Preliminary Test Estimation for Multi-sample Principal Components

Davy Paindaveine, Rondrotiana Joséa Rasoafaraniaina and Thomas Verdebout

ECARES and Département de Mathématique
Université libre de Bruxelles (ULB)
Avenue F.D. Roosevelt, 50 - ECARES, CP 114/04
and Campus Plaine, CP 210, bvd du triomphe
B-1050 Bruxelles, BELGIUM

Abstract

In this paper, we consider point estimation in a multi-sample principal components setup, in a situation where it is suspected that the hypothesis of *common principal components* (CPC) holds. We propose preliminary test estimators of the various principal eigenvectors. We derive their asymptotic distributions (i) under the CPC hypothesis, (ii) under sequences of hypotheses that are contiguous to the CPC hypothesis, and (iii) away from the CPC hypothesis. We conduct a Monte-Carlo study that shows that the proposed estimators perform well, particularly so in the Gaussian case.

Keywords: Preliminary test estimation, Common Principal Components

1. Introduction

Principal Component Analysis (PCA) is arguably one of the most popular multivariate methods. In this paper, we consider PCA in a multi-sample context. Consider $m(> 1)$ mutually independent samples of p-vectors $\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i}$, $i = 1, \ldots, m$, with respective sample sizes $n_1, \ldots, n_m$, such that for any i, the $\mathbf{X}_{ij}$'s form a random sample from a distribution with mean $\boldsymbol{\theta}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. In the ith population, the rth principal component scores are

$$(\boldsymbol{\beta}_i^{(r)})'\mathbf{X}_{i1}, \ldots, (\boldsymbol{\beta}_i^{(r)})'\mathbf{X}_{in_i}, \tag{1.1}$$

where $\boldsymbol{\beta}_i^{(r)}$ is the unit eigenvector associated with the rth largest eigenvalue of $\boldsymbol{\Sigma}_i$. In other words, $\boldsymbol{\beta}_i^{(r)}$ is the rth column vector in the matrix $\boldsymbol{\beta}_i$ from the factorization $\boldsymbol{\Sigma}_i = \boldsymbol{\beta}_i \boldsymbol{\Lambda}_i \boldsymbol{\beta}_i'$, where $\boldsymbol{\beta}_i \in \mathcal{SO}_p := \{\mathbf{O} \in \mathbb{R}^{p\times p} : \det(\mathbf{O}) = 1 \text{ and } \mathbf{O}^{-1} = \mathbf{O}\}$ and $\boldsymbol{\Lambda}_i := \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$ (with $\lambda_{i1} \geq \lambda_{i2} \geq \ldots \geq \lambda_{ip}$).

If no extra assumptions are adopted, then m eigenvectors matrices, namely $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$, are to be estimated, each on the basis of the observations from the corresponding population, which typically leads to running m PCAs independently. However, it is quite common that the covariance matrices $\mathbf{\Sigma}_i$, $i = 1, \ldots, m$ are linked in some way. Possible links include (i) *Homogeneity*: $\mathbf{\Sigma}_1 = \ldots = \mathbf{\Sigma}_m$, (ii) *Proportionality*: $\mathbf{\Sigma}_i = \rho_i \mathbf{\Sigma}_1$, $i = 2, \ldots, m$ for some positive scalar factors $\rho_1, \ldots, \rho_m$, or (iii) *Common Principal Components (CPC)*: $\mathbf{\Sigma}_i = \boldsymbol{\beta} \mathbf{\Lambda}_i \boldsymbol{\beta}'$, $i = 1, \ldots, m$, with a common $\boldsymbol{\beta} \in \mathcal{SO}_p$. In (i)-(iii), the $\mathbf{\Sigma}_i$'s share the same eigenvectors, leading to a common eigenvectors matrix $\boldsymbol{\beta} \in \mathcal{SO}_p$; therefore, it is natural to base estimation of $\boldsymbol{\beta}$ on the pooled sample collecting all $\mathbf{X}_{ij}$'s, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$.

The CPC model introduced by Flury (1984) is the most flexible model among those in (i)-(iii) above. This flexibility has made the model quite popular in the past decades : Flury (1984, 1986) considered Gaussian maximum likelihood estimators and derived the likelihood ratio test for the CPC structure. Estimation of common subspaces was considered in Schott (1988) and Fujioka (1993). Boente and Orellana (2001), Boente (2002, 2006) proposed inference methods that are robust to possible outliers, whereas Hallin *et al.* (2008, 2010a,b, 2013, 2014) proposed various parametric (*pseudo-Gaussian*) and nonparametric (rank-based) procedures that combine validity- and efficiency-robustness. Inference for functional CPC has recently been considered in Benko (2009).

In this paper, we tackle the problem of estimating $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ in a situation where it is suspected (but not certain) that the CPC hypothesis holds. In such a setup, one may be tempted to estimate $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ by a constrained estimator $\hat{\boldsymbol{\beta}}$ of the common eigenvectors matrix $\boldsymbol{\beta} := \boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_m$ under the CPC hypothesis. While this will increase efficiency over the unconstrained estimator $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_m$ if the CPC hypothesis indeed holds, it will introduce some bias if this hypothesis does not hold. We propose here a *preliminary test estimator (PTE)* of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$, that achieves a trade-off between the constrained and unconstrained estimators. PTEs were first introduced by Bancroft (1944) in an ANOVA setup. Since then, PTEs have been used in various contexts, including covariance matrices estimation, linear regression models, and time series analysis; see Toyoda and Wallace (1975), Sen and Saleh (1987), and Maeyama (2011), respectively. We refer to the monograph Saleh (2006) for a modern account on PTEs.

Quite naturally, the proposed PTE will select the constrained estimator of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ or an unconstrained one based on the outcome of a test for the null hypothesis of CPC. In the sequel, the unconstrained estimator will be the eigenvectors matrices $(\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_m)$ obtained from the various empirical covariance matrices $\mathbf{S}_1, \ldots, \mathbf{S}_m$, the constrained one will be based on the Flury (1984, 1986) Gaussian maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of the common eigenvectors matrix $\boldsymbol{\beta}$, and the CPC hypothesis will be tested by using a pseudo-Gaussian test $\phi^{(n)}$ that is similar in spirit to the one proposed in Hallin (2010a); see Section 2 for details. Writing $\mathbb{I}[A]$ for the indicator function associated with the event A and using the classical notation for test functions (that is, $\phi^{(n)} = 1$ corresponds to rejection of the null, while $\phi^{(n)} = 0$ indicates non-rejection), the proposed PTE is of the form

$$(\hat{\boldsymbol{\beta}}_1^{\mathrm{PT}}, \ldots, \hat{\boldsymbol{\beta}}_m^{\mathrm{PT}}) := (\hat{\boldsymbol{\beta}}, \ldots, \hat{\boldsymbol{\beta}})\mathbb{I}[\phi^{(n)} = 0] + (\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_m)\mathbb{I}[\phi^{(n)} = 1]. \tag{1.2}$$

In other words, the PTE coincides with the unconstrained estimator when the null of CPC is rejected and with the constrained estimator when the null is not rejected.

Below, we study the asymptotic properties of this PTE in various elliptically symmetric scenarios, namely (i) under the CPC hypothesis, (ii) under sequences of hypotheses that are contiguous to the CPC hypothesis, and (iii) under fixed distributions that do not satisfy the CPC structure. As we show through simulations, the proposed PTE provides a nice trade-off, in terms of efficiency, between the popular constrained and unconstrained estimators.

The outline of the paper is as follows. In Section 2, we describe the Gaussian and pseudo-Gaussian procedures for fitting a CPC model and for testing the CPC hypothesis. We also define the proposed PTE there. In Section 3, we derive the asymptotic properties of this PTE under various scenarios. In Section 4, we report the results of Monte-Carlo simulations that compare the performances of the PTE with those of the constrained and unconstrained estimators. Finally, an appendix collects the proofs of the technical results.

2. Preliminary test estimator

2.1. Gaussian and Pseudo-Gaussian inference for CPC

Let $(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})$, $i = 1, \ldots, m$, be mutually independent random samples from the p-variate normal distribution with respective mean vectors $\boldsymbol{\theta}_i$ and positive definite covariance matrices $\boldsymbol{\Sigma}_i$, $i = 1, \ldots, m$. Flury (1984) derived the corresponding (Gaussian) maximum likelihood estimators of the common eigenvectors matrix $\boldsymbol{\beta}$ in the CPC model described in the introduction, as well as the (Gaussian) likelihood ratio test (LRT) for the null hypothesis $\mathcal{H}^{\mathrm{CPC}}$ of CPC.

For point estimation in the CPC model, the likelihood equations for $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(p)})$ and the corresponding eigenvalues $\lambda_{ir} = (\boldsymbol{\beta}^{(r)})'\boldsymbol{\Sigma}_i\boldsymbol{\beta}^{(r)}$, $i = 1, \ldots, m$, $r = 1, \ldots, p$, are

$$(\boldsymbol{\beta}^{(r)})'\bigg(\sum_{i=1}^{m} n_i \frac{\lambda_{ir} - \lambda_{is}}{\lambda_{ir}\lambda_{is}} \mathbf{S}_i\bigg)\boldsymbol{\beta}^{(s)} = 0, \quad r, s = 1, \ldots, p \ \text{ with } \ r \neq s, \tag{2.3}$$

$$(\boldsymbol{\beta}^{(r)})'\mathbf{S}_i\boldsymbol{\beta}^{(r)} = \lambda_{ir}, \quad i = 1, \ldots, m, \ r = 1, \ldots, p, \qquad (\boldsymbol{\beta}^{(r)})'\boldsymbol{\beta}^{(s)} = \delta_{rs}, \quad r, s = 1, \ldots, p, \tag{2.4}$$

where δ_{rs} is the usual Kronecker symbol and

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

is the empirical covariance matrix associated with the ith sample, $i = 1, \ldots, m$. The likelihood equations (2.3)-(2.4) cannot be solved explicitly, but an algorithm for solving them numerically has been proposed by Flury and Gautschi (1986).

For hypothesis testing, the LRT, in the present Gaussian setup, rejects the null of CPC for small values of

$$\Lambda^{(n)} = \prod_{i=1}^{m} \left(\frac{\det(\hat{\boldsymbol{\beta}}' \mathbf{S}_i \hat{\boldsymbol{\beta}})}{\det(\mathrm{diag}(\hat{\boldsymbol{\beta}}' \mathbf{S}_i \hat{\boldsymbol{\beta}}))} \right)^{n_i/2}, \tag{2.5}$$

where $\hat{\boldsymbol{\beta}}$ denotes the Gaussian maximum likelihood estimator of the common eigenvectors matrix under the null and $\mathrm{diag}(\mathbf{A})$ is the diagonal matrix having the same diagonal as $\mathbf{A}$. The intuition behind (2.5) is clear : under the null of CPC, $\hat{\boldsymbol{\beta}}' \mathbf{S}_i \hat{\boldsymbol{\beta}}$ should be close to diagonal, so that $\det(\hat{\boldsymbol{\beta}}' \mathbf{S}_i \hat{\boldsymbol{\beta}})$ and $\det(\mathrm{diag}(\hat{\boldsymbol{\beta}}' \mathbf{S}_i \hat{\boldsymbol{\beta}}))$ should be approximately equal, leading to a large (that is, close to one) value of the likelihood ratio statistic $\Lambda^{(n)}$. On the contrary, small values of $\Lambda^{(n)}$ provide evidence against the null of CPC. Flury (1986) shows that, under the null, $-2 \log \Lambda^{(n)}$ is asymptotically chi-square with $(m-1)s$ degrees of freedom, where we let $s := p(p-1)/2$. Consequently, the LRT rejects the null of CPC at asymptotic level α whenever

$$-2 \log \Lambda^{(n)} > \chi^2_{(m-1)s, 1-\alpha},$$

where $\chi^2_{\ell, 1-\alpha}$ denotes the α-upper quantile of the chi-square distribution with ℓ degrees of freedom. It follows from the results in Hallin *et al.* (2013) that, in the Gaussian case, this test is asymptotically optimal in the Le Cam sense.

As it is often the case, however, this Gaussian LRT unfortunately is highly sensitive to violations of the Gaussian assumptions. Away from the multinormal case, this test may be either overly conservative or extremely liberal. As a reaction to this lack of *validity-robustness*, we propose using the following pseudo-Gaussian test $\phi^{(n)}$. Writing $d_{ij}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) := \big((\mathbf{X}^{(n)}_{ij} - \boldsymbol{\theta})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}^{(n)}_{ij} - \boldsymbol{\theta})\big)^{1/2}$ for the Mahalanobis distance between $\mathbf{X}^{(n)}_{ij}$ and $\boldsymbol{\theta}$ in the metric associated with $\boldsymbol{\Sigma}$, let

$$\hat{\kappa}_i := \frac{p}{p+2} \frac{\frac{1}{n_i} \sum_{j=1}^{n_i} d^4_{ij}(\bar{\mathbf{X}}_i, \mathbf{S}_i)}{\big(\frac{1}{n_i} \sum_{j=1}^{n_i} d^2_{ij}(\bar{\mathbf{X}}_i, \mathbf{S}_i)\big)^2} - 1, \quad i = 1, \ldots, m$$

(note that $\hat{\kappa}_i$ is a consistent estimator of the kurtosis coefficient $\kappa_p(f_i) := p\mathrm{E}[d^4_{i1}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)]/((p+2)\mathrm{E}^2[d^2_{i1}(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)]) - 1$ associated with the ith population). Further define $\hat{\mathbf{\Lambda}}_i = \mathrm{diag}(\hat{\lambda}_{i1}, \ldots, \hat{\lambda}_{ip})$, $i = 1, \ldots, m$, based on the eigenvalues of $\mathbf{S}_i$, $i = 1, \ldots, m$, and put $\hat{\boldsymbol{\nu}}_i := \mathrm{diag}(\hat{\nu}_{12}^{(i)}, \hat{\nu}_{23}^{(i)}, \ldots, \hat{\nu}_{(p-1)p}^{(i)})$, $i = 1, \ldots, m$, with

$$\hat{\nu}_{j\ell}^{(i)} := \frac{n(1+\hat{\kappa}_i)\hat{\lambda}_{ij}\hat{\lambda}_{i\ell}}{n_i(\hat{\lambda}_{ij} - \hat{\lambda}_{i\ell})^2}.$$

Then the test $\phi^{(n)}$ rejects the null of CPC at asymptotic level α whenever

$$Q_n := \sum_{i,i'=1}^{m} (\mathrm{ovec}(\mathbf{T}_i))' \left[\delta_{ii'}\mathbf{I}_s - \hat{\boldsymbol{\nu}}_i^{-1/2}\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}_{i'}^{-1/2}\right] \mathrm{ovec}(\mathbf{T}_{i'}) > \chi^2_{(m-1)s, 1-\alpha}, \tag{2.6}$$

where $\mathbf{T}_i := n_i^{1/2}(1+\hat{\kappa}_i)^{-1/2}\hat{\mathbf{\Lambda}}_i^{-1/2}\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}\hat{\mathbf{\Lambda}}_i^{-1/2}$, $\hat{\boldsymbol{\nu}} := (\sum_{i=1}^m \hat{\boldsymbol{\nu}}_i^{-1})^{-1}$ and $\mathrm{ovec}(\mathbf{A})$ is the $p(p-1)/2$-vector stacking the upper-diagonal elements of the $p \times p$ matrix $\mathbf{A}$ on top of each other. Under the CPC hypothesis, the upper-triangular elements of $\hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}$ should be close to zero for any $i = 1, \ldots, m$, so that Q_n should then be small. On the contrary, large values of Q_n indicate deviations from the CPC hypothesis.

Hallin *et al.* (2010a) considered a slightly different version $\tilde{Q}_n$ of Q_n, that is obtained by replacing in Q_n the unconstrained estimators of the eigenvalues $\hat{\lambda}_{ij}$ above by the ML eigenvalue estimators under the null of CPC, that is, by the eigenvalues solutions of (2.3)-(2.4). By using the same techniques as in Hallin *et al.* (2010a), one can easily show that, in the Gaussian case, $Q_n = \tilde{Q}_n + o_{\mathrm{P}}(1) = -2\log\Lambda^{(n)} + o_{\mathrm{P}}(1)$ as $n \to \infty$ under the null of CPC (hence, also under sequences of contiguous alternatives), but that, away from the Gaussian case, both Q_n and $\tilde{Q}_n$ remain asymptotically chi-square with $(m-1)s$ degrees of freedom under the null; more precisely, they remain chi-square with $(m-1)s$ degrees of freedom provided that observations are sampled from m elliptical distributions with finite fourth-order moments. It readily follows that $\phi^{(n)}$ is, as announced, a *pseudo-Gaussian* version of the Gaussian LRT, in the sense that it is asymptotically equivalent to the latter (hence, is asymptotically optimal in the Le Cam sense) in the Gaussian case, while extending the validity of the Gaussian LRT much beyond the Gaussian case.

2.2. The proposed preliminary test estimator

As explained in the Introduction, we consider here the estimation of $\underline{\boldsymbol{\beta}} := ((\mathrm{vec}\,\boldsymbol{\beta}_1)', \ldots, (\mathrm{vec}\,\boldsymbol{\beta}_m)')'$, where $\boldsymbol{\beta}_i$ is the eigenvectors matrix associated with $\mathbf{\Sigma}_i$, $i = 1, \ldots, m$. If the various covariance matrices do not share a particular structure, the Gaussian maximum likelihood estimator of $\underline{\boldsymbol{\beta}}$ is $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} := ((\mathrm{vec}\,\hat{\boldsymbol{\beta}}_1)', \ldots, (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_m)')'$, where $\hat{\boldsymbol{\beta}}_i$ is the eigenvectors matrix associated with the empirical covariance matrix $\mathbf{S}_i$, $i = 1, \ldots, m$. We refer to this estimator as the *unconstrained* estimator. Under the CPC hypothesis, $\underline{\boldsymbol{\beta}}$ is of the form $\underline{\boldsymbol{\beta}} = ((\mathrm{vec}\,\boldsymbol{\beta})', \ldots, \mathrm{vec}\,\boldsymbol{\beta})')'$ for some $\boldsymbol{\beta} \in SO_p$, and the corresponding Gaussian

MLE is the *constrained* estimator $\underline{\hat{\boldsymbol{\beta}}}^{\mathrm{C}} := ((\mathrm{vec}\,\hat{\boldsymbol{\beta}})', \ldots, (\mathrm{vec}\,\hat{\boldsymbol{\beta}})')'$, where the eigenvectors matrix $\hat{\boldsymbol{\beta}}$ is the solution of the likelihood equations (2.3)-(2.4).

If one suspects that the CPC hypothesis may hold, then it is natural to adopt, in the spirit of Saleh (2006), a *preliminary test estimator* (PTE) of $\underline{\boldsymbol{\beta}}$, that will be the unconstrained (rep., constrained) estimator when a test for CPC does (resp., does not) lead to rejection of the null. When based on the test $\phi^{(n)}$ above, the proposed PTE is thus

$$\begin{aligned} \underline{\hat{\boldsymbol{\beta}}}^{\mathrm{PT}} &= ((\mathrm{vec}\,\hat{\boldsymbol{\beta}}_1^{\mathrm{PT}})', \ldots, (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_m^{\mathrm{PT}})')' \\ &= \underline{\hat{\boldsymbol{\beta}}}^{\mathrm{C}}\, \mathbb{I}[Q_n \leq \chi^2_{(m-1)s,1-\alpha}] + \underline{\hat{\boldsymbol{\beta}}}^{\mathrm{U}}\, \mathbb{I}[Q_n > \chi^2_{(m-1)s,1-\alpha}]. \end{aligned} \tag{2.7}$$

In Section 3, we study the asymptotic properties of this PTE under elliptical distributions with finite fourth-order moments, while in Section 4, we compare the finite-sample performances of $\underline{\hat{\boldsymbol{\beta}}}^{\mathrm{U}}$, $\underline{\hat{\boldsymbol{\beta}}}^{\mathrm{C}}$ and $\underline{\hat{\boldsymbol{\beta}}}^{\mathrm{PT}}$ via Monte-Carlo simulations.

3. Asymptotics

The aim of the present section is to investigate the asymptotic behavior of the estimator $\underline{\hat{\boldsymbol{\beta}}}^{\mathrm{PT}}$ in (2.7). To do so, we consider triangular arrays of observations of the form

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

where the sequences $n_i = n_i^{(n)}$ are such that $n_i \to \infty$ as $n \to \infty$ for all $i = 1, \ldots, m$. The following distributional assumption will be required.

Assumption (A). (i) The p-variate observations $\mathbf{X}_{ij}^{(n)}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$ are mutually independent. (ii) For any $i = 1, \ldots, m$, $\mathbf{X}_{i1}^{(n)}, \ldots, \mathbf{X}_{in_i}^{(n)}$ have a common elliptical distribution with location $\boldsymbol{\theta}_i$, positive-definite covariance matrix $\boldsymbol{\Sigma}_i$ and radial density f_i such that $\mathrm{E}[\|\mathbf{X}_{i1}^{(n)}\|^4] < \infty$. (iii) The m-tuple of radial densities $\underline{f} := (f_1, \ldots, f_m)$ is satisfies $\kappa_p(f_1) = \ldots = \kappa_p(f_m) (=: \kappa_p(\underline{f}))$.

Assumption (A)(iii) implies that we restrict to the *homokurtic* case. This assumption is adopted for ease of presentation only, as the results of the paper could be extended to the more general *heterokurtic* case in a rather straightforward way.

The eigenvectors in each $\boldsymbol{\beta}_i$, $i = \ldots, m$ have to be properly identified, which is guaranteed by the following assumption.

Assumption (B). For any $i = 1, \ldots, m$, the eigenvalues of $\boldsymbol{\Sigma}_i$ are pairwise different.

This assumption properly identifies, for any $i = 1, \ldots, m$, the eigenvalues matrices $\mathbf{\Lambda}_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$, with $\lambda_{i1} > \ldots > \lambda_{ip}$, and the corresponding eigenvectors matrices $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_i^{(1)}, \ldots, \boldsymbol{\beta}_i^{(p)}) \in \mathcal{SO}_p$ (up to unimportant sign changes in each column). While Assumption (B) is classical in (multi-sample) PCA, it is not in the CPC model; (there, the common eigenvectors will indeed be identified as soon as for any $r, s = 1, \ldots, p$ ($r \neq s$), there is at least one i such that $\lambda_{ir} \neq \lambda_{is}$, which leaves space for equality of the corresponding eigenvalues in some populations). However, since preliminary test estimation requires considering both types of eigenstructures (CPC and unconstrained multi-sample PCA), we need to adopt Assumption (B) throughout.

Finally, our asymptotic results also require controlling the sample sizes $n_i^{(n)}$.

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

Clearly, the asymptotic behavior of $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}}$, equivalently of

$$\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}} - \underline{\boldsymbol{\beta}}), \quad \text{with} \;\; \mathbf{R}^{(n)} := \mathrm{diag}\Big(\sqrt{r_1^{(n)}}, \ldots, \sqrt{r_m^{(n)}}\,\Big) \otimes \mathbf{I}_{p^2}$$

($\otimes$ denotes the classical Kronecker product and $\mathbf{I}_\ell$ is the $\ell \times \ell$ identity matrix), will crucially depend on the asymptotic behaviors of $\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}})$ and $\sqrt{n}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}})$, which are summarized in Theorem 1 below. Stating this result requires introducing the following notation. For any $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(p)}) \in \mathcal{SO}_p$, let $\mathbf{G}^{\boldsymbol{\beta}}_{p;rs} := \mathbf{e}_r \otimes \boldsymbol{\beta}^{(s)} - \mathbf{e}_s \otimes \boldsymbol{\beta}^{(r)}$ and $\mathbf{G}^{\boldsymbol{\beta}}_p := (\mathbf{G}^{\boldsymbol{\beta}}_{p;12}\, \mathbf{G}^{\boldsymbol{\beta}}_{p;13} \ldots \mathbf{G}^{\boldsymbol{\beta}}_{p;(p-1)p})$, where $\mathbf{e}_\ell$ stands for the ℓth vector of the canonical basis of $\mathbb{R}^p$. Letting $\nu_{rs}^{(i)} := (1 + \kappa_p(\underline{f}))^{-1}\lambda_{ir}\lambda_{is}/(\lambda_{ir} - \lambda_{is})^2$, further define

$$\boldsymbol{\nu}^{(i)} := \mathrm{diag}(\nu_{12}^{(i)}, \nu_{13}^{(i)}, \ldots, \nu_{(p-1)p}^{(i)}) \quad \text{and} \quad \underline{\boldsymbol{\nu}} := \left(\sum_{i=1}^m r_i (\boldsymbol{\nu}^{(i)})^{-1}\right)^{-1}.$$

For the sake of readability, the joint distribution of the $\mathbf{X}_{ij}^{(n)}$'s, under Assumptions (A)-(B), will throughout be denoted as $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}, \underline{f}}$ (rather than the more correct $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m, \mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_m, \underline{f}}$). In the sequel, the notation "$\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$" will mean that $\underline{\boldsymbol{\beta}}$ is of the form $\underline{\boldsymbol{\beta}} = (\boldsymbol{\beta}, \ldots, \boldsymbol{\beta})$ for some $\boldsymbol{\beta}$ in $\mathcal{SO}(p)$. The following result then readily follows from Hallin (2008).

Proposition 1. *Let Assumptions (A), (B) and (C) hold. Then,*

(i) under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}, \underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$, $\sqrt{n}\,\mathrm{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges weakly to a Gaussian random vector with mean zero and covariance matrix $\mathbf{G}^{\boldsymbol{\beta}}_p \underline{\boldsymbol{\nu}} (\mathbf{G}^{\boldsymbol{\beta}}_p)'$ (so that $\sqrt{n}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}})$ converges weakly to a Gaussian random vector with mean zero and covariance matrix $\mathbf{\Gamma}^{\mathrm{C}}_{\underline{f}}(\boldsymbol{\beta}, \underline{\mathbf{\Lambda}}) := \mathrm{diag}(\mathbf{G}^{\boldsymbol{\beta}}_p \underline{\boldsymbol{\nu}}^{(1)} (\mathbf{G}^{\boldsymbol{\beta}}_p)', \ldots, \mathbf{G}^{\boldsymbol{\beta}}_p \underline{\boldsymbol{\nu}}^{(m)} (\mathbf{G}^{\boldsymbol{\beta}}_p)'))$;

(ii) under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$, $\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}})$ *converges weakly to a Gaussian random vector with mean zero and covariance matrix* $\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) := \mathrm{diag}(\mathbf{G}_p^{\boldsymbol{\beta}_1}\underline{\boldsymbol{\nu}}^{(1)}(\mathbf{G}_p^{\boldsymbol{\beta}_1})', \ldots, \mathbf{G}_p^{\boldsymbol{\beta}_m}\underline{\boldsymbol{\nu}}^{(m)}(\mathbf{G}_p^{\boldsymbol{\beta}_m})')$.

The asymptotic behavior of $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}}$ does not only depend on the asymptotic behaviors of $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}}$ and $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}}$, but also on that of the test statistic Q_n in (2.6). It follows from Section 6 of Hallin *et al.* (2013) that, as $n \to \infty$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$,

$$Q_n = (\mathbf{S}^{(n)})'\big((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-} - \mathbf{R}\boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\mathbf{R}\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})\mathbf{R}\boldsymbol{\Upsilon})^{-}\boldsymbol{\Upsilon}'\mathbf{R}\big)\mathbf{S}^{(n)} + o_{\mathrm{P}}(1), \tag{3.8}$$

where $\boldsymbol{\Upsilon} := \mathbf{1}_m \otimes \mathbf{I}_{p^2}$ with $\mathbf{1}_m := (1, \ldots, 1)' \in \mathbb{R}^m$, $\mathbf{R} := \lim_{n\to\infty} \mathbf{R}^{(n)}$, $\mathbf{A}^{-}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{A}$, and where $\mathbf{S}^{(n)}$ is the *central sequence* associated with the model defined in Proposition 5.1 of Hallin *et al.* (2013). Still under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$, this central sequence is asymptotically normal with mean zero and covariance matrix $\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})$, so that, in view of Theorem 9.2.1 in Rao and Mitra (1971), Q_n is then asymptotically chi-square with $(m-1)s$ degrees of freedom.

Our goal is to obtain asymptotic results in three different scenarios :

(i) *the CPC hypothesis* $\mathcal{H}^{\mathrm{CPC}}$, that is, $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$, or equivalently, $\underline{\boldsymbol{\beta}} \in \mathcal{M}(\boldsymbol{\Upsilon}) \cap (\mathrm{vec}\,\mathcal{SO}_p)^m$ (throughout, $\mathcal{M}(\mathbf{A})$ stands for the vector subspace spanned by the columns of $\mathbf{A}$);

(ii) *the vicinity of* $\mathcal{H}^{\mathrm{CPC}}$, characterized by eigenvectors matrices of the form

$$\underline{\boldsymbol{\beta}}^{(n)} := \underline{\boldsymbol{\beta}} + (\sqrt{n}\,\mathbf{R}^{(n)})^{-1}\mathbf{b}^{(n)} \quad (\in (\mathrm{vec}\,\mathcal{SO}_p)^m), \quad \text{with } \underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}, \tag{3.9}$$

where $\mathbf{b}^{(n)}$ is a bounded sequence such that $\underline{\boldsymbol{\beta}}^{(n)} \notin \mathcal{M}(\boldsymbol{\Upsilon})$. It follows from the ULAN result in Hallin *et al.* (2013) that the corresponding distributions are actually contiguous to the CPC hypothesis;

(iii) *away from* $\mathcal{H}^{\mathrm{CPC}}$, characterized by eigenvectors matrices of the form

$$\underline{\boldsymbol{\beta}}^{(n)} := \underline{\boldsymbol{\beta}} + \mathbf{b}^{(n)} \quad (\in (\mathrm{vec}\,\mathcal{SO}_p)^m), \quad \text{with } \underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}, \tag{3.10}$$

where $\mathbf{b}^{(n)}$ is a bounded sequence such that $\underline{\boldsymbol{\beta}}^{(n)} \notin \mathcal{M}(\boldsymbol{\Upsilon})$;

The following result describes the asymptotic behavior of $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}}$ away from $\mathcal{H}^{\mathrm{CPC}}$ (see the appendix for the proof).

Theorem 1. *Let Assumptions (A), (B) and (C) hold. Then, under* $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$ *with* $\underline{\boldsymbol{\beta}}^{(n)}$ *away from* $\mathcal{H}^{\mathrm{CPC}}$ *(as in (3.10)),* $\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}} - \underline{\boldsymbol{\beta}}^{(n)}) = \sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) + o_{\mathrm{P}}(1)$ *as* $n \to \infty$.

As expected, since the test $\phi^{(n)}$ is consistent away from $\mathcal{H}^{\rm CPC}$, the preliminary test estimator $\hat{\underline{\boldsymbol{\beta}}}^{\rm PT}$, still away from $\mathcal{H}^{\rm CPC}$, is asymptotically equivalent to the unconstrained estimator $\hat{\underline{\boldsymbol{\beta}}}^{\rm U}$. The other main results of the paper concern the asymptotic behavior of $\hat{\underline{\boldsymbol{\beta}}}^{\rm PT}$ under $\mathcal{H}^{\rm CPC}$ and in the vicinity of $\mathcal{H}^{\rm CPC}$. To state those results, we let

$$\begin{aligned}\mathbf{D}^{(n)} &:= \Big(\mathbf{I}_{mp^2} - \mathrm{proj}\Big((\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{1/2}\mathbf{R}\boldsymbol{\Upsilon}\Big)\Big)((\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{S}^{(n)} \\ &=: \mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})((\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{S}^{(n)},\end{aligned} \tag{3.11}$$

with $\mathrm{proj}(\mathbf{A}) := \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}$. Since $(\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-}\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) = \frac{1}{2}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)')$ and $\frac{1}{2}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)')\mathbf{S}^{(n)} = \mathbf{S}^{(n)}$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\rm CPC}$ (see Hallin *et al.* (2013)), it follows from (3.8) that, as $n \to \infty$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\rm CPC}$,

$$Q_n = \|\mathbf{D}^{(n)}\|^2 + o_{\rm P}(1).$$

Putting $\gamma(Q) := \mathbb{I}[Q \leq \chi^2_{(m-1)s,1-\alpha}]$, we then have the following result

Theorem 2. *Let Assumptions (A), (B) and (C) hold. Then, under* $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$ *with* $\underline{\boldsymbol{\beta}}^{(n)}$ *in the vicinity of* $\mathcal{H}^{\rm CPC}$ *(as in (3.9)),* $\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\rm PT} - \underline{\boldsymbol{\beta}}^{(n)})$*, conditional on* $\mathbf{D}^{(n)} = \mathbf{D}$*, converges weakly to a Gaussian random vector with mean vector*

$$\begin{aligned}\boldsymbol{\mu}^{\rm PT}_{\rm Vic}(\mathbf{D}) := (1-\gamma(\|\mathbf{D}\|^2))((\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})(\mathbf{D} - (\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{1/2}\mathbf{b}) \\ +\gamma(\|\mathbf{D}\|^2)(\mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) - \mathbf{I}_{mp^2})\mathbf{b}\end{aligned}$$

and covariance matrix

$$\begin{aligned}\boldsymbol{\Gamma}^{\rm PT}_{\rm Vic}(\mathbf{D}) := (1-\gamma(\|\mathbf{D}\|^2))^2\Big((\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-} - ((\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})((\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\Big) \\ +\frac{1}{2}\gamma(\|\mathbf{D}\|^2)(1-\gamma(\|\mathbf{D}\|^2))(\mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)') + (\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)'))(\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})})'\mathbf{R}) \\ +\gamma^2(\|\mathbf{D}\|^2)\mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})(\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})})'\mathbf{R},\end{aligned}$$

with $\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})} := \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}'\mathbf{R}\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})\mathbf{R}\boldsymbol{\Upsilon})^{-}\boldsymbol{\Upsilon}'\mathbf{R}$*,* $\mathbf{R} := \lim_{n\to\infty}\mathbf{R}^{(n)}$ *and* $\mathbf{b} := \lim_{n\to\infty}\mathbf{b}^{(n)}$*.*

The corresponding unconditional asymptotic distribution of $\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\rm PT} - \underline{\boldsymbol{\beta}}^{(n)})$ can then be obtained from the fact that, under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$ with $\underline{\boldsymbol{\beta}}^{(n)}$ as in (3.9), $\mathbf{D}^{(n)}$ is asymptotically normal with mean vector $\boldsymbol{\xi} := \mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})(\boldsymbol{\Gamma}^{\rm U}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{1/2}\mathbf{b}$ and covariance matrix $\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})$. It directly follows from Theorem 2 that the asymptotic distribution of

$\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}} - \underline{\boldsymbol{\beta}}^{(n)})$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$ with $\underline{\boldsymbol{\beta}}^{(n)}$ as in (3.9) has density

$$\mathbf{y} \mapsto \int_{\mathbb{R}^{mp^2}} \phi^{(mp^2)}_{\boldsymbol{\mu}^{\mathrm{PT}}_{\mathrm{Vic}}(\mathbf{x}),\boldsymbol{\Gamma}^{\mathrm{PT}}_{\mathrm{Vic}}(\mathbf{x})}(\mathbf{y})\ \phi^{(mp^2)}_{\boldsymbol{\xi},\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}(\mathbf{x})\ \ d\mathbf{x}, \tag{3.12}$$

where $\phi^{(\ell)}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ stands for the probability density function of a Gaussian random ℓ-vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

It is easy to show that the proof of Theorem 2 covers the case $\mathbf{b}^{(n)} \equiv 0$, which provides the following conditional asymptotic distribution of the PTE under the CPC hypothesis (of course, unconditioning can be performed as above, based this time on the *central* multinormal asymptotic distribution of $\mathbf{D}^{(n)}$ under the CPC hypothesis).

Corollary 1. *Let Assumptions (A), (B) and (C) hold. Then, under the CPC hypothesis* ($\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$*, with* $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$), $\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}} - \underline{\boldsymbol{\beta}})$*, conditional on* $\mathbf{D}^{(n)} = \mathbf{D}$*, converges weakly to a Gaussian random vector with mean vector* $(\gamma(\|\mathbf{D}\|^2) - 1)((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^-)^{1/2}\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})\mathbf{D}$ *and covariance matrix* $\boldsymbol{\Gamma}^{\mathrm{PT}}_{\mathrm{Vic}}(\mathbf{D})$*.*

The complicated asymptotic distributions above make it difficult to compare the proposed PTE to its constrained and unconstrained antecedent through asymptotic relative efficiencies, and we therefore rather focus on finite-sample comparisons based on simulations.

4. Simulations

In this section, we compare the estimators $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}}$, $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}}$ and $\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}}$ via Monte-Carlo simulations. First we considered a bivariate setup ($p = 2$) involving two populations ($m = 2$), with 200 observations in each group ($n_1 = n_2 = 200$). We started with a Gaussian setup for which

- observations in the first group are randomly sampled from the bivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\beta}_1\boldsymbol{\Lambda}\boldsymbol{\beta}_1'$, with $\boldsymbol{\Lambda} = \mathrm{diag}(2,1)$ and $\boldsymbol{\beta}_1 = \mathbf{I}_2$, and
- observations in the second group are randomly sampled from the bivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\beta}_2\boldsymbol{\Lambda}\boldsymbol{\beta}_2'$, with either

$$\boldsymbol{\beta}_2 = \mathbf{I}_2 \qquad\qquad \text{(CPC)},$$

$$\boldsymbol{\beta}_2 = \begin{pmatrix} \cos\left(\frac{2\pi}{\sqrt{n_2}}\right) & -\sin\left(\frac{2\pi}{\sqrt{n_2}}\right) \\ \sin\left(\frac{2\pi}{\sqrt{n_2}}\right) & \cos\left(\frac{2\pi}{\sqrt{n_2}}\right) \end{pmatrix} \qquad \text{(vicinity of CPC)},$$

and

$$\boldsymbol{\beta}_2 = \begin{pmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & \cos(\frac{\pi}{3}) \end{pmatrix} \qquad \text{(away from CPC).}$$

We generated $2,500$ independent samples of this form and computed in each replication the constrained, unconstrained and preliminary test estimators (for the latter, the tests of CPC were conducted at level 5%). In the qth replication, we evaluated, for each estimator $\underline{\hat{\boldsymbol{\beta}}}(q) = ((\mathrm{vec}\,\hat{\boldsymbol{\beta}}_1(q))', (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_2(q))')'$, the corresponding estimation accuracy measure

$$\omega^{(1)}_{\underline{\hat{\boldsymbol{\beta}}}}(q) := \frac{\big((\boldsymbol{\beta}^{(1)}_1)'\hat{\boldsymbol{\beta}}^{(1)}_1(q)\big)^2 + \big((\boldsymbol{\beta}^{(1)}_2)'\hat{\boldsymbol{\beta}}^{(1)}_2(q)\big)^2}{2}.$$

Clearly, the average of cosines $\omega^{(1)}_{\underline{\hat{\boldsymbol{\beta}}}}(q)$ takes value in $[0,1]$ with large (resp., small) values indicating good (resp., poor) estimation.

In a second simulation, we considered a four-dimensional setup ($p = 4$) involving four populations ($m = 4$) with 200 observations in each group ($n_i = 200$, $i = 1,2,3,4$). We similarly started with the Gaussian case for which

- observations in the first two groups are randomly sampled from the four-dimensional normal distribution with mean zero and covariance matrix $\boldsymbol{\beta}_1\boldsymbol{\Lambda}\boldsymbol{\beta}_1'$, with $\boldsymbol{\Lambda} = \mathrm{diag}(4,3,2,1)$ and $\boldsymbol{\beta}_1 = \mathbf{I}_4$, and
- observations in the third and fourth groups are randomly sampled from the four-dimensional normal distribution with mean zero and covariance matrix $\boldsymbol{\beta}_2\boldsymbol{\Lambda}\boldsymbol{\beta}_2'$, with either

$$\boldsymbol{\beta}_2 = \mathbf{I}_4 \qquad \text{(CPC),}$$

$$\boldsymbol{\beta}_2 = \mathrm{diag}\left(\begin{pmatrix} \cos\left(\frac{2\pi}{\sqrt{n_4}}\right) & -\sin\left(\frac{2\pi}{\sqrt{n_4}}\right) \\ \sin\left(\frac{2\pi}{\sqrt{n_4}}\right) & \cos\left(\frac{2\pi}{\sqrt{n_4}}\right) \end{pmatrix}, \begin{pmatrix} \cos\left(\frac{2\pi}{\sqrt{n_4}}\right) & -\sin\left(\frac{2\pi}{\sqrt{n_4}}\right) \\ \sin\left(\frac{2\pi}{\sqrt{n_4}}\right) & \cos\left(\frac{2\pi}{\sqrt{n_4}}\right) \end{pmatrix}\right) \qquad \text{(vicinity of CPC),}$$

$$\boldsymbol{\beta}_2 = \mathrm{diag}\left(\begin{pmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & \cos(\frac{\pi}{3}) \end{pmatrix}, \begin{pmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & \cos(\frac{\pi}{3}) \end{pmatrix}\right) \qquad \text{(away from CPC).}$$

We also generated $2,500$ independent samples of this form and computed in each replication the constrained, unconstrained and preliminary test estimators (PTEs are still based on 5%-level tests of CPC). In the qth replication, we evaluated, for each estimator $\underline{\hat{\boldsymbol{\beta}}}(q) = ((\mathrm{vec}\,\hat{\boldsymbol{\beta}}_1(q))', (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_2(q))', (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_3(q))', (\mathrm{vec}\,\hat{\boldsymbol{\beta}}_4(q))')'$, the corresponding estimation accuracy measure

$$\omega^{(2)}_{\underline{\hat{\boldsymbol{\beta}}}}(q) := \frac{\sum_{j=1}^4 \big((\boldsymbol{\beta}^{(1)}_j)'\hat{\boldsymbol{\beta}}^{(1)}_j(q)\big)^2}{4}.$$

Figure 1 and Figure 3 provide the resulting boxplots of the $\omega^{(\ell)}_{\underline{\hat{\boldsymbol{\beta}}}}(q)$'s ($\ell = 1, 2$), $q = 1, \ldots, 2,500$ for the constrained, unconstrained and preliminary test estimators, while Figure 2 and Figure 4 show the corresponding boxplots for the t_6 distribution. Inspection of these figures reveals that, as expected, the PTE behaves (i) almost like the unconstrained estimator away from CPC, (ii) almost like the constrained estimator under CPC and (iii) achieves a balance between both estimators in the vicinity of CPC. Note that in the vicinity of CPC, the PTE is closer to the unconstrained estimator in the Gaussian case than in the t_6 case. This comes from the fact that the pseudo-Gaussian test $\phi^{(n)}$ is more powerful in the Gaussian case than under Student distributions (see, e.g. Hallin *et al.* 2010a). Note also that still in the vicinity of CPC, the constrained estimator performs better in the $(m = 4, p = 4)$ (second simulation) case than in the $(m = 2, p = 2)$ (first simulation) case. This only comes from the fact that the chosen local alternatives in the $(m = 4, p = 4)$ case are less severe than those in the $(m = 2, p = 2)$ case.

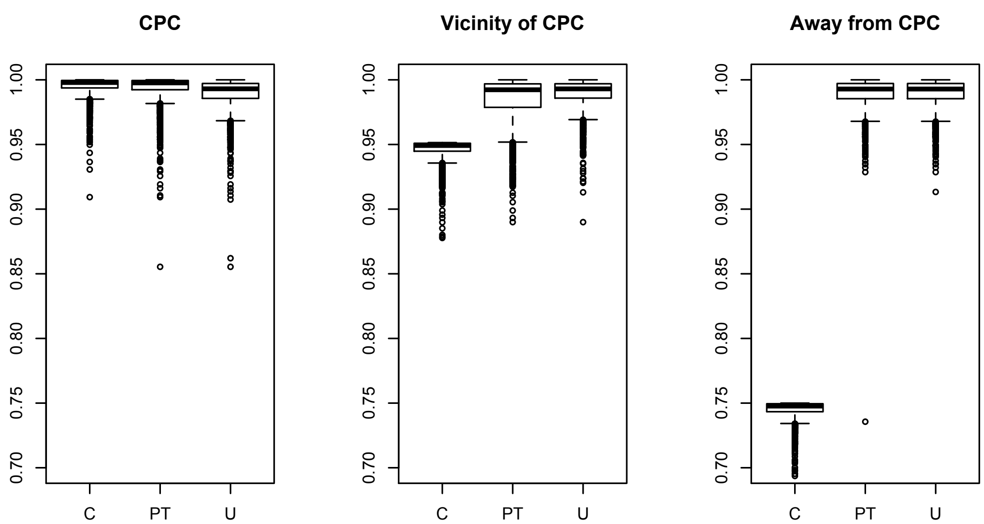

Figure 1: Boxplots of the estimation accuracy measures $\omega^{(1)}_{\underline{\hat{\boldsymbol{\beta}}}}(q)$ associated with the constrained, preliminary test estimation, and unconstrained estimators in $2,500$ replications of the bivariate Gaussian setup described in Section 4; the left panel corresponds to the CPC hypothesis, the middle one to the vicinity of CPC, and the right one to a distribution that is away from CPC.

Appendix A. Proofs

We start with the following preliminary result.

Lemma 1. *Let Assumptions (A), (B) and (C) hold. Then under* $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)}, \underline{f}}$ *away from* $\mathcal{H}^{\mathrm{CPC}}$*,* $n/Q_n = O_{\mathrm{P}}(1)$ *as* $n \to \infty$*.*

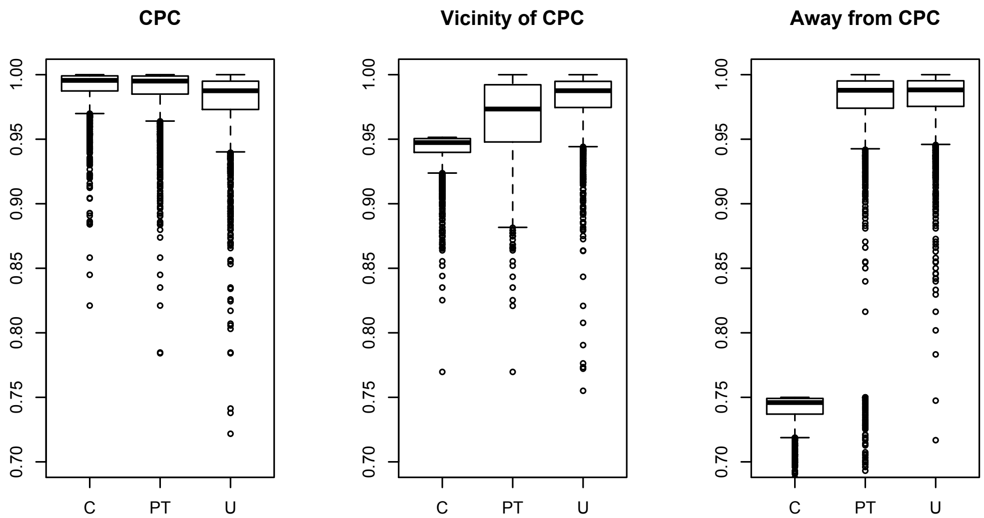

Figure 2: Boxplots of the estimation accuracy measures $\omega^{(1)}_{\underline{\hat{\boldsymbol{\beta}}}}(q)$ associated with the constrained, preliminary test estimation, and unconstrained estimators in $2,500$ replications of the bivariate t_6 setup described in Section 4; the left panel corresponds to the CPC hypothesis, the middle one to the vicinity of CPC, and the right one to a distribution that is away from CPC.

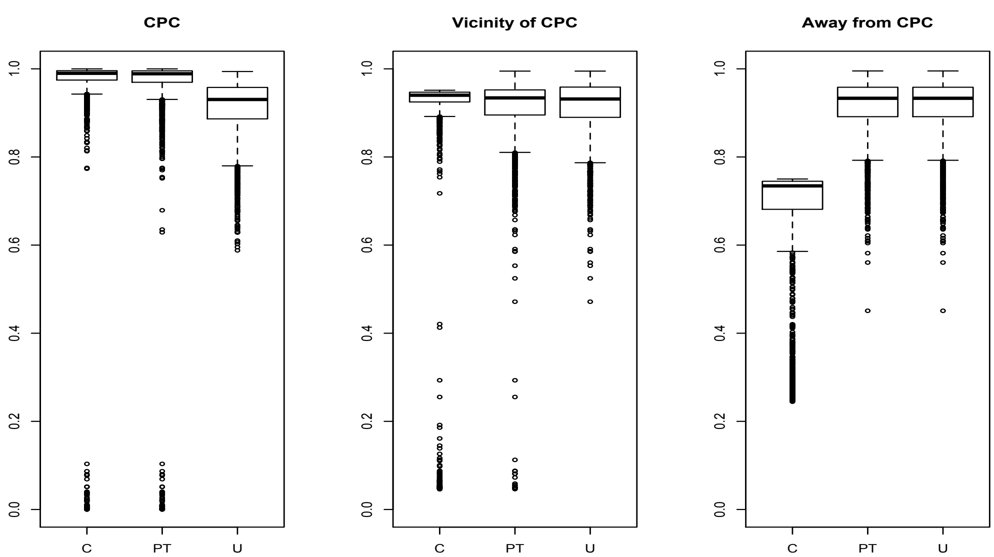

Figure 3: Boxplots of the estimation accuracy measures $\omega^{(2)}_{\underline{\hat{\boldsymbol{\beta}}}}(q)$ associated with the constrained, preliminary test estimation, and unconstrained estimators in $2,500$ replications of the 4-dimensional Gaussian setup described in Section 4; the left panel corresponds to the CPC hypothesis, the middle one to the vicinity of CPC, and the right one to a distribution that is away from CPC.

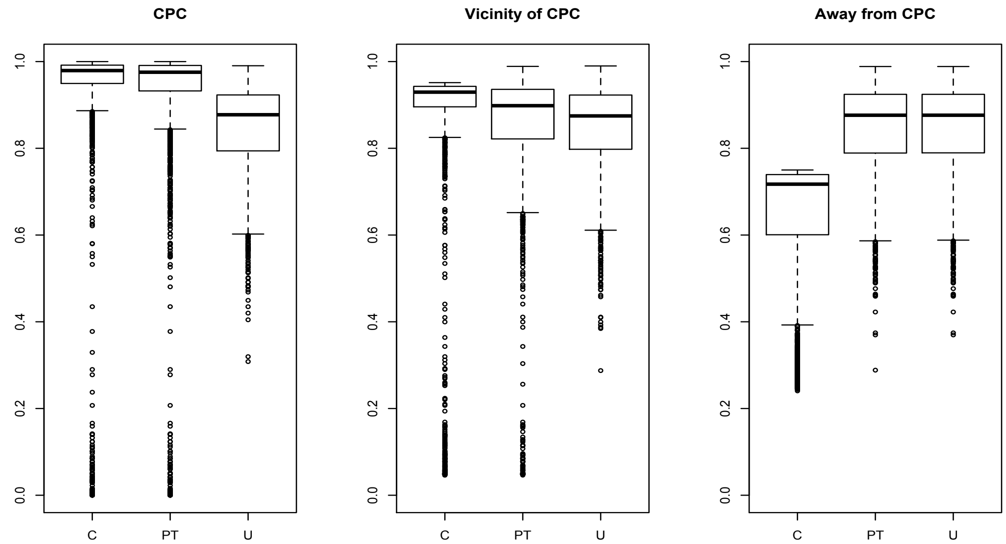

Figure 4: Boxplots of the estimation accuracy measures $\omega^{(2)}_{\underline{\hat{\boldsymbol{\beta}}}}(q)$ associated with the constrained, preliminary test estimation, and unconstrained estimators in $2,500$ replications of the 4-dimensional t_6 setup described in Section 4; the left panel corresponds to the CPC hypothesis, the middle one to the vicinity of CPC, and the right one to a distribution that is away from CPC.

Proof of Lemma 1. Unless otherwise mentioned, all infima/suprema in $\boldsymbol{\beta}$ in this proof are over $\mathcal{SO}_p$. For any $\boldsymbol{\beta} \in \mathcal{SO}_p$, denote as $Q_n(\boldsymbol{\beta})$ the random variable obtained by replacing $\hat{\boldsymbol{\beta}}$ with $\boldsymbol{\beta}$ in Q_n; see (2.6). For any $\boldsymbol{\beta}$, we have that

$$Q_n(\boldsymbol{\beta})/n = q(\boldsymbol{\beta}) + o_{\mathrm{P}}(1) \tag{A.1}$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$, where

$$\begin{aligned} q(\boldsymbol{\beta}) \quad &:= \quad \sum_{i,i'=1}^{m} (\mathrm{ovec}(\mathbf{t}_i(\boldsymbol{\beta})))' \left[\delta_{ii'}\mathbf{I}_s - \boldsymbol{\nu}_i^{-1/2}\boldsymbol{\nu}\boldsymbol{\nu}_{i'}^{-1/2}\right] \mathrm{ovec}(\mathbf{t}_{i'}(\boldsymbol{\beta})) \\ &=: \quad \begin{pmatrix} \mathrm{ovec}(\mathbf{t}_1(\boldsymbol{\beta})) \\ \vdots \\ \mathrm{ovec}(\mathbf{t}_m(\boldsymbol{\beta})) \end{pmatrix}' \mathbf{A} \begin{pmatrix} \mathrm{ovec}(\mathbf{t}_1(\boldsymbol{\beta})) \\ \vdots \\ \mathrm{ovec}(\mathbf{t}_m(\boldsymbol{\beta})) \end{pmatrix} \end{aligned}$$

is based on $\mathbf{t}_i(\boldsymbol{\beta}) := r_i^{1/2}(1+\kappa_i)^{-1/2}\boldsymbol{\Lambda}_i^{-1/2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_i\boldsymbol{\beta}\boldsymbol{\Lambda}_i^{-1/2}$, $\boldsymbol{\Lambda}_i = \mathrm{diag}(\lambda_{i1},\ldots,\lambda_{ip})$, $\boldsymbol{\nu} := (\sum_{i=1}^m \boldsymbol{\nu}_i^{-1})^{-1}$, $\boldsymbol{\nu}_i := \mathrm{diag}(\nu_{12}^{(i)},\nu_{23}^{(i)},\ldots,\nu_{(p-1)p}^{(i)})$, with $\nu_{j\ell}^{(i)} := (1+\kappa_i)\lambda_{ij}\lambda_{i\ell}/[r_i(\lambda_{ij}-\lambda_{i\ell})^2]$. Since $\mathbf{A}$ is positive definite and since, away from the CPC hypothesis, no eigenvectors matrix $\boldsymbol{\beta}$ can simultaneously diagonalize $\boldsymbol{\Sigma}_i$, $i = 1,\ldots,m$, we have that $q(\boldsymbol{\beta}) > 0$ for any $\boldsymbol{\beta} \in \mathcal{SO}_p$. The compacity of $\mathcal{SO}_p$ then implies that

$$\inf_{\boldsymbol{\beta}} q(\boldsymbol{\beta}) > 0. \tag{A.2}$$

Now, letting $R_n(\boldsymbol{\beta}) := Q_n(\boldsymbol{\beta})/n - q(\boldsymbol{\beta})$, it is easy to show that $\{R_n(\boldsymbol{\beta}) : n = 1, 2, \ldots\}$ is stochastically equicontinuous (see Newey (1991)) over $\mathcal{SO}_p$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)}, \underline{f}}$, in the sense that, for any $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that ($\|.\|$ denotes the Frobenius norm)

$$\limsup_{n\to\infty} \mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)}, \underline{f}} \Big[\sup_{\boldsymbol{\beta}} \sup_{\tilde{\boldsymbol{\beta}}:\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\|<\delta} |R_n(\tilde{\boldsymbol{\beta}}) - R_n(\boldsymbol{\beta})| > \eta \Big] < \varepsilon.$$

Jointly with (A.1), this implies that

$$\sup_{\boldsymbol{\beta}} |Q_n(\boldsymbol{\beta})/n - q(\boldsymbol{\beta})| = o_{\mathrm{P}}(1), \tag{A.3}$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)}, \underline{f}}$; see Newey (1991). Since

$$\begin{aligned} n/Q_n &\leq \Big| \frac{1}{Q_n/n} - \frac{1}{q(\hat{\boldsymbol{\beta}})} \Big| + \frac{1}{q(\hat{\boldsymbol{\beta}})} = \frac{|Q_n/n - q(\hat{\boldsymbol{\beta}})|}{q(\hat{\boldsymbol{\beta}}) Q_n/n} + \frac{1}{q(\hat{\boldsymbol{\beta}})} \\ &\leq \frac{\sup_{\boldsymbol{\beta}} |Q_n/n - q(\boldsymbol{\beta})|}{\inf_{\boldsymbol{\beta}} q(\boldsymbol{\beta}) \big(\inf_{\boldsymbol{\beta}} q(\boldsymbol{\beta}) - \sup_{\boldsymbol{\beta}} |Q_n(\boldsymbol{\beta})/n - q(\boldsymbol{\beta})| \big)} + \frac{1}{\inf_{\boldsymbol{\beta}} q(\boldsymbol{\beta})}, \end{aligned}$$

the result then follows from (A.2)-(A.3). □

Proof of Theorem 1. Letting $\gamma(t) := \mathbb{I}[t \leq \chi^2_{(m-1)s, 1-\alpha}]$, we have that

$$\begin{aligned} \sqrt{n}\, \mathbf{R}^{(n)} (\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}} - \underline{\boldsymbol{\beta}}^{(n)}) &= \sqrt{n}\, \mathbf{R}^{(n)} \big((1 - \gamma(Q_n)) \hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} + \gamma(Q_n) \hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)} \big) \\ &= \sqrt{n}\, \mathbf{R}^{(n)} (\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) + \sqrt{n}\, \mathbf{R}^{(n)} \gamma(Q_n) (\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}}), \end{aligned} \tag{A.4}$$

so that it remains to show that $\sqrt{n}\, \mathbf{R}^{(n)} \gamma(Q_n) (\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}})$ is $o_{\mathrm{P}}(1)$ as $n \to \infty$. Write then

$$\sqrt{n}\, \mathbf{R}^{(n)} \gamma(Q_n) (\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}}) = Q_n^{1/2} \gamma(Q_n) (n/Q_n)^{1/2} \mathbf{R}^{(n)} (\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}}).$$

Since $\mathbf{R}^{(n)}$ converges to $\mathbf{R}$ and since $(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}})$ is trivially $O_{\mathrm{P}}(1)$, it thus remains to show, in view of Lemma 1, that $Q_n^{1/2} \gamma(Q_n)$ is $o_{\mathrm{P}}(1)$ as $n \to \infty$. Since the positive quantity Q_n is not a $O_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)}, \underline{f}}$ away from CPC and $\lim_{t\to\infty} t^{1/2} \gamma(t) = 0$, the result follows. □

Proof of Theorem 2. First note that by using the fact that $Q_n - \|\mathbf{D}^{(n)}\|^2$ is $o_{\mathrm{P}}(1)$ under

$\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$ as $n \to \infty$, it is easy to show that under the same sequence of hypotheses

$$\begin{aligned}\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{PT}} - \underline{\boldsymbol{\beta}}^{(n)}) &= \sqrt{n}\,\mathbf{R}^{(n)}\big((1-\gamma(Q_n))\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} + \gamma(Q_n)\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)}\big) \\ &= (1-\gamma(Q_n))\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) + \gamma(Q_n)\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)}) \\ &= (1-\gamma(\|\mathbf{D}^{(n)}\|^2))\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) + \gamma(\|\mathbf{D}^{(n)}\|^2)\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)}) + o_{\mathrm{P}}(1)\end{aligned} \tag{A.5}$$

as $n \to \infty$. By contiguity, (A.5) also holds under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$. Therefore we need to obtain the asymptotic distribution (conditional to $\mathbf{D}^{(n)} = \mathbf{D}$) of

$$\begin{pmatrix} \sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)}) \\ \sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) \end{pmatrix}$$

under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$.

As $n \to \infty$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$, it follows from [11] and [15] that

$$\sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}) = \mathbf{R}^{(n)}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\mathbf{S}^{(n)} + o_{\mathrm{P}}(1) \tag{A.6}$$

and

$$\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}) = (\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-}\mathbf{S}^{(n)} + o_{\mathrm{P}}(1), \tag{A.7}$$

where $\mathbf{S}^{(n)}$ is the random vector in (3.11) and $\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}$ was defined in the statement of Theorem 2. Therefore, with $\underline{\boldsymbol{\beta}}^{(n)}$ in the vicinity of $\mathcal{H}^{\mathrm{CPC}}$ as in (3.9), we readily obtain

$$\begin{aligned}\sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)}) &= \sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}) + \sqrt{n}\mathbf{R}^{(n)}(\underline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}^{(n)}) \\ &= \mathbf{R}^{(n)}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\mathbf{S}^{(n)} - \mathbf{b}^{(n)} + o_{\mathrm{P}}(1)\end{aligned} \tag{A.8}$$

and

$$\begin{aligned}\sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) &= \sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}) + \sqrt{n}\,\mathbf{R}^{(n)}(\underline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}^{(n)}) \\ &= (\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-}\mathbf{S}^{(n)} - \mathbf{b}^{(n)} + o_{\mathrm{P}}(1),\end{aligned} \tag{A.9}$$

as $n \to \infty$ under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$, hence also (from contiguity) under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$.

Under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}},\underline{f}}$ with $\underline{\boldsymbol{\beta}} \in \mathcal{H}_{\mathrm{CPC}}$, $\mathbf{S}^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))$, where $\xrightarrow{\mathcal{D}}$ denotes weak convergence. Le Cam's third Lemma then implies that under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$,

$$\mathbf{S}^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})\mathbf{b}, \boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})), \tag{A.10}$$

with $\mathbf{b} := \lim_{n\to\infty} \mathbf{b}^{(n)}$. Recalling that $\mathbf{D}^{(n)} = \mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{S}^{(n)}$ (see (3.11)), (A.9), (A.8) and (A.10) then yield that, under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$,

$$\begin{pmatrix} \sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)}) \\ \sqrt{n}\,\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) \\ \mathbf{D}^{(n)} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix} (\mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) - \mathbf{I}_{mp^2})\mathbf{b} \\ \mathbf{0} \\ \mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})(\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{1/2}\mathbf{b} \end{pmatrix}, \mathbf{C}_1\right),$$

with

$$\mathbf{C}_1 := \begin{pmatrix} \mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})(\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})})'\mathbf{R} & \frac{1}{2}\mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)') & \mathbf{0} \\ \frac{1}{2}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)')(\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})})'\mathbf{R} & (\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-} & ((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) \\ \mathbf{0} & \mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2} & \mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) \end{pmatrix},$$

where we used in the computations the facts that $(\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-}\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) = \frac{1}{2}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)')$ and that $\frac{1}{2}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)')\mathbf{b} = \mathbf{b}$ (see Lemma A.1 of Hallin *et al.* (2013)). It follows that, still under $\mathrm{P}^{(n)}_{\underline{\boldsymbol{\beta}}^{(n)},\underline{f}}$, we have that, conditional to $\mathbf{D}^{(n)} = \mathbf{D}$,

$$\begin{pmatrix} \sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{C}} - \underline{\boldsymbol{\beta}}^{(n)}) \\ \sqrt{n}\mathbf{R}^{(n)}(\hat{\underline{\boldsymbol{\beta}}}^{\mathrm{U}} - \underline{\boldsymbol{\beta}}^{(n)}) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix} (\mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}) - \mathbf{I}_{mp^2})\mathbf{b} \\ ((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})(\mathbf{D} - (\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{1/2}\mathbf{b}) \end{pmatrix}, \mathbf{C}_2\right),$$

with

$$\mathbf{C}_2 := \begin{pmatrix} \mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})(\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})})'\mathbf{R} & \frac{1}{2}\mathbf{R}\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)') \\ \frac{1}{2}(\mathbf{I}_m \otimes \mathbf{G}^{\boldsymbol{\beta}}_p(\mathbf{G}^{\boldsymbol{\beta}}_p)')(\boldsymbol{\Upsilon}^{\boldsymbol{\Gamma}}_{(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})})'\mathbf{R} & (\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-} - ((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2}\mathbf{P}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}})((\boldsymbol{\Gamma}^{\mathrm{U}}_{\underline{f}}(\underline{\boldsymbol{\beta}},\underline{\boldsymbol{\Lambda}}))^{-})^{1/2} \end{pmatrix}.$$

The result then follows directly from (A.5).

□

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

[2] Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *The Annals of Mathematical Statistics* **15**, 190-204.

[3] Benko, M., Hardle, W., and Kneip, A. (2009). Common functional principal components. *The Annals of Statistics* **37**, 1-34.

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

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

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

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

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

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

[10] Fujioka, T. (1993). An approximate test for common principal component subspaces in two groups. *Annals of the Institute of Statistical Mathematics* **45**, 147-158.

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

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

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

[14] Hallin, M., Paindaveine, D. and Verdebout, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli* **19**, 2524-2556.

[15] Hallin, M., Paindaveine, D. and Verdebout, T. (2014). Efficient R-estimation of principal and common principal components. *Journal of the American Statistical Association* **109**, 1071-1083

[16] Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica* **59**, 1161-1167.

[17] Maeyama, Y., Tamaki, K., and Taniguchi, M. (2011). Preliminary test estimation for spectra. *Statistics and Probability Letters* **81**, 1580-1587.

[18] Rao, C. R., and Mitra, S. K. (1971). *Generalized Inverses of Matrices and its Applications*. John Wiley, New York.

[19] Saleh, A. M. E. (2006). *Theory of preliminary test and Stein-type estimation with applications*. John Wiley and Sons.

[20] Sen, P. K., and Saleh, A. E. (1987). On preliminary test and shrinkage M-estimation in linear models. *The Annals of Statistics* **15**, 1580-1592.

[21] Schott, J.R. (1988). Common principal component subspaces in two groups. *Biometrika* **75**, 229-236.

[22] Toyoda, T., and Wallace, T. D. (1975). Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test. *Journal of Econometrics* **3**, 395-404.