J. R. Statist. Soc. B (1986)
48, *No.* 2, *pp.* 133–169

Assessment of Local Influence

By R. DENNIS COOK†

University of Minnesota, USA

[*Read before the* Royal Statistical Society, *at a meeting organized by the* Research Section *on Wednesday, January 15th, 1986*, Professor A. F. M. Smith *in the Chair*]

SUMMARY

Statistical models usually involve some degree of approximation and therefore are nearly always wrong. Because of this inexactness, an assessment of the influence of minor perturbations of the model is important. We discuss a method for carrying out such an assessment. The method is not restricted to linear regression models, and it seems to provide a relatively simple, unified approach for handling a variety of problems.

Keywords: COLLINEARITY; CURVATURE; DIAGNOSTICS; INFLUENCE GRAPHS; INFLUENTIAL OBSERVATIONS; LOGISTIC REGRESSION

1. INTRODUCTION

Statistical models are extremely useful devices for extracting and understanding the essential features of a set of data. Models, however, are nearly always approximate descriptions of more complicated processes and therefore are nearly always wrong. Because of this inexactness, the study of the variation in the results of an analysis under modest modifications of the problem formulation becomes important. If a minor modification of an approximate description seriously influences key results of an analysis, there is surely cause for concern. On the other hand, if such modifications are found to be unimportant, the sample is robust with respect to the induced perturbations and our ignorance of the precise model will do no harm (Barnard 1980).

Although an assessment of the influence of a model perturbation is generally considered to be useful, few general methods are available for carrying out such an assessment in contexts other than normal linear regression, and much of the recent work is concerned with only the perturbation scheme in which the weights attached to individual or groups of cases are modified. Cook (1977, 1979) and Belsley, Kuh and Welsch (1980) propose diagnostics for assessing the influence of case-weight perturbations in linear regression. For the most part, the case-weights are restricted to be either 0 or 1 so that a case is either deleted or retained at full weight. These ideas are adapted for use in logistic regression by Pregibon (1981). In recent years, deleting cases has become a popular basis for studying sensitivity in statistical problems: Moolgavkar, Lustbader and Venson (1984) give a number of useful results for general exponential families, and Storer and Crowley (1985) study the change in parameter estimates from general conditional likelihoods. Cook and Wang (1983) investigate the change in the transformation parameter for a linear regression response variable, while Lustbader and Moolgavkar (1985) concentrate on the score statistic. Oman (1984) develops measures for assessing the influence of individual cases in calibration problems. Extensions of D_i (Cook, 1977) for use in the errors-in-variables problem are proposed by Kelly (1984).

Andrews and Pregibon (1978), Atkinson (1982) and Johnson and Geisser (1982) also propose diagnostics based on case deletion schemes. For a review of these works and related literature, see Cook and Weisberg (1982).

† *Present address*: Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108, USA.

 0035–9246/86/47133 $2.00

Attempts to provide a firm foundation for diagnostics based on case-weight perturbation schemes are described in Cook and Weisberg (1982) and Welsch (1982). These attempts are based on the influence curve, a construction that relies on an appropriate functional of the true underlying distribution function. The influence curve has been of value in the formulation of robust estimators, but it may be more of a hindrance than a help in the present context. To employ this idea for the construction of an influence diagnostic we must construct the influence curve, choose one of the many sample versions and then select a suitable norm. Even in normal linear regression this process seems to obscure rather than illuminate the problem at hand. The difficulty involved in carrying out the program for more complicated settings is a further annoyance.

Case-deletion is not the only paradigm that has been used to investigate the consequences of modest modifications of a statistical model. Hodges and Moore (1972), Davies and Hutton (1975), and Belsley (1984), for example, consider various aspects of perturbing the explanatory variables in linear regression. Polasek (1984) and Leamer (1984) investigate the sensitivity of estimated coefficients in changes in the entire covariance matrix of the errors in linear regression. Emerson, Hoaglin and Kempthorne (1984) use perturbations of the response variable to study leverage in an additive-plus-multiplicative model for a two-way table of data. Although these and similar past investigations differ in many important respects, they all reflect a common concern for the inevitable imprecision of a statistical model, and have greatly increased our awareness of the importance of various aspects of the data, particularly with regard to individual cases.

Other types of methods have been proposed for dealing with inexactness in a statistical model. In a Bayesian context, Box (1980) adds a discrepancy parameter α to the target model and then uses the predictive distribution conditional on α to measure deviations from a target value α^*. He also provides a forceful discussion on the interplay between scientific learning and statistical practice, with emphasis on the role of model criticism. Robust methods, on the other hand, are designed to be insensitive to selected aspects of the model or data. Robust methods are surely important, particularly for automated analysis, and may occasionally provide useful diagnostic information, but their stage of development does not match the concerns that arise in practice (Carroll, 1983). In many situations, the exclusive use of robust methods can obscure important substantive problems. Carefully constructed models in combination with appropriate diagnostic methods still provide a useful basis for thorough statistical analysis. For further discussion on the interplay between various methods, see Cook and Weisberg (1983a) and Carroll and Ruppert (1985).

This paper presents a general method for assessing the local influence of minor perturbations of a statistical model. The method relies on a well-behaved likelihood and certain elementary ideas from differential geometry, and seems to provide a relatively simple, unified approach for handling a variety of problems. A distinguishing feature of this method is its use of log-likelihood contours to gauge influence. Pregibon (1981), Cook and Weisberg (1982) and Cook and Wang (1983) use this same idea in combination with case-deletion diagnostics. Barnard (1980) gives a brief general discussion on using the likelihood to assess the consequences of model perturbations. Although this paper is concerned primarily with local influence, some discussion of assessing global influence, which is a significantly more difficult problem, will be given also.

In the next Section, we introduce the idea of an influence graph, a notion which seems fundamental to the study of influence as described earlier in this Section. In Section 3, we discuss numerical summaries of influence graphs. Several illustrations are given in the remaining Sections.

2. INFLUENCE GRAPHS

2.1. *Motivation*

Consider the standard linear regression model

$$Y = X\beta + \epsilon \tag{1}$$

where the elements ϵ_i of the $n \times 1$ vector ϵ are assumed to be independent normal random variables with mean zero and known variance σ^2. Collectively, the ith observation y_i on the response variable in combination with the associated values for the explanatory variables will be referred to as the ith *case*. To motivate the developments of this section, we use model (1) and the following form of the influence statistic D_i proposed by Cook (1977),

$$D_i = \| \hat{Y} - \hat{Y}_{(i)} \|^2 / p\sigma^2 \qquad (2)$$

where $\hat{Y}$ and $\hat{Y}_{(i)}$ are the $n \times 1$ vectors of fitted values based on the full data and the data without case i, respectively, and p is the dimension of β. A similar motivation can be constructed by using other case-deletion diagnostics.

The statistic D_i can be usefully viewed as a basis for detecting cases that should be carefully inspected for gross errors. The finding of a gross error must necessarily force the removal or correction of the corresponding case, and such actions may cause a substantial change in the results of an analysis if D_i is large.

Generally, case-deletion diagnostics allow for only one of two possibilities: a case is either as specified by the model or totally unreliable (variance $\to \infty$). Other reasonable and equally important concerns are not reflected by such diagnostics. For example, we might postulate a model with constant variance but admit that the true variances could range between $\sigma^2/2$ and $2\sigma^2$, a level of heteroscedasticity that will often go undetected in practice. To investigate this specific concern, we use the following slightly more general version of D_i,

$$D_i(\omega) = \| \hat{Y} - \hat{Y}_\omega \|^2 / p\sigma^2 \qquad (3)$$

where $\hat{Y}_\omega$ is the vector of fitted values obtained when the ith case has weight ω and the remaining cases have weight 1. Of course, as $\omega \to 0$, var $(\epsilon_i) \to \infty$ and $D_i(\omega) \to D_i$. If $D_i(\omega)$ is large then the stipulation that the ith case has variance σ^2/ω rather than σ^2 will lead to substantial changes in the results of the analysis.

At first glance, it might seem that D_i and $D_i(\omega)$ would always give essentially the same information. This does not seem to be the situation, however. Figure 1 gives plots of $pD_i(\omega)$

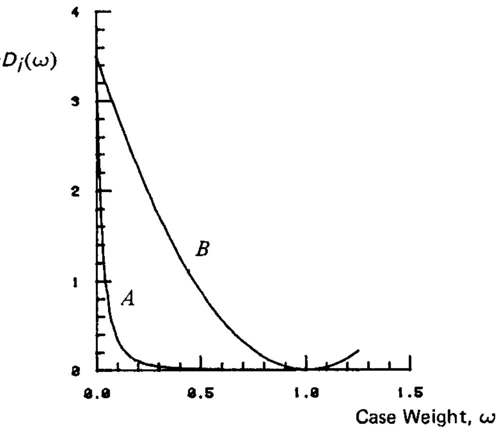

Fig. 1. $pD_i(\omega)$ versus ω for two possible cases A and B for model (1)

versus ω for two possible cases A and B from model (1). The details behind Fig. 1 will be presented in Sections 3 and 4.2. For now we note that the analysis is clearly more sensitive to perturbations in the weight attached to case B since $D_B(\omega) - D_A(\omega) \geq 0$ and for some ω this difference is substantial. We must have $D_A(1) = D_B(1) = 0$, of course. However, in Fig. 1, $D_A(0) = D_B(0)$ so that the two cases will be judged to be equally influential when using D_i. It seems clear that case deletion diagnostics alone are not sufficient to handle concerns other than gross errors. In particular, for a more complete understanding of the influence of a single case it is necessary to investigate the behaviour of $D_i(\omega)$ at values of ω other than $\omega = 0$.

In the next section we extend these ideas to general models in which ω can be used to perturb model components other than case-weights. This extension is partially motivated by the following relationship between $D_i(\omega)$ and the log-likelihood $L(\beta)$ for model (1),

$$pD_i(\omega) = [\,\| Y - \hat{Y}_\omega \|^2 - \| Y - \hat{Y} \|^2\,]/\sigma^2$$

$$= 2[L(\hat{\beta}) - L(\hat{\beta}_\omega)] \tag{4}$$

where $\hat{\beta} = \hat{\beta}_{\omega=1}$ and $\hat{\beta}_\omega$ is the maximum likelihood estimator of β when the ith case has weight ω. The form of this relationship is a consequence of the statistical structure assumed for the errors in model (1).

2.2. *Development*

For a given set of observed data, let $L(\theta)$ denote the log-likelihood corresponding to the postulated model, where θ is a $p \times 1$ vector of unknown parameters. We introduce perturbations into the model through the $q \times 1$ vector ω which is restricted to some open subset Ω of R^q. Generally, ω can reflect any well-defined perturbation scheme and thus is not restricted to be a collection of case weights. For example, ω might be used to induce a minor modification of the explanatory variables in a generalized linear model, or to perturb the entire covariance matrix of the errors in a normal linear model. As illustrated in later examples, ω must be chosen carefully so that the application is sensible. For now we assume this choice to have been made.

Let $L(\theta \mid \omega)$ denote the log-likelihood corresponding to the perturbed model for a given ω in Ω. We assume that there is an ω_0 in Ω such that $L(\theta) = L(\theta \mid \omega_0)$ for all θ. Finally, let $\hat{\theta}$ and $\hat{\theta}_\omega$ denote the maximum likelihood estimators under $L(\theta)$ and $L(\theta \mid \omega)$, respectively, and assume that $L(\theta \mid \omega)$ is twice continuously differentiable in (θ^T, ω^T).

To assess the influence of varying ω throughout Ω, we initially consider the *likelihood displacement*

$$LD(\omega) = 2\,[L(\hat{\theta}) - L(\hat{\theta}_\omega)]. \tag{5}$$

In a particular problem, specific characteristics of $\{\hat{\theta}_\omega \mid \omega \in \Omega\}$ might be relevant, but $LD(\omega)$ is a useful universally applicable feature that can be interpreted in terms of the large sample confidence region for θ (Cox and Hinkley, 1974, Chapter 9)

$$\{\theta \mid 2[L(\hat{\theta}) - L(\theta)] < X^2_\alpha(p)\}.$$

Here, $X^2_\alpha(p)$ is the upper α probability point of a chi-squared distribution with p degrees of freedom. The motivation for (5) comes largely from (4), but some alternatives will be discussed later. For further discussion see Cook and Weisberg (1982, Chapter 5) and Pregibon (1981).

From this perspective, a graph of $LD(\omega)$ versus ω contains essential information on the influence of the perturbation scheme in question. It is useful to view this graph as the geometric surface formed by the values of the $(q + 1) \times 1$ vector

$$\alpha(\omega) = \begin{pmatrix} \omega \\ LD(\omega) \end{pmatrix} \tag{6}$$

as ω varies throughout Ω. In differential geometry a surface of this form is frequently called a Monge patch (Millman and Parker, 1977). We will refer to $\alpha(\omega)$ as an *influence graph* since it is the graph of $LD(\omega)$ that displays the influence of the perturbation scheme. In this terminology, Fig. 1 displays two possible influence graphs for the scheme in which the weight attached to a single case in linear regression is varied.

The rationale that led to the influence graph $\alpha(\omega)$ is not the only reasonable approach, of course. Suppose that we partition $\theta^T = (\theta_1^T, \theta_2^T)$, where θ_i is $p_i \times 1$, and agree that only θ_1 is of interest. In this situation the analogue of (6) is

$$\alpha_s(\omega) = \begin{pmatrix} \omega \\ LD_s(\omega) \end{pmatrix} \tag{7}$$

where

$$LD_s(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_{1\omega}, g(\hat{\theta}_{1\omega}))],$$

$g(\theta_1)$ is the function that maximizes $L(\theta_1, \theta_2)$ for each fixed θ_1, $\hat{\theta}_{1\omega}$ is determined from the partition $\hat{\theta}_\omega^T = (\hat{\theta}_{1\omega}^T, \hat{\theta}_{2\omega}^T)$, and $L(\theta_1, g(\theta_1))$ is the likelihood profile for θ_1. The motivation behind (7) comes in part from the large sample confidence region for θ_1, (Cox and Hinkley, 1974, Chapter 9)

$$\{\theta_1 \mid 2[L(\hat{\theta}) - L(\theta_1, g(\theta_1))] < X_\alpha^2(p_1)\}.$$

The influence graph defined at (6) and its counterpart for parameter subsets defined at (7) are based on using the contours of the postulated log-likelihood $L(\theta)$ to measure the amount that $\hat{\theta}_\omega$ is displaced from $\hat{\theta}$, but other formulations are possible. For example, we might consider using the contours of the perturbed log-likelihood $L(\theta \mid \omega)$ to compare $\hat{\theta}$ and $\hat{\theta}_\omega$. This approach leads to the influence graph

$$\alpha'(\omega) = \begin{pmatrix} \omega \\ LD'(\omega) \end{pmatrix} \tag{8}$$

where $LD'(\omega) = 2[L(\hat{\theta}_\omega \mid \omega) - L(\hat{\theta} \mid \omega)]$.

In the construction of $\alpha'(\omega)$, the moving frame of reference $L(\theta \mid \omega)$ is used to compare $\hat{\theta}_\omega$ and $\hat{\theta}$, while $\alpha(\omega)$ was constructed by using the fixed frame of reference $L(\theta)$ for the same comparison. Influence diagnostics based on variable frames of reference have been proposed in the statistical literature. Notably, Belsley, Kuh and Welsch (1980) use $\hat{\sigma}_{(i)}$, the sample standard deviation based on the reduced data, as a scale factor in their measure DFFITS of the influence of the ith case in linear regression. Since $\hat{\sigma}_{(i)}$ changes from case to case, the reference frame is variable and in this sense DFFITS is similar to the influence graph based on $L(\theta \mid \omega)$. Although diagnostics based on fixed reference frames seem easier to motivate and interpret, it will be shown in Section 3.2 that α and α' are equivalent for the local approach considered here.

Ideally, we would like a complete influence graph, such as those displayed in Fig. 1, to assess influence in a particular problem. Clearly, this is possible in only the simplest situations so that it becomes necessary to consider other methods for extracting the information contained in an influence graph. Global measures of influence, which characterize the behaviour of an influence graph over all of Ω, are generally much more difficult to construct in practice than local measures which characterize behaviour in a neighbourhood of a selected ω, say ω^*.

The various influence diagnostics that rely on case-deletion can be regarded as global measures since they are designed to measure total change at various corners of $\Omega = (0; 1)^n$, where n is the sample size. Systematically searching the corners of Ω, however, can be a difficult task. Single case-deletion diagnostics can be computationally intensive and suffer from a form of masking. Group deletion methods are not easily implemented or well understood, although the recent work by Gray and Ling (1984) may be useful in linear regression (Carroll and Ruppert, 1985). In contrast to case-deletion diagnostics, the methodology developed in the following Sections is relatively easy to use for the identification of groups of cases that may require special attention. In addition, from Fig. 1 and the discussion of Section 2.1, it is clear that the behaviour of an influence graph around $\omega^* = \omega_0 = 1$ is certainly relevant.

In the next section we suggest a local measure of influence for characterizing the behaviour of an influence graph around $\omega^* = \omega_0$.

3. LOCAL INFLUENCE

In this Section we use geometric normal curvatures to characterize the behaviour of an influence graph around ω_0, although the essential results can be obtained by using less descriptive but more standard methods of analysis. The normal curvature of a surface ($\alpha(\omega)$ in this application) should be discussed in any first text on differential geometry. Some background information is available in Bates and Watts (1980). For convenience we use $\alpha(\omega)$ as defined in (6) to develop normal curvatures. The other types of influence graphs discussed in Section 2.2 will be compared later in this section. Also we concentrate on studying curvatures at ω_0. Curvatures at points other than ω_0 may be of some value in assessing the global behaviour of an influence graph, and a few results along these lines are briefly mentioned at the end of Section 3.1.

3.1. *Curvatures for* $\alpha(\omega)$

When $q = 1$, the influence graph $\alpha(\omega) = \{\alpha_i(\omega)\}$ reduces to a *plain curve* as illustrated in Fig. 1. The curvature of such plane curves at ω_0 is (Stoker 1969, p. 26; Goetz 1968, p. 84)

$$C = |\dot{\alpha}_1 \ddot{\alpha}_2 - \dot{\alpha}_2 \ddot{\alpha}_1| / (\dot{\alpha}_1^2 + \dot{\alpha}_2^2)^{3/2} \tag{9}$$

where the first and second derivatives $\dot{\alpha}_i$ and $\ddot{\alpha}_i$ are evaluated at ω_0. This curvature can be viewed as the inverse of the radius of the circle which best approximates α at ω_0, or as the rate of change of the angle that the tangent vector makes with the horizontal axis with respect to arc length along the curve. In standard differential geometry texts, curvature is developed in the arc length parameterization and a sign is often attached, but for our purposes (9) will be sufficient.

Since $\dot{\alpha}_1 = 1$ and $\dot{\alpha}_2 = \ddot{\alpha}_1 = 0$, C reduces to

$$C = \ddot{\alpha}_2 = \ddot{LD}(\omega_0)$$

which must necessarily be positive since $LD(\omega)$ achieves a local minimum at ω_0. As developed in Section 4.2, this curvature can be evaluated for the graphs of Fig. 1: $C = 2(0.05)^2 pD_i$ for case A and $C = 2(0.99)^2 pD_i$ for case B. Clearly, C easily distinguishes between the two influence graphs in this figure.

When $q > 1$, an influence graph is a surface in R^{q+1} and the notion of curvature becomes a bit more complicated. We are specifically interested in a description of how the surface $\alpha(\omega)$ deviates from its tangent plane at ω_0. Such a description can be obtained by studying the curvature of certain curves on the surface that pass through $\alpha(\omega_0)$. Visualized in R^3, the required curves are the *normal sections* formed by the intersection of the surface with planes containing the vector that is normal (orthogonal) to the tangent plane at ω_0 (Stoker 1969, p. 88). The curvatures of these normal sections are called *normal curvatures* and these are the curvatures that we use to characterize the behaviour of an influence graph around ω_0.

Because $LD(\omega)$ achieves a local minimum at ω_0, a simple representation of normal sections is possible. To construct a normal section, consider a straight line in Ω passing through ω_0. Such a line can be represented by

$$\omega(a) = \omega_0 + al \tag{10}$$

where $a \in R^1$ and l is a fixed nonzero vector of unit length in R^q. This line generated a *lifted line* on the influence graph $\alpha(\omega)$ passing through $\alpha(\omega_0)$. Each direction l specifies such a lifted line and each lifted line corresponds to a normal section. The equivalence of lifted lines and normal sections in this application can be seen as follows. Without loss of generality, we first shift the problem so that ω_0 corresponds to the origin, $\omega_0 = 0$. The tangent plane at ω_0 is spanned by the columns of the $(q+1) \times q$ matrix V with elements $\partial\alpha_i(\omega)/\partial\omega_j$, $i = 1, \ldots, q+1, j = 1, \ldots, q$, where all derivatives are evaluated at ω_0. Since $\partial LD(\omega)/\partial\omega_j = 0$, V has the simple form $V^T = (I_q, 0)$ and thus the subspace orthogonal to the tangent plane is spanned by b_{q+1}, the basis vector for R^{q+1} with a 1 in the last position and zeros elsewhere. A normal section can now be seen to be that portion of the influence graph cut out by the plane spanned by the vectors b_{q+1} and $(l^T, 0)$. It follows that each lifted line $\alpha\{\omega(a)\}$ is a normal section.

The normal curvature C_l of the lifted line in the direction l can now be obtained by applying (9) to the plane curve $\rho^T(a) = (a, LD\{\omega(a)\})$ at $a = 0$. This plane curve is just the lifted line

$\alpha\{\omega(a)\}$ in rotated coordinates. The individual curvatures in the family of plane curves ρ obtained by letting l range over all unit vectors in R^q form the basis for our characterization of α.

Clearly $\dot{\rho}_1 = 1$ and $\dot{\rho}_2 = \ddot{\rho}_1 = 0$ and thus $C_l = |\ddot{\rho}_2| = |\ddot{LD}\{\omega(a)\}|$. Using the chain rule for differentiation C_l can be evaluated further,

$$C_l = 2\,|\,l^{\mathrm{T}} \ddot{F} l\,| \tag{11}$$

where $\|l\| = 1$ and $\ddot{F}$ is the $q \times q$ matrix with elements $\partial^2 L(\hat{\theta}_\omega)/\partial\omega_k \partial\omega_j, j, k = 1, \ldots, q$. This is the basic form for normal curvatures that will be used in this paper.

For (11) to be useful we should have a straightforward way to evaluate $\ddot{F}$. Using the chain rule for differentiation, $\ddot{F}$ can be expressed as

$$\ddot{F} = J^{\mathrm{T}} \ddot{L} J \tag{12}$$

where $-\ddot{L}$ is the observed information for the postulated model ($\omega = \omega_0$) and J is the $p \times q$ matrix with elements $\partial\hat{\theta}_{i\omega}/\partial\omega_j, i = 1, 2, \ldots, p, j = 1, 2, \ldots, q$, where $\hat{\theta}_{i\omega}$ is the ith component of $\hat{\theta}_\omega$. Next, to evaluate J we use the fact that

$$\left.\frac{\partial L(\theta \mid \omega)}{\partial \theta_j}\right|_{\theta = \hat{\theta}_\omega} = 0 \tag{13}$$

for $j = 1, 2, \ldots, p$ and all ω in Ω. Differentiating both sides of (13) with respect to ω and evaluating at ω_0, it follows that

$$J = -(\ddot{L})^{-1}\Delta \tag{14}$$

where Δ is the $p \times q$ matrix with elements

$$\Delta_{ij} = \frac{\partial^2 L(\theta \mid \omega)}{\partial\theta_i\,\partial\omega_j}$$

evaluated at $\theta = \hat{\theta}$ and $\omega = \omega_0$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$. Substituting (14) into (12) we obtain

$$\ddot{F} = \Delta^{\mathrm{T}} (\ddot{L})^{-1} \Delta \tag{15}$$

and therefore from (11)

$$C_l = 2\,|\,l^{\mathrm{T}} \Delta^{\mathrm{T}} (\ddot{L})^{-1} \Delta l\,|. \tag{16}$$

where $\|l\| = 1$. The individual components of (16) are often straightforward to obtain once the perturbation scheme has been defined.

There are several ways in which (16) might be used to study $\alpha(\omega)$ in practice. The extremes $C_{\max} = \max_l C_l$ and $C_{\min} = \min_l C_l$ are two possible options. Of course, $C_{\max}$ and $C_{\min}$ correspond to the maximum and minimum absolute eigenvalues of $\ddot{F}$ in (15). Another option is the average curvature $\bar{C}$ obtained by averaging (16) with respect to a uniform distribution on the surface of the unit sphere in q dimensions.

The eigenvector $l_{\max}$ associated with $C_{\max}$ can be used to set the directions in (10) which can then be used to construct plots, similar to those in Fig. 1, of the lifted line $\alpha\{\omega(a)\}$. This vector indicates how to perturb the postulated model to obtain the greatest local change in the likelihood displacement, and may be the most important diagnostic to come from this approach. When simultaneously perturbing all case-weights in linear regression, for example, suppose that the ith element of $l_{\max}$ is found to be relatively large. This indicates that perturbations in the weight ω_i of the ith case may lead to substantial changes in the results of the analysis and thus that ω_i is relatively influential. In such situations, it will, of course, be important to investigate the ith case to find the specific cause of the sensitivity. Similarly, the eigenvectors associated with intermediate eigenvalues can be used to investigate the behaviour of $\alpha(\omega)$ in directions corresponding to less extreme curvatures. These and other ideas will be illustrated in later sections.

Recall that the developments thus far have been guided by the goal of understanding the behaviour of an influence graph around ω_0. These developments do not apply when studying a graph around a different point ω^*. Additional complication arises because lifted lines generated by using $\omega(a) = \omega^* + al$ no longer corresponds to normal sections in general. However, curvatures of normal sections can still be obtained as follows.

Let $\dot{F}^{\mathrm{T}} = (\partial LD(\omega)/\partial\omega_j) = (-2\partial L(\hat{\theta}_\omega)/\partial\omega_j)$. Here and in the remainder of this Section all derivatives are evaluated at ω^* rather than ω_0 so that V, for example, now has the form $V^{\mathrm{T}} = (I_q, \dot{F})$. Further, let w_{jk} denote the $(q+1) \times 1$ vector with elements $\partial^2\alpha_i(\omega)/\partial\omega_j\partial\omega_k$, $i = 1, \ldots, q+1$. Then the velocity and acceleration vectors of the lifted line $\alpha\{\omega(a)\}$ at ω^* are respectively

$$\dot{\alpha}_l = Vl \tag{17}$$

and

$$\begin{aligned}\ddot{\alpha}_l &= \Sigma_j \Sigma_k w_{jk} l_j l_k \\ &= -2 l^{\mathrm{T}} \ddot{F} l b_{q+1}\end{aligned} \tag{18}$$

where $l = (l_j)$. The normal curvature C_l^* associated with the direction l can be written as

$$C_l^* = \| P_V' \ddot{\alpha}_l \| / \| \dot{\alpha}_l \|^2 \tag{19}$$

where P_V is the projection operator for the column space of V and $P_V' = I - P_V$. The projection in (19) is the essential step for obtaining the curvature of a normal section from a lifted line. The curvature C_l^* can be further evaluated by making use of the forms for the velocity and acceleration vectors given above. First, from the form of V, $\| \dot{\alpha}_l \|^2 = l^{\mathrm{T}} (I + \dot{F}\dot{F}^{\mathrm{T}})\, l$. Next, using (18) the numerator of C_l^* is just $2\,|\, l^{\mathrm{T}}\ddot{F}l \,|$ times the length of the last column of P_V'. Combining these results we obtain,

$$C_l^* = \frac{2\,|\, l^{\mathrm{T}} \ddot{F} l \,|}{(1 + \| \dot{F} \|^2)^{1/2}\, l^{\mathrm{T}} (I + \dot{F}\dot{F}^{\mathrm{T}})\, l} \tag{20}$$

Equation (20) in addition to the tangent planes at selected points ω^* may be of value in studying the global behaviour of an influence graph, but this idea requires further study. Recall that $\dot{F} = 0$ at ω_0 so that (20) reduces to (11) when $\omega^* = \omega_0$. In the remainder of this paper we focus on the behaviour of various influence graphs around ω_0. The developments leading to (19) provide a convenient way to obtain normal curvatures for other influence graphs.

3.2. *Other Influence Graphs*

In this Section we investigate the influence graphs $\alpha_s(\omega)$ and $\alpha'(\omega)$ defined in equations (7) and (8), respectively.

By replacing α and α' in the development that led to (20), and by using the chain rule for differentation, it can be verified that the velocity and acceleration vectors at ω_0 for α' are the same as those for α. Thus, α and α' have identical curvatures at ω_0, although the two influence graphs can differ considerably in global behaviour. Since we are primarily interested in assessing local influence around ω_0, α and the analogous graph α_s for subsets will be used in the remainder of this paper.

To develop the curvatures for $\alpha_s(\omega)$, we first note that the development leading to (20) is valid with $L(\hat{\theta}_\omega)$ replaced by $L[\gamma(\hat{\theta}_{1\omega})]$ where $\gamma^{\mathrm{T}} = (\hat{\theta}_{1\omega}^{\mathrm{T}}, g^{\mathrm{T}}(\hat{\theta}_{1\omega}))$ and g is defined following (7). It follows that (20) can be adapted for α_s by replacing $\dot{F}$ and $\ddot{F}$ by $\dot{G} = 2\partial L(\gamma)/\partial\omega$ and $\ddot{G} = \partial^2 L(\gamma)/\partial\omega^2$, respectively. Since $\dot{G} = 0$ at ω_0, (11) is also valid with $\ddot{F}$ replaced by $\ddot{G}$. To find a useful expression for $\ddot{G}$, we again use the chain rule and obtain

$$\ddot{G} = K^{\mathrm{T}} \ddot{L} K \tag{21}$$

where $\ddot{L}$ is as defined following (12) and K is the $p \times q$ matrix with elements $\partial\gamma_i(\hat{\theta}_{1\omega})/\partial\omega_j$, $i = 1, 2, \ldots, p, j = 1, 2, \ldots, q$, evaluated at ω_0.

We next need to find a useful representation for K. Let K_1 denote the $p_1 \times q$ matrix $\partial\hat{\theta}_{1\omega}/\partial\omega$ and let K_2 denote the $p_2 \times p_1$ matrix $\partial g(\theta_1)/\partial\theta_1$ evaluated at $\theta_1 = \hat{\theta}_1$. Then

$$K = \begin{pmatrix} I \\ K_2 \end{pmatrix} K_1. \tag{22}$$

The matrix K_1 consists of the first p_1 rows of J defined in (14). To evaluate K_2 we make use of the fact that

$$\frac{\partial}{\partial g_i} L[\theta_1, g(\theta_1)] = 0 \quad \text{for all} \quad \theta_1 \tag{23}$$

where g_i is the ith component of g, and the derivative is evaluated at $g = g(\theta_1), i = 1, 2, \ldots, p_2$. Differentiating (23) with respect to θ_1 and evaluating at $\hat{\theta}_1$ we find

$$K_2 = -(L_{22})^{-1} L_{21} \tag{24}$$

where L_{22} and L_{21} are determined from the partition

$$\ddot{L} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix}. \tag{25}$$

Finally, combining (11), (21) and (24) with the form of K_1 mentioned above, we obtain the normal curvature for subsets,

$$C_l(\theta_1) = 2 \mid l^{\mathrm{T}} \Delta^{\mathrm{T}} (\ddot{L}^{-1} - B_{22}) \Delta l \mid \tag{26}$$

where $\| l \| = 1$ and

$$B_{22} = \begin{pmatrix} 0 & 0 \\ 0 & L_{22}^{-1} \end{pmatrix}.$$

The techniques discussed at the end of Section 3.1 are applicable to (26), of course.

In the following Sections we describe several applications of these ideas. Our intent is to illustrate the range of possible use and to develop selected applications in some detail. If the proposed methodology is found to yield useful results in relatively well understood situations then we might expect it to yield similar results in more complicated settings where few, if any, methods are available. For this reason selected applications involving model (1) will be given special attention.

4. CASE-WEIGHTS IN NORMAL LINEAR REGRESSION

Let ω denote the $n \times 1$ vector of case-weights for the regression model (1) and again assume that σ^2 is known. The relevant part of the log-likelihood for the perturbed model is

$$L(\beta \mid \omega) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \omega_i (y_i - x_i^{\mathrm{T}} \beta)^2 \tag{27}$$

where ω_i and y_i are the ith components of ω and Y, respectively, and x_i^{T} is the ith row of X. Differentiating (27) with respect to β and ω, and evaluating at $\hat{\beta}$ and $\omega_0 = 1$, we find

$$\Delta = X^{\mathrm{T}} D(e)/\sigma^2 \tag{28}$$

where $e = (e_i)$ is the $n \times 1$ vector of ordinary residuals when $\omega = 1$ and $D(e) = \mathrm{diag}(e_1, \ldots, e_n)$. Since $\ddot{L}(\hat{\beta}) = -X^{\mathrm{T}}X/\sigma^2$,

$$C_l = 2 \mid l^{\mathrm{T}} \Delta^{\mathrm{T}} L^{-1} \Delta l \mid$$

$$= 2\, l^{\mathrm{T}} D(e)\, P_X D(e)\, l/\sigma^2 \tag{29}$$

where $P_X = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$, and $\| l \| = 1$. In what follows, P_M will be used to denote the projection operator for the column space of the matrix M, and $P_M' = I - P_M$.

When σ^2 is unknown, a similar calculation for $\theta^{\mathrm{T}} = (\beta^{\mathrm{T}}, \sigma^2)$ yields

$$\Delta = \begin{pmatrix} X^{\mathrm{T}}D(e)/\hat{\sigma}^2 \\ e_{sq}^{\mathrm{T}}/2\hat{\sigma}^4 \end{pmatrix} \tag{30}$$

where $\hat{\sigma}^2$ is the maximum likelihood estimator of σ^2 and e_{sq} is the $n \times 1$ vector with elements e_i^2. Since

$$\ddot{L}(\hat{\theta}) = -\begin{pmatrix} X^{\mathrm{T}}X/\hat{\sigma}^2 & 0 \\ 0 & n/2\hat{\sigma}^4 \end{pmatrix}$$

the analogous result for θ is

$$C_l = 2l^{\mathrm{T}}\,[D(e)\,P_X D(e) + e_{sq}e_{sq}^{\mathrm{T}}/2n\hat{\sigma}^2\,]\, l/\hat{\sigma}^2. \tag{31}$$

General analytic expressions for $l_{\max}$ are not known for (29) or (31).

If only β is of interest, the above results in combination with (26) shows that the curvature is given by (29) with σ^2 replaced with $\hat{\sigma}^2$. The three special situations described in Sections 4.1–4.3 should furnish some insight into the behaviour of the curvature and the interpretation of $l_{\max}$ when only β is of interest.

4.1. *Simple Random Samples*

For a simple random sample, $\ddot{F}$ has only one nonzero eigenvalue, $C_{\max}$, with corresponding eigenvector $l_{\max} = e/\| e \|$. Thus, the local changes in $\hat{\beta}$ will be zero when $\omega_0 = 1$ is perturbed in any direction that is orthogonal to e. In this simple situation, the maximum curvature is $C_{\max} = 2$ which is independent of the data. For this reason a curvature of 2 serves as a useful general reference, with curvatures much larger than 2 indicating notable local sensitivity. Perturbations of the case weights in a simple random sample can therefore never result in serious local changes, although global changes resulting from gross errors can be serious, of course. It is well-known that a gross error in a simple random sample is indicated by a relatively large element of e. An important general implication of this is that even if $C_{\max}$ is small an inspection of $l_{\max}$ may reveal the presence of gross errors. In other words, $C_{\max}$ is a useful indicator of serious local problems, but it may miss global concerns that are not manifest locally. Experience has shown that, regardless of the size of $C_{\max}$, an inspection of $l_{\max}$ is worthwhile. This idea will be illustrated further in later examples.

4.2. *Individual Cases*

The curvature for the influence graph obtained by modifying the weight attached to a single case, say the ith, is

$$C = 2e_i^2 h_{ii}/\hat{\sigma}^2 = 2p(1 - h_{ii})^2 D_i \tag{32}$$

where h_{ij} is the (i,j)th element of P_X and D_i is given by (2). Form (32) was used to construct Fig. 1. For case A, $h_{ii} = 0.95$ and for case B, $h_{ii} = 0.01$. Thus, case A corresponds to a high leverage point with a relatively small residual while B corresponds to a low leverage point with a large residual. For both cases $pD_i = 3.5$ so that the curvature for case A is about $C = 0.02$ while $C = 6.9$ for case B. In this example, perturbing the weight attached to case B would lead to changes in $\hat{\beta}$ that are uniformly larger than those obtained when the weight attached to case A is similarly modified, although the two cases would appear from Fig. 1 equally influential when deleted.

Generally, high leverage points with relatively small residuals are influential only when considering the possibility of a gross error so that the case contains no relevant information about β. In the example of Fig. 1, the variance of case A might be set at ten times the variances of the remaining cases without any serious consequences while a similar modification of the variance of case B could lead to substantial changes.

As mentioned previously, the curvature given in (32) and the analogous case-deletion diagnostics are simply summaries of different characteristics of the influence graph obtained by modifying a single case-weight: C_l measures the influence of local changes in the case weight, while D_i measures global changes. Clearly, a case that is locally influential must be globally influential, but the reverse need not be true. Both types of information can be useful, depending on the concerns of the investigator. However, when considering many cases simultaneously, it will be easier to use (29) to characterize the local behaviour of an influence graph than to use multiple-case-deletion diagnostics to characterize global behaviour by using the corners of Ω.

4.3. *Individual Coefficients*

Individual coefficients in a linear model are often of special interest. In such situations, (26) can be used to assess influence on a selected coefficient of simultaneously modifying all case weights. To develop this result, rearrange the columns of $X = (X_1, X_2)$ so that the first column X_1 corresponds to the coefficient β_1 of interest, and let r denote the residuals from the regression of X_1 on X_2,

$$r = (r_i) = P'_{X_2} X_1.$$

An evaluation of (26) in this situation requires Δ as given in (30), $\ddot{L}(\hat{\theta})$ as given near (31), and

$$L_{22} = \begin{pmatrix} X_2^{\mathrm{T}} X_2 & 0 \\ 0 & n/2\hat{\sigma}^4 \end{pmatrix}$$

With these results and a little simplification, $C_l(\beta_1)$ can be written in the intermediate form

$$C_l(\beta_1) = 2\,|\, l^{\mathrm{T}} D(e)\, X[(X^{\mathrm{T}}X)^{-1} - A_{22}]\, X^{\mathrm{T}} D(e)\, l\,|/\hat{\sigma}^2$$

where

$$A_{22} = \begin{pmatrix} 0 & 0 \\ 0 & (X_2^{\mathrm{T}} X_2)^{-1} \end{pmatrix}$$

Next, using standard results to obtain the inverse of the partitioned form of $(X^{\mathrm{T}}X)^{-1}$ and simplifying,

$$C_l(\beta_1) = 2\,|\, l^{\mathrm{T}} D(e)\, r r^{\mathrm{T}} D(e)\, l\,|/\|\hat{r}\|^2 \hat{\sigma}^2. \tag{33}$$

It follows that

$$C_{\max}(\beta_1) = 2n\, \Sigma\, r_j^2 e_j^2 / \{\Sigma\, r_j^2\ \Sigma\, e_j^2\}. \tag{34}$$

This curvature is bounded above by $2n$ and will tend to be large when corresponding elements of r and e are large.

The maximum curvature occurs in the direction

$$l_{\max} = D(e)\, r = (e_j r_j). \tag{35}$$

Relatively large or small elements of $l_{\max}$ correspond to cases that have $|r_j|$ and $|e_j|$ large simultaneously. Thus, an inspection of $l_{\max}$ will identify cases, if any, that are contributing substantially to $C_{\max}$.

The recognition of the importance of cases with $|r_j e_j|$ relatively large is not new. Such cases will stand out in a standard added variable plot of the residuals from the regression of Y on X_2

versus r, whose specific form is $P'_{X_2} Y = \hat{\beta}r + e$ versus r. The term $\hat{\beta}r$ produces the linear trend in an added variable plot and e produces the scatter, (see, for example, Cook and Weisberg, 1982). For illustration, Fig. 2 gives a schematic representation of a de-trended added variable plot of e versus r. The sloping lines are not relevant to the present discussion; these will be discussed in Section 5.1. The central elliptical portion of Fig. 2 is assumed to contain the bulk of the data. The points designated by A to H represent the kinds of cases that generally attract special attention, but not all of these cases will be influential for $\hat{\beta}_1$ under the perturbation scheme described above. The estimate $\hat{\beta}_1$ will be most sensitive to perturbations of the weights attached to cases like A, C, E and G, since $|r_j e_j|$ will be relatively large for these cases. On the other hand, $\hat{\beta}_1$ will be relatively insensitive to the weights for the remaining cases, B, D, F and H, since for these cases $|r_j e_j|$ is small.

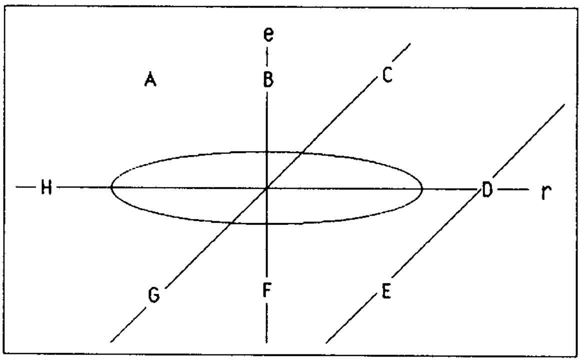

Fig. 2. Schematic representation of a detrended added variable plot of e versus r.

To illustrate why cases like B, D, F and H are not influential under case-weight perturbations, consider simple regression through the origin. The results (34) and (35) apply in this situation with r interpreted as the vector of values of the single explanatory variable and β_1 interpreted as the corresponding regression coefficient. Clearly, $\hat{\beta}_1$ will be independent of the weight attached to any observation at the origin. Observations at the origin have $r_j = 0$ and thus will not and should not stand out regardless of the size of the associated residual e_j. These comments may seem to present a little dilemma since experience with added variable plots clearly indicates that cases like B, D, F and H can exert considerable influence on $\hat{\beta}_1$. A resolution of this will be suggested in Section 5.

4.4. *Geese Data*

As a first numerical illustration, consider the geese data for observer 1 as reported in Weisberg (1980, p. 95). The data consist of observations on y = true flock size as obtained by count from aerial photographs and x = visually estimated flock size for a sample of $n = 45$ flocks of snow geese. These data were collected in an effort to determine how well flock sizes could be visually estimated during a census of the population, and were instrumental in the decision to base the actual census counts on aerial photographs. Part of the rationale for this decision can be seen from Fig. 3 which is a plot y versus x; the different markings for the points are discussed below. We use these data in combination with a simple linear regression model to illustrate the behaviour of l_{max} in the presence of heteroscedasticity.

The plot of the data given in Fig. 3 shows strong evidence of heteroscedasticity. From this it seems reasonable to expect that the coefficients of the fitted line will be sensitive to minor case weight modifications, particularly the weights corresponding to cases with large observed counts. This is confirmed by the curvature $C_{max} = 14.37$ computed from (29) with $\sigma^2 = \hat{\sigma}^2$ so that only

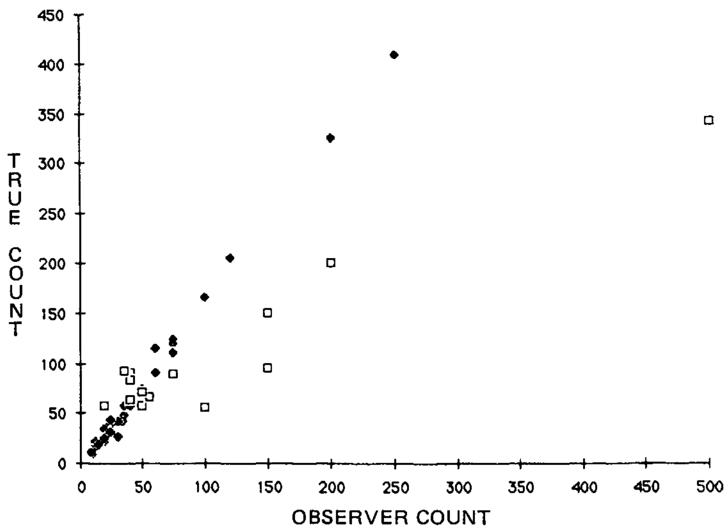

Fig. 3. Scatter plot of the geese data; the different point marks correspond to the signs of the elements of l_{max}.

coefficients are of interest. Further, to get maximum movement from the fitted line, we might expect to increase the weights of cases with large positive residuals, while decreasing the weights of cases with large negative residuals, or vice versa. This is just what an inspection of l_{max} indicates; the different markings for the points in Fig. 3 correspond to the signs of the elements of l_{max}.

Figure 4 gives a plot of l_{max} versus x. Clearly, l_{max} is responding to the essential heteroscedasticity in the data. Generally, such inspections of l_{max} can provide useful diagnostic information that may be used to guide subsequent analysis and future experimentation.

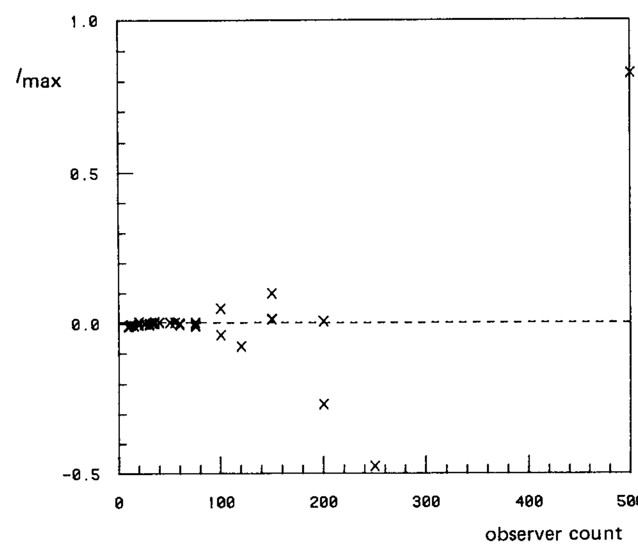

Fig. 4. Geese data: Scatter plot of l_{max} versus observer count.

For these data there is a strong relationship between the ordinary residuals e from the simple linear regression of y on x and the elements of l_{max}. For example, a plot (not shown) of the absolute elements of l_{max} versus e shows a very strong U-shaped pattern. Cases with relatively large absolute residuals tend to be emphasized by l_{max}, while cases with small absolute residuals are de-emphasized. In this way l_{max} directs attention to cases where specification of weights is most important. In general, the relationship between l_{max} and e is not monotonic since it depends on P_X as shown in (29).

4.5. *Rat Data*

For a second numerical illustration, we use the rat data and corresponding model as reported in Weisberg (1980, p.110–113). The data consist of 19 cases and the regression model contains 4 explanatory variables, x_0 = constant, x_1 = body weight, x_2 = liver weight and x_3 = relative dose. The response variable is y = percentage of dose in the liver. Weisberg found that case 3 is influential, $D_3 = 0.93$ and $h_{33} = 0.85$, and that the relative dose for this case is apparently anomolous.

Again considering only coefficients, the maximum curvature for the rat data is $C_{max} = 3.58$. This curvature does not indicate extreme local sensitivity, but further investigation is certainly in order. A scatter plot of the absolute values of the elements of l_{max} versus x_2 is given in Fig. 5.

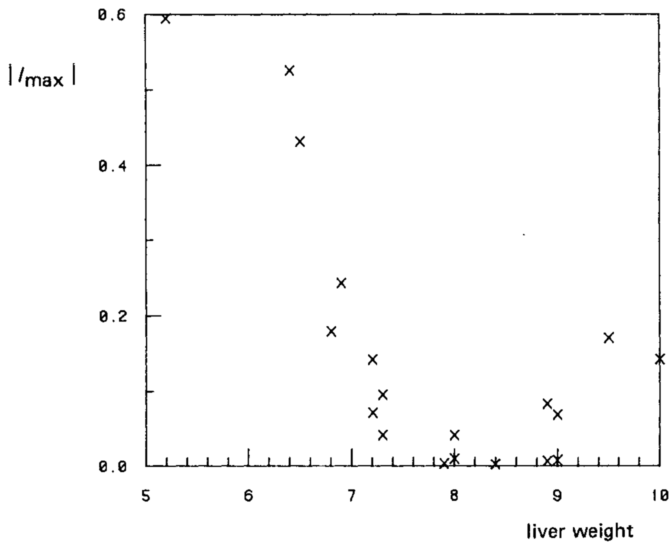

Fig. 5. Rat Data: Scatter plot of the absolute elements of l_{max} versus liver weight.

The pattern in this plot suggests heteroscedasticity. Since analogous plots involving x_1 and x_3 do not exhibit a clear pattern, any further analysis of these data should probably give consideration to the possibility of heteroscedasticity as a function of x_2. For example, the pattern in Fig. 5 indicates that the error variances may be a function of $|x_2 - m|$, where m is a measure of central tendency for x_2. The score test (Cook and Weisberg, 1983b) supports this indication.

Further information might be obtained by looking in directions that correspond to smaller nonzero eigenvalues of $\ddot{F}$. Since P_X has rank p there will be at most p such directions. This serves as a reminder that the sensitivity of an analysis to case-weight perturbations can be expected to increase with p for fixed n.

5. EXPLANATORY VARIABLES IN NORMAL LINEAR REGRESSION

It is well known that minor perturbations of the explanatory variables in linear regression can seriously influence the results of a least squares analysis when collinearity is present. Such results may also be influenced by a few isolated errors in the values of the explanatory variables or by a few values that are widely separated from the remaining data. We are not thinking of the errors-in-variables problem where enough information may be available to allow X to be modelled stochastically. For convenience we again assume that σ^2 is known. The following results can be easily adapted for the situation in which σ^2 is unknown and only β is of interest by replacing σ^2 with $\hat{\sigma}^2$.

Let s_j, $j = 1, \ldots, p$, denote scale factors to account for the different measurement units associated with the columns of X. Then the perturbed log-likelihood $L(\beta \mid \omega)$ is constructed from (1) with X replaced by

$$X_\omega = X + WS \tag{36}$$

where $W = (\omega_{ij})$ is an $n \times p$ matrix of perturbations and $S = \text{diag}(s_1, \ldots, s_p)$. The diagonal element s_j of S converts the generic perturbation ω_{ij} to the appropriate size and units so that $\omega_{ij} s_j$ is compatible with the ijth element of X. Next partition the $p \times np$ matrix Δ as $\Delta = (\Delta_1, \ldots, \Delta_p)$ where the elements of the $p \times n$ matrix Δ_k are $\partial^2 L(\beta \mid \omega)/\partial\beta_i\, \partial\omega_{jk}$, $i = 1, 2, \ldots, p, j = 1, 2, \ldots, n$. Then

$$\Delta_k = s_k\,(d_k e^{\mathrm{T}} - \hat{\beta}_k X^{\mathrm{T}})/\sigma^2 \tag{37}$$

where d_k is a $p \times 1$ vector with a 1 in the kth position and zeros elsewhere.

In this application, $\ddot{F}$ is a potentially large $np \times np$ matrix and determining its eigenvalues may be an unpleasant task. However, the eigenvalues of $\ddot{F} = \Delta^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\Delta/\sigma^2$ can be determined by replacing $(X^{\mathrm{T}}X)^{-1}$ with $(X^{\mathrm{T}}X)^{-½}(X^{\mathrm{T}}X)^{-½}$ and using the fact that the nonzero eigenvalues of $A^{\mathrm{T}}A$ are the same as those of AA^{T} which will be a manageable $p \times p$ matrix in this situation. Using this method, it can be shown that the nonzero eigenvalues of $\ddot{F}$ are

$$e^{\mathrm{T}}e\delta_i/\sigma^2 + \sum_j \hat{\beta}_j^2 s_j^2/\sigma^2 \tag{38}$$

where δ_i is the ith eigenvalue of $S(X^{\mathrm{T}}X)^{-1}S$, $i = 1, 2, \ldots, p$. Thus,

$$C_{\max} = 2e^{\mathrm{T}}e\,\delta_{\max}/\sigma^2 + 2\sum_j \hat{\beta}_j^2 s_j^2/\sigma^2. \tag{39}$$

Except in the special situation discussed below, an analytic form for $l_{\max}$ is unknown.

5.1. *Individual Columns*

The above results can be applied to situations in which less than p explanatory variables are perturbed by setting $s_j = 0$ for the unperturbed variables. In particular, when only the first column of X is perturbed, $s_j = 0$ for $j \neq 1$ and $\ddot{F} = \Delta_1^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\Delta_1\sigma^2$ where Δ_1 is given by (37) with $k = 1$. Using the identity

$$d_1^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}d_1 = \| r \|^{-2} \tag{40}$$

$C_{\max}$ can be obtained from (39) or directly from $\ddot{F}$,

$$C_{\max} = 2s_1^2\,(\| e \|^2\, \| r \|^{-2} + \hat{\beta}_1^2)/\sigma^2 \tag{41}$$

where r is defined near (33). Clearly, $C_{\max}$ depends on the value chosen for s_1. In practice it is convenient to leave s_1 unspecified until the remaining part of a $C_{\max}$ has been determined. The magnitude of s_1 necessary for a large curvature can then be determined and compared to prior knowledge on the size of potential errors.

Next, using the relation

$$X(X^{\mathrm{T}}X)^{-1}d_1 = r/\| r \|^2 \tag{42}$$

and $\ddot{F}$ as given above, it can be verified that the direction of maximum curvature is

$$l_{max} = e - \hat{\beta}_1 r \tag{43}$$

A particular element x_{i1} of X_1 will tend to be influential if the corresponding element $(e_i - \hat{\beta} r_i)$ of l_{max} is relatively large. As in Section 4.3, l_{max} depends on the building blocks of a detrended added variable plot for $\hat{\beta}_1$ and it would be convenient if this plot could be used to assess the relative sizes of the elements of l_{max}. A single plot could then be used to assess the effects of two perturbation schemes-(case-weight perturbations as described in Section 4.3 and explanatory variable perturbations as described here)-in addition to furnishing other useful diagnostic information; see, for example, Chambers, *et al.* 1983.

The ith plotted point in a detrended added variable plot is (r_i, e_i). If we pass a line with slope $\hat{\beta}_1$ through this point, the intercept of this line will be the ith element of l_{max}, $(e_i - \hat{\beta} r_i)$. In practice it is often sufficient to visualize a single line with slope $\hat{\beta}_1$ on the plot, and then mentally shift the line to assess the absolute sizes of the intercepts for each plotted point. Of course, special plots may be necessary if this procedure does not yield a clear interpretation.

The two diagonal lines in Fig. 2 were constructed under the assumption that $\hat{\beta}_1 = 1$ so that $|e_i - \hat{\beta} r_i| = 0$ for cases C and G while $|e_i - \hat{\beta} r_i|$ is relatively large for the remaining outlying cases. These ideas should resolve the dilemma mentioned in Section 4.3: Depending on $\hat{\beta}_1$, cases like B, D, F and H will be important under perturbations of an explanatory variable, but not under perturbations of the case-weights.

5.2. *Rat Data Again*

For a first numerical illustration we again use the rat data. The perturbation scheme is characterized by $S = \text{diag}(s_0, s_1, s_2, s_3) = \text{diag}(0, 1, 0, s_3)$, with $s_3 = 0.01\ (0.01)\ 0.04$. Thus only body weight and relative dose are to be perturbed, although for a thorough analysis perturbations in liver weight should probably be considered also. The value $s_1 = 1$ indicates that perturbations of body weight will be on the order of 1 gram. Similar interpretations hold for the various values of s_3. For $s_3 = 0.01\ (0.01)\ 0.04$ the maximum curvatures obtained by setting $\sigma^2 = \hat{\sigma}^2$ in (39), are $C_{max} = 2.8$, 9.4, 20.5 and 36.0, respectively. The curvature for $s_3 = 0.01$ is relatively small while the curvatures for the remaining s_3s are cause for concern. At the very least, further investigation is indicated.

For example, a plot of $LD\{\omega(a)\}$ in the direction of the eigenvector corresponding to C_{max} is given in Fig. 6 for $s_3 = 0.03$. Interestingly, the largest element of l_{max} always corresponds to the relative dose for case 3 which is the anomalous value Weisberg (1980) identified. The scale on the x-axis in Fig. 6 is the amount that the relative dose for case 3 is perturbed in units of s_3. Thus, for example, $a = 0.5$ indicates $as_3 = 0.015$ was added to the relative dose for case 3 while the relative doses for the remaining cases were perturbed by an amount less than 0.015 in absolute value. Clearly, the influence of perturbations for $s_3 = 0.03$ is very strong. In particular, the value of LD at $a = -0.5$ shows that $\hat{\beta}_\omega$ will be moved outside of a 95 per cent confidence region for $\hat{\beta}$ when perturbing each element of X by an amount that is not greater than $\frac{1}{2}$ of the respective s_js. The nonmonotonic behaviour in Fig. 6 arises since the lifted line $\alpha\{\omega(a)\}$ need not correspond to a path of monotonic increase.

When perturbing explanatory variables to detect a few isolated errors or outlying values that seriously influence the coefficient estimates, l_{max} is of primary importance. If we perturb only relative dose and concentrate attention on the corresponding coefficient, information on the relative sizes of the elements of l_{max} can be obtained from the de-trended added variable plot of e versus r, as discussed in Section 5.1. This plot is shown in Fig. 7. The common slope of the two diagonal lines in Fig. 7 is $\hat{\beta}_3 = 4.18$. The intercepts of these lines serve as bounds on the intercepts obtained by passing lines with common slope $\hat{\beta}_3$ through each of the points on the plot. In view of the discussion given in Section 5.1, we see immediately that $\hat{\beta}_3$ will be most sensitive to perturbations in the dose for case 3 since the absolute intercept of the line passing through (r_3, e_3) is the largest.

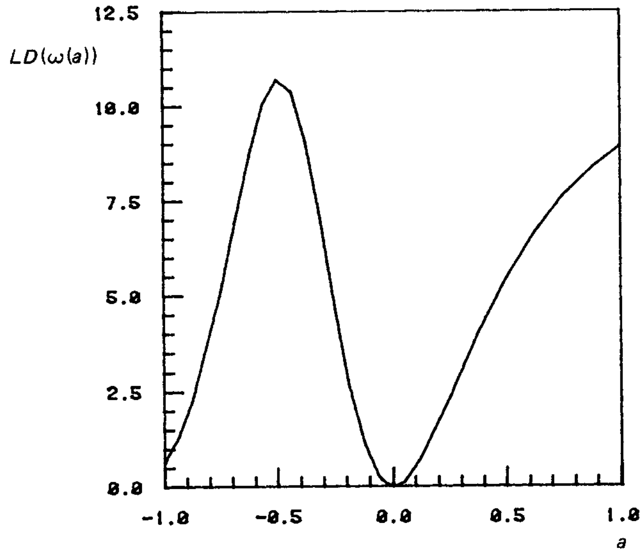

Fig. 6. Plot of the likelihood displacement *LD* in the direction of maximum curvature for the rat data.

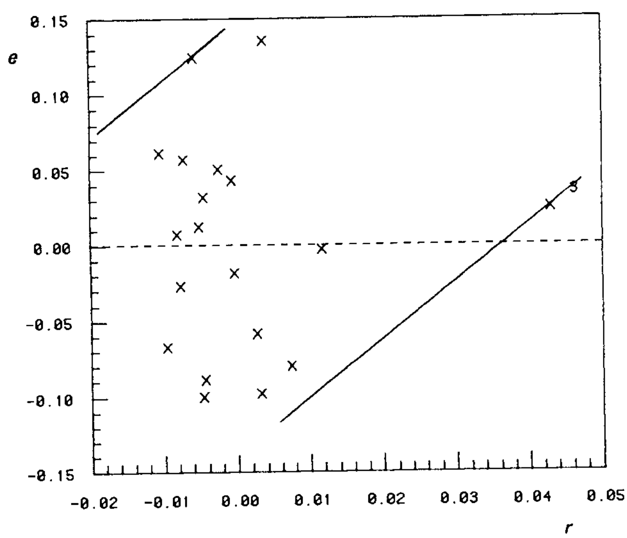

Fig. 7. Rat data: Added variable plot of *e* versus *r* for X_3 = relative does. The diagonal line has slope $\hat{\beta}_3$ = 4.18.

5.3. *Longley Data*

For a second numerical illustration we use the perturbation scheme for the Longley data described in Weisberg (1980, p. 70-72). For this setup, which consists essentially of using the s_ks to represent round-off errors in the last digit of the explanatory variables, evaluating (39) with $\sigma^2 = \hat{\sigma}^2$ gives $C_{max} = 0.18$. Weisberg found that only one significant digit in the $\hat{\beta}$s would be stable under his perturbation scheme. However, the small maximum curvature indicates that this instability does not reflect important local changes in the estimates when judged against the log-likelihood. This illustrates that seemingly substantial changes in estimates may, in fact, be inconsequential if the log-likelihood is sufficiently flat.

5.4. *Acetylene Data*

Finally, we consider the acetylene data as reported by Snee and Marquardt (1975). There are 16 observations and the model includes all linear, cross-product and quadratic terms in X_1 = reactor temperature, X_2 = mole ratio, and X_3 = contact time. The addition of a constant X_0 gives a total of 10 coefficients in the model. For these data, perturbing a single column of X may not be appropriate since any anomalies associated with a base variable will also be present in the associated quadratic and cross-product terms. The primary intent of this example then is to indicate briefly how the perturbation scheme described in (36) can be modified to handle polynomial models.

For illustration, suppose that we choose to perturb only X_2. Then

$$X_\omega = (X_0, X_1, X_{2\omega}, X_3, X_1X_{2\omega}, X_1X_3, X_{2\omega}X_3, X_1^2, X_{2\omega}^2, X_3^2)$$

where $X_{2\omega} = X_2 + \omega s_2$. The maximum curvature for this scheme is $C_{max} = 17s_2^2$, and an index plot of l_{max} is given in Fig. 8. Clearly, the value of X_2 for case 13 deserves careful consideration

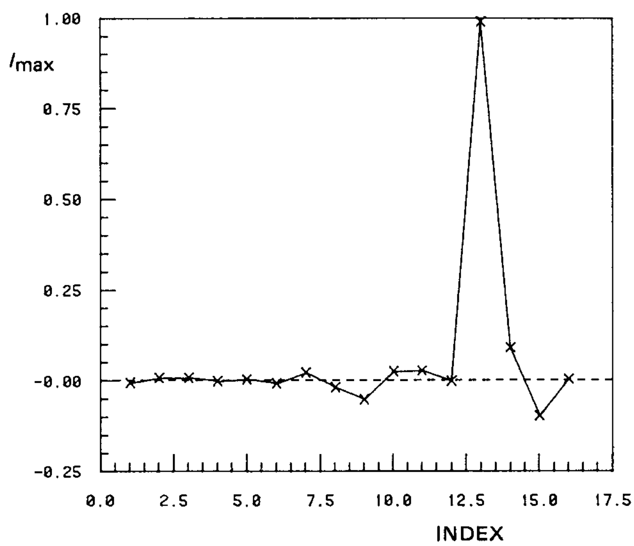

Fig. 8. Acetylene Data: Index plot for l_{max}.

regardless of the size of s_2^2. Similar perturbations of X_1 and X_3 do not indicate that case 13 is of special interest, but further investigation using case-deletion diagnostics does support this conclusion. Case-deletion diagnostics, however, can only direct attention to an entire case, while perturbing a base variable can isolate a particular component of a case.

6. EXTENSIONS

6.1. *Case-Weights in Exponential Families*

In this and the following Section, we indicate how previous results can be extended beyond normal linear models.

Let y_i denote an observation from a regular curved exponential family with minimal representaition

$$f(y \mid \theta) = \exp\{y\eta_i(\theta) - \psi_i(\eta_i(\theta))\}.$$

For a series $y_1, \ldots, y_n$ of independent observations the log-likelihood is therefore

$$L(\theta) = \sum_i (y_i\eta_i - \psi_i(\eta_i)). \tag{44}$$

Often, the log-likelihood obtained by attaching a weight ω_i to the ith case can be written simply as

$$L(\theta \mid \omega) = \sum_i \omega_i(y_i\eta_i - \psi_i(\eta_i)). \tag{45}$$

Pregibon (1981) used a likelihood of this form to derive various diagnostics for logistic regression.

Let

$$\eta = (\eta_i),$$

$$\dot{\eta} = \partial\eta/\partial\theta \qquad (n \times p)$$

$$\ddot{\eta}_i = \partial^2\eta_i(\theta)/\partial\theta^2 \qquad (p \times p)$$

and

$$\ddot{\psi} = \mathrm{diag}(\partial^2\psi_i/\partial\eta_i^2) \qquad (n \times n)$$

where all derivatives are evaluated at $\hat{\theta}$, the maximum likelihood estimator of θ. Applying the results of Section 3.1,

$$\ddot{F} = D(u)\,\dot{\eta}\,[\sum_i u_i\ddot{\eta}_i - \dot{\eta}^{\mathrm{T}}\,\ddot{\psi}\,\dot{\eta}]^{-1}\,\dot{\eta}^{\mathrm{T}}D(u) \tag{46}$$

where $D(u)$ is an $n \times n$ diagonal matrix with the score residuals

$$u_i = (y_i - \partial\psi_i/\partial\eta_i) \tag{47}$$

as the diagonal entries.

Many generalized linear models are special cases of (44) with $\eta_i(\theta) = K(x_i^{\mathrm{T}}\theta)$ where K is the link function. Further,

$$\dot{\eta} = \mathrm{diag}\,(\dot{K}_i)\,X$$

and

$$\ddot{\eta}_i = \ddot{K}_i\,x_ix_i^{\mathrm{T}},$$

where $\dot{K}_i$ and $\ddot{K}_i$ are the first and second derivatives of K evaluated at $x_i^{\mathrm{T}}\hat{\theta}$, respectively. In particular, $\ddot{\eta} = 0$ when the canonical link is used.

With a few simple modifications, the results of Section 4.3 can be applied in logistic regression where $y_i \sim$ binomial (m_i, p_i) and $\log\{p_i/(1-p_i)\} = x_i^{\mathrm{T}}\theta$. Let $w_i = [m_i\hat{p}_i\,(1-\hat{p}_i)]$, $i = 1, \ldots, n$, let $D(w) = \mathrm{diag}(w_i)$, and let $D(\chi)$ denote a diagonal matrix with the components of χ^2, $\chi_i = (y_i - m_i\hat{p}_i)/w_i^{1/2}$, as the diagonal entries. Then

$$\ddot{F} = D(\chi)\,P_W D(\chi)$$

where $W = D(w)^{1/2}X$. Similarly, when θ_1 is of special interest, the direction of maximum curvature is

$$l_{\max} = D(\chi)r_w,$$

where r_w is the residual vector from the regression of the first column of W on the remaining columns. Figure 9 gives a plot of χ_i versus the elements of r_w for the coefficient of log(volume) from Finney's data as reported in Pregibon (1981); the outstanding cases are the same as those identified by Pregibon.

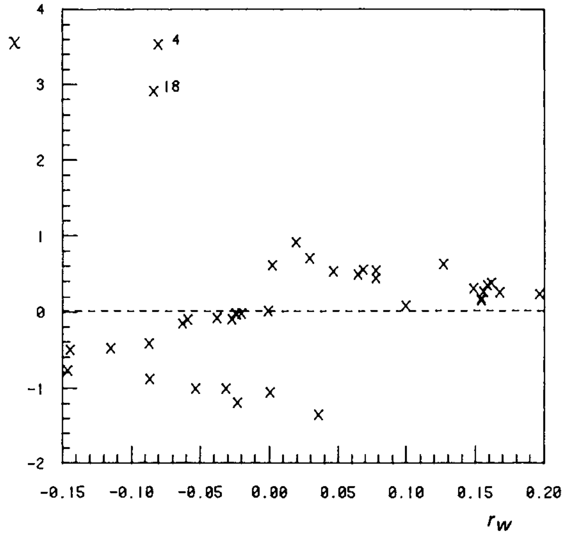

Fig. 9. Finney's Data: Added variable plot of χ_i versus r_w.

6.2. *Explanatory Variables in Generalized Linear Models*

Consider the log-likelihood (44) with $\eta_i = K(x_i^{\mathrm{T}}\theta)$. The log-likelihood $L(\theta \mid \omega)$ obtained after the explanatory variables have been perturbed by an amount ω can be constructed by replacing x_i^{T} with $x_{i\omega}^{\mathrm{T}}$, the ith row of X_ω defined in (36). From this it can be verified that Δ has the same structure as described in Section 5 and that

$$\Delta_k = s_k \{d_k u^{\mathrm{T}} \mathrm{diag}(\dot{K}_i) + \hat{\theta}_k X^{\mathrm{T}} \mathrm{diag}(u_i \ddot{K}_i - \ddot{\psi}_i \dot{K}_i^2)\} \tag{48}$$

where d_k is defined following (37) and $u = (u_i)$. Further, the observed information matrix is

$$-\ddot{L} = -X^{\mathrm{T}} \mathrm{diag}(u_i \ddot{K}_i - \ddot{\psi}_i \dot{K}_i^2)\, X. \tag{49}$$

For a concrete illustration we use the leukemia data as reported in Cook and Weisberg (1982, p. 179). Here, a patient's survival time in weeks y_i, $i = 1, 2, \ldots, 17$, is assumed to follow a one parameter exponential distribution with mean $\exp\{\theta_1 + \theta_2 x_i\}$ where $x_i = \log_{10}(WBC_i)$ and WBC_i is the white blood cell count for the ith patient.

The log-likelihood for the original data is of the form given in (44) with

$$\eta_i = K(\theta_1 + \theta_2 x_i) = -\exp[-(\theta_1 + \theta_2 x_i)]$$

and $\psi_i(\eta_i) = -\log(-\eta_i)$. From this it follows that

$$\dot{K}_i = \exp\,[-(\hat{\theta}_1 + \hat{\theta}_2 x_i)] = (\hat{E}Y_i)^{-1}$$

$$\ddot{K}_i = -\dot{K}_i$$

$$\dot{\psi}_i = -\eta_i^{-1} = \exp\,[\hat{\theta}_1 + \hat{\theta}_2 x_i] = \dot{K}_i^{-1}$$

$$\ddot{\psi}_i = \eta_i^{-2} = \exp\,[2(\hat{\theta}_1 + \hat{\theta}_2 x_i)] = \hat{\mathrm{var}}\,(Y_i)$$

$$u_i = y_i - \exp\,[\hat{\theta}_1 + \hat{\theta}_2 x_i]$$

and thus that

$$u_i \ddot{K}_i - \ddot{\psi}_i \dot{K}_i^2 = -y_i \dot{K}_i.$$

These calculations along with (48) and (49) can now be used to construct $\ddot{F}$ as given in (15).

To assess the influence of potential anomalies in *WBC*, we perturb $x = \log_{10}(WBC)$ rather than *WBC* itself. This implies that errors associated with *WBC* are multiplicative rather than additive, which seems to be a reasonable implication.

The maximum curvature for this perturbation scheme is $C_{max} = 17.014 s_x^2$ where s_x is the scale of the error associated with x. Clearly, s_x^2 must be substantial for the local influence to be large. An index plot of l_{max} is given in Fig. 10. Although the curvature is small, an inspection of the direction of maximum curvature does direct attention to case 17 which is the case that Cook and Weisberg (1982, p. 185) identified as influential by using case-deletion diagnostics.

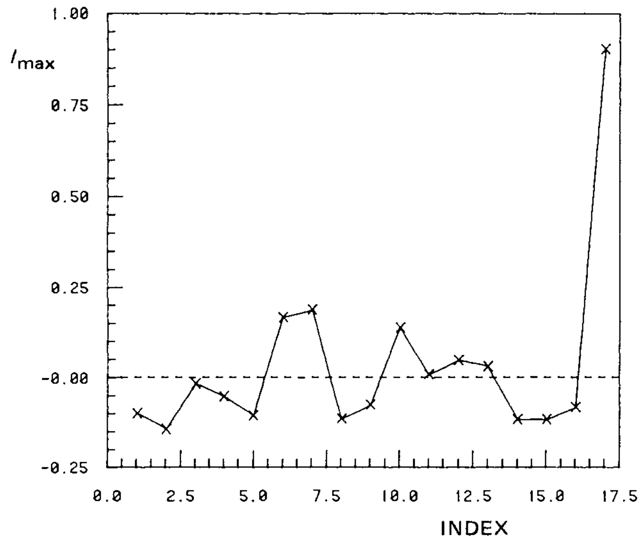

Fig. 10. Leukaemia Data: Index plot of l_{max}.

7. DISCUSSION

Statistical conclusions can be viewed as the end result of a synthesis of the relevant information provided by the observed data and the prior information provided by the model which is usually a plausible, but necessarily imprecise, description of the actual process that generated the data. The developments in the proceeding Sections are based on the informal notion that important conclusions should not depend critically on the hypothesized model or unusual aspects of the data. If our conclusions do depend critically on the model or the data, there is surely cause for concern and the knowledge of such dependence must become a part of the conclusions. Otherwise, our ignorance of the precise process that generated the data should do no harm.

An obvious way to see if perturbations of the model influence key results of the analysis is to compare the results derived from the original and perturbed models. The influence graphs introduced in Section 2 are simply devices to facilitate such comparisons when the behaviour of the parameter estimates is of interest. These graphs are not designed to display all of the possible consequences of an induced perturbation. For example, they may not respond to situations in which the shapes of the original and perturbed log likelihoods are substantially different, while the corresponding parameter estimates are nearly identical. Nevertheless, as illustrated in the discussion of linear regression, the proposed graphs do provide important diagnostic information and a relatively simple, unified approach for handling a variety of problems.

For a complete understanding of the influence of a particular perturbation scheme, it is probably necessary to know the full behaviour of an influence graph over Ω. The central methodology of this paper seems to be a reasonable way of characterizing the local behaviour of an influence graph around ω_0. The maximum curvature $C_{\max}$ is a useful indicator of extreme local behaviour, and the plot of the corresponding lifted line provides a straightforward way to confirm such indications. Experience has shown, however, that the direction of maximum curvature $l_{\max}$ contains the most important diagnostic information. This approach leads to new interpretations for outlying points in added variable plots from linear regression and provides a way to extend these plots to more complicated situations, including generalized linear models.

In some problems such as case-weight perturbations, it may be worthwhile to supplement a local analysis of an influence graph by studying its behaviour at the boundaries of Ω. In other problems it may not be possible to specify Ω precisely and in such situations a local analysis is most natural. This happens, for example, when perturbing explanatory variables.

The proposed methodology was developed in a likelihood framework, but the basic ideas are equally applicable in Bayesian analyses. For example, the likelihood displacement might be replaced by the Kullback–Leibler divergence between the original and perturbed predictive densities and ω might be used to assess the sensitivity of the analysis to perturbations in the prior parameters. Geisser (1985) briefly describes several related ideas that might be used in this context.

There is an apparent danger associated with the use of diagnostic methods: can we too often end up pursuing meaningless shadows in the data and thereby neglecting important substantive issues that should be our primary concern? This danger may seem particularly real for the proposed methodology since the diagnostic statistics are obtained by maximizing over a potentially large number of dimensions. Figure 5 and the associated discussion may be a good example of how we can be misled. There seems to be no sure answer to such concerns, but there are a few ways in which they may be alleviated. First, simulation methods can be used to aid the interpretation of plots such as that in Fig. 5; Atkinson's (1981) envelopes is one method that might be adaptable. Second, in all situations where explicit expressions for $C_{\max}$ and $l_{\max}$ have been obtained, the proposed methodology agrees well with standard procedures. And lastly, few analyses are intended to yield definitive answers. Recalling the comments by Box (1980), results from one analysis will direct our attention in future studies, with persistent results leading to scientific learning.

Finally, this work has centered on the identification of influential aspects of the model and this leads naturally to the identification of relevant and perhaps unexpected characteristics of the data. Inevitably, we will be faced with the issue of accommodation, or how to proceed once potential problems have been identified. Accommodation is important, of course, but specific universal recommendations are elusive since appropriate procedures depend strongly on context. In any event, identification must necessarily precede accommodation.

ACKNOWLEDGEMENTS

I am indebted to the referees and several colleagues, particularly Sanford Weisberg and Morris DeGroot, for many helpful suggestions.

This research was supported by the United States Army under contract DAAG29-80-C-0041.

REFERENCES

Andrews, D. F. and Pregibon, D. (1978) Finding the outliers that matter. *J. R. Statist. Soc.* B, **40**, 85–93.

Atkinson, A. C. (1981) Robustness, transformations and two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13–20.

——— (1982) Regression diagnostics, transformations and constructed variables. *J. R. Statist. Soc.* B, **44**, 1–36.

Barnard, G. A. (1980) Discussion of Professor Box's paper. *J. R. Statist. Soc.* A, **143**, 404–406.

Bates, D. M. and Watts, D. G. (1980) Relative curvature measures of non-linearity. *J. R. Statist. Soc.* B, 42, 1–25.

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

——— (1984) Demeaning conditioning diagnostics through centering. *Amer. Statist'n*, **38**, 73–77.

Box, G. E. P. (1980) Sampling and Bayes' inference in scientific modelling and robustness. *J. R. Statist. Soc.* A, **143**, 383–430.

Carroll, R. J. (1983) Comment on minimax aspects of bounded-influence regression by P. J. Huber. *J. Amer. Statist. Ass.*, 78, 78–79.

Carroll, R. J. and Ruppert, D. (1985) Transformations in regression: A robust analysis. *Technometrics*, **27**, 1–12.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) *Graphical Methods for Data Analysis*, Boston: Duxbury Press.

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

——— (1979) Influential observations in linear regression. *J. Amer. Statist. Ass.*, 74, 169–174.

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

Cook, R. D. and Wang, P. C. (1983) Transformations and influential cases in regression. *Technometrics*, **25**, 337–343.

Cook, R. D. and Weisberg, S. (1983a) Comment on minimax aspects of bounded-influence regression by P. J. Huber. *J. Amer. Statist. Ass.*, **78**, 74.

——— (1983b) Diagnostics for heteroscedasticity in regression. *Biometrika*, **70**, 1–10.

Cox, D. R. and Hinkley, D. V. (1974) *Theoretical Statistics*. London: Chapman and Hall.

Davies, R. B. and Hutton, B. (1975) The effect of errors in the independent variables in linear regression. *Biometrika*, **62**, 383–391.

Emerson, H. D., Hoaglin, D. C. and Kempthorne, P. J. (1984) Leverage in least squares additive-plus-multiplicative fits for two-way tables. *J. Amer. Statist. Ass.*, 79, 329–335.

Geisser, S. (1985) On the prediction of observables: A selective update. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds), pp. 203–230. Amsterdam: North-Holland.

Goetz, A. (1970) *Introduction to Differential Geometry*. Reading, MA.: Addison Wesley.

Hodges, S. D. and Moore, P. G. (1972) Data uncertainties and least squares regression. *Appl. Statist.*, **21**, 185–195.

Johnson, W. and Geisser, S. (1983) A predictive view of the detection and characterization of influential observations in regression analysis. *J. Amer. Statist. Ass.*, **78**, 137–144.

Kelly, G. (1984) The influence function in the errors in variables problem. *Ann. Statist.*, **12**, 87–100.

Leamer, E. E. (1984) Global sensitivity results for generalized least squares estimates. *J. Amer. Statist. Ass.*, 79, 867–870.

Lustbader, E. D. and Moolgavkar, S. H. (1985) A diagnostic statistic for the score test. *J. Amer. Statist. Ass.*, to appear.

Marquardt, D. W. and Snee, R. D. (1975) Ridge regression in practice. *Amer. Statist'n*, **29**, 3–19.

Millman, R. S. and Parker, G. D. (1977) *Elements of Differential Geometry*. Englewood Cliffs, NJ: Prentice-Hall.

Moolgavkar, S. H., Lustbader, E. D. and Venzon, D. J. (1984) A geometric approach to non-linear regression diagnostics with application to matched case-control studies. *Ann. Statist.*, **12**, 816–826.

Oman, S. D. (1984) Analyzing residuals in calibration problems. *Technometrics*, **26**, 347–354.

Polasek, W. (1984) Regression diagnostics for general linear regression models. *J. Amer. Statist. Ass.*, 79, 336–340.

Pregibon, D. (1981) Logistic regression diagnostics. *Ann. Statist.*, **9**, 705–724.

Stoker, J. J. (1969) *Differential Geometry*. New York: Wiley.

Storer, B. E. and Crowley, J. (1985) A diagnostic for Cox regression and general conditional likelihoods. *J. Amer. Statist. Ass.*, **80**, 139–147.

Weisberg, S. (1980) *Applied Linear Regression*, New York: Wiley.

Welsch, R. E. (1982) Influence functions and regression diagnostics. In *Modern Data Analysis* (R. L. Launer, and A. F. Siegel, eds). New York: Academic Press.

DISCUSSION OF PROFESSOR COOK'S PAPER

Professor R. M. Loynes (University of Sheffield): Ideas rarely spring fully-armed from the brow of their begetter, and ideas of diagnosis of problems in regression can be traced back a long way if one tries, but there is no doubt that Professor Cook's introduction of what is now called Cook's distance (Cook, 1977) seemed a remarkable advance at the time, and it is therefore a considerable pleasure to listen to him speaking tonight on a related topic. Now, of course, everyone calculates this, or some very closely related, measure when fitting a regression model, and, as Professor Cook observes, the general idea, of deleting cases, has been adapted for use in other contexts—in logistic regression and calibration problems for example. A further application, incidentally, is in subset selection (Loynes 1983; Chatterjee and Hadi, 1984).

The object of tonight's paper is rather different: assessment of the effect of small perturbations of the model. Before coming to the main topic, by the way, I'm not sure I entirely agree that case deletion is merely a reflection of model perturbation, as seems to be assumed: one particular kind of model perturbation happens to lead to case deletion in its associated estimator, but logically they don't seem to be tied very closely together—a large D_i, for example, can indicate either error in the model (should be non-linear) or error in the data. In a similar way, I would not consider, in most circumstances, a perturbation of the explanatory variables as a perturbation of the model: moreover, is the likelihood displacement necessarily the right criterion in this context? For subset selection problems (which might be regarded as an extreme kind of explanatory variable perturbation), for example, the unadjusted likelihood is certainly not appropriate.

However, these are mostly rather minor matters of terminology. The underlying tenet, that, if a minor perturbation in the model leads to a major change in essential parts of the results of the analysis, then there is evidence of a difficulty, is very appealing and, indeed, I would say completely convincing. It suggests measuring the sensitivity of the analysis to change in the model by some kind of derivative, and this is exactly what is suggested: in fact what is used is essentially the second derivative at the proposed model of the likelihood displacement as defined at (5)—more precisely it is the normal curvature of the likelihood displacement surface.

This is an interesting and valuable idea, and some of its consequences and applications are worked out in the paper—I particularly liked the possibility of focussing on single columns of X, as in sections 5.1 and 5.4—but I think some further development is needed: I see two apparently different difficulties, though they may merely be variations on a theme, with the present formulation. In the first place, neither the curvature nor the direction of maximum curvature is invariant with respect to reparameterization. Suppose that the case weights for normal linear regression are taken as $(1 + w_i)/2$ rather than w_i: then the original C_l is multiplied by $\frac{1}{4}$. (Of course this is not a very natural parametrization in the present situation, but I don't know how to ensure or even define a natural one.) This means that we have no idea of the appropriate scale for C_l. (And even within the terms of the paper, why is the value 2 necessarily appropriate? Should the yardstick perhaps depend on X, at least to the extent of involving its column rank p?.) The second difficulty is that the measure C_l depends on the details of the perturbation permitted. Suppose that, as in Section 4.2, we consider the effect of perturbing a single case, but instead of varying the weight directly we take the disturbance term in the ith observation to have a distribution which is $N(0, \sigma^2)$ with probability w_i and $N(0, k\sigma^2)$ with probability $1 - w_i$. Then we find that

$$C = 2h_{ii} \frac{(k-1)^2}{k^3} \frac{e_i^2}{\sigma^2} \exp\left\{\frac{k-1}{k} \frac{e_i^2}{\sigma^2}\right\},$$

which is to be compared with (32). Both models of perturbation predict larger variance for y_i if $w_i < 1$, and so in some rather general sense are similar, and I'm not sure how often I would be able to plump for one rather than the other, but it seems to me that the message to be got from one might be quite different from that derived from the other. However it may be that this is not in fact a serious problem.

As another application, which again may throw some light on the question, take the normal linear model and assume the variance-covariance matrix of y is $\sigma^2 (I + W)^{-1}$, with σ^2 assumed known for convenience. The relevant part of the log-likelihood is

$$L(\beta, w) = -(2\sigma^2)^{-1} (y - X\beta)' (I + W) (y - X\beta)$$

where W represents a perturbation in both variances and covariances. Although W is in fact symmetric the expression for $L(\beta, w)$ is unchanged if no such restriction is imposed and as it is slightly simpler to do so we take a general W. The vector l in this case is most easily thought of as a matrix L, and the analysis to find the curvature in direction L can be carried through either by using suffices explicitly or by using the "vec" operator. In either case the equivalent of (2.9) is found to be $2e'L'P_xLe/\sigma^2$ with constraint $\mathrm{tr}(L'L) = 1$. The equation for a maximum is thus $P_xLee' = \lambda L$ for some λ: postmultiplying by e shows that Le is an eigenvector of P_x, and since it must obviously correspond to a non-zero eigenvalue we have $Le = Xt$ for some t. Substituting this into the previous equation we find $L = Xse'$. Then the constraint gives $e'e\,s'X'Xs = 1$, and the maximum curvature is then $2e'es'X'P_xXse'e/\sigma^2 = 2e'e/\sigma^2$, for any s. Again, if only β is of interest and σ^2 is estimated we replace σ^2 by $\hat{\sigma}^2$ and find $C_{\max} = 2n$ which is both independent of X and large. Assuming my algebra is correct, does this result tell us that the perturbation scheme is unreasonable?

Returning to the first problem, I don't have a particularly compelling solution. As far as showing that local influence is different to global influence, and that the behaviour of the likelihood displacement surface near the postulated (unperturbed) model is of considerable interest and probably importance I think that both the general argument and Fig. 1 are convincing. It would be possible to use the ratio of C_l to the square of the slope of the line joining the case deletion value to the unperturbed value, but this seems clumsy. Another approach would be to take the position that it is a comparison that is sought, for example between C_l for different realizations of the same model with the same perturbation scheme, that is needed; this may show that the value 2 (or rather a value derived in just that way), on which some doubt was thrown above, is in fact both valuable and sensible.

But the basic idea is surely important and, as our chairman reminded us in a similar situation a year or two ago—it is the first step which counts. I have therefore much pleasure in proposing the vote of thanks.

Dr A. J. Lawrance (University of Birmingham): I would like to welcome Professor Cook to our meeting tonight. His paper develops an important new view of influence based on likelihood displacement. One of the main statistics from Professor Cook's earlier deletion approach is also equivalently a likelihood displacement, but is now usually called by the name of Cook's Distance; it is available in many statistical packages. I have found out that Cook's distance tonight is actually 4,015 miles, thus giving a further demonstration that it is not intended to measure local influence.

Tonight's paper is primarily concerned with local influence, and its chief conceptual tool is that of differential geometry. I was delighted to see the subject put to such good statistical use. Also I congratulate Professor Cook on having found what must be the last new thing to perturb in a statistical analysis, that is perhaps surprisingly, the parameters of the model. Previous perturbing has been to the data and to the explanatory variables, and tonight's paper has comments on these as well.

I find emphasis in the paper on case-weight analysis to be a little overdone, and regard case-weights as rather artificial parameters of the model—if you want to equivalently regard them as reciprocals of non-constant variances in a least squares analysis, then heteroscedasticity can be analysed in its own right. This appears to be the *de facto* use in the examples anyway. I wonder if there is any advantage to be had from describing the non-constancy in terms of external variables Z and a perturbing parameter λ, perhaps in the form $\sigma^2 \exp(\lambda Z)$. If case-weight local analysis is to be undertaken for its heteroscedastic implications, then I would have thought that plots such as Fig. 1 should be superimposed for each and every data case, and perhaps be studied for their outlying or clustering effects, or absence of such effects. Incidentally, the slopes at $\omega = 0$ in Fig. 1 are another informative feature and are given by

$$-2\,\frac{h_{ii}}{1-h_{ii}}\left[\frac{y_i - x_i\hat{\beta}_{(i)}}{\hat{\sigma}_{(i)}}\right]^2$$

This is the result for the ith-case curve with a general regression model when only β is being considered. Here $h_{ii}/(1-h_{ii})$ represents the distance of the ith point in the explanatory space from the average of the other points and $y_i - x_i\hat{\beta}_{(i)}$ is the usual predictive residual; $\hat{\sigma}^2_{(i)}$ is a predictive residual mean square over all responses, but based on fitted values which have been calculated without using the ith case of the data. The slope result may interestingly be compared with (32) for the curvature at zero, and in practice should be used in conjunction with $LD_s(0)$ which is $n \log(\hat{\sigma}^2_{(i)}/\hat{\sigma}^2)$.

The core of the paper is in Sections 2.2 and 3.1 with the introduction of the likelihood displacement (5). I am particularly grateful to the author for his careful discussion of curvature, and in his presentation, for the three-dimensional illustration of an influence surface concerning the rat data. Also, by indicating that curvature refers to the ordinary curvature of the curve formed by a lifted line in a specified direction, we have a clear intuitive idea of its meaning. The author takes the second derivative form of curvature in a particular direction and shows that it can be neatly reexpressed as (16). The calculation does, however, need explicit knowledge of the maximum likelihood estimates $\hat{\theta}$ of θ, but only at the non-perturbation value ω_0 of ω. This is perhaps just minimally inconvenient. From (16) the directions of maximum curvature follow. All this is very nice, although the need to explicitly calculate the estimates and eigenvalues can lead to intractability in theoretical developments. Life is also more complicated if curvatures are required at points away from the non-perturbing value ω_0. The relevant derivations (17) to (20) are certainly skilful but it is a pity that there are no illustrations in the paper to demonstrate their tractability or to justify their usefulness.

Turning next to the case-weight local analysis of Section 4, we are advised to inspect the elements of $l_{\max}$, the over-all direction which gives maximum curvature. The relevance of individual elements is however an empirical observation and has not yet been justified by their explicit interplay with the basic quantities of regression analysis; indeed, this would appear to require the intractable eigenvalue analysis. I am also slightly mystified by the sanctity of two for the maximum curvature. Its value depends on the parameterization of the perturbing. What other evidence is there that this represents a universally small curvature? The results for case-weight local analysis of individual regression coefficients are elegant, and interesting in their relation to added variable plots, or individual coefficient plots, as I sometimes prefer to call them. I wonder about $C_{\max}$ and $l_{\max}$ at deletion ($\omega = 0$) instead of at inclusion ($\omega = 1$), but note that in presenting the paper, the author appeared to play down the analysis of maximum curvature, while emphasizing the importance of its direction. This needs further theoretical clarification.

I should like briefly to take up the author's invitation to consider other possible applications, and in view of our overlapping interests in regression transformation, I will briefly consider this area. In the spirit of the Box–Cox approach one might first think of perturbing the regression model (4) of the paper using the transformation parameter λ. This would be convenient since λ does not have to be estimated but is used at $\lambda = 1$, its non-perturbing value. However, the likelihood displacement (5) is measuring changes in parameter values. The regression parameters β and σ^2 clearly change enormously with changes in scale of measurement, and curvatures would not be meaningful. Thus, the author's comments about the careful and sensible choice of ω are to be heeded; scale transformation is not an operation which causes "interesting" changes in regression parameter estimates. I think I am coming round to case-weight local analysis! Suppose we regard λ as a parameter and eliminate β and σ^2 from the unperturbed likelihood, since changes in β and σ^2 are not relevant to our initial interest in λ; this is the parameter-subsets approach at (7). We have a likelihood displacement $LD_s(\omega)$ involving $\hat{\lambda}(\omega)$, the maximum likelihood estimate of λ for the full and ω-displaced likelihood. This will be obtained by minimizing

$$\tilde{y}^{(\lambda)T}(W - WX(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W)\,\tilde{y}^{(\lambda)}$$

where $W = \operatorname{diag}(\omega_1, \omega_2, \ldots, \omega_n)$ and $\tilde{y}^{(\lambda)}$ is the usual scale-corrected version of $y^{(\lambda)}$. The likelihood displacement is then given using the unperturbed profile likelihood for λ as

$$LD_s(\omega) = \text{constant} + n \log\{\tilde{y}^{(\lambda(\omega))T} P'_X\, \tilde{y}^{(\lambda(\omega))}\}.$$

The use I see for this is in the situation of Fig. 1 in which just the ith case is perturbed and ω is scalar. The computing of $\hat{\lambda}(\omega)$ is of course the drawback for explicit results, but the curvature at $\omega = 1$ and the slope at $\omega = 0$ should be tractable, and I hope to publish results later. If this is indeed the case, one could determine which are the individually most influential data cases as far as λ is concerned. If we are just interested in maximum curvatures, we may go to (2.6) of the paper; for this we again only require $\hat{\lambda} = \lambda(1)$, and the observed information matrix $\ddot{L}$ for λ, β and σ^2 which is not impossible. But it may be easier to use (10) directly.

I will conclude by venturing to suggest that regression diagnostics are here to stay, and that tonight's paper is a valuable contribution to the area by one of its most innovative contributors. Thus I am not in the least bit perturbed by this paper, and have great pleasure in seconding the vote of thanks.

The vote of thanks was passed by acclamation.

Dr F. Critchley (University of Warwick): It is a pleasure to congratulate the author on a major contribution to an important area. The spirit of my comments is two-fold: to balance some of the points made and to suggest some extensions to what is, I am sure, going to prove a seminal paper.

In his Introduction, the author raises three objections to an existing device, the influence *curve*: (1) difficulties in its construction limit its scope; (2) a norm must be selected; and (3) there is too much choice (from among the several sample versions). It appears that similar objections apply equally to the influence *graph* and the related constructs described in the present paper: (1) their construction is not immediate, especially when $\hat{\theta}$ and/or $l_{\max}$ are not known explicitly. Lack of an analytic expression also limits theoretical insight; (2) this is not a problem here precisely because only *one* response (the likelihood displacement) is being monitored. As the author fairly states in the second paragraph of the Discussion, this is not a virtue. Whereas $LD(\omega)$ has merits when viewed as a norm for the change in the p-dimensional parameter estimate, there are, as with the influence curve, a variety of other measures of the divergence of $\hat{\theta}$ and $\hat{\theta}_{\omega}$. Moreover, here we have the general challenge of comparing two entire log-likelihood functions and not just their modal values; and (3) in the penultimate paragraph of the paper, the author himself fairly discusses the problems associated with the q-dimensional choice of ω.

It is also instructive to enquire how strong these objections really are. The answer appears to be "not very". The author presents a strong case for the value of the influence graph. For the influence curve, we note that: (1) at yesterday's meeting of the Multivariate Study Group, Patricia Calder, Ian Jolliffe and the current writer demonstrated the tractability and utility of the influence curve in multivariate analysis. Indeed, in Principal Components Analysis, there is an *additive* measure which therefore avoids the masking and computational difficulties which arise in regression (see the end of Section 2); (2) Selection of a norm is necessary whenever more than one response is monitored. This does not seem to be a major difficulty. Indeed, as with (3), it seems rather to be an advantage in that a variety of measures provide a variety of information. Finally, concerning (3), we note that Critchley (1985, Section 4) contains some general theoretical results comparing three standard sample versions of the influence curve. These are in terms of the *curvature* of a certain functional and so are very much in the same spirit as tonight's paper.

Overall, I believe we should welcome the influence graph as a valuable supplement to the influence curve rather than as a replacement for it. Finally, three possible extensions.

Uncertainty: currently the state of the art in influence analysis is point estimation. Influence curve measures are based on a point estimate $\hat{F}$ of the underlying distribution and influence graph measures on (highly nonlinear) functions of $\hat{\theta}$ and $\hat{\theta}_{\omega}$. Obvious benefits will flow from incorporating uncertainty into influence analysis. There are many ways in which this could be done. One example for the present paper would be to use interval estimates based on the null ($\omega = \omega_0$) distribution of $C_{\max}$ and $l_{\max}$. Generally, it seems that uncertainty will increase with q.

Global measures: The tough challenge which these present seems less difficult when $LD(\omega)$ is known explicitly. For instance in the motivating example in Section 2.1 we can show that $LD(\omega) = \alpha_i^2 LD(0)$ where

$$\alpha_i = \frac{(1-h_{ii})(1-\omega)}{\{1-h_{ii}(1-\omega)\}} \text{ and } LD(0) = \frac{h_{ii}}{(1-h_{ii})^2}\,\frac{e_i^2}{\sigma^2}$$

with $LD(0)$ corresponding to deleting the ith case completely. Apart from the known constant σ^2, the whole graph of $LD(.)$ depends only upon two quantities h_{ii} and e_i and so can be described globally by just *two* summary measures, say its curvature at ω_0 (as proposed) together with $LD(0)$. As an aside, this is a small example of the fruitful supplementing of old ideas with new as advocated above.

Data analysis without a probability model: in principle there is no barrier to extending influence analysis beyond the realm of probability models. For example, the influence curve can be used in both scaling and clustering methods by replacing F with the observed dissimilarity matrix in the appropriate functional. For the influence graph, we need only replace maximized log-likelihoods by the negatives of the discrepancy measures (e.g. sum of squares) minimized by the data analytic procedure.

Professor A. C. Atkinson (Imperial College, London): It is a pleasure to congratulate Professor Cook both on a paper which suggests many interesting new avenues for research and on his lucid presentation this evening of the key ideas behind his paper.

I have two points. The first is to ask for clarification of the relationship between the scheme for model perturbation and the score test for the related elaboration of the model.

The second point is more in the nature of a challenge. The standard works on diagnostic regression analysis (Belsley, Kuh and Welsch, 1980; Cook and Weisberg, 1982; Atkinson, 1985 and Weisberg, 1985, Chapters 5 and 6) describe single deletion diagnostics in appreciable detail, but give little guidance on how to detect multiple outliers in the presence of masking. One exploratory method which does work is least median of squares regression (Rousseeuw, 1984). Let the residual $r_i = y_i - x_i^{\mathrm{T}} b$. Two criteria for the estimation of b are

Ordinary Least Squares $\quad \underset{b}{\text{Minimize}} \ \Sigma r_i^2$

Least Median of Squares $\quad \underset{b}{\text{Minimize}} \ \underset{i}{\text{median}} \ r_i^2.$

The least median of squares estimate of b is found by a Monte-Carlo algorithm in which the model is fitted to randomly chosen subsets of p cases (Hawkins, Bradu and Kass, 1984; Rousseeuw and Leroy in an unpublished technical report). This exploratory first stage of fitting to reveal outliers is followed by a confirmatory stage. Rousseeuw (1984) uses robust estimation; Atkinson (1986) least squares with omission of suspect cases followed by single case deletion and prediction diagnostics.

Rousseeuw (1984) motivates his discussion with a simulated example in which there are 30 "good" observations and 20 "bad" ones. Atkinson (1986) shows that single deletion diagnostics fail to reveal this structure. I would like to see what Professor Cook's methods make of this extreme example.

Professor Sir David Cox (Department of Mathematics, Imperial College, London): It is perhaps only the most sophisticated of the theoretical physicists investigating the very foundations of the Universe who consider that their mathematical representations are in some sense exact: others regard mathematical formulations as idealized approximate representations of aspects of the system under study. Therefore investigation of sensitivity to model formulation is very important. Therefore the present paper is extremely welcome.

Professor Cook's approach is interesting and fruitful. There are other formulations. The following is a sketch of one alternative. Let $f(y;\theta)$ denote a working model and $\pi(y)$ an unknown truer model. We fit the working model say by maximum likelihood, obtaining a maximum likelihood estimate $\hat\theta$ and an estimated covariance matrix $(n\hat{\imath})^{-1}$ from the observed information. Now under the model π, it is known that the estimate converges to θ_π, say, that the estimated variance converges to $(ni_\pi)^{-1}$ and that the asymptotic variance of $\hat\theta$ is $(n\tilde{i}_\pi)^{-1}$, say, where simple formulae (Cox, 1961) are available for these vectors and matrices.

In the first place, two questions arise. Is θ_π meaningful? What is the relation between apparent and true covariance matrices, i.e. between i_π and $\tilde{i}_\pi$?

Provided that the parameter is regarded merely as a specifier of a model and not as something of intrinsic interest, the first question has a simple answer in that θ_π gives the model within the family in question that is closest to π in minimizing a natural measure of distance, $\| f - \pi \|$; the best-fitting model within the family is obtained. This leaves, of course, the issue as to whether the model family can be improved. The second question leads to a constrained calculus of variations problem to find the kind of departure with respect to which there is greatest sensitivity; this is somewhat analogous to the treatment in the paper, although aimed at variance of estimation rather than at the point estimate itself.

If, however, we regard the parameter, or more realistically a component θ_s, say, as of intrinsic interest, the question becomes one of whether $\theta_{s\pi}$ keeps its physical interpretation as a slope, a quantile, or whatever. Of course, as Professor Cook notes, similar questions arise in his discussion, both as regards the specification of the parameter of interest and as regards the perturbing parameter ω; in particular it is necessary that his ω is chosen so that the Euclidean metric used to determine curvature is meaningful. Such questions about parameterization are addressed to some extent in as yet unpublished work of N. Reid's and mine.

It will be clear that I found Professor Cook's paper very stimulating.

Dr P. Prescott (Southampton University): May I express my appreciation to Professor Cook for a stimulating paper which I have enjoyed very much. I am certain that close examination of influence graphs will form a most useful generalization of diagnostic data analysis procedures and it is likely that this work will prompt further research into ways of assessing the shapes of influence graphs in addition to consideration of its direction of maximum curvature. There are, however two matters about which I am concerned.

The first point which worries me concerns the evaluation of *maximum* curvature or of examination of the shape of the lifted line in the direction of maximum curvature. The relevance of any numerical "benchmark" selected to assess "largeness" for the maximum curvature, and therefore providing cause for concern about the analysis, will change with the dimensionality of ω. We have seen that it is possible to extend ω to include case weights, explanatory variable perturbations, changes in the correlation structure and other sensible perturbations to the data or to the model. In the most general case ω could contain a large number of elements, well in excess of the number of observations. Analytical evaluation of the maximum curvature will no doubt be very difficult in these cases but numerical methods may be feasible. Care needs to be exercised in assessing the importance of a specific perturbation identified in this way from amongst many others.

I used the term "benchmark" earlier rather than "critical value" in this context since I believe that it is inappropriate to assess the "significance" of a large curvature, or of the peakedness of a lifted line plot, by means of a formal significance test. Professor Cook does not mention critical values but does suggest that further research could involve simulations and this causes me some concern. There will always be an element of choice in the form of perturbation to be considered. In fact we are advised by the author that "ω must be chosen carefully so that the application is sensible". I suspect that in practice the form of ω will also depend on an initial examination of the data. Professor Cook has warned of these difficulties and dangers in his conclusions but it should be stressed that anyone considering using simulation methods to determine critical values for maximum curvature should keep in mind that the element of selection of the form of ω could lead to inappropriate conclusions.

The following contributions were received in writing, after the meeting.

Mr R. Beckman (Los Alamos Nat. Lab., USA): Professor Cook is to be congratulated; the paper he presented here is outstanding. It is destined to become a classic. While the author devotes most of his examples to linear regression, the real worth of the paper is the notion that local influence may be assessed for most likelihood based procedures. Influence is no longer restricted to linear models.

There are four minor points which I would like to make. First, it is not clear that $C_{\max} = 2$ "serves as a useful general reference". The maximum curvature is a function of the perturbation scheme, and only experience will allow one to judge its relative size. This is made clear by the author's scale functions S for the perturbation of the explanatory variables in linear regression. From the author's equation (41) the curvature is directly proportional to the square of the scale s_j. Therefore, as demonstrated in the rat example, the curvature may take on any value.

In the explanatory variables case Professor Cook has left the choice of s_j arbitrary. I think it best at this point to require the s_j to be either the range or standard deviation of the jth explanatory variable, as either of these choices will leave the curvature scale invariant. Experience will then dictate the size of "significant" curvatures.

Implementation of these procedures in moderate sized problems will lead to some numerical headaches. The matrix F with moderate n can become so large that the computation of eigenvalues and eigenvectors becomes a problem. Even the author's mathematical "tricks" between equations (37) and (38) do not lead to the eigenvector $l_{\max}$; it is critical to know $l_{\max}$ and hence the direction of the maximum curvature.

The choice of the perturbation scheme may be a problem. The author has provided the machinery to assess local influence in so many ways—perturbation of the *ys*, perturbation of the *xs*, perturbation of the case weights—that one is left with the possibility of using different perturbation schemes until some significant local influence factor is found. Is it possible that all data sets will exhibit some form of local influence?

The last problem I see is not confined to the procedures given here. The author has stated, and I reiterate, that accommodation of influential cases is not addressed in this work. With all the procedures given to us in the last few years, the practitioners are left with the uneasy feelings of estimation errors when the influence flag is raised. Until now however all influence measures dealt with case deletion. The logical mode of accommodation therefore was removal of the influential case. I am not sure what accommodation means in the case of local influence.

Once again, Professor Cook is to be congratulated on such a significant work.

Dr R. W. Farebrother (University of Manchester): The maximum likelihood estimator of β in the model

$$y_j = x_j'\beta + \epsilon_j \qquad \epsilon_j \sim IN(0, \sigma_j^2)$$

is given by

$$b(\iota) = (X'X)^{-1} X'y$$

if $\sigma_j^2 = \sigma^2$ and by

$$b(w) = [X'D(w)X]^{-1} X'D(w)y$$

if $\sigma_j^2 = \sigma^2/w_j$. The difference between these two estimators is

$$b(w) - b(\iota) = [X'D(w)X]^{-1} X'D(w)\,[y - Xb(\iota)]$$

and thus

$$K_v = e'D(w)X\,[X'D(w)X]^{-1} X'D(w)e/(\sigma^2 v'v)$$

is a measure of the proportionate effect on $b(w)$ of a small change in w from $w = \iota$ to $w = \iota + v$.

Now $D(w) = D(\iota) + D(v)$, $D(\iota) = I_n$ and $X'e = 0$ so that we may approximate $2K_v$ by

$$C_v = 2e'D(v)X(X'X)^{-1} X'D(v)e/(\sigma^2 v'v)$$

or

$$C_v = 2v'D(e)X(X'X)^{-1} X'D(e)v/(\sigma^2 v'v)$$

which is identical to the author's C_l statistic (29) if we set $w_0 = \iota$, $a = \| v \|$ and $l = v/\| v \|$ in his equation (10).

The values of v which minimize C_v for all y may readily be obtained by solving $X'D(v)Q = 0$ or

$$\sum_{j=1}^{n} x_{ji} q_{jk} v_j = 0 \qquad i = 1, 2, \ldots, p, \quad k = 1, 2, \ldots, n - p,$$

for v where Q is an $n \times (n - p)$ matrix satisfying $Q'X = 0$ and $Q'Q = I_{n-p}$.

In general the value of v which maximizes C_v is not known but if $p = 1$ and X is an $n \times 1$ column of ones then it is given by $v = e$. If we set $v = e$ and $\hat{\sigma}^2 = e'e/n$ when $p > 1$ then we obtain the statistic

$$\hat{C}_e = 2e_{sq}'X(X'X)^{-1} X'e_{sq}/(n\hat{\sigma}^4)$$

which is closely related to Breusch and Pagan's (1979), p. 1289) statistic

$$BP = e_{sq}'Z(Z'Z)^{-1} Z'e_{sq}/(2\hat{\sigma}^4) - n/2$$

for testing for heteroscedasticity of the form $\sigma_t^2 = h(z_t'\alpha)$.

Professor S. Geisser (University of Minnesota, USA): I would like to briefly sketch a Bayesian approach to the perturbation problem that has been discussed by Professor Cook. Let $L(d(y), \theta)$ be the loss in making decision d on observing y when θ is true. Assume that $P_\omega(\theta \mid y)$ is the "perturbed" posterior distribution of θ obtained from the perturbed likelihood $l_\omega(\theta \mid y)$ and prior distribution $g_\omega(\theta)$. Then for each $\omega \in \Omega$ let $\bar{L}_\omega(d(y))$ be the expected loss and d_ω^* minimize this loss so that

$$\bar{L}(d_\omega^*) = \min_d \bar{L}_\omega(d).$$

Hence perturbations in ω show how decision d_ω^* will vary from $d_{\omega_0}^*$ with a corresponding differential loss $\bar{L}(d_\omega^*) - \bar{L}(d_{\omega_0}^*)$. The latter is presumably expressed in a well defined and understood unit of measurement. This formal approach can be used for a subset of the parameter set or more importantly for predictions where a predictive loss function and a predictive distribution replace the parametric loss and distribution functions respectively.

In many situations where a decision is either not the immediate goal or is embedded in the reporting of the posterior or predictive distribution, i.e., a clearly informative summary which can be used for inferences or decisions by others at some future time, Kullback–Leibler divergences are useful measures of the effects of model perturbation.

For example one such divergence is

$$I(p_\omega, p_{\omega_0}) = E\left[\ln \frac{p_\omega(\theta \mid y)}{p_{\omega_0}(\theta \mid y)}\right],$$

where the expectation is taken over $p_\omega(\theta \mid y)$, the perturbed posterior density of θ. Another potentially useful divergence is its counterpart $I(p_{\omega_0}, p_\omega)$. Other measures could also be entertained. For example

$$\delta_\omega = \max_\theta \mid P_\omega(\theta \mid y) - P_{\omega_0}(\theta \mid y) \mid$$

will also be informative. In each of these cases a perturbation graph of the influence of ω on the measure provides a revealing summary of the potential perturbation effect.

Similarly for prediction one would use the predictive distribution $F_\omega(z \mid y)$ in place of the posterior distribution. Generally perturbations will have much less of an effect on prediction than they do on estimation, see for example my rejoinder to discussion, Geisser (1982). In addition one would also expect that the smaller the sample size the less the effect of the perturbation. Thus the interpretation of $LD(\omega)$ as an asymptotically useful feature will, it appears, overemphasize the effect of the perturbation for small or moderate sample sizes. This, of course, would not happen under the Bayesian approach discussed here.

Professor C. J. Nachtsheim (University of Minnesota, USA): Reading this paper has been enjoyable and thought provoking. Professor Cook has developed an amazingly general method for assessing the stability of statistical models to perturbations in assumptions. As indicated in the paper, applications in regression and general linear models seem unlimited. Part of my enthusiasm for the method, however, springs from the potential I see for its use in statistics when the objective function is other than the likelihood.

In optimal experimental design, for instance, the objective function is often the determinant of the information matrix. For approximate (continuous) optimal designs, this matrix is a function of the support points and the variances and the design weights associated with each point. Using Professor Cook's approach, one might assess the criticality of assumptions concerning the variance of the response at each design point. Alternatively, one might assess the sensitivity of the determinant to perturbations in the weights or the supports. Since, in practice, it may be difficult to precisely achieve the desired weights and treatment combinations, such diagnostic information could be of use. Some preliminary work concerning perturbations of design weights suggests that many standard optimal designs are relatively insensitive to perturbations of the optimal weights. Assessment of the influence of perturbations to the support points is slightly more complicated, since many are often on the boundary of the design space. The derivative of the criterion function with respect to the coordinates of such points is usually nonzero; thus, without modification (i.e., transformation of the design space if possible) the method will not strictly apply.

In light of current industrial emphasis on off-line quality control, practitioners are often concerned that the response, $y(x)$, be insensitive to minor perturbations from the solution point x^*, found via response surface analysis. The impact of worst-case local perturbations is easily assessed, if x^* is an unconstrained local maximum. In this case F is simply the Hessian of the regression function $y(x)$ evaluated at x^* and $l_{\max}$ is the eigenvector associated with the largest eigenvalue of F. Inspection of the lifted line generated by $l_{\max}$, then, would reveal (locally) the worst case decline in the predicted response, with accompanying implications for on-line quality control or further research. Of course, canonical analysis of F is often standard procedure when analysing response surfaces. Professor Cook's work suggests that additional, useful diagnostic information can be obtained in this situation from an inspection of $C_{\max}$ and the lifted line generated by $l_{\max}$.

Professor D. Peña (Universidad Politécnica de Madrid): This lucid paper helps to clarify important issues on the assessment of influence in statistical models. In fact, the perturbation scheme used in the paper points out that the key issue in the study of influence is not the case-deletion technique, but the assumption of lack of knowledge about the reliability of one observation or subset.

Thus, the assessment of global influence should be made assuming that the observation is *missing*. Of course, for the linear model, both approaches are equivalent but they are not for dependent data as time series. It has been shown (Peña, 1984, 1986) that this missing value approach leads to sensible diagnostic measures of influence for *ARIMA* models, whereas the case-deleting does not. Suppose, for instance, that the data are generated by the dynamic linear model:

$$y_t = \mu + \mathbf{b}'\mathbf{x}_t + \phi' \mathbf{Y}_{t-1} + \theta' \mathbf{a}_{t-1} + a_t$$

where μ is a global level parameter, $\mathbf{b}$, ϕ and θ are vectors of unknown parameters and $\mathbf{x}_t$ is a vector of explanatory variables, $\mathbf{Y}'_{t-1} = (y_{t-1}, \ldots, y_{t-p})$ a vector of lagged values of the response and $\mathbf{a}_{t-1} = (a_{t-1}, \ldots, a_{t-q})$ a vector of white noise. This model includes the standard linear model $(\phi' = \theta' = 0')$, *ARIMA* models $(\mathbf{b}' = \mathbf{0})$ and dynamic transfer function model. The model has a level parameter, μ, a scale parameter σ^2, and a vector of structural parameters $\beta' = (\mathbf{b}'\ \phi'\ \theta')$. Let $\hat{\mu}$, $\hat{\beta}$ be the *ML* estimators of these parameters; $\hat{\sigma}^2_{\mu}$ and $\hat{\Sigma}_{\beta}$ the variance and covariance matrix of these estimators; $\hat{\mu}_{(i)}$, $\hat{\beta}_{(i)}$ the *ML* estimators treating the ith observation as missing; $\hat{\sigma}^2$ the usual unbiased estimator of the variance σ^2; and $\hat{\sigma}^2_{(i)}$ this same estimator when the ith observation is missing. Then, we define:

(1) The influence on the global level as

$$D_L(i) = (\hat{\mu} - \hat{\mu}(i))^2 / \hat{\sigma}^2_{\mu},$$

(2) The influence on the structural parameters as

$$D_S(i) = (\hat{\beta} - \hat{\beta}_{(i)})^1\ \hat{\Sigma}_{\beta}^{-1}\ (\hat{\beta} - \hat{\beta}_{(i)}).$$

this statistic can be decomposed to separate the effects on the components of β.

(3) The influence on the scale is given by

$$D_V(i) = (\hat{\sigma}^2 - \hat{\sigma}^2_{(i)}) / \hat{\sigma}^2_{(i)}.$$

The application of these definitions to regression models leads to statistics similar to those proposed previously: the statistic suggested by Cook (1977) is $(D_S + D_L)/p$, whereas *DFFITS*—advocated by Belsley *et al.* (1980)—is $(D_S + D_L)\ (1 + D_V)$, being a mixture of the three kinds of effects. Note that D_V is equivalent to the likelihood ratio test for outliers in linear models.

However, the application of this missing value approach to dependent data leads to ***substitute each observation by its forecast using the rest of the data, instead of deleting it***, that seems a much more sensible procedure. Therefore, the missing value approach is more general, and includes as particular case for independent data the well known case-deletion procedure. I believe that this same idea can be used in some other fields—as Jacknife or cross-validation—that are based on the case-deletion technique.

In contrast from the deleted approach, the procedure advocated in this paper to study local influence based on the likelihood displacement, seems to provide sensible solutions for dependent as well as independent data. For instance, for the $AR(1)$ process:

$$y_t = \phi y_{t-1} + a_t$$

it is straightforward to show, using the conditional likelihood $l(\phi \mid y_1)$, that the curvature is

$$C = 2\, e_i^2\, z_{i-1}^2/n,$$

where $z_{i-1} = y_{i-1}/s_y$, $s_y^2 = \Sigma\, y_t^2/n$, $e_i = y_t - \hat{\phi} y_{t-1}$.

In closing, I would like to congratulate Professor Cook for this very interesting and stimulating paper.

Mr G. J. S. Ross (Rothamsted Experimental Station): The author has provided us with some very interesting new diagnostic tools which are important not only for evaluating past analyses but for designing future investigations. However there may be simpler methods of arriving at the same conclusions which are worth considering.

In the discussion of Atkinson (1982) I drew attention to the fact that reciprocals of the diagonal elements of P_x, widely known as the Hat-matrix, are numbers between 1 and n which may be interpreted as replication factors, being the size of a sample of independent observations with given x which estimate $E(y)$ with the same precision as the fitted model. The quantity $r_i = 1/h_{ii}$ I therefore termed "effective replication", by analogy with results from completely randomized designs. Effective replication turns out to be a useful idea in the present context, that of assessing local influence in linear and nonlinear regression. A paper on this subject has been submitted (Ross, 1986).

Consider first the Rat Data, fitting all three variables. The effective replication for Case 3 is only 1.175 which implies that the observation is nearly self-estimating, and that whatever value of y was observed would be fitted with a small residual. This occurs in linear regression when the value of x is a design outlier, and so a visual examination of the data quickly reveals that $x_3 = x_1/200$ except for Case 3, so that almost a whole degree of freedom is available to estimate Y. In nonlinear regression effective replication may also be computed using the local design matrix, and it shows that extreme points are more self estimating if they are on the steepest part of the fitted function.

The effect of changing the weighting of a given point may be expressed as the relative change in effective replication at other points. Effective replication may also be calculated at other values of x, so that curves or surfaces may be drawn showing which range of x constitutes the set of "neighbours" of a given point, whose expectations are influenced by that point. Thus the idea is equally useful for determining the effect of additional observations (design implications) as of removing observations.

Dr C. L. Tsai (New York University, USA): This interesting paper clearly demonstrates the value of geometry in statistics. Professor Cook proposes a normal curvature measure for the influence graph and shows how this can be used to assess the consequences of certain model perturbations. This assessment requires an objective criterion for judging whether the curvature measure is large or not. One possibility would be to compare the observed curvature calculated from the given data set with a bootstrap distribution of curvature found using simulation.

An alternative approach for judging the significance of model perturbations is based on hypothesis testing. Let $LD^*(w) = L(\hat{\theta}) - L(\hat{\theta}_w \mid w)$. This is the log likelihood ratio test statistic for testing the null hypothesis $w = w_0$. Therefore, $-2LD^*(\hat{w})$ has an asymptotic chi-squared distribution with degrees of freedom q. In order to avoid calculating $\hat{w}$, Cox and Hinkley (1974, Chapter 9) suggest the score test given by

$$S = \left(\frac{\partial L(\theta \mid w)}{\partial w}\right)^{\mathrm{T}} \left(E(I_w) - E(\Delta)^{\mathrm{T}} E(-\ddot{L})^{-1}\, E(\Delta)\right)^{-1} \left(\frac{\partial L(\theta \mid w)}{\partial w}\right)\Bigg|_{\substack{w = w_0 \\ \theta = \hat{\theta},}}$$

where $I_w = (-\partial^2 L(\theta \mid w)/\partial w \partial w^{\mathrm{T}})$, $\ddot{L}$ and Δ are respectively defined in equations (12) and (14), and E denotes the expectation operator. The statistic S is also distributed asymptotically chi-squared. If the expectation E is removed and the resulting statistic denoted S^*, then, by comparing S^* and $C_{\min}$, we see that if $C_{\min}$ is large, then S^* will also tend to be large, providing evidence against the null hypothesis that $w = w_0$. Perhaps Professor Cook would comment further on this connection between his geometrical approach and the hypothesis testing approach.

The **author** replied later in writing, as follows.

I am grateful to the discussants for their kind reception of this paper and for their many stimulating and penetrating comments. The overlap between the contributors remarks seems sufficient to justify arranging my reply by rather broad topics.

Invariance and benchmarks

Professor Loynes perceives two difficulties with the proposed methodology: (1) neither C_l nor l_{max} is invariant with respect to reparameterization and (2) C_l depends on the details of the perturbation permitted. To address these issues it is necessary to distinguish between a reparameterization of an influence graph and its definition. According to most texts on differential geometry (e.g. Millman and Parker, 1977), a reparameterization of an influence graph $\alpha: \Omega \to R^{q+1}$ is a smooth one-to-one function $g: \Omega^* \to \Omega$. The reparameterized influence graph is then $\alpha^* = \alpha(g(\omega^*)) = (g^T(\omega^*), LD(g(\omega^*))$ for ω^* in Ω^*. Under this defintion, the maximum curvature is in fact invariant with respect to reparameterization, a standard result in differential geometry. In contrast, Professor Loynes is concerned about invariance with respect to certain modifications of the perturbation scheme itself. This can be illustrated by using Professor Loynes' example where the case-weights ω are taken as $\omega = g(\omega^*) = (1 + \omega^*)/2$. In this situation, the case weights are perturbed through the parameterization provided by ω^*, but the elements of $g(\omega^*)$ are still case-weights no matter how they are parameterized. In showing that the curvature under ω^* is $\frac{1}{4}$ of the original curvature, Professor Loynes implicitly used the influence graph $\beta^T = (\omega^{*T}, LD(\omega^*))$ which is distinct from α since it is not defined in terms of case-weights. But this disagreement may be largely a matter of terminology.

The concern that neither the curvature nor the direction of maximum curvature is invariant with respect to seemingly unimportant changes in the perturbation scheme.does appear to be important. For progress it is necessary to specify the invariance properties that are required. Suppose, for example, that two case-weights are to be perturbed so that $\omega^T = (\omega_1, \omega_2)$. Further, let ω^* denote parameters defined by $\omega^T = (10\omega_1^*, 100\omega_2^*)$. Should we require that the influence graphs $\alpha^T = (\omega^T, LD(\omega))$ and $\beta^T = (\omega^{*T}, LD(\omega^*))$ have identical values for C_l and l_{max}? The answer seems to be "definitely not" since the perturbation schemes are fundamentally different: α is based on case-weight perturbations while β is based on perturbations of different fractional case-weights, ω_1^* and ω_2^*. The answer may be clearer if we view the ω's as perturbations of explanatory variables rather than case-weights.

On the other hand, it may be worthwhile to investigate the invariance properties of C_l and l_{max} under *identical* coordinate transformations of ω. Specifically, do the influence graphs $\alpha^T = (\omega^T, LD(\omega))$ and $\beta^T = (\omega^{*T}, LD(\omega^*))$ have the same values for C_l and l_{max} when $\omega_i = k(\omega_i^*)$, $i = 1, \ldots, q$, where k is a smooth one-to-one function? For this situation, α and β yield identical values for l_{max}, but the curvature for β is $[\partial k/\partial\omega^*]^2$ times the curvature for α, where the derivative of k is evaluated at the value of ω^* which yields the postulated model. When perturbing case-weights and $k(\omega^*) = (1 + \omega^*)/2$, $[\partial k/\partial\omega^*]^2 = \frac{1}{4}$, as mentioned by Professor Loynes. Thus, l_{max} is invariant under identical coordinate transformations, but C_l depends on the selected coordinate form. This dependence does not seem to be a serious problem, since the functional relationship between C_l's for different coordinate forms can be used to transfer interpretation from one form to the other. Dr Beckman describes two possible approaches when perturbing explanatory variables and k is restricted to scale transformations. With experience either approach should work. More generally, the comparative approach mentioned by Professor Loynes should prove useful. In fact, this is essentially how I determined the benchmark $C_{max} = 2$: My experience indicates that $C_{max} = 2$ is a useful rough guide when perturbing case-weights in linear models, but its usefulness may not extend to other models or other perturbation schemes. When working with parameters ω^* rather than case-weights, *per se*, the corresponding benchmark is $2\,[\partial k/\partial\omega^*]^2$, as indicated above. Recall, however, that an inspection of l_{max} may be worthwhile regardless of the size of C_{max} since l_{max} can identify global problems that are not manifest locally.

Formulation

A harder question is how to formulate a new perturbation scheme. Professor Loynes demonstrates that the proposed methodology allows great flexibility and that perturbing the entire covariance matrix W in a normal linear model can lead to rather curious results when the symmetry

W is neglected. I expect that allowing the covariance matrix to be nonsymmetric has serious consequences: without symmetry $vec(W)$ can lie in any direction in R^{n^2}, but with symmetry $vec(W)$ must lie in a lower dimensional subspace and this can have important consequences for C_{max} and l_{max}. Of course, symmetry is not the only condition that is needed for W to be a valid covariance matrix.

The schemes developed in the main paper are based on perturbing case-weights and explanatory variables, ideas which seem to be fairly well accepted. New schemes will have to be developed carefully, perhaps following the principles set forth by Weisberg (1983), and grudgingly implemented if we are to avoid the pitfall of overuse, as described by Dr Beckman. I expect that Dr Geisser's remarks on sample size are true qualitatively, but available evidence indicates that the quantitative impact of sample size should not be serious. In linear models, for example, the proposed methodology leads to standard graphical methods where the effect of sample size is not particularly worrisome.

Professor Cox appropriately warns that ω must be chosen so that the Euclidean metric is meaningful, although this present necessity can be removed with a more general development. The appropriate choice of a metric seems to be at the heart of the problems that Dr Lawrance encountered when considering the possibility of perturbing the transformation parameter λ. The Euclidean metric is meaningful when perturbing case-weights and I look forward to seeing Dr Lawrance's future results on the transformation $\lambda(\omega)$.

These general comments also apply to Dr Lawrance's query about the advantage to be had from introducing case-weight perturbations through a variance model that incorporates external variables Z. Such an approach can be useful provided that the variance model reflects firm knowledge and influence graph is defined in terms of case-weights, however they are modelled. The variance model will serve to concentrate our attention in relevant directions while defining the influence graph in terms of case-weights will insure that the Euclidean metric is reasonable.

I am pleased that Dr Pena found the likelihood displacement to provide sensible solutions for dependent as well as independent data. His distinction between a missing value approach and data deletion does seem to be useful.

Diagnostic value

I am not sure what kind of additional information Dr Lawrance requires to demonstrate the relevance of the individual elements of l_{max}. As mentioned in the main paper, these elements show how to introduce perturbations to achieve the largest local change in the likelihood displacement and thus indicate the relative importance of the individual elements of ω and provide important diagnostic information. I find the illustrations in the main paper to be compelling in this regard. Dr Lawrance's comments on the usefulness of considering the slope of an influence graph at $\omega = 0$ when perturbing case-weights are quite appropriate. Generally, the tangent plane in addition to (20) may be of value for studying the behaviour of an influence graph away from the non-perturbing value ω_0. The value of the tangent plane seems clear, particularly in view of the illustrations provided by Dr Lawrance. The value of (20) is under study. However, when considering case-weights, (20) will probably add little to what is already known. When perturbing a single case-weight, for example, Dr Critchley observes that the entire influence graph depends on only two quantities so that the curvature at $\omega = 0$ can recover only a function of what is already known to be relevant.

Turning to Professor Atkinson's challenge, it is clear that any reasonable method that involves plotting against x will find the clusters of 30 "good" and 20 "bad" cases in simulated simple linear regression data sets constructed according to the method in Rousseeuw (1984). In particular, application of the added variable plots discussed in Section 4.3 leads to identification of the clusters. (Statistical trivia: with Dr Lawrance's reference to individual coefficient plots, the number of distinct names in the literature on this graphical method is at least eight.) If we disallow plotting against x, as I expect Professor Atkinson had in mind, and instead attempt to identify the clusters by using l_{max} alone the situation becomes less clear. When perturbing case weights, the largest absolute elements of l_{max} always corresponded to the bad cases in data sets with 30 good cases and a smaller number (1 to 5 or 6) of bad cases, each data set being generated according to the method in Rousseeuw (1984). Thus, l_{max} might be expected to identify small clusters of outliers, but the method seems to break down as the fraction of outliers becomes large. Recall, however, that there is a clear difference between outliers and influential cases and that the proposed methods were designed to detect sensitivity rather than complicated structure in the data or lack of fit. This serves as a useful reminder that omnipotent diagnostic methods have yet to be developed. Similarly, Dr Critchley's comments on the balance between alternative influence methodologies are quite appropriate. I was fortunate to be in attendance at the meeting of the Multivariate Study Group which he mentions and I share his view on the tractability and utility of the influence curve in multivariate analysis. Nevertheless, when the likelihood is fundamental for inference, it should be studied directly. The influence curve takes a circuitous route, first concentrating on parameters and then on the selection of an appropriate norm. The likelihood seems to take a back seat in this approach.

Alternative approaches

Professor Loynes asks if the likelihood displacement is the right criterion. The answer to this question will surely depend on context and the statistical philosophy of the investigator. From a Bayesian perspective, I am sure that Professor Geisser would answer "no", preferring instead an approach based on comparing posterior or predictive densities. This point of view is important and was anticipated in the Discussion of the main paper. The essential change necessary to adapt the methods for application in Bayesian analyses consists of replacing the likelihood displacement with an alternative criterion. For example, an influence graph might be defined as $(\omega^{\mathrm{T}}, I(p_{\omega}, p_{\omega_0}))$ where I is a Kullback-Leibler divergence, as defined by Professor Geisser. When ω is a vector of case-weights, this approach will connect with the case-deletion methods of Johnson and Geisser (1983) in the same way that an influence graph based on the likelihood displacement connects with D_i. Some work along these lines has been completed by Michael Lavine at the University of Minnesota and the results seem promising.

In a likelihood framework, we should strive for a meaningful overall comparison of the postulated and perturbed likelihoods as ω varies in Ω. Such a comparison should be relevant in Bayesian analyses in as much as the likelihood is a key ingredient in the Bayesian paradigm. The likelihood displacement provides a summary of the postulated and perturbed likelihoods when interest centres on $\hat{\theta}$ and, as demonstrated in the main paper, it does yield useful data analytic information. The alternative formulative formulation sketched by Professor Cox is very interesting and I look forward to seeing additional details. Some of the questions posed by Professor Cox have been addressed in a recent Minnesota Ph.D. dissertation by Robert McCulloch.

An answer to Professor Atkinson's question on the relationship between model perturbations and the score test for the related elaboration of the model is given by Dr Tsai. Similarly, Dr Farebrother develops an interesting connection between the curvature under case-weight perturbations and a test statistic for heteroscedasticity. While an understanding of such relationships may be useful, I view perturbing the model and testing model elaborations as distinct diagnostic methodologies. In regression models, for example, deleting a single case to assess its influence is philosophically and practically different than performing an outlier test based on an elaborated model. Similarly, formal significance tests based on the null $(\omega = \omega_0)$ distributions of $C_{\max}$ and $l_{\max}$ seem largely irrelevant for the purpose of assessing influence, a view which agrees with that expressed by Dr Prescott but which may be opposed to the views expressed by Drs Critchley and Tsai. My comments on simulation apply when there is doubt about seriousness of potentially important patterns discovered in the data. For example, if any data analytic method leads to the discovery of apparent heteroscedasticity, simulation may be useful when there is concern that the result may be due to chance in combination with a vigorous inspection of the data.

Design

Dr Nachtsheim and Mr Ross both suggest the possibility of adapting portions of the proposed methodology to address problems in experimental design. The potential here seems great. Some connections between influence methods and experimental design are available in the literature. Box and Draper (1975), for example, ensure insensitivity to outliers by minimizing Σh_{ii}^2 through choice of design. Huber (1977, p. 37) first drew my attention to the idea that $1/h_{ii}$ might be viewed as "the effective number of observations entering into the determination of $\hat{y}_i$," and I look forward to reading Mr Ross' development of this idea, particularly in nonlinear regression where large intrinsic curvatures may cast doubt on the appropriateness of the local design matrix. Returning to analysis, Dr Nachtsheim's comments suggest the possibility of using local influence methods to assess the stability of a solution point x^* from an estimated response surface. I expect that this will lead to a nonlinear model and may overlap with the ideas of Mr Ross.

REFERENCES IN THE DISCUSSION

Atkinson, A. C. (1985) *Plots, Transformations and Regression*. Oxford: University Press.
——— (1986) *Masking Unmasked*. Paper submitted for publication.
Box, G. E. P. and Draper, N. R. (1975) Robust designs. *Biometrika*, **62**, 347–352.
Breusch, T. S. and Pagan, A. R. (1979) A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294.
Chatterjee, S. and Hadi, A. S. (1984) Impact of simultaneous deletion of a variable and an observation on a linear regression equation. Unpublished manuscript.
Cox, D. R. (1961) Tests of separate families of hypotheses. In *Proc. 4th Berkeley Symp.*, **1**, 105–123.
Cox, R. D. and Hinkley, D. V. (1974) *Theoretical Statistics*. London: Chapman and Hall.
Geisser, S. (1982) Aspects of the predictive and estimative approaches in the determination of probabilities (with Discussion). *Biometrics Suppl.*, 75–93.
Critchley, F. (1985) Influence in principal components analysis. *Biometrika*, **72**, 627–636.
Hawkins, D. M., Bradu, D. and Kass, G. V. (1984) Location of several outliers in multiple regression data using elemental sets. *Technometrics*, **26**, 197–208.
Huber, P. (1977) Robust statistical procedures. Regional Conference Series in Applied Mathematics, No. 27. Philadelphia: SIAM.
Loynes, R. M. (1983) The influence of individual observations on subset selection. Contributed paper in 44th Session of the ISI, pp. 901–903.
Pena, D. (1984) Influential observations in time series. Mathematics Research Center Technical Report No. 2712, University of Wisconsin at Madison.
——— (1986) Measuring the importance of outliers in ARIMA models. In *New Perspectives in Theoretical and Applied Statistics* (M. L. Puri *et al.*, eds). New York: Wiley (in press).
Ross, G. J. S. (1986) Effective replication and the prediction range. Paper submitted for publication.
Rousseeuw, P. J. (1984) Least median of squares regression. *J. Amer. Statist. Ass.*, 79, 871–880.
Weisberg, S. (1983) In discussion of "Developments in linear regression" by R. R. Hocking. *Technometrics*, **25**, 240–244.
——— (1985) *Applied Linear Regression*, 2nd ed. New York: Wiley.