Statistics for Industry and Technology, 235–246
© 2004 Birkhäuser Verlag Basel/Switzerland

Robust Redundancy Analysis by Alternating Regression

M.R. Oliveira, J.A. Branco, C. Croux and P. Filzmoser

Abstract. Given two groups of variables redundancy analysis searches for linear combinations of variables in one group that maximizes the variance of the other group that is explained by the linear combination. The method is important as an alternative to canonical correlation analysis, and can be seen as an alternative to multivariate regression when there are collinearity problems in the dependent set of variables. Principal component analysis is itself a special case of redundancy analysis.

In this work we propose a new robust method to estimate the redundancy analysis parameters based on alternating regressions. These estimators are compared with the classical estimator as well as other robust estimators based on robust covariance matrices. The behavior of the proposed estimators is also investigated under large contamination by the analysis of the empirical breakdown point.

Mathematics Subject Classification (2000). 62F35; 62J99.

Keywords. Redundancy analysis, alternating regression, robust regression.

1. Introduction

The problem of finding relationships between groups of variables is central in multivariate analysis. A number of methods have been suggested to pursue this objective but canonical correlation analysis is by far the most commonly used. Given two sets of variables the goal of canonical correlation analysis is to construct pairs of canonical variates (linear combinations of the original variables, one in each set) such that they have maximum correlation. The correlations between canonical variates of the same pair are important to the study of the correlations between the two sets but they cannot be interpreted as the degree of relation between the sets of variables. In particular the squared canonical correlations represent the variance shared by the two canonical variates of the same pair but not the variance shared by the two sets of observed random variables.

To overcome the difficulty in using the squared canonical correlations as a measure of the shared variance between the two sets, a redundancy index was proposed by Stewart and Love (1968). This index is a measure of the proportion of the variance in one set $\boldsymbol{y} = (y_1, \ldots, y_q)^t$ (dependent or criterion set) that is accounted for by the other set $\boldsymbol{x} = (x_1, \ldots, x_p)^t$ (independent or predictor set). The redundancy analysis model, proposed by van den Wollenberg (1977), searches for the linear combination, $u_1 = \boldsymbol{\alpha}_1^t \boldsymbol{x}$ (the first redundancy variate), of the independent set that maximizes the redundancy index, R_y, defined as

$$R_y = \sum_{j=1}^{q} Corr(\boldsymbol{\alpha}^t \boldsymbol{x}, y_j)^2 / q, \tag{1.1}$$

under the restriction $Var(\boldsymbol{\alpha}_1^t \boldsymbol{x}) = 1$. The second redundancy variate, $u_2 = \boldsymbol{\alpha}_2^t \boldsymbol{x}$, is defined as the linear combination of the independent set, non-correlated with u_1, that maximizes R_y under the restriction $Var(\boldsymbol{\alpha}_2^t \boldsymbol{x}) = 1$. A maximum of $r = \min(p, q)$ redundancy variates can be sequentially defined, following the scheme as above. The objective of this paper is to provide robust estimators for $\boldsymbol{\alpha}$ and R_y. Robust estimation is particularly useful in the analysis of multivariate data where the presence of outlying observations is common.

In Section 2, the relationship between redundancy and canonical correlation analysis is highlighted. This will help to address the problem of robust estimation in redundancy analysis and in particular to introduce the robust method based on alternating regressions. The algorithm to estimate the redundancy variates by alternating regressions is also presented in Section 2, and in Section 3 a simulation study is developed in order to compare the performance of the proposed estimators with others based on robust covariance matrices. In the last section a brief discussion about the results obtained is made and links with canonical correlation analysis and suggestions for future work are discussed.

2. Robust Estimation

The classical solution to (1.1) comes down to an eigenvector/eigenvalue problem, as was shown by van den Wollenberg (1977). The coefficients $\boldsymbol{\alpha}$ are the eigenvectors of the matrix $R_{11}^{-1} R_{12} R_{21}$, where R_{ij} $(i, j = 1, 2)$ are the usual partition matrices of the correlation matrix associated with the two sets of variables. A simple approach to robust estimation is to robustify the correlation matrix and then apply the traditional methods of estimation. So given a robust estimate of the correlation matrix the eigenvectors of $R_{11}^{-1} R_{12} R_{21}$ are calculated in order to estimate the coefficients $\boldsymbol{\alpha}$. This approach will be considered in Section 3. For the estimation of the robust correlation matrix we will use the M-estimator (M) as outlined by Maronna (1976). As an alternative we will consider the minimum covariance determinant (MCD) estimator of Rousseeuw (1985).

The above approaches were studied for the first latent variate (redundancy variate) in Oliveira and Branco (2002). Additionally, a method based on robust al-

ternating regression was considered. This technique, originally suggested by Wold (1966), has recently been used in the context of robust factor analysis (Croux et al., 2003) and robust canonical correlation analysis (Branco et al., 2003). Like Tenenhaus (1998) has pointed out, the algorithm is quite similar to the one proposed for canonical correlation analysis, initially discussed by Wold (1966), Lyttkens (1972) and later on mentioned by Tenenhaus (1998). For higher-order redundancy variates, the algorithm is based on repeating the idea underlying the construction of the first redundancy variate in successive residual spaces. Since these spaces have reduced rank, problems occur in the regression procedure.

In this paper we will focus on estimating higher order redundancy variates by robust alternating regressions. This procedure has a main advantage over the approach based on robust correlation matrix estimation: While the latter method discards an outlying observation completely, robust alternating regression is still using the information of the non-outlying cells for parameter estimation. For this reason, the method based on robust alternating regressions can also deal with missing values (Croux et al., 2003).

The general idea of the algorithm takes advantage of the links between redundancy analysis and canonical correlation analysis. In the context of canonical correlation analysis, the redundancy index (1.1) can be written as in Rencher (1998) by:

$$R_y = \rho^2 \sum_{j=1}^{q} Corr(\boldsymbol{\beta}^t \boldsymbol{y}, y_j)^2/q, \tag{2.1}$$

where ρ is the canonical correlation coefficient and $\boldsymbol{\beta}$ the canonical coefficient associated with the $\boldsymbol{y}'s$. Looking at the redundancy coefficient from these two different perspectives, an alternating procedure based on regression models can be built. To clarify this idea, let us consider that an initial value $\boldsymbol{\beta}$ is given. If the redundancy index is written in the form (2.1), to maximize R_y we need to determine the vector $\boldsymbol{\alpha}$ that maximizes $\rho^2 = Corr(\boldsymbol{\alpha}^t \boldsymbol{x}, \boldsymbol{\beta}^t \boldsymbol{y})^2$. From standard results on multiple regression, it follows that $\boldsymbol{\alpha}$ is proportional to the regression coefficient $\boldsymbol{a}$ in the model

$$\boldsymbol{\beta}^t \boldsymbol{y} = \boldsymbol{a}^t \boldsymbol{x} + \gamma_1 + \varepsilon_1. \tag{2.2}$$

This step is analogous to the procedure followed in canonical correlation analysis (Branco et al., 2003).

Once $\boldsymbol{\alpha}$ has been obtained, the value $\boldsymbol{\beta}$ has to be updated. Taking into account that R_y is a measure of the proportion of the variance in set $\boldsymbol{y}$ explained by $\boldsymbol{\alpha}^t \boldsymbol{x}$, Tenenhaus (1998) suggested that $\boldsymbol{\beta}$ can be chosen proportional to $\boldsymbol{b} = (b_1, \ldots, b_q)^t$, where its components are the coefficients of the regression equations

$$y_j = b_j \boldsymbol{\alpha}^t \boldsymbol{x} + \gamma_{2j} + \varepsilon_{2j}, \tag{2.3}$$

that maximizes the variance of y_j explained by $\boldsymbol{\alpha}^t \boldsymbol{x}$, i.e., $\hat{b}_j$ is such that $Corr\big(y_j, \hat{b}_j \boldsymbol{\alpha}^t \boldsymbol{x}\big)^2$ is maximum. From regression standard results $Corr(y_j, \hat{y}_j)^2 =$

$Corr(y_j, \hat{b}_j \boldsymbol{\alpha}^t \boldsymbol{x})^2 = Corr(y_j, \boldsymbol{\alpha}^t \boldsymbol{x})^2$. Choosing $\hat{b}_j$ in s way leads to an updated value of β, that maximizes the variance of y_j explained by $\boldsymbol{\alpha}^t \boldsymbol{x}$. By (1.1)

$$\sum\nolimits_{j=1}^{q} Corr(\hat{y}_j, y_j)^2/q = R_y,$$

where R_y is the value of the redundancy index obtained in the previous step.

This scheme can be implemented using least squares regression, but it leads to non robust estimates. In order to robustify the parameters in redundancy analysis we can simply use robust regression estimators. In the case of least squares estimators convergence of the algorithm is fast. However, using a robust estimator yields more local minima, and less smooth objective functions, so higher risks of lack of convergence appear. One benefits from choosing weighted L_1 regression estimators since they have bounded influence functions and the algorithm is fast to compute, once the weights are properly calculated. These weights are defined smoothly according to an appropriate distance measure associated with each observation. Although, weighted L_1 regression is a good choice other robust regression estimators could be used.

In the next subsection we give a general description of the algorithm to estimate the redundancy variates by alternating regressions.

2.1. Algorithm

It is convenient to center the observations, so that the intercept terms, γ_1 and γ_{2j}, can be eliminated from the equations (2.2) and (2.3). The observations are centered using robust estimators of location. In the present work, the median of each variable, ($\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$), has been chosen for robust estimators of location. So, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ represent the data matrices of size $(n \times p)$ and $(n \times q)$, respectively, the centered data are

$$\boldsymbol{X}_0 = \boldsymbol{X} - \mathbf{1}_p \tilde{\boldsymbol{x}}^t \quad \text{and} \quad \boldsymbol{Y}_0 = \boldsymbol{Y} - \mathbf{1}_q \tilde{\boldsymbol{y}}^t,$$

where $\mathbf{1}_k$ is a k-vector of ones.

2.1.1. First Redundancy Variate. The algorithm starts by choosing an initial value, $\boldsymbol{\beta}^{(0)}$, followed by the estimation of the regression coefficients in (2.2).

Starting values: In the context of robust estimation, the starting values can be of crucial importance. For the classical version of the algorithm Tenenhaus (1998) suggested to use the vector $\boldsymbol{\beta}^{(0)} = (1, 0, \ldots, 0)^t$. However, we have built the starting value using the first robust principal component of $\boldsymbol{X}_0$, $\boldsymbol{z}_1$ (see e.g. Croux and Ruiz-Gazen, 1996). Let $\hat{b}_j^{(0)}$ $(j = 1, \ldots, q)$ be the estimated regression coefficient associated with the model

$$\boldsymbol{y}_{0j} = b_j \boldsymbol{z}_1 + \boldsymbol{\varepsilon}_{3j}, \tag{2.4}$$

where $\boldsymbol{y}_{0j}$ is the j^{th} column of $\boldsymbol{Y}_0$. The starting value is defined as

$$\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{b}}^{(0)} / ||\hat{\boldsymbol{b}}^{(0)}||, \tag{2.5}$$

where $\hat{\boldsymbol{b}}^{(0)} = (\hat{b}_1^{(0)}, \ldots, \hat{b}_q^{(0)})^t$, and the corresponding latent variate is $\boldsymbol{v}^{(0)} = \boldsymbol{Y}_0 \boldsymbol{\beta}^{(0)}$.

Step s: Given the values $\boldsymbol{\beta}^{(s-1)}$ and $\boldsymbol{v}^{(s-1)}$ (for $s > 1$), the $\hat{\boldsymbol{a}}^{(s)}$ is the estimated regression coefficient of the model

$$\boldsymbol{v}^{(s-1)} = \boldsymbol{X}_0 \boldsymbol{a}^{(s)} + \boldsymbol{\varepsilon}_4. \tag{2.6}$$

The estimated vector of redundancy coefficients is

$$\boldsymbol{\alpha}^{(s)} = \hat{\boldsymbol{a}}^{(s)} / ||\hat{\boldsymbol{a}}^{(s)}||, \tag{2.7}$$

and the associated redundancy variate is $\boldsymbol{u}^{(s)} = \boldsymbol{X}_0 \boldsymbol{\alpha}^{(s)}$.

Given these new estimated values, $\boldsymbol{\beta}^{(s-1)}$ and $\boldsymbol{v}^{(s-1)}$ have to be updated. Let $\hat{b}_j^{(s)}$ $(j = 1, \ldots, q)$ be the estimate of the regression coefficient associated with the model

$$\boldsymbol{y}_{0j} = b_j^{(s)} \boldsymbol{u}^{(s)} + \boldsymbol{\varepsilon}_{5j}. \tag{2.8}$$

So, the updated vector is defined as

$$\boldsymbol{\beta}^{(s)} = \hat{\boldsymbol{b}}^{(s)} / ||\hat{\boldsymbol{b}}^{(s)}||, \tag{2.9}$$

where $\hat{\boldsymbol{b}}^{(s)} = (\hat{b}_1^{(s)}, \ldots, \hat{b}_q^{(s)})^t$, and the corresponding latent variate is $\boldsymbol{v}^{(s)} = \boldsymbol{Y}_0 \boldsymbol{\beta}^{(s)}$.

Repeat this procedure until convergence of the algorithm, and let $\boldsymbol{\alpha}_1$ and $\boldsymbol{u}_1$ be proportional to the last values of $\boldsymbol{\alpha}^{(s)}$ and $\boldsymbol{u}^{(s)}$, respectively, where the proportional constant is such that $\boldsymbol{u}_1$ has norm equal to one.

The first redundancy coefficient, R_{y1}, is estimated using a robust estimator of the correlation coefficient (in the present case, reweighed MCD estimator, RMCD, Rousseeuw and Van Driessen, 1999).

$$R_{y1} = \sum_{j=1}^{q} Corr(\boldsymbol{u}_1, \boldsymbol{y}_{0j})^2 / q. \tag{2.10}$$

We recall that RMCD is the empirical covariance matrix computed from the subset of size $h \approx 0.75n$ with the smallest determinant, where n is the sample size.

2.1.2. Redundancy Variates of Higher Order. The next redundancy variate, $\boldsymbol{u}_2$, has to be uncorrelated with the previous one. In other words, we can say that $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ have to be orthogonal. This restriction can be fulfilled if the data matrix, $\boldsymbol{X}_0$, is projected into the space orthogonal to $\boldsymbol{u}_1$. Let us consider the following regression model

$$\boldsymbol{X}_0 = \boldsymbol{u}_1 \boldsymbol{c}^t + \boldsymbol{\varepsilon}_6 \tag{2.11}$$

where $\boldsymbol{X}_1 = \boldsymbol{X}_0 - \hat{\boldsymbol{X}}_0$ are the corresponding estimated residuals. Due to standard results from multiple linear regression, the residuals, $\boldsymbol{X}_1$, are orthogonal to $\boldsymbol{u}_1$. Hence, repeating the procedure to obtain the first redundancy variate using the residuals, $\boldsymbol{X}_1$, instead of the original data, $\boldsymbol{X}_0$, we can guarantee that the next redundancy variate (a linear combination of the columns of $\boldsymbol{X}_1$) is, in fact, orthogonal to $\boldsymbol{u}_1$.

However, this idea raises some difficulties, since $\boldsymbol{X}_1$ is orthogonal to $\boldsymbol{u}_1$, and therefore, it has rank $(p-1)$, because $rank\,(\boldsymbol{X}_0) = p$. In order to overcome this

collinearity problem we consider, as in Branco et al. (2003), the singular value decomposition

$$\boldsymbol{X}_1 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^t = \boldsymbol{U}^*\boldsymbol{D}^*\boldsymbol{V}^{*t}, \tag{2.12}$$

where $\boldsymbol{D}$ is a diagonal matrix of the p singular values, $\boldsymbol{D}^*$ is diagonal matrix reduced by one row and column associated with the null singular value. A similar idea is used to define the matrices $\boldsymbol{U}^*$ and $\boldsymbol{V}^*$. Using (2.12), we can define $\boldsymbol{X}_1^* = \boldsymbol{U}^{*t}\boldsymbol{X}_1 = \boldsymbol{D}^*\boldsymbol{V}^{*t}$, and this matrix of size $(n \times (p-1))$ has full rank. Finally, the procedure proposed to estimate the first redundancy variate can be applied to $\boldsymbol{X}_1^*$ and $\boldsymbol{Y}_0$. As opposed to the estimation method developed for canonical correlation analysis (Branco et al., 2003), in the redundancy analysis there is no need to transform $\boldsymbol{Y}_0$.

Let $\boldsymbol{u}^*$ and $\boldsymbol{\alpha}^*$ be the results, after convergence, of the iterative procedure developed to estimate the second redundancy variate. These quantities are defined in residual spaces. Hence, a transformation to the original space has to be done, under the orthogonality condition, that is, $\boldsymbol{u}_2 = \boldsymbol{X}_0\boldsymbol{\alpha}_2$ has to be orthogonal to $\boldsymbol{u}_1$. Let us consider the regression model

$$\boldsymbol{u}^* = e\boldsymbol{u}_1 + \boldsymbol{\varepsilon}_7, \tag{2.13}$$

and let $\tilde{\boldsymbol{u}}$ be the estimated residuals. This vector, $\tilde{\boldsymbol{u}}$, is orthogonal to the previous estimated redundancy variate, $\boldsymbol{u}_1$. Moreover, in order to express the redundancy variate as a linear combination of $\boldsymbol{X}_0$ with coefficients $\boldsymbol{\alpha}_2$ we need to regress $\tilde{\boldsymbol{u}}$ on $\boldsymbol{X}_0$

$$\tilde{\boldsymbol{u}} = \boldsymbol{X}_0\boldsymbol{f} + \boldsymbol{\varepsilon}_8. \tag{2.14}$$

Then $\boldsymbol{\alpha}_2 = k\hat{\boldsymbol{f}}$, where k is such that $\boldsymbol{u}_2 = \boldsymbol{X}_0\boldsymbol{\alpha}_2$ has norm equal to one.

To obtain redundancy variates of order $l \in \{3, \ldots, r\}$, with $r = \min\{p, q\}$, a similar procedure has to be developed, but to transform the estimates into the original space, instead of using (2.13), a slightly different regression model has to be considered

$$\boldsymbol{u}^* = \boldsymbol{U}\boldsymbol{e} + \boldsymbol{\varepsilon}_9, \tag{2.15}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{l-1}]$. This model guarantees that $\boldsymbol{u}_l$ is orthogonal to $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{l-1}$.

2.2. Robust Alternating Regressions

As suggested in Croux et al. (2003) and Branco et al. (2003), the regression models considered in the iterative procedure are estimated using weighted L_1 regression, with weights $w_i(\boldsymbol{X}_{l-1}^*)$ defined by

$$w_i(\boldsymbol{X}_{l-1}^*) = \min\left(1, \frac{\chi^2_{p^*,0.95}}{D_i^2(\boldsymbol{X}_{l-1}^*)}\right),\ i = 1, \ldots, n, \tag{2.16}$$

where $\chi^2_{p^*,0.95}$ is the upper 5% critical value of a chi-squared distribution with $p^* = p - l + 1$ degrees of freedom (the number of columns of $\boldsymbol{X}_{l-1}^*$), and

$$D_i(\boldsymbol{X}_{l-1}^*) = \sqrt{(\boldsymbol{x}_i^{*(l-1)} - T(\boldsymbol{X}_{l-1}^*))^t C(\boldsymbol{X}_{l-1}^*)^{-1}(\boldsymbol{x}_i^{*(l-1)} - T(\boldsymbol{X}_{l-1}^*))},\ i = 1, \ldots, n,$$

where $\boldsymbol{x}_i^{*(l-1)}$ is the i^{th} column of $\boldsymbol{X}_{l-1}^*$ and D_i is a robust distance. By the same reasons pointed out in Croux et al. (2003) and Branco et al. (2003) T and C were chosen to be MVE estimators of location and scatter, respectively. The regressions used to rewrite $\boldsymbol{u}^*$ in the original space were estimated using LTS regression (Rousseeuw, 1984). However, other robust estimators of regression can also be used.

3. Simulation Study

It is convenient to keep the simulation conditions as in Branco et al. (2003). By doing so, it is possible to compare the performance of the robust estimators based on alternating regressions not only with other robust estimators but also with the robust estimators of canonical correlation analysis based on alternating regression. This makes sense since the correlation matrices chosen are such that the redundancy analysis and the canonical correlation analysis lead to the same linear combination of the observed variables, i.e. the true values for the canonical coefficients are equal to the true values for the redundancy coefficients.

The performance of the robust estimators based on alternating regressions (RAR) are compared with the performance of the estimators based on robust correlation matrices (RMCD - $h \approx 0.75n$ and M-Estimator (M), using Huber psi-function: $w(s) = \min\left(1, \chi^2_{(p+q),0.95}/|s|\right)$) as well as with the classical estimator (Class). To do so, samples of size 500 were generated from four different distributions: normal distribution (NOR), with zero mean and covariance matrix Σ ($N(\mathbf{0}, \Sigma)$); symmetric contaminated normal (SCN) where each vector of observations is generated from a normal distribution ($N(\mathbf{0}, \Sigma)$) with probability 0.9 and from $N(\mathbf{0}, 9\Sigma)$ with probability 0.1; t-distribution with 3 degrees of freedom (T3) and asymmetric contamination (ACN) where 90% of the data are generated from a $N(\mathbf{0}, \Sigma)$ and the other 10% of the observations are equal to the vector $tr(\Sigma)\mathbf{1}^t_{(p+q)}$.

For each type of distribution and each estimation method $m = 300$ samples of 500 observations were produced. To assess the performance of each estimation method the following measures of MSE, were defined

$$\begin{aligned} \text{MSE}(\hat{\boldsymbol{\alpha}}_l) &= \frac{1}{m}\sum_{j=1}^{m} \cos^{-1}\left(\frac{|\boldsymbol{\alpha}^t\hat{\boldsymbol{\alpha}}_l^j|}{||\hat{\boldsymbol{\alpha}}_l^j|| \cdot ||\boldsymbol{\alpha}_l||}\right), \\ \text{MSE}(\hat{R}_{yl}) &= \frac{1}{m}\sum_{j=1}^{m}\left(\hat{R}_{yl}^j - R_{yl}^j\right)^2, \end{aligned}$$

where $\hat{\boldsymbol{\alpha}}_l^j$ is the j^{th} estimate of the l^{th} redundancy coefficient, $\boldsymbol{\alpha}_l$, and $\hat{R}_{yl}^j$ is the j^{th} estimate of the l^{th} redundancy index, R_{yl}.

Table 1 shows the values of p and q that were considered, together with the correlation matrix between the two groups. The correlation matrix of each group was taken as the identity matrix.

TABLE 1. Simulation setup. $R_{11} = \boldsymbol{I}_p$ and $R_{22} = \boldsymbol{I}_q$.

n	m	p	q	R_{12}
500	300	2	2	$\begin{bmatrix} 0.9 & 0 \\ 0 & 1/2 \end{bmatrix}$
500	300	2	4	$\begin{bmatrix} 0.9 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \end{bmatrix}$
500	300	4	4	$\begin{bmatrix} 0.9 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix}$

The results for the case $p = q = 2$ are summarized in Figure 1 and 2 (see Figure 5(a) to identify each contamination scheme). In this case, it can be said that the contamination influences the vectors of redundancy coefficients equally. ACN is the contamination scheme that produces more damages in the estimated values, and only the estimates of redundancy coefficients based on RMCD and RAR are robust for this kind of contamination. ACN has also a serious effect on the first redundancy index when the classical method is used. In the case $p = 2$ and $q = 4$ something quite similar happens, except that RAR reveals to be even better to estimate the first redundancy coefficient. Because of lack of space the figures associated with this case are omitted.

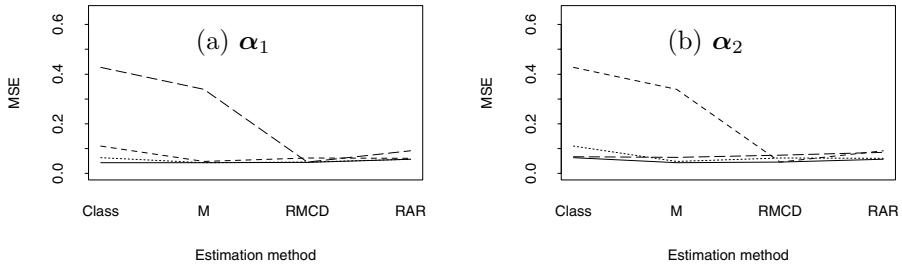

FIGURE 1. Mean square error of the redundancy coefficients, $p = 2$, $q = 2$.

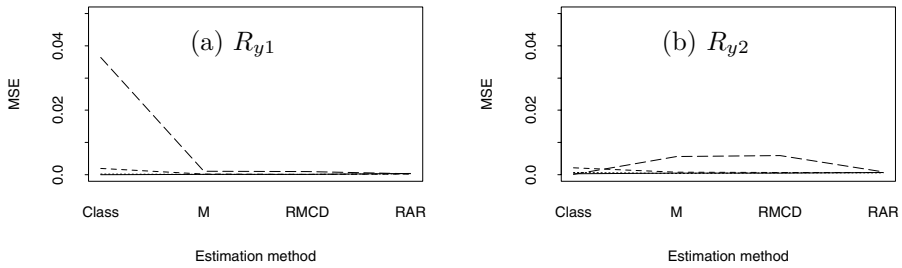

FIGURE 2. Mean square error of the redundancy indexes, $p = 2$, $q = 2$.

For $p = q = 4$ we obtain very similar results for the first redundancy coefficient and index as for $p = q = 2$. However, the effect of the contamination increases with the order of the redundancy coefficients (see Figures 3, 4, and Figure 5(a) to identify each contamination scheme). The method based on alternating regression performs slightly worse than RMCD for the two last redundancy coefficients ($\boldsymbol{\alpha_3}$ and $\boldsymbol{\alpha_4}$), and for the two last redundancy indexes (R_{y3} and R_{y4}).

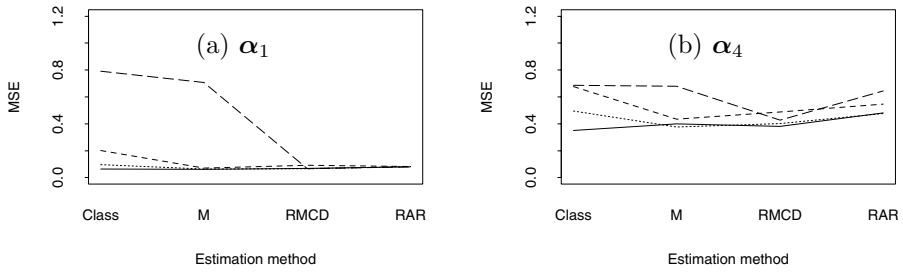

FIGURE 3. Mean square error of the first and last redundancy coefficients, $p = 4$, $q = 4$.

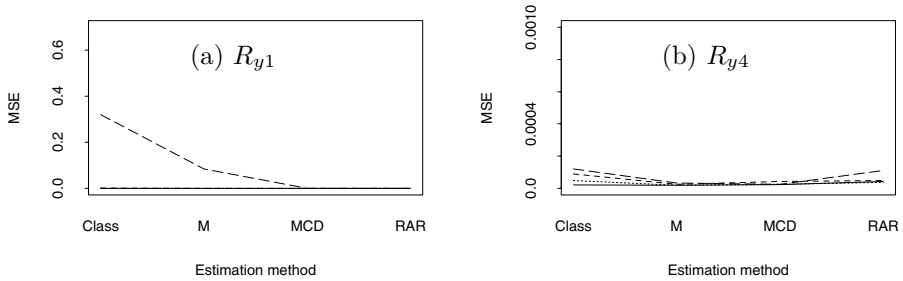

FIGURE 4. Mean square error of the first and last redundancy indexes, $p = 4$, $q = 4$.

FIGURE 5. (a) On the left – legend of Figures 1, 2, 3 and 4. (b) On the right – legend of the figures summarizing the performance and the empirical breakdown point of the various estimators.

3.1. Empirical Breakdown Point

Another criteria to compare the performance of the estimators is based on the empirical breakdown point. To do so, a simulation study was carried out, with the same setup that was used in Branco et al. (2003), where each group has 3 variables ($p = q = 3$). For each sample, $(100 - \epsilon)\%$ of the points were generated from a normal distribution with zero mean and correlation matrix, R, with $R_{11} = \boldsymbol{I}_3$ and $R_{22} = \boldsymbol{I}_3$ and

$$R_{12} = \begin{bmatrix} 0.9 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix},$$

and the other $\epsilon\%$ of the observations are equal $tr(\Sigma)\mathbf{1}^t$, (in the present case, $tr(\Sigma) = 6$). The values of ϵ were chosen from zero (no contamination) to 25 (25% of contamination), i.e., $\epsilon \in \{0, 1, \ldots, 25\}$. As before we chose $n = 500$. The procedure was repeated 200 times for each estimation method. The results, for the first and last redundancy coefficients and indexes, are summarized in Figures 6 and 7 (see Figure 5(b) to identify each estimator considered). In Figure 7(a), the MSE associated with the M, RMCD and RAR estimators seems to be equally low due to the apparent large magnitude of the MSE of the Class estimator. Nevertheless, the empirical breakdown point for R_{y2} and R_{y3} (see Figure 7(b)) show that the RAR has lower empirical breakdown point than RMCD but higher than the estimators based on the M-estimator (M). Similar conclusions apply to $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$, as showed in Figure 6.

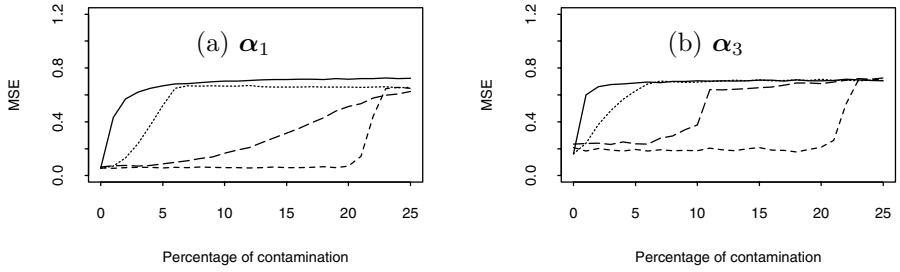

FIGURE 6. Empirical breakdown point, redundancy coefficients.

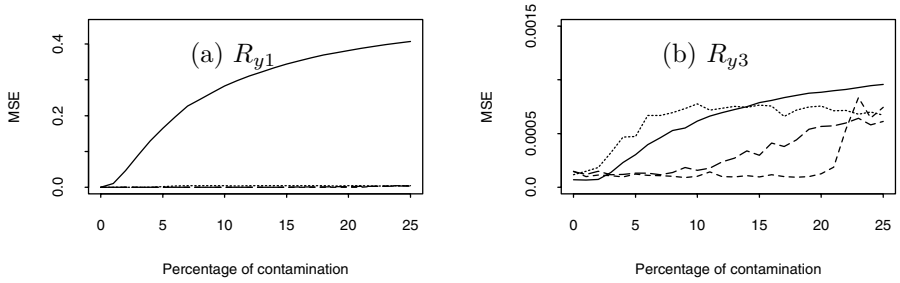

FIGURE 7. Empirical breakdown point, redundancy indexes.

4. Discussion and Future Work

The RAR procedure produced useful results and it has the advantage that it can be carried out in the case of more variables than observations. It is also capable of dealing with missing values and it can cope with outlying cells (Croux et al., 2003). Moreover, RAR aims at directly maximizing the redundancy index, and is therefore in a way intuitively more appealing than a covariance matrix based procedure. However, the search for other robust estimators should be pursued. In Oliveira and Branco (2002) various robust estimators have been studied for the first redundancy variate, where the estimators based on projection pursuit revealed promising results. This estimator has to be developed for higher order variates.

The simulation developed in this study was designed to facilitate the comparison with the study in Branco et al. (2003). So far, it can be said that the behavior of RAR is similar in both canonical correlation analysis and redundancy analysis. Having resolved the problem of robustifying (by alternating regression) the canonical correlation analysis and redundancy analysis we are compelled to consider the generalization of these two methods proposed by DeSarbo (1981).

As it was pointed out by van den Wollenberg (1977), principal components analysis is a special case of redundancy analysis. If we choose the dependent variables, $\boldsymbol{y}$, equal to the independent variables $\boldsymbol{x}$, the principal components differ from the redundancy variates by a constant term. In principal components analysis, we search for a linear combination of the $\boldsymbol{x}$, $\boldsymbol{a}$, that maximizes $\boldsymbol{a}^t R \boldsymbol{a}$ and $\boldsymbol{a}^t \boldsymbol{a} = 1$ ($R = Corr(\boldsymbol{x})$). In redundancy analysis we seek for a linear combination of $\boldsymbol{x}$, $\boldsymbol{\alpha}$, that maximizes $\boldsymbol{\alpha}^t R^2 \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}^t R \boldsymbol{\alpha} = 1$. Both the solutions of the two problems have the directions of the eigenvectors of R. Therefore, the algorithm based on alternating regressions can be used to estimate principal components. It would be interesting to compare this approach with other suggestions made in the literature, like the procedure based on alternating regressions to estimate principal components proposed in Wold (1966), and other robust estimators, e.g., based on projection pursuit.

References

[1] J.A. Branco, C. Croux, P. Filzmoser, and M.R. Oliveira, *Robust Canonical Correlations: A Comparative Study.* Preprint, 2003, available at *http://www.statistik.tuwien.ac.at/public/filz/publications/.*

[2] C. Croux and A. Ruiz-Gazen, *A fast Algorithm for Robust Principal Components Based on Projection Pursuit.* In A. Prat, editor, COMPSTAT: Proceedings in Computational Statistics, pages 211–216. Physica-Verlag, Heidelberg, 1996.

[3] C. Croux, P. Filzmoser, G. Pison, and P.J. Rousseeuw, *Fitting Multiplicative Models by Robust Alternating Regressions.* Statist. Comput. **13** (2003), 23–36.

[4] W.S. DeSarbo, *Canonical/Redundancy Factoring Analysis.* Psychometrika **46** (1981), 307–329.

[5] E. Lyttkens, *Regression Aspects of Canonical Correlation.* J. Multivariate Anal. **2** (1972), 418–439.

[6] R.A. Maronna, *Robust M-Estimators of Multivariate Location and Scatter.* Ann. Statist. **4** (1976), 51–67.

[7] M.R. Oliveira and J.A. Branco, *Comparison of Three Methods for Robust Redundancy Analysis.* In R. Dutter, P. Filzmoser, U. Gather, and P. J. Rousseeuw, editors, Developments in Robust Statistics, pages 287–295. Physica-Verlag, Heidelberg, 2003.

[8] A.C. Rencher, *Multivariate statistical inference and applications.* Wiley Series in Probability and Statistics, New York, 1998.

[9] P.J. Rousseeuw, *Multivariate Estimation With High Breakdown Point.* In W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, editors, Mathematical Statistics and Applications, Vol. B, pages 283–297. Reidel, Dordrecht, The Netherlands, 1985.

[10] P.J. Rousseeuw and K. Van Driessen, *A Fast Algorithm For the Minimum Covariance Determinant Estimator.* Technometrics **41** (1999), 212–223.

[11] P.J. Rousseeuw, *Least Median of Squares Regression.* J. Amer. Statist. Assoc. **79** (1984), 871–881.

[12] D.K. Stewart and W.A. Love, *A General Canonical Correlation Index.* Psychol. Bullet. **70** (1968), 160–163.

[13] M. Tenenhaus, *La Régression PLS. Théorie et pratique.* Éditions Technip, Paris, 1998.

[14] H. Wold, *Nonlinear Estimation by Iterative Least Squares Procedures.* In F.N. David, editor, A Festschrift for J. Neyman, pages 411–444. John Wiley and Sons, New York, 1966.

[15] A.L. van den Wollenberg, *Redundancy Analysis: An Alternative for Canonical Correlation Analysis.* Psychometrika **42** (1977), 207–219.

M.R. Oliveira and J.A. Branco
Instituto Superior Técnico
Av. Rovisco Pais
1049-001 Lisboa
Portugal
e-mail: {rosario.oliveira, jbranco}@math.ist.utl.pt

C. Croux
Katholieke Universiteit Leuven
Naamsestraat 69
B-3000 Leuven
Belgium
e-mail: Christophe.Croux@econ.kuleuven.ac.be

P. Filzmoser
Vienna University of Technology
Wiedner Hauptstr. 8–1
A-1040 Vienna
Austria
e-mail: P.Filzmoser@tuwien.ac.at