper is a study of that relation. A more detailed description of the problems we are considering however requires a more precise definition of the parametric and semiparametric models under study.

1.2 Shape: from parametric to semiparametric inference

Inference on shape, for any given S , naturally takes place in parametric families of the form

$$\mathcal{P}_{S;f}^{(n)} := \left\{ \mathbf{P}_{\boldsymbol{\theta}, \mathbf{V}, f}^{(n)} : \boldsymbol{\theta} \in \mathbb{R}^k, \mathbf{V} \in \mathcal{V}_k^S \right\},$$

where the radial density $f \in \mathcal{F}$ (hence, also the scale σ , which is the unique solution of $\int_0^\sigma r^{k-1} f(r) dr = \int_\sigma^\infty r^{k-1} f(r) dr$) is completely specified: call this case (a). Such full specification of f in practice is too much of an assumption, and the classical parametric approach consists in specifying f up to a scale parameter only, that is, specifying f_1 . The resulting family of densities, still of a parametric nature, is of the form (call this case (b))

$$\mathcal{P}_{S;f_1}^{(n)} := \left\{ \mathbf{P}_{\boldsymbol{\theta}, \sigma^2, \mathbf{V}, f_1}^{(n)} : \boldsymbol{\theta} \in \mathbb{R}^k, \sigma^2 \in \mathbb{R}_0^+, \mathbf{V} \in \mathcal{V}_k^S \right\}.$$

Specifying the radial density up to a scale parameter still may be unrealistic, and one may prefer considering the same problems of inference on \mathbf{V} in the semiparametric model under which f remains completely unspecified: call it case (c). The corresponding family can be described either as

$$\mathcal{P}_S^{(n)} := \left\{ \mathbf{P}_{\boldsymbol{\theta}, \mathbf{V}, f}^{(n)} : \boldsymbol{\theta} \in \mathbb{R}^k, \mathbf{V} \in \mathcal{V}_k^S; f \in \mathcal{F} \right\} = \bigcup_{f \in \mathcal{F}} \mathcal{P}_{S;f}^{(n)}$$

(call this case (c1)), or as

$$\mathcal{P}_S^{(n)} := \left\{ \mathbf{P}_{\boldsymbol{\theta}, \sigma^2, \mathbf{V}, f_1}^{(n)} : \boldsymbol{\theta} \in \mathbb{R}^k, \sigma^2 \in \mathbb{R}_0^+, \mathbf{V} \in \mathcal{V}_k^S; f_1 \in \mathcal{F}_1 \right\} = \bigcup_{f_1 \in \mathcal{F}_1} \mathcal{P}_{S;f_1}^{(n)}$$

(call this case (c2)). Although perfectly equivalent, cases (c1) and (c2) correspond to distinct semiparametric descriptions of the same family $\mathcal{P}_S^{(n)}$: in case (c1), the parameter is $(\boldsymbol{\theta}, \mathbf{V}) \in \mathbb{R}^k \times \mathcal{V}_k^S$, and the nonparametric nuisance is $f \in \mathcal{F}$, whereas in case (c2), the parameter is $(\boldsymbol{\theta}, \sigma^2, \mathbf{V}) \in \mathbb{R}^k \times \mathbb{R}_0^+ \times \mathcal{V}_k^S$, and the nonparametric nuisance is $f_1 \in \mathcal{F}_1$. This, as we shall see, leads to distinct definitions of adaptivity. In all cases ((a), (b), and (c)), the $d_i(\boldsymbol{\theta}; \mathbf{V})$'s are i.i.d., with density $\tilde{f}(r) = \sigma^{-1} \tilde{f}_1(r/\sigma)$ such that $\int_0^1 \tilde{f}_1(r) dr = 1/2$.

What is to be meant by *optimal* inference on $\mathbf{V} \in \mathcal{V}_k^S$, at given $\mathbf{P}_{\boldsymbol{\theta}, \mathbf{V}, f}^{(n)}$, as well as the corresponding optimal performance (efficiency), depends on the model (the family $\mathcal{P}_{S;f}^{(n)}$, $\mathcal{P}_{S;f_1}^{(n)}$, or $\mathcal{P}_S^{(n)}$) adopted, which includes the choice of a scale functional S . Clearly, for

given S , the optimal performance achievable in $\mathcal{P}_{S;f}^{(n)}$ (case (a)), where f is completely specified, is highest (*full* parametric efficiency), followed by the performance in $\mathcal{P}_{S;f_1}^{(n)}$ (case (b)), where f is only partially specified (parametric efficiency in the presence of unknown scale), and the performance in $\mathcal{P}_S^{(n)}$, where f is completely unspecified (semiparametric efficiency).

It can be shown that, under mild regularity assumptions on f and S (ensuring *local asymptotic normality* (LAN)),

- (i) the location parameter θ has no influence on any of these efficiencies, which are the same whether θ is known or not; in practice, any root- n consistent estimator $\theta^{(n)}$ thus safely can be substituted for an unspecified θ , whereas in theoretical developments, we can assume without any loss of generality that $\theta = \mathbf{0}$ (see [6]);
- (ii) at given S and f_1 , efficiencies in $\mathcal{P}_{S;f_1}^{(n)}$ (parametric efficiency in the presence of unknown scale) and in $\mathcal{P}_S^{(n)}$ (semiparametric efficiency) coincide; the difference between full parametric efficiency and semiparametric efficiency at $\mathbf{P}_{\theta, \mathbf{V}, f}^{(n)}$ ($\mathbf{V} \in \mathcal{V}_k^S$) thus is entirely due to the non-specification of scale—a fact that was already established in [1] for the inverse \mathbf{V}^{-1} of the shape matrix and a normalization based on the trace; the same proof however holds for arbitrary (homogeneous) scale functionals;
- (iii) if adaptivity (at given S) means that the semiparametric efficiency bounds (case (c)) for \mathbf{V} and the parametric ones (in the presence of unknown scale: case (b)) coincide, then (ii) implies that the model, as far as \mathbf{V} is concerned, is *adaptive*, or that \mathbf{V} is *adaptively estimable*, for any choice of S ; this property corresponds to the (c1)-description of $\mathcal{P}_S^{(n)}$ —call it *restricted adaptivity*;
- (iv) a stronger adaptivity property however has been established in [15], where it is shown that iff the determinant-based scale functional is adopted (namely, iff $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$), the semiparametric efficiency bounds (case (c)) for \mathbf{V} and the parametric *full*-efficiency ones (case (a)) coincide: even the non-specification of the scale here has no asymptotic cost when performing inference on \mathbf{V} , with respect to which the model is thus *fully adaptive*; this property corresponds to the (c2)-description of $\mathcal{P}_S^{(n)}$;
- (v) for given S , parametric and semiparametric efficiencies at $\mathbf{P}_{\theta, \mathbf{V}, f}^{(n)}$ ($\mathbf{V} \in \mathcal{V}_k^S$) do not depend on the actual values of θ and σ : it makes sense, thus, to speak about *f_1 -parametric efficiency* in case (a), about *f_1 -semiparametric efficiency* in cases (b) and (c) (see [15]).

Natural questions in this context are: how do these f_1 -parametric and f_1 -semiparametric efficiencies compare to each other, that is, how large is, under given f_1 and S , the cost of not knowing the scale (not knowing f at all) when performing inference on shape? Are there density types f_1 for which this cost would be minimal or maximal or zero? How does that cost depend on the definition of shape, that is, on the choice of the scale functional S (we know from [15] that this cost is nil for the determinant-based normalization)? These are the questions we are addressing here.

In the hypothesis testing context, efficiencies are measured in terms of local powers, which depend on quadratics in the parametric and semiparametric information matrices for \mathbf{V} . Those information matrices, and answers to the above questions, are explicitly provided in [6], for the scale functional $S(\boldsymbol{\Sigma}) = \Sigma_{11}$. Here, we extend the discussion to arbitrary S . We then turn to point estimation, where performance are measured in terms of asymptotic covariance matrices. These matrices, for efficient estimators, are the *inverses* of the information matrices involved in local powers, which are rather complex; obtaining their inverses under closed form (Lemma 5.2) is far from trivial. Those closed forms allow for the desired efficiency comparisons; they also are required in the definition of optimal test statistics for various related problems, such as testing the hypothesis of homogeneity of scatter matrices: see [7].

The main conclusions of these efficiency comparisons are that efficiency losses, for any given scale functional S , are maximum at arbitrarily light tails whereas, under arbitrarily heavy tails, they are arbitrarily close to zero. It follows that, under very heavy tails, ignoring the scale (ignoring the exact density) asymptotically does not harm inference on shape. Semiparametric efficiencies, on the other hand, only depend on the standardized radial density f_1 , not on the scale functional S . Finally, the result of [15] indicating that the efficiency loss is nil, irrespective of f_1 , when the scale functional is the k -th root of the scatter determinant, is confirmed.

1.3 Outline of the paper

The paper is organized as follows. Section 2 provides a local asymptotic normality (LAN) result for an arbitrary scale functional S , thus extending the earlier result obtained by [6] for $S(\boldsymbol{\Sigma}) = \Sigma_{11}$. The resulting parametric and semiparametric information matrices are derived in Section 3. The case of hypothesis testing problems is considered in Section 4, where the differences between parametric efficiency and semiparametric efficiency, measured by the variations of the noncentrality parameters of optimal tests, are evaluated as functions of f_1 and S . The same problem is considered, in Section 5, in the point estimation context. Technical results are proved in the appendix.

2 LAN for general shape and scale

The following notation will be used. For any $k \times k$ matrix \mathbf{A} , let $\text{vec } \mathbf{A}$ denote the k^2 -dimensional vector resulting from stacking the columns of \mathbf{A} on top of each other. If \mathbf{A} moreover is symmetric, write $\text{vech } \mathbf{A} := (A_{11}, (\text{vech } \mathbf{A})')'$ for the $(K + 1)$ -dimensional vector (throughout, $K := k(k + 1)/2 - 1$) obtained by stacking the upper-triangular elements of $\mathbf{A} = (A_{ij})$: $\text{vech } \mathbf{A}$ thus stands for $\text{vec } \mathbf{A}$ deprived of its first component A_{11} .

On the scale functional S we make the following assumption.

Assumption 2.1 *The scale functional $S : \mathcal{S}_k \rightarrow \mathbb{R}_0^+$*

- (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 $S(\mathbf{I}_k) = 1$, where \mathbf{I}_k denotes the $k \times k$ identity matrix.*

Clearly, one can also look at $\Sigma \mapsto S(\Sigma)$ as a function of $\text{vech } \Sigma$: with a slight abuse of notation, we indifferently write $S(\Sigma)$ or $S(\text{vech } \Sigma)$ in the sequel, and denote by $\nabla S(\text{vech } \Sigma)$ the gradient $\text{grad}_{\text{vech } \Sigma} S(\text{vech } \Sigma)$. Under Assumption 2.1, V_{11} , for any $\mathbf{V} = (V_{ij}) \in \mathcal{V}_k^S$, can be recovered from $(\text{vech } \mathbf{V})$. This special role of Σ_{11} in Assumption 2.1(ii) could have been played by any other entry of Σ . Assuming that some other component of ∇S is non-zero would allow, for instance, for dealing with scale functionals such as $S(\Sigma) = \Sigma_{22}$ or $S(\Sigma) = (\prod_{i=2}^k \Sigma_{ii})^{1/(k-1)}$ —with appropriate redefinition of the vech operator. As the extension of our results to such cases is trivial, we stick to Assumption 2.1 in the sequel.

Now, for any $\Sigma \in S_k$ and any S satisfying Assumption 2.1, define $\mathbf{C}_S^\Sigma := \mathbf{C}_{S,k}^\Sigma$ as the upper-triangular $k \times k$ matrix such that $\text{vech } \mathbf{C}_S^\Sigma = \nabla S(\text{vech } \Sigma)$, and let $\mathbf{D}_S^\Sigma := (\mathbf{C}_S^\Sigma + (\mathbf{C}_S^\Sigma)')/2$. Clearly, for any symmetric $k \times k$ matrix \mathbf{v} , $(\text{vec } \mathbf{D}_S^\Sigma)'(\text{vec } \mathbf{v}) = (\text{vech } \mathbf{C}_S^\Sigma)'(\text{vech } \mathbf{v})$. Define $\mathbf{M}_S^\Sigma := \mathbf{M}_{S,k}^\Sigma$ as the $K \times k^2$ matrix such that $(\mathbf{M}_S^\Sigma)'(\text{vech } \mathbf{v}) = \text{vec } \mathbf{v}$ for any symmetric $k \times k$ matrix \mathbf{v} satisfying $(\nabla S(\text{vech } \Sigma))'(\text{vech } \mathbf{v}) = 0$ (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 S and $\mathbf{V} \in \mathcal{V}_k^S$, define $\mathcal{E}_k^{\mathbf{V}} := \text{tr}(\mathbf{D}_S^{\mathbf{V}} \mathbf{V}^2)$. For $S(\Sigma) = \Sigma_{11}$, $S(\Sigma) = (\text{tr } \Sigma)/k$, and $S(\Sigma) = |\Sigma|^{1/k}$, one has $\mathbf{D}_S(\Sigma) = \mathbf{e}_1 \mathbf{e}_1'$ (where \mathbf{e}_1 denotes the first vector of the canonical basis of \mathbb{R}^k), $\mathbf{D}_S^\Sigma = \frac{1}{k} \mathbf{I}_k$, and $\mathbf{D}_S^\Sigma = \frac{1}{k} |\Sigma|^{1/k} \Sigma^{-1}$ —hence $\mathcal{E}_k^{\mathbf{V}} = 1$, $\mathcal{E}_k^{\mathbf{V}} = \frac{1}{k^2} \text{tr}(\mathbf{V}^2)$, and $\mathcal{E}_k^{\mathbf{V}} = \frac{1}{k}$, respectively.

The scale functional S , in the definition of the various parametric and semiparametric models $\mathcal{P}_{S:f}^{(n)}$, $\mathcal{P}_{S:f_1}^{(n)}$, and $\mathcal{P}_S^{(n)}$ (see Section 1.2) plays a role through the definition of the shape-parameter space \mathcal{V}_k^S only. Important as it is, this role is somewhat indirect. For the sake of clarity, we systematically, if artificially, emphasize it in the sequel, by writing \mathbf{V}_S for $\mathbf{V} \in \mathcal{V}_k^S$ and σ_S for σ .

For given S satisfying Assumption 2.1, the scatter parameter Σ (in vector form, $\text{vech } \Sigma$), as in the introduction, decomposes into scale and shape through $\Sigma = \sigma_S^2 \mathbf{V}_S$, where $\sigma_S^2 = S(\Sigma)$ and $\mathbf{V}_S := \Sigma/S(\Sigma) \in \mathcal{V}_k^S$. In vector form, dropping $(\mathbf{V}_S)_{11}$, the new parameter is thus $\boldsymbol{\vartheta}_S := (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')' \in \Theta_S := \mathbb{R}^k \times \mathbb{R}_0^+ \times \text{vech } \mathcal{V}_k^S$. The result below (see Theorem 2.1 in [15]) states that, under mild conditions, the families of distributions $\mathcal{P}_{S:f_1}^{(n)} := \{\mathbf{P}_{\boldsymbol{\vartheta}, f_1}^{(n)} : \boldsymbol{\vartheta} \in \Theta_S\}$ are *locally asymptotically normal* (LAN; see [10]). Of course, LAN requires some regularity condition on the radial density f_1 . A minimal assumption is given in [6], where only $S(\Sigma) = \Sigma_{11}$ is considered. Here, for the sake of simplicity, we rather provide the following sufficient one.

Assumption 2.2 *The standardized radial density f_1 belongs to the subset \mathcal{F}_1^* of all absolutely continuous functions $f_1 \in \mathcal{F}_1$ such that, denoting by \dot{f}_1 their a.e.-derivative, and letting $\varphi_{f_1} := -\dot{f}_1/f_1$, the quantities $\mathcal{I}_k(f_1) := \int_0^\infty (\varphi_{f_1}(r))^2 \tilde{f}_{1k}(r) dr$ and $\mathcal{J}_k(f_1) := \int_0^\infty r^2 (\varphi_{f_1}(r))^2 \tilde{f}_{1k}(r) dr$ are finite.*

This assumption is extremely mild, and does not imply any moment conditions; $\mathcal{I}_k(f_1)$ and $\mathcal{J}_k(f_1)$ can be interpreted as radial Fisher information for location and scale, respectively. It can be checked that—provided that $k \geq 2$ (the problem under consideration is void for $k = 1$)—Assumption 2.2 is satisfied at Gaussian densities, at all Student densities (including the Cauchy ones), as well as at all power-exponential densities.

Using the notation of the previous section, the corresponding radial Fisher information values are given, for the Gaussian, the Student with ν degrees of freedom, and the power-exponential with parameter η , by

$$\mathcal{I}_k(\phi_1) = a_k k, \quad \mathcal{I}_k(f_{1,\nu}^t) = a_{k,\nu} \frac{k(k+\nu)}{k+\nu+2}, \quad \mathcal{I}_k(f_{1,\eta}^e) = 4\eta^2 b_{k,\eta} \frac{\Gamma\left(\frac{4\eta+k-2}{2\eta}\right)}{\Gamma\left(\frac{k}{2\eta}\right)},$$

and

$$\mathcal{J}_k(\phi_1) = k(k+2), \quad \mathcal{J}_k(f_{1,\nu}^t) = \frac{k(k+2)(k+\nu)}{k+\nu+2}, \quad \mathcal{J}_k(f_{1,\eta}^e) = k(k+2\eta),$$

respectively, where Γ stands for Euler’s Gamma function. For all k , $\inf_{f_1 \in \mathcal{F}_1^*} \mathcal{J}_k(f_1) = k^2$ (see [6]), but no f_1 in \mathcal{F}_1^* achieves this lower bound. However, it is achieved at arbitrarily heavy tails, that is, as $\nu \rightarrow 0$ and $\eta \rightarrow 0$ in the classes of k -variate Student and power-exponential distributions, respectively.

The following notation is needed in the statement of LAN and will be used throughout the paper. Write $\mathbf{V}^{\otimes 2}$ for the Kronecker product $\mathbf{V} \otimes \mathbf{V}$. Denoting by \mathbf{e}_ℓ the ℓ -th 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 $\mathbf{J}_k := \sum_{i,j=1}^k (\mathbf{e}_i \mathbf{e}_j') \otimes (\mathbf{e}_i \mathbf{e}_j') = (\text{vec } \mathbf{I}_k)(\text{vec } \mathbf{I}_k)'$. Also let \mathbf{N}_k be such that $\mathbf{N}_k(\text{vec } \mathbf{v}) = (\text{vech } \mathbf{v})$ for any $k \times k$ matrix \mathbf{v} . Finally, although any square root $\mathbf{V}^{1/2}$ of $\mathbf{V} \in \mathcal{S}_k$ (satisfying $\mathbf{V}^{1/2}(\mathbf{V}^{1/2})' = \mathbf{V}$) can be used in the results below (provided, of course, it is used in a consistent way), we will use the symmetric root in order to save superfluous primes. We then have the following LAN result (see [15] for a proof).

Proposition 2.3 *Under Assumptions 2.1 and 2.2, the family $\mathcal{P}_{\vartheta_S, f_1}^{(n)} = \{\mathbf{P}_{\vartheta_S, f_1}^{(n)} : \vartheta_S \in \Theta_S\}$ is LAN. More precisely, for any $\vartheta_S = (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')$ and any bounded sequence $\boldsymbol{\tau}_n \in \mathbb{R}^{k+K+1}$, we have, under $\mathbf{P}_{\vartheta_S, f_1}^{(n)}$, as $n \rightarrow \infty$,*

$$(i) \quad \log \left(d\mathbf{P}_{\vartheta_S+n^{-1/2}\boldsymbol{\tau}_n, f_1}^{(n)} / d\mathbf{P}_{\vartheta_S, f_1}^{(n)} \right) = \boldsymbol{\tau}_n' \boldsymbol{\Delta}_{\vartheta_S, f_1}^{(n)} - \frac{1}{2} \boldsymbol{\tau}_n' \boldsymbol{\Gamma}_{\vartheta_S, f_1} \boldsymbol{\tau}_n + o_P(1),$$

where, letting $d_i := d(\mathbf{X}_i, \boldsymbol{\theta}; \mathbf{V}_S)$ and $\mathbf{U}_i := \mathbf{V}_S^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})/d_i$,

$$\boldsymbol{\Delta}_{\vartheta_S, f_1}^{(n)} := \left(\left(\boldsymbol{\Delta}_{\vartheta_S, f_1;1}^{(n)} \right)', \boldsymbol{\Delta}_{\vartheta_S, f_1;2}^{(n)}, \left(\boldsymbol{\Delta}_{\vartheta_S, f_1;3}^{(n)} \right)' \right)'$$

with

$$\begin{aligned} \boldsymbol{\Delta}_{\vartheta_S, f_1;1}^{(n)} &:= \frac{1}{\sigma_S \sqrt{n}} \sum_{i=1}^n \varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \mathbf{V}_S^{-1/2} \mathbf{U}_i, \\ \boldsymbol{\Delta}_{\vartheta_S, f_1;2}^{(n)} &:= \frac{1}{2\sigma_S^2 \sqrt{n}} \sum_{i=1}^n \left(\varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} - k \right), \end{aligned} \tag{2.1}$$

and

$$\boldsymbol{\Delta}_{\vartheta_S, f_1;3}^{(n)} := \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \sum_{i=1}^n \text{vec} \left(\varphi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} \mathbf{U}_i \mathbf{U}_i' - \mathbf{I}_k \right); \tag{2.2}$$

(ii) the central sequence $\mathbf{\Delta}_{\vartheta_S, f_1}^{(n)}$, still under $\mathbf{P}_{\vartheta_S, f_1}^{(n)}$, is asymptotically normal with mean zero and covariance matrix

$$\mathbf{\Gamma}_{\vartheta_S, f_1} := \begin{pmatrix} \mathbf{\Gamma}_{\vartheta_S, f_1;11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{\vartheta_S, f_1;22} & \mathbf{\Gamma}'_{\vartheta_S, f_1;32} \\ \mathbf{0} & \mathbf{\Gamma}_{\vartheta_S, f_1;32} & \mathbf{\Gamma}_{\vartheta_S, f_1;33} \end{pmatrix}, \tag{2.3}$$

where

$$\begin{aligned} \mathbf{\Gamma}_{\vartheta_S, f_1;11} &:= \frac{\mathcal{I}_k(f_1)}{k\sigma_S^2} \mathbf{V}^{-1}, \\ \mathbf{\Gamma}_{\vartheta_S, f_1;22} &:= \frac{\mathcal{J}_k(f_1) - k^2}{4\sigma_S^4}, \quad \mathbf{\Gamma}_{\vartheta_S, f_1;32} := \frac{\mathcal{J}_k(f_1) - k^2}{4k\sigma_S^2} \mathbf{M}_S^{\mathbf{V}_S} (\text{vec } \mathbf{V}_S^{-1}), \end{aligned}$$

and

$$\begin{aligned} \mathbf{\Gamma}_{\vartheta_S, f_1;33} & \tag{2.4} \\ &:= \frac{1}{4} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \left[\frac{\mathcal{J}_k(f_1)}{k(k+2)} (\mathbf{I}_{k^2} + \mathbf{K}_k + \mathbf{J}_k) - \mathbf{J}_k \right] (\mathbf{V}_S^{\otimes 2})^{-1/2} (\mathbf{M}_S^{\mathbf{V}_S})'. \end{aligned}$$

The block-diagonal structure of the information matrix (2.3) implies that the non-specification of the location centre $\boldsymbol{\theta}$ does not affect optimal parametric performances when estimating \mathbf{V}_S and/or σ_S^2 , or when performing tests about the same. More precisely, when estimating \mathbf{V}_S for instance, the optimal asymptotic covariance matrix that can be achieved (at $\mathbf{P}_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^{(n)}$) by an estimator of \mathbf{V}_S is the same (namely, $(\mathbf{\Gamma}_{\vartheta_S, f_1;33})^{-1}$) in $\mathcal{P}_{S; \sigma_S^2, f_1}^{(n)} := \left\{ \mathbf{P}_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^{(n)} : \boldsymbol{\theta} \in \mathbb{R}^k, \mathbf{V}_S \in \mathcal{V}_k^S \right\}$ as in $\mathcal{P}_{S; \sigma_S^2, f_1}^{(n)} := \left\{ \mathbf{P}_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^{(n)} : \mathbf{V}_S \in \mathcal{V}_k^S \right\}$. Since this asymptotic covariance only depends on \mathbf{V}_S and f_1 , so does this optimal performance.

On the contrary, a non-zero covariance $\mathbf{\Gamma}_{\vartheta_S, f_1;32}$ between the scale and shape parts of the central sequences implies that the optimal parametric performance when estimating \mathbf{V}_S is affected by the non-specification of the scale σ_S . The following result, which is proved in [15], shows that there is an important exception to this rule.

Theorem 2.4 *Let Assumptions 2.1 and 2.2 hold. Then $\mathbf{\Gamma}_{\vartheta_S, f_1;32} = \mathbf{0}$ for all $\vartheta_S \in \Theta_S$ and $f_1 \in \mathcal{F}_1^*$ iff $S = S_d$, where $S_d(\boldsymbol{\Sigma}) := |\boldsymbol{\Sigma}|^{1/k}$.*

This shows that the decomposition of scatter into scale and shape through the determinant-based scale functional is, in a sense, canonical; the determinant-based scale functional is the only one that guarantees orthogonality of the scale and shape parts of the central sequence, and, consequently, full adaptivity in the estimation of shape. For any other scale functional S , the non-specification of σ_S has a strictly positive cost when estimating \mathbf{V}_S . Our objective here is to quantify, both for hypothesis testing and point estimation, this cost for each f_1 and S , by comparing the parametric efficiency bounds under specified and unspecified scale (as already mentioned, the latter actually coincide with the semiparametric efficiency bounds).

3 Parametric and semiparametric efficiency bounds

The LAN result of Proposition 2.3 is about the “unspecified scale” model $\mathcal{P}_{S;f_1}^{(n)}$, but automatically entails LAN for the “specified scale” models $\mathcal{P}_{S;f}^{(n)}$, the information matrices of which are obtained by deleting the row and the column corresponding to σ_S^2 in (2.3). Parametric efficiency at $\mathbf{P}_{\vartheta_{S,f_1}}^{(n)}$ thus is characterized by the *parametric information matrix for shape* $\Gamma_{S;f_1}(\mathbf{V}) := \Gamma_{\vartheta_{S,f_1};33}$ in (2.4), which does not depend on σ_S nor on $\boldsymbol{\theta}$ (whence the notation).

Now, in the more realistic setup where $\boldsymbol{\theta}$, σ_S^2 , and f_1 remain unspecified and play the role of a nuisance, LAN and the convergence of local experiments to the Gaussian shift experiments

$$\begin{pmatrix} \Delta_2 \\ \Delta_3 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \Gamma_{\vartheta_{S,f_1};22} \Gamma'_{\vartheta_{S,f_1};32} \\ \Gamma_{\vartheta_{S,f_1};32} \Gamma_{\vartheta_{S,f_1};33} \end{pmatrix} \begin{pmatrix} \tau_2 \\ \tau_3 \end{pmatrix}, \begin{pmatrix} \Gamma_{\vartheta_{S,f_1};22} \Gamma'_{\vartheta_{S,f_1};32} \\ \Gamma_{\vartheta_{S,f_1};32} \Gamma_{\vartheta_{S,f_1};33} \end{pmatrix} \right) \quad (3.1)$$

(where $(\tau_2, \tau_3)' \in \mathbb{R}^{k(k+1)/2}$), imply that locally optimal inference on shape should be based on the residual of the regression (in (3.1)) of Δ_3 with respect to Δ_2 (that is, $\Delta_3 - \Gamma_{\vartheta_{S,f_1};32} \Gamma_{\vartheta_{S,f_1};22}^{-1} \Delta_2$), computed at $\Delta_{\vartheta_{S,f_1};3}^{(n)}$ and $\Delta_{\vartheta_{S,f_1};2}^{(n)}$. The resulting f_1 -efficient central sequence for shape is then

$$\begin{aligned} \Delta_{\vartheta_{S,f_1};3}^{(n)*} &:= \Delta_{\vartheta_{S,f_1};3}^{(n)} - \Gamma_{\vartheta_{S,f_1};32} \Gamma_{\vartheta_{S,f_1};22}^{-1} \Delta_{\vartheta_{S,f_1};2}^{(n)} \\ &= \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \sum_{i=1}^n \psi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} \text{vec} \left(\mathbf{U}_i \mathbf{U}_i' - \frac{1}{k} \mathbf{I}_k \right), \end{aligned}$$

which (unlike the original central sequence for shape; compare with $\Delta_{\vartheta_{S,f_1};3}^{(n)}$ in (2.2)) remains centered under $\cup_{g_1 \in \mathcal{F}_1} \mathbf{P}_{\vartheta_{S,g_1}}^{(n)}$. This efficient central sequence under $\mathbf{P}_{\vartheta_{S,f_1}}^{(n)}$ is asymptotically normal, with mean zero and covariance (the *semiparametrically efficient Fisher information for shape* under radial density f_1)

$$\begin{aligned} \Gamma_{S;f_1}^*(\mathbf{V}_S) &= \Gamma_{\vartheta_{S,f_1};33} - \Gamma_{\vartheta_{S,f_1};32} \Gamma_{\vartheta_{S,f_1};22}^{-1} \Gamma'_{\vartheta_{S,f_1};32} \\ &= \frac{\mathcal{J}_k(f_1)}{4k(k+2)} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k \right] (\mathbf{V}_S^{\otimes 2})^{-1/2} (\mathbf{M}_S^{\mathbf{V}_S})'. \end{aligned} \quad (3.2)$$

Clearly, Lemma A.1(iv) entails that, for $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$, one has $\Gamma_{S;f_1}^*(\mathbf{V}_S) = \Gamma_{S;f_1}(\mathbf{V}_S)$ and

$$\Delta_{\vartheta_{S,f_1};3}^{(n)*} = \Delta_{\vartheta_{S,f_1};3}^{(n)} = \frac{1}{2\sqrt{n}} \mathbf{M}_S^{\mathbf{V}_S} (\mathbf{V}_S^{\otimes 2})^{-1/2} \sum_{i=1}^n \psi_{f_1} \left(\frac{d_i}{\sigma_S} \right) \frac{d_i}{\sigma_S} \text{vec}(\mathbf{U}_i \mathbf{U}_i'),$$

and Theorem 2.4 shows that this definition of scale/shape is the only one for which parametric and semiparametric efficiency bounds do coincide (actually, it is the only parametrization for which $\Delta_{\vartheta_{S,f_1};3}^{(n)*}$ and $\Delta_{\vartheta_{S,f_1};3}^{(n)}$ are equal up to $o_P(1)$ terms.

The efficient Fisher information for shape $\Gamma_{S;f_1}^*(\mathbf{V}_S)$ provides the semiparametric efficiency bound at f_1 . More precisely, a test $\phi^{(n)}$ (with asymptotic level $\alpha \in (0, 1)$) of

$\mathcal{H}_0^{(n)} : \text{vech } \mathbf{V}_S = \text{vech } \mathbf{V}_S^0$ against $\mathcal{H}_1^{(n)} : \text{vech } \mathbf{V}_S \neq \text{vech } \mathbf{V}_S^0$ is said to be *f₁-semiparametrically optimal* iff, along any sequence of local alternatives of the form $\mathcal{H}_1^{(n)} : \text{vech } \mathbf{V}_S = (\text{vech } \mathbf{V}_S^0) + n^{-1/2} \boldsymbol{\tau}$, the power of $\phi^{(n)}$ tends to

$$1 - \Psi_K \left(\Psi_K^{-1}(1 - \alpha) ; \boldsymbol{\tau}' \boldsymbol{\Gamma}_{S;f_1}^* (\mathbf{V}_S^0) \boldsymbol{\tau} \right)$$

where $\Psi_\ell(\cdot ; c)$ stands for the distribution function of a noncentral chi-square variable with ℓ degrees of freedom and noncentrality parameter c , and $\Psi_\ell(\cdot) := \Psi_\ell(\cdot ; 0)$.

Similarly, an estimator $\mathbf{V}_S^{(n)}$ of $\mathbf{V}_S \in \mathcal{V}_k^S$ is said to be *f₁-semiparametrically efficient* iff, for all $\boldsymbol{\vartheta}_S \in \boldsymbol{\Theta}_S$,

$$n^{1/2} \text{vech} \left(\mathbf{V}_S^{(n)} - \mathbf{V}_S \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}, \left(\boldsymbol{\Gamma}_{S;f_1}^* (\mathbf{V}_S) \right)^{-1} \right), \tag{3.3}$$

under $\mathbf{P}_{\boldsymbol{\vartheta}_S, f_1}^{(n)}$, as $n \rightarrow \infty$. Bickel ([1, Example 4]) has shown that this *f₁-semiparametric efficiency* coincides with parametric efficiency under *f₁* and unspecified σ_S .

4 Hypothesis testing

In this section, we consider in further detail the testing problem

$$\mathcal{H}_0^{(n)} : \text{vech } \mathbf{V}_S = \text{vech } \mathbf{V}_S^0 \quad \text{versus} \quad \mathcal{H}_1^{(n)} : \text{vech } \mathbf{V}_S \neq \text{vech } \mathbf{V}_S^0, \tag{4.1}$$

where $\mathbf{V}_S, \mathbf{V}_S^0 \in \mathcal{V}_k^S$ (\mathbf{V}_S^0 fixed). A test $\phi^{(n)}$ (with asymptotic level $\alpha \in (0, 1)$) is said to be *f₁-parametrically optimal* (resp., *f₁-semiparametrically optimal*) iff, along any sequence of local alternatives of the form $\mathcal{H}_1^{(n)} : \text{vech } \mathbf{V}_S = (\text{vech } \mathbf{V}_S^0) + n^{-1/2} \boldsymbol{\tau}$, the asymptotic power of $\phi^{(n)}$ is

$$1 - \Psi_K \left(\Psi_K^{-1}(1 - \alpha) ; \boldsymbol{\tau}' \mathbf{G} \boldsymbol{\tau} \right), \tag{4.2}$$

where $\mathbf{G} := \boldsymbol{\Gamma}_{S;f_1} (\mathbf{V}_S^0)$ (resp., $\mathbf{G} := \boldsymbol{\Gamma}_{S;f_1}^* (\mathbf{V}_S^0)$).

Of course, the testing problem (4.1) is equivalent to

$$\mathcal{H}_0^{(n)} : \mathbf{V}_S = \mathbf{V}_S^0 \quad \text{versus} \quad \mathcal{H}_1^{(n)} : \mathbf{V}_S \neq \mathbf{V}_S^0,$$

where $\mathbf{V}_S, \mathbf{V}_S^0 \in \mathcal{V}_k^S$; the sequence of alternatives $\text{vech } \mathbf{V}_S = (\text{vech } \mathbf{V}_S^0) + n^{-1/2} \boldsymbol{\tau}$ also naturally writes $\mathbf{V}_S = \mathbf{V}_S^0 + n^{-1/2} \mathbf{v}^{(n)}$, where the entry $(\mathbf{v}^{(n)})_{11}$ is defined in such a way that $S(\mathbf{V}_S^0 + n^{-1/2} \mathbf{v}^{(n)}) = 1$ for all n (other entries are determined by $(\text{vech } \mathbf{v}^{(n)}) = \boldsymbol{\tau}$, and do not depend on n). Since

$$0 = S(\mathbf{V}_S^0 + n^{-1/2} \mathbf{v}^{(n)}) - S(\mathbf{V}_S^0) = n^{-1/2} [\nabla S(\text{vech } \mathbf{V}_S^0)]' (\text{vech } \mathbf{v}^{(n)}) + o(n^{-1/2})$$

as $n \rightarrow \infty$, we have

$$0 = \lim_{n \rightarrow \infty} [\nabla S(\text{vech } \mathbf{V}_S^0)]' (\text{vech } \mathbf{v}^{(n)}) = \frac{\partial S}{\partial \Sigma_{11}} (\mathbf{V}_S^0) \lim_{n \rightarrow \infty} (\mathbf{v}^{(n)})_{11} + (\text{vech } \mathbf{C}_S^{\mathbf{V}_S^0})' \boldsymbol{\tau},$$

so that Assumption 2.1 guarantees that $\lim_{n \rightarrow \infty} (\mathbf{v}^{(n)})_{11}$, hence also $\mathbf{v} := \lim_{n \rightarrow \infty} (\mathbf{v}^{(n)})$, exist. This shows that sequences of local alternatives under matrix form must satisfy

$$[\nabla S(\text{vech } \mathbf{V}_S^0)]' (\text{vech } \mathbf{v}) = \left(\text{vec } \mathbf{D}_S^{\mathbf{V}_S^0} \right)' (\text{vec } \mathbf{v}) = \text{tr} \left(\mathbf{D}_S^{\mathbf{V}_S^0} \mathbf{v} \right) = 0. \tag{4.3}$$

4.1 The cost of unspecified scale

We now evaluate, for any given scale functional S , the difference between the local powers in (4.2) associated with parametrically and semiparametrically optimal tests, respectively. More precisely, we measure the efficiency loss by comparing the noncentrality parameters $\boldsymbol{\tau}' \boldsymbol{\Gamma}_{S;f_1}(\mathbf{V}_S^0) \boldsymbol{\tau}$ and $\boldsymbol{\tau}' \boldsymbol{\Gamma}_{S;f_1}^*(\mathbf{V}_S^0) \boldsymbol{\tau}$.

Theorem 4.1 *Let Assumptions 2.1 and 2.2 hold. Then, $\boldsymbol{\tau}' \boldsymbol{\Gamma}_{S;f_1}(\mathbf{V}_S^0) \boldsymbol{\tau}$ and $\boldsymbol{\tau}' \boldsymbol{\Gamma}_{S;f_1}^*(\mathbf{V}_S^0) \boldsymbol{\tau}$ are given by*

$$\frac{\mathcal{J}_k(f_1)}{2k(k+2)} \operatorname{tr} \left(\left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right)^2 \right) + \frac{\mathcal{J}_k(f_1) - k(k+2)}{4k(k+2)} \left(\operatorname{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) \right)^2 \tag{4.4}$$

and

$$\frac{\mathcal{J}_k(f_1)}{2k(k+2)} \left\{ \operatorname{tr} \left(\left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right)^2 \right) - \frac{1}{k} \left(\operatorname{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) \right)^2 \right\}, \tag{4.5}$$

respectively.

Hence, a measure of the loss due to the unspecified scale σ_S is the difference

$$\delta_{S;f_1}(\mathbf{V}_S^0, \mathbf{v}) = \boldsymbol{\tau}' \left[\boldsymbol{\Gamma}_{S;f_1}(\mathbf{V}_S^0) - \boldsymbol{\Gamma}_{S;f_1}^*(\mathbf{V}_S^0) \right] \boldsymbol{\tau} = \frac{\mathcal{J}_k(f_1) - k^2}{4k^2} \left(\operatorname{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) \right)^2$$

between the noncentrality parameters in (4.4) and (4.5). Another measure of the same loss, with interpretation in terms of asymptotic relative efficiency, hence in observation numbers, is the ratio $\operatorname{ARE}_{S;f_1}(\mathbf{V}_S^0, \mathbf{v})$ of (4.5) to (4.4).

The difference $\delta_{S;f_1}(\mathbf{V}_S^0, \mathbf{v})$ of course is always nonnegative (recall indeed that $\inf_{f_1 \in \mathcal{F}_1^*} \mathcal{J}_k(f_1) = k^2$), whereas the ratio is always smaller than or equal to one. It follows from (4.3) that, for $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$, the loss is nil: $\delta_{S;f_1}(\mathbf{V}_S^0, \mathbf{v}) = 0$ for all $\mathbf{V}_S^0, \mathbf{v}$, and f_1 . Theorem 2.4 moreover shows that this definition of scale/shape is the only one for which this holds. Note however that, for $S(\boldsymbol{\Sigma}) = (\operatorname{tr} \boldsymbol{\Sigma})/k$, no loss is incurred when testing for sphericity: indeed, $\delta_{S;f_1}(\mathbf{I}_k, \mathbf{v}) = 0$ for all \mathbf{v} and f_1 .

Irrespective of the definition of scale/shape, the difference $\delta_{S;f_1}(\mathbf{V}_S^0, \mathbf{v})$ converges to zero for all \mathbf{V}_S^0 and \mathbf{v} when the Fisher information for scale/shape $\mathcal{J}_k(f_1)$ tends to its infimum k^2 , which occurs at arbitrarily heavy tails, that is, as $\nu \rightarrow 0$ and $\eta \rightarrow 0$ in the classes of k -variate Student and power-exponential distributions, respectively. On the contrary, the same difference $\delta_{S;f_1}(\mathbf{V}_S^0, \mathbf{v})$ (unless of course $\operatorname{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) = 0$) becomes arbitrarily large when $\mathcal{J}_k(f_1)$ tends to infinity, that is, under arbitrarily light tails—for instance, as $\eta \rightarrow \infty$ in the class of k -variate power-exponential distributions. Whether this means that the corresponding loss *tends to infinity* is a matter of definition; a rapid calculation shows that asymptotic relative efficiencies, under the same conditions, converge to

$$\inf_{f_1 \in \mathcal{F}_1^*} \operatorname{ARE}_{S;f_1}(\mathbf{V}_S^0, \mathbf{v}) = \frac{\operatorname{tr} \left(\left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right)^2 \right) - \frac{1}{k} \left(\operatorname{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) \right)^2}{\operatorname{tr} \left(\left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right)^2 \right) + \frac{1}{2} \left(\operatorname{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) \right)^2}.$$

4.2 Comparisons between the various S -local powers

Efficiency losses in Section 4.1 all are considered for a *fixed scale functional* S , and various density types f_1 . The null hypothesis in (4.1) also can be written, in $\mathcal{P}_{T;f}^{(n)}$ or $\mathcal{P}_T^{(n)}$, as $\mathcal{H}_0^{(n)} : \mathbf{V}_T = \mathbf{V}_T^0$, for any other scale functional $T \neq S$, where $\mathbf{V}_T^0 = \mathbf{V}_S^0/T(\mathbf{V}_S^0)$. In this section, we show that while the parametrically optimal performances associated with S and T may be quite different, the semiparametrically optimal ones, at any f_1 , all are the same—hence coincide with the parametrically optimal performance for the determinant-based scale functional, which therefore can be considered as the “least favorable” one.

To be more specific, we consider the semiparametric performance that can be achieved when testing $\mathcal{H}_0^{(n)} : \text{vech } \mathbf{V}_S = \text{vech } \mathbf{V}_S^0$ against local alternatives of the form

$$\text{vech } \mathbf{V}_S = \text{vech } \mathbf{V}_S^0 + n^{-1/2} \boldsymbol{\tau}_S = \text{vech } \left(\mathbf{V}_S^0 + n^{-1/2} \mathbf{v}_S^{(n)} \right), \tag{4.6}$$

where $(\mathbf{v}_S^{(n)})_{11}$ is such that $\mathbf{V}_S^0 + n^{-1/2} \mathbf{v}_S^{(n)}$ is an (S -)shape matrix, which implies—as in (4.3)—that $\text{tr}(\mathbf{D}_S^{\mathbf{V}_S^0} \mathbf{v}_S) = 0$ (with $\mathbf{v}_S := \lim_{n \rightarrow \infty} \mathbf{v}_S^{(n)}$). This performance, which is measured by the noncentrality parameter $\boldsymbol{\tau}'_S \boldsymbol{\Gamma}_{S;f_1}^* (\mathbf{V}_S^0) \boldsymbol{\tau}_S$, is to be compared with the semiparametrically optimal one obtained when testing $\mathcal{H}_0^{(n)} : \text{vech } \mathbf{V}_T = (\text{vech } \mathbf{V}_T^0)$, where $\mathbf{V}_T^0 = \mathbf{V}_S^0/T(\mathbf{V}_S^0)$, against local alternatives of the form

$$\text{vech } \mathbf{V}_T = \text{vech } \mathbf{V}_T^0 + n^{-1/2} \boldsymbol{\tau}_T = \text{vech } \left(\mathbf{V}_T^0 + n^{-1/2} \mathbf{v}_T^{(n)} \right), \tag{4.7}$$

where $\boldsymbol{\tau}_T$ and $\mathbf{v}_T^{(n)}$ are such that the local alternatives (4.6) and (4.7) coincide. This requires that

$$\text{vech } \left(\mathbf{V}_T^0 + n^{-1/2} \mathbf{v}_T^{(n)} \right) = \text{vech } \left(\frac{\mathbf{V}_S^0 + n^{-1/2} \mathbf{v}_S^{(n)}}{T(\mathbf{V}_S^0 + n^{-1/2} \mathbf{v}_S^{(n)})} \right) + o(n^{-1/2}), \tag{4.8}$$

where $\text{tr}(\mathbf{D}_T^{\mathbf{V}_T^0} \mathbf{v}_T) = 0$, with $\mathbf{v}_T := \lim_{n \rightarrow \infty} (\mathbf{v}_T^{(n)})$ (again a consequence of the fact that $(\mathbf{v}_T^{(n)})_{11}$ is chosen in such a way that $\mathbf{V}_T^0 + n^{-1/2} \mathbf{v}_T^{(n)}$ is a T -shape matrix). The resulting semiparametric performance is then measured by the noncentrality parameter $\boldsymbol{\tau}'_T \boldsymbol{\Gamma}_{T;f_1}^* (\mathbf{V}_T^0) \boldsymbol{\tau}_T$. The following result shows that, as announced, the semiparametric performances for \mathbf{V}_S and \mathbf{V}_T do coincide.

Theorem 4.2 *Let Assumptions 2.1 and 2.2 hold. Then, with the same notation as above,*

(i) *the quantities \mathbf{V}_T^0 , $\mathbf{v}_T^{(n)}$, and $\boldsymbol{\tau}_T$ are, in terms of \mathbf{V}_S^0 and \mathbf{v}_S ,*

$$\begin{aligned} \mathbf{V}_T^0 &= \mathbf{V}_S^0/T(\mathbf{V}_S^0), \quad \boldsymbol{\tau}_T = (\text{vech } \mathbf{v}_T), \quad \text{and} \\ \mathbf{v}_T &= \frac{1}{T(\mathbf{V}_S^0)} \left[\mathbf{v}_S - \left(\text{tr} \left(\mathbf{D}_T^{\mathbf{V}_T^0} \mathbf{v}_S \right) \right) \mathbf{V}_T^0 \right]. \end{aligned} \tag{4.9}$$

(ii) *The S - and T -semiparametric performances do coincide, that is,*

$$\boldsymbol{\tau}'_S \boldsymbol{\Gamma}_{S;f_1}^* (\mathbf{V}_S^0) \boldsymbol{\tau}_S = \boldsymbol{\tau}'_T \boldsymbol{\Gamma}_{T;f_1}^* (\mathbf{V}_T^0) \boldsymbol{\tau}_T$$

at any $f_1 \in \mathcal{F}_1^$.*

See the appendix for the proof.

5 Point estimation

5.1 Some definitions and a key lemma

An estimator $\mathbf{V}_S^{(n)}$ of $\mathbf{V}_S \in \mathcal{V}_k^S$ is said to be f_1 -parametrically efficient (resp., f_1 -semi-parametrically efficient) iff, for all $\boldsymbol{\vartheta}_S = (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')' \in \boldsymbol{\Theta}_S$, we have

$$n^{1/2} \text{vech} \left(\mathbf{V}_S^{(n)} - \mathbf{V}_S \right) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{G}^{-1}), \tag{5.1}$$

under $\mathbb{P}_{\boldsymbol{\vartheta}_S, f_1}^{(n)}$, as $n \rightarrow \infty$, with $\mathbf{G} := \boldsymbol{\Gamma}_{S; f_1}(\mathbf{V}_S)$ (resp., $\mathbf{G} := \boldsymbol{\Gamma}_{S; f_1}^*(\mathbf{V}_S)$), or, in terms of $\text{vec } \mathbf{V}_S$, iff

$$n^{1/2} \text{vec} \left(\mathbf{V}_S^{(n)} - \mathbf{V}_S \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}, \left(\mathbf{M}_S^{\mathbf{V}_S} \right)' \mathbf{G}^{-1} \mathbf{M}_S^{\mathbf{V}_S} \right), \tag{5.2}$$

under $\mathbb{P}_{\boldsymbol{\vartheta}_S, f_1}^{(n)}$, as $n \rightarrow \infty$. This result (5.2) for the vec form follows from the following lemma.

Lemma 5.1 *Let Assumption 2.1 hold. Then, under $\mathbb{P}_{\boldsymbol{\vartheta}_S, f_1}^{(n)}$,*

$$n^{1/2} \text{vec} \left(\mathbf{V}_S^{(n)} - \mathbf{V}_S \right) = n^{1/2} \left(\mathbf{M}_S^{\mathbf{V}_S} \right)' \text{vech} \left(\mathbf{V}_S^{(n)} - \mathbf{V}_S \right) + o_{\mathbb{P}}(1) \text{ as } n \rightarrow \infty.$$

In this section, we provide—for any scale functional S —an explicit ($\mathbf{M}_S^{\mathbf{V}_S}$ -free) expression for the asymptotic covariance in (5.2), allowing for a comparison with the asymptotic covariance matrix achievable in the more realistic unspecified scale setup, that is, in the family $\mathcal{P}_{S; f_1}^{(n)} := \{ \mathbb{P}_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, f_1}^{(n)}, \boldsymbol{\theta} \in \mathbb{R}^k, \sigma_S^2 \in \mathbb{R}_0^+, \mathbf{V}_S \in \mathcal{V}_k^S \}$ —a performance that coincides (as shown in [1]) with the f_1 -semiparametric one, achievable in $\mathcal{P}_S^{(n)}$.

Let

$$\mathbf{Q}_{k; r, s}^{\mathbf{V}_S} := \mathbf{P}_k^{\mathbf{V}_S} \left\{ r \left[\mathbf{I}_{k^2} + \mathbf{K}_k \right] \left(\mathbf{V}_S^{\otimes 2} \right) + s \left(\text{vec } \mathbf{V}_S \right) \left(\text{vec } \mathbf{V}_S \right)' \right\} \left(\mathbf{P}_k^{\mathbf{V}_S} \right)', \tag{5.3}$$

where

$$\mathbf{P}_k^{\mathbf{V}_S} := \mathbf{I}_{k^2} - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} \left(\mathbf{V}_S^{\otimes 2} \right) \left(\text{vec } \mathbf{D}_S^{\mathbf{V}_S} \right) \left(\text{vec } \mathbf{D}_S^{\mathbf{V}_S} \right)'. \tag{5.4}$$

The following lemma constitutes the key result in the derivation of an explicit expression of the asymptotic covariance matrix in (5.2) (see the appendix for the proof).

Lemma 5.2 *Let $\mathbf{V}_S \in \mathcal{V}_k^S$ and $a, b \in \mathbb{R}$ such that $a \neq 0$ and $b \neq (2a + bk)\mathcal{E}_k^{\mathbf{V}_S}$. Then,*

$$(i) \quad \left\{ \frac{1}{4} \mathbf{M}_S^{\mathbf{V}_S} \left(\mathbf{V}_S^{\otimes 2} \right)^{-1/2} \left[a \left(\mathbf{I}_{k^2} + \mathbf{K}_k \right) + b \mathbf{J}_k \right] \left(\mathbf{V}_S^{\otimes 2} \right)^{-1/2} \left(\mathbf{M}_S^{\mathbf{V}_S} \right)' \right\}^{-1} \tag{5.5}$$

$$= \mathbf{N}_k \mathbf{Q}_{k; A, B}^{\mathbf{V}_S} \mathbf{N}_k',$$

with $A := a^{-1}$ and $B := 2a^{-1} b \mathcal{E}_k^{\mathbf{V}_S} / (b - (2a + bk)\mathcal{E}_k^{\mathbf{V}_S})$;

(ii) for all $r, s \in \mathbb{R}$, $\left(\mathbf{M}_S^{\mathbf{V}_S} \right)' \mathbf{N}_k \mathbf{Q}_{k; r, s}^{\mathbf{V}_S} = \mathbf{Q}_{k; r, s}^{\mathbf{V}_S} = \mathbf{Q}_{k; r, s}^{\mathbf{V}_S} \mathbf{N}_k' \mathbf{M}_S^{\mathbf{V}_S}$.

5.2 Parametric and semiparametric performances

Lemma 5.2 and (5.2) directly yield the following result.

Proposition 5.3 *Let Assumptions 2.1 and 2.2 hold. Then, the asymptotic (under $\mathbb{P}_{\vartheta_S, f_1}^{(n)}$, as $n \rightarrow \infty$) covariance matrix of (the vec form of) f_1 -parametrically efficient estimators of \mathbf{V}_S is, for all $\vartheta_S = (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')' \in \Theta_S$,*

$$\left(\mathbf{M}_S^{\mathbf{V}_S}\right)' \left(\boldsymbol{\Gamma}_{S; f_1}(\mathbf{V}_S)\right)^{-1} \mathbf{M}_S^{\mathbf{V}_S} = \frac{k(k+2)}{\mathcal{J}_k(f_1)} \mathbf{Q}_{k; 1, 2, \mathcal{M}_{k, f_1}^{\mathbf{V}_S}}, \tag{5.6}$$

where

$$\mathcal{M}_{k, f_1}^{\mathbf{V}_S} := \frac{(\mathcal{J}_k(f_1) - k(k+2))\mathcal{E}_k^{\mathbf{V}_S}}{\mathcal{J}_k(f_1) - k(k+2) - (k+2)(\mathcal{J}_k(f_1) - k^2)\mathcal{E}_k^{\mathbf{V}_S}}. \tag{5.7}$$

Note that, since $\mathcal{J}_k(f_1) \geq k^2$, the quantity $\mathcal{M}_{k, f_1}^{\mathbf{V}_S}$ in (5.7) satisfies

$$-\mathcal{E}_k^{\mathbf{V}_S} / ((k+2)\mathcal{E}_k^{\mathbf{V}_S} - 1) \leq \mathcal{M}_{k, f_1}^{\mathbf{V}_S} \leq \mathcal{E}_k^{\mathbf{V}_S}. \tag{5.8}$$

These lower and upper bounds for $\mathcal{M}_{k, f_1}^{\mathbf{V}_S}$ are achieved, for instance, within the class of k -variate power-exponential densities $f_{1, \eta}^e$, by letting $\eta \rightarrow \infty$ and $\eta \rightarrow 0$, respectively.

As an illustration, we now provide an estimator $\mathbf{V}_{S, \mathcal{N}}^{(n)}$ that is ϕ_1 -parametrically efficient, that is, parametrically efficient in the multinormal case with specified scale. Under $\mathbb{P}_{\vartheta_S, \phi_1}^{(n)}$, the regular sample covariance matrix $\boldsymbol{\Sigma}^{(n)} := (n-1)^{-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$ (with $\bar{\mathbf{X}} := n^{-1} \sum_{i=1}^n \mathbf{X}_i$) is consistent for $\boldsymbol{\Sigma}_k := a_k^{-1} \sigma_S^2 \mathbf{V}_S$ (where a_k was defined in Section 1.1). Actually, it is easy to show that

$$n^{1/2} \text{vec}(\boldsymbol{\Sigma}^{(n)} - \boldsymbol{\Sigma}_k) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, [\mathbf{I}_{k^2} + \mathbf{K}_k](\boldsymbol{\Sigma}_k^{\otimes 2})),$$

under $\mathbb{P}_{\vartheta_S, \phi_1}^{(n)}$, as $n \rightarrow \infty$. If the scale parameter $\sigma_S^2 = a_k S(\boldsymbol{\Sigma}_k)$ is known, one can define the estimator of shape

$$\begin{aligned} \mathbf{V}_{S, \mathcal{N}}^{(n)} &:= \frac{\boldsymbol{\Sigma}^{(n)}}{S(\boldsymbol{\Sigma}_k)} - \frac{1}{\mathcal{E}_k(\mathbf{V}_{S, \mathcal{N}}^{*(n)})} \left(\frac{S(\boldsymbol{\Sigma}^{(n)})}{S(\boldsymbol{\Sigma}_k)} - 1 \right) \mathbf{V}_{S, \mathcal{N}}^{*(n)} \mathbf{D}_S^{\mathbf{V}_{S, \mathcal{N}}^{*(n)}} \mathbf{V}_{S, \mathcal{N}}^{*(n)} \\ &= \mathbf{V}_{S, \mathcal{N}}^{*(n)} + \left(\frac{S(\boldsymbol{\Sigma}^{(n)})}{S(\boldsymbol{\Sigma}_k)} - 1 \right) \left\{ \mathbf{V}_{S, \mathcal{N}}^{*(n)} - \frac{1}{\mathcal{E}_k(\mathbf{V}_{S, \mathcal{N}}^{*(n)})} \mathbf{V}_{S, \mathcal{N}}^{*(n)} \mathbf{D}_S^{\mathbf{V}_{S, \mathcal{N}}^{*(n)}} \mathbf{V}_{S, \mathcal{N}}^{*(n)} \right\}, \end{aligned} \tag{5.9}$$

where $\mathbf{V}_{S, \mathcal{N}}^{*(n)} := \boldsymbol{\Sigma}^{(n)} / S(\boldsymbol{\Sigma}^{(n)})$ does not take advantage from the specified scale. The difference $\mathbf{V}_{S, \mathcal{N}}^{(n)} - \mathbf{V}_{S, \mathcal{N}}^{*(n)}$ can be interpreted as the improvement (in the estimation of \mathbf{V}_S) that can be achieved when knowing the value of the scale σ_S ; indeed, as we will show, $\mathbf{V}_{S, \mathcal{N}}^{(n)}$ (resp., $\mathbf{V}_{S, \mathcal{N}}^{*(n)}$) is parametrically efficient in the multinormal case with specified (resp., unspecified) scale. That such an improvement is possible is intuitively clear from

the non-diagonal form of the information matrix in (2.3). However, in the particular case where $S(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/k}$, that information matrix is block-diagonal and no improvement can be expected; actually, one easily checks that $\mathbf{V}_{S,\mathcal{N}}^{(n)} = \mathbf{V}_{S,\mathcal{N}}^{*(n)}$ for this choice of S .

Now, let us show that the estimator $\mathbf{V}_{S,\mathcal{N}}^{(n)}$ in (5.9) is indeed parametrically efficient in the multinormal case with specified scale. Applying Slutsky's Lemma, we obtain, under $\bigcup_{\boldsymbol{\theta}} \{P_{\boldsymbol{\theta}, \sigma_S^2, \mathbf{V}_S, \phi_1}^{(n)}\}$,

$$\begin{aligned} n^{1/2} \text{vec}(\mathbf{V}_{S,\mathcal{N}}^{(n)} - \mathbf{V}_S) &= \frac{1}{S(\boldsymbol{\Sigma}_k)} \mathbf{P}_k^{\mathbf{V}_S} \left[n^{1/2} \text{vec}(\boldsymbol{\Sigma}^{(n)} - \boldsymbol{\Sigma}_k) \right] + o_P(1) \\ &\xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k;1,0}^{\mathbf{V}_S}), \end{aligned}$$

as $n \rightarrow \infty$. The resulting asymptotic covariance matrix $\mathbf{Q}_{k;1,0}^{\mathbf{V}_S}$ is the value of the asymptotic covariance matrix in (5.6) at normal radial densities ($f_1 = \phi_1$), which establishes the ϕ_1 -parametric efficiency of $\mathbf{V}_{S,\mathcal{N}}^{(n)}$.

Using Lemma 5.2 again, we obtain

Proposition 5.4 *Let Assumptions 2.1 and 2.2 hold. Then, the asymptotic (under $P_{\boldsymbol{\vartheta}_S, f_1}^{(n)}$, as $n \rightarrow \infty$) covariance matrix of (the vec form of) f_1 -semiparametrically efficient estimators of \mathbf{V}_S is, for all $\boldsymbol{\vartheta}_S = (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')' \in \Theta_S$,*

$$\left(\mathbf{M}_S^{\mathbf{V}_S}\right)' \left(\boldsymbol{\Gamma}_{S;f_1}^*(\mathbf{V}_S)\right)^{-1} \mathbf{M}_S^{\mathbf{V}_S} = \frac{k(k+2)}{\mathcal{J}_k(f_1)} \mathbf{Q}_{k;1,2\varepsilon_k}^{\mathbf{V}_S}, \tag{5.10}$$

where $\mathbf{Q}_{k;r,s}^{\mathbf{V}_S}$ is defined in (5.3).

Most estimators of shape $\mathbf{V}^{(n)}$, once properly normalized into $\mathbf{V}_S^{(n)} := \mathbf{V}^{(n)}/S(\mathbf{V}^{(n)})$, are such that $\text{vec } \mathbf{V}_S^{(n)}$ has an asymptotic covariance matrix (at $P_{\boldsymbol{\vartheta}, f_1}^{(n)}$) of the form $\lambda_{k, f_1} \times \mathbf{Q}_{k;1,2\varepsilon_k}^{\mathbf{V}_S}$, for some $\lambda_{k, f_1} > 0$; see, e.g., [2], [11], [13], or [20]. The ratio $k(k+2)/[\mathcal{J}_k(f_1)\lambda_{k, f_1}]$ then can be used as a measure of their relative efficiency with respect to the semiparametric efficiency bound at f_1 .

We conclude this section by showing that $\mathbf{V}_{S,\mathcal{N}}^{*(n)} = \boldsymbol{\Sigma}^{(n)}/S(\boldsymbol{\Sigma}^{(n)})$, as announced above, is parametrically efficient in the multinormal case with unspecified scale (at $f_1 = \phi_1$), hence semiparametrically efficient at the same. Slutsky's Lemma yields

$$\begin{aligned} n^{1/2} \text{vec}(\mathbf{V}_{S,\mathcal{N}}^{*(n)} - \mathbf{V}_S) &= \frac{1}{S(\boldsymbol{\Sigma}_k)} \left[\mathbf{I}_{k^2} - (\text{vec } \mathbf{V}_S) (\text{vec } \mathbf{D}_S^{\mathbf{V}_S})' \right] \left[n^{1/2} \text{vec}(\boldsymbol{\Sigma}^{(n)} - \boldsymbol{\Sigma}_k) \right] + o_P(1) \\ &\xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \left[\mathbf{I}_{k^2} - (\text{vec } \mathbf{V}_S) (\text{vec } \mathbf{D}_S^{\mathbf{V}_S})' \right] \left[\mathbf{I}_{k^2} + \mathbf{K}_k \right] \right. \\ &\quad \left. \times (\mathbf{V}_S^{\otimes 2}) \left[\mathbf{I}_{k^2} - (\text{vec } \mathbf{V}_S) (\text{vec } \mathbf{D}_S^{\mathbf{V}_S})' \right]' \right), \end{aligned}$$

under $\bigcup_{\theta} \{P_{\theta, \sigma_S^2, \mathbf{V}_S, \phi_1}^{(n)}\}$, as $n \rightarrow \infty$, where the asymptotic covariance matrix, after standard algebra, reduces to $\mathbf{Q}_{k; 1, 2\mathcal{E}_k^{\mathbf{V}_S}}^{\mathbf{V}_S}$, which is the value, at $f_1 = \phi_1$, of the asymptotic covariance matrix in (5.10). The semiparametric efficiency of $\mathbf{V}_{S, \mathcal{N}}^{*(n)}$ in the multinormal case follows.

5.3 The cost of unspecified scale

We now are able to compare the asymptotic performances of f_1 -parametrically and f_1 -semiparametrically efficient estimators for shape, and to quantify the corresponding efficiency losses. The main result of this paper is stated in the next theorem; see the appendix for the proof. In order to improve readability, we write $\text{AVar}[S^{(n)}]$ and $\text{ACov}[S^{(n)}, T^{(n)}]$ for the asymptotic variance of $S^{(n)}$ and asymptotic covariance of $S^{(n)}$ and $T^{(n)}$, respectively.

Theorem 5.5 *Let Assumptions 2.1 and 2.2 hold. Then, under $\bigcup_{\theta} \bigcup_{\sigma_S} \{P_{\theta, \sigma_S^2, \mathbf{V}_S, f_1}^{(n)}\}$, for any f_1 -parametrically efficient (resp., f_1 -semiparametrically efficient) estimator $\mathbf{V}_S^{(n)}$ of \mathbf{V}_S ,*

$$\begin{aligned} & \text{ACov}[n^{1/2}(\mathbf{V}_S^{(n)} - \mathbf{V}_S)_{ij}, n^{1/2}(\mathbf{V}_S^{(n)} - \mathbf{V}_S)_{rs}] \\ &= \frac{k(k+2)}{\mathcal{J}_k(f_1)} \left\{ \left((\mathbf{V}_S)_{ir}(\mathbf{V}_S)_{js} + (\mathbf{V}_S)_{is}(\mathbf{V}_S)_{jr} - \frac{2}{\mathcal{E}_k^{\mathbf{V}_S}} (\mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S)_{ij} (\mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S)_{rs} \right) \right. \\ & \quad \left. + 2\mathcal{G}_{k, f_1}^{\mathbf{V}_S} \left(\mathbf{V}_S - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} \mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S \right)_{ij} \left(\mathbf{V}_S - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} \mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S \right)_{rs} \right\}, \end{aligned} \tag{5.11}$$

with $\mathcal{G}_{k, f_1}^{\mathbf{V}_S} = \mathcal{M}_{k, f_1}^{\mathbf{V}_S}$ (resp., $\mathcal{G}_{k, f_1}^{\mathbf{V}_S} = \mathcal{E}_k^{\mathbf{V}_S}$).

A measure of the efficiency loss due to an unspecified scale when estimating the shape is provided by the differences between the semiparametrically and parametrically optimal values in (5.11), namely

$$\frac{2k(k+2)}{\mathcal{J}_k(f_1)} \left(\mathcal{E}_k^{\mathbf{V}_S} - \mathcal{M}_{k, f_1}^{\mathbf{V}_S} \right) \left(\mathbf{V}_S - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} \mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S \right)_{ij} \left(\mathbf{V}_S - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} \mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S \right)_{rs}. \tag{5.12}$$

For $S(\Sigma) = |\Sigma|^{1/k}$, $\mathcal{E}_k^{\mathbf{V}_S} \mathbf{V}_S = \mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S$ for all \mathbf{V}_S , so that no loss of efficiency is incurred: again, the k -th root of the determinant is the only scale functional for which this holds uniformly in \mathbf{V}_S . For $S(\Sigma) = (\text{tr } \Sigma)/k$, no loss is encountered at $\mathbf{V}_S = \mathbf{I}_k$ —a result which is the estimation counterpart of the result on testing for sphericity.

An alternative way of measuring efficiency losses is in terms of AREs, that is, in terms of ratios rather than differences. Entrywise ratios of asymptotic covariance matrices however do not make much sense. Therefore, we rather consider the performances achieved

in the estimation of linear functionals of the \mathbf{V}_S 's. To be more specific, assume we want to estimate $(\text{vec } \mathbf{v})'(\text{vec } \mathbf{V}_S)$, where \mathbf{v} is some fixed symmetric $k \times k$ matrix. Denote by $\mathbf{V}_S^{(n)}$ and $\mathbf{V}_S^{*(n)}$, respectively, some f_1 -parametrically and f_1 -semiparametrically efficient estimators of shape. It readily follows from Propositions 5.3 and 5.4 that the asymptotic variances of $n^{1/2}(\text{vec } \mathbf{v})'(\text{vec } \mathbf{V}_S^{(n)})$ and $n^{1/2}(\text{vec } \mathbf{v})'(\text{vec } \mathbf{V}_S^{*(n)})$ are

$$\frac{k(k+2)}{\mathcal{J}_k(f_1)} (\text{vec } \mathbf{v})' \mathbf{Q}_{k;1,2\mathcal{M}_{k,f_1}^{\mathbf{V}_S}} (\text{vec } \mathbf{v}) \quad \text{and} \quad \frac{k(k+2)}{\mathcal{J}_k(f_1)} (\text{vec } \mathbf{v})' \mathbf{Q}_{k;1,2\mathcal{E}_k^{\mathbf{V}_S}} (\text{vec } \mathbf{v}),$$

respectively, so that

$$\frac{(\text{vec } \mathbf{v})' \mathbf{Q}_{k;1,2\mathcal{M}_{k,f_1}^{\mathbf{V}_S}} (\text{vec } \mathbf{v})}{(\text{vec } \mathbf{v})' \mathbf{Q}_{k;1,2\mathcal{E}_k^{\mathbf{V}_S}} (\text{vec } \mathbf{v})} = 1 - \frac{(\text{vec } \mathbf{v})' \left(\mathbf{Q}_{k;1,2\mathcal{E}_k^{\mathbf{V}_S}} - \mathbf{Q}_{k;1,2\mathcal{M}_{k,f_1}^{\mathbf{V}_S}} \right) (\text{vec } \mathbf{v})}{(\text{vec } \mathbf{v})' \mathbf{Q}_{k;1,2\mathcal{E}_k^{\mathbf{V}_S}} (\text{vec } \mathbf{v})}$$

can be considered as a measure of asymptotic relative efficiency. Since

$$\begin{aligned} 0 < \mathcal{E}_k^{\mathbf{V}_S} - \mathcal{M}_{k,f_1}^{\mathbf{V}_S} \\ &= \frac{(k+2)(\mathcal{J}_k(f_1) - k^2) (\mathcal{E}_k^{\mathbf{V}_S})^2}{(k+2)(\mathcal{J}_k(f_1) - k^2) \mathcal{E}_k^{\mathbf{V}_S} - (\mathcal{J}_k(f_1) - k(k+2))} \leq \frac{(k+2) (\mathcal{E}_k^{\mathbf{V}_S})^2}{(k+2) \mathcal{E}_k^{\mathbf{V}_S} - 1}, \end{aligned}$$

the efficiency loss vanishes iff the Fisher information for scale/shape $\mathcal{J}_k(f_1)$ tends to its infimum k^2 , which occurs at arbitrarily heavy tails, that is, as $\nu \rightarrow 0$ and $\eta \rightarrow 0$ in the classes of k -variate Student and power-exponential distributions, respectively. On the contrary, the efficiency loss is maximal when $\mathcal{J}_k(f_1)$ goes to infinity, that is under arbitrarily light tails, namely, as $\eta \rightarrow \infty$ in the classes of k -variate power-exponential distributions.

5.4 Comparisons between the various S -efficiencies

Similarly as in Section 4.2, we now proceed to show that semiparametric efficiencies do not depend on the scale functional S . To be more specific, Proposition 5.4 shows that, if $\mathbf{V}_S^{*(n)}$ is a f_1 -semiparametrically efficient estimators of \mathbf{V}_S ,

$$n^{1/2} \text{vec} \left(\mathbf{V}_S^{*(n)} - \mathbf{V}_S \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}, \frac{k(k+2)}{\mathcal{J}_k(f_1)} \mathbf{Q}_{k;1,2\mathcal{E}_k^{\mathbf{V}_S}} \right),$$

as $n \rightarrow \infty$, under $\mathbb{P}_{\boldsymbol{\vartheta}_S, f_1}^{(n)}$ (as usual, $\boldsymbol{\vartheta}_S = (\boldsymbol{\theta}', \sigma_S^2, (\text{vech } \mathbf{V}_S)')' \in \Theta_S$). If the scale functional T is used instead of S , the quantity to be estimated under the corresponding distribution in $\mathcal{P}_T^{(n)}$, namely $\mathbb{P}_{\boldsymbol{\vartheta}_T, f_1}^{(n)}$, where

$$\boldsymbol{\vartheta}_T = \left(\boldsymbol{\theta}', \sigma_T^2, (\text{vech } \mathbf{V}_T)' \right)' = \left(\boldsymbol{\theta}', T(\mathbf{V}_S), (\text{vech } \mathbf{V}_S)' / T(\mathbf{V}_S) \right)' \in \Theta_T,$$

is clearly $\mathbf{V}_T = \mathbf{V}_S/T(\mathbf{V}_S)$. An obvious estimator for \mathbf{V}_T is $\mathbf{V}_T^{*(n)} = \mathbf{V}_S^{*(n)}/T(\mathbf{V}_S^{*(n)})$. Using Slutsky’s Lemma, we obtain that

$$\begin{aligned} & n^{1/2}\text{vec}\left(\mathbf{V}_T^{*(n)} - \mathbf{V}_T\right) \\ &= \frac{1}{T(\mathbf{V}_S)} \left[\mathbf{I}_{k^2} - (\text{vec}\mathbf{V}_T)(\text{vec}\mathbf{D}_T^{\mathbf{V}_S})' \right] \left[n^{1/2}\text{vec}\left(\mathbf{V}_S^{*(n)} - \mathbf{V}_S\right) \right] + o_{\mathbb{P}}(1) \\ &\xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \frac{k(k+2)}{\mathcal{J}_k(f_1)(T(\mathbf{V}_S))^2} \left[\mathbf{I}_{k^2} - (\text{vec}\mathbf{V}_T)(\text{vec}\mathbf{D}_T^{\mathbf{V}_S})' \right] \right. \\ &\quad \left. \times \mathbf{Q}_{k;1,2\varepsilon_k}^{\mathbf{V}_S} \left[\mathbf{I}_{k^2} - (\text{vec}\mathbf{V}_T)(\text{vec}\mathbf{D}_T^{\mathbf{V}_S})' \right]' \right), \end{aligned} \tag{5.13}$$

under $\mathbb{P}_{\vartheta_T, f_1}^{(n)}$, as $n \rightarrow \infty$. Painful, though standard, computations show that the asymptotic covariance matrix in (5.13) reduces to $\frac{k(k+2)}{\mathcal{J}_k(f_1)} \mathbf{Q}_{k;1,2\varepsilon_k}^{\mathbf{V}_T}$, which shows (Proposition 5.4 again) that $\mathbf{V}_T^{*(n)} = \mathbf{V}_S^{*(n)}/T(\mathbf{V}_S^{*(n)})$ is f_1 -semiparametrically efficient for \mathbf{V}_T at $\mathbb{P}_{\vartheta_T, f_1}^{(n)}$, as $n \rightarrow \infty$. This clearly is an evidence that semiparametric efficiency bounds do not depend on the choice of the scale functional.

A Appendix

In this final section, we prove Lemmas 5.1 and 5.2, as well as Theorems 4.1, 4.2, and 5.5.

For any S satisfying Assumption 2.1, consider the mapping $V_{11}^S : \text{vec } \mathcal{S}^k \rightarrow \mathbb{R}$ defined by $S((V_{11}^S(\text{vec } \mathbf{V}), (\text{vec } \mathbf{V})')) = 1$, the existence of which—locally around any $\mathbf{V}_S \in \mathcal{V}_k^S$ —is guaranteed by Assumption 2.1 and the implicit function Theorem. We will need the following result, which states some important properties of $\mathbf{M}_S^{\mathbf{V}_S}$ and $\mathbf{D}_S^{\mathbf{V}_S}$.

Lemma A.1 *Let S satisfy Assumption 2.1 and fix $\mathbf{V}_S \in \mathcal{V}_k^S$. Let also \mathbf{P}_k be the $(K+1) \times k^2$ matrix such that $\mathbf{P}'_k(\text{vec } \mathbf{v}) = \text{vec } \mathbf{v}$ for any symmetric $k \times k$ matrix \mathbf{v} . Then,*

- (i) $\mathbf{M}_S^{\mathbf{V}_S} = (\nabla V_{11}^S(\text{vec } \mathbf{V}_S) : \mathbf{I}_K) \mathbf{P}_k$;
- (ii) $\mathbf{D}_S^{\mathbf{V}_S} = \mathbf{D}_S^{\lambda \mathbf{V}_S}$ for all $\lambda > 0$;
- (iii) $\text{tr}(\mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S) = S(\mathbf{V}_S) = 1$;
- (iv) denoting by $\ker \mathbf{A}$ the null space of a matrix \mathbf{A} , $(\ker \mathbf{M}_S^{\mathbf{V}_S}) \cap (\text{vec } \mathcal{S}_k) = \{\lambda(\text{vec } \mathbf{D}_S^{\mathbf{V}_S}) : \lambda \in \mathbb{R}\}$;
- (v) $\varepsilon_k^{\mathbf{V}_S} \geq 1/k$.

Proof of Lemma A.1: Parts (i)–(iv) are proved in Lemmas 4.1 and 4.2 of [15]. Let $\lambda(\mathbf{V}_S) = (\lambda_1(\mathbf{V}_S), \dots, \lambda_k(\mathbf{V}_S))'$, where the $\lambda_i(\mathbf{V}_S)$ ’s are the eigenvalues of the symmetric matrix $\mathbf{V}_S^{1/2} \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S^{1/2}$ (the ordering of components in $\lambda(\mathbf{V}_S)$ is irrelevant in the sequel). Part (iii) of the lemma shows that $\mathbf{1}'_k \lambda(\mathbf{V}_S) = \text{tr}(\mathbf{V}_S^{1/2} \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S^{1/2}) = \text{tr}(\mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S) = 1$, where $\mathbf{1}_k := (1, \dots, 1)' \in \mathbb{R}^k$.

Part (v) follows since the Cauchy–Schwarz inequality yields $1 = (\mathbf{1}'\boldsymbol{\lambda}(\mathbf{V}_S))^2 \leq k (\boldsymbol{\lambda}(\mathbf{V}_S))'\boldsymbol{\lambda}(\mathbf{V}_S) = k \times \text{tr}((\mathbf{V}_S^{1/2}\mathbf{D}_S^{\mathbf{V}_S}\mathbf{V}_S^{1/2})^2) = k \mathcal{E}_k^{\mathbf{V}_S}$. \square

Proof of Theorem 4.1: The noncentrality parameters $\boldsymbol{\tau}'\boldsymbol{\Gamma}_{S;f_1}(\mathbf{V}_S^0)\boldsymbol{\tau}$ and $\boldsymbol{\tau}'\boldsymbol{\Gamma}_{S;f_1}^*(\mathbf{V}_S^0)\boldsymbol{\tau}$ to be computed are given by

$$(\text{vech } \mathbf{v})'\mathbf{M}_S^{\mathbf{V}_S^0}((\mathbf{V}_S^0))^{-1/2} \left[a(\mathbf{I}_{k^2} + \mathbf{K}_k) + b\mathbf{J}_k \right] ((\mathbf{V}_S^0))^{-1/2} \left(\mathbf{M}_S^{\mathbf{V}_S^0} \right)' (\text{vech } \mathbf{v}), \quad (\text{A.1})$$

with $(a, b) = \left(\frac{\mathcal{J}_k(f_1)}{4k(k+2)}, \frac{\mathcal{J}_k(f_1)}{4k(k+2)} - \frac{1}{4} \right)$ and $(a, b) = \left(\frac{\mathcal{J}_k(f_1)}{4k(k+2)}, -\frac{\mathcal{J}_k(f_1)}{2k^2(k+2)} \right)$, respectively.

Now, from (4.3) and the definition of $\mathbf{M}_S^{\mathbf{V}_S^0}$, note that $(\mathbf{M}_S^{\mathbf{V}_S^0})'(\text{vech } \mathbf{v}) = (\text{vec } \mathbf{v})$. Hence, by using that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})(\text{vec } \mathbf{B})$, $(\text{vec } \mathbf{A})'(\text{vec } \mathbf{B}) = \text{tr}(\mathbf{A}'\mathbf{B})$, and $\mathbf{K}_k(\text{vec } \mathbf{A}) = (\text{vec } \mathbf{A}')$, equation (A.1) can easily be rewritten as

$$\begin{aligned} & \left(\text{vec} \left((\mathbf{V}_S^0)^{-1/2} \mathbf{v} (\mathbf{V}_S^0)^{-1/2} \right) \right)' \left[a(\mathbf{I}_{k^2} + \mathbf{K}_k) + b\mathbf{J}_k \right] \left(\text{vec} \left((\mathbf{V}_S^0)^{-1/2} \mathbf{v} (\mathbf{V}_S^0)^{-1/2} \right) \right) \\ & = 2a \left(\text{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) \right)^2 + b \left(\text{tr} \left((\mathbf{V}_S^0)^{-1} \mathbf{v} \right) \right)^2, \end{aligned}$$

which establishes the result. \square

Proof of Theorem 4.2: (i) Clearly, the null hypothesis $\mathcal{H}_0 : \mathbf{V}_S = \mathbf{V}_S^0$ translates, in the T -parametrization, into $\mathcal{H}_0 : \mathbf{V}_T = \mathbf{V}_T^0$, where $\mathbf{V}_T^0 = \mathbf{V}_S^0/T(\mathbf{V}_S^0)$. Now, considering local alternatives,

$$\begin{aligned} & \text{vech} \left(\frac{\mathbf{V}_S^0 + n^{-1/2}\mathbf{v}_S^{(n)}}{T(\mathbf{V}_S^0 + n^{-1/2}\mathbf{v}_S^{(n)})} \right) - \text{vech}(\mathbf{V}_T^0) \\ & = \text{vech} \left(\frac{\mathbf{V}_S^0 + n^{-1/2}\mathbf{v}_S^{(n)} - \mathbf{V}_S^0}{T(\mathbf{V}_S^0 + n^{-1/2}\mathbf{v}_S^{(n)})} + \left[\frac{1}{T(\mathbf{V}_S^0 + n^{-1/2}\mathbf{v}_S^{(n)})} - \frac{1}{T(\mathbf{V}_S^0)} \right] \mathbf{V}_S^0 \right) \\ & = \frac{n^{-1/2}}{T(\mathbf{V}_S^0)} \text{vech} \left(\mathbf{v}_S^{(n)} - \frac{n^{1/2}}{T(\mathbf{V}_S^0)} \left(T(\mathbf{V}_S^0 + n^{-1/2}\mathbf{v}_S^{(n)}) - T(\mathbf{V}_S^0) \right) \mathbf{V}_S^0 \right) + o(n^{-1/2}). \end{aligned} \quad (\text{A.2})$$

Since Lemma A.1(ii) yields

$$\begin{aligned} n^{1/2} \left(T(\mathbf{V}_S^0 + n^{-1/2}\mathbf{v}_S^{(n)}) - T(\mathbf{V}_S^0) \right) & = \left[\nabla T(\text{vech } \mathbf{V}_S^0) \right]' (\text{vech } \mathbf{v}_S^{(n)}) + o(1) \\ & = \text{tr} \left(\mathbf{D}_T^{\mathbf{V}_S^0} \mathbf{v}_S^{(n)} \right) + o(1) = \text{tr} \left(\mathbf{D}_T^{\mathbf{V}_T^0} \mathbf{v}_S \right) + o(1) \end{aligned}$$

as $n \rightarrow \infty$, equality (4.9) follows from (4.8) and (A.2). Note that Lemma A.1(iii) entails that $\text{tr}(\mathbf{D}_T^{\mathbf{V}_T^0} \mathbf{v}_T) = 0$.

(ii) The last remark in the proof of (i) and the definition of $\mathbf{M}_T^{\mathbf{V}_T^0}$ imply that the noncentrality parameter $\boldsymbol{\tau}'_T \boldsymbol{\Gamma}_{T; f_1}^* (\mathbf{V}_T^0) \boldsymbol{\tau}_T$ is given by

$$\begin{aligned} & \frac{\mathcal{J}_k(f_1)}{4k(k+2)} (\text{vech } \mathbf{v}_T)' \mathbf{M}_T^{\mathbf{V}_T^0} ((\mathbf{V}_T^0))^{-1/2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k \right] \\ & \quad \times ((\mathbf{V}_T^0))^{-1/2} \left(\mathbf{M}_T^{\mathbf{V}_T^0} \right)' (\text{vech } \mathbf{v}_T) \\ & = \frac{\mathcal{J}_k(f_1)}{4k(k+2)} (\text{vec } \mathbf{v}_T)' ((\mathbf{V}_T^0))^{-1/2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k \right] ((\mathbf{V}_T^0))^{-1/2} (\text{vec } \mathbf{v}_T). \end{aligned} \tag{A.3}$$

Hence, by using (4.9), the identity $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})(\text{vec } \mathbf{B})$, and the fact that $[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k](\text{vec } \mathbf{I}_k) = \mathbf{0}$, this noncentrality parameter can be written as

$$\begin{aligned} & \frac{\mathcal{J}_k(f_1)}{4k(k+2)} \frac{1}{(T(\mathbf{V}_S^0))^2} (\text{vec } \mathbf{v}_S)' ((\mathbf{V}_T^0))^{-1/2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k \right] ((\mathbf{V}_T^0))^{-1/2} (\text{vec } \mathbf{v}_S) \\ & = \frac{\mathcal{J}_k(f_1)}{4k(k+2)} (\text{vec } \mathbf{v}_S)' ((\mathbf{V}_S^0))^{-1/2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k \right] ((\mathbf{V}_S^0))^{-1/2} (\text{vec } \mathbf{v}_S) \\ & = \frac{\mathcal{J}_k(f_1)}{4k(k+2)} (\text{vec } \mathbf{v}_S)' \mathbf{M}_S^{\mathbf{V}_S^0} ((\mathbf{V}_S^0))^{-1/2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k} \mathbf{J}_k \right] \\ & \quad \times ((\mathbf{V}_S^0))^{-1/2} \left(\mathbf{M}_S^{\mathbf{V}_S^0} \right)' (\text{vec } \mathbf{v}_S), \end{aligned} \tag{A.4}$$

which is the noncentrality parameter $\boldsymbol{\tau}'_S \boldsymbol{\Gamma}_{S; f_1}^* (\mathbf{V}_S^0) \boldsymbol{\tau}_S$ associated with the S -parametrization (in order to obtain (A.4), we used the definition of $\mathbf{M}_S^{\mathbf{V}_S^0}$ and the fact that $\text{tr}(\mathbf{D}_S^{\mathbf{V}_S^0} \mathbf{v}_S) = 0$). □

Proof of Lemma 5.1: The delta method yields, under $\mathbf{P}_{\boldsymbol{\theta}, f_1}^{(n)}$, as $n \rightarrow \infty$,

$$\begin{aligned} & n^{1/2} \text{vec}(\mathbf{V}_S^{(n)} - \mathbf{V}_S) \\ & = n^{1/2} \mathbf{P}'_k \text{vech}(\mathbf{V}_S^{(n)} - \mathbf{V}_S) \\ & = n^{1/2} \mathbf{P}'_k (V_{11}^S (\text{vech } \mathbf{V}_S^{(n)}) - V_{11}^S (\text{vech } \mathbf{V}_S), (\text{vech } (\mathbf{V}_S^{(n)} - \mathbf{V}_S)))' \\ & = n^{1/2} \mathbf{P}'_k \left(\left[\nabla V_{11}^S (\text{vech } \mathbf{V}_S) \right]'_{\mathbf{I}_K} \right) \text{vech}(\mathbf{V}_S^{(n)} - \mathbf{V}_S) + o_P(1). \end{aligned}$$

The result then follows from Part (i) of Lemma A.1. □

Proof of Lemma 5.2: (i) Write $\Gamma_{a,b}$ for the matrix in braces in the left-hand side of (5.5). Using part (ii) of the lemma (as we shall see, the proof of (ii) does not require (i)) and the identities $\mathbf{K}_k^2 = \mathbf{I}_{k^2}$, $\mathbf{K}_k \mathbf{J}_k = \mathbf{K}_k = \mathbf{J}_k \mathbf{K}_k$, $\mathbf{J}_k^2 = k \mathbf{J}_k$, $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})(\text{vec } \mathbf{B})$, $(\text{vec } \mathbf{A})'(\text{vec } \mathbf{B}) = \text{tr}(\mathbf{A}'\mathbf{B})$, $\mathbf{K}_k(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}_k$, $\mathbf{K}_k(\text{vec } \mathbf{A}) = \text{vec}(\mathbf{A}')$ (holding for all $k \times k$ matrices \mathbf{A} , \mathbf{B} , and \mathbf{C}), lengthy but straightforward calculations yield

$$\begin{aligned} \Gamma_{a,b}(\mathbf{N}_k \mathbf{Q}_{k;r,s}^{\mathbf{V}_S} \mathbf{N}_k') &= \frac{ar}{2} \mathbf{M}_k [\mathbf{I}_{k^2} + \mathbf{K}_k] \mathbf{N}_k' + \frac{1}{4} \left(2as + 2br + bks - \frac{bs}{\mathcal{E}_k^{\mathbf{V}_S}} \right) \\ &\quad \times \mathbf{M}_S^{\mathbf{V}_S} (\text{vec } \mathbf{V}_S^{-1}) \left[\text{vec} \left(\mathbf{V}_S^{-1} - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} \mathbf{D}_S^{\mathbf{V}_S} \right) \right]' (\mathbf{V}_S^{\otimes 2}) \mathbf{N}_k'. \end{aligned}$$

Now, since

- (a) $\mathbf{M}_S^{\mathbf{V}_S} (\text{vec } \mathbf{D}_S^{\mathbf{V}_S}) = \mathbf{0}$ (see Lemma A.1(iv)),
- (b) $\mathbf{M}_S^{\mathbf{V}_S} \mathbf{K}_k = \mathbf{M}_S^{\mathbf{V}_S}$ (since, for all symmetric $k \times k$ matrix \mathbf{w} , $(\text{vec } \mathbf{D}_S^{\mathbf{V}_S})'(\text{vec } \mathbf{w}) = 0$ implies that $\mathbf{K}_k(\mathbf{M}_S^{\mathbf{V}_S})'(\text{vec } \mathbf{w}) = \mathbf{K}_k(\text{vec } \mathbf{w}) = \text{vec } \mathbf{w} = (\mathbf{M}_S^{\mathbf{V}_S})'(\text{vec } \mathbf{w})$), and
- (c) $\mathbf{M}_S^{\mathbf{V}_S} \mathbf{N}_k' = \mathbf{I}_k$ (since $\mathbf{N}_k(\mathbf{M}_S^{\mathbf{V}_S})'(\text{vec } \mathbf{w}) = \mathbf{N}_k(\text{vec } \mathbf{w}) = \text{vec } \mathbf{w}$ for all symmetric $k \times k$ matrix \mathbf{w} such that $(\text{vec } \mathbf{D}_S^{\mathbf{V}_S})'(\text{vec } \mathbf{w}) = 0$),

we obtain that $\mathbf{N}_k \mathbf{Q}_{k;A,B}^{\mathbf{V}_S} \mathbf{N}_k'$ is a right-inverse of $\Gamma_{a,b}$. Since both $\mathbf{N}_k \mathbf{Q}_{k;A,B}^{\mathbf{V}_S} \mathbf{N}_k'$ and $\Gamma_{a,b}$ are symmetric, this right-inverse is also a left-inverse.

(ii) Since $\mathbf{Q}_{k;r,s}^{\mathbf{V}_S}$ is symmetric, it is clearly sufficient to prove that $\mathbf{M}_k' \mathbf{N}_k \mathbf{Q}_{k;r,s}^{\mathbf{V}_S} = \mathbf{Q}_{k;r,s}^{\mathbf{V}_S}$. It follows from the definition of $\mathbf{Q}_{k;r,s}^{\mathbf{V}_S}$ (see (5.3)) that it is sufficient to show that

$$\mathbf{M}_k' \mathbf{N}_k \mathbf{P}_k^{\mathbf{V}_S} [\mathbf{I}_{k^2} + \mathbf{K}_k] = \mathbf{P}_k^{\mathbf{V}_S} [\mathbf{I}_{k^2} + \mathbf{K}_k] \quad (\text{A.5})$$

and

$$\mathbf{M}_k' \mathbf{N}_k \mathbf{P}_k^{\mathbf{V}_S} (\text{vec } \mathbf{V}_S) = \mathbf{P}_k^{\mathbf{V}_S} (\text{vec } \mathbf{V}_S). \quad (\text{A.6})$$

But, letting $\mathbf{E}_{ij} := \mathbf{e}_i \mathbf{e}_j' + \mathbf{e}_j \mathbf{e}_i'$ and using the fact that $\mathbf{D}_S^{\mathbf{V}_S} = \frac{1}{2} \sum_{i,j=1}^k (\mathbf{D}_S^{\mathbf{V}_S})_{ij} \mathbf{E}_{ij}$, we have

$$\begin{aligned} \mathbf{P}_k^{\mathbf{V}_S} [\mathbf{I}_{k^2} + \mathbf{K}_k] &= \mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{\mathcal{E}_k^{\mathbf{V}_S}} (\mathbf{V}_S^{\otimes 2}) (\text{vec } \mathbf{D}_S^{\mathbf{V}_S}) (\text{vec } \mathbf{D}_S^{\mathbf{V}_S})' \\ &= \frac{1}{2} \sum_{i,j=1}^k (\text{vec } \mathbf{E}_{ij}) (\text{vec } \mathbf{E}_{ij})' - \frac{2}{\mathcal{E}_k^{\mathbf{V}_S}} (\text{vec} (\mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S)) (\text{vec } \mathbf{D}_S^{\mathbf{V}_S})' \\ &= \sum_{i,j=1}^k \text{vec} \left(\frac{1}{2} \mathbf{E}_{ij} - \frac{(\mathbf{D}_S^{\mathbf{V}_S})_{ij}}{\mathcal{E}_k^{\mathbf{V}_S}} \mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S \right) (\text{vec } \mathbf{E}_{ij})', \end{aligned}$$

where

$$\begin{aligned} & \operatorname{tr} \left(\frac{1}{2} \mathbf{D}_S^{\mathbf{V}_S} \mathbf{E}_{ij} - \frac{(\mathbf{D}_S^{\mathbf{V}_S})_{ij}}{\mathcal{E}_k^{\mathbf{V}_S}} (\mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S)^2 \right) \\ &= \frac{1}{2} \operatorname{tr} (\mathbf{e}'_j \mathbf{D}_S^{\mathbf{V}_S} \mathbf{e}_i) + \frac{1}{2} \operatorname{tr} (\mathbf{e}'_i \mathbf{D}_S^{\mathbf{V}_S} \mathbf{e}_j) - (\mathbf{D}_S^{\mathbf{V}_S})_{ij} = 0. \end{aligned}$$

This establishes (A.5) since $\mathbf{M}'_k \mathbf{N}_k (\operatorname{vec} \mathbf{w}) = (\operatorname{vec} \mathbf{w})$ for all symmetric $k \times k$ matrix \mathbf{w} satisfying $\operatorname{tr}(\mathbf{D}_S^{\mathbf{V}_S} \mathbf{w}) = 0$. As for (A.6), it follows in a similar way from noting that

$$\begin{aligned} \mathbf{P}_k^{\mathbf{V}_S} (\operatorname{vec} \mathbf{V}_S) &= (\operatorname{vec} \mathbf{V}_S) - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} (\mathbf{V}_S^{\otimes 2}) (\operatorname{vec} \mathbf{D}_S^{\mathbf{V}_S}) \\ &= \operatorname{vec} \left(\mathbf{V}_S - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} \mathbf{V}_S \mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S \right), \end{aligned}$$

where $\operatorname{tr} \left(\mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S - \frac{1}{\mathcal{E}_k^{\mathbf{V}_S}} (\mathbf{D}_S^{\mathbf{V}_S} \mathbf{V}_S)^2 \right) = 0$. □

Proof of Theorem 5.5: Clearly, if $n^{1/2} \operatorname{vec} (\mathbf{V}_S^{(n)} - \mathbf{V}_S)$ is asymptotically multinormal with mean zero and covariance matrix \mathbf{H} , one has

$$\operatorname{ACov} [n^{1/2} (\mathbf{V}_S^{(n)} - \mathbf{V}_S)_{ij}, n^{1/2} (\mathbf{V}_S^{(n)} - \mathbf{V}_S)_{rs}] = (\operatorname{vec} (\mathbf{e}_i \mathbf{e}'_j))' \mathbf{H} \operatorname{vec} (\mathbf{e}_r \mathbf{e}'_s).$$

The result then follows from standard computations by using Propositions 5.3 and 5.4. □

Acknowledgements This research was supported by a P.A.I. contract of the Belgian Federal Government and an Action de Recherche Concertée of the Communauté française de Belgique. The authors gratefully acknowledge the comments of two referees and an associate editor, which greatly helped improving the scope and the structure of the paper.

References

- [1] P. J. Bickel. On adaptive estimation. *Annals of Statistics*, 10:647–671, 1982.
- [2] C. Croux and G. Haesbroeck. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87:603–618, 2000.

- [3] L. Dümbgen and D. E. Tyler. On the breakdown properties of some multivariate M -Functionals. *Scandinavian Journal of Statistics*, 32:247–264, 2005.
- [4] L. Dümbgen. On Tyler's M -functional of scatter in high dimension. *Annals of the Institute of Statistical Mathematics*, 50:471–491, 1998.
- [5] M. Hallin, H. Oja, and D. Paindaveine. Semiparametrically efficient rank-based inference for shape: II. Optimal R -estimation of shape. *Annals of Statistics*, to appear, 34, 2006.
- [6] M. Hallin and D. Paindaveine. Semiparametrically efficient rank-based inference for shape: I. Optimal rank-based tests for sphericity. *Annals of Statistics*, to appear, 34, 2006.
- [7] M. Hallin and D. Paindaveine. Optimal rank-based tests for homogeneity of scatter. Submitted, 2006.
- [8] M. Hallin and B. J. M. Werker. Semiparametric efficiency, distribution-freeness, and invariance. *Bernoulli*, 9:55–65, 2003.
- [9] T. P. Hettmansperger and R. H. Randles. A practical affine equivariant multivariate median. *Biometrika*, 89:851–860, 2002.
- [10] L. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York, 1986.
- [11] H. P. Lopuhaä. On the relation between S -estimators and M -estimators of multivariate location and covariance. *Annals of Statistics*, 17:1662–1683, 1989.
- [12] E. Ollila, C. Croux, and H. Oja. Influence function and asymptotic efficiency of the affine equivariant rank covariance matrix. *Statistica Sinica*, 14:297–316, 2004.
- [13] E. Ollila, T. P. Hettmansperger, and H. Oja. Affine equivariant multivariate sign methods. Preprint, University of Jyväskylä, 2004.
- [14] E. Ollila, H. Oja, and C. Croux. The affine equivariant sign covariance matrix: Asymptotic behavior and efficiencies. *Journal of Multivariate Analysis*, 87:328–355, 2003.
- [15] D. Paindaveine. A canonical definition of shape. Submitted, 2006.
- [16] R. H. Randles. A simpler, affine-invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Association*, 95:1263–1268, 2000.
- [17] M. Salibián-Barrera, S. Van Aelst, and G. Willems. Principal components analysis based on multivariate MM -estimators with fast and robust bootstrap. *Journal of the American Statistical Association*, 101:1198–1211, 2006.
- [18] S. Taskinen, C. Croux, A. Kankainen, E. Ollila, and H. Oja. Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *Journal of Multivariate Analysis*, 97:359–384, 2006.

- [19] K. S. Tatsuoka and D. E. Tyler. On the uniqueness of S-Functionals and M-functionals under nonelliptical distributions. *Annals of Statistics*, 28:1219–1243, 2000.
- [20] D. E. Tyler. A distribution-free M-estimator of multivariate scatter. *Annals of Statistics*, 15:234–251, 1987.

Marc Hallin
Institut de Recherche en Statistique,
E.C.A.R.E.S, and
and Département de Mathématique,
Université Libre de Bruxelles
Campus de la Plaine CP 210
1050 Bruxelles
Belgium
mhallin@ulb.ac.be

Davy Paindaveine
Institut de Recherche en Statistique,
E.C.A.R.E.S, and
Département de Mathématique,
Université Libre de Bruxelles
Campus de la Plaine CP 210
1050 Bruxelles
Belgium
dpaindav@ulb.ac.be