Biometrika (2002), **89**, 4, *pp.* 861–875

Printed in Great Britain

Influence functions and outlier detection under the common principal components model: A robust approach

By GRACIELA BOENTE

Conicet, Departamento de Matemática and Instituto de Cálculo, Ciudad Universitaria, Pabellón 1, Buenos Aires, C1428EHA, Argentina

gboente@mate.dm.uba.ar

ANA M. PIRES and ISABEL M. RODRIGUES

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

ana.pires@math.ist.utl.pt   isabel.rodrigues@math.ist.utl.pt

Summary

The common principal components model for several groups of multivariate observations assumes equal principal axes but different variances along these axes among the groups. Influence functions for plug-in and projection-pursuit estimates under a common principal component model are obtained. Asymptotic variances are derived from them. Outlier detection is possible using partial influence functions.

Some key words: Asymptotic variance; Common principal components; Partial influence function; Projection-pursuit; Robust estimation; Robust scatter matrix

1. Introduction

If in multivariate analysis involving several populations the assumption of equality of covariance matrices is not adequate, problems may arise because of an excessive number of parameters if we estimate the covariance 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), generalises equality by allowing the matrices to have different eigenvalues but identical eigenvectors; that is

$$\Sigma_i = \beta \Lambda_i \beta' \quad (1 \leqslant i \leqslant k), \tag{1}$$

where the Λ_i are diagonal matrices and β is the orthogonal matrix of the common eigenvectors. Flury (1984) described how to obtain the maximum likelihood estimators of β and the Λ_i under normality. Their asymptotic distribution was studied by Flury (1986).

As in the one-population setting, one of the aims when performing a common principal components analysis is to reduce the dimensionality of the data by imposing a common structure on the k populations and by retaining as much as possible of the variability present in each set. Even if the principal axes are the same for all populations, the amount of variability explained by each of them may vary among populations.

In biometric applications, principal components are frequently interpreted as independent factors determining the growth, size or shape of an organism. It seems therefore reasonable to consider a model in which the same factors arise in different, but related, species. The common principal components model clearly serves this purpose. Applications of this and some other hierarchical models are discussed, for instance, in Flury (1988), Klingerberg & Neuenschwander (1996), Arnold & Phillips (1999) and Phillips & Arnold (1999).

Let $x_{i1}, \ldots, x_{in_i}$ $(1 \leqslant i \leqslant k)$ be independent observations from k independent samples in R^p with location parameter μ_i and scatter matrix Σ_i. For the sake of technical simplicity and without changing the final results, when computing the partial influence functions it will be assumed that $\mu_i = 0_p$, where 0_p denotes a p-dimensional vector of zeros. Let $N = \sum_{i=1}^{k} n_i$ and $\tau_i = n_i/N$. As in principal component analysis, classical common principal components analysis can be affected by outliers in the sample, so it is of interest to study robust estimators of the common eigenvectors and the related eigenvalues for each sample, under the common principal components model (1). As in the one-population setting, robust affine-equivariant estimators of the covariance matrices Σ_i, can be considered; see Novi Inverardi & Flury (1992) and Boente & Orellana (2001). Projection-pursuit estimators, introduced by these last authors, are an alternative whose main advantage is that they can be computed for datasets with more variables than observations.

Influence functions are measures of robustness with respect to single outliers. When we deal with several populations, partial influence functions measure resistance towards pointwise contaminations of each population.

Several authors, such as Critchley (1985), Jaupi & Saporta (1993), Shi (1997), Croux & Haesbroeck (1999) and C. Croux and A. Ruiz-Gazen, in a Université Libre de Bruxelles technical report, have suggested statistical diagnostics and graphical displays for detecting outliers in prinicpal component analysis for one population, such as side-by-side boxplots of the scores obtained from a robust principal component analysis and index plots based on empirical influence functions. Under a common principal components model, partial influence functions can also be used to detect influential observations in a sample.

In § 2, partial influence functions are derived for the estimates obtained by plugging equivariant scatter matrix estimators, possessing an influence function, into the equations defining the maximum likelihood estimators, and for those defined through a projection-pursuit approach by using a scale estimator with differentiable influence function. Asymptotic variances are then computed for the projection-pursuit approach. In § 3, outlier detection is discussed, while in § 4 a biological example is analysed. All proofs are given in the Appendix.

2. Influence functions

2·1. *Preamble*

Partial influence functions were introduced by Pires & Branco (2002) 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, $F = F_1 \times \ldots \times F_k$. Partial influence functions of a functional $T(F)$ are then defined as

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

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

In Pires & Branco (2002), it is shown that the following expansion holds:

$$N^{\frac{1}{2}}\{T(F_N)-T(F)\} = \sum_{i=1}^{k} \frac{1}{(\tau_i n_i)^{\frac{1}{2}}} \sum_{j=1}^{n_i} \text{PIF}_i(x_{ij}, T, F) + o_p(1),$$

where F_N denotes the empirical distribution of the k independent samples x_{ij} $(1 \leqslant j \leqslant n_i, 1 \leqslant i \leqslant k)$. Therefore, the asymptotic variance of the estimators can be evaluated as

$$\text{avar}(T, F) = \sum_{i=1}^{k} \tau_i^{-1} E_{F_i}\{\text{PIF}_i(x_{i1}, T, F)\, \text{PIF}_i(x_{i1}, T, F)'\}. \tag{2}$$

2·2. *Influence functions for plug-in estimators*

These estimators are obtained by plugging robust scatter matrices into the equations defining the maximum likelihood estimators for normal data. They are defined as the solution of

$$\text{diag}(\hat{\beta}' V_i \hat{\beta}) = \hat{\Lambda}_i,$$
$$\hat{\beta}_m' \left(\sum_{i=1}^{k} n_i \frac{\hat{\lambda}_{im} - \hat{\lambda}_{ij}}{\hat{\lambda}_{im}\hat{\lambda}_{ij}} V_i \right) \hat{\beta}_j = 0 \quad (m \neq j), \tag{3}$$
$$\hat{\beta}_m' \hat{\beta}_j = \delta_{mj},$$

where the V_i are robust consistent scatter matrices and $\delta_{mj} = 1$ is the Kronecker delta. As with maximum likelihood estimation, a solution for (3) always exists, since the group of orthogonal matrices is compact. Uniqueness conditions are similar to those given in Flury (1988, pp. 194–200) for the maximum likelihood estimators.

For a given distribution $F = F_1 \times \ldots \times F_k$, let $V_i = V_i(F_i)$ be a robust scatter functional evaluated at the distribution of the ith sample. We will thus define the functionals $\beta_V(F)$, $\Lambda_{V,i}(F)$ $(1 \leqslant i \leqslant k)$, related to $V = (V_1, \ldots, V_k)$, as the solution of

$$\text{diag}\{\beta_V(F)' V_i(F_i) \beta_V(F)\} = \Lambda_{V,i}(F), \tag{4}$$

$$\beta_{V,m}(F)' \left\{ \sum_{i=1}^{k} \tau_i \frac{\lambda_{V,im}(F) - \lambda_{V,ij}(F)}{\lambda_{V,im}(F)\lambda_{V,ij}(F)} V_i(F_i) \right\} \beta_{V,j}(F) = 0 \quad (m \neq j), \tag{5}$$

$$\beta_{V,m}(F)' \beta_{V,j}(F) = \delta_{mj}. \tag{6}$$

When V_i provides Fisher-consistent estimators, that is $V_i(F_i) = \Sigma_i$, the solutions $(\Lambda_{V,i}(F), \beta_V(F))$ are Fisher-consistent for (Λ_i, β).

The following theorem gives the values of the partial influence functions for the plug-in estimators.

THEOREM 1. *Let $V_i(F)$ be a scatter functional such that $V_i(F_i) = \Sigma_i$. Denote by $\beta_1, \ldots, \beta_p$, $\lambda_{i1}, \ldots, \lambda_{ip}$ the common eigenvectors and the eigenvalues of Σ_i. Assume that the influence function* IF(x, V_i, F_i) *exists and that $\lambda_{11} > \ldots > \lambda_{1p}$. Then the partial influence functions of the solution $\beta_V(F)$, $\Lambda_{V,i}(F)$ $(1 \leqslant i \leqslant k)$ of* (4) *to* (6) *are given by*

$$\text{PIF}_i(x, \lambda_{V,lj}, F) = \delta_{li} \beta_j' \,\text{IF}(x, V_i, F_i) \beta_j, \tag{7}$$

$$\text{PIF}_i(x, \beta_{V,j}, F) = \tau_i \sum_{m \neq j} \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im}\lambda_{ij}} \left\{ \sum_{l=1}^{k} \tau_l \frac{(\lambda_{lm} - \lambda_{lj})^2}{\lambda_{lm}\lambda_{lj}} \right\}^{-1} \{\beta_j' \,\text{IF}(x, V_i, F_i) \beta_m\} \beta_m. \tag{8}$$

Remark 1. Theorem 1 entails that, for the sample covariance matrix, the partial influence functions are given by

$$\mathrm{PIF}_i(x, \lambda_{S,lj}, F) = \delta_{li}\beta_j'(xx' - \Sigma_i)\beta_j = \delta_{li}\{(\beta_j'x)^2 - \lambda_{ij}\}, \quad (9)$$

$$\mathrm{PIF}_i(x, \beta_{S,j}, F) = \tau_i \sum_{m \neq j} \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im}\lambda_{ij}} \left\{ \sum_{l=1}^{k} \tau_l \frac{(\lambda_{lm} - \lambda_{lj})^2}{\lambda_{lm}\lambda_{lj}} \right\}^{-1} \{\beta_j'(xx' - \Sigma_i)\beta_m\}\beta_m$$

$$= \tau_i \sum_{m \neq j} \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im}\lambda_{ij}} \left\{ \sum_{l=1}^{k} \tau_l \frac{(\lambda_{lm} - \lambda_{lj})^2}{\lambda_{lm}\lambda_{lj}} \right\}^{-1} \beta_j' x \beta_m' x \beta_m, \quad (10)$$

and are therefore unbounded.

On the other hand, the partial influence functions for the eigenvalues given by (7) are analogous to the influence function obtained by Croux & Haesbroeck (2000) in the one-population setting. Plots for the classical and the *S*-estimates are given therein. The partial influence functions for the eigenvectors include weights depending on the eigenvalues of the k populations. The behaviour of the partial influence function will be analogous to that described by Croux & Haesbroeck (2000).

From the expressions given for the partial influence functions and using (2), one obtains the expressions for the asymptotic variance of the plug-in estimators solution of (3), which were derived in Theorems 2 and 3 of Boente & Orellana (2001), when the estimators of the scatter matrix Σ_i are asymptotically normally distributed and spherically invariant.

2·3. *Influence functions for projection-pursuit estimators*

Let x_i be independent vectors such that $x_i \sim F_i$, where F_i has location parameter μ_i and scatter matrix $\Sigma_i = C_i C_i'$ satisfying (1). Denote by $F_i[b]$ the distribution of $b'x_i$ and write $F[b] = F_1[b] \times \ldots \times F_k[b]$.

Let $\varsigma(b) = \varsigma(F[b]) = \sum_{i=1}^{k} \tau_i \sigma^2(F_i[b])$, where $\sigma(.)$ is a univariate scale estimator which is equivariant under scale transformations.

The projection-pursuit functional for the common directions, $\beta_\sigma(F)$, with columns $\beta_{\sigma,1}(F), \ldots, \beta_{\sigma,p}(F)$, is the solution of

$$\varsigma\{\beta_{\sigma,1}(F)\} = \sup_{\|b\|=1} \varsigma(b), \quad \varsigma\{\beta_{\sigma,j}(F)\} = \sup_{b \in \mathscr{B}_j} \varsigma(b) \quad (2 \leqslant j \leqslant p),$$

where $\mathscr{B}_j = \{b : \|b\| = 1, b'\beta_{\sigma,m}(F) = 0 \text{ for } 1 \leqslant m \leqslant j-1\}$, while the functionals related to the estimators of the eigenvalues and of the covariance matrix are defined as

$$\lambda_{\sigma,ij}(F) = \sigma^2\{F_i[\beta_{\sigma,j}(F)]\}, \quad V_{\sigma,i}(F) = \sum_{j=1}^{p} \lambda_{\sigma,ij}(F)\beta_{\sigma,j}(F)\beta_{\sigma,j}(F)'.$$

A different definition arises if we take infimum instead of supremum, but both will have the same partial influence functions. When $\Sigma = \sum_{i=1}^{k} \tau_i \Sigma_i$ has no multiple root, Boente & Orellana (2001) have shown that the functionals $\beta_\sigma(F)$ and $\lambda_{\sigma,ij}(F)$ will be Fisher-consistent at any distribution F such that $z_i = C_i^{-1} x_i$ has the same spherical distribution G for all $1 \leqslant i \leqslant k$, if $\sigma(G_0) = 1$, where G_0 is the distribution of z_{11}; that is, $\beta_\sigma(F) = \beta$ and $\lambda_{\sigma,ij}(F) = \lambda_{ij}$.

The following theorem gives the values of the partial influence functions for the projection estimators.

Theorem 2. *Let x_i be independent random vectors with ellipsoidal distribution F_i, with location parameters $\mu_i = 0_p$ and scatter matrices $\Sigma_i = C_i C_i'$ satisfying* (1) *and such* *that* $C_i^{-1}x_i = z_i$ *has the same spherical distribution* G *for all* $1 \leqslant i \leqslant k$. *Assume that* $\sigma(G_0) = 1$, *where* G_0 *is the distribution of* z_{11}, *and that* $\Sigma = \beta\, \mathrm{diag}(v_1, \ldots, v_p)\beta'$, *where* $v_1 > v_2 > \ldots > v_p$.

Then, if the function $(\varepsilon, y) \to \sigma\{(1-\varepsilon)G_0 + \varepsilon\delta_y\}$ *is twice continuously differentiable at* $(0, y)$, *we have that for any* x

$$\mathrm{PIF}_i(x, \lambda_{\sigma,lj}, F) = 2\delta_{li}\lambda_{ij}\, \mathrm{IF}\left(\frac{x'\beta_j}{\lambda_{ij}^{1/2}}, \sigma, G_0\right), \tag{11}$$

$$\begin{aligned}\mathrm{PIF}_i(x, \beta_{\sigma,j}, F) = {}& \tau_i\lambda_{ij}^{1/2}\, \mathrm{DIF}\left(\frac{x'\beta_j}{\lambda_{ij}^{1/2}}, \sigma, G_0\right) \sum_{s=j+1}^{p} \frac{1}{v_j - v_s}\beta_s(x'\beta_s) \\ & + \tau_i \sum_{s=1}^{j-1} \frac{1}{v_j - v_s}\beta_s\lambda_{is}^{1/2}\, \mathrm{DIF}\left(\frac{x'\beta_s}{\lambda_{is}^{1/2}}, \sigma, G_0\right)(x'\beta_j),\end{aligned} \tag{12}$$

$$\begin{aligned}\mathrm{PIF}_i(x, V_{\sigma,l}, F) = {}& 2\delta_{li} \sum_{j=1}^{p} \lambda_{ij}\, \mathrm{IF}\left(\frac{x'\beta_j}{\lambda_{ij}^{1/2}}, \sigma, G_0\right)\beta_j\beta_j' \\ & + \sum_{j=2}^{p}\sum_{s=1}^{j-1} \frac{\lambda_{lj} - \lambda_{ls}}{v_j - v_s}\tau_i\lambda_{is}^{1/2}\, \mathrm{DIF}\left(\frac{x'\beta_s}{\lambda_{is}^{1/2}}, \sigma, G_0\right)(x'\beta_j)(\beta_s\beta_j' + \beta_j\beta_s'),\end{aligned} \tag{13}$$

where $\mathrm{DIF}(y, \sigma, G)$ *denotes the derivative of* $\mathrm{IF}(y, \sigma, G)$ *with respect to* y.

Remark 2. As in the one-population setting note that, if we use a scale estimator with bounded influence, the eigenvalues will have bounded influence. However, the influence function for the eigenvectors may be unbounded, since the term $x'\beta_j$ will remain unbounded.

Remark 3. As in Croux and Ruiz-Gazen's technical report, one can consider the case when $\sigma^2(F) = \mathrm{var}(F)$. In our situation, this choice will not lead to the maximum likelihood estimators but to the eigenvalues and eigenvectors of the pooled matrix.

Since $\mathrm{IF}(y, \sigma, G) = \frac{1}{2}\{y^2 - \mathrm{var}(G)\}$, and if we assume that $\mathrm{var}(G_0) = 1$, Theorem 2 yields

$$\mathrm{PIF}_i(x, \lambda_{\mathrm{var},lj}, F) = \delta_{li}\{(x'\beta_j)^2 - \lambda_{ij}\}, \tag{14}$$

$$\mathrm{PIF}_i(x, \beta_{\mathrm{var},j}, F) = \tau_i \sum_{s \neq j} \frac{1}{v_j - v_s} x'\beta_s x'\beta_j \beta_s. \tag{15}$$

Plots for the influence functions of the eigenvalues and eigenvectors based on this scale estimator can be seen in Croux and Ruiz-Gazen's technical report. Moreover, the plots corresponding to the Q-dispersion measure are analogous to those shown therein.

The asymptotic variance of the projection-pursuit estimators of the common eigenvectors and of the eigenvalues can be obtained heuristically using (2).

Corollary 1. *Let* $x_{i1}, \ldots, x_{in_i}$ $(1 \leqslant i \leqslant k)$ *be independent observations from* k *independent samples with distributions* F_i, *location parameters* $\mu_i = 0_p$ *and scatter matrices* $\Sigma_i = \Lambda_i$; *that is* Σ_i *satisfies* (1) *with* $\beta = I_p$. *Assume that* $n_i = \tau_i N$, *with* $0 < \tau_i < 1$ *and* $\sum_{i=1}^{k}\tau_i = 1$. *Moreover, assume that* $\Lambda_i^{-1/2}x_{i1} = z_i$ *has the same spherical distribution* G *for all* $1 \leqslant i \leqslant k$. *Assume that* $\sigma(G_0) = 1$, *where* G_0 *is the distribution of* z_{11}, *and that* $\Sigma = \mathrm{diag}(v_1, \ldots, v_p)$, *where* $v_1 > v_2 > \ldots > v_p$. *Let* $s(.)$ *be the univariate robust scale statistic related to the scale functional* $\sigma(.)$ *and let* $X_i = (x_{i1}, \ldots, x_{in_i})$ *and* $\tau_i = n_i/N$ *for* $1 \leqslant i \leqslant k$.

Define the common principal axes by solving iteratively

$$r(\hat{\beta}_1) = \sup_{\|b\|=1} \sum_{i=1}^{k} \tau_i s^2(X_i' b), \quad r(\hat{\beta}_j) = \sup_{b \in \mathscr{B}_j} \sum_{i=1}^{k} \tau_i s^2(X_i' b) \quad (2 \leqslant j \leqslant p), \tag{16}$$

where $\mathscr{B}_j = \{b : \|b\| = 1, b'\hat{\beta}_m = 0$ *for* $1 \leqslant m \leqslant j-1\}$.

The estimators of the eigenvalues and the covariance matrix of the ith population are computed as

$$\hat{\lambda}_{ij} = s^2(X_i' \hat{\beta}_j) \quad (1 \leqslant j \leqslant p), \quad V_i = \sum_{j=1}^{p} \hat{\lambda}_{ij} \hat{\beta}_j \hat{\beta}_j'. \tag{17}$$

Then, if the function $(\varepsilon, y) \to \sigma\{(1-\varepsilon)G_0 + \varepsilon\delta_y\}$ *is twice continuously differentiable at* $(0, y)$, *the asymptotic variances of the projection-pursuit estimators solution of* (16) *are given by*

$$\operatorname{avar}(\hat{\lambda}_{lj}) = 4\lambda_{lj}^2 \frac{1}{\tau_l} \operatorname{avar}(\sigma, G_0),$$

$$\operatorname{avar}(\hat{\beta}_{jm}) = \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}\lambda_{im}}{(\nu_j - \nu_m)^2} E_G\{\text{DIF}(z_{1j}, \sigma, G_0) z_{1m}\}^2 \quad (m \neq j).$$

In particular, when $G = N(0_p, I_p)$, *we have that*

$$\operatorname{avar}(\hat{\lambda}_{lj}) = 4\lambda_{lj}^2 \frac{1}{\tau_l} \operatorname{avar}(\sigma, \Phi),$$

$$\operatorname{avar}(\hat{\beta}_{jm}) = \sum_{i=1}^{k} \frac{\lambda_{ij}\lambda_{im}}{(\nu_j - \nu_m)^2} E_\Phi\{\text{DIF}(Y, \sigma, \Phi)\}^2 \quad (m \neq j),$$

$$\operatorname{acov}(\hat{\beta}_{jm}, \hat{\beta}_{jr}) = 0 \quad (m \neq j, m \neq r, r \neq j).$$

Remark 4. Note that the asymptotic variances of the projection-pursuit estimators of the eigenvalues obtained using $\sigma^2(F) = \operatorname{var}(F)$ equal those of the maximum likelihood estimators. On the other hand, the asymptotic variances of the projection-pursuit estimators of the eigenvectors will not be those of the maximum likelihood estimators, but as expected they are those of the eigenvectors of the pooled matrix obtained in Boente & Orellana (2001).

3. Outlier detection

An immediate application of the influence functions is in the detection of influential/outlying observations. For the one-population case, Croux & Haesbroeck (1999) discussed the use of the empirical influence functions based on the sample covariance matrix considered by Critchley (1985) and Shi (1997) and that based on the one-step reweighted minimum covariance determinant estimator (Rousseeuw, 1985). As expected, the empirical influence function of the robust estimator hardly changes when contaminated data points are included in the sample, because outliers usually have small influence on robust estimators, while, if we consider the empirical influence of the classical estimators, a masking effect may appear and so outlying observations are not detected. An alternative approach is to consider a robust empirical influence function for the classical estimators, where the parameters are estimated through a robust procedure in order to avoid masking; see also Pison et al. (2000). This procedure is analogous to the use of the robustified version of Mahalanobis distance introduced by Rousseeuw & van Zomeren (1990).

Since there are a large number of influence functions, one for each parameter, a smaller number of aggregate measures is more suitable for analysis. We consider just two measures, one for the eigenvalues and one for the eigenvectors. We consider standardised robust empirical functions to avoid the problem of the different sizes of the eigenvalues and to be able to measure the absolute influence of each observation. For the case of the eigenvectors, the diagnostics must be invariant through orthogonal transformations. This can be achieved by transforming the problem to the diagonal case and by noting that, from equivariance, $\text{PIF}_i\{x, \beta(F), F\} = \beta\, \text{PIF}_i\{\beta' x, \beta(F), F_0\}$, where $F_0 = F_{1,0} \times \ldots \times F_{k,0}$, $F_{i,0}$ is the distribution of $\beta' y$ when $y \sim F_i$ and $\beta(F)$ is an equivariant eigenvector functional, such as $\beta_V(F)$ or $\beta_\sigma(F)$. These considerations lead to the following definition.

Given an observation x from the ith population, define

$$\text{IML}_i(x, \beta, \lambda) = \left\{ \sum_{r=1}^{p} \frac{\text{PIF}_i(x, \lambda_{S,ir}, F)^2}{v_{ir}(\beta, \lambda)} \right\}^{\frac{1}{2}},$$

$$\text{IMB}_i(x, \beta, \lambda) = \left[\sum_{r=1}^{p} \{\text{PIF}_i(\beta' x, \beta_{S,r}^{(r)}, F_0)\}' A_{ir}^{-1}(\beta, \lambda) \{\text{PIF}_i(\beta' x, \beta_{S,r}^{(r)}, F_0)\} \right]^{\frac{1}{2}},$$

where $\lambda_{s,il}$ and $\beta_{S,r}$ denote the classical functional estimators, $z^{(r)}$ the vector z without the rth component, β and $\lambda = (\lambda_{11}, \ldots, \lambda_{1p}, \ldots, \lambda_{k1}, \ldots, \lambda_{kp})'$ are the unknown parameters and

$$A_{ir}(\beta, \lambda) = E_{F_i}\{\text{PIF}_i(u, \beta_{S,r}^{(r)}, F_0)\, \text{PIF}_i(u, \beta_{S,r}^{(r)}, F_0)'\},$$

$$v_{ir}(\beta, \lambda) = E_{F_i}\{\text{PIF}_i(u, \lambda_{S,ir}, F)\}^2.$$

The rth coordinate is not included in the expressions for $\text{IMB}_i(x, \beta, \lambda)$, since both its partial influence function and its asymptotic variance are equal to zero when transforming the data to the diagonal case.

The outlier detection measures are now defined as $\text{IML}_i(x, \hat{\beta}, \hat{\lambda})$ and $\text{IMB}_i(x, \hat{\beta}, \hat{\lambda})$, where the 'hat' denotes replacement of the unknown parameters by their robust estimators. Our proposal is analogous to the one considered by Pison et al. (2000) for principal factor analysis in order to avoid the masking effect. As those authors mentioned, if one computes $\text{IML}_i(x, \beta, \lambda)$ and $\text{IMB}_i(x, \beta, \lambda)$ using the partial influence function of a robust functional and then the diagnostic measures $\text{IML}_i(x, \hat{\beta}, \hat{\lambda})$ and $\text{IMB}_i(x, \hat{\beta}, \hat{\lambda})$ at robust estimators, one will not achieve the desired property of detecting influential points.

Another approach could be to consider the partial influence functions of the projection-pursuit estimators obtained by using the sample variance. However, from (9) and (10), and (14) and (15), we see that both expressions are equivalent.

When $F_{i,0} = N(0_p, \Lambda_i)$ from (9) and (10) the expressions for the diagnostics simplify to

$$\text{IML}_i(x, \hat{\beta}, \hat{\lambda}) = \left[\sum_{r=1}^{p} \frac{\{(\hat{\beta}_r' x)^2 - \hat{\lambda}_{ir}\}^2}{2\hat{\lambda}_{ir}^2} \right]^{\frac{1}{2}},$$

$$\text{IMB}_i(x, \hat{\beta}, \hat{\lambda}) = \left[\sum_{r=1}^{p} \sum_{s \neq r} \frac{\{(\hat{\beta}_r' x)(\hat{\beta}_s' x)\}^2}{\hat{\lambda}_{ir} \hat{\lambda}_{is}} \right]^{\frac{1}{2}},$$

where $\hat{\beta}$ and $\hat{\lambda}_{ij}$ are the estimators derived from (3) for the plug-in proposal or through (16) and (17) for the projection-pursuit approach. Note that both IMB and IML are simple functions of the standardised robust scores. The advantage of IML and IMB with respect to the separate influence plots for each parameter is that the number of plots is reduced from $\{p + \frac{1}{2}p(p-1)\}k$ to $2k$.

In order to detect influential observations, we must compare the observed values of IML and IMB with the high percentiles of

$$G_\lambda = \left\{ \sum_{r=1}^{p} \frac{(z_r^2 - 1)^2}{2} \right\}^{\frac{1}{2}}, \quad G_\beta = \left(\sum_{r=1}^{p} \sum_{s \neq r} z_r^2 z_s^2 \right)^{\frac{1}{2}} = \left\{ \left(\sum_{r=1}^{p} z_r^2 \right)^2 - \sum_{r=1}^{p} z_r^4 \right\}^{\frac{1}{2}},$$

respectively, where $z_1, \ldots, z_p$ are independent and identically distributed $N(0, 1)$ random variables.

Table 1 gives estimates for the 95%, 97·5% and 99% points of G_λ and G_β, denoted by $\text{IML}_{0\cdot95}$, $\text{IML}_{0\cdot975}$, $\text{IML}_{0\cdot99}$, $\text{IMB}_{0\cdot95}$, $\text{IMB}_{0\cdot975}$ and $\text{IMB}_{0\cdot99}$, respectively, and obtained through a simulation study. For each dimension p, we have performed 100 replications generating samples of size $m = 10\,000$. The values tabulated are the medians over the 100 replications of the corresponding empirical percentiles for $G_{\lambda,1}, \ldots, G_{\lambda,m}$ and $G_{\beta,1}, \ldots, G_{\beta,m}$ with $m = 10\,000$.

Table 1. *Estimates for the* 95%, 97·5% *and* 99% *points of* G_λ *and* G_β

	$p=2$	$p=3$	$p=5$	$p=10$		$p=2$	$p=3$	$p=5$	$p=10$
$\text{IML}_{0\cdot95}$	2·927	3·514	4·316	5·604	$\text{IMB}_{0\cdot95}$	3·088	5·058	8·463	15·910
$\text{IML}_{0\cdot975}$	3·788	4·381	5·215	6·494	$\text{IMB}_{0\cdot975}$	3·933	6·131	9·846	17·190
$\text{IML}_{0\cdot99}$	4·974	5·558	6·375	7·637	$\text{IMB}_{0\cdot99}$	5·088	7·535	11·630	20·190

Note that, since the principal components model is a special case of the common principal components model with $k = 1$, the diagnostic measures proposed here can be used in the usual one-population setting and deserve to be compared with other measures appearing in the literature, but this is beyond the scope of the paper. To illustrate the different influential observations that are detected by the three measures, Fig. 1 shows for $p = 2$ the outlier detection regions defined using the Mahalanobis distance with $\Sigma = \text{diag}(4, 1)$, $\text{IML}_1(x, \beta, \lambda)$ and $\text{IMB}_1(x, \beta, \lambda)$ with $\beta = I_2$ and $\lambda = (4, 1)$, together with 500 normally distributed observations with mean $(0, 0)'$ and $\Sigma = \text{diag}(4, 1)$. Ninety-five percent detection limits are plotted as solid curves while, for IML_1 and IMB_1, 97·5% detection limits are shown dashed. Note that a fairer comparison, using Bonferroni, should use the 97·5% quantile for IML_1 and IMB_1 and the 95% quantile for the Mahalanobis distance.

4. An example

We consider a dataset, with $k = 2$ and five variables, which is part of a larger dataset described in a 1995 Master's thesis from the University of Lisbon by I. Oliveira, who performed a principal component analysis. The data correspond to two varieties, *lada* and *longal*, of chestnut tree leaves of the genus *Castanea*. The sample sizes were $n_1 = 100$ and $n_2 = 47$ and the variables were x_1, the petiole length in mm, x_2, the number of nervures from the right-hand side of the leaf, x_3, the number of nervures from the left-hand side of the leaf, x_4, the number of teeth from the right-hand side of the leaf and x_5, the number of teeth from the left-hand side of the leaf.

The robust principal component analysis of each variety showed similar principal axes with different amounts of variability. Therefore, a common principal components model was judged adequate. This conclusion does not hold with the classical principal component

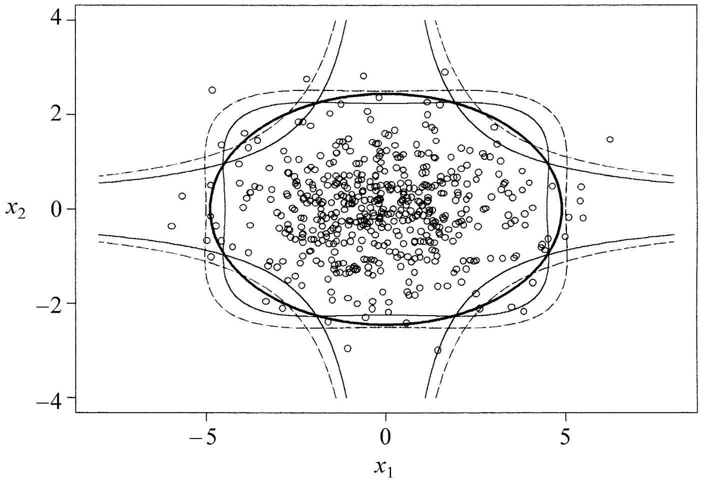

Fig. 1. Detection regions obtained with Mahalanobis distance (ellipse), IML$_1$ (closed curves) and IMB$_1$ (open curves). The 95% detection limits are the solid curves and the 97·5% detection limits are the dashed curves.

analysis, which is not surprising, since several outliers were detected in the boxplots for each variable.

The robust common principal components obtained using plug-in estimates with the reweighted minimum covariance determinant scatter matrices, with $h = [0{\cdot}75n]$, are given in Table 2 together with the classical ones and the projection-pursuit estimates, based on two different scale estimates, the median of the absolute deviations and an M-estimate, and calculated using an algorithm similar to that considered by Croux & Ruiz-Gazen (1996).

As expected there are differences between the classical and the robust estimates, the most noticeable being in the eigenvalues of the second group. Note that the projection-pursuit estimates are quite different from the robust plug-in ones, because there are inliers in the projected observations. On the other hand, the projection-pursuit estimates obtained by minimising the scale measure give similar results to those of the robust plug-in, and are therefore not reported here.

Influence plots with the two proposed measures show that there are several influential observations in each group, but some differences appear among estimation procedures. The labels of the observations detected as possible outliers by IML and IMB for the plug-in and the two projection-pursuit methods are given in Table 3, at the 95%, 97·5% and 99% points. In Table 3 we also report the labels of the influential observations obtained by three other methods, the robustified Mahalanobis distance (Rousseeuw & van Zomeren, 1990), classical versions of IML and IMB and classical Mahalanobis distance. The labels are ordered by increasing values of the corresponding measure and we separated by | | those between the percentiles mentioned above.

The masking effect in the classical methods is confirmed by the above results; only six observations are detected in the first group and one in the second, and this does not have the highest robust Mahalanobis distance value. When comparing the proposed measures IML and IMB using a plug-in procedure with the robust Mahalanobis distance, we see that the highest influential observations are detected by both methods. There are however

Table 2. *Plug-in*, PI, *and projection-pursuit*, PP, *eigenvalues and eigenvectors estimates for the chestnut trees, Castanea lada and longal, leaves data*

(*a*) *Plug-in estimates*

	CPC_1	CPC_2	CPC_3	CPC_4	CPC_5	CPC_1	CPC_2	CPC_3	CPC_4	CPC_5
	Eigenvalues by PI (RMCD matrices)					Eigenvalues by PI (Classical matrices)				
lada	26·904	7·392	1·993	1·949	0·637	26·968	11·138	2·047	2·362	0·876
longal	65·318	20·391	2·734	0·588	0·496	54·673	47·508	3·665	1·548	0·552
	Eigenvectors by PI (RMCD matrices)					Eigenvectors by PI (Classical matrices)				
x_1	0·661	0·750	0·022	0·012	−0·002	0·579	0·812	−0·039	−0·051	−0·032
x_2	−0·463	0·429	−0·444	−0·409	0·488	−0·497	0·329	−0·492	−0·348	0·530
x_3	−0·417	0·355	0·585	−0·359	−0·478	−0·454	0·293	0·442	−0·493	−0·520
x_4	−0·325	0·290	−0·469	0·576	−0·508	−0·355	0·250	−0·441	0·607	−0·499
x_5	−0·262	0·208	0·491	0·610	0·525	−0·295	0·288	0·606	0·515	0·445

(*b*) *Projection-pursuit estimates*

	CPC_1	CPC_2	CPC_3	CPC_4	CPC_5	CPC_1	CPC_2	CPC_3	CPC_4	CPC_5
	Eigenvalues by PP (MAD)					Eigenvalues by PP (*M*-estimates)				
lada	33·405	12·673	2·584	2·357	2·066	32·547	10·358	2·416	2·351	0·966
longal	102·388	29·096	7·969	3·844	1·145	88·693	21·916	3·307	1·832	1·019
	Eigenvectors by PP (MAD)					Eigenvectors by PP (*M*-estimates)				
x_1	0·903	0·269	0·332	0·033	−0·019	0·886	0·451	0·085	−0·061	−0·027
x_2	−0·177	0·701	−0·012	−0·686	0·083	−0·299	0·612	0·344	0·457	0·458
x_3	−0·189	0·453	0·064	0·418	−0·762	−0·289	0·381	0·385	−0·754	−0·235
x_4	−0·341	0·114	0·833	0·233	0·351	−0·182	0·425	−0·366	0·328	−0·739
x_5	0·013	0·468	−0·438	0·547	0·538	−0·095	0·311	−0·769	−0·335	0·435

RMCD, reweighted minimum covariance determinant; MAD, median of the absolute deviations; CPC_j, jth common principal direction.

Table 3. *Labels of the observations detected as possible outliers in the chestnut trees, Castanea lada and longal, leaves data*

Method	*lada*	*longal*
IML-PI	29 | 94 79 77 | 24 12 87 88 30	29 33 | 14 | 34 42 39 44 32 22 1 6 24
IMB-PI	98 72 45 85 94 12 | 88 | 30	* | 32 | 42 14 34 44 39 6 22 24 1
IML-PP (MAD)	45 | 30 87 | 88	6 | * | 44 22 1 24
IMB-PP (MAD)	85 100 72 93 45 | 94 88 87 | 30	* | * | 22 6 1 24
IML-PP (*M*-est)	18 79 77 45 12 | * | 87 88 30	10 21 | 7 | 24 14 22 1
IMB-PP (*M*-est)	99 93 88 | 85 | 30	45 39 42 | 24 | 14 22 1
RMD	79 45 98 | 94 | 12 87 88 30	* | * | 42 14 34 32 44 39 6 22 24 1
Classical IML	24 79 | 12 | 87 88 30	32 | * | *
Classical IMB	* | * | 30	24 | * | *
MD	87 | 88 | 30	24 | * | *

PI, plug-in estimates; PP, projection-pursuit estimates; MAD, median of the absolute deviations; RMD, robust Mahalanobis distance; MD, Mahalanobis distance.
* no observation in the relevant range.

some discrepancies for the observations near the detection limit, which is to be expected since (IML, IMB) and the robust Mahalanobis distance are measuring different effects. The influence measures IML and IMB using projection-pursuit estimates show some differences in outlier detection because of the presence of inliers as mentioned above. However, most of the observations with higher values of the robust Mahalanobis distance, and of the influence measures IML and IMB computed with the plug-in estimates, are also detected with the projection-pursuit procedure.

5. Discussion

Detecting outlying observations is an important step in any analysis, even when robust estimates are used. The classical diagnostic measure based on the Mahalanobis distance focuses on detecting outlying data with respect to the confidence ellipsoid, while our proposal tends to discover anomalous data with respect to the estimation of the principal axes and their sizes, that is, for the spectral decomposition of the scatter matrices, instead of the matrices themselves. It is worth noticing that the proposed summary diagnostic measures can also be applied when dealing with just one multivariate population which will lead to a reduced number of comparisons.

For theoretical reasons and based on previous studies, we recommend the diagnostic measures based on projection-pursuit estimators obtained through an M-estimator of scale with a differentiable score function. This can be justified by the fact that estimates based on the median of the absolute deviations would be expected to have a lower convergence rate just as least median of squares does in regression. The use of an M-estimator of scale such as the one considered combines reasonable efficiency with a good breakdown point. As noted by Boente & Orellana (2001), plug-in estimators are quite sensitive to contamination in the direction of the lower eigenvalues. For this reason, projection-pursuit procedures should be preferred. Also, as is well known, projection-pursuit methods can be applied even when dealing with more variables than observations. In this situation, plug-in estimators cannot be applied. Another advantage of projection-pursuit is that the common directions are obtained consecutively, and thus, if the goal is to obtain dimension-reduction with just a few components, computation time can be saved.

Acknowledgement

This research was partially supported by grants from the Universidad de Buenos Aires, from the Consejo Nacional de Investigaciones Científicas y Técnicas and from the Agencia Nacional de Promoción Científica y Tecnológica in Buenos Aires, Argentina, and by the Center for Mathematics and its Applications, Lisbon, Portugal. This research was partially developed while Prof. Graciela Boente was visiting the Departamento de Matemática at the Instituto Superior Técnico. The research of Graciela Boente was also partially supported by a Guggenheim Fellowship. We thank Eng. M. J. Pimentel Pereira from the Department of Biological and Environmental Sciences, University of Trás-os-Montes and Alto Douro, Portugal, for authorising the use of the chestnut dataset. 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

Proofs

Proof of Theorem 1. The proof follows the same steps as those given in the proof of Lemma 3 in Croux & Haesbroeck (2000), which can be found in a Université Libre de Bruxelles technical report by those authors.

Let $F_{i,\varepsilon,x} = (1-\varepsilon)F_i + \varepsilon\delta_x$ and let $F_{\varepsilon,x,i} = F_1 \times \ldots \times F_{i-1} \times F_{i,\varepsilon,x} \times F_{i+1} \times \ldots \times F_k$. Let $\beta_{j,\varepsilon,i} = \beta_{V,j}(F_{\varepsilon,x,i})$, $\lambda_{lj,\varepsilon,i} = \lambda_{V,lj}(F_{\varepsilon,x,i})$, $V_{i,\varepsilon} = V_i(F_{i,\varepsilon,x})$ and $V_l = V_l(F_l)$. Then we have that

$$\lambda_{ij,\varepsilon,i} = \beta'_{j,\varepsilon,i} V_{i,\varepsilon} \beta_{j,\varepsilon,i}, \tag{A1}$$

$$\lambda_{lj,\varepsilon,i} = \beta'_{j,\varepsilon,i} V_l \beta_{j,\varepsilon,i} \quad (l \neq i), \tag{A2}$$

$$\begin{aligned} 0 = \beta'_{m,\varepsilon,i} &\left(\sum_{l=1, l\neq i}^{k} \tau_l \frac{\lambda_{lm,\varepsilon,i} - \lambda_{lj,\varepsilon,i}}{\lambda_{lm,\varepsilon,i}\lambda_{lj,\varepsilon,i}} V_l \right) \beta_{j,\varepsilon,i} \\ &+ \beta'_{m,\varepsilon,i} \left(\tau_i \frac{\lambda_{im,\varepsilon,i} - \lambda_{ij,\varepsilon,i}}{\lambda_{im,\varepsilon,i}\lambda_{ij,\varepsilon,i}} V_{i,\varepsilon} \right) \beta_{j,\varepsilon,i} \quad (m \neq j), \end{aligned} \tag{A3}$$

$$\delta_{mj} = \beta'_{m,\varepsilon,i} \beta_{j,\varepsilon,i}. \tag{A4}$$

Therefore, differentiating (A4) with respect to ε, we obtain that

$$\text{PIF}_i(x, \beta_{V,m}, F)'\beta_m = 0, \tag{A5}$$

$$\text{PIF}_i(x, \beta_{V,m}, F)'\beta_j + \text{PIF}_i(x, \beta_{V,j}, F)'\beta_m = 0. \tag{A6}$$

Differentiating (A1) and (A2), using (A5) and the fact that $V_l \beta_j = \Sigma_l \beta_j = \lambda_{lj}\beta_j$, we easily obtain (7).

It remains to show (8). Differentiating (A3) leads to

$$\begin{aligned} 0 = \text{PIF}_i(x, \beta_{V,m}, F)' &\left(\sum_{l=1}^{k} \tau_l \frac{\lambda_{lm} - \lambda_{lj}}{\lambda_{lm}\lambda_{lj}} V_l \right) \beta_j + \beta'_m \left(\sum_{l=1}^{k} \tau_l \frac{\lambda_{lm} - \lambda_{lj}}{\lambda_{lm}\lambda_{lj}} V_l \right) \text{PIF}_i(x, \beta_{V,j}, F) \\ &+ \beta'_m \left\{ \sum_{l=1}^{k} \tau_l \frac{\partial}{\partial \varepsilon} \left(\frac{\lambda_{lm,\varepsilon,i} - \lambda_{lj,\varepsilon,i}}{\lambda_{lm,\varepsilon,i}\lambda_{lj,\varepsilon,i}} \right) \Bigg|_{\varepsilon=0} V_l + \tau_i \frac{\lambda_{im} - \lambda_{ij}}{\lambda_{im}\lambda_{ij}} \text{IF}(x, V_i, F_i) \right\} \beta_j \quad (m \neq j). \end{aligned} \tag{A7}$$

Using again the fact that $\Sigma_l \beta_j = \lambda_{lj}\beta_j$ in (A7), the orthogonality condition $\beta'_m \beta_j = 0$ for $m \neq j$ and (A6), we obtain, after some algebra,

$$\text{PIF}_i(x, \beta_{V,j}, F)'\beta_m = \tau_i \frac{\lambda_{ij} - \lambda_{im}}{\lambda_{im}\lambda_{ij}} \left\{ \sum_{l=1}^{k} \tau_l \frac{(\lambda_{lm} - \lambda_{lj})^2}{\lambda_{lm}\lambda_{lj}} \right\}^{-1} \{\beta'_j \,\text{IF}(x, V_i, F_i)\beta_m\} \quad (m \neq j).$$

□

Proof of Theorem 2. The proof follows the ideas given in C. Croux and A. Ruiz-Gazen's technical report. Let $F_{\varepsilon,x,i} = F_1 \times \ldots \times F_{i-1} \times F_{i,\varepsilon,x} \times F_{i+1} \times \ldots \times F_k$, where $F_{i,\varepsilon,x} = (1-\varepsilon)F_i + \varepsilon\delta_x$. Let $\beta_{j,\varepsilon,i} = \beta_{\sigma,j}(F_{\varepsilon,x,i})$, $\lambda_{lj,\varepsilon,i} = \sigma^2(F_l[\beta_{j,\varepsilon,i}])$, for $l \neq i$, and $\lambda_{ij,\varepsilon,i} = \sigma^2(F_{i,\varepsilon,x}[\beta_{j,\varepsilon,i}])$. Let $V_{l,\varepsilon,i} = V_{\sigma,l}(F_{\varepsilon,x,i}) = \sum_{j=1}^{p} \lambda_{lj,\varepsilon,i} \beta_{j,\varepsilon,i} \beta'_{j,\varepsilon,i}$. Now, $\beta_{j,\varepsilon,i}$ maximises $\varsigma(F_{\varepsilon,x,i}[b])$ under the constraints $\beta'_{j,\varepsilon,i}\beta_{j,\varepsilon,i} = 1$ and $\beta'_{s,\varepsilon,i}\beta_{j,\varepsilon,i} = 0$ for $1 \leqslant s \leqslant j-1$. Therefore, $\beta_{j,\varepsilon,i}$ maximises

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

and so it should satisfy

$$0 = \frac{\partial}{\partial b} L(b, \gamma, \alpha) \Bigg|_{b = \beta_{j,\varepsilon,i}} = \psi(\varepsilon) - 2\gamma\beta_{j,\varepsilon,i} - \sum_{s=1}^{j-1} \alpha_s \beta_{s,\varepsilon,i}, \tag{A8}$$

with

$$\psi(\varepsilon) = \tau_i \frac{\partial}{\partial b} \sigma^2(F_{i,\varepsilon,x}[b]) \Bigg|_{b=\beta_{j,\varepsilon,i}} + \sum_{i_0 \neq i} \tau_{i_0} \frac{\partial}{\partial b} \sigma^2(F_{i_0}[b]) \Bigg|_{b=\beta_{j,\varepsilon,i}}. \tag{A9}$$

Since $\beta'_{j,\varepsilon,i}\beta_{j,\varepsilon,i}=1$ and $\beta'_{s,\varepsilon,i}\beta_{j,\varepsilon,i}=0$ for $1\leqslant s\leqslant j-1$ we have that $\psi(\varepsilon)'\beta_{j,\varepsilon,i}=2\gamma$, $\psi(\varepsilon)'\beta_{s,\varepsilon,i}=\alpha_s$, for $1\leqslant s\leqslant j-1$. Using this in (A8) and differentiating with respect to ε, we obtain

$$\frac{\partial}{\partial\varepsilon}\psi(\varepsilon)\bigg|_{\varepsilon=0}=\sum_{s=1}^{j}\left[\{\psi(0)'\,\text{PIF}_i(x,\beta_{\sigma,s},F)\}\beta_s+\left\{\beta_s'\frac{\partial}{\partial\varepsilon}\psi(\varepsilon)\bigg|_{\varepsilon=0}\right\}\beta_s+\{\psi(0)'\beta_s\}\,\text{PIF}_i(x,\beta_{\sigma,s},F)\right]. \tag{A10}$$

Since F_i is an elliptical distribution and $\sigma(G_0)=1$, using the fact that $\sigma^2(F_i[b])=b'\Sigma_i b_i$ we obtain $\psi(0)=2v_j\beta_j$, which implies that $\psi(0)'\beta_s=0$ for $1\leqslant s\leqslant j-1$. Write $P_{j+1}=I_p-\sum_{s=1}^{j}\beta_s\beta_s'$. Then (A10) can be written as

$$P_{j+1}\frac{\partial}{\partial\varepsilon}\psi(\varepsilon)\bigg|_{\varepsilon=0}=2v_j\sum_{s=1}^{j}\beta_j'\,\text{PIF}_i(x,\beta_{\sigma,s},F)\beta_s+2v_j\,\text{PIF}_i(x,\beta_{\sigma,j},F). \tag{A11}$$

On the other hand, from (A9) and since $\varsigma(F[b])=b'\Sigma b$, we have that

$$\frac{\partial}{\partial\varepsilon}\psi(\varepsilon)\bigg|_{\varepsilon=0}=2\Sigma\,\text{PIF}_i(x,\beta_{\sigma,j},F)+\tau_i\frac{\partial}{\partial b}\text{IF}(b'x,\sigma^2,F_i[b])\bigg|_{b=\beta_j}. \tag{A12}$$

Again from the equivariance of the scale estimator, we have that

$$\frac{\partial}{\partial b}\text{IF}(b'x,\sigma^2,F_i[b])|_{b=\beta_j}=2\lambda_{ij}\beta_j\,\text{IF}\left(\frac{\beta_j'x}{\lambda_{ij}^{1/2}},\sigma^2,G_0\right)+\lambda_{ij}\,\text{DIF}\left(\frac{\beta_j'x}{\lambda_{ij}^{1/2}},\sigma^2,G_0\right)\left(\frac{x}{\lambda_{ij}^{1/2}}-\frac{\beta_j'x}{\lambda_{ij}^{1/2}}\beta_j\right). \tag{A13}$$

From (A11), (A12) and (A13) and using that $\text{PIF}_i(x,\beta_{\sigma,j},F)'\beta_j=0$, we obtain

$$2(P_{j+1}\Sigma-v_jI_p)\,\text{PIF}_i(x,\beta_{\sigma,j},F)=2v_j\sum_{s=1}^{j-1}\beta_j'\,\text{PIF}_i(x,\beta_{\sigma,s},F)\beta_s-\tau_i\lambda_{ij}^{1/2}\,\text{DIF}\left(\frac{\beta_j'x}{\lambda_{ij}^{1/2}},\sigma^2,G_0\right)P_{j+1}x. \tag{A14}$$

The matrix $P_{j+1}\Sigma-v_jI_p=\sum_{s=j+1}^{p}v_s\beta_s\beta_s'-v_jI_p$ is a full-rank matrix with inverse

$$(P_{j+1}\Sigma-v_jI_p)^{-1}=\sum_{s=j+1}^{p}\frac{1}{v_s-v_j}\beta_s\beta_s'-\sum_{s=1}^{j}\frac{1}{v_j}\beta_s\beta_s',$$

so that

$$(P_{j+1}\Sigma-v_jI_p)^{-1}\beta_s=-\frac{1}{v_j}\beta_s\quad(1\leqslant s\leqslant j-1),\quad (P_{j+1}\Sigma-v_jI_p)^{-1}P_{j+1}=\sum_{s=j+1}^{p}\frac{1}{v_s-v_j}\beta_s\beta_s'.$$

Thus, from (A14), after some calculations, we obtain for any $s\geqslant j+1$ that

$$\text{PIF}_i(x,\beta_{\sigma,j},F)'\beta_s=\frac{1}{2(v_j-v_s)}\tau_i\lambda_{ij}^{1/2}\,\text{DIF}\left(\frac{\beta_j'x}{\lambda_{ij}^{1/2}},\sigma^2,G_0\right)\beta_s'x,$$

which implies (12), using the fact that $\text{IF}(y,\sigma^2,G_0)=2\,\text{IF}(y,\sigma,G_0)$.

Since $\lambda_{ij,\varepsilon,i}=\sigma^2(F_{i,\varepsilon,x}[\beta_{j,\varepsilon,i}])$ and, for $l\neq i$, $\lambda_{lj,\varepsilon,i}=\sigma^2(F_l[\beta_{j,\varepsilon,i}])=\beta'_{j,\varepsilon,i}\Sigma_l\beta_{j,\varepsilon,i}$, the chain rule easily yields (11). Since $V_{l,\varepsilon,i}=V_{\sigma,l}(F_{\varepsilon,x,i})=\sum_{j=1}^{p}\lambda_{lj,\varepsilon,i}\beta_{j,\varepsilon,i}\beta'_{j,\varepsilon,i}$, and with the help of the expressions for the partial influence functions derived for the eigenvectors and eigenvalues, straightforward calculations lead to (13). □

Proof of Corollary 1. From (2), using (11) and the fact that $x'\beta_j/\lambda_{ij}^{1/2}\sim G_0$ when $x\sim F_i$, we obtain

$$\text{avar}(\hat{\lambda}_{lj})=\sum_{i=1}^{k}\frac{1}{\tau_i}E_{F_i}\left[\left\{\delta_{li}2\lambda_{ij}\,\text{IF}\left(\frac{x'\beta_j}{\lambda_{ij}^{1/2}},\sigma,G_0\right)\right\}^2\right]=4\lambda_{lj}^2\frac{1}{\tau_l}\text{avar}(\sigma,G_0).$$

From (12) and since $\beta = I_p$, we obtain

$$\begin{aligned}\text{avar}(\hat{\beta}_{jm}) &= \delta_{m>j} \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}\lambda_{im}}{(v_j - v_m)^2} E_G\{\text{DIF}(z_{1j}, \sigma, G_0) z_{1m}\}^2 \\ &\quad + \delta_{m<j} \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}\lambda_{im}}{(v_j - v_m)^2} E_G\{\text{DIF}(z_{1m}, \sigma, G_0) z_{1j}\}^2 \\ &= \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}\lambda_{im}}{(v_j - v_m)^2} E_G\{\text{DIF}(z_{1j}, \sigma, G_0) z_{1m}\}^2,\end{aligned}$$

where $\delta_{m>j} = 0$ if $m \leqslant j$, and $\delta_{m>j} = 1$ if $m > j$.

Moreover, when $G = N(0_p, I_p)$, since z_{1j}, z_{1m} and z_{1r} are independent for $j \neq r$, $j \neq m$ and $m \neq r$ and $E_\Phi\{\text{DIF}(Y, \sigma, \Phi)\} = 0$, after straightforward calculations, we have that

$$\begin{aligned}\text{acov}(\hat{\beta}_{jm}, \hat{\beta}_{jr}) &= \delta_{m>j}\delta_{r>j} \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}(\lambda_{im}\lambda_{ir})^{\frac{1}{2}}}{(v_j - v_m)(v_j - v_r)} E_G[\{\text{DIF}(z_{1j}, \sigma, G_0)\}^2 z_{1m} z_{1r}] \\ &\quad + \delta_{m<j}\delta_{r<j} \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}(\lambda_{im}\lambda_{ir})^{\frac{1}{2}}}{(v_j - v_m)(v_j - v_r)} E_G\{\text{DIF}(z_{1m}, \sigma, G_0)\, \text{DIF}(z_{1r}, \sigma, G_0)(z_{1j})^2\} \\ &\quad + \delta_{m>j}\delta_{r<j} \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}(\lambda_{im}\lambda_{ir})^{\frac{1}{2}}}{(v_j - v_m)(v_j - v_r)} E_G\{\text{DIF}(z_{1j}, \sigma, G_0)\, \text{DIF}(z_{1r}, \sigma, G_0) z_{1m} z_{1j}\} \\ &\quad + \delta_{m<j}\delta_{r>j} \sum_{i=1}^{k} \tau_i \frac{\lambda_{ij}(\lambda_{im}\lambda_{ir})^{\frac{1}{2}}}{(v_j - v_m)(v_j - v_r)} E_G\{\text{DIF}(z_{1j}, \sigma, G_0)\, \text{DIF}(z_{1m}, \sigma, G_0) z_{1r} z_{1j}\},\end{aligned}$$

and thus $\text{acov}(\hat{\beta}_{jm}, \hat{\beta}_{jr}) = 0$, which concludes the proof. □

REFERENCES

ARNOLD, S. J. & PHILLIPS, P. C. (1999). Hierarchical comparison of genetic variance-covariance matrices. II. Coastal-inland divergence in the garter snake *Thamnophis elegans*. *Evolution* **53**, 1516–27.

BOENTE, G. & ORELLANA, L. (2001). A robust approach to common principal components. In *Statistics in Genetics and in the Environmental Sciences*, Ed. L. T. Fernholz, S. Morgenthaler and W. Stahel, pp. 117–47. Basel: Birkhauser.

CRITCHLEY, F. (1985). Influence in principal components analysis. *Biometrika* **72**, 627–36.

CROUX, C. & HAESBROECK, G. (1999). Empirical influence functions for robust principal component analysis. In *Proc. Statist. Comp. Sect., Am. Statist. Assoc.*, pp. 201–6. Alexandria, VA: American Statistical Association.

CROUX, C. & HAESBROECK, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika* **87**, 603–18.

CROUX, C. & RUIZ-GAZEN, A. (1996). A fast algorithm for robust principal components based on projection-pursuit. In *Compstat: Proceedings in Computational Statistics*, Ed. A. Prat, pp. 211–7. Heidelberg; Physica-Verlag.

FLURY, B. K. (1984). Common principal components in *k* groups. *J. Am. Statist. Assoc.* **79**, 892–8.

FLURY, B. K. (1986). Asymptotic theory for common principal components. *Ann. Statist.* **14**, 418–30.

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

JAUPI, L. & SAPORTA, G. (1993). Using the influence function in robust principal components analysis. In *New Directions in Statistical Data Analysis and Robustness*, Ed. S. Morgenthaler, E. Ronchetti and W. Stahel, pp. 147–56. Basel: Birkhauser.

KLINGERBERG, C. P. & NEUENSCHWANDER, B. E. (1996). Ontogeny and individual variation: analysis of patterned covariance matrices with common principal components. *Syst. Biol.* **45**, 135–7.

NOVI INVERARDI, P. L. & FLURY, B. (1992). Robust estimation of common principal components. *Quaderni Statist. Mat. Appl. Sci. Econ. Social* **14**, 49–79.

PHILLIPS, P. C. & ARNOLD, S. J. (1999). Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution* **53**, 1506–15.

PIRES, A. M. & BRANCO, J. (2002). Partial influence functions. *J. Mult. Anal.* To appear.

PISON, G., ROUSSEEUW, P. J., FILZMOSER, P. & CROUX, C. (2000). A robust version of principal factor analysis. In *Compstat: Proceedings in Computational Statistics*, Ed. J. Bethlehem and P. van der Heijden, pp. 385–90. Heidelberg: Physica-Verlag.

ROUSSEEUW, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications*, Vol. B, Ed. W. Grossmann, G. Pflug, I. Vincze and W. Wertz, pp. 283–97. Dordrecht: Reidel.
ROUSSEEUW, P. J. & VAN ZOMEREN, B. C. (1990). Unmasking multivariate outliers and leverage points. *J. Am. Statist. Assoc.* **85**, 633–9.
SHI, L. (1997). Local influence in principal components analysis. *Biometrika* **84**, 175–86.

[*Received August* 2001. *Revised February* 2002]