Pseudo-Gaussian Inference in Heterokurtic Elliptical Common Principal Components Models

Marc Hallin*, Davy Paindaveine*†
and
Thomas Verdebout*

Abstract

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), who develops Gaussian parametric inference methods for this model (Gaussian maximum likelihood estimation and Gaussian likelihood ratio testing). A key result in that context is the joint asymptotic normality of the Gaussian maximum likelihood estimators of the common eigenvectors and the corresponding eigenvalues. Flury's derivation of that result is based on a soft argument and is valid under Gaussian (hence, also homokurtic) conditions only. In this paper, we provide a formal proof of the same result under more general assumptions of elliptical, possibly heterokurtic, densities with finite fourth-order moments. This allows for a pseudo-Gaussian solution to all inference problems about eigenvectors and eigenvalues in CPC models. As an application, we consider inference about the proportion of total variance explained by a given subset of common principal components. More precisely, we test the null hypothesis $\mathcal{H}_0$ that this proportion is smaller than some fixed value $p_0 \in (0, 1)$ in each population. Based on our result, we provide a pseudo-Gaussian test which, contrary to Flury's Gaussian one, is valid under arbitrary m-tuples of elliptical densities with finite fourth-order moments, while remaining asymptotically equivalent to Flury's under multinormal distributions.

1 Introduction.

Principal Components are one of the oldest and most widespread methods of multivariate analysis. Originally proposed by Pearson (1901), Principal Component techniques were developed mainly after a seminal paper by Hotelling (1933), in a one-sample context, essentially. Inference problems in that context have been thoroughly investigated, and the asymptotics of sample principal components, in particular, can be found in most standard textbooks: see, e.g, Anderson (2003).

Multi-sample generalizations have been considered much later. In 1984, Flury introduced the so-called Common Principal Components (CPC) model. This model typically addresses datasets in which the same k variables are measured from m distinct populations, and has a

*Institut de Recherche en Statistique, E.C.A.R.E.S., and Département de Mathématique, Université Libre de Bruxelles, Belgium. Hallin and Paindaveine are also members of ECORE, the recently created association between CORE and ECARES.

†Research supported by a Crédit d'Impulsion of the Fonds National de la Recherche Scientifique, Communauté française de Belgique.

number of biometrical applications. CPC models are characterized by the structural assumption that the $(k \times k)$ covariance matrices $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_m$ of m populations under study are diagonalized by a single matrix $\boldsymbol{\beta}$ of *common eigenvectors*. More precisely, the hypothesis

$$\mathcal{H}_{\text{CPC}} : \boldsymbol{\beta}' \mathbf{\Sigma}_i \boldsymbol{\beta} = \mathbf{\Lambda}_i, \quad i = 1, \ldots, m, \tag{1.1}$$

where $\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_m$ are unspecified diagonal matrices, holds for some unspecified orthogonal matrix $\boldsymbol{\beta}$.

Flury (1984 and 1986) develops, under Gaussian assumptions, the Gaussian maximum likelihood estimation of the parameters of the CPC model and the corresponding asymptotic distribution theory. His derivation of the asymptotic normality of the estimates $\hat{\lambda}_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ of the diagonal elements of the $\mathbf{\Lambda}_i$'s, however, is a bit sketchy, with a disputable application of Slutzky's argument, and cannot be considered as a formal proof. Our purpose in this paper is to provide a formal derivation of Flury's result on the asymptotic distribution of the $\hat{\lambda}_{ij}$'s as a special case of a much more general result, where Flury's strict Gaussian assumptions are weakened into an assumption of m elliptical, possibly heterokurtic densities with finite fourth-order moments. While Gaussian densities are highly unrealistic in most applications of CPC models, all available asymptotic results indeed not only require multinormality assumptions, but moreover are quite sensitive to their violations: see Hallin, Paindaveine, and Verdebout (2007).

Our asymptotic normality result involves the kurtosis coefficients of each population, and relies on a non-trivial derivation of the asymptotic covariances between the m empirical covariance matrices $\mathbf{S}_i$, $i = 1, \ldots, m$ and the Gaussian maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of the orthogonal matrix $\boldsymbol{\beta}$ of common eigenvectors.

As an application of this result, we provide a pseudo-Gaussian test of the null hypothesis

$$\mathcal{H}_0 : \max_{i=1,\ldots,m} \Big(\sum_{j=q+1}^{k} \lambda_{ij} \,/ \sum_{j=1}^{k} \lambda_{ij} \Big) \leq p_0 \tag{1.2}$$

or, equivalently,

$$\mathcal{H}_0 : \ (1 - p_0) \sum_{j=q+1}^{k} \lambda_{ij} - p_0 \sum_{j=1}^{q} \lambda_{ij} \leq 0, \quad \text{for all } i \in \{1, \ldots, m\}, \tag{1.3}$$

under which some specified $(k - q)$-tuple of common principal components accounts for a proportion of the total variance smaller than or equal to some given $p_0 \in (0, 1)$ in all populations. Since the matrix $\boldsymbol{\beta}$ in (1.1) clearly is defined up to a permutation of its columns, the labels $j = 1, \ldots, k$ without loss of generality can be chosen in such a way that the $(k - q)$ principal components involved in $\mathcal{H}_0$ are the last $(k - q)$ ones.

Note that this hypothesis $\mathcal{H}_0$ only makes sense if the columns of $\boldsymbol{\beta}$—that is, the common eigenvectors—can be labelled in some meaningful way (an assumption which is also made by Flury). They may be ordered, for instance, in such a way that $\hat{\lambda}_{11} > \hat{\lambda}_{12} > \ldots > \hat{\lambda}_{1k}$, or $\sum_{i=1}^{m} \hat{\lambda}_{i1} > \sum_{i=1}^{m} \hat{\lambda}_{i2} > \ldots > \sum_{i=1}^{m} \hat{\lambda}_{ik}$. Preliminary knowledge of the data at hand also may allow for unambiguous identification of the $(k - q)$ eigenvectors involved in $\mathcal{H}_0$.

Flury (1984) is proposing a test of this null hypothesis, which is based on the Gaussian asymptotic distribution of the Gaussian maximum likelihood estimators $\hat{\lambda}_{ij}$, and is not robust against violations of strict Gaussian assumptions. Quite on the contrary, the pseudo-Gaussian procedure we are proposing here is valid under any possibly heterokurtic m-tuple of elliptical distributions with finite fourth-order moments, while remaining asymptotically equivalent to Flury's test under multinormal densities.

The paper is organized as follows. Section 2 describes the model and collects the assumptions needed throughout. The main result of the paper is in Section 3, where we derive (Theorem 3.1) the asymptotic joint distribution of the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\lambda}_{ij}$ of $\boldsymbol{\beta}$ and the λ_{ij}'s. In Section 4, we show how this result allows for constructing a pseudo-Gaussian test of the null hypothesis (1.2) which, contrary to Flury's Gaussian test, resists non-normality and heterokurticity. Section 5 is devoted to simulations. The appendix collects the proofs of technical results.

2 Main assumptions and notation.

2.1 Gaussian likelihood-based inference for Gaussian CPC models.

We start with a more precise statement of Flury's assumptions. 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 $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, m$. Flury (1984 and 1986) mainly deals with the Gaussian maximum likelihood estimators (MLEs) $\hat{\boldsymbol{\beta}} =: (\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_k)$ and $\hat{\lambda}_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, k$ of the common eigenvectors $\boldsymbol{\beta} =: (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)$ and the corresponding eigenvalues λ_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, k$ of $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m$ under the hypothesis, equivalent to (1.1),

$$\mathcal{H}_{\text{CPC}} : \boldsymbol{\Sigma}_i \boldsymbol{\beta}_j = \lambda_{ij} \boldsymbol{\beta}_j \quad i = 1, \ldots, m, \ j = 1, \ldots, k$$

of common principal components. Call this Assumption ($\mathrm{A}_{\mathcal{N}}$).

If the common principal components are to be identifiable, however, an assumption has to be made on eigenvalues.

Assumption (B). For all couple $j \neq l$, there is at least one population i such that $\lambda_{ij} \neq \lambda_{il}$.

Note that this Assumption (B) is implied by the slightly stronger assumption made by Flury, namely

Assumption (B′). For all $j = 1, \ldots, k$, there is at least one population in which the jth eigenvalue is distinct from all other ones—namely, there exists at least one i such that $\lambda_{il} \neq \lambda_{ij}$ for all $l \neq j$.

which in turn readily follows if we assume that

Assumption (B″). For all $i = 1, \ldots, m$, $\lambda_{i1} > \lambda_{i2} > \ldots > \lambda_{ik}$.

Denoting 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, the empirical mean vector and covariance matrix in population i, the MLEs $\hat{\boldsymbol{\beta}}$ and $\hat{\lambda}_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, k$, are solutions of the likelihood equations

$$\boldsymbol{\beta}_j' \Big(\sum_{i=1}^{m} n_i \frac{\lambda_{ij} - \lambda_{il}}{\lambda_{ij} \lambda_{il}} \mathbf{S}_i \Big) \boldsymbol{\beta}_l = 0, \quad j \neq l = 1, \ldots, k, \tag{2.1}$$

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

(see Flury 1984), where δ_{lj} stands for the usual Kronecker symbol. An explicit solution of the likelihood equations (2.1) does not exist, but an algorithm for solving them numerically has been proposed by Flury and Gautschi (1986).

Writing vec($\mathbf{A}$) (resp., dvec($\mathbf{A}$)) for the vector obtained by stacking the columns (resp., the diagonal elements) of a matrix $\mathbf{A}$, let $\mathbf{H}_k$ be 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 $\mathbf{A}$. Define diag$(\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_m)$ as the block-diagonal matrix with diagonal blocks $\mathbf{B}_1$, $\mathbf{B}_2, \ldots, \mathbf{B}_m$, and denote by $\mathbf{A}^{\otimes 2}$ the Kronecker product $\mathbf{A} \otimes \mathbf{A}$. Finally, denoting by $(\mathbf{e}_1, \ldots, \mathbf{e}_k)$ the canonical basis in $\mathbb{R}^k$, write $\mathbf{K}_k := \sum_{j,l=1}^k (\mathbf{e}_j\mathbf{e}_l') \otimes (\mathbf{e}_l\mathbf{e}_j')$ for the $k^2 \times k^2$ *commutation matrix* (see Magnus and Neudecker 1999).

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

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

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

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

Useless $^{(n)}$ supercripts however will be dropped for the sake of simplicity.

The asymptotic distribution—under the Gaussian assumptions just described—of the maximum likelihood estimates (the solutions $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Lambda}}_1, \ldots, \hat{\boldsymbol{\Lambda}}_k$ of (2.1)) then is given in the following theorem.

Theorem 2.1 (Flury 1986) *Assume that ($A_{\mathcal{N}}$), (B), and (C) hold. Then,*

$$\left(n^{1/2}(\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))', \; n_1^{1/2}(\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1))', \ldots, n_m^{1/2}(\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_m - \boldsymbol{\Lambda}_m))'\right)'$$

is asymptotically normal with mean $\mathbf{0}$ *and covariance matrix* diag$(\mathbf{V}_{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\Lambda}_1}, \ldots, \mathbf{V}_{\boldsymbol{\Lambda}_m})$, *where*

$$\mathbf{V}_{\boldsymbol{\beta}} := (\mathbf{I}_k \otimes \boldsymbol{\beta}) \left\{ \sum_{\substack{j,l=1 \\ j \neq l}}^{k} \nu_{jl} \left[(\mathbf{e}_j\mathbf{e}_j') \otimes (\mathbf{e}_l\mathbf{e}_l') - (\mathbf{e}_j\mathbf{e}_l') \otimes (\mathbf{e}_l\mathbf{e}_j') \right] \right\} (\mathbf{I}_k \otimes \boldsymbol{\beta}'),$$

with

$$\nu_{jl} := \left(\sum_{i=1}^{m} r_i \, \frac{(\lambda_{ij} - \lambda_{il})^2}{\lambda_{ij}\lambda_{il}} \right)^{-1}, \quad j \neq l = 1, \ldots, k, \tag{2.2}$$

and $\mathbf{V}_{\boldsymbol{\Lambda}_i} := \mathbf{H}_k(\mathbf{I}_{k^2} + \mathbf{K}_k)\boldsymbol{\Lambda}_i^{\otimes 2}\mathbf{H}_k'$.

Note that Assumption (B) ensures that the quantities ν_{jl} in (2.2) are finite. Flury's derivation of the $\boldsymbol{\Lambda}$-part of Theorem 2.1 however is based on an invalid application of Slutzky's classical argument. One of the goals of this paper is to give a rigorous proof of this result as a particular case of a more general one, which we establish under possibly heterokurtic elliptical densities with finite moments of order four.

2.2 Elliptical CPC: main assumptions.

In this section, we relax Flury's Gaussian Assumption ($\mathrm{A}_{\mathcal{N}}$) by assuming that all populations are elliptically symmetric, with unspecified and possibly distinct radial densities. More precisely, defining, for $r \geq 2$,

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

where $\mu_{\ell;h} := \int_0^\infty r^\ell h(r)\, dr$, we assume that the following elliptical CPC hypothesis holds.

Assumption ($\mathrm{A}_{\mathcal{E}}$). The mutually independent i.i.d. samples $(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})$, have probability densities

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

for some k-dimensional vectors $\boldsymbol{\theta}_i$ (*location*), some positive definite $(k \times k)$ covariance matrices $\boldsymbol{\Sigma}_i$ satisfying $\mathcal{H}_{\mathrm{CPC}}$ (1.1), and some f_i in the class $\mathcal{F}_1^4$ of *standardized radial densities* with finite fourth-order moments.

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

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

Under Assumption ($\mathrm{A}_{\mathcal{E}}$), 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. Since $\mathcal{F}_1^4 \subset \mathcal{F}_1^2$, the condition that $f_i \in \mathcal{F}_1^4$ implies that $\mathrm{E}[d_{ij}^2(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)] = k$—thus justifying the terminology *standardized radial density* for f_i. It also follows that $\boldsymbol{\Sigma}_i = \mathrm{Var}[\mathbf{X}_{ij}]$ under Assumption ($\mathrm{A}_{\mathcal{E}}$) 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 $f_i \in \mathcal{F}_1^4$.

Multiple-in-all-populations roots being excluded under Assumption (B), the common (under $\mathcal{H}_{\mathrm{CPC}}$) eigenvector matrix $\boldsymbol{\beta}$ is, up to irrelevant sign changes, a well identified element of the orthogonal group $\mathcal{O}(k)$, and hence contains $s := k(k-1)/2$ functionally independent parameters. These parameters can be stored in a vector $\boldsymbol{\beta}^\star$ in such a way that

$$n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{G}_k n^{1/2}(\hat{\boldsymbol{\beta}}^\star - \boldsymbol{\beta}^\star) + o_{\mathrm{P}}(1), \tag{2.3}$$

with $\mathbf{G}_k = (\mathbf{G}_{k;12}\ \mathbf{G}_{k;13} \ldots \mathbf{G}_{k;(k-1)k})$ and $\mathbf{G}_{k;jl} := (\mathbf{e}_j \otimes \boldsymbol{\beta}_l - \mathbf{e}_l \otimes \boldsymbol{\beta}_j)/\sqrt{2}$ (see Flury 1986 for details). Note that the matrix $\mathbf{G}_k$ is composed of s orthonormal column vectors.

The vector

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

contains the parameters of the model. For any $\underline{f} = (f_1, \ldots, f_m) \in (\mathcal{F}_1^4)^m$, we denote by $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$ the joint distribution of the observations under 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.

For any $\underline{f} = (f_1, \ldots, f_m) \in (\mathcal{F}_1^4)^m$ (which implies finite fourth-order moments), let $E_k(f_i) := \mathrm{E}_{\boldsymbol{\vartheta};\underline{f}}[d_{ij}^4(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)] = \int_0^1 (\tilde{F}_{ik}^{-1}(u))^4 du$. 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 the Gaussian density ϕ, $E_k(\phi) = k(k+2)$, so that $\kappa_k(\phi) = 0$.

3 Asymptotic distribution of $\hat{\boldsymbol{\beta}}$ and the $\hat{\lambda}_{ij}$'s under elliptical CPC.

As mentioned before, we plan to extend Theorem 2.1 by deriving the asymptotic distribution of $(n^{1/2}(\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))',\ n_1^{1/2}(\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1))', \ldots, n_m^{1/2}(\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_m - \boldsymbol{\Lambda}_m))')'$ under any m-tuple of elliptical densities satisfying the elliptical CPC assumption ($\mathrm{A}_{\mathcal{E}}$), thereby also providing a valid proof of the Gaussian result.

First write $\mathbf{R} := \mathrm{diag}(r_1, \ldots, r_m)$ and define the $s \times (mk^2)$ matrix $\tilde{\boldsymbol{\beta}}_k := (\tilde{\boldsymbol{\beta}}_k^{(1)}, \ldots, \tilde{\boldsymbol{\beta}}_k^{(m)})$, where $\tilde{\boldsymbol{\beta}}_k^{(i)} := (\tilde{\boldsymbol{\beta}}_{k;12}^{(i)}, \tilde{\boldsymbol{\beta}}_{k;13}^{(i)}, \ldots, \tilde{\boldsymbol{\beta}}_{k;(k-1)k}^{(i)})'$, with $\tilde{\boldsymbol{\beta}}_{k;jl}^{(i)} := \nu_{jl}(\lambda_{il}^{-1} - \lambda_{ij}^{-1})(\boldsymbol{\beta}_l \otimes \boldsymbol{\beta}_j)$. Also let

$$\mathbf{T}^{(n)} := (n_1^{1/2}(\mathrm{vec}\,(\mathbf{S}_1 - \boldsymbol{\Sigma}_1))', \ldots, n_m^{1/2}(\mathrm{vec}\,(\mathbf{S}_m - \boldsymbol{\Sigma}_m))')'. \tag{3.4}$$

We then have the following key result (see the appendix for a proof).

Proposition 3.1 *Let Assumptions ($\mathrm{A}_{\mathcal{E}}$), (B), and (C) hold. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, as $n \to \infty$,*

$$(i) \quad n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{2}\mathbf{G}_k\tilde{\boldsymbol{\beta}}_k(\mathbf{R}^{1/2} \otimes \mathbf{I}_{k^2})\mathbf{T}^{(n)} + o_{\mathrm{P}}(1)$$

and

$$(ii) \quad n_i^{1/2}\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_i - \boldsymbol{\Lambda}_i) = \mathbf{H}_k\big[(\boldsymbol{\beta}'^{\otimes 2})\,(\mathbf{e}_i^{(m)\prime} \otimes \mathbf{I}_{k^2}) + \sqrt{2r_i}\,(\mathbf{I}_{k^2} + \mathbf{K}_k)(\mathbf{I}_k \otimes \boldsymbol{\beta}'\boldsymbol{\Sigma}_i)\mathbf{G}_k\tilde{\boldsymbol{\beta}}_k(\mathbf{R}^{1/2} \otimes \mathbf{I}_{k^2})\big]\mathbf{T}^{(n)} + o_{\mathrm{P}}(1),$$

where $\mathbf{e}_i^{(m)}$ denotes the ith vector of the canonical basis of $\mathbb{R}^m$.

The asymptotic joint multinormality of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Lambda}}_i$, $i = 1, \ldots, m$ under any (possibly heterokurtic) elliptical densities follows trivially from Proposition 3.1 and the asymptotic multinormality of $\mathbf{T}^{(n)}$ under such densities (see Lemma A.1). Of course, the derivation of the corresponding asymptotic covariance matrix requires tedious (yet simple) algebraic computations.

Theorem 3.1 *Assume that ($\mathrm{A}_{\mathcal{E}}$), (B), and (C) hold. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$,*

$$\Big(n^{1/2}(\mathrm{vec}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))',\ n_1^{1/2}(\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1))', \ldots, n_m^{1/2}(\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_m - \boldsymbol{\Lambda}_m))'\Big)'$$

is asymptotically normal with mean $\mathbf{0}$ *and covariance matrix* $\mathrm{diag}(\mathbf{V}_{\boldsymbol{\beta}}(\underline{f}), \mathbf{V}_{\boldsymbol{\Lambda}_1}(\underline{f}), \ldots, \mathbf{V}_{\boldsymbol{\Lambda}_m}(\underline{f}))$, *where*

$$\mathbf{V}_{\boldsymbol{\beta}}(\underline{f}) := (\mathbf{I}_k \otimes \boldsymbol{\beta})\left\{\sum_{\substack{j,l=1 \\ j \neq l}}^{k} \frac{\nu_{jl}^2}{\nu_{jl}(\underline{f})}\left[(\mathbf{e}_j\mathbf{e}_j') \otimes (\mathbf{e}_l\mathbf{e}_l') - (\mathbf{e}_j\mathbf{e}_l') \otimes (\mathbf{e}_l\mathbf{e}_j')\right]\right\}(\mathbf{I}_k \otimes \boldsymbol{\beta}'), \tag{3.5}$$

with

$$\nu_{jl}(\underline{f}) := \left(\sum_{i=1}^{m} r_i (1+\kappa_k(f_i)) \frac{(\lambda_{ij}-\lambda_{il})^2}{\lambda_{ij}\lambda_{il}} \right)^{-1}, \quad j \neq l = 1, \ldots, k,$$

and $\mathbf{V}_{\mathbf{\Lambda}_i}(\underline{f}) := \mathbf{H}_k\big[(1+\kappa_k(f_i))(\mathbf{I}_{k^2}+\mathbf{K}_k)\mathbf{\Lambda}_i^{\otimes 2} + \kappa_k(f_i)\mathrm{vec}(\mathbf{\Lambda}_i)(\mathrm{vec}(\mathbf{\Lambda}_i))'\big]\mathbf{H}_k'$, $i = 1, \ldots, m$.

At the multinormal ($\underline{f} = \underline{\phi} := (\phi, \ldots, \phi)$), $\kappa(\phi) = 0$ (hence we retrieve $\nu_{jl}(\underline{\phi}) = \nu_{jl}$ given in (2.2)), so that $\mathbf{V}_{\boldsymbol{\beta}}(\underline{\phi}) = \mathbf{V}_{\boldsymbol{\beta}}$ and $\mathbf{V}_{\mathbf{\Lambda}_i}(\underline{\phi}) = \mathbf{V}_{\mathbf{\Lambda}_i}$, $i = 1, \ldots, m$. Theorem 3.1 thus extends to the (possibly heterokurtic) elliptical case the Gaussian result of Theorem 2.1.

4 A pseudo-Gaussian test for $\mathcal{H}_0$.

In this section, we show how Theorem 3.1 can be used to build a pseudo-Gaussian test for the null hypothesis $\mathcal{H}_0$ in (1.2). In view of (1.3), it is quite natural to base the test on the statistics

$$Y_i^{(n)} := n_i^{1/2}\left[(1-p_0)\sum_{j=q+1}^{k}\hat{\lambda}_{ij} - p_0\sum_{j=1}^{q}\hat{\lambda}_{ij}\right], \quad i = 1, \ldots, m$$

(after due standardization). Letting $Y_{i;0}^{(n)} := Y_i^{(n)} - n_i^{1/2}\big[(1-p_0)\sum_{j=q+1}^{k}\lambda_{ij} - p_0\sum_{j=1}^{q}\lambda_{ij}\big]$, $i = 1, \ldots, m$, the following key result easily follows from Theorem 3.1 (see the appendix for the proof).

Lemma 4.1 *Assume that* ($A_{\mathcal{E}}$), (B), *and* (C) *hold. Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, $(Y_{1;0}^{(n)}, \ldots, Y_{m;0}^{(n)})'$ *is asymptotically normal with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{V}(\underline{f}) := \mathrm{diag}(V_1(\underline{f}), \ldots, V_m(\underline{f}))$, *where*

$$V_i(\underline{f}) = (2+3\kappa_k(f_i))\left(p_0^2\sum_{j=1}^{q}\lambda_{ij}^2 + (1-p_0)^2\sum_{j=q+1}^{k}\lambda_{ij}^2\right) + \kappa_k(f_i)\left(p_0^2\sum_{\substack{j,l=1\\ j\neq l}}^{q}\lambda_{ij}\lambda_{il} - p_0(1-p_0)\sum_{j=q+1}^{k}\sum_{l=1}^{q}\lambda_{ij}\lambda_{il}\right). \tag{4.6}$$

At the multinormal, the asymptotic variances (4.6) of the $Y_i^{(n)}$'s reduce to

$$V_i(\underline{\phi}) := 2p_0^2\sum_{j=1}^{q}\lambda_{ij}^2 + 2(1-p_0)^2\sum_{j=q+1}^{k}\lambda_{ij}^2.$$

Replacing the λ_{ij}'s with their Gaussian maximum likelihood estimates $\hat{\lambda}_{ij}$ yields an estimator $\hat{V}_i^{(n)}$ of $V_i(\underline{\phi})$ which is consistent under Gaussian densities. Flury's test $\phi^{(n)}_{\mathrm{Flury}}$ (Flury 1986) then is a multiple test for location, performed on the studentized $Y_i^{(n)}$'s, and rejects $\mathcal{H}_0$ (at asymptotic level α) as soon as

$$T^{(n)}_{\mathrm{Flury}} := \max_{i=1,\ldots,m}\left(Y_i^{(n)}/(\hat{V}_i^{(n)})^{1/2}\right) > \Phi^{-1}((1-\alpha)^{1/m}), \tag{4.7}$$

where Φ denotes the standard normal distribution function.

That $\phi^{(n)}_{\text{Flury}}$ is invalid under non-Gaussian densities (more precisely, under densities with non-Gaussian kurtoses) readily follows from Lemma 4.1: the studentization of $Y_i^{(n)}$ by $\hat{V}_i^{(n)}$ indeed is correct under Gaussian kurtoses only. More interestingly, Lemma 4.1 also clearly shows how $\phi^{(n)}_{\text{Flury}}$ should be modified in order to retain its validity under non-Gaussian densities: instead of studentizing $Y_i^{(n)}$ with $\hat{V}_i^{(n)}$ based on the Gaussian variance $V_i(\underline{\phi})$, one simply should use the estimator

$$\hat{V}_{i\dagger}^{(n)} := (2+3\hat{\kappa}_k^{(i)})\left(p_0^2\sum_{j=1}^{q}\hat{\lambda}_{ij}^2 + (1-p_0)^2\sum_{j=q+1}^{k}\hat{\lambda}_{ij}^2\right) \qquad (4.8)$$
$$+\hat{\kappa}_k^{(i)}\left(p_0^2\sum_{\substack{j,l=1\\ j\neq l}}^{q}\hat{\lambda}_{ij}\hat{\lambda}_{il} - p_0(1-p_0)\sum_{j=q+1}^{k}\sum_{l=1}^{q}\hat{\lambda}_{ij}\hat{\lambda}_{il}\right),$$

based on the general variance (4.6) given in Lemma 4.1, where $\hat{\kappa}_k^{(i)}$ denotes a consistent (under any $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, with $\boldsymbol{\vartheta}\in\mathcal{H}_{\text{CPC}}$ and $\underline{f}\in(\mathcal{F}_1^4)^m$) estimator of the kurtosis $\kappa_k(f_i)$—an obvious choice being

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

The resulting pseudo-Gaussian test ($\phi_\dagger^{(n)}$, say) rejects $\mathcal{H}_0$ (at asymptotic level α) whenever

$$T_\dagger^{(n)} := \max_{i=1,\ldots,m}\left(Y_i^{(n)}/(\hat{V}_{i\dagger}^{(n)})^{1/2}\right) > \Phi^{-1}\left((1-\alpha)^{1/m}\right).$$

Summing up, we have established the following result, which summarizes the properties of our pseudo-Gaussian test $\phi_\dagger^{(n)}$.

Proposition 4.1 *Assume that ($A_{\mathcal{E}}$), (B), and (C) hold. Then,*

(i) under $\bigcup_{\boldsymbol{\vartheta}\in\mathcal{H}_0}\bigcup_{\underline{f}\in(\mathcal{F}_1^4)^m}\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, *the sequence of tests* $\phi_\dagger^{(n)}$ *has asymptotic level* α*;*

(ii) under $\bigcup_{\boldsymbol{\vartheta}\in\mathcal{H}_0}\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{\phi}}$, $T_\dagger^{(n)} = T^{(n)}_{\text{Flury}} + o_{\mathrm{P}}(1)$, *as* $n\to\infty$.

Part (i) of the Proposition confirms the (asymptotic) validity of $\phi_\dagger^{(n)}$ under any m-tuple of elliptical distributions with finite fourth-order moments. As for Part (ii), it states that $\phi_\dagger^{(n)}$ and $\phi^{(n)}_{\text{Flury}}$ are asymptotically equivalent in the Gaussian case.

The form of the test statistic $T_\dagger^{(n)}$ in Proposition 4.1, as well as that of Flury's $T^{(n)}_{\text{Flury}}$, calls for some comments. Both test statistics indeed follow from considering an m-tuple of (asymptotically) standard normal variables (the $Y_i^{(n)}/(\hat{V}_{i\dagger}^{(n)})^{1/2}$'s for $\phi_\dagger^{(n)}$, the $Y_i^{(n)}/(\hat{V}_i^{(n)})^{1/2}$'s for $\phi^{(n)}_{\text{Flury}}$) having negative means iff the null hypothesis $\mathcal{H}_0$ holds. From a decisional point of view, this reduces the original problem (involving eigenvalues) to the one-sided problem

$$\mathcal{H}_0 : \mu_i \leq 0,\ i=1,\ldots,m \quad \text{against} \quad \mathcal{H}_1 : \mu_i > 0 \quad \text{for at least one } i$$

on the means $\mu_1,\ldots,\mu_m$ of an m-tuple of mutually independent normal observations $\eta_1,\ldots,\eta_m$, say, all with variance one. This one-sided testing problem involving a cone-shaped null hypothesis on a multidimensional parameter is a classical one which, despite a long history, has not found any fully satisfactory solution yet: see Section 4.1 of Akharif and Hallin (2003) for a brief survey and references. The solution we are adopting here is based on a test statistic of the form $\max_{i=1,\ldots,m} \eta_i$. The resulting test has size α at $(\mu_1, \ldots, \mu_m) = (0, \ldots, 0)$, which is the least favorable point in $\mathcal{H}_0$, and size strictly less than α everywhere else in $\mathcal{H}_0$. In particular, its size is less than α at boundary points of the form $(0, \ldots, 0, \mu_{i_0}, 0, \ldots, 0)$, with the consequence that it is non-similar and inherently biased. Other solutions have been proposed in the literature (such as Schaafsma and Smid (1966)'s *most stringent somewhere most powerful tests*); but none of them is able to produce power functions that remain constant over the surface of a cone-shaped null hypothesis—which implies that they all suffer the same non-similarity and biasedness problems as $\phi_\dagger^{(n)}$ and $\phi_{\text{Flury}}^{(n)}$.

5 Simulations

In this section, we are comparing Flury's Gaussian test $\phi_{\text{Flury}}^{(n)}$ with the pseudo-Gaussian test $\phi_\dagger^{(n)}$ proposed in Section 4. In order to do so, we conducted three distinct simulation studies. In each of them, we generated $N = 10,000$ independent replications of three pairs ($m = 2$) of mutually independent samples of bivariate ($k = 2$) random vectors

$$\boldsymbol{\varepsilon}_{\ell;1j} \quad \text{and} \quad \boldsymbol{\varepsilon}_{\ell;2j}, \quad \ell = 1,\, 2,\, 3, \quad j = 1, \ldots, n_1 = n_2 = 100,$$

with—in each case, zero-mean unit-covariance—(a) Gaussian densities ($\boldsymbol{\varepsilon}_{1;1j}$ and $\boldsymbol{\varepsilon}_{1;2j}$: Gaussian case), (b) t_5 densities ($\boldsymbol{\varepsilon}_{2;1j}$ and $\boldsymbol{\varepsilon}_{2;2j}$: non-Gaussian homokurtic case), (c) Gaussian ($\boldsymbol{\varepsilon}_{3;1j}$) and t_5 ($\boldsymbol{\varepsilon}_{3;2j}$) densities (heterokurtic case), respectively. Each replication of the $\boldsymbol{\varepsilon}_{\ell;1j}$'s was transformed into

$$\mathbf{X}_{\ell;1j} = \boldsymbol{\beta}\boldsymbol{\Lambda}_1^{1/2}\boldsymbol{\varepsilon}_{\ell;1j}, \quad \ell = 1,\, 2,\, 3, \quad j = 1, \ldots, n_1, \tag{5.9}$$

with

$$\boldsymbol{\beta} = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Lambda}_1 = \begin{pmatrix} 4 & 0 \\ 0 & \lambda_{12} \end{pmatrix},$$

while each replication of the $\boldsymbol{\varepsilon}_{\ell;2j}$'s was subjected to a sequence of 21 linear transformations of the form

$$\mathbf{X}_{\ell;2j;\xi} = \boldsymbol{\beta}\boldsymbol{\Lambda}_{2;\xi}^{1/2}\boldsymbol{\varepsilon}_{\ell;2j}, \quad \ell = 1,\, 2,\, 3, \quad j = 1, \ldots, n_2, \quad \xi = 0, 1, 2, \ldots, 20, \tag{5.10}$$

with

$$\boldsymbol{\Lambda}_{2;\xi} = \begin{pmatrix} 9 - \frac{\xi}{3} & 0 \\ 0 & 2 \end{pmatrix}.$$

Clearly, $\mathbf{X}_{\ell;1j}$ and $\mathbf{X}_{\ell;2j;\xi}$ have common eigenvectors $\boldsymbol{\beta}$, but distinct eigenvalue matrices $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_{2;\xi}$. The null hypothesis considered throughout is $\mathcal{H}_0 : \max_{i=1,2} \left(\lambda_{i2}/(\lambda_{i1} + \lambda_{i2})\right) \leq 2/9$ ($q = 1$).

Experiment 1. The first simulation experiment was performed with $\lambda_{12} = 8/7$. The proportion of variability associated with $\mathbf{X}_{\ell;1j}$'s second eigenvector is $2/9$; the same proportion, for $\mathbf{X}_{\ell;2j;\xi}$'s second eigenvector, is less than $2/9$ for $\xi = 0, 1, \ldots, 5$, equal to $2/9$ for $\xi = 6$, and larger than $2/9$ for $\xi = 7, 8, \ldots, 20$, thus characterizing alternatives that are increasingly distant from the null, while $\xi = 6$ yields the least favorable point in $\mathcal{H}_0$. Rejection frequencies (at nominal 5% level) are reported in the three graphs of Figure 1 for densities (a), (b), and (c), respectively. Inspection of these graphs confirms the theoretical results of the previous section:

(i) under Gaussian densities (a), the performances of $\phi_{\dagger}^{(n)}$ roughly coincide with those of $\phi_{\rm Flury}^{(n)}$, with a slight advantage for the latter (not surprisingly, the Gaussian test is slightly more powerful, under Gaussian densities and for finite sample sizes, than the asymptotically equivalent pseudo-Gaussian one);

(ii) the pseudo-Gaussian procedure $\phi_{\dagger}^{(n)}$, contrary to the Gaussian one $\phi_{\rm Flury}^{(n)}$ which severely overrejects, remains valid in (b) the non-Gaussian homokurtic as well as in (c) the heterokurtic case.

Note that $\phi_{\dagger}^{(n)}$, in accordance with the closing comments of Section 4, has size α at the least favorable point $\xi = 6$.

Experiment 2. The second experiment was conducted with $\lambda_{12} = 9/8$. The proportion of variability associated with $\mathbf{X}_{\ell;1j}$'s second eigenvector now is $9/41 < 2/9$; the $\mathbf{X}_{\ell;2j;\xi}$'s are the same as in Experiment 1. Here, $\xi = 6$ yields a boundary point of $\mathcal{H}_0$, but not the least favorable one. Rejection frequencies (at nominal 5% level) are reported in the three graphs of Figure 2. Our test $\phi_{\dagger}^{(n)}$, for the reasons explained in Section 4, is slightly biased, but Flury's $\phi_{\rm Flury}^{(n)}$ is still invalid under non-Gaussian densities (b) and (c).

Experiment 3. The third experiment was conducted with $\lambda_{12} = 1$. The proportion of variability associated with $\mathbf{X}_{\ell;1j}$'s second eigenvector is $1/5 < 9/41 < 2/9$; the $\mathbf{X}_{\ell;2j;\xi}$'s are the same as in Experiments 1 and 2. Again, $\xi = 6$ here yields a boundary point of $\mathcal{H}_0$ which is not the least favorable one. Rejection frequencies (still at nominal 5% level) are reported in the three graphs of Figure 3. Our test $\phi_{\dagger}^{(n)}$ is more severely biased than it was under Experiment 2, while the bias of Flury's $\phi_{\rm Flury}^{(n)}$ under heterokurtic non-Gaussian densities (c) compensates overrejection; $\phi_{\rm Flury}^{(n)}$ nevertheless fails to be valid, as shown in Experiments 1 and 2.

A Appendix.

We will need the following lemma, which, in view of the mutual independence of the m populations, readily follows from the well-known result on the asymptotic normality of the empirical covariance matrix under elliptical densities with finite fourth-order moments (see, e.g., Bilodeau and Brenner 1999, Page 212).

Lemma A.1 *Assume that* ($A_{\mathcal{E}}$) *and (C) hold. Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, $\mathbf{T}^{(n)}$ *in (3.4) is asymptotically normal with mean* $\mathbf{0}$ *and covariance matrix* $\mathrm{diag}(\mathbf{B}_1,\ldots,\mathbf{B}_m)$, *where* $\mathbf{B}_i := (1+\kappa_k(f_i))(\mathbf{I}_{k^2} + \mathbf{K}_k)\boldsymbol{\Sigma}_i^{\otimes 2} + \kappa_k(f_i)\mathrm{vec}(\boldsymbol{\Sigma}_i)(\mathrm{vec}(\boldsymbol{\Sigma}_i))'$.

Proof of Proposition 3.1. In this proof, all $o_{\mathrm{P}}(1)$ and $o(1)$ quantities, all stochastic convergences (as $n \to \infty$) and all expectations are to be taken under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\underline{f}}$, for some fixed $\underline{f} \in (\mathcal{F}_1^4)^m$.

(i) Letting $\mathbf{D}_1^{(n)} := \sqrt{2}\mathbf{G}_k\tilde{\boldsymbol{\beta}}_k(\mathbf{R}^{1/2}\otimes\mathbf{I}_{k^2})\mathbf{T}^{(n)}$ and $\mathbf{D}_2^{(n)} := n^{1/2}\mathrm{vec}\,(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$, we prove the result by showing that, as $n \to \infty$,

$$
\begin{aligned}
&\text{(a)} \quad \mathrm{E}[\mathbf{D}_1^{(n)}\mathbf{D}_1^{(n)\prime}] &= \mathbf{V}_{\boldsymbol{\beta}}(\underline{f}) + o(1),\\
&\text{(b)} \quad \mathrm{E}[\mathbf{D}_2^{(n)}\mathbf{D}_1^{(n)\prime}] &= \mathbf{V}_{\boldsymbol{\beta}}(\underline{f}) + o(1), \quad \text{and}\\
&\text{(c)} \quad \mathrm{E}[\mathbf{D}_2^{(n)}\mathbf{D}_2^{(n)\prime}] &= \mathbf{V}_{\boldsymbol{\beta}}(\underline{f}) + o(1),
\end{aligned}
$$

where $\mathbf{V}_{\boldsymbol{\beta}}(\underline{f})$ is given in (3.5); note indeed that if (a), (b), and (c) hold,

$$\mathrm{E}[(\mathbf{D}_2^{(n)} - \mathbf{D}_1^{(n)})(\mathbf{D}_2^{(n)} - \mathbf{D}_1^{(n)})'] = o(1),$$

so that $\mathbf{D}_2^{(n)} - \mathbf{D}_1^{(n)}$ asymptotically vanishes in quadratic mean, hence also in probability.

(a) This follows trivially from Lemma A.1 via painful but quite straightforward algebra, once it is noticed that $\mathbf{V}_{\boldsymbol{\beta}}(\underline{f}) = 2\mathbf{G}_k \,\mathrm{diag}(\nu_{12}^2/\nu_{12}(\underline{f}), \ldots, \nu_{(k-1)k}^2/\nu_{(k-1)k}(\underline{f}))\mathbf{G}_k'$.

(b) From (2.1), we have

$$\mathrm{vec}\left[\hat{\boldsymbol{\beta}}_j' \left(\sum_{i=1}^m n_i(\frac{1}{\hat{\lambda}_{il}} - \frac{1}{\hat{\lambda}_{ij}})(\mathbf{S}_i - \boldsymbol{\Sigma}_i)\right)\hat{\boldsymbol{\beta}}_l\right] = -\mathrm{vec}\left[\hat{\boldsymbol{\beta}}_j' \left(\sum_{i=1}^m n_i(\frac{1}{\hat{\lambda}_{il}} - \frac{1}{\hat{\lambda}_{ij}})\boldsymbol{\Sigma}_i\right)\hat{\boldsymbol{\beta}}_l\right],\ l \neq j = 1, \ldots, k. \tag{A.1}$$

Postmultiplying both sides by $(\mathrm{vec}(\mathbf{S}_h - \boldsymbol{\Sigma}_h))'$, we obtain

$$(\hat{\boldsymbol{\beta}}_l' \otimes \hat{\boldsymbol{\beta}}_j') \sum_{i=1}^m \left(\frac{r_i^{(n)}}{r_h^{(n)}}\right)^{1/2} (\frac{1}{\hat{\lambda}_{il}} - \frac{1}{\hat{\lambda}_{ij}})\, n_i^{1/2}\mathrm{vec}(\mathbf{S}_i - \boldsymbol{\Sigma}_i) n_h^{1/2}(\mathrm{vec}(\mathbf{S}_h - \boldsymbol{\Sigma}_h))'$$
$$= -\sum_{i=1}^m \frac{r_i^{(n)} n^{1/2}}{(r_h^{(n)})^{1/2}}(\frac{1}{\hat{\lambda}_{il}} - \frac{1}{\hat{\lambda}_{ij}})\hat{\boldsymbol{\beta}}_j'\boldsymbol{\Sigma}_i\hat{\boldsymbol{\beta}}_l(n_h^{1/2}\mathrm{vec}(\mathbf{S}_h - \boldsymbol{\Sigma}_h))'.$$

Decomposing, in the right hand side, $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\beta}}_l$ into $\hat{\boldsymbol{\beta}}_j = \boldsymbol{\beta}_j + (\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)$ and $\hat{\boldsymbol{\beta}}_l = \boldsymbol{\beta}_l + (\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l)$, respectively, and using Slutzky's lemma yields

$$(\boldsymbol{\beta}_l' \otimes \boldsymbol{\beta}_j') \sum_{i=1}^m \left(\frac{r_i^{(n)}}{r_h^{(n)}}\right)^{1/2} (\frac{1}{\lambda_{il}} - \frac{1}{\lambda_{ij}})\, n_i^{1/2}\mathrm{vec}(\mathbf{S}_i - \boldsymbol{\Sigma}_i) n_h^{1/2}(\mathrm{vec}(\mathbf{S}_h - \boldsymbol{\Sigma}_h))'$$
$$= -\sum_{i=1}^m \frac{r_i^{(n)} n^{1/2}}{(r_h^{(n)})^{1/2}}(\frac{1}{\lambda_{il}} - \frac{1}{\lambda_{ij}})(\boldsymbol{\beta}_j'\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l) + \boldsymbol{\beta}_l'\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j))(n_h^{1/2}\mathrm{vec}(\mathbf{S}_h - \boldsymbol{\Sigma}_h))' + o_{\mathrm{P}}(1).$$

Taking expectations and limits as $n \to \infty$, we obtain

$$(\frac{1}{\lambda_{hl}} - \frac{1}{\lambda_{hj}})(\boldsymbol{\beta}_l' \otimes \boldsymbol{\beta}_j')\left[(1 + \kappa_k(f_h))(\mathbf{I}_{k^2} + \mathbf{K}_k)\boldsymbol{\Sigma}_h^{\otimes 2} + \kappa_k(f_h)\mathrm{vec}(\boldsymbol{\Sigma}_h)(\mathrm{vec}(\boldsymbol{\Sigma}_h))'\right] \tag{A.2}$$
$$= -\left[\sum_{i=1}^m r_i(\frac{1}{\lambda_{il}} - \frac{1}{\lambda_{ij}})\left(\lambda_{il}\boldsymbol{\beta}_l' \vdots \lambda_{ij}\boldsymbol{\beta}_j'\right)\right] \lim_{n\to\infty} \mathrm{E}\left[n^{1/2}\begin{pmatrix} \hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j \\ \hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l \end{pmatrix}(n^{1/2}\mathrm{vec}(\mathbf{S}_h - \boldsymbol{\Sigma}_h))'\right].$$

Since $(\boldsymbol{\beta}_l' \otimes \boldsymbol{\beta}_j')\mathrm{vec}(\boldsymbol{\Sigma}_h) = 0$ for $j \neq l$, (A.2) takes the form

$$(1 + \kappa_k(f_h))(\frac{1}{\lambda_{hl}} - \frac{1}{\lambda_{hj}})(\boldsymbol{\beta}_l' \otimes \boldsymbol{\beta}_j')(\mathbf{I}_{k^2} + \mathbf{K}_k)\boldsymbol{\Sigma}_h^{\otimes 2} \tag{A.3}$$
$$= -\left[\sum_{i=1}^m r_i(\frac{1}{\lambda_{il}} - \frac{1}{\lambda_{ij}})\left(\lambda_{il}\boldsymbol{\beta}_l' \vdots \lambda_{ij}\boldsymbol{\beta}_j'\right)\right] \lim_{n\to\infty} \mathrm{E}\left[n^{1/2}\begin{pmatrix} \hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j \\ \hat{\boldsymbol{\beta}}_l - \boldsymbol{\beta}_l \end{pmatrix}(n^{1/2}\mathrm{vec}(\mathbf{S}_h - \boldsymbol{\Sigma}_h))'\right],$$

which provides an equation for all distinct pairs of eigenvectors. Solving the resulting system of s equations with respect to the s functionally independent parameters characterizing $\boldsymbol{\beta}$ and transforming back by means of the matrix $\mathbf{G}_k$ (see (2.3)), we obtain that

$$\lim_{n\to\infty} \mathrm{E}\Big[n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(n^{1/2}\mathrm{vec}(\mathbf{S}_h-\boldsymbol{\Sigma}_h))'\Big] = (1+\kappa_k(f_h))\sqrt{2}\mathbf{G}_k\tilde{\boldsymbol{\beta}}_k^{(i)}(\mathbf{I}_{k^2}+\mathbf{K}_k)\boldsymbol{\Sigma}_h^{\otimes 2}, \quad \text{(A.4)}$$

which, in view of the definition of $\mathbf{D}_1^{(n)}$, yields

$$\lim_{n\to\infty} \mathrm{E}\Big[\mathbf{D}_2^{(n)}\mathbf{D}_1^{(n)\prime}\Big] = 2\mathbf{G}_k\sum_{i=1}^m(1+\kappa_k(f_i))r_i\tilde{\boldsymbol{\beta}}_k^{(i)}(\mathbf{I}_{k^2}+\mathbf{K}_k)\boldsymbol{\Sigma}_i^{\otimes 2}(\tilde{\boldsymbol{\beta}}_k^{(i)})'\mathbf{G}_k' = \mathbf{V}_{\boldsymbol{\beta}}(\underline{f}).$$

Part (b) of the proof follows.

(c) Postmultiplying both side of (A.1) by $((\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j)'\vdots(\hat{\boldsymbol{\beta}}_l-\boldsymbol{\beta}_l)')$ and using Slutzky's lemma again, we obtain

$$(\boldsymbol{\beta}_l'\otimes\boldsymbol{\beta}_j')\sum_{i=1}^m r_i^{(n)}(\frac{1}{\lambda_{il}}-\frac{1}{\lambda_{ij}})\,n^{1/2}\mathrm{vec}(\mathbf{S}_i-\boldsymbol{\Sigma}_i)\;n^{1/2}\begin{pmatrix}\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\\ \hat{\boldsymbol{\beta}}_l-\boldsymbol{\beta}_l\end{pmatrix}'$$

$$=-\sum_{i=1}^m r_i^{(n)}n^{1/2}(\frac{1}{\lambda_{il}}-\frac{1}{\lambda_{ij}})(\boldsymbol{\beta}_j'\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\beta}}_l-\boldsymbol{\beta}_l)+\boldsymbol{\beta}_l'\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j))n^{1/2}\begin{pmatrix}\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\\ \hat{\boldsymbol{\beta}}_l-\boldsymbol{\beta}_l\end{pmatrix}'+o_{\mathrm{P}}(1).$$

Proceeding along the same lines as above and using (A.4), we obtain, after some lengthy computations,

$$\lim_{n\to\infty} \mathrm{E}\Big[\mathbf{D}_2^{(n)}\mathbf{D}_2^{(n)\prime}\Big] = 2\mathbf{G}_k\sum_{i=1}^m(1+\kappa_k(f_i))r_i\tilde{\boldsymbol{\beta}}_k^{(i)}(\mathbf{I}_{k^2}+\mathbf{K}_k)\boldsymbol{\Sigma}_i^{\otimes 2}(\tilde{\boldsymbol{\beta}}_k^{(i)})'\mathbf{G}_k' = \mathbf{V}_{\boldsymbol{\beta}}(\underline{f}).$$

Part (i) of the lemma thus follows.

(ii) The likelihood equations (2.1) and the definition of $\mathbf{H}_k$ entail that

$$\begin{aligned} n_i^{1/2}\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_i-\boldsymbol{\Lambda}_i) &= n_i^{1/2}\mathrm{dvec}\,(\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}'\boldsymbol{\Sigma}_i\boldsymbol{\beta})\\ &= n_i^{1/2}\mathbf{H}_k\mathrm{vec}\,(\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}'\boldsymbol{\Sigma}_i\boldsymbol{\beta})\\ &= n_i^{1/2}\mathbf{H}_k\big[(\hat{\boldsymbol{\beta}}'^{\otimes 2})\,\mathrm{vec}(\mathbf{S}_i-\boldsymbol{\Sigma}_i)+(\hat{\boldsymbol{\beta}}'^{\otimes 2}-\boldsymbol{\beta}'^{\otimes 2})\,\mathrm{vec}(\boldsymbol{\Sigma}_i)\big]\\ &=: \mathbf{E}_1^{(n)}+\mathbf{E}_2^{(n)}.\end{aligned}$$

Applying Slutzky's lemma again yields

$$\mathbf{E}_1^{(n)} = \mathbf{H}_k(\boldsymbol{\beta}'^{\otimes 2})\,n_i^{1/2}\mathrm{vec}(\mathbf{S}_i-\boldsymbol{\Sigma}_i)+o_{\mathrm{P}}(1)$$

and

$$\begin{aligned}\mathbf{E}_2^{(n)} &= n_i^{1/2}\big[(\mathbf{I}_k\otimes\boldsymbol{\beta}'\boldsymbol{\Sigma}_i)\mathrm{vec}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+(\mathbf{I}_k\otimes(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{\Sigma}_i)\mathrm{vec}(\boldsymbol{\beta})\big]+o_{\mathrm{P}}(1)\\ &= r_i^{1/2}(\mathbf{I}_{k^2}+\mathbf{K}_k)(\mathbf{I}_k\otimes\boldsymbol{\beta}'\boldsymbol{\Sigma}_i)\,n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+o_{\mathrm{P}}(1).\end{aligned} \quad \text{(A.5)}$$

Hence, we obtain that

$$n_i^{1/2}\mathrm{dvec}\,(\hat{\boldsymbol{\Lambda}}_i-\boldsymbol{\Lambda}_i) = \mathbf{H}_k(\boldsymbol{\beta}'^{\otimes 2})\,n_i^{1/2}\mathrm{vec}(\mathbf{S}_i-\boldsymbol{\Sigma}_i)+r_i^{1/2}\mathbf{H}_k(\mathbf{I}_{k^2}+\mathbf{K}_k)(\mathbf{I}_k\otimes\boldsymbol{\beta}'\boldsymbol{\Sigma}_i)\,n^{1/2}\mathrm{vec}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+o_{\mathrm{P}}(1),$$

which (by using Part (i) of the Proposition) establishes the desired result. □

Proof of Lemma 4.1. The result trivially follows from Theorem 3.1 and the identity

$$\begin{pmatrix} Y_{1;0}^{(n)} \\ \vdots \\ Y_{m;0}^{(n)} \end{pmatrix} = \Big[\mathbf{I}_m \otimes (-p_0\mathbf{1}_q', (1-p_0)\mathbf{1}_{k-q}')\Big] \begin{pmatrix} n_1^{1/2}\text{dvec}\,(\hat{\mathbf{\Lambda}}_1 - \mathbf{\Lambda}_1) \\ \vdots \\ n_m^{1/2}\text{dvec}\,(\hat{\mathbf{\Lambda}}_m - \mathbf{\Lambda}_m) \end{pmatrix},$$

where $\mathbf{1}_\ell = (1, \ldots, 1)' \in \mathbb{R}^\ell$. □

References

[1] Akharif, A. and M. Hallin (2003). Efficient detection of random coefficients in AR(p) models. *Annals of Statistics* **31**, 675-704.

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

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

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

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

[6] Flury, B. and W. Gautschi (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.

[7] Hallin, M., D. Paindaveine, and T. Verdebout (2007). Pseudo-Gaussian tests for common principal components, submitted.

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

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

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

[11] Schaafsma, W. and L.J. Smid (1966). Most stringent somewhere most powerful tests against alternatives restricted by a number of linear alternatives. *Annals of Mathematical Statististics* **37**, 1161-1172.

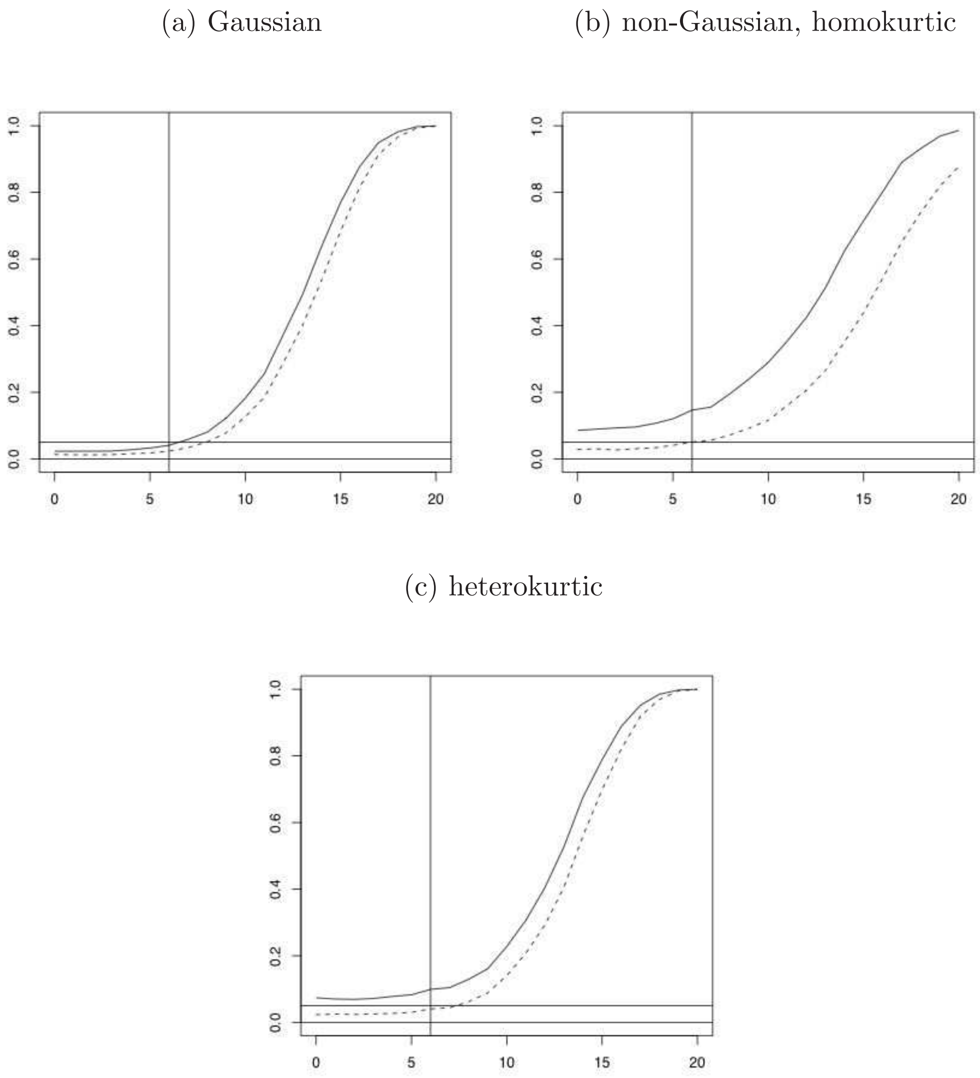
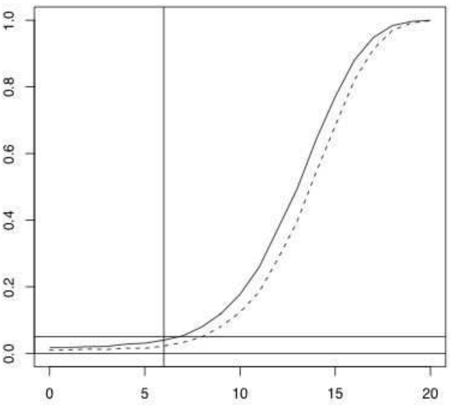
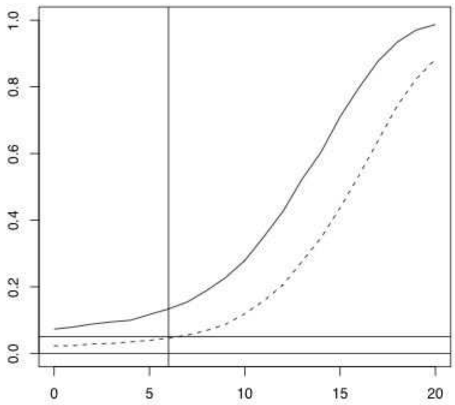


Figure 1: Rejection frequencies (out of $N = 10,000$ replications), in Experiment 1 (see Section 5), of Flury's Gaussian test $\phi^{(n)}_{\text{Flury}}$ (solid line) and our pseudo-Gaussian test $\phi^{(n)}_{\dagger}$ (dotted line). The asymptotic nominal size is $\alpha = 5\%$; $k = 2$, $m = 2$, $q = 1$. Sample sizes are $n_1 = n_2 = 100$.

(a) Gaussian



(b) non-Gaussian, homokurtic



(c) heterokurtic

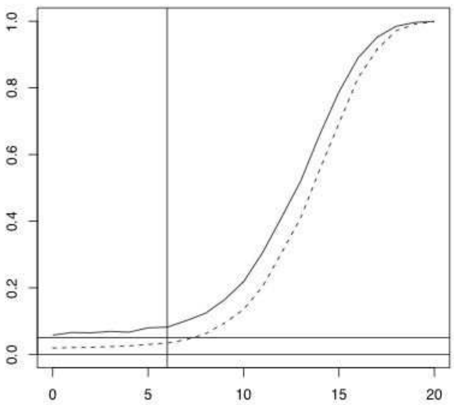


Figure 2: Rejection frequencies (out of $N = 10,000$ replications), in Experiment 2 (see Section 5), of Flury's Gaussian test $\phi^{(n)}_{\text{Flury}}$ (solid line) and our pseudo-Gaussian test $\phi^{(n)}_{\dagger}$ (dotted line). The asymptotic nominal size is $\alpha = 5\%$; $k = 2$, $m = 2$, $q = 1$. Sample sizes are $n_1 = n_2 = 100$.

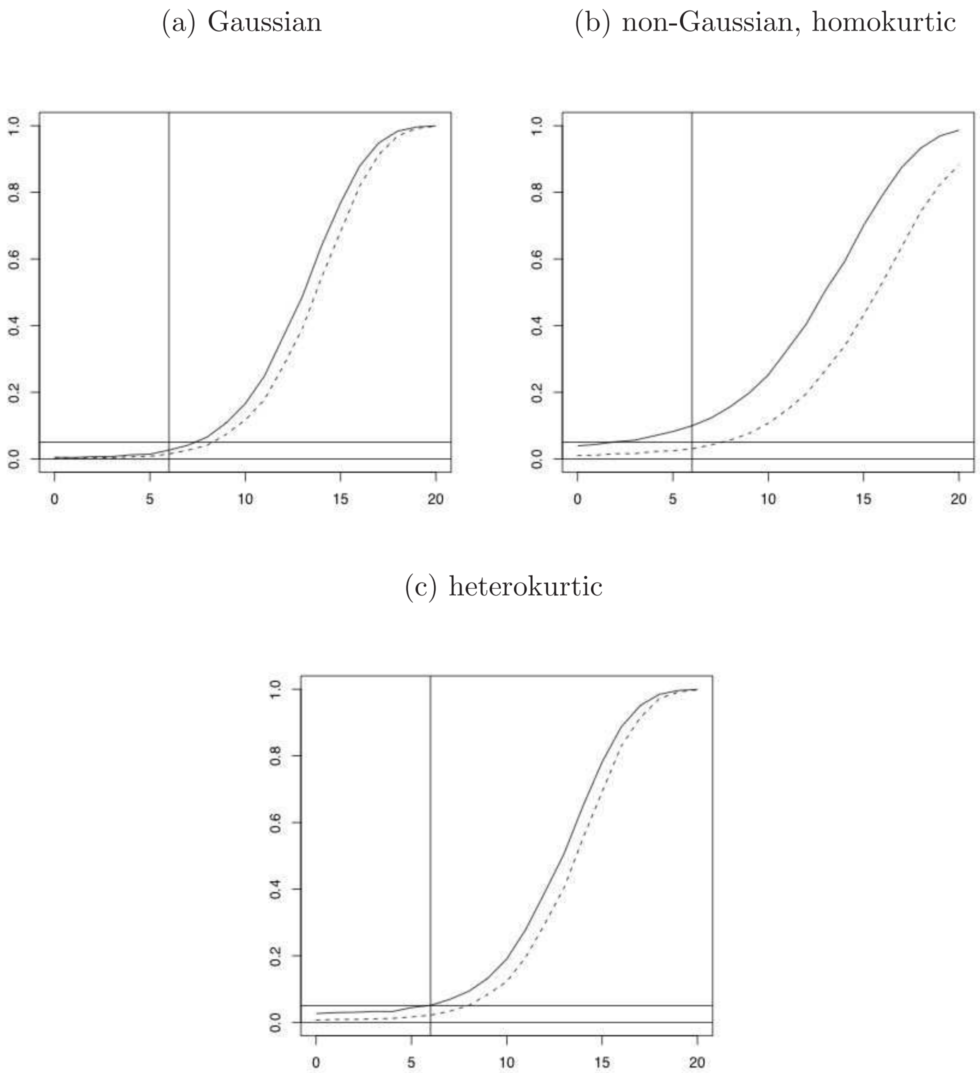


Figure 3: Rejection frequencies (out of $N = 10,000$ replications), in Experiment 3 (see Section 5), of Flury's Gaussian test $\phi^{(n)}_{\text{Flury}}$ (solid line) and our pseudo-Gaussian test $\phi^{(n)}_{\dagger}$ (dotted line). The asymptotic nominal size is $\alpha = 5\%$; $k = 2$, $m = 2$, $q = 1$. Sample sizes are $n_1 = n_2 = 100$.