Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Multivariate Analysis 97 (2006) 124 – 147
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

General projection-pursuit estimators for the common principal components model: influence functions and Monte Carlo study

Graciela Boente[a,*], Ana M. Pires[b], Isabel M. Rodrigues[b]

[a] *Facultad de Ciencias Exactas y Naturales, Departamento de Matemática and Instituto de Cálculo, Universidad de Buenos Aires and CONICET, Ciudad Universitaria, Pabellón 1, Buenos Aires C1428EHA, Argentina*

[b] *Departamento de Matemática, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*

Received 18 August 2002

Available online 20 January 2005

Abstract

The common principal components (CPC) model for several groups of multivariate observations assumes equal principal axes but possibly different variances along these axes among the groups. Under a CPC model, generalized projection-pursuit estimators are defined by using score functions on the dispersion measure considered. Their partial influence functions are obtained and asymptotic variances are derived from them. When the score function is taken equal to the logarithm, it is shown that, under a proportionality model, the eigenvector estimators are optimal in the sense of minimizing the asymptotic variance of the eigenvectors, for a given scale measure.

Keywords: Asymptotic variances; Common principal components; Partial influence function; Projection-pursuit; Robust estimation

* Corresponding author. Tel.: +54 11 45763375; fax: +54 11 45763375.
E-mail addresses: gboente@mate.dm.uba.ar (G. Boente), ana.pires@math.ist.utl.pt (A.M. Pires), isabel.rodrigues@math.ist.utl.pt (I.M. Rodrigues).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2004.11.007

1. Introduction

Several authors, as [10], have studied models for common structure dispersion. As it is well known, those models have been introduced to overcome the problem of an excessive number of parameters, when dealing with several populations, in multivariate analysis. One such basic common structure assumes that the k covariance matrices have possibly different eigenvalues but identical eigenvectors, i.e.,

$$\mathbf{\Sigma}_i = \boldsymbol{\beta}\mathbf{\Lambda}_i\boldsymbol{\beta}', \quad 1 \leqslant i \leqslant k, \tag{1}$$

where $\mathbf{\Lambda}_i$ are diagonal matrices, $\boldsymbol{\beta}$ is the orthogonal matrix of the common eigenvectors and $\mathbf{\Sigma}_i$ is the covariance matrix of the ith population. The more restrictive proportionality model assumes that the scatter matrices are equal up to a proportionality constant, i.e.,

$$\mathbf{\Sigma}_i = \rho_i\mathbf{\Sigma}_1 \quad \text{for } 1 \leqslant i \leqslant k \text{ and } \rho_1 = 1. \tag{2}$$

Model (1) was proposed in [9] and became known as the *common principal components* (CPC) model. The maximum likelihood estimators of $\boldsymbol{\beta}$ and $\mathbf{\Lambda}_i$ are derived in [9], assuming multivariate normality of the original variables. In [10] a unified study of the maximum likelihood estimators under a CPC model and, in particular, under a proportionality model is given.

Let $\left(\mathbf{x}_{ij}\right)_{1 \leqslant j \leqslant n_i, 1 \leqslant i \leqslant k}$ be independent observations from k independent samples in $\mathbb{R}^p$ with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\mathbf{\Sigma}_i$. Let $N = \sum_{i=1}^{k} n_i$, $\tau_i = \frac{n_i}{N}$ and $\mathbf{X}_i = \left(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}\right)$. For the sake of simplicity and without loss of generality, we will also assume that $\boldsymbol{\mu}_i = \mathbf{0}_p$.

It is well known that, in practice, the classical CPC analysis can be affected by the existence of outliers in a sample. In order to obtain robust estimators, in [3,4], an approach based on robust affine equivariant estimators of the covariance matrices $\mathbf{\Sigma}_i$, $1 \leqslant i \leqslant k$ is considered. These authors also studied an approach based on projection-pursuit principles in which, the estimator of $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p\right)$ are the solution of

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_1 &= \underset{\|\mathbf{b}\|=1}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i\, s^2(\mathbf{X}_i'\mathbf{b}),\\ \widehat{\boldsymbol{\beta}}_j &= \underset{\mathbf{b}\in\mathcal{B}_j}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i s^2(\mathbf{X}_i'\mathbf{b}), \quad 2 \leqslant j \leqslant p,\end{aligned} \tag{3}$$

where $\mathcal{B}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}'\widehat{\boldsymbol{\beta}}_m = 0 \text{ for } 1 \leqslant m \leqslant j-1\}$ and s is a univariate scale estimator.

In this paper, we adopt a more general approach which consists of applying a score function to the scale estimator. The paper is organized as follows. In Section 2, we motivate and introduce our proposal while, in Section 3, partial influence functions are computed and asymptotic variances are derived heuristically from them. Also, it is shown that the choice $f(t) = \ln(t)$ leads to an optimal score function since it minimizes the asymptotic variance of the common directions, under proportionality of the scatter matrices. In Section 4, through a simulation study the proposed estimators are compared with those defined through (3) for normal and contaminated samples. All proofs are given in the Appendix.

2. General projection-pursuit estimators

2.1. Definition

In robust principal component analysis, an alternative to using the eigenvalues and eigenvectors of a robust scatter matrix, was first considered in [12] who proposed projection-pursuit estimators maximizing (or minimizing) a robust scale of a one-dimensional projection of the data. A fast algorithm for computing their proposal was developed in [6]. Later, [7] studied the breakdown point of the scatter matrix related to the estimators obtained with this algorithm. These authors also derived influence functions for the projection-pursuit estimators. A rigorous proof of the asymptotic distribution of these estimators has been given recently in [8]. Maximum biases under contaminated models can be found in [13,1].

For the CPC model, the common decomposition given in (1) implies that for any $\mathbf{b} \in \mathbb{R}^p$, and $1 \leqslant i \leqslant k$, $\mathrm{VAR}\left(\mathbf{b}'\mathbf{x}_{i1}\right) = \mathbf{b}'\boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}'\mathbf{b}$. Therefore, the first (or the last) axis could be obtained by maximizing (or minimizing) $\sum_{i=1}^{k} \tau_i \mathrm{VAR}\left(\mathbf{b}'\mathbf{x}_{i1}\right)$ over $\mathbf{b} \in \mathbb{R}^p$ with $\|\mathbf{b}\| = 1$ if the matrix $\sum_{i=1}^{k} \tau_i \boldsymbol{\Sigma}_i$ has no multiple eigenvalues. By considering orthogonal directions to $\boldsymbol{\beta}_1$, the second axis is defined and so on. The eigenvalues for the ith population are clearly $\boldsymbol{\beta}_j'\boldsymbol{\Sigma}_i\boldsymbol{\beta}_j$ and finally, the eigenvectors can be arranged according to a decreasing order of the eigenvalues of the first population. This allows to define the projection-pursuit estimators under a CPC model as in (3).

On the other hand, it is well known (see [10]) that the maximum likelihood estimator of $\boldsymbol{\beta}$ in a normal model minimizes

$$\prod_{i=1}^{k}\left[\frac{\det\{diag\,(\mathbf{F}_i)\}}{\det(\mathbf{F}_i)}\right]^{n_i},$$

where $\mathbf{F}_i = \boldsymbol{\beta}'\mathbf{S}_i\boldsymbol{\beta}$, and $\mathbf{S}_i$ is the sample covariance matrix of the ith population. This is equivalent to minimizing

$$\begin{aligned}
\ln\left(\prod_{i=1}^{k}\left[\frac{\det\{diag\,(\mathbf{F}_i)\}}{\det(\mathbf{F}_i)}\right]^{n_i}\right) &= \sum_{i=1}^{k} n_i\left(\ln\left[\det\{diag\,(\mathbf{F}_i)\}\right] - \ln\{\det(\mathbf{F}_i)\}\right)\\
&= \sum_{i=1}^{k} n_i\left(\ln\left[\det\{diag\,(\mathbf{F}_i)\}\right] - \ln\{\det(\mathbf{S}_i)\}\right)\\
&= \sum_{i=1}^{k} n_i \sum_{j=1}^{p} \ln\left(\ell_{ij}\right) - \ln\{\det(\mathbf{S}_i)\},
\end{aligned}$$

where ℓ_{ij} are the diagonal elements of $\mathbf{F}_i$. Therefore, the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ can be viewed as the minimizer of $\sum_{j=1}^{p}\sum_{i=1}^{k} n_i \ln\left(\ell_{ij}\right)$. By noting that $\widehat{\lambda}_{ij} = \widehat{\boldsymbol{\beta}}_j'\mathbf{S}_i\widehat{\boldsymbol{\beta}}_j$ equals the sample variance of the projected vector, a robust projection–pursuit procedure can be obtained by solving iteratively

$$\widehat{\boldsymbol{\beta}}_p = \underset{\|\mathbf{b}\|=1}{\operatorname{argmin}} \sum_{i=1}^{k} \tau_i \ln\left\{s^2(\mathbf{X}_i'\mathbf{b})\right\},$$

$$\widehat{\boldsymbol{\beta}}_{p-j} = \underset{\mathbf{b}\in\mathcal{A}_j}{\operatorname{argmin}} \sum_{i=1}^{k} \tau_i \ln\left\{s^2(\mathbf{X}_i'\mathbf{b})\right\}, \quad 1\leqslant j\leqslant p-1,$$

where $\mathcal{A}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}'\widehat{\boldsymbol{\beta}}_{p-m} = 0 \text{ for } 0\leqslant m\leqslant j-1\}$. Another consistent solution can be obtained by maximizing instead of minimizing in the above definition, just as in the one population case (see [12]).

A more general framework considers a general increasing score function $f : \mathbb{R}^+ \to \mathbb{R}$ and thus, we propose to estimate the common directions as

$$\begin{cases} \widehat{\boldsymbol{\beta}}_1 = \underset{\|\mathbf{b}\|=1}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i\, f\left\{s^2(\mathbf{X}_i'\mathbf{b})\right\}, \\ \widehat{\boldsymbol{\beta}}_j = \underset{\mathbf{b}\in\mathcal{B}_j}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i\, f\left\{s^2(\mathbf{X}_i'\mathbf{b})\right\}, \quad 2\leqslant j\leqslant p. \end{cases} \tag{4}$$

The estimators of the eigenvalues of the ith population are then computed as $\widehat{\lambda}_{ij} = s^2(\mathbf{X}_i'\widehat{\boldsymbol{\beta}}_j)$ for $1\leqslant j\leqslant p$. A different definition arises by taking minimum instead of maximum, which lead to different solutions (beyond the order) due to the use of a robust scale. However, both will have the same partial influence functions and thus, the same asymptotic variances.

2.2. Notation and assumptions

From now on, $\mathbf{x}_i$ will denote independent vectors such that $\mathbf{x}_i \sim F_i$, where F_i has location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i = \mathbf{C}_i\mathbf{C}_i'$ satisfying (1). As in [4], without loss of generality, we will assume that $\boldsymbol{\mu}_i = \mathbf{0}$. Denote by $F_i[\mathbf{b}]$ the distribution of $\mathbf{b}'\mathbf{x}_i$ and by F the product measure, $F = F_1 \times \cdots \times F_k$.

Let $\varsigma(\mathbf{b}) = \sum_{i=1}^{k} \tau_i\, f\{\sigma^2\,(F_i[\mathbf{b}])\}$ where $\sigma(\cdot)$ is a univariate scale functional related to the scale estimator s, which is assumed to be equivariant under scale transformations.

Throughout this paper we will consider the following set of assumptions:

A1. F_i is an ellipsoidal distribution with location parameter $\boldsymbol{\mu}_i = \mathbf{0}_p$ and scatter matrix $\boldsymbol{\Sigma}_i = \mathbf{C}_i\mathbf{C}_i'$ satisfying (1). Moreover, when $\mathbf{x}_i \sim F_i$, $\mathbf{C}_i^{-1}\mathbf{x}_i = \mathbf{z}_i$ has the same spherical distribution G for all $1\leqslant i\leqslant k$.

A2. $\sigma(\cdot)$ is a robust scale functional, equivariant under scale transformations, such that $\sigma(G_0) = 1$, with G_0 the distribution of z_{11}.

A3. The function $(\varepsilon, y) \rightarrow \sigma\left((1-\varepsilon)G_0 + \varepsilon\Delta_y\right)$ is twice continuously differentiable in $\{(0, y), y \in \mathbb{R}\}$ where Δ_y denotes the point mass at y.
A4. f is a twice continuously differentiable function.
A5. $n_i = \tau_i N$, $\sum_{i=1}^{k} \tau_i = 1$, with $0 < \tau_i < 1$ fixed numbers.

2.3. Fisher-consistency

In this Section, we provide conditions under which the common directions projection-pursuit estimates will be Fisher-consistent, that is, they will estimate the right quantities at the idealized model (see, [11]). For this purpose let us introduce the statistical functionals corresponding to (4). The common directions projection-pursuit functional $\boldsymbol{\beta}_\sigma(F) = (\boldsymbol{\beta}_{\sigma,1}(F), \ldots, \boldsymbol{\beta}_{\sigma,p}(F))$ is defined as the solution of

$$\begin{cases} \boldsymbol{\beta}_{\sigma,1}(F) = \underset{\|\mathbf{b}\|=1}{\operatorname{argmax}}\ \varsigma(\mathbf{b}), \\[2ex] \boldsymbol{\beta}_{\sigma,j}(F) = \underset{\mathbf{b}\in\mathcal{B}_j}{\operatorname{argmax}}\ \varsigma(\mathbf{b}), \quad 2 \leqslant j \leqslant p, \end{cases} \tag{5}$$

where $\mathcal{B}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}'\boldsymbol{\beta}_m(F) = 0 \text{ for } 1 \leqslant m \leqslant j-1\}$. As in [12] an alternative, though less popular, is based on stepwise minimization, defined as

$$\begin{cases} \boldsymbol{\beta}_{\sigma,p}(F) \quad = \underset{\|\mathbf{b}\|=1}{\operatorname{argmin}}\, \varsigma(\mathbf{b}), \\[2ex] \boldsymbol{\beta}_{\sigma,p-j}(F) = \underset{\mathbf{b}\in\mathcal{A}_j}{\operatorname{argmin}}\, \varsigma(\mathbf{b}), \quad 1 \leqslant j \leqslant p-1, \end{cases} \tag{6}$$

where $\mathcal{A}_j = \{\mathbf{b} : \|\mathbf{b}\| = 1, \mathbf{b}'\boldsymbol{\beta}_{\sigma,p-m} = 0 \text{ for } 0 \leqslant m \leqslant j-1\}$. In both cases the eigenvalues and the covariance matrix functionals are defined as

$$\lambda_{\sigma,ij}(F) = \sigma^2\left(F_i\left[\boldsymbol{\beta}_{\sigma,j}(F)\right]\right), \tag{7}$$

$$\mathbf{V}_{\sigma,i}(F) = \sum_{j=1}^{p} \lambda_{\sigma,ij}(F)\boldsymbol{\beta}_{\sigma,j}(F)\boldsymbol{\beta}_{\sigma,j}(F)'. \tag{8}$$

When $f(t) = \mathrm{id}(t) = t$, conditions under which the functional defined through (5) will be Fisher-consistent were obtained in [3]. The following Proposition, which is easily derived from the optimality properties of the eigenvectors of a symmetric positive definite matrix, establishes the conditions under which the functionals (5), (6) and (7) are Fisher-consistent for a general f.

Proposition 1. *Let* $\mathbf{x}_i$ *be independent random vectors with distribution* F_i *satisfying* A1 *and* $\sigma(\cdot)$ *a robust scale functional satisfying* A2. *Denote by* $\boldsymbol{\beta}_\sigma(F) = (\boldsymbol{\beta}_{\sigma,1}(F), \ldots, \boldsymbol{\beta}_{\sigma,p}(F))$

the solution of (5) *and let* $\lambda_{\sigma,ij}(F) = \sigma^2\left(F_i[\boldsymbol{\beta}_{\sigma,j}(F)]\right)$. *Then, we have that*

(a) *If* $\boldsymbol{\Lambda}_i = diag\left(\lambda_{i1}, \ldots, \lambda_{ip}\right)$ *are such that* $\lambda_{i1} \geqslant \cdots \geqslant \lambda_{ip}$ *for* $1 \leqslant i \leqslant k$ *and, for any* $1 \leqslant j \leqslant p-1$, *there exists an* $i_0 = i_0(j)$ *such that* $\lambda_{i_0 j} > \lambda_{i_0\,(j+1)}$, *the functionals* $\boldsymbol{\beta}_\sigma(F)$, *defined either by* (5) *or* (6), *and* $\lambda_{\sigma,ij}(F)$ *are Fisher consistent for any strictly increasing function* $f(t)$.

(b) *Denote* $\boldsymbol{\Sigma}_f = \boldsymbol{\beta}' \Lambda_f \boldsymbol{\beta}$ *where* $\Lambda_f = diag\left(\xi_1, \ldots, \xi_p\right)$ *with* $\xi_j = \sum_{i=1}^k \tau_i f\left(\lambda_{ij}\right)$ *and* $\xi_1 \geqslant \cdots \geqslant \xi_p$. *If* f *is strictly increasing and strictly convex* (*respectively*, *strictly concave*) *and* $\xi_1, \ldots \xi_p$ *are pairwise different or* $\forall j \neq m\ \exists i_0 : \lambda_{i_0 j} \neq \lambda_{i_0 m}$, *then* $\boldsymbol{\beta}$ *is the unique solution of the system defined by* (5) (*respectively*, *by* (6)) *and so*, *the functionals* $\boldsymbol{\beta}_\sigma(F)$ *and* $\lambda_{\sigma,ij}(F)$ *are Fisher-consistent.*

Remark 1. The case $f(t) = \ln(t)$, closely related to the maximum likelihood approach and for which some optimality properties are derived later (Propositions 2 and 3) is included in (b). Note that $\boldsymbol{\Sigma}_{\ln} = \ln\left(\prod_{i=1}^k \boldsymbol{\Sigma}_i^{\tau_i}\right) = \boldsymbol{\beta}' diag\left(\ln\left(\prod_{i=1}^k \lambda_{i1}^{\tau_i}\right), \ldots, \ln\left(\prod_{i=1}^k \lambda_{ip}^{\tau_i}\right)\right)\boldsymbol{\beta}$.

3. Influence functions and asymptotic variances

Partial influence functions were introduced by [14] in order to ensure that the usual properties of the influence function for the one-population case are satisfied when dealing with several populations. Denote by F the product measure, $F = F_1 \times \cdots \times F_k$. Partial influence functions of a functional $T(F)$ are then defined as

$$PIF_i(\mathbf{x}, T, F) = \lim_{\varepsilon \to 0} \frac{T(F_{\varepsilon,\mathbf{x},i}) - T(F)}{\varepsilon},$$

where $F_{\varepsilon,\mathbf{x},i} = F_1 \times \cdots \times F_{i-1} \times F_{i,\varepsilon,\mathbf{x}} \times F_{i+1} \times \cdots F_k$, with $F_{i,\varepsilon,\mathbf{x}} = (1-\varepsilon)F_i + \varepsilon \Delta_{\mathbf{x}}$, and $\Delta_{\mathbf{x}}$ the point mass at $\mathbf{x}$.

With the general results in [14] one can show that the following expansion holds:

$$N^{\frac{1}{2}}\{T(F_N) - T(F)\} = \sum_{i=1}^k \frac{1}{(\tau_i\, n_i)^{\frac{1}{2}}} \sum_{j=1}^{n_i} PIF_i\left(\mathbf{x}_{ij}, T, F\right) + o_p(1),$$

where F_N denotes the empirical distribution of the k independent samples $\mathbf{x}_{ij}$, $1 \leqslant j \leqslant n_i$, $1 \leqslant i \leqslant k$. Therefore, the asymptotic variance of the estimates, i.e., the variance of the approximating normal distribution, can be evaluated as

$$ASVAR(T, F) = \sum_{i=1}^k {\tau_i}^{-1} E_{F_i}\left\{PIF_i(\mathbf{x}_{i1}, T, F)\, PIF_i(\mathbf{x}_{i1}, T, F)'\right\}. \tag{9}$$

The following Theorem gives the partial influence functions for the general projection-pursuit functionals.

Theorem 1. *Let* $\mathbf{x}_i$ *be independent random vectors with distribution* F_i *and* $\sigma(\cdot)$ *a robust scale functional. Then, under* A1 *to* A4, *we have that for any* $\mathbf{x}$, *the partial influence functions of the functionals defined through* (5), (7) *and* (8) *are given by*

$$PIF_i(\mathbf{x}, \lambda_{\sigma,\ell j}, F) = 2\ \delta_{\ell i}\ \lambda_{ij}\ IF\left(\frac{\mathbf{x}'\boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}}, \sigma, G_0\right), \tag{10}$$

$$\begin{aligned} PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = {} & \tau_i\ \boldsymbol{\beta}_j'\mathbf{x} \sum_{s=1}^{j-1} \frac{1}{v_{sj} - v_s} \sqrt{\lambda_{is}} f'(\lambda_{is})\ DIF\left(\frac{\boldsymbol{\beta}_s'\mathbf{x}}{\sqrt{\lambda_{is}}}, \sigma, G_0\right)\ \boldsymbol{\beta}_s \\ & + \tau_i \sqrt{\lambda_{ij}} f'\left(\lambda_{ij}\right) DIF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma, G_0\right) \\ & \times \sum_{s=j+1}^{p} \frac{1}{v_j - v_{js}} \boldsymbol{\beta}_s'\mathbf{x}\ \boldsymbol{\beta}_s, \end{aligned} \tag{11}$$

$$\begin{aligned} PIF_i(\mathbf{x}, \mathbf{V}_{\sigma,\ell}, F) = {} & \delta_{\ell i}\ 2 \sum_{j=1}^{p} \lambda_{ij} IF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}_i}{\sqrt{\lambda_{ij}}}, \sigma, G_0\right) \boldsymbol{\beta}_j \boldsymbol{\beta}_j' \\ & + \sum_{j=2}^{p} \sum_{s=1}^{j-1} \frac{\lambda_{\ell j} - \lambda_{\ell s}}{v_{sj} - v_s} \tau_i \sqrt{\lambda_{is}}\ f'(\lambda_{is}) \\ & \times DIF\left(\frac{\mathbf{x}'\boldsymbol{\beta}_s}{\sqrt{\lambda_{is}}}, \sigma, G_0\right) \left(\mathbf{x}'\boldsymbol{\beta}_j\right) \left(\boldsymbol{\beta}_s\boldsymbol{\beta}_j' + \boldsymbol{\beta}_j\boldsymbol{\beta}_s'\right), \end{aligned}$$

where $DIF(y, \sigma, G)$ *denotes the derivative of the influence function,* $IF(y, \sigma, G)$, *of the scale functional* σ *with respect to* y, $v_{js} = \sum_{i=1}^{k} \tau_i f'\left(\lambda_{ij}\right) \lambda_{is}$, $v_j = v_{jj}$ *and* $v_{js} \neq v_{jj}$ *for* $s \neq j$.

Remark 2. Under a proportional model $v_{js} = \lambda_s \sum_{i=1}^{k} \tau_i f'\left(\lambda_{ij}\right) \rho_i$. Therefore, $v_{js} \neq v_{jj}$ for $s \neq j$ if and only if the eigenvalues of the first population are different, which is a usual assumption in order to identify the common directions. It is worthwhile noticing that the partial influence functions of the functionals defined through (6) are also given by (11).

Remark 3. As for the projection-pursuit estimators considered in [4], one notices that by using a scale estimator with bounded influence, the eigenvalues will have bounded influence, and its influence function does not depend on the score function f considered. However, the influence function for the eigenvectors may be unbounded, since the term $\mathbf{x}'\boldsymbol{\beta}_j$ will still remain unbounded.

The asymptotic variance of the projection-pursuit estimators of the common eigenvectors and of the eigenvalues can be obtained heuristically using (9).

Corollary 1. *Let $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}$, $1 \leqslant i \leqslant k$, be independent observations from k independent samples with distribution F_i and $\sigma(\cdot)$ a robust scale functional. Assume that* A1 *to* A5 *hold and that $\boldsymbol{\beta} = \mathbf{I}_p$, i.e., $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i$.*

Let $s(\cdot)$ be the univariate scale statistic related to σ and $\mathbf{X}_i = \left(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}\right)$, for $1 \leqslant i \leqslant k$. Define the common principal axes as the solution of (4) *and the estimators of the eigenvalues and of the covariance matrix of the ith population as $\widehat{\lambda}_{ij} = s^2(\mathbf{X}_i'\widehat{\boldsymbol{\beta}}_j)$, for $1 \leqslant j \leqslant p$ and $\mathbf{V}_i = \sum_{j=1}^p \widehat{\lambda}_{ij}\widehat{\boldsymbol{\beta}}_j\widehat{\boldsymbol{\beta}}_j'$.*

Then, the asymptotic variances of the general projection-pursuit estimators are given by

$$ASVAR\left(\widehat{\lambda}_{ij}\right) = 4\lambda_{ij}^2 \frac{1}{\tau_i} ASVAR\left(s, G_0\right),$$

$$\begin{aligned} ASVAR\left(\widehat{\boldsymbol{\beta}}_{jm}\right) &= \sum_{i=1}^k \tau_i \lambda_{ij}\lambda_{im}\left\{\delta_{m>j}\frac{\left\{f'\left(\lambda_{ij}\right)\right\}^2}{\left(v_j - v_{jm}\right)^2} + \delta_{m<j}\frac{\left\{f'\left(\lambda_{im}\right)\right\}^2}{\left(v_{mj} - v_m\right)^2}\right\} \\ &\quad \times E_G\left\{DIF\left(z_{1j}, \sigma, G_0\right) z_{1m}\right\}^2. \end{aligned} \tag{12}$$

In particular, when $G = N\left(\mathbf{0}, \mathbf{I}_p\right)$, if Φ denotes the standard normal distribution function, we have that $ASVAR\left(\widehat{\lambda}_{ij}\right) = 4\lambda_{ij}^2\frac{1}{\tau_i}ASVAR\left(s, \Phi\right)$, for $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant p$, $ASCOV\left(\widehat{\boldsymbol{\beta}}_{jm}, \widehat{\boldsymbol{\beta}}_{jr}\right) = 0$, for $m \neq j$, $m \neq r$ and $r \neq j$ and that

$$\begin{aligned} ASVAR\left(\widehat{\boldsymbol{\beta}}_{jm}\right) &= \sum_{i=1}^k \tau_i \lambda_{ij}\lambda_{im}\left\{\delta_{m>j}\frac{\left\{f'\left(\lambda_{ij}\right)\right\}^2}{\left(v_j - v_{jm}\right)^2} + \delta_{m<j}\frac{\left\{f'\left(\lambda_{im}\right)\right\}^2}{\left(v_{mj} - v_m\right)^2}\right\} \\ &\quad \times E_\Phi\left[DIF\left(Y, \sigma, \Phi\right)\right]^2 \text{ for } m \neq j, \end{aligned}$$

where $v_{js} = \sum_{i=1}^k \tau_i f'\left(\lambda_{ij}\right)\lambda_{is}$, $v_j = v_{jj}$ and $v_{jm} \neq v_{jj}$ for $m \neq j$.

Remark 4. Note that when all the populations have the same scatter matrices, the asymptotic variance of the projection–pursuit estimators of the eigenvectors does not depend on the score function f. From Remark 2, it follows that the asymptotic variance of the estimators defined by minimizing instead of maximizing $\varsigma(\mathbf{b})$, is also given by (12).

The following Proposition states the optimality of $f(t) = \ln(t)$, under a proportional model, but when equality of all the scatter matrices does not hold.

Proposition 2. *Let $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}$, $1 \leqslant i \leqslant k$, be independent observations from k independent samples with distribution F_i and $\sigma(\cdot)$ a robust scale functional. Assume that* A1 *to* A3 *and* A5 *hold and that $\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i = \rho_i \boldsymbol{\Lambda}_1$, i.e., $\boldsymbol{\Sigma}_i$ satisfies* (2) *with $\boldsymbol{\beta} = \mathbf{I}_p$, $\rho_1 = 1$. Let $\boldsymbol{\Lambda}_1 = diag\left(\lambda_1, \ldots, \lambda_p\right)$ and assume that $\lambda_1 > \cdots > \lambda_p$.*

Let $s(\cdot)$ be the fixed univariate scale statistic related to the scale functional $\sigma(G)$ and $\mathcal{F}$ the class of all strictly increasing score functions f satisfying A4. *Then,*

(i) $f(t) = \ln(t)$ *minimizes the asymptotic variance given by* (12) *of the general projection–pursuit estimators in the class $\mathcal{F}$.*
(ii) *Moreover, if at least one $\rho_i \neq 1$, and $f \in \mathcal{F}$ minimizes* (12) *for arbitrary $F = F_1 \times \cdots \times F_k$ as above, then there exists $a \neq 0$ and $b \in \mathbb{R}$ such that $f(t) = a\ln(t) + b$.*

Remark 5. Under a proportional model, once the principal directions have been derived, projection-pursuit functionals for the proportionality constant and for the eigenvalues of the first population can be obtained as in [2]. As in the CPC model, these eigenvalue functionals will have partial influence functions which will be independent of the choice of the function f to be considered. As mentioned above, Proposition 2 entails the optimality of $f(t) = \ln(t)$, since we can always assume $a = 1$ and $b = 0$ and so, we get eigenvector estimators with the lowest variances, given a scale $\sigma(G)$, while the estimators for the eigenvalues and proportionality constants will behave as those with the identity score function. Note that this optimality property extends the well known result for the maximum likelihood estimators, which minimize the asymptotic variance of any asymptotically normally distributed estimator and, in particular, of those defined through (4) with s^2 the sample variance, to estimates defined through a general scale function satisfying A2.

Moreover, under the proportionality model and for normally distributed data, the asymptotic variance of the mth element of the maximum likelihood estimator of the common direction $\boldsymbol{\beta}_j$ is given by $\frac{\lambda_m \lambda_j}{(\lambda_m - \lambda_j)^2}$. Therefore, when considering the score function $f(t) = \ln(t)$, the relative efficiency of the projection-pursuit estimator is given by $E_{\Phi}\,[DIF\,(Y, \sigma, \Phi)]^2$, and thus, it does not depend on the proportionality constants as it does when $f(t) = t$. It is worthwhile noticing that, in this case, when $\sigma^2(G) = var(G)$ and $f(t) = \ln(t)$, the asymptotic variance of the projection-pursuit estimator of $\boldsymbol{\beta}_j$ equals that of the maximum likelihood estimator.

The following Proposition states a partial result regarding the optimality of $f(t) = \ln(t)$, under a CPC model where the scatter matrices are not proportional but when their eigenvalues preserve the order among populations. However, the class of objective functions need to be restricted in order to obtain the optimality. We have considered the well-known Box and Cox class of functions.

Proposition 3. *Let $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}$, $1 \leqslant i \leqslant k$, be independent observations from k independent samples with distribution F_i and $\sigma(\cdot)$ a robust scale functional. Assume that* A1 *to* A3 *and* A5 *hold with $\boldsymbol{\beta} = \mathbf{I}_p$. Let $\boldsymbol{\Lambda}_i = diag\left(\lambda_{i1}, \ldots, \lambda_{ip}\right)$ and $s(\cdot)$ be the fixed univariate scale statistic related to the scale functional $\sigma(G)$.*

Denote by $\mathcal{F} = \{f(t) : f : \mathbb{R}^+ \to \mathbb{R},\ f'(t) = t^{\alpha-1}, \alpha \geqslant 0\}$. Let $1 \leqslant j \leqslant p$ be fixed, and denote by $a_{i,mj} = \frac{\lambda_{im}}{\lambda_{ij}}$.

(i) *Given $m > j$, assume that $\lambda_{ij} \geqslant \lambda_{im}$ for $1 \leqslant i \leqslant k$ and there exists i_0 such that $\lambda_{i_0 j} > \lambda_{i_0 m}$, if $\lambda_{1j} \leqslant \cdots \leqslant \lambda_{kj}$ and $a_{1,mj} \leqslant \cdots \leqslant a_{k,mj}$, then $f(t) = \ln(t)$ minimizes the* *asymptotic variance* (12) *of* $\widehat{\boldsymbol{\beta}}_{jm}$ *in the class* $\mathcal{F}$. *Moreover if there exists,* $\ell < i$ *such that* $\lambda_{\ell j} < \lambda_{ij}$ *and* $a_{\ell, mj} < a_{i, mj}$, *then* $f(t) = \ln(t)$ *is the unique minimizer within* $\mathcal{F}$.

(ii) *Given* $m < j$, *assume that* $\lambda_{ij} \leqslant \lambda_{im}$ *for* $1 \leqslant i \leqslant k$ *and there exists* i_0 *such that* $\lambda_{i_0 j} < \lambda_{i_0 m}$, *if* $\lambda_{1m} \leqslant \cdots \leqslant \lambda_{km}$ *and* $a_{1, jm} \leqslant \cdots \leqslant a_{k, jm}$, *then* $f(t) = \ln(t)$ *minimizes the asymptotic variance* (12) *of* $\widehat{\boldsymbol{\beta}}_{jm}$ *in the class* $\mathcal{F}$. *Moreover if there exists,* $\ell < i$ *such that* $\lambda_{\ell m} < \lambda_{im}$ *and* $a_{\ell, jm} < a_{i, jm}$, *then* $f(t) = \ln(t)$ *is the unique minimizer within* $\mathcal{F}$.

Remark 6. A similar result holds, for instance, for $m > j$, if we assume that $\lambda_{ij} \geqslant \lambda_{im}$ for $1 \leqslant i \leqslant k$ and there exists i_0 such that $\lambda_{i_0 j} \geqslant \lambda_{i_0 m}$, and if $\lambda_{1j} \geqslant \cdots \geqslant \lambda_{kj}$ and $a_{1, mj} \geqslant \cdots \geqslant a_{k, mj}$.

4. Monte Carlo study

We performed a simulation study in dimension 4. A more extensive Monte Carlo study including a simulation study in dimension 2 can be found in [15].

We evaluated the estimators defined in Section 2 with $f(t) = \ln(t)$ and $f(t) = t$. In all Figures, LPP_1 and LPP_2 will denote the estimates corresponding to the first choice, while PP_1 and PP_2 correspond to those related to the second one. The index 1 indicates that the scale σ considered is the MAD (median of the absolute deviations from the median) while the index 2 is for the M-scale estimator with score function $\chi(t) = \min\left(\frac{t^2}{c^2}, 1\right) - \frac{1}{2}$ and $c = 1.041$, which gives a scale estimator with breakdown point $\frac{1}{2}$ and efficiency 0.509.

Two procedures were considered to compute the projection-pursuit estimates. They will be denoted DD and RD in Figures. The first one corresponds to the procedure considered in [3] which adapts the proposal given in [6] for the one-population case. By RD we will denote the following *random direction* procedure for searching the common principal directions:

(i) We first generate 1000 random directions $\mathbf{z}_i / \|\mathbf{z}_i\|$, $\mathbf{z}_i \sim N\left(\mathbf{0}_p, \mathbf{I}_p\right)$

(ii) Let $\mathbf{P}_0 = \mathbf{I}_p$ and denote $\mathbf{P}_q$ the projection matrix over the linear space orthogonal to the already computed q common principal directions $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_q$ for $1 \leqslant q \leqslant p-1$. Following the same steps as in [6], for $0 \leqslant q \leqslant p-1$, we search the common principal direction $\widehat{\boldsymbol{\beta}}_{q+1}$ among $\frac{\mathbf{P}_q\left(\mathbf{z}_i - \hat{\boldsymbol{\mu}}\right)}{\left\|\mathbf{P}_q\left(\mathbf{z}_i - \hat{\boldsymbol{\mu}}\right)\right\|}$, $1 \leqslant i \leqslant 1000$, where $\hat{\boldsymbol{\mu}}$ is the Donoho–Stahel location estimator of $(\mathbf{z}_1, \ldots, \mathbf{z}_{1000})$.

This procedure can be helpful when the number of observations in each sample is small and thus the procedure described in [3] does not provide a good algorithm to search for the projection-pursuit directions.

We have considered the following three models

- **Model 1.** $k = 2$ populations where $\boldsymbol{\Sigma}_1 = diag(4, 3, 2, 1)$ and $\boldsymbol{\Sigma}_2 = diag(6, 8, 2.5, 5)$. Then, $\boldsymbol{\Sigma}_{\mathrm{id}} = \sum_{i=1}^{k} \tau_i \boldsymbol{\Sigma}_i = diag(5, 5.5, 2.25, 3)$ while $\boldsymbol{\Sigma}_{\ln} = \ln\left(diag(\sqrt{24}, \sqrt{24}, \sqrt{5}, \sqrt{5})\right)$.

- **Model 2.** $k = 2$ populations with proportional covariance matrices $\mathbf{\Sigma}_1 = diag(4, 3, 2, 1)$ and $\mathbf{\Sigma}_2 = 4\,\mathbf{\Sigma}_1$. Therefore, $\mathbf{\Sigma}_{\text{id}} = diag(10, 7.5, 5, 2.5)$ and $\mathbf{\Sigma}_{\ln} = \ln(diag(8, 6, 4, 2))$.
- **Model 3.** $k = 3$ populations with covariance matrices $\mathbf{\Sigma}_1 = diag(8, 4, 2, 1)$, $\mathbf{\Sigma}_2 = diag(16, 12, 8, 4)$ and $\mathbf{\Sigma}_3 = 5\ \mathbf{\Sigma}_1$. Thus, $\mathbf{\Sigma}_{\text{id}} = diag(\frac{64}{3}, 12, \frac{20}{3}, \frac{10}{3})$ while $\mathbf{\Sigma}_{\ln} = \frac{1}{3}\ln(diag(5120, 960, 160, 20))$.

These models have been selected since, Model 1 does not preserve the order of the eigenvalues among populations and has an identification problem of the directions when using $f(t) = \ln(t)$. More precisely, any basis of the subspace generated by $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, could be considered by this procedure as good as the two first canonical vectors. Model 2 is a proportional model where the optimality of $f(t) = \ln(t)$ holds and our aim is to illustrate the results of Proposition 2. On the other hand, in Model 3, neither proportionality among the three populations nor the assumptions of Proposition 3 hold. However, straightforward calculations show that the general projection-pursuit estimates using $f(t) = \ln(t)$ have smaller asymptotic variance than those using $f(t) = t$. Moreover, for instance, for $m = 3, 4$, the asymptotic variance $ASVAR\left(\widehat{\boldsymbol{\beta}}_{1m}\right)$ given by (12) is minimized when $f(t) = \ln(t)$ over the class of functions $\mathcal{F} = \{f(t) : f : \mathbb{R}^+ \to \mathbb{R}, f'(t) = t^{\alpha-1}, \alpha \geqslant 0\}$, considered in Proposition 3.

In all models, we performed 1000 replications generating k independent samples of size $n_i = n = 100$. The true common principal axes are thus the original $\mathbf{x}$-axes given by the unit vectors $\mathbf{e}_j$. The eigenvectors were ordered according to a decreasing order of the eigenvalues of the first population and so, $\boldsymbol{\beta}_j = \mathbf{e}_j$.

The results for normal data sets will be indicated by C_0, while $C_{1,\varepsilon}$ denote the following contamination: $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in}$ are i.i.d. $(1 - \varepsilon)N(\mathbf{0}, \mathbf{\Sigma}_i) + \varepsilon N(\boldsymbol{\mu}, \mathbf{\Sigma}_i)$ with $\boldsymbol{\mu} = 10\,\mathbf{e}_4 = (0, 0, 0, 10)'$. We present the results for $\varepsilon = 0.05$ and $\varepsilon = 0.10$. This case corresponds to contaminating both populations in the direction of the smallest eigenvalue of the first population. The aim is to study changes in the estimation of the principal directions.

For simplicity, we report only the results corresponding to the common eigenvectors and to the eigenvalues of the first population. Moreover, we present only a subset of the results in a graphical way. More detailed results are available in [15].

Figs. 1 and 2 give the boxplots of $\log\left(\frac{\widehat{\lambda}_{1j}}{\lambda_{1j}}\right)$ where $\widehat{\lambda}_{1j}$ denote the eigenvalue estimates of $\mathbf{\Sigma}_1$ for Models 1 and 2. The results for Model 3 are similar and thus, they are not reported here but are described in [15].

The best performance for estimating the eigenvalues under contamination is obtained by the estimates based on the M-scale, see Figs. 1 and 2. Contrary to our expectations, there is not a great improvement when we use 1000 random directions instead of the normalized data. An alternative procedure could be to combine both the random generated directions and the directions given by the sample data at each replication. Except for the smallest eigenvalue, in most cases the estimates based on $f(t) = t$ have similar or smaller dispersion than those based on $f(t) = \ln(t)$. The poor efficiency of the eigenvalue estimates obtained using both projection-pursuit procedures is related to the low efficiency of the MAD and of the M-scale

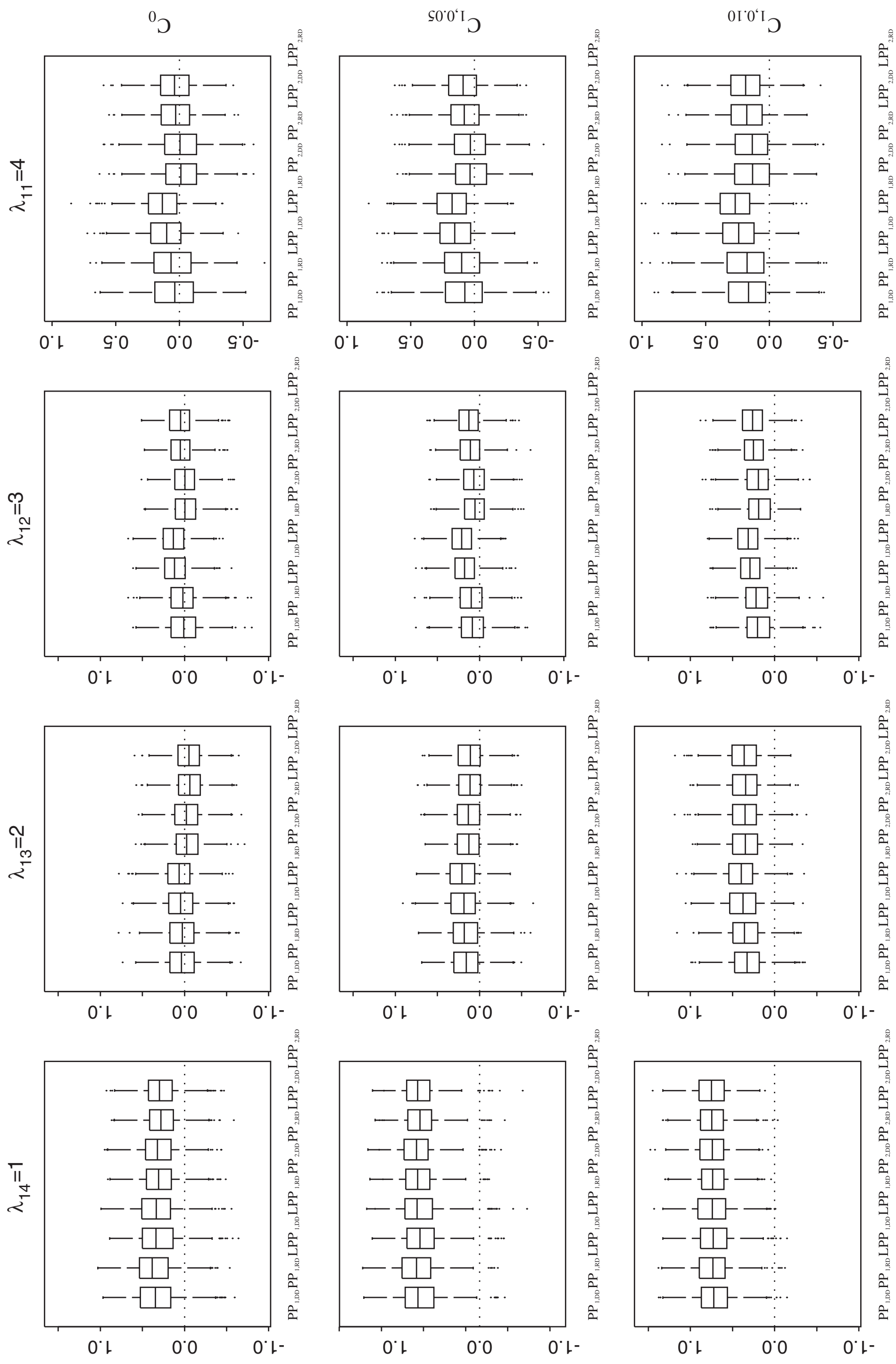

Fig. 1. Boxplots of $\log\left(\frac{\widehat{\lambda}_{1j}}{\lambda_{1j}}\right)$ for Model 1. The dotted lines correspond to 0. The boxplots correspond in order to $PP_{1,\mathrm{DD}}$, $PP_{1,\mathrm{RD}}$, $LPP_{1,\mathrm{DD}}$, $LPP_{1,\mathrm{RD}}$, $PP_{2,\mathrm{DD}}$, $PP_{2,\mathrm{RD}}$, $LPP_{2,\mathrm{DD}}$ and $LPP_{2,\mathrm{RD}}$.

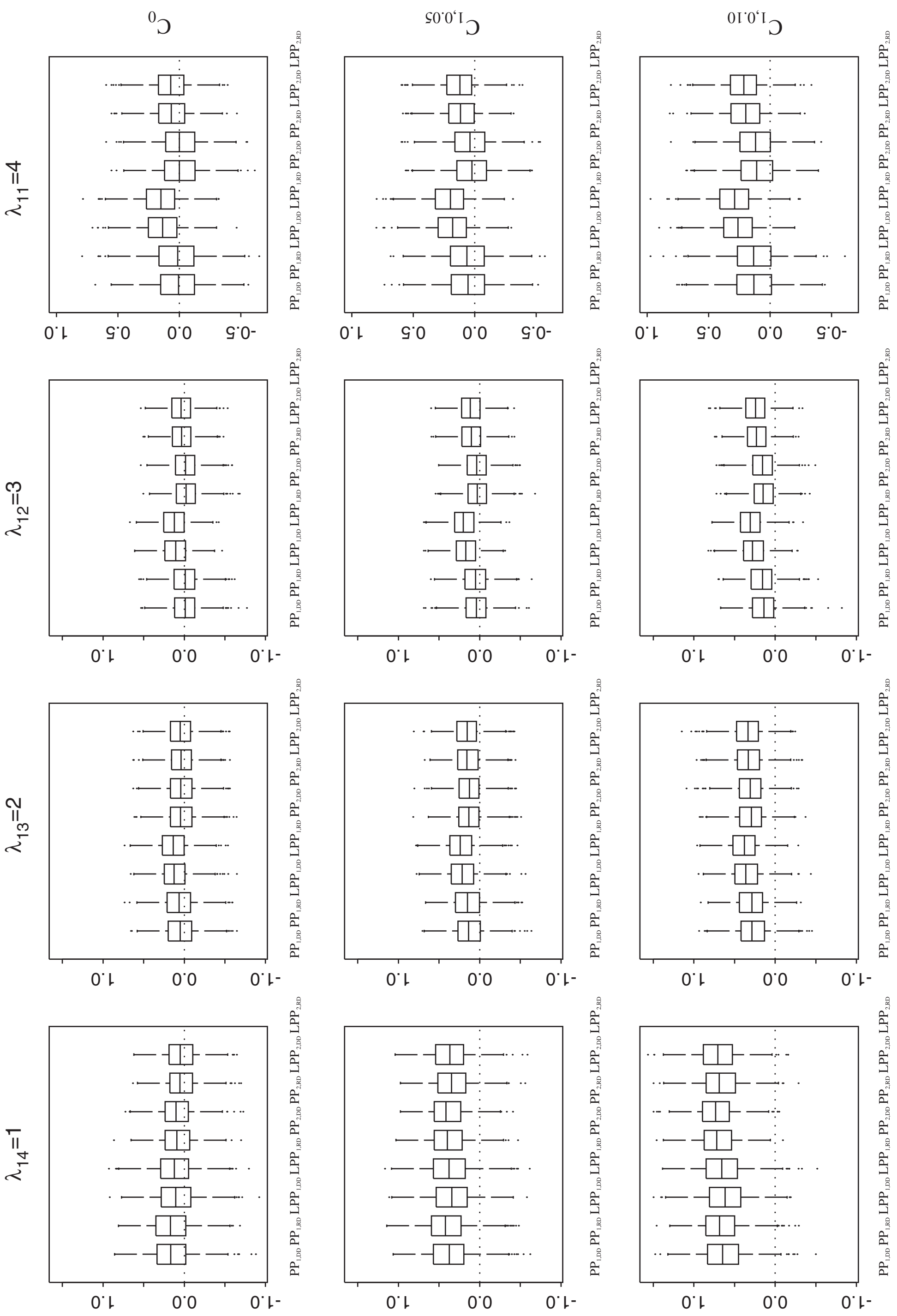

Fig. 2. Boxplots of $\log\left(\frac{\widehat{\lambda}_{1j}}{\lambda_{1j}}\right)$ for Model 2. The dotted lines correspond to 0. The boxplots correspond in order to $PP_{1,\mathrm{DD}}$, $PP_{1,\mathrm{RD}}$, $LPP_{1,\mathrm{DD}}$, $LPP_{1,\mathrm{RD}}$, $PP_{2,\mathrm{DD}}$, $PP_{2,\mathrm{RD}}$, $LPP_{2,\mathrm{DD}}$ and $LPP_{2,\mathrm{RD}}$.

estimator. A better performance could be reached by using, for instance, the χ-function proposed in [5] or a τ-scale estimator.

With regard to eigenvector estimation, in Figs. 3–5 a density estimator (using the normal kernel with a bandwidth equal to 0.3) of the cosines of the angle between the true and the estimated direction related to the smallest eigenvalue of the first population, $\cos(\widehat{\theta}_4)$, is plotted. The solid lines correspond to the densities of $\cos(\widehat{\theta}_4)$ evaluated over the 1000 normally distributed samples, while the dashed and dotted correspond to the asymmetric contaminated samples generated according to $C_{1,0.05}$ and $C_{1,0.10}$, respectively. The vertical lines indicate the corresponding estimated medians.

Under $C_{1,0.10}$ the projection-pursuit estimates perform differently for Model 1 than for Models 2 and 3. Note that the direction corresponding to the smallest eigenvalue of the first population was the one we have chosen to contaminate the samples. A huge sensitivity is observed for all the proposals under Model 1, while for Models 2 and 3 the estimates are more stable under contamination. This can be explained by the fact that for Model 1, the pooled matrix becomes $\mathbf{\Sigma}_{\text{id}} = diag(5, 5.5, 2.25, 3)$ and thus, we are using a distribution for the observations for which the projection-pursuit estimates with $f(t) = t$ will breakdown with a 10% of contamination even by using the MAD scale estimator (see [3]). Moreover, since the larger eigenvalues of $\mathbf{\Sigma}_{\text{id}}$ are quite close, the sample size does not allow to distinguish between them and thus, the projection-pursuit estimates produce a systematic bias for normal data (see [15]). On the other hand, due to the structure of $\mathbf{\Sigma}_{\ln}$, the projection-pursuit procedure based on $f(t) = \ln(t)$ will not be able to distinguish among the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and among $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$, leading to the bias observed in Fig. 3. Moreover, any basis of the subspace generated by $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, could be considered by this procedure as good as the two first canonical vectors. In order to evaluate the performance of the projection-pursuit method based on $f(t) = \ln(t)$, an alternative in this case, could be to evaluate the distance between the linear space generated by $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and that spanned by $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$. Since our aim is to discover the common structure among the scatter matrices, we have only compared the true and the estimated directions. Note that, under the asymmetric contamination $C_{1,\varepsilon}$, these estimates have a slightly better performance than those based on $f(t) = t$.

On the other hand, under Models 2 and 3 all the populations have the common directions related to increasing eigenvalues and since both $\mathbf{\Sigma}_{\text{id}}$ and $\mathbf{\Sigma}_{\ln}$ have different and well separated eigenvalues the projection-pursuit estimates perform much better. This structure helps in the identification of the common direction, as observed in the plots. As expected, in Model 2, the estimates based on $f(t) = \ln(t)$, have a slightly better performance that the raw projection-pursuit ones, a behavior which was already observed in Fig. 4. It is worthwhile noticing, that in Model 3, even if we are not under a proportional model, the estimates based on $f(t) = \ln(t)$ perform similarly or better than those computed with $f(t) = t$, see Fig. 4 under the central model. For a more detailed description see [15], where the performance for other directions is described.

As mentioned above, no improvement in the eigenvector estimation is obtained by using 1000 random directions instead of the normalized data.

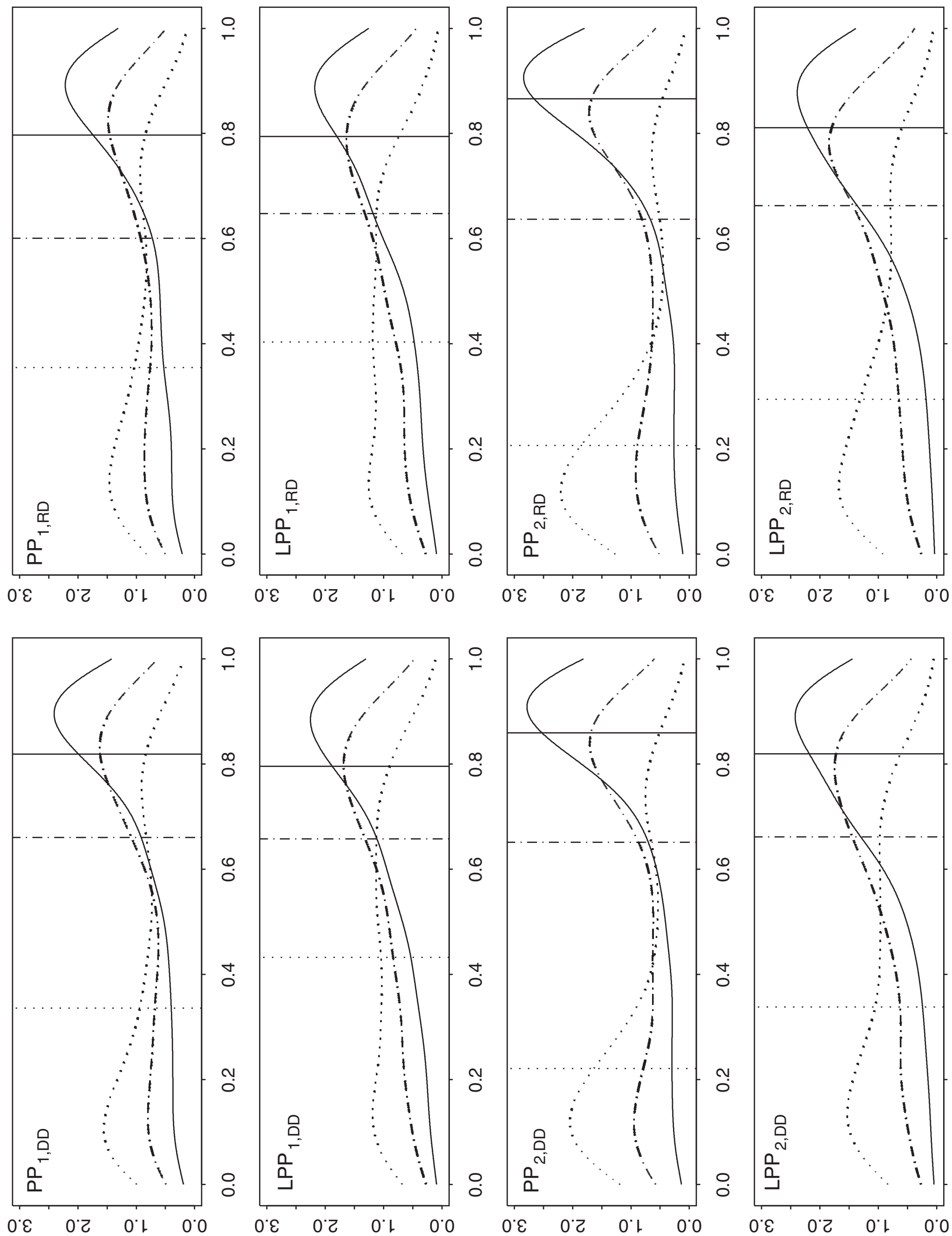

Fig. 3. Density estimates of the cosines of the angle between the true and estimated direction related to the smallest eigenvalue of the first population for Model 1. The plots given in solid lines correspond to C_0, while those in dashed and dotted lines correspond to $C_{1,0.05}$ and $C_{1,0.10}$, respectively. The vertical lines indicate the corresponding estimated medians.

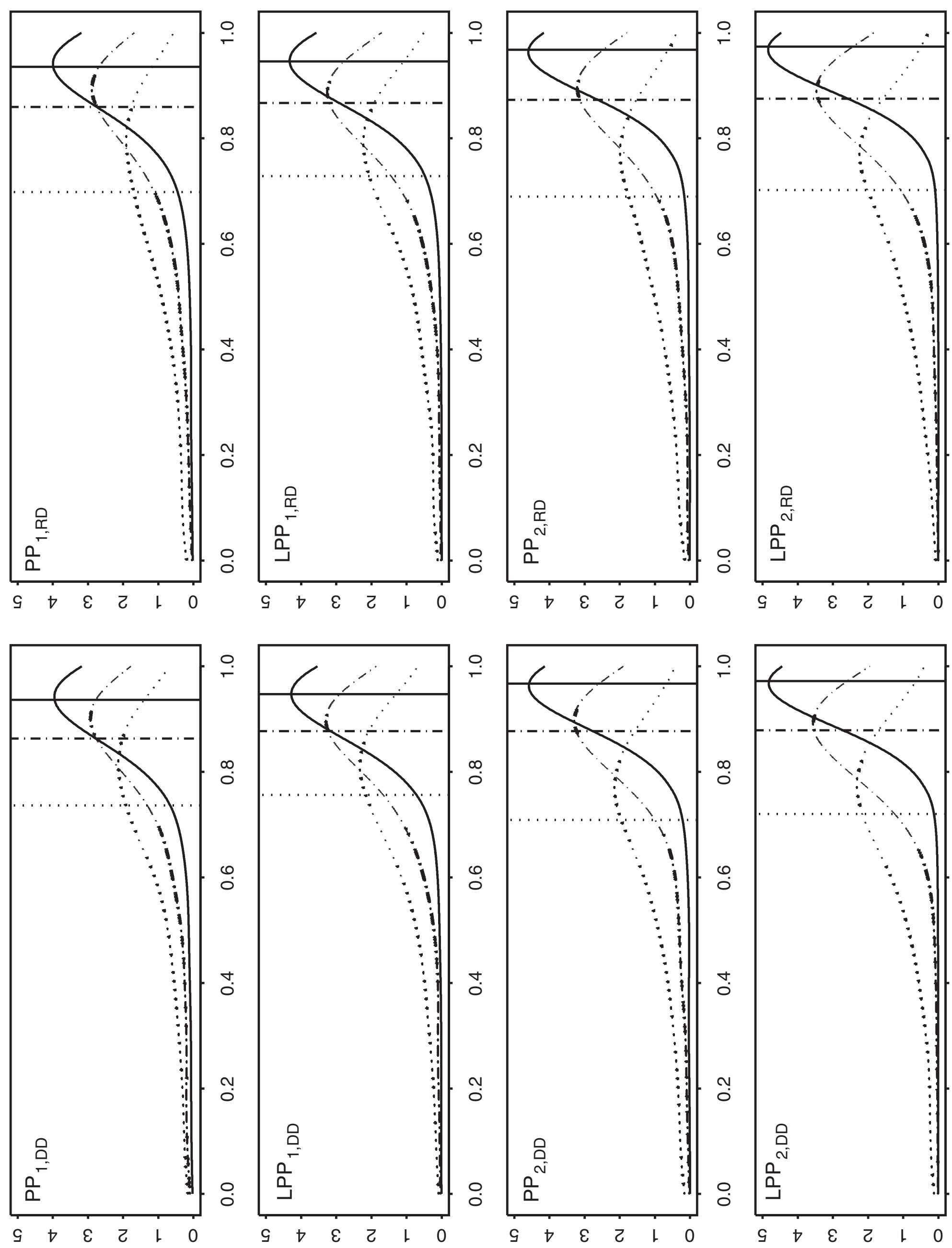

Fig. 4. Density estimates of the cosines of the angle between the true and estimated direction related to the smallest eigenvalue for Model 2. The plots given in solid lines correspond to C_0, while those in dashed and dotted lines correspond to $C_{1,0.05}$ and $C_{1,0.10}$, respectively. The vertical lines indicate the corresponding estimated medians.

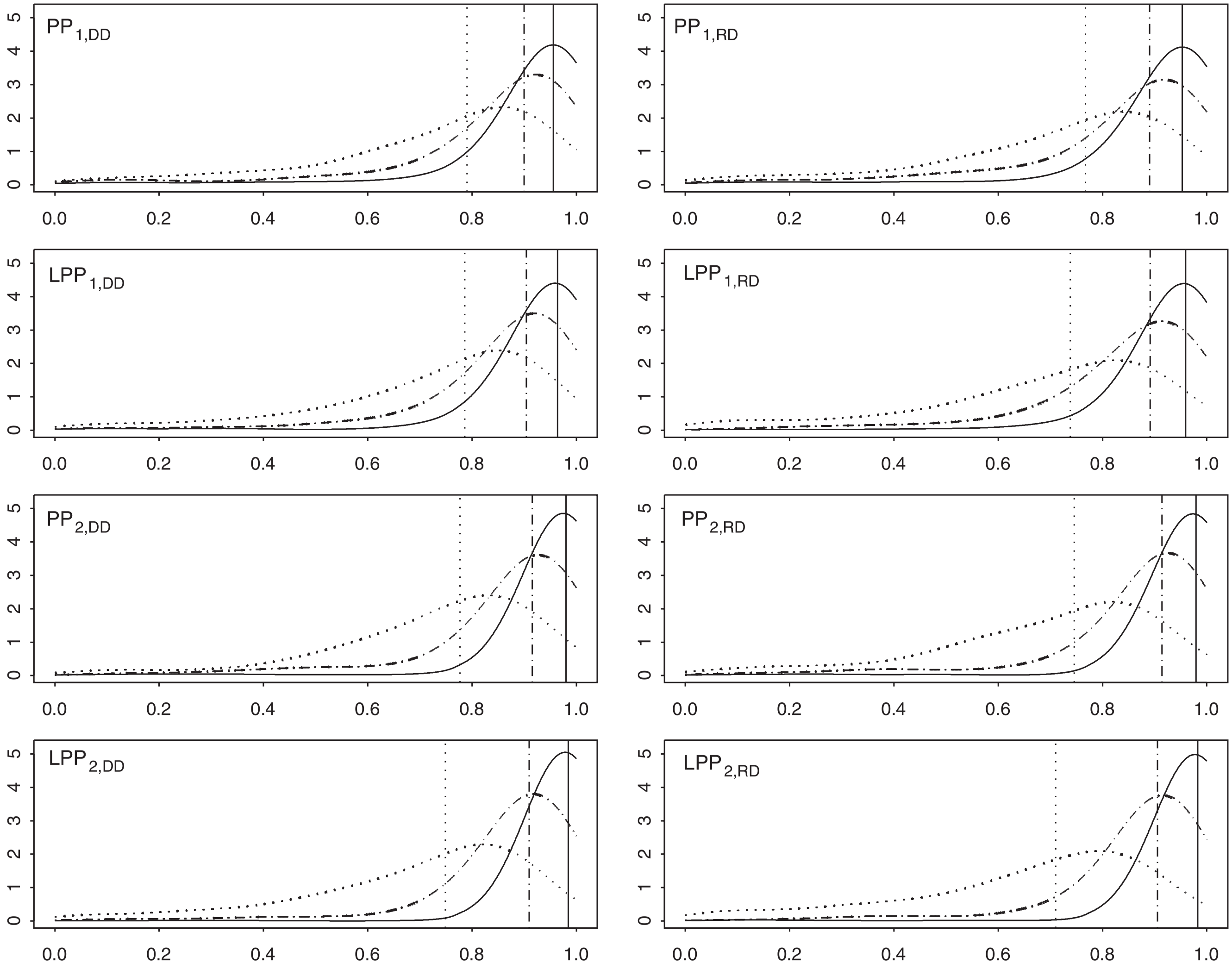

Fig. 5. Density estimates of the cosines of the angle between the true and estimated direction related to the smallest eigenvalue for Model 3. The plots given in solid lines correspond to C_0, while those in dashed and dotted lines correspond to $C_{1,0.05}$ and $C_{1,0.10}$, respectively. The vertical lines indicate the corresponding estimated medians.

5. Conclusions

The proposed generalized projection-pursuit procedure showed its advantage with respect to the usual one, when estimating the common directions. Under an ellipsoidal distribution and a proportional model, the common direction estimators can attain the lowest asymptotic variance for a given scale estimator, using the logarithmic function. Moreover, under a normal distribution, the efficiency of the projection-pursuit common direction estimates evaluated with $f(t) = \ln(t)$ does not depend on the proportionality constants. Proposition 2 also shows that $f(t) = \ln(t)$ minimizes the asymptotic expected value of the square distance between the estimated and the true jth common direction. Also, with respect to the eigenvalue estimates, the generalized projection-pursuit procedure give, for any function f, estimates with the same partial influence functions and thus, with the same asymptotic variance than that corresponding to the identity function.

Our simulation study shows that, with respect to eigenvalues, none of the considered methods has uniformly better performance over all the distributions considered. On the other hand, the random direction selection did not give a great improvement with respect to that considered in [3]. However, it can be an alternative for small data sets. Finally, with respect to the common principal directions, the estimates based on $f(t) = \ln(t)$ show their advantage, in most cases.

Acknowledgments

This research was partially supported by grants from the Agencia Nacional de Promoción Científica y Tecnológica and from the Fundación Antorchas in 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 partially developed while Graciela Boente was visiting the Departamento de Matemática at the Instituto Superior Técnico. The research of Graciela Boente was begun while she was granted with a Guggenheim Fellowship.

We also wish to thank an anonymous referee and the editor for valuable comments which led to an improved version of the original paper.

Appendix A

Proof of Proposition 1. First note that $\sigma^2(F_i[\mathbf{b}]) = \mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}$.
(a) Since $\lambda_{i1} \geqslant \cdots \geqslant \lambda_{ip}$ for $i = 1, \ldots, k$ using that f is strictly increasing and

$$\sup_{\|\mathbf{b}\|=1} (\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}) = \lambda_{i1} = \boldsymbol{\beta}_1'\boldsymbol{\Sigma}_i\boldsymbol{\beta}_1,$$
$$\sup_{\mathbf{b}\in\mathcal{B}_r} (\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}) = \lambda_{ir} = \boldsymbol{\beta}_r'\boldsymbol{\Sigma}_i\boldsymbol{\beta}_r,$$

where $\mathcal{B}_r = \{\mathbf{b} : \|\mathbf{b}\| = 1,\ \boldsymbol{\beta}_\ell'\mathbf{b} = 0,\ \forall 1 \leqslant \ell \leqslant r-1\}$, we get

$$\sup_{\|\mathbf{b}\|=1} \sum_{i=1}^{k} \tau_i f\left(\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}\right) = \sum_{i=1}^{k} \tau_i f(\lambda_{i1}) = \sum_{i=1}^{k} \tau_i f\left(\boldsymbol{\beta}_1'\boldsymbol{\Sigma}_i\boldsymbol{\beta}_1\right),$$
$$\sup_{\mathbf{b}\in\mathcal{B}_r} \sum_{i=1}^{k} \tau_i f\left(\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}\right) = \sum_{i=1}^{k} \tau_i f(\lambda_{ir}) = \sum_{i=1}^{k} \tau_i f\left(\boldsymbol{\beta}_r'\boldsymbol{\Sigma}_i\boldsymbol{\beta}_r\right).$$

Moreover, since for any $j = 1, \ldots, p-1$ there exists $i_0 = i_0(j)$ such that $\lambda_{i_0 j} > \lambda_{i_0, j+1}$, we have that $\mathbf{b}'\boldsymbol{\Sigma}_{i_0}\mathbf{b} < \lambda_{i_0 j}$ for any $\mathbf{b} \in \mathcal{B}_j$ and thus, using again the fact that f is strictly increasing we get

$$\sum_{i=1}^{k} \tau_i f\left(\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}\right) < \sum_{i=1}^{k} \tau_i f(\lambda_{i1}) \quad \forall \|\mathbf{b}\| = 1,$$
$$\sum_{i=1}^{k} \tau_i f\left(\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}\right) < \sum_{i=1}^{k} \tau_i f(\lambda_{ir}) \quad \forall \mathbf{b} \in \mathcal{B}_r.$$

Therefore, for any $j = 1, \ldots, p$, $\boldsymbol{\beta}_{\sigma,j}(F) = \boldsymbol{\beta}_j$ are uniquely defined and the functional is Fisher-consistent. The Fisher-consistency of $\lambda_{\sigma,ij}(F)$ follows immediately.
(b) We can assume without lack of generality that $\boldsymbol{\beta} = \mathbf{I}_p$ and $\xi_1 \geqslant \cdots \geqslant \xi_p$.
Let us first assume that $\xi_1, \ldots, \xi_p$ are pairwise different, then we have $\xi_1 > \cdots > \xi_p$. Given $\mathbf{b} = \left(b_1, \ldots, b_p\right)'$ such that $\|\mathbf{b}\| = 1$, using the fact that f is strictly convex, we have that for all $i = 1, \ldots, k$

$$f\left(\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}\right) = f\left(\sum_{j=1}^{p} b_j^2 \lambda_{ij}\right) \leqslant \sum_{j=1}^{p} b_j^2 f\left(\lambda_{ij}\right) \tag{13}$$

and equality holds if and only if $\forall j \neq m$, such that $b_j \neq 0$ and $b_m \neq 0$, $\lambda_{ij} = \lambda_{im}$.

Using (13) for all $i = 1, \ldots, k$, we get

$$\sum_{i=1}^{k} \tau_i f\left(\sum_{j=1}^{p} b_j^2 \lambda_{ij}\right) \leqslant \sum_{i=1}^{k} \tau_i \sum_{j=1}^{p} b_j^2 f\left(\lambda_{ij}\right) = \mathbf{b}'\Lambda_f\mathbf{b},$$

which together with

$$\begin{cases} \sup_{\|\mathbf{b}\|=1} \left(\mathbf{b}'\Lambda_f\mathbf{b}\right) = \xi_1 = \boldsymbol{\beta}_1'\Lambda_f\boldsymbol{\beta}_1, \\ \sup_{\mathbf{b}\in\mathcal{B}_r} \left(\mathbf{b}'\Lambda_f\mathbf{b}\right) = \xi_r = \boldsymbol{\beta}_r'\Lambda_f\boldsymbol{\beta}_r \end{cases} \tag{14}$$

and $\xi_1 > \cdots > \xi_p$, entails that for $\mathbf{b} \neq \boldsymbol{\beta}_1$

$$\varsigma(\mathbf{b}) = \sum_{i=1}^{k} \tau_i f \left(\sum_{j=1}^{p} b_j^2 \lambda_{ij} \right) \leqslant \mathbf{b}' \Lambda_f \mathbf{b} < \sum_{i=1}^{k} \tau_i f\left(\lambda_{i1}\right) = \xi_1 = \varsigma(\boldsymbol{\beta}_1)$$

and similarly, using induction, given $\mathbf{b} \in \mathcal{B}_r$, $\mathbf{b} \neq \boldsymbol{\beta}_r$ we have

$$\sum_{i=1}^{k} \tau_i f\left(\mathbf{b}'\boldsymbol{\Sigma}_i \mathbf{b}\right) \mathbf{b}' \Lambda_f \mathbf{b} < \xi_r = \varsigma(\boldsymbol{\beta}_r)$$

showing that $\boldsymbol{\beta}_\sigma(F) = \mathbf{I}_p$ is the unique solution of (5). The case for f strictly concave follows in an analogous way.

For the second assumption, $\forall j \neq m\ \exists i_0 : \lambda_{i_0 j} \neq \lambda_{i_0 m}$, it is easy to show using (13) and (14), that $\boldsymbol{\beta}_\sigma(F) = \mathbf{I}_p$ is a solution of (5). To ensure Fisher-consistency, we need to prove uniqueness of solution. Assume that if $\forall j \neq m$ there exists $i_0 = i_0(j, m)$ such that $\lambda_{i_0 j} \neq \lambda_{i_0 m}$. Let $\mathbf{b}$ be such that $\|\mathbf{b}\| = 1, \mathbf{b} \neq \mathbf{e}_j$ for all $j = 1, \ldots, p$, then, we have that there exists $j \neq m$ such that $b_j \neq 0$ and $b_m \neq 0$. Therefore, since there exists $i_0 = i_0(j, m)$ such that $\lambda_{i_0 j} \neq \lambda_{i_0 m}$, the strictly convexity of f entails that $f\left(\mathbf{b}'\boldsymbol{\Sigma}_{i_0}\mathbf{b}\right) < \sum_{j=1}^{p} b_j^2 f\left(\lambda_{i_0 j}\right)$. Now using (13) for $i \neq i_0$ we get that $\varsigma(\mathbf{b}) < \mathbf{b}'\Lambda_f \mathbf{b} \leqslant \xi_1$ and so, $\varsigma(\mathbf{b}) < \varsigma(\boldsymbol{\beta}_1)$ implying that $\boldsymbol{\beta}_{\sigma,1}(F) = \boldsymbol{\beta}_1$ is the unique solution.

Similarly, given $\mathbf{b}$ such that $\|\mathbf{b}\| = 1$, $\mathbf{b} \neq \boldsymbol{\beta}_j$ for all $j = 1, \ldots, p$ and $\mathbf{b} \in \mathcal{B}_r$, we have that $b_j = 0$ for all $1 \leqslant j \leqslant r-1$ and there exists $j \neq m\ j \geqslant r, m \geqslant r$ such that $b_j \neq 0$ and $b_m \neq 0$. Therefore, there exists $i_0 = i_0(j, m)$ such that $\lambda_{i_0 j} \neq \lambda_{i_0 m}$ and so, the strictly convexity of f entails that $f\left(\mathbf{b}'\boldsymbol{\Sigma}_{i_0}\mathbf{b}\right) < \sum_{j=1}^{p} b_j^2 f\left(\lambda_{i_0 j}\right)$. Now using (13) for $i \neq i_0$ we get that $\varsigma(\mathbf{b}) < \mathbf{b}'\Lambda_f \mathbf{b} \leqslant \xi_r$, since $\mathbf{b} \in \mathcal{B}_r$ and so, $\varsigma(\mathbf{b}) < \varsigma(\boldsymbol{\beta}_r)$ implying that $\boldsymbol{\beta}_{\sigma,r}(F) = \boldsymbol{\beta}_r$ is also the unique solution. $\square$

Proof of Theorem 1. Since the proof follows the ideas given in [4], we will only derive the partial influence functions for the eigenvectors. Denote by $F_{i,\varepsilon,\mathbf{x}} = (1-\varepsilon)F_i + \varepsilon\Delta_{\mathbf{x}}$ and by $F_{\varepsilon,\mathbf{x},i} = F_1 \times \cdots \times F_{i-1} \times F_{i,\varepsilon,\mathbf{x}} \times F_{i+1} \times \cdots \times F_k$. Let $\boldsymbol{\beta}_{j,\varepsilon,i} = \boldsymbol{\beta}_{\sigma,j}\left(F_{\varepsilon,\mathbf{x},i}\right)$, $\lambda_{\ell j,\varepsilon,i} = \sigma^2\left(F_\ell\left[\boldsymbol{\beta}_{j,\varepsilon,i}\right]\right)$, for $\ell \neq i$ and $\lambda_{ij,\varepsilon,i} = \sigma^2\left(F_{i,\varepsilon,\mathbf{x}}\left[\boldsymbol{\beta}_{j,\varepsilon,i}\right]\right)$. Denote $\mathbf{V}_{\ell,\varepsilon,i} = \mathbf{V}_{\sigma,\ell}\left(F_{\varepsilon,\mathbf{x},i}\right) = \sum_{j=1}^{p} \lambda_{\ell j,\varepsilon,i} \boldsymbol{\beta}_{j,\varepsilon,i} \boldsymbol{\beta}'_{j,\varepsilon,i}$.

$\boldsymbol{\beta}_{j,\varepsilon,i}$ maximizes $\varsigma\left(F_{\varepsilon,\mathbf{x},i}\left[\mathbf{b}\right]\right)$ under the constraints $\boldsymbol{\beta}'_{j,\varepsilon,i}\boldsymbol{\beta}_{j,\varepsilon,i} = 1$ and $\boldsymbol{\beta}'_{s,\varepsilon,i}\boldsymbol{\beta}_{j,\varepsilon,i} = 0$ for $1 \leqslant s \leqslant j-1$. Therefore, $\boldsymbol{\beta}_{j,\varepsilon,i}$ maximizes the function

$$L(\mathbf{b}, \gamma, \boldsymbol{\alpha}) = \tau_i f\left\{\sigma^2\left(F_{i,\varepsilon,\mathbf{x}}\left[\mathbf{b}\right]\right)\right\} + \sum_{i_0 \neq i} \tau_{i_0} f\left\{\sigma^2\left(F_{i_0}\left[\mathbf{b}\right]\right)\right\}$$
$$-\gamma\left(\mathbf{b}'\mathbf{b} - 1\right) - \sum_{s=1}^{j-1} \alpha_s \mathbf{b}'\boldsymbol{\beta}_{s,\varepsilon,i}$$

and so, it should satisfy

$$0 = \frac{\partial}{\partial \mathbf{b}} L(\mathbf{b}, \gamma, \boldsymbol{\alpha})\bigg|_{\mathbf{b}=\boldsymbol{\beta}_{j,\varepsilon,i}} = \psi(\varepsilon) - 2\gamma \boldsymbol{\beta}_{j,\varepsilon,i} - \sum_{s=1}^{j-1} \alpha_s \boldsymbol{\beta}_{s,\varepsilon,i} \tag{15}$$

with

$$\begin{aligned}\psi(\varepsilon) &= \frac{\partial}{\partial \mathbf{b}} \varsigma\left(F_{\varepsilon,\mathbf{x},i}[\mathbf{b}]\right)\bigg|_{\mathbf{b}=\boldsymbol{\beta}_{j,\varepsilon,i}} \\ &= \tau_i f'\left\{\sigma^2\left(F_{i,\varepsilon,\mathbf{x}}\left[\boldsymbol{\beta}_{j,\varepsilon,i}\right]\right)\right\} \frac{\partial}{\partial \mathbf{b}} \left\{\sigma^2\left(F_{i,\varepsilon,\mathbf{x}}\left[\mathbf{b}\right]\right)\right\} |_{\mathbf{b}=\boldsymbol{\beta}_{j,\varepsilon,i}} \\ &\quad + \sum_{i_0 \neq i} \tau_{i_0} f'\left\{\sigma^2\left(\boldsymbol{\beta}'_{j,\varepsilon,i} \boldsymbol{\Sigma}_{i_0} \boldsymbol{\beta}_{j,\varepsilon,i}\right)\right\} \frac{\partial}{\partial \mathbf{b}} \left\{\sigma^2\left(F_{i_0}\left[\mathbf{b}\right]\right)\right\} |_{\mathbf{b}=\boldsymbol{\beta}_{j,\varepsilon,i_0}}.\end{aligned} \tag{16}$$

Since $\boldsymbol{\beta}'_{j,\varepsilon,i}\boldsymbol{\beta}_{j,\varepsilon,i} = 1$ and $\boldsymbol{\beta}'_{s,\varepsilon,i}\boldsymbol{\beta}_{j,\varepsilon,i} = 0$ for $1 \leqslant s \leqslant j-1$, we have that $\psi(\varepsilon)'\boldsymbol{\beta}_{j,\varepsilon,i} = 2\gamma$ and $\psi(\varepsilon)'\boldsymbol{\beta}_{s,\varepsilon,i} = \alpha_s$ for $1 \leqslant s \leqslant j-1$. Therefore, (15) can be written as $\psi(\varepsilon) = \sum_{s=1}^{j}\left(\psi(\varepsilon)'\boldsymbol{\beta}_{s,\varepsilon,i}\right)\boldsymbol{\beta}_{s,\varepsilon,i}$. Assumptions A3 and A4 entail that $\psi(\varepsilon)$ is differentiable. Thus, differentiating this last expression with respect to ε, we get

$$\begin{aligned}\frac{\partial}{\partial \varepsilon}\psi(\varepsilon)\bigg|_{\varepsilon=0} &= \sum_{s=1}^{j}\Bigg[\left(\psi(0)' PIF_i\left(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F\right)\right)\boldsymbol{\beta}_s + \left(\boldsymbol{\beta}'_s \frac{\partial}{\partial \varepsilon}\psi(\varepsilon)|_{\varepsilon=0}\right)\boldsymbol{\beta}_s \\ &\quad + \left(\psi(0)'\boldsymbol{\beta}_s\right) PIF_i\left(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F\right)\Bigg].\end{aligned} \tag{17}$$

Since F_i is an elliptical distribution and $\sigma(G_0) = 1$, using that $\sigma^2\left(F_i[\mathbf{b}]\right) = \mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}$, we obtain that $\psi(0) = \frac{\partial}{\partial \mathbf{b}}\varsigma\left(F_{0,\mathbf{x},i}[\mathbf{b}]\right)|_{\mathbf{b}=\boldsymbol{\beta}_j} = 2\nu_j\boldsymbol{\beta}_j$, which implies that $\psi(0)'\boldsymbol{\beta}_s = 0$ for $1 \leqslant s \leqslant j-1$. Denote $\mathbf{P}_{j+1} = \mathbf{I}_p - \sum_{s=1}^{j}\boldsymbol{\beta}_s\boldsymbol{\beta}'_s$, then (17) can be written as

$$\mathbf{P}_{j+1}\frac{\partial}{\partial \varepsilon}\psi(\varepsilon)\bigg|_{\varepsilon=0} = 2\nu_j\sum_{s=1}^{j}\boldsymbol{\beta}'_j PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F)\boldsymbol{\beta}_s + 2\nu_j PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F). \tag{18}$$

On the other hand, from (16) and since $\varsigma(F[\mathbf{b}]) = \sum_{i=1}^{k}\tau_i f\left(\mathbf{b}'\boldsymbol{\Sigma}_i\mathbf{b}\right)$, we have that

$$\begin{aligned}\frac{\partial}{\partial \varepsilon}\psi(\varepsilon)|_{\varepsilon=0} &= 2\,\widetilde{\boldsymbol{\Sigma}}_j\, PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) \\ &\quad + \tau_i\frac{\partial}{\partial \mathbf{b}}\left[f'\left\{\sigma^2\left(F_i[b]\right)\right\} IF\left(\mathbf{b}'\mathbf{x}, \sigma^2, F_i[\mathbf{b}]\right)\right]\bigg|_{\mathbf{b}=\boldsymbol{\beta}_j},\end{aligned} \tag{19}$$

where $\widetilde{\mathbf{\Sigma}}_j = \sum_{i=1}^k \tau_i f'\left(\lambda_{ij}\right)\mathbf{\Sigma}_i = \sum_{s=1}^p v_{js}\boldsymbol{\beta}_s'\boldsymbol{\beta}_s$. Again from the equivariance of the scale estimator, we have

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{b}}\left[f'\left\{\sigma^2\left(\mathbf{b}'\mathbf{\Sigma}_i\mathbf{b}\right)\right\} IF\left(\mathbf{b}'\mathbf{x}, \sigma^2, F_i[\mathbf{b}]\right)\right]\Bigg|_{\mathbf{b}=\boldsymbol{\beta}_j}\\
&= 2\lambda_{ij} f''\left(\lambda_{ij}\right)\boldsymbol{\beta}_j\ \lambda_{ij} IF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right)\\
&\quad + f'\left(\lambda_{ij}\right)\left\{2\lambda_{ij}\boldsymbol{\beta}_j IF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right)\right.\\
&\quad \left. + \lambda_{ij} DIF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right)\left(\frac{\mathbf{x}}{\sqrt{\lambda_{ij}}} - \frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}\boldsymbol{\beta}_j\right)\right\}.
\end{aligned} \tag{20}$$

From (19) and (20), we get

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon}\psi(\varepsilon)\Bigg|_{\varepsilon=0} &= 2\,\widetilde{\mathbf{\Sigma}}_j\, PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)\\
&\quad + 2\tau_i\left\{f''\left(\lambda_{ij}\right)\lambda_{ij} + f'\left(\lambda_{ij}\right)\right\}\lambda_{ij}\boldsymbol{\beta}_j IF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right)\\
&\quad + \tau_i f'\left(\lambda_{ij}\right)\sqrt{\lambda_{ij}} DIF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right)\left(\mathbf{x} - \boldsymbol{\beta}_j'\mathbf{x}\ \boldsymbol{\beta}_j\right).
\end{aligned} \tag{21}$$

Therefore, from (21), (18) can be written as

$$\begin{aligned}
&2v_j \sum_{s=1}^{j} \boldsymbol{\beta}_j'\, PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F)\boldsymbol{\beta}_s + 2v_j PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)\\
&= 2\,\mathbf{P}_{j+1}\,\widetilde{\mathbf{\Sigma}}_j\, PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)\\
&\quad + \tau_i\sqrt{\lambda_{ij}} f'\left(\lambda_{ij}\right) DIF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right)\mathbf{P}_{j+1}\mathbf{x}.
\end{aligned} \tag{22}$$

Note that $\boldsymbol{\beta}_{j,\varepsilon}'\boldsymbol{\beta}_{j,\varepsilon} = 1$ entails that $PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)'\boldsymbol{\beta}_j = 0$, therefore (22) is equivalent to

$$\begin{aligned}
2\left(\mathbf{P}_{j+1}\widetilde{\mathbf{\Sigma}}_j - v_j\mathbf{I}_p\right) PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) &= 2v_j \sum_{s=1}^{j-1} \boldsymbol{\beta}_j' PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F)\boldsymbol{\beta}_s\\
&\quad - \tau_i\sqrt{\lambda_{ij}} f'\left(\lambda_{ij}\right) DIF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right)\mathbf{P}_{j+1}\mathbf{x}.
\end{aligned}$$

The matrix $\mathbf{P}_{j+1}\widetilde{\mathbf{\Sigma}}_j - v_j\mathbf{I}_p = \sum_{s=j+1}^p v_{js}\boldsymbol{\beta}_s\boldsymbol{\beta}_s' - v_j\mathbf{I}_p$ is a full rank matrix since $v_{js} \neq v_{jj}$ for $s \neq j$, with inverse $\left(\mathbf{P}_{j+1}\widetilde{\mathbf{\Sigma}}_j - v_j\mathbf{I}_p\right)^{-1} = \sum_{s=j+1}^p \frac{1}{v_{js}-v_j}\boldsymbol{\beta}_s\boldsymbol{\beta}_s' - \sum_{s=1}^j \frac{1}{v_j}\boldsymbol{\beta}_s\boldsymbol{\beta}_s'$. Then, for $1 \leqslant s \leqslant j-1$, we have $\left(\mathbf{P}_{j+1}\widetilde{\mathbf{\Sigma}}_j - v_j\mathbf{I}_p\right)^{-1}\boldsymbol{\beta}_s = -\frac{1}{v_j}\boldsymbol{\beta}_s$ and $\left(\mathbf{P}_{j+1}\widetilde{\mathbf{\Sigma}}_j - v_j\mathbf{I}_p\right)^{-1}\mathbf{P}_{j+1} =$

$\sum_{s=j+1}^{p} \frac{1}{v_{js}-v_j} \boldsymbol{\beta}_s \boldsymbol{\beta}_s'$. Hence, we obtain

$$PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F) = -\sum_{s=1}^{j-1} \boldsymbol{\beta}_j' PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,s}, F)\boldsymbol{\beta}_s$$
$$-\frac{1}{2}\sum_{s=j+1}^{p} \frac{1}{v_{js}-v_j} \boldsymbol{\beta}_s \tau_i \sqrt{\lambda_{ij}} f'\left(\lambda_{ij}\right) DIF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right) \boldsymbol{\beta}_s'\mathbf{x},$$

which implies that

$$PIF_i(\mathbf{x}, \boldsymbol{\beta}_{\sigma,j}, F)'\boldsymbol{\beta}_s = \frac{1}{2(v_j - v_{js})} \tau_i \sqrt{\lambda_{ij}} f'\left(\lambda_{ij}\right) DIF\left(\frac{\boldsymbol{\beta}_j'\mathbf{x}}{\sqrt{\lambda_{ij}}}, \sigma^2, G_0\right) \boldsymbol{\beta}_s'\mathbf{x},$$

for any $s \geqslant j+1$. Finally, using that $IF\left(y, \sigma^2, G_0\right) = 2IF\left(y, \sigma, G_0\right)$, we get (11). □

Proof of Corollary 1. It follows as in [4] using (9), (10), (11) and that $\frac{\mathbf{x}'\boldsymbol{\beta}_j}{\sqrt{\lambda_{ij}}} \sim G_0$ when $\mathbf{x} \sim F_i$. □

Proof of Proposition 2. For the sake of simplicity, we will only consider the case when $m > j$, the other one being analogous. From (12), we have that the minimum variance will be attained for a strictly increasing function f that minimizes $V(f) = \sum_{i=1}^{k} \left\{\tau_i^{\frac{1}{2}} \rho_i f'\left(\rho_i\lambda_j\right)\right\}^2 \left\{\sum_{i=1}^{k} \tau_i^{\frac{1}{2}}\, \tau_i^{\frac{1}{2}} \rho_i f'\left(\rho_i \lambda_j\right)\right\}^{-2} = \|\mathbf{y}_j\|^2 \left(\mathbf{y}_j'\boldsymbol{\xi}\right)^{-2}$, where $\mathbf{y}_j = \left(\tau_1^{\frac{1}{2}} \rho_1 f'\left(\rho_1\lambda_j\right), \ldots, \tau_k^{\frac{1}{2}} \rho_k f'\left(\rho_k\lambda_j\right)\right)'$ and $\boldsymbol{\xi} = \left(\tau_1^{\frac{1}{2}}, \ldots, \tau_k^{\frac{1}{2}}\right)'$. The Cauchy–Schwartz inequality entails $\left(\mathbf{y}_j'\boldsymbol{\xi}\right)^2 \leqslant \|\mathbf{y}_j\|^2 \|\boldsymbol{\xi}\|^2$ and thus $V(f) \geqslant \|\boldsymbol{\xi}\|^{-2} = 1 = V(\ln)$.

Moreover the minimum value of $V(f)$ is attained if and only if $\mathbf{y}_j$ and $\boldsymbol{\xi}$ are collinear, i.e., if and only if $\rho_i f'\left(\rho_i\lambda_j\right) = f'\left(\lambda_j\right)$, for all i, j. Since there exists at least one $\rho_i \neq 1$ this is verified for arbitrary ρ_i and λ_j if and only if there exists $a \neq 0$ and $b \in \mathbb{R}$ such that $f(\cdot) = a\ln(\cdot) + b$. □

Proof of Proposition 3. For the sake of simplicity, we will only consider the case when $m > j$, the proof of (ii) being analogous. From (12), we have that the minimum variance will be attained for the function $f \in \mathcal{F}$ that minimizes

$$H(\alpha) = \frac{\sum_{i=1}^{k} \tau_i a_{i,mj} \lambda_{ij}^{2\alpha}}{\left\{\sum_{i=1}^{k} \tau_i \left(1 - a_{i,mj}\right) \lambda_{ij}^{\alpha}\right\}^2}.$$

Note that since $\lambda_{i_0 j} > \lambda_{i_0 m}$, $a_{i_0,mj} \neq 1$ and the assumption $v_{jm} \neq v_j$ holds.

Differentiating with respect to α, straightforward calculations lead to

$$\begin{aligned}
&H'(\alpha)\\
&= \frac{\sum_{i=1}^{k} \tau_i \lambda_{ij}^{\alpha} \sum_{\ell=1}^{i-1} \tau_\ell \lambda_{\ell j}^{\alpha} \left[\ln\left(\lambda_{ij}^2\right) - \ln\left(\lambda_{\ell j}^2\right)\right]\left[a_{i,mj}\left(1-a_{\ell,mj}\right)\lambda_{ij}^{\alpha} - a_{\ell,mj}\left(1-a_{i,mj}\right)\lambda_{\ell j}^{\alpha}\right]}{\left\{\sum_{i=1}^{k} \tau_i \left(1-a_{i,mj}\right)\lambda_{ij}^{\alpha}\right\}^3}\\
&= \frac{N(\alpha)}{D(\alpha)}.
\end{aligned}$$

Since $\lambda_{ij} \geqslant \lambda_{im}$ and $\lambda_{i_0 j} > \lambda_{i_0 m}$, we have $D(\alpha) > 0$. On the other hand, the order assumptions made among the eigenvalues of the different populations entail that

$$\left[\ln\left(\lambda_{ij}^2\right) - \ln\left(\lambda_{\ell j}^2\right)\right]\left[a_{i,mj}\left(1-a_{\ell,mj}\right)\lambda_{ij}^{\alpha} - a_{\ell,mj}\left(1-a_{i,mj}\right)\lambda_{\ell j}^{\alpha}\right] \geqslant 0,$$

which implies that $N(\alpha) \geqslant 0$ and thus $H(\alpha)$ is an increasing function. Hence, its minimum is attained at $\alpha = 0$ which corresponds to $f(t) = \ln(t)$, as desired. Moreover if there exists, $\ell < i$ such that $\lambda_{\ell j} < \lambda_{ij}$ and $a_{\ell,mj} < a_{i,mj}$, then $N(\alpha) > 0$ implying that $H(\alpha)$ is a strictly increasing function and thus $f(t) = \ln(t)$ is the unique minimizer of $H(\alpha)$. □

References

[1] J.R. Berrendero, Contribuciones a la teoría de la robustez respecto al sesgo, Unpublished Ph.D. Thesis, Universidad Carlos III de Madrid, 1996 (in Spanish).

[2] G. Boente, F. Critchley, L. Orellana, Influence functions for robust estimators under proportional scatter matrices, Working paper, Universidad de Buenos Aires.

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

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

[5] C. Croux, Efficient high-breakdown M-estimators of scale, Statist. Probab. Lett. 19 (1994) 371–379.

[6] C. Croux, A. Ruiz-Gazen, A fast algorithm for robust principal components based on projection pursuit, in: A. Prat (Ed.), Compstat: Proceedings in Computational Statistics, Physica-Verlag, Heidelberg, 1996, pp. 211–217.

[7] C. Croux, A. Ruiz-Gazen, High breakdown estimators for principal components: the projection-pursuit approach revisited, Working paper, Université Libre de Bruxelles, 2000.

[8] H. Cui, X. He, K.W. Ng, Asymptotic distribution of principal components based on robust dispersions, Biometrika 90 (2003) 953–966.

[9] B.K. Flury, Common principal components in k groups, J. Amer. Statist. Assoc. 79 (1984) 892–898.

[10] B.K. Flury, Common Principal Components and Related Multivariate Models, Wiley, New York, 1988.

[11] P. Huber, Robust Statistics, Wiley, New York, 1981.

[12] G. Li, Z. Chen, Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo, J. Amer. Statist. Assoc. 80 (1985) 759–766.

[13] Z. Patak, Robust principal components, Master Thesis of the Department of Statistics, University of British Columbia, Vancouver, 1991.

[14] A.M. Pires, J. Branco, Partial influence functions, J. Multivariate Anal. 83 (2002) 458–468.

[15] I.M. Rodrigues, Métodos robustos em análise de componentes principais comuns, Unpublished Ph.D. Thesis,Universidade Técnica de Lisboa, 2003 (in Portuguese). Available on http://www.math.ist.utl.pt/ãpires/phd.html.