Journal of Applied Statistics
Vol. 34, No. 8, 997–1009, October 2007

Local Conditional Influence

WAI-YIN POON* & YAT SUN POON**

**The Chinese University of Hong Kong, Hong Kong; **Department of Mathematics, University of California at Riverside, USA*

ABSTRACT *Through an investigation of normal curvature functions for influence graphs of a family of perturbed models, we develop the concept of local conditional influence. This concept can be used to study masking and boosting effects in local influence. We identify the situation under which the influence graph of the unperturbed model contains all the information on these effects. The linear regression model is used for illustration and it is shown that the concept developed is consistent with Lawrance's (1995) approach of conditional influence in Cook's distance.*

KEY WORDS: Normal curvature, curvature function, local conditional influence, masking, linear regression

Introduction

The local influence approach is a popular approach in influence analysis, it develops diagnostic measures by examining the effect of an infinitesimal perturbation on relevant quantity. A general method for assessing the influence of local perturbation by differential geometric techniques was proposed by Cook (1986). Let $L(\theta)$ denote the log-likelihood for a postulated model, where θ is a $p \times 1$ vector of unknown parameters. Let $L(\theta \mid \omega)$ be the log-likelihood corresponding to the perturbed model with perturbation parameter vector ω, where $\omega^T = (\omega_1, \ldots, \omega_n)$ is a $n \times 1$ vector in Ξ of R^n, and Ξ represents the set of relevant perturbations. It is assumed that there is a point of null perturbation ω_0 in Ξ such that $L(\theta \mid \omega_0) = L(\theta)$ for all θ. Under some mild regularity conditions, the maximum likelihood estimate $\hat{\theta}_\omega$ with respect to any fixed perturbation parameter ω is a solution to the equation

$$\frac{\partial L}{\partial \theta_a}_{|\theta=\hat{\theta}_\omega} = 0, \text{ for } 1 \leq a \leq p \tag{1}$$

The likelihood displacement function (Cook, 1986) is given by

$$f(\omega) = 2(L(\hat{\theta}_{\omega_0} \mid \omega_0) - L(\hat{\theta}_\omega \mid \omega_0)) \tag{2}$$

Clearly, this function has a minimum at ω_0. Cook (1986) proposed using normal curvature of the influence graph α for the function f at ω_0 to assess local influence of the perturbation. This normal curvature is sometimes called the Cook's curvature.

Correspondence Address: Wai-Yin Poon, Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong. Email: wypoon@cuhk.edu.hk

0266-4763 Print/1360-0532 Online/07/080997–13 © 2007 Taylor & Francis
DOI: 10.1080/02664760600744371

Geometrically, normal curvature is a function of the first fundamental form I and the second fundamental form Π (Thorpe, 1979) of the graph of the likelihood displacement function f. They are symmetric matrices whose entries are respectively

$$\mathrm{I}_{ij} = \delta_{ij} + \frac{\partial f}{\partial \omega_i}\frac{\partial f}{\partial \omega_j} \quad \text{and} \quad \Pi_{ij} = \frac{1}{(1+ \mid \nabla_f \mid^2)^{1/2}} \frac{\partial^2 f}{\partial \omega_i \partial \omega_j} \tag{3}$$

where δ_{ij} equals 1 when $i = j$, and is zero otherwise, and $\mid \nabla_f \mid$ represents the norm of the gradient vector of f. These two forms are evaluated on vectors v and w by $\mathrm{I}(v, w) = v^T \mathrm{I} w$, and $\Pi(v, w) = v^T \Pi w$. The normal curvature of the graph α in the direction ℓ at the point ω_0 is

$$C_\ell = \frac{\Pi(\ell, \ell)}{\mathrm{I}(\ell, \ell)} = \frac{\ell^T H_f \ell}{\ell^T (I_n + \nabla_f \nabla_f^T)\ell \; (1+ \mid \nabla_f \mid^2)^{1/2}}_{|\omega=\omega_0} \tag{4}$$

where I_n is the $n \times n$ identity matrix and $H_f = (\partial^2 f / \partial \omega_i \partial \omega_j)$ is the Hessian matrix. At ω_0, the function $f(\omega)$ is minimized, therefore, $\nabla_f = 0$. The first fundamental form is equal to the standard inner product, and the second fundamental is the Hessian of the function f. It can be shown that C_{max}, the maximum of C_ℓ, is given by the maximum eigenvalue of the Hessian matrix and is achieved by the direction ℓ_{max}, that is the eigenvector of the maximum eigenvalue. Cook proposed using ℓ_{max} as a diagnostic measure and paying special attention to those elements with large magnitude in ℓ_{max}. When $\{e_i : 1 \leq i \leq n\}$ denotes the standard basis in R^n, then

$$C_{e_i} = C_i = \frac{\partial^2 f}{\partial \omega_i^2}_{|\omega=\omega_0} \tag{5}$$

Geometrically, this quantity measures the curvature of the cross section of the graph cut out by the vector e_{n+1} and the vector e_i at ω_0. From Cook's point of view, this quantity measures the local sensitivity of the model with respect to the ith perturbation parameter. We call it the ith Cook's curvature. A relation between C_i and the eigenvectors of H_f at ω_0 has been established by Poon & Poon (1999, Theorem 4) using the conformal normal curvature, which is a one–one function of the normal curvature. Specifically, their result implies that if C_{max} is sufficiently large, then the element i with large magnitude in ℓ_{max} will possess large C_i values and vice versa. In effect, if the perturbation ω_i is applied to the ith observation, then the group of observations with large C_i value is a group of influential cases.

One feature of the Cook's local influence approach is its ability to handle cases simultaneously, hence there might have an impression that masking is not an issue in the local influence approach. However, as pointed out by Lawrance (1995), although masking is broadly concerned with the limitations imposed by the use of individual cases, two notions of masking have emerged in the literature. One is associated with the idea of joint influence and the other is associated with the idea of conditional influence. Lawrance discussed these two kinds of influences and their differences in the context of the linear regression model and suggested describing the possible effects in joint influence as reducing, enhancing and swamping; and in conditional influence as masking and boosting. Specifically, let D_j be the Cook's distance obtained by deleting the jth case, then joint influence is associated with the difference between D_{ij} and D_j, where D_{ij} denotes the Cook's distance obtained by deleting both the ith and the jth case, and conditional influence is associated with the difference between $D_{j(i)}$ and D_j, where $D_{j(i)}$ is the Cook's distance for the jth case conditional on the deletion of the ith case. If $D_{j(i)} > D_j$, the effect of case i on case j is masking and if $D_{j(i)} < D_j$, the effect of case i on case j is boosting.

In terms of Lawrance's concept, the local influence approach is capable of addressing effects under the category of joint influence because the local influence approach examines the effect of perturbing a group of cases simultaneously. However, the masking effects from a conditional perspective need further attention. With this consideration, we develop the concept of local conditional influence (Poon & Poon, 2001). The concept can be used to compare the change in Cook's curvature due to slight modification of the postulated model. As a result, it can be used to study the masking and boosting effects in local influence.

We will demonstrate that while it seems necessary to investigate the influence graphs of a family of perturbed models in order to examine local conditional influence, there exists a situation under which the influence graph of the unperturbed model contains all the information on the conditional effects. In this case, local conditional influence can be studied in a much simpler manner.

Curvature Function and Conditional Effects

Curvature Function

Let the postulation parameter vector ζ be a perturbation of ω_0. Consider $L(\theta \mid \zeta)$ as a new postulated model, then the likelihood displacement function for the new postulated model is

$$f(\omega \mid \zeta) = 2(L(\hat{\theta}_\zeta \mid \zeta) - L(\hat{\theta}_\omega \mid \zeta)) \tag{6}$$

For every fixed postulation parameter ζ, we compute the Cook's curvature through evaluations of differentiations of the function $f(\omega \mid \zeta)$ with respect to $\omega^T = (\omega_1, \ldots, \omega_n)$ at $\omega = \zeta$. For each perturbation parameter, we have the Cook's normal curvature

$$C_j(\zeta) = \left(\frac{\partial^2 f}{\partial \omega_j^2}\right)_{|\omega=\zeta} \tag{7}$$

where all derivatives are with respect to ω_j.

To understand the conditional effects on the measures C_j, we need to investigate how the normal curvature function $C_j(\zeta)$ varies when the postulation parameter ζ changes in a neighbourhood of ω_0. Therefore, we calculate the partial derivatives of C_j with respect to ζ. It will be demonstrated that these partial derivatives are effective measures for assessing the masking and boosting effects under the context of local influence and the following definition will be further justified later.

Definition 1 *The jth perturbation parameter is masked by the ith parameter if $(\partial C_j/\partial \zeta_i)_{|\zeta=\omega_0} < 0$. The jth perturbation parameter is boosted by the ith parameter if $(\partial C_j/\partial \zeta_i)_{|\zeta=\omega_0} > 0$.*

Computation of the Curvature Function and a Simplification

In this section, we find straightforward ways to evaluate the normal curvature function and its derivatives with reference to the computation given in Cook (1986). From the definition of the displacement function $f(\omega \mid \zeta)$, we have

$$\frac{\partial f}{\partial \omega_i} = -2\frac{\partial L(\hat{\theta}_\omega \mid \zeta)}{\partial \omega_i} = -2\sum_{a=1}^{p} \frac{\partial L(\theta \mid \zeta)}{\partial \theta_a}_{|\theta=\hat{\theta}_\omega} \frac{\partial \hat{\theta}_{\omega,a}}{\partial \omega_i} \tag{8}$$

It follows from chain rule that

$$\frac{\partial^2 f}{\partial \omega_i \partial \omega_j} = -2 \sum_{a,b=1}^{p} \frac{\partial^2 L(\theta \mid \zeta)}{\partial \theta_a \partial \theta_b}_{|\theta=\hat{\theta}_\omega} \frac{\partial \hat{\theta}_{\omega,a}}{\partial \omega_i} \frac{\partial \hat{\theta}_{\omega,b}}{\partial \omega_j} - 2 \sum_{a=1}^{p} \frac{\partial L(\theta \mid \zeta)}{\partial \theta_a}_{|\theta=\hat{\theta}_\omega} \frac{\partial^2 \hat{\theta}_{\omega,a}}{\partial \omega_i \partial \omega_j} \tag{9}$$

Due to the definition of $\hat{\theta}_\zeta$,

$$C_{ij}(\zeta) = \frac{\partial^2 f}{\partial \omega_i \partial \omega_j}_{|\omega=\zeta} = -2 \sum_{a,b=1}^{p} \frac{\partial^2 L(\theta \mid \omega)}{\partial \theta_a \partial \theta_b}_{|\theta=\hat{\theta}_\zeta,\omega=\zeta} \frac{\partial \hat{\theta}_{\omega,a}}{\partial \omega_i}_{|\omega=\zeta} \frac{\partial \hat{\theta}_{\omega,b}}{\partial \omega_j}_{|\omega=\zeta} \tag{10}$$

Taking the differentiation of the identity (1) with respect to ω_j, we have

$$\sum_{b=1}^{p} \frac{\partial^2 L(\theta \mid \omega)}{\partial \theta_a \partial \theta_b}_{|\theta=\hat{\theta}_\omega} \frac{\partial \hat{\theta}_{\omega,b}}{\partial \omega_j} + \frac{\partial^2 L(\theta \mid \omega)}{\partial \omega_j \partial \theta_a}_{|\theta=\hat{\theta}_\omega} = 0 \tag{11}$$

Following Cook (1986), we define

$$\ddot{L} = \left(\frac{\partial^2 L(\theta \mid \omega)}{\partial \theta_a \partial \theta_b}_{|\theta=\hat{\theta}_\zeta,\omega=\zeta} \right) \triangle = \left(\frac{\partial^2 L(\theta \mid \omega)}{\partial \theta_a \partial \omega_i}_{|\theta=\hat{\theta}_\zeta,\omega=\zeta} \right) J = \left(\frac{\partial \hat{\theta}_{\omega,b}}{\partial \omega_i}_{|\theta=\hat{\theta}_\zeta,\omega=\zeta} \right) \tag{12}$$

Then $\ddot{L}$ is a $p \times p$-matrix, J is a $p \times n$-matrix, and $\triangle$ is a $p \times n$-matrix. The entries of these matrices are functions of ζ. Equation (11) is

$$\ddot{L}J + \triangle = 0, \text{ or } J = -\ddot{L}^{-1}\triangle \tag{13}$$

Therefore, the matrix $C(\zeta)$ with entries $C_{ij}(\zeta)$ is

$$C = -2\triangle^T \ddot{L}^{-1} \ddot{L} \ddot{L}^{-1} \triangle = -2\triangle^T \ddot{L}^{-1} \triangle \tag{14}$$

Now we calculate the derivatives of the matrix of functions $C(\zeta)$. By the chain rule,

$$\frac{\partial C_{ij}(\zeta)}{\partial \zeta_k} = \frac{\partial^3 f}{\partial \omega_i \partial \omega_j \partial \omega_k}_{|\omega=\zeta} + \frac{\partial^3 f}{\partial \omega_i \partial \omega_j \partial \zeta_k}_{|\omega=\zeta} \tag{15}$$

To calculate these two terms, one simply takes the appropriate derivatives of equation (9), applies the chain rule, and takes implicit derivatives of equation (11) to seek terms for substitution. After a long and tedious computation, we find that

$$\begin{aligned} \frac{\partial^3 f}{\partial \omega_i \partial \omega_j \partial \omega_k}_{|\omega=\zeta} &= 4 \sum_{a,b,c=1}^{p} \frac{\partial^3 L(\theta \mid \omega)}{\partial \theta_a \partial \theta_b \partial \theta_c}_{|\theta=\hat{\theta}_\zeta,\omega=\zeta} J_{ai} J_{bj} J_{ck} \\ &+ 4 \sum_{a,b=1}^{p} \left(\frac{\partial^3 L(\theta \mid \omega)}{\partial \theta_a \partial \theta_b \partial \omega_k} J_{ai} J_{bj} + \frac{\partial^3 L(\theta \mid \omega)}{\partial \omega_i \partial \theta_b \partial \theta_a} J_{bj} J_{ak} \right. \\ &\left. + \frac{\partial^3 L(\theta \mid \omega)}{\partial \theta_a \partial \omega_j \partial \theta_b} J_{ai} J_{bk} \right)_{|\theta=\hat{\theta}_\zeta,\omega=\zeta} + 2 \sum_{a=1}^{p} \left(\frac{\partial^3 L(\theta \mid \omega)}{\partial \theta_a \partial \omega_j \partial \omega_k} J_{ai} \right. \\ &\left. + \frac{\partial^3 L(\theta \mid \omega)}{\partial \omega_i \partial \theta_a \partial \omega_k} J_{aj} + \frac{\partial^3 L(\theta \mid \omega)}{\partial \omega_i \partial \omega_j \partial \theta_a} J_{ak} \right)_{|\theta=\hat{\theta}_\zeta,\omega=\zeta} \end{aligned} \tag{16}$$

and

$$\frac{\partial^3 f(\omega \mid \zeta)}{\partial\omega_i \partial\omega_j \partial\zeta_k}_{|\omega=\zeta} = -2 \sum_{a,b,c=1}^{p} \frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\theta_b \partial\theta_c}_{|\theta=\hat{\theta}_\zeta, \omega=\zeta} J_{ak} J_{bi} J_{cj}$$

$$- 2 \sum_{a=1}^{p} \frac{\partial^3 L(\theta \mid \omega)}{\partial\omega_i \partial\omega_j \partial\theta_a}_{|\theta=\hat{\theta}_\zeta, \omega=\zeta} J_{ak} - 2 \sum_{a,b=1}^{p} \left(\frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\theta_b \partial\omega_k} J_{ai} J_{bj} \right.$$

$$\left. + \frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\omega_i \partial\theta_b} J_{ak} J_{bj} + \frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\omega_j \partial\theta_b} J_{ak} J_{bi} \right)_{|\theta=\hat{\theta}_\zeta, \omega=\zeta} \quad (17)$$

Combine these results, we find that

$$\frac{\partial C_{ij}(\zeta)}{\partial\zeta_k} = 2 \sum_{a,b,c=1}^{p} \frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\theta_b \partial\theta_c} J_{ai} J_{bj} J_{ck} + 2 \sum_{a=1}^{p} \left(\frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\omega_j \partial\omega_k} J_{ai} + \frac{\partial^3 L(\theta \mid \omega)}{\partial\omega_i \partial\theta_a \partial\omega_k} J_{aj} \right)$$

$$+ 2 \sum_{a,b=1}^{p} \left(\frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\theta_b \partial\omega_k} J_{ai} J_{bj} + \frac{\partial^3 L(\theta \mid \omega)}{\partial\omega_i \partial\theta_b \partial\theta_a} J_{bj} J_{ak} + \frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a \partial\omega_j \partial\theta_b} J_{ai} J_{bk} \right) \quad (18)$$

where the partial derivatives are evaluated at $\theta = \hat{\theta}_\zeta, \omega = \zeta$. As a result, the measure $\partial C_j / \partial\zeta_i$ given in Definition 1 for assessing the masking and boosting effects can be obtained accordingly.

It is very useful to note the following by comparing equation (16) to equation (18).

PROPOSITION 1 *Suppose that the log-likelihood $L(\theta)$ depends on θ up to order 2, and suppose that the perturbation scheme $L(\theta \mid \omega)$ is linear in ω, then*

$$2 \frac{\partial C_{ij}(\zeta)}{\partial\zeta_k}_{|\zeta=\omega_0} = \frac{\partial^3 f(\omega \mid \omega_0)}{\partial\omega_i \partial\omega_j \partial\omega_k}_{|\omega=\omega_0} \quad (19)$$

It is worthy of note that when we study masking of the Cook's curvature, we compare the measure before and after perturbation. Therefore, the infinitesimal computation involves the Cook's curvature of different graphs. The significance of Proposition 1 is that when the log-likelihood and the perturbation is as simple as prescribed, the influence graph of the preliminary postulated model contains the information on the change of Cook's curvature at nearby models.

Case Weight Perturbation in Linear Regression

In this section, we examine our theory through case-weight perturbation of the linear regression model given by

$$y = X\theta + \epsilon, \quad E(\epsilon) = 0, \quad Var(\epsilon) = \sigma^2 I \quad (20)$$

For $1 \le i \le n$, the ith row is $y_i = x_i^T\theta + \epsilon_i = \sum_{a=1}^{p} x_{ia}\theta_a + \epsilon_i$. Here we treat x_i as a column p-vector so that x_i and θ have the same dimension. The log-likelihood corresponding to case-weight perturbation is

$$L(\theta \mid \omega) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\omega_i(y_i - x_i^T\theta)^2 \tag{21}$$

and the un-perturbed model is given by $\omega_0 = (1, \ldots, 1)^T$. From equation (21), it is clear that Proposition 1 is applicable.

The Curvature Function

Let Ω be the $n \times n$-diagonal matrix whose entries are ω_i. Let Z be the $n \times n$ diagonal matrix whose entries are ζ_i. The log-likelihood corresponding to case-weight perturbation is now given in a matrix form

$$L(\theta \mid \omega) = -\frac{1}{2\sigma^2}(y - X\theta)^T\Omega(y - X\theta) \tag{22}$$

For given ω, the weighted least square estimators are $\hat{\theta}_\omega = (X^T\Omega X)^{-1}X^T\Omega y$. When $\omega = \zeta$, the vector of residuals is $y - X(X^TZX)^{-1}X^TZy$. Let $D(\zeta)$ be the diagonal matrix whose entries are the coordinates of the vector of these residuals. Then at $\theta = \hat{\theta}_\zeta$, $\omega = \zeta$,

$$\ddot{L} = -\frac{1}{\sigma^2}X^TZX, \quad \text{and} \quad \triangle = \frac{1}{\sigma^2}X^TD(\zeta) \tag{23}$$

Therefore, for a given ζ, the second fundamental form of the graph of $f(\omega \mid \zeta)$ at $\omega = \zeta$ is

$$C(\zeta) = -2\triangle^T\ddot{L}^{-1}\triangle = \frac{2}{\sigma^2}D(\zeta)X(X^TZX)^{-1}X^TD(\zeta) \tag{24}$$

Since the model under investigation satisfies the assumptions of Proposition 1, we may calculate the changes of curvatures near the original postulated model by calculating the changes of the normal curvature of the influence graph of the original postulated model. At $\theta = \hat{\theta}_{\omega_0}$ and $\omega = \omega_0$,

$$J = (X^TX)^{-1}X^TD(\omega_0) \quad \text{and} \quad \frac{\partial^3 L(\theta \mid \omega)}{\partial\theta_a\partial\theta_b\partial\omega_i} = -\frac{1}{\sigma^2}x_{ia}x_{ib} \tag{25}$$

Let h_{ij} be the (i, j)th entry of the matrix $H = X(X^TX)^{-1}X^T$ and $r = (r_1, \ldots, r_n)^T$ be the vector of unperturbed residuals. Since $XJ = HD(\omega_0)$,

$$\begin{aligned}\frac{\partial C_{ij}}{\partial\zeta_k} &= -\frac{2}{\sigma^2}\sum_{a,b=1}^{p}\left(x_{ka}x_{kb}J_{ai}J_{bj} + x_{ia}x_{ib}J_{bj}J_{ak} + x_{ja}x_{jb}J_{ai}J_{bk}\right)\\ &= -\frac{2}{\sigma^2}\left(h_{ki}r_ih_{kj}r_j + h_{ik}r_kh_{ij}r_j + h_{ji}r_ih_{jk}r_k\right)\end{aligned} \tag{26}$$

As a result, the measure $\partial C_j/\partial\zeta_{i\,|\zeta=\omega_0}$ given in Definition 1 becomes

$$\frac{\partial C_j}{\partial\zeta_i} = -\frac{2}{\sigma^2}\Big(2r_ih_{ij}h_{jj}r_j + r_j^2h_{ij}^2\Big) \tag{27}$$

Masking Matrix

In the computations of the previous sections, the space of perturbation parameters Ξ is part of the vector space R^n. When ω_0 designates a postulated model and ζ designates a competing postulated model in the perturbation space, then any perturbation ω for ω_0 is also considered as a perturbation of ζ. For example, when we consider the case-weight perturbation in linear regression, $\omega_0 = (1, \ldots, 1, 1, 1)^T$ is the ordinary least square model. A competing postulation could be a deletion of the last case. Then $\zeta = (1, \ldots, 1, 1, 0)^T$. Now the deletion of the last two cases $\omega = (1, \ldots, 1, 0, 0)$ is a perturbation of both cases.

To calculate the change of $C_j(\zeta)$ from $\zeta = \omega_0$ to $\omega_1 = (1, \ldots, 1, 0, 1, \ldots, 1)$ when 0 is in the ith slot, we consider $C_j(\zeta(t))$ when $\zeta(t) = \omega_0 - te_i$ and $0 \le t \le 1$ where $\{e_1, \ldots, e_n\}$ is the standard basis for R^n. When t increases, that is when ω_i decreases, the weight of the ith case decreases. If $C_j(\zeta(t))$ decreases at the same time, inspired by Lawrance's concept (Lawrance, 1995), we consider the ith case boosts the effect of the jth case. Infinitesimally,

$$W_{ij} = -\frac{\partial^3 f(\omega|\omega_0)}{\partial\omega_j \partial\omega_j \partial\omega_i}_{|\omega=\omega_0} \tag{28}$$

is negative. As a contrast, when W_{ij} is positive, $C_j(\zeta(t))$ increases when the weight of the ith case decreases. It means that the ith case masks the influence of the jth case. This observation justifies our definition of masking and boosting in Definition 1. Given the formula in equation (27), the following definitions are established.

DEFINITION 2 *For case-weight perturbation for linear regression, the masking matrix W is an $n \times n$-matrix whose (i, j)th entry, when $i \neq j$ is*

$$W_{ij} = \frac{4}{\sigma^2}(2r_i h_{ij} h_{jj} r_j + r_j^2 h_{ij}^2) \tag{29}$$

and $W_{ii} = 0$.

Using Definition 1, the ith case locally masks the jth case if $W_{ij} > 0$, and the ith case locally boosts the jth case if $W_{ij} < 0$.

To identify cases with large masking or boosting effects, the 'natural gap' approach (Lawrance, 1991) can be applied. For example, an index plot of the values in the matrix W versus its coordinate (see Figures 2 and 4 later) in general can reveal W_{ij} values with large magnitudes.

Comparison with Lawrance's Approach to Cook's Distance

Lawrance (1995) discussed deletion influences under the context of linear regression, and concluded that the masking or boosting effects of Cook's distance should be assessed using a conditional perspective, namely, by assessing the change of the measure $D_{j(i)}$ relative to D_j, where $D_{j(i)}$ is the Cook's distance for the jth case after deletion of the ith case and D_j is the Cook's distance for the jth case. Define

$$H_\Omega = X(X^T \Omega X)^{-1} X^T \Omega \tag{30}$$

It can be shown that $f(\omega \mid \omega_0) = pD(\omega)$ where $D(\omega)$ is the Cook's distance function

$$D(\omega) = \frac{1}{p\sigma^2}(\hat{\theta}_\omega - \hat{\theta}_{\omega_0})^T X^T X(\hat{\theta}_\omega - \hat{\theta}_{\omega_0}) = \frac{1}{p\sigma^2}|Hy - H_\Omega y|^2 \tag{31}$$

We next demonstrate that our approach to masking is consistent with Lawrance's approach by showing that the proposed conditional displacement function $f(\omega \mid \zeta)$ has a similar property.

Note that the pair of vectors $Z^{1/2}y - Z^{1/2}H_Z y$ and $Z^{1/2}H_Z y - Z^{1/2}H_\Omega y$ is orthogonal. It follows that

$$f(\omega \mid \zeta) = \frac{1}{\sigma^2}\left((y - X\hat{\theta}_\omega)^T Z(y - X\hat{\theta}_\omega) - (y - X\hat{\theta}_\zeta)^T Z(y - X\hat{\theta}_\zeta)\right)$$
$$= \frac{1}{\sigma^2}|Z^{1/2}H_Z y - Z^{1/2}H_\Omega y|^2 = \frac{1}{\sigma^2}|(Z^{1/2}X)(\hat{\theta}_\omega - \hat{\theta}_\zeta)|^2 \quad (32)$$

Let $\zeta = \omega_0 - e_i$ and $\omega = \omega_0 - e_i - e_j$. Let $X_{(i)}$ be the matrix X with its ith row deleted, $y_{(i)}$ be the vector y with its ith coordinate deleted, and so on. Then $X^T ZX = (Z^{1/2}X)^T(Z^{1/2}X) = X_{(i)}^T X_{(i)}$. The ith coordinate of the vector $Z^{1/2}X\hat{\theta}_\zeta = (Z^{1/2}X)((Z^{1/2}X)^T(Z^{1/2}X))^{-1}(Z^{1/2}X)^T(Z^{1/2}y)$ is equal to zero. When its ith coordinate is deleted, it is the $(n-1)$-vector $X_{(i)}\hat{\theta}_{(i)} = X_{(i)}(X_{(i)}^T X_{(i)})^{-1}X_{(i)}y_{(i)}$. Similarly, ith coordinate of the vector $Z^{1/2}X\hat{\theta}_\omega = (Z^{1/2}X)((\Omega^{1/2}X)^T(\Omega^{1/2}X))^{-1}(\Omega^{1/2}X)(\Omega^{1/2}y)$ is equal to zero. If its ith coordinate is deleted, it is equal to the $(n-1)$-vector $X_{(i)}\hat{\theta}_{(i,j)} = X_{(i)}(X_{(i,j)}^T X_{(i,j)})^{-1}X_{(i,j)}y_{(i,j)}$. It follows that

$$\frac{1}{p}f(\omega_0 - e_i - e_j \mid \omega_0 - e_i) = \frac{1}{p\sigma^2}|Z^{1/2}X(\hat{\theta}_\omega - \hat{\theta}_\zeta)|^2 = \frac{1}{p\sigma^2}|X_{(i)}(\hat{\theta}_{(i,j)} - \hat{\theta}_{(i)})|^2$$
$$= \frac{1}{p\sigma^2}(\hat{\theta}_{(i,j)} - \hat{\theta}_{(i)})^T X_{(i)}^T X_{(i)}(\hat{\theta}_{(i,j)} - \hat{\theta}_{(i)}) = D_{j(i)}$$

where $D_{j(i)}$ is the Cook's distance of the jth case subject to the condition that the ith case is deleted.

These computations demonstrate that our choice for $f(\omega \mid \zeta)$ is consistent with Lawrance's approach to conditional influence.

Examples

For illustration, we first consider the three data sets (a), (b) and (c) given by Pena & Yohai (1995, Example 1). The scatter plots of these data sets are presented in Figure 1. In all three data sets, cases 1 through 8 are good data points and cases 9 and 10 are unusual. In (a), both cases 9 and 10 produce the same effect; in (b), the two cases produce opposite effects; and in (c), case 9 appears as an outlier due to the effect of case 10. For each data set, the values of C_i as described by equation (5) and the coefficients in the direction ℓ_{max} which maximizes equation (4) were computed and presented in Table 1. In data set (a), the value of C_{10} and the 10th element in ℓ_{max} are relatively large, suggesting case 10 is a very influential one. Moreover, although the magnitudes of the values for case 9 are relatively large, indicating cases 9 and 10 may share joint influences, a similar phenomenon is also observed for other cases, such as case 7 and case 8; and the strong masking effect between cases 9 and 10 cannot be highlighted by studying only the values of ℓ_{max} and C_i. However, the strong masking effects between these two cases can be identified easily by studying the masking matrix W in equation (28). Figure 2 presents the index plots of W_{ij} with $j = 1, \ldots, 10$, $j \neq i$ for each fixed i. In (a), the values $W_{9,10} = 1.932$ and $W_{10,9} = 1.266$ are clearly larger than other W_{ij}'s, suggesting that cases 9 and 10 mask each other substantially. For the data set (b), we see from Table 1 that all the entries for cases 1 to 8 are very small but those for cases 9 and 10 are large in magnitude, showing cases 9 and 10 share joint effects. The values of

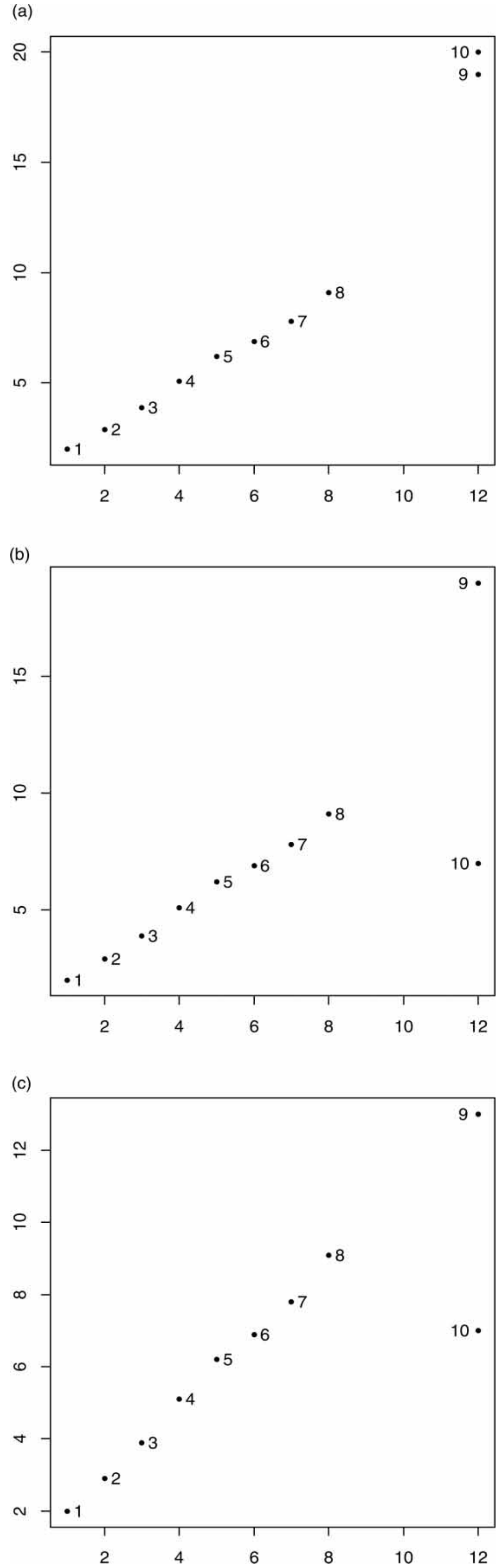

Figure 1. Pena & Yohai's data set

Table 1. Normal curvature of Pena & Yohai's data sets

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

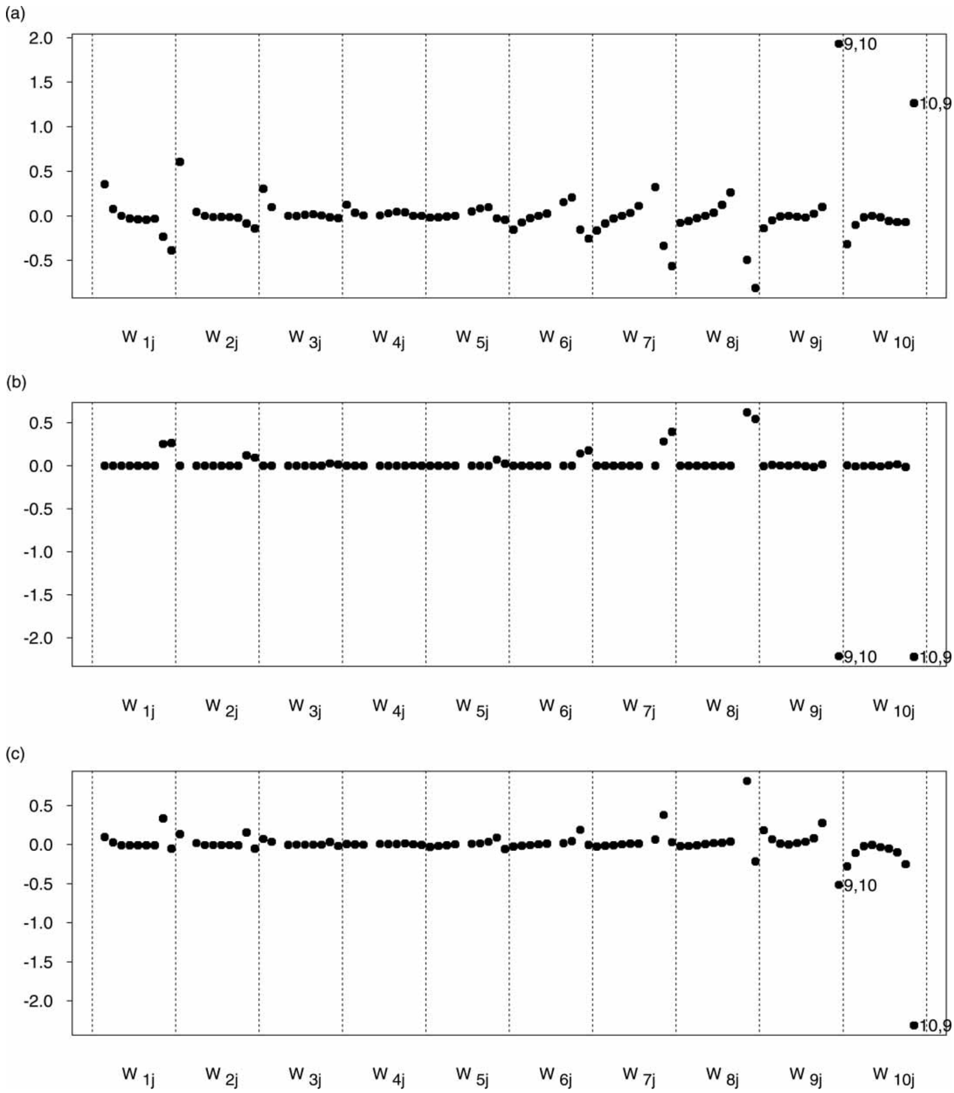

Figure 2. W_{ij} for Pena & Yohai's data

$W_{9,10}$ and $W_{10,9}$ are -2.216 and -2.222 respectively. The magnitudes of these values are relatively large as evidenced by the index plot of W_{ij} given in Figure 2(b), suggesting cases 9 and 10 boost the effects of each other. For the data set (c), it is possible to conclude from the values of ℓ_{max} and C_i that cases 9 and 10 share joint effects; however, the index plot of W_{ij} provides the additional information that there is a strong boosting effect of case 10 on case 9 ($W_{10,9} = -2.313$) but the boosting effect of case 9 on case 10 is comparatively mild ($W_{9,10} = -0.514$).

The second data set is a real data set with 35 observations 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. The data set had been fully analyzed by Atkinson and 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 (1991). Based on the coefficients in ℓ_{max} which are plotted in Figure 3(a), he proposed paying

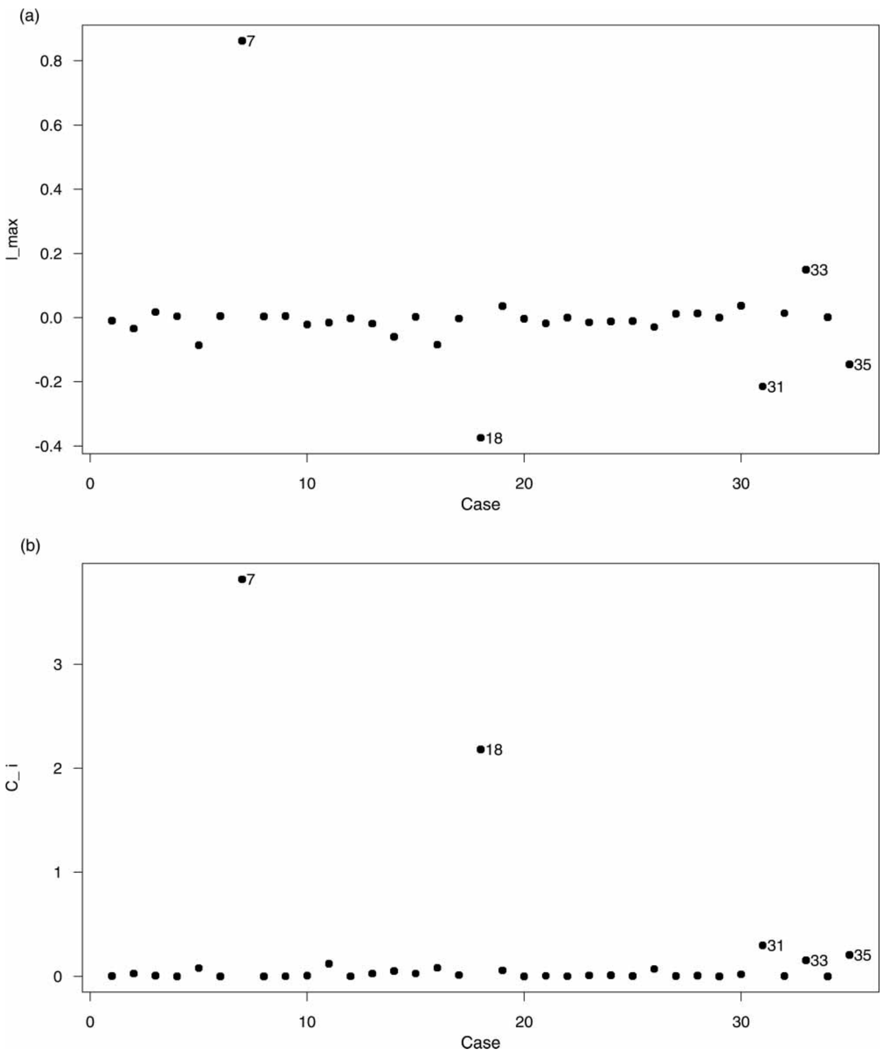

Figure 3. Index plot of ℓ_{max} and C_i for Hill's Data

special attention to cases 7, 18, 31, 33 and 35. We computed the values of C_i and W_{ij} and presented the results respectively in Figure 3(b) and Figure 4. Figure 3(b) provides similar information to Figure 3a, namely, the cases 7, 18, 31, 35 and 33 are influential cases. From Figure 4, we found several W_{ij} values which are large in magnitude and they are related to those cases already identified as influential. We presented these values in Table 2 and noted the following phenomena. There are strong masking effects between case 7 and case 33 ($W_{33,7} = 1.988$, $W_{7,33} = 0.531$). Case 7 is a very influential case but its

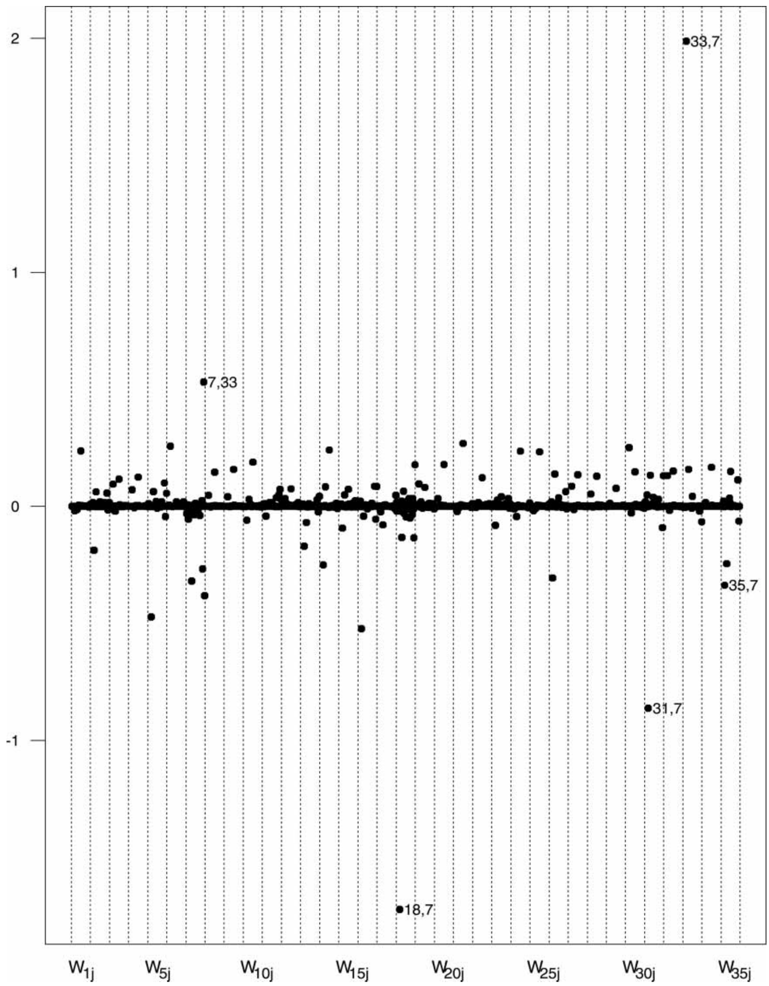

Figure 4. W_{ij} for Hill's data

Table 2. W_{ij} for Hill's data set

$i \backslash j$	7	18	31	33	35
7		0.013	−0.267	0.531	−0.382
18	−1.723		0.033	−0.134	0.178
31	−0.863	0.019		−0.091	0.131
33	1.988	0.043	−0.020		−0.066
35	−0.337	0.149	0.113	−0.063	

effect is in fact boosted strongly by case 18 ($W_{18,7} = -1.723$), and to a lesser extent by case 31 ($W_{31,7} = -0.863$). Finally, the effect of case 18 on case 33 is a mild boosting effect ($W_{18,33} = -0.134$), this observation adds further insight to Atkinson's (1986) conclusion that case 33 is 'masked' by case 18.

Concluding Remark

We use the linear regression model as an example to illustrate the details of the proposed theory and its compatibility with relevant concepts given in the literature. The linear regression model was used because of its popularity and because Lawrance's concept of conditional influence (Lawrance, 1995) was developed under the context of a linear regression model. Generalization of the proposed theory to other models is possible and sensible applications of the proposed theory to other situations so to facilitate the studies of conditional influence are interesting topics for further research.

Acknowledgements

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 (Project No. CUHK4242/03H).

References

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

Cook, R.D. (1986) Assessment of local influence (with discussion), *Journal of Royal Statistical Society, B*, 2, pp. 133–169.

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).

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

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

Poon, W.Y. & Poon, Y.S. (1999) Conformal normal curvature and assessment of local influence, *Journal of Royal Statistical Society, B*, 61, pp. 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, pp. 177–191.

Thorpe, J.A. (1979) *Elementary Topics in Differential Geometry* (New York: Springer).