Journal of Nonparametric Statistics
Vol. 00, No. 00, Month 200x, 1–21

RESEARCH ARTICLE

Testing for Common Principal Components under Heterokurticity

Marc Hallin[a,b], Davy Paindaveine[a,c] and Thomas Verdebout[d]

[a]*Institut de Recherche en Statistique, E.C.A.R.E.S., Université Libre de Bruxelles, Belgium.* [b]*Académie Royale de Belgique.* [c]*Département de Mathématique, Université Libre de Bruxelles, Belgium.* [d]*EQUIPPE-GREMARS, Université Lille Nord de France, France.*
(*Received 00 Month 200x; in final form 00 Month 200x*)

The so-called Common Principal Components (CPC) model, in which the covariance matrices $\mathbf{\Sigma}_i$ of m populations are assumed to have identical eigenvectors, was introduced by Flury (1984). Gaussian parametric inference methods (Gaussian maximum likelihood estimation; Gaussian likelihood ratio testing) have been fully developed for this model, but their validity does not extend beyond the case of elliptical densities with common Gaussian kurtosis. A non-Gaussian (but still homokurtic) extension of Flury's Gaussian likelihood ratio test (LRT) for the hypothesis of CPC (Flury (1984)) is proposed in Boik (2002); see also Boente et al. (2001, 2009) for robust versions. In this paper, we show how Flury's LRT can be modified into a pseudo-Gaussian test which remains valid under arbitrary, hence possibly heterokurtic, elliptical densities with finite fourth-order moments, while retaining its optimality features at the Gaussian.

Keywords: common principal components, elliptical distributions, pseudo-Gaussian inference, robust inference

1. Introduction.

1.1. *The Common Principal Components model.*

One of the most widely used tools of classical multivariate analysis, Principal Components (PC) also ranks among the oldest, since the one-sample version of that popular technique goes back as far as Pearson (1901). Its multisample versions appeared much later. In 1984, Flury introduced the so-called *common principal components* (CPC) model, which since then has been used in a number of applications, mainly in a biometric context. Under such a model, m k-dimensional populations, with covariance matrices $\mathbf{\Sigma}_i$, $i = 1, \ldots, m$, are

ISSN: 1048-5252 print/ISSN 1029-0311 online
© 200x Taylor & Francis
DOI: 10.1080/1048525YYxxxxxxxx
http://www.informaworld.com

assumed to share the same PCs, with possibly different eigenvalues: the $\mathbf{\Sigma}_i$'s thus are such that $\boldsymbol{\beta}'\mathbf{\Sigma}_i\boldsymbol{\beta} = \mathbf{\Lambda}_i$ for some m-tuple of diagonal matrices $\mathbf{\Lambda}_i$, $i = 1, \ldots, m$, and some unique orthogonal matrix $\boldsymbol{\beta}$. Those CPC models later on have been generalized (Flury 1988) into *partial* CPC models, in which only a subset of $q < k$ principal components are common to the m populations. More recently, a broader class of models, which includes CPC and partial CPC, but also common space models with simultaneous sphericity, has been considered by Boik (2002).

Inference for CPC models has been developed, still by Flury (1984, 1986, and 1988); Krzanowski (1979)'s general results on m-sample PC testing problems also apply. The procedures they are proposing, however, are of a Gaussian nature and do not resist violations of Gaussian assumptions—which is hardly surprising, as it is well known that Gaussian likelihood ratio tests (Gaussian LRTs) for structured covariance matrices are quite sensitive to such violations. This phenomenon, in the context of CPC, is largely confirmed by the simulations of Section 5.1. Unless such assumptions can be relaxed, CPC methods remain strongly Gaussian, which very severely restricts their applicability.

This lack of robustness of inference procedures for CPC recently received much interest in the literature, and several attempts were made in order to extend the validity of Flury's inference methods beyond the Gaussian case. In the sequel, we concentrate on the LRT statistic for the null hypothesis $\mathcal{H}_0$ of CPC.

Assuming that the m populations admit the same kurtosis κ, Boik (2002), applying a general result by Shapiro and Browne (1987), shows that dividing Flury's LRT statistic by a factor $(1+\hat{\kappa})$, where $\hat{\kappa}$, under $\mathcal{H}_0$, is a consistent estimator of κ, yields an adjusted LRT valid under any homokurtic m-tuple of elliptical densities. In a series of papers, Boente et al. (2001, 2002, and 2009) adopt a robustness approach, replacing, in Flury's LRT statistic, empirical covariance matrices with more general matrices of scatter. Dividing the resulting statistic by a factor $(1+\hat{\sigma}_1)$, where $\hat{\sigma}_1$ is a consistent estimator of a quantity σ_1 itself related with the efficiency of the off-diagonal elements of the scatter matrices considered, they obtain an adjusted LRT, which is valid under any m-tuple of densities with common σ_1 value. The homokurticity assumption thus is replaced with an assumption of homogenous σ_1's.

Our purpose here is to provide an adjusted LRT that remains valid in the absence of any homogeneity assumption. The test we are proposing (for the null hypothesis of CPC) is valid under arbitrary elliptical densities with finite fourth-order moments.

Deriving such a test requires a delicate analysis of the quadratic form underlying the LRT statistic. That analysis so far had been avoided: Flury indeed basically relies on the general result by Wilks (1938) on the asymptotic chi-square null distribution of $-2\log\Lambda$ under Gaussian assumptions, whereas Boik(2002) essentially checks that, under homokurticity, the necessary and sufficient condition of Shapiro and Browne (1987) is satisfied.

1.2. *Adjusted LRT statistics versus adjusted null asymptotic distributions.*

Gaussian likelihood ratio tests typically rely on the asymptotically chi-square distribution, under Gaussian assumptions, of the logarithm $-2\log\Lambda$ of the likelihood ratio statistic Λ—a classical asymptotic result of Wilks (1938). Beyond immediate intuition, this chi-square asymptotic behavior of likelihood ratio test statistics is strongly associated with the local asymptotic optimality, in the Le Cam sense and under Gaussian assumptions, of the corresponding tests—formalizing this would be too long; we refer to Chapter 11 of Le Cam (1986) for general results on locally asymptotically optimal tests, or to Hallin, Paindaveine and Verdebout (2008), where a general method is proposed for the construction of adjusted LRT statistics.

It is a well-known fact that LRT statistics, under non-Gaussian densities, in general are no longer asymptotically chi-square, but behave as a weighted sum of independent chi-square variables with one degree of freedom: see Boik (2002) for the CPC problem, Yanagihara et al. (2005) for other problems involving structured covariance matrices, or Hallin and Paindaveine (2009) for the problem of testing equality of m covariance matrices. In problems involving covariance matrices, the coefficients in that weighted sum are related to the kurtosis coefficients κ_i, $i = 1, \ldots, m$, of the underlying populations. Gaussian densities (with $\kappa_i = 0$) yield equal weights, hence global chi-square distributions. Those κ_i's can be consistently estimated; based on these estimations, weighted sums of chi-square variables can be simulated in order to obtain approximate critical values. Bootstrap estimates of the same critical values also can be produced (see, e.g., Zhang and Boos 1992, Goodnight and Schwartz 1997, or Zhu et al. 2002). Other authors (Yuan and Bentler 1997; Boik 2002) recommend approximating the distribution of the weighted sum of chi squares by modifying the number of degrees of freedom of the Gaussian chi-square limit distribution. In all these approaches, the LRT statistic remains unmodified while its asymptotic distribution is adjusted. This is highly unsatisfactory from a decision-theoretic point of view. The weights in the adjusted asymptotic distribution indeed have the undesirable effect of weighting the directions in the alternative in a manner that does not follow from any sound optimality property or decision-theoretic principle, but rather depends, in a totally uncontrolled way, on the unknown underlying population kurtoses. Therefore, we feel that the resulting tests, although asymptotically valid, are intrinsically flawed, and should be abandoned in favor of a modification of the LRT statistic itself, preserving its asymptotically chi-square behavior.

This issue of preserving the asymptotic behavior of LRT statistics has a long history, and is not limited to problems involving structured covariance matrices: see also Muirhead and Waternaux (1980), Browne (1984), Satorra and Bentler (1988), Hu et al. (1992), Bentler and Dudgeon (1996), to quote only a few.

1.3. *Outline of the paper.*

The paper is organized as follows. Section 2 collects the assumptions needed throughout. In Section 3, we introduce a quadratic asymptotic representation $Q^{(n)}$ of the Gaussian LRT statistic $-2 \log \Lambda$, and formally derive its asymptotic null distribution. In Section 4, we show how the same test statistic $Q^{(n)}$ can be turned into a pseudo-Gaussian one $Q^{(n)}_{\dagger}$, resisting both non-normality and heterokurticity. Section 5 is devoted to an empirical illustration of the method and some simulation results. Finally, the appendix collects the proofs of technical results.

2. Notation and main assumptions.

2.1. *Gaussian likelihood inference for CPC models.*

Denote by $(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})$, $i = 1, \ldots, m$ a collection of m mutually independent samples of i.i.d. k-dimensional Gaussian random vectors with location parameters $\boldsymbol{\theta}_i$ and positive definite covariance matrices $\mathbf{\Sigma}_i$, $i = 1, \ldots, m$. The null hypothesis $\mathcal{H}_0$ of common principal components takes the form

$$\mathcal{H}_0 : \mathbf{\Sigma}_i \boldsymbol{\beta}_j = \lambda_{ij} \boldsymbol{\beta}_j \quad i = 1, \ldots, m, \ j = 1, \ldots, k, \tag{2.1}$$

for some $k \times k$ orthogonal matrix $\boldsymbol{\beta} := (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)$ of common eigenvectors and some mk-tuple $\lambda_{11}, \ldots, \lambda_{mk}$ of eigenvalues. Some notational care, and some additional assumptions, however, are required here, in order for $\boldsymbol{\beta}$ to be well-defined and fully identifiable.

One way to deal with that identifiability problem consists in imposing, as in Boente et al. (2009), the following assumption.

Assumption (H′). The covariance matrix $\mathbf{\Sigma}_1$ has eigenvalues $\lambda_{11} > \lambda_{12} > \ldots > \lambda_{1k} > 0$.

Under Assumption (H′) and the null hypothesis (2.1), the well-defined (up to irrelevant sign changes, as usual) eigenvectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ of $\mathbf{\Sigma}_1$ are common to all $\mathbf{\Sigma}_i$'s; $\boldsymbol{\beta}_j$ in $\mathbf{\Sigma}_i$ is associated with eigenvalue λ_{ij}, but, for given $i > 1$, the λ_{ij}'s are not necessarily ordered from largest to smallest, and are not necessarily mutually distinct.

By requiring one at least of the m populations (arbitrarily labeled $i = 1$) to have m distinct eigenvalues, however, Assumption (H′) is unnecessarily restrictive. A sufficient condition for the unordered collection of common principal directions $\{\boldsymbol{\beta}_j, \, j = 1, \ldots, k\}$ to be properly identified is

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

Under Assumption (H), multiple eigenvalues, hence partial sphericities, are allowed provided that the intersections over the m populations of the corresponding spaces have at most dimension one. The matrix $\boldsymbol{\beta} := (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)$ of common eigenvectors then is identified up to the ordering of its columns (still forgetting about irrelevant sign changes of the $\boldsymbol{\beta}_j$'s). Fixing that ordering in some way thus completely identifies $\boldsymbol{\beta}$—hence also the corresponding λ_{ij}'s. Contrary to Assumption (H), the choice of a particular ordering criterion has the nature of an identification constraint, and therefore is largely arbitrary.

For instance, one can require (a lexicographical ordering) that $\lambda_{11} \geq \lambda_{12} \geq \ldots \geq \lambda_{1k} (> 0)$, and that, for any sequence of the form $\lambda_{1j} = \lambda_{1,j+1} = \ldots = \lambda_{1,j+\ell}$, one has $\lambda_{2j} \geq \lambda_{2,j+1} \geq \ldots \geq \lambda_{2,j+\ell}$. Recursively, if new ties occur among those $\lambda_{2,j}$'s, the ranking is then based on the $\lambda_{3,j}$'s, etc. Clearly, Assumption (H) ensures that this defines a unique ordering of the common principal directions and corresponding eigenvalues. Throughout, Assumption (H) will be tacitly considered part of the null hypothesis $\mathcal{H}_0$, and we adopt the lexicographic ordering just described.

Flury (1984) mainly deals with

(i) the Gaussian maximum likelihood estimators (MLEs) $(\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_k) =: \hat{\boldsymbol{\beta}}$ and $\hat{\lambda}_{ij}$, $i = 1, \ldots, m, \ j = 1, \ldots, k$ of the common eigenvectors and the corresponding eigenvalues under $\mathcal{H}_0$, and
(ii) the Gaussian likelihood ratio test for the same $\mathcal{H}_0$.

Denoting the empirical means and covariance matrices by

$$\bar{\mathbf{X}}_i := \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{X}_{ij} \quad \text{and} \quad \mathbf{S}_i := \frac{1}{n_i-1}\sum_{j=1}^{n_i}(\mathbf{X}_{ij}-\bar{\mathbf{X}}_i)(\mathbf{X}_{ij}-\bar{\mathbf{X}}_i)',$$

respectively, $\hat{\boldsymbol{\beta}}$ and the $\hat{\lambda}_{ij}$'s are solutions of the likelihood equations

$$\boldsymbol{\beta}_j'\Big(\sum_{i=1}^{m} n_i \frac{\lambda_{ij}-\lambda_{il}}{\lambda_{ij}\lambda_{il}}\mathbf{S}_i\Big)\boldsymbol{\beta}_l = 0, \quad j \neq l = 1,\ldots,k, \tag{2.2}$$

$$\boldsymbol{\beta}_j'\mathbf{S}_i\boldsymbol{\beta}_j = \lambda_{ij}, \quad i=1,\ldots,m,\ j=1,\ldots,k, \qquad \boldsymbol{\beta}_j'\boldsymbol{\beta}_l = \delta_{jl}, \quad l,j=1,\ldots,k,$$

where δ_{jl} is the usual Kronecker symbol. An explicit solution of equations (2.2) does not exist, but an algorithm for solving them numerically has been proposed by Flury and Gautschi (1986).

In practice, ties among the $\hat{\lambda}_{ij}$'s a.s. never occur. The lexicographic ordering just described thus a.s. leads to labeling the solutions of equations (2.2) in such a way that $\hat{\lambda}_{11} > \ldots > \hat{\lambda}_{1k}$, which, in case of tied λ_{1j}'s may lead to inconsistent estimators. This is easily overcome, however, by duly discretizing the λ_{ij}'s prior to ranking them. For any real z, consider a sequence $(z)_{\#}^{(n)}$ of truncations involving a finite number of decimal digits (increasing with n) such that $\sup_{z\in\mathbb{R}}|(z)_{\#}^{(n)} - z| = o_{\mathrm{P}}(n^{-1/2})$. Denoting by $(\hat{\lambda}_{ij}$ the unique solution of equations (2.2) with truncated values $\lambda_{ij})_{\#}^{(n)}$ satisfying the lexicographic ordering constraint, consistency under $\mathcal{H}_0$ is restored. Such truncation, however, has little practical impact, as the number of digital units in the numerical resolution of equations (2.2) remains finite.

Letting $n := \sum_{i=1}^m n_i$ tend to infinity in such a way that $n_i/n \to r_i$ for some $r_i \in (0,1)$, $i=1,\ldots,m$ (see Assumption (B) below), Flury (1986) shows that the MLE $\hat{\boldsymbol{\beta}}$ is root-n consistent and asymptotically normally distributed under the CPC assumption and Gaussian densities as soon as the matrix $\boldsymbol{\beta}$ of common eigenvectors is well defined. These consistency and asymptotic normality results extend to the case of elliptical distributions with finite fourth-order moments (see Section 3 of Hallin et al. (2008)).

We stress that we mainly adopt Assumption (H) to avoid cases where the $\boldsymbol{\beta}_j$'s are defined up to orthogonal transformations only, which probably leads, in case of tied λ_{ij}'s, parallel to the Gaussian one-sample case, to a much more complicated asymptotic theory. Recall indeed that if the covariance matrix $\boldsymbol{\Sigma}$ in the one-sample setup factorizes into

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}\begin{pmatrix} l_1\mathbf{I}_{k_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & l_r\mathbf{I}_{k_r}\end{pmatrix}\boldsymbol{\beta}', \qquad k_1+\ldots+k_r = k, \quad l_1 > \ldots > l_r (>0),$$

where $\mathbf{I}_\ell$ denotes the ℓ-dimensional identity matrix, then the eigenvectors $\boldsymbol{\beta}_j$, $j=1,\ldots,k$ of $\boldsymbol{\Sigma}$ are not well defined (only the linear spaces generated by the eigenvectors related with identical eigenvalues are), and the diagonal blocks $\hat{\mathbf{B}}_{ss}$ (of dimension k_s), $s=1,\ldots,r$, of the Gaussian MLE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$

(i) converge at a rate faster than root-n (in the sense that, under $\boldsymbol{\beta} = \mathbf{I}_k$, $n^{1/2}(\hat{\mathbf{B}}_{ss}\hat{\mathbf{B}}'_{ss} - \mathbf{I}_{k_s})$ is $o_{\mathrm{P}}(1)$), and
(ii) have a limiting conditional Haar invariant distribution; see Anderson (1963).

The Gaussian likelihood ratio test (LRT) statistic for testing $\mathcal{H}_0$ against the alternative that the $\boldsymbol{\Sigma}_i$'s have at least one pair of distinct eigenvectors is

$$\Lambda = \prod_{i=1}^{m} \left(\frac{|\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}|}{|\mathrm{diag}(\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}})|} \right)^{n_i/2}, \tag{2.3}$$

where $\mathrm{diag}(\mathbf{A})$ denotes the diagonal matrix having the same diagonal as $\mathbf{A}$. The intuition behind (2.3) is clear: under $\mathcal{H}_0$, $\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}$ should be diagonal or nearly diagonal, so that $|\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}|$ and $|\mathrm{diag}(\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}})|$ should be approximately equal, hence Λ close to one. The LRT rejects $\mathcal{H}_0$ when Λ is "too small". It follows from Wilks (1938)'s classical result that, under $\mathcal{H}_0$, $-2\log\Lambda$ is asymptotically chi-square, with $(m-1)s$ degrees of freedom, where $s := k(k-1)/2$. The test rejecting $\mathcal{H}_0$ whenever $-2\log\Lambda$ exceeds the corresponding chi-square α-upper quantile is thus a LRT with asymptotic level α.

As mentioned before, this test unfortunately is highly sensitive to violations of Gaussian assumptions. Our objective is to show how an adequate modification of test statistic (2.3) yields a *pseudo-Gaussian* version of this LRT—namely, a test statistic which is asymptotically equivalent to the LRT statistic Λ under Gaussian assumptions, but, unlike Λ, remains asymptotically chi-square, with $(m-1)s$ degrees of freedom, under possibly non-Gaussian and heterokurtic elliptical densities with finite fourth-order moments.

2.2. *Main assumptions.*

We henceforth throughout assume that all populations are elliptically symmetric. More precisely, defining, for $q \geq 2$,

$$\mathcal{F}^q := \left\{ h : \mathbb{R}_0^+ \to \mathbb{R}^+ : \mu_{k+q-1;h} < \infty \right\} \quad \text{and} \quad \mathcal{F}_1^q := \left\{ h \in \mathcal{F}^q \ : \frac{\mu_{k+1;h}}{\mu_{k-1;h}} = k \right\},$$

where $\mu_{\ell;h} := \int_0^\infty r^\ell h(r)\,dr$, we require the following.

Assumption (A). The observations $\mathbf{X}_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$ are mutually independent; $\mathbf{X}_{ij}$, $j = 1, \ldots, n_i$ have probability density function

$$\mathbf{x} \mapsto c_{k,f_i} |\boldsymbol{\Sigma}_i|^{-1/2} f_i\Big(\big((\mathbf{x}-\boldsymbol{\theta}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\theta}_i) \big)^{1/2} \Big), \quad i = 1, \ldots, m, \tag{2.4}$$

for some k-dimensional vector $\boldsymbol{\theta}_i$ (*location*), some positive definite $(k \times k)$ covariance matrix $\boldsymbol{\Sigma}_i$, and some f_i in the class $\mathcal{F}_1^4$ of *standardized radial densities* with finite fourth-order moments; c_{k,f_i} is a normalizing constant.

Writing $\mathbf{M}^{1/2}$ for the symmetric root of the positive semi-definite symmetric matrix $\mathbf{M}$, define the *elliptical coordinates*

$$\mathbf{U}_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i) := \frac{\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)}{\|\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)\|} \quad \text{and} \quad d_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i) := \|\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)\|. \tag{2.5}$$

Under Assumption (A), the $\mathbf{U}_{ij}$'s, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$ are i.i.d., uniformly distributed over the unit sphere in $\mathbb{R}^k$, so that $\mathrm{E}[\mathbf{U}_{ij}\mathbf{U}'_{ij}] = (1/k)\mathbf{I}_k$; as for the *standardized elliptical distances* d_{ij}, they are mutually independent, with density $\tilde{f}_{ik}(r) := (\mu_{k-1;f_i})^{-1} r^{k-1} f_i(r)$ and distribution function $\tilde{F}_{ik}$, and independent of the $\mathbf{U}_{ij}$'s. The condition that $f_i \in \mathcal{F}^2$ is therefore equivalent to the finiteness of second-order moments, while $f_i \in \mathcal{F}_1^4$ implies the finiteness of fourth-order moments, and is standardized in such a way that $\mathrm{E}[d^2_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)] = k$—thus justifying the terminology *standardized radial density* for f_i. It follows that, under Assumption (A), $\boldsymbol{\Sigma}_i = \mathrm{Var}[\mathbf{X}_{ij}]$ is the covariance matrix in population i.

Special instances of such elliptical densities are the k-variate multinormal distribution, with radial densities $f_i(r) = \phi(r) := \exp(-r^2/2)$, the k-variate Student distributions, with radial densities (for $\nu > 4$ degrees of freedom) $f_i(r) := (1 + a_{k,\nu} r^2/\nu)^{-(k+\nu)/2}$, and the k-variate power-exponential distributions, with radial densities of the form $f_i(r) := \exp(-b_{k,\eta} r^{2\eta})$, $\eta \in \mathbb{R}_0^+$; the positive constants $a_{k,\nu}$ and $b_{k,\eta}$ are such that $f_i \in \mathcal{F}_1^4$.

Under Assumption (H), the matrix $\boldsymbol{\beta}$ of common (under $\mathcal{H}_0$) eigenvectors is (up to the usual sign changes) a well identified element of the orthogonal group $\mathcal{O}(k)$, and hence contains s functionally independent parameters. These parameters can be stored in a vector $\boldsymbol{\beta}^\star$ in such a way that

$$n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{G}_k n^{1/2}(\hat{\boldsymbol{\beta}}^\star - \boldsymbol{\beta}^\star) + o_{\mathrm{P}}(1), \tag{2.6}$$

with $\mathbf{G}_k = (\mathbf{G}_{k;12}\ \mathbf{G}_{k;13} \ldots \mathbf{G}_{k;(k-1)k})$ and, denoting by $\mathbf{e}_j$ the jth unit vector in the canonical basis of $\mathbb{R}^k$, $\mathbf{G}_{k;jh} := (\mathbf{e}_j \otimes \boldsymbol{\beta}_h - \mathbf{e}_h \otimes \boldsymbol{\beta}_j)/\sqrt{2}$ (see Flury 1986 for details). Note that this matrix $\mathbf{G}_k$ is composed of s orthonormal column vectors.

Denote by $\boldsymbol{\Lambda}_i$ the diagonalized form of $\boldsymbol{\Sigma}_i$ (under $\mathcal{H}_0$, thus, $\boldsymbol{\Lambda}_i = \boldsymbol{\beta}'\boldsymbol{\Sigma}_i\boldsymbol{\beta}$). Writing $\mathrm{dvec}(\mathbf{A})$ for the vector obtained by stacking the diagonal elements of the matrix $\mathbf{A}$, the vector

$$\boldsymbol{\vartheta} := (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_m, (\mathrm{dvec}\,(\boldsymbol{\Lambda}_1))', \ldots, (\mathrm{dvec}\,(\boldsymbol{\Lambda}_m))', \boldsymbol{\beta}^{\star\prime})'$$

remains unspecified under $\mathcal{H}_0$ and thus constitutes a nuisance parameter in our problem. For any $\underline{f} = (f_1, \ldots, f_m) \in (\mathcal{F}_1^4)^m$ (which implies that $f_1, \ldots, f_m$ have finite fourth-order moments), denote by $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$ the joint distribution of the observations under the parameter value $\boldsymbol{\vartheta}$ and the m-tuple $\underline{f}$ of radial densities, and write $\mathrm{E}_{\boldsymbol{\vartheta};\underline{f}}$ for the corresponding expectations.

Now, we are doing asymptotics, and hence actually consider sequences of statistical experiments, with triangular arrays of observations of the form

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

indexed by the total sample size $n (:= \sum_i n_i^{(n)}) \in \mathbb{N}$. Most asymptotic results below are valid under the simple assumption that $n_i^{(n)} \to \infty$ for all i, but it will be convenient to assume that the sample sizes n_i satisfy the following assumption.

Assumption (B). For all $i = 1, \ldots, m$, there exists $r_i \in (0, 1)$ such that $r_i^{(n)} := (n_i^{(n)}/n) \to r_i$ as $n \to \infty$.

Useless $^{(n)}$ supercripts however will be avoided in the sequel. The following notation will be used throughout. For any $\underline{f} = (f_1, \ldots, f_m) \in (\mathcal{F}_1^4)^m$, let

$$E_k(f_i) := \mathrm{E}_{\boldsymbol{\vartheta};\underline{f}}[d_{ij}^4(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)] = \int_0^1 (\tilde{F}_{ik}^{-1}(u))^4\, du$$

(where $\tilde{F}_{ik}$ stands for the distribution function associated with the density $\tilde{f}_{ik}$ of $d_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$). Under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, the parameter $\kappa_k(f_i) := (k(k+2))^{-1}E_k(f_i) - 1$ is the *kurtosis* of the ith elliptic population (see, e.g., page 54 of Anderson 2003); note that no population-specific standardization of this kurtosis measure is required since $\mathrm{E}_{\boldsymbol{\vartheta};\underline{f}}[d_{ij}^2(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)] = k$ for all i. For Gaussian densities, $E_k(\phi) = k(k+2)$ and $\kappa_k(\phi) = 0$.

3. The Gaussian test statistic.

Let us consider the traditional Gaussian case first. That is, let us assume, as in Flury (1984), that m mutually independent samples are observed, where the ith random sample $\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i}$ is drawn from a multinormal distribution with location parameter $\boldsymbol{\theta}_i$ and positive definite covariance matrix $\boldsymbol{\Sigma}_i$.

As mentioned above, the LRT proposed by Flury (1984) is based on the asymptotically chi-square distribution of $-2\log\Lambda$. The problem with that approach, which is entirely correct as long as Gaussian assumptions hold, is that it does not bring any insight into the behavior of $-2\log\Lambda$ under possibly non-Gaussian observations. As a preparation for the pseudo-Gaussian approach of Section 4, we therefore provide here a direct derivation of the same result, based on a careful study of an appropriate quadratic representation of $-2\log\Lambda$.

Letting $\mathbf{T}_i := n_i^{1/2}\hat{\boldsymbol{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Lambda}}_i^{-1/2}$ (here, $\hat{\boldsymbol{\Lambda}}_i$ and $\hat{\boldsymbol{\beta}}$ stand for the Gaussian MLEs obtained from (2.2)), consider the quadratic form

$$Q^{(n)} := \frac{1}{2}\sum_{i=1}^m \big(\mathrm{tr}[\mathbf{T}_i^2] - \mathrm{tr}[(\mathrm{diag}(\mathbf{T}_i))^2]\big) = \sum_{i=1}^m \big(\mathrm{ovec}\,\mathbf{T}_i\big)'\big(\mathrm{ovec}\,\mathbf{T}_i\big), \tag{3.1}$$

where $\mathrm{ovec}(\mathbf{A})$ stands for the vector stacking the s entries lying *strictly* above the diagonal of the $k \times k$ matrix $\mathbf{A}$. Under this latter form, $Q^{(n)}$ has an appealing intuitive interpretation as a test statistic for $\mathcal{H}_0$. Indeed, it is summing, over the m samples, the (half) sums of squared off-diagonal elements of the random matrices $\mathbf{T}_i$ which, under $\mathcal{H}_0$, clearly should be close to zero (since, under $\mathcal{H}_0$, $\boldsymbol{\Lambda}_i^{-1/2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_i\boldsymbol{\beta}\boldsymbol{\Lambda}_i^{-1/2} = \mathbf{I}_k$).

The following result shows that, under the null and *any m-tuple of radial densities with finite fourth-order moments*, $Q^{(n)}$ is asymptotically equivalent to $-2\log\Lambda$ (see the appendix for the proof).

Proposition 3.1: *Assume that Assumptions (A) and (B) hold. Then, under $\mathcal{H}_0$ and finite fourth-order moments, $Q^{(n)} = -2\log\Lambda + o_{\mathrm{P}}(1)$, as $n \to \infty$.*

This asymptotic equivalence implies that $Q^{(n)}$ and the Gaussian LRT statistic have the same asymptotic null distribution and asymptotically yield the same test.

Let $\boldsymbol{\nu}^{(i)} = \mathrm{diag}(\nu_{12}^{(i)}, \nu_{13}^{(i)}, \ldots, \nu_{(k-1)k}^{(i)})$, $i = 1, \ldots, m$, where

$$\nu_{jl}^{(i)} := \frac{\lambda_{ij}\lambda_{il}}{r_i^{(n)}(\lambda_{ij} - \lambda_{il})^2}, \quad j, l = 1, \ldots, k, \quad j < l, \tag{3.2}$$

and define $\boldsymbol{\nu} := (\sum_{i=1}^m (\boldsymbol{\nu}^{(i)})^{-1})^{-1}$ (with the notational convention that $\frac{1}{+\infty} = 0$ and $\frac{c}{0} = +\infty$ for any $c > 0$). Note that, under Assumption (H), the entries of $\boldsymbol{\nu}$ are finite. Write $\hat{\boldsymbol{\nu}}^{(i)}$, $i = 1, \ldots, m$ and $\hat{\boldsymbol{\nu}}$ for the corresponding quantities computed from the $\hat{\lambda}_{ij}$'s. It then easily follows from (2.2) that

$$\sum_{i=1}^m (\hat{\boldsymbol{\nu}}^{(i)})^{-1/2} \big(\mathrm{ovec}\, \mathbf{T}_i\big) = \mathbf{0}, \tag{3.3}$$

so that $Q^{(n)}$ can be rewritten as

$$Q^{(n)} = \sum_{i,i'=1}^m \big(\mathrm{ovec}\, \mathbf{T}_i\big)' \Big[\delta_{ii'}\mathbf{I}_s - (\hat{\boldsymbol{\nu}}^{(i)})^{-1/2} \hat{\boldsymbol{\nu}}\, (\hat{\boldsymbol{\nu}}^{(i')})^{-1/2}\Big] \big(\mathrm{ovec}\, \mathbf{T}_{i'}\big). \tag{3.4}$$

Under this latter form, the asymptotically chi-square distribution of $Q^{(n)}$ (hence also that of $-2 \log \Lambda$) under the multinormal version of $\mathcal{H}_0$ is a direct consequence of the more general Theorem 4.1 stated in Section 4.

Theorem 3.2: *Assume that Assumptions (A) and (B) hold. Then, under $\mathcal{H}_0$ and Gaussian densities, $Q^{(n)}$ and $-2 \log \Lambda$ are asymptotically chi-square, with $(m-1)s$ degrees of freedom.*

The resulting Gaussian test $\phi^{(n)}$ consists in rejecting (at asymptotic level α) the null hypothesis of multinormal CPC whenever $Q^{(n)}$ exceeds the α-upper quantile of the chi-square distribution with $(m-1)s$ degrees of freedom.

4. A pseudo-Gaussian test statistic.

Section 3 was merely developing an alternative proof of Flury's Gaussian results, as a preparation for this section, where we are turning the Gaussian LRT statistic $-2 \log \Lambda$ or, equivalently, $Q^{(n)}$, into a pseudo-Gaussian test statistic $Q_\dagger^{(n)}$, under the general heterokurtic assumptions of Section 2.

Letting $\mathbf{T}_i(\underline{f}) := (1 + \kappa_k(f_i))^{-1/2}\mathbf{T}_i$, define the adjusted or robustified version $Q_\dagger^{(n)}(\underline{f})$ of the Gaussian test statistic $Q^{(n)}$ as

$$Q_\dagger^{(n)}(\underline{f}) = \sum_{i,i'=1}^m \big(\mathrm{ovec}\, \mathbf{T}_i(\underline{f})\big)' \Big[\delta_{ii'}\mathbf{I}_s - (\boldsymbol{\nu}^{(i)}(\underline{f}))^{-1/2} \boldsymbol{\nu}(\underline{f}) (\boldsymbol{\nu}^{(i')}(\underline{f}))^{-1/2}\Big] \big(\mathrm{ovec}\, \mathbf{T}_{i'}(\underline{f})\big), \tag{4.5}$$

with $\boldsymbol{\nu}^{(i)}(\underline{f}) := \text{diag}(\nu^{(i)}_{12}(\underline{f}), \nu^{(i)}_{13}(\underline{f}), \ldots, \nu^{(i)}_{(k-1)k}(\underline{f}))$, $i = 1, \ldots, m$, where

$$\nu^{(i)}_{jl}(\underline{f}) := \frac{(1 + \kappa_k(f_i))\lambda_{ij}\lambda_{il}}{r_i^{(n)}(\lambda_{ij} - \lambda_{il})^2}, \quad j, l = 1, \ldots, k, \quad j < l,$$

and $\boldsymbol{\nu}(\underline{f}) := (\sum_{i=1}^m (\boldsymbol{\nu}^{(i)}(\underline{f}))^{-1})^{-1}$ (same notational convention as in (3.2)). Note that, at the multinormal ($\underline{f} = \underline{\phi} := (\phi, \ldots, \phi)$), one has $\mathbf{T}_i(\underline{\phi}) = \mathbf{T}_i$, $\boldsymbol{\nu}^{(i)}(\underline{\phi}) = \boldsymbol{\nu}^{(i)}$, $i = 1, \ldots, m$, and $\boldsymbol{\nu}(\underline{\phi}) = \boldsymbol{\nu}$, so that $Q^{(n)}_\dagger(\underline{\phi})$ takes the form $Q^{(n)}_\dagger(\underline{\phi}) = \sum_{i,i'=1}^m (\text{ovec}\,\mathbf{T}_i)'[\delta_{ii'}\mathbf{I}_s - (\boldsymbol{\nu}^{(i)})^{-1/2}\boldsymbol{\nu}(\boldsymbol{\nu}^{(i')})^{-1/2}](\text{ovec}\,\mathbf{T}_{i'})$ which, computed at the estimated λ_{ij}'s, clearly coincides with $Q^{(n)}$ (see (3.4)).

Now, in order to obtain a genuine test statistic (that is, a random variable that does not depend on $\underline{f}$ anymore) which nevertheless, under any $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$ (with $\underline{f} \in (\mathcal{F}_1^4)^m$), is asymptotically equivalent to $Q^{(n)}_\dagger(\underline{f})$, it is sufficient to replace in (4.5) the $\kappa_k(f_i)$'s and the λ_{ij}'s with consistent (still under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, $\underline{f} \in (\mathcal{F}_1^4)^m$) estimators $\hat{\kappa}_i$ and $\hat{\lambda}_{ij}$. An obvious choice for $\hat{\kappa}_i$ is

$$\hat{\kappa}_i^{(n)} := (k(k+2))^{-1}\Big(n_i^{-1}\sum_{j=1}^{n_i} d^4_{ij}(\bar{\mathbf{X}}_i, \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\beta}}')\Big) - 1, \qquad i = 1, \ldots, m;$$

write $Q^{(n)}_\dagger$ for the resulting statistic.

We are now able to state the main result of this paper. The following theorem establishes the pseudo-Gaussian nature of the test $\phi^{(n)}_\dagger$ which consists in rejecting (at asymptotic level α) the null hypothesis of CPC whenever $Q^{(n)}_\dagger$ exceeds the α-upper quantile of the chi-square distribution with $(m-1)s$ degrees of freedom (see the appendix for a proof).

Theorem 4.1 : *Assume that Assumptions (A), (B), and the null hypothesis $\mathcal{H}_0$, hold. Then,*

(i) under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$ with $\underline{f} \in (\mathcal{F}_1^4)^m$, $Q^{(n)}_\dagger$ is asymptotically chi-square with $(m-1)s$ degrees of freedom;

(ii) under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{\phi}}$ (the multinormal case), $Q^{(n)}_\dagger = Q^{(n)} + o_\mathrm{P}(1)$, as $n \to \infty$.

It follows that the test $\phi^{(n)}_\dagger$ rejecting $\mathcal{H}_0$ whenever $Q^{(n)}_\dagger$" exceeds the α-upper quantile of a chi-square distribution with $(m-1)s$ degrees of freedom is a pseudo-Gaussian test with asymptotic level α. Note that we do not assume that the m populations under $\mathcal{H}_0$ have the same radial density, or even the same kurtosis. Each term in $Q^{(n)}_\dagger$ originating from population i indeed is weighted by a factor taking possibly heterogeneous kurtoses into account.

If, however, homokurticity can be assumed—that is, if the underlying $\underline{f}$ belongs to the collection $(\mathcal{F}_1^4)^m_{\text{homo}}$ of m-tuples of radial densities in $(\mathcal{F}_1^4)^m$ for which $\kappa_k(f_i)$ does not depend on i, one may rely on a much simpler test statistic. Indeed, for $\underline{f} \in (\mathcal{F}_1^4)^m_{\text{homo}}$,

$Q_{\dagger}^{(n)}(\underline{f})$ reduces to

$$Q_{*}^{(n)}(\underline{f}) := \frac{1}{1+\kappa_k(f_1)} \sum_{i,i'=1}^{m} (\text{ovec } \mathbf{T}_i)' \left[\delta_{ii'}\mathbf{I}_s - (\boldsymbol{\nu}^{(i)})^{-1/2}\boldsymbol{\nu}\,(\boldsymbol{\nu}^{(i')})^{-1/2}\right] (\text{ovec } \mathbf{T}_{i'})\,.$$

Consequently, a homokurtic version ($\phi_{*}^{(n)}$, say) of our pseudo-Gaussian test can be based on the asymptotically chi-square distribution (with $(m-1)s$ degrees of freedom) of the statistic

$$Q_{*}^{(n)} := \frac{1}{1+\hat{\kappa}^{(n)}} \sum_{i=1}^{m} (\text{ovec } \mathbf{T}_i)'\,(\text{ovec } \mathbf{T}_i)$$

resulting from $Q_{*}^{(n)}(\underline{f})$ (i) by replacing the λ_{ij}'s with their Gaussian MLEs and $\kappa_k(f_1)$ with a consistent estimate $\hat{\kappa}^{(n)}$, such as

$$\hat{\kappa}^{(n)} := (k(k+2))^{-1}\Big(n^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} d_{ij}^4(\bar{\mathbf{X}}_i, \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Lambda}}_i\boldsymbol{\beta}')\Big) - 1,$$

then (ii) by using (3.3) again.

This shows that, under $\mathrm{P}_{\boldsymbol{\vartheta};\underline{f}}^{(n)}$ with $\underline{f} \in (\mathcal{F}_1^4)_{\text{homo}}^m$, the Gaussian test statistic $Q^{(n)}$ simply has to be corrected by a factor depending on the common kurtosis value in order to remain valid, which confirms earlier results by Shapiro and Browne (1987) and Boik (2002).

5. Numerical results.

5.1. *Simulations.*

In this section, we perform a Monte Carlo study to asses the finite-sample performances of the tests considered in the previous sections.. We concentrate on comparing Flury's traditional LRT ($\phi_{\text{LRT}}^{(n)}$) for testing the null hypothesis of Common Principal Components with $\phi^{(n)}$ defined in Section 3 and with $\phi_{*}^{(n)}$ and $\phi_{\dagger}^{(n)}$ defined in Section 4.

We generated $N = 10,000$ independent replications of three pairs ($m = 2$) of mutually independent samples (with respective sizes n_1 and n_2) of random vectors

$$\boldsymbol{\varepsilon}_{\ell;1j_1} \quad \text{and} \quad \boldsymbol{\varepsilon}_{\ell;2j_2}, \qquad j_i = 1,\ldots,n_i, \quad i = 1,2, \quad \ell = 1,\,2,\,3$$

with zero mean and unit covariance, and

($\ell = 1$) (Gaussian case) $\boldsymbol{\varepsilon}_{1;1j_1}$ and $\boldsymbol{\varepsilon}_{1;2j_2}$ both Gaussian;
($\ell = 2$) (non-Gaussian homokurtic case) $\boldsymbol{\varepsilon}_{2;1j_1}$ and $\boldsymbol{\varepsilon}_{2;2j_2}$ both Student with 5 degrees of freedom;
($\ell = 3$) (heterokurtic case) $\boldsymbol{\varepsilon}_{3;1j_1}$ Gaussian and $\boldsymbol{\varepsilon}_{3;2j_2}$ Student with 5 degrees of freedom,

respectively. We considered sample sizes $n_1 = 40$, $n_2 = 60$, and $n_1 = n_2 = 100$. Two different designs were studied.

In Design (A), the random vectors $\boldsymbol{\varepsilon}_{\ell;1j_1}$ and $\boldsymbol{\varepsilon}_{\ell;2j_2}$ are bivariate ($k=2$), whereas they are four-dimensional vectors ($k=4$) in Design (B). Each replication of the $\boldsymbol{\varepsilon}_{\ell;1j_1}$'s was transformed into

$$\mathbf{X}_{\ell;1j_1} = \boldsymbol{\beta}\boldsymbol{\Lambda}_1^{1/2}\boldsymbol{\varepsilon}_{\ell;1j_1}, \quad j_1 = 1,\ldots,n_1, \quad \ell = 1,\,2,\,3,$$

where

$$\boldsymbol{\beta} = \boldsymbol{\beta}^{(\mathrm{A})} = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_1^{(\mathrm{A})} = \begin{pmatrix} 16 & 0 \\ 0 & 1 \end{pmatrix}$$

in Design (A), and

$$\boldsymbol{\beta} = \boldsymbol{\beta}^{(\mathrm{B})} = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) & \mathbf{0} \\ \sin(\pi/6) & \cos(\pi/6) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_1^{(\mathrm{B})} = \begin{pmatrix} 16 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/16 \end{pmatrix}$$

in Design (B). Then each replication of the $\boldsymbol{\varepsilon}_{\ell;2j_2}$'s was subjected to a sequence of 21 linear transformations

$$\mathbf{X}_{\ell;2j_2;\xi} = \mathbf{B}_\xi\boldsymbol{\beta}\boldsymbol{\Lambda}_2^{1/2}\boldsymbol{\varepsilon}_{\ell;2j_2}, \quad j_2 = 1,\ldots,n_2, \quad \ell = 1,\,2,\,3, \quad \xi = 0,\ldots,20,$$

where

$$\mathbf{B}_\xi = \mathbf{B}_\xi^{(\mathrm{A})} = \begin{pmatrix} \cos(\pi\xi/80) & -\sin(\pi\xi/80) \\ \sin(\pi\xi/80) & \cos(\pi\xi/80) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}_2^{(\mathrm{A})} = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$$

in Design (A), and

$$\mathbf{B}_\xi = \mathbf{B}_\xi^{(\mathrm{B})} = \begin{pmatrix} \cos(\pi\xi/80) & -\sin(\pi\xi/80) & \mathbf{0} \\ \sin(\pi\xi/80) & \cos(\pi\xi/80) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}_2^{(\mathrm{B})} = \begin{pmatrix} 9 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

in Design (B).

Clearly, the covariance matrices of $\mathbf{X}_{\ell;1j_1}$ and $\mathbf{X}_{\ell;2j_2;0}$ have common eigenvectors $\boldsymbol{\beta}$, with distinct eigenvalue matrices $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$, while the eigenvectors associated with $\mathbf{X}_{\ell;2j_2;\xi}$, $\xi = 1,\ldots,20$ increasingly differ from those of $\mathbf{X}_{\ell;1j_1}$, thus characterizing increasingly heterogeneous alternatives to the null hypothesis of CPC. Note that, in Design (B), the second population shows an eigenvalue with multiplicity larger than one (partial sphericity); Assumption (H) nevertheless is satisfied.

Rejection frequencies for $\phi^{(n)}_{\mathrm{LRT}}$, $\phi^{(n)}$, $\phi^{(n)}_*$ and $\phi^{(n)}_\dagger$ (based on the asymptotic chi-square critical values, at nominal 5% level) are plotted against ξ in the six graphs of Figure 1 (Design (A)) and the six graphs of Figure 2 (Design (B)). Inspection of these graphs reveals (more clearly in the large-sample case $n_1 = n_2 = 100$ than in the small-sample case) several well expected facts:

(i) $\phi_{\rm LRT}^{(n)}$ and $\phi^{(n)}$ exhibit identical performances irrespective of the underlying distribution: both are valid under Gaussian densities, and both strongly overreject in the two other cases, with Type I risks as high as 20%;
(ii) under Gaussian densities, the performances of $\phi_*^{(n)}$ and $\phi_\dagger^{(n)}$ coincide with those of $\phi_{\rm LRT}^{(n)}$ and $\phi^{(n)}$;
(iii) the pseudo-Gaussian procedures $\phi_*^{(n)}$ and $\phi_\dagger^{(n)}$ both are valid in the non-Gaussian homokurtic case, but $\phi_\dagger^{(n)}$ remains valid in the heterokurtic case, while $\phi_*^{(n)}$ is either invalid (Design (A)) or severely biased (Design (B)).

Moreover, both $\phi_*^{(n)}$ (under homokurticity) and $\phi_\dagger^{(n)}$ (also under heterokurticity) appear to be *efficiency-robust*, in the sense that their power curves are roughly parallel to that of the invalid $\phi_{\rm LRT}^{(n)}$ and $\phi^{(n)}$: robustification against non-Gaussian and possibly heterokurtic observations (*validity-robustness*) thus has not been obtained at the expense of increased Type II risks.

5.2. *A real data application.*

The following application is based on a dataset collected by Airoldi and Hoffman (1984), who took various skull measurements on two species of voles: *Microtus californicus* and *Microtus ochrogaster*. This dataset was studied by Flury and Riedwyl (1988), who only take into account the males of those two species of mammals. The three variables measured on $n = 110$ ($n_1 = n_2 = 55$) individuals are (a) the condylo-incisive length, (b) the zygomatic width, and (c) the skull height. The empirical covariance matrices of the two samples are

$$\mathbf{S}_1 = \begin{pmatrix} 138.15 & 133.17 & 69.42 \\ 133.17 & 136.29 & 72.88 \\ 69.42 & 72.88 & 44.16 \end{pmatrix} \quad \text{and} \quad \mathbf{S}_2 = \begin{pmatrix} 70.61 & 65.25 & 29.22 \\ 65.25 & 66.52 & 28.99 \\ 29.22 & 28.99 & 19.87 \end{pmatrix},$$

respectively.

Before starting a statistical analysis of this dataset based on the CPC model, Flury and Riedwyl propose to check whether the CPC assumption is appropriate. The value of Flury's LRT test statistic (three degrees of freedom) is $-2\log\Lambda = 6.535$, yielding p-value 0.088. The CPC model is thus rejected at probability level 10%.

Even a rapid glance at the data however reveals that they hardly look Gaussian, and Mardia (1974)'s test of skewness is rejecting multinormality at probability level 5%. Flury's likelihood ratio procedure thus seems questionable in this situation. Turning to the more reliable pseudo-Gaussian tests we are proposing here, we obtain $Q_*^{(n)} = 5.840$, with p-value 0.120 and $Q_\dagger^{(n)} = 5.521$ with p-value 0.137. Even at nominal level 10%, there is nothing wrong, thus, with the null hypothesis. The data do not produce evidence against Common Principal Components assumptions, and the spuriously low p-value of Flury's test is probably due to the non-Gaussian and heterokurtic nature of the observations (with estimated radial kurtoses $\hat{\kappa}_1 = 0.007$ for Microtus californicus and $\hat{\kappa}_2 = 0.227$ for Microtus ochrogaster).

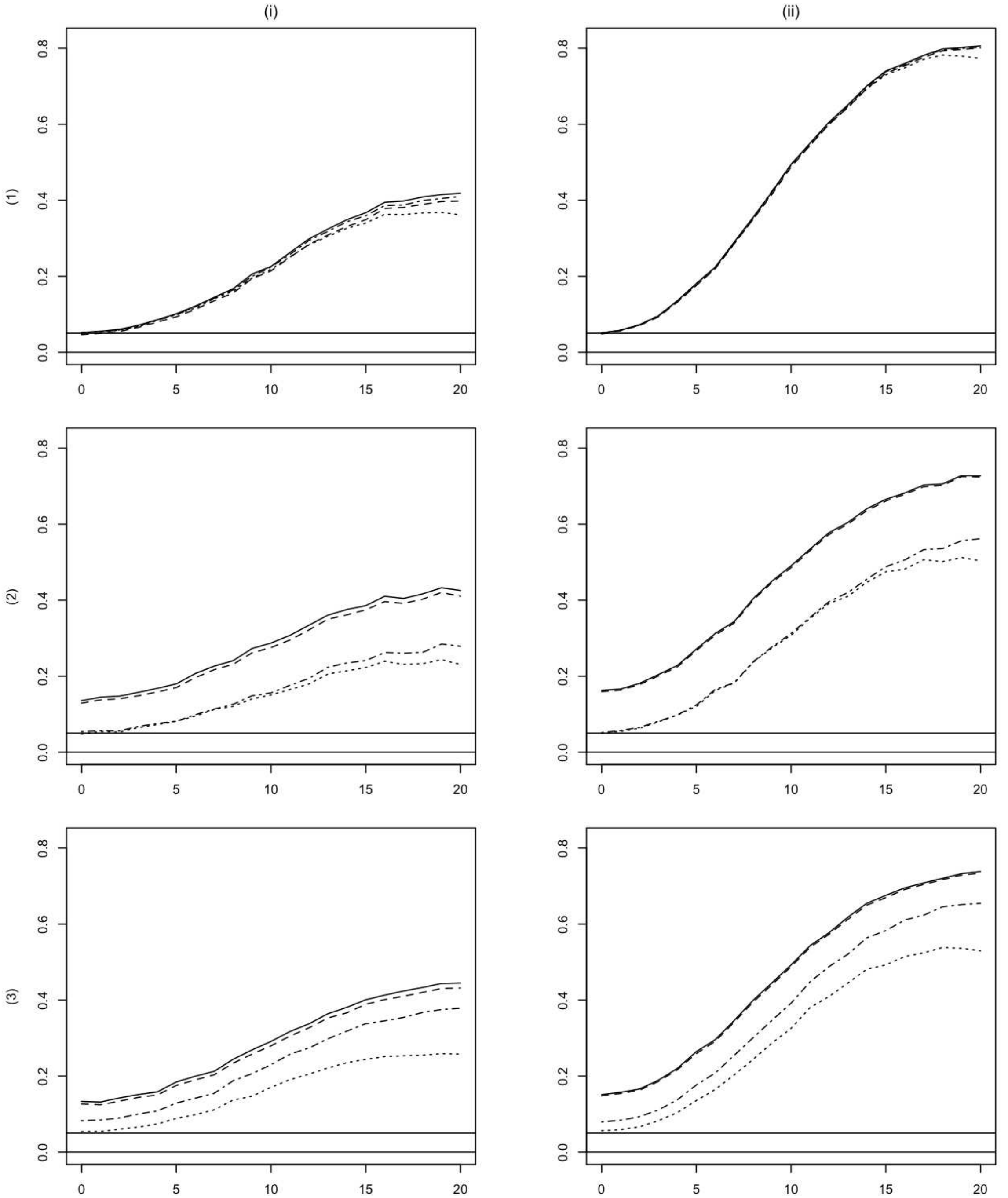

Figure 1. Rejection frequencies (out of $N = 10,000$ replications in Design (A)), under the null and various alternatives associated with the Gaussian case (1), the non-Gaussian homokurtic case (2), or the heterokurtic case (3) (see Section 5.1 for details), of the Gaussian LRT $\phi^{(n)}_{\mathrm{LRT}}$ (solid line), our Gaussian test $\phi^{(n)}$ (dashed line), and its pseudo-Gaussian versions $\phi^{(n)}_{*}$ and $\phi^{(n)}_{\dagger}$ (dot-dash line and dotted line, respectively). The asymptotic nominal size α is 5%; $k = 2$, $m = 2$. Sample sizes are (i) $n_1 = 40$ and $n_2 = 60$, or (ii) $n_1 = n_2 = 100$.

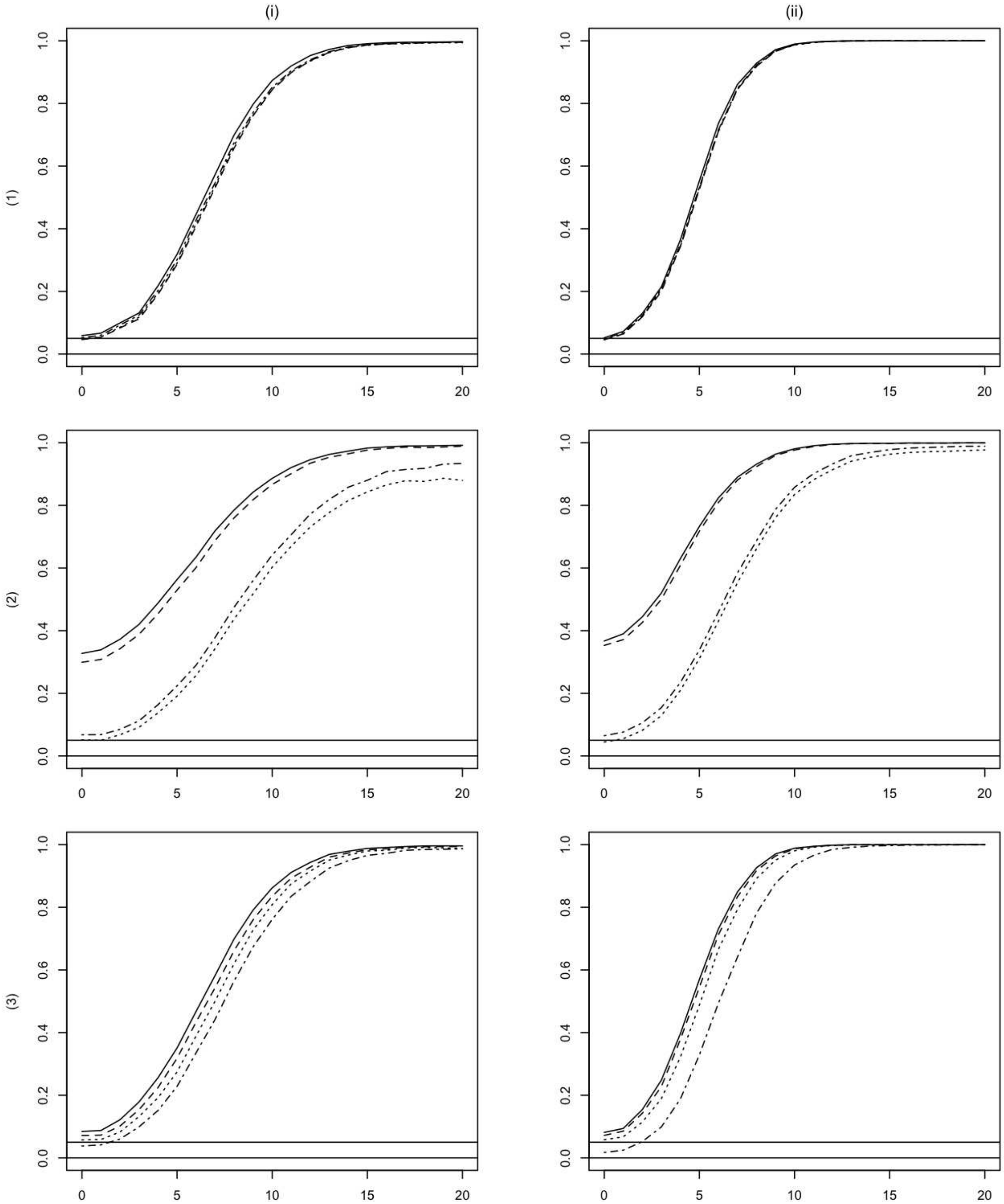

Figure 2. Rejection frequencies (out of $N = 10,000$ replications in Design (B)), under the null and various alternatives associated with the Gaussian case (1), the non-Gaussian homokurtic case (2), or the heterokurtic case (3) (see Section 5.1 for details), of the Gaussian LRT $\phi^{(n)}_{\mathrm{LRT}}$ (solid line), our Gaussian test $\phi^{(n)}$ (dashed line), and its pseudo-Gaussian versions $\phi^{(n)}_{*}$ and $\phi^{(n)}_{\dagger}$ (dot-dash line and dotted line, respectively). The asymptotic nominal size α is 5%; $k = 4$, $m = 2$. Sample sizes are (i) $n_1 = 40$ and $n_2 = 60$, or (ii) $n_1 = n_2 = 100$.

6. Final comments.

In this paper, we provide a pseudo-Gaussian test for the null hypothesis of Common Principal Components (CPC). This test is asymptotically equivalent to the Gaussian LRT under Gaussian densities but, contrary to the Gaussian LRT, remains asymptotically valid under possibly heterokurtic elliptical distributions (with finite fourth-order moments).

A number of related problems deserve to be investigated further. In the hierarchy of similarities between covariance matrices given by Flury (1988) is the hypothesis of "partial CPC"; see Section 1.1. Since, usually, only the first few principal directions are considered in practice, the hypothesis of partial CPC is appealing for the applications. Of course, it is also much less stringent than the hypothesis of *full* CPC, hence is possibly compatible with many more real data sets. Therefore, it would be interesting to extend the tests for CPC developed in this paper into tests for the null hypothesis of partial CPC or into tests for the null hypothesis of CPC against an alternative of partial CPC.

Finally, we indicate that it is possible to study the asymptotic behavior of the proposed tests under sequences of local alternatives, and to obtain closed-form expressions for the corresponding local asymptotic powers. This, however, requires somewhat technical results related to *uniform local and asymptotic normality* (ULAN) in *curved* experiments (see Hallin et al. 2009), that are beyond the scope of the present work.

Acknowledgements

We thank the referees for their very careful reading of the original version of this paper. Their pertinent comments helped us fixing some not so minor details in the exposition, which greatly improved the quality of the final version.

Marc Hallin is also Extra-muros Fellow of CentER, Tilburg University, the Nederlands. His research is supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, and by a Discovery Grant of the Australian Research Council. Davy Paindaveine's research is supported by a Mandat d'Impulsion Scientifique of the Fonds National de la Recherche Scientifique, Communauté française de Belgique.

7. Appendix.

Proof of Proposition 3.1. Recalling that $\mathbf{T}_i = n_i^{1/2}\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}$, we have

$$-2\log\Lambda = \sum_{i=1}^{m}\Big\{\log|\mathrm{diag}(\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}})| - \log|\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}|\Big\} = \sum_{i=1}^{m}\Big\{\log|n_i^{-1/2}\mathrm{diag}(\mathbf{T}_i)| - \log|n_i^{-1/2}\mathbf{T}_i|\Big\}.$$

Since $\log|\mathbf{I}_k+\mathbf{A}| = \mathrm{tr}[\mathbf{A}] - \frac{1}{2}\mathrm{tr}[\mathbf{A}^2] + o(\|\mathbf{A}\|^2)$ as $\|\mathbf{A}\| \to 0$, and noting that $\mathrm{ovec}(\mathbf{I}_k) = \mathbf{0}$, we obtain (as $n \to \infty$, under $\mathcal{H}_0$ and finite fourth-order moments)

$$-2\log\Lambda = \frac{1}{2}\sum_{i=1}^m n_i \Big(\mathrm{tr}[(n_i^{-1/2}\mathbf{T}_i - \mathbf{I}_k)^2] - \mathrm{tr}[(\mathrm{diag}(n_i^{-1/2}\mathbf{T}_i - \mathbf{I}_k))^2]\Big) + o_{\mathrm{P}}(1) \quad \text{(A.1)}$$

$$= \sum_{i=1}^m n_i \big(\mathrm{ovec}(n_i^{-1/2}\mathbf{T}_i - \mathbf{I}_k)\big)'\big(\mathrm{ovec}(n_i^{-1/2}\mathbf{T}_i - \mathbf{I}_k)\big) + o_{\mathrm{P}}(1)$$

$$= Q^{(n)} + o_{\mathrm{P}}(1),$$

as was to be proved. □

Let $\mathbf{D}_k$ be the $(s \times k^2)$ matrix such that $\mathbf{D}_k(\mathrm{vec}\,\mathbf{A}) = \mathrm{ovec}(\mathbf{A})$ and denote by $\mathcal{T}_k$ the collection of $(k \times k)$ matrices $\mathbf{A} = (A_{jl})$ such that $A_{jl} = 0$ for any $j \geq l$. Also, let $\mathbf{J}_k := (\mathrm{vec}\,\mathbf{I}_k)(\mathrm{vec}\,\mathbf{I}_k)'$ and write $\mathbf{K}_k := \sum_{i,j=1}^k (\mathbf{e}_i\mathbf{e}_j') \otimes (\mathbf{e}_j\mathbf{e}_i')$ for the $k^2 \times k^2$ *commutation matrix* (see Magnus and Neudecker 1999).

Lemma 7.1: *(i) For any* $\mathbf{A} \in \mathcal{T}_k$, $\mathbf{D}_k'(\mathrm{ovec}\,\mathbf{A}) = \mathrm{vec}(\mathbf{A})$*; (ii)* $\mathbf{D}_k\mathbf{D}_k' = \mathbf{I}_s$*; (iii)* $\mathbf{D}_k\mathbf{K}_k\mathbf{D}_k' = \mathbf{0}$*; (iv)* $\mathbf{D}_k\mathbf{J}_k\mathbf{D}_k' = \mathbf{0}$.

Proof of Lemma 7.1. Fix $\mathbf{A} \in \mathcal{T}_k$. Then, by using successively the definition of $\mathbf{D}_k$ and the fact that $\mathbf{A} \in \mathcal{T}_k$, we have, for any $(k \times k)$ matrix $\mathbf{B}$, $(\mathrm{vec}\,\mathbf{B})'\mathbf{D}_k'(\mathrm{ovec}\,\mathbf{A}) = (\mathrm{ovec}\,\mathbf{B})'(\mathrm{ovec}\,\mathbf{A}) = (\mathrm{vec}\,\mathbf{B})'(\mathrm{vec}\,\mathbf{A})$, which proves Part (i). Part (ii) follows since, for any $\mathbf{A} \in \mathcal{T}_k$, $\mathbf{D}_k\mathbf{D}_k'(\mathrm{ovec}\,\mathbf{A}) = \mathbf{D}_k\mathrm{vec}(\mathbf{A}) = \mathrm{ovec}(\mathbf{A})$. Now, using (i) again, we obtain, for any $\mathbf{A} \in \mathcal{T}_k$, $\mathbf{D}_k\mathbf{K}_k\mathbf{D}_k'(\mathrm{ovec}\,\mathbf{A}) = \mathbf{D}_k\mathbf{K}_k(\mathrm{vec}\,\mathbf{A}) = \mathbf{D}_k(\mathrm{vec}\,\mathbf{A}') = \mathbf{0}$, which proves (iii). As for Part (iv), it is obtained by noting that $\mathbf{D}_k(\mathrm{vec}\,\mathbf{I}_k) = \mathrm{ovec}(\mathbf{I}_k) = \mathbf{0}$. □

Next, define $\tilde{\mathbf{Z}}(\underline{f}) := ((\mathrm{ovec}\,\mathbf{T}_1(\underline{f}))', \ldots, (\mathrm{ovec}\,\mathbf{T}_m(\underline{f}))')'$, $\tilde{\mathbf{Z}}_a(\underline{f}) := (\mathbf{Z}_{a1}'(\underline{f}), \ldots, \mathbf{Z}_{am}'(\underline{f}))'$, and $\tilde{\mathbf{Z}}_b(\underline{f}) := (n/2)^{1/2}\boldsymbol{\xi}(\underline{f})(\hat{\boldsymbol{\beta}}^\star - \boldsymbol{\beta}^\star)$, with

$$\mathbf{Z}_{ai}(\underline{f}) := \frac{n_i^{1/2}}{(1+\kappa_k(f_i))^{1/2}}\,\mathrm{ovec}\,(\boldsymbol{\Lambda}_i^{-1/2}\boldsymbol{\beta}'(\mathbf{S}_i - \boldsymbol{\Sigma}_i)\boldsymbol{\beta}\boldsymbol{\Lambda}_i^{-1/2}),$$

and the $(ms \times s)$ matrix $\boldsymbol{\xi}(\underline{f}) := ((\boldsymbol{\nu}^{(1)}(\underline{f}))^{-1/2}, \ldots, (\boldsymbol{\nu}^{(m)}(\underline{f}))^{-1/2})'$.

Lemma 7.2: *For any* $\underline{f} \in (\mathcal{F}_1^4)^m$*, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$,

(i) $\tilde{\mathbf{Z}}(\underline{f}) = \tilde{\mathbf{Z}}_a(\underline{f}) + \tilde{\mathbf{Z}}_b(\underline{f}) + o_{\mathrm{P}}(1)$*, as* $n \to \infty$*, where*
(ii) $\tilde{\mathbf{Z}}_a(\underline{f})$ *is asymptotically normal, with zero mean and unit covariance.*

Proof of Lemma 7.2. Throughout this proof, we fix $\underline{f} \in (\mathcal{F}_1^4)^m$. All $o_{\mathrm{P}}(1)$ quantities are taken under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, as $n \to \infty$.

(i) Decomposing $(\mathrm{ovec}\,\mathbf{T}_i(\underline{f}))$ into

$$\frac{n_i^{1/2}}{(1+\kappa_k(f_i))^{1/2}}\,\mathrm{ovec}\,\big(\hat{\boldsymbol{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'(\mathbf{S}_i - \boldsymbol{\Sigma}_i)\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Lambda}}_i^{-1/2}\big) + \frac{n_i^{1/2}}{(1+\kappa_k(f_i))^{1/2}}\,\mathrm{ovec}\,\big(\hat{\boldsymbol{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_i\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Lambda}}_i^{-1/2}\big)$$

and using the continuous mapping theorem and the definition of $\mathbf{D}_k$, we obtain

$$\text{ovec}\,(\mathbf{T}_i(\underline{f})) = \mathbf{Z}_{ai}(\underline{f}) + \frac{n_i^{1/2}}{(1+\kappa_k(f_i))^{1/2}}\,\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{\Sigma}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}\big) + o_{\mathrm{P}}(1),$$

so that it only remains to prove that

$$\frac{n_i^{1/2}}{(1+\kappa_k(f_i))^{1/2}}\,\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{\Sigma}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = (n/2)^{1/2}(\boldsymbol{\nu}^{(i)}(\underline{f}))^{-1/2}(\hat{\boldsymbol{\beta}}^{\star} - \boldsymbol{\beta}^{\star}) + o_{\mathrm{P}}(1),$$

or, equivalently, that

$$n_i^{1/2}\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{\Sigma}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = (n/2)^{1/2}(\boldsymbol{\nu}^{(i)})^{-1/2}(\hat{\boldsymbol{\beta}}^{\star} - \boldsymbol{\beta}^{\star}) + o_{\mathrm{P}}(1). \tag{A.2}$$

Now, splitting $\hat{\boldsymbol{\beta}}$ into $\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and using the continuous mapping theorem again, we obtain

$$n_i^{1/2}\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{\Sigma}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = n_i^{1/2}\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\boldsymbol{\beta}'\mathbf{\Sigma}_i\boldsymbol{\beta}\hat{\mathbf{\Lambda}}_i^{-1/2}\big)$$
$$+n_i^{1/2}\mathbf{D}_k\text{vec}\,\big(\mathbf{\Lambda}_i^{-1/2}\boldsymbol{\beta}'\mathbf{\Sigma}_i(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\mathbf{\Lambda}_i^{-1/2}\big) + n_i^{1/2}\mathbf{D}_k\text{vec}\,\big(\mathbf{\Lambda}_i^{-1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{\Sigma}_i\boldsymbol{\beta}\mathbf{\Lambda}_i^{-1/2}\big) + o_{\mathrm{P}}(1).$$

Since $\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\boldsymbol{\beta}'\mathbf{\Sigma}_i\boldsymbol{\beta}\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = \mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\mathbf{\Lambda}_i\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = \text{ovec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\mathbf{\Lambda}_i\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = \mathbf{0}$ under $\mathcal{H}_0$ and since $\mathbf{K}_k(\text{vec}\,\mathbf{A}) = \text{vec}\,(\mathbf{A}')$, this yields

$$n_i^{1/2}\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{\Sigma}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = n_i^{1/2}\mathbf{D}_k[\mathbf{I}_{k^2}+\mathbf{K}_k]\text{vec}\,\big(\mathbf{\Lambda}_i^{-1/2}\boldsymbol{\beta}'\mathbf{\Sigma}_i(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\mathbf{\Lambda}_i^{-1/2}\big) + o_{\mathrm{P}}(1).$$

Well-known properties of Kronecker products and (2.6) imply that

$$n_i^{1/2}\mathbf{D}_k\text{vec}\,\big(\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{\Sigma}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}\big) = n_i^{1/2}\mathbf{D}_k[\mathbf{I}_{k^2}+\mathbf{K}_k]\big(\mathbf{\Lambda}_i^{-1/2}\otimes(\mathbf{\Lambda}_i^{1/2}\boldsymbol{\beta}')\big)\mathbf{G}_k(\hat{\boldsymbol{\beta}}^{\star}-\boldsymbol{\beta}^{\star})+o_{\mathrm{P}}(1).$$

Equality (A.2)—hence also Part (i) of the lemma—then follows easily by checking that

$$n_i^{1/2}\mathbf{D}_k[\mathbf{I}_{k^2}+\mathbf{K}_k]\big(\mathbf{\Lambda}_i^{-1/2}\otimes(\mathbf{\Lambda}_i^{1/2}\boldsymbol{\beta}')\big)\mathbf{G}_k = (n/2)^{1/2}(\boldsymbol{\nu}^{(i)})^{-1/2}.$$

(ii) The definition of $\mathbf{D}_k$ entails that

$$\mathbf{Z}_{ai}(\underline{f}) = \frac{n_i^{1/2}}{(1+\kappa_k(f_i))^{1/2}}\,\text{ovec}\,(\mathbf{\Lambda}_i^{-1/2}\boldsymbol{\beta}'(\mathbf{S}_i-\mathbf{\Sigma}_i)\boldsymbol{\beta}\mathbf{\Lambda}_i^{-1/2})$$
$$= \frac{n_i^{1/2}}{(1+\kappa_k(f_i))^{1/2}}\,\mathbf{D}_k\big((\mathbf{\Lambda}_i^{-1/2}\boldsymbol{\beta}')\otimes(\mathbf{\Lambda}_i^{-1/2}\boldsymbol{\beta}')\big)\text{vec}\,(\mathbf{S}_i-\mathbf{\Sigma}_i).$$

Since $n_i^{1/2}\text{vec}\,(\mathbf{S}_i-\mathbf{\Sigma}_i)$ is, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$ with $\underline{f}\in(\mathcal{F}_1^4)^m$, asymptotically normal with mean zero and covariance matrix $(1+\kappa_k(f_i))(\mathbf{I}_{k^2}+\mathbf{K}_k)(\mathbf{\Sigma}_i\otimes\mathbf{\Sigma}_i)+\kappa_k(f_i)\text{vec}\,(\mathbf{\Sigma}_i)(\text{vec}\,(\mathbf{\Sigma}_i))'$ (see, e.g., page 212 in Bilodeau and Brenner 1999), it follows, by using Lemma 7.1(ii)-(iv), that $\mathbf{Z}_{ai}(\underline{f})$ is asymptotically normal with mean zero and covariance matrix

$$\mathbf{D}_k(\mathbf{I}_{k^2} + \mathbf{K}_k)\mathbf{D}_k' + \frac{\kappa_k(f_i)}{1+\kappa_k(f_i)}\mathbf{D}_k\mathbf{J}_k\mathbf{D}_k' = \mathbf{I}_s,$$

which, along with the mutual independence of the $\mathbf{S}_i$'s, establishes the result. □

Proof of Theorem 4.1. The consistency of the $\hat{\kappa}_i$'s and the $\hat{\lambda}_{ij}$'s entails that $Q_\dagger^{(n)} = Q_\dagger^{(n)}(\underline{f}) + o_{\rm P}(1)$, as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, $\underline{f} \in (\mathcal{F}_1^4)^m$. Now, $Q_\dagger^{(n)}(\underline{f}) = \tilde{\mathbf{Z}}'(\underline{f})\mathbf{P}_{m,k}(\underline{f})\tilde{\mathbf{Z}}(\underline{f}) + o_{\rm P}(1)$, where

$$\mathbf{P}_{m,k}(\underline{f}) := \begin{pmatrix} \mathbf{P}_{m,k;1,1}(\underline{f}) & \dots & \mathbf{P}_{m,k;1,m}(\underline{f}) \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{m,k;m,1}(\underline{f}) & \dots & \mathbf{P}_{m,k;m,m}(\underline{f}) \end{pmatrix}$$

with

$$\mathbf{P}_{m,k;i,i'}(\underline{f}) := \delta_{ii'}\mathbf{I}_s - (\boldsymbol{\nu}^{(i)}(\underline{f}))^{-1/2}\boldsymbol{\nu}(\underline{f})(\boldsymbol{\nu}^{(i')}(\underline{f}))^{-1/2}.$$

Noting that $\mathbf{P}_{m,k}(\underline{f}) = \mathbf{I}_{ms} - \boldsymbol{\xi}(\underline{f})\left[\boldsymbol{\xi}'(\underline{f})\boldsymbol{\xi}(\underline{f})\right]^{-1}\boldsymbol{\xi}'(\underline{f})$, it directly follows from Lemma 7.2(i) that $Q_\dagger^{(n)} = \tilde{\mathbf{Z}}_a'(\underline{f})\mathbf{P}_{m,k}(\underline{f})\tilde{\mathbf{Z}}_a(\underline{f}) + o_{\rm P}(1)$ (still as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$). Since $\mathbf{P}_{m,k}(\underline{f})$ is idempotent, with trace $(m-1)s$, Part (i) of the theorem follows from Lemma 7.2(ii). Now, by using the consistency of the $\hat{\kappa}_i$'s and the $\hat{\lambda}_{ij}$'s, we obtain that, at the multinormal, $Q_\dagger^{(n)} = Q_\dagger^{(n)}(\underline{\phi}) + o_{\rm P}(1) = Q^{(n)} + o_{\rm P}(1)$ as $n \to \infty$ (see Section 4), which establishes Part (ii). □

References

[1] Airoldi, J.P., and Hoffmann, R.S. (1984), "Age variation in voles (Microtus californicus and Microtus ochrogaster) and its significance for systematic studies", *Occasional Papers of the Museum of the Natural History*, University of Kansas, Lawrence, 111, 1–45.

[2] Anderson, T.W. (1963), "Asymptotic theory for principal components", *Annals of Mathematical Statistics*, 34, 122–148.

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

[4] Bentler, P.M., and Dudgeon, P. (1996), "Covariance structure analysis: statistical practice, theory, and directions", *Annual Review of Psychology*, 47, 563–592.

[5] Bilodeau, M., and Brenner, D. (1999), *Theory of Multivariate Analysis*, New York: Springer-Verlag.

[6] Boente, G., and Orellana, L. (2001), "A robust approach to common principal components", in *Statistics in Genetics and in the Environmental Sciences*, ed. Sciences-Fernholz, L.T., Morgenthaler, S., and Stahel, W., Basel: Birkhauser, 117–147.

[7] 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.

[8] Boente, G., Pires, A.M., and Rodrigues I.M. (2009), "Robust tests for the common principal components model", *Journal of Statistical Planning and Inference*, 139, 1332–1347.

[9] Boik, J.R. (2002), "Spectral models for covariance matrices", *Biometrika*, 89, 159–182.

[10] Browne, M.W. (1984), "The decomposition of multitrait-multimethod matrices", *British Journal of Mathematical and Statistical Psychology*, 37, 1–21.

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

[12] Flury, B. (1986), "Asymptotic theory for common principal components analysis", *Annals of Statistics*, 14, 418–430.

[13] Flury, B. (1988), "Two generalizations of the common principal component model", *Biometrika*, 74, 59–69.

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

[15] 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.

[16] Goodnight, C.J., and Schwartz, J.M. (1997), "A bootstrap comparison of genetic covariance matrices", *Biometrics*, 53, 1026—1035.

[17] Hallin, M., and Paindaveine, D. (2008), "A general method for constructing pseudo-Gaussian tests", *Journal of the Japan Statistical Society*, 38 (Celebration Volume for Hirotogu Akaike), 27–40.

[20] Hallin, M., and Paindaveine, D. (2009), "Optimal tests for homogeneity of covariance, scale, and shape", *Journal of Multivariate Analysis*, 100, 422–444.

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

[20] Hallin, M., Paindaveine, D., and Verdebout, T. (2009), "Optimal rank-based testing for principal components", Submitted.

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

[22] Hu, L.T., Bentler, P.M., and Kano, Y. (1992), "Can test statistics in covariance structure analysis be trusted?", *Psychological Bulletin*, 112, 351–362.

[23] Krzanowski, W.J. (1979), "Between-group comparison of principal components", *Journal of the American Statistical Association*, 74, 703–707.

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

[25] Magnus, J.R., and Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: Wiley.

[26] Mardia, K.V. (1974), "Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies", *Sankhyā Ser. B*, 36, 115–128.

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

[28] Pearson, K. (1901), "On lines and planes of closest fit to system of points in space",

Philosophical Magazine, 2, 559–572.

[29] Satorra, A., and Bentler, P.M. (1988), “Scaling corrections for chi-square statistics in covariance structure analysis”, *ASA Proceedings of the Business and Economic Statistics Section, Alexandria: American Statistical Association*, 308–313.

[30] Shapiro, A., and Browne, M.W. (1987), “Analysis of covariance structures under elliptical distributions”, *Journal of the American Statistical Association*, 82, 1092–1097.

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

[32] Wilks, S.S. (1938), “The large-sample distribution of the likelihood ratio for testing composite hypotheses”, *Annals of Mathematical Statistics*, 9, 60–62.

[33] Yanagihara, H., Tonda, T., and Matsumoto, C. (2005), “The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption”, *Journal of Multivariate Analysis*, 96, 237–264.

[34] Yuan, K.H., and Bentler, P.M. (1997), “Mean and covariance structure analysis: theoretical and practical improvements”, *Journal of the American Statistical Association*, 92, 766–773.

[35] Zhang, J., and Boos, D.D. (1992), “Bootstrap critical values for testing homogeneity of covariance matrices”, *Journal of the American Statistical Association*, 87, 425–429.

[36] Zhu, L.X., Ng, K.W., and Jing, P. (2002). “Resampling methods for homogeneity tests of covariance matrices”, *Statistica Sinica*, 12, 769–783.