T E C H N I C A L
R E P O R T

0430

OPTIMAL R-ESTIMATION OF SHAPE

HALLIN, M., OJA, H. and D. PAINDAVEINE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Optimal R-Estimation of Shape

Marc Hallin[*]
Université Libre de Bruxelles
Brussels, Belgium

Hannu Oja
University of Jyvåskylå
Finland

and

Davy Paindaveine[*]
Université Libre de Bruxelles
Brussels, Belgium

Abstract

A class of R-estimators, based on the concepts of multivariate signed ranks and the optimal rank-based tests developed in Hallin and Paindaveine (2004b), is proposed for the estimation of the shape matrix of an elliptical distribution. These R-estimators are root-n consistent under any radial density g, without any moment assumptions, and semiparametrically efficient at some prespecified density f. When based on normal scores, they are uniformly more efficient than the traditional normal-theory estimator, based on empirical covariance matrices (the asymptotic normality of which moreover requires finite moments of order four), irrespective of the actual underlying elliptical density. They rely on an original rank-based version of Le Cam's one-step methodology, which avoids the unpleasant nonparametric estimation of cross-information quantities that is generally required in the context of R-estimation. Although they are not strictly equivariant, they are shown to be equivariant in a weak asymptotic sense. Simulations confirm their feasability and excellent finite-sample performances.

AMS 1980 subject classification : 62M15, 62G35.

Key words and phrases : Elliptical densities, Shape matrix, Multivariate ranks and signs, R-estimation, Local asymptotic normality, Semiparametric efficiency, One-step estimation, Affine equivariance.

1 Introduction.

1.1 Rank-based inference for elliptical families.

An elliptical density is determined by a location centre $\boldsymbol{\theta}$, a positive definite symmetric $k \times k$ matrix $\mathbf{V} = (V_{ij})$ with $V_{11} = 1$, the *shape matrix*, and the so-called *radial density* g; for a precise

[*]Research supported by by a P.A.I. contract of the Belgian Federal Government and an Action de Recherche Concertée of the Communauté française de Belgique.

definition, see Section 2. The shape (matrix) $\mathbf{V}$ is a genuinely multivariate concept, which describes the shape and orientation of the distribution. The traditional covariance matrix, if it exists, is proportional to the shape matrix, but the latter does not require any moment assumptions.

Elliptical families have been introduced in multivariate analysis as a reaction against pervasive Gaussian assumptions. Most classical multivariate analysis procedures—principal component, canonical correlations, multivariate regression, etc.—readily extend to elliptical models, with the shape matrix playing the role of traditional covariances or correlations. All these methods crucially depend on an estimation $\widehat{\mathbf{V}}$ of the shape matrix. So does inference on the location parameter $\boldsymbol{\theta}$, or on the parameters of interest in more complex models (multiple output regression, multivariate analysis of variance, VARMA models, etc.) involving elliptical rather than traditional Gaussian noise; see, e.g., Hallin and Paindaveine (2002a, 2002b, 2004a, and 2005). It is reasonable, in most applications, to treat the radial density g as a nuisance. Therefore, it is essential for $\widehat{\mathbf{V}}$ to enjoy good properties under arbitrary g, including the heavy-tailed ones.

The traditional sample covariance matrix in this respect clearly does not qualify, and new estimators have to be considered. One of them is the celebrated Tyler estimator $\mathbf{V}_T$ (Tyler 1987a, b), which is extremely robust, but relies entirely on *multivariate signs*, and thus does not fully exploit the available information. When g is unspecified, invariance arguments suggest that this information is contained in the ranks of radial distances, and that these ranks should allow for improving over $\mathbf{V}_T$.

Similar arguments, in the hypothesis testing context, were used by Hallin and Paindaveine (2004b), who developed a class of optimal signed rank tests for null hypotheses of the form $\mathbf{V} = \mathbf{V}_0$. Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be a random sample from an elliptical distribution with unknown shape matrix $\mathbf{V}$. Hallin and Paindaveine (2004b)'s test statistics are constructed as follows. For a known $\boldsymbol{\theta}$ (which can be replaced by a root-n consistent estimate), denote by $\mathbf{Z}_i := \mathbf{V}_0^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})$ the $\mathbf{V}_0$-standardized observations, by $d_i := \|\mathbf{Z}_i\|$ their lengths, and by $\mathbf{U}_i := \|\mathbf{Z}_i\|^{-1}\mathbf{Z}_i$ their multivariate signs, $i = 1, ..., n$. Let R_i, $i = 1, ..., n$ be the rank of d_i among $d_1, ..., d_n$. For a chosen score function $K_f : (0, 1) \to \mathbb{R}$ (ensuring optimality at some chosen radial density f), consider the matrix-valued signed rank statistic

$$\mathbf{S}_f(\mathbf{V}_0) := \frac{1}{n}\sum_{i=1}^{n} K_f\left(\frac{R_i}{n+1}\right)\mathbf{U}_i\mathbf{U}_i'.$$

The (distribution-free under the null hypothesis) test statistic is then $Q_f := Q(\mathbf{S}_f(\mathbf{V}_0))$, where

$$Q(\mathbf{S}) := \mathrm{tr}(\mathbf{S}^2) - \frac{1}{k}(\mathrm{tr}\,\mathbf{S})^2$$

is k times the variance of the eigenvalues of the positive definite symmetric $k \times k$ matrix $\mathbf{S}$ (another possibility would be the ratio $k|\mathbf{S}|^{1/k}/(\mathrm{tr}\,\mathbf{S})$ of the geometric mean of the eigenvalues to their arithmetic mean).

These test procedures enjoy a series of attractive features: (i) they are valid under arbitrary radial density g, irrespective of any moment assumptions; (ii) they are nevertheless (semiparametrically) efficient at some prespecified radial density f (more specifically, at the class of densities with standardized form f_1; see Section 2.1 for a precise definition); (iii) they exhibit surprisingly high asymptotic relative efficiencies, with respect to classical Gaussian procedures, under non Gaussian g's; and (iv), quite remarkably, when Gaussian (van der Waerden) scores are adopted, their AREs with respect to the classical Gaussian tests (John (1971, 1972)'s test, or equivalently, the extension by Muirhead and Waternaux (1980) of the Mauchly (1940) Gaussian likelihood ratio test) are uniformly larger than one—see Paindaveine (2004) for this extension to shape matrices of the celebrated Chernoff-Savage (1958) result.

All these nice properties, actually, are properties of the noncentrality parameters of the asymptotically noncentral chi-square distributions, under local alternatives, of the rank-based test statistic considered. When the radial density, under such alternatives, is g, these noncentrality parameters depend on a symmetric positive definite matrix, of the form $\mathcal{J}_k(f_1, g_1)\boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})$ (where $\boldsymbol{\Upsilon}_k$ does not depend on f_1 nor g_1) which, for $g_1 = f_1$ coincides with the efficient (at f_1) information matrix for $\mathbf{V}$.

An immediate question is: do such tests have any natural counterparts in the context of point estimation, that is, can we construct estimators $\widehat{\mathbf{V}}^{(n)}$ for the shape matrix that match the performances of those rank-based tests, in the sense of (i) being root-n consistent under any radial density g, irrespective of any moment assumptions—in sharp contrast with the Gaussian estimators, which require finite second-order moments for consistency, and finite fourth-order ones for asymptotic normality; (ii) being nevertheless (semiparametrically) efficient at some prespecified standardized radial density f_1; and (iii) exhibiting the same asymptotic relative efficiencies, with respect to classical Gaussian estimators, including (iv) the Chernoff-Savage property of Paindaveine (2004). Such estimators would improve the performance of the existing ones that satisfy the consistency requirement (i), such as Tyler (1987a)'s celebrated affine-equivariant estimator of shape (*scatter*, in Tyler's terminology), or the estimator of shape based on the *Oja signs* developed in Ollila, Hettmansperger, and Oja (2004). These estimators indeed are root-n consistent under extremely general conditions (finite second-order moments, however, for Ollila et al.), but they are not efficient. Tyler's estimate as well as the associate sign test (Tyler 1987b) actually are efficient in the so-called *angular central Gaussian model*—the model obtained by applying the transformation $\mathbf{x} \mapsto \mathbf{x}/\|\mathbf{x}\|$ to the model considered here, that is, by taking the trace $\mathbf{U}_i$ of our observations $\mathbf{X}_i$ on the unit sphere; see Tyler (1987b) for details.

The answer, as we shall see, is positive, and the estimators achieving the required performances, as expected, are R-estimators, based on the same concepts of multivariate ranks and signs as the test statistics in Hallin and Paindaveine (2004b). Actually, when adequately defined, the R-estimator which is efficient at standardized density f_1 (that is, has an asymptotic covariance matrix, under density f, coinciding with the inverse of the efficient information matrix) has, under radial density g, asymptotic covariance $\mathcal{J}_k^{-1}(f_1, g_1)\boldsymbol{\Upsilon}_k(\mathbf{V})$. This implies, for instance, that the R-estimator corresponding to van der Waerden tests inherits the attractive property of dominating, irrespective of the actual radial density, its Gaussian parametric competitor.

1.2 R-estimation.

The derivation of such R-estimators however is by no means straightforward. Traditional R-estimators are defined (and computed) via the minimization of some rank-based objective function; see Hodges and Lehmann (1963a), Adichie (1967), Jurečková (1971), Koul (1971), or the review paper by Draper (1988). Such "argmin" definitions are numerically costly, especially when the dimension of the parameter is high (in dimension k, a shape matrix has dimension $k(k+1)/2 - 1$), since only numerical implementation is possible (the only example we know of a closed-form is the Hodges-Lehmann, Wilcoxon type, estimator for univariate location and regression). Neither is the "argmin" form convenient from a theoretical point of view, since the objective function, being a function of ranks, is discontinuous, and uniform monotonicity in the vicinity of the parameter is not always an obvious issue; this issue remains unsolved, for instance, in the serial context (estimation of autoregressive parameters, see Allal et al. 2001—note that Koul and Saleh (1993)'s estimator, which escapes this difficulty, is not genuinely rank-based and, as such, does not qualify as an R-estimator).

In the present context, this intuitively appealing approach would consist in defining an R-estimator as the shape matrix $\mathbf{V}$ minimizing the test statistic $Q(\mathbf{S}_f(\mathbf{V}))$. Unfortunately, the uniqueness, consistency and other asymptotic properties of such an estimator are unknown for a general choice of K_f. Another serious inconvenience of R-estimators stems from the fact that estimating their asymptotic variances is another hard problem, even in the simplest cases. Indeed, this variance, under density g, involves a cross-information quantity $\mathcal{J}(g)$ (below taking the form $\mathcal{J}_k(f_1, g_1)$ when the scores are the optimal ones associated with density f; see (3.4)), which depends on the unknown g. Simple consistent estimators of this quantity have been proposed by Lehmann (1963b) and Sen (1966) for one- and two-sample location problems; these estimators are based on comparisons of confidence interval lengths, a method which involves the arbitrary choice of a confidence level $1-\alpha$—a choice which has quite an impact on the final result. Another simple method for constructing consistent estimators of $\mathcal{J}(g)$ can be obtained from the asymptotic linearity property of rank statistics (see Kraft and van Eeden 1972, Antille 1974, or page 321 of Jurečková and Sen 1996 for location and regression; see Section 3.3 of the present paper for a more elaborated version). These "naive" estimators however involve the arbitrary choice of a "small" perturbation of the parameter (generally set to $n^{-1/2}$; but it could be $2n^{-1/2}$ or $3n^{-1/2}$ as well...). Theory again provides no guidelines for this choice, which has a dramatic impact on the output. The resulting estimators thus are likely to be poor. More elaborated approaches rely on a kernel estimate of g—hence cannot be expected to perform well under small and moderate sample sizes. Such kernel methods have been considered, for the Wilcoxon case, by Schweder (1975) (see also Cheng and Serfling 1981, Bickel and Ritov 1988, and Fan 1991) and, in a more general setting, by Koul (2002) (Section 4.5). They also require arbitrary choices (window width and kernel; or, as in Koul 2002, the choice of the order α of an empirical quantile) for which universal recommendations seem hardly possible (see Koul, Sievers, and Mc Kean 1987 for an empirical investigation). Moreover, estimating the actual underlying density is somewhat incompatible with the spirit and basic invariance principles of rank-based inference.

Motivated by these problems, still in the context of univariate regression models, Kraft and van Eeden (1972) have proposed a class of closed form *linearized* R-estimators, which are defined as the sum of a preliminary root-n consistent estimator and an adequate rank-based correction (see also Antille 1974). A severe objection against linearized R-estimators is that the influence of the preliminary estimator in general does not fade away as $n \to \infty$, so that the resulting estimator cannot be considered, even asymptotically, as a genuine R-estimator. There is however one noteworthy exception: when the rank-based correction has a very specific form, proportional to $\mathcal{J}(g)$ or to some consistent estimator thereof, then the influence of the preliminary estimator is $o_{\mathrm{P}}(n^{-1/2})$, and linearized R-estimators can be made asymptotically equivalent to genuine R-estimators. This however brings us back to the difficult problem of estimating $\mathcal{J}(g)$.

These difficulties probably are the reason why R-estimation has never become as popular in applications as rank tests.

1.3 R-estimation of shape.

In the context of elliptical models, and for the multivariate concepts of signs and ranks considered in this paper, we propose an R-estimator that avoids the drawbacks of "classical" univariate

R-estimators we just described. The R-estimators we are proposing are closed-form estimators (in the spirit of Kraft and van Eeden's linearized ones) depending on multivariate ranks and signs only, which achieve the objectives (i)-(iv) mentioned in Section 1.1 without having recourse to kernel-based estimation of such quantities as $\mathcal{J}(g)$ (here, $\mathcal{J}_k(f_1, g_1)$). These estimators are *one-step estimators* in the sense of Le Cam, with Tyler's scatter matrix as a preliminary estimator, and a rank-based version of the *semiparametrically efficient central sequences* derived in Hallin and Paindaveine (2004b) as a linear correction; the influence of the preliminary estimator however disappears as $n \to \infty$. As can be expected, the one-step correction term again involves the cross-information quantity $\mathcal{J}_k(f_1, g_1)$, or at least some consistent estimator thereof. A sophisticated version of the traditional "naive" estimator of this quantity is briefly described, but suffers the same weaknesses as in the traditional setting. We therefore propose another, more elaborated construction which, by fully exploiting the local uniform asymptotic normality (ULAN) structure of the model under study, produces an indirect estimation of $\mathcal{J}_k(f_1, g_1)$.

Summing up, we are breaking with the tradition of an "argmin definition" of R-estimators, and rather use the Le Cam theory of LAN experiments in order to construct an explicit one-step R-estimator entirely based on (multivariate concepts of) ranks and signs, enjoying the robustness (no moment assumptions), optimality, and relative efficiency properties of the corresponding rank tests (listed under (i)-(iv) in Section 1.1), as well as the computational advantages related to the one-step structure.

1.4 Outline of the paper.

The outline of the paper is as follows. First, we need some preparation and notation: in Section 2.1, we recall the main definitions related with elliptical symmetry, location, scale, and shape. Section 2.2 restates the LAN result derived in Hallin and Paindaveine (2004b), and Section 2.3 briefly explains the relation between ranks and signs on one hand, semiparametric efficiency on the other. Section 3 is devoted to the presentation and asymptotic properties of our one-step R-estimators. We start, in Section 3.1, with the rank-based version of semiparametrically efficient central sequences. Then, in Section 3.2, we investigate the asymptotic behavior of a pseudo-estimator involving the unknown cross-information quantity $\mathcal{J}_k(f_1, g_1)$. Finally, we show how this quantity can be estimated consistently, in a naive way first (Section 3.3), but also (Sections 3.4 and 3.5) in a more subtle way, without turning to kernel estimation of g and its derivative, as in the methods that have been proposed before (in the restricted context of univariate Hodges-Lehmann estimation of location). The estimation of the asymptotic covariance of the resulting R-estimator, which is essential for confidence estimation, follows as a by-product. The resulting estimators enjoy all the asymptotic properties expected from R-estimation (see the unusually high AREs figures listed in Table 1), but remain unsatisfactory on one point: for fixed sample size n, they are not affine-equivariant. They are nevertheless equivariant in a *weak asymptotic* sense, as shown in Section 4. A numerical study (Section 5) confirms the excellent performances of the method. An appendix (Section 6) collects technical proofs.

1.5 Notation.

As usual, $\mathrm{vec}\,(\mathbf{A})$ stands for the $k^2 \times 1$ vector resulting from stacking the columns of a $k \times k$ matrix $\mathbf{A}$ on top of each other, and, whenever $\mathbf{A}$ is symmetric, $\mathrm{vech}\,(\mathbf{A})$ for the $k(k+1)/2$-dimensional vector obtained by stacking $\mathbf{A}$'s upper-triangular elements. Now, if $\mathbf{A} = (A_{ij})$ is a shape matrix, A_{11} is automatically one, and we write $\overset{\circ}{\mathrm{vech}}\,(\mathbf{A})$ for the $(k(k+1)/2 - 1)$-dimensional vector obtained by omitting $\mathrm{vech}\,(\mathbf{A})$'s first component. Denoting by $\mathbf{e}_\ell$ the ℓth vector in the canonical basis of $\mathbb{R}^k$ and by $\mathbf{I}_k$ the $k \times k$ unit matrix, let

$$\mathbf{K}_k := \sum_{i,j=1}^{k} (\mathbf{e}_i\mathbf{e}_j') \otimes (\mathbf{e}_j\mathbf{e}_i') \quad \text{and} \quad \mathbf{J}_k := \sum_{i,j=1}^{k} (\mathbf{e}_i\mathbf{e}_j') \otimes (\mathbf{e}_i\mathbf{e}_j') = (\mathrm{vec}\,\mathbf{I}_k)(\mathrm{vec}\,\mathbf{I}_k)';$$

$\mathbf{K}_k$ is the $k^2 \times k^2$ *commutation matrix*. With this notation, note that $\mathbf{K}_k\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\mathbf{A}')$, and $\mathbf{J}_k\mathrm{vec}(\mathbf{A}) = (\mathrm{tr}\,\mathbf{A})(\mathrm{vec}\,\mathbf{I}_k)$. Also note that $(1/k)\mathbf{J}_k$ and $\mathbf{I}_k - (1/k)\mathbf{J}_k$ are the matrices of the mutually orthogonal projections on the subspaces $\{\lambda(\mathrm{vec}\,\mathbf{I}_k) \,|\, \lambda \in \mathbb{R}\}$ and $\{\mathrm{vec}(\mathbf{A}) \,|\, \mathrm{tr}\,\mathbf{A} = 0\}$, respectively. Define $\mathbf{M}_k$ as the $(k(k+1)/2-1) \times k^2$ matrix such that $\mathbf{M}_k'(\overset{\circ}{\mathrm{vech}}(\mathbf{v})) = \mathrm{vec}\,(\mathbf{v})$ for any symmetric $k \times k$ matrix $\mathbf{v} = (v_{ij})$ such that $v_{11} = 0$, and let $\mathbf{N}_k$ be the $(k(k+1)/2-1) \times k^2$ real matrix such that $\mathbf{N}_k(\mathrm{vec}\,\mathbf{v}) = \overset{\circ}{\mathrm{vech}}\,\mathbf{v}$ for any symmetric $k \times k$ matrix $\mathbf{v}$. For any symmetric and positive semi-definite matrix $\mathbf{A}$, we will denote by $\mathbf{A}^{1/2}$ the symmetric root of $\mathbf{A}$; however, any other square root $\mathbf{B}$ of $\mathbf{A}$ (satisfying $\mathbf{B}\mathbf{B}' = \mathbf{A}$) could also be used—provided, of course, it is used in a consistent way. Finally, we write $\mathbf{V}^{\otimes 2}$ for the Kronecker product $\mathbf{V} \otimes \mathbf{V}$.

Throughout the paper, we had to face very serious notational problems. Indeed, for each score function (each radial density f), we have to distinguish between sequences of estimators of $\mathbf{V}$ combining several of the following features: non-rank-based/rank-based, estimators/pseudo-estimators, discretized/nondiscretized, etc. Although in practice we will be interested in non-discretized R-estimators only, the proofs of asymptotic properties unfortunately require various clumsy discretization steps. We deliberately adopted a heavy, but explicit and systematic notation. As a rule, tilde ($\underset{\sim}{\mathbf{V}}$, $\underset{\sim}{\mathbf{\Delta}}$, ...) is used for R-estimators and rank-based quantities; subscript $_\#$ indicates discretization (at some point, further discretization is indicated by $_{\#\!\#}$); hats distinguish estimators from pseudo-estimators, etc. For instance, $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$ stands for a discretized rank-based pseudo-R-estimator with f-scores, $\hat{\underset{\sim}{\mathbf{V}}}^{(n)}_{f\#}$ and $\hat{\underset{\sim}{\mathbf{V}}}^{(n)}_{f}$ for the corresponding discretized and undiscretized estimators, and so on. We apologize for such awkward style (fortunately, it mainly appears in the proofs), but we feel this is the only way we can avoid ambiguous or erroneous statements.

2 Elliptical symmetry.

2.1 Location, scale, and shape.

Denote by $\mathbf{X}^{(n)} := (\mathbf{X}_1^{(n)\prime}, \ldots, \mathbf{X}_n^{(n)\prime})'$, $n \in \mathbb{N}$ a triangular array of k-dimensional observations. Throughout, $\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)}$ are assumed to be i.i.d., with *elliptical density*

$$\underline{f}_{\boldsymbol{\theta},\mathbf{V};f}(\mathbf{x}) := c_{k,f}\,\frac{1}{|\mathbf{V}|^{1/2}}\, f\left(\left((\mathbf{x}-\boldsymbol{\theta})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\theta})\right)^{1/2}\right), \quad \mathbf{x} \in \mathbb{R}^k, \tag{2.1}$$

where $\boldsymbol{\theta} \in \mathbb{R}^k$ is a *location parameter* and $\mathbf{V} := (V_{ij})$, a symmetric positive definite real $k \times k$ matrix with $V_{11} = 1$, is a *shape parameter*. The infinite-dimensional parameter $f : \mathbb{R}_0^+ \longrightarrow \mathbb{R}^+$, the so-called *radial density*, is an a.e. strictly positive function, the constant $c_{k,f}$ a normalization factor depending on the dimension k and f.

It is convenient to rewrite the radial density f into $r \mapsto f(r) := \frac{1}{\sigma} f_1(\frac{r}{\sigma})$, where $\sigma^2 > 0$ and f_1 are the *scale parameter* and *standardized radial density* associated with f, respectively. Note that f and f_1 are quite improperly called *radial densities*. Denote indeed by $d_i^{(n)} = d_i^{(n)}(\boldsymbol{\theta}, \mathbf{V}) := \|\mathbf{Z}_i^{(n)}(\boldsymbol{\theta}, \mathbf{V})\|$ the modulus of the centered and *sphericized* observations $\mathbf{Z}_i^{(n)} = \mathbf{Z}_i^{(n)}(\boldsymbol{\theta}, \mathbf{V}) :=$

$\mathbf{V}^{-1/2}(\mathbf{X}_i^{(n)} - \boldsymbol{\theta})$, $i = 1, \ldots, n$. If the $\mathbf{X}_i^{(n)}$'s have density (2.1), these moduli are i.i.d., with density and distribution function

$$r \mapsto \frac{1}{\sigma} \tilde{f}_{1k}\left(\frac{r}{\sigma}\right) := \frac{1}{\sigma \mu_{k-1;f_1}} \left(\frac{r}{\sigma}\right)^{k-1} f_1\left(\frac{r}{\sigma}\right) I_{[r>0]} \quad \text{and} \quad r \mapsto \tilde{F}_{1k}(r/\sigma) := \int_0^{r/\sigma} \tilde{f}_{1k}(s)\, ds,$$

respectively—provided, however, that

$$\mu_{k-1;f} := \int_0^\infty r^{k-1} f(r)\, dr = \sigma^{k-1} \mu_{k-1;f_1} < \infty, \tag{2.2}$$

an assumption we henceforth always make on f. This function $\tilde{f}_{1k}$ is the *actual* standardized radial density, and (2.2) thus merely ensures that it be a probability density function; in particular, it does not imply any moment restriction on $\tilde{f}_{1k}$, the $d_i^{(n)}$'s, nor the $\mathbf{X}_i^{(n)}$'s.

In order for σ and f_1 to be identifiable, a scale constraint is required. Still in order to avoid moment restrictions, we impose that the $d_i^{(n)}$'s, under (2.1), have common median σ, i.e., that

$$\tilde{F}_{1k}(1) = 1/2, \quad \text{or, equivalently,} \quad (\mu_{k-1;f})^{-1} \int_0^\sigma r^{k-1} f(r) dr = 1/2. \tag{2.3}$$

The k-variate multinormal distribution and k-variate Student distribution with $\nu > 0$ degrees of freedom are associated with radial densities $f_1(r) = \phi_1(r) := \exp(-a_k r^2/2)$ and $f_1(r) = f_{1,\nu}^t(r) := (1 + a_{k,\nu} r^2/\nu)^{-(k+\nu)/2}$, respectively; the constants $a_k > 0$ and $a_{k,\nu} > 0$ are such that (2.3) is satisfied.

Denote by $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};f}$ or $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\mathbf{V};f_1}$ the distribution of $\mathbf{X}^{(n)}$ under parameter value $(\boldsymbol{\theta}', (\mathrm{v\mathring{e}ch}\mathbf{V})')'$ and radial density f (where f_1 satisfies (2.2) and (2.3)). The parameter space is thus $\boldsymbol{\Theta} := \mathbb{R}^k \times \mathcal{V}_k$, where $\mathcal{V}_k$ either stands for the set of all $k \times k$ symmetric positive definite matrices $\mathbf{V}$ such that $V_{11} = 1$, or for the corresponding set (in $\mathbb{R}^{(k(k+1)/2)-1}$) of values of $\mathrm{v\mathring{e}ch}\mathbf{V}$.

2.2 Uniform local asymptotic normality.

Our objective is the estimation of the shape parameter $\mathbf{V}$ under unspecified location $\boldsymbol{\theta}$ and radial density f. In this problem, $\mathbf{V}$ is the parameter of interest, whereas $\boldsymbol{\theta}$ and f (equivalently, $\boldsymbol{\theta}$, σ^2, and f_1) play the role of a nuisance.

The relevant statistical experiment is the nonparametric family

$$\mathcal{P}^{(n)} := \bigcup_{f \in \mathcal{F}_\mathrm{A}} \mathcal{P}_f^{(n)} := \bigcup_{f \in \mathcal{F}_\mathrm{A}} \left\{ \mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};f} \middle| \boldsymbol{\theta} \in \mathbb{R}^k, \mathbf{V} \in \mathcal{V}_k \right\} \tag{2.4}$$

(f ranges over the set $\mathcal{F}_\mathrm{A}$ of densities satisfying Assumption (A) below), in which the partition of $\mathcal{P}^{(n)}$ into a collection of parametric subexperiments $\mathcal{P}_f^{(n)}$, all indexed by the same parameters $\boldsymbol{\theta}$ and $\mathbf{V}$, induces a semiparametric structure. Uniform local asymptotic normality (ULAN, for fixed f, with respect to $\boldsymbol{\theta}$ and $\mathbf{V}$) of the parametric subexperiments $\mathcal{P}_f^{(n)}$, $f \in \mathcal{F}_\mathrm{A}$ readily follows from the ULAN property established, in Proposition 1 of Hallin and Paindaveine (2004b), for $\mathcal{P}_{f_1}^{(n)} := \left\{ \mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma^2,\mathbf{V};f_1} \middle| (\boldsymbol{\theta}', (\mathrm{v\mathring{e}ch}\mathbf{V})')' \in \boldsymbol{\Theta}, \sigma^2 \in \mathbb{R}_0^+ \right\}$ with respect to $\boldsymbol{\theta}$, σ^2, and $\mathbf{V}$. The latter is established under extremely mild assumptions on f_1; in order to save space, we do not restate these assumptions here, and only give the following simple sufficient condition, which is satisfied by all densities considered in practice (it is satisfied, for instance, by all multivariate Student radial densities, including the Cauchy ones).

ASSUMPTION (A). The radial density f is absolutely continuous, with a.e.-derivative $\dot{f}$; letting $\varphi_{f_1}(r) := -\dot{f}_1(r)/f_1(r)$, $r \in \mathbb{R}_0^+$ and $K_{f_1}(u) := \varphi_{f_1}(\tilde{F}_{1k}^{-1}(u))(\tilde{F}_{1k}^{-1}(u))$, $u \in (0,1)$, the integrals

$$\mathcal{I}_k(f_1) := \int_0^1 \varphi_{f_1}^2(\tilde{F}_{1k}^{-1}(u))\,du \quad \text{and} \quad \mathcal{J}_k(f_1) := \int_0^1 K_{f_1}^2(u)\,du$$

are finite.

Note that, under this assumption, $\int_0^1 \varphi_{f_1}(\tilde{F}_{1k}^{-1}(u))\,du = 0$ and $\int_0^1 K_{f_1}(u)\,du = k$, so that $\mathrm{E}[\varphi_{f_1}(d_i/\sigma)] = 0$ and $\mathrm{E}[\varphi_{f_1}(d_i/\sigma)\,d_i/\sigma] = k$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};f}$.

Proposition 2.1 *Assume that f satisfies Assumption (A). Then, the sequence of experiments $\mathcal{P}_f^{(n)}$ is ULAN, with (writing d_i and $\mathbf{U}_i$ for $d_i^{(n)}(\boldsymbol{\theta},\mathbf{V})$ and $\mathbf{U}_i^{(n)}(\boldsymbol{\theta},\mathbf{V}) := \mathbf{Z}_i^{(n)}(\boldsymbol{\theta},\mathbf{V})/d_i^{(n)}(\boldsymbol{\theta},\mathbf{V})$, respectively) central sequence*

$$\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta},\mathbf{V}) := \begin{pmatrix} \boldsymbol{\Delta}_{f;I}^{(n)}(\boldsymbol{\theta},\mathbf{V}) \\ \boldsymbol{\Delta}_{f;II}^{(n)}(\boldsymbol{\theta},\mathbf{V}) \end{pmatrix} := n^{-1/2} \begin{pmatrix} \frac{1}{\sigma}\mathbf{V}^{-1/2}\sum_{i=1}^n \varphi_{f_1}\left(\frac{d_i}{\sigma}\right)\mathbf{U}_i \\ \frac{1}{2}\mathbf{M}_k\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\sum_{i=1}^n \mathrm{vec}\left(\varphi_{f_1}\left(\frac{d_i}{\sigma}\right)\frac{d_i}{\sigma}\mathbf{U}_i\mathbf{U}_i' - \mathbf{I}_k\right) \end{pmatrix} \tag{2.5}$$

and full-rank information matrix

$$\boldsymbol{\Gamma}_f(\boldsymbol{\theta},\mathbf{V}) \quad := \quad \begin{pmatrix} \boldsymbol{\Gamma}_{f;I}(\boldsymbol{\theta},\mathbf{V}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{f;II}(\boldsymbol{\theta},\mathbf{V}) \end{pmatrix}, \tag{2.6}$$

where

$$\boldsymbol{\Gamma}_{f;I}(\boldsymbol{\theta},\mathbf{V}) := \frac{1}{k\sigma^2}\mathcal{I}_k(f_1)\mathbf{V}^{-1},$$

and

$$\boldsymbol{\Gamma}_{f;II}(\boldsymbol{\theta},\mathbf{V}) := \frac{1}{4}\mathbf{M}_k\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\left[\frac{\mathcal{J}_k(f_1)}{k(k+2)}\left(\mathbf{I}_{k^2} + \mathbf{K}_k + \mathbf{J}_k\right) - \mathbf{J}_k\right]\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\mathbf{M}_k'. \tag{2.7}$$

The block-diagonal structure of the information matrix (2.6) and ULAN imply that substituting (in principle, after adequate discretization) a root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$ for the unknown location $\boldsymbol{\theta}$ has no influence, asymptotically, on the $\mathbf{V}$-part $\boldsymbol{\Delta}_{f;II}^{(n)}$ of the central sequence. Hence, optimal inference about $\mathbf{V}$ can be based, without any loss of (asymptotic) efficiency, on $\boldsymbol{\Delta}_{f;II}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{V})$ as if $\hat{\boldsymbol{\theta}}^{(n)}$ were the actual location parameter: this actually follows from the asymptotic linearity property of Section 6.1. Therefore, in the derivation of theoretical results, we tacitly may assume, without loss of generality, that $\boldsymbol{\theta} = \mathbf{0}$. The notation $\mathrm{P}^{(n)}_{\mathbf{V};f}$, $d_i^{(n)}(\mathbf{V})$, $\mathbf{U}_i^{(n)}(\mathbf{V})$, $\boldsymbol{\Delta}_f^{(n)}(\mathbf{V})$, $\boldsymbol{\Gamma}_f(\mathbf{V})$, $\ldots$ will be used in an obvious way instead of $\mathrm{P}^{(n)}_{\mathbf{0},\mathbf{V};f}$, $d_i^{(n)}(\mathbf{0},\mathbf{V})$, $\mathbf{U}_i^{(n)}(\mathbf{0},\mathbf{V})$, $\boldsymbol{\Delta}_{f;II}^{(n)}(\mathbf{0},\mathbf{V})$, $\boldsymbol{\Gamma}_{f;II}(\mathbf{0},\mathbf{V})$, etc. The experiment (2.4) under study now takes the form

$$\mathcal{P}^{(n)} := \bigcup_{f\in\mathcal{F}_\mathrm{A}} \mathcal{P}_f^{(n)} := \bigcup_{f\in\mathcal{F}_\mathrm{A}} \left\{ \mathrm{P}^{(n)}_{\mathbf{V};f} \,\middle|\, \mathbf{V} \in \mathcal{V}_k \right\} \tag{2.8}$$

Although any root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$ could be used, we suggest adopting the multivariate affine-equivariant median introduced by Hettmansperger and Randles (2002), which is itself a "sign-based" estimator. The multivariate signs to be considered then are the $\mathbf{U}_i^{(n)}(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{V})$'s, and the ranks those of the $d_i^{(n)}(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{V})$'s.

2.3 Semiparametric efficiency, ranks, and signs.

As already mentioned, the partition of $\mathcal{P}^{(n)}$ into a collection of parametric subexperiments $\mathcal{P}_f^{(n)}$, all indexed by $\mathbf{V}$, induces a semiparametric structure, where $\mathbf{V}$ is the parameter of interest, whereas f or, equivalently, (σ^2, f_1) plays the role of a nuisance. Except for the unavoidable loss of efficiency resulting from the presence of this nuisance, we would like our estimators to be optimal, i.e., to reach semiparametric efficiency bounds, either at some prespecified radial density f_1, or at any density belonging to some class $\mathcal{F}$ of radial densities. The semiparametric efficiency bound at f is provided by the so-called *efficient information matrix*

$$\mathbf{\Gamma}^*_{f_1}(\mathbf{V}) := \frac{\mathcal{J}_k(f_1)}{4k(k+2)}\mathbf{M}_k\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\left[\mathbf{I}_{k^2}+\mathbf{K}_k-\frac{2}{k}\mathbf{J}_k\right]\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\mathbf{M}'_k =: \mathcal{J}_k(f_1)\mathbf{\Upsilon}_k^{-1}(\mathbf{V}), \tag{2.9}$$

which is the asymptotic covariance matrix of the *efficient central sequence*

$$\mathbf{\Delta}_f^{(n)*}(\mathbf{V}) := \frac{1}{2}n^{-1/2}\mathbf{M}_k\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\left[\mathbf{I}_{k^2}-\frac{1}{k}\mathbf{J}_k\right]\sum_{i=1}^{n}\varphi_{f_1}\left(\frac{d_i}{\sigma}\right)\frac{d_i}{\sigma}\,\mathrm{vec}\left(\mathbf{U}_i\mathbf{U}'_i\right) \tag{2.10}$$

(see Hallin and Paindaveine 2004b) and does not depend on σ (whence the notation). An estimator $\mathbf{V}^{(n)}$ of $\mathbf{V}$ is semiparametrically efficient at f iff the asymptotic distribution under $\mathrm{P}^{(n)}_{\mathbf{V};f}$ of $n^{1/2}\left(\overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)})-\overset{\circ}{\mathrm{vech}}(\mathbf{V})\right)$ is the same as that of $\left(\mathbf{\Gamma}^*_{f_1}(\mathbf{V})\right)^{-1}\mathbf{\Delta}_f^{(n)*}(\mathbf{V})$, that is, iff, under $\mathrm{P}^{(n)}_{\mathbf{V};f}$,

$$n^{1/2}\left(\overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)})-\overset{\circ}{\mathrm{vech}}\left(\mathbf{V}\right)\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(\mathbf{0},\left(\mathbf{\Gamma}^*_{f_1}(\mathbf{V})\right)^{-1}\right). \tag{2.11}$$

The difference $\mathbf{\Gamma}_f(\mathbf{V})-\mathbf{\Gamma}^*_{f_1}(\mathbf{V})$ quantifies the loss of information on $\mathbf{V}$ which is due to the non-specification of (σ^2, f_1)—actually, to the non-specification of σ^2 (as far as f_1 is concerned, the model is *adaptive*); see Section 3.1 of Hallin and Paindaveine (2004b) for details.

A general result by Hallin and Werker (2003) indicates that, in case

(i) the parametric subexperiments $\mathcal{P}_f^{(n)}$, $f\in\mathcal{F}_{\mathrm{A}}$ are ULAN, with central sequences $\mathbf{\Delta}_f^{(n)}(\mathbf{V})$ and information matrices $\mathbf{\Gamma}_f(\mathbf{V})$, and

(ii) the nonparametric subexperiments $\mathcal{P}_{\mathbf{V}}^{(n)} := \left\{\mathrm{P}^{(n)}_{\mathbf{V};f}\,\middle|\, f\in\mathcal{F}_{\mathrm{A}}\right\}$, $\mathbf{V}\in\mathcal{V}_k$ are generated by groups of transformations $\mathcal{G}_{\mathbf{V}}^{(n)}$ with maximal invariant σ-fields $\mathcal{B}_{\mathbf{V}}^{(n)}$,

then the projection $\mathrm{E}\left[\mathbf{\Delta}_f^{(n)}(\mathbf{V})\,\middle|\,\mathcal{B}_{\mathbf{V}}^{(n)}\right]$ of $\mathbf{\Delta}_f^{(n)}(\mathbf{V})$ onto $\mathcal{B}_{\mathbf{V}}^{(n)}$ yields a distribution-free version of the *semiparametrically efficient central sequence* (2.10), in the sense that, under $\mathrm{P}^{(n)}_{\mathbf{V};f}$,

$$\mathrm{E}\left[\mathbf{\Delta}_f^{(n)}(\mathbf{V})\,\middle|\,\mathcal{B}_{\mathbf{V}}^{(n)}\right]-\mathbf{\Delta}_f^{(n)*}(\mathbf{V}) = o_{\mathrm{P}}(1)$$

as $n\to\infty$.

In the present context, this double structure exists: (i) is an immediate consequence of Proposition 2.1, and the generating groups $\mathcal{G}_{\mathbf{V}}^{(n)}$ are the groups of order-preserving radial transformations (see Section 4.1 of Hallin and Paindaveine (2004b) for a precise description), which admit the ranks $R_i^{(n)} = R_i^{(n)}(\mathbf{V})$ of the distances $d_i^{(n)}(\mathbf{V})$ and the multivariate signs $\mathbf{U}_i^{(n)} = \mathbf{U}_i^{(n)}(\mathbf{V})$ as maximal invariants. Locally asymptotically optimal estimation of $\mathbf{V}$ thus in principle can be based on ranks and signs.

Note that, under $\mathcal{P}_{\mathbf{V}}^{(n)}$, the vector of ranks $(R_1^{(n)}(\mathbf{V}), \ldots, R_n^{(n)}(\mathbf{V}))$ is uniformly distributed over the $n!$ permutations of $(1, \ldots, n)$, while the signs $\mathbf{U}_i^{(n)}(\mathbf{V})$ are i.i.d. and uniformly distributed over the unit sphere; ranks and signs moreover are mutually independent. Since however $\mathbf{V}$ is unknown, those ranks and signs associated with $\mathbf{V}$ are not computable from the observations, and the construction of our R-estimators will be based on the ranks $R_i^{(n)}(\mathbf{V}_T^{(n)})$ and the signs $\mathbf{U}_i^{(n)}(\mathbf{V}_T^{(n)})$ associated with Tyler's estimator $\mathbf{V}_T^{(n)}$ (see Section 3.2 for a precise definition)—call them *Tyler ranks* and *Tyler signs*, respectively.

3 One-step optimal R-estimation of shape.

3.1 A central sequence for shape based on ranks and signs.

Rather than $\mathrm{E}\left[\mathbf{\Delta}_f^{(n)}(\mathbf{V}) \middle| R_1^{(n)}(\mathbf{V}), \ldots, R_n^{(n)}(\mathbf{V}), \mathbf{U}_1^{(n)}(\mathbf{V}), \ldots, \mathbf{U}_n^{(n)}(\mathbf{V})\right]$, the semiparametrically efficient rank-based central sequences we plan to consider for our construction of R-estimators are of the (asymptotically equivalent: see (3.2)) form

$$\begin{aligned} \underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\mathbf{V}) &:= \frac{1}{2} n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \left[\mathbf{I}_{k^2} - \frac{1}{k}\mathbf{J}_k\right] \sum_{i=1}^n K_{f_1}\left(\frac{R_i}{n+1}\right) \mathrm{vec}\left(\mathbf{U}_i \mathbf{U}_i'\right) \qquad (3.1) \\ &= \frac{1}{2} n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \sum_{i=1}^n \left[K_{f_1}\left(\frac{R_i}{n+1}\right) \mathrm{vec}\left(\mathbf{U}_i \mathbf{U}_i'\right) - \frac{m_{f_1}^{(n)}}{k} \mathrm{vec}\left(\mathbf{I}_k\right)\right], \end{aligned}$$

with $R_i = R_i^{(n)}(\mathbf{V})$, $\mathbf{U}_i = \mathbf{U}_i^{(n)}(\mathbf{V})$, and the exact centering constants $m_{f_1}^{(n)} := \frac{1}{n}\sum_{i=1}^n K_{f_1}\left(\frac{i}{n+1}\right)$. It follows from part (ii) of Proposition 3.1 below that

$$\mathrm{E}\left[\mathbf{\Delta}_f^{(n)}(\mathbf{V}) \middle| R_1^{(n)}(\mathbf{V}), \ldots, R_n^{(n)}(\mathbf{V}), \mathbf{U}_1^{(n)}(\mathbf{V}), \ldots, \mathbf{U}_n^{(n)}(\mathbf{V})\right] - \underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\mathbf{V}) = o_{\mathrm{P}}(1) \qquad (3.2)$$

under $\mathrm{P}_{\mathbf{V};f}^{(n)}$, as $n \to \infty$: $\underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\mathbf{V})$ thus also qualifies as a rank-based version of the semiparametrically efficient (at $\mathrm{P}_{\mathbf{V};f}^{(n)}$) central sequence (2.10). For any $g \in \mathcal{F}_{\mathrm{A}}$, define

$$\mathbf{\Gamma}_{f_1,g_1}^*(\mathbf{V}) := \mathcal{J}_k(f_1, g_1)\mathbf{\Upsilon}_k^{-1}(\mathbf{V}), \qquad (3.3)$$

where (using the notation $\tilde{G}_{1k}$, φ_{g_1} in an obvious way)

$$\mathcal{J}_k(f_1, g_1) := \int_0^1 K_{f_1}(u) K_{g_1}(u)\, du = \int_0^1 \tilde{F}_{1k}^{-1}(u)\varphi_{f_1}(\tilde{F}_{1k}^{-1}(u))\tilde{G}_{1k}^{-1}(u)\varphi_{g_1}(\tilde{G}_{1k}^{-1}(u))\, du \qquad (3.4)$$

(a cross-information quantity); note that, when f and g coincide up to the scale (that is, for $f_1 = g_1$), $\mathcal{J}_k(f_1, g_1) = \mathcal{J}_k(f_1)$, and that $\mathbf{\Gamma}_{f_1,f_1}^*(\mathbf{V})$ reduces to $\mathbf{\Gamma}_{f_1}^*(\mathbf{V})$ defined in (2.9). The properties of $\underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\mathbf{V})$ are summarized in the following proposition (in which we let $d_i := d_i^{(n)}(\mathbf{V})$).

Proposition 3.1 *For any $f \in \mathcal{F}_{\mathrm{A}}$, the rank-based random vector $\underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\mathbf{V})$ defined in (3.1)*

(i) is distribution-free under $\left\{\mathrm{P}_{\mathbf{V};g}^{(n)} \middle| g \in \mathcal{F}\right\}$, where $\mathcal{F}$ denotes the class of all possible radial densities;

(ii) is asymptotically equivalent, in $\mathrm{P}^{(n)}_{\mathbf{V};g}$*-probability for any* $g \in \mathcal{F}$*, to*

$$\boldsymbol{\Delta}^{(n)*}_{f,g}(\mathbf{V}) := \frac{1}{2} n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \left[\mathbf{I}_{k^2} - \frac{1}{k}\mathbf{J}_k\right] \sum_{i=1}^{n} K_{f_1}\left(\tilde{G}_{1k}\left(\frac{d_i}{\sigma}\right)\right) \mathrm{vec}\left(\mathbf{U}_i \mathbf{U}_i'\right) \quad (3.5)$$

(σ here stands for the scale parameter associated with g), hence, in $\mathrm{P}^{(n)}_{\mathbf{V};f}$*-probability, to the semiparametrically efficient (at f, for any σ) central sequence for shape (2.10);*

(iii) is asymptotically normal under $\left\{ \mathrm{P}^{(n)}_{\mathbf{V};g} \middle| \, g \in \mathcal{F} \right\}$*, with mean zero and covariance matrix* $\boldsymbol{\Gamma}^*_{f_1}(\mathbf{V})$*;*

(iv) is asymptotically normal under $\mathrm{P}^{(n)}_{\mathbf{V}+n^{-1/2}\mathbf{v};g}$*, as $n \to \infty$, with mean* $\boldsymbol{\Gamma}^*_{f_1,g_1}(\mathbf{V})\overset{\circ}{\mathrm{vech}}(\mathbf{v})$ *and covariance matrix* $\boldsymbol{\Gamma}^*_{f_1}(\mathbf{V})$*, for any symmetric matrix $\mathbf{v}$ such that $v_{11} = 0$ and any $g \in \mathcal{F}_{\mathrm{A}}$;*

(v) satisfies under $\mathrm{P}^{(n)}_{\mathbf{V};g}$*, as $n \to \infty$, the asymptotic linearity property*

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f}(\mathbf{V} + n^{-1/2}\mathbf{v}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f}(\mathbf{V}) = -\boldsymbol{\Gamma}^*_{f_1,g_1}(\mathbf{V})\, \overset{\circ}{\mathrm{vech}}(\mathbf{v}^{(n)}) + o_{\mathrm{P}}(1) \quad (3.6)$$

for any bounded sequence $\mathbf{v}^{(n)}$ of symmetric matrices such that $v^{(n)}_{11} = 0$, and any $g \in \mathcal{F}_{\mathrm{A}}$.

Proof. Part (i): the distribution-freeness of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f}(\mathbf{V})$ readily follows from the distribution-freeness, under ellipticity, of the ranks $R^{(n)}_i(\mathbf{V})$ and the signs $\mathbf{U}^{(n)}_i(\mathbf{V})$ with respect to which $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f}(\mathbf{V})$ is measurable. Parts (ii) and (iii) are consequences of the more general asymptotic representation result given in Lemma 2 of Hallin and Paindaveine (2004b) (see the proof of their Proposition 3). Part (iv) is a direct application of Le Cam's Third Lemma (Proposition 4 of Hallin and Paindaveine 2004b). Finally, the asymptotic linearity property of part (v) follows from the more general property given in Proposition 6.1). □

3.2 An optimal one-step rank-based pseudo-estimator.

Tyler's celebrated estimator of shape $\mathbf{V}^{(n)}_T$ was introduced by Tyler (1987a) from the very simple idea that, if $\mathbf{X}$ is elliptical with location $\boldsymbol{\theta}$, then its shape $\mathbf{V}$ is entirely characterized by the fact that $\mathbf{U}(\boldsymbol{\theta}, \mathbf{V}) := \mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\theta})/\|\mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\theta})\|$ is centered, with covariance $(1/k)\mathbf{I}_k$. He accordingly defines (assume $\boldsymbol{\theta}$ is known) $\mathbf{V}^{(n)}_T$ as the unique shape matrix satisfying $\frac{1}{n}\sum_{i=1}^{n} \mathbf{U}^{(n)}_i(\boldsymbol{\theta}, \mathbf{V})\big(\mathbf{U}^{(n)}_i(\boldsymbol{\theta}, \mathbf{V})\big)' = \frac{1}{k}\mathbf{I}_k$.

Denote by $\mathbf{V}^{(n)}_{\#}$ a discretized version of $\mathbf{V}^{(n)}_T$. Such a discretization can be obtained, for instance, by mapping each component $v^{(n)}_i$ of the $\overset{\circ}{\mathrm{vech}}$ form of Tyler's original estimate onto $v^{(n)}_{\#i} := c_0^{-1}\mathrm{sign}(v^{(n)}_i)n^{-1/2}\lceil n^{1/2} c_0 |v^{(n)}_i| \rceil$, where $\lceil n^{1/2}|v^{(n)}_i| \rceil$ denotes the smallest integer larger than or equal to $n^{1/2}|v^{(n)}_i|$ and c_0 an arbitrary positive constant that does not depend on n. Clearly, this discretization does not affect root-n consistency; in practice (where $n = n_0$ is fixed), it is not required, as c_0 can be arbitrarily large, and actually makes little sense, as one can always pretend starting discretization at $n = n_0 + 1$; see Section 3.5 for practical implementation.

Since $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f}(\mathbf{V})$ is a version of the efficient central sequence for shape, Le Cam's classical one-step method suggests estimating $\overset{\circ}{\mathrm{vech}}(\mathbf{V})$ by means of

$$\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}) := \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) + n^{-1/2}\left(\boldsymbol{\Gamma}^*_{f_1,g_1}(\mathbf{V}^{(n)}_{\#})\right)^{-1} \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f}(\mathbf{V}^{(n)}_{\#}). \quad (3.7)$$

Such an estimator is semiparametrically efficient at $\mathcal{P}_f^{(n)}$, in the sense of (2.11). Indeed, in view of Proposition 3.1 and the continuity of $\mathbf{v} \mapsto \mathbf{\Gamma}^*_{f_1,g_1}(\mathbf{v})$, we have, under $\mathrm{P}^{(n)}_{\mathbf{V};g}$,

$$
\begin{aligned}
n^{1/2}\Big(\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}) - \overset{\circ}{\mathrm{vech}}(\mathbf{V})\Big) &= n^{1/2}\left(\overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) - \overset{\circ}{\mathrm{vech}}(\mathbf{V})\right) + \left(\mathbf{\Gamma}^*_{f_1,g_1}(\mathbf{V}^{(n)}_{\#})\right)^{-1} \underset{\sim}{\mathbf{\Delta}}^{(n)}_{f}(\mathbf{V}^{(n)}_{\#}) \\
&= n^{1/2}\left(\overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) - \overset{\circ}{\mathrm{vech}}(\mathbf{V})\right) + \left(\mathbf{\Gamma}^*_{f_1,g_1}(\mathbf{V}^{(n)}_{\#})\right)^{-1} \\
&\quad \times\Big(\mathbf{\Delta}^{(n)*}_{f,g}(\mathbf{V}) - \mathbf{\Gamma}^*_{f_1,g_1}(\mathbf{V})\, n^{1/2}\Big(\overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) - \overset{\circ}{\mathrm{vech}}(\mathbf{V})\Big)\Big) + o_{\mathrm{P}}(1) \\
&= \left(\mathbf{\Gamma}^*_{f_1,g_1}(\mathbf{V})\right)^{-1} \mathbf{\Delta}^{(n)*}_{f,g}(\mathbf{V}) + o_{\mathrm{P}}(1) \qquad (3.8) \\
&= \left(\mathbf{\Gamma}^*_{f_1,g_1}(\mathbf{V})\right)^{-1} \underset{\sim}{\mathbf{\Delta}}^{(n)}_{f}(\mathbf{V}) + o_{\mathrm{P}}(1), \qquad (3.9)
\end{aligned}
$$

where the application to $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{f}(\mathbf{V}^{(n)}_{\#})$ of the asymptotic linearity property (3.6) is made possible, as usual, by the local discreteness of $\mathbf{V}^{(n)}_{\#}$; the asymptotic representation (3.8) implies, for $g = f$, the efficiency of $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$, whereas (3.9), by providing for $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$ an asymptotic representation as a genuine signed rank quantity, justifies its status as an R-estimator.

A major problem however is that (3.7), via $\mathbf{\Gamma}^*_{f_1,g_1}(\mathbf{V}^{(n)}_{\#})$, involves the unknown cross-information quantity $\mathcal{J}_k(f_1, g_1)$ defined in (3.4); $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$ thus is just a *pseudo-estimator*, as it cannot be computed from the observations. The presence of this unknown coefficient $\mathcal{J}_k(f_1, g_1)$ probably explains why one-step R-estimation methods seldom have been considered so far in practice (this problem is completely overlooked, for instance, in Allal et al. 2001).

Clearly, any $\mathrm{P}^{(n)}_{\mathbf{V};g}$-consistent estimator $\hat{\mathcal{J}}^{(n)}_f$ of $\mathcal{J}_k(f_1, g_1)$ can be substituted for $\mathcal{J}_k(f_1, g_1)$ itself. The same estimation of $\mathcal{J}_k(f_1, g_1)$ will be needed whenever the asymptotic variance of $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$ is to be computed. This estimation thus plays a crucial role.

Estimating $\mathcal{J}_k(f_1, g_1)$—the expectation of a function that depends on the unknown g—however is not an obvious task, as explained in Section 1.2.

"Naive" consistent estimators can be obtained (Section 3.3) from the asymptotic linearity property (3.6), but cannot be expected to be very accurate. And, although plain consistency in theory is sufficient, a poor estimation of $\mathcal{J}_k(f_1, g_1)$ is likely to ruin the finite sample performance of (3.7). More accuracy can be expected, for large sample size, from a kernel estimation of g, along with a fair amount of technical work (see Schweder 1975 and Bickel and Ritov 1988 for the simple Wilcoxon case). From a practical perspective, this approach however is rather heavy (with the usual kernel and bandwidth choices), and is hopeless for small or moderate values of n. Moreover, from a decision-theoretical point of view, estimating g is somewhat incompatible with the group-invariance spirit of the rank-based approach: if indeed the unknown density g eventually is to be estimated by some $\hat{g}$, why not simply adopt a more traditional estimated-score approach, based on the asymptotic reconstruction, via $\mathbf{\Delta}^{(n)*}_{\hat{g}}$, of the efficient central sequence $\mathbf{\Delta}^{(n)*}_{g}$?

We therefore propose, in Section 3.4, an original solution to this problem of estimating $\mathcal{J}_k(f_1, g_1)$. Our solution, inspired by local likelihood maximization ideas, yields a genuine one-step R-estimator $\widehat{\underset{\sim}{\mathbf{V}}}^{(n)}_{f\#}$ asymptotically equivalent (under any $\mathrm{P}^{(n)}_{\mathbf{V};g}$) to the pseudo-estimator $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$, hence also to the signed rank statistic (3.9) based on the "genuine ranks".

The asymptotic properties of the resulting R-estimators are thus the same as those of $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$, which we summarize in the following proposition: (i) they are asymptotically equivalent to a function of the genuine ranks and signs, asymptotically normal, and their covariance matrix is the inverse of the covariance matrix characterizing the local powers of the optimal rank tests derived in Hallin and Paindaveine (2004b); (ii) when based on f_1-scores, they are semiparametrically efficient at radial density f; (iii) for finite n, they can be expressed as a linear combination of the Tyler shape matrix and a rank-based shape matrix involving the Tyler ranks and signs; (iv) their asymptotic covariance matrix, under any density, is proportional to the asymptotic covariance matrices of the Gaussian estimator and of Tyler's estimator (the Gaussian estimator is defined in (iv)); the proportionality constant, which can be considered as a measure of asymptotic relative efficiency, is provided in (v).

In order to obtain a simpler "$\mathbf{M}_k$-free" expression for the asymptotic covariance matrix of $\mathrm{vec}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#})$ (cf. 3.14), define $\mathbf{Q}_k(\mathbf{V}) := [k(k+2)]^{-1}\mathbf{M}'_k\boldsymbol{\Upsilon}_k(\mathbf{V})\mathbf{M}_k$. As shown in the proof of Lemma 1 in Hallin and Paindaveine (2004b),

$$\boldsymbol{\Upsilon}_k(\mathbf{V}) = k(k+2)\,\mathbf{N}_k\mathbf{Q}_k(\mathbf{V})\mathbf{N}'_k, \tag{3.10}$$

where $\mathbf{N}_k$ is defined in Section 1.5.

Proposition 3.2 *Let f and g belong to $\mathcal{F}_\mathrm{A}$. Then,*

(i) under $\mathrm{P}^{(n)}_{\mathbf{V};g}$, as $n \to \infty$,

$$n^{1/2}\overset{\circ}{\mathrm{vech}}\left(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#} - \mathbf{V}\right) = \left(\boldsymbol{\Gamma}^*_{f_1,g_1}(\mathbf{V})\right)^{-1}\boldsymbol{\Delta}^{(n)*}_{f,g}(\mathbf{V}) + o_\mathrm{P}(1) \tag{3.11}$$

$$= \left(\boldsymbol{\Gamma}^*_{f_1,g_1}(\mathbf{V})\right)^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f}(\mathbf{V}) + o_\mathrm{P}(1) \tag{3.12}$$

$$\overset{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\mathbf{0}, \left(\mathcal{J}_k(f_1)/\mathcal{J}^2_k(f_1,g_1)\right)\boldsymbol{\Upsilon}_k(\mathbf{V})\right), \tag{3.13}$$

or, in terms of $\mathrm{vec}\,\mathbf{V}$,

$$n^{1/2}\mathrm{vec}\left(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#} - \mathbf{V}\right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\mathbf{0}, \left(k(k+2)\mathcal{J}_k(f_1)/\mathcal{J}^2_k(f_1,g_1)\right)\mathbf{Q}_k(\mathbf{V})\right); \tag{3.14}$$

(ii) $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$ (equivalently, $\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#})$ or $\mathrm{vec}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#})$) is semiparametrically efficient at $\left\{\mathrm{P}^{(n)}_{\mathbf{V};f} \,\middle|\, \mathbf{V} \in \mathcal{V}_k\right\}$;

(iii) for all n,

$$\underset{\sim}{\mathbf{V}}^{(n)}_{f\#} = \left(1 - \frac{k(k+2)}{\mathcal{J}_k(f_1,g_1)}(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#})_{11}\right)\mathbf{V}^{(n)}_{\#} + \left(\frac{k(k+2)}{\mathcal{J}_k(f_1,g_1)}(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#})_{11}\right)\underset{\sim}{\mathbf{W}}^{(n)}_{f\#}\Big/(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#})_{11}, \tag{3.15}$$

where $\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} := \underset{\sim}{\mathbf{W}}^{(n)}_{f}(\mathbf{V}^{(n)}_{\#})$, with

$$\underset{\sim}{\mathbf{W}}^{(n)}_{f}(\mathbf{V}) := (\mathbf{V})^{1/2}\left[\frac{1}{n}\sum_{i=1}^n K_{f_1}\left(\frac{R^{(n)}_i(\mathbf{V})}{n+1}\right)\mathbf{U}^{(n)}_i(\mathbf{V})\mathbf{U}'^{(n)}_i(\mathbf{V})\right](\mathbf{V})^{1/2}; \tag{3.16}$$

(iv) the parametric Gaussian estimator is $\mathbf{V}^{(n)}_{\mathcal{G}} := \boldsymbol{\Sigma}^{(n)}/(\boldsymbol{\Sigma}^{(n)})_{11}$, with $\boldsymbol{\Sigma}^{(n)} := n^{-1}\sum_{i=1}^n(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$; provided that the kurtosis coefficient $\kappa_k(g_1) := (kE_k(g_1))/((k+2)D^2_k(g_1)) - 1$ (where we let $E_k(g_1) := \int_0^1(\tilde{G}_1^{-1}(u))^4\,du$ and $D_k(g_1) := \int_0^1(\tilde{G}_1^{-1}(u))^2\,du$) associated with g_1 is finite,

$$n^{1/2}\mathrm{vec}\left(\mathbf{V}^{(n)}_{\mathcal{G}} - \mathbf{V}\right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(\mathbf{0}, (1+\kappa_k(g_1))\,\mathbf{Q}_k(\mathbf{V})\right)$$

under $\mathrm{P}^{(n)}_{\mathbf{V};g}$ as $n \to \infty$;

(v) the ARE (i.e., the inverse ratio of the corresponding asymptotic variances), under $\mathrm{P}^{(n)}_{\mathbf{V};g}$ *where* g_1 *has finite moments of order four (resp., without any moment assumption on* g_1*), of* $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$ *with respect to* $\mathbf{V}^{(n)}_{\mathcal{G}}$ *(resp.,* $\mathbf{V}^{(n)}_{T}$*) is* $\frac{1+\kappa_k(g_1)}{k(k+2)}\frac{\mathcal{J}^2_k(f_1,g_1)}{\mathcal{J}_k(f_1)}$ *(resp.,* $\frac{1}{k^2}\frac{\mathcal{J}^2_k(f_1,g_1)}{\mathcal{J}_k(f_1)}$*).*

Proof. See Appendix (Section 6.2). □

Remark that the terminology "asymptotic relative efficiency" in part (v) of the proposition can be used unambiguously, despite the multivariate setting, as all pseudo-estimators $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$, the Gaussian estimator $\mathbf{V}^{(n)}_{\mathcal{G}}$, and Tyler's estimator $\mathbf{V}^{(n)}_{T}$, have an asymptotic covariance matrix proportional to $\mathbf{\Upsilon}_k(\mathbf{V})$ (see the proof of part (v) for the asymptotic distribution of $\mathbf{V}^{(n)}_{T}$). The relative performances of these estimators of shape thus can be described by a single number, a fact that was already observed in Ollila, Hettmansperger, and Oja (2004); the situation is entirely different for covariance matrices, where two numbers are required (see Tyler 1982, Ollila, Oja, and Croux 2003, and Ollila, Croux, and Oja 2004).

Table 1 provides some numerical values, under various Student (t_ν) and normal ($\mathcal{N}$) radial densities g_1, of the AREs obtained in part (v) of Proposition 3.2. Note that, under Student densities with less than 4 degrees of freedom, the ARE of $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}$ with respect to $\mathbf{V}^{(n)}_{\mathcal{G}}$ is infinite, as $n^{1/2}(\mathbf{V}^{(n)}_{\mathcal{G}}-\mathbf{V})$ is not even $O_{\mathrm{P}}(1)$. On the other hand, also note that under Student densities with 0.5 degrees of freedom, the AREs, with respect to Tyler's $\mathbf{V}^{(n)}_{T}$, of $\underset{\sim}{\mathbf{V}}_{\nu\#}$ are relatively modest. This is explained by the fact that, roughly speaking, "$\mathbf{V}^{(n)}_{T}$ is optimal at t_0 (the Student with zero degrees of freedom)." In more rigorous terms, we have that, for any fixed n,

$$\underset{\sim}{\mathbf{V}}^{(n)}_{\nu\#}-\mathbf{V}^{(n)}_{T}=o(1)\quad \mathcal{P}^{(n)}\text{-a.s., as } \nu\to 0. \tag{3.17}$$

Indeed, the scores K_ν associated with the k-dimensional Student t_ν take the form

$$K_\nu(u)=k(k+\nu)G^{-1}_{k,\nu}(u)/(\nu+kG^{-1}_{k,\nu}(u))\quad u\in(0,1),$$

where $G_{k,\nu}$ stands for the Fisher-Snedecor distribution function with k and ν degrees of freedom. It is easily checked that $G^{-1}_{k,\nu}(u)/\nu\to\infty$ as $\nu\to 0$, so that $\lim_{\nu\to 0}K_\nu(u)=k$, for all $u\in(0,1)$. It follows that (with obvious notation) $\underset{\sim}{\mathbf{W}}^{(n)}_{\nu\#}-\mathbf{V}^{(n)}_{\#}=o(1)$, $\mathcal{P}^{(n)}$-a.s. as $\nu\to 0$. This, in view of (3.15), implies (3.17).

Using similar arguments, it can easily be shown that, for all fixed n and ν, $\underset{\sim}{\mathbf{V}}^{(n)}_{\nu\#}-\mathbf{V}^{(n)}_{T}=o(1)$ $\mathcal{P}^{(n)}$-a.s., as $k\to\infty$. This explains the fact that, for all fixed ν, the ARE of $\underset{\sim}{\mathbf{V}}^{(n)}_{\nu\#}$ with respect to $\mathbf{V}^{(n)}_{T}$ goes to 1 as $k\to\infty$. Incidentally, this also holds for the van der Waerden—that is, the Gaussian-score ($K_{f_1}=K_{\phi_1}$)—version of our pseudo-estimators: as the dimension k of the observation space goes to infinity, the information contained in the radii d_i becomes negligible when compared with that contained in the directions $\mathbf{U}_i$.

3.3 Naive estimation of cross-information coefficients.

A simple consistent estimate of $\mathcal{J}_k(f_1,g_1)$ can be obtained through the asymptotic linearity property in part (v) of Proposition 3.1. The idea we are exploiting here goes back, under its simplest form (for traditional univariate ranks, in location and regression models), to Kraft and van Eeden (1972) and Antille (1974); see page 321 of Jurečková and Sen (1996).

		underlying density			
	k	$t_{0.5}$	t_3	t_{10}	$\mathcal{N}$
	2	1.111 (∞)	1.246 (∞)	1.280 (0.853)	1.296 (0.648)
	3	1.061 (∞)	1.145 (∞)	1.173 (0.939)	1.189 (0.713)
$\underset{\sim}{\mathbf{V}}^{(n)}_{0.5\#}$	4	1.038 (∞)	1.098 (∞)	1.121 (0.996)	1.136 (0.757)
	6	1.020 (∞)	1.054 (∞)	1.070 (1.070)	1.083 (0.813)
	10	1.008 (∞)	1.024 (∞)	1.034 (1.149)	1.044 (0.870)
	2	0.969 (∞)	1.429 (∞)	1.651 (1.101)	1.792 (0.896)
	3	0.972 (∞)	1.250 (∞)	1.400 (1.120)	1.507 (0.904)
$\underset{\sim}{\mathbf{V}}^{(n)}_{3\#}$	4	0.977 (∞)	1.667 (∞)	1.278 (1.136)	1.366 (0.911)
	6	0.985 (∞)	1.091 (∞)	1.162 (1.162)	1.229 (0.921)
	10	0.992 (∞)	1.040 (∞)	1.078 (1.198)	1.123 (0.936)
	2	0.829 (∞)	1.376 (∞)	1.714 (1.143)	1.961 (0.980)
	3	0.861 (∞)	1.212 (∞)	1.444 (1.156)	1.633 (0.979)
$\underset{\sim}{\mathbf{V}}^{(n)}_{10\#}$	4	0.887 (∞)	1.136 (∞)	1.313 (1.167)	1.468 (0.979)
	6	0.921 (∞)	1.070 (∞)	1.185 (1.185)	1.304 (0.978)
	10	0.955 (∞)	1.027 (∞)	1.091 (1.212)	1.174 (0.978)
	2	0.720 (∞)	1.280 (∞)	1.681 (1.120)	2.000 (1.000)
	3	0.757 (∞)	1.130 (∞)	1.415 (1.132)	1.667 (1.000)
$\underset{\sim}{\mathbf{V}}^{(n)}_{\mathrm{vdW}\#}$	4	0.786 (∞)	1.063 (∞)	1.285 (1.142)	1.500 (1.000)
	6	0.829 (∞)	1.005 (∞)	1.159 (1.159)	1.333 (1.000)
	10	0.877 (∞)	0.973 (∞)	1.067 (1.186)	1.200 (1.000)

Table 1: AREs of the rank-based pseudo-estimators $\underset{\sim}{\mathbf{V}}^{(n)}_{0.5\#}$, $\underset{\sim}{\mathbf{V}}^{(n)}_{3\#}$, $\underset{\sim}{\mathbf{V}}^{(n)}_{10\#}$, and $\underset{\sim}{\mathbf{V}}^{(n)}_{\mathrm{vdW}\#}$ (associated with $t_{0.5}$, t_3, t_{10}, and Gaussian scores, respectively) with respect to Tyler's estimator $\mathbf{V}^{(n)}_T$ and, in parentheses, with respect to the Gaussian estimator $\mathbf{V}^{(n)}_{\mathcal{G}}$, under k-dimensional Student (with 0.5, 3, and 10 degrees of freedom) and normal densities, respectively, for $k = 2,\ 3,\ 4,\ 6,$ and 10. The same figures also hold for the R-estimators $\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{f\#}$ described in Section 3.4.

Asymptotic linearity implies that, for all $f, g \in \mathcal{F}_{\mathrm{A}}$ and $k \times k$ symmetric matrix $\mathbf{v}$ such that $v_{11} = 0$,

$$\begin{aligned}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_\# + n^{-1/2}\mathbf{v}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_\#) &= \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V} + n^{-1/2}\mathbf{v}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}) + o_{\mathrm{P}}(1) \qquad (3.18)\\ &=: -\mathcal{J}_k(f_1, g_1)\boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\overset{\circ}{\mathrm{vech}}(\mathbf{v}) + o_{\mathrm{P}}(1),\end{aligned}$$

with $\boldsymbol{\Upsilon}_k(\mathbf{V})$ defined in (2.9). Thus, for any $\mathbf{v}$,

$$\mathcal{J}^{(n)}_k(f_1; \mathbf{v}) := \left\| \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_\# + n^{-1/2}\mathbf{v}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_\#) \right\| \Big/ \left\| \boldsymbol{\Upsilon}_k^{-1}(\mathbf{V}^{(n)}_\#)\overset{\circ}{\mathrm{vech}}(\mathbf{v}) \right\| \qquad (3.19)$$

is a consistent estimate, under $\mathrm{P}^{(n)}_{\mathbf{V};g}$, of $\mathcal{J}_k(f_1, g_1)$.

Theoretical guidelines for the choice of a particular $\mathbf{v}$ would require some information on the higher-order behavior of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V} + n^{-1/2}\mathbf{v}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V})$. Since this information is not available, we suggest the following heuristic choice. Let $\mathbf{v}^*$ be such that $\pm\overset{\circ}{\mathrm{vech}}(\mathbf{v}^*)$ are the eigenvectors associated with $\boldsymbol{\Upsilon}_k^{-1}(\mathbf{V}^{(n)}_\#)$'s largest eigenvalue. Denoting by $r^-(\mathbf{V}^{(n)}_\#)$ and $r^+(\mathbf{V}^{(n)}_\#)$ the smallest integers such that

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_\# + n^{-1/2}c_0^{-1}r^+\mathbf{v}^*) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_\#) \neq \mathbf{0} \quad \text{and} \quad \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_T - n^{-1/2}c_0^{-1}r^-\mathbf{v}^*) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_\#) \neq \mathbf{0}$$

(c_0 is the constant that has been used in discretizing $\mathbf{V}^{(n)}_T$ into $\mathbf{V}^{(n)}_\#$), choose

$$\mathcal{J}^{(n)}_k(f_1) := \frac{1}{2}\left(\mathcal{J}^{(n)}_k(f_1; r^-\mathbf{v}^*) + \mathcal{J}^{(n)}_k(f_1; r^+\mathbf{v}^*)\right).$$

3.4 Optimal one-step R-estimation: consistency and efficiency.

A more sophisticated way of dealing with the estimation of $\mathcal{J}_k(f_1, g_1)$ can be obtained from better exploiting the ULAN structure of the model. The basic intuition is that of solving a local likelihood equation. Consistency however requires somewhat confusing discretization steps which, as usual, are needed in formal proofs only. We therefore provide two descriptions of our estimators. This section carefully goes through the details of discretization, and establishes the asymptotic equivalence of the resulting estimator with the pseudo-estimator of Section 3.2, while Section 3.5, where discretization is skipped, can be used for practical implementation.

Consider the sequence of (random) half-lines

$$\mathcal{D}_{\#}^{(n)} = \mathcal{D}_{\#}^{(n)}(\mathbf{V}_{\#}^{(n)}; \underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\mathbf{V}_{\#}^{(n)})) = \left\{ \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha)) \,\middle|\, \alpha \in \mathbb{R}^+ \right\}, \quad n \in \mathbb{N}$$

with equation

$$\begin{aligned} \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha)) &:= \overset{\circ}{\mathrm{vech}}(\mathbf{V}_{\#}^{(n)}) + n^{-1/2}\alpha \boldsymbol{\Upsilon}_k(\mathbf{V}_{\#}^{(n)}) \underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\mathbf{V}_{\#}^{(n)}) \qquad (3.20) \\ &= \overset{\circ}{\mathrm{vech}}(\mathbf{V}_{\#}^{(n)}) + \alpha k(k+2)\, \mathbf{N}_k \left[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V}_{\#}^{(n)})(\mathbf{e}_{k^2,1})' \right] \mathrm{vec}(\underset{\sim}{\mathbf{W}}_{f\#}^{(n)}), \end{aligned}$$

where $\mathbf{e}_{k^2,1}$ stands for the first vector of the canonical basis in $\mathbb{R}^{k^2}$, and $\underset{\sim}{\mathbf{W}}_{f\#}^{(n)} = \underset{\sim}{\mathbf{W}}_f^{(n)}(\mathbf{V}_{\#}^{(n)})$; the last equality is obtained exactly as in the proof of Proposition 3.2(iii). Each value of α defines on $\mathcal{D}_{\#}^{(n)}$ a sequence of root-n consistent estimators $\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha)$ of $\mathbf{V}$; one of them, namely $\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\mathcal{J}_k^{-1}(f_1, g_1))$, coincides with $\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}$ in (3.7), and is efficient at $\mathcal{P}_f^{(n)}$ (actually, an estimator $\widehat{\mathbf{V}}^{(n)}$ is efficient iff $\widehat{\mathbf{V}}^{(n)} - \underset{\sim}{\mathbf{V}}_{f\#}^{(n)} = o_{\mathrm{P}}(n^{-1/2})$ under $\mathcal{P}_f^{(n)}$).

These estimators $\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha)$ however are not locally discrete, as the multivariate signs $\mathbf{U}_i^{(n)}$ in $\underset{\sim}{\mathbf{W}}_{f\#}^{(n)}$ are not discretized (even though they are evaluated at $\mathbf{V}_{\#}^{(n)}$); we therefore discretize further the estimators $\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha)$ by discretizing $\underset{\sim}{\mathbf{W}}_{f\#}^{(n)}$. Similarly as for the discretization of Tyler's estimator $\mathbf{V}_T^{(n)}$ in Section 3.2, let $\underset{\sim}{\mathbf{W}}_{f\#}^{(n)}$ be the $k \times k$ matrix obtained by mapping each component $w_{i\#}^{(n)}$ of $\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{W}}_{f\#}^{(n)})$ onto $w_{i\#}^{(n)} := c_1^{-1}\mathrm{sign}(w_{i\#}^{(n)}) n^{-1/2} \lceil n^{1/2} c_1 |w_{i\#}^{(n)}| \rceil$, where $c_1 > 0$ is some arbitrarily large constant. Replacing (3.20) with

$$\begin{aligned} \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha)) &:= \overset{\circ}{\mathrm{vech}}(\mathbf{V}_{\#}^{(n)}) + c_2^{-1}\ell\, k(k+2)\, \mathbf{N}_k \left[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V}_{\#}^{(n)})(\mathbf{e}_{k^2,1})' \right] \mathrm{vec}(\underset{\sim}{\mathbf{W}}_{f\#}^{(n)}) \qquad (3.21) \\ &=: \overset{\circ}{\mathrm{vech}}(\mathbf{V}_{\#}^{(n)}) + n^{-1/2} c_2^{-1} \ell\, \boldsymbol{\Upsilon}_k(\mathbf{V}_{\#}^{(n)}) \underset{\sim}{\boldsymbol{\Delta}}_{f\#}^{(n)}(\mathbf{V}_{\#}^{(n)}), \quad \ell \in \mathbb{N}, \end{aligned}$$

with $\alpha_\ell := \ell/c_2$, where $c_2 > 0$ is some arbitrarily large constant (but keeping the same notation for the sake of simplicity) yields estimators $\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell)$ that are root-n consistent and locally discrete, in the sense that the number of possible values of $\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell))$ in balls with $O(n^{-1/2})$ radius centered at $\overset{\circ}{\mathrm{vech}}(\mathbf{V})$ is bounded as $n \to \infty$ (below we denote by $\mathcal{D}_{\#}^{(n)}$ this new sequence $\mathcal{D}_{\#}^{(n)}(\mathbf{V}_{\#}^{(n)}; \underset{\sim}{\boldsymbol{\Delta}}_{f\#}^{(n)}(\mathbf{V}_{\#}^{(n)}))$ of *fully-discretized* half-lines). Hence, for any $\ell \in \mathbb{N}$, $\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell)$ again can serve as the preliminary estimator in a rank-based one-step procedure: letting

$$\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell; \delta)) := \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell)) + n^{-1/2}\delta\, \boldsymbol{\Upsilon}_k(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell)) \underset{\sim}{\boldsymbol{\Delta}}_{f\#}^{(n)}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell)),$$

$\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}_{f\#}^{(n)}(\alpha_\ell; \mathcal{J}_k^{-1}(f_1, g_1)))$ thus is such that

$$\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha_\ell; \mathcal{J}_k^{-1}(f_1, g_1))) - \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}) = o_{\mathrm{P}}(n^{-1/2}) \tag{3.22}$$

under $\mathrm{P}^{(n)}_{\mathbf{V};g}$. However, $\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha_\ell; \mathcal{J}_k^{-1}(f_1, g_1)))$ still cannot be computed from the observations.

Denote by $\mathbf{u}_{\mathcal{D}}$ the unit vector along $\mathcal{D}^{(n)}_{\#}$ (corresponding to $\mathcal{D}^{(n)}_{\#}$'s natural orientation as a half-line), and define

$$\ell^+ := \min\left\{\ell \in \mathbb{N}_0 \;\middle|\; h^{\#}(\alpha_\ell) := \mathbf{u}'_{\mathcal{D}} \boldsymbol{\Upsilon}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha_\ell))\, \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f\#}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha_\ell)) \leq 0\right\}, \tag{3.23}$$

$\ell^- := \ell^+ - 1$, and $\alpha^{\pm} := \alpha_{\ell^{\pm}}$. The integers $\ell^{\pm}$ are random; in order for $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha^{\pm})$ to remain root-n consistent and locally discrete, it is sufficient to check that $\ell^{\pm}$ is $O_{\mathrm{P}}(1)$. This indeed does imply that, for any $\epsilon > 0$, there exist integers L_ϵ and N_ϵ such that, for all $n \geq N_\epsilon$, the minimization in (3.23) with probability larger than $1 - \epsilon$ only runs over the finite set $\ell \in \{1, \ldots, L_\epsilon\}$ (equivalently, over the finite set $\alpha \in \{\alpha_1, \ldots, \alpha_{L_\epsilon}\}$). In order to show this, let us assume that $\ell^{\pm}$ is not $O_{\mathrm{P}}(1)$. Then, there exists $\epsilon > 0$ and a sequence $n_i \uparrow \infty$ such that, for all $L \in \mathbb{N}$, $\mathrm{P}^{(n_i)}_{\mathbf{V};g}[\ell^- > L] > \epsilon$. This and Pythagoras' Theorem implies that, for $L > c_2 \mathcal{J}_k^{-1}(f_1, g_1)$, with $\mathrm{P}^{(n_i)}_{\mathbf{V};g}$-probability larger than ϵ,

$$\begin{aligned}\left\|\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n_i)}_{f\#}(\alpha_L; \mathcal{J}_k^{-1}(f_1, g_1))) - \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n_i)}_{f\#})\right\| &\geq \left\|\overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n_i)}_{\#}(\alpha_L)) - \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n_i)}_{f\#})\right\| \\ &= n_i^{-1/2}(c_2^{-1}L - \mathcal{J}_k^{-1}(f_1, g_1)) \left\|\boldsymbol{\Upsilon}_k(\mathbf{V}^{(n_i)}_{\#})\, \underset{\sim}{\boldsymbol{\Delta}}^{(n_i)}_{f\#}(\mathbf{V}^{(n_i)}_{\#})\right\|,\end{aligned}$$

which is clearly incompatible with the fact that (3.22) holds for $\ell = L$. Thus, $\ell^{\pm}$ are $O_{\mathrm{P}}(1)$, and $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha^{\pm})$ also can serve as initial estimators in a one-step strategy.

The final step in the construction of our estimator $\widehat{\underset{\sim}{\mathbf{V}}}^{(n)}_{f\#}$ then is a "fine tuning" step, which consists in selecting an intermediate point between α^- and α^+. This intermediate value, as we shall see, turns out to consistently estimate $\mathcal{J}^{-1}(f_1, g_1)$. Denote by $\boldsymbol{\pi}^{(n)}_{\pm}(\delta)$ the projection on $\mathcal{D}^{(n)}_{\#}$ of $\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha^{\pm}; \delta))$, and let $\pi^{(n)}_{\pm}(\delta) := \|\boldsymbol{\pi}^{(n)}_{\pm}(\delta) - \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#})\|$. Note that $\delta \mapsto \pi^{(n)}_{-}(\delta)$ (resp., $\delta \mapsto \pi^{(n)}_{+}(\delta)$) is $\mathcal{P}^{(n)}$-a.e. continuous and strictly monotone increasing (resp., decreasing). Therefore, there exists a unique δ^* such that $\boldsymbol{\pi}^{(n)}_{-}(\delta^*) = \boldsymbol{\pi}^{(n)}_{+}(\delta^*)$. The proposed R-estimator of $\mathbf{V}$ is the shape matrix $\widehat{\underset{\sim}{\mathbf{V}}}^{(n)}_{f\#}$ characterized by $\overset{\circ}{\mathrm{vech}}(\widehat{\underset{\sim}{\mathbf{V}}}^{(n)}_{f\#}) := \boldsymbol{\pi}^{(n)}_{\pm}(\delta^*)$.

Let us show indeed that $\boldsymbol{\pi}^{(n)}_{\pm}(\delta^*) - \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}) = o_{\mathrm{P}}(n^{-1/2})$. Either $\pi^{(n)}_{-}(\mathcal{J}_k^{-1}(f_1, g_1)) \leq \pi^{(n)}_{+}(\mathcal{J}_k^{-1}(f_1, g_1))$, and $\pi^{(n)}_{-}(\mathcal{J}_k^{-1}(f_1, g_1)) \leq \pi^{(n)}_{\pm}(\delta^*) \leq \pi^{(n)}_{+}(\mathcal{J}_k^{-1}(f_1, g_1))$; or, $\pi^{(n)}_{-}(\mathcal{J}_k^{-1}(f_1, g_1)) > \pi^{(n)}_{+}(\mathcal{J}_k^{-1}(f_1, g_1))$, and $\pi^{(n)}_{+}(\mathcal{J}_k^{-1}(f_1, g_1)) < \pi^{(n)}_{\pm}(\delta^*) \leq \pi^{(n)}_{-}(\mathcal{J}_k^{-1}(f_1, g_1))$. In both cases, thus, $\boldsymbol{\pi}^{(n)}_{\pm}(\delta^*)$ belongs to the interval $[\boldsymbol{\pi}^{(n)}_{-}(\mathcal{J}_k^{-1}(f_1, g_1)), \boldsymbol{\pi}^{(n)}_{+}(\mathcal{J}_k^{-1}(f_1, g_1))]$. Now, both $\boldsymbol{\pi}^{(n)}_{-}(\mathcal{J}_k^{-1}(f_1, g_1))$ and $\boldsymbol{\pi}^{(n)}_{+}(\mathcal{J}_k^{-1}(f_1, g_1))$ are efficient estimators satisfying (3.8) and (3.9). Indeed, from Pythagoras' Theorem,

$$\left\|\boldsymbol{\pi}^{(n)}_{\pm}(\mathcal{J}_k^{-1}(f_1, g_1)) - \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#})\right\| \quad\leq\quad \left\|\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(\alpha_{\ell^{\pm}}; \mathcal{J}_k^{-1}(f_1, g_1))) - \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#})\right\| = o_{\mathrm{P}}(n^{-1/2}).$$

As a convex linear combination of $\boldsymbol{\pi}^{(n)}_{-}(\mathcal{J}_k^{-1}(f_1, g_1))$ and $\boldsymbol{\pi}^{(n)}_{+}(\mathcal{J}_k^{-1}(f_1, g_1))$, $\overset{\circ}{\mathrm{vech}}(\widehat{\underset{\sim}{\mathbf{V}}}^{(n)}_{f\#}) = \boldsymbol{\pi}^{(n)}_{\pm}(\delta^*)$ thus also is an efficient estimator satisfying (3.8) and (3.9). And, contrary to $\boldsymbol{\pi}^{(n)}_{\pm}(\mathcal{J}_k^{-1}(f_1, g_1))$, it is computable from the sample. As a by-product,

$$\hat{\mathcal{J}}^{(n)}_k(f_1) := n^{-1/2}\|\boldsymbol{\Upsilon}_k(\mathbf{V}^{(n)}_{\#})\, \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f\#}(\mathbf{V}^{(n)}_{\#})\| \Big/ \left\|\boldsymbol{\pi}^{(n)}_{\pm}(\delta^*) - \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#})\right\| \tag{3.24}$$

and $\left(\mathcal{J}_k(f_1)/(\hat{\mathcal{J}}_k^{(n)}(f_1))^2\right)\boldsymbol{\Upsilon}_k(\hat{\underset{\sim}{\mathbf{V}}}{}^{(n)}_{f\#})$ yield consistent estimators of $\mathcal{J}_k(f_1, g_1)$ and the asymptotic covariance matrix of the R-estimate $\overset{\circ}{\text{vech}}(\hat{\underset{\sim}{\mathbf{V}}}{}^{(n)}_{f\#})$, respectively.

3.5 Optimal one-step R-estimation: practical implementation.

As usual, the discretization technique which complicates the proofs of asymptotic results and obscures the definition of the estimator makes little sense in practice, where n is fixed. Also recall that the ranks $R_i^{(n)} = R_i^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \mathbf{V})$ and the signs $\mathbf{U}_i^{(n)} = \mathbf{U}_i^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \mathbf{V})$ in practice should take into account the estimation of location.

Discretization in the previous sections was achieved in three steps: discretization of Tyler's $\mathbf{V}_T^{(n)}$ into $\mathbf{V}_\#^{(n)}$ (based on c_0), discretization of $\underset{\sim}{\boldsymbol{\Delta}}{}_f^{(n)}(\mathbf{V}_\#^{(n)})$ into $\underset{\sim}{\boldsymbol{\Delta}}{}_{f\#}^{(n)}(\mathbf{V}_\#^{(n)})$ (based on c_1), and discretization of α into α_ℓ (based on c_2). The "undiscretized version" $\hat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$ of $\hat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}$ corresponds to arbitrarily large values of these three discretization constants, leaving $\mathbf{V}_T^{(n)}$ and $\underset{\sim}{\boldsymbol{\Delta}}{}_f^{(n)}$ unchanged, and bringing (for the sample size at hand) α^+ and α^- so close to each other that the final tuning (involving the solution δ^* of $\boldsymbol{\pi}_-^{(n)}(\delta) = \boldsymbol{\pi}_+^{(n)}(\delta)$) becomes numerically meaningless. Alternatively, denoting by $\hat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}(\mathbf{c})$ the estimator associated with the discretization constants $\mathbf{c} = (c_0, c_1, c_2)$, we have $\hat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)} := \lim_{\mathbf{c}\to\infty} \hat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}(\mathbf{c})$, where $\mathbf{c} \to \infty$ means that $c_i \to \infty$ for $i = 0$, 1, and 2.

This practical implementation $\hat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$ of $\hat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}$ can be obtained more directly as follows. Letting

$$\overset{\circ}{\text{vech}}(\underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha)) := \overset{\circ}{\text{vech}}(\mathbf{V}_T^{(n)}) + n^{-1/2}\alpha\, \boldsymbol{\Upsilon}_k(\mathbf{V}_T^{(n)})\, \underset{\sim}{\boldsymbol{\Delta}}{}_f^{(n)}(\mathbf{V}_T^{(n)}), \quad \alpha \in \mathbb{R}^+$$

(the undiscretized version of $\overset{\circ}{\text{vech}}(\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(\alpha_\ell))$), consider the $\mathcal{P}^{(n)}$-a.e. piecewise continuous function

$$\alpha \mapsto h(\alpha) := \left(\underset{\sim}{\boldsymbol{\Delta}}{}_f^{(n)}(\mathbf{V}_T^{(n)})\right)' \boldsymbol{\Upsilon}_k(\mathbf{V}_T^{(n)})\, \boldsymbol{\Upsilon}_k(\underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha))\, \underset{\sim}{\boldsymbol{\Delta}}{}_f^{(n)}(\underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha)), \quad \alpha \in \mathbb{R}^+, \tag{3.25}$$

and put $\alpha^* := \inf\{\alpha > 0 \,|\, h(\alpha) \le 0\}$, $\alpha^{*-} := \alpha^* - 0$, and $\alpha^{*+} := \alpha^* + 0$. The matrices $\underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha^{*-})$ and $\underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha^{*+})$ clearly are the "undiscretized counterparts" of $\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(\alpha^-)$ and $\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(\alpha^+)$, respectively. However, $\alpha \mapsto \underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha)$ being continuous, $\underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha^{*-}) = \underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha^{*+})$. The proposed estimator, in Section 3.4, lies between $\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(\alpha^-)$ and $\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(\alpha^+)$; accordingly, the R-estimator we are proposing in practice is $\hat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)} := \underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha^*) = \underset{\sim}{\mathbf{V}}{}_f^{(n)}(\alpha^{*\pm})$, while α^* provides the corresponding estimator of $\mathcal{J}_k^{-1}(f_1, g_1)$—the "undiscretized" version of (3.24).

Let us stress however the fact that all asymptotic properties—among which asymptotic optimality—belong to the discretized estimators $\hat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}$, whereas nothing can be said about the asymptotics of the practical implementation $\hat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$.

4 Asymptotic affine-equivariance.

An estimator $\hat{\mathbf{V}}^{(n)}$ of the shape matrix $\mathbf{V}$ is said to be (strictly, that is, for any fixed n) affine-equivariant iff, for all invertible $k \times k$ matrix $\mathbf{M}$,

$$\hat{\mathbf{V}}^{(n)}(\mathbf{M}) = \left(\mathbf{M}\hat{\mathbf{V}}^{(n)}\mathbf{M}'\right) \big/ \left(\mathbf{M}\hat{\mathbf{V}}^{(n)}\mathbf{M}'\right)_{11}, \tag{4.1}$$

where we denote by $\widehat{\mathbf{V}}^{(n)}(\mathbf{M})$ the value of the statistic $\widehat{\mathbf{V}}^{(n)}$ computed from the transformed sample $\mathbf{MX}_1, \ldots, \mathbf{MX}_n$. Note that, if $\widehat{\mathbf{V}}^{(n)}$ is affine-equivariant, then its square-root satisfies the equivariance relation

$$(\widehat{\mathbf{V}}^{(n)}(\mathbf{M}))^{1/2} = d\,\mathbf{M}(\widehat{\mathbf{V}}^{(n)})^{1/2}\mathbf{O} \tag{4.2}$$

for some positive scalar d and some $k \times k$ orthogonal matrix $\mathbf{O}$ (see Randles 2000, page 1267 for a proof). Both Tyler's estimator $\mathbf{V}_T^{(n)}$ and the Gaussian estimator $\mathbf{V}_{\mathcal{G}}^{(n)}$ are affine-equivariant. Unfortunately, the final estimator $\widehat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$ proposed in Section 3.5 is not.

One could wonder whether $\widehat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$ at least is *asymptotically* affine-equivariant, that is, whether $\widehat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$ is asymptotically equivalent to some stricly affine-equivariant sequence—not necessarily a sequence of estimators: for all practical purposes, a sequence of pseudo-estimators, or simply a sequence of random shape matrices would be fine. A closer inspection of this idea however reveals a major conceptual problem. Recall indeed that all asymptotic results belong to the discretized estimators $\widehat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}$, while nothing can be said about the asymptotics of $\widehat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$: a definition of asymptotic equivariance relying on the asymptotic behavior of $\widehat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$ is thus totally ineffective.

We therefore propose the following, somewhat weaker, definition. Denote by

$$\mathcal{S}^{(n)} := \left\{\mathbf{S}_m^{(n)}(\mathbf{X}^{(n)})\,|\,m \in \mathbb{N}\right\} \quad \text{and} \quad \mathcal{T}^{(n)} := \left\{\mathbf{T}_m^{(n)}(\mathbf{X}^{(n)})\,|\,m \in \mathbb{N}\right\}, \quad n \in \mathbb{N},$$

two countable classes of sequences of $\mathbf{X}^{(n)}$-measurable random vectors. Assume that both classes are asymptotically equivalent as $n \to \infty$, meaning that

(i) for all m, $\mathbf{S}_m^{(n)}(\mathbf{X}^{(n)}) - \mathbf{T}_m^{(n)}(\mathbf{X}^{(n)}) = o_{\mathrm{P}}(n^{-1/2})$ as $n \to \infty$.

Assume moreover that

(ii) the almost sure limits $\mathbf{S}^{(n)} := \lim_{m\to\infty} \mathbf{S}_m^{(n)}(\mathbf{X}^{(n)})$ and $\mathbf{T}^{(n)} := \lim_{m\to\infty} \mathbf{T}_m^{(n)}(\mathbf{X}^{(n)})$ exist for all fixed n.

Then, if, for fixed n, $\mathbf{S}^{(n)}$ is equivariant (in the sense, for instance, of (4.1)), we may consider that $\mathbf{T}^{(n)}$ somehow inherits, in an approximate or asymptotic form, this equivariance property: we say that $\mathbf{T}^{(n)}$ is *weakly asymptotically equivariant*.

In order to show that the proposed estimators $\widehat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)} := \lim_{\mathbf{c}\to\infty} \widehat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}(\mathbf{c})$ are weakly asymptotically affine-equivariant, consider the class $\mathcal{T}^{(n)} := \{\widehat{\underset{\sim}{\mathbf{V}}}{}_{f\#}^{(n)}(\mathbf{c}_m) | m \in \mathbb{N}\}$, where the sequence $\mathbf{c}_m = (c_{m,0}, c_{m,1}, c_{m,2})$ is such that $\lim_{m\to\infty} \mathbf{c}_m = \infty$, and let us construct a class $\mathcal{S}^{(n)}$ such that conditions (i) and (ii) for weak asymptotic equivariance are satisfied. Incidently, note that a choice of the form $\mathcal{S}^{(n)} := \{\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(c_{0,m}) | m \in \mathbb{N}\}$ (with $c_{0,m} \to \infty$), where $\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(c_0)$ denotes the pseudo-estimator defined in (3.7), is not suitable, since the corresponding practical implementation $\underset{\sim}{\mathbf{V}}{}_f^{(n)} := \lim_{c_0\to\infty} \underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}(c_0)$ is not strictly affine-equivariant.

Inspired by $\underset{\sim}{\mathbf{V}}{}_{f\#}^{(n)}$'s representation (3.15) as a linear combination of $\mathbf{V}_{\#}^{(n)}$ and the rank-based shape matrix $\underset{\sim}{\mathbf{W}}{}_{f\#}^{(n)}$ defined in (3.16), rather consider

$$\underset{\approx}{\mathbf{V}}{}_{f\#}^{(n)} = \underset{\approx}{\mathbf{V}}{}_{f\#}^{(n)}(c_0) := \underset{\sim}{\mathbf{B}}{}_{f\#}^{(n)} \Big/ (\underset{\sim}{\mathbf{B}}{}_{f\#}^{(n)})_{11}, \text{ with } \underset{\sim}{\mathbf{B}}{}_{f\#}^{(n)} := \left(1 - \frac{k(k+2)}{\mathcal{J}_k(f_1, g_1)}\right)\mathbf{V}_{\#}^{(n)} + \frac{k(k+2)}{\mathcal{J}_k(f_1, g_1)}\,\underset{\sim}{\mathbf{W}}{}_{f\#}^{(n)}, \tag{4.3}$$

where c_0 denotes the constant used in the discretization of Tyler's estimator. The class $\mathcal{S}^{(n)}$ allowing to establish the weak asymptotic affine-equivariance of $\widehat{\underset{\sim}{\mathbf{V}}}{}_f^{(n)}$ is then $\mathcal{S}^{(n)} := \{\underset{\approx}{\mathbf{V}}{}_{f\#}^{(n)}(c_{0,m})\}$.

Because of discretization, neither $\mathbf{V}_{\#}^{(n)}$ nor $\underset{\approx}{\mathbf{V}}_{f\#}^{(n)}$ are affine-equivariant for fixed n. However, in view of (4.2), one easily can check that the practical implementation $\underset{\approx}{\mathbf{V}}_{f}^{(n)} := \lim_{m\to\infty} \underset{\approx}{\mathbf{V}}_{f\#}^{(n)}(c_{0,m})$ (which is based on $\mathbf{V}_{T}^{(n)}$ and $\underset{\sim}{\mathbf{W}}_{f}^{(n)}(\mathbf{V}_{T}^{(n)})$ instead of $\mathbf{V}_{\#}^{(n)}$ and $\underset{\sim}{\mathbf{W}}_{f\#}^{(n)}$) is. The weak asymptotic affine-equivariance of $\hat{\underset{\sim}{\mathbf{V}}}_{f}^{(n)}$ thus readily follows from the following proposition.

Proposition 4.1 *Denote by* $\underset{\approx}{\mathbf{V}}_{f\#}^{(n)} := \underset{\approx}{\mathbf{V}}_{f\#}^{(n)}(c_0)$ *and by* $\hat{\underset{\sim}{\mathbf{V}}}_{f\#}^{(n)} := \hat{\underset{\sim}{\mathbf{V}}}_{f\#}^{(n)}(\mathbf{c})$ *the pseudo-estimator defined in (4.3) and the estimator defined in Section 3.4, respectively. Then,* $\underset{\approx}{\mathbf{V}}_{f\#}^{(n)} - \hat{\underset{\sim}{\mathbf{V}}}_{f\#}^{(n)} = o_{\mathrm{P}}(n^{-1/2})$ *under* $\mathcal{P}^{(n)}$*, as* $n\to\infty$.

Proof. See Section 6.3. □

Whether weak asymptotic equivariance is a satisfactory property or not is a matter of statistical taste. If it is, this section shows that $\hat{\underset{\sim}{\mathbf{V}}}_{f}^{(n)}$ is the estimator to be used. The reader who feels that strict equivariance of the practical implementation is an essential requirement is referred to Hallin, Oja and Paindaveine (2004), where we show that an adequate modification of $\hat{\underset{\sim}{\mathbf{V}}}_{f}^{(n)}$ into a strictly equivariant $\hat{\underset{\approx}{\mathbf{V}}}_{f}^{(n)}$ is possible—at the price, however, of some technicalities, and a weakening of the relation to the class of optimal discretized estimators $\left\{\hat{\underset{\sim}{\mathbf{V}}}_{f\#}^{(n)}(\mathbf{c}) \,|\, \mathbf{c}\in(\mathbb{R}_0^+)^3\right\}$.

5 Simulations.

In this section, we conduct a Monte-Carlo study in order to compare the finite-sample performances of the one-step R-estimators $\hat{\underset{\sim}{\mathbf{V}}}_{f}^{(n)}$ proposed in Section 3.5 to those of Tyler's estimator $\mathbf{V}_{T}^{(n)}$ and the Gaussian estimator $\mathbf{V}_{\mathcal{G}}^{(n)}$. We restrict to the bivariate spherical case ($\mathbf{V}=\mathbf{I}_2$). We generated $M = 1,000$ samples of i.i.d. observations $\mathbf{X}_1,\ldots,\mathbf{X}_n$ with sizes $n = 250$ and $n = 50$, from the bivariate standard normal ($\mathcal{N}$), Student distributions ($t_{0.5}$), (t_3), and (t_{10}) (with 0.5, 3, and 10 degrees of freedom), and power-exponential distributions (e_3) and (e_5) (with parameters $\eta = 3$ and 5); recall that power-exponential distributions are associated with standardized radial densities of the form $f_1(r) = f_{1,\eta}^{e}(r) := \exp(-b_{k,\eta} r^{2\eta})$, where $b_{k,\eta} > 0$ is such that (2.3) is satisfied. Student and power-exponential distributions allow for considering heavier-than-normal and lighter-than-normal tail distributions, respectively.

For each replication, we computed $\mathbf{V}_{T}^{(n)}$, $\mathbf{V}_{\mathcal{G}}^{(n)}$, and the $\mathbf{V}_{T}^{(n)}$- and $\mathbf{V}_{\mathcal{G}}^{(n)}$-based one-step R-estimators $\hat{\underset{\sim}{\mathbf{V}}}_{\mathrm{vdW}}^{(n)}$, $\hat{\underset{\sim}{\mathbf{V}}}_{0.5}^{(n)}$, $\hat{\underset{\sim}{\mathbf{V}}}_{3}^{(n)}$, and $\hat{\underset{\sim}{\mathbf{V}}}_{10}^{(n)}$ corresponding to semiparametric efficiency at Gaussian and Student densities with .5, 3, and 10 degrees of freedom, respectively. In Tables 2 (sample size $n = 250$) and 3 (sample size $n = 50$), we report, for each of these estimates, the two components of the average bias

$$\mathrm{BIAS}^{(n)} := \frac{1}{M}\sum_{i=1}^{M} \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}(i)) = \frac{1}{M}\sum_{i=1}^{M}\left(V_{12}^{(n)}(i), V_{22}^{(n)}(i)-1\right)'$$

and the two components of the mean square error

$$\mathrm{MSE}^{(n)} := \frac{1}{M}\sum_{i=1}^{M}\left((V_{12}^{(n)}(i))^2, (V_{22}^{(n)}(i)-1)^2\right)'.$$

These simulations show that the proposed rank-based estimators behave remarkably well under all distributions under consideration and significantly improve on Tyler's estimator. They confirm the optimality of the Tyler-based f-score R-estimators under density f, and essentially agree with the ARE rankings presented in Table 1. Also, the van der Waerden rank-based estimator (based on preliminary estimator $\mathbf{V}_T^{(n)}$ or $\mathbf{V}_{\mathcal{G}}^{(n)}$) uniformly dominates the parametric Gaussian estimator $\mathbf{V}_{\mathcal{G}}^{(n)}$, and competes evenly with it in the normal case; this dominance over $\mathbf{V}_{\mathcal{G}}^{(n)}$, which is observed both under lighter-than-normal and under heavier-than-normal tail distributions, provides an empirical validation of the Chernoff-Savage result established in Paindaveine (2004).

The behavior of one-step rank-based estimators does not seem to depend very much on the preliminary estimator used ($\mathbf{V}_T^{(n)}$ or $\mathbf{V}_{\mathcal{G}}^{(n)}$), which confirms that the influence of the preliminary estimator is asymptotically nil. More surprising is the fact that R-estimator based on $\mathbf{V}_{\mathcal{G}}^{(n)}$ behave reasonably well under heavy tails (under $t_{0.5}$), although $\mathbf{V}_{\mathcal{G}}^{(n)}$ is not even root-n consistent there (which explains the total collapse under $t_{0.5}$ of $\mathbf{V}_{\mathcal{G}}^{(n)}$).

These conclusions are equally valid for small ($n = 50$) as for large ($n = 250$) sample sizes.

6 Appendix.

6.1 Local asymptotic linearity.

Rather than Proposition 3.1(v), we actually prove in this section a more general asymptotic linearity result in which not only the shape, but also the location parameter, are perturbed locally.

Proposition 6.1 *For any bounded sequence of k-dimensional vectors $\mathbf{t}^{(n)}$ and symmetric matrices $\mathbf{v}^{(n)}$ satisfying $v_{11}^{(n)} = 0$, and for any $g \in \mathcal{F}_{\mathrm{A}}$, the central sequence $\underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\boldsymbol{\theta}, \mathbf{V})$ satisfies, under $\mathrm{P}_{\boldsymbol{\theta},\mathbf{V};g}^{(n)}$, as $n \to \infty$, the asymptotic linearity property*

$$\underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}^{(n)}, \mathbf{V} + n^{-1/2}\mathbf{v}^{(n)}) - \underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\boldsymbol{\theta}, \mathbf{V}) = -\mathbf{\Gamma}_{f_1,g_1}^*(\mathbf{V})\, \overset{\circ}{\mathrm{vech}}(\mathbf{v}^{(n)}) + o_{\mathrm{P}}(1). \tag{6.4}$$

The proof of Proposition 6.1 relies on a series of lemmas. In this section, we let $\boldsymbol{\theta}^n := \boldsymbol{\theta} + n^{-1/2}\mathbf{t}^{(n)}$ and $\mathbf{V}^n := \mathbf{V} + n^{-1/2}\mathbf{v}^{(n)}$. Accordingly, let $\mathbf{Z}_i^0 := \mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})$, $d_i^0 := \|\mathbf{Z}_i^0\|$, $\mathbf{U}_i^0 := \mathbf{Z}_i^0/d_i^0$, $\mathbf{Z}_i^n := (\mathbf{V}^n)^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta}^n)$, $d_i^n := \|\mathbf{Z}_i^n\|$, and $\mathbf{U}_i^n := \mathbf{Z}_i^n/d_i^n$. We begin with the following preliminary result.

Lemma 6.1 *For all i, as $n \to \infty$, under $\mathrm{P}_{\boldsymbol{\theta},\mathbf{V};g}^{(n)}$*

(i) $|d_i^n - d_i^0| = o_{\mathrm{P}}(1)$ and

(ii) $\|\mathbf{U}_i^n - \mathbf{U}_i^0\| = o_{\mathrm{P}}(1)$.

Proof of Lemma 6.1. First note that, denoting by $\|\mathbf{M}\|_{\mathcal{L}} := \sup\{\|\mathbf{M}\mathbf{x}\| \mid \|\mathbf{x}\| = 1\}$ the operator norm of the square matrix $\mathbf{M}$, we have

$$\begin{aligned}
\|\mathbf{Z}_i^n - \mathbf{Z}_i^0\| &\leq \|(\mathbf{V}^n)^{-1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}^n)\| + \|((\mathbf{V}^n)^{-1/2} - \mathbf{V}^{-1/2})(\mathbf{X} - \boldsymbol{\theta})\| \\
&\leq n^{-1/2}\|(\mathbf{V}^n)^{-1/2}\|_{\mathcal{L}}\, \|\mathbf{t}^{(n)}\| + \|(\mathbf{V}^n)^{-1/2} - \mathbf{V}^{-1/2}\|_{\mathcal{L}}\, \|\mathbf{V}^{1/2}\|_{\mathcal{L}}\, d_i^0, \\
&\leq C(n)\,(1 + d_i^0),
\end{aligned}$$

for some positive sequence $C(n)$, with $C(n) = o(1)$ as $n \to \infty$. Now, since, for all $\delta > 0$, $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}[C(n)(d^0_i)^a > \delta] = o(1)$ as $n \to \infty$ ($a = -1, 0, 1$), we obtain that $\|\mathbf{Z}^n_i - \mathbf{Z}^0_i\|$ and $\|\mathbf{Z}^n_i - \mathbf{Z}^0_i\|/d^0_i$ are $o_\mathrm{P}(1)$ as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$. The result follows since (i) $|d^n_i - d^0_i| \leq \|\mathbf{Z}^n_i - \mathbf{Z}^0_i\|$ and (ii) $\|\mathbf{U}^n_i - \mathbf{U}^0_i\| \leq |(1/d^n_i - 1/d^0_i)|\,\|\mathbf{Z}^n_i\| + \|\mathbf{Z}^n_i - \mathbf{Z}^0_i\|/d^0_i \leq 2\,\|\mathbf{Z}^n_i - \mathbf{Z}^0_i\|/d^0_i$. □

Proof of Proposition 6.1. We first consider the following truncation of the score function K_{f_1}. For all $\ell \in \mathbb{N}_0$, define

$$K^{(\ell)}_{f_1}(u) := K_{f_1}\left(\frac{1}{\ell}\right) I_{[u\leq\frac{1}{\ell}]} + K_{f_1}(u)\, I_{[\frac{1}{\ell}<u\leq 1-\frac{1}{\ell}]} + K_{f_1}\left(1-\frac{1}{\ell}\right) I_{[u>1-\frac{1}{\ell}]},$$

where I_A denotes the indicator function of A. Since $u \mapsto K_{f_1}(u)$ is continuous, the functions $u \mapsto K^{(\ell)}_{f_1}(u)$ are also continuous on $(0, 1)$. It follows that—even for unbounded scores K_{f_1}—the truncated scores $K^{(\ell)}_{f_1}$ are bounded for all ℓ. Clearly, it can safely be assumed that K_{f_1} is a monotone increasing function (rather than the difference of two monotone increasing functions), so that (at least for ℓ sufficiently large) $|K^{(\ell)}_{f_1}|$ is bounded by $|K_{f_1}|$ uniformly in ℓ and u, i.e., there exists some L such that $|K^{(\ell)}_{f_1}(u)| \leq |K_{f_1}(u)|$ for all $u \in (0, 1)$ and all $\ell \geq L$.

We have to prove that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, as $n \to \infty$,

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta}^n, \mathbf{V}^n) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta}, \mathbf{V}) + \mathcal{J}_k(f_1, g_1)\, \boldsymbol{\Upsilon}^{-1}_k(\mathbf{V})\, \overset{\circ}{\mathrm{vech}}(\mathbf{v}^{(n)}) \tag{6.5}$$

is $o_\mathrm{P}(1)$. Proposition 3.1*(ii)* shows that $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta}, \mathbf{V}) - \boldsymbol{\Delta}^{(n)*}_{f,g}(\boldsymbol{\theta}, \mathbf{V})$ is $o_\mathrm{P}(1)$, as $n \to \infty$, under the same sequence of hypotheses. Similarly,

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta}^n, \mathbf{V}^n) - \boldsymbol{\Delta}^{(n)*}_{f,g}(\boldsymbol{\theta}^n, \mathbf{V}^n) \tag{6.6}$$

is $o_\mathrm{P}(1)$ as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^n,\mathbf{V}^n;g}$. It follows from contiguity that (6.6) is also $o_\mathrm{P}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, as $n \to \infty$. Consequently, the difference (6.5) is asymptotically equivalent, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, to

$$\boldsymbol{\Delta}^{(n)*}_{f,g}(\boldsymbol{\theta}^n, \mathbf{V}^n) - \boldsymbol{\Delta}^{(n)*}_{f,g}(\boldsymbol{\theta}, \mathbf{V}) + \mathcal{J}_k(f_1, g_1)\, \boldsymbol{\Upsilon}^{-1}_k(\mathbf{V})\, \overset{\circ}{\mathrm{vech}}(\mathbf{v}^{(n)}). \tag{6.7}$$

Now, putting $\mathbf{J}^\perp_k := \mathbf{I}_{k^2} - (1/k)\mathbf{J}_k$, $n^{-1/2}\mathbf{J}^\perp_k\, \mathrm{vec}\left[\sum_{i=1}^n K_{f_1}(\tilde{G}_{1k}(d^n_i/\sigma))\, \mathbf{U}^n_i \mathbf{U}^{n\prime}_i\right]$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^n,\mathbf{V}^n;g}$, is asymptotically k^2-normal as $n \to \infty$, with mean zero and covariance matrix $(k(k+2))^{-1}$ $\mathcal{J}_k(f_1)[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k}\mathbf{J}_k]$, so that

$$\frac{1}{2}\, n^{-1/2}\mathbf{M}_k \left[\left((\mathbf{V}^n)^{\otimes 2}\right)^{-1/2} - \left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\right] \mathbf{J}^\perp_k\, \mathrm{vec}\left[\sum_{i=1}^n K_{f_1}(\tilde{G}_{1k}(d^n_i/\sigma))\, \mathbf{U}^n_i \mathbf{U}^{n\prime}_i\right]$$

is $o_\mathrm{P}(1)$, as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^n,\mathbf{V}^n;g}$, as well as under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$ (by contiguity). Consequently, (6.7) is asymptotically equivalent, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, to

$$\mathbf{C}^{(n)} := \frac{1}{2}\, n^{-1/2}\mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}^\perp_k\, \mathrm{vec}\left[\sum_{i=1}^n K_{f_1}(\tilde{G}_{1k}(d^n_i/\sigma))\, \mathbf{U}^n_i \mathbf{U}^{n\prime}_i\right] \tag{6.8}$$

$$-\frac{1}{2}\, n^{-1/2}\mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}^\perp_k\, \mathrm{vec}\left[\sum_{i=1}^n K_{f_1}(\tilde{G}_{1k}(d^0_i/\sigma))\, \mathbf{U}^0_i \mathbf{U}^{0\prime}_i\right] + \mathcal{J}_k(f_1, g_1)\, \boldsymbol{\Upsilon}^{-1}_k(\mathbf{V})\, (\overset{\circ}{\mathrm{vech}}\, \mathbf{v}^{(n)}),$$

and we only have to prove that $\mathbf{C}^{(n)}$ is $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, as $n \to \infty$. Now, decompose $\mathbf{C}^{(n)}$ into $\mathbf{C}^{(n)} = \mathbf{D}_1^{(n;\ell)} + \mathbf{D}_2^{(n;\ell)} - \mathbf{R}_1^{(n;\ell)} + \mathbf{R}_2^{(n;\ell)} + \mathbf{R}_3^{(n;\ell)}$ where, denoting by E_0 the expectation under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$ and defining $\mathcal{J}_k^{(\ell)}(f_1; g_1) := \int_0^1 K_{f_1}^{(\ell)}(u) K_{g_1}(u)\, du$,

$$\mathbf{D}_1^{(n;\ell)} := \frac{1}{2}\, n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp} \operatorname{vec}\left[\sum_{i=1}^n K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^n/\sigma))\, \mathbf{U}_i^n \mathbf{U}_i^{n\prime}\right]$$
$$- \frac{1}{2}\, n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp} \operatorname{vec}\left[\sum_{i=1}^n K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^0/\sigma))\, \mathbf{U}_i^0 \mathbf{U}_i^{0\prime}\right]$$
$$- \frac{1}{2}\, n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp}\, \mathrm{E}_0\left[\operatorname{vec}\left[\sum_{i=1}^n K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^n/\sigma))\, \mathbf{U}_i^n \mathbf{U}_i^{n\prime}\right]\right],$$

$$\mathbf{D}_2^{(n;\ell)} := \frac{1}{2}\, n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp}\, \mathrm{E}_0\left[\operatorname{vec}\left[\sum_{i=1}^n K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^n/\sigma))\, \mathbf{U}_i^n \mathbf{U}_i^{n\prime}\right]\right]$$
$$+ \mathcal{J}_k^{(\ell)}(f_1; g_1)\, \boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\, (\overset{\circ}{\operatorname{vech}}\, \mathbf{v}^{(n)}),$$

$$\mathbf{R}_1^{(n;\ell)} := \frac{1}{2}\, n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp} \operatorname{vec}\left[\sum_{i=1}^n \left[K_{f_1}(\tilde{G}_{1k}(d_i^0/\sigma)) - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^0/\sigma))\right] \mathbf{U}_i^0 \mathbf{U}_i^{0\prime}\right],$$

$$\mathbf{R}_2^{(n;\ell)} := \frac{1}{2}\, n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp} \operatorname{vec}\left[\sum_{i=1}^n \left[K_{f_1}(\tilde{G}_{1k}(d_i^n/\sigma)) - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^n/\sigma))\right] \mathbf{U}_i^n \mathbf{U}_i^{n\prime}\right],$$

and

$$\mathbf{R}_3^{(n;\ell)} := \left(\mathcal{J}_k(f_1, g_1) - \mathcal{J}_k^{(\ell)}(f_1; g_1)\right) \boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\, (\overset{\circ}{\operatorname{vech}}\, \mathbf{v}^{(n)}).$$

We prove that $\mathbf{C}^{(n)} = o_{\mathrm{P}}(1)$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, as $n \to \infty$ (thus completing the proof of Proposition 6.1) by establishing that $\mathbf{D}_1^{(n;\ell)}$ and $\mathbf{D}_2^{(n;\ell)}$ are $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, as $n \to \infty$, for fixed ℓ, and that $\mathbf{R}_1^{(n;\ell)}$, $\mathbf{R}_2^{(n;\ell)}$ and $\mathbf{R}_3^{(n;\ell)}$ are $o_{\mathrm{P}}(1)$ under the same sequence of hypotheses, as $\ell \to \infty$, uniformly in n. For the sake of convenience, these three results are treated as separate lemmas (Lemmas 6.2 and 6.3, and Lemma 6.4, respectively).

Lemma 6.2 *For any fixed* ℓ, $\mathrm{E}_0\left[\left\|\mathbf{D}_1^{(n;\ell)}\right\|^2\right] = o(1)$ *as* $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$.

Lemma 6.3 *For any fixed* ℓ, $\mathbf{D}_2^{(n;\ell)} = o(1)$ *as* $n \to \infty$.

Lemma 6.4 (i) *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, $\mathbf{R}_1^{(n;\ell)}$ *is* $o_{\mathrm{P}}(1)$ *as* $\ell \to \infty$, *uniformly in* n.

(ii) *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, $\mathbf{R}_2^{(n;\ell)}$ *is* $o_{\mathrm{P}}(1)$ *as* $\ell \to \infty$, *uniformly in* n (*for* n *sufficiently large*).

(iii) $\mathbf{R}_3^{(n;\ell)}$ *is* $o(1)$ *as* $\ell \to \infty$, *uniformly in* n.

Proof of Lemma 6.2. First note that

$$\mathbf{D}_1^{(n;\ell)} = \frac{1}{2}\, n^{-1/2}\mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp} \sum_{i=1}^{n} \left[\mathbf{T}_i^{(n;\ell)} - \mathrm{E}_0[\mathbf{T}_i^{(n;\ell)}]\right],$$

where $\mathbf{T}_i := \mathrm{vec}\left[K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^n/\sigma))\,\mathbf{U}_i^n\mathbf{U}_i^{n\prime} - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^0/\sigma))\,\mathbf{U}_i^0\mathbf{U}_i^{0\prime}\right]$, $i = 1, \ldots, n$ are i.i.d. under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$. Consequently, $\mathbf{D}_1^{(n;\ell)}$ is centered under the same sequence of hypotheses, and therefore (writing Var_0 for variances under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$)

$$\begin{aligned}
\mathrm{E}_0\left[\left\|\mathbf{D}_1^{(n;\ell)}\right\|^2\right] &= \mathrm{tr}\left[\mathrm{E}_0\left[\left(\mathbf{D}_1^{(n;\ell)}\right)\left(\mathbf{D}_1^{(n;\ell)}\right)'\right]\right] = \mathrm{tr}\left[\mathrm{Var}_0\left[\mathbf{D}_1^{(n;\ell)}\right]\right] \\
&\le C n^{-1}\mathrm{tr}\left[\mathrm{Var}_0\left[\sum_{i=1}^{n}\mathbf{T}_i^{(n;\ell)}\right]\right] = C\,\mathrm{tr}\left[\mathrm{Var}_0\left[\mathbf{T}_1^{(n;\ell)}\right]\right] \\
&= C\,\mathrm{E}_0[\,(\mathbf{T}_1 - \mathrm{E}_0[\mathbf{T}_1])'\,(\mathbf{T}_1 - \mathrm{E}_0[\mathbf{T}_1])] \le C\,\mathrm{E}_0[\|\mathbf{T}_1\|^2\,],
\end{aligned}$$

so that it remains to show that

$$\mathrm{E}_0\left[\|\mathbf{T}_1\|^2\right] = \mathrm{E}_0\left[\left\|K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^n/\sigma))\,\mathrm{vec}\left[\mathbf{U}_1^n\mathbf{U}_1^{n\prime}\right] - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^0/\sigma))\,\mathrm{vec}\left[\mathbf{U}_1^0\mathbf{U}_1^{0\prime}\right]\right\|^2\right] = o(1) \tag{6.9}$$

as $n \to \infty$.

Now, noting that $\|\mathrm{vec}\,(\mathbf{u}\mathbf{v}')\| = \|\mathbf{u}\|\|\mathbf{v}\|$, we have

$$\begin{aligned}
&\left\|K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^n/\sigma))\,\mathrm{vec}\,[\mathbf{U}_1^n\mathbf{U}_1^{n\prime}] - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^0/\sigma))\,\mathrm{vec}\,[\mathbf{U}_1^0\mathbf{U}_1^{0\prime}]\right\|^2 \\
&\quad\le 2\,|K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^n/\sigma)) - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^0/\sigma))|^2\,\|\mathrm{vec}\,[\mathbf{U}_1^n\mathbf{U}_1^{n\prime}]\|^2 \\
&\qquad\qquad + 2\,|K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^0/\sigma))|^2\,\|\mathrm{vec}\,[\mathbf{U}_1^n\mathbf{U}_1^{n\prime} - \mathbf{U}_1^0\mathbf{U}_1^{0\prime}]\|^2 \\
&\quad\le C\,|K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^n/\sigma)) - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^0/\sigma))|^2 + C\,\|\mathbf{U}_1^n - \mathbf{U}_1^0\|^2,
\end{aligned}$$

for some constant C. Lemma 6.1(i) and the continuity of $K_{f_1}^{(\ell)} \circ \tilde{G}_{1k}$ imply that $K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^n/\sigma)) - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_1^0/\sigma)) = o_{\mathrm{P}}(1)$ as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$. Since $K_{f_1}^{(\ell)}$ is bounded, this convergence to zero also holds in quadratic mean. Similarly, using Lemma 6.1(ii) and the boundedness of $\mathbf{U}_1^0$ and $\mathbf{U}_1^n$, we obtain that $\|\mathbf{U}_1^n - \mathbf{U}_1^0\|$ is $o(1)$ in quadratic mean, as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$. The convergence in (6.9) follows. □

Proof of Lemma 6.3. Letting

$$\mathbf{B}_1^{(n;\ell)} := \frac{1}{2}\, n^{-1/2}\mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp}\,\mathrm{vec}\left[\sum_{i=1}^{n} K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^0/\sigma))\,\mathbf{U}_i^0\mathbf{U}_i^{0\prime}\right],$$

one can show that

$$\mathbf{B}_1^{(n;\ell)} \xrightarrow{\mathcal{L}} \mathcal{N}_{k^2}\left(\mathbf{0}, \mathrm{E}[(K_{f_1}^{(\ell)}(U))^2]\,\boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\right) \tag{6.10}$$

as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$. Under the sequence of local alternatives $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^n,\mathbf{V}^n;g}$, as $n \to \infty$,

$$\mathbf{B}_1^{(n;\ell)} - \mathcal{J}_k^{(\ell)}(f_1, g_1)\,\boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\,(\overset{\circ}{\mathrm{vech}}\,\mathbf{v}^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}_{k^2}\left(\mathbf{0}, \mathrm{E}[(K_{f_1}^{(\ell)}(U))^2]\,\boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\right).$$

Defining $\mathbf{B}_2^{(n;\ell)} := \frac{1}{2}\, n^{-1/2} \mathbf{M}_k \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \mathbf{J}_k^{\perp} \operatorname{vec}\left[\sum_{i=1}^n K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^n/\sigma))\, \mathbf{U}_i^n \mathbf{U}_i^{n\prime}\right]$, it follows from uniform local asymptotic normality that

$$\mathbf{B}_2^{(n;\ell)} + \mathcal{J}_k^{(\ell)}(f_1, g_1)\, \boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\, (\overset{\circ}{\operatorname{vech}}\, \mathbf{v}^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}_{k^2}\left(\mathbf{0}, \mathrm{E}[(K_{f_1}^{(\ell)}(U))^2]\, \boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\right), \tag{6.11}$$

as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$.

Now, from (6.10) and the fact that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, $\mathbf{D}_1^{(n;\ell)} = \mathbf{B}_2^{(n;\ell)} - \mathbf{B}_1^{(n;\ell)} - \mathrm{E}_0[\mathbf{B}_2^{(n;\ell)}] = o_{\mathrm{P}}(1)$ as $n \to \infty$ (Lemma 6.2), we obtain that

$$\mathbf{B}_2^{(n;\ell)} - \mathrm{E}_0[\mathbf{B}_2^{(n;\ell)}] \xrightarrow{\mathcal{L}} \mathcal{N}_{k^2}\left(\mathbf{0}, \mathrm{E}[(K_{f_1}^{(\ell)}(U))^2]\, \boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\right),$$

as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$. Comparing with (6.11), it follows that

$$\mathbf{D}_2^{(n;\ell)} = \mathrm{E}_0\left[\mathbf{B}_2^{(n;\ell)}\right] + \mathcal{J}_k^{(\ell)}(f_1, g_1)\, \boldsymbol{\Upsilon}_k^{-1}(\mathbf{V})\, (\overset{\circ}{\operatorname{vech}}\, \mathbf{v}^{(n)})$$

is $o(1)$, as $n \to \infty$, as was to be proved. □

We now complete the proof of Proposition 6.1 by proving Lemma 6.4.

Proof of Lemma 6.4. (i) In view of the independence, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$, between the d_i^0's and the $\mathbf{U}_i^0$'s, we obtain, for all n,

$$\begin{aligned}
\mathrm{E}_0[\|\mathbf{R}_1^{(n;\ell)}\|^2] &\leq \frac{C}{n} \sum_{i=1}^n \mathrm{E}_0\left[\left[K_{f_1}(\tilde{G}_{1k}(d_i^0/\sigma)) - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^0/\sigma))\right]^2\right] \mathrm{E}_0\left[\left[\operatorname{vec} \mathbf{U}_i^0 \mathbf{U}_i^{0\prime}\right]' \mathbf{J}_k^{\perp} \left[\operatorname{vec} \mathbf{U}_i^0 \mathbf{U}_i^{0\prime}\right]\right] \\
&= \frac{2C(k-1)}{kn} \sum_{i=1}^n \mathrm{E}_0\left[\left[K_{f_1}(\tilde{G}_{1k}(d_i^0/\sigma)) - K_{f_1}^{(\ell)}(\tilde{G}_{1k}(d_i^0/\sigma))\right]^2\right] \\
&= \frac{2C(k-1)}{k} \int_0^1 \left[K_{f_1}(u) - K_{f_1}^{(\ell)}(u)\right]^2 du. \qquad (6.12)
\end{aligned}$$

Now, $K_{f_1}^{(\ell)}(u)$ converges to $K_{f_1}(u)$, for all $u \in (0,1)$. Also, since $|K_{f_1}^{(\ell)}(u)| \leq |K_{f_1}(u)|$, for all $\ell \geq L$, the integrand in (6.12) is bounded (uniformly in ℓ) by $4\,|K_{f_1}(u)|^2$, which is integrable on $(0,1)$. Consequently, the Lebesgue dominated convergence theorem yields that $\mathrm{E}_0[\|\mathbf{R}_1^{(n;\ell)}\|^2] = o(1)$, as $\ell \to \infty$. This convergence is of course uniform in n, since the constant C above does not depend on n.

(ii) The claim in (ii) is the same as in (i), except that d_i^n and $\mathbf{U}_i^n$ replace d_i^0 and $\mathbf{U}_i^0$, respectively. Accordingly, (ii) holds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}^n,\mathbf{V}^n;g}$. That it also holds under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V};g}$ follows from Lemma 3.5 in Jurečková (1969).

(iii) Note that

$$\begin{aligned}
|\mathcal{J}_k(f_1, g_1) - \mathcal{J}_k^{(\ell)}(f_1; g_1)|^2 &= \left|\int_0^1 \left(K_{f_1}(u) - K_{f_1}^{(\ell)}(u)\right) K_{g_1}(u)\, du\right|^2 \\
&\leq \mathcal{J}_k(g_1) \int_0^1 \left|K_{f_1}(u) - K_{f_1}^{(\ell)}(u)\right|^2 du.
\end{aligned}$$

Again, $|K_{f_1}^{(\ell)}(u) - K_{f_1}(u)|^2 \leq 4|K_{f_1}(u)|^2$, with $\int_0^1 |K_{f_1}(u)|^2\, du < \infty$. Consequently, the pointwise convergence of $(K_{f_1}^{(\ell)})$ to K implies that $\mathcal{J}_k(f_1, g_1) - \mathcal{J}_k^{(\ell)}(f_1; g_1) = o(1)$ as $\ell \to \infty$. The result then follows from the boundedness of the sequence $(\mathbf{v}^{(n)})$. □

6.2 Proof of Proposition 3.2.

Proof of Proposition 3.2. (i) The asymptotic representations (3.11) are just a restatement of (3.8) and (3.9), where we refer to for the proof; (3.13) then readily results from part (iii) of Proposition 3.1. As for (3.14), it directly follows from the fact that $\mathrm{vec}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}-\mathbf{V}) = \mathbf{M}'_k \overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}-\mathbf{V})$ and the definition of $\mathbf{Q}_k(\mathbf{V})$.

(ii) Semiparametric efficiency is a consequence of the fact that $\mathcal{J}_k(f_1,f_1)=\mathcal{J}_k(f_1)$, so that under $\mathrm{P}^{(n)}_{\mathbf{V};f}$, the asymptotic variance in (3.13) reduces to $\mathcal{J}_k(f_1)^{-1}\mathbf{\Upsilon}_k(\mathbf{V}) = (\mathbf{\Gamma}^*_{f_1}(\mathbf{V}))^{-1}$, i.e., to the efficient information matrix.

(iii) By using (3.3) and (3.10) in (3.7), we obtain

$$\begin{aligned}\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}) &= \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) + \frac{k(k+2)}{\sqrt{n}\mathcal{J}_k(f_1,g_1)}\,\mathbf{N}_k\mathbf{Q}_k(\mathbf{V}^{(n)}_{\#})\mathbf{N}'_k\ \underset{\sim}{\mathbf{\Delta}}^{(n)}_f(\mathbf{V}^{(n)}_{\#})\\ &= \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) + \frac{k(k+2)}{2n\mathcal{J}_k(f_1,g_1)}\,\mathbf{N}_k\mathbf{Q}_k(\mathbf{V}^{(n)}_{\#})\Big((\mathbf{V}^{(n)}_{\#})^{\otimes 2}\Big)^{-1/2}\\ &\qquad \times\sum_{i=1}^n\left[K_{f_1}\left(\frac{R_i}{n+1}\right)\mathrm{vec}\left(\mathbf{U}_i\mathbf{U}'_i\right) - \frac{m^{(n)}_{f_1}}{k}\mathrm{vec}\left(\mathbf{I}_k\right)\right],\end{aligned}$$

where we used the fact that (see Section 3.4 for the definition of $\mathbf{e}_{k^2,1}$)

$$\begin{aligned}\mathbf{Q}_k(\mathbf{V})\mathbf{N}'_k\mathbf{M}_k = \mathbf{Q}_k(\mathbf{V}) &= \Big[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V})\,(\mathbf{e}_{k^2,1})'\Big]\,[\mathbf{I}_{k^2}+\mathbf{K}_k]\,(\mathbf{V}^{\otimes 2})\,\Big[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V})\,(\mathbf{e}_{k^2,1})'\Big]'\\ &= [\mathbf{I}_{k^2}+\mathbf{K}_k]\,(\mathbf{V}^{\otimes 2}) - 2(\mathbf{V}^{\otimes 2})\,\mathbf{e}_{k^2,1}\,(\mathrm{vec}\mathbf{V})' - 2\,(\mathrm{vec}\mathbf{V})\,(\mathbf{e}_{k^2,1})'(\mathbf{V}^{\otimes 2}) + 2\,(\mathrm{vec}\mathbf{V})\,(\mathrm{vec}\mathbf{V})';\end{aligned}$$

see the proof of Lemma 1 in Hallin and Paindaveine (2004b). Now, routine algebra yields

$$\begin{aligned}\overset{\circ}{\mathrm{vech}}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#}) &= \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) + \frac{k(k+2)}{\mathcal{J}_k(f_1,g_1)}\,\mathbf{N}_k\Big[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V}^{(n)}_{\#})(\mathbf{e}_{k^2,1})'\Big]\Big((\mathbf{V}^{(n)}_{\#})^{\otimes 2}\Big)^{1/2}\\ &\qquad \times\left(\frac{1}{n}\sum_{i=1}^n K_{f_1}\left(\frac{R_i}{n+1}\right)\mathrm{vec}\left(\mathbf{U}_i\mathbf{U}'_i\right)\right)\\ &= \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) + \frac{k(k+2)}{\mathcal{J}_k(f_1,g_1)}\,\mathbf{N}_k\Big[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V}^{(n)}_{\#})(\mathbf{e}_{k^2,1})'\Big]\,\mathrm{vec}(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#})\\ &= \overset{\circ}{\mathrm{vech}}(\mathbf{V}^{(n)}_{\#}) + \frac{k(k+2)}{\mathcal{J}_k(f_1,g_1)}\,\mathbf{N}_k\mathrm{vec}\left(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} - (\underset{\sim}{\mathbf{W}}^{(n)}_{f\#})_{11}\mathbf{V}^{(n)}_{\#}\right),\end{aligned} \tag{6.1}$$

which establishes the result, since $\overset{\circ}{\mathrm{vech}}\mathbf{v} = \overset{\circ}{\mathrm{vech}}\mathbf{w}$ if and only if $\mathbf{v}=\mathbf{w}$, for all $k\times k$ symmetric matrices $\mathbf{v}=(v_{ij})$, $\mathbf{w}=(w_{ij})$ such that $v_{11}=w_{11}$.

(iv) Due to the identification constraints adopted, the population covariance matrix under $\mathrm{P}^{(n)}_{\mathbf{V};g} = \mathrm{P}^{(n)}_{\mathbf{V};\sigma^2,g_1}$ with finite second-order moments, is not $\mathbf{\Sigma} := \sigma^2\mathbf{V}$, but $\eta\mathbf{\Sigma} := k^{-1}\sigma^2 D_k(g_1)\mathbf{V}$ (hence $\eta = k^{-1}D_k(g_1) = k^{-1}\int_0^1\big(\tilde{G}_1^{-1}(u)\big)^2du$). Provided that $\kappa_k(g_1)<\infty$, the multivariate Central Limit Theorem yields $n^{1/2}\mathrm{vec}\left(\mathbf{\Sigma}^{(n)} - \eta\mathbf{\Sigma}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0},\mathbf{A}\right)$, where

$$\mathbf{A} := \frac{\sigma^4 E_k(g_1)}{k(k+2)}\,[\mathbf{I}_{k^2}+\mathbf{K}_k]\,(\mathbf{V}^{\otimes 2}) + \frac{\sigma^4\kappa_k(g_1)D_k^2(g_1)}{k^2}\,(\mathrm{vec}\mathbf{V})\,(\mathrm{vec}\mathbf{V})'.$$

Now, applying Slutzky's Lemma, we obtain, as $n \to \infty$, under $\mathrm{P}^{(n)}_{\mathbf{V};g}$,

$$n^{1/2}\mathrm{vec}\left(\mathbf{V}^{(n)}_{\mathcal{G}} - \mathbf{V}\right) = \frac{1}{\eta\Sigma_{11}}\left[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V})\,(\mathbf{e}_{k^2,1})'\right]\left[n^{1/2}\mathrm{vec}\left(\mathbf{\Sigma}^{(n)} - \eta\mathbf{\Sigma}\right)\right] + o_{\mathrm{P}}(1)$$

$$\xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \frac{1}{\eta^2\sigma^4}\left[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V})\,(\mathbf{e}_{k^2,1})'\right]\mathbf{A}\left[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V})\,(\mathbf{e}_{k^2,1})'\right]'\right),$$

where the covariance matrix, after lengthy but standard algebra, reduces to $(1+\kappa_k(g_1))\,\mathbf{Q}_k(\mathbf{V})$, yielding the desired result; see also Ollila, Hettmansperger, and Oja (2004).

(v) The asymptotic covariance matrices of $\mathrm{vec}(\underset{\sim}{\mathbf{V}}^{(n)}_{f\#})$ and $\mathrm{vec}(\mathbf{V}^{(n)}_{\mathcal{G}})$ (in (3.14) and (iv), respectively) are proportional. The ARE with respect to $\mathbf{V}^{(n)}_{\mathcal{G}}$ in (v) thus directly follow by computing the inverse of the corresponding ratio of proportionality factors. As for AREs with respect to $\mathbf{V}^{(n)}_{T}$, they follow from the fact that, in the adopted normalization (for which $(\mathbf{V}^{(n)}_{T})_{11} = 1$), $\sqrt{n}\,\mathrm{vec}(\mathbf{V}^{(n)}_{T} - \mathbf{V})$ is asymptotically normal with mean zero and covariance matrix $((k+2)/k)\mathbf{Q}_k(\mathbf{V})$. □

6.3 Proof of Proposition 4.1.

Proof of Proposition 4.1. We first prove that

$$\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} - \mathbf{V}^{(n)}_{\#} = O_{\mathrm{P}}(n^{-1/2}), \tag{6.2}$$

under $\mathcal{P}^{(n)}$, as $n \to \infty$ (recall that $\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} := \underset{\sim}{\mathbf{W}}^{(n)}_{f}(\mathbf{V}^{(n)}_{\#})$). To this end, define

$$\underset{\sim}{\mathbf{T}}^{(n)}_{f}(\mathbf{V}) := n^{-1/2}\left(\mathbf{V}^{\otimes 2}\right)^{1/2}\sum_{i=1}^{n}\left[K_{f_1}\left(\frac{R_i}{n+1}\right)\mathrm{vec}\left(\mathbf{U}_i\mathbf{U}_i'\right) - \frac{m^{(n)}_{f_1}}{k}\mathrm{vec}\left(\mathbf{I}_k\right)\right]$$

(with $R_i = R^{(n)}_i(\mathbf{V})$ and $\mathbf{U}_i = \mathbf{U}^{(n)}_i(\mathbf{V})$), which is asymptotically normal with mean zero and covariance matrix $\mathcal{J}_k(f_1)\mathbf{H}_k(\mathbf{V})$, where

$$\mathbf{H}_k(\mathbf{V}) := \frac{1}{k(k+2)}\left(\mathbf{V}^{\otimes 2}\right)^{1/2}\left[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k}\mathbf{J}_k\right]\left(\mathbf{V}^{\otimes 2}\right)^{1/2}.$$

Working exactly as in the proof of Proposition 6.1, we obtain that, for any bounded sequence $\mathbf{v}^{(n)}$ of symmetric matrices such that $v^{(n)}_{11} = 0$,

$$\underset{\sim}{\mathbf{T}}^{(n)}_{f}(\mathbf{V} + n^{-1/2}\mathbf{v}^{(n)}) - \underset{\sim}{\mathbf{T}}^{(n)}_{f}(\mathbf{V}) + \frac{1}{2}\mathcal{J}_k(f_1, g_1)\,\mathbf{H}_k(\mathbf{V})\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\mathrm{vec}(\mathbf{v}^{(n)}) = o_{\mathrm{P}}(1) \tag{6.3}$$

under $\mathrm{P}^{(n)}_{\mathbf{V};g}$, as $n \to \infty$. The local discreteness of $\mathbf{V}^{(n)}_{\#}$ allows to replace the nonrandom quantity $\mathbf{V}^{(n)} = \mathbf{V} + n^{-1/2}\mathbf{v}^{(n)}$ with the random one $\mathbf{V}^{(n)}_{\#}$ in (6.3) (see, e.g., Kreiss 1987, Lemma 4.4), yielding

$$\underset{\sim}{\mathbf{T}}^{(n)}_{f}(\mathbf{V}^{(n)}_{\#}) - \underset{\sim}{\mathbf{T}}^{(n)}_{f}(\mathbf{V}) + \frac{1}{2}\mathcal{J}_k(f_1, g_1)\,\mathbf{H}_k(\mathbf{V})\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\mathrm{vec}\left(\sqrt{n}(\mathbf{V}^{(n)}_{\#} - \mathbf{V})\right) = o_{\mathrm{P}}(1),$$

under $\mathrm{P}^{(n)}_{\mathbf{V};g}$, as $n \to \infty$. This establishes (6.2), since

$$
\begin{aligned}
\sqrt{n}\,\mathrm{vec}\left(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} - \mathbf{V}^{(n)}_{\#}\right) &= \underset{\sim}{\mathbf{T}}^{(n)}_{f}(\mathbf{V}^{(n)}_{\#}) + \sqrt{n}\,k^{-1}\left(m^{(n)}_{f_1} - k\right)\mathrm{vec}\left(\mathbf{V}^{(n)}_{\#}\right) \\
&= \sqrt{n}\,k^{-1}\left(m^{(n)}_{f_1} - k\right)\mathrm{vec}\left(\mathbf{V}^{(n)}_{\#}\right) + \underset{\sim}{\mathbf{T}}^{(n)}_{f}(\mathbf{V}) \\
&\quad -\frac{1}{2}\mathcal{J}_k(f_1,g_1)\,\mathbf{H}_k(\mathbf{V})\left(\mathbf{V}^{\otimes 2}\right)^{-1/2}\mathrm{vec}\left(\sqrt{n}(\mathbf{V}^{(n)}_{\#} - \mathbf{V})\right) + o_{\mathrm{P}}(1),
\end{aligned} \tag{6.4}
$$

(still under $\mathrm{P}^{(n)}_{\mathbf{V};g}$, as $n \to \infty$), and since the square-integrability of K_{f_1} on $(0,1)$ implies that

$$
m^{(n)}_{f_1} - k = m^{(n)}_{f_1} - \int_0^1 K_{f_1}(u)\,du = o(n^{-1/2})
$$

(see the proof of part (i) of Proposition 3.2 in Hallin, Vermandele, and Werker 2003).

Now, denoting by $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#} := \underset{\sim}{\mathbf{V}}^{(n)}_{f\#}(c_0)$ the pseudo-estimator defined in (3.7), it follows, from (6.2), that

$$
\mathrm{vec}(\underset{\approx}{\mathbf{V}}^{(n)}_{f\#} - \underset{\sim}{\mathbf{V}}^{(n)}_{f\#}) = \frac{-b^2(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} - \mathbf{V}^{(n)}_{\#})_{11}}{1 + b\,(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} - \mathbf{V}^{(n)}_{\#})_{11}}\left[\mathbf{I}_{k^2} - (\mathrm{vec}\mathbf{V}^{(n)}_{\#})(\mathbf{e}_{k^2,1})'\right]\mathrm{vec}(\underset{\sim}{\mathbf{W}}^{(n)}_{f\#} - \mathbf{V}^{(n)}_{\#})
$$

(with $b := k(k+2)\mathcal{J}_k^{-1}(f_1,g_1)$) is $o_{\mathrm{P}}(n^{-1/2})$ under $\mathcal{P}^{(n)}$ as $n \to \infty$. This yields the result, since we proved in Section 3.4 that $\underset{\sim}{\mathbf{V}}^{(n)}_{f\#} - \underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{f\#} = o_{\mathrm{P}}(n^{-1/2})$ under $\mathcal{P}^{(n)}$, as $n \to \infty$. □

References

[1] Adichie, J. N. (1967). Estimates of regression parameters based on rank tests, *Annals of Mathematical Statistics* **38**, 894-904.

[2] Allal, J., A. Kaaouchi, and D. Paindaveine, D. (2001). R-estimation for ARMA models, *Journal of Nonparametric Statistics* **13**, 815-831.

[3] Antille, A. (1974). A linearized version of the Hodges-Lehmann estimator, *Annals of Statistics* **2**, 1308-1313.

[4] Bickel, P.J. and Y. Ritov (1988). Estimating integrated squared density derivatives. *Sankhya* A-**50**, 381-393.

[5] Cheng, K.F. And R.J. Serfling (1981). On estimation of a class of efficiency-related parameters. *Scandinavian Actuarial Journal* **8**, 83-92.

[6] Chernoff, H., et I. R. Savage (1958). Asymptotic normality and efficiency of certain nonparametric tests, *Ann. Math. Statist.* **29**, 972-994.

[7] Draper, D. (1988). Rank-based robust analysis of linear models. I. Exposition and review, *Statistical Science* **3**, 239-257.

[8] Fan, J. (1991). On the estimation of quadratic functionals, *Annals of Statistics* **19**, 1273-1294.

[9] Hallin, M., H. Oja, and D. Paindaveine (2004). Affine-equivariant R-estimation of shape. Manuscript in preparation.

[10] Hallin, M. and D. Paindaveine (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks, *Annals of Statistics* **30**, 1103-1133.

[11] Hallin, M. and D. Paindaveine (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence, *Bernoulli* **8**, 787-816.

[12] Hallin, M. and D. Paindaveine (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model,*Annals of Statistics* **32**, 2642-2678.

[13] Hallin, M. and D. Paindaveine (2004b). Optimal rank-based tests for sphericity, submitted.

[14] Hallin, M. and D. Paindaveine (2005). Affine invariant aligned rank tests for the multivariate general linear model with ARMA errors, *Journal of Multivariate Analysis* **93**, 122-163.

[15] Hallin, M., C. Vermandele, and B.J.M. Werker (2003). Serial and nonserial sign-and-rank statistics: asymptotic representation and asymptotic normality, submitted.

[16] Hallin, M., and B. J. M. Werker (2003). Semiparametric efficiency, distribution-freeness, and invariance, *Bernoulli* **9**, 55-65.

[17] Hettmansperger, T. P. and R. Randles (2002). A practical affine equivariant multivariate median, *Biometrika* **89**, 851-860.

[18] Hodges, J.L. and E.L. Lehmann (1963). Estimates of location based on rank tests, *Annals of Mathematical Statistics* **34**, 598-611.

[19] John, S. (1971). Some optimal multivariate tests, *Biometrika* **58**, 123-127.

[20] John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions, *Biometrika* **59**, 169-174.

[21] Jurečková, J. (1969). Asymptotic linearity of a rank statistic in regression parameter, *Ann. Math. Statist.* **40**, 1889-1900.

[22] Jurečková, J. (1971). Nonparametric estimate of regression coefficients, *Annals of Mathematical Statistics* **42**, 1328-1338.

[23] Jurečková, J. and P.K. Sen (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*, J. Wiley, New York.

[24] Kreiss, J.P. (1987). On adaptive estimation in stationary ARMA processes, *Annals of Statistics* **15**, 112-133.

[25] Koul, H. (1971). Asymptotic behavior of a class of confidence regions based on ranks in regression, *Annals of Mathematical Statistics* **42**, 466-476.

[26] Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edition. Springer Verlag, New York.

[27] Koul, H.L. and A. K. Md. E. Saleh (1993). R-Estimation of the parameters of autoregressive AR(p) models, *Annals of Statistics* **21**, 534-551.

[28] Koul, H.L., Sievers, G.L., and Mc Kean, J.W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scandinavian Journal of Statistics* **14**, 131-141.

[29] Kraft, C.H. and C. van Eeden (1972). Linearized rank estimates and signed-rank estimates for the general linear hypothesis, *Annals of Mathematical Statistics* **43**, 42-57.

[30] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

[31] Le Cam, L. and G.L. Yang (2000). *Asymptotics in Statistics*, 2nd edition. Springer-Verlag, New York.

[32] Lehmann, E.L. (1963a). Robust estimation in the analysis of variance, *Annals of Mathematical Statistics* **34**, 957-966.

[33] Lehmann, E.L. (1963b). Nonparametric confidence intervals for a shift parameter. *Annals of Mathematical Statistics* **34**, 1507-1512.

[34] Mauchly, J.W. (1940). Test for sphericity of a normal n-variate distribution, *Annals of Mathematical Statistics* **11**, 204-209.

[35] Muirhead, R. J. and C. M. Waternaux (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika* **67**, 31-43.

[36] Ollila, E., C. Croux, and H. Oja (2004). Influence function and asymptotic efficiency of the affine equivariant rank covariance matrix. *Statistica Sinica* **24**, 297-316.

[37] Ollila, E., T.P. Hettmansperger, and H. Oja (2004). Affine equivariant multivariate sign methods. Preprint, University of Jyväskylä.

[38] Ollila, E., H. Oja, and C. Croux (2003). The Affine equivariant sign covariance matrix: asymptotic behaviour and efficiencies, *Journal of Multivariate Analysis* **87**, 328-355.

[39] Paindaveine, D. (2004). A Chernoff-Savage result for shape: on the non-admissibility of pseudo-Gaussian methods. Submitted.

[40] Randles, R.H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test, *Journal of the American Statististical Association* **95**, 1263-1268.

[41] Schweder, T. (1975). Window estimation of the asymptotic variance of rank estimators of location, *Scandinavian Journal of Statistics* **2**, 113-126.

[42] Sen, P.K. (1966). On a distribution-free method of estimating asymptotic efficiency of a class of nonparametric tests. *Annals of Mathematical Statistics* **37**, 1759-1770.

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

[44] Tyler, D. E. (1987a). A distribution-free M-estimator of multivariate scatter, *Annals of Statistics* **15**, 234-251.

[45] Tyler, D. E. (1987b). Statistical analysis for the angular central Gaussian distribution on the sphere, *Biometrika* **74**, 579-589.

Marc Hallin
Département de Mathématique, I.S.R.O., and E.C.A.R.E.S.
Université Libre de Bruxelles
Campus de la Plaine CP 210
B-1050 Bruxelles Belgium
mhallin@ulb.ac.be

Hannu Oja
Department of Mathematics and Statistics
University of Jyväskylä P.O. Box 35, MaD
FIN-40351 Jyväskylä Finland
Hannu.Oja@maths.jyu.fi

Davy Paindaveine
Département de Mathématique, I.S.R.O., and E.C.A.R.E.S.
Université Libre de Bruxelles
Campus de la Plaine CP 210 B-1050 Bruxelles Belgium
dpaindav@ulb.ac.be

	preliminary estimator	BIAS^{250}					
		$t_{0.5}$	$t_{0.3}$	t_{10}	$\mathcal{N}$	e_3	e_5
–	$\mathbf{V}^{(n)}_T$	-0.0043	-0.0043	0.0003	-0.0030	0.0070	-0.0023
		0.0207	0.0219	0.0062	0.0024	0.0201	0.0072
–	$\mathbf{V}^{(n)}_{\mathcal{G}}$	-0.0522	-0.0005	-0.0010	0.0005	0.0021	-0.0021
		20.6781	0.0410	0.0058	0.0024	0.0041	0.0006
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{0.5}$	$\mathbf{V}^{(n)}_T$	-0.0024	-0.0031	-0.0006	-0.0019	0.0043	-0.0026
		0.0183	0.0180	0.0031	0.0030	0.0115	0.0037
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0021	-0.0030	-0.0006	-0.0019	0.0043	-0.0026
		0.0171	0.0178	0.0032	0.0032	0.0116	0.0037
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{3}$	$\mathbf{V}^{(n)}_T$	-0.0014	-0.0017	-0.0009	-0.0004	0.0022	-0.0022
		0.0216	0.0142	0.0024	0.0028	0.0047	0.0006
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0051	-0.0017	-0.0009	-0.0004	0.0023	-0.0021
		0.0219	0.0140	0.0023	0.0030	0.0043	0.0006
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{10}$	$\mathbf{V}^{(n)}_T$	-0.0008	-0.0015	-0.0008	0.0001	0.0014	-0.0021
		0.0250	-0.0261	0.0029	0.0026	0.0032	0.0000
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0075	-0.0014	-0.0008	0.0001	0.0016	-0.0020
		0.0254	0.0128	0.0028	0.0028	0.0032	-0.0000
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{\mathrm{vdW}}$	$\mathbf{V}^{(n)}_T$	-0.0003	-0.0014	-0.0007	0.0003	0.0011	-0.0020
		0.0281	0.0124	0.0036	0.0025	0.0026	-0.0000
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0091	-0.0013	-0.0007	0.0004	0.0013	-0.0019
		0.0284	0.0122	0.0036	0.0026	0.0026	-0.0001

		MSE^{250}					
		$t_{0.5}$	$t_{0.3}$	t_{10}	$\mathcal{N}$	e_3	e_5
–	$\mathbf{V}^{(n)}_T$	0.0083	0.0081	0.0075	0.0075	0.0080	0.0085
		0.0392	0.0357	0.0337	0.0369	0.0337	0.0320
–	$\mathbf{V}^{(n)}_{\mathcal{G}}$	11.3416	0.0329	0.0050	0.0038	0.0028	0.0029
		42,948	0.2358	0.0211	0.0175	0.0115	0.0109
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{0.5}$	$\mathbf{V}^{(n)}_T$	0.0075	0.0065	0.0058	0.0058	0.0057	0.0061
		0.0339	0.0285	0.0258	0.0282	0.0233	0.0223
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0278	0.0065	0.0057	0.0058	0.0057	0.0061
		0.0566	0.0284	0.0258	0.0281	0.0233	0.0223
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{3}$	$\mathbf{V}^{(n)}_T$	0.0090	0.0057	0.0044	0.0042	0.0031	0.0030
		0.0371	0.0247	0.0199	0.0198	0.0127	0.0112
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0295	0.0058	0.0044	0.0042	0.0031	0.0030
		0.0598	0.0247	0.0199	0.0197	0.0128	0.0112
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{10}$	$\mathbf{V}^{(n)}_T$	0.0106	0.0060	0.0043	0.0039	0.0025	0.0022
		0.0428	1.5539	0.0191	0.0180	0.0102	0.0084
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0339	0.0060	0.0043	0.0039	0.0025	0.0022
		0.0662	0.0253	0.0191	0.0180	0.0101	0.0083
$\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{\mathrm{vdW}}$	$\mathbf{V}^{(n)}_T$	0.0121	0.0064	0.0044	0.0039	0.0022	0.0019
		0.0486	0.0267	0.0192	0.0176	0.0092	0.0073
	$\mathbf{V}^{(n)}_{\mathcal{G}}$	0.0377	0.0064	0.0044	0.0039	0.0022	0.0018
		0.0726	0.0266	0.0192	0.0175	0.0092	0.0072

Table 2: Empirical bias and mean-square error, under various bivariate t-, power-exponential, and normal densities, of the preliminary estimators $\mathbf{V}^{(n)}_{\mathcal{G}}$ and $\mathbf{V}^{(n)}_T$, and the corresponding one-step R-estimators $\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{0.5}$, $\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{3}$, $\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{10}$, and $\underset{\sim}{\widehat{\mathbf{V}}}^{(n)}_{\mathrm{vdW}}$. The simulation is based on 1000 replications; sample size is $n = 250$.

	preliminary estimator	BIAS^{50}					
		$t_{0.5}$	$t_{0.3}$	t_{10}	$\mathcal{N}$	e_3	e_5
–	$\mathbf{V}_T^{(n)}$	0.0042	-0.0038	-0.0016	0.0006	0.0067	-0.0070
		0.0830	0.0973	0.0865	0.0895	0.1118	0.0906
–	$\mathbf{V}_{\mathcal{G}}^{(n)}$	-0.6148	0.0012	-0.0003	-0.0058	0.0025	-0.0024
		310.8334	0.1782	0.0497	0.0375	0.0484	0.0308
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{0.5}$	$\mathbf{V}_T^{(n)}$	0.0034	-0.0004	0.0004	-0.0006	0.0039	-0.0030
		0.0771	0.0806	0.0619	0.0674	0.0821	0.0664
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.0001	0.0004	-0.0005	-0.0007	0.0033	-0.0036
		0.0798	0.0782	0.0612	0.0671	0.0820	0.0661
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{3}$	$\mathbf{V}_T^{(n)}$	0.0002	0.0019	0.0005	-0.0024	0.0023	-0.0017
		0.0861	0.0680	0.0438	0.0444	0.0533	0.0338
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.0014	0.0028	0.0002	-0.0021	0.0023	-0.0019
		0.1717	0.0665	0.0433	0.0442	0.0531	0.0336
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{10}$	$\mathbf{V}_T^{(n)}$	-0.0001	0.0025	0.0004	-0.0036	0.0023	-0.0019
		0.0962	0.0681	0.0427	0.0395	0.0441	0.0253
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.0037	0.0034	0.0001	-0.0031	0.0023	-0.0019
		0.1074	0.0672	0.0419	0.0398	0.0440	0.0250
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{\mathrm{vdW}}$	$\mathbf{V}_T^{(n)}$	0.0005	0.0027	0.0005	-0.0044	0.0024	-0.0024
		0.1057	0.0702	0.0441	0.0387	0.0404	0.0217
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.0034	0.0035	-0.0001	-0.0041	0.0024	-0.0022
		0.1164	0.0696	0.0435	0.0392	0.0402	0.0211

		MSE^{50}					
		$t_{0.5}$	$t_{0.3}$	t_{10}	$\mathcal{N}$	e_3	e_5
–	$\mathbf{V}_T^{(n)}$	0.0410	0.0407	0.0408	0.0404	0.0444	0.0423
		0.2009	0.2467	0.2192	0.2311	0.2163	0.2031
–	$\mathbf{V}_{\mathcal{G}}^{(n)}$	298.8463	0.1033	0.0265	0.0183	0.0155	0.0138
		80,313,350	0.7141	0.1247	0.0941	0.0624	0.0617
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{0.5}$	$\mathbf{V}_T^{(n)}$	0.0368	0.0328	0.0312	0.0307	0.0320	0.0296
		0.1862	0.1879	0.1629	0.1701	0.1425	0.1411
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.1152	0.0337	0.0308	0.0309	0.0318	0.0294
		0.2700	0.1852	0.1614	0.1686	0.1416	0.1398
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{3}$	$\mathbf{V}_T^{(n)}$	0.0419	0.0290	0.0238	0.0208	0.0178	0.0149
		0.2239	0.1546	0.1169	0.1138	0.0715	0.0676
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.1184	0.0296	0.0235	0.0209	0.0175	0.0146
		5.6092	0.1537	0.1162	0.1132	0.0709	0.0668
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{10}$	$\mathbf{V}_T^{(n)}$	0.0490	0.0300	0.0234	0.0191	0.0147	0.0118
		0.2701	0.1579	0.1117	0.1005	0.0568	0.0519
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.1307	0.0306	0.0232	0.0190	0.0143	0.0114
		0.3796	0.1583	0.1108	0.1006	0.0562	0.0511
$\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{\mathrm{vdW}}$	$\mathbf{V}_T^{(n)}$	0.0552	0.0316	0.0238	0.0187	0.0135	0.0106
		0.3134	0.1652	0.1129	0.0964	0.0518	0.0457
	$\mathbf{V}_{\mathcal{G}}^{(n)}$	0.1406	0.0322	0.0238	0.0185	0.0131	0.0102
		0.4237	0.1665	0.1121	0.0967	0.0511	0.0449

Table 3: Empirical bias and mean-square error, under various bivariate t-, power-exponential, and normal densities, of the preliminary estimators $\mathbf{V}_{\mathcal{G}}^{(n)}$ and $\mathbf{V}_T^{(n)}$, and the corresponding one-step R-estimators $\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{0.5}$, $\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{3}$, $\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{10}$, and $\underset{\sim}{\widehat{\mathbf{V}}}{}^{(n)}_{\mathrm{vdW}}$. The simulation is based on 1000 replications; sample size is $n = 50$.