Robust tests for the common principal components model*

By Graciela Boente
Universidad de Buenos Aires and CONICET, Argentina
gboente@mate.dm.uba.ar
Ana M. Pires and Isabel M. Rodrigues
Departamento de Matemática and CEMAT, Instituto Superior Técnico, Lisboa, Portugal
ana.pires@math.ist.utl.pt isabel.rodrigues@math.ist.utl.pt

Abstract

The common principal components model for several groups of multivariate observations assumes equal principal axes but different variances along these axes among the groups. In this paper, the null hypothesis of a CPC model versus no restrictions on the scatter matrices is studied. Also, two statistics for testing proportionality against a common principal components model are considered. Their asymptotic distribution under the null hypothesis and their partial influence functions are derived.

Some key words: Common principal components; Log–likelihood test; Plug–in methods; Proportional scatter matrices; Robust estimation; Wald–type test.

Running Head: Tests for the CPC model

Corresponding Author

Graciela Boente

Moldes 1855, 3^o A

Buenos Aires, C1428CRA, Argentina

email: gboente@mate.dm.uba.ar

PHONE AND FAX 54-11-45763375

*This research was partially supported by Grant X-094 from the Universidad de Buenos Aires, Grant # 5505 from CONICETand Grant 13900-6 of the Fundación Antorchas at Buenos Aires, Argentina and also by the Center for Mathematics and its Applications, Lisbon, Portugal. This research was partially developed while Graciela Boente was visiting the Departamento de Matemática at the Instituto Superior Técnico supported by the project ANCYT–GRICES PO/PA#04-EIII/020.

1 Introduction

Assume that we are dealing with independent observations from k independent samples in $\mathbb{R}^p$ with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i$. It is usual in multivariate analysis to treat the scatter matrices $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$ as unrelated if an overall test of equality tell us that they are not identical. As mentioned in Flury (1988) *"In contrast to the univariate situation, inequality is not just inequality - there are indeed many ways in which covariance matrices can differ"*. To avoid this problem he considered the following general relations among scatter matrices

- **Level 1.** All scatter matrices $\boldsymbol{\Sigma}_i$ are equal.
- **Level 2.** The matrices are proportional to each other, i.e., $\boldsymbol{\Sigma}_i = \rho_i \boldsymbol{\Sigma}_1$, for $2 \leq i \leq k$.
- **Level 3.** The matrices satisfy a CPC model, i.e., $\boldsymbol{\Sigma}_i = \boldsymbol{\beta} \boldsymbol{\Lambda}_i \boldsymbol{\beta}^{\mathrm{T}}$, $1 \leq i \leq k$.
- **Level 4.** $\boldsymbol{\Sigma}_i$ are arbitrary scatter matrices.

The number of parameters for each level is $p(p+1)/2$, $k-1+p(p+1)/2$, $kp+p(p-1)/2$ and $k\,p(p+1)/2$, respectively. Therefore, the difference between the number of parameters in level 1 and 4 is $(k-1)p(p+1)/2$ which is, generally, bad in practice, specially when we are dealing with a large number of populations. This question suggested the decomposition of the log–likelihood ratio statistics for equality of covariance matrices, according to his hierarchy, described in Flury (1988).

It is well known that likelihood ratio test are in most situations, affected by anomalous observations. A robust statistic to test equality against proportionality was studied by Boente and Orellana (2004). On the other hand, an approximate test, based on eigenprojections, for testing the hypothesis that the subspaces spanned by the first q principal components of several different covariances matrices are identical, was derived by Schott (1991) who also considered a plug–in version of his test by using the $M-$estimator proposed by Tyler (1987).

In this paper, we go further and we will deal with the hypothesis involving level 2 versus level 3 and level 3 versus the hierarchically lower model given in level 4. An approach based on estimators of the eigenvalues under a CPC model is considered to test the first hypothesis, while the second one is tested using a robustified version of the log–likelihood statistic. We also define a robust plug–in log–likelihood statistic to test proportionality against CPC.

This paper is organized as follows. In Section 2, we introduce the criterion for testing CPC against arbitrary scatter matrices and we derive its asymptotic behavior under the null hypothesis. In Section 3, we consider the test statistics for proportionality versus CPC while in Section 4 the partial influence functions of all test statistics is derived. Finally, the conclusions of a simulation study are given in Section 5 and a real example is studied in Section 6. Proofs are given in the Appendix.

2 Testing CPC against arbitrary scatter matrices

Let us assume that $(\mathbf{x}_{ij})_{1\leq j\leq n_i, 1\leq i\leq k}$ are independent observations from k independent samples in $\mathbb{R}^p$ with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i$. Let $N = \sum_{i=1}^{k} n_i$, $\tau_i = \dfrac{n_i}{N}$ and $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$.

A basic common structure, described as level 3 in the Introduction, assumes that the k covariance matrices have different eigenvalues but identical eigenvectors, i.e.,

$$\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}} \,, \quad 1 \le i \le k \,, \tag{1}$$

where $\boldsymbol{\Lambda}_i$ are diagonal matrices, $\boldsymbol{\beta}$ is the orthogonal matrix of the common eigenvectors and $\boldsymbol{\Sigma}_i$ is the covariance matrix of the $i-$th population. Model (1) was proposed in Flury (1984) and became known as the *Common Principal Components* (CPC) model. The maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_i$, assuming multivariate normality of the original variables, are derived in Flury (1984).

The hierarchy of test statisitcs discussed in Flury (1988) includes to test level 3 against level 4, i.e,

$$H_{\mathrm{CPC}} : \; \boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}} \;, \text{ for } 1 \le i \le k$$

versus

$$H_1 : \; \boldsymbol{\Sigma}_i \quad \text{are arbitrary positive definite scatter matrices} \;, \quad 1 \le i \le k \;.$$

The classical log–likelihood test statistic to test this hypothesis is given by

$$T_{\mathrm{ML,CPC}} = \sum_{i=1}^{k} n_i \log \left[\frac{\det\left(\widehat{\boldsymbol{\Lambda}}_{i,\mathrm{ML}}\right)}{\det\left(\mathbf{S}_i\right)} \right] \;,$$

where $\widehat{\boldsymbol{\Lambda}}_{i,\mathrm{ML}}$ is the diagonal matrix of the eigenvalues estimated under H_{CPC} and $\mathbf{S}_i$ is the sample covariance matrix.

The idea beyond this statistic is that, under H_{CPC}, it should be expected that $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}^{\mathrm{T}}\mathbf{S}_i\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$ will be aproximately a diagonal matrix where $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$ are the maximum likelihood estimators of the common directions. This idea can be used to robustify the test statistic by plugging–in independent robust affine equivariant scatter estimates, $\mathbf{V}_i$, into the log–likelihood ratio. Hence, the robust statistic can be defined as

$$T_{\mathrm{CPC}} = \sum_{i=1}^{k} n_i \log \left[\frac{\det\left(\mathrm{diag}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}}\right)\right)}{\det\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}}\right)} \right] = \sum_{i=1}^{k} n_i \log \left[\frac{\det\left(\widehat{\boldsymbol{\Lambda}}_i\right)}{\det\left(\mathbf{V}_i\right)} \right] \;,$$

where now $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Lambda}}_i = \mathrm{diag}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}}\right)$ are, respectively, the plug–in estimators of the common directions and of the eigenvalue matrices $\boldsymbol{\Lambda}_i$ related to the scatter estimates $\mathbf{V}_i$.

A standard framework to derive the asymptotic behavior in robust principal component analysis is to assume that the estimators of the scatter matrix are asymptotically normally distributed and spherically invariant. For that reason, and since the samples of the k populations are independent, we will assume, throughout this paper, that for $1 \le i \le k$, the estimators, $\mathbf{V}_i$, of the scatter matrix $\boldsymbol{\Sigma}_i$ are independent and satisfy the following assumption

A1. $\sqrt{n_i}\left(\mathbf{V}_i - \boldsymbol{\Sigma}_i\right) \xrightarrow{\mathcal{D}} \mathbf{Z}_i$ where $\mathbf{Z}_i$ has a multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Xi}_i$ such that

$$\boldsymbol{\Xi}_i = \sigma_1\left(\mathbf{I} + K_{pp}\right)\left(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i\right) + \sigma_2 \;\mathrm{vec}\left(\boldsymbol{\Sigma}_i\right)\; \mathrm{vec}\left(\boldsymbol{\Sigma}_i\right)^{\mathrm{T}} \;, \tag{2}$$

with K_{pp} the $p^2 \times p^2$ block matrix with the $(l,m)-$block equal to a $p \times p$ matrix with a 1 at entry (l,m) and 0 everywhere else.

where, from now on, when dealing with random matrices $\mathbf{Z}_n \xrightarrow{\mathcal{D}} \mathbf{Z}$ means that $\text{vec}(\mathbf{Z}_n) \xrightarrow{\mathcal{D}} \text{vec}(\mathbf{Z})$.

Remark 1. It is well known that, for elliptically distributed observations, MCD, M, S and $\tau-$estimators are asymptotically normally distributed and spherically invariant. If the k populations have ellipsoidal distributions that only differ on their scatter matrix and if the same robust scatter estimate is considered for each population, these estimators will satisfy **A1** (see, Tyler, 1982). Explicit forms for the constants σ_1 and σ_2 are given in Tyler (1982), for M–estimators, and in Lopuhaä (1991), for S and $\tau-$estimators.

Implicitely, we are thus assuming that all the scatter estimates are related to the same functional $\mathbf{V}$ and that all the populations have the same elliptical distribution except for changes in the scatter matrices so that **A1** will hold.

Theorem 1. *Let* $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}$, $1 \le i \le k$, *be independent observations from* k *independent samples with location parameter* $\boldsymbol{\mu}_i$ *and scatter matrix* $\boldsymbol{\Sigma}_i$. *Assume that* H_{CPC} *holds, i.e.,* $\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\text{T}} = \boldsymbol{\beta} diag(\lambda_{i1}, \ldots, \lambda_{ip})\boldsymbol{\beta}^{\text{T}}$ *and that* $\lambda_{11} > \ldots > \lambda_{1p}$, $n_i = \tau_i N$ *with* $0 < \tau_i < 1$ *fixed numbers such that* $\sum_{i=1}^{k} \tau_i = 1$.

Let $\mathbf{V}_i$ *be robust affine equivariant estimates of the scatter matrices* $\boldsymbol{\Sigma}_i$, *satisfying* ***A1***. *Let us consider the plug–in estimates of the common axes and their size given by*

$$\widehat{\boldsymbol{\Lambda}}_i = diag\left(\widehat{\boldsymbol{\beta}}^{\text{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}}\right) \tag{3}$$

$$\widehat{\boldsymbol{\beta}}_m^{\text{T}}\left[\sum_{i=1}^{k} n_i \frac{\widehat{\lambda}_{im} - \widehat{\lambda}_{ij}}{\widehat{\lambda}_{im}\widehat{\lambda}_{ij}}\mathbf{V}_i\right]\widehat{\boldsymbol{\beta}}_j = 0 \qquad \text{for } m \neq j \tag{4}$$

$$\widehat{\boldsymbol{\beta}}_m^{\text{T}}\widehat{\boldsymbol{\beta}}_j = \delta_{mj} \, . \tag{5}$$

Then, we have that $T_{\text{CPC}} \xrightarrow{\mathcal{D}} \sigma_1 \chi^2_{\frac{(k-1)p(p-1)}{2}}$.

It is woth noticing that, under normality, the asymptotic distribution of the test statistic is that of the classical log–likelihood test, except for a multiplicative constant which is related to the efficiencies of the off–diagonal elements of the scatter estimates considered.

3 Testing proportionality against CPC

The proportionality model is more restrictive that the CPC model and assumes that the scatter matrices are equal up to a proportionality constant, i.e.,

$$\boldsymbol{\Sigma}_i = \rho_i \boldsymbol{\Sigma}_1 \, , \text{ for } 2 \le i \le k \, . \tag{6}$$

In this section, we will provide two statistics for testing level 2 versus level 3, i.e.,

$$H_{\text{PROP}} : \boldsymbol{\Sigma}_i = \rho_i \boldsymbol{\Sigma}_1 \, , \text{ for } 2 \le i \le k \quad \text{versus} \quad H_{\text{CPC}} : \boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\text{T}} \, , \quad 1 \le i \le k \, .$$

Let $c_{ij} = \dfrac{\lambda_{ij}}{\lambda_{i1}}$, and $\mathbf{c}_i = (c_{i2}, \ldots, c_{ip})^{\text{T}}$, then under H_{PROP} $c_{ij} = c_{1j}$, for all i and j. On the other hand, if H_{CPC} holds, $c_{ij} = c_{1j}$ for all i and j entail that H_{PROP} is true. The first test statistic will be based on robust estimators of the eigenvalues of the matrices $\boldsymbol{\Sigma}_i$ assuming that the CPC model holds.

Let $\widehat{\mathbf{\Lambda}}_i$ be the plug–in estimators of the eigenvalues matrix $\mathbf{\Lambda}_i$ related to robust scatter estimators $\mathbf{V}_i$, defined in (3).

In this case, it is well known (see Boente and Orellana, 2001 and Boente, Pires and Rodrigues, 2002) that the estimators of the eigenvalues are asymptotically normally distributed and their asymptotic variances and covariances denoted ASVAR and ASCOV, respectively, are

$$\begin{aligned} \text{ASCOV}\left(\widehat{\lambda}_{\ell j}, \widehat{\lambda}_{im}\right) &= 0 \text{ for } \ell \neq i \\ \text{ASVAR}\left(\widehat{\lambda}_{ij}\right) &= \frac{1}{\tau_i}(2\sigma_1 + \sigma_2)\lambda_{ij}^2 , \quad 1 \le j \le p \\ \text{ASCOV}\left(\widehat{\lambda}_{ij}, \widehat{\lambda}_{im}\right) &= \frac{1}{\tau_i}\sigma_2 \lambda_{ij}\lambda_{im} \quad \text{ for } m \neq j \ . \end{aligned} \tag{7}$$

A natural estimate of c_{ij} is $\widehat{c}_{ij} = \dfrac{\widehat{\lambda}_{ij}}{\widehat{\lambda}_{i1}}$. The statistic to test H_{PROP} versus H_{CPC} based on $\widehat{c}_{ij}$ is a Wald–type statistic and is defined by

$$\mathcal{W}_{\text{PROP}} = N \operatorname{vec}\left(\widehat{\mathbf{D}}\right)^{\mathrm{T}} \widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{D}}}^{-1} \operatorname{vec}\left(\widehat{\mathbf{D}}\right)$$

where

$$\begin{aligned} \widehat{\mathbf{D}} &= (\widehat{\mathbf{c}}_2 - \widehat{\mathbf{c}}_1, \dots, \widehat{\mathbf{c}}_k - \widehat{\mathbf{c}}_1) \\ \widehat{\mathbf{\Sigma}}_{\widehat{\mathbf{D}}} &= \widehat{\mathbf{B}}_1 \otimes \left\{\frac{1}{\tau_1}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right\} + \operatorname{diag}\left(\frac{1}{\tau_2}\widehat{\mathbf{B}}_2, \dots, \frac{1}{\tau_k}\widehat{\mathbf{B}}_k\right) \\ \widehat{\mathbf{B}}_i &= 2\sigma_1\left(\widehat{\mathbf{c}}_i\widehat{\mathbf{c}}_i^{\mathrm{T}} + \operatorname{diag}\left(\widehat{c}_{i2}^2, \dots, \widehat{c}_{ip}^2\right)\right) \ . \end{aligned}$$

The following Theorem gives its asymptotic distribution under the null hypothesis.

Theorem 2. *Let $\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$, $1 \le i \le k$, be independent observations from k independent samples with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\mathbf{\Sigma}_i = \boldsymbol{\beta}\mathbf{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}} = \boldsymbol{\beta}\text{diag}(\lambda_{i1}, \dots, \lambda_{ip})\boldsymbol{\beta}^{\mathrm{T}}$. Assume that H_{PROP} holds, that the eigenvalues of the first population satify $\lambda_1 > \ldots > \lambda_p$ and that $n_i = \tau_i N$ with $0 < \tau_i < 1$ fixed numbers such that $\sum_{i=1}^{k} \tau_i = 1$.*

*Let $\mathbf{V}_i$ be robust affine equivariant estimates of the scatter matrices $\mathbf{\Sigma}_i$, satisfying **A1**. Let us consider the plug–in estimates of the common axes and their size defined through (3) to (5). Then, we have that $\mathcal{W}_{\text{PROP}} \xrightarrow{\mathcal{D}} \chi^2_{(k-1)(p-1)}$.*

Another possibility to test H_{PROP} against H_{CPC} is to use a robust plug–in version of the log–likelihood statistic, i.e.,

$$\sum_{i=1}^{k} n_i \log\left[\frac{\det\left(\widehat{\rho}_i \text{diag}\left(\widehat{\lambda}_1, \dots, \widehat{\lambda}_p\right)\right)}{\det\left(\text{diag}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}}\right)\right)}\right]$$

where $\mathbf{V}_i$ are independent robust affine equivariant scatter estimates, $\widehat{\boldsymbol{\beta}}$ are the plug–in estimators of the common directions related to the scatter estimates $\mathbf{V}_i$, $\widehat{\rho}_i$ and $\widehat{\lambda}_j$ are the plug–in estimators of the proportionality constants and of the eigenvalues of the first population related to $\mathbf{V}_i$, as defined in Boente and Orellana (2004) and also studied in Boente, Critchley and Orellana (2004). However, the asymptotic distribution of this statistic cannot be easily derived. Instead, the eigenvalues of the first population and the proportionality constants should be estimated related to the prior estimation of the common direction. Thus, we will consider the statistic defined through

$$T_{\text{PROP}} = \sum_{i=1}^{k} n_i \log \left[\frac{\det\left(\widehat{\rho}_i \text{diag}\left(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p\right)\right)}{\det\left(\text{diag}\left(\widehat{\lambda}_{i1}, \ldots, \widehat{\lambda}_{ip}\right)\right)} \right] ,$$

where $\widehat{\lambda}_{ij} = \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}} \mathbf{V}_i \widehat{\boldsymbol{\beta}}_j$ are defined in (3) and $\widehat{\rho}_i$ and $\widehat{\lambda}_j$ solve the following equations

$$\widehat{\rho}_i = \frac{1}{p} \sum_{j=1}^{p} \frac{\widehat{\lambda}_{ij}}{\widehat{\lambda}_j} \qquad 2 \le i \le k \tag{8}$$

$$\widehat{\lambda}_j = \frac{1}{N} \sum_{i=1}^{k} \frac{n_i}{\widehat{\rho}_i} \widehat{\lambda}_{ij} \qquad 1 \le j \le p \, . \tag{9}$$

Theorem 3. *Let $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}$, $1 \le i \le k$, be independent observations from k independent samples with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i = \boldsymbol{\beta} \boldsymbol{\Lambda}_i \boldsymbol{\beta}^{\mathrm{T}} = \boldsymbol{\beta} \text{diag}(\lambda_{i1}, \ldots, \lambda_{ip}) \boldsymbol{\beta}^{\mathrm{T}}$. Assume that H_{PROP} holds, that the eigenvalues of the first population satify $\lambda_1 > \ldots > \lambda_p$ and that $n_i = \tau_i N$ with $0 < \tau_i < 1$ fixed numbers such that $\sum_{i=1}^{k} \tau_i = 1$. Let $\mathbf{V}_i$ be robust affine equivariant estimates of the scatter matrices $\boldsymbol{\Sigma}_i$, satisfying **A1**. Let us consider the plug–in estimates of the common axes and their size defined by (3) to (5) and the related estimates of the proportionality constants and their size solution of (8) and (9). Then, we have that $T_{\text{PROP}} \xrightarrow{\mathcal{D}} \sigma_1 \chi^2_{(k-1)(p-1)}$.*

4 Partial Influence Functions

Denote by F the product measure, $F = F_1 \times \ldots \times F_k$. Partial influence functions of a functional $T(F)$, introduced in Pires and Branco (2002), are then defined as

$$\text{PIF}_{i_0}(\mathbf{x}, T, F) = \lim_{\epsilon \to 0} \frac{T(F_{\epsilon, \mathbf{x}, i_0}) - T(F)}{\epsilon} ,$$

where $F_{\epsilon, \mathbf{x}, i_0} = F_1 \times \ldots \times F_{i_0 - 1} \times F_{i_0, \epsilon, \mathbf{x}} \times F_{i_0+1} \times \ldots \times F_k$ and $F_{i, \epsilon, \mathbf{x}} = (1-\epsilon) F_i + \epsilon \delta_{\mathbf{x}}$.

In this Section, we will derive the partial influence functions of the functionals related to the test statistics defined in the previous Sections.

Let $\mathbf{V}_i(G)$ be a Fisher–consistent scatter functional such that $\mathbf{V}_i(F_i) = \boldsymbol{\Sigma}_i$. Denote by $\lambda_j\left(\mathbf{V}_\ell(F)\right)$ the eigenvalues of $\mathbf{V}_\ell(F)$ and by $\boldsymbol{\beta}_{\mathbf{V}}(F)$ and $\lambda_{\mathbf{V}, \ell j}$ the plug–in functionals related to the scatter functionals $\mathbf{V}(F) = (\mathbf{V}_1(F_1), \ldots, \mathbf{V}_k(F_k))$, i.e., the solution of

$$\text{diag}\left\{ \boldsymbol{\beta}_{\mathbf{V}}(F)^{\mathrm{T}} \mathbf{V}_i(F_i) \boldsymbol{\beta}_{\mathbf{V}}(F) \right\} = \boldsymbol{\Lambda}_{\mathbf{V}, i}(F) \tag{10}$$

$$\boldsymbol{\beta}_{\mathbf{V}, m}(F)^{\mathrm{T}} \left\{ \sum_{i=1}^{k} \tau_i \frac{\lambda_{\mathbf{V}, im}(F) - \lambda_{\mathbf{V}, ij}(F)}{\lambda_{\mathbf{V}, im}(F) \lambda_{\mathbf{V}, ij}(F)} \mathbf{V}_i(F_i) \right\} \boldsymbol{\beta}_{\mathbf{V}, j}(F) = 0 \qquad \text{for } m \ne j \tag{11}$$

$$\boldsymbol{\beta}_{\mathbf{V}, m}(F)^{\mathrm{T}} \boldsymbol{\beta}_{\mathbf{V}, j}(F) = \delta_{mj} \, . \tag{12}$$

The test functional related to the statistic used to test H_{CPC} versus H_1 is given by

$$T_{\mathbf{V},\text{CPC}}(F) = \sum_{\ell=1}^{k} \tau_\ell \sum_{j=1}^{p} \log\left(\lambda_{\mathbf{V},\ell j}(F)\right) - \log\left(\lambda_j\left(\mathbf{V}_\ell(F_\ell)\right)\right) \ ,$$

while those related to T_{PROP} and $\mathcal{W}_{\text{PROP}}$ are respectively defined through

$$\begin{aligned} T_{\mathbf{V},\text{PROP}}(F) &= \sum_{\ell=1}^{k} \tau_\ell \sum_{j=1}^{p} \log\left(\rho_{\mathbf{V},\ell}(F)\lambda_{\mathbf{V},j}(F)\right) - \log\left(\lambda_{\mathbf{V},\ell j}(F)\right) \ , \\ \mathcal{W}_{\mathbf{V},\text{PROP}}(F) &= \text{vec}\left(\mathbf{D}(F)\right)^{\text{T}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1}(F)\text{vec}\left(\mathbf{D}(F)\right) \end{aligned}$$

where $\lambda_{\mathbf{V},ij}$ is defined through (10) to (12) and

$$\begin{aligned} \rho_{\mathbf{V},i}(F) &= \frac{1}{p}\sum_{j=1}^{p} \frac{\lambda_{\mathbf{V},ij}}{\lambda_{\mathbf{V},j}(F)} \quad i = 2,\ldots,k \\ \lambda_{\mathbf{V},j}(F) &= \sum_{i=1}^{k} \tau_i \frac{\lambda_{\mathbf{V},ij}(F)}{\rho_{\mathbf{V},i}(F)} \qquad 1 \le j \le p \\ \mathbf{D}(F) &= \left(\mathbf{c}_2(F) - \mathbf{c}_1(F), \ldots, \mathbf{c}_k(F) - \mathbf{c}_1(F)\right) = \left(\mathbf{D}_2(F), \ldots, \mathbf{D}_k(F)\right) \\ \boldsymbol{\Sigma}_{\mathbf{D}}(F) &= \mathbf{B}_1(F) \otimes \left\{\frac{1}{\tau_1}\mathbf{1}\mathbf{1}^{\text{T}}\right\} + \text{diag}\left(\frac{1}{\tau_2}\mathbf{B}_2(F), \ldots, \frac{1}{\tau_k}\mathbf{B}_k(F)\right) \\ \mathbf{B}_i(F) &= 2\sigma_1\left(\mathbf{c}_i(F)\mathbf{c}_i(F)^{\text{T}} + \text{diag}\left(c_{i2}^2(F), \ldots, c_{ip}^2(F)\right)\right) \\ c_{ij}(F) &= \frac{\lambda_{\mathbf{V},ij}(F)}{\lambda_{\mathbf{V},i1}(F)} \\ \mathbf{c}_i(F) &= \left(c_{i2}(F), \ldots, c_{ip}(F)\right)^{\text{T}} \ . \end{aligned}$$

It is easy to see that, since $\mathbf{V}_i(F_i) = \boldsymbol{\Sigma}_i$, under H_{CPC}, $\text{PIF}_i(\mathbf{x}, T_{\mathbf{V},\text{CPC}}, F) = 0$ while under H_{PROP}, $\text{PIF}_i(\mathbf{x}, T_{\mathbf{V},\text{PROP}}, F) = 0$ and $\text{PIF}_i(\mathbf{x}, \mathcal{W}_{\mathbf{V},\text{PROP}}, F)$ are also equal to 0. So, as in Hampel et al. (1986), we consider as test statistics $S_{\mathbf{V},\text{CPC}}(F) = T_{\mathbf{V},\text{CPC}}(F)^{\frac{1}{2}}$, $S_{\mathbf{V},\text{PROP}}(F) = T_{\mathbf{V},\text{PROP}}(F)^{\frac{1}{2}}$ and $R_{\mathbf{V},\text{PROP}}(F) = \mathcal{W}_{\mathbf{V},\text{PROP}}(F)^{\frac{1}{2}}$. As for the linear model, using that $T_{\mathbf{V},\text{CPC}}(F) = 0$, it is easy to see that

$$\text{PIF}_i(\mathbf{x}, S_{\mathbf{V},\text{CPC}}, F) = \left(\frac{1}{2}\frac{\partial^2}{\partial\epsilon^2} T_{\mathbf{V},\text{CPC}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0}\right)^{\frac{1}{2}}$$

and similarly for $S_{\mathbf{V},\text{PROP}}(F)$ and $R_{\mathbf{V},\text{PROP}}(F)$.

The following Theorem gives the values of the partial influence functions of the test statistic $S_{\mathbf{V},\text{CPC}}(F)$.

Theorem 4. *Let $\mathbf{V}_i(F)$ be a scatter functional such that $\mathbf{V}_i(F_i) = \boldsymbol{\Sigma}_i$. Denote by $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$, $\lambda_{i1}, \ldots, \lambda_{ip}$ the common eigenvectors and the eigenvalues of $\boldsymbol{\Sigma}_i$, i.e., assume that $H_{\text{CPC}}: \ \boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\text{T}}$, for $1 \le i \le k$ holds. Assume that the influence function $IF(\mathbf{x}, \mathbf{V}_i, F_i)$ and $\frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}$ exist and that $\lambda_{11} > \ldots > \lambda_{1p}$. Then, the partial influence functions of $S_{\mathbf{V},\text{CPC}}$ are given by*

$$PIF_i(\mathbf{x}, S_{\mathbf{V},\text{CPC}}, F)^2 \ = \ \tau_i^2 \sum_{j=1}^{p}\sum_{m\neq j} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\text{T}} IF(\mathbf{x}, \mathbf{V}_i, F_i)\, \boldsymbol{\beta}_m\right]^2 \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij}^2\lambda_{im}^2} \sum_{\ell=1}^{k} \tau_\ell \frac{\lambda_{\ell m} - \lambda_{\ell j}}{\lambda_{\ell j}} +$$

$$+\tau_i \sum_{j=1}^{p} \sum_{m\neq j} \left[\boldsymbol{\beta}_j^{\mathrm{T}} IF\left(\mathbf{x}, \mathbf{V}_i, F_i\right) \boldsymbol{\beta}_m\right]^2 \frac{1}{\lambda_{ij}^2 \lambda_{im} \left(\lambda_{ij} - \lambda_{im}\right)} \left[2\tau_i A_{mj} \left(\lambda_{im} - \lambda_{ij}\right)^2 - \lambda_{ij}\lambda_{im}\right] \quad (13)$$

$$= \tau_i \sum_{j=1}^{p} \sum_{m>j} \frac{\left[\boldsymbol{\beta}_j^{\mathrm{T}} IF\left(\mathbf{x}, \mathbf{V}_i, F_i\right) \boldsymbol{\beta}_m\right]^2}{\lambda_{im}\lambda_{ij}} \left[1 - A_{mj}\, \tau_i \frac{(\lambda_{ij} - \lambda_{im})^2}{\lambda_{ij}\lambda_{im}}\right], \quad (14)$$

where $A_{mj} = \left\{ \sum_{\ell=1}^{k} \tau_\ell \frac{(\lambda_{\ell m} - \lambda_{\ell j})^2}{\lambda_{\ell m}\lambda_{\ell j}} \right\}^{-1}$.

Remark 2. In the particular case of a proportional model, i.e., if $\boldsymbol{\Lambda}_i = \rho_i \boldsymbol{\Lambda}_1$ and we denote $\lambda_j = \lambda_{1j}$ we have that

$$\mathrm{PIF}_i(\mathbf{x}, S_{\mathbf{V},\mathrm{CPC}}, F)^2 = \frac{\tau_i(1-\tau_i)}{\rho_i^2} \sum_{j=1}^{p} \sum_{m>j} \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x}, \mathbf{V}_i, F_i\right) \boldsymbol{\beta}_m\right]^2 \frac{1}{\lambda_j \lambda_m} .$$

Note that the partial influence functions of the test statistic are unbounded if the scatter matrices have unbounded influence functions. The behaviour of the partial influence function of the test statistic will be analogous to that described for the principal components by Croux & Haesbroeck (2000). For instance, in dimension 2, the largest values of the partial influence function are obtained along the bisectors.

The expression (14) allows to see easily that the right hand term is positive.

The following theorems state the partial influence functions of the functionals related to the test statistics used to test H_{PROP} against H_{CPC}.

Theorem 5. *Let* $\mathbf{V}_i(F)$ *be a scatter functional such that* $\mathbf{V}_i(F_i) = \boldsymbol{\Sigma}_i$. *Denote by* $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$, $\lambda_{i1}, \ldots, \lambda_{ip}$ *the common eigenvectors and the eigenvalues of* $\boldsymbol{\Sigma}_i$, *i.e., assume that* $H_{\mathrm{PROP}} : \boldsymbol{\Sigma}_i = \rho_i \boldsymbol{\beta} \boldsymbol{\Lambda}_1 \boldsymbol{\beta}^{\mathrm{T}}$, *for* $1 \leq i \leq k$ *holds with* $\rho_1 = 1$. *Assume that the influence function* $IF(\mathbf{x}, \mathbf{V}_i, F_i)$ *and* $\frac{\partial^2}{\partial \epsilon^2} \mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}$ *exist and that* $\lambda_1 > \ldots > \lambda_p$. *Then, the partial influence functions of* $R_{\mathbf{V},\mathrm{PROP}}$ *are given by*

$$PIF_i(\mathbf{x}, R_{\mathbf{V},\mathrm{PROP}}, F)^2 = vec\left[PIF_i(\mathbf{x}, \mathbf{D}, F)\right]^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{D}}^{-1} vec\left[PIF_i(\mathbf{x}, \mathbf{D}, F)\right] \quad (15)$$

where

$$PIF_i(\mathbf{x}, \mathbf{D}_i, F) = \frac{1}{\lambda_{i1}} \left[\mathbf{J}_i - \boldsymbol{\beta}_1^{\mathrm{T}} IF(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_1\, \mathbf{c}_i\right] \qquad i \neq 1 \quad (16)$$

$$PIF_1(\mathbf{x}, \mathbf{D}_i, F) = -\frac{1}{\lambda_{11}} \left[\mathbf{J}_1 - \boldsymbol{\beta}_1^{\mathrm{T}} IF(\mathbf{x}, \mathbf{V}_1, F_1) \boldsymbol{\beta}_1\, \mathbf{c}_1\right] \quad (17)$$

$$\mathbf{J}_i = \left(\boldsymbol{\beta}_2^{\mathrm{T}} IF(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p^{\mathrm{T}} IF(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_p\right)^{\mathrm{T}} .$$

The following Theorem, whose proof follows easily using the same arguments as above, gives the values of the partial influence functions of $S_{\mathbf{V},\mathrm{PROP}}(F)$.

Theorem 6. *Let* $\mathbf{V}_i(F)$ *be a scatter functional such that* $\mathbf{V}_i(F_i) = \boldsymbol{\Sigma}_i$. *Denote by* $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$, $\lambda_{i1}, \ldots, \lambda_{ip}$ *the common eigenvectors and the eigenvalues of* $\boldsymbol{\Sigma}_i$, *i.e., assume that* $H_{\mathrm{PROP}} : \boldsymbol{\Sigma}_i = \rho_i \boldsymbol{\beta} \boldsymbol{\Lambda}_1 \boldsymbol{\beta}^{\mathrm{T}}$, *for* $1 \leq$

$i \leq k$ holds with $\rho_1 = 1$. Assume that the influence function $IF(\mathbf{x}, \mathbf{V}_i, F_i)$ and $\frac{\partial^2}{\partial \epsilon^2} \mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}$ exist and that $\lambda_1 > \ldots > \lambda_p$. Moreover, assume that the influence function $IF(\mathbf{x}, \mathbf{V}_i, F_i)$ and $\frac{\partial^2}{\partial \epsilon^2} \mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}$ exist and that $\lambda_1 > \ldots > \lambda_p$. Then, under H_{PROP}, the partial influence functions of $S_{\mathbf{V},\mathrm{PROP}}$ are given by

$$PIF_i(\mathbf{x}, S_{\mathbf{V},\mathrm{PROP}}, F)^2 = \frac{\tau_i (1 - \tau_i)}{2\rho_i^2} \left\{ \zeta_i - \frac{1}{p} \gamma_i^2 \right\} , \tag{18}$$

where

$$\zeta_i = \sum_{j=1}^{p} \frac{\left[\boldsymbol{\beta}_j^{\mathrm{T}} IF(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j \right]^2}{\lambda_j^2} \qquad \gamma_i = \sum_{j=1}^{p} \frac{\boldsymbol{\beta}_j^{\mathrm{T}} IF(\mathbf{x}, \mathbf{V}_i, F_i) \boldsymbol{\beta}_j}{\lambda_j} .$$

Note that Cauchy–Schwartz inequality imply that $\zeta_i p \geq \gamma_i^2$ and thus, $\mathrm{PIF}_i(\mathbf{x}, S_{\mathbf{V},\mathrm{PROP}}, F)$ is well defined.

Remark 3. It is worth noticing that

$$\begin{aligned}
\sum_{i=1}^{k} \frac{1}{\tau_i} E\left(\mathrm{PIF}_i(\mathbf{x}, S_{\mathbf{V},\mathrm{CPC}}, F)^2 \right) &= \sigma_1 \frac{p(p-1)(k-1)}{2} \\
\sum_{i=1}^{k} \frac{1}{\tau_i} E\left(\mathrm{PIF}_i(\mathbf{x}, S_{\mathbf{V},\mathrm{PROP}}, F)^2 \right) &= \sigma_1 (p-1)(k-1)
\end{aligned}$$

as expected.

Figures 1 to 3 give the plots of the partial influence function PIF_1 of the three functionals when $p = 2$ at $F = F_1 \times F_2$ with $F_1 = N(\mathbf{0}, \mathrm{diag}(2, 1))$ and $F_2 = N(\mathbf{0}, 4\ \mathrm{diag}(2, 1))$, respectively. We have considered as scatter matrices estimators the sample covariance matrix and the Donoho–Stahel estimator using as weight function the Huber's function with constant $\sqrt{\chi_p^2(0.95)} = 2.4477$. The univariate location and scale functionals are the median and the MAD, median of the absolute deviations with respect to the median. An expression for the influence function of the Donoho–Stahel scatter functional can be found in Gervini (2002).

In all cases, the shape of the partial influence functions of the robust estimates is comparable to that of their classical relatives at the center of the distribution, while the influence at points further away is downweighted for the robust estimates while it is much more larger for the classical ones. However, it should be noticed that the robust functionals have a discontinuity at 0, due to the discontinuity of the influence function of the Donoho–Stahel scatter functional. On the other hand, as mentioned in Remark 2, for the $\mathrm{PIF}_i(\mathbf{x}, S_{\mathbf{V},\mathrm{CPC}}, F)$ the largest values of the partial influence function are obtained along the bisectors while the largest values of $\mathrm{PIF}_i(\mathbf{x}, R_{\mathbf{V},\mathrm{PROP}}, F)$ and $\mathrm{PIF}_i(\mathbf{x}, S_{\mathbf{V},\mathrm{PROP}}, F)$ are attained along the axis, except for $(0, 0)$.

5 Monte Carlo study

5.1 Testing H_{CPC} against arbitrary scatter matrices

We performed a simulation study in dimension $p = 4$ to compare $T_{\mathrm{ML,CPC}}$, i.e., the log–likelihood test statistic and T_{CPC} with the Donoho(1982)–Stahel(1981) estimator, indicated with thick and dashed lines in the plots respectively. The Donoho–Stahel scatter estimator (DS estimator) was computed using the Huber weights with tunning constant $\sqrt{\chi^2_p(0.95)} = 3.0803$ and as univariate location and scale estimators, the sample median and the MAD, median of the absolute deviations with respect to the median. Note that for the classical test $\sigma_1 = 1$ when the observations are normally distributed. Then, to provide fair comparisons, the value of σ_1 was numerically computed for the Donoho–Stahel estimator in $p = 4$, under normality, and it result equal to 1.0246.

We have considered two populations with $\mathbf{\Sigma}_1 = \text{diag}(16, 8, 2, 1)$,

$$\mathbf{\Sigma}_2 = 4\ \mathbf{\Sigma}_1 + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A \\ 0 & 0 & 0 & 0 \\ 0 & A & 0 & 0 \end{pmatrix}$$

and $A = 0, 3, 6$ and 9.

In all models, we performed 500 replications generating two independent samples of size $n_i = n = 100$.

The results for normal data sets will be indicated by C_0, while C_1, C_2 and C_3 will denote the following contaminations.

- C_1: $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in}$ are i.i.d. $0.9N(\mathbf{0}, \mathbf{\Sigma}_i) + 0.1N(\mathbf{0}, 9\mathbf{\Sigma}_i)$.
- C_2: $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n}$ are i.i.d. $0.9N(\mathbf{0}, \mathbf{\Sigma}_1) + 0.1N(\mathbf{0}, 9\mathbf{\Sigma}_1)$ and $\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n}$ are i.i.d. $0.9N(\mathbf{0}, \mathbf{\Sigma}_2) + 0.1N(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$ with $\boldsymbol{\mu}_2 = \mathbf{e}_4 - 100\mathbf{e}_2$. The aim of this contamination is to see how the bias of parameter estimates affects the level of the test.
- C_3: $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n}$ are i.i.d. $0.9N(\mathbf{0}, \mathbf{\Sigma}_1) + 0.1N(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ with $\boldsymbol{\mu}_1 = 300\,\mathbf{\Sigma}_1^{1/2}\mathbf{e}_4$ and $\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n}$ are i.i.d. $0.9N(\mathbf{0}, \mathbf{\Sigma}_2) + 0.1N(\boldsymbol{\mu}_2, 4\mathbf{\Sigma}_1)$ with $\boldsymbol{\mu}_2 = 2\boldsymbol{\mu}_1$. The aim of this contamination is to breakdown the power of the test for large values of A.

To summarize the results, we evaluated the power of the test, with fixed size $\alpha = 5\%$, by computing the percentage of rejections, over the replications. Fig. 4 displays these results as a function of A.

The performance of T_{CPC} with the DS estimator is similar to that of $T_{\text{ML,CPC}}$ under normality, but it is better under contamination. Both the power and the size of the classical test can be seriously affected.

5.2 Testing H_{PROP} against H_{CPC}

We performed a simulation study in dimension 4 to compare the behavior of the Wald type statistics computed with the sample covariance matrices, $\mathcal{W}_{\text{ML,CPC}}$, and with the Donoho–Stahel estimator, $\mathcal{W}_{\text{CPC}}$, plotted with thick and dashed lines in the figures, respectively. We have considered two populations with $\mathbf{\Sigma}_1 = \text{diag}(16, 8, 2, 1)$,

$$\mathbf{\Sigma}_2 = 4\ \mathbf{\Sigma}_1 + \begin{pmatrix} 16\,A & 0 & 0 & 0 \\ 0 & 8\,A & 0 & 0 \\ 0 & 0 & 2\,A & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and $A = 0, 1.5, 3, 4.5, 6, 9$ and 15.

In all models, we performed 500 replications generating two independent samples of size $n_i = n = 100$. As in the previous Section, we show the results for normal data sets (C_0), for C_1 and C_3 and also for the following contamination

- C_4: $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n}$ are i.i.d.$N(\mathbf{0}, \mathbf{\Sigma}_1)$ and $\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n}$ are i.i.d. $0.9N(\mathbf{0}, \mathbf{\Sigma}_2) + 0.1N(\mathbf{0}, \mathbf{\Sigma}_3)$ with $\mathbf{\Sigma}_3 = \text{diag}(160, 80, 20, 145)^{\mathrm{T}}$.

To summarize the results we evaluated the power of the test, with fixed size $\alpha = 5\%$, by computing again the percentage of rejections, over the replications, for both statistics. Fig. 5 displays these results as a function of A.

The performance of the $\mathcal{W}_{\text{PROP}}$ with the DS estimator is similar to the one of $\mathcal{W}_{\text{ML,PROP}}$ under normality, but it is better under contamination. The same behavior is observed with the plug–in test statistic T_{PROP} and $T_{\text{ML,PROP}}$ which are plotted in Fig. 6. However, it should be noticed that the rate of convergence, under the null hypothesis, of the log–likelihood test seems to be slower since the frequency of rejection is 0.0820 againts 0.030 for the Wald test statistic. For the robust ones, the frequencies are 0.1020 and 0.0380, respectively. With both type of statistics, the size of the classical test can be seriously affected under the contaminations considered. Note that the power of the classical test is highly sensitive to the contamination C_4, while that computed with the Donoho–Stahel estimator shows some sensitivity for lower alternatives but recovers its performance as A increases. Under this contamination the robust plug–in statistic performs much better than the Wald–type one. Contamination C_4 also breakdown the level for all methods. It is worth noticing that contamination C_4 is a difficult one to be detected by the robust procedure since it corresponds to mild outliers showing that more research should be done in this direction to avoid this effect.

6 Example

The following example shows that, similar problems to those described when estimating the principal components, arise when testing a CPC model by introducing a few atypical data.

We have considered the data on the petal and sepal width of two species of Iris, *Iris versicolor* and *Iris virginica,* given in Fisher (1936) and studied in Boente and Orellana (2001). The log–likelihood test statistic assuming normality for H_{CPC} against arbitrary scatter matrices has a value of 1.7463 that gives a $p-$value of 18.63%, not rejecting the null hypothesis. In order to show the effect of a small number of outliers we modified four data points in *Iris versicolor* and three data points in *Iris virginica* as shown in Figure 1 in Boente and Orellana (2001). The value of the test statistic is now 4.7880 and so the null hypothesis is now rejected at level 5%. Moreover, the $p-$value is 2.86%.

In order to solve the problems observed with the maximum likelihood procedure, we have considered the test statistic described in Section 2 computed with the Donoho–Stahel estimator. As in the Monte Carlo study, the value of σ_1 was set equal to 1.0241 for the robust scale estimator while $\sigma_1 = 1$ for the sample covariance matrix. The test statistic for the original data is now 2.6485 while for the modified one is 2.2012 giving as $p-$values 10.78% and 14.26%, respectively. This shows that, when using the robust procedure, the introduced outliers do not change the decision.

When testing H_{PROP} against H_{CPC}, we get a different situation: the classical and the robust procedures lead to different conclusions. The $p-$values for the Wald–type test based on the sample covariance matrices are 11.98% and 6.90% for the original and modified data sets, while for that based on the

Donoho–Stahel statistics we get 3.52% and 2.19%, respectively. The results given in Table 1 in Boente and Orellana (2001), make us suspect from the non–validity of H_{PROP}. Thus, we may suspect that the result obtained for the classical test could be distorted by some masked outlier or by a small number of data points. We thus performed the log–likelihood tests obtaining as $p-$value 3.36% and 1.19% for the classical and robust procedure, respectively, for the original data set. Note that this results are consistent with the result obtained for the robust Wald proposal.

In summary, using the robust tests we would conclude that level 3 is adequate (with or without the outliers), while when using the classical tests the conclusions would be different (indicating level 4 when the outliers are present) and depending on the the test statistic used to decide between levels 2 and 3.

7 Final Comments

If in multivariate analysis involving k independent populations the assumption of equality of scatter matrices is not adequate, problems may arise because of an excessive number of parameters if we estimate the scatter matrices separately for each population. Such problems can often be avoided if the different covariance matrices exhibit some common structure. The common principal components model, introduced by Flury (1984), generalizes equality by allowing the matrices to have different eigenvalues but identical eigenvectors. This explains why it is often of interest to decide on the equality of their scatter matrices or on how they can differ. Moreover, since multivariate outliers are very difficult to detect, it is important to use robust procedures to take these decisions.

In this paper, we have considered the four level hierarchy of relations between scatter matrices introduced by Flury (1988) and we have proposed several robust tests to decide between level 2 (proportional model), level 3 (CPC model) and level 4 (arbitrary matrices) of that system. We have derived the asymptotic distribution of each test statistic under the corresponding null hypothesis. As in the classical case, this distribution is, in every case, a chi-square distribution (or a multiple of it) with a number of degrees of freedom equal to the difference between the number of parameters in the appropriate levels. The multiplicative constant is related to the asymptotic efficiencies of the common eigenvectors, which are the efficiencies of the off–diagonal elements of the scatter estimates considered. We have also obtained the partial influence functions of the test statistics which show how contaminating observations affect the test statistics.

A small simulation study complements the theoretical results obtained. It shows first that, for the number of observations considered, the desired level is almost attained when using the asymptotic distribution. It also shows that the loss of power of the robust procedure for non contaminated samples is negligible when compared to the classical procedure. Finally, it is obvious from the results given, that contamination of the samples can destroy both the level and the power of the classical versions of the tests, whereas the robust tests still lead to reliable results. A real example also illustrates some of these aspects.

Acknowledgment

This research was partially supported by Grant X-094 from the Universidad de Buenos Aires, Grant # 5505 from CONICET and Grant 13900-6 of the Fundación Antorchas at Buenos Aires, Argentina and by the Center for Mathematics and its Applications, Lisbon, Portugal, through *Programa Operacional "Ciência,* *Tecnologia, Inovação"* (POCTI) of the *Fundação para a Ciência e a Tecnologia* (FCT), cofinanced by the European Community fund FEDER. This research was also supported by the project ANCYT–GRICES PO/PA#04-EIII/020.

8 Appendix

PROOF OF THEOREM 1. Without loss of generality, we can assume that, under H_{CPC}, $\boldsymbol{\beta} = \mathbf{I}_p$, i.e, $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i = \text{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$.

First of all, note that $\widehat{\mathbf{Z}}_i = \sqrt{N}\,(\mathbf{V}_i - \boldsymbol{\Lambda}_i) \xrightarrow{\mathcal{D}} \mathbf{Z}_i$ where $\mathbf{Z}_i \sim N\left(\mathbf{0}, \frac{1}{\tau_i}\boldsymbol{\Xi}_i\right)$, with $\boldsymbol{\Xi}_i = \sigma_1\,(\mathbf{I} + K_{pp})\,(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Lambda}_i) + \sigma_2 \text{ vec}\,(\boldsymbol{\Lambda}_i) \text{ vec}\,(\boldsymbol{\Lambda}_i)^{\mathrm{T}}$. Note that when dealing with the maximum likelihood estimators, and when $G = N(\mathbf{0}, \mathbf{I}_p)$, $\sigma_2 = 0$ and $\sigma_1 = 1$, i.e., $\widehat{\mathbf{Z}}_{i,\text{ML}} = \sqrt{N}\,(\mathbf{S}_i - \boldsymbol{\Lambda}_i) \xrightarrow{\mathcal{D}} \widetilde{\mathbf{Z}}_i$ where $\widetilde{\mathbf{Z}}_i \sim N\left(\mathbf{0}, \frac{1}{\tau_i}\,(\mathbf{I} + K_{pp})\,(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Lambda}_i)\right)$.

Denote by $\widehat{\mathbf{U}}_i = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{\Lambda}_i\right)$. We will show that $\widehat{\mathbf{U}}_i$ are asymptotically normally distributed. The proof follows similar steps as those given in Theorem 2 in Boente and Orellana (2001). Denote $\mathbf{e}_j$ the j-th vector of the canonical basis and $\widehat{\mathbf{f}}_j = \left(\widehat{f}_{j1}, \ldots, \widehat{f}_{jp}\right)^{\mathrm{T}}$, with $\widehat{f}_{jj} = \sqrt{N}\left(\widehat{\beta}_{jj} - 1\right)$ and $\widehat{f}_{js} = \sqrt{N}\widehat{\beta}_{sj}$. Then, using that $\widehat{f}_{jj} = o_p(1)$ and that $\widehat{f}_{js} = O_p(1)$, for $s \neq j$, straighforward calculations lead to that

$$\begin{aligned}
\widehat{U}_{i,ss} &= \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_s^{\mathrm{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}}_s - \lambda_{is}\right) = \widehat{Z}_{i,ss} + o_p(1) \\
\widehat{U}_{i,sj} &= \sqrt{N}\widehat{\boldsymbol{\beta}}_s^{\mathrm{T}}\mathbf{V}_i\widehat{\boldsymbol{\beta}}_j = \widehat{Z}_{i,sj} + \lambda_{is}\widehat{f}_{js} + \lambda_{ij}\widehat{f}_{sj} + o_p(1)\ .
\end{aligned}$$

Note that as in the proof of Theorem 2 in Boente and Orellana (2001), $\widehat{\mathbf{f}} = \text{vec}\left(\widehat{\mathbf{f}}_1, \ldots, \widehat{\mathbf{f}}_p\right)$ can be written as $\widehat{\mathbf{f}} = \mathbf{B}^{-1}\widehat{\mathbf{d}} + o_p(1)$, with $\mathbf{B}$ a non–singular matrix and $\widehat{\mathbf{d}}$ with his first $p(p+1)/2$ rows equal to 0 and his last rows equal to vec$(\widehat{\mathbf{W}})$ where $\widehat{\mathbf{W}} = (\widehat{w}_{sj})_{1 \leq s < j \leq p}$

$$\widehat{w}_{sj} = -\widehat{w}_{js} = \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij} - \lambda_{is}}{\lambda_{ij}\lambda_{is}} \widehat{Z}_{i,sj} \quad \text{for } s \neq j\ .$$

The non–singular matrix $\mathbf{B}$ is related to the system of equations

$$\begin{aligned}
f_{js} + f_{sj} &= \theta_{js} \quad \text{for} \quad 1 \leq s \leq j \leq p \\
\mathbf{e}_j^{\mathrm{T}}\mathbf{A}_{js}\mathbf{f}_s + \mathbf{e}_s^{\mathrm{T}}\mathbf{A}_{js}\mathbf{f}_j &= \nu_{js} \quad \text{for} \quad 1 \leq s < j \leq p
\end{aligned}$$

where $\mathbf{A}_{js} = \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij} - \lambda_{is}}{\lambda_{ij}\lambda_{is}} \boldsymbol{\Lambda}_i$, with solution given by

$$\begin{aligned}
f_{jj} &= \frac{\theta_{jj}}{2} \quad \text{for} \quad 1 \leq j \leq p \\
f_{js} = \theta_{js} - f_{sj} &= -\frac{\nu_{js} - \theta_{js} \sum_{\ell=1}^{k} \tau_\ell \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\lambda_{\ell s}}}{\sum_{\ell=1}^{k} \tau_\ell \frac{(\lambda_{\ell j} - \lambda_{\ell s})^2}{\lambda_{\ell j}\lambda_{\ell s}}} \quad \text{for} \quad 1 \leq s < j \leq p\ .
\end{aligned}$$

which implies that

$$\widehat{f}_{jj} = o_p(1) \quad \text{for} \quad 1 \le m \le p$$

$$\widehat{f}_{js} = -\widehat{f}_{sj} + o_p(1) = -\frac{\widehat{w}_{js}}{\displaystyle\sum_{\ell=1}^{k} \tau_\ell \frac{(\lambda_{\ell j} - \lambda_{\ell s})^2}{\lambda_{\ell j}\lambda_{\ell s}}} + o_p(1) = \frac{\displaystyle\sum_{\ell=1}^{k} \tau_\ell \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\lambda_{\ell j}\lambda_{\ell s}} \widehat{Z}_{\ell,sj}}{\displaystyle\sum_{\ell=1}^{k} \tau_\ell \frac{(\lambda_{\ell j} - \lambda_{\ell s})^2}{\lambda_{\ell j}\lambda_{\ell s}}} + o_p(1) \quad \text{for} \quad 1 \le s < j \le p\,.$$

Putting the above expansions together we get that

$$\widehat{U}_{i,ss} = \widehat{Z}_{i,ss} + o_p(1) \tag{19}$$

$$\widehat{U}_{i,sj} = \widehat{Z}_{i,sj} + \frac{\lambda_{is} - \lambda_{ij}}{\displaystyle\sum_{\ell=1}^{k} \tau_\ell \frac{(\lambda_{\ell j} - \lambda_{\ell s})^2}{\lambda_{\ell j}\lambda_{\ell s}}} \sum_{\ell=1}^{k} \tau_\ell \frac{\lambda_{\ell j} - \lambda_{\ell s}}{\lambda_{\ell j}\lambda_{\ell s}} \widehat{Z}_{\ell,sj} + o_p(1) \quad 1 \le s < j \le p\,, \tag{20}$$

which shows that $\widehat{\mathbf{U}}_i = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{V}_i \widehat{\boldsymbol{\beta}} - \boldsymbol{\Lambda}_i\right)$ are jointly asymptotically normally distributed.

In Muirhead and Waternaux (1980), it is shown that given a sample of size n and a scatter statistic $\mathbf{S} = \boldsymbol{\Sigma} + n^{-\frac{1}{2}}\mathbf{Z}$ with $\mathbf{Z}$ an asymptotically normally distributed matrix, under $H_{04} : \boldsymbol{\Sigma} = \operatorname{diag}(\sigma_{11}, \ldots, \sigma_{pp})$ the test statistic $-2\log\Delta_4$ with $\Delta_4^{\frac{2}{n}} = \det(\mathbf{S}) \left[\prod_{i=1}^{p} s_{ii}\right]^{-1}$ can be expanded as

$$-2\log\Delta_4 = n\log\left(\frac{\det(\operatorname{diag}(\mathbf{S}))}{\det(\mathbf{S})}\right) = \sum_{1\le i<j\le p} \frac{z_{ij}^2}{\sigma_{ii}\sigma_{jj}} + O_p(n^{-\frac{1}{2}})$$

This expansion can be applied to each term in T_{CPC}, since H_{CPC} holds and we have assumed $\boldsymbol{\beta} = \mathbf{I}_p$. Therefore, using that $n_i = \tau_i N$ with $0 < \tau_i < 1$ fixed numbers, we have that for $1 \le i \le k$

$$n_i \log\left(\frac{\det\left(\operatorname{diag}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{V}_i \widehat{\boldsymbol{\beta}}\right)\right)}{\det\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{V}_i \widehat{\boldsymbol{\beta}}\right)}\right) = \tau_i \sum_{1\le s<j\le p} \frac{\widehat{U}_{i,sj}^2}{\sigma_{i,ss}\sigma_{i,jj}} + O_p(N^{-\frac{1}{2}})\,,$$

which entails that

$$T_{\mathrm{CPC}} = \sum_{i=1}^{k} \tau_i \sum_{1\le s<j\le p} \left[\widehat{U}_{i,sj}\right]^2 \frac{1}{\lambda_{is}\lambda_{ij}} + O_p(N^{-\frac{1}{2}})$$

since $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i$.

Therefore, the asymptotic behavior of T_{CPC} depends on the asymptotic distribution of $\widetilde{\mathbf{U}} = \left(\widehat{U}_{i,sj}\right)_{\substack{1\le i\le k\\ 1\le s<j\le p}}$ and therefore using (20) on that of $\widehat{\mathbf{Y}}_i = \left(\widehat{Z}_{i,sj}\right)_{1\le s<j\le p}$, $1 \le i \le k$. Note that $\widehat{\mathbf{Y}}_i$ are independent and asymptotically normally distributed.

Denote $\mathbf{Y}_i = \left(\widetilde{\mathbf{Z}}_{i,sj}\right)_{1\le s<j\le p}$ with $\widetilde{\mathbf{Z}}_i \sim N\left(\mathbf{0}, \frac{1}{\tau_i}(\mathbf{I} + K_{pp})(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Lambda}_i)\right)$ independent for $1 \le i \le k$. Then, $\widehat{\mathbf{Y}}_i \xrightarrow{\mathcal{D}} \sigma_1^{\frac{1}{2}} \mathbf{Y}_i$.

This implies that the asymptotic distribution of $T_{\rm CPC}$ is that of

$$T_1 = \sigma_1 \sum_{i=1}^k \tau_i \sum_{1\le s<j\le p} [U_{i,sj}]^2 \frac{1}{\lambda_{is}\lambda_{ij}}$$

with

$$U_{i,sj} \quad = \quad \widetilde{Z}_{i,sj} + \frac{\lambda_{is}-\lambda_{ij}}{\displaystyle\sum_{\ell=1}^k \tau_\ell \frac{(\lambda_{\ell j}-\lambda_{\ell s})^2}{\lambda_{\ell j}\lambda_{\ell s}}} \sum_{\ell=1}^k \tau_\ell \frac{\lambda_{\ell j}-\lambda_{\ell s}}{\lambda_{\ell j}\lambda_{\ell s}} \widetilde{Z}_{\ell,sj} \quad 1\le s<j\le p\,,$$

The distribution of $\frac{T_1}{\sigma_1}$ is the distribution of the likelihood ratio test for normal populations, since the above expansions hold in this particular case.

Therefore, $T_{\rm CPC}$ has asymptotic distribution $\sigma_1 \chi^2_{\frac{(k-1)p(p-1)}{2}}$. $\square$

Proof of Theorem 2. From the equality

$$\widehat{c}_{ij} - c_{ij} = \frac{\widehat{\lambda}_{ij}-\lambda_{ij}}{\widehat{\lambda}_{i1}} - \frac{\lambda_{ij}}{\widehat{\lambda}_{i1}\lambda_{i1}}\left(\widehat{\lambda}_{i1}-\lambda_{i1}\right)$$

we get easily,

$$\text{ASCOV}\left(\widehat{c}_{ij},\widehat{c}_{is}\right) \quad = \quad \frac{1}{\lambda_{i1}^2}\left[\text{ASCOV}\left(\widehat{\lambda}_{ij},\widehat{\lambda}_{is}\right) - \frac{\lambda_{ij}}{\lambda_{i1}}\text{ASCOV}\left(\widehat{\lambda}_{i1},\widehat{\lambda}_{is}\right) - \frac{\lambda_{is}}{\lambda_{i1}}\text{ASCOV}\left(\widehat{\lambda}_{i1},\widehat{\lambda}_{ij}\right) + \frac{\lambda_{ij}\lambda_{is}}{\lambda_{i1}^2}\text{ASVAR}\left(\widehat{\lambda}_{i1}\right)\right]$$

and thus, replacing by the expressions given in (7), we obtain

$$\begin{aligned} \text{ASCOV}\left(\widehat{c}_{ij},\widehat{c}_{is}\right) &= 2\sigma_1 \frac{1}{\tau_i}\frac{\lambda_{ij}\lambda_{is}}{\lambda_{i1}^2} \text{ for } j\ne s \\ \text{ASVAR}\left(\widehat{c}_{ij}\right) &= 4\sigma_1 \frac{1}{\tau_i}\frac{\lambda_{ij}^2}{\lambda_{i1}^2} \end{aligned}$$

Therefore, we have that under $H_{\rm CPC}$ $\sqrt{N}\left(\widehat{\mathbf{c}}_i - \mathbf{c}_i\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, 2\sigma_1 \frac{1}{\tau_i}\left(\mathbf{c}_i\mathbf{c}_i^{\rm T} + \text{diag}\left(c_{i2}^2,\dots,c_{ip}^2\right)\right)\right)$ and asymptotically independent for different i which entails that, under $H_{\rm CPC}$

$$\sqrt{N}\left(\widehat{\mathbf{c}}_i - \mathbf{c}_i - (\widehat{\mathbf{c}}_1 - \mathbf{c}_1)\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, 2\sigma_1\left[\frac{1}{\tau_i}\left(\mathbf{c}_i\mathbf{c}_i^{\rm T} + \text{diag}\left(c_{i2}^2,\dots,c_{ip}^2\right)\right) + \frac{1}{\tau_1}\left(\mathbf{c}_1\mathbf{c}_1^{\rm T} + \text{diag}\left(c_{12}^2,\dots,c_{1p}^2\right)\right)\right]\right)$$

and so under $H_{\rm PROP}$

$$\sqrt{N}\left(\widehat{\mathbf{c}}_i - \widehat{\mathbf{c}}_1\right) \quad \xrightarrow{\mathcal{D}} \quad N\left(\mathbf{0}, 2\sigma_1\left[\frac{1}{\tau_i}+\frac{1}{\tau_1}\right]\left(\mathbf{c}_1\mathbf{c}_1^{\rm T} + \text{diag}\left(c_{12}^2,\dots,c_{1p}^2\right)\right)\right)$$

Denote $\mathbf{B}_1 = 2\sigma_1\left(\mathbf{c}_1\mathbf{c}_1^{\rm T} + \text{diag}\left(c_{12}^2,\dots,c_{1p}^2\right)\right)$, $\mathbf{C} = (\mathbf{c}_1,\dots,\mathbf{c}_k)$ and $\widehat{\mathbf{C}} = (\widehat{\mathbf{c}}_1,\dots,\widehat{\mathbf{c}}_k)$. Using the asymptotic independence of the columns of $\widehat{\mathbf{C}}$ and that $\text{ASCOV}\left(\widehat{\mathbf{c}}_i - \widehat{\mathbf{c}}_1, \widehat{\mathbf{c}}_\ell - \widehat{\mathbf{c}}_1\right) = \text{ASVAR}\left(\widehat{\mathbf{c}}_1\right) = \frac{1}{\tau_1}\mathbf{B}_1$, we get that under $H_{\rm PROP}$,

$$\sqrt{N}\widehat{\mathbf{D}} \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \mathbf{\Sigma}_{\widehat{\mathbf{D}}} = \mathbf{B}_1 \otimes \left\{\frac{1}{\tau_1}\mathbf{1}\mathbf{1}^{\rm T}\right\} + \text{diag}\left(\frac{1}{\tau_2}\mathbf{B}_2,\dots,\frac{1}{\tau_k}\mathbf{B}_k\right)\right)$$

where $\mathbf{B}_i = 2\sigma_1 \left(\mathbf{c}_i \mathbf{c}_i^{\mathrm{T}} + \mathrm{diag}\left(c_{i2}^2, \ldots, c_{ip}^2\right)\right)$.

The matrix $\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbf{D}}}$ is a consistent estimate of $\boldsymbol{\Sigma}_{\widehat{\mathbf{D}}}$. Therefore, from the above discussion we have that, under H_{PROP}, $\mathcal{W}_{\mathrm{PROP}} \xrightarrow{\mathcal{D}} \chi^2_{(p-1)(k-1)}$.$\square$

Proof of Theorem 3. As in Theorem 1, without loss of generality, we can assume that, under H_{PROP}, $\boldsymbol{\beta} = \mathbf{I}_p$, i.e, $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i = \rho_i \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$.

Notice that $\widehat{\mathbf{Z}}_i = \sqrt{N}\left(\mathbf{V}_i - \boldsymbol{\Lambda}_i\right) \xrightarrow{\mathcal{D}} \mathbf{Z}_i$ where $\mathbf{Z}_i \sim N\left(\mathbf{0}, \frac{1}{\tau_i}\boldsymbol{\Xi}_i\right)$, with $\boldsymbol{\Xi}_i = \sigma_1\left(\mathbf{I} + K_{pp}\right)\left(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Lambda}_i\right) + \sigma_2 \ \mathrm{vec}\left(\boldsymbol{\Lambda}_i\right) \ \mathrm{vec}\left(\boldsymbol{\Lambda}_i\right)^{\mathrm{T}}$. Moreover, from the proof of Theorem 1, $\widehat{\mathbf{U}}_i = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{V}_i \widehat{\boldsymbol{\beta}} - \boldsymbol{\Lambda}_i\right)$ is asymptotically normally distributed and satisfies that

$$\widehat{U}_{i,ss} = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_s^{\mathrm{T}} \mathbf{V}_i \widehat{\boldsymbol{\beta}}_s - \lambda_{is}\right) = \widehat{Z}_{i,ss} + o_p(1) .$$

Using a Taylor's expansion of order 2 and since the eigenvalue and the proportionality constant estimators are asymptotically normally distributed, we get

$$\log\left(\widehat{\lambda}_{ij}\right) - \log\left(\widehat{\rho}_i \widehat{\lambda}_j\right) = \frac{1}{\widehat{\rho}_i \widehat{\lambda}_j}\left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right) - \frac{1}{2\widehat{\rho}_i^2 \widehat{\lambda}_j^2}\left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right)^2 + o_p\left(N^{-1}\right) .$$

Then, we get

$$\begin{aligned}
T_{\mathrm{PROP}} &= -\sum_{i=1}^k n_i \sum_{j=1}^p \log\left(\widehat{\lambda}_{ij}\right) - \log\left(\widehat{\rho}_i \widehat{\lambda}_j\right) \\
&= -\sum_{i=1}^k n_i \sum_{j=1}^p \frac{1}{\widehat{\rho}_i \widehat{\lambda}_j}\left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right) + \sum_{i=1}^k n_i \sum_{j=1}^p \frac{1}{2\widehat{\rho}_i^2 \widehat{\lambda}_j^2}\left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right)^2 + o_p(1) \\
&= -\sum_{i=1}^k n_i \sum_{j=1}^p \left(\frac{\widehat{\lambda}_{ij}}{\widehat{\rho}_i \widehat{\lambda}_j} - 1\right) + \sum_{i=1}^k n_i \sum_{j=1}^p \frac{1}{2\widehat{\rho}_i^2 \widehat{\lambda}_j^2}\left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right)^2 + o_p(1) \\
&= pN - \sum_{i=1}^k \frac{n_i}{\widehat{\rho}_i} \sum_{j=1}^p \frac{\widehat{\lambda}_{ij}}{\widehat{\lambda}_j} + \sum_{i=1}^k n_i \sum_{j=1}^p \frac{1}{2\widehat{\rho}_i^2 \widehat{\lambda}_j^2}\left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right)^2 + o_p(1) \\
&= pN - \sum_{i=1}^k \frac{n_i}{\widehat{\rho}_i} p \widehat{\rho}_i + \sum_{i=1}^k n_i \sum_{j=1}^p \frac{1}{2\widehat{\rho}_i^2 \widehat{\lambda}_j^2}\left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right)^2 + o_p(1) \\
&= \frac{1}{2}\sum_{i=1}^k \tau_i \sum_{j=1}^p \frac{1}{\widehat{\rho}_i^2 \widehat{\lambda}_j^2} \ N \ \left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right)^2 + o_p(1) \\
&= \frac{1}{2}\sum_{i=1}^k \tau_i \sum_{j=1}^p \frac{1}{\rho_i^2 \lambda_j^2} \ N \ \left(\widehat{\lambda}_{ij} - \widehat{\rho}_i \widehat{\lambda}_j\right)^2 + o_p(1) .
\end{aligned}$$

Using that

$$\widehat{\rho}_i = \frac{1}{p}\sum_{j=1}^p \frac{\widehat{\lambda}_{ij}}{\widehat{\lambda}_j} = \frac{1}{p}\sum_{j=1}^p \frac{\widehat{U}_{i,jj}}{\sqrt{N}\widehat{\lambda}_j} + \frac{\rho_i}{p}\sum_{j=1}^p \frac{\lambda_j}{\widehat{\lambda}_j}$$

$$\widehat{\lambda}_j = \sum_{i=1}^k \frac{\tau_i \widehat{\lambda}_{ij}}{\widehat{\rho}_i} = \sum_{i=1}^k \frac{\tau_i \widehat{U}_{i,jj}}{\sqrt{N}\widehat{\rho}_i} + \lambda_j \sum_{i=1}^k \frac{\tau_i \rho_i}{\widehat{\rho}_i} \ ,$$

we get

$$\sqrt{N}\left(\widehat{\rho}_i \widehat{\lambda}_j - \widehat{\lambda}_{ij}\right) = \sum_{\ell=1}^k \tau_\ell W_{\ell j} + \frac{1}{p}\sum_{s=1}^p W_{is} - \frac{1}{p}\sum_{s=1}^p \sum_{\ell=1}^k \tau_\ell W_{\ell s} - W_{ij} + o_p(1) \ ,$$

where $W_{ij} = \widehat{U}_{i,jj}/\lambda_{ij}$ is such that $\mathbf{W}_i \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{B})$ independent where $\mathbf{B} = 2\sigma_1 \mathbf{I}_p + \sigma_2 \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}$, i.e, $\mathbf{W} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}_k \otimes \mathbf{B})$. Thus,

$$\begin{aligned} T_{\mathrm{PROP}} &= \frac{1}{2}\sum_{i=1}^k \tau_i \sum_{j=1}^p \left(W_{ij} + \frac{V}{p} - M_j - N_i\right)^2 + o_p(1) \\ &= \frac{1}{2}\sum_{i=1}^k \tau_i \sum_{j=1}^p W_{ij}^2 + \frac{V^2}{p} - \sum_{j=1}^p M_j^2 - p\sum_{i=1}^k \tau_i N_i^2 + o_p(1) \ , \end{aligned}$$

where

$$M_j = \sum_{i=1}^k \tau_i W_{ij} \qquad N_i = \frac{1}{p}\sum_{j=1}^p W_{ij} \qquad V = \sum_{j=1}^p M_j = p \sum_{i=1}^k \tau_i N_i \ .$$

After some algebra, we get that the asymptotic distribution of T_{PROP} is that of $L_1 = \frac{1}{2}\mathbf{W}^{\mathrm{T}}\boldsymbol{\Upsilon}\mathbf{W}$ where $\boldsymbol{\Upsilon}$ is given by

$$\boldsymbol{\Upsilon} = \boldsymbol{\Delta} + \frac{1}{p}\mathbf{a}\mathbf{a}^{\mathrm{T}} - \mathbf{C} - \frac{1}{p}\mathbf{T}$$

with

$$\begin{aligned} \mathbf{T} &= \mathrm{diag}\left(\tau_1 \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}, \dots, \tau_k \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}\right) = \mathrm{diag}(\tau_1, \dots, \tau_k) \otimes \left(\mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}\right) \\ \mathbf{C} &= \sum_{j=1}^p \mathbf{B}_j \mathbf{1}_{kp} \mathbf{1}_{kp}^{\mathrm{T}} \mathbf{B}_j = \boldsymbol{\tau}\boldsymbol{\tau}^{\mathrm{T}} \otimes \mathbf{I}_p \\ \mathbf{B}_j &= \mathrm{diag}\left(\tau_1 \mathbf{e}_j^{\mathrm{T}}, \dots, \tau_k \mathbf{e}_j^{\mathrm{T}}\right) \\ \boldsymbol{\tau} &= (\tau_1, \dots, \tau_k)^{\mathrm{T}} \\ \mathbf{a} &= \left(\tau_1 \mathbf{1}_p^{\mathrm{T}}, \dots, \tau_k \mathbf{1}_p^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{\tau} \otimes \mathbf{1}_p \\ \boldsymbol{\Delta} &= \mathrm{diag}\left(\tau_1 \mathbf{1}_p^{\mathrm{T}}, \dots, \tau_k \mathbf{1}_p^{\mathrm{T}}\right)^{\mathrm{T}} = \mathrm{diag}(\tau_1, \dots, \tau_k) \otimes \mathbf{I}_p \end{aligned}$$

It is easy to see that $\boldsymbol{\Upsilon}\left(\mathbf{I}_k \otimes \left(\mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}\right)\right) = 0$, thus the asymptotic distribution of T_{PROP} is the distribution of $L_2 = \mathbf{W}^{*\mathrm{T}}\boldsymbol{\Upsilon}\mathbf{W}^*$ with $\mathbf{W}^* \sim N(\mathbf{0}, \sigma_1 \mathbf{I}_{kp})$. Using that $\boldsymbol{\Upsilon}$ is idempotent with rank $(p-1)(k-1)$ we get the desired result. $\square$

The following Lemma will be useful to prove Theorem 4.

Lemma 8.1. *Let $\mathbf{V}_i(F)$ be a scatter functional such that $\mathbf{V}_i(F_i) = \boldsymbol{\Sigma}_i$. Denote by $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p$, $\lambda_{i1}, \dots, \lambda_{ip}$ the common eigenvectors and the eigenvalues of $\boldsymbol{\Sigma}_i$, i.e., assume that $H_{\mathrm{CPC}}:\ \boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}$, for $1 \le i \le k$ holds. Assume that the influence function $IF(\mathbf{x}, \mathbf{V}_i, F_i)$ and $\frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}$ exist and that*

$\lambda_{11} > \ldots > \lambda_{1p}$. Let $\boldsymbol{\beta}_{\mathbf{V}}(F)$ and $\lambda_{\mathbf{V},\ell j}$ denote the plug–in functionals related to the scatter $\mathbf{V}(F) = (\mathbf{V}_1(F_1), \ldots, \mathbf{V}_k(F_k))$ and let $F_{\epsilon,\mathbf{x},i} = F_1 \times \ldots \times F_{i-1} \times F_{i,\epsilon,\mathbf{x}} \times F_{i+1} \times \ldots \times F_k$ where $F_{i,\epsilon,\mathbf{x}} = (1-\epsilon)F_i + \epsilon\delta_{\mathbf{x}}$. Moreover, define $\lambda^{(i)}_{\ell j,\epsilon} = \lambda_{\mathbf{V},\ell j}\left(F_{\epsilon,\mathbf{x},i}\right)$, then we have that

$$\frac{\partial^2}{\partial\epsilon^2}\lambda^{(i)}_{ij,\epsilon}\Big|_{\epsilon=0} = 4\tau_i \sum_{m\neq j} \frac{\lambda_{ij}-\lambda_{im}}{\lambda_{im}\lambda_{ij}} A_{mj}\left[\boldsymbol{\beta}_j^{\mathrm{T}} IF\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_m\right]^2 + + 2\,\tau_i^2 \sum_{m\neq j} \frac{\left(\lambda_{im}-\lambda_{ij}\right)^3}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} IF\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_m\right]^2 + \boldsymbol{\beta}_j^{\mathrm{T}} \frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j \quad (21)$$

$$\frac{\partial^2}{\partial\epsilon^2}\lambda^{(i)}_{\ell j,\epsilon}\Big|_{\epsilon=0} = 2\,\tau_i^2 \sum_{m\neq j} \left(\lambda_{\ell m}-\lambda_{\ell j}\right) \frac{\left(\lambda_{ij}-\lambda_{im}\right)^2}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} IF\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_m\right]^2 \qquad \ell \neq i \quad (22)$$

where $A_{mj} = \left\{\sum_{\ell=1}^k \tau_\ell \frac{\left(\lambda_{\ell m}-\lambda_{\ell j}\right)^2}{\lambda_{\ell m}\lambda_{\ell j}}\right\}^{-1}$.

Proof. Using (10), we have that $\lambda^{(i)}_{\ell j,\epsilon} = \lambda_{\mathbf{V},\ell j}\left(F_{\epsilon,\mathbf{x},i}\right) = \boldsymbol{\beta}_{j,\epsilon}^{\mathrm{T}}\mathbf{V}_\ell\boldsymbol{\beta}_{j,\epsilon}$ for $\ell \neq i$ and $\lambda^{(i)}_{ij,\epsilon} = \lambda_{\mathbf{V},ij}\left(F_{\epsilon,\mathbf{x},i}\right) = \boldsymbol{\beta}_{j,\epsilon}^{\mathrm{T}}\mathbf{V}_{i,\epsilon}\boldsymbol{\beta}_{j,\epsilon}$ with $\boldsymbol{\beta}_{j,\epsilon} = \boldsymbol{\beta}_{\mathbf{V},j}\left(F_{\epsilon,\mathbf{x},i}\right)$. Thus, straighforward calculations, lead to

$$\begin{aligned}
\frac{\partial^2}{\partial\epsilon^2}\lambda^{(i)}_{ij,\epsilon}\Big|_{\epsilon=0} &= 2\frac{\partial^2}{\partial\epsilon^2}\boldsymbol{\beta}_{j,\epsilon}\Big|_{\epsilon=0}^{\mathrm{T}} \boldsymbol{\Sigma}_i\boldsymbol{\beta}_j + 4\ \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_j + \\
&+ 2\ \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)^{\mathrm{T}} \boldsymbol{\Sigma}_i \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right) + \boldsymbol{\beta}_j^{\mathrm{T}} \frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j \\
&= 2\,\lambda_{ij}\frac{\partial^2}{\partial\epsilon^2}\boldsymbol{\beta}_{j,\epsilon}\Big|_{\epsilon=0}^{\mathrm{T}} \boldsymbol{\beta}_j + 4\ \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_j + \\
&+ 2\ \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)^{\mathrm{T}} \boldsymbol{\Sigma}_i\ \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right) + \boldsymbol{\beta}_j^{\mathrm{T}} \frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j
\end{aligned}$$

and

$$\frac{\partial^2}{\partial\epsilon^2}\lambda^{(i)}_{\ell j,\epsilon}\Big|_{\epsilon=0} = 2\,\lambda_{\ell j}\frac{\partial^2}{\partial\epsilon^2}\boldsymbol{\beta}_{j,\epsilon}\Big|_{\epsilon=0}^{\mathrm{T}} \boldsymbol{\beta}_j + 2\ \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)^{\mathrm{T}} \boldsymbol{\Sigma}_\ell\ \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)$$

Using the orthogonality conditions we get that $\frac{\partial^2}{\partial\epsilon^2}\boldsymbol{\beta}_{j,\epsilon}\Big|_{\epsilon=0}^{\mathrm{T}} \boldsymbol{\beta}_j = -\mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)^{\mathrm{T}} \mathrm{PIF}_i\left(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},j}, F\right)$. From Theorem 1 in Boente, Pires and Rodrigues (2002), we have that

$$\mathrm{PIF}_i(\mathbf{x}, \boldsymbol{\beta}_{\mathbf{V},\,j}, F) = \tau_i \sum_{m\neq j} \frac{\lambda_{ij}-\lambda_{im}}{\lambda_{im}\lambda_{ij}} A_{mj}\left\{\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_m\right\}\boldsymbol{\beta}_m\ ,$$

and so

$$\begin{aligned}
\frac{\partial^2}{\partial\epsilon^2}\lambda^{(i)}_{ij,\epsilon}\Big|_{\epsilon=0} &= -2\,\lambda_{ij}\tau_i^2 \sum_{m\neq j} \frac{\left(\lambda_{ij}-\lambda_{im}\right)^2}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_m\right]^2 + \\
&+ 4\tau_i \sum_{m\neq j} \frac{\lambda_{ij}-\lambda_{im}}{\lambda_{im}\lambda_{ij}} A_{mj}\left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_m\right]^2 + \\
&+ 2\,\tau_i^2 \sum_{m\neq j} \lambda_{im} \frac{\left(\lambda_{ij}-\lambda_{im}\right)^2}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x}, \mathbf{V}_i, F_i\right)\boldsymbol{\beta}_m\right]^2 + \boldsymbol{\beta}_j^{\mathrm{T}} \frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j
\end{aligned}$$

$$
\begin{aligned}
&= 4\tau_i \sum_{m\neq j} \frac{\lambda_{ij}-\lambda_{im}}{\lambda_{im}\lambda_{ij}} A_{mj} \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 + \\
&+ 2\,\tau_i^2 \sum_{m\neq j} \frac{\left(\lambda_{im}-\lambda_{ij}\right)^3}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 + \boldsymbol{\beta}_j^{\mathrm{T}} \frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2}{\partial\epsilon^2}\lambda_{\ell j,\epsilon}\Big|_{\epsilon=0} &= -2\,\lambda_{\ell j}\tau_i^2 \sum_{m\neq j} \frac{\left(\lambda_{ij}-\lambda_{im}\right)^2}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 + \\
&+ 2\,\tau_i^2 \sum_{m\neq j} \lambda_{\ell m} \frac{\left(\lambda_{ij}-\lambda_{im}\right)^2}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 \\
&= 2\,\tau_i^2 \sum_{m\neq j} \left(\lambda_{\ell m}-\lambda_{\ell j}\right) \frac{\left(\lambda_{ij}-\lambda_{im}\right)^2}{\lambda_{im}^2\lambda_{ij}^2} A_{mj}^2 \left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 ,
\end{aligned}
$$

concluding the proof. □

Proof of Theorem 4. We need to compute $\frac{\partial^2}{\partial\epsilon^2} T_{\mathbf{V},\mathrm{CPC}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0}$. Note that if $\ell \neq i$, then, $\frac{\partial}{\partial\epsilon}\log\left(\lambda_j\left(\mathbf{V}_\ell(F_\ell)\right)\right) = 0$ for all j.

$$
\begin{aligned}
\frac{\partial^2}{\partial\epsilon^2} T_{\mathbf{V},\mathrm{CPC}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} &= \frac{\partial^2}{\partial\epsilon^2} \sum_{\ell\neq i}\tau_\ell \sum_{j=1}^p \log\left(\lambda_{\mathbf{V},\ell j}(F_{\epsilon,\mathbf{x},i})\right) - \log\left(\lambda_j\left(\mathbf{V}_\ell(F_\ell)\right)\right)\Big|_{\epsilon=0} + \\
&+ \frac{\partial^2}{\partial\epsilon^2}\tau_i \sum_{j=1}^p \log\left(\lambda_{\mathbf{V},ij}(F_{\epsilon,\mathbf{x},i})\right) - \log\left(\lambda_j\left(\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\right)\right)\Big|_{\epsilon=0} \\
&= \sum_{\ell\neq i}\tau_\ell \sum_{j=1}^p \frac{\partial^2}{\partial\epsilon^2}\log\left(\lambda_{\mathbf{V},\ell j}(F_{\epsilon,\mathbf{x},i})\right)\Big|_{\epsilon=0} + \\
&+ \tau_i \sum_{j=1}^p \frac{\partial^2}{\partial\epsilon^2}\left[\log\left(\lambda_{\mathbf{V},ij}(F_{\epsilon,\mathbf{x},i})\right) - \log\left(\lambda_j\left(\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\right)\right)\right]\Big|_{\epsilon=0} \\
&= \sum_{\ell\neq i}\tau_\ell \sum_{j=1}^p \frac{-\left[\mathrm{PIF}_i(\mathbf{x},\lambda_{\mathbf{V},\,\ell j},F)\right]^2}{\lambda_{\ell j}^2} + \frac{1}{\lambda_{\ell j}}\frac{\partial^2}{\partial\epsilon^2}\lambda_{\mathbf{V},\ell j}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} + \\
&+ \tau_i \sum_{j=1}^p \frac{-\left[\mathrm{PIF}_i(\mathbf{x},\lambda_{\mathbf{V},\,ij},F)\right]^2}{\lambda_{ij}^2} + \frac{1}{\lambda_{ij}}\frac{\partial^2}{\partial\epsilon^2}\lambda_{\mathbf{V},ij}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} \\
&- \left\{\frac{-\left[\mathrm{IF}(\mathbf{x},\lambda_j\left(\mathbf{V}_i\right),F_i)\right]^2}{\lambda_{ij}^2} + \frac{1}{\lambda_{ij}}\frac{\partial^2}{\partial\epsilon^2}\lambda_j\left(\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\right)\right\}\Big|_{\epsilon=0}
\end{aligned}
$$

Using Theorem 1 in Boente, Pires and Rodrigues (2002) we have that $\mathrm{PIF}_i(\mathbf{x},\lambda_{\mathbf{V},\,\ell j},F) = \delta_{\ell i}\,\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_j$, while Lemma 3 in Croux and Haesbroek (2000) which gives an expression for the influence function of a robust scatter functional entails that $\mathrm{IF}(\mathbf{x},\lambda_j\left(\mathbf{V}_i\right),F_i) = \boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_j$. Therefore

$$
\begin{aligned}
\frac{\partial^2}{\partial\epsilon^2} T_{\mathbf{V},\mathrm{CPC}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} &= \sum_{\ell\neq i}\tau_\ell \sum_{j=1}^p \frac{1}{\lambda_{\ell j}}\frac{\partial^2}{\partial\epsilon^2}\lambda_{\mathbf{V},\ell j}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} + \\
&+ \tau_i \sum_{j=1}^p \frac{1}{\lambda_{ij}}\left[\frac{\partial^2}{\partial\epsilon^2}\lambda_{\mathbf{V},ij}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} - \frac{\partial^2}{\partial\epsilon^2}\lambda_j\left(\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\right)\Big|_{\epsilon=0}\right] \qquad (23)
\end{aligned}
$$

Let $\boldsymbol{\beta}_{j,i,\epsilon}$ be the eigenvector of $\mathbf{V}_{i,\epsilon} = \mathbf{V}_i\left(F_{i,\epsilon,\mathbf{x}}\right)$, then $\boldsymbol{\beta}_{j,i,0} = \boldsymbol{\beta}_j$ and $\lambda_{j,i,\epsilon} = \lambda_j\left(\mathbf{V}_i\left(F_{i,\epsilon,\mathbf{x}}\right)\right) = \boldsymbol{\beta}_{j,i,\epsilon}^{\mathrm{T}}\mathbf{V}_{i,\epsilon}\boldsymbol{\beta}_{j,i,\epsilon}$ which entails that $\frac{\partial}{\partial\epsilon}\lambda_{j,i,\epsilon} = 2\left(\frac{\partial}{\partial\epsilon}\boldsymbol{\beta}_{j,i,\epsilon}\right)^{\mathrm{T}}\mathbf{V}_{i,\epsilon}\boldsymbol{\beta}_{j,i,\epsilon} + \boldsymbol{\beta}_{j,i,\epsilon}^{\mathrm{T}}\left(\frac{\partial}{\partial\epsilon}\mathbf{V}_{i,\epsilon}\right)\boldsymbol{\beta}_{j,i,\epsilon}$, thus

$$\begin{aligned}\frac{\partial^2}{\partial\epsilon^2}\lambda_j\left(\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\right)\Big|_{\epsilon=0} &= 2\frac{\partial^2}{\partial\epsilon^2}\boldsymbol{\beta}_{j,i,\epsilon}\Big|_{\epsilon=0}^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\beta}_j + 4\,\mathrm{IF}\left(\mathbf{x},\boldsymbol{\beta}_j\left(\mathbf{V}_i\right),F_i\right)^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_j + \\ &+ 2\,\mathrm{IF}\left(\mathbf{x},\boldsymbol{\beta}_j\left(\mathbf{V}_i\right),F_i\right)^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathrm{IF}\left(\mathbf{x},\boldsymbol{\beta}_j\left(\mathbf{V}_i\right),F_i\right) + \boldsymbol{\beta}_j^{\mathrm{T}}\frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j\end{aligned}$$

Using that $\boldsymbol{\Sigma}_i\boldsymbol{\beta}_j = \lambda_{ij}\boldsymbol{\beta}_j$, the expression for the influence function of an equivariant scatter matrix given in Lemma 3 in Croux and Haesbroek (2000)

$$\mathrm{IF}\left(\mathbf{x},\boldsymbol{\beta}_j\left(\mathbf{V}_i\right),F_i\right) = \sum_{m\neq j}\frac{1}{\lambda_{ij}-\lambda_{im}}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]\boldsymbol{\beta}_m$$

and the fact that $\boldsymbol{\beta}_{j,i,\epsilon}^{\mathrm{T}}\boldsymbol{\beta}_{j,i,\epsilon} = 1$ which entails that $\frac{\partial}{\partial\epsilon}\boldsymbol{\beta}_{j,i,\epsilon}^{\mathrm{T}}\boldsymbol{\beta}_j = 0$ and so

$$\frac{\partial^2}{\partial\epsilon^2}\boldsymbol{\beta}_{j,i,\epsilon}\Big|_{\epsilon=0}^{\mathrm{T}}\boldsymbol{\beta}_j = -\mathrm{IF}\left(\mathbf{x},\boldsymbol{\beta}_j\left(\mathbf{V}_i\right),F_i\right)^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\boldsymbol{\beta}_j\left(\mathbf{V}_i\right),F_i\right) = -\sum_{m\neq j}\frac{1}{\left(\lambda_{ij}-\lambda_{im}\right)^2}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2$$

we get

$$\begin{aligned}\frac{\partial^2}{\partial\epsilon^2}\lambda_j\left(\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\right)\Big|_{\epsilon=0} &= -2\lambda_{ij}\sum_{m\neq j}\frac{1}{\left(\lambda_{ij}-\lambda_{im}\right)^2}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 + \\ &+ 4\sum_{m\neq j}\frac{1}{\lambda_{ij}-\lambda_{im}}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 + \\ &+ 2\sum_{m\neq j}\frac{\lambda_{im}}{\left(\lambda_{ij}-\lambda_{im}\right)^2}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 + \boldsymbol{\beta}_j^{\mathrm{T}}\frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j \\ &= \boldsymbol{\beta}_j^{\mathrm{T}}\frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j + \\ &+ 2\sum_{m\neq j}\frac{1}{\left(\lambda_{ij}-\lambda_{im}\right)^2}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2\left[\lambda_{im}-\lambda_{ij}+2\left(\lambda_{ij}-\lambda_{im}\right)\right] \\ &= \boldsymbol{\beta}_j^{\mathrm{T}}\frac{\partial^2}{\partial\epsilon^2}\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\Big|_{\epsilon=0}\boldsymbol{\beta}_j + 2\sum_{m\neq j}\frac{1}{\left(\lambda_{ij}-\lambda_{im}\right)}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 . \quad (24)\end{aligned}$$

Thus, from Lemma 8.1 using (24) and (21), we obtain that

$$\begin{aligned}\frac{\partial^2}{\partial\epsilon^2}\left[\lambda_{\mathbf{V},ij}(F_{\epsilon,\mathbf{x},i}) - \lambda_j\left(\mathbf{V}_i(F_{i,\epsilon,\mathbf{x}})\right)\right]\Big|_{\epsilon=0} &= 4\tau_i\sum_{m\neq j}\frac{\lambda_{ij}-\lambda_{im}}{\lambda_{im}\lambda_{ij}}A_{mj}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 + \\ &+ 2\,\tau_i^2\sum_{m\neq j}\frac{\left(\lambda_{im}-\lambda_{ij}\right)^3}{\lambda_{im}^2\lambda_{ij}^2}A_{mj}^2\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2 \\ &- 2\sum_{m\neq j}\frac{1}{\left(\lambda_{ij}-\lambda_{im}\right)}\left[\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}\left(\mathbf{x},\mathbf{V}_i,F_i\right)\boldsymbol{\beta}_m\right]^2\end{aligned}$$

The equality (13) follows now using (23) and the expression (22) given in Lemma 8.1. Straightforward calculations allow to derive (14) from (13). □

Proof of Theorem 5. We need to compute $\frac{\partial^2}{\partial\epsilon^2}\mathcal{W}_{\mathbf{V},\text{PROP}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0}$. Since,

$$\frac{\partial^2}{\partial\epsilon^2}\mathcal{W}_{\mathbf{V},\text{PROP}}(F_{\epsilon,\mathbf{x},i}) = 2vec\left(\frac{\partial}{\partial\epsilon}\mathbf{D}(F_{\epsilon,\mathbf{x},i})\right)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{D}}^{-1}(F_{\epsilon,\mathbf{x},i})vec\left(\mathbf{D}(F_{\epsilon,\mathbf{x},i})\right)$$

using that under H_{PROP}, $\mathbf{D}(F)=0$ we get (16). On the other hand, using that $\text{PIF}_i\left(\mathbf{x},\lambda_{\ell j}(F),F\right)=0$ if $i\neq\ell$ and $\text{PIF}_i\left(\mathbf{x},\lambda_{\ell j}(F),F\right)=\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j$ the proof follows easily. □

Proof of Theorem 6. We need to compute $\frac{\partial^2}{\partial\epsilon^2}T_{\mathbf{V},\text{PROP}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0}$. Since

$$T_{\mathbf{V},\text{PROP}}(F_{\epsilon,\mathbf{x},i}) = \sum_{\ell=1}^{k}\tau_\ell\sum_{j=1}^{p}\log\left(\rho_{\mathbf{V},\ell}(F)\right)+\log\left(\lambda_{\mathbf{V},j}(F)\right)-\log\left(\lambda_{\mathbf{V},\ell j}(F)\right)$$

and $\text{PIF}_i\left(\mathbf{x},\lambda_{\mathbf{V},\ell j},F\right)=\delta_{\ell i}\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j$, we have

$$\begin{aligned}
\frac{\partial^2}{\partial\epsilon^2}T_{\mathbf{V},\text{PROP}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} &= \sum_{\ell=1}^{k}\tau_\ell\sum_{j=1}^{p}\frac{\partial^2}{\partial\epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0}\frac{1}{\rho_\ell}+\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{j,\epsilon}\right)\Big|_{\epsilon=0}\frac{1}{\lambda_j}-\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0}\frac{1}{\lambda_j\rho_\ell}\\
&+ \sum_{\ell=1}^{k}\tau_\ell\sum_{j=1}^{p}\frac{\left[\text{PIF}_i\left(\mathbf{x},\lambda_{\mathbf{V},\ell j},F\right)\right]^2}{\lambda_{\ell j}^2}-\frac{\left[\text{PIF}_i\left(\mathbf{x},\rho_{\mathbf{V},\ell},F\right)\right]^2}{\rho_\ell^2}-\frac{\left[\text{PIF}_i\left(\mathbf{x},\lambda_{\mathbf{V},j},F\right)\right]^2}{\lambda_j^2}\\
&= S_1+S_2\\
S_1 &= p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell}\frac{\partial^2}{\partial\epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0}+\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{j,\epsilon}\right)\Big|_{\epsilon=0}-\sum_{\ell=1}^{k}\tau_\ell\sum_{j=1}^{p}\frac{1}{\rho_\ell\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0}\\
S_2 &= \tau_i\sum_{j=1}^{p}\frac{\left[\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j\right]^2}{\rho_i^2\lambda_j^2}-p\sum_{\ell=2}^{k}\frac{\tau_\ell A_i^2}{\rho_\ell^2}\left[\delta_{\ell i}(1-\delta_{i1})+\rho_\ell^2\delta_{i1}(1-\delta_{\ell i})\right]\\
&- \sum_{j=1}^{p}\frac{1}{\lambda_j^2}\left[\frac{\tau_i}{\rho_i}\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j-\frac{\tau_i}{\rho_i}\lambda_jA_i+\lambda_jA_1\delta_{i1}\right]^2\\
A_i &= A_i(\mathbf{x})=\frac{1}{p}\sum_{j=1}^{p}\frac{\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j}{\lambda_j}
\end{aligned}$$

where the last equalities follow from Lemma 7.1 in Boente, Crichtley and Orellana (2005), since

$$\begin{aligned}
\text{PIF}_i\left(\mathbf{x},\rho_{\mathbf{V},\ell},F\right) &= A_i(\mathbf{x})\ (1-\delta_{i1})\,\delta_{\ell i}-\rho_\ell A_1(\mathbf{x})\delta_{i1}\,(1-\delta_{\ell i}) \qquad 2\le\ell\le k\\
\text{PIF}_i\left(\mathbf{x},\lambda_{\mathbf{V},j},F\right) &= \frac{\tau_i}{\rho_i}\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j-\frac{\tau_i}{\rho_i}\lambda_jA_i(\mathbf{x})+\lambda_jA_1(\mathbf{x})\delta_{i1} \qquad 1\le j\le p
\end{aligned}$$

a) We begin by computing S_2.
<u>For $i\neq 1$</u>

$$\begin{aligned}
S_2 &= \tau_i\sum_{j=1}^{p}\frac{\left[\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j\right]^2}{\rho_i^2\lambda_j^2}-p\frac{\tau_iA_i^2}{\rho_i^2}-\frac{\tau_i^2}{\rho_i^2}\sum_{j=1}^{p}\frac{1}{\lambda_j^2}\left[\left(\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j\right)^2\right.\\
&- \left.2\lambda_jA_i\boldsymbol{\beta}_j^{\mathrm{T}}\text{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j+\lambda_j^2A_i^2\right]\\
&= \frac{\tau_i(1-\tau_i)}{\rho_i^2}\xi_i-pA_i^2\frac{\tau_i}{\rho_i^2}+pA_i^2\frac{\tau_i^2}{\rho_i^2}=\frac{\tau_i(1-\tau_i)}{\rho_i^2}\left(\xi_i-pA_i^2\right)=\frac{\tau_i(1-\tau_i)}{\rho_i^2}\left(\xi_i-\frac{1}{p}\gamma_i^2\right)
\end{aligned}$$

For $i = 1$

$$
\begin{aligned}
S_2 &= \tau_1 \sum_{j=1}^{p} \frac{\left[\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}(\mathbf{x}, \mathbf{V}_1, F)\boldsymbol{\beta}_j\right]^2}{\lambda_j^2} - pA_1^2(1-\tau_1) \\
&- \sum_{j=1}^{p} \frac{1}{\lambda_j^2}\left[\tau_1^2\left(\boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}(\mathbf{x}, \mathbf{V}_1, F)\boldsymbol{\beta}_j\right)^2 + 2\tau_1(1-\tau_1)\lambda_j A_1 \boldsymbol{\beta}_j^{\mathrm{T}} \mathrm{IF}(\mathbf{x}, \mathbf{V}_1, F)\boldsymbol{\beta}_j + (1-\tau_1)^2\lambda_j^2 A_1^2\right] \\
&= \tau_1(1-\tau_1)\xi_1 - pA_1^2(1-\tau_1) - pA_1^2(1-\tau_1^2)
\end{aligned}
$$

b) We compute now S_1. Using that

$$
\rho_{\mathbf{V},\ell}(F) = \frac{1}{p}\sum_{j=1}^{p} \frac{\lambda_{\mathbf{V},\ell j}}{\lambda_{\mathbf{V},j}(F)} \quad \ell = 2,\dots,k \qquad \lambda_{\mathbf{V},j}(F) = \sum_{i=1}^{k} \tau_i \frac{\lambda_{\mathbf{V},ij}(F)}{\rho_{\mathbf{V},i}(F)} \quad 1 \le j \le p
$$

we get that

$$
\begin{aligned}
\frac{\partial^2}{\partial \epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0} &= \frac{1}{p}\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} - \rho_\ell \frac{1}{p}\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{j,\epsilon}\right)\Big|_{\epsilon=0} \\
&- \frac{2}{p}\sum_{j=1}^{p}\frac{1}{\lambda_j^2}\mathrm{PIF}_i\left(\mathbf{x}, \lambda_{\mathbf{V},j}, F\right)\mathrm{PIF}_i\left(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F\right) + \frac{2}{p}\rho_\ell\sum_{j=1}^{p}\frac{1}{\lambda_j^2}\mathrm{PIF}_i\left(\mathbf{x}, \lambda_{\mathbf{V},j}, F\right)^2 \\
\frac{\partial^2}{\partial \epsilon^2}\left(\lambda_{j,\epsilon}\right)\Big|_{\epsilon=0} &= \sum_{\ell=1}^{k}\frac{\tau_\ell}{\rho_\ell}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\mathrm{PIF}_i\left(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F\right)\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right) \\
&+ 2\lambda_j\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right)^2 - \lambda_j\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell}\frac{\partial^2}{\partial\epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0}
\end{aligned}
$$

which implies that

$$
\begin{aligned}
\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial \epsilon^2}\left(\lambda_{j,\epsilon}\right)\Big|_{\epsilon=0} &= \sum_{\ell=1}^{k}\frac{\tau_\ell}{\rho_\ell}\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\mathrm{PIF}_i\left(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F\right)\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right) \\
&+ 2\sum_{j=1}^{p}\frac{1}{\lambda_j}\lambda_j\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right)^2 - \sum_{j=1}^{p}\frac{1}{\lambda_j}\lambda_j\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell}\frac{\partial^2}{\partial\epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0} \\
&= \sum_{\ell=1}^{k}\frac{\tau_\ell}{\rho_\ell}\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\mathrm{PIF}_i\left(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F\right)\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right) \\
&+ 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right)^2 - p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell}\frac{\partial^2}{\partial\epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial \epsilon^2}\left(\lambda_{j,\epsilon}\right)\Big|_{\epsilon=0} &+ p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell}\frac{\partial^2}{\partial\epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0} = \sum_{\ell=1}^{k}\frac{\tau_\ell}{\rho_\ell}\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} \\
&- 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\mathrm{PIF}_i\left(\mathbf{x}, \lambda_{\mathbf{V},\ell j}, F\right)\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right) + 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\mathrm{PIF}_i\left(\mathbf{x}, \rho_{\mathbf{V},\ell}, F\right)^2
\end{aligned}
$$

$$
\begin{aligned}
S_1 &= p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell}\frac{\partial^2}{\partial\epsilon^2}\left(\rho_{\ell,\epsilon}\right)\Big|_{\epsilon=0} + \sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{j,\epsilon}\right)\Big|_{\epsilon=0} - \sum_{\ell=1}^{k}\tau_\ell\sum_{j=1}^{p}\frac{1}{\rho_\ell\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} \\
&= \sum_{\ell=1}^{k}\frac{\tau_\ell}{\rho_\ell}\sum_{j=1}^{p}\frac{1}{\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\mathrm{PIF}_i\left(\mathbf{x},\lambda_{\mathbf{V},\ell j},F\right)\mathrm{PIF}_i\left(\mathbf{x},\rho_{\mathbf{V},\ell},F\right) \\
&+ 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\mathrm{PIF}_i\left(\mathbf{x},\rho_{\mathbf{V},\ell},F\right)^2 - \sum_{\ell=1}^{k}\tau_\ell\sum_{j=1}^{p}\frac{1}{\rho_\ell\lambda_j}\frac{\partial^2}{\partial\epsilon^2}\left(\lambda_{\ell j,\epsilon}\right)\Big|_{\epsilon=0} \\
&= 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\mathrm{PIF}_i\left(\mathbf{x},\rho_{\mathbf{V},\ell},F\right)^2 - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\mathrm{PIF}_i\left(\mathbf{x},\lambda_{\mathbf{V},\ell j},F\right)\mathrm{PIF}_i\left(\mathbf{x},\rho_{\mathbf{V},\ell},F\right) \\
&= 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}A_i^2\left[\left(1-\delta_{i1}\right)\delta_{\ell i} + \rho_\ell^2\delta_{i1}\left(1-\delta_{\ell i}\right)\right] \\
&\quad -2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\mathrm{PIF}_i\left(\mathbf{x},\lambda_{\mathbf{V},\ell j},F\right)A_i\left[\left(1-\delta_{i1}\right)\delta_{\ell i} - \rho_\ell\delta_{i1}\left(1-\delta_{\ell i}\right)\right] \\
&= 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}A_i^2\left[\left(1-\delta_{i1}\right)\delta_{\ell i} + \rho_\ell^2\delta_{i1}\left(1-\delta_{\ell i}\right)\right] - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\delta_{\ell i}\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j A_i\left(1-\delta_{i1}\right)
\end{aligned}
$$

Thus, <u>for $i \neq 1$</u>

$$
\begin{aligned}
S_1 &= 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}A_i^2\delta_{\ell i} - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\delta_{\ell i}\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j A_i \\
&= 2p\frac{\tau_i}{\rho_i^2}A_i^2 - 2\frac{\tau_i}{\rho_i^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}(\mathbf{x},\mathbf{V}_i,F)\boldsymbol{\beta}_j A_i = 2p\frac{\tau_i}{\rho_i^2}A_i^2 - 2\frac{\tau_i}{\rho_i^2}pA_i^2 = 0
\end{aligned}
$$

which entails together with a)

$$
\frac{\partial^2}{\partial\epsilon^2}T_{\mathbf{V},\mathrm{PROP}}(F_{\epsilon,\mathbf{x},i})\Big|_{\epsilon=0} = S_2 = \frac{\tau_i(1-\tau_i)}{\rho_i^2}\left(\xi_i - \frac{1}{p}\gamma_i^2\right)
$$

<u>For $i = 1$</u>

$$
\begin{aligned}
S_1 &= 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}A_1^2\left[\rho_\ell^2\left(1-\delta_{\ell i}\right)\right] - 2\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}\sum_{j=1}^{p}\frac{1}{\lambda_j}\delta_{\ell i}\boldsymbol{\beta}_j^{\mathrm{T}}\mathrm{IF}(\mathbf{x},\mathbf{V}_1,F)\boldsymbol{\beta}_j A_1\left(1-\delta_{i1}\right) \\
&= 2p\sum_{\ell=2}^{k}\frac{\tau_\ell}{\rho_\ell^2}A_1^2\rho_\ell^2 = 2p(1-\tau_1)A_1^2
\end{aligned}
$$

and so using a)

$$
\begin{aligned}
\frac{\partial^2}{\partial\epsilon^2}T_{\mathbf{V},\mathrm{PROP}}(F_{\epsilon,\mathbf{x},1})\Big|_{\epsilon=0} &= S_1 + S_2 = 2p(1-\tau_1)A_1^2 + \tau_1(1-\tau_1)\xi_1 - pA_1^2(1-\tau_1) - pA_1^2(1-\tau_1^2) \\
&= p(1-\tau_1)A_1^2 + \tau_1(1-\tau_1)\xi_1 - pA_1^2(1-\tau_1^2) \\
&= pA_1^2\tau_1^2 - \tau_1 pA_1^2 + \tau_1(1-\tau_1)\xi_1 = \tau_1(1-\tau_1)\left(\xi_1 - \frac{\gamma_1^2}{p}\right)\Box
\end{aligned}
$$

References

[1] Boente, G., Critchley, F and Orellana, L. (2004). Influence functions for robust estimators under proportional scatter matrices. http://www.ic.fcen.uba.ar/preprints/boecriore.pdf

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

[3] Boente, G. and Orellana, L. (2004). Robust Plug–in Estimators in Proportional Scatter Models. *J. Statist. Plann. Inf.*, **122**, 95-110.

[4] 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.

[5] Croux, C. and Haesbroeck, G. (2000). Principal Component Analysis based on Robust estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies. *Biometrika*, **87**, 603–618.

[6] Donoho, D. L. (1982). *Breakdown Properties of Multivariate Location Estimators.* Ph.D. qualifying paper, Harvard University.

[7] Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics,* **7**, 179-188.

[8] Flury, B. K. (1984). Common Principal Components in Groups. *J. Amer. Statist. Assoc.,* **79**, 892-898.

[9] Flury, B. K. (1988). *Common Principal Components and Related Multivariate Models.* New York: John Wiley.

[10] Gervini, D. (2002). The influence function of the Donoho–Stahel estimator of multivariate location and scale. *Statistics and Probability Letters*, **60**, 425-435.

[11] Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[12] Lopuhaä, H. P. (1991). Breakdown Points and Asymptotic Properties of Multivariate S–estimators and $\tau-$estimators: A Summary. In *Directions in Robust Statistics and Diagnostics* (W. Stahel and S. Weisberg, eds.), Part I, 167-182, New York: Springer-Verlag.

[13] Muirhead, R. J. and Waterneaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, **63**, 31-43.

[14] Pires, A. M. and Branco, J. (2002). Partial influence functions. *Journal of Multivariate Analysis*, **83**, 451-468.

[15] Stahel, W. (1981). *Robust estimation: Infinitesimal Optimality and Covariance Matrix Estimation.* Thesis (in German), ETH, Zurich.

[16] Schott, J. R. (1991). Some tests for common principal component subspaces in several groups. *Biometrika,* **78**, 771-777.

[17] Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika,* **69**, 429-436.

[18] Tyler, D. E. (1987). A distribution–free $M-$estimator of multivariate scatter. *Ann. Statist.,* **15**, 234-251.

a) b)

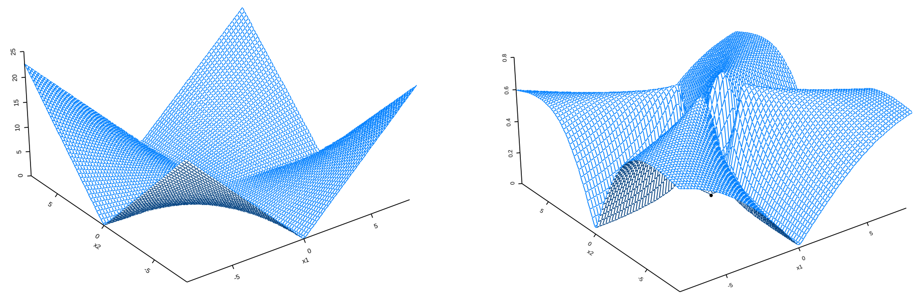

Figure 1: $\text{PIF}_i(\mathbf{x}, S_{\mathbf{V},\text{CPC}}, F)$ at $F = F_1 \times F_2$ with $F_1 = N(\mathbf{0}, \text{diag}(2,1))$ and $F_2 = N(\mathbf{0}, 4\text{diag}(2,1))$. a) Sample Covariance Matrix b) Donoho–Stahel Scatter Matrix

a) b)

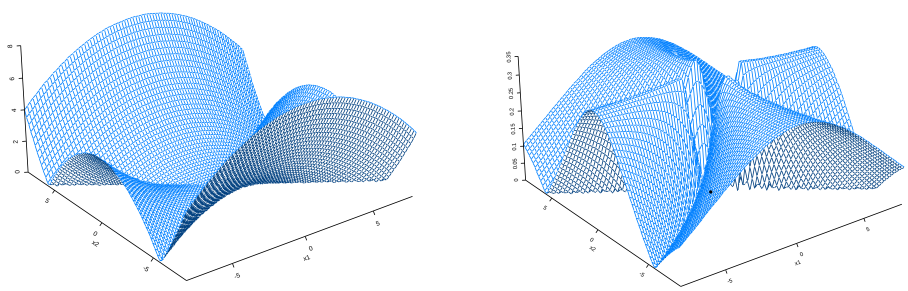

Figure 2: $\text{PIF}_i(\mathbf{x}, R_{\mathbf{V},\text{PROP}}, F)$ at $F = F_1 \times F_2$ with $F_1 = N(\mathbf{0}, \text{diag}(2,1))$ and $F_2 = N(\mathbf{0}, 4\text{diag}(2,1))$. a) Sample Covariance Matrix b) Donoho–Stahel Scatter Matrix

a) b)

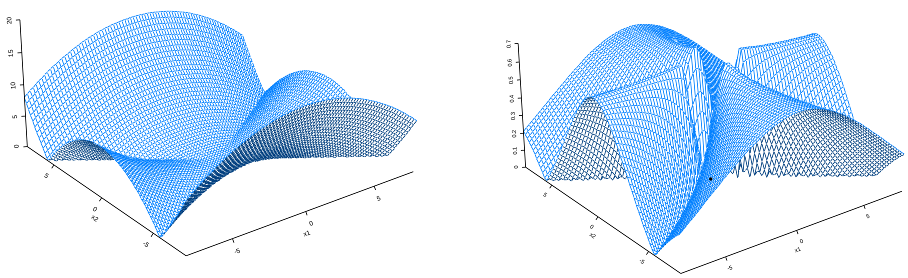

Figure 3: $\text{PIF}_i(\mathbf{x}, S_{\mathbf{V},\text{PROP}}, F)$ at $F = F_1 \times F_2$ with $F_1 = N(\mathbf{0}, \text{diag}(2,1))$ and $F_2 = N(\mathbf{0}, 4\text{diag}(2,1))$. a) Sample Covariance Matrix b) Donoho–Stahel Scatter Matrix

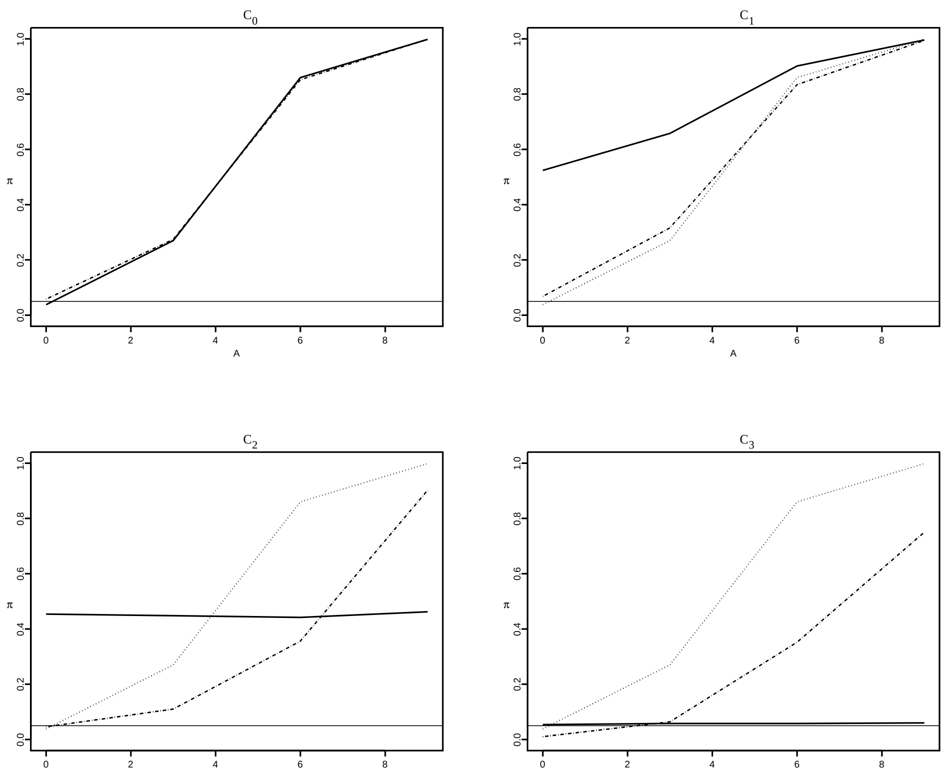

Figure 4: Frequency of rejection (π) of the log–likelihood test for testing model CPC against arbitrary scatter matrices (thick line) and of the robust plug–in test (dashed lines $-\cdot-$) under normal data and under contaminations C_1, C_2 and C_3. The dotted lines correspond to the frequency of rejection of the classical test under C_0. The horizontal line corresponds to the fixed 5% level

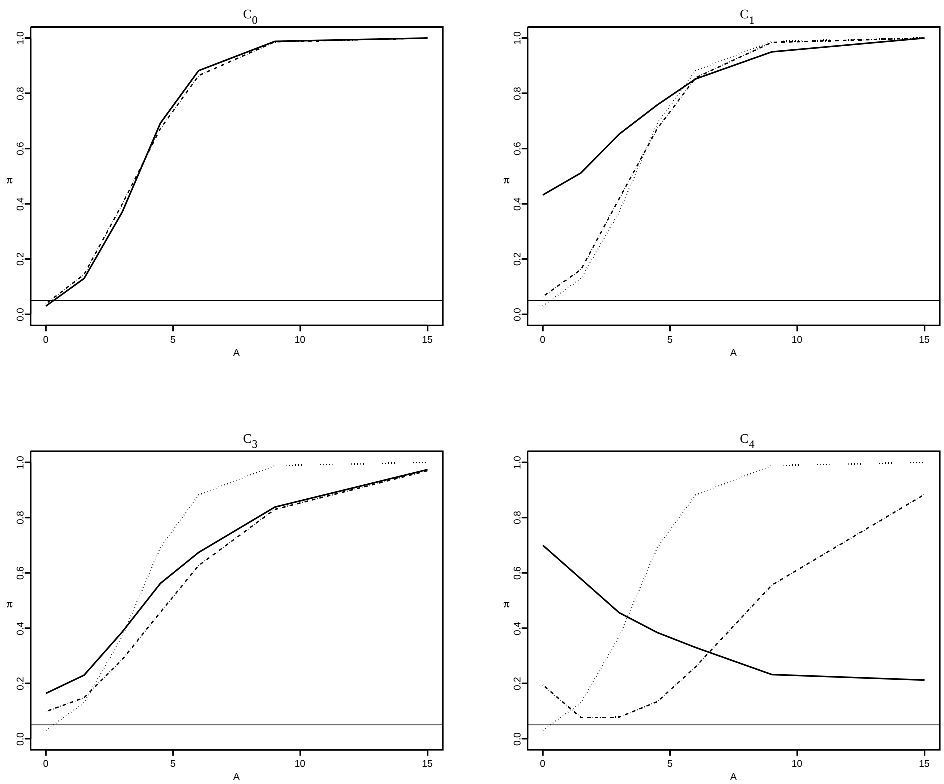

Figure 5: Frequency of rejection (π), for testing proportionality against model CPC, of the classical (thick line) and of the robust (dashed lines $-\cdot-$) Wald test, under normal data and under contaminations C_1, C_3 and C_4. The dotted lines correspond to the frequency of rejection of the classical test under C_0. The horizontal line corresponds to the fixed 5% level

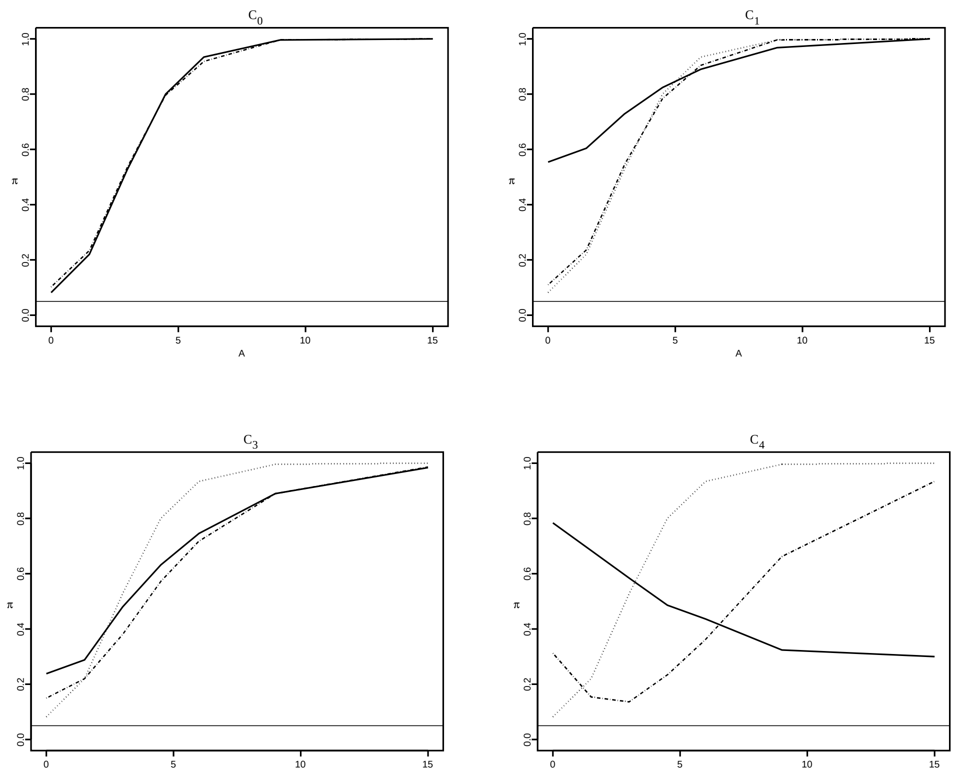

Figure 6: Frequency of rejection (π), for testing proportionality against model CPC, of the classical log–likelihood test for proportionality against CPC (thick line) and of the robust plug–in one (dashed lines $- \cdot -$) under normal data and under contaminations C_1, C_3 and C_4. The dotted lines correspond to the frequency of rejection of the classical test under C_0. The horizontal line corresponds to the fixed 5% level