British Journal of Mathematical and Statistical Psychology (2002), 55, 177–192

www.bps.org.uk

Influential observations in the estimation of mean vector and covariance matrix

Wai-Yin Poon[1]* and Yat Sun Poon[2]
[1]The Chinese University of Hong Kong
[2]University of California at Riverside, USA

Statistical procedures designed for analysing multivariate data sets often emphasize different sample statistics. While some procedures emphasize the estimates of both the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, others may emphasize only one of these two sample quantities. In effect, while an unusual observation in a data set has a deleterious impact on the results from an analysis that depends heavily on the covariance matrix, its effect when dependence is on the mean vector may be minimal. The aim of this paper is to develop diagnostic measures for identifying influential observations of different kinds. Three diagnostic measures, based on the local influence approach, are constructed to identify observations that exercise undue influence on the estimate of $\boldsymbol{\mu}$, of $\boldsymbol{\Sigma}$, and of both together. Real data sets are analysed and results are presented to illustrate the effectiveness of the proposed measures.

1. Introduction

The multivariate normal distribution is a common distribution in analysing multivariate data. The distribution is specified by a mean vector $\boldsymbol{\mu}$ together with a symmetric and positive definite matrix Σ. The estimates of these parameters are usually obtained by the method of maximum likelihood, and various multivariate statistical procedures have been developed based on one or both of these sample quantities. However, it is well known that the accuracy of the estimates is affected by unusual observations in the data set, so-called outliers.

Outlier identification is an important task in data analysis because outlying observations can have a disproportionate influence on statistical analysis. When such distortion occurs, the outliers are regarded as influential observations. However, an observation

* *Requests for reprints should be addressed to Wai-Yin Poon, Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: wypoon@cuhk.edu.hk).*

that substantially affects the results of one analysis may have little influence on another because the statistical procedures may emphasize different sample quantities. Therefore, influential observations must be identified according to context.

The identification of influential observations or outliers of different kinds is well documented in the literature on the regression model; however, less has been achieved for multivariate analysis. Although the detection of multivariate outliers has received considerable attention (see Rousseeuw & van Zomeren, 1990; Hadi, 1992, 1994; Fung, 1993; Atkinson & Mulira, 1993; Atkinson, 1994; Rocke & Woodruff, 1996; Poon, Lew, & Poon, 2000), the emphasis has been on the identification of location outliers that influence the estimate of $\boldsymbol{\mu}$. Observations that influence the estimate of Σ are addressed by studying their effect on the measures used to identify the location outliers. In view of the fact that statistical procedures may depend heavily on different sample quantities, methods for identifying influential observations of different kinds in multivariate data sets are necessary, and the aim of this paper is to develop diagnostic measures for this purpose.

There are two major paradigms in the influence analysis literature: the deletion approach and the local influence approach. The deletion approach assesses the effect of dropping a case on a chosen statistical quantity, and a typical diagnostic measure is Cook's distance (Cook, 1977). The concept of Cook's distance was first introduced in the context of linear regression and was subsequently generalized to other statistical models (McCullagh & Nelder, 1983; Bruce & Martin, 1989). While intuitively convincing measures of this kind have become very popular in influence analysis, it is also well known that they are vulnerable to masking effects that arise in the presence of several unusual observations. Diagnostic measures derived from deleting a group of cases are also well documented in the literature, but their practicality is in doubt because of combinatorial and computational problems.

On the other hand, the local influence approach is well known for its ability to detect joint influence. In this approach diagnostic measures are derived by examining the consequence of an infinitesimal perturbation on the relevant quantity (Belsley, Kuh, & Welsch, 1980; Pregibon, 1981); a general method for assessing the influence of local perturbation was proposed by Cook (1986). The approach starts with a carefully chosen perturbation on the model under consideration and then uses differential geometry techniques to assess the behaviour of the influence graph of the induced likelihood displacement function. In particular, the normal curvature $C_{\mathbf{l}}$ along a direction $\mathbf{l}$, where $\mathbf{l}^T\mathbf{l} = 1$, at the critical point of the influence graph of the displacement function is used as the diagnostic quantity. A large value of $C_{\mathbf{l}}$ is an indication of strong local influence and the corresponding direction $\mathbf{l}$ indicates how to perturb the postulated model to obtain the greatest change in the likelihood displacement. Moreover, the conformal normal curvature $B_{\mathbf{l}}$ transforms the normal curvature onto the unit interval and has been demonstrated to be another effective influence measure (Poon & Poon, 1999).

The current study uses the local influence approach to develop diagnostic measures for identifying influential observations that affect the estimate of the mean, of the covariance matrix, and of both. These measures inherit the nice property of the local influence approach in its ability in detecting joint influence, hence multiple outliers that share joint influence can be identified simultaneously. It is worth noting that while a 'masking' effect arises in the presence of multiple outliers, there are two distinct notions of masking effect, namely the joint influence and the conditional influence (Lawrance, 1995; Poon & Poon, 2001). The proposed diagnostic measures, like other measures developed by the local influence approach, are capable of addressing masking effects under the category of joint influence but not of conditional influence.

Particulars about the local influence approach will be given in the next section where we will also use the approach to develop measures for identifying influential observations of different kinds The results of analysing real data sets using the proposed measures are then presented as illustrations.

2. Influential observations in the estimation of normal distribution parameters

2.1. Case-weights perturbation

Let $\mathbf{X}$ be a $p \times 1$ random vector distributed as $N(\boldsymbol{\mu}, \Sigma)$, where $\Sigma = \{\sigma_{\alpha\beta}\}$ is a symmetric and positive definite matrix, and let $\{\mathbf{x}_i, i = 1, \ldots, n\}$ be a random sample. The maximum likelihood estimates $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ and $\hat{\Sigma}$ of Σ are obtained by maximizing the log-likelihood given by

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(-\frac{1}{2}p\ \log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right), \tag{1}$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}^{\mathrm{T}}, \boldsymbol{\sigma}^{\mathrm{T}})^{\mathrm{T}}$ is a $q = p + p^*$ vector storing the elements in $\boldsymbol{\mu}$ and the lower triangular elements of Σ, and $p^* = p(p+1)/2$. It can be shown that the maximum likelihod estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\mu}}^{\mathrm{T}}, \hat{\boldsymbol{\sigma}}^{\mathrm{T}})^{\mathrm{T}}$ of $\boldsymbol{\theta}$ is given by (Anderson, 1958)

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{\Sigma_{i=1}^{n}\mathbf{x}_i}{n}, \qquad \hat{\Sigma} = \mathbf{S} = \frac{\Sigma_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}}}{n}. \tag{2}$$

In order to assess the influence of individual observations on the estimate of $\boldsymbol{\theta}$, we follow Cook (1986) and introduce the case-weights perturbation to the log-likelihood. The resulting perturbed likelihood is given by

$$L(\boldsymbol{\theta}, |\boldsymbol{\omega}) = \sum_{i=1}^{n}\omega_i\left(-\frac{1}{2}p\ \log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right), \tag{3}$$

where $\omega_i, i = 1, \ldots, n$, are perturbation parameters and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\mathrm{T}}$ is defined on a relevant perturbation space Ω of $\mathbb{R}^n$. For example, Ω may be the space such that $0 \leq \omega_i \leq 1$ for all i. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ be the quantities that maximize (1) and (3), respectively; then the discrepancy between them can be measured by the likelihood displacement function

$$f(\boldsymbol{\omega}) = 2(L(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0) - L(\hat{\boldsymbol{\theta}}_{\omega}|\boldsymbol{\omega}_0)). \tag{4}$$

This function has its minimum value at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and we have $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$ if $\boldsymbol{\omega}_0 = \mathbf{1}$, which is an $n \times 1$ vector of 1s. When the perturbation specified by $\boldsymbol{\omega}$ causes a substantial deviation of $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ from $\hat{\boldsymbol{\theta}}$, a substantial deviation of the function $f(\boldsymbol{\omega})$ from $f(\boldsymbol{\omega}_0)$ is induced. Therefore, examining the changes in $f(\boldsymbol{\omega})$ as a function of $\boldsymbol{\omega}$ will enable the identification of influential perturbations that in turn will disclose influential observations. Cook (1986) proposed quantifying such changes by the normal curvature C_l of the influence graph g of $f(\boldsymbol{\omega})$ along a direction $\mathbf{l}$ at the optimal point $\boldsymbol{\omega}_0$, where $\mathbf{l}$ with $\mathbf{l}^{\mathrm{T}}\mathbf{l} = 1$ defines a direction for a straight line in Ω passing through $\boldsymbol{\omega}_0$. A large value of C_l is an indication that the perturbation along the corresponding direction $\mathbf{l}$ induces a considerable change in the likelihood displacement.

2.2. Normal curvature as an influential measure

Let $\ddot{\mathbf{L}}$ and Δ be $q \times q$ and $q \times n$ matrices with elements respectively given by

$$\ddot{L}_{ij} = \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega}_0)}{\partial\theta_i\partial\theta_j}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \qquad \Delta_{ij} = \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\theta_i\partial\omega_j}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \tag{5}$$

Cook (1986) noted that the normal curvature of g in a direction $\mathbf{l}$ at the point $\boldsymbol{\omega}_0$ could be computed by

$$C_{\mathbf{l}} = -2\mathbf{l}^{\mathrm{T}}(\Delta^{\mathrm{T}}\ddot{\mathbf{L}}^{-1}\Delta)\mathbf{l}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -2\mathbf{l}^{\mathrm{T}}\ddot{\mathbf{F}}\mathbf{l}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \tag{6}$$

The direction $\mathbf{l}_{max}$ along which the greatest change in the likelihood displacement is observed identifies the most influential observations. The direction $\mathbf{l}_{max}$ is that which gives $C_{max} = \max_{\mathbf{l}} C_{\mathbf{l}}$, and C_{max} together with $\mathbf{l}_{max}$ are the largest eigenvalue and associated eigenvector of the symmetric matrix $-2\ddot{\mathbf{F}}$ in (6).

When $C_{\mathbf{l}}$ is defined on an unbounded interval and it may be difficult to judge its magnitude, the conformal normal curvature is a one-to-one transformation of the normal curvature onto the unit interval. Along the direction $\mathbf{l}$ at the critical point $\boldsymbol{\omega}_0$, the conformal normal curvature is given by

$$B_{\mathbf{l}} = -\frac{\mathbf{l}^{\mathrm{T}}\Delta^{\mathrm{T}}\ddot{\mathbf{L}}^{-1}\Delta\mathbf{l}}{\sqrt{\mathrm{tr}\,(\Delta^{\mathrm{T}}\ddot{\mathbf{L}}^{-1}\Delta)^2}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{\mathbf{l}^{\mathrm{T}}\ddot{\mathbf{F}}\mathbf{l}}{\sqrt{\mathrm{tr}\,\ddot{\mathbf{F}}^2}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \tag{7}$$

Let $\mathbf{E}_j$, $j = 1,\ldots,n$, be vectors of the n-dimensional standard basis. Poon and Poon (1999) demonstrated that $B_{\mathbf{E}_j} = B_j$, $j = 1,\ldots,n$, are effective measures for identifying the influential perturbation parameters when C_{max} is sufficiently large. Moreoever, the computation of B_j, $j = 1,\ldots,n$, is easy because B_j is the jth diagonal element of the matrix

$$-\frac{\ddot{\mathbf{F}}}{\sqrt{\mathrm{tr}\,\ddot{\mathbf{F}}^2}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{\Delta^{\mathrm{T}}\ddot{\mathbf{L}}^{-1}\Delta}{\sqrt{\mathrm{tr}\,(\Delta^{\mathrm{T}}\ddot{\mathbf{L}}^{-1}\Delta)^2}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \tag{8}$$

Clearly, to develop diagnostic measures for uncovering observations that influence the estimates of both $\boldsymbol{\mu}$ and Σ, it is necessary to compute the matrix $\ddot{\mathbf{F}}$. We discuss such computation in the next subsection.

2.3. Observations influencing the estimates of $\boldsymbol{\mu}$ and Σ

There are two components in the matrix $\ddot{\mathbf{F}}$: the $q \times q$ matrix $\ddot{\mathbf{L}}$ and the $q \times n$ matrix Δ. Because $-\ddot{\mathbf{L}}$ is the observed information for the postulated model and the maximum likelihood estimates $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ and $\hat{\boldsymbol{\sigma}}$ of $\boldsymbol{\sigma}$ are statistically independent, $\ddot{\mathbf{L}}^{-1}$ is a diagonal block matrix given by

$$\ddot{\mathbf{L}}^{-1} = \begin{pmatrix} -\mathrm{Cov}(\hat{\boldsymbol{\mu}}) & \mathbf{0} \\ \mathbf{0} & -\mathrm{Cov}(\hat{\boldsymbol{\sigma}}) \end{pmatrix} = \begin{pmatrix} \{\ddot{\mathbf{L}}_{ab}^{-1}\} & \mathbf{0} \\ \mathbf{0} & \{\ddot{\mathbf{L}}_{(\alpha\beta)(\gamma\rho)}^{-1}\} \end{pmatrix}, \tag{9}$$

where $\mathrm{Cov}(\hat{\boldsymbol{\mu}})$ is a $p \times p$ matrix storing the covariance matrix of $\hat{\boldsymbol{\mu}}$ and $\mathrm{Cov}(\hat{\boldsymbol{\sigma}})$ is a $p^* \times p^*$ matrix storing the covariance matrix of $\hat{\boldsymbol{\sigma}}$. For the sake of clarity, we denote an element in $-\mathrm{Cov}(\hat{\boldsymbol{\mu}})$ by $\ddot{\mathbf{L}}_{ab}^{-1}$ when it relates to the covariance between the ath and bth elements $\hat{\mu}_a$ and $\hat{\mu}_b$ of $\hat{\boldsymbol{\mu}}$, and an element in $-\mathrm{Cov}(\hat{\boldsymbol{\sigma}})$ by $\ddot{\mathbf{L}}_{(\alpha\beta)(\gamma\rho)}^{-1}$ when it relates to the covariance between $\hat{\sigma}_{\alpha\beta}$ and $\hat{\sigma}_{\gamma\rho}$. Using such notation, we have (Anderson, 1958)

$$\ddot{\mathbf{L}}^{-1}_{ab} = -(\mathrm{Cov}(\hat{\boldsymbol{\mu}}))_{ab} = -\mathrm{Cov}(\hat{\mu}_a, \hat{\mu}_b) = -\frac{1}{n}\sigma_{ab}, \tag{10}$$

and

$$\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)} = -(\mathrm{Cov}(\hat{\boldsymbol{\sigma}}))_{(\alpha\beta)(\gamma\rho)} = -\mathrm{Cov}(\hat{\sigma}_{\alpha\beta}, \hat{\sigma}_{\gamma\rho}) = -\frac{1}{n}(\sigma_{\alpha\gamma}\sigma_{\beta\rho} + \sigma_{\alpha\rho}\sigma_{\beta\gamma}). \tag{11}$$

Moreover, we denote the elements in the ith column of Δ by

$$\Delta_{ai} = \frac{\partial^2 L}{\partial \mu_a \partial \omega_i} \quad \text{or} \quad \Delta_{(\alpha\beta)i} = \frac{\partial^2 L}{\partial \sigma_{\alpha\beta}\partial\omega_i}, \tag{12}$$

depending on whether they correspond to μ_a in $\boldsymbol{\mu}$ or $\sigma_{\alpha\beta}$ in Σ, respectively. Using (6) and (9), the (i, j)th element of the matrix $\ddot{\mathbf{F}}$ becomes

$$F_{ij} = \sum_{1\le a,b\le p} \Delta_{ai}\ddot{\mathbf{L}}^{-1}_{ab}\Delta_{bj} + \sum_{\substack{1\le\beta\le\alpha\le p\\ 1\le\rho\le\gamma\le p}} \Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)}\Delta_{(\gamma\rho)j}\,. \tag{13}$$

By (3),

$$\frac{\partial L}{\partial\omega_i} = -\frac{1}{2}p\,\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu}). \tag{14}$$

Let $\mathbf{e}_a$ be a $p \times 1$ vector with 1 as its ath element and zeros elsewhere, let $\mathbf{y}_i = \mathbf{x}_i - \boldsymbol{\mu}$ be a $p \times 1$ vector with y_{bi} as its bth coordinate, and denote the (a, b)th element of Σ^{-1} by σ^{-1}_{ab}. From (14),

$$\Delta_{ai} = \frac{\partial^2 L}{\partial\mu_a\partial\omega_i} = \mathbf{y}_i^{\mathrm{T}}\Sigma^{-1}\mathbf{e}_a = \sum_b y_{bi}\sigma^{-1}_{ba} = \sum_b \sigma^{-1}_{ab}\,y_{bi}\,. \tag{15}$$

Therefore, the first summand on the right-hand side of (13) becomes

$$\sum_{a,b,c,d}\sigma^{-1}_{ac}\,y_{ci}\left(-\frac{1}{n}\sigma_{ab}\right)\sigma^{-1}_{bd}\,y_{dj} = -\frac{1}{n}\sum_{c,d} y_{ci}\sigma^{-1}_{cd}\,y_{dj} = -\frac{1}{n}\,\mathbf{y}_i^{\mathrm{T}}\Sigma^{-1}\,\mathbf{y}_j\,. \tag{16}$$

The second summand on the right-hand side of (13) can be obtained numerically by the expression in (11) and the following expression derived from (14):

$$\Delta_{(\alpha\beta)i} = \frac{\partial^2 L}{\partial\sigma_{\alpha\beta}\partial\omega_i} = -\frac{1}{2}\frac{\partial\,\log|\Sigma|}{\partial\sigma_{\alpha\beta}} - \frac{1}{2}\mathbf{y}_i^{\mathrm{T}}\frac{\partial\Sigma^{-1}}{\partial\sigma_{\alpha\beta}}\,\mathbf{y}_i = -\left(\sigma^{-1}_{\alpha\beta} - \sum_{a,b}\sigma_{\alpha a}\,y_{ai}\sigma^{-1}_{\beta b}\,y_{bi}\right). \tag{17}$$

It is also possible to compute the second summand as follows:

$$\begin{aligned}
\sum_{\beta\le\alpha,\rho\le\gamma}\Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)}\Delta_{(\gamma\rho)j} &= \sum_{\alpha,\gamma}\Delta_{(\alpha\alpha)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\alpha)(\gamma\gamma)}\Delta_{(\gamma\gamma)j} + \sum_{\beta<\alpha,\rho<\gamma}\Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)}\Delta_{(\gamma\rho)j}\\
&\quad + \sum_{\alpha,\rho<\gamma}\Delta_{(\alpha\alpha)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\alpha)(\gamma\rho)}\Delta_{(\gamma\rho)j} + \sum_{\beta<\alpha,\gamma}\Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\gamma)}\Delta_{(\gamma\gamma)j}\\
&= \frac{1}{4}\left\{\sum_{\alpha,\gamma}\Delta_{(\alpha\alpha)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\alpha)(\gamma\gamma)}\Delta_{(\gamma\gamma)j} + \sum_{\alpha,\gamma,\rho}\Delta_{(\alpha\alpha)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\alpha)(\gamma\rho)}\Delta_{(\gamma\rho)j}\right.\\
&\quad \left. + \sum_{\alpha,\beta,\gamma}\Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\gamma)}\Delta_{(\gamma\gamma)j} + \sum_{\alpha,\beta,\gamma,\rho}\Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)}\Delta_{(\gamma\rho)j}\right\}.
\end{aligned} \tag{18}$$

Using (11) as well as (17), we obtain the following expressions for the summands of (18):

$$\sum_{\alpha,\gamma} \Delta_{(\alpha\alpha)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\alpha)(\gamma\gamma)}\Delta_{(\gamma\gamma)j} = -\frac{2}{n}\sum_{\alpha}\sum_{\gamma}\left\{\sigma^{-1}_{\alpha\alpha}\sigma^{2}_{\alpha\gamma}\sigma^{-1}_{\gamma\gamma} - \sigma^{-1}_{\alpha\alpha}\sigma^{2}_{\alpha\gamma}\left[\left(\sum_{a}\sigma^{-1}_{\gamma a}\,y_{aj}\right)^{2} + \left(\sum_{a}\sigma^{-1}_{\gamma a}\,y_{ai}\right)^{2}\right] + \sigma^{2}_{\alpha\gamma}\left(\sum_{a}\sigma^{-1}_{\alpha a}\,y_{ai}\right)^{2}\left(\sum_{a}\sigma^{-1}_{\gamma a}\,y_{aj}\right)^{2}\right\}; \tag{19}$$

$$\sum_{\alpha,\gamma,\rho} \Delta_{(\alpha\alpha)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\alpha)(\gamma\rho)}\Delta_{(\gamma\rho)j} = -\frac{2}{n}\sum_{\alpha}\left\{[\sigma^{-1}_{\alpha\alpha} - y^{2}_{\alpha j}]\left[\sigma^{-1}_{\alpha\alpha} - \left(\sum_{a}\sigma^{-1}_{\alpha a}\,y_{ai}\right)^{2}\right]\right\}; \tag{20}$$

$$\sum_{\alpha,\beta,\gamma} \Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\gamma)}\Delta_{(\gamma\gamma)j} = -\frac{2}{n}\sum_{\gamma}\left\{[\sigma^{-1}_{\gamma\gamma} - y^{2}_{\gamma i}]\left[\sigma^{-1}_{\gamma\gamma} - \left(\sum_{a}\sigma^{-1}_{\gamma a}\,y_{aj}\right)^{2}\right]\right\}; \tag{21}$$

$$\sum_{\alpha,\beta,\gamma,\rho} \Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)}\Delta_{(\gamma\rho)j} = -\frac{2}{n}\{p - \mathbf{y}_j^{\mathrm{T}}\Sigma^{-1}\mathbf{y}_j - \mathbf{y}_i^{\mathrm{T}}\Sigma^{-1}\mathbf{y}_i + (\mathbf{y}_i^{\mathrm{T}}\Sigma^{-1}\mathbf{y}_j)^2\}. \tag{22}$$

Using (13), (16), and (19) to (22), it is possible to compute $\ddot{\mathbf{F}}$ and hence the diagnostic measures. Specifically, observations that unduly influence the estimates of the mean and the covariance matrix can be isolated by examining $\mathbf{l}_{\max}$ or B_j, $j = 1, \ldots, n$. That is, the elements with large B_j values or large magnitudes in $\mathbf{l}_{\max}$ are the group of influential observations.

Note that observations so identified are influential on the estimates of both $\boldsymbol{\mu}$ and Σ. However, as there may be observations in the data set that influence the estimate of $\boldsymbol{\mu}$ but not the estimate of Σ, or vice versa, it is also of interest to develop diagnostic measures for identifying these.

2.4. Observations influencing the estimate of $\boldsymbol{\mu}$ or Σ

The diagnostic measures developed in Section 2.3 are developed based on the influence graph given in (4) and hence the effects of the perturbation on estimates of all parameters in $\boldsymbol{\theta}$ are taken into account. When the effects on only a subset of the parameters is of interest, Cook (1986) demonstrated that the effects could be assessed by examining the normal curvature of the influence graph of an objective function deduced from (4). Specifically, if one is only interested in the effects on the estimate of $\boldsymbol{\mu}$, one can examine the normal curvature given by

$$C^{\boldsymbol{\mu}}_{\mathbf{l}} = -2\mathbf{l}^{\mathrm{T}}(\Delta^{\mathrm{T}}\ddot{\mathbf{L}}^{\boldsymbol{\mu}})^{-1}\Delta)\mathbf{l}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -2\mathbf{l}^{\mathrm{T}}\ddot{\mathbf{F}}^{\boldsymbol{\mu}}\mathbf{l}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}, \tag{23}$$

where

$$(\ddot{\mathbf{L}}^{\boldsymbol{\mu}})^{-1} = \begin{pmatrix} -\mathrm{Cov}(\hat{\boldsymbol{\mu}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \{\ddot{\mathbf{L}}^{-1}_{ab}\} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{24}$$

and Δ is as in Section 2.3 (see (12)). Using (10), (12), (15) and (16), the (i, j)th element of $\ddot{\mathbf{F}}^{\boldsymbol{\mu}}$ is found to be

$$\ddot{\mathbf{F}}^{\boldsymbol{\mu}}_{ij} = \sum_{1\le a,b\le p} \Delta_{ai}\ddot{\mathbf{L}}^{-1}_{ab}\Delta_{bj} = -\frac{1}{n}\mathbf{y}_i^{\mathrm{T}}\Sigma^{-1}\mathbf{y}_j. \tag{25}$$

By similar arguments to those given in Section 2.2, observations that exert a disproportionate influence to the estimate of $\boldsymbol{\mu}$ can be located by examining the eigenvector $\mathbf{l}^{\boldsymbol{\mu}}_{\max}$ associated with the largest eigenvalue of $-2\ddot{\mathbf{F}}^{\boldsymbol{\mu}}$ or by examining the diagonal elements $B^{\boldsymbol{\mu}}_j, j = 1, \ldots, n$, of the matrix $-\ddot{\mathbf{F}}^{\boldsymbol{\mu}}/\sqrt{\mathrm{tr}\,(\ddot{\mathbf{F}}^{\boldsymbol{\mu}})^2}$, where the unknowns in $\ddot{\mathbf{F}}^{\boldsymbol{\mu}}$ are evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

Similarly, let

$$(\ddot{\mathbf{L}}^{\boldsymbol{\sigma}})^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & -\mathrm{Cov}(\hat{\boldsymbol{\sigma}}) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \{\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)}\} \end{pmatrix}. \tag{26}$$

The normal curvature that reflects the influences of the perturbation on the estimates of Σ is given by

$$C^{\boldsymbol{\sigma}}_{\mathbf{l}} = -2\mathbf{l}^{\mathrm{T}}(\Delta^{\mathrm{T}}(\ddot{\mathbf{L}}^{\boldsymbol{\sigma}})^{-1}\Delta)\mathbf{l}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}, = -2\mathbf{l}^{\mathrm{T}}\ddot{\mathbf{F}}^{\boldsymbol{\sigma}}\mathbf{l}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}, \tag{27}$$

where

$$\ddot{\mathbf{F}}^{\boldsymbol{\sigma}}_{ij} = \sum_{\substack{1\le\beta\le\alpha\le p \\ 1\le\rho\le\gamma\le p}} \Delta_{(\alpha\beta)i}\ddot{\mathbf{L}}^{-1}_{(\alpha\beta)(\gamma\rho)}\Delta_{(\gamma\rho)j},$$

which can be computed using (18) to (22). Let $\mathbf{l}^{\boldsymbol{\sigma}}_{\max}$ be the eigenvector associated with the largest eigenvalue of $-2\ddot{\mathbf{F}}^{\boldsymbol{\sigma}}{}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}$ and $B^{\boldsymbol{\sigma}}_j$ be the jth diagonal element of $-\ddot{\mathbf{F}}^{\boldsymbol{\sigma}}/\sqrt{\mathrm{tr}(\ddot{\mathbf{F}}^{\boldsymbol{\sigma}})^2}{}_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}$; observations that influence the estimates of Σ can be detected by $\mathbf{l}^{\boldsymbol{\sigma}}_{\max}$ or $B^{\boldsymbol{\sigma}}_j, j = 1, \ldots, n$.

3. Examples

Example 1: Brain and body weight data set

As an illustration, we first consider the brain and body weight data set (in logarithms to base 10) which is available in Rousseeuw and Leroy (1987, p. 58). The data set consists of observations for 28 species on two variables: body weight and brain weight. From a scatter plot of the data (Fig. 1) it can be seen that the two variables exhibit a positively correlated pattern. Many analyses of this data set have been carried out in the context of outlier identification. For example, Rousseeuw and van Zomeren (1990) used a robust version of the Mahalanobis distance to conclude that cases 25, 6, 16, 14 and 17, in this order, are outlying observations and that the effect of case 17 is marginal. Atkinson and Mulira (1993), on the other hand, reached a similar conclusion using the Mahalanobis distance in a forward identification technique with results summarized visually in a stalactite plot. Poon *et al.* (2000) used the local influence approach to identify location outliers under various metrics. Different cases were identified as outliers under different metrics, but cases 25, 6 and 16 were always flagged as outliers when the metric used had taken into account the shape of the data set.

We reanalysed the data set using the proposed procedure and plotted in Fig. 2 the values of B_j, $B^{\boldsymbol{\mu}}_j$ and $B^{\boldsymbol{\sigma}}_j$, $j = 1, \ldots, n$, computed from $\ddot{\mathbf{F}}$, $\ddot{\mathbf{F}}^{\boldsymbol{\mu}}$ and $\ddot{\mathbf{F}}^{\boldsymbol{\sigma}}$ respectively. The results show that cases 25, 6 and 16, in this order, are most influential on the estimates of $\boldsymbol{\mu}$ and Σ. From Figs. 2b and 2c, we conclude that case 20 affects the estimate of $\boldsymbol{\mu}$ substantially but its effect on the estimate of Σ is less pronounced. Note from Fig. 1 that as case 20 lies at the lower left-hand corner with smallest values in both variables, it therefore affects the location of the data set substantially but its effect on the dispersions or correlation of the variables is not conspicuous. A similar phenomenon can also be

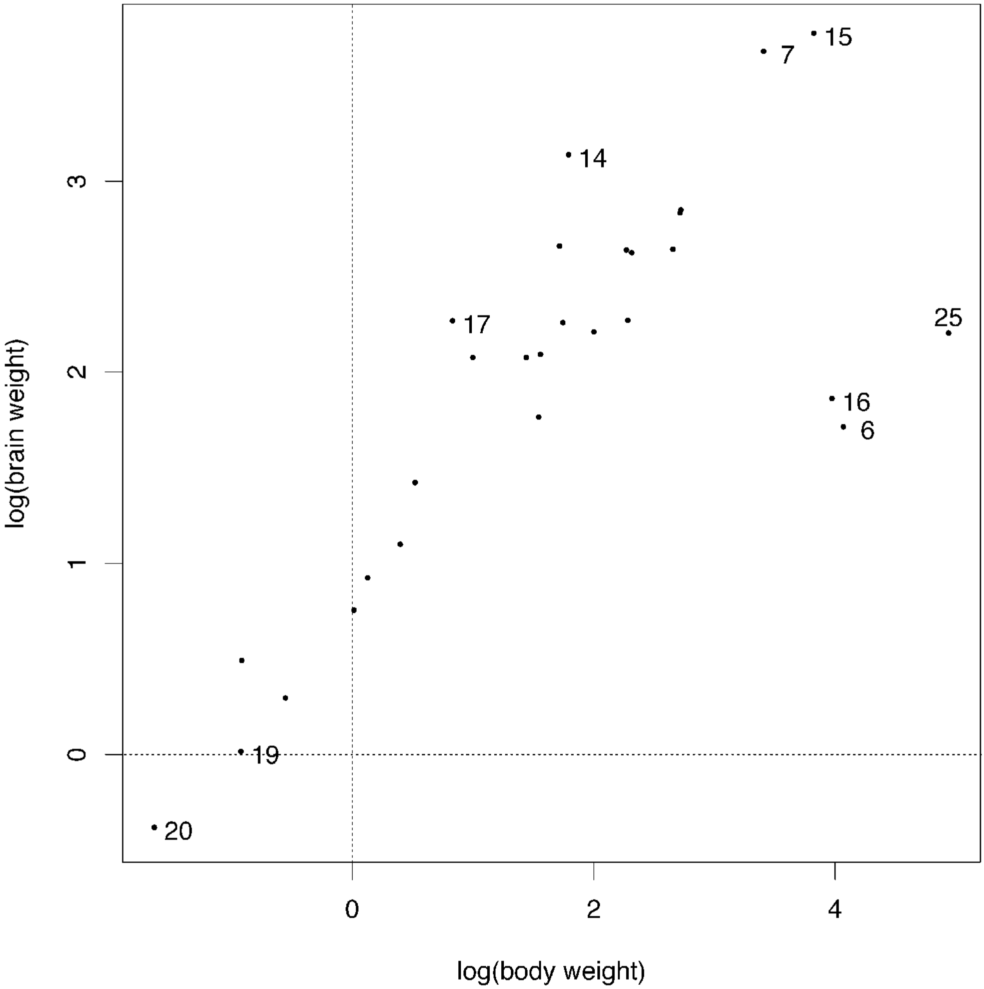

Figure 1. Scatter plot of the brain and body weight data set.

observed for cases 15, 19 and 7. On the other hand, the three extreme cases, 25, 6 and 16, are located outside the ellipse formed by the majority of the data; while they are influential on the estimates of both $\boldsymbol{\mu}$ and Σ, their effects on Σ are more pronounced than those on $\boldsymbol{\mu}$. These results, therefore, indicate that the proposed measures have identified what they are supposed to identify.

Example 2: Head data set

The second data set, originally from Frets (1921), is also contained in Seber (1984, p. 263). It consists of measurements of head lengths and breadths of the first and second adult sons in 25 families. Pairwise scatter plots among the four variables are given in Fig. 3, and several cases are marked for further discussion. The computed values of B_j, $B_j^{\boldsymbol{\mu}}$ and B_j^{σ}, $j = 1, \ldots, n$, from $\ddot{\mathbf{F}}$, $\ddot{\mathbf{F}}^{\boldsymbol{\mu}}$ and $\ddot{\mathbf{F}}^{\sigma}$ are presented in the form of index plots in Fig. 4. Three cases, 2, 17 and 6, are classified as influential on the estimates of both $\boldsymbol{\mu}$ and Σ. From Fig. 3, we see that these three observations are usually located at the boundaries of

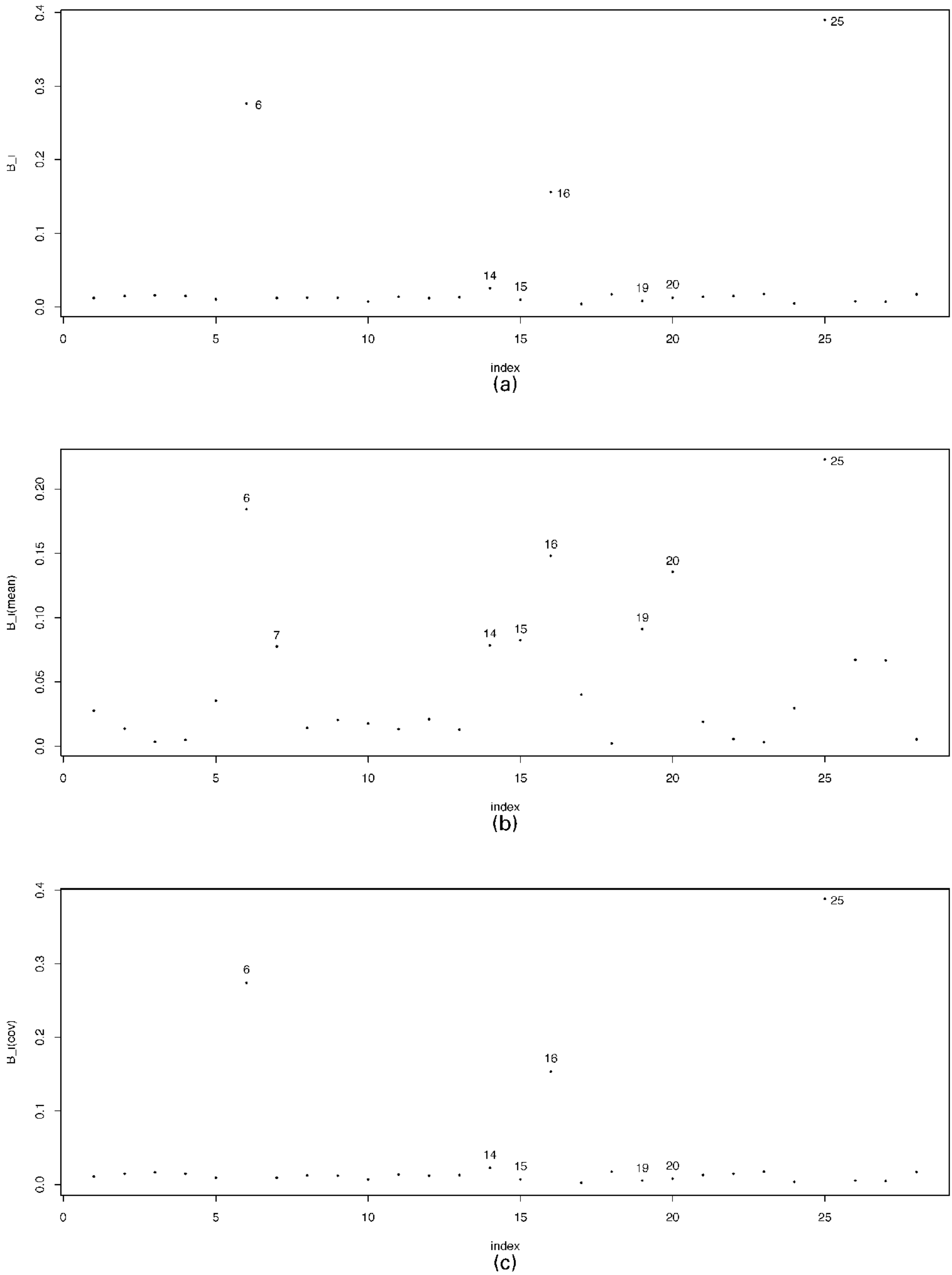

Figure 2. Index plots of the influential measures for the brain and body weight data set: (a) B_j; (b) $B_j^{\boldsymbol{\mu}}$; (c) $B_j^{\boldsymbol{\sigma}}$.

the data point clouds formed by different pairs of variables, hence they are influential on both the estimates of $\boldsymbol{\mu}$ and Σ. On the other hand, Figs. 4b and 4c show that $B_{16}^{\boldsymbol{\mu}}$ is relatively large but $B_{16}^{\boldsymbol{\sigma}}$ is not, which leads to the conclusion that the influence of case 16 on the estimate of $\boldsymbol{\mu}$ is substantial but on Σ is not noticeable. This finding makes sense

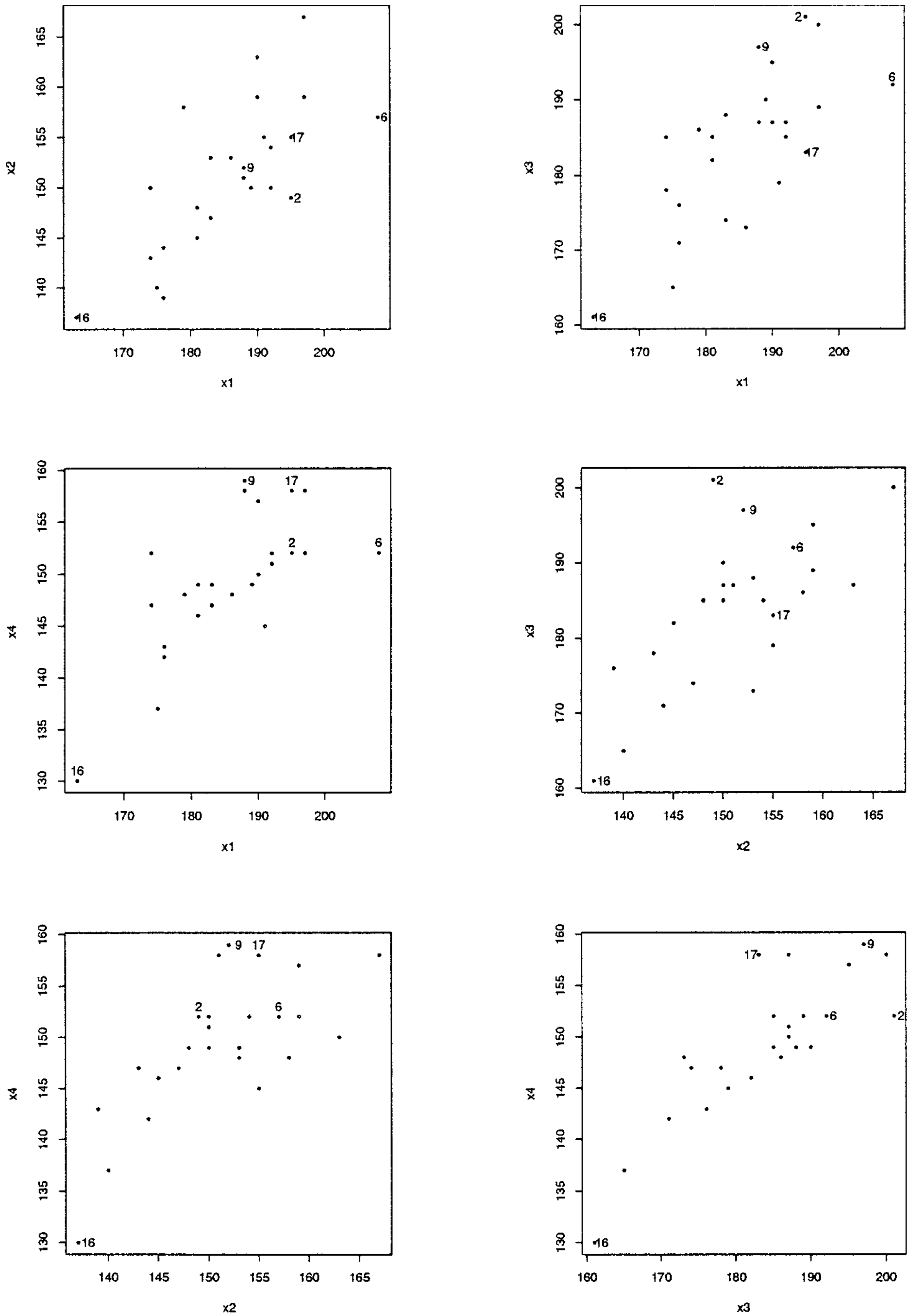

Figure 3. Scatter plots for the head data set.

because it can be seen from Fig. 3 that case 16, which has the smallest observed values on all four variables, always lies in the lower left-hand corner of the scatter plots; its influence on the location of the data set is therefore considerable.

Following the suggestion of a reviewer, we reduced the dimensionality to $p = 2$ by

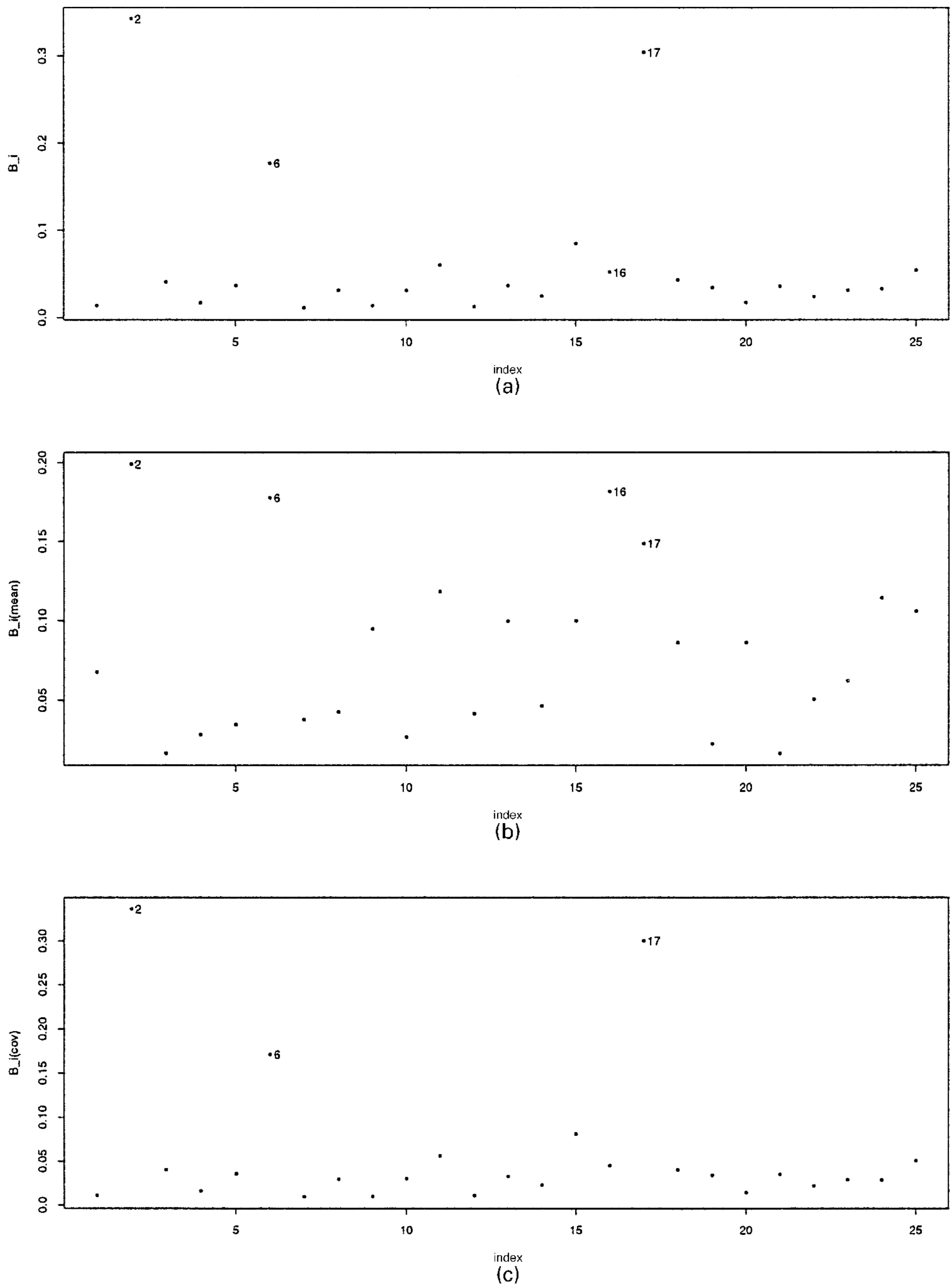

Figure 4. Index plots of the influential measures for the head data set: (a) B_j; (b) B_j^{μ}; (c) B_j^{σ}

defining the measure of the 'first son' as the average of the first two variables and that of the 'second son' as the average of the last two variables. The scatter plot of the reduced data is presented in Fig. 5 and the index plots of the diagnostic measures are given in Fig. 6. Cases 9, 6, 1, 16, in this order, are identified as the observations that influence the estimates of $\boldsymbol{\mu}$ and Σ simultaneously. Case 20, which is the second extreme observation

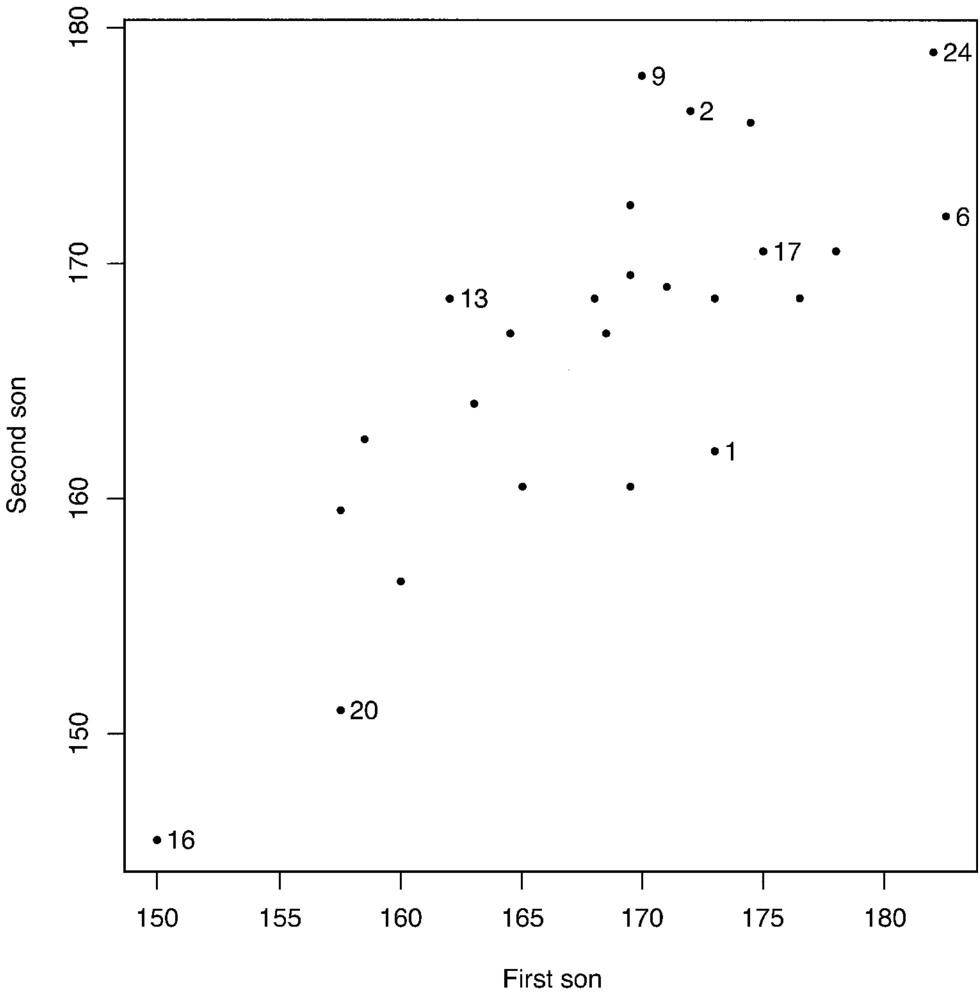

Figure 5. Scatter plot for the head data set with $p = 2$.

in the lower left-hand corner of Fig. 5, is an influential observation in estimating $\boldsymbol{\mu}$ but its effect on the estimate of Σ is less pronounced. Case 9, which is also marked in the plots of Fig. 3 and is not a noticeable influential point when all four variables are used in the analysis, is now located at the boundary of the data point cloud as depicted in Fig. 5 and becomes a very influential point, especially with respect to the estimate of the covariance matrix. On the other hand, cases 2 and 17, which are extreme in the plots in Fig. 4, are no longer influential.

Example 3: Open–closed book data set

The open–closed book data set, taken from Mardia, Kent, and Bibby (1979, p. 3), consists of five measurements on 88 students. Influence analyses in the context of factor analysis have been performed by several authors (Tanaka, Watadani, & Moon, 1991; Lee & Wang, 1996). Case 81 has been identified by all analyses as an extreme influential point, and cases 3, 28 and 56 have been identified by Tanaka *et al.* (1991) and case 87 by

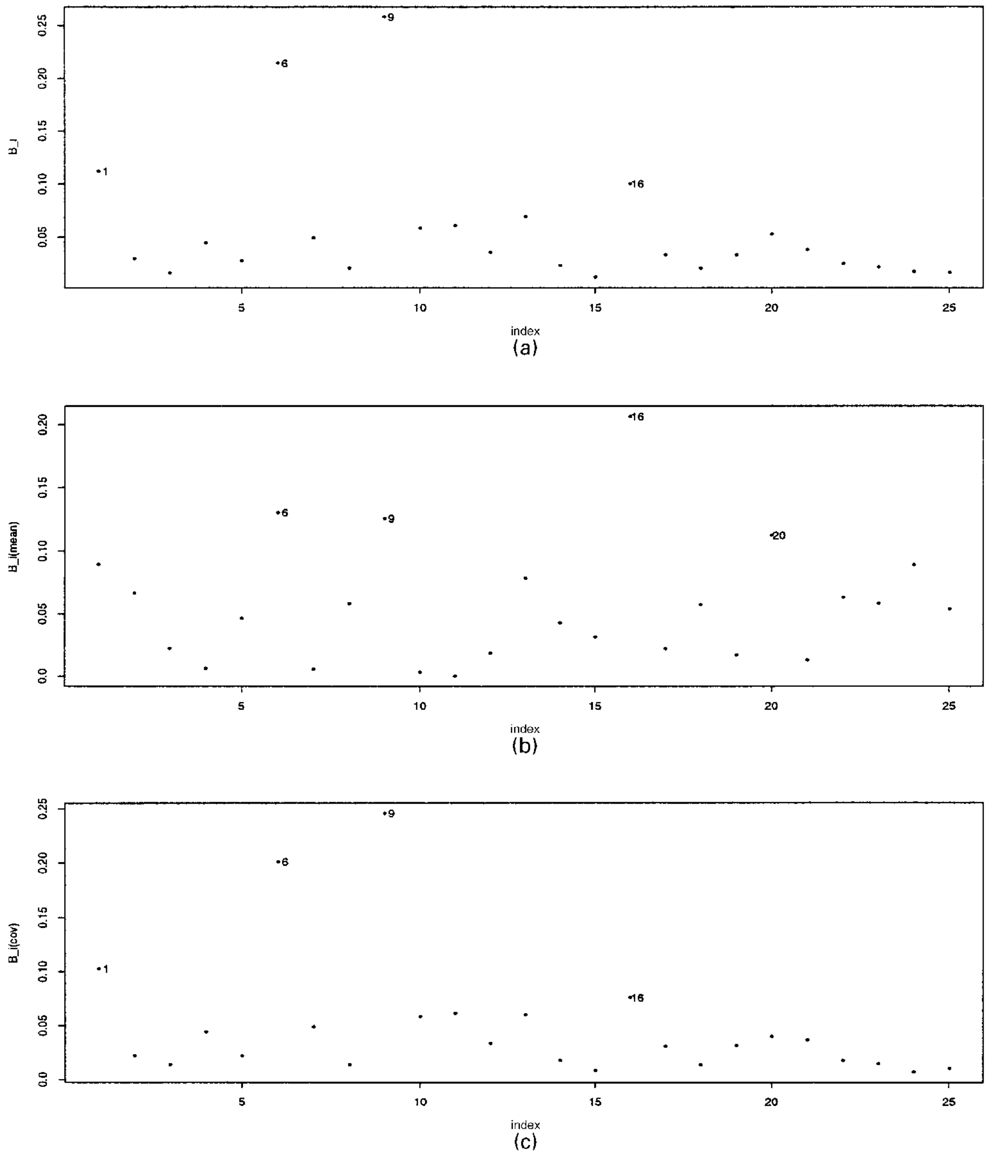

Figure 6. Index plots of the influential measures for the head data set with $p = 2$: (a) B_j; (b) B_j^{μ}; (c) B_j^{σ}.

Lee and Wang (1996) as other observations worth special attention. Poon *et al.* (2000) used different metrics to identify location outliers in the data set, and cases 28, 54, 56, 61, 81, 87 and 88 were flagged. Our proposed diagnostic statistics are presented as index plots in Fig. 7. Cases 81 and 87 are seen to be the most influential observations in estimating both $\boldsymbol{\mu}$ and Σ, and cases 54, 28 and 56 are other influential observations. These cases are also influential in estimating $\boldsymbol{\mu}$ or Σ, but case 88, which has substantial effect on estimating $\boldsymbol{\mu}$, does not have a similar effect in estimating Σ. Examining the original data shows that case 88 possesses the smallest scores in variables 1 and 4 and the second smallest scores in variables 3 and 5, and therefore affects the estimate of the mean vector substantially.

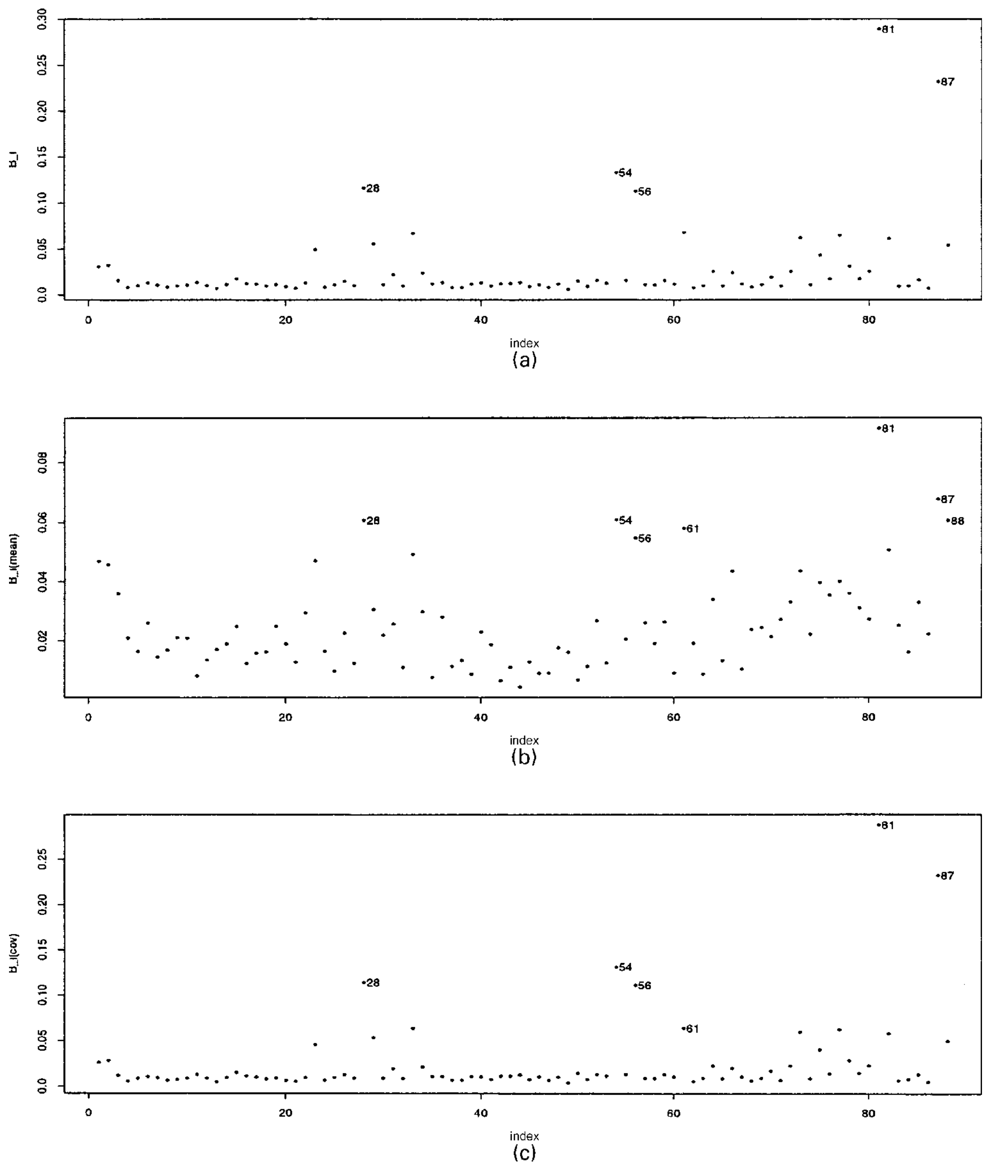

Figure 7. Index plots of the influential measures for the open–closed book data set: (a) B_j; (b) B_j^{μ}; (c) B_j^{σ}.

4. Discussion

Although our development relies on the multivariate normal distribution and the likelihood function in (1), the procedure can in fact be generalized to other multivariate distributions. The normal distribution is chosen in the current study because of its popularity and because many multivariate techniques make use of the sample mean and/or the sample covariance matrix which are in effect the maximum likelihood estimates of the normal model parameters. When a data set does not follow the normal distribution but its location or dispersion is estimated using (2), the proposed diagnostic measures can still be applied to identify those observations that exert undue influence on the estimates.

In the examples, we used B_j, $B_j^{\boldsymbol{\mu}}$ and $B_j^{\boldsymbol{\sigma}}$ as diagnostic measures to identify observations which are influential on the estimates of $\boldsymbol{\mu}$ and Σ, the estimate of $\boldsymbol{\mu}$, and the estimate of Σ, respectively. The results of our illustrative examples indicate that the proposed measures work well and can successfully identify the observations that they are supposed to identify. We chose to use B_j, $B_j^{\boldsymbol{\mu}}$ and $B_j^{\boldsymbol{\sigma}}$ rather than $\mathbf{l}_{\max}$, $\mathbf{l}_{\max}^{\boldsymbol{\mu}}$ and $\mathbf{l}_{\max}^{\boldsymbol{\sigma}}$ as diagnostic measures because it is easier to compute the B_j. Moreover, letting $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ be the $p \times n$ matrices with jth column given by $\mathbf{y}_j = \mathbf{x}_j - \boldsymbol{\mu}$ and $\hat{\mathbf{y}}_j = \mathbf{x}_j - \hat{\boldsymbol{\mu}} = \mathbf{x}_j - \bar{\mathbf{x}}$ respectively, we have (see (25)) $\ddot{\mathbf{F}}^{\boldsymbol{\mu}} = -\frac{1}{n}\mathbf{Y}^{\mathrm{T}}\Sigma^{-1}\mathbf{Y}$. When the matrix is evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, it becomes (see (2)) $-(1/n)\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{S}^{-1}\hat{\mathbf{Y}} = -\hat{\mathbf{Y}}^{\mathrm{T}}(\hat{\mathbf{Y}}\hat{\mathbf{Y}}^{\mathrm{T}})^{-1}\hat{\mathbf{Y}}$. Therefore, the matrix is a constant multiple of a projection matrix. It has a single non-zero eigenvalue and geometric multiplicity equal to p. Choosing eigenvectors with maximum eigenvalues amounts to finding an (orthonormal) basis for the eigenspace for the curvature matrix. We then have to analyse contributions to the basis. Therefore, we choose to use the measure $B_j^{\boldsymbol{\mu}}$, $j = 1, \ldots, n$, which aggregates the coefficients in the eigenvectors corresponding to non-zero eigenvalues to detect influential observations.

In determining when a measure is large enough to be worthy of further notice, we simply employ the natural gap approach and use an index plot to help detect large values. In most cases, such a simple method can efficiently reveal observations that require special attention. When we consider the identification of influential observations to be an exploratory rather than confirmatory goal, and ensure that any observations so identified are followed by thorough analysis of the facts underlying the observations, strict adherence to a critical value for identification does not seem necessary. If objectivity is desired, for example in automating the implementation of the proposed measures, the reference constant proposed by Poon and Poon (1999), which is rationalized by the geometric concept of mean curvature, can be used to establish a benchmark for judging the largeness of a measure.

When the matrix $-\ddot{\mathbf{F}}^{\boldsymbol{\mu}}$ is evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, it becomes $(1/n)\hat{\mathbf{Y}}^{\mathrm{T}}\mathbf{S}^{-1}\hat{\mathbf{Y}}$, and its jth diagonal element is equivalent to the Mahalanobis distance

$$MD_j(\hat{\boldsymbol{\mu}}, \hat{\Sigma}) = MD_j(\bar{\mathbf{x}}, \mathbf{S}) = \sqrt{(\mathbf{x}_j - \bar{\mathbf{x}})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})} \tag{28}$$

for case j. Since the Mahalanobis distance measures the distance between an individual observation and the location of the data set, its value can be used to flag outlying observations. However, when there are many outliers, a simple one-step outlier identification procedure based on MD_j is not satisfactory and various procedures have been developed to improve the use of the Mahalanobis distance for identification purposes. There are two main directions for improvement: the first is the use of robust distance which is obtained by replacing $\bar{\mathbf{x}}$ and $\mathbf{S}$ in (28) by other robust estimates (Rousseeuw & van Zomeren, 1990); and the second is to employ stepping techniques. Stepping techniques usually start with a subset of the data set, which is updated step by step in order to exclude outliers; the observations in the outlier-free subset are then used to construct distance measures for outlier identification (Atkinson & Mulira, 1993; Atkinson, 1994; Hadi, 1994). In this paper, we have proposed three measures to identify observations that affect respectively the estimates of $\boldsymbol{\mu}$ and Σ, the estimate of $\boldsymbol{\mu}$ and the estimate of Σ; the possible improvement of these measures along the lines mentioned above is an interesting topic for further study.

Acknowledgement

The work described in this paper was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (RGC ref. CUHK4186/98P and RGC Direct Allocation ID2060178). Y. S. Poon thanks the Department of Statistics at the Chinese University of Hong Kong for hospitality. Both authors thank Mr. Mannon Wong for his assistance in preparing the figures and an anonymous referee together with the Editor for valuable comments.

References

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. New York: Wiley.

Atkinson, A. C. (1994). Fast very robust methods for the detection of multiple outliers. *Journal of the American Statistical Association*, *89*, 1329–1339.

Atkinson, A. C., & Mulira, H. M. (1993). The stalactite plot for the detection of multivariate outliers. *Statistics and Computing*, *3*, 27–35.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics*. New York: Wiley.

Bruce, A. G., & Martin, R. D. (1989). Leave-*k*-out diagnostics for time series (with discussion). *Journal of the Royal Statistical Society Series B*, *51*, 363–424.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*, 15–18.

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society Series B*, *48*, 133–169.

Frets, G. P. (1921). Heredity of head form in man. *Genetica*, *3*, 193–384.

Fung, W. K. (1993). Unmasking outliers and leverage points: A confirmation. *Journal of the American Statistical Association*, *88*, 515–519.

Hadi, A. S. (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society Series B*, *54*, 761–771.

Hadi, A. S. (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society Series B*, *56*, 393–396.

Lawrance, A. J. (1995). Deletion influence and masking in regression. *Journal of the Royal Statistical Society Series B*, *57*, 181–189.

Lee, S. Y., & Wang, S. J. (1996). Sensitivity analysis of structural equation models. *Psychometrika*, *61*, 93–108.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London: Chapman & Hall.

Poon, W.-Y., Lew, S. F., & Poon, Y. S. (2000). A local influence approach to identify multiple multivariate outliers. *British Journal of Mathematical and Statistical Psychology*, *53*, 255–273.

Poon, W.-Y., & Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society Series B*, *61*, 51–61.

Poon, W.-Y., & Poon, Y. S. (2001). Conditional local influence in case-weights linear regression. *British Journal of Mathematical and Statistical Psychology*, *54*, 177–191.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, *9*, 705–724.

Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*, *91*, 1047–1061.

Rousseeuw, P. J., & Leroy, A. (1987). *Robust regression and outlier detection*. New York: Wiley.

Rousseeuw, P. J., & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, *85*, 633–639.

Seber, G. A. F. (1984). *Multivariate observations*. New York: Wiley.

Tanaka, Y., Watadani, S., & Moon, S. H. (1991). Influence in covariance structure analysis: With an application to confirmatory factor analysis. *Communications in Statistics—Theory & Methods*, *20*, 3805–3821.

Received 22 November 2000; revised version received 21 May 2001