Efficient R-estimation of Principal and Common Principal Components

Marc Hallin, Davy Paindaveine, and Thomas Verdebout*

Abstract

We propose rank-based estimators of principal components, both in the one-sample and, under the assumption of *common principal components*, in the m-sample cases. Those estimators are obtained via a rank-based version of Le Cam's one-step method, combined with an estimation of *cross-information quantities*. Under arbitrary elliptical distributions with, in the m-sample case, possibly heterogeneous radial densities, those R-estimators remain root-n consistent and asymptotically normal, while achieving asymptotic efficiency under correctly specified radial densities. Contrary to their traditional counterparts computed from empirical covariances, they do not require any moment conditions. When based on Gaussian score functions, in the one-sample case, they moreover uniformly dominate their classical competitors in the Pitman sense. Their AREs with respect to other robust procedures are quite high—up to 30, in the Gaussian case, with respect to minimum covariance determinant estimators. Their finite-sample performances are investigated via a Monte-Carlo study.

*Marc Hallin is Professor of Statistics, Université libre de Bruxelles, ECARES, Avenue F. D. Roosevelt, 50, CP 114/04, B-1050 Bruxelles, Belgium and ORFE, Princeton University, Sherrerd Hall, Princeton, NJ 08544, USA (E-mail: mhallin@ulb.ac.be). Davy Paindaveine is Professor of Statistics, Université libre de Bruxelles, ECARES and Département de Mathématique, Avenue F. D. Roosevelt, 50, CP 114/04, B-1050 Bruxelles, Belgium (E-mail: dpaindav@ulb.ac.be). Thomas Verdebout is Professor of Statistics, EQUIPPE and INRIA, Université Lille 3, Domaine universitaire du "Pont de Bois", Rue du Barreau, BP 60149, 59653 Villeneuve-d'Ascq CEDEX, France (E-mail: thomas.verdebout@univ-lille3.fr).

Keywords and phrases: Common Principal Components, elliptical densities, Uniform local asymptotic normality, principal components, ranks, R-estimation, robustness.

1 Introduction

Principal component analysis (PCA) arguably constitutes one of the most useful and popular techniques of multivariate analysis. Introduced by Pearson (1901) and rediscovered by Hotelling (1933), PCA is a powerful dimension reduction tool, by which the k (k typically large) marginals of a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ get replaced with (typically, a few) appropriately chosen mutually orthogonal random variables, called the *principal components* (PCs), in such a way that most of the variability in $\mathbf{X}$ still is accounted for. Assuming that the original random vector $\mathbf{X}$ has finite second-order moments, traditional PCs are obtained by projecting $\mathbf{X}$ onto the eigenvectors of its covariance matrix; the variances of those projections then are the corresponding eigenvalues.

The multisample version of principal components came much later, when Flury (1984) introduced the Common Principal Components (CPC) model as a parcimonious way of parametrizing an m-tuple of covariance matrices. CPC models since then have been used in a variety of applications (see Flury and Riedwyl 1988). Under CPC, $m \geq 2$ populations of dimension k, with covariance matrices $\mathbf{\Sigma}_i^{\text{Cov}}$, $i = 1, \ldots, m$, share, with possibly different eigenvalues, the same eigenvectors: namely, the m covariance matrices $\mathbf{\Sigma}_i^{\text{Cov}}$ factorize into $\mathbf{\Sigma}_i^{\text{Cov}} = \boldsymbol{\beta}\mathbf{\Lambda}_i^{\text{Cov}}\boldsymbol{\beta}'$ for some m-tuple of positive diagonal matrices $\mathbf{\Lambda}_i^{\text{Cov}}$, $i = 1, \ldots, m$, and some orthogonal matrix $\boldsymbol{\beta}$—the matrix of *common eigenvectors*, which does not depend on i and characterizes the *common principal components*.

In his 1984 paper, Flury also deals, under the hypothesis of CPC, with the Gaussian maximum likelihood estimators (MLEs) $(\hat{\boldsymbol{\beta}}_1^{\text{MLE}}, \ldots, \hat{\boldsymbol{\beta}}_k^{\text{MLE}}) =: \hat{\boldsymbol{\beta}}^{\text{MLE}}$ and $\hat{\lambda}_{ij}^{\text{MLE}}$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ of the common eigenvectors $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k) =: \boldsymbol{\beta}$ and the corresponding eigenvalues λ_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, k$ of $\boldsymbol{\Sigma}_1^{\text{Cov}}, \ldots, \boldsymbol{\Sigma}_m^{\text{Cov}}$. Denoting by $\bar{\mathbf{X}}_i$ and the empirical mean and covariance matrix (unbiased versions) in sample i, $i = 1, \ldots, m$, he shows that those MLEs are solutions of the likelihood equations

$$\boldsymbol{\beta}_j' \Big(\sum_{i=1}^{m} n_i \frac{\lambda_{ij} - \lambda_{il}}{\lambda_{ij}\lambda_{il}} \mathbf{S}_i \Big) \boldsymbol{\beta}_l = 0, \quad j \neq l = 1, \ldots, k,$$
$$\boldsymbol{\beta}_j' \mathbf{S}_i \boldsymbol{\beta}_j = \lambda_{ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, k, \qquad \boldsymbol{\beta}_j' \boldsymbol{\beta}_l = \delta_{jl}, \quad j, l = 1, \ldots, k, \tag{1.1}$$

where δ_{jl} stands for the usual Kronecker symbol. An explicit solution of equations (1.1) does not exist, but an algorithm providing a numerical solution has been proposed by Flury and Gautschi (1986).

Traditional PCA and CPC methods are based on Gaussian assumptions, and their implementation is based on empirical covariance matrices (as in (1.1) above). This Gaussian approach puts regrettable limitations on the applicability of the method. Principal components, indeed, intuitively only depend on the elliptical geometry of the underlying distributions, irrespective of any moment conditions. And covariance-based methods are known to be poorly robust. More resistant PCA and CPC methods, remaining valid under arbitrary elliptical densities, are thus highly desirable. This is the motivation behind the *projection-poursuit* techniques developed by Croux and Ruiz-Gazen (2005), which are based on robust scale functionals. Under elliptical symmetry with scatter matrix $\boldsymbol{\Sigma}$ (reducing to a covariance matrix only under finite moments of order two), all "reasonable" (we refer to Croux and Ruiz-Gazen 2005 for a precise statement) equivariant scale functionals lead to the same concept of principal components, namely the one associated with the eigenvectors of $\mathbf{\Sigma}$. The PC estimators obtained by Croux and Ruiz-Gazen have high finite-sample breakdown points. Delvin *et al.* (1981), Croux and Haesbroeck (2000) and Taskinen *et al.* (2012) also consider PCA techniques based on robust estimators of the covariance matrix. In the CPC context, Boente *et al.* (2001, 2002) similarly proposed to replace the empirical covariances $\mathbf{S}_i$ in (1.1) with more robust estimators. Projection pursuit techniques for CPC also have been considered by Boente *et al.* (2006).

Robust methods, as a rule, suffer from a loss of efficiency, and those robust PCA and CPC methods are no exceptions to that rule. To improve on this, Hallin *et al.* (2010b and 2013) recently provided locally asymptotically optimal (in the Le Cam sense) rank tests for PCA and CPC, respectively. A major advantage of these tests is that they are not only *validity-robust*, in the sense of surviving arbitrary (possibly very heavy-tailed) elliptical densities: unlike their pseudo-Gaussian and robust competitors, they also are *efficiency-robust*, in the sense that their local powers do not deteriorate away from the reference density at which they are optimal. Their normal-score versions, moreover, uniformly dominate, in the Pitman sense, the (pseudo-)Gaussian methods, based on sample covariance matrices.

Daily practice in PCA and CPC, however, is about estimation rather than hypothesis testing, which raises a natural question: do the rank tests in Hallin *et al.* (2010b and 2013) have any estimation counterparts? That is, can we construct rank-based estimators for the (common) eigenvectors that match the performances of those rank-based tests?

In this paper, we provide a positive answer to that question by constructing rank-based estimators (R-estimators) that (i) are root-n consistent and asymptotically normal under any elliptical density (for CPC, any m-tuple of elliptical densities), irrespective of any moment assumptions; (ii) are efficient at some prespecified elliptical density (for CPC, some prespecified m-tuple of them); (iii) exhibit the same asymptotic relative efficiencies as the rank tests from Hallin *et al.* (2010b and 2013) with respect to classical Gaussian procedures; as a corollary, the Gaussian-score rank-based estimators will uniformly dominate, in the one-sample case and in terms of Pitman efficiencies, the classical estimators based on sample covariance matrices.

Traditional R-estimators in principle are obtained via the minimization of some rank-based objective function. From a practical point of view, this is known to be numerically costly, or even infeasible, especially in the multiparameter case, hence in the present context of (common) principal components: rank-based objective functions indeed are piecewise constant, hence discontinuous and non-convex. Instead, we use a rank-based version of Le Cam's one-step methodology. Letting $\hat{\boldsymbol{\beta}}$ stand for a preliminary root-n consistent estimator, our estimators are of the form $\mathrm{vec}(\underset{\sim}{\boldsymbol{\beta}}) = \mathrm{vec}(\hat{\boldsymbol{\beta}}) + n^{-1/2}\underset{\sim}{\boldsymbol{\Gamma}}^{-}\underset{\sim}{\boldsymbol{\Delta}}$, where $\underset{\sim}{\boldsymbol{\Delta}}$ is a rank-based central sequence and $\underset{\sim}{\boldsymbol{\Gamma}}^{-}$ the Moore-Penrose inverse of some estimated cross-information matrix.

The outline of the paper is as follows. In Section 2, we introduce the notation needed in the sequel. In Section 3.1, we describe the proposed estimators for the common eigenvectors under CPC. We then study their asymptotic properties in Section 3.2. In Section 4, we consider estimation of eigenvectors in the one-sample case, that is, for PCA. A Monte-

Carlo simulation is performed in Section 5 to investigate the finite-sample behavior of our estimators. Finally, an appendix collects the technical proofs.

2 Notation and main assumptions

2.1 Elliptical densities

Throughout the paper, $(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})$, $i = 1, \ldots, m$ form a collection of m mutually independent samples of i.i.d. k-dimensional random vectors with elliptically symmetric densities. More precisely, we assume that $\mathbf{X}_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$ are mutually independent, with elliptical probability densities of the form

$$\underline{f_i}(\mathbf{x}) = c_{k,f_i} \, (\det(\mathbf{\Sigma}_i))^{-1/2} \, f_i\Big(((\mathbf{x} - \boldsymbol{\theta}_i)' \mathbf{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\theta}_i))^{1/2} \Big) \tag{2.1}$$

for some k-dimensional *location* parameter $\boldsymbol{\theta}_i$, some symmetric positive definite *scatter* matrix $\mathbf{\Sigma}_i$ and some *radial density* function $f_i : \mathbb{R}_0^+ \mapsto \mathbb{R}^+$; c_{k,f_i} is a normalization constant. Note that the radial density f_i is not a probability density since it does not integrate to one; but the function $\tilde{f}_i := r \mapsto \mu_{k-1;f_i}^{-1} r^{k-1} f_i(r)$ (for simplicity, we write $\tilde{f}_i$ instead of $\tilde{f}_{ik}$), where $\mu_{\ell;f} := \int_0^\infty r^\ell f(r) \, \mathrm{d}r$, is. Define

$$\mathcal{F} := \big\{ f : f(r) > 0 \text{ a.e. and } \mu_{k-1;f} < \infty \big\} \text{ and } \mathcal{F}_1 := \Big\{ f \in \mathcal{F} : \mu_{k-1;f}^{-1} \int_0^1 r^{k-1} f(r) \, \mathrm{d}r = 1/2 \Big\};$$

the family $\mathcal{F}_1$ is a class of nowhere vanishing *standardized* radial densities, in the sense that, for any radial density $f \in \mathcal{F}_1$, the probability density $\tilde{f} := r \mapsto \mu_{k-1;f}^{-1} r^{k-1} f(r)$ is a properly standardized probability density. By "standardized", here, we mean that the corresponding median is one; the median, for a nonvanishing density over $\mathbb{R}_0^+$, indeed, is a scale parameter—the existence of which does not require any moment conditions. Classical examples of elliptical distributions are the k-variate multinormal distributions ($\mathcal{N}$), with standardized radial densities $f_i(r) = \phi(r) := \exp(-a_k r^2/2)$, the k-variate Student distributions (t_ν), with standardized radial densities $f_i(r) = f^t_\nu(r) := (1 + a_{k,\nu} r^2/\nu)^{-(k+\nu)/2}$, $\nu > 0$, and the k-variate power-exponential distributions ($\mathcal{E}_\eta$) with standardized radial densities of the form $f_i(r) = f^e_\eta(r) := \exp(-b_{k,\eta} r^{2\eta})$, $\eta > 0$; the positive constants a_k, $a_{k,\nu}$, and $b_{k,\eta}$ are such that $f_i \in \mathcal{F}_1$. Summarizing this, we throughout assume that the following assumption holds.

Assumption (A1). The observations $\mathbf{X}_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$ are mutually independent, with probability densities $\underline{f_i}$ given in (2.1), for some m-tuple of (possibly distinct) radial densities $\mathrm{f} := (f_1, \ldots, f_m)$ such that $f_i \in \mathcal{F}_1$, $i = 1, \ldots, m$.

Under Assumption (A1), the distances $d_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i) := \|\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)\|$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$ have probability density $\tilde{f}_i$, with median one, which identifies the *scatter* matrices $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, m$ also in the absence of any moments (throughout, $\mathbf{A}^{1/2}$ stands for the symmetric and positive definite root of the symmetric and positive definite matrix $\mathbf{A}$). Under finite second-order moments, however, $\boldsymbol{\Sigma}_i$ is proportional to the covariance matrix $\boldsymbol{\Sigma}_i^{\mathrm{Cov}}$ of $\mathbf{X}_{ij}$. Note that the observations $\mathbf{X}_{ij}$ then decompose into $\mathbf{X}_{ij} = \boldsymbol{\theta}_i + d_{ij}\boldsymbol{\Sigma}_i^{1/2}\mathbf{U}_{ij}$, where, under Assumption (A1), the *multivariate signs* $\mathbf{U}_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i) := \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)/d_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$ are i.i.d. uniform over the unit sphere of $\mathbb{R}^k$ and the *standardized radial distances* $d_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$ just defined are independent of the $\mathbf{U}_{ij}$'s, with standardized probability density $\tilde{f}_i$ over $\mathbb{R}^+$ and distribution function $\tilde{F}_i$.

The derivation of asymptotically efficient estimators at a given m-tuple f $= (f_1, \ldots, f_m)$ of radial densities will be based on the *uniform local and asymptotic normality* (ULAN) of the CPC model; a precise statement, with explicit forms of the *central sequence* $\mathbf{\Delta}^{(n)}_{\boldsymbol{\vartheta};\mathrm{f}}$ and the *information matrix* $\mathbf{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}$, is provided in Proposition A.1. That ULAN property holds under mild regularity conditions on the f_i's. More precisely, it requires the f_i's to belong to the collection $\mathcal{F}_a$ of those radial densities $f \in \mathcal{F}_1$ that are absolutely continuous, with almost everywhere derivative $\dot{f}$ such that, letting $\varphi_f := -\dot{f}/f$ and denoting by $\tilde{F}$ the distribution function associated with $\tilde{f}$, the integrals

$$\mathcal{I}_k(f) := \int_0^1 \varphi_f^2(\tilde{F}^{-1}(u))\, \mathrm{d}u \quad \text{and} \quad \mathcal{J}_k(f) := \int_0^1 \varphi_f^2(\tilde{F}^{-1}(u))(\tilde{F}^{-1}(u))^2\, \mathrm{d}u$$

are finite. The quantities $\mathcal{I}_k(f_i)$ and $\mathcal{J}_k(f_i)$ play the roles of *radial Fisher information* for location and shape/scale, respectively, in population i, $i = 1, \ldots, m$ (see Hallin and Paindaveine 2006).

2.2 Parametrization

Since the common eigenvectors $\boldsymbol{\beta} := (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)$ of $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_m$ are scale-free functions of the $\mathbf{\Sigma}_i$'s, it is appropriate to decompose each $\mathbf{\Sigma}_i$ into a product $\mathbf{\Sigma}_i = \sigma_i^2 \mathbf{V}_i$, where $\sigma_i > 0$ is a *scale* parameter and $\mathbf{V}_i$ is a *shape* matrix for population i (see Hallin and Paindaveine (2006) for details). Paindaveine (2008) shows the advantage of doing so by defining σ_i^2 as $(\det \mathbf{\Sigma}_i)^{1/k}$. This definition, which is the one we are adopting here, implies that the eigenvalues $\lambda^{\mathbf{V}}_{ij}$ of the shape matrices $\mathbf{V}_i$ are such that $\prod_{j=1}^k \lambda^{\mathbf{V}}_{ij} = 1$ for all $i = 1, \ldots, m$; clearly, $\mathbf{V}_i$ and $\mathbf{\Sigma}_i$ share the same eigenvectors. Obviously, the shape matrices in turn factorize into $\mathbf{V}_i = \boldsymbol{\beta}\mathbf{\Lambda}^{\mathbf{V}}_i\boldsymbol{\beta}'$. In the CPC case, the following assumption moreover ensures the identifiability of the common eigenvectors $\boldsymbol{\beta}$:

Assumption (A2). For any $1 \leq j \neq j' \leq k$, there exists $i \in \{1, \ldots, m\}$ such that $\lambda_{ij}^{\mathbf{V}} \neq \lambda_{ij'}^{\mathbf{V}}$.

Under the hypothesis of CPC and Assumption (A2), the matrix $\boldsymbol{\beta}$ of common eigenvectors is identified up to an arbitrary permutation of its columns (we forget about the irrelevant sign changes of the $\boldsymbol{\beta}_j$'s). However, it is easy to fix an ordering, hence to make the $\boldsymbol{\beta}_j$'s—hence also the corresponding $\lambda_{ij}^{\mathbf{V}}$'s—(individually) identifiable.

We then adopt the following parametrization. Denoting by $\mathrm{dvec}\,(\mathbf{A})$ the vector obtained by stacking the diagonal elements of a square matrix $\mathbf{A}$, and by $\mathrm{d\mathring{v}ec}\,\mathbf{A}$ the same vector deprived of its first element A_{11} so that $\mathrm{dvec}\,(\mathbf{A}) = (A_{11}, (\mathrm{d\mathring{v}ec}\,\mathbf{A})')'$, our parameter is the vector

$$\boldsymbol{\vartheta} \;:=\; (\boldsymbol{\vartheta}'_{I}, \boldsymbol{\vartheta}'_{II}, \boldsymbol{\vartheta}'_{III}, \boldsymbol{\vartheta}'_{IV})' := (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_m, \sigma_1^2, \ldots, \sigma_m^2, (\mathrm{d\mathring{v}ec}\,\boldsymbol{\Lambda}_1^{\mathbf{V}})', \ldots, (\mathrm{d\mathring{v}ec}\,\boldsymbol{\Lambda}_m^{\mathbf{V}})', (\mathrm{vec}\,\boldsymbol{\beta})')',$$

where $\boldsymbol{\theta}_i$ and σ_i^2 are the location and scale parameters, $\boldsymbol{\Lambda}_i^{\mathbf{V}} := \mathrm{diag}(\lambda_{i1}^{\mathbf{V}}, \ldots, \lambda_{ik}^{\mathbf{V}})$, the diagonal matrix of eigenvalues in population i, $i = 1, \ldots, m$, and $\boldsymbol{\beta}$ the matrix of common eigenvectors. The reason why the $\lambda_{i1}^{\mathbf{V}}$'s are omitted in the parametrization is that, $\mathbf{V}_i$ being a shape matrix, we have $\lambda_{i1}^{\mathbf{V}} = 1/\prod_{j=2}^{k} \lambda_{ij}^{\mathbf{V}}$. The parameter space is thus $\boldsymbol{\Theta} := \mathbb{R}^{mk} \times (\mathbb{R}_0^+)^m \times (\mathcal{C}^{k-1})^m \times (\mathrm{vec}\,\mathcal{S}O_k)$, where $\mathcal{C}^{k-1}$ is the open positive orthant of $\mathbb{R}^{k-1}$ and $\mathcal{S}O_k$ stands for the class of $k \times k$ real orthogonal matrices with determinant one. Write $\mathrm{P}_{\boldsymbol{\vartheta};\mathrm{f}}^{(n)}$ for the joint distribution of the n observations under parameter value $\boldsymbol{\vartheta}$ and standardized radial densities $\mathrm{f} = (f_1, \ldots, f_m)$; note that Assumption (A2) is explicitly incorporated in the definition of $\boldsymbol{\Theta}$.

2.3 Asymptotic behavior of sample sizes and score functions

Asymptotics in this paper are considered for triangular arrays of observations of the form

$$(\mathbf{X}_{11}^{(n)}, \ldots, \mathbf{X}_{1n_1^{(n)}}^{(n)}, \mathbf{X}_{21}^{(n)}, \ldots, \mathbf{X}_{2n_2^{(n)}}^{(n)}, \ldots, \mathbf{X}_{m1}^{(n)}, \ldots, \mathbf{X}_{mn_m^{(n)}}^{(n)}),$$

indexed by the total sample size $n := \sum_{i=1}^m n_i^{(n)}$, where the sequences $n_i^{(n)}$ of sizes in each sample satisfy the following assumption.

ASSUMPTION (A3). For all $i = 1, \ldots, m$, $r_i^{(n)} := n_i^{(n)}/n \to r_i \in (0,1)$ as $n \to \infty$.

Letting $\mathbf{r}^{(n)} := \mathrm{diag}((r_1^{(n)})^{-1/2}, \ldots, (r_m^{(n)})^{-1/2})$, define

$$\boldsymbol{\varsigma}^{(n)} := \mathrm{diag}\big(\boldsymbol{\varsigma}_{I}^{(n)}, \boldsymbol{\varsigma}_{II}^{(n)}, \boldsymbol{\varsigma}_{III}^{(n)}, \boldsymbol{\varsigma}_{IV}^{(n)}\big) := \mathrm{diag}\big(\mathbf{r}^{(n)} \otimes \mathbf{I}_k, \mathbf{r}^{(n)}, \mathbf{r}^{(n)} \otimes \mathbf{I}_{k-1}, \mathbf{I}_{k^2}\big). \qquad (2.2)$$

The *consistency (contiguity) rates* for $\boldsymbol{\vartheta}$ throughout then will be of the form $n^{1/2}(\boldsymbol{\varsigma}^{(n)})^{-1}$.

Finally, the R-estimators considered in Section 3.1 are based on m-tuples $\mathrm{K} = (K_1, \ldots, K_m)$ of *score functions*, that are assumed to satisfy the following regularity conditions.

ASSUMPTION (A4). For any $i = 1, \ldots, m$, the mapping (from $(0,1)$ to $\mathbb{R}$) $u \mapsto K_i(u)$ (i) is continuous and square-integrable, (ii) can be expressed as the difference of two monotone increasing functions, and (iii) satisfies $\int_0^1 K_i(u)\,\mathrm{d}u = k$.

Assumption (A4)(iii) is a normalization constraint that is automatically satisfied by the score functions $K_i(u) = K_{f_i}(u) := \varphi_{f_i}(\tilde{F}_i^{-1}(u))\tilde{F}_i^{-1}(u)$ leading to asymptotic efficiency at m-tuples of radial densities $\mathrm{f} = (f_1, \ldots, f_m)$ for which ULAN holds; see Section 3.2.

For score functions K, K_1, K_2 satisfying Assumption (A4), let (throughout, U stands for a random variable uniformly distributed over $(0,1)$), $\mathcal{J}_k(K_1, K_2) := \mathrm{E}[K_1(U)K_2(U)]$. For simplicity, we write $\mathcal{J}_k(K)$ for $\mathcal{J}_k(K,K)$, $\mathcal{J}_k(K,f)$ for $\mathrm{E}[K(U)K_f(U)]$, etc.

Among the possible score functions (Laplace, Wilcoxon, etc) satisfying Assumption (A4), an important particular case of score functions of the form K_{f_i} is that of van der Waerden or normal scores, obtained for $f_i = \phi$. Denoting by Ψ_k the chi-square distribution function with k degrees of freedom, we have $K_\phi(u) = \Psi_k^{-1}(u)$, and $\mathcal{J}_k(\phi) = k(k+2)$. Similarly, Student densities $f_i = f_\nu^t$ yield

$$K_{f_\nu^t}(u) = k(k+\nu)G_{k,\nu}^{-1}(u)/(\nu + kG_{k,\nu}^{-1}(u)) \quad \text{and} \quad \mathcal{J}_k(f_\nu^t) = k(k+2)(k+\nu)/(k+\nu+2),$$

where $G_{k,\nu}$ stands for the Fisher-Snedecor distribution function with k and ν degrees of freedom.

3 R-estimation of common principal components (CPC)

3.1 One-step R-estimators

As explained in the introduction, our R-estimators $\underset{\sim}{\boldsymbol{\beta}}$ are (after vectorization) of the one-step form

$$\text{vec}(\underset{\sim}{\boldsymbol{\beta}}) = \text{vec}(\hat{\boldsymbol{\beta}}) + n^{-1/2}\underset{\sim}{\boldsymbol{\Gamma}}^{-}\underset{\sim}{\boldsymbol{\Delta}},$$

where $\hat{\boldsymbol{\beta}}$ is part of a preliminary estimator

$$\hat{\boldsymbol{\vartheta}} = \Big(\hat{\boldsymbol{\theta}}_1', \ldots, \hat{\boldsymbol{\theta}}_m', \hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2, (\text{d}\overset{\circ}{\text{vec}}\,\hat{\boldsymbol{\Lambda}}_1^{\mathbf{V}})', \ldots, (\text{d}\overset{\circ}{\text{vec}}\,\hat{\boldsymbol{\Lambda}}_m^{\mathbf{V}})', (\text{vec}\,\hat{\boldsymbol{\beta}})'\Big)', \tag{3.1}$$

$\underset{\sim}{\boldsymbol{\Delta}}$ is some rank-based form of central sequence, and $\underset{\sim}{\boldsymbol{\Gamma}}^{-}$ is the Moore-Penrose inverse of some estimated cross-information matrix, both involving the preliminary $\hat{\boldsymbol{\vartheta}}$. Here, we describe their construction, deferring technical details and justifications to the Appendix.

Consider the *multivariate signs* $(\mathbf{U}_{11}, \ldots, \mathbf{U}_{mn_m})$ and the *ranks* $(R_{11}, \ldots, R_{mn_m})$, where, letting $\hat{\mathbf{V}}_i := \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Lambda}}_i^{\mathbf{V}}\hat{\boldsymbol{\beta}}'$, $\mathbf{U}_{ij} := \mathbf{U}_{ij}(\hat{\boldsymbol{\theta}}_i, \hat{\mathbf{V}}_i)$, while $R_{ij} := R_{ij}(\hat{\boldsymbol{\theta}}_i, \hat{\mathbf{V}}_i)$ denotes the rank of $d_{ij} := d_{ij}(\hat{\boldsymbol{\theta}}_i, \hat{\mathbf{V}}_i)$ among $d_{i1}, \ldots, d_{in_i}$. Based on those signs and ranks and the m-tuple of score functions $\mathrm{K} := (K_1, \ldots, K_m)$, we introduce the rank-based statistics

$$\underset{\sim}{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\vartheta}};\mathrm{K}} := \frac{1}{2n^{1/2}} \sum_{i=1}^{m} \mathbf{G}_k^{\hat{\boldsymbol{\beta}}} \mathbf{L}_k^{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Lambda}}_i^{\mathbf{V}}} \left(\hat{\mathbf{V}}_i^{\otimes 2} \right)^{-1/2} \sum_{j=1}^{n_i} K_i \left(\frac{R_{ij}}{n_i + 1} \right) \mathrm{vec}\, (\mathbf{U}_{ij} \mathbf{U}'_{ij}), \tag{3.2}$$

where $\mathbf{A}^{\otimes 2}$ stands for the Kronecker product $\mathbf{A} \otimes \mathbf{A}$, and where the matrices $\mathbf{G}_k^{\boldsymbol{\beta}}$ and $\mathbf{L}_k^{\boldsymbol{\beta}}$ are defined in Appendix A.1. When $\mathrm{K} := (K_1, \ldots, K_m)$ denotes the m-tuple of score functions associated with the densities $\mathrm{f} = (f_1, \ldots, f_m)$, this vector $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}}$ is a rank-based version, computed at $\hat{\boldsymbol{\vartheta}}$, of the $\boldsymbol{\beta}$-part $\boldsymbol{\Delta}^{IV}_{\boldsymbol{\vartheta};\mathrm{f}}$ of the central sequence appearing in Proposition A.1. Proposition A.2 (in Appendix A.2) summarizes its asymptotic properties.

The *preliminary estimator* $\hat{\boldsymbol{\vartheta}}$, however, should satisfy the following assumption.

Assumption (A5). The estimator

$$\hat{\boldsymbol{\vartheta}} = \left(\hat{\boldsymbol{\theta}}'_1, \ldots, \hat{\boldsymbol{\theta}}'_m, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2, (\mathrm{d}\overset{\circ}{\mathrm{ve}}\mathrm{c}\, \hat{\boldsymbol{\Lambda}}_1^{\mathbf{V}})', \ldots, (\mathrm{d}\overset{\circ}{\mathrm{ve}}\mathrm{c}\, \hat{\boldsymbol{\Lambda}}_m^{\mathbf{V}})', (\mathrm{vec}\, \hat{\boldsymbol{\beta}})' \right)'$$

is such that (i) $\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta} = O_{\mathrm{P}}(n^{-1/2} \boldsymbol{\varsigma}^{(n)})$ under $\bigcup_{\mathrm{g} \in (\mathcal{F}_a)^m} \{\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}\}$, and (ii) $\hat{\boldsymbol{\vartheta}}$ is *locally and asymptotically discrete*, that is, it only takes a bounded number of distinct values in balls with $O(n^{-1/2} \boldsymbol{\varsigma}^{(n)})$ radius centered at $\boldsymbol{\vartheta}$.

Assumption (A5)(i) requires $\hat{\boldsymbol{\vartheta}}$ to be root-n consistent under the whole set $(\mathcal{F}_a)^m$ of m-tuples g of standardized radial densities ensuring ULAN. As for Assumption (A5)(ii), it is the traditional assumption of local asymptotic discreteness, which is easily enforced by discretizing $\hat{\boldsymbol{\vartheta}}$ in an adequate way. Such discretization, however, is a purely technical requirement, with no practical consequences, and is only required in asymptotic statements (see, for instance, Hallin *et al.* 2006).

Estimators satisfying Assumption (A5) are easily obtained. The following one, based on the Hettmansperger and Randles median and Tyler's estimator of shape (see also, in a slightly different context, Luo *et al.* 2009), has quite attractive properties. To start with, compute the Hettmansperger and Randles (2002) affine-equivariant medians $\hat{\boldsymbol{\theta}}_1^{\text{HR}}, \ldots, \hat{\boldsymbol{\theta}}_m^{\text{HR}}$, and the (normalized; that is, with determinant one) shape estimators $\hat{\mathbf{V}}_1^{\text{Tyler}}, \ldots, \hat{\mathbf{V}}_m^{\text{Tyler}}$ of Tyler (1987) in each sample. Those estimators are implicitly defined by

$$\frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{U}_{ij}(\hat{\boldsymbol{\theta}}_i^{\text{HR}}, \hat{\mathbf{V}}_i^{\text{Tyler}}) = \mathbf{0} \quad \text{and} \quad \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{U}_{ij}(\hat{\boldsymbol{\theta}}_i^{\text{HR}}, \hat{\mathbf{V}}_i^{\text{Tyler}})\mathbf{U}'_{ij}(\hat{\boldsymbol{\theta}}_i^{\text{HR}}, \hat{\mathbf{V}}_i^{\text{Tyler}}) = \frac{1}{k}\mathbf{I}_k,$$

$i = 1, \ldots, m$, a system of equations for which good numerical solutions exist. The preliminary estimators $\text{d}\overset{\circ}{\text{ve}}\text{c}\,(\hat{\boldsymbol{\Lambda}}_1^{\mathbf{V}}), \ldots, \text{d}\overset{\circ}{\text{ve}}\text{c}\,(\hat{\boldsymbol{\Lambda}}_m^{\mathbf{V}}), \text{vec}\,\hat{\boldsymbol{\beta}}$ then are obtained by plugging the values of $\hat{\boldsymbol{\theta}}_1^{\text{HR}}, \ldots, \hat{\boldsymbol{\theta}}_m^{\text{HR}}, \hat{\mathbf{V}}_1^{\text{Tyler}}, \ldots, \hat{\mathbf{V}}_m^{\text{Tyler}}$ into Flury's Gaussian likelihood equations (1.1). Denote by $\hat{\boldsymbol{\vartheta}}_{\text{Tyler}}$ the resulting estimator (note that the scales σ_i^2, $i = 1, \ldots, m$ are not involved in $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\text{K}}$, hence do not need be estimated). That preliminary estimator $\hat{\boldsymbol{\vartheta}}_{\text{Tyler}}$ satisfies the required consistency assumption: see Boente *et al.* (2002) for details.

Many other choices for $\hat{\boldsymbol{\vartheta}}$ are possible, though. In the Monte-Carlo study of Section 5 below, we also consider the preliminary estimator $\hat{\boldsymbol{\vartheta}}_{\text{MCD}}$ obtained from the robust Minimum Covariance Determinant (MCD) estimators of location/shape described, e.g., in Rousseuw and Leroy (1987). Note, however, that, contrary to $\hat{\boldsymbol{\vartheta}}_{\text{Tyler}}$ and $\hat{\boldsymbol{\vartheta}}_{\text{MCD}}$ (for the asymptotic behavior of the latter, see Cantor and Lopuhaä (2010)), Flury's covariance-based estimator $\hat{\boldsymbol{\vartheta}}_{\text{MLE}}$ does not satisfy the consistency requirements, as it loses root-n consistency under non-Gaussian densities. Asymptotically, the choice of $\hat{\boldsymbol{\vartheta}}$ does not affect the asymptotic properties of our R-estimators; it seems, from the simulations in Section 5, that the impact of that choice on their finite-sample behavior is quite limited as well.

It follows from Proposition A.2 in the Appendix that a natural estimator for $\boldsymbol{\beta}$ would be the matrix $\tilde{\boldsymbol{\beta}}_{\text{K};\mathcal{J}_k(\text{K,g})}$ defined by

$$\text{vec}(\tilde{\boldsymbol{\beta}}_{\text{K};\mathcal{J}_k(\text{K,g})}) := \text{vec}(\hat{\boldsymbol{\beta}}) + n^{-1/2}(\boldsymbol{\Gamma}_{\hat{\boldsymbol{\vartheta}};\text{K,g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\vartheta}};\text{K}}, \tag{3.3}$$

where $\mathbf{A}^-$ stands for the Moore-Penrose inverse of $\mathbf{A}$ and (see Section 2.3 and Appendix A.1 for the definitions of $\mathcal{J}_k(K_i, g_i)$ and $\boldsymbol{\nu}^{(i)}$, respectively)

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}} := \frac{1}{4k(k+2)} \mathbf{G}_k^{\boldsymbol{\beta}} \left(\sum_{i=1}^{m} r_i \mathcal{J}_k(K_i, g_i) (\boldsymbol{\nu}^{(i)})^{-1} \right) (\mathbf{G}_k^{\boldsymbol{\beta}})'. \tag{3.4}$$

However, $\tilde{\boldsymbol{\beta}}_{\mathrm{K};\mathcal{J}_k(\mathrm{K},\mathrm{g})}$ still suffers two majors drawbacks: (i) it is not a genuine statistic, since it still depends on the cross-information quantities $\mathcal{J}_k(K_1, f_1), \ldots, \mathcal{J}_k(K_m, f_m)$, and (ii) in general, it does not belong to $\mathcal{SO}_k$.

Point (i) is easily taken care of by plugging into $\boldsymbol{\Gamma}_{\hat{\boldsymbol{\vartheta}};\mathrm{K},\mathrm{g}}$ the consistent estimators

$$\widehat{\mathcal{J}}_k(\mathrm{K}, \mathrm{g}) := (\hat{\mathcal{J}}_k(K_1, g_1), \ldots, \hat{\mathcal{J}}_k(K_m, g_m))$$

of $\mathcal{J}_k(K_1, f_1), \ldots, \mathcal{J}_k(K_m, f_m)$ defined in Section 7 of Hallin *et al.* (2013), where we refer to for details. The notation $\widehat{\mathcal{J}}_k(\mathrm{K}, \mathrm{g})$ indicates an estimator of $\mathcal{J}_k(\mathrm{K}, \mathrm{g})$, where g is the actual, unspecified, m-tuple of radial densities—not a dependence on the unspecified g.

As for point (ii), we propose to bring $\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})}$ back to $\mathcal{SO}_k$ by means of the following simple Gram-Schmidt orthogonalization procedure. First, standardize $\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});1}$ into $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});1} := \tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});1} / \|\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});1}\|$; then, recursively, put

$$\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});l} := \frac{\left(\mathbf{I}_k - \sum_{j=1}^{l-1} \underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});j} \underset{\sim}{\boldsymbol{\beta}}'_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});j}\right) \tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});l}}{\left\| \left(\mathbf{I}_k - \sum_{j=1}^{l-1} \underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});j} \underset{\sim}{\boldsymbol{\beta}}'_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});j}\right) \tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});l} \right\|}, \quad l = 2, \ldots, k.$$

This eventually yields an R-estimator $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})} := \left(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});1}, \ldots, \underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g});k} \right)$ that belongs to $\mathcal{SO}_k$.

3.2 Asymptotic properties and AREs

It remains to justify the use of the estimators constructed in the previous section, by showing that they do enjoy the appealing properties announced in the Introduction.

In this section, we establish those properties. In particular, we prove that $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})}$ is root-n consistent and asymptotically normal, and that, when based on the score functions $\mathrm{K}_{\mathrm{f}} = (K_{f_1}, \ldots, K_{f_m})$ associated with the m-tuple of radial densities $\mathrm{f} = (f_1, \ldots, f_m)$, it is asymptotically efficient under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{f}}$.

Using the consistency of $\widehat{\mathcal{J}}_k(\mathrm{K}, \mathrm{g})$, Proposition A.2(iii), and the fact that

$$(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}})^{-} = k(k+2)\mathbf{G}_k^{\boldsymbol{\beta}}\Big(\sum_{i=1}^{m} r_i \mathcal{J}_k(K_i, g_i)(\boldsymbol{\nu}^{(i)})^{-1}\Big)^{-1}(\mathbf{G}_k^{\boldsymbol{\beta}})', \tag{3.5}$$

we obtain (see (3.3) for the definition of $\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})}$) that

$$\begin{aligned}
\underset{\sim}{\mathbf{T}}^{(n)} := n^{1/2}\mathrm{vec}(\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})} - \boldsymbol{\beta}) &= n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\boldsymbol{\Gamma}_{\hat{\boldsymbol{\vartheta}};\mathrm{K},\mathrm{g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\vartheta}};\mathrm{K}} \\
&= n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}})^{-}\Big(\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} - \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}} n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\Big) + o_{\mathrm{P}}(1) \\
&= n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} - \frac{1}{2}\mathbf{G}_k^{\boldsymbol{\beta}}(\mathbf{G}_k^{\boldsymbol{\beta}})' n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_{\mathrm{P}}(1),
\end{aligned} \tag{3.6}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$ as $n \to \infty$. The column vectors of the $k^2 \times k(k-1)/2$ matrix $\mathbf{G}_k^{\boldsymbol{\beta}}$ form a basis of the tangent space to $\mathrm{vec}(\mathcal{SO}_k)$ at $\mathrm{vec}(\boldsymbol{\beta})$. Lemma A.1 in Appendix A.3, which is of independent interest, shows that projecting $n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ onto this tangent space does not modify its asymptotic behavior; applying this to (3.6) directly yields that

$$n^{1/2}\mathrm{vec}(\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})} - \boldsymbol{\beta}) = (\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} + o_{\mathrm{P}}(1), \tag{3.7}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$ as $n \to \infty$. The asymptotic behavior of the proposed R-estimator $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})}$ then easily follows from applying Lemma A.2 in Appendix A.3 to (3.7), yielding, in view of (3.5), under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$ as $n \to \infty$,

$$\begin{aligned}
n^{1/2}\mathrm{vec}(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})} - \boldsymbol{\beta}) &= \mathbf{J}_k^{\boldsymbol{\beta}} n^{1/2}\mathrm{vec}(\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})} - \boldsymbol{\beta}) + o_{\mathrm{P}}(1) \\
&= \mathbf{J}_k^{\boldsymbol{\beta}}(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} + o_{\mathrm{P}}(1) = (\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} + o_{\mathrm{P}}(1).
\end{aligned} \tag{3.8}$$

The asymptotic properties of $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})}$, summarized in the following proposition, now follow from those of $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}}$ (Proposition A.2). Note that (3.8), by establishing the asymptotic equivalence of $n^{1/2}\mathrm{vec}(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})}-\boldsymbol{\beta})$ and the rank-measurable random vector $(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K,g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}}$, fully justifies calling $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})}$ an "R-estimator".

Proposition 3.1 *Let Assumptions (A1)-(A4) hold and let* $\hat{\boldsymbol{\vartheta}}$ *satisfy Assumption (A5). Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$, $\mathrm{g}\in(\mathcal{F}_a)^m$,

$$n^{1/2}\mathrm{vec}(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})}-\boldsymbol{\beta}) = (\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K,g}})^{-}\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} + o_{\mathrm{P}}(1)$$

is asymptotically normal with mean zero and covariance matrix

$$(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K,g}})^{-}\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K}}(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K,g}})^{-} = k(k+2)\mathbf{G}_k^{\boldsymbol{\beta}}\Big(\sum_{i=1}^m r_i\mathcal{J}_k(K_i,g_i)(\boldsymbol{\nu}^{(i)})^{-1}\Big)^{-1} \times\Big(\sum_{i=1}^m r_i\mathcal{J}_k(K_i)(\boldsymbol{\nu}^{(i)})^{-1}\Big)\Big(\sum_{i=1}^m r_i\mathcal{J}_k(K_i,g_i)(\boldsymbol{\nu}^{(i)})^{-1}\Big)^{-1}(\mathbf{G}_k^{\boldsymbol{\beta}})'. \quad (3.9)$$

If $\mathrm{g}=(g_1,\ldots,g_1)$ (homogeneous elliptical densities), and if the same score function, $K_1:(0,1)\to\mathbb{R}$, say, is used for the m rankings, then the covariance matrix (3.9) reduces to

$$(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K,g}})^{-}\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K}}(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K,g}})^{-} = k(k+2)\frac{\mathcal{J}_k(K_1)}{\mathcal{J}_k^2(K_1,g_1)}\,\mathbf{G}_k^{\boldsymbol{\beta}}\Big(\sum_{i=1}^m r_i(\boldsymbol{\nu}^{(i)})^{-1}\Big)^{-1}(\mathbf{G}_k^{\boldsymbol{\beta}})'.$$

Under the additional assumption of finite fourth-order moments, letting

$$\kappa_k(f_i) := \frac{k}{k+2}\frac{\int_0^1(\tilde{F}_{ik}^{-1}(u))^4\,du}{\big(\int_0^1(\tilde{F}_{ik}^{-1}(u))^2\,du\big)^2}-1$$

denote the *kurtosis* of the ith elliptic population (see, e.g., page 54 of Anderson 2003), the asymptotic relative efficiency of $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})}$ with respect to the Flury (1984) Gaussian MLE $\hat{\boldsymbol{\beta}}$ in (1.1) takes the simple form (see Hallin *et al.* (2008) for the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ in that case)

$$\mathrm{ARE}_{k,\mathrm{g}}(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})}/\hat{\boldsymbol{\beta}}) = \frac{(1+\kappa_k(g_1))}{k(k+2)}\frac{\mathcal{J}_k^2(K_1,g_1)}{\mathcal{J}_k(K_1)}. \quad (3.10)$$

The AREs in (3.10) coincide with those obtained in one-sample shape problems: see Hallin and Paindaveine (2006), and Hallin *et al.* (2006, 2010b). The Chernoff-Savage property of Paindaveine (2006) therefore extends to the present CPC context: denoting by $\underset{\sim}{\boldsymbol{\beta}}_{\rm vdW}$ the van der Waerden estimator (based on the Gaussian scores $K_1 = \ldots = K_m := \Psi_k^{-1}$; see Section 2.3), we have

$$\text{ARE}_{k,\mathrm{g}}(\underset{\sim}{\boldsymbol{\beta}}_{\rm vdW}/\hat{\boldsymbol{\beta}}) \geq 1 \tag{3.11}$$

for all homogeneous $\mathrm{g} \in (\mathcal{F}_a^4)^m$, with equality in the Gaussian case only. Our van der Waerden estimator of CPC thus is not just more robust than Flury's MLE, it also uniformly outperforms it, in the Pitman sense, under homogeneous elliptical densities.

Denote by $\hat{\boldsymbol{\beta}}_A$ the estimator of $\boldsymbol{\beta}$ obtained by replacing, in the Gaussian likelihood equations (1.1), the covariance matrices $\mathbf{S}_1, \ldots, \mathbf{S}_m$ by root-n consistent estimators of shape $\hat{\mathbf{V}}_{A,1}, \ldots, \hat{\mathbf{V}}_{A,m}$ (typically, robust ones). It follows from Boente *et al.* (2002) that $n^{1/2}\text{vec}(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta})$ is asymptotically normal (still in the homogeneous elliptical case $\mathrm{g} = (g_1, \ldots, g_1)$), with mean zero and covariance matrix

$$\rho(A, g_1)\, \mathbf{G}_k^{\boldsymbol{\beta}} \Big(\sum_{i=1}^m r_i (\boldsymbol{\nu}^{(i)})^{-1} \Big)^{-1} \mathbf{G}_k^{\boldsymbol{\beta}\prime},$$

for some scalar $\rho(A, g_1)$ governing the efficiency properties of the off-diagonal elements of $\hat{\mathbf{V}}_A$ (their role is comparable to that of our cross-information quantities: see Croux and Haesbroeck 2000 for similar results in the PCA context). It follows that the asymptotic relative efficiency, in the homogeneous elliptical case $\mathrm{g} = (g_1, \ldots, g_1)$, of $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})}$ with respect to $\hat{\boldsymbol{\beta}}_A$ is

$$\text{ARE}_{k,\mathrm{g}}(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K},\mathrm{g})}/\hat{\boldsymbol{\beta}}_A) = \frac{\rho(A, g_1)}{k(k+2)} \frac{\mathcal{J}_k^2(K_1, g_1)}{\mathcal{J}_k(K_1)}. \tag{3.12}$$

Some numerical values of (3.10) are provided in Table 1, which also provides AREs with respect to the (50% breakdown point) MCD shape estimator $\hat{\mathbf{V}}_{\rm MCD}$. Note that

Table 1: AREs of the R-estimators $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})}$ based on van der Waerden (vdW), Wilcoxon (W), and t_5 scores with respect to Flury's Gaussian estimator $\hat{\boldsymbol{\beta}}$ (in brackets, with respect to the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{MCD}}$ obtained from the MCD estimator of shape), under k-dimensional Student (with 5, 8, and 12 degrees of freedom), and Gaussian densities, for $k = 2$, 3, 4, 6, 10, and 250.

		underlying density			
K	k	t_5	t_8	t_{12}	$\mathcal{N}$
vdW	2	2.204 (13.721)	1.215 (17.623)	1.078 (20.664)	1.000 (30.018)
	3	2.270 (7.617)	1.233 (9.453)	1.086 (10.935)	1.000 (15.835)
	4	2.326 (5.587)	1.249 (6.747)	1.093 (7.710)	1.000 (11.114)
	6	2.413 (4.051)	1.275 (4.698)	1.106 (5.262)	1.000 (7.504)
	10	2.531 (3.113)	1.312 (3.438)	1.126 (3.745)	1.000 (5.223)
	250	2.959 (2.194)	1.480 (2.149)	1.234 (2.128)	1.000 (2.331)
W	2	2.258 (14.056)	1.174 (17.023)	1.001 (19.197)	0.844 (25.328)
	3	2.386 (8.004)	1.246 (9.557)	1.068 (10.756)	0.913 (14.457)
	4	2.432 (5.843)	1.273 (6.881)	1.094 (7.716)	0.945 (10.506)
	6	2.451 (4.113)	1.283 (4.729)	1.105 (5.256)	0.969 (7.272)
	10	2.426 (2.983)	1.264 (3.313)	1.088 (3.619)	0.970 (5.069)
	250	2.262 (1.677)	1.135 (1.648)	0.950 (1.637)	0.821 (1.913)
t_5	2	2.333 (14.526)	1.244 (18.039)	1.078 (20.676)	0.945 (28.355)
	3	2.400 (8.052)	1.264 (9.689)	1.089 (10.967)	0.946 (14.980)
	4	2.455 (5.896)	1.281 (6.921)	1.099 (7.749)	0.948 (10.531)
	6	2.538 (4.261)	1.309 (4.824)	1.115 (5.305)	0.951 (7.134)
	10	2.647 (3.255)	1.347 (3.531)	1.139 (3.788)	0.956 (4.995)
	250	2.977 (2.207)	1.488 (2.161)	1.240 (2.138)	0.994 (2.317)

the 50% breakdown point of the MCD estimator implies a very high cost in terms of efficiency, with AREs of the order of 30 in dimension 2, under Gaussian densities.

Finally, note that, when $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K_f};\widehat{\mathcal{J}}_k(\mathrm{K_f},\mathrm{g})}$ is based on the score functions $\mathrm{K_f} = (K_{f_1}, \ldots, K_{f_m})$ with $K_{f_i}(u) := \varphi_{f_i}(\tilde{F}_i^{-1}(u))\tilde{F}_i^{-1}(u)$, then $n^{1/2}\mathrm{vec}(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K_f};\widehat{\mathcal{J}}_k(\mathrm{K_f},\mathrm{g})} - \boldsymbol{\beta})$ is, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{f}}$ with $\mathrm{f} = (f_1, \ldots, f_m)$, asymptotically normal with mean zero and covariance matrix

$$k(k+2)\mathbf{G}_k^{\boldsymbol{\beta}}\Big(\sum_{i=1}^{m} r_i \mathcal{J}_k(K_{f_i})(\boldsymbol{\nu}^{(i)})^{-1}\Big)^{-1}(\mathbf{G}_k^{\boldsymbol{\beta}})' = k(k+2)\mathbf{G}_k^{\boldsymbol{\beta}}\Big(\sum_{i=1}^{m} r_i \mathcal{J}_k(f_i)(\boldsymbol{\nu}^{(i)})^{-1}\Big)^{-1}(\mathbf{G}_k^{\boldsymbol{\beta}})',$$

where the right-hand side is nothing else but the Moore-Penrose inverse of the Fisher information for $\boldsymbol{\beta}$ at $\mathrm{f} = (f_1, \ldots, f_m)$. It follows that the R-estimator $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K_f};\widehat{\mathcal{J}}_k(\mathrm{K_f},\mathrm{g})}$ is asymptotically efficient under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{f}}$ (it achieves the parametric efficiency bound).

4 Rank-based PCA

In the one-sample setup ($m = 1$), common principal components reduce to ordinary principal components, and it can be expected that the methodology just described yields estimators enjoying the same type of asymptotic properties as in Section 3.2. We show in this section that this is indeed the case.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from an elliptical distribution with location $\boldsymbol{\theta}$, scale σ, shape matrix $\mathbf{V} = \boldsymbol{\beta}\boldsymbol{\Lambda}^{\mathbf{V}}\boldsymbol{\beta}'$, and radial density f_1. Put $\mathbf{U}_i := \mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})/d_i$, where $d_i := d_i(\boldsymbol{\theta}, \mathbf{V}) := \|\mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})\|$, $i = 1, \ldots, n$, and write $R_i := R_i(\boldsymbol{\theta}, \mathbf{V})$ for the rank of d_i among $d_1, \ldots, d_n$. In this one-sample setup, we write $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};f}$ for the joint distribution of the $\mathbf{X}_i$'s under parameter value $\boldsymbol{\vartheta} := (\boldsymbol{\theta}', \sigma^2, (\mathrm{d}\overset{\circ}{\mathrm{vec}}\,\boldsymbol{\Lambda}^{\mathbf{V}})', (\mathrm{vec}\,\boldsymbol{\beta})')'$ and radial density f_1.

The one-sample versions of the rank-based central sequence in (3.2) and the cross-information matrix in (3.4) are (for a score function K satisfying Assumption (A4))

$$\underset{\sim}{\mathbf{\Delta}}_{\hat{\boldsymbol{\vartheta}};K}, \quad \text{with} \quad \underset{\sim}{\mathbf{\Delta}}_{\boldsymbol{\vartheta};K} = \frac{1}{2n^{1/2}} \mathbf{G}_k^{\boldsymbol{\beta}} \mathbf{L}_k^{\boldsymbol{\beta},\mathbf{\Lambda}^{\mathbf{V}}} \left(\mathbf{V}^{\otimes 2}\right)^{-1/2} \sum_{i=1}^{n} K\left(\frac{R_i}{n+1}\right) \mathrm{vec}\,(\mathbf{U}_i \mathbf{U}_i'),$$

and

$$\mathbf{\Gamma}_{\boldsymbol{\vartheta};K,g_1} = \frac{\mathcal{J}_k(K,g_1)}{4k(k+2)} \mathbf{G}_k^{\boldsymbol{\beta}} \boldsymbol{\nu}^{-1} (\mathbf{G}_k^{\boldsymbol{\beta}})',$$

respectively, where $\boldsymbol{\nu} := \mathrm{diag}(\nu_{12}, \nu_{13}, \ldots, \nu_{(k-1)k})$, with $\nu_{jh} := \lambda_j^{\mathbf{V}} \lambda_h^{\mathbf{V}} / (\lambda_j^{\mathbf{V}} - \lambda_h^{\mathbf{V}})^2$. Working along the same lines as in Section 3.1, define

$$\mathrm{vec}(\tilde{\boldsymbol{\beta}}_{K;\mathcal{J}_k(K,g_1)}) = \mathrm{vec}(\hat{\boldsymbol{\beta}}) + n^{-1/2} (\mathbf{\Gamma}_{\hat{\boldsymbol{\vartheta}};K,g_1})^{-} \underset{\sim}{\mathbf{\Delta}}_{\hat{\boldsymbol{\vartheta}};K},$$

where $\hat{\boldsymbol{\vartheta}} := (\hat{\boldsymbol{\theta}}', \hat{\sigma}^2, (\mathrm{d}\overset{\circ}{\mathrm{ve}}\mathrm{c}\,\hat{\mathbf{\Lambda}}^{\mathbf{V}})', (\mathrm{vec}\,\hat{\boldsymbol{\beta}})')'$ is a (adequately discretized) root-n consistent preliminary estimator. Letting $\widehat{\mathcal{J}}_k(K,g_1)$ be a consistent estimator of the cross-information quantity $\mathcal{J}_k(K,g_1)$, the final estimator is

$$\underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1)} := \left(\underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);1}, \ldots, \underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);k}\right),$$

where

$$\underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);1} := \tilde{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);1} / \|\tilde{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);1}\|$$

and, recursively,

$$\underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);l} := \frac{\left(\mathbf{I}_k - \sum_{j=1}^{l-1} \underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);j} \underset{\sim}{\boldsymbol{\beta}}'_{K;\widehat{\mathcal{J}}_k(K,g_1);j}\right) \tilde{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);l}}{\|\left(\mathbf{I}_k - \sum_{j=1}^{l-1} \underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);j} \underset{\sim}{\boldsymbol{\beta}}'_{K;\widehat{\mathcal{J}}_k(K,g_1);j}\right) \tilde{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1);l}\|}, \quad l = 2, \ldots, k.$$

As the following result shows, this PCA R-estimator $\underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1)}$ has the same asymptotic properties as its CPC counterpart: root-n consistency, asymptotic normality, and asymptotic efficiency under correctly specified radial densities.

Proposition 4.1 *Let $\hat{\boldsymbol{\vartheta}}$ stand for a locally and asymptotically discrete estimator (see Asumption (A5)) such that $\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta} = O_{\mathrm{P}}(n^{-1/2})$ under $\bigcup_{g_1 \in \mathcal{F}_a} \mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$ and K be a score* *function satisfying Assumption (A4). Furthermore let (the one sample versions of) Assumptions (A1)-(A2) hold. Then,*

(i) $n^{1/2}\mathrm{vec}(\underset{\sim}{\boldsymbol{\beta}}_{K;\widehat{\mathcal{J}}_k(K,g_1)} - \boldsymbol{\beta})$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};g_1}$ *is asymptotically normal with mean zero and covariance matrix*

$$\frac{k(k+2)\mathcal{J}_k(K)}{\mathcal{J}_k^2(K,g_1)}\mathbf{G}_k^{\boldsymbol{\beta}}\boldsymbol{\nu}(\mathbf{G}_k^{\boldsymbol{\beta}})';$$

(ii) when based on the score function $K_{f_1}(u) := \varphi_{f_1}(\tilde{F}_1^{-1}(u))\tilde{F}_1^{-1}(u)$*, the R-estimator* $\underset{\sim}{\boldsymbol{\beta}}_{K_{f_1};\widehat{\mathcal{J}}_k(K_{f_1},g_1)}$ *is asymptotically efficient under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};f_1}$.

The AREs in (3.10) thus remain valid under finite fourth-order moments, and the Chernoff-Savage result in (3.11) still holds, since $m = 1$ trivially implies homogeneity of radial densities.

5 Monte-Carlo study

This section presents a numerical study of the finite-sample performances of our R-estimators under various light- and heavy-tailed population densities, for various scores and preliminary estimators, both for CPC and PCA.

5.1 CPC

We considered two distinct CPC setups: (i) a "proportional" CPC setup, that involves eigenvalues matrices that are proportional to each other, and (ii) a "non-proportional" CPC setup that does not exhibit such proportionality structure. In both cases, we generated $N = 1,500$ independent replications of four pairs ($m = 2$) of mutually independent samples with respective (and relatively small) sizes $n_1 = 100$ and $n_2 = 150$

$$\boldsymbol{\varepsilon}_{\ell;1j}, \quad j = 1, \ldots, n_1 = 100, \quad \text{and} \quad \boldsymbol{\varepsilon}_{\ell;2j}, \quad j = 1, \ldots, n_2 = 150, \quad \ell = a, b, c, d$$

of bivariate ($k = 2$) spherical random vectors, where

- (power-exponential/Gaussian case) the $\boldsymbol{\varepsilon}_{a;1j}$'s have a power-exponential radial density with parameter $\eta = 10$ ($\mathcal{E}_{10}$), and the $\boldsymbol{\varepsilon}_{a;2j}$'s are standard normal;
- (Gaussian/Gaussian case) the $\boldsymbol{\varepsilon}_{b;1j}$'s and the $\boldsymbol{\varepsilon}_{b;2j}$'s are standard normal;
- (Gaussian/Student t_5 case) the $\boldsymbol{\varepsilon}_{c;1j}$ are standard normal and the $\boldsymbol{\varepsilon}_{c;2j}$'s have a t_5 radial density;
- (Student t_5/Cauchy case) the $\boldsymbol{\varepsilon}_{d;1j}$'s have a t_5 radial density, and the $\boldsymbol{\varepsilon}_{d;2j}$'s have a t_1 radial density;

Note that both lighter-than-Gaussian tails ($\mathcal{E}_{10}$) and heavier-than-Gaussian tails (t_5, t_1) are considered.

In the first setup, each replication of the $\boldsymbol{\varepsilon}_{\ell;1j}$'s was linearly transformed into

$$\mathbf{X}_{\ell;1j} = \boldsymbol{\beta}\boldsymbol{\Lambda}_1^{1/2}\boldsymbol{\varepsilon}_{\ell;1j}, \quad \ell = a, b, c, d, \quad j = 1, \ldots, n_1 = 100,$$

with $\boldsymbol{\beta} = \mathbf{I}_2$ and $\boldsymbol{\Lambda}_1 = \text{diag}(2, 1)$, each replication of the $\boldsymbol{\varepsilon}_{\ell;2j}$'s into

$$\mathbf{X}_{\ell;2j} = \boldsymbol{\beta}\boldsymbol{\Lambda}_2^{1/2}\boldsymbol{\varepsilon}_{\ell;2j}, \quad \ell = a, b, c, d, \quad j = 1, \ldots, n_2 = 150, \;\; \text{with } \boldsymbol{\Lambda}_2 := 2\boldsymbol{\Lambda}_1 = \text{diag}(4, 2).$$

The second setup rather uses $\boldsymbol{\Lambda}_2 := \text{diag}(3, 1)$.

For each replication, we computed the preliminary estimators $\hat{\boldsymbol{\beta}}_{\text{MLE}}$, $\hat{\boldsymbol{\beta}}_{\text{Tyler}}$ and $\hat{\boldsymbol{\beta}}_{\text{MCD}}$, along with the resulting one-step van der Waerden R-estimators $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{vdW}}$ (Gaussian scores in each sample), one-step Wilcoxon R-estimators $\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{W}}$ (Wilcoxon scores in each sample), one-step R-estimators $\underset{\sim}{\boldsymbol{\beta}}_{(\mathcal{N},t_5)}$ (Gaussian scores in the first sample, t_5 scores in the second one) and $\underset{\sim}{\boldsymbol{\beta}}_{(t_5,t_1)}$ (t_5 scores in the first sample, t_1 scores in the second one). For each of those R-estimators $\underset{\sim}{\boldsymbol{\beta}} = (\underset{\sim}{\boldsymbol{\beta}}_1, \underset{\sim}{\boldsymbol{\beta}}_2)$, taking values $\underset{\sim}{\boldsymbol{\beta}}^{(\nu)} = (\underset{\sim}{\boldsymbol{\beta}}_1^{(\nu)}, \underset{\sim}{\boldsymbol{\beta}}_2^{(\nu)})$ in replication ν, we computed the mean squared errors

$$\gamma_\nu := n^{-1} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left\| (\mathbf{X}'_{\ell;ij} \underset{\sim}{\boldsymbol{\beta}}_1^{(\nu)}) \underset{\sim}{\boldsymbol{\beta}}_1^{(\nu)} - (\mathbf{X}'_{\ell;ij} \boldsymbol{\beta}_1) \boldsymbol{\beta}_1 \right\|^2, \quad \nu = 1, \ldots, N = 1,500. \tag{5.1}$$

Those γ_ν's provide measures of the performances of the various $\underset{\sim}{\boldsymbol{\beta}}_1^{(\nu)}$'s in the estimation of the first common eigenvector $\boldsymbol{\beta}_1$ in replication ν. Tables 2 and 3 report boxplots for those γ_ν's in the first and second setups, respectively; since γ_ν is intrinsically nonnegative, those boxplots, reporting side quantiles only, are one-sided (from the bottom upwards: first quartile, median, third quartile, and a whisker at the .95 quantile).

Inspection of these tables reveals that the results are uniformly good, and that one-step R-estimators, as a rule, do improve over the corresponding preliminary estimators. The performances in Table 2 being very similar to those in Table 3, our discussion concentrates on Table 2.

Flury's Gaussian MLE, as expected, produces excellent results in the light-tailed cases (a) and (b). In the Gaussian case (b), the impact of the one-step improvement is essentially nil, irrespective of the scores considered: in case (b), no improvement is possible asymptotically while, in the power-exponential case (a), improvement is almost imperceptible. However, the performance of $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ rapidly deteriorates as tails get heavier. Under the t_5/t_1 case (d), the mean squared error for $\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ explodes (in agreement with the fact that root-n consistency does not hold anymore), a situation the one-step R-estimators only partially manage to straighten out—although dividing the median squared error by two. One should thus avoid considering Flury's $\hat{\boldsymbol{\beta}}_{\text{MLE}}$ as a preliminary as soon as one of the samples involved in the CPC analysis is likely to exhibit heavy tails.

Although $\hat{\boldsymbol{\beta}}_{\text{MCD}}$ and $\hat{\boldsymbol{\beta}}_{\text{Tyler}}$ have very similar behaviors under light-tailed densities, $\hat{\boldsymbol{\beta}}_{\text{Tyler}}$ clearly dominates $\hat{\boldsymbol{\beta}}_{\text{MCD}}$ under the heavy-tailed ones. The second column of Table 2 leads to the following conclusions for the choice of $\hat{\boldsymbol{\beta}}_{\text{MCD}}$ as a preliminary: in the presence (t_5/t_1 case (d)) of heavy tails in one of the samples, and although root-n consistency still does hold, its median performance is not that bad, but its mean squared errors is quite poor in the upper tail, a behavior for which the one-step R-estimators only partly compensate. A Tyler preliminary $\hat{\boldsymbol{\beta}}_{\text{Tyler}}$, along with van der Waerden or Wilcoxon scores, thus seems to be the safest choice, yielding, in the Gaussian case (b), a moderate increase of about 30% over the optimal Gaussian MLE of the median of mean squared errors, but dividing it by a factor eight in the t_5/t_1 case (d).

5.2 PCA

In the one-sample case, we similarly generated $N = 1,500$ independent replications of four independent samples (with small sample size $n = 150$) of ($k = 4$)-dimensional spherical random vectors

$$\boldsymbol{\varepsilon}_{\ell;j}, \quad j = 1, \ldots, n = 150, \quad \ell = a, b, c, d,$$

where

- (power-exponential case) the $\boldsymbol{\varepsilon}_{a;j}$'s have a power-exponential ($\mathcal{E}_{10}$) radial density;
- (Gaussian case) the $\boldsymbol{\varepsilon}_{b;j}$'s are standard normal;

– (Student t_5 case) the $\boldsymbol{\varepsilon}_{c;j}$'s have a t_5 radial density;

– (Cauchy t_1 case) the $\boldsymbol{\varepsilon}_{d;j}$'s have a t_1 radial density.

Each replication of the $\boldsymbol{\varepsilon}_{\ell;j}$'s was transformed into

$$\mathbf{X}_{\ell;j} = \boldsymbol{\beta}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\varepsilon}_{\ell;j}, \quad j = 1, \ldots, 150, \quad \ell = a, b, c, d,$$

with $\boldsymbol{\Lambda} := \text{diag}(4, 3, 2, 1)$, and $\boldsymbol{\beta} = \mathbf{I}_4$. For each replication, we computed the eigenvectors $\hat{\boldsymbol{\beta}}_{\text{MLE}}$, $\hat{\boldsymbol{\beta}}_{\text{MCD}}$, $\hat{\boldsymbol{\beta}}_{\text{Tyler}}$ of the empirical covariance, the MCD and the Tyler matrices, respectively. Based on these, we also computed the one-step van der Waerden, Wilcoxon, and Student R-estimators $\underset{\sim}{\boldsymbol{\beta}}_{\text{vdW}}$ (Gaussian scores), $\underset{\sim}{\boldsymbol{\beta}}_{\text{W}}$ (Wilcoxon scores), $\underset{\sim}{\boldsymbol{\beta}}_{(t_5)}$ and $\underset{\sim}{\boldsymbol{\beta}}_{(t_1)}$ (t_5 and t_1 scores, respectively). For each of those R-estimators $\underset{\sim}{\boldsymbol{\beta}} = (\underset{\sim}{\boldsymbol{\beta}}_1, \underset{\sim}{\boldsymbol{\beta}}_2, \underset{\sim}{\boldsymbol{\beta}}_3, \underset{\sim}{\boldsymbol{\beta}}_4)$, taking value $\underset{\sim}{\boldsymbol{\beta}}^{(\nu)} = (\underset{\sim}{\boldsymbol{\beta}}_1^{(\nu)}, \underset{\sim}{\boldsymbol{\beta}}_2^{(\nu)}, \underset{\sim}{\boldsymbol{\beta}}_3^{(\nu)}, \underset{\sim}{\boldsymbol{\beta}}_4^{(\nu)})$ in replication ν, and for each replication, we evaluated the estimation performance via the mean squared error

$$\gamma_\nu := n^{-1} \sum_{i=1}^{n} \left\| (\mathbf{X}'_{\ell;i}\underset{\sim}{\boldsymbol{\beta}}_1^{(\nu)})\underset{\sim}{\boldsymbol{\beta}}_1^{(\nu)} - (\mathbf{X}'_{\ell;i}\boldsymbol{\beta}_1)\boldsymbol{\beta}_1 \right\|^2, \quad \nu = 1, \ldots, N = 1,500. \tag{5.2}$$

One-sided boxplots (from the bottom upwards: first quartile, median, third quartile, and a whisker at the .95 quantile) of the γ_ν's are provided in Table 4. Inspection of those boxplots calls for very similar comments as Tables 2-3: the Gaussian MLE preliminary is definitely dangerous, while the MCD one behaves rather poorly, under heavy-tailed distributions such as the Cauchy. The best overall performance seems to be that of a Tyler preliminary, along with van der Waerden or Wilcoxon scores.

Acknowledgments

The research of Marc Hallin and Davy Paindaveine is supported by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). Davy Paindaveine moreover benefits from an A.R.C. contract of the Communauté Française de Belgique. We gratefully acknowledge the constructive remarks of an Associate Editor and three anonymous referees.

A Appendix

A.1 ULAN

Consider an arbitrary *local sequence*

$$\begin{aligned}\boldsymbol{\vartheta}^{(n)} &:= (\boldsymbol{\vartheta}_{I}^{(n)\prime}, \boldsymbol{\vartheta}_{II}^{(n)\prime}, \boldsymbol{\vartheta}_{III}^{(n)\prime}, \boldsymbol{\vartheta}_{IV}^{(n)\prime})' := (\boldsymbol{\theta}_1^{(n)\prime}, \ldots, \boldsymbol{\theta}_m^{(n)\prime}, \\ &\qquad \sigma_1^{2(n)}, \ldots, \sigma_m^{2(n)}, (\mathrm{d\overset{\circ}{ve}c}\, \boldsymbol{\Lambda}_1^{\mathbf{V}(n)})', \ldots, (\mathrm{d\overset{\circ}{ve}c}\, \boldsymbol{\Lambda}_m^{\mathbf{V}(n)})', (\mathrm{vec}\, \boldsymbol{\beta}^{(n)})')' \in \boldsymbol{\Theta},\end{aligned}$$

where $\boldsymbol{\vartheta}^{(n)} - \boldsymbol{\vartheta} = O(n^{-1/2})$, and further sequences of the form $\boldsymbol{\vartheta}^{(n)} + n^{-1/2}\boldsymbol{\varsigma}^{(n)}\boldsymbol{\tau}^{(n)}$, where

$$\boldsymbol{\tau}^{(n)} = (\boldsymbol{\tau}_{I}^{(n)\prime}, \boldsymbol{\tau}_{II}^{(n)\prime}, \boldsymbol{\tau}_{III}^{(n)\prime}, \boldsymbol{\tau}_{IV}^{(n)\prime})' = (\mathbf{t}_1^{(n)\prime}, \ldots, \mathbf{t}_m^{(n)\prime}, s_1^{(n)}, \ldots, s_m^{(n)}, \mathbf{l}_1^{(n)\prime}, \ldots, \mathbf{l}_m^{(n)\prime}, (\mathrm{vec}\, \mathbf{b}^{(n)})')'$$

is such that $\sup_n \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\tau}^{(n)} < \infty$ and $\boldsymbol{\vartheta}^{(n)} + n^{-1/2}\boldsymbol{\varsigma}^{(n)}\boldsymbol{\tau}^{(n)} \in \boldsymbol{\Theta}$. Strong restrictions are required on $\boldsymbol{\tau}^{(n)} = (\boldsymbol{\tau}_{I}^{(n)\prime}, \boldsymbol{\tau}_{II}^{(n)\prime}, \boldsymbol{\tau}_{III}^{(n)\prime}, \boldsymbol{\tau}_{IV}^{(n)\prime})'$ if the perturbed parameter values $\boldsymbol{\vartheta}^{(n)} + n^{-1/2}\boldsymbol{\varsigma}^{(n)}\boldsymbol{\tau}^{(n)}$ are to belong to $\boldsymbol{\Theta}$. In particular, the perturbed orthogonal matrix should remain orthogonal; we refer to Hallin et al. (2010b) for details.

Denoting by $\mathbf{e}_\ell$ the ℓth vector of the canonical basis of $\mathbb{R}^k$, let $\mathbf{K}_k := \sum_{i,j=1}^k (\mathbf{e}_i\mathbf{e}_j') \otimes (\mathbf{e}_j\mathbf{e}_i')$ denote the classical $(k^2 \times k^2)$ *commutation matrix*. Define $\mathbf{H}_k$ as the $k \times k^2$ matrix such that $\mathbf{H}_k \mathrm{vec}\,(\mathbf{A}) = \mathrm{dvec}\,(\mathbf{A})$ for any $k \times k$ matrix $\mathbf{A}$. For any $k \times k$ diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$, write $\mathbf{M}_k^{\boldsymbol{\Lambda}}$ for the $(k-1) \times k$ matrix $\big(-\lambda_1(\lambda_2^{-1}, \ldots, \lambda_k^{-1})' \,\vdots\, \mathbf{I}_{k-1}\big)$ and $\mathbf{L}_k^{\boldsymbol{\beta}, \boldsymbol{\Lambda}_i^{\mathbf{V}}}$ for $(\mathbf{L}_{k;12}^{\boldsymbol{\beta}, \boldsymbol{\Lambda}_i^{\mathbf{V}}}\ \mathbf{L}_{k;13}^{\boldsymbol{\beta}, \boldsymbol{\Lambda}_i^{\mathbf{V}}} \ldots \mathbf{L}_{k;(k-1)k}^{\boldsymbol{\beta}, \boldsymbol{\Lambda}_i^{\mathbf{V}}})'$, with $\mathbf{L}_{k;jh}^{\boldsymbol{\beta}, \boldsymbol{\Lambda}_i^{\mathbf{V}}} := (\lambda_{ih}^{\mathbf{V}} - \lambda_{ij}^{\mathbf{V}})(\boldsymbol{\beta}_h \otimes \boldsymbol{\beta}_j)$. Finally, let $\mathbf{G}_k^{\boldsymbol{\beta}} := (\mathbf{G}_{k;12}^{\boldsymbol{\beta}}\ \mathbf{G}_{k;13}^{\boldsymbol{\beta}} \ldots \mathbf{G}_{k;(k-1)k}^{\boldsymbol{\beta}})$, with $\mathbf{G}_{k;jh}^{\boldsymbol{\beta}} := \mathbf{e}_j \otimes \boldsymbol{\beta}_h - \mathbf{e}_h \otimes \boldsymbol{\beta}_j$, $\nu_{jh}^{(i)} := \lambda_{ij}^{\mathbf{V}}\lambda_{ih}^{\mathbf{V}}/(\lambda_{ij}^{\mathbf{V}} - \lambda_{ih}^{\mathbf{V}})^2$, and $\boldsymbol{\nu}^{(i)} := \mathrm{diag}(\nu_{12}^{(i)}, \nu_{13}^{(i)}, \ldots, \nu_{(k-1)k}^{(i)})$. We then have the following ULAN result.

Proposition A.1 *(ULAN) Let Assumptions (A1) (with* $\mathrm{f} = (f_1, \ldots, f_m) \in (\mathcal{F}_a)^m$*), (A2) and (A3) hold. Then, the family* $\mathcal{P}_{\mathrm{f}}^{(n)} := \left\{ \mathrm{P}_{\boldsymbol{\vartheta};\mathrm{f}}^{(n)} \,|\, \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \right\}$ *is ULAN, with central sequence*

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}} = \boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{(n)} := \left(\boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{I(n)\prime},\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{II(n)\prime},\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{III(n)\prime},\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{IV(n)\prime} \right)',$$

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{I} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\vartheta};f_1}^{I,1} \\ \vdots \\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta};f_m}^{I,m} \end{pmatrix},\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{II} = \begin{pmatrix} \Delta_{\boldsymbol{\vartheta};f_1}^{II,1} \\ \vdots \\ \Delta_{\boldsymbol{\vartheta};f_m}^{II,m} \end{pmatrix},\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{III} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\vartheta};f_1}^{III,1} \\ \vdots \\ \boldsymbol{\Delta}_{\boldsymbol{\vartheta};f_m}^{III,m} \end{pmatrix},$$

where (with $d_{ij} = d_{ij}(\boldsymbol{\theta}_i, \mathbf{V}_i)$ *and* $\mathbf{U}_{ij} = \mathbf{U}_{ij}(\boldsymbol{\theta}_i, \mathbf{V}_i)$*)*

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta};f_i}^{I,i} := \frac{1}{\sqrt{n_i}\sigma_i} \sum_{j=1}^{n_i} \varphi_{f_i}\left(\frac{d_{ij}}{\sigma_i}\right) \mathbf{V}_i^{-1/2}\mathbf{U}_{ij}, \qquad \Delta_{\boldsymbol{\vartheta};f_i}^{II,i} := \frac{1}{2\sqrt{n_i}\sigma_i^2} \sum_{j=1}^{n_i} \left(\varphi_{f_i}\left(\frac{d_{ij}}{\sigma_i}\right) \frac{d_{ij}}{\sigma_i} - k \right),$$

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta};f_i}^{III,i} := \frac{1}{2\sqrt{n_i}} \mathbf{M}_k^{\boldsymbol{\Lambda}_i^{\mathbf{V}}} \mathbf{H}_k \left((\boldsymbol{\Lambda}_i^{\mathbf{V}})^{-1/2} \boldsymbol{\beta}' \right)^{\otimes 2} \sum_{j=1}^{n_i} \varphi_{f_i}\left(\frac{d_{ij}}{\sigma_i}\right) \frac{d_{ij}}{\sigma_i} \operatorname{vec}\left(\mathbf{U}_{ij}\mathbf{U}_{ij}' \right),$$

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{f}}^{IV} := \frac{1}{2n^{1/2}} \sum_{i=1}^{m} \mathbf{G}_k^{\boldsymbol{\beta}} \mathbf{L}_k^{\boldsymbol{\beta},\boldsymbol{\Lambda}_i^{\mathbf{V}}} \left(\mathbf{V}_i^{\otimes 2} \right)^{-1/2} \sum_{j=1}^{n_i} \varphi_{f_i}\left(\frac{d_{ij}}{\sigma_i}\right) \frac{d_{ij}}{\sigma_i} \operatorname{vec}\left(\mathbf{U}_{ij}\mathbf{U}_{ij}' \right),$$

$i = 1, \ldots, m$*, and with block-diagonal information matrix*

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}} := \operatorname{diag}(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{I}, \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{II}, \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{III}, \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{IV}), \tag{A.1}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{I} = \operatorname{diag}(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};f_1}^{I,1}, \ldots, \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};f_m}^{I,m})$, $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{II} = \operatorname{diag}(\Gamma_{\boldsymbol{\vartheta};f_1}^{II,1}, \ldots, \Gamma_{\boldsymbol{\vartheta};f_m}^{II,m})$, $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{III} = \operatorname{diag}(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};f_1}^{III,1}, \ldots, \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};f_m}^{III,m})$*, with*

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};f_i}^{I,i} := \frac{\mathcal{I}_k(f_i)}{k\sigma_i^2} \mathbf{V}_i^{-1}, \quad \Gamma_{\boldsymbol{\vartheta};f_i}^{II,i} := \frac{\mathcal{J}_k(f_i) - k^2}{4\sigma_i^4},$$

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};f_i}^{III,i} := \frac{\mathcal{J}_k(f_i)}{4k(k+2)} \mathbf{M}_k^{\boldsymbol{\Lambda}_i^{\mathbf{V}}} \mathbf{H}_k((\boldsymbol{\Lambda}_i^{\mathbf{V}})^{-1})^{\otimes 2} \left[\mathbf{I}_{k^2} + \mathbf{K}_k \right] \mathbf{H}_k' (\mathbf{M}_k^{\boldsymbol{\Lambda}_i^{\mathbf{V}}})',$$

and

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}}^{IV} = \frac{1}{4k(k+2)} \mathbf{G}_k^{\boldsymbol{\beta}} \left(\sum_{i=1}^{m} r_i \mathcal{J}_k(f_i) (\boldsymbol{\nu}^{(i)})^{-1} \right) (\mathbf{G}_k^{\boldsymbol{\beta}})'.$$

More precisely, for any $\boldsymbol{\vartheta}^{(n)} = \boldsymbol{\vartheta} + O(n^{-1/2}) \in \boldsymbol{\Theta}$ *and any bounded sequence* $\boldsymbol{\tau}^{(n)}$ *such that* $\boldsymbol{\vartheta}^{(n)} + n^{-1/2}\boldsymbol{\varsigma}^{(n)}\boldsymbol{\tau}^{(n)} \in \boldsymbol{\Theta}$*, we have, under* $\mathrm{P}_{\boldsymbol{\vartheta}^{(n)};\mathrm{f}}^{(n)}$*,*

$$\begin{aligned} \Lambda_{\boldsymbol{\vartheta}^{(n)}+n^{-1/2}\boldsymbol{\varsigma}^{(n)}\boldsymbol{\tau}^{(n)}/\boldsymbol{\vartheta}^{(n)};\mathrm{f}}^{(n)} &:= \log\left(d\mathrm{P}_{\boldsymbol{\vartheta}^{(n)}+n^{-1/2}\boldsymbol{\varsigma}^{(n)}\boldsymbol{\tau}^{(n)};\mathrm{f}}^{(n)} / d\mathrm{P}_{\boldsymbol{\vartheta}^{(n)};\mathrm{f}}^{(n)} \right) \\ &= (\boldsymbol{\tau}^{(n)})' \boldsymbol{\Delta}_{\boldsymbol{\vartheta}^{(n)};\mathrm{f}}^{(n)} - \frac{1}{2} (\boldsymbol{\tau}^{(n)})' \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}} \boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1) \end{aligned}$$

and $\boldsymbol{\Delta}_{\boldsymbol{\vartheta}^{(n)};\mathrm{f}} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{f}})$*, as* $n \to \infty$*.*

Although this ULAN result is distinct from the one in Hallin *et al.* (2013) (where perturbations of the CPC hypothesis are considered), its proof follows along the same lines, and is therefore omitted.

A.2 Asymptotic properties of $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}}$ and $\underset{\sim}{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\vartheta}};\mathrm{K}}$

The following proposition provides (i) asymptotic representation, (ii) asymptotic normality, and (iii) asymptotic linearity results for $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}}$.

Proposition A.2 *Let Assumptions (A1)-(A4) hold and let* $\hat{\boldsymbol{\vartheta}}$ *satisfy Assumption (A5). Fix* $\mathrm{g} \in (\mathcal{F}_1)^m$. *Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$, *as* $n \to \infty$,

(i) $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} = \boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{K};\mathrm{g}} + o_{L^2}(1)$, *where (recall that* $\tilde{G}_i$ *stands for the cumulative distribution function under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$ *of* d_{ij}*; see Section 2.1)*

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{K};\mathrm{g}} := \frac{1}{2n^{1/2}} \sum_{i=1}^{m} \mathbf{G}_k^{\boldsymbol{\beta}} \mathbf{L}_k^{\boldsymbol{\beta},\boldsymbol{\Lambda}_i^{\mathbf{V}}} \left(\mathbf{V}_i^{\otimes 2}\right)^{-1/2} \sum_{j=1}^{n_i} K_i\big(\tilde{G}_i(d_{ij})\big) \mathrm{vec}\,(\mathbf{U}_{ij}\mathbf{U}'_{ij});$$

(ii) $\boldsymbol{\Delta}_{\boldsymbol{\vartheta};\mathrm{K};\mathrm{g}}$ *is asymptotically normal with mean zero and covariance matrix*

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K}} := \frac{1}{4k(k+2)} \mathbf{G}_k^{\boldsymbol{\beta}} \left(\sum_{i=1}^{m} \mathcal{J}_k(K_i)(\boldsymbol{\nu}^{(i)})^{-1} \right) (\mathbf{G}_k^{\boldsymbol{\beta}})';$$

(iii) $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}}$ *is locally and asymptotically linear in the sense that*

$$\underset{\sim}{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\vartheta}};\mathrm{K}} - \underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\vartheta};\mathrm{K}} = -\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}}\, n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_{\mathrm{P}}(1),$$

(see (3.4) for a definition of $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\mathrm{K},\mathrm{g}}$*); this latter result requires* $\mathrm{g} \in (\mathcal{F}_a)^m$.

Proof. Part (i) of the result follows from more or less standard application of Hájek's classical projection theorem, Part (ii) from the multivariate central limit theorem. We thus focus on Part (iii). Let $\mathbf{J}_k^{\perp} := \mathbf{I}_{k^2} - k^{-2}(\mathrm{vec}\,\mathbf{I}_k)(\mathrm{vec}\,\mathbf{I}_k)'$ and

$$\underset{\sim}{\mathbf{S}}^{(n)}_{\boldsymbol{\vartheta};K_i} := n_i^{-1} \sum_{j=1}^{n_i} K_i\left(\frac{R_{ij}^{(n)}(\boldsymbol{\theta}_i, \mathbf{V}_i)}{n_i+1}\right) \mathbf{U}_{ij}(\boldsymbol{\theta}_i, \mathbf{V}_i)\mathbf{U}'_{ij}(\boldsymbol{\theta}_i, \mathbf{V}_i).$$

Lemma A.1 in Hallin *et al.* (2006) and Lemma 4.4 in Kreiss (1987) entail that

$$\mathbf{J}_k^{\perp} n_i{}^{1/2} \,\mathrm{vec}\,(\underset{\sim}{\mathbf{S}}{}^{(n)}_{\hat{\boldsymbol{\vartheta}};K_i} - \underset{\sim}{\mathbf{S}}{}^{(n)}_{\boldsymbol{\vartheta};K_i})$$

$$+\frac{\mathcal{J}_k(K_i, g_i)}{4k(k+2)}\Big[\mathbf{I}_{k^2} + \mathbf{K}_k - \frac{2}{k}\mathbf{J}_k\Big](\mathbf{V}_i{}^{-1/2})^{\otimes 2} n_i^{1/2}\mathrm{vec}\,(\hat{\mathbf{V}}_i - \mathbf{V}_i) = o_{\mathrm{P}}(1) \qquad \text{(A.2)}$$

as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$. This and the fact that $\mathbf{L}_k^{\boldsymbol{\beta},\boldsymbol{\Lambda}_i^{\mathbf{V}}}(\mathbf{V}_i^{-1/2})^{\otimes 2}\mathbf{J}_k = \mathbf{0}$ directly imply that, still under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$,

$$\underset{\sim}{\boldsymbol{\Delta}}{}^{IV}_{\hat{\boldsymbol{\vartheta}};\mathrm{K}} - \underset{\sim}{\boldsymbol{\Delta}}{}^{IV}_{\boldsymbol{\vartheta};\mathrm{K}} = \sum_{i=1}^{m} r_i \frac{\mathcal{J}_k(K_i, g_i)}{4k(k+2)} \mathbf{G}_k^{\boldsymbol{\beta}} \mathbf{L}_k^{\boldsymbol{\beta},\boldsymbol{\Lambda}_i^{\mathbf{V}}} \left(\mathbf{V}_i^{\otimes 2}\right)^{-1} \Big[\mathbf{I}_{k^2} + \mathbf{K}_k\Big] n_i^{1/2}\mathrm{vec}\,(\hat{\mathbf{V}}_i - \mathbf{V}_i) + o_{\mathrm{P}}(1). \qquad \text{(A.3)}$$

Following the same argument as in the proof of Lemma 4.2 in Hallin *et al.* (2010b), we obtain that

$$n_i^{1/2}\mathrm{vec}\,(\hat{\mathbf{V}}_i - \mathbf{V}_i) = (\mathbf{L}_k^{\boldsymbol{\beta},\boldsymbol{\Lambda}_i^{\mathbf{V}}})'(\mathbf{G}_k^{\boldsymbol{\beta}})' n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \boldsymbol{\beta}^{\otimes 2}\mathbf{H}_k' n_i^{1/2}\mathrm{dvec}(\hat{\boldsymbol{\Lambda}}_i^{\mathbf{V}} - \boldsymbol{\Lambda}_i^{\mathbf{V}}) + o_{\mathrm{P}}(1) \quad \text{(A.4)}$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$. The result then follows by plugging (A.4) into (A.3), taking into account the fact that $(\mathbf{L}_k^{\boldsymbol{\beta},\boldsymbol{\Lambda}_i^{\mathbf{V}}})' \left(\mathbf{V}_i^{\otimes 2}\right)^{-1} \big[\mathbf{I}_{k^2} + \mathbf{K}_k\big]\boldsymbol{\beta}^{\otimes 2}\mathbf{H}_k' = \mathbf{0}$. □

A.3 Two lemmas

This appendix states and proves two lemmas used in Section 3.2.

Lemma A.1 *Let $\hat{\boldsymbol{\beta}}$ (with values in $\mathcal{S}O_k$) be any estimator of $\boldsymbol{\beta} \in \mathcal{S}O_k$ such that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}$, say, as $n \to \infty$. Then, denoting by $\mathrm{proj}(\mathbf{A}) := \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ the projection onto the column space of $\mathbf{A}$,*

$$\Big[\mathbf{I}_{k^2} - \mathrm{proj}(\mathbf{G}_k^{\boldsymbol{\beta}})\Big] n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \Big[\mathbf{I}_{k^2} - \frac{1}{2}\mathbf{G}_k^{\boldsymbol{\beta}}\mathbf{G}_k^{\boldsymbol{\beta}\prime}\Big] n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_{\mathrm{P}}(1),$$

under $\mathrm{P}^{(n)}$ as $n \to \infty$.

Proof. Since $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$ are elements of $\mathcal{S}O_k$, it is trivial that

$$n^{1/2}\boldsymbol{\beta}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\beta} + n^{1/2}\boldsymbol{\beta}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\beta} = \mathbf{0}.$$

Root-n consistency of $\hat{\boldsymbol{\beta}}$ yields $n^{1/2}\boldsymbol{\beta}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\beta} = o_{\mathrm{P}}(1)$; since $n^{1/2}\boldsymbol{\beta}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\beta} = \mathbf{0}$ implies that $n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \in \mathcal{M}(\mathbf{G}_k^{\boldsymbol{\beta}}(\mathbf{G}_k^{\boldsymbol{\beta}})')$, we deduce that

$$\left[\mathbf{I}_{k^2} - \mathrm{proj}(\mathbf{G}_k^{\boldsymbol{\beta}}(\mathbf{G}_k^{\boldsymbol{\beta}})')\right] n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_{\mathrm{P}}(1).$$

Now, using the fact that $(\mathbf{G}_k^{\boldsymbol{\beta}})'\mathbf{G}_k^{\boldsymbol{\beta}} = 2\mathbf{I}_{k(k-1)/2}$, the result follows easily from the standard properties of Moore-Penrose inverses. □

Lemma A.2 *Let Assumptions (A1)-(A4) hold and let $\hat{\boldsymbol{\vartheta}}$ satisfy Assumption (A5). Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\mathrm{g}}$ as $n \to \infty$,*

$$n^{1/2}\mathrm{vec}(\underset{\sim}{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})} - \boldsymbol{\beta}) = \mathbf{J}_k^{\boldsymbol{\beta}} n^{1/2}\mathrm{vec}(\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K,g})} - \boldsymbol{\beta}) + o_{\mathrm{P}}(1), \tag{A.5}$$

where $\mathbf{J}_k^{\boldsymbol{\beta}}$ is a $k^2 \times k^2$ matrix such that $\mathbf{J}_k^{\boldsymbol{\beta}}\mathbf{G}_k^{\boldsymbol{\beta}} = \mathbf{G}_k^{\boldsymbol{\beta}}$.

Proof. The mapping from $\hat{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K})}$ to $\tilde{\boldsymbol{\beta}}_{\mathrm{K};\widehat{\mathcal{J}}_k(\mathrm{K})}$ is continuously differentiable. Denoting by $\mathbf{J}_k^{\boldsymbol{\beta}}$ its Jacobian matrix at $\mathrm{vec}(\boldsymbol{\beta})$, the result follows from an application of the Delta method. Now, it is easily shown that

$$\mathbf{J}_k^{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{I}_k - \boldsymbol{\beta}_1\boldsymbol{\beta}_1' & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \boldsymbol{\beta}_1\boldsymbol{\beta}_2' & \mathbf{I}_k - \boldsymbol{\beta}_1\boldsymbol{\beta}_1' - \boldsymbol{\beta}_2\boldsymbol{\beta}_2' & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \boldsymbol{\beta}_1\boldsymbol{\beta}_2' & \boldsymbol{\beta}_1\boldsymbol{\beta}_3' & \mathbf{I}_k - \boldsymbol{\beta}_1\boldsymbol{\beta}_1' - \boldsymbol{\beta}_2\boldsymbol{\beta}_2' - \boldsymbol{\beta}_3\boldsymbol{\beta}_3' & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \ddots & \vdots \\ \boldsymbol{\beta}_1\boldsymbol{\beta}_2' & \boldsymbol{\beta}_1\boldsymbol{\beta}_3' & \cdots & \cdots & \boldsymbol{\beta}_1\boldsymbol{\beta}_{k-1}' & \mathbf{0} \end{pmatrix}.$$

The identity $\mathbf{J}_k^{\boldsymbol{\beta}}\mathbf{G}_k^{\boldsymbol{\beta}} = \mathbf{G}_k^{\boldsymbol{\beta}}$ then follows from elementary algebra. □

References

[1] Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, New York: Wiley.

[2] Boente, G. and Orellana, L. (2001). A robust approach to common principal components. In: Fernholz, L.T., Morgenthaler, S., and Stahel, W. (Eds.): *Statistics in Genetics and in the Environmental Sciences*. Birkhäuser Verlag AG, Basel, Switzerland, 117–145.

[3] Boente, G., Pires, A.M. and Rodrigues, I.M. (2002). Influence functions and outlier detection under the common principal components model: a robust approach. *Biometrika* **89**, 861–875.

[4] Boente G., Pires A. and Rodrigues I. (2006) General projection-pursuit estimators for the common principal components model: influence functions and Monte Carlo study. *Journal of Multivariate Analysis* **97**, 124-147.

[5] Cator, E.A. and Lopuhaä, H.P. (2010). Asymptotic expansion of the minimum covariance determinant estimators. *Journal of Multivariate Analysis* **101**, 2372–2388.

[6] Croux, C., and Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika* **87**, 603–618.

[7] Croux, C., and Ruiz-Gazen, A. (2005). High breakdown estimators for principal components: the projection-pursuit approach revisited. *Journal of Multivariate Analysis* **95**, 206–226.

[8] Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association* **76**, 354–362.

[9] Flury, B. (1984). Common principal components in k groups. *Journal of the American Statistical Association* **79**, 892–898.

[10] Flury, B. and Gautschi, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *SIAM Journal on Scientific and Statistical Computing* **7**, 169–184.

[11] Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics: a practical approach.* Chapman and Hall, New York.

[12] Hallin, M., Oja, H. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *Annals of Statistics* **34**, 2757–2789.

[13] Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Annals of Statistics* **34**, 2707–2756.

[14] Hallin, M., Paindaveine, D. and Verdebout, T. (2008). Pseudo-Gaussian inference in heterokurtic elliptical common principal components models. *Annales de l'ISUP* LII, 9–24.

[15] Hallin, M., Paindaveine, D. and Verdebout, T. (2010a). Testing for common principal components under heterokurticity. *Journal of Nonparametric Statistics* **22**, 879–895.

[16] Hallin, M., Paindaveine, D. and Verdebout, T. (2010b). Optimal rank-based testing for principal components. *Annals of Statistics* **38**, 3245–3299.

[17] Hallin, M., Paindaveine, D. and Verdebout, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli*, to appear.

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

[19] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24**, 417–441.

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

[21] Luo, R., Wang, H., and Tsai, C.L. (2009). Sufficient dimension reduction by contour projection. *Annals of Statistics* **37**, 3743–3778

[22] Paindaveine, D. (2006). A Chernoff-Savage result for shape. On the non-admissibility of pseudo-Gaussian methods. *Journal of Multivariate Analysis* **97**, 2206–2220.

[23] Paindaveine, D. (2008). A canonical definition of shape. *Statistics and Probability Letters* **78**, 2240–2247.

[24] Pearson, K. (1901). On lines and planes of closest fit to system of points in space. *Philosophical Magazine* **2**, 559–572.

[25] Rousseuw, P.J., and Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

[26] Taskinen, S., Koch, I., and Oja, H. (2012). Robustifying principal component analysis with spatial sign vectors. *Statistics and Probability letters* **82**, 765–774.

[27] Tyler, D. (1987). A distribution-free M-estimate of multivariate scatter. *Annals of Statistics* **15**, 234–251.

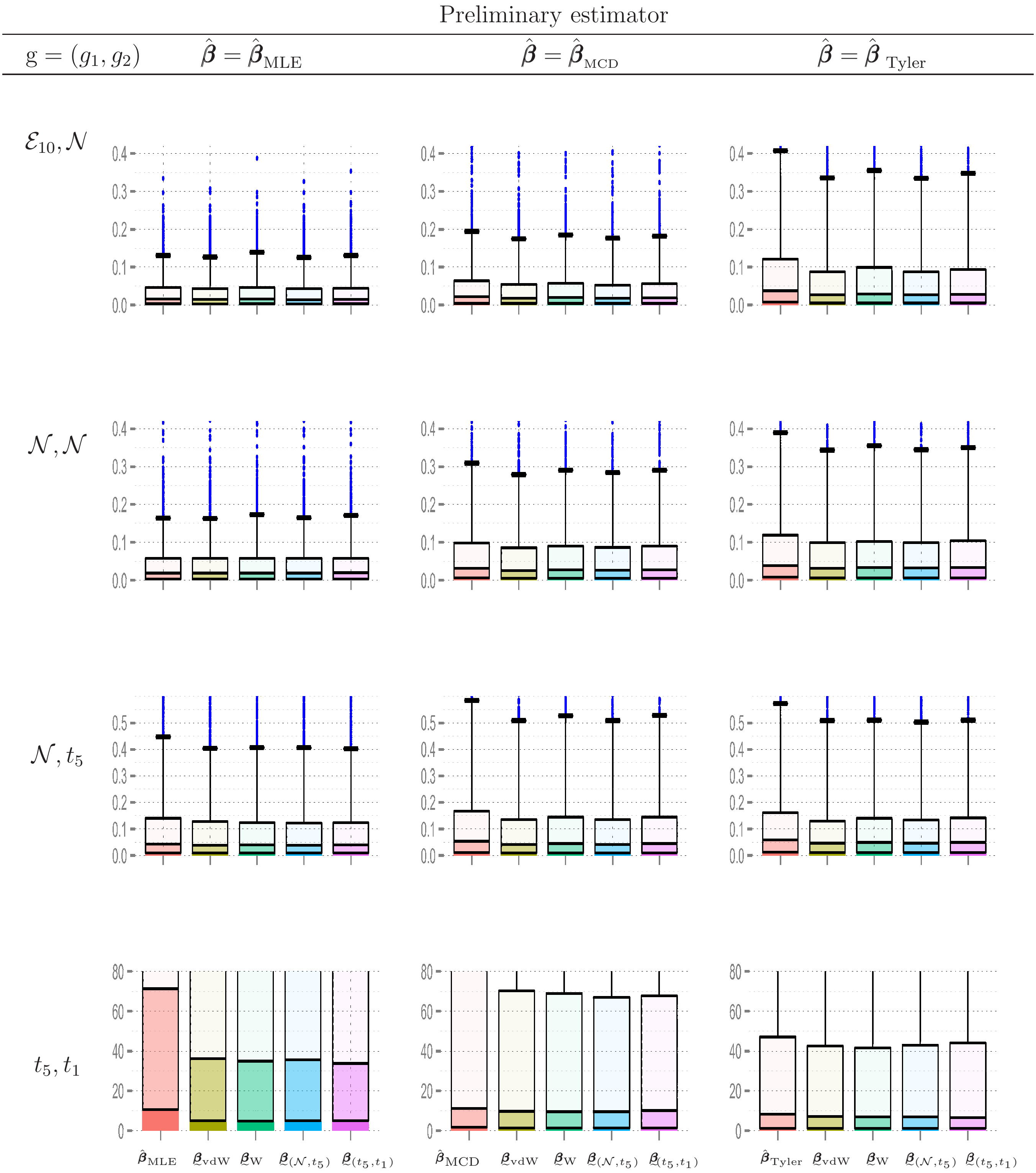

Figure 1: **Finite-sample performance of R-estimators for CPC.** One-sided boxplots of mean squared errors, under various couples of elliptical densities (power-exponential $\mathcal{E}_{10}$/Gaussian, Gaussian/Gaussian, Gaussian/t_5, t_5/t_1, in rows) and different preliminary estimators ($\hat{\boldsymbol{\beta}}_{\mathrm{MLE}}$, $\hat{\boldsymbol{\beta}}_{\mathrm{MCD}}$, $\hat{\boldsymbol{\beta}}_{\mathrm{Tyler}}$, in columns), of R-estimators of the first principal component based on the following scores: van der Waerden, Wilcoxon, van der Waerden in sample 1 and t_5 in sample 2, t_5 in sample 1 and t_1 in sample 2. Results are obtained from $N = 1,500$ replications of the bivariate two-sample "proportional" CPC model described in Section 5.1.

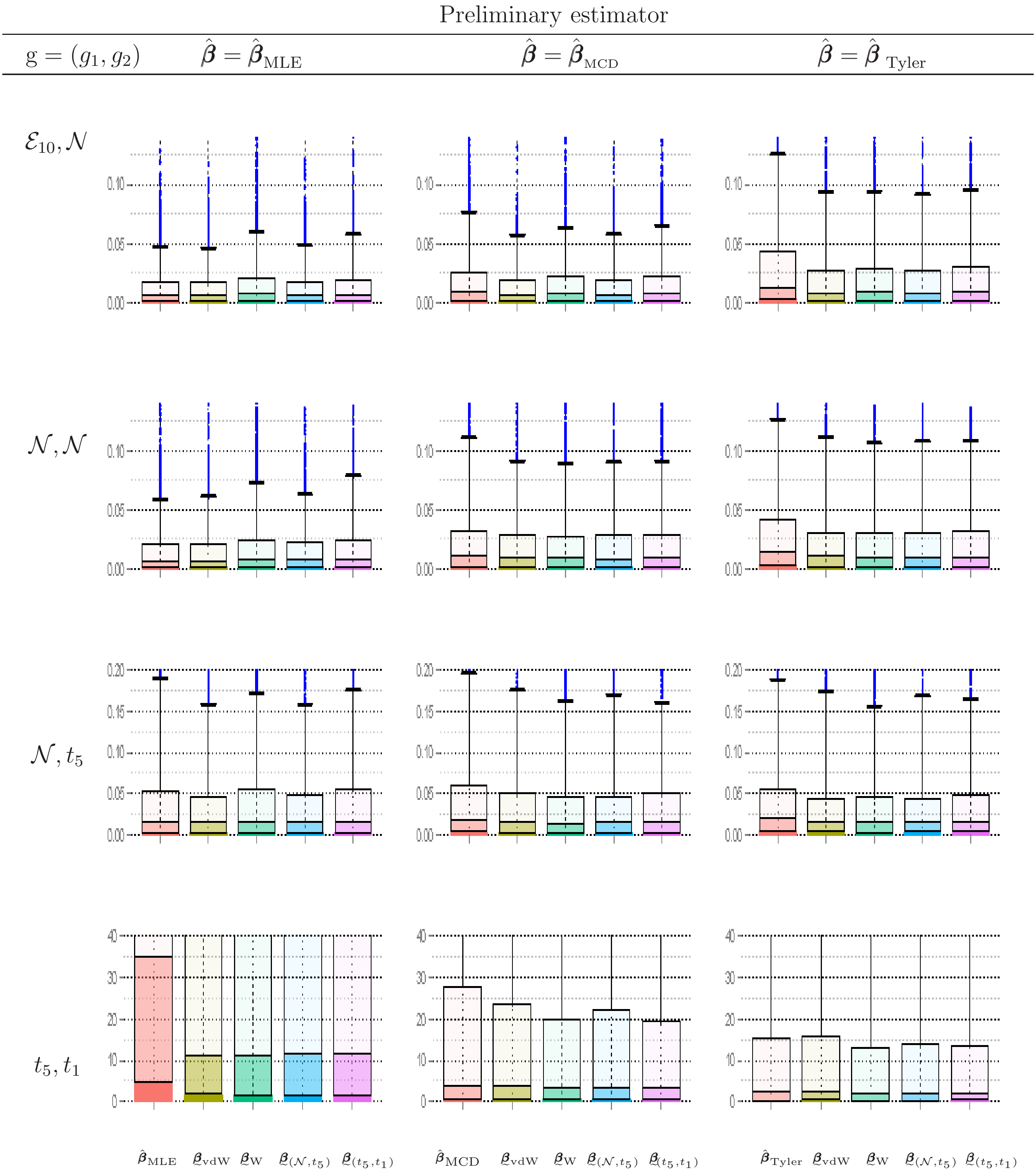

Figure 2: **Finite-sample performance of R-estimators for CPC.** One-sided boxplots of mean squared errors, under various couples of elliptical densities (power-exponential $\mathcal{E}_{10}$/Gaussian, Gaussian/Gaussian, Gaussian/t_5, t_5/t_1, in rows) and different preliminary estimators ($\hat{\boldsymbol{\beta}}_{\text{MLE}}$, $\hat{\boldsymbol{\beta}}_{\text{MCD}}$, $\hat{\boldsymbol{\beta}}_{\text{Tyler}}$, in columns), of R-estimators of the first principal component based on the following scores: van der Waerden, Wilcoxon, van der Waerden in sample 1 and t_5 in sample 2, t_5 in sample 1 and t_1 in sample 2. Results are obtained from $N = 1,500$ replications of the bivariate two-sample "non-proportional" CPC model described in Section 5.1.

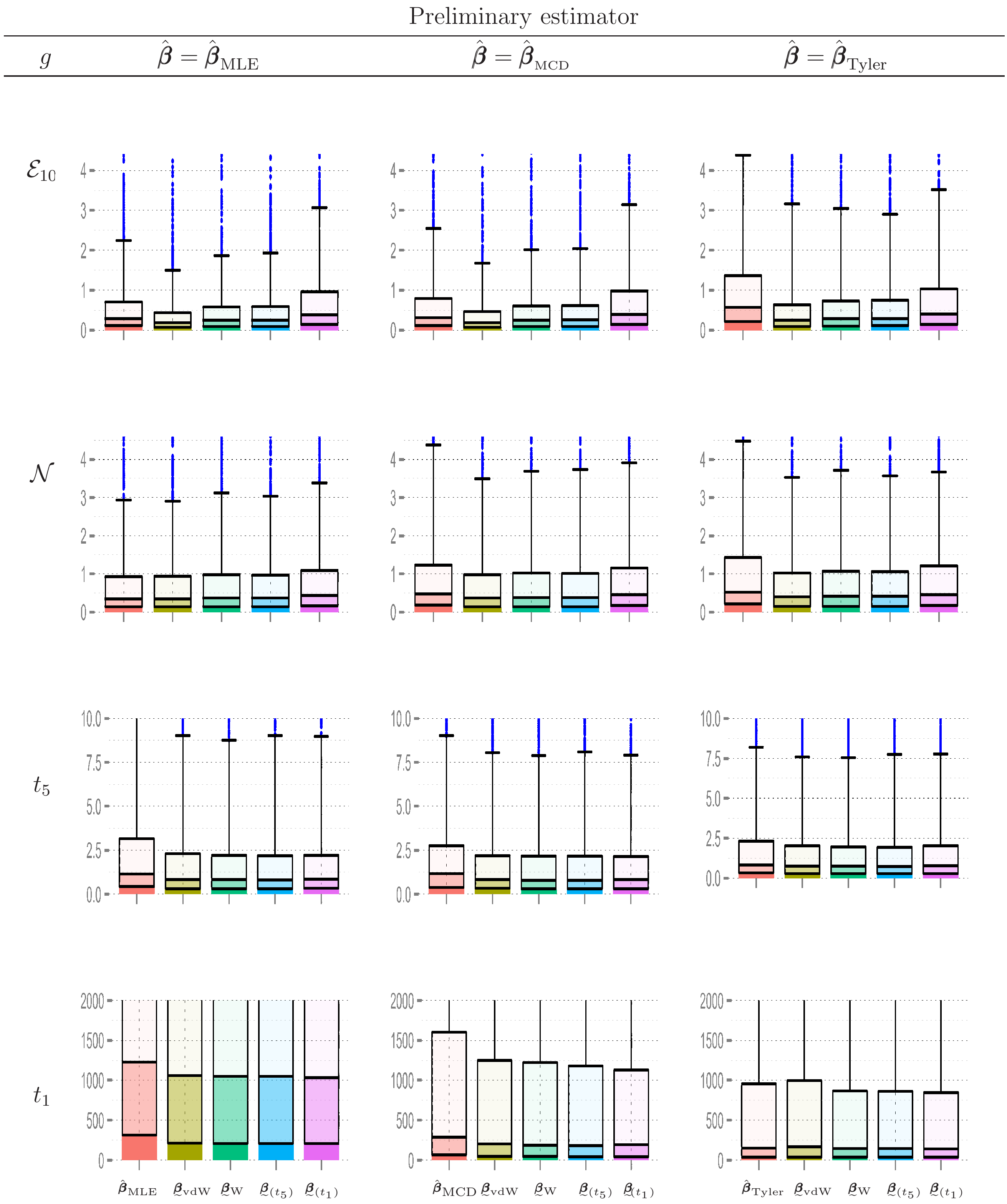

Figure 3: **Finite-sample performance of R-estimators for PCA.** One-sided boxplots of mean squared errors, under various elliptical densities (power-exponential $\mathcal{E}_{10}$, Gaussian, t_5, t_1, in rows) and different preliminary estimators ($\hat{\boldsymbol{\beta}}_{\text{MLE}}$, $\hat{\boldsymbol{\beta}}_{\text{MCD}}$, $\hat{\boldsymbol{\beta}}_{\text{Tyler}}$, in columns), of R-estimators of the first principal component based on the following scores: van der Waerden, Wilcoxon, van der Waerden, t_5 and t_1. Results are obtained from $N = 1,500$ replications of the 4-dimensional model described in Section 5.2.