British Journal of Mathematical and Statistical Psychology (2001), **54**, 177–191 *Printed in Great Britain*

Conditional local influence in case-weights linear regression

Wai-Yin Poon*

The Chinese University of Hong Kong, Hong Kong

Yat Sun Poon

University of California at Riverside, USA

The local influence approach proposed by Cook (1986) makes use of the normal curvature and the direction achieving the maximum curvature to assess the local influence of minor perturbation of statistical models. When the approach is applied to the linear regression model, the result provides information concerning the data structure different from that contributed by Cook's distance. One of the main advantages of the local influence approach is its ability to handle the simultaneous effect of several cases, namely, the ability to address the problem of 'masking'. However, Lawrance (1995) points out that there are two notions of 'masking' effects, the joint influence and the conditional influence, which are distinct in nature. The normal curvature and the direction of maximum curvature are capable of addressing effects under the category of joint influences but not conditional influences. We construct a new measure to define and detect conditional local influences and use the linear regression model for illustration. Several reported data sets are used to demonstrate that new information can be revealed by this proposed measure.

1. Introduction

The local influence perturbation approach plays an important role in diagnostic and influence analyses. In effect, diagnostics are deduced from local changes of relevant measures caused by small perturbations (see, for example, Pregibon, 1981). Cook (1986) developed a unified approach, based on likelihood displacement, to assess local influence and, in particular, discussed the application to case-weights perturbation in linear regression (Cook, 1986, equations (27)–(32)). The basic diagnostic measures are the maximum normal curvature and the direction giving the maximum normal curvature. Cook's work has stimulated various applications. Examples range from the diagnostics and influence analyses in nonlinear regression (St Laurent & Cook, 1993), in linear mixed models (Lesaffre & Verbeke, 1998), in factor analysis (Kwan & Fung, 1998) and in multivariate analysis (Poon, Lew & Poon, 2000), to structural equation models (Cadigan, 1995; Lee & Wang, 1996; Poon, Wang & Lee, 1999).

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

One major advantage of the local influence approach lies in its ability to handle cases simultaneously (Lawrance, 1991). When a group of cases is involved, the problem of 'masking' emerges. Although 'masking' has been well known as being broadly concerned with the limitations imposed by the use of individual observations, there is in the literature a degree of ambiguity. A very useful discussion of the ambiguities and a way to remove them was given by Lawrance (1995) based on Cook's distance in the linear regression model. He pointed out that two notions of 'masking' had emerged in the literature: the joint influence and the conditional influence, which are distinct in nature. While the former is concerned with the simultaneous effect of several cases, the latter is concerned with the difference before and after the deletion of one or more cases. Lawrance clearly defined different kinds of possible effects arising from these two situations by describing the joint influence effects as reducing, enhancing and swamping and the conditional influence effects as masking and boosting. Various measures were developed to assess these effects. Such a differentiation of effects was found to be very useful in understanding the structure of a data set.

The local influence approach can be employed to study the simultaneous effects of a group of cases, that is, it is inherently good at addressing joint local influence, using a concept similar to Lawrance's. However, since a common, although not necessarily appropriate, method of accommodating an influential observation is to remove it from subsequent analyses, it is advantageous to know in advance the effect of this on diagnostic measures. With this in mind, it is interesting to examine the effect of the removal of a specific case on a diagnostic measure—this is the notion of conditional influence as defined by Lawrance. When the diagnostic measure is constructed using the local influence approach, the concept of conditional local influence is induced.

Within the framework of a case-weights linear regression model, we develop a measure for studying conditional local influence. We begin with a summary of the local influence approach and its application in case-weights linear regression in Section 2, and then we develop a measure for studying conditional local influence in Section 3. The details are illustrated using several reported data sets and are presented in Section 4.

2. Local influence in case-weights linear regression

Consider the linear regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{X}$ is an $n \times p$ matrix, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ normal random vector with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$. Let $\boldsymbol{x}_i$ be a column $p \times 1$ vector storing the ith row of $\boldsymbol{X}$; then the log-likelihood function is given by

$$L(\boldsymbol{\beta}) = -\frac{1}{2\sigma^2} \sum_i (y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta})^2. \tag{2}$$

To examine the effect of individual observations on the estimates of the unknown parameters, Cook (1986) proposes a case-weights perturbation to the likelihood, and the perturbed likelihood becomes

$$L(\boldsymbol{\beta}|\boldsymbol{\omega}) = -\frac{1}{2\sigma^2} \sum_i \omega_i (y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta})^2, \tag{3}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\mathrm{T}}$ is an $n \times 1$ vector in Ω of $\mathbb{R}^n$ storing the case weights and Ω represents the set of relevant perturbations. For example, Ω may be the set defined by $0 \le \omega_i \le 1$ for $i = 1, \ldots, n$. When $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (1, \ldots 1)^{\mathrm{T}}$, $L(\boldsymbol{\beta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\beta})$ for all $\boldsymbol{\beta}$. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_\omega$ be the maximum likelihood estimators of $\boldsymbol{\beta}$ under $L(\boldsymbol{\beta}|\boldsymbol{\omega}_0)$ and $L(\boldsymbol{\beta}|\boldsymbol{\omega})$ respectively; then their difference can be measured by the likelihood displacement function (Cook, 1986) given by

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

This likelihood displacement, together with the case-weights perturbation scheme, can be considered as a natural generalization of Cook's distance (see Cook, 1986; Cook, Pena & Weisberg, 1988; and Lawrance, 1991). At the point $\boldsymbol{\omega}_0, f$ achieves its minimum. A straight line in Ω passing through $\boldsymbol{\omega}_0$ is given by $\boldsymbol{\omega}(\mathbf{l}) = \boldsymbol{\omega}_0 + a\mathbf{l}$, where a is a scalar and $\mathbf{l}$ is a fixed column vector in $\mathbb{R}^n$ with $\mathbf{l}^{\mathrm{T}}\mathbf{l} = 1$. For example, if $a = -1$ and $\mathbf{l}$ is a vector with one in the ith position and zeros elsewhere, then the perturbation becomes $\boldsymbol{\omega} = (1, \ldots, 1, 0, 1, \ldots, 1)^{\mathrm{T}}$ with 0 in the ith position and (3) becomes the likelihood when case i is deleted. If the ith case is an influential case, then the resulting $\hat{\boldsymbol{\beta}}_\omega$ will be substantially different from $\hat{\boldsymbol{\beta}}$, thereby making the value of $f(\boldsymbol{\omega})$ increase considerably from that of $f(\boldsymbol{\omega}_0)$. In view of this, influential cases can be revealed by studying the characteristics of $f(\boldsymbol{\omega})$.

Cook (1986) suggested using the normal curvature $C_{\mathbf{l}}$ of the graph of the likelihood displacement function along a direction $\mathbf{l}$ at the optimal point $\boldsymbol{\omega}_0$ to examine the local behaviour of $f(\boldsymbol{\omega})$ for assessing the local influence of the perturbation in the direction $\mathbf{l}$. It is recalled that the normal curvature is determined by the first fundamental form I and the second fundamental form $\mathcal{S}$ of the graph of the function f near the point $\boldsymbol{\omega}_0$ (see Poon & Poon, 1999). These are symmetric matrices given by

$$I_{ij} = \delta_{ij} + \frac{\partial f}{\partial \omega_i}\frac{\partial f}{\partial \omega_j} \text{ and } \mathcal{S}_{ij} = \frac{1}{(1 + |\nabla f|^2)^{1/2}}\frac{\partial^2 f}{\partial \omega_i \partial \omega_j},$$

where δ_{ij} is 1 when $i = j$, and zero otherwise, and $|\nabla f|$ represents the norm of the gradient vector of f. These matrices are evaluated on vectors $\mathbf{v}$ and $\mathbf{w}$ by $I(\mathbf{v}, \mathbf{w}) = \mathbf{v}^{\mathrm{T}} I\mathbf{w}$, and $\mathcal{S}(\mathbf{v}, \mathbf{w}) = \mathbf{v}^{\mathrm{T}}\mathcal{S}\mathbf{w}$, respectively. In particular, the normal curvature of the graph in a direction $\mathbf{l}$ at the point $\boldsymbol{\omega}_0$ is

$$C_{\mathbf{l}} = C(\mathbf{l}, \mathbf{l}) = \mathbf{l}^{\mathrm{T}}\mathbf{C}\mathbf{l} = \frac{\mathcal{S}(\mathbf{l}, \mathbf{l})}{I(\mathbf{l}, \mathbf{l})}, \tag{5}$$

where all functions and their derivatives are evaluated at $\boldsymbol{\omega}_0$. We call the matrix $\mathbf{C}$ the normal curvature matrix.

Let $\ddot{\mathbf{L}}$ and $\boldsymbol{\Delta}$ be $p \times p$ and $p \times n$ matrices with elements

$$\ddot{L}_{ij} = \frac{\partial^2 L(\boldsymbol{\beta}|\boldsymbol{\omega}_0)}{\partial \beta_i \partial \beta_j}_{|\beta = \hat{\beta}} \text{ and } \Delta_{ij} = \frac{\partial^2 L(\boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \beta_i \partial \omega_j}_{|\beta = \hat{\beta},\, \omega = \omega_0}, \tag{6}$$

respectively, Cook (1986) deduced an expression for computing the normal curvature matrix which is given by

$$\mathbf{C} = -2(\boldsymbol{\Delta}^{\mathrm{T}}\ddot{\mathbf{L}}^{-1}\boldsymbol{\Delta})_{|\beta = \hat{\beta},\, \omega = \omega_0}. \tag{7}$$

Let $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$ be the hat matrix with entries h_{ij} and diagonal elements $h_j = h_{jj}$, r_j the residual of the jth case, and diag(r) the diagonal matrix whose jth diagonal element is r_j. Assuming that σ^2 is known, it can be shown from (2) and (3) that

$$\ddot{\mathbf{L}} = -\mathbf{X}^{\mathrm{T}}\mathbf{X}/\sigma^2 \text{ and } \boldsymbol{\Delta} = \mathbf{X}^{\mathrm{T}}\mathrm{diag}(r)/\sigma^2. \tag{8}$$

As a result, the normal curvature matrix $\mathbf{C}$ is given by (Cook, 1986, eq. (29))

$$\mathbf{C} = \frac{2}{\sigma^2}\mathrm{diag(r)}\mathbf{H}\mathrm{diag}(r). \tag{9}$$

For a fixed direction $\mathbf{l}_0$, $C_{\mathbf{l}0} = \mathbf{l}_0^{\mathrm{T}}\mathbf{C}\mathbf{l}_0$. A large value of $C_{\mathbf{l}0}$ indicates strong local influence along the direction $\mathbf{l}_0$, that is to say, the displacement function $f(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ will change substantially along the direction $\mathbf{l}_0$. For example, if $\mathbf{l}_0$ is a vector with one in the ith position and zero elsewhere, large value of $C_{\mathbf{l}0}$ indicates that $f(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ will change substantially along the direction which changes only the weight of case i. In other words, changing the weight of case i will have a considerable impact on $\hat{\boldsymbol{\beta}}$ and hence case i is an influential observation. Similarly, a general direction $\mathbf{l}$ takes account of all cases simultaneously; when $C_{\mathbf{l}}$ is large, changing the weights of the cases that possess large magnitudes in $\mathbf{l}$ will bring about a noteworthy change in $\hat{\boldsymbol{\beta}}$, and such cases exert an influence on the parameter estimates.

Cook (1986) proposed paying special attention to the directions associated with large $C_{\mathbf{l}}$ values, and in particular the direction $\mathbf{l}_{\max}$ corresponding to the maximum curvature $C_{\max} = \max_{\mathbf{l}} C_{\mathbf{l}}$. In effect, $\mathbf{l}_{\max}$ becomes a very popular diagnostic quantity in the influence and diagnostic analysis literature. It can be shown that $C_{\max}$ and $\mathbf{l}_{\max}$ are the largest eigenvalue and the associated eigenvector of the matrix $\boldsymbol{C}$. However, since $\mathbf{l}_{\max}$ can rarely be obtained explicitly, any further development that requires the analytical form of the elements in $\mathbf{l}_{\max}$ becomes difficult. For example, suppose it is decided to accommodate an unusual observation by removing it from subsequent analysis; then $\mathbf{l}_{\max(i)}$ will be obtained by the remaining $n-1$ observations of the data set. The influence of case i on the diagnostic measure $\mathbf{l}_{\max}$ can be examined by comparing $\mathbf{l}_{\max}$ to $\mathbf{l}_{\max(i)}$, and such comparison induces the concept of conditional influence. When $\mathbf{l}_{\max}$ cannot be obtained explicitly, n replications must be implemented, each with one observation removed in turn, so as to give a thorough picture of all the conditional effects. Such a procedure becomes infeasible with large n. As a result, conditional local influence remains unexplored, in spite of extensive contributions in the literature in the area of local influence.

The recent work of Poon & Poon (1999) lessens the dependence of the local influence perturbation approach on $\mathbf{l}_{\max}$ or the directions corresponding to large normal curvatures. Let $\boldsymbol{e}_j, j = 1, \ldots, n$, be basic perturbation vectors of the perturbation space, that is, the vectors of the standard basis; by (5), $C_j = C_{\boldsymbol{e}_j} = C(\boldsymbol{e}_j, \boldsymbol{e}_j)$ is given by the jth diagonal element of the normal curvature matrix $\mathbf{C}$. From (9), we have

$$C_j = \frac{2}{\sigma^2} r_j^2 h_j. \tag{10}$$

Although C_j is not very different from Cook's distance in the quantities it involves, Cook (1986, near eq. (32)) demonstrated that C_j carries different information than Cook's distance. Moreover, based on the conformal normal curvature, Poon & Poon (1999, Theorem 4) developed a relation between the C_j and the influential eigenvector directions at $\boldsymbol{\omega}_0$, and established as a special case the relation between $\mathbf{l}_{\max}$ and the C_j. Specifically, if $C_{\max}$ is sufficiently large, then the cases with large magnitudes in $\mathbf{l}_{\max}$ will possess large C_j values and vice versa. In effect, the group of cases with large C_j values is a group of influential cases.

A very nice feature of C_j is that it can be obtained explicitly as the jth diagonal element of the matrix $\mathbf{C}$ (see (10)). When the analytical expression of this diagnostic quantity is available, further studies based on C_j become possible. In particular, we can study conditional local influence by assessing the changes in the C_j when a case is removed from analysis.

3. Conditional local influence

Since the C_j are effective measures to characterize local influence, conditional local influence can be examined by assessing the changes in the C_j when a case is removed from analysis. We follow the common practice of using the subscript (i) to indicate that case i is deleted. From (9), we see that

$$\mathbf{C}_{(i)} = \frac{2}{\sigma^2}\mathrm{diag}(r_{(i)})\mathbf{H}_{(i)}\mathrm{diag}(r_{(i)}), \tag{11}$$

where $\mathbf{H}_{(i)} = \mathbf{X}_{(i)}(\mathbf{X}_{(i)}^{\mathrm{T}}\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}^{\mathrm{T}}$, and $r_{(i)} = (\mathbf{I}_{(i)} - \mathbf{H}_{(i)})y_{(i)}$. By Atkinson (1985, Section 2.2),

$$(\mathbf{X}_{(i)}^{\mathrm{T}}\mathbf{X}_{(i)})^{-1} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i(1 - \mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i)^{-1}\mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{T}\mathbf{X})^{-1}.$$

Therefore, the (j, k)th entry, for $j, k \neq i$, of the new hat matrix and the jth residual are

$$H_{(i)jk} = h_{jk} + \frac{h_{ji}h_{ki}}{1 - h_i} \text{ and } r_{(i)j} = r_j + \frac{h_{ji}r_i}{1 - h_i},$$

respectively. It follows from (11) that the (j, k)th entry of the new normal curvature matrix is

$$C_{(i)jk} = \frac{2}{\sigma^2}r_{(i)j}H_{(i)jk}r_{(i)k} = \frac{2}{\sigma^2}\left(r_j + \frac{h_{ji}r_i}{1 - h_i}\right)\left(h_{jk} + \frac{h_{ji}h_{ki}}{1 - h_i}\right)\left(r_k + \frac{h_{ki}r_i}{1 - h_i}\right).$$

Then the difference in the curvature of the jth case before and after the deletion of case i is

$$E_{ij} = C_{(i)j} - C_j = \frac{2}{\sigma^2}\left(\frac{2h_jr_jh_{ji}r_i + r_j^2h_{ji}^2}{1 - h_i} + \frac{2h_{ji}^2(r_jh_{ji}r_i) + h_jh_{ji}^2r_i^2}{(1 - h_i)^2} + \frac{h_{ji}^4r_i^2}{(1 - h_i)^3}\right). \tag{12}$$

Definition 1. The $n \times n$ matrix $\mathbf{E} = (E_{ij})$, where E_{ij} is given by (12) *when $i \neq j$ and $E_{ii} = 0$, is called the local masking matrix in case-weights linear regression.*

Let D_j and $D_{(i)j}$ be Cook's distance for case j before and after the deletion of case i respectively. Lawrance (1995, eq. (4.3)) considers the ratio $D_{(i)j}/D_j$ and defines the effect of case i on case j as masking if this ratio is greater than 1, and as boosting if it is less than 1. Equivalently, the effects may be described by the difference $D_{(i)j} - D_j$. Inspired by this observation, we propose the following.

Definition 2. The ith case masks the jth case locally if $E_{ij} > 0$. The ith case boosts the jth case locally if $E_{ij} < 0$.

We prefer to use the difference rather than ratio because that the ratio $C_{(i)j}/C_j$ can inflate dramatically when C_j is small, leading to many unstable elements. As a result, it is difficult to judge the significance of an effect from the magnitude of $C_{(i)j}/C_j$.

One nice feature of the masking matrix **E** lies in its ability to produce measures of all conditional effects in a single pass, and replications, each with one observation removed in turn, are not required. Elements in **E** are functions of the error variance, the residuals and the elements of the hat matrix. These quantities are readily available after a standard regression analysis, so very little additional effort is required to compute **E**.

The next question is when a particular element in **E** is large enough in magnitude to be worthy of further analysis. This problem of determining how large is large for a diagnostic measure has led to different paradigms in the regression literature. Although external scaling methods (Belsley, Kuh & Welsch, 1980, p. 27) which determine cut-off values with reference to statistical distribution theories have been suggested, other authors (such as Bruce & Martin, 1989) call for flexible approaches and point out that the testing framework may not be appropriate (Lawrance, 1989). The natural gap approach (Belsley, *et al.* 1980, p. 27; Lawrance, 1991) has also been considered as a useful alternative; in such an approach, an index plot can always reveal useful information.

4. Examples

Several examples are used to demonstrate that new information can be revealed by the matrix **E**.

4.1. Pena and Yohai's data sets

We first analysed the three data sets taken from Pena & Yohai (1995, Example 1). Each data set consists of 10 observations on one explanatory and one dependent variable. Scatterplots of these data sets are given in Fig. 1. The first eight cases are ordinary ones, but cases 9 and 10 are unusual. In Fig. 1(a), cases 9 and 10 are located at about the same place and produce

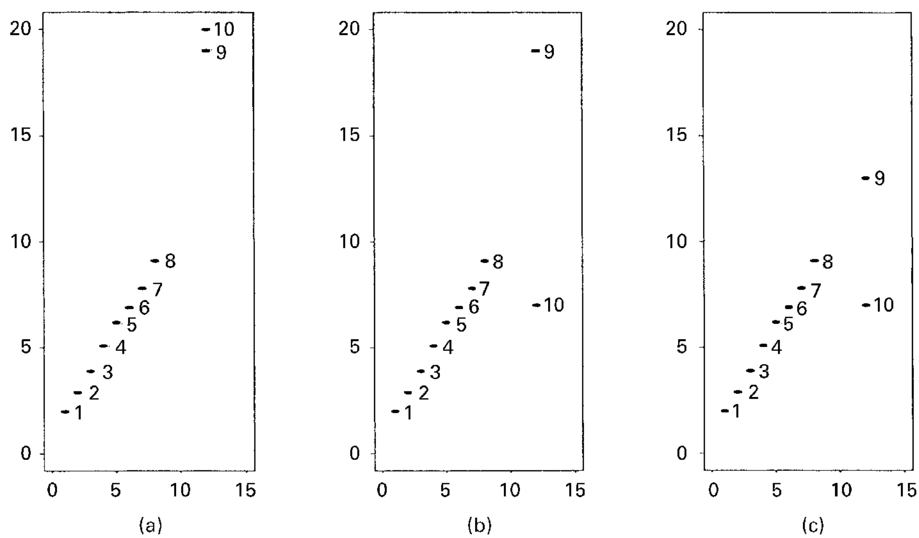

Figure 1. Scatterplots of Pena and Yohai's data sets.

Table 1. Pena and Yohai's data sets: $\mathbf{l}_{max}$ and C_j

j	1	2	3	4	5	6	7	8	9	10
					Data set (a)					
$\mathbf{l}_{max}$	−0.116	−0.032	0.002	−0.000	−0.038	−0.158	−0.314	−0.445	0.384	0.718
C_j	0.590	0.155	0.019	0.000	0.019	0.140	0.339	0.534	0.358	1.252
					Data set (b)					
$\mathbf{l}_{max}$	0.001	−0.002	−0.001	−0.000	−0.004	0.003	0.009	−0.006	−0.707	0.708
C_j	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	2.973	2.980
					Data set (c)					
$\mathbf{l}_{max}$	−0.056	−0.028	−0.006	−0.001	−0.018	−0.031	−0.060	−0.145	−0.502	0.847
C_j	0.112	0.051	0.011	0.003	0.022	0.018	0.035	0.143	1.307	3.720

similar effects; in Fig. 1(b), cases 9 and 10 produce opposite effects; and in (c), case 9 appears as an outlier due to the presence of case 10. The computed $\mathbf{l}_{max}$ and C_j are presented in Table 1. From $\mathbf{l}_{max}$ or the C_j, it is possible to conclude for each data set that cases 9 and 10 are very influential and that they share joint influence. However, the values in $\mathbf{l}_{max}$ or C_j cannot reveal the strong conditional effects between these two cases. To assess the conditional effects, we compute the matrix $\mathbf{E}$ for each data set. The results are presented in Fig. 2, where the values of E_{ij} are plotted for $j = 1, \ldots, 10, j \neq i$ with each i. The strong masking or boosting effects between cases 9 and 10 in these data sets are nicely revealed by the extremely large magnitudes of the values of E_{ij} associated with them. In Fig. 2(a), we have $E_{9,10} = 2.186$ and $E_{10,9} = 2.214$, indicating that cases 9 and 10 mask each other. In Fig. 2(b), we have $E_{9,10} = -2.195$ and $E_{10,9} = -2.195$, suggesting that cases 9 and 10 boost each other to about the same extent. In Fig. 2(c), we find $E_{10,9} = -1.307$ and $E_{9,10} = -1.231$, showing that cases 9 and 10 boost each other and the effect of case 10 on case 9 is stronger than that of case 9 on case 10.

4.2. Paul and Fung's data set

The second data set is taken from Paul & Fung (1991, Example 2). Figure 3 presents the scatterplot of the data, and it is observed that cases 7, 8 and 9 are atypical. The vector $\mathbf{l}_{max}$ and the C_j are given in Table 2. Note that the values of $\mathbf{l}_{max}$ and C_j for the 8th element are the smallest among all elements and the mutual effect between cases 8 and 9 therefore cannot be revealed. The plot of the elements in $\boldsymbol{E}$ is given in Fig. 4. We find very strong boosting effects between cases 7 and 9 ($E_{7,9} = -1.638, E_{9,7} = -1.720$) and a strong masking effect of case 9 on case 8 ($E_{9,8} = 1.321$).

4.3. Atkinson's data

The third data set with more than one explanatory variable is taken from Atkinson (1986). The dependent variable is the record time for a hill race and the explanatory variables are the distance in miles and the climb in feet. There were a total of 35 observations and the data set was fully analyzed by Atkinson (1986). He concluded that cases 7, 18 and 33 were different from the rest of the data and that the outlying nature of observation 33 was masked (in a broad sense) by observations 7 and 18. Another analysis of the data was provided by Lawrance

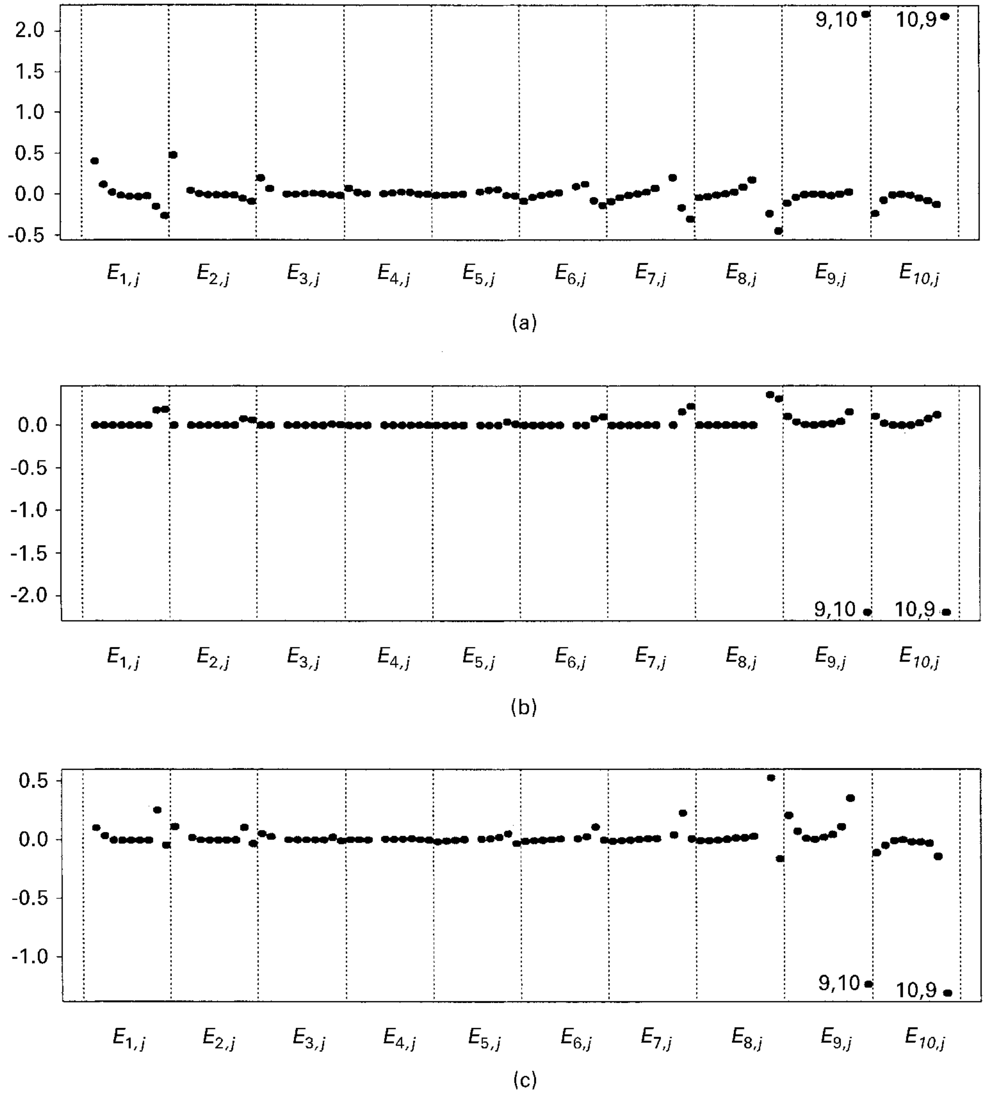

Figure 2. E_{ij} for Pena and Yohai's data sets.

(1991). Based on $\mathbf{l}_{\max}$ (see the index plot in Fig. 5(a)), he proposed paying attention to cases 7, 18, 31, 33 and 35. He also pointed out that cases 7 and 18 would be sensitive to opposite types of perturbation since they had oppositely signed elements in $\mathbf{l}_{\max}$. We have computed the values of C_j, and the results are presented in Fig. 5(b). Figure 5(b) carries similar information to Fig. 5(a), that is to say, cases 7, 18, 31, 33 and 35 in this order require special attention. We examined the conditional local influence by computing the masking matrix $\mathbf{E}$, and the result is given in Fig. 6. The following extreme effects are observed: $E_{7,33} = 0.998$, $E_{33,7} = 1.325$, $E_{18,7} = -0.861$ and $E_{31,7} = -0.545$. A strong masking effect exists between case 7 and case 33. Though case 7 is very influential, its effect is highly boosted by cases 18 and 31. Finally,

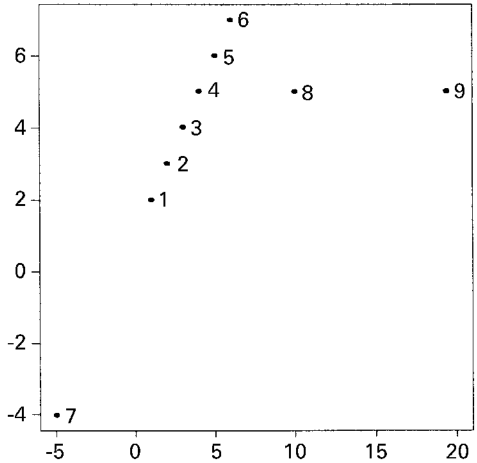

Figure 3. Scatter plot of Paul & Fung's data set.

case 18 has a mild boosting effect on case 33 ($E_{18,33} = -0.063$, not marked in the index plot), and this observation adds further insight to Atkinson's (1986) conclusion that case 33 is 'masked' by case 18.

4.4. Hawkins, Bradu and Kass's data

In order to give some idea of the performance of the measure when there are several unusual observations located at about the same place, we have analysed the data set constructed by Hawkins, Bradu & Kass (1984). The data set consists of 75 observations on four variables, of which one is the dependent variable and the other three are independent variables. The observations are constructed such that the first 10 are high-leverage outliers located at about the same place, the next four are high-leverage inliers, and the rest are ordinary observations. The index plots of $\mathbf{l}_{max}$ and the C_j are given in Fig. 7. According to the magnitudes of the coefficients in $\mathbf{l}_{max}$, the cases fall into three groups. The first group (with the largest magnitudes) is formed by cases 11–14, the second group by cases 1–10, and the third (with the smallest magnitudes) by other ordinary cases. The plot of the values of C_j given in Fig. 7(b) shows a similar pattern. The result indicates that, when there are groups of cases sharing joint influences, $\mathbf{l}_{max}$ or the C_j can be used to identify these cases. In order to facilitate comparison with other results in the literature, the values of the elements in $\mathbf{l}_{max}$ and the

Table 2. Paul and Fung's data: $\mathbf{l}_{max}$ and C_j

j	1	2	3	4	5	6	7	8	9
$\mathbf{l}_{max}$	0.036	–0.017	–0.050	–0.061	–0.051	–0.021	0.783	–0.006	–0.613
C_j	0.009	0.003	0.032	0.089	0.173	0.294	2.299	0.001	1.850

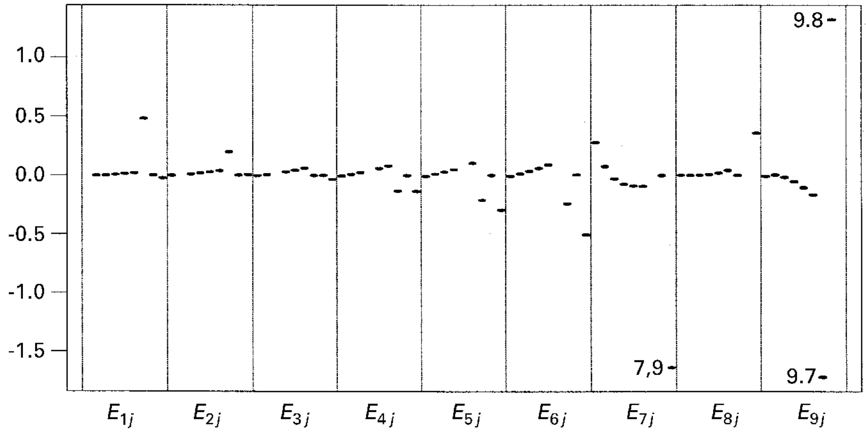

Figure 4. E_{ij} for Paul & Fung's data set.

values of the C_j for the 14 high-leverage observations are presented in Table 3. For the other observations, the magnitudes in $\mathbf{l}_{\max}$ are less than 0.018 and the values of C_j are less than 0.072. The masking matrix **E** has also been computed and the result is presented in Fig. 8. Although unusual observations are located at nearly the same place and removing just one of

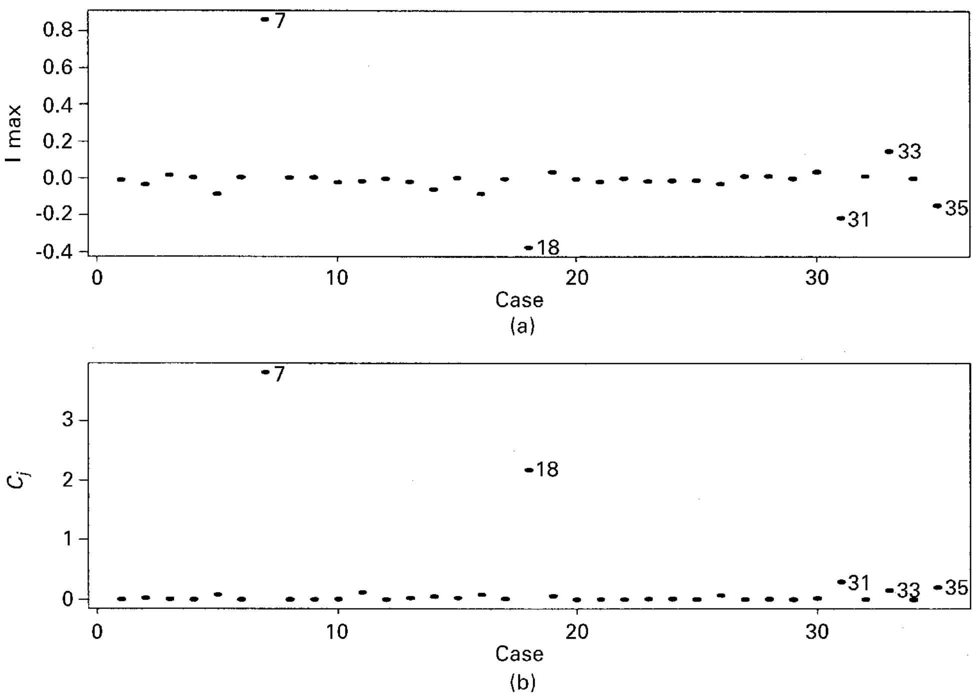

Figure 5. $\mathbf{l}_{\max}$ and C_j for Atkinson's data.

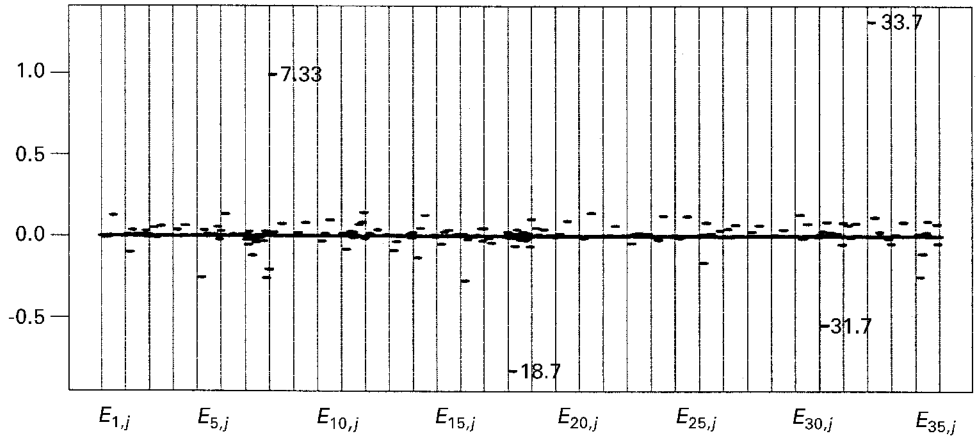

Figure 6. E_{ij} for Atkinson's data.

them is not expected to cause a drastic effect on the conditional influence measure, the values of the elements in **E** are still relatively large for the 14 unusual observations. We magnify the part of the plot corresponding to the first 20 observations and present the result in Fig. 9. Strong masking effects can be observed for case 13 on case 11, case 14 on case 13, case 11 on

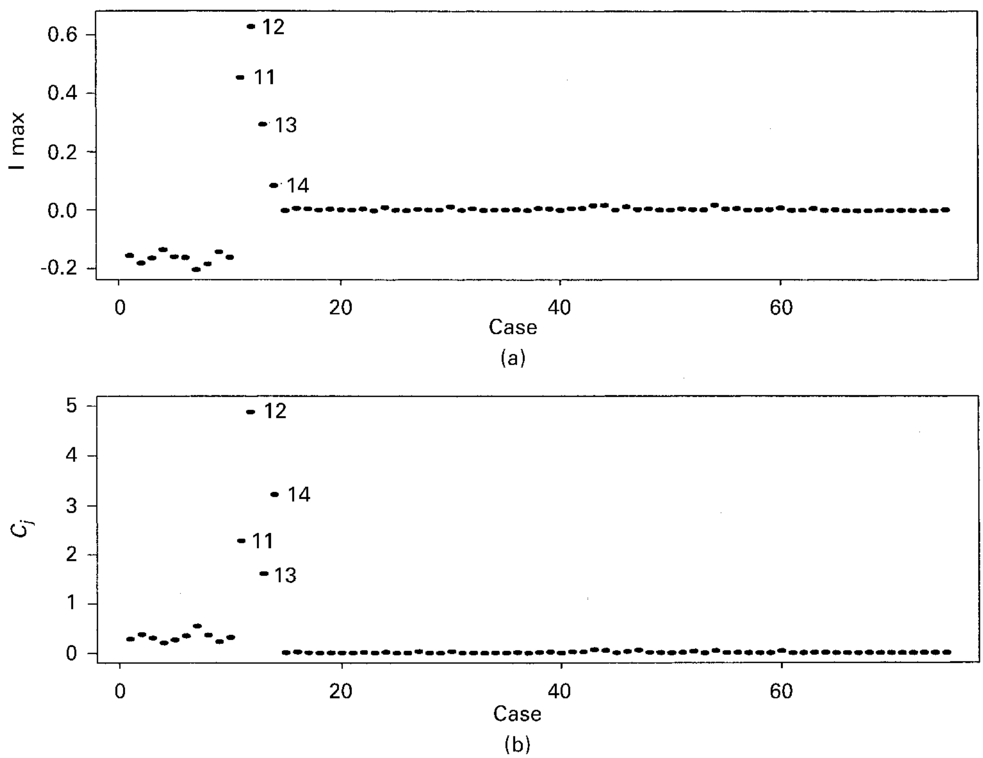

Figure 7. $\mathbf{l}_{\max}$ and C_j for Hawkins, Bradu and Kass's data.

Table 3. Hawkins, Bradu and Kass's data: $\mathbf{l}_{max}$ and C_j

j	1	2	3	4	5	6	7
$\mathbf{l}_{max}$	−0.157	−0.182	−0.165	−0.136	−0.160	−0.163	−0.205
C_j	0.284	0.377	0.305	0.209	0.270	0.352	0.547
j	8	9	10	11	12	13	14
$\mathbf{l}_{max}$	−0.185	−0.143	−0.162	0.454	0.648	0.294	0.084
C_j	0.367	0.232	0.317	2.283	4.992	1.613	3.219

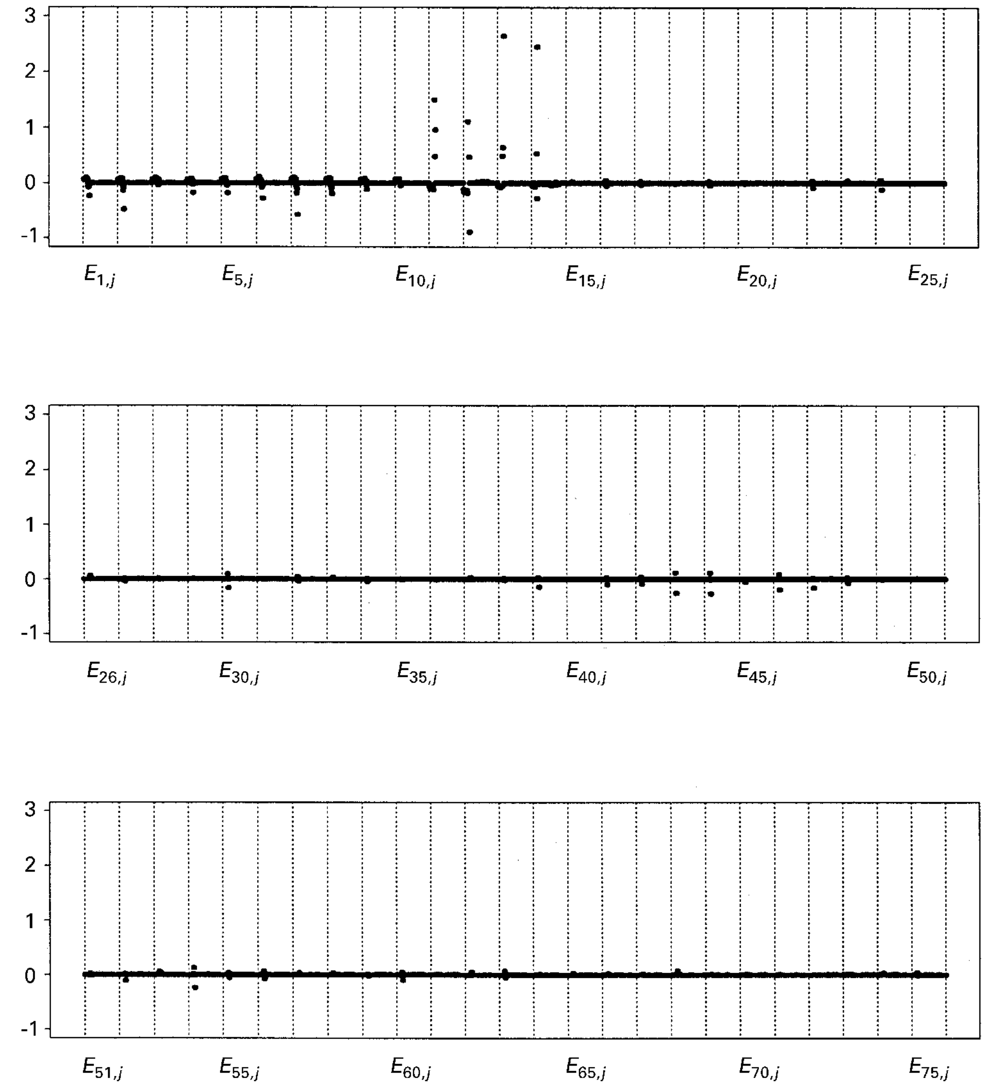

Figure 8. E_{ij} for Hawkins, Bradu & Kass's data.

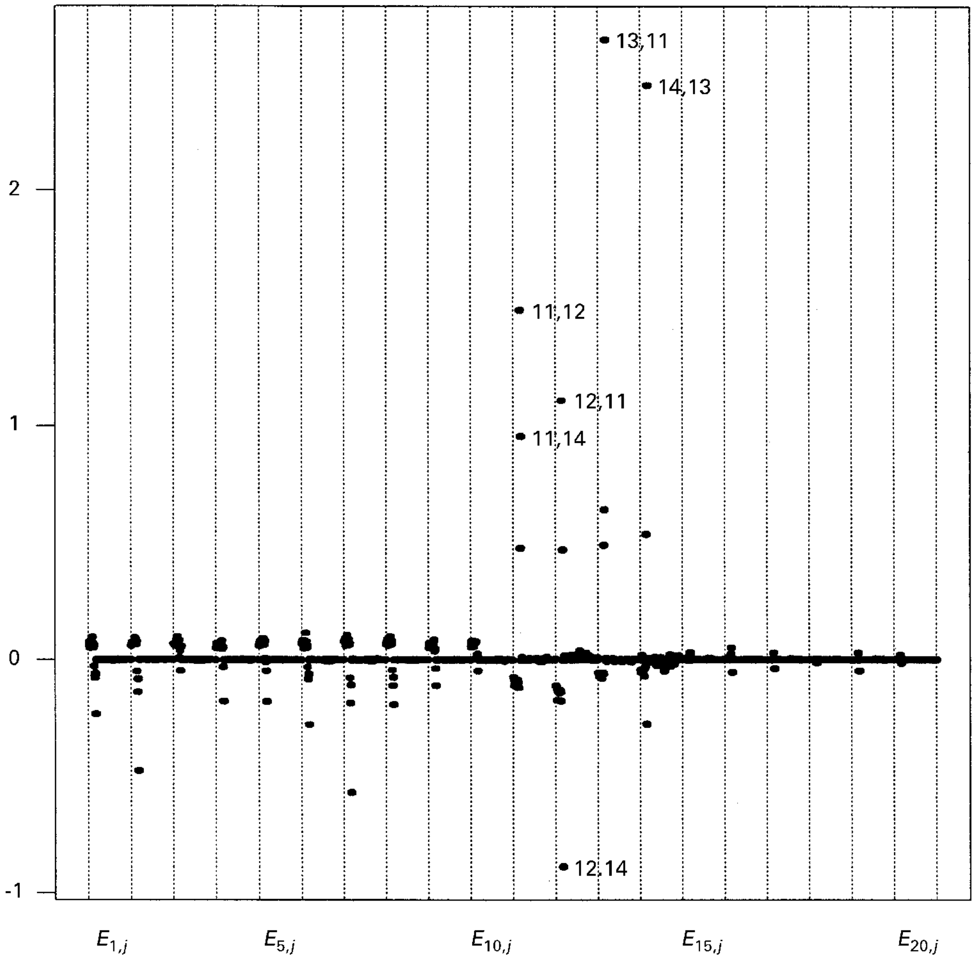

Figure 9. $E_{ij}, i,j = 1, \ldots, 20, i \neq j$ for Hawkins, Bradu and Kass's data.

case 12, case 12 on case 11 and case 11 on case 14, and a strong boosting effect can be observed for case 12 on case 14. The massing or boosting effects among cases 11–14 are relatively stronger than those among cases 1–10.

5. Discussion

The measure **E** is derived under the assumption that the variance σ^2 is known. When σ^2 is not known, we replace it by the maximum likelihood estimate computed from the original data set. From (12), we see that changes in the estimate of σ^2 lead to changes in the values of the elements of **E** in proportion. In effect, whether or not E_{ij} is relatively large is independent of the choice of the estimate of σ^2.

Pena & Yohai (1995) have used the eigenstructure of an influence matrix **M** to identify sets of potential influential observations. The influence matrix **M** is very similar to the normal curvature matrix **C** defined in (9), except that **M** uses standardized residuals instead of residuals. Pena and Yohai have also pointed out that if there are no high-leverage points and the h_j values are similar for all points, then **M** will be similar to ***C***. On the other hand, when the leverage values of the observations are very different, the eigenvector of **M** will give more weight to the high-leverage observations. Their assertion is supported by the results presented in our Tables 1 and 3. The figures in these tables can be compared to those resulting from the analyses produced by Pena & Yohai (1995, Tables 3 and 5). Nevertheless, since the objective of Pena and Yohai's procedure is to identify a set of influential observations, it addresses effects that fall within the realm of joint influence rather than conditional influence.

We apply Lawrance's concept and consider conditional local influence as the difference between the values of a diagnostic measure before and after the deletion of a particular case, where the diagnostic measure is developed from the local influence approach. Unlike diagnostic measures developed from a deletion paradigm that assess the changes of a relevant measure after the deletion of one or more cases, diagnostic measures developed using the local influence approach consider the perturbation instead of the deletion of cases. With this consideration, it is possible to carry the 'localness' over to the conditional aspect and consider local conditional influence, that is, the conditional case is perturbed rather than removed. In fact, such an approach may be more relevant if a researcher chooses to perturb rather than to remove an influential observation in a data set. However, the development of such theory is much more extensive and technical than that proposed here. We have studied such an approach using the concept of a family of influence graphs, and the results will be presented in a forthcoming paper.

Although the case-weights linear regression model provided a convenient way of illustrating the notion of conditional local influence in a concrete fashion, the concept can be generalized to other models or perturbation schemes such as the factor analysis model and the structural equation model.

When there is a group of cases that are candidates for removal, generalization of the present work to conditionality on more than one case is of interest. The generalization is theoretically straightforward but the development will be more intensive. Moreover, when the conditionality is on a group of candidates, combinatorial and hence computational problems may emerge. Nevertheless, when the conditionality concept arises from assessing the influence conditional on the deletion of the other cases (Lawrance, 1995), the problem seems to be one that is inherent to the concept.

Acknowledgement

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (RCG Ref. No. CUHK4186/98P). Y. S. Poon thanks the Department of Statistics at the Chinese University of Hong Kong for its hospitality. Both authors thank Mr Mannon Wong for his assistance in preparing the figures, and the referees together with the Editor for valuable comments.

References

Atkinson, A. C. (1985). *Plots, transformations and regression.* Oxford: Clarendon Press.

Atkinson, A. C. (1986). In discussion of 'Influential observations, high leverage points, and outliers in linear regression' by S. Chatterjee and A. S. Hadi. *Statistical Science, 1,* 379–416.

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

Bruce, A. G. and Martin, R. D. (1989). Leave-*k*-out diagnostics for time series (with discussion). *Journal of Royal Statistical Society, Series B, 51,* 363–464.

Cadigan, N. G. (1995). Local influence in structural equation models. *Structural Equation Modeling, 2,* 13–30.

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

Cook, R. D., Pena, D., & Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communication in Statistics—Theory and Methods, 17,* 623–640.

Hawkins, D. M., Bradu, D., & Kass, G. V. (1984). Location of several outliers in multiple regression data using elemental sets. *Technometrics, 26,* 197–208.

Kwan, C. W., & Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: Application to factor analysis. *Psychometrika, 63,* 35–46.

Lawrance, A. J. (1989). In discussion of 'Leave-*k*-out diagnostics for time series' by A. G. Bruce and R. D. Martin. *Journal of Royal Statistic Society, Series B, 51,* 363–424.

Lawrance, A. J. (1991). Local and deletion influence. In W. Stahel & S. Weisberg (Eds.), *Directions in Robust Statistics and Diagnostics, Part 1* (pp. 141–157). Berlin, Springer-Verlag.

Lawrance, A. J. (1995). Deletion influence and masking in regression. *Journal of 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.

Lesaffre, E., & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics, 54,* 570–582.

Paul, S. R., & Fung, K. Y. (1991). A generalized extreme studentized residual multiple-outlier-detection procedure in linear regression. *Technometrics, 33,* 339–348.

Pena, D., & Yohai, V. J. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of Royal Statistical Society, Series B, 57,* 145–156.

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

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., Wang, S. J., & Lee, S.-Y. (1999). Influence analysis of structural equation models with polytomous variables. *Psychometrika, 64,* 461–473.

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

St. Laurent, R. Y., & Cook, R. D. (1993). Leverage, local influence and curvature in nonlinear regression. *Biometrika, 80,* 99–106.

Received 10 August 1999; revised version received 2 June 2000