Ann. Inst. Statist. Math.
Vol. 50, No. 4, 755–772 (1998)

ASSESSING LOCAL INFLUENCE IN CANONICAL CORRELATION ANALYSIS

HONG GU AND WING K. FUNG

Department of Statistics, The University of Hong Kong, Pokfulam Road, Hong Kong, China

(Received March 18, 1997; revised September 26, 1997)

Abstract. The first order local influence approach is adopted in this paper to assess the local influence of observations to canonical correlation coefficients, canonical vectors and several relevant test statistics in canonical correlation analysis. This approach can detect different aspects of influence due to different perturbation schemes. In this paper, we consider two different kinds, namely, the additive perturbation scheme and the case-weights perturbation scheme. It is found that, under the additive perturbation scheme, the influence analysis of any canonical correlation coefficient can be simplified to just observing two predicted residuals. To do the influence analysis for canonical vectors, a scale invariant norm is proposed. Furthermore, by choosing proper perturbation scales on different variables, we can compare the different influential effects of perturbations on different variables under the additive perturbation scheme. An example is presented to illustrate the effectiveness of the first order local influence approach.

Key words and phrases: Canonical correlation analysis, local influence, diagnostics, perturbation, tests of independence.

1. Introduction

Canonical correlation analysis, or shortly, CCA, as an important method for reducing the correlation structure between two sets of variables, is extremely sensitive to outliers just as that in principal component analysis. However, influence analysis in CCA has been less developed except several articles (see Radhakrishnan and Kshirsagar (1981) and Romanazzi (1992)), in which only the influence function approach has been considered. We know, two sample versions of influence function, the empirical influence function EIC and the deleted empirical influence function $\mathrm{EIC}_{(i)}$ (Cook and Weisberg (1982)), are indeed some kinds of local influence under certain perturbation schemes (e.g., see Tanaka (1994)). However, the basic idea of influence function analysis is usually to assess the influence of each single observation. Local influence analysis, by perturbing all cases simultaneously, may disclose the most sensitive perturbation ways which may be used to detect some joint influential effect.

The local influence approach was first proposed by Cook (1986) as a general method for assessing the effect of minor perturbations of a statistical model. It suggests measuring the sensitivity of the analysis through the normal curvature of the likelihood displacement surface $\mathrm{LD}(\omega) = 2[\ell(\hat{\theta}) - \ell(\hat{\theta}_\omega)]$ when minor perturbations were introduced in the postulated model. This method extracts a synthesis information from the first derivative of the parameter estimate $\hat{\theta}_\omega$ with respect to perturbation ω and the second derivative of log-likelihood function $\ell(\theta)$ with respect to the parameter θ. The former is actually the influence arisen by the introduced perturbation to the parameter estimate $\hat{\theta}$. So it is reasonable to just use the former as an influence assessment when only the parameter is of interest. In fact, if we use other objective functions rather than the likelihood displacement for local influence analysis, the normal curvature and the second order derivative are relatively less important compared to the first order derivative which is non-zero. Moreover, Fung and Kwan (1997) showed that the normal curvature is not scale invariant when the first derivative of the objective function is non-zero. Thus the gradient direction of an objective function is used for diagnostic analysis in our paper, which is also termed the first order local influence approach by Wu and Luo (1993).

One advantage of the local influence approach is its ability to handle cases simultaneously. Unlike the deletion influence of multiple cases which in general arises a lot of computational problems, local influence approach employs a differential comparison of parameter estimates before and after perturbation which is simple in computation. Besides, different kinds of perturbation schemes can be chosen to perform the local influence analysis based on different special concerns. Lawrance (1991) logically distinguished different perturbation schemes into three sorts, they are respectively perturbations to model assumptions, perturbations to data values and perturbations to case weights. More motivation of local influence method can be referred to Cook (1986). In this paper, we will examine the local influence in canonical correlation analysis under two perturbation schemes, namely, the additive perturbation scheme and the case-weights perturbation scheme.

The organization of the rest of the paper is as follows. In Section 2, we present the first order local influence results for canonical correlation coefficients and canonical vectors and their corresponding interpretations under the additive perturbation scheme. It is found that the local influence of any canonical correlation can be simplified to just observing two predicted residuals. When the parameter of interest is a vector rather than a scalar, a proper norm is suggested to measure the influence to the parameter vector. The comparison of the different influential effects of different variables is also included in this section. The local influence of canonical correlations and canonical vectors under the case-weights perturbation scheme is outlined in Section 3, of which the basic results are actually equivalent to those of the corresponding empirical influence functions. Lee and Zhao (1996) recently investigated the local influence on the Pearson's goodness-of-fit statistic in generalized linear models, however, the local influence approach has been less adopted to study the influence on test statistics. In Section 4, we consider the local influence analysis of four common test statistics in CCA under the above two perturbation schemes. An illustrative example is presented in

Section 5. Some concluding remarks are included in the last section.

2. Local influence under additive perturbation

Suppose $X_{n\times p}$ is a matrix of n observations of a p-dimensional random vector x and $Y_{n\times q}$ is the corresponding n observations of a q-dimensional random vector y $(p \leq q)$. Let the sample mean and sample covariance matrix be $(\bar{x}, \bar{y})$ and $\hat{\Sigma}$ respectively, where $\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}$ is the maximum likelihood estimator of the population covariance matrix for (x^T, y^T) under the normal assumption. The sample squared canonical correlations $r_1^2, \ldots, r_p^2$ are the eigenvalues of matrices $\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$ and $\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$, which are distinct and nonzero with probability 1. For convenience, we assume $1 > r_1^2 > r_2^2 > \cdots > r_p^2 > 0$. The i-th canonical vectors $\hat{a}_i, \hat{b}_i$ are respectively the eigenvectors of matrices $\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$ and $\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$ corresponding to the i-th eigenvalue r_i^2, which satisfy the constraints

$$\hat{a}_i^T\hat{\Sigma}_{11}\hat{a}_j = \delta_{ij} \tag{2.1}$$

and

$$\hat{b}_i^T\hat{\Sigma}_{22}\hat{b}_j = \delta_{ij} \tag{2.2}$$

where δ_{ij} is the Kronecker delta.

Now we investigate the sensitivity of canonical correlation analysis to the measurement errors in each variable by perturbing all the observations in all variables simultaneously. Under such a perturbation scheme, the original data matrices X and Y will be replaced by

$$X_\omega = X + \omega^X S^X \tag{2.3}$$

and

$$Y_\omega = Y + \omega^Y S^Y \tag{2.4}$$

where $\omega = (\omega_{n\times p}^X, \omega_{n\times q}^Y)$ represents a well defined perturbation scheme and is restricted to some open subset Ω of $R^{n(p+q)}$. As indicated by Cook (1986) that global measures of influence which characterize the behavior of an influence graph over all of Ω are generally much more difficult to construct than local measures which characterize behavior in a neighborhood of a selected ω. Here we consider a local measure of influence around $\omega_0 = (\omega_0^X, \omega_0^Y) = (0, 0)$, which represents no perturbation on X and Y. $S^X = \mathrm{diag}(s_1^x, \ldots, s_p^x)$ and $S^Y = \mathrm{diag}(s_1^y, \ldots, s_q^y)$ convert the perturbation ω^X and ω^Y to the appropriate sizes and units, so that the elements of $\omega^X S^X$ and $\omega^Y S^Y$ are compatible with the corresponding elements of X and Y respectively. If we denote the perturbed covariance matrix and the i-th perturbed canonical correlation coefficient and canonical vectors as $\hat{\Sigma}(\omega)$, $r_i(\omega)$, $\hat{a}_i(\omega)$ and $\hat{b}_i(\omega)$, then we have

$$\hat{\Sigma}_{11}^{-1}(\omega)\hat{\Sigma}_{12}(\omega)\hat{\Sigma}_{22}^{-1}(\omega)\hat{\Sigma}_{21}(\omega)\hat{a}_i(\omega) = r_i^2(\omega)\hat{a}_i(\omega), \tag{2.5}$$

$$\hat{a}_i^T(\omega)\hat{\Sigma}_{11}(\omega)\hat{a}_i(\omega) = 1. \tag{2.6}$$

To find the gradient direction which can indicate how to perturb the data under such perturbation scheme to obtain the greatest local change in the concerned parameters, we need to calculate the first derivative of concerned parameters with respect to ω. For ease of manipulation, we denote $\omega^X = \omega_0^X + \eta L_1$ and $\omega^Y = \omega_0^Y + \eta L_2$, where $l = \mathrm{vec}(L_1, L_2)$ is a direction vector and η represents the perturbation scale with $\eta = 0$ corresponds to no perturbation. By taking derivatives with respect to η on both sides of equations (2.5) and (2.6) and then evaluating them at ω_0, we get a system of equations. Solving the resulting equations, we obtain

$$\frac{\partial r_i(\omega)}{\partial \eta}\bigg|_{\omega=\omega_0} = [(S^X \hat{a}_i) \otimes (V_i - r_i U_i)]^T \mathrm{vec}(\tilde{L}_1) + [(S^Y \hat{b}_i) \otimes (U_i - r_i V_i)]^T \mathrm{vec}(\tilde{L}_2) \tag{2.7}$$

and

$$\frac{\partial \hat{a}_i(\omega)}{\partial \eta}\bigg|_{\omega=\omega_0} = f_1 \hat{a}_1 + \cdots + f_p \hat{a}_p, \tag{2.8}$$

where $\hat{a}_i$ and $\hat{b}_i$ are the i-th pair of canonical vectors, $U_i = \tilde{X}\hat{a}_i$ and $V_i = \tilde{Y}\hat{b}_i$ are two $n \times 1$ vectors, and $\tilde{X}$, $\tilde{Y}$, $\tilde{L}_1$ and $\tilde{L}_2$ are centered X, Y, L_1 and L_2, respectively. The $f_k's$ $(k = 1, \ldots, p)$ are defined as follows

$$f_k = \begin{cases} -[(S^X \hat{a}_i) \otimes U_k]^T \mathrm{vec}(\tilde{L}_1) \\ \quad + \dfrac{1}{r_i^2 - r_k^2} \{[r_i((S^X \hat{a}_k) \otimes (V_i - r_i U_i))^T \\ \qquad + r_k((S^X \hat{a}_i) \otimes (V_k - r_k U_k))^T] \mathrm{vec}(\tilde{L}_1) \\ \qquad + [r_k((S^Y \hat{b}_k) \otimes (U_i - r_i V_i))^T \\ \qquad + r_i((S^Y \hat{b}_i) \otimes (U_k - r_k V_k))^T] \mathrm{vec}(\tilde{L}_2)\}, & k \neq i, \\ -[(S^X \hat{a}_i) \otimes U_i]^T \mathrm{vec}(\tilde{L}_1), & k = i. \end{cases} \tag{2.9}$$

The corresponding expressions for $\partial \hat{b}_i(\omega)/\partial \eta|_{\omega=\omega_0}$ are similar to (2.8) and (2.9). For brevity, they are omitted.

2.1 *Identifying influential points for canonical correlations*

According to the above definitions, the n-vectors U_i and V_i are the centered scores of the n individuals on the i-th canonical variables for x and y, and r_i is the simple sample correlation coefficient of U_i and V_i. If the x and y variables are interpreted as the "predictor" and "predicted" variables, the U_i score can be used to predict a value of the V_i score using the least squares regression. Since the sample variances of U_i and V_i are both 1, the regression coefficient will be r_i when regressing V_i on U_i. Thus, we may term $(V_i - r_i U_i)$ as the predicted residual of V_i on U_i and vice versa, $(U_i - r_i V_i)$ the predicted residual of U_i on V_i.

If the two scale factors S^X and S^Y are given, as indicated by (2.7), the extreme local sensitivity direction $\tilde{l} = \mathrm{vec}(\tilde{L}_1, \tilde{L}_2)$ for r_i corresponds to the gradient direction which is the unit vector proportional to $([(S^X \hat{a}_i) \otimes (V_i - r_i U_i)]^T, [(S^Y \hat{b}_i) \otimes (U_i - r_i V_i)]^T)^T$, and the norm of $([(S^X \hat{a}_i) \otimes (V_i - r_i U_i)]^T, [(S^Y \hat{b}_i) \otimes (U_i - r_i V_i)]^T)^T$

can serve as a measurement of possible maximum local influence. This vector $\tilde{l}$ indicates how to perturb the data to obtain the greatest local change in r_i. And further, it should be noted that in this $\tilde{l}$ every column of $\tilde{L}_1$ is proportional to $(V_i - r_iU_i)$ and every column of $\tilde{L}_2$ is proportional to $(U_i - r_iV_i)$. That means any observation which is the most influential in one variable of x is also the most influential one for the other variables in x and similarly for y. Thus it is enough to just look at the two predicted residuals $(V_i - r_iU_i)$ and $(U_i - r_iV_i)$ if we are only concerned about which observations are influential in the sense of additive perturbation. In order to specify influential observations, it is better to give reference points or cut-off points. To set such reference points for influence analysis, we need to know the distributions of $\tilde{Y}_k\hat{b}_i - r_i\tilde{X}_k\hat{a}_i$ and $\tilde{X}_k\hat{a}_i - r_i\tilde{Y}_k\hat{b}_i$, where $\tilde{X}^T = (\tilde{X}_1^T, \ldots, \tilde{X}_n^T)$ and $\tilde{Y}^T = (\tilde{Y}_1^T, \ldots, \tilde{Y}_n^T)$. However, these distributions seem to be intractable even under the normality assumption since they involve the estimates of canonical vectors, whose distributions are intractable themselves. Nevertheless, we may use the normal distribution to approximate the distributions, the plausible reasons are given as follows.

LEMMA 1. *If the joint distribution of (x^T, y^T) is $N((\mu_x^T, \mu_y^T), \Sigma)$, then $(y - \mu_y)^T b_i - \rho_i(x - \mu_x)^T a_i$ and $(x - \mu_x)^T a_i - \rho_i(y - \mu_y)^T b_i$ follow $N(0, 1 - \rho_i^2)$, where ρ_i, a_i and b_i are respectively the i-th population canonical correlation and canonical vectors.*

It seems to be reasonable that this lemma is approximately true when replacing μ_x, μ_y, ρ_i, a_i and b_i by their respective maximum likelihood estimators $\bar{x}$, $\bar{y}$, r_i, $\hat{a}_i$ and $\hat{b}_i$. That is, $(y-\bar{y})^T\hat{b}_i - r_i(x-\bar{x})^T\hat{a}_i$ and $(x-\bar{x})^T\hat{a}_i - r_i(y-\bar{y})^T\hat{b}_i$ approximately follow $N(0, 1 - r_i^2)$. Furthermore, the sample variance of $\tilde{Y}_k\hat{b}_i - r_i\tilde{X}_k\hat{a}_i$ can be calculated directly as $\Sigma_{k=1}^n(\tilde{Y}_k\hat{b}_i - r_i\tilde{X}_k\hat{a}_i)^2/n$. By the constraints (2.1) and (2.2), we know both sample variances of $\tilde{Y}_k\hat{b}_i$ and $\tilde{X}_k\hat{a}_i$ are 1, and since the sample correlation of $\tilde{Y}_k\hat{b}_i$ and $\tilde{X}_k\hat{a}_i$ is r_i, the sample variance of $\tilde{Y}_k\hat{b}_i - r_i\tilde{X}_k\hat{a}_i$ will be exactly $(1 - r_i^2)$. Similarly, we know the sample variance of $\tilde{X}_k\hat{a}_i - r_i\tilde{Y}_k\hat{b}_i$ is also $(1 - r_i^2)$. Thus we may use the 2σ rule or 3σ rule to set up the reference cut-off points.

As for which columns play more important roles under such perturbation, we should investigate the vectors $(S^X\hat{a}_i)$ and $(S^Y\hat{b}_i)$. To do this, we need to specify S^X and S^Y. One reasonable choice for S^X and S^Y is the sample standard deviation of each variable, that is, $S^X = [\mathrm{diag}(\hat{\Sigma}_{11})]^{1/2}$ and $S^Y = [\mathrm{diag}(\hat{\Sigma}_{22})]^{1/2}$. Such a choice has some nice properties. Firstly, it makes $(S^X\hat{a}_i)$ and $(S^Y\hat{b}_i)$ scale invariant to the rescaling of original data which is commonly found in practice. Secondly, $(S^X\hat{a}_i)$ and $(S^Y\hat{b}_i)$ under such a choice are the i-th pair of canonical vectors if the correlation matrix rather than the covariance matrix is used to perform CCA.

Thus we may identify not only which observation has a large influence on the canonical correlation, but also which variables of this observation should be further investigated to find the special cause of the sensitivity.

2.2 *Identifying influential points for canonical vectors*

If we want to consider the influence analysis of canonical vectors, (2.8) and (2.9) present some basic results. Since here a vector is monitored, a selection of norm is necessary to measure the influence to the vector. We define the scale invariant relative norms as

$$\text{RN}\left(\frac{\partial \hat{a}_i(\omega)}{\partial \eta}\right) = \left(\frac{\partial \hat{a}_i(\omega)}{\partial \eta}\right)^T \hat{\Sigma}_{11} \left(\frac{\partial \hat{a}_i(\omega)}{\partial \eta}\right), \tag{2.10}$$

and

$$\text{RN}\left(\frac{\partial \hat{b}_i(\omega)}{\partial \eta}\right) = \left(\frac{\partial \hat{b}_i(\omega)}{\partial \eta}\right)^T \hat{\Sigma}_{22} \left(\frac{\partial \hat{b}_i(\omega)}{\partial \eta}\right), \tag{2.11}$$

which are the norms of vectors $\partial \hat{a}_i(\omega)/\partial \eta$ and $\partial \hat{b}_i(\omega)/\partial \eta$ in affine spaces spanned respectively by $(\hat{a}_1, \ldots, \hat{a}_p)$ and $(\hat{b}_1, \ldots, \hat{b}_q)$. Thus the norm of vector $(\partial \hat{a}_i(\omega)/\partial \eta|_{\omega=\omega_0})$ is

$$\text{RN}\left(\frac{\partial \hat{a}_i(\omega)}{\partial \eta}\bigg|_{\omega=\omega_0}\right) = (f_1\hat{a}_1 + \cdots + f_p\hat{a}_p)^T \hat{\Sigma}_{11} (f_1\hat{a}_1 + \cdots + f_p\hat{a}_p). \tag{2.12}$$

Now we put f_k's $(k = 1, \ldots, p)$ in (2.9) into form $f_k = \alpha_k^T \operatorname{vec}(\tilde{L}_1) + \beta_k^T \operatorname{vec}(\tilde{L}_2)$, where α_k is an $np \times 1$ vector and β_k is an $nq \times 1$ vector. Making use of the constraint in (2.1), we have

$$\text{RN}\left(\frac{\partial \hat{a}_i(\omega)}{\partial \eta}\bigg|_{\omega=\omega_0}\right) = \sum_{k=1}^{p} f_k^2 = (\operatorname{vec}(\tilde{L}_1)^T, \operatorname{vec}(\tilde{L}_2)^T) F \begin{pmatrix} \operatorname{vec}(\tilde{L}_1) \\ \operatorname{vec}(\tilde{L}_2) \end{pmatrix}, \tag{2.13}$$

where $F = \sum_{k=1}^{p} \binom{\alpha_k}{\beta_k}(\alpha_k^T, \beta_k^T)$. Similarly, $\text{RN}(\partial \hat{b}_i(\omega)/\partial \eta|_{\omega=\omega_0})$ can be rewritten in the same form as (2.13).

Similar to the method proposed by Cook (1986) for studying the influence graph, we may also use the eigenvector $l_{\max}$ associated with the largest eigenvalue of the matrix F to determine the direction of $(\operatorname{vec}(\tilde{L}_1)^T, \operatorname{vec}(\tilde{L}_2)^T)$ in (2.13), which can then be used to indicate what kind of perturbation may arise the greatest local change in such defined norms of canonical vectors. However, in practical applications, it may not be easy to use this method directly due to two main reasons. One is that the F matrix involves the definitions of S^X and S^Y which is usually difficult to be given in advance. The other is that when n or $(p+q)$ are not so small, the eigen-analysis of $F_{n(p+q)\times n(p+q)}$ will be very time consuming or even impossible. A better way to do this is to perturb one column only, and to do so, we can have the $n \times n$ influence matrix F if we set the value corresponding to the perturbed column as 1 and other values as 0 in matrices S^X and S^Y in (2.9). The eigen-analysis result of F can provide the most significant local perturbation, from which we may indicate not only one or more influential cases by looking at which elements in $l_{\max}$ are relatively large, but also the group effects according to the different signs in $l_{\max}$. Another important information directly from the

F matrix is the diagonal elements of F, which can be used to construct plots to indicate the influential effect of each single observation in each variable.

Unlike the conclusion for canonical correlations, that is, if one observation is the most influential in one column, it must be the most influential one for other columns, however, for canonical vectors, the most influential observation in one column is not necessarily the most influential one for other columns. But we may only need to investigate carefully part of the columns for any fixed canonical vector, because perturbations on different columns under the same perturbation scheme lead to different magnitude of influence to the above defined relative norm. Thus only those columns which have large magnitude of influence are noteworthy. We may make use of the maximum possible influence magnitude of each column, which is the maximum eigenvalue of the corresponding F matrix, to do the comparison. But these maximum eigenvalues are not directly comparable because we have not yet chosen a proper perturbation scale on different columns. A reasonable choice for this scale is to make the perturbation scale on each column proportional to the sample standard deviation of that column, which results in the new maximum eigenvalues as the original maximum eigenvalues multiplied by the sample variances of corresponding columns. Besides, such choice makes the maximum eigenvalues invariant to the rescaling of original data. These ideas will be further illustrated in Section 5.

3. Local influence under case-weights perturbation

The case-weights perturbation scheme has been used in many different areas for influence analysis in which, quite often, a model must be specified, e.g., in linear regression. However, it seems no such model as linear regression is involved in canonical correlation analysis. What we are concerned about in CCA is just the sensitivity of canonical correlation analysis to the observed data. Based on the sense that case-weights perturbation is actually perturbing the weights of cases in model or estimation, we set up the perturbation scheme as follows.

In CCA with n observations of $(p+q)$ dimensional random vector (x_i^T, y_i^T), $i = 1, \ldots, n$, by giving weight ω_i to the i-th observation rather than one, we have

$$\hat{\Sigma}(\omega) = \frac{1}{\sum_{i=1}^n \omega_i} \sum_{i=1}^n \omega_i \begin{pmatrix} x_i - \bar{x}(\omega) \\ y_i - \bar{y}(\omega) \end{pmatrix} \begin{pmatrix} x_i - \bar{x}(\omega) \\ y_i - \bar{y}(\omega) \end{pmatrix}^T, \tag{3.1}$$

where

$$\bar{x}(\omega) = \frac{1}{\sum_{i=1}^n \omega_i} \sum_{i=1}^n \omega_i x_i, \tag{3.2}$$

$$\bar{y}(\omega) = \frac{1}{\sum_{i=1}^n \omega_i} \sum_{i=1}^n \omega_i y_i, \tag{3.3}$$

and $\omega = (\omega_1, \ldots, \omega_n)^T$ denotes the vector of case-weights, obviously, $\omega_0 = (1, \ldots, 1)^T$ corresponds to the case with no perturbation. We denote the i-th squared canonical correlation coefficient and the corresponding canonical vectors under this perturbation as $r_i^2(\omega)$, $\hat{a}_i(\omega)$ and $\hat{b}_i(\omega)$, then they satisfy the following constraints,

$$\hat{\Sigma}_{11}^{-1}(\omega)\hat{\Sigma}_{12}(\omega)\hat{\Sigma}_{22}^{-1}(\omega)\hat{\Sigma}_{21}(\omega)\hat{a}_i(\omega) = r_i^2(\omega)\hat{a}_i(\omega), \tag{3.4}$$

$$\hat{a}_i^T(\omega)\hat{\Sigma}_{11}(\omega)\hat{a}_i(\omega) = 1, \tag{3.5}$$

and similar constraints for $\hat{b}_i(\omega)$.

By taking derivatives with respect to ω_j ($j = 1, \ldots, n$) on both sides of (3.4) and (3.5), solving these equations and then evaluating them at $\omega_0 = (1, \ldots, 1)$, we obtain

$$\left.\frac{\partial r_i(\omega)}{\partial \omega_j}\right|_{\omega=\omega_0} = \frac{1}{2}[u_i^j(v_i^j - r_i u_i^j) + v_i^j(u_i^j - r_i v_i^j)], \tag{3.6}$$

$$\left.\frac{\partial \hat{a}_i(\omega)}{\partial \omega_j}\right|_{\omega=\omega_0} = f_1^j \hat{a}_1 + \cdots + f_p^j \hat{a}_p, \tag{3.7}$$

where U_i and V_i are defined as before and $U_i = (u_i^1, \ldots, u_i^n)^T$, $V_i = (v_i^1, \ldots, v_i^n)^T$. The f_k^j's ($k = 1, \ldots, p$) are defined as

$$f_k^j = \begin{cases} \dfrac{1}{r_i^2 - r_k^2}[r_i u_k^j(v_i^j - r_i u_i^j) + r_k v_k^j(u_i^j - r_i v_i^j)], & k \neq i, \\ \dfrac{1}{2}[1 - (u_i^j)^2], & k = i. \end{cases} \tag{3.8}$$

As indicated by Tanaka (1994) that the $\mathrm{EIC}(x_i, \hat{\theta})$ is equivalent to the partial derivative of $\hat{\theta}$ with respect to ω_i evaluated at $\omega_i = 1$, we can find that $\partial r_i(\omega)/\partial \omega_j|_{\omega=\omega_0}$ and $\partial \hat{a}_i(\omega)/\partial \omega_j|_{\omega=\omega_0}$ actually are equivalent to the empirical influence functions $\mathrm{EIC}((x_j^T, y_j^T), r_i)$ and $\mathrm{EIC}((x_j^T, y_j^T), \hat{a}_i)$ given by Romanazzi (1992), but the expressions here are simpler.

The vector $\{\partial r_i(\omega)/\partial \omega_j|_{\omega=\omega_0}, j = 1, \ldots, n\}$ indicates the direction in which the greatest local change for r_i occurs. From this direction, we can observe not only the influence of a single case but also the possible local group effect of several observations.

Let the $p \times n$ matrix M_i be $(\partial \hat{a}_i(\omega)/\partial \omega_1|_{\omega=\omega_0}, \ldots, \partial \hat{a}_i(\omega)/\partial \omega_n|_{\omega=\omega_0})$, denote the $n \times 1$ vector $(f_k^1, \ldots, f_k^n)^T$ as f_k, then $M_i = (\hat{a}_1, \ldots, \hat{a}_p)(f_1, \ldots, f_p)^T$. Suppose l_ω is any nonzero vector of unit length stemmed from $\omega_0 = (1, \ldots, 1)^T$ in R^n, then the local influence of case-weights for $\hat{a}_i$ in direction l_ω is $M_i l_\omega$. To assess the local influence in direction l_ω, we may adopt the same definition of relative norm as that in Section 2, that is

$$\begin{aligned}
\mathrm{RN}(M_i l_\omega) &= (M_i l_\omega)^T \hat{\Sigma}_{11}(M_i l_\omega) \\
&= l_\omega^T (M_i^T \hat{\Sigma}_{11} M_i) l_\omega \\
&= l_\omega^T (f_1, \ldots, f_p)(\hat{a}_1, \ldots, \hat{a}_p)^T \hat{\Sigma}_{11}(\hat{a}_1, \ldots, \hat{a}_p)(f_1, \ldots, f_p)^T l_\omega \\
&= l_\omega^T (f_1, \ldots, f_p)(f_1, \ldots, f_p)^T l_\omega \\
&= l_\omega^T F_{n\times n} l_\omega.
\end{aligned}$$

Thus we can use the influence matrix $F_{n\times n}$ in the same way as before. That is, the eigenvector corresponding to the largest eigenvalue of F is to be used for finding the most sensitive case-weights perturbation way which will lead to the greatest local influence, and each diagonal element of F indicates the local change of canonical vector arisen by the case-weight change of the corresponding observation.

4. Local influence for several test statistics

The study of influence on test statistics is not common in the literature. Jolliffe and Lukudu (1993) studied the case-deletion diagnostics on simple one-sample tests for means and variance. The influence of observations on the Pearson's goodness-of-fit and goodness-of-link tests in generalized linear models was recently investigated by Lee and Zhao (1996, 1997). In this section, we will investigate the local influence on several test statistics in CCA.

It is well known that canonical correlation analysis is concerned with the study of dependence and correlation structure between two sets of variables. When we test the hypothesis that the two sets of variables x and y are independent, four test statistics are commonly used (Muirhead (1982)). They are,

(1) Wilks' Λ statistic,

$$W = \frac{\det(\hat{\Sigma})}{\det(\hat{\Sigma}_{11})\det(\hat{\Sigma}_{22})} = \prod_{i=1}^{p}(1-r_i^2).$$

An approximate test of significance level α is to reject H_0 if $-n\rho\log(W) > c_f(\alpha)$, where $c_f(\alpha)$ denotes the upper $100\alpha\%$ quantile of a χ_f^2 distribution with degrees of freedom $f = pq$, n is the sample size and $\rho = 1-(p+q+3)/(2n)$. Denote $-n\rho\log(W)$ by T, which is given as $-n\rho\sum_{i=1}^{p}\log(1-r_i^2)$.

(2) Hotelling's trace statistic,

$$L_1 = \sum\frac{r_i^2}{1-r_i^2}.$$

(3) Pillai's trace statistic,

$$L_2 = \sum r_i^2.$$

(4) Roy's largest root statistic,

$$L_3 = r_1^2.$$

All the above test statistics are functions of canonical correlations. The local influence analysis of these statistics is not difficult using the previous results about canonical correlations.

4.1 *Additive perturbation scheme*

For Wilks' Λ statistic, we consider the local influence under the additive perturbation scheme for statistic $T = -n\rho\log(W) = -n\rho\sum_{i=1}^{p}\log(1-r_i^2)$. After some algebra, we obtain

$$\left.\frac{\partial T(\omega)}{\partial\eta}\right|_{\omega=\omega_0} = n\rho\left[\sum_{i=1}^{p}\frac{2r_i}{1-r_i^2}(\hat{a}_i^T\otimes(V_i - r_iU_i)^T)\right](S^X\otimes I)\operatorname{vec}(\tilde{L}_1) + n\rho\left[\sum_{i=1}^{p}\frac{2r_i}{1-r_i^2}(\hat{b}_i^T\otimes(U_i - r_iV_i)^T)\right](S^Y\otimes I)\operatorname{vec}(\tilde{L}_2). \tag{4.1}$$

For Hotelling's trace and Pillai's trace statistics, similar results can be obtained and they are presented as follows.

$$\left.\frac{\partial L_1(\omega)}{\partial\eta}\right|_{\omega=\omega_0} = \left[\sum_{i=1}^{p}\frac{2r_i}{(1-r_i^2)^2}(\hat{a}_i^T\otimes(V_i - r_iU_i)^T)\right](S^X\otimes I)\operatorname{vec}(\tilde{L}_1) + \left[\sum_{i=1}^{p}\frac{2r_i}{(1-r_i^2)^2}(\hat{b}_i^T\otimes(U_i - r_iV_i)^T)\right](S^Y\otimes I)\operatorname{vec}(\tilde{L}_2), \tag{4.2}$$

$$\left.\frac{\partial L_2(\omega)}{\partial\eta}\right|_{\omega=\omega_0} = \left[\sum_{i=1}^{p}(2r_i)(\hat{a}_i^T\otimes(V_i - r_iU_i)^T)\right](S^X\otimes I)\operatorname{vec}(\tilde{L}_1) + \left[\sum_{i=1}^{p}(2r_i)(\hat{b}_i^T\otimes(U_i - r_iV_i)^T)\right](S^Y\otimes I)\operatorname{vec}(\tilde{L}_2). \tag{4.3}$$

Roy's largest root statistic is the largest squared canonical correlation coefficient, and the local influence analysis can be found in previous two sections. We will not repeat it here.

Making use of (4.1), (4.2) and (4.3), we can detect the influential cases column by column. To compare the influence effects of different columns, we can take $S^X = [\operatorname{diag}(\hat{\Sigma}_{11})]^{1/2}$ and $S^Y = [\operatorname{diag}(\hat{\Sigma}_{22})]^{1/2}$, and then compare the maximum possible influence magnitude of each column.

4.2 *Case-weights perturbation scheme*

Under the case-weights perturbation scheme, the first order local influence of these test statistics also equal the empirical influence functions (EIC's) of them. We just list the results as follows:

$$\left.\frac{\partial T(\omega)}{\partial\omega_j}\right|_{\omega=\omega_0} = n\rho\sum_{i=1}^{p}\frac{r_iu_i^j(v_i^j - r_iu_i^j) + r_iv_i^j(u_i^j - r_iv_i^j)}{1-r_i^2}, \tag{4.4}$$

$$\left.\frac{\partial L_1(\omega)}{\partial\omega_j}\right|_{\omega=\omega_0} = \sum_{i=1}^{p}\frac{r_iu_i^j(v_i^j - r_iu_i^j) + r_iv_i^j(u_i^j - r_iv_i^j)}{(1-r_i^2)^2}, \tag{4.5}$$

$$\left.\frac{\partial L_2(\omega)}{\partial\omega_j}\right|_{\omega=\omega_0} = \sum_{i=1}^{p} r_iu_i^j(v_i^j - r_iu_i^j) + r_iv_i^j(u_i^j - r_iv_i^j). \tag{4.6}$$

These results will also be illustrated in the next section.

5. Example: fitness club data

The fitness club data is a typical example used for demonstrating canonical correlation analysis in the SAS manual (1990). In this data set, three physiological and three exercise variables were measured on 20 middle-aged men in a fitness club, we denote these physiological and exercise variables respectively as x_1, x_2, x_3 and y_1, y_2, y_3. Here we will do the influence analysis of the canonical correlations, canonical vectors and relevant test statistics.

For this data set, the canonical correlations are respectively $r_1 = 0.796$, $r_2 = 0.200$ and $r_3 = 0.071$. The third one is so small that we may regard it as unimportant and omit the influence analysis of it. Firstly, we will check which observations are influential to r_1 and r_2.

Under the additive perturbation scheme, Fig. 1 and Fig. 2 present the comparisons of two predicted residuals $VUi = V_i - r_iU_i$ and $UVi = U_i - r_iV_i$ for the first two canonical correlations r_1 and r_2. By the 2σ rule, it can be seen that for r_1, the physiological variables of case 19 and the exercise variables of case 14 have large local influence; for r_2, the influence of physiological variables of case 10 is not negligible. As for which columns of these observations are more crucial in bringing about a large influence to r_1 and r_2, we should investigate the vectors $(S^X\hat{a}_i)$ and $(S^Y\hat{b}_i)$. In this example, we have $(S^X\hat{a}_1)^T = (-0.775, 1.579, -0.059)$, $(S^Y\hat{b}_1)^T = (-0.349, -1.054, 0.716)$, $(S^X\hat{a}_2)^T = (1.884, -1.181, 0.231)$ and $(S^Y\hat{b}_2)^T = (0.376, -0.123, -1.062)$. From these results, we may conclude that the x_2 variable of case 19 and y_2 variable of case 14 exert the largest additive local influence to r_1, while both of the x_1 variable and x_2 variable of case 10 have a large such influence to r_2.

On the other hand, under the case-weights perturbation scheme, from Fig. 3, the most outstanding observations for r_1 are case 9 and case 14. Although cases 13, 15 and 19 are not as large as cases 9 and 14, it should be noted that they may have a joint influence effect with case 9 and case 14. Figure 3 indicates that the increase of weights on cases 9 and 14, and the decrease of weights on cases 13, 15 and 19 simultaneously will cause the greatest influence. The opposite effect of these two groups of observations can also be observed from Fig. 1. It can be easily verified that the sample correlation of $VU1$ and $UV1$ is exactly $-r_1$, thus from Fig. 1, it is obvious that existence of cases 13, 15 and 19 makes the linear relationship increase, while the existence of cases 9 and 14 makes it decrease. Furthermore we can see from Fig. 1 that the influence of case 9 to r_1 is mainly from the exercise variables of case 9, that is, from y_2 variable of case 9.

Figure 4 presents the influence analysis result of r_2 under the case-weights perturbation scheme. It indicates the same most influential observation as Fig. 2 does, thus as for r_2, physiological variables of case 10 should be investigated carefully.

Now we will conduct the influence analysis for canonical vector $\hat{a}_1$. Under the additive perturbation scheme, we first calculate the previously mentioned scaled maximum eigenvalues of F matrices. When the perturbation scale on every column is the same, the maximum possible influence magnitudes of different columns are respectively 0.105, 7.829, 0.279, 0.166, 0.010 and 0.007. But if the perturba-

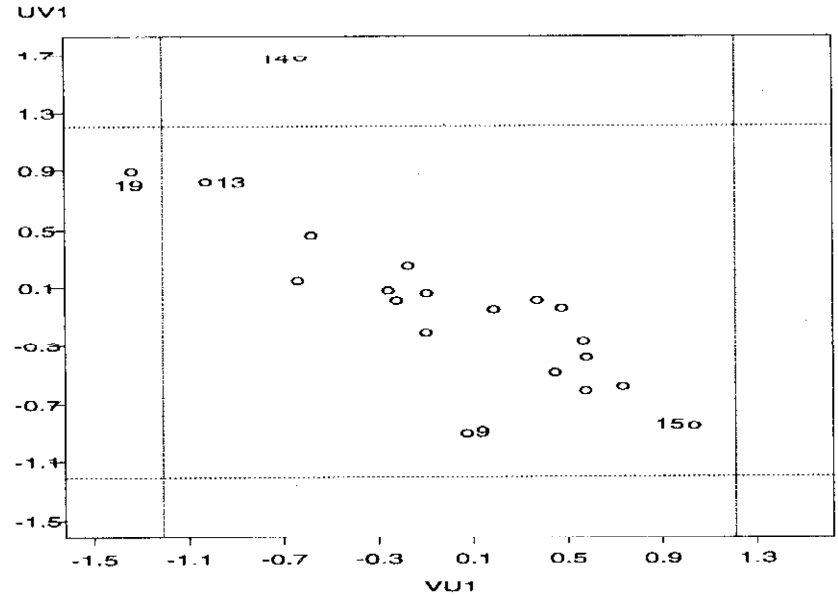

Fig. 1. Plot of two predicted residuals: UV1 and VU1 for the influence analysis of r_1 under the additive perturbation.

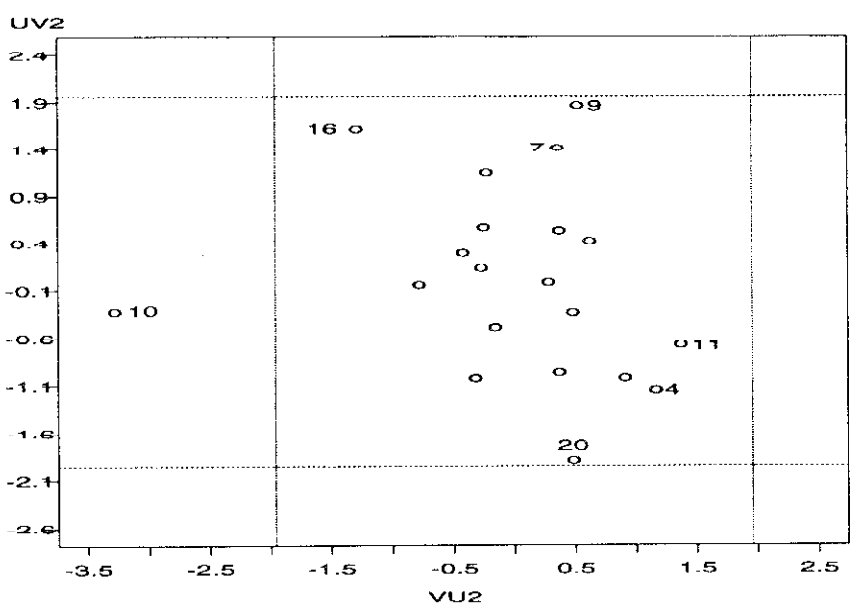

Fig. 2. Plot of two predicted residuals: UV2 and VU2 for the influence analysis of r_2 under the additive perturbation.

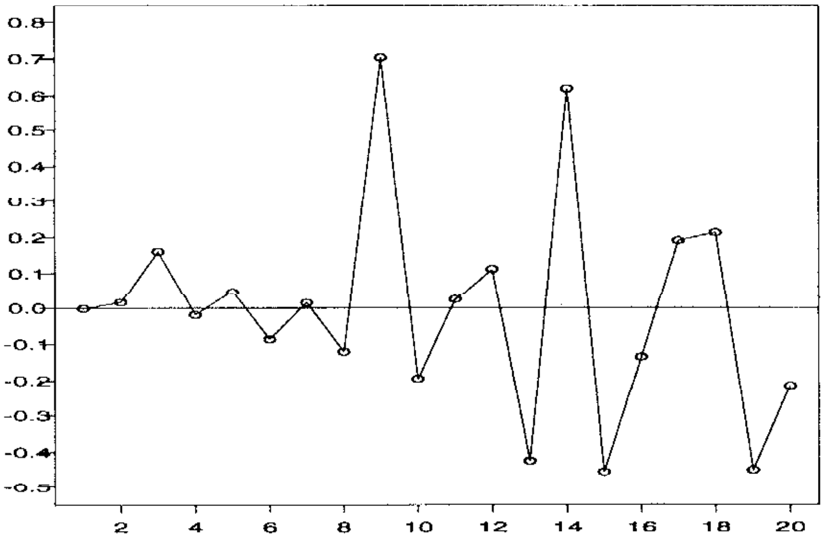

Fig. 3. Index plot for the influence analysis of r_1 under the case-weights perturbation.

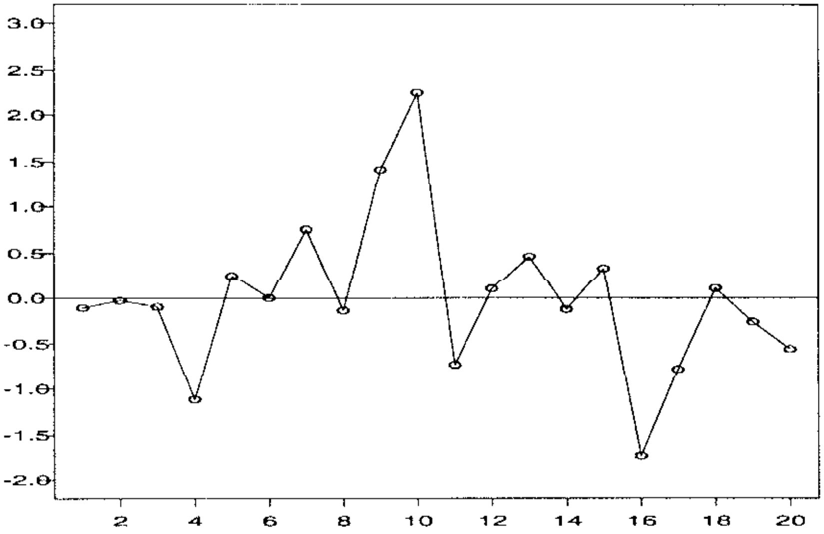

Fig. 4. Index plot for the influence analysis of r_2 under the case-weights perturbation.

tion scale on every column is proportional to the sample standard deviation of that column, the comparable scaled maximum eigenvalues are respectively 60.719, 76.253, 13.797, 4.401, 38.430 and 18.704. It seems that to $\hat{a}_1$ analyzing the influence of additive perturbation on x_1 and x_2 is much more important, the influence of other columns to $\hat{a}_1$ is comparatively minor. For perturbation on column x_1,

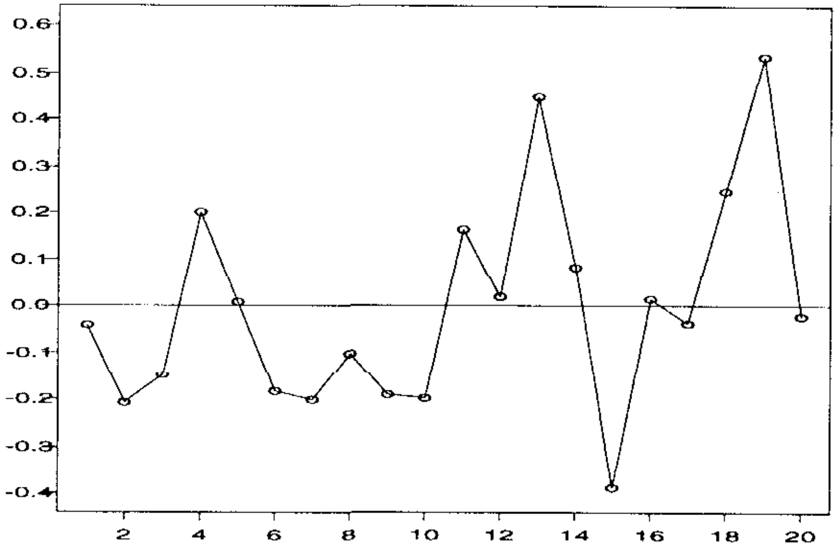

Fig. 5. The influence of additive perturbation on x_1 for $\hat{a}_1$: index plot of the eigenvector of F matrix.

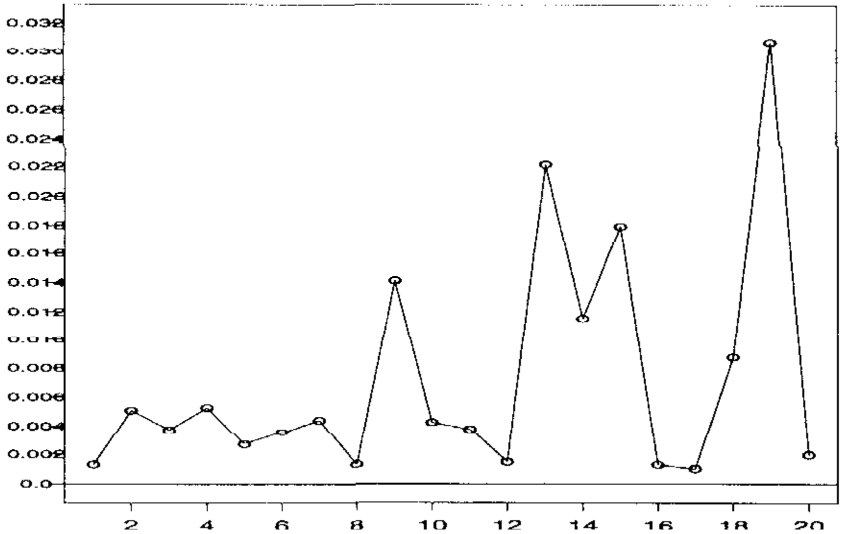

Fig. 6. The influence of additive perturbation on x_1 for $\hat{a}_1$: index plot of the diagonal elements of F matrix.

Fig. 5 gives out the index plot of the first eigenvector $l_{\max}$ of the F matrix. Cases 19, 13 and 15 look extreme. It can also be noted that case 15 has an opposite effect with case 19 and case 13. Figure 6 is the index plot of diagonal elements of the same F matrix for perturbation on column x_1. The influence of every single

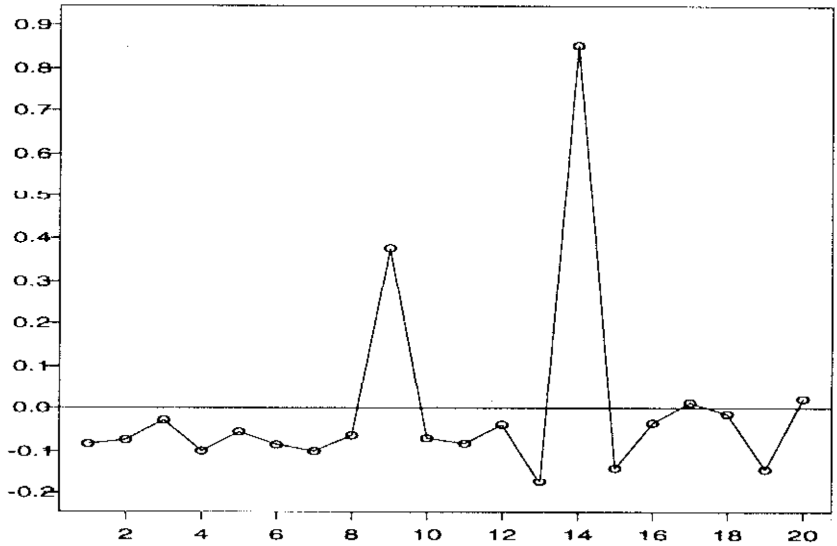

Fig. 7. Influence analysis for $\hat{a}_1$ under the case-weights perturbation: index plot of the eigenvector of F matrix.

observation is displayed and it can be seen that cases 19, 13 and 15 stand out from the other cases. Figure 6 can be regarded as a confirmation to Fig. 5 because they are consistent in this situation.

But for perturbation on x_2, this consistency is not present. From the first eigenvector of the F matrix which is obtained from adding perturbation on x_2, cases 13, 15 and 19 are found to be the most outstanding observations. But these cases are not the same as those indicated by the diagonal elements of the F matrix for perturbation on column x_2. The most outstanding cases detected by the diagonal elements method are cases 9 and 14. In such situation, it seems reasonable to regard case 9 and case 14 as most influential, because the first eigenvector of the F matrix may be less informative unless the measurement error happened to be in the way as indicated by the first eigenvector. Nevertheless, these two groups of observations $(13, 15, 19)$ and $(9, 14)$ are found to be influential under different aspects as discussed above. For brevity, the index plots are omitted.

Under the case-weights perturbation scheme, we also get an influence matrix F. The conclusions from the analyses by the first eigenvector and the diagonal elements of F matrix are quite consistent. Here we just present the index plot of the first eigenvector in Fig. 7. Both the first eigenvector and the diagonal elements of F matrix indicate that case 14 is no doubt the most influential point to $\hat{a}_1$. From Fig. 7, it is also obvious that cases 14 and 9 have a great joint influence effect to $\hat{a}_1$. It can be anticipated that increasing or decreasing simultaneously the weights of cases 14 and 9 in the estimation of covariance matrix will give rise to a great change to the result of $\hat{a}_1$.

When determining if the physiological variables are related in any way to the exercise variables is our objective for this data set, we find that the independence

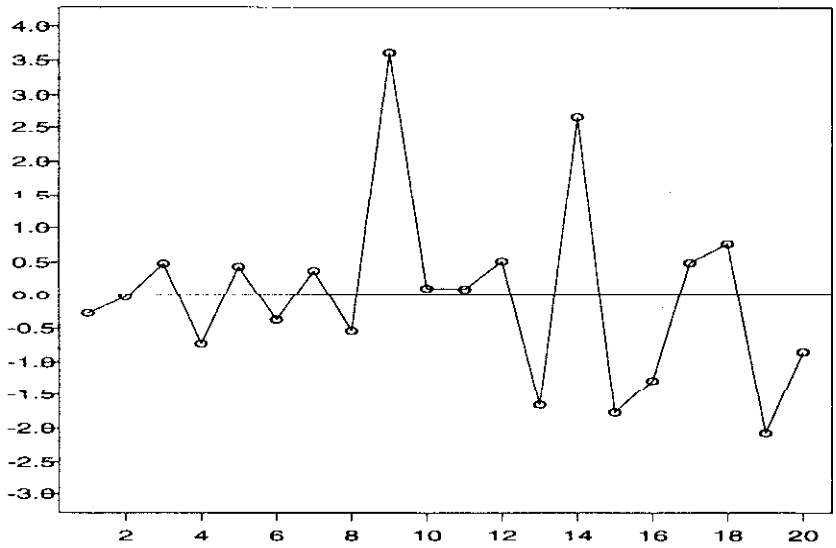

Fig. 8. Index plot for the influence analysis of Wilks' Λ statistic under the case-weights perturbation.

test results given by the four test statistics are not all the same. From the Wilks' Λ, the p value is about 0.0635, which results in marginally accepting the null hypothesis at the 5% significance level. That is, the two group of variables are independent. The p-values for the Pillai's trace and Hotelling's trace statistics are given as 0.1551 and 0.0238 respectively. However, the lower bound of the p-value given by the Roy's largest root is as small as 0.0009.

The influence of local perturbation to the test statistics is also examined. Under the case-weights perturbation scheme, all the influence plots of the Wilks' Λ, Hotelling's trace, Pillai's trace and Roy's largest root statistics indicate the same most influential observations. Here we just present the influence plot of the Wilks' Λ statistic in Fig. 8, and the influence plot of Roy's largest root statistic may be referred to Fig. 3. From these plots, it is clear that case 9 (and case 14) is commonly the most influential case to every statistic. In this situation, case 9 is worthy to be investigated carefully.

If we assign zero weight to the case of interest and weight one to the other cases, the case-weights perturbation scheme is just equivalent to case deletion. So we recompute the statistics without case 9. The test results obtained from the Wilks' Λ, Hotelling's trace and Pillai's trace statistics are quite consistent, the p-values corresponding to them are respectively 0.179, 0.113 and 0.282, which means we cannot reject the null hypothesis at the 5% significance level. The lower bound of the p-value given by Roy's largest root increases from 0.0009 to 0.0047. It seems that case 9 is more influential to the conclusion of the Wilks' Λ, Hotelling's trace and Pillai's trace tests, but less so to that of the Roy's largest root.

6. Concluding remarks

We have derived the first order local influence diagnostics under the additive perturbation scheme and the case-weights perturbation scheme for the canonical correlations, canonical vectors and several test statistics in CCA. In the case when the objective function is not a scalar, such as the canonical vectors, we define a scale invariant norm to measure the influence. Both the direction cosines corresponding to the largest norm and the diagonal elements of the same influence matrix prove to be useful in assessing local sensitivity.

The diagnostics under the case-weights perturbation actually give the same expressions as the empirical influence curves. From this point of view, we may interpret such diagnostics as not only the influence of every case but also the most sensitive direction of case-weights. Under the additive perturbation scheme, the influence analysis of canonical correlations is simplified to just observing two predicted residuals, and by comparing the different influential effects of properly scaled perturbation on different variables, the influence analysis for each canonical vector is also reduced to just investigating the influence of perturbation on some variables.

However, it is noted that, based on the notion that reasonable conclusions should not depend critically on the unusual aspects of the data, the above proposed methods are centered on the identification of various influential aspects of the data. Since diagnostic methods, especially the local influence approach, are often exploratory, further comparison procedures may be needed in practical applications. For example, if one sort of perturbation is identified as influential, assessing the importance and possibility of this specific perturbation is necessary. More specifically, say, if a specified perturbation involves slightly increasing the weights of a large part of points and simultaneously largely decreasing the weights of several points which are sporadically distributed far from the bulk of other points, it will, of course, be important to investigate the actual influence of such perturbation due to some fixed perturbation scales. An obvious way for dealing with this is to compare the results derived from the original and perturbed data. Generally, it seems impossible to set up some specific universal cut-off points for the proposed local influence measures, although we did set up some reference cut-off points in the case of identifying influential points for canonical correlations under additive perturbation scheme. More discussion about the 'benchmark' or 'critical value' in the context of local influence is referred to the discussion of Cook's seminar paper (Cook (1986), pp. 156–169).

Acknowledgements

The authors are most grateful to the editor and the referee for valuable comments and suggestions which improved the presentation of the paper considerably. The second author was partly supported by a RGC grant (HKU 7134/98H).

References

Cook, R. D. (1986). Assessment of local influence (with discussion), *J. Roy. Statist. Soc. Ser. B*, **48**, 133–169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, New York.

Fung, W. K. and Kwan, C. W. (1997). A note on local influence based on the normal curvature, *J. Roy. Statist. Soc. Ser. B*, **59**, 839–843.

Jolliffe, I. T. and Lukudu, S. G. (1993). The influence of a single observation on some standard test statistics, *Journal of Applied Statistics*, **20**, 143–151.

Lawrance, A. J. (1991). Local and deletion influence, *Direction in Robust Statistics and Diagnostics, Part I*, 141–157.

Lee, A. H. and Zhao, Y. (1996). Sensitivity of Pearson's goodness-of-fit statistic in generalized linear models, *Comm. Statist. Theory Methods*, **25**, 143–157.

Lee, A. H. and Zhao, Y. (1997). Assessing influence on the goodness-of-link in generalized linear models, *Statist. Probab. Lett.* (to appear).

Muirhead, R. M. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Radhakrishnan, R. and Kshirsagar, A. M. (1981). Influence functions for certain parameters in multivariate analysis, *Comm. Statist. Theory Methods*, **10**, 515–529.

Romanazzi, M. (1992). Influence in canonical correlation analysis, *Psychometrika*, **57**, 237–259.

SAS/STAT User's Guide (1990). Version 6, Vol. 1, SAS Institute Inc., Cary, North Carolina.

Tanaka, Y. (1994). Recent advance in sensitivity analysis in multivariate statistical methods, *J. Japanese Soc. Comput. Statist.*, **7**, 1–25.

Wu, X. and Luo, Z. (1993). Second-order approach to local influence, *J. Roy. Statist. Soc. Ser. B*, **55**, 929–936.