J. Japan Statist. Soc.
Vol. 38 No. 1 2008 27–39

A GENERAL METHOD FOR CONSTRUCTING PSEUDO-GAUSSIAN TESTS

Marc Hallin*† and Davy Paindaveine*

A general method for constructing *pseudo-Gaussian tests*—reducing to traditional Gaussian tests under Gaussian densities but remaining valid under non-Gaussian ones—is proposed. This method provides a solution to several open problems in classical multivariate analysis. One of them is the test of the homogeneity of covariance matrices, an assumption that plays a crucial role in multivariate analysis of variance, under elliptical, and possibly heterokurtic densities with finite fourth-order moments.

Key words and phrases: Elliptical symmetry, homogeneity of covariances, local asymptotic normality, multivariate analysis of variance, pseudo-Gaussian tests.

1. Introduction

1.1. Gaussian and pseudo-Gaussian procedures

Classical statistical inference has been shaped under the influence of explicit or implicit Gaussian assumptions, and much of everyday practice is still deeply rooted in that traditional Gaussian vision. Such areas as multivariate analysis, time series or spatial statistics, where least squares, Gaussian maximum likelihood estimators and Gaussian likelihood ratio tests, periodograms, variograms, correlograms, or covariance-based methods are present at each step, are particularly marked. Robust and distribution-free methods have been introduced as a reaction to the pervasiveness of Gaussian influences but, despite their developments, and despite a widespread consensus that Gaussian assumptions are hardly realistic, most practitioners still adhere to traditional Gaussian methods.

This unfailing popularity is not just a symptom of scientific conservatism: Gaussian methods, with all their drawbacks, also yield quite attractive features which, to a large extent, account for their maintained success. The main reason for the success of Gaussian methods in multivariate analysis is probably due to the strong relation between the geometry of multinormal distributions and classical Euclidean linear structures, which provide simple and easily understandable geometric interpretations of orthogonality, linear projections, sums of squares, and analysis of variance tables. In the time series context, the Hilbert-space-based Cramér-Wold theory, which yields the simple, complete and fully coherent theoretical framework of classical forecasting, and constitutes the cornerstone of everyday time series practice, is also tailor-made for Gaussian processes. A

Accepted November 12, 2007.

*E.C.A.R.E.S., Institut de Recherche en Statistique, and Département de Mathématique, Université Libre de Bruxelles, CP 114, B-1050 Bruxelles, Belgium; the authors are also members of ECORE, the recently created association between CORE and ECARES.

†Royal Academy of Belgium.

more subtle and less visible argument is that, among all "regular" densities, the Gaussian ones are those for which inference for location is "most difficult", in the sense that Fisher information for location or regression, under given scale, reaches a minimum at the Gaussian—meaning that Gaussian methods may be an optimal choice, in a maximin sense, in the presence of observations with unspecified densities.

Since everybody nevertheless does agree that Gaussian assumptions are unrealistic, both arguments very strongly plead in favor of *pseudo-Gaussian* methods—namely, methods that remain (asymptotically) valid under a "broad class" of densities, while being (asymptotically) equivalent, in case the actual density happens to be Gaussian, to the "standard Gaussian procedure".

This definition of a pseudo-Gaussian method is rather vague, and provides little hint on how such methods should be constructed. Sometimes Gaussian methods themselves *ne varietur* can be considered as pseudo-Gaussian ones. For instance (denoting by $t_{n-1;\alpha}$ the $(1-\alpha)$-quantile of the Student distribution with $(n-1)$ degrees of freedom), the classical one-sided Student test for location

$$\varphi^{(n)}_{\text{Student}} := I[S^{(n)}_{\text{Student}} > t_{n-1;\alpha}], \qquad \text{with} \quad S^{(n)}_{\text{Student}} := \sqrt{n-1}(\bar{X} - \mu_0)/s \tag{1.1}$$

where $\bar{X} := n^{-1}\sum_{i=1}^n X_i$ and $s^2 := n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$, remains asymptotically valid (for $\mathcal{H}_0 : \mu := \mathrm{E}[X_1] = \mu_0$, at asymptotic level α) under any i.i.d. n-tuple $X_1, \ldots, X_n$ with finite variance, and is most powerful against $\mathcal{H}_1 : X_1, \ldots, X_n$ i.i.d. normal, with $\mu > \mu_0$.

It also often happens that turning a Gaussian procedure into a pseudo-Gaussian only requires a very simple and obvious modification of the test statistic. For instance, the Gaussian large-sample test $\varphi^{(n)}_{\text{scale}}$ of $\mathcal{H}_0 : \sigma^2 := \mathrm{Var}[X_1] = \sigma_0^2$ against $\mathcal{H}_1 : \sigma^2 > \sigma_0^2$ is based on the asymptotic standard normal distribution, under $\mathcal{H}_0$ and i.i.d. Gaussian observations $X_1, \ldots, X_n$, of

$$S^{(n)}_{\text{scale}} := \sqrt{n/2}(s^2 - \sigma_0^2)/\sigma_0^2.$$

This test under non-Gaussian X_i's is not valid anymore, as the asymptotic variance of s^2 then involves the kurtosis parameter $\kappa := (\mu_4/3\sigma^4) - 1$ of the X_i's which, under Gaussian conditions, is zero. More precisely, under the assumption that $X_1, \ldots, X_n$ are i.i.d. with variance σ^2 and finite fourth-order moment μ_4, $\sqrt{n}(s^2 - \sigma^2)$ is asymptotically normal, with mean zero and variance $\sigma^4(3\kappa + 2)$. An obvious pseudo-Gaussian version of $\varphi^{(n)}_{\text{scale}}$ is then

$$\varphi^{(n)}_{\dagger} := I\left[\sqrt{n}(s^2 - \sigma_0^2) > z_\alpha \sigma_0^2 \sqrt{3\hat{\kappa}^{(n)} + 2}\right], \tag{1.2}$$

where z_α denotes the $(1-\alpha)$ standard normal quantile and $\hat{\kappa}^{(n)}$ is an arbitrary consistent (under $\mathcal{H}_0$) estimator of κ; see Shapiro and Browne (1987) for a much more general instance of the same phenomenon.

Under our purposely vague definition, pseudo-Gaussian methods for a given problem cannot be expected to be unique, and several distinct pseudo-Gaussian versions of the same Gaussian method may exist, even in the very simple case of the Student test for location. Denoting by $R_+^{(n)}(\mu_0)$ the rank of $|X_i - \mu_0|$ among $|X_1 - \mu_0|, \ldots, |X_n - \mu_0|$, the van der Waerden or normal-score signed rank version of $\varphi^{(n)}_{\mathrm{Student}}$ is $\varphi^{(n)}_{\mathrm{rank}} := I[S^{(n)}_{\mathrm{rank}} > z_\alpha]$, with

$$S^{(n)}_{\mathrm{rank}} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{sign}(X_i - \mu_0)\Phi^{-1}\left(\frac{1}{2} + \frac{R_+^{(n)}(\mu_0)}{2(n+1)}\right). \tag{1.3}$$

This test is valid (for $\mathcal{H}_0 : \mu = \mu_0$, at asymptotic level α) under any i.i.d. n-tuple $X_1, \ldots, X_n$ with density f such that $f(\mu - x) = f(\mu + x)$ (note that $\mu = \mathrm{E}[X_1]$ as soon as f has finite first-order moments); under Gaussian f, it is asymptotically equivalent to $\varphi^{(n)}_{\mathrm{Student}}$, and hence inherits, asymptotically, the Gaussian optimality properties of the latter. This van der Waerden test $\varphi^{(n)}_{\mathrm{rank}}$ thus qualifies, in the sense of our general definition, as a pseudo-Gaussian version of $\varphi^{(n)}_{\mathrm{Student}}$.

A third variant of the same Student test is the *permutation t-test*

$$\varphi^{(n)}_{\mathrm{permutation}} := I[S^{(n)}_{\mathrm{Student}} > t^{(n)}_{1-\alpha}(|X_1 - \mu_0|, \ldots, |X_n - \mu_0|)], \tag{1.4}$$

(see, e.g., Lehmann and Casella (1998), or Efron (1969)) obtained by comparing $S^{(n)}_{\mathrm{Student}}$ with the $(1-\alpha)$-quantile $t^{(n)}_{1-\alpha}(|X_1 - \mu_0|, \ldots, |X_n - \mu_0|)$ of its conditional (on the n-tuple of absolute values $|X_1 - \mu_0|, \ldots, |X_n - \mu_0|$) distribution. This test, which is valid under the same conditions as the van der Waerden test $\varphi^{(n)}_{\mathrm{rank}}$, is also asymptotically equivalent to $\varphi^{(n)}_{\mathrm{Student}}$ under Gaussian assumptions—hence also qualifies, under the above definition, as a pseudo-Gaussian version of the Student test. More generally, normal-score rank test statistics are obtained (under *exact score* form: see Hájek *et al.* (1999)) by conditioning a Gaussian test statistic on some appropriate *maximal invariant* (residual ranks or signed ranks, or some adequate generalization thereof), while Gaussian permutation tests consist in comparing the original Gaussian test statistic with its $(1-\alpha)$-quantile conditional on some $\mathcal{H}_0$-sufficient and -complete statistic (provided, of course, that such a statistic exists).

Similar definitions also apply to point estimation. Validity then is to be understood as root-n consistency; the analogues of normal-score rank tests are R-estimators based on normal-score objective functions, those of the Gaussian permutation tests are the U-statistics resulting from the Rao-Blackwellisation, based on the same sufficient and complete statistic as the corresponding permutation test, of standard Gaussian estimates.

In the sequel, we exclusively focus on testing, and concentrate on pseudo-Gaussian tests of type (1.1) or (1.2) that directly result from the "validity-robustification" of some original Gaussian test statistic.

1.2. Pseudo-Gaussian tests

Transforming a Gaussian test into a pseudo-Gaussian one is not always as easy and straightforward as it is for examples (1.1) or (1.2). First of all, such a transformation is not always possible, as it requires a form of *adaptivity* of the underlying model. Indeed, the fact that the same (asymptotic) performance can be attained, under Gaussian densities, whether Gaussianity is assumed or not, is an indication that the Gaussian semiparametric efficiency bounds (obtained in the semiparametric model where densities remain unspecified within some "broad class" $\mathcal{F}$ of densities) are the same as the parametric ones (obtained in the parametric model where densities are assumed to be Gaussian): call this *$\mathcal{F}$-adaptivity at the Gaussian*. Even in such favorable cases, the construction of pseudo-Gaussian tests may be all but obvious. In a classical reference, Muirhead and Waternaux (1980) provide an insightful study of the problem of turning standard Gaussian tests of hypotheses involving covariance matrices into pseudo-Gaussian ones remaining valid under elliptical densities with finite fourth-order moments. They clearly distinguish some "easy" problems—tests of sphericity, tests of the equality of a subset of the characteristic roots of the covariance matrix (i.e., *subspace sphericity*), tests of block-diagonality—and some "harder" ones, among which the (apparently simpler) one-sample test of the hypothesis that the covariance matrix $\mathbf{\Sigma}$ takes some given value $\mathbf{\Sigma}_0$, the two-sample test of equality of covariance matrices, and the corresponding m-sample test of covariance homogeneity. Another "hard" case is the test of the hypothesis of common principal components treated in Hallin *et al.* (2007). For those "hard" problems, Muirhead and Waternaux (1980) conclude that "*it is not possible in the more general elliptical case to adjust the* (*Gaussian likelihood ratio*) *test so that its limiting distribution agrees with that obtained under the normality assumption*"; see also Section 3 of Tyler (1983) and Shapiro and Browne (1987).

The objective of this paper is to show how the Le Cam approach, which is now quite standard in the modern treatment of asymptotic inference (see Le Cam and Yang (2000), or Taniguchi and Kakizawa (2000) in the time series context), provides a general method for transforming a Gaussian test (typically, of the likelihood ratio or Lagrange multiplier type) into a pseudo-Gaussian one. This method, which considerably extends the Shapiro and Browne (1987) one, can handle—without any tangent space computation—most of Muirhead and Waternaux's "hard cases".

The problem of covariance homogeneity is treated as an illustration.

2. A general method

2.1. Assumptions

Throughout, we are considering sequences of statistical models or *experiments*, that is, sequences of triples of the form $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_{\boldsymbol{g}}^{(n)})$, $n \in \mathbb{N}$, where $\mathcal{P}_{\boldsymbol{g}}^{(n)} := \{\mathrm{P}_{\boldsymbol{\vartheta};\boldsymbol{g}}^{(n)} \mid \boldsymbol{\vartheta} \in \mathbf{\Theta}\}$ denotes a family of probability measures over $(\mathcal{X}^{(n)}$, $\mathcal{A}^{(n)})$, indexed by a parameter $\boldsymbol{\vartheta} \in \mathbf{\Theta} \subseteq \mathbb{R}^p$ and characterized (typically, in m-sample problems) by some m-tuple $\boldsymbol{g} := (g_1, \ldots, g_m)$, $m \in \mathbb{N}$, of standardized probability densities g_i defined over $(\mathbb{R}^\ell, \mathcal{B}^\ell)$, $\ell \in \mathbb{N}$. Standardization here can be based on any choice of location and scale parameters—which in turn can be included in $\boldsymbol{\vartheta}$. The $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)})$-measurable observation described by $\mathrm{P}_{\boldsymbol{\vartheta};\boldsymbol{g}}^{(n)}$ is denoted by $\boldsymbol{X}^{(n)}$.

Let $\Lambda^{(n)}_{\boldsymbol{\vartheta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}/\boldsymbol{\vartheta};\boldsymbol{g}} := \log(d\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+\boldsymbol{\nu}(n)\boldsymbol{\tau};\boldsymbol{g}}/d\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}(\boldsymbol{X}^{(n)}))$, where $\boldsymbol{\nu}(n)$ is a sequence of full-rank matrices such that $\lim_{n\to\infty}\boldsymbol{\nu}(n) = \mathbf{0}$; in the traditional case of root-n contiguity, $\boldsymbol{\nu}(n)$ takes the simple form $\boldsymbol{\nu}(n) = n^{-1/2}\boldsymbol{I}_p$, where $\boldsymbol{I}_p$ denotes the p-dimensional identity matrix. We say that the family $\mathcal{P}^{(n)}_{\boldsymbol{g}}$ is ULAN, with *contiguity rate* $\boldsymbol{\nu}(n)$, *central sequence* $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, and positive definite *information matrix* $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{g}}$ if, for any bounded sequence $\boldsymbol{\tau}^{(n)}$ and any $\boldsymbol{\vartheta}^{(n)}$ such that $\boldsymbol{\nu}^{-1}(n)(\boldsymbol{\vartheta}^{(n)} - \boldsymbol{\vartheta})$ is $O(1)$ for some $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, we have, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^{(n)};\boldsymbol{g}}$, as $n \to \infty$,

(i) $\Lambda^{(n)}_{\boldsymbol{\vartheta}^{(n)}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}/\boldsymbol{\vartheta}^{(n)};\boldsymbol{g}} = \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta}^{(n)};\boldsymbol{g}} - \frac{1}{2}\boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{g}}\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1)$, and

(ii) $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta}^{(n)};\boldsymbol{g}}$ is asymptotically $\mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{g}})$.

In our definition of ULAN, we moreover include the following convenient continuity requirement:

(iii) $\boldsymbol{\vartheta} \mapsto \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{g}}$ is continuous.

Focusing on the Gaussian case, denote by $\mathcal{P}^{(n)}_{\boldsymbol{\phi}} = \{\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \mid \boldsymbol{\vartheta} \in \boldsymbol{\Theta}\}$ the family associated with the m-tuple $\boldsymbol{g} = \boldsymbol{\phi} := (\phi, \dots, \phi)$ of standardized Gaussian densities (where standardization does not necessarily imply zero mean and unit variance). We assume the following.

Assumption (A1). The Gaussian family $\mathcal{P}^{(n)}_{\boldsymbol{\phi}}$ is ULAN, with contiguity rate $\boldsymbol{\nu}(n)$, central sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, and information matrix $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$.

Consider the problem of testing the null hypothesis $\mathcal{H}_0$ under which

$$\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon}) := \{\boldsymbol{\vartheta}_0 + \boldsymbol{\Upsilon}\boldsymbol{\ell} \mid \boldsymbol{\ell} \in \mathbb{R}^q\},$$

where $\boldsymbol{\Upsilon}$ is a $p \times q$ matrix of full rank $0 \le q < p$, and $\boldsymbol{\vartheta}_0 \in \boldsymbol{\Theta}$ (for $q = 0$, put $\mathcal{M}(\boldsymbol{\Upsilon}) := \{\mathbf{0}\}$; $\mathcal{H}_0$ then reduces to $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$). The null hypothesis $\mathcal{H}_0$ thus places $(p - q)$ independent linear constraints on $\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0$ (for the sake of simplicity, we avoid writing $\mathcal{H}_0 : \boldsymbol{\vartheta} \in (\boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})) \bigcap \boldsymbol{\Theta}$, as we should). Denote by $\mathcal{F}_\dagger$ the class of all m-tuples $\boldsymbol{g}$ such that, for any $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$,

(i) for some positive definite $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ is asymptotically $\mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$ as $n \to \infty$;

(ii) $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ is *asymptotically linear* under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, that is, there exist $p \times p$ matrices $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}$ such that

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta}^{(n)};\boldsymbol{\phi}} - \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = -\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\nu}^{-1}(n)(\boldsymbol{\vartheta}^{(n)} - \boldsymbol{\vartheta}) + o_{\mathrm{P}}(1), \tag{2.1}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, for any sequence $\boldsymbol{\vartheta}^{(n)}$ such that $\boldsymbol{\nu}^{-1}(n)(\boldsymbol{\vartheta}^{(n)} - \boldsymbol{\vartheta}) = O(1)$;

(iii) the mappings $\boldsymbol{\vartheta} \mapsto \boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ and $\boldsymbol{\vartheta} \mapsto \boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}$ are continuous.

Note that, in case the family $\mathcal{P}^{(n)}_{\boldsymbol{g}}$ itself is ULAN, with central sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, Le Cam's third Lemma implies that $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}$ is the asymptotic covariance, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, of $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ and $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$. The Gaussian m-tuple $\boldsymbol{\phi}$ under Assumption (A1) automatically belongs to $\mathcal{F}_\dagger$. Indeed, under $\boldsymbol{g} = \boldsymbol{\phi}$, conditions (i) and (iii) respectively coincide with parts (ii) and (iii) of the definition of ULAN, of which condition (ii) is a direct consequence (with $\boldsymbol{\Gamma}^{\boldsymbol{\phi}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \boldsymbol{\Gamma}^{\boldsymbol{\phi}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{\phi}} = \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$). As we shall see, the sequence of semiparametric experiments $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \bigcup_{g\in\mathcal{F}_\dagger} \mathcal{P}^{(n)}_{\boldsymbol{g}})$ is $\mathcal{F}_\dagger$-adaptive at $\boldsymbol{g} = \boldsymbol{\phi}$, in the sense defined in Subsection 1.2—provided of course that the following assumption is satisfied.

Assumption (A2). The class $\mathcal{F}_\dagger$ does not reduce to $\{\boldsymbol{\phi}\}$.

2.2. The proposed pseudo-Gaussian test

Recall that the matrix of the projection (in $\mathbb{R}^p$) onto $\mathcal{M}(\boldsymbol{A})$, where $\boldsymbol{A}$ is $p \times q$, with rank $0 < q \leq p$, is $\text{proj}(\boldsymbol{A}) := \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$; this definition naturally extends to $q = 0$ by putting $\text{proj}(\boldsymbol{A}) := \boldsymbol{0}$ for $\boldsymbol{A}$ with rank $q = 0$. An optimal (namely, *locally and asymptotically most stringent* if $q > 0$, and *locally and asymptotically maximin* if $q = 0$: see Section 11.9 of Le Cam (1986)) Gaussian test for $\mathcal{H}_0$ is based on the asymptotically (under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, with $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$) chi-squared (with $p-q$ degrees of freedom) null distribution of $Q^{(n)}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\vartheta}}^{(n)})$, where

$$Q^{(n)}_{\boldsymbol{\phi}}(\boldsymbol{\vartheta}) := \|[\boldsymbol{I} - \text{proj}(\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})]\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\|^2, \tag{2.2}$$

and $\hat{\boldsymbol{\vartheta}}^{(n)}$ is a sequence of estimators satisfying the following assumption.

Assumption (B1). For all $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$,
- (i) $\hat{\boldsymbol{\vartheta}}^{(n)} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$ with $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$-probability one for all n;
- (ii) $\|\boldsymbol{\nu}^{-1}(n)(\hat{\boldsymbol{\vartheta}}^{(n)} - \boldsymbol{\vartheta})\|$ is $O_{\mathrm{P}}(1)$ as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ for all $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$;
- (iii) the number of distinct values of $\hat{\boldsymbol{\vartheta}}^{(n)}$ in balls of the form $\{\boldsymbol{t} \in \mathbb{R}^p \mid \|\boldsymbol{\nu}^{-1}(n)(\boldsymbol{t} - \boldsymbol{\vartheta})\| \leq C\}$, $C \in \mathbb{R}^+$ fixed, is bounded as $n \to \infty$.

In the sequel, we often simply write $\hat{\boldsymbol{\vartheta}}$ for $\hat{\boldsymbol{\vartheta}}^{(n)}$. Part (iii) of Assumption (B1) is easily met by discretizing each component of any estimator $\tilde{\boldsymbol{\vartheta}}^{(n)}$ satisfying parts (i) and (ii), that is, by letting

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}^{(n)} &:= c_0^{-1}\boldsymbol{\nu}(n)(\text{sign}((\boldsymbol{\nu}^{-1}(n)\tilde{\boldsymbol{\vartheta}}^{(n)})_1)\lceil c_0|(\boldsymbol{\nu}^{-1}(n)\tilde{\boldsymbol{\vartheta}}^{(n)})_1|\rceil, \ldots, \\ &\qquad \ldots, c_0^{-1}\boldsymbol{\nu}(n)(\text{sign}((\boldsymbol{\nu}^{-1}(n)\tilde{\boldsymbol{\vartheta}}^{(n)})_p)\lceil c_0|(\boldsymbol{\nu}^{-1}(n)\tilde{\boldsymbol{\vartheta}}^{(n)})_p|\rceil),\end{aligned}$$

where c_0 is an arbitrary positive constant. Such discretized estimators indeed trivially satisfy part (iii) of Assumption (B1). Discretization is standard in the context (see page 125 of Le Cam and Yang (2000)), and allows for replacing in (2.1) the deterministic quantity $\boldsymbol{\vartheta}^{(n)}$ with the estimator $\hat{\boldsymbol{\vartheta}}$ (see, e.g., Lemma 4.4 in Kreiss (1987)). In practice (where $n = n_0$ is fixed), however, the importance of discretization should not be overemphasized: indeed, c_0 can be chosen arbitrarily large, and one always can pretend to start discretizing from $n = n_0 + 1$ on.

Moreover, in view of the strong regularity features of Gaussian estimators, one often can dispense with part (iii) of Assumption (B1) (see Hallin and Paindaveine (2007)).

The matrix in brackets in (2.2) is the matrix of the projection onto the linear space $\mathcal{M}^{\perp}(\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$ orthogonal to $\mathcal{M}(\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}(\boldsymbol{\vartheta})\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$. In view of (2.1), the continuity of $\boldsymbol{\vartheta} \mapsto \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, and Assumption (B1), this projection, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ with $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$, maps $\boldsymbol{\Gamma}^{-1/2}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}(\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}} - \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})$ to a $o_{\mathrm{P}}(1)$ quantity. As a consequence, the effect (still under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$) of substituting an estimator $\hat{\boldsymbol{\vartheta}}$ satisfying (B1) (such as the discretized null Gaussian MLE) for $\boldsymbol{\vartheta}$ in $Q^{(n)}_{\boldsymbol{\phi}}(\boldsymbol{\vartheta})$ is asymptotically nil (in probability). In practice, $Q^{(n)}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\vartheta}}^{(n)})$ corresponds, up to o_{P}'s, to the likelihood ratio, Wald, or Lagrange Multipliers test statistics.

If such tests are to remain (asymptotically) valid under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$ with $\boldsymbol{g} \neq \boldsymbol{\phi}$,

(i) the matrix of the quadratic form should remain a projection matrix "neutralizing" (asymptotically, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, with $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$) the substitution of $\hat{\boldsymbol{\vartheta}}$ for $\boldsymbol{\vartheta}$, and

(ii) the argument of the quadratic form should remain asymptotically standard multinormal.

This latter argument thus should be of the form $(\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ rather than $\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$. Now, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$, $(\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}})^{-1/2}(\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}} - \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})$, still in view of the asymptotic linearity property holding for $\boldsymbol{g} \in \mathcal{F}_{\dagger}$, belongs (up to o_{P}'s) to $\mathcal{M}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$. Hence, the projection should be onto $\mathcal{M}^{\perp}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$ rather than $\mathcal{M}^{\perp}(\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$, which leads to

$$Q^{(n)}_{\dagger}(\boldsymbol{\vartheta}) := \|[\boldsymbol{I} - \mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})](\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\|^2,$$

still to be computed at some $\hat{\boldsymbol{\vartheta}}$. This estimator $\hat{\boldsymbol{\vartheta}}$ however is to satisfy Assumption (B1) under any $\boldsymbol{g} \in \mathcal{F}_{\dagger}$: denote by (B1†) this reinforcement of (B1).

Note that for $\boldsymbol{g} = \boldsymbol{\phi}$, we retrieve the Gaussian statistic: $Q^{(n)}_{\dagger}(\hat{\boldsymbol{\vartheta}}) = Q^{(n)}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\vartheta}})$. Contrary to $Q^{(n)}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\vartheta}})$, however, $Q^{(n)}_{\dagger}(\hat{\boldsymbol{\vartheta}})$ is asymptotically chi-square under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$, for *any* $\boldsymbol{g} \in \mathcal{F}_{\dagger}$. Finally, let us assume that $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ and $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}$ can be consistently estimated (under the null):

ASSUMPTION (B2). There exists a couple of estimators $\hat{\boldsymbol{\Gamma}}_{\mathrm{Var}}$ and $\hat{\boldsymbol{\Gamma}}_{\mathrm{Cov}}$ that, for any $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$ and any $\boldsymbol{g} \in \mathcal{F}_{\dagger}$, converge in $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$-probability, as $n \to \infty$, to $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ and $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}$, respectively.

One would expect $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}$ to involve "functional covariance coefficients", that would typically be difficult to estimate. Actually, in most cases (as in the application in Section 3 below), $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}$ does not depend on $\boldsymbol{g}$ at all, so that $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}} = \boldsymbol{\Gamma}^{\boldsymbol{\phi}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{\phi}} = \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$. The continuity of the mapping $\boldsymbol{\vartheta} \mapsto \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ then implies that $\boldsymbol{\Gamma}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}$ is an admissible choice for $\hat{\boldsymbol{\Gamma}}_{\mathrm{Cov}}$. As for $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, it does in general depend on $\boldsymbol{g}$, but in a way that makes consistent estimation (relying on a law of large numbers argument) quite straightforward.

Summing this up, we can state the following result.

THEOREM. *Let Assumptions* (A1), (A2), (B1†), *and* (B2) *hold, and denote by* $\chi^2_{p-q;1-\alpha}$ *the* $(1-\alpha)$*-quantile of the chi-square distribution with* $(p-q)$ *degrees of freedom. Then, the test* $\varphi^{(n)}_{\dagger} := I[Q^{(n)}_{\dagger}(\hat{\boldsymbol{\vartheta}}) > \chi^2_{p-q;1-\alpha}]$, *with*

$$Q^{(n)}_{\dagger}(\hat{\boldsymbol{\vartheta}}) := \|[\boldsymbol{I} - \mathrm{proj}(\hat{\boldsymbol{\Gamma}}^{-1/2}_{\mathrm{Var}}\hat{\boldsymbol{\Gamma}}_{\mathrm{Cov}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})]\hat{\boldsymbol{\Gamma}}^{-1/2}_{\mathrm{Var}}\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}\|^2,$$

is a pseudo-Gaussian version, valid under any $\boldsymbol{g} \in \mathcal{F}_{\dagger}$, *of the locally and asymptotically optimal Gaussian test* $\varphi^{(n)}_{\boldsymbol{\phi}} := I[Q^{(n)}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\vartheta}}) > \chi^2_{p-q;1-\alpha}]$.

PROOF. Throughout this proof, statements are made under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, for arbitrary $\boldsymbol{\vartheta} \in \boldsymbol{\vartheta}_0 + \mathcal{M}(\boldsymbol{\Upsilon})$ and $\boldsymbol{g} \in \mathcal{F}_{\dagger}$. Assumption (B2) implies that

$$\begin{aligned}&[\boldsymbol{I} - \mathrm{proj}(\hat{\boldsymbol{\Gamma}}^{-1/2}_{\mathrm{Var}}\hat{\boldsymbol{\Gamma}}_{\mathrm{Cov}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})]\hat{\boldsymbol{\Gamma}}^{-1/2}_{\mathrm{Var}}\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}\\ &\quad= [\boldsymbol{I} - \mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})](\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}} + o_{\mathrm{P}}(1).\end{aligned} \tag{2.3}$$

It further follows from Assumption (B1†) that the effect of substituting $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ for $\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}$ in (2.3) also is $o_{\mathrm{P}}(1)$. The asymptotically chi-square distribution, with $p-q$ degrees of freedom, of $Q^{(n)}_{\dagger}$ then readily follows from the asymptotically standard multinormal distribution of $(\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ and the fact that $[\boldsymbol{I}-\mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})]$ is symmetric and idempotent with rank $p-q$. Validity of $\varphi^{(n)}_{\dagger}$ under any $\boldsymbol{g} \in \mathcal{F}_{\dagger}$ follows. Its asymptotic equivalence, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, with $\varphi^{(n)}_{\boldsymbol{\phi}}$ is a direct consequence of the fact that $\hat{\boldsymbol{\Gamma}}_{\mathrm{Var}}$ and $\hat{\boldsymbol{\Gamma}}_{\mathrm{Cov}}$ both converge to $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, yielding $[\boldsymbol{I} - \mathrm{proj}(\boldsymbol{\Gamma}^{1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})]\boldsymbol{\Gamma}^{-1/2}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ in (2.3) and the Gaussian test statistic (2.2). The test $\varphi^{(n)}_{\dagger}$ is thus the pseudo-Gaussian test we are looking for.

Noncentrality parameters under local alternatives are easily derived via Le Cam's third Lemma: under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}+\boldsymbol{\nu}(n)\boldsymbol{\tau};\boldsymbol{g}}$, where $\boldsymbol{\nu}(n)\boldsymbol{\tau} \notin \mathcal{M}(\boldsymbol{\Upsilon})$ and $\boldsymbol{g} \in \mathcal{F}_{\dagger}$ is such that the family $\mathcal{P}^{(n)}_{\boldsymbol{g}}$ is ULAN, the pseudo-Gaussian test statistic $Q^{(n)}_{\dagger}$ is asymptotically noncentral chi-square, with $(p-q)$ degrees of freedom and noncentrality parameter

$$\|[\boldsymbol{I} - \mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})](\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}}\boldsymbol{\tau}\|^2.$$

3. An application: Testing for covariance homogeneity

As an illustration, let us consider the problem of testing for covariance homogeneity in a k-dimensional m-sample location-scale and possibly heterokurtic model. This problem, which belongs to Muirhead and Waternaux (1980)'s list of "hard problems", has remained unsolved for about half a century. The classical

Gaussian solution is Bartlett's modified likelihood ratio test, and is notoriously sensitive to violations of the Gaussian assumptions. We refer to Hallin and Paindaveine (2007) for a complete pseudo-Gaussian solution and a bibliography of the subject.

The observation in that problem consists of an m-tuple of mutually independent samples $(\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i})$, $i = 1, \ldots, m$, of i.i.d. k-dimensional observations. Instead of the classical assumption of multinormality, we assume that these observations are *elliptical* with finite fourth-order moments. More precisely, considering the class $\mathcal{F}_0$ of all functions $h : \mathbb{R}_0^+ \to \mathbb{R}^+$ such that

(i) $\mu_{k+3;h} < \infty$ and $\mu_{k+1;h}/\mu_{k-1;h} = k$, where $\mu_{\ell;h} := \int_0^\infty r^\ell h(r) dr$,

(ii) h is absolutely continuous, with a.e. derivative $\dot{h}$, and

(iii) the integrals $\mathcal{I}_k(h) := \int_0^\infty \varphi_h^2(r) r^{k-1} h(r) dr$ and $\mathcal{J}_k(h) := \int_0^\infty \varphi_h^2(r) r^{k+1} h(r) dz$, where $\varphi_h := -\dot{h}/h$, are finite,

we assume that $\boldsymbol{X}_{ij}$, $j = 1, \ldots, n_i$ are mutually independent, with probability density function of the form

$$\boldsymbol{x} \mapsto c_{k,g_i} |\boldsymbol{\Sigma}_i|^{-1/2} g_i(((\boldsymbol{x} - \boldsymbol{\theta}_i)' \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x} - \boldsymbol{\theta}_i))^{1/2}), \qquad i = 1, \ldots, m, \tag{3.1}$$

for some k-dimensional location vector $\boldsymbol{\theta}_i$, some positive definite $(k \times k)$ covariance matrix $\boldsymbol{\Sigma}_i$, and some $g_i \in \mathcal{F}_0$ (c_{k,g_i} is a norming constant).

The functions g_i are not, *stricto sensu*, probability density functions. However, defining (throughout, $\boldsymbol{\Sigma}^{1/2}$ stands for the symmetric root of $\boldsymbol{\Sigma}$) the *elliptical coordinates*

$$\begin{aligned} \boldsymbol{U}_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i) &:= \frac{\boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{X}_{ij} - \boldsymbol{\theta}_i)}{\|\boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{X}_{ij} - \boldsymbol{\theta}_i)\|} \qquad \text{and} \\ d_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i) &:= \|\boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{X}_{ij} - \boldsymbol{\theta}_i)\|, \end{aligned} \tag{3.2}$$

the *radial distances* d_{ij} have density $\tilde{g}_i(r) := (\mu_{k-1;g_i})^{-1} r^{k-1} g_i(r)$. Condition (i) therefore implies that d_{ij} has finite fourth-order moments, and is standardized in such a way that $\mathrm{E}[d_{ij}^2(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)] = k$—so that $\boldsymbol{\Sigma}_i$ indeed is the covariance matrix $\mathrm{Var}(\boldsymbol{X}_{ij})$ in population i.

Write $\mathrm{diag}(\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_m)$ for the block-diagonal matrix with diagonal blocks $\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_m$. Conditions (ii) and (iii), along with the additional assumption that $\lambda_i^{(n)} := n_i^{(n)}/n \to \lambda_i \in (0, 1)$, as $n \to \infty$, for all $i = 1, \ldots, m$, entail (see Proposition 4.1 of Hallin and Paindaveine (2007)) ULAN under any m-tuple $\boldsymbol{g} := (g_1, \ldots, g_m) \in \mathcal{F}_0^m (:= \mathcal{F}_0 \times \cdots \times \mathcal{F}_0$, m times), with contiguity rate

$$\begin{aligned} \boldsymbol{\nu}(n) &:= \mathrm{diag}(\boldsymbol{\nu}_{I}(n), \boldsymbol{\nu}_{II}(n), \boldsymbol{\nu}_{III}(n)) \\ &:= n^{-1/2}\,\mathrm{diag}(\boldsymbol{\Lambda}^{(n)} \otimes \boldsymbol{I}_k, \boldsymbol{\Lambda}^{(n)}, \boldsymbol{\Lambda}^{(n)} \otimes \boldsymbol{I}_{k_0}), \end{aligned} \tag{3.3}$$

where $k_0 := k(k+1)/2 - 1$ and $\boldsymbol{\Lambda}^{(n)} := \mathrm{diag}((\lambda_1^{(n)})^{-1/2}, \ldots, (\lambda_m^{(n)})^{-1/2})$ ($\otimes$ stands for the Kronecker product).

Before proceeding further, some clarification about the parametrization is necessary (which, in particular, will make the partitioned contiguity rate in (3.3) less mysterious), and some additional notation is needed. For given radial densities $\boldsymbol{g} \in \mathcal{F}_0^m$, the parameters of the model are the m location vectors $\boldsymbol{\theta}_i$ and the m covariance matrices $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, m$. Decomposing the latter into $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{V}_i$, where $\sigma_i := |\boldsymbol{\Sigma}_i|^{1/2k}$ is a scale and $\boldsymbol{V}_i$ a *shape matrix* (that is, a $k \times k$ symmetric positive definite matrix such that $|\boldsymbol{V}_i| = 1$), this parameter takes the form of a L-dimensional vector

$$\boldsymbol{\vartheta} := (\boldsymbol{\vartheta}'_I, \boldsymbol{\vartheta}'_{II}, \boldsymbol{\vartheta}'_{III})' := (\boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_m, \sigma_1^2, \ldots, \sigma_m^2, (\overset{\circ}{\text{vech}}\, \boldsymbol{V}_1)', \ldots, (\overset{\circ}{\text{vech}}\, \boldsymbol{V}_m)')',$$

where $L = mk(k+3)/2$ and $\overset{\circ}{\text{vech}}(\boldsymbol{V})$ is defined by $\text{vech}(\boldsymbol{V}) =: ((\boldsymbol{V})_{11}, (\overset{\circ}{\text{vech}}\, \boldsymbol{V})')'$: $\boldsymbol{\Sigma}_i$ indeed is entirely determined by σ_i^2 and $\overset{\circ}{\text{vech}}(\boldsymbol{V}_i)$ (see Hallin and Paindaveine (2007)). Write $\boldsymbol{\Theta}$ for the set of admissible $\boldsymbol{\vartheta}$ values, and $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$ (instead of $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\tilde{\boldsymbol{g}}}$ with $\tilde{\boldsymbol{g}} := (\tilde{g}_1, \ldots, \tilde{g}_m)$) for the joint distribution of the $n := \sum_{i=1}^m n_i$ observations under parameter value $\boldsymbol{\vartheta}$ and the m-tuple of radial densities $\boldsymbol{g} \in \mathcal{F}_0^m$; the notation $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ however will be maintained in the Gaussian case.

The null hypothesis $\mathcal{H}_0 : \sigma_1^2 \boldsymbol{V}_1 = \cdots = \sigma_m^2 \boldsymbol{V}_m$ of covariance homogeneity then can be written as $\mathcal{H}_0 : \boldsymbol{\vartheta} \in \mathcal{M}(\boldsymbol{\Upsilon})$, with

$$\boldsymbol{\Upsilon} := \text{diag}(\boldsymbol{\Upsilon}_I, \boldsymbol{\Upsilon}_{II}, \boldsymbol{\Upsilon}_{III}) := \text{diag}(\boldsymbol{I}_{mk}, \boldsymbol{1}_m, \boldsymbol{1}_m \otimes \boldsymbol{I}_{k_0}), \tag{3.4}$$

where $\boldsymbol{1}_m := (1, \ldots, 1)' \in \mathbb{R}^m$.

The following notation will be used in the sequel. Denoting by $\boldsymbol{e}_\ell$ the ℓ-th vector of the canonical basis of $\mathbb{R}^k$, let $\boldsymbol{K}_k := \sum_{i,j=1}^k (\boldsymbol{e}_i \boldsymbol{e}'_j) \otimes (\boldsymbol{e}_j \boldsymbol{e}'_i)$ be the $k^2 \times k^2$ *commutation matrix*. Define $\boldsymbol{M}_k(\boldsymbol{V})$ as the $(k_0 \times k^2)$ matrix such that $(\boldsymbol{M}_k(\boldsymbol{V}))'(\overset{\circ}{\text{vech}}\, \boldsymbol{v}) = (\text{vec}\, \boldsymbol{v})$ for any symmetric $k \times k$ matrix $\boldsymbol{v}$ such that $\text{tr}(\boldsymbol{V}^{-1}\boldsymbol{v}) = 0$. Let further $\boldsymbol{V}^{\otimes 2} := \boldsymbol{V} \otimes \boldsymbol{V}$, $\underline{E_k}(\boldsymbol{g}) := \text{diag}(E_k(g_1), \ldots, E_k(g_m))$, with $E_k(g_i) := \int_0^\infty r^4 \tilde{g}_i(r)dr$, $\underline{C_k}(\boldsymbol{g}) := \text{diag}(E_k(g_1) - k^2, \ldots, E_k(g_m) - k^2)$, $\underline{\boldsymbol{V}} := \text{diag}(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_m)$, $\underline{\boldsymbol{\sigma}} := \text{diag}(\sigma_1, \ldots, \sigma_m)$, and

$$\boldsymbol{H}_k(\boldsymbol{V}) := \frac{1}{4k(k+2)} \boldsymbol{M}_k(\boldsymbol{V})[\boldsymbol{I}_{k^2} + \boldsymbol{K}_k](\boldsymbol{V}^{\otimes 2})^{-1}(\boldsymbol{M}_k(\boldsymbol{V}))'.$$

It then follows from Proposition 4.1 of Hallin and Paindaveine (2007) that Assumption (A1) is satisfied: the Gaussian family $\mathcal{P}^{(n)}_{\boldsymbol{\phi}} := \{\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \mid \boldsymbol{\vartheta} \in \boldsymbol{\Theta}\}$ is ULAN, with contiguity rate $\boldsymbol{\nu}(n)$, central sequence

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := \begin{pmatrix} \boldsymbol{\Delta}^{I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \\ \boldsymbol{\Delta}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \\ \boldsymbol{\Delta}^{III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \end{pmatrix}, \quad \boldsymbol{\Delta}^{I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \begin{pmatrix} \boldsymbol{\Delta}^{I,1}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \\ \vdots \\ \boldsymbol{\Delta}^{I,m}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \end{pmatrix},$$

$$\boldsymbol{\Delta}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \begin{pmatrix} \Delta^{II,1}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \\ \vdots \\ \Delta^{II,m}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \end{pmatrix}, \quad \boldsymbol{\Delta}^{III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \begin{pmatrix} \boldsymbol{\Delta}^{III,1}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \\ \vdots \\ \boldsymbol{\Delta}^{III,m}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} \end{pmatrix},$$

where (letting $d_{ij} := d_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$ and $\boldsymbol{U}_{ij} := \boldsymbol{U}_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$)

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

$$\boldsymbol{\Delta}^{III,i}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := \frac{n_i^{-1/2}}{2}\boldsymbol{M}_k(\boldsymbol{V}_i)(\boldsymbol{V}_i^{\otimes 2})^{-1/2}\sum_{j=1}^{n_i} d_{ij}^2 \operatorname{vec}(\boldsymbol{U}_{ij}\boldsymbol{U}_{ij}') \qquad i = 1, \ldots, m,$$

and full-rank block-diagonal information matrix $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := \operatorname{diag}(\boldsymbol{\Gamma}^{I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}, \boldsymbol{\Gamma}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}, \boldsymbol{\Gamma}^{III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})$, where $\boldsymbol{\Gamma}^{I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := (\underline{\boldsymbol{\sigma}}^{-2} \otimes \boldsymbol{I}_k)\underline{\boldsymbol{V}}^{-1}$, $\boldsymbol{\Gamma}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := (k/2)\underline{\boldsymbol{\sigma}}^{-4}$, and

$$\boldsymbol{\Gamma}^{III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := k(k+2)\operatorname{diag}(\boldsymbol{H}_k(\boldsymbol{V}_1), \ldots, \boldsymbol{H}_k(\boldsymbol{V}_m))$$

(see Proposition 4.1 of Hallin and Paindaveine (2007), with $\mathcal{I}_k(\phi) = k$, $\mathcal{J}_k(\phi) = k(k+2)$, and $\mathcal{L}_k(\phi) := \mathcal{J}_k(\phi) - k^2 = 2k$), reducing, for $\sigma_1 = \cdots = \sigma_m = \sigma$ and $\boldsymbol{V}_1 = \cdots = \boldsymbol{V}_m = \boldsymbol{V}$, to

$$\begin{aligned} &\boldsymbol{\Gamma}^{I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := \sigma^{-2}(\boldsymbol{I}_m \otimes \boldsymbol{V}^{-1}), \qquad \boldsymbol{\Gamma}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := \frac{k}{2\sigma^4}\boldsymbol{I}_m, \qquad \text{and} \\ &\boldsymbol{\Gamma}^{III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} := k(k+2)(\boldsymbol{I}_m \otimes \boldsymbol{H}_k(\boldsymbol{V})). \end{aligned} \tag{3.5}$$

Assumption (A1) thus is satisfied. Turning to Assumption (A2), Lemmas 5.1 and 5.2 of Hallin and Paindaveine (2007) imply that, for any $\boldsymbol{g} \in \mathcal{F}_0^m$, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$,

(i) $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \operatorname{diag}(\boldsymbol{\Gamma}^{\boldsymbol{g},I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}, \boldsymbol{\Gamma}^{\boldsymbol{g},II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}, \boldsymbol{\Gamma}^{\boldsymbol{g},III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})$, where

$$\boldsymbol{\Gamma}^{\boldsymbol{g},I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = (\underline{\boldsymbol{\sigma}}^{-2} \otimes \boldsymbol{I}_k)\underline{\boldsymbol{V}}^{-1}, \qquad \boldsymbol{\Gamma}^{\boldsymbol{g},II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \frac{1}{4}\underline{C_k}(\boldsymbol{g})\underline{\boldsymbol{\sigma}}^{-4},$$

and

$$\boldsymbol{\Gamma}^{\boldsymbol{g},III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = (\underline{E_k}(\boldsymbol{g}) \otimes \boldsymbol{I}_{k_0})\operatorname{diag}(\boldsymbol{H}_k(\boldsymbol{V}_1), \ldots, \boldsymbol{H}_k(\boldsymbol{V}_m)),$$

reducing, for $\sigma_1 = \cdots = \sigma_m = \sigma$ and $\boldsymbol{V}_1 = \cdots = \boldsymbol{V}_m = \boldsymbol{V}$, to

$$\begin{aligned} &\boldsymbol{\Gamma}^{\boldsymbol{g},I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \sigma^{-2}(\boldsymbol{I}_m \otimes \boldsymbol{V}^{-1}), \qquad \boldsymbol{\Gamma}^{\boldsymbol{g},II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \frac{1}{4\sigma^4}\underline{C_k}(\boldsymbol{g}), \qquad \text{and} \\ &\boldsymbol{\Gamma}^{\boldsymbol{g},III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}} = \underline{E_k}(\boldsymbol{g}) \otimes \boldsymbol{H}_k(\boldsymbol{V}); \end{aligned} \tag{3.6}$$

(ii) $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ is *asymptotically linear* (see (2.1)), with $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}} = \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ given in (3.5).

Assumption (A2) thus is also satisfied, with $\mathcal{F}_\dagger = \mathcal{F}_0^m$, and the main result of Section 2 applies, provided that estimators satisfying Assumption (B2) are available. Clearly, $\hat{\boldsymbol{\Gamma}}_{\mathrm{Cov}} := \boldsymbol{\Gamma}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}$ is such an estimator for $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi},\boldsymbol{g}} = \boldsymbol{\Gamma}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, while the matrix $\hat{\boldsymbol{\Gamma}}_{\mathrm{Var}} =: \hat{\boldsymbol{\Gamma}}$ obtained by replacing, in $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}$, with $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ given in (3.6), the $E_k(g_i)$'s with their empirical counterparts, is another one for $\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta},\boldsymbol{\phi}}$.

The pseudo-Gaussian test $\varphi^{(n)}_{\mathrm{HP};\dagger}$ proposed in Section 5.2 of Hallin and Paindaveine (2007) is based on the asymptotically chi-squared distribution of

$$Q^{(n)}_{\mathrm{HP};\dagger} := \|[\boldsymbol{I} - \operatorname{proj}(\hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})]\hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}\|^2.$$

Clearly, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, with $\boldsymbol{\vartheta} \in \mathcal{M}(\boldsymbol{\Upsilon})$, $\boldsymbol{g} \in \mathcal{F}_0^m$, the projection in this quadratic form, up to a $o_{\mathrm{P}}(1)$ quantity, is onto $\mathcal{M}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$. Now, a robustification of the Gaussian quadratic form (2.2) would rather have been expected to be based on

$$\|[\boldsymbol{I} - \mathrm{proj}(\hat{\boldsymbol{\Gamma}}^{1/2}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})]\hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{\Delta}^{(n)}_{\hat{\boldsymbol{\vartheta}};\boldsymbol{\phi}}\|^2,$$

in which the projection—still up to $o_{\mathrm{P}}(1)$ quantities, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta};\boldsymbol{g}}$, with $\boldsymbol{\vartheta} \in \mathcal{M}(\boldsymbol{\Upsilon})$, $\boldsymbol{g} \in \mathcal{F}_0^m$—is onto $\mathcal{M}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{1/2}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$, hence not onto $\mathcal{M}((\boldsymbol{\Gamma}^{\boldsymbol{g}}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\nu}^{-1}(n)\boldsymbol{\Upsilon})$ as in $Q^{(n)}_{\mathrm{HP};\dagger}$. This structural difference between these two projections is all but intuitive, and even somewhat puzzling. A technical proof that $\varphi^{(n)}_{\mathrm{HP};\dagger}$ indeed produces a valid pseudo-Gaussian test is given in Hallin and Paindaveine (2007), but no intuitive justification is provided.

The results of Section 2 allow for solving that puzzle, by showing that $\varphi^{(n)}_{\mathrm{HP};\dagger}$ actually coincides with the test $\varphi^{(n)}_{\dagger}$ resulting from applying the theorem of Section 2. All the matrices involved being block-diagonal, the quadratic forms considered decompose into a sum of three separate ones, in $\boldsymbol{\Delta}^{I}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, $\boldsymbol{\Delta}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, and $\boldsymbol{\Delta}^{III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$, respectively. Each of those three quadratic forms involves a matrix of the form $[\boldsymbol{I} - \mathrm{proj}(\ldots)]$: $[\boldsymbol{I} - \mathrm{proj}^{I}_{\mathrm{HP};\dagger}]$, $[\boldsymbol{I} - \mathrm{proj}^{II}_{\mathrm{HP};\dagger}]$, and $[\boldsymbol{I} - \mathrm{proj}^{III}_{\mathrm{HP};\dagger}]$ for $Q^{(n)}_{\mathrm{HP};\dagger}$ and $[\boldsymbol{I} - \mathrm{proj}^{I}_{\dagger}]$, $[\boldsymbol{I} - \mathrm{proj}^{II}_{\dagger}]$, and $[\boldsymbol{I} - \mathrm{proj}^{III}_{\dagger}]$ for $Q^{(n)}_{\dagger}$, say.

Clearly, $\mathrm{proj}^{I}_{\dagger} = \boldsymbol{I}_{mk} = \mathrm{proj}^{I}_{\mathrm{HP};\dagger}$. Since $\boldsymbol{\Gamma}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}$ is proportional to the identity matrix,

$$\begin{aligned}\mathrm{proj}^{II}_{\dagger} &= \mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g},II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\nu}^{-1}_{II}(n)\boldsymbol{\Upsilon}_{II})\\ &= \mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g},II}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\nu}^{-1}_{II}(n)\boldsymbol{\Upsilon}_{II}) = \mathrm{proj}^{II}_{\mathrm{HP};\dagger}.\end{aligned}$$

Finally,

$$\begin{aligned}\mathrm{proj}^{III}_{\dagger} &= \mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g},III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\Gamma}^{III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}}\boldsymbol{\nu}^{-1}_{III}(n)\boldsymbol{\Upsilon}_{III})\\ &= \mathrm{proj}((\underline{E_k}(\boldsymbol{g}) \otimes \boldsymbol{H}_k(\boldsymbol{V}))^{-1/2}(\boldsymbol{I}_m \otimes \boldsymbol{H}_k(\boldsymbol{V}))\boldsymbol{\nu}^{-1}_{III}(n)\boldsymbol{\Upsilon}_{III})\\ &= \mathrm{proj}((\underline{E_k}(\boldsymbol{g}) \otimes \boldsymbol{H}_k(\boldsymbol{V}))^{-1/2}(\boldsymbol{I}_m \otimes \boldsymbol{H}_k(\boldsymbol{V}))(\boldsymbol{\Lambda}^{(n)} \otimes \boldsymbol{I}_{k_0})^{-1}(\boldsymbol{1}_m \otimes \boldsymbol{I}_{k_0}))\\ &= \mathrm{proj}([(\underline{E_k}(\boldsymbol{g}))^{-1/2}(\boldsymbol{\Lambda}^{(n)})^{-1}\boldsymbol{1}_m] \otimes (\boldsymbol{H}_k(\boldsymbol{V}))^{1/2}),\end{aligned}$$

which, since $\mathrm{proj}(\boldsymbol{A}) = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, and since $\boldsymbol{H}_k(\boldsymbol{V})$ has full rank, yields

$$\begin{aligned}\mathrm{proj}^{III}_{\dagger} &= \mathrm{proj}([(\underline{E_k}(\boldsymbol{g}))^{-1/2}(\boldsymbol{\Lambda}^{(n)})^{-1}\boldsymbol{1}_m] \otimes (\boldsymbol{H}_k(\boldsymbol{V}))^{-1/2})\\ &= \mathrm{proj}((\underline{E_k}(\boldsymbol{g}) \otimes \boldsymbol{H}_k(\boldsymbol{V}))^{-1/2}(\boldsymbol{\Lambda}^{(n)} \otimes \boldsymbol{I}_{k_0})^{-1}(\boldsymbol{1}_m \otimes \boldsymbol{I}_{k_0}))\\ &= \mathrm{proj}((\boldsymbol{\Gamma}^{\boldsymbol{g},III}_{\boldsymbol{\vartheta};\boldsymbol{\phi}})^{-1/2}\boldsymbol{\nu}^{-1}_{III}(n)\boldsymbol{\Upsilon}_{III}) = \mathrm{proj}^{III}_{\mathrm{HP};\dagger},\end{aligned}$$

which establishes the desired result.

Acknowledgements

The research of Davy Paindaveine is supported by a Mandat d'Impulsion Scientifique of the Fonds National de la Recherche Scientifique, Communauté française de Belgique.

References

Efron, B. (1969). Student's t-test under symmetry conditions, *Journal of American Statistical Association*, **63**, 1278–1302.

Hájek, I., Šidák, Z. and Sen, P. K. (1999). *Theory of Rank Tests*, 2nd edition, Academic Press, New York.

Hallin, M. and Paindaveine, D. (2007). Optimal tests for homogeneity of covariance, scale, and shape, *Journal of Multivariate Analysis* (to appear).

Hallin, M., Paindaveine, D. and Verdebout, T. (2007). Optimal pseudo-Gaussian tests for common principal components (submitted).

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

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

Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics*: *Some Basic Concepts*, 2nd edition, Springer-Verlag, New York.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition, Springer-Verlag, New York.

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.

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

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*, Springer-Verlag, New York.

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