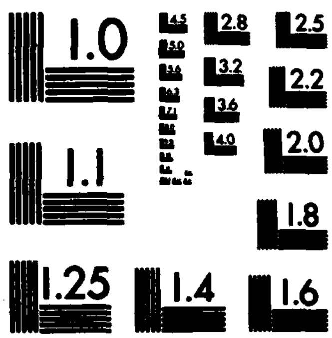
AD-A143 002 ASSESSMENT OF LOCAL INFLUENCE(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER R D COOK MAY 84 MRC-TSR-2694 DAAG29-80-C-0041 1/1

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A143 002

MRC Technical Summary Report # 2694

ASSESSMENT OF LOCAL INFLUENCE

R. Dennis Cook

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

May 1984

(Received March 16, 1984)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ASSESSMENT OF LOCAL INFLUENCE

R. Dennis Cook

Technical Summary Report #2694

May 1984

ABSTRACT

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 a particular class of models, and it seems to provide a relatively simple, unified approach for handling a variety of problems.

AMS (MOS) Subject Classification: 62-07

Key Words: Collinearity, curvature, diagnostics, influence graphs, influential observations.

Work Unit Number 4 - Statistics and Probability

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

The statistical analysis of a collection of data is usually based on a specified model, a mathematical formula describing the behavior of the data up to a few unknown parameters which are to be estimated from the data. The specification of a model often involves making assumptions that may have little prior support. In such situations it becomes necessary to understand if important results of an analysis are strongly dependent on the validity of assumptions underlying the model. The purpose of this paper is to provide a method for assessing the influence of minor model perturbations on the results of an analysis.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

ASSESSMENT OF LOCAL INFLUENCE

R. Dennis Cook

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 the 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 (Barnard 1980) and our ignorance of the precise model will do no harm.

Although an assessment of the influence of a model perturbation is generally considered to be important, few general methods are available for carrying out such an assessment in contexts other than normal linear regression, and much of the past 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). Moolgavkar, Lustbader and Venzon (1984) give a number of useful results on

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

case deletion diagnostics for general exponential families, and Lustbader and Moolgavkar (1984) investigate the change in the score test on deletion of cases. Oman (1984) develops measures for assessing the influence of individual cases in calibration problems.

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).

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.

This paper presents a general method for assessing the local influence of minor perturbations of a statistical model. The method relies on a well-defined likelihood and certain elementary ideas from differential geometry, and seems to provide a relatively simple, unified approach for handling a variety of problems. 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 quantity 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 section 4 and section 5 contains concluding comments.

2. INFLUENCE GRAPHS

2.1 Motivation

Consider the standard linear regression model

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

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. 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),

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

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. For example, since σ^2 is known, $pD_i = (\text{DFFITS})^2$ from Belsley, Kuh and Welsch (1980).

The statistic D_i in (2) 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,

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

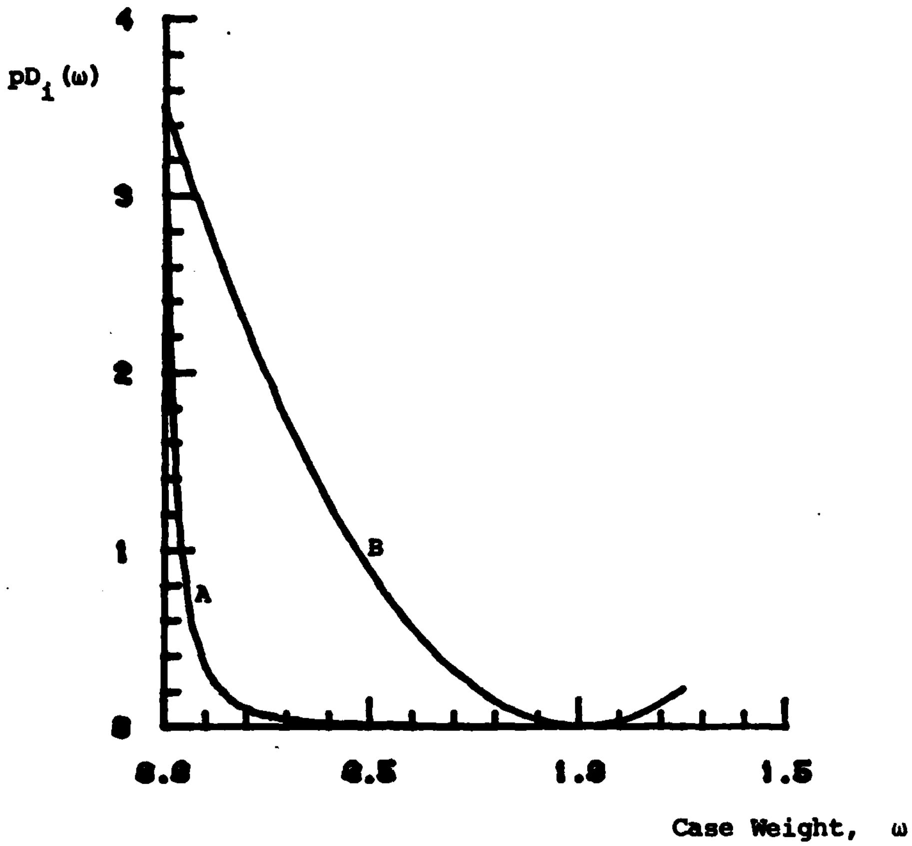
where $\hat{Y}_\omega$ is the vector of fitted values obtained when the i-th case has weight ω and the remaining cases have weight 1. Of course, as $\omega \to 0$, $\mathrm{var}(\varepsilon_i) \to \infty$ and $D_i = D_i(0)$. If $D_i(\omega)$ is large then the stipulation that the i-th case has variance 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 case, however. Figure 1 gives plots of $pD_i(\omega)$ versus ω for two possible cases A and B from model (1). The details behind Figure 1 will be presented later. For now we note that the analysis is clearly more sensitive to alterations in the weight attached to case B since $D_B(\omega) - D_A(\omega) > 0$ and for some ω this difference is substantial. We must have $D_A(1) = D_B(1)$, of course. However, the fact that $D_A(0) = D_B(0)$ means 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 behavior 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 based on the following relationship between $D_i(\omega)$ and the log likelihood $L(\beta)$ for model (1),

$$\begin{aligned} pD_i(\omega) &= [\|Y - \hat{Y}_\omega\|^2 - \|Y - \hat{Y}\|^2]/\sigma^2 \\ &= 2[L(\hat{\beta}) - L(\hat{\beta}_\omega)] \end{aligned} \tag{4}$$

where $\hat{\beta} = \hat{\beta}_{\omega=1}$ and $\hat{\beta}_\omega$ is the maximum likelihood estimator of β when the i-th case has weight ω. This relationship was pointed out in the special



<u>**Figure 1**</u>

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

case $\omega = 0$ by Cook and Weisberg (1982, Chapter 5).

2.2 Development

For a given statistical problem, let θ denote the $p \times 1$ vector of unknown parameters and let $L(\theta)$ denote the log likelihood corresponding to the postulated model. 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 the examples of section 4, ω must be chosen carefully so that the application is sensible. For now we assume this choice to have been made.

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

To assess the influence of varying ω throughout Ω, we initially consider the <u>likelihood difference</u>

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

In a particular problem, specific characteristics of $\{\hat{\theta}_\omega | \omega \in \Omega\}$ 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)

$$\{\theta | 2[L(\hat{\theta}) - L(\theta)] \le \chi^2_\alpha(p)\} \quad .$$

Here, $\chi^2_\alpha(p)$ 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

$$\alpha(\omega) = \binom{LD(\omega)}{\omega} \quad . \tag{6}$$

In differential geometry a surface of this form is frequently called a Monge patch. 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 retrospect, Figure 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 if of interest. In this situation the analog of (6) is

$$\alpha_1(\omega) = \binom{LD_1(\omega)}{\omega} \tag{7}$$

where

$$LD_1(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_{1\omega}, g(\hat{\theta}_{1\omega}))] \quad ,$$

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

$$\{\theta_1 | 2[L(\hat{\theta}) - L(\theta_1, g(\theta_1))] \le \chi^2_\alpha(p_1)\} \quad .$$

The influence graph defined in (7) reflects a special interest. On the other hand, a somewhat different but related perspective leads to the influence graph

$$\alpha'(\omega) = \binom{LD'(\omega)}{\omega} \tag{8}$$

where

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

In the construction of this graph, the moving frame of reference $L(\theta|\omega)$ 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. Both α and α' may be useful for assessing influence.

Ideally, we would like a complete influence graph, such as those displayed in Figure 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 behavior of an influence graph over all ω in Ω, are generally much more difficult to construct in practice than local measures which characterize behavior in a neighborhood of a selected ω, say ω^*.

In normal linear regression, the various influence diagnostics that rely on case deletion (D_i for example) can be regarded as local measures since they are designed to measure influence on various "corners" of $\Omega = (0,1)^n$, where n is the sample size. However, from Figure 1 and the discussion of section 2.1, it is clear that the behavior of an influence graph around $\omega^* = \omega_0 = 1$ may be as relevant as the behavior at the corners of Ω.

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

3. LOCAL INFLUENCE

The behavior of an influence graph around ω_0 is accurately reflected by the geometric normal curvature at ω_0. For $q = 1$ this curvature can be viewed as the inverse of the radius of the circle which best approximates an influence graph at ω_0, or as the rate of change of the angle between the tangent vector at ω_0 and the horizontal axis. This curvature easily distinguishes between the two influence graphs shown in Figure 1: the curvature is $2(.05)^2 pD_i$ for case A and $2(.99)^2 pD_i$ for case B.

In this section we use normal curvatures to characterize the behavior of an influence graph around ω_0. The normal curvature of a surface ($\alpha(\omega)$ in this application) should be discussed in any first text on differential geometry. Sufficient 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 will initially develop the normal curvatures at an arbitrary ω^*, although our primary interest is in the case $\omega^* = \omega_0$. Curvatures at points other than ω_0 may be of some value in assessing the global behavior of an influence graph.

3.1 Curvatures for $\alpha(\omega)$

For $q > 1$ consider a straight line in Ω passing through ω^*. Such a line can be represented by

$$\omega(a) = \omega^* + a\ell \qquad (9)$$

where $a \in R^1$ and ℓ is a fixed nonzero vector in R^q. This line generates a <u>lifted line</u> on the influence graph $\alpha(\omega)$ passing through $\alpha(\omega^*)$. Each direction ℓ specifies such a lifted line and for each lifted line we can imagine a normal curvature as discussed in connection with Figure 1. For a given direction $\boldsymbol{\ell}$, let $C_{\ell}(\boldsymbol{\omega}^*)$ denote the normal curvature of $\alpha(\boldsymbol{\omega})$ at $\boldsymbol{\omega}^*$.

Let $\mathbf{V}$ denote the $(q+1) \times q$ matrix with elements $\partial\alpha_i(\omega)/\partial\omega_j$, $i = 1,2,\ldots,q+1$, $j = 1,2,\ldots,q$. Here α_i is the i-th component of α and all derivatives are evaluated at $\boldsymbol{\omega}^*$. Further, let $\mathbf{w}_{jk}$ denote the $(q+1) \times 1$ vector with elements $\partial^2\alpha_i(\boldsymbol{\omega})/\partial\omega_j\partial\omega_k$, $i = 1,2,\ldots,q+1$. Then the velocity and acceleration vectors in the direction $\boldsymbol{\ell}$ are respectively

$$\dot{\boldsymbol{\alpha}}_{\ell} = \mathbf{V}\boldsymbol{\ell} \tag{10}$$

and

$$\ddot{\boldsymbol{\alpha}}_{\ell} = \sum_j \sum_k \mathbf{w}_{jk}\ell_j\ell_k \tag{11}$$

where $\boldsymbol{\ell} = (\ell_k)$. The normal curvature $C_{\ell}(\boldsymbol{\omega}^*)$ can now be written as

$$C_{\ell}(\boldsymbol{\omega}^*) = \|\mathbf{P}_V'\ddot{\boldsymbol{\alpha}}_{\ell}\|/\|\dot{\boldsymbol{\alpha}}_{\ell}\|^2 \tag{12}$$

where $\mathbf{P}_V'$ is the projection operator for the null space of $\mathbf{V}$. Carrying out the operations indicated in (12) we find

$$C_{\ell}(\boldsymbol{\omega}^*) = \frac{2|\boldsymbol{\ell}^T\ddot{\mathbf{F}}\boldsymbol{\ell}|}{(1+\|\dot{\mathbf{F}}\|^2)^{1/2}\,\boldsymbol{\ell}^T(\mathbf{I}+\dot{\mathbf{F}}\dot{\mathbf{F}}^T)\boldsymbol{\ell}} \tag{13}$$

where $\dot{\mathbf{F}}$ is the $q \times 1$ vector with elements $2\partial L(\hat{\theta}_{\omega})/\partial\omega_j$, $j = 1,2,\ldots,q$, and $\ddot{\mathbf{F}}$ is the $q \times q$ matrix with elements $\partial^2 L(\hat{\theta}_{\omega})/\partial\omega_k\partial\omega_j$, $j,k = 1,2,\ldots,q$. Since $\dot{\mathbf{F}} = 0$ at $\boldsymbol{\omega}^* = \boldsymbol{\omega}_0$,

$$C_{\ell} = C_{\ell}(\boldsymbol{\omega}_0) = 2\,\frac{|\boldsymbol{\ell}^T\ddot{\mathbf{F}}\boldsymbol{\ell}|}{\boldsymbol{\ell}^T\boldsymbol{\ell}} \; . \tag{14}$$

This simple form appears since $\boldsymbol{\omega}_0$ is a global minimum and thus the velocity and acceleration vectors are orthogonal; that is, every acceleration vector is orthogonal to the tangent plane at $\boldsymbol{\omega}_0$. Unless indicated otherwise, we take $\boldsymbol{\omega}^* = \boldsymbol{\omega}_0$ in the remainder of this section.

For (14) to be useful we should have a straightforward way to evaluate $\ddot{F}$. Using the chain rule for differentiation, it is not difficult to verify that

$$\ddot{F} = J^T \ddot{L} J \tag{15}$$

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 i-th component of $\hat{\theta}_\omega$. Next, to evaluate J we use the fact that

$$\left.\frac{\partial L(\theta|\omega)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_\omega} = 0 \tag{16}$$

for $j = 1,2,\ldots,p$ and all ω in Ω. Differentiating both sides of (16) with respect to ω and evaluating at ω_0, it follows that

$$J = -(\ddot{L})^{-1}\Delta \tag{17}$$

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

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

evaluated at $\theta = \hat{\theta}$ and $\omega = \omega_0$, $i = 1,2,\ldots,p$, $j = 1,2,\ldots,q$. Substituting (17) into (15) we obtain

$$\ddot{F} = \Delta^T(\ddot{L})^{-1}\Delta \tag{18}$$

and therefore

$$C_\ell = 2|\ell^T\Delta^T(\ddot{L})^{-1}\Delta\ell|/\ell^T\ell \quad . \tag{19}$$

The individual components of (19) are usually straightforward to obtain once the perturbation scheme has been defined.

There are several obvious ways in which (19) might be used to study $\alpha(\omega)$ in practice. The extremes $C_{max} = \max_\ell C_\ell$ and $C_{min} = \min_\ell C_\ell$ are two useful options. Of course, C_{max} and C_{min} correspond to the maximum and minimum absolute eigenvalues of $\ddot{F}$ in (18). The eigenvectors associated with these eigenvalues can be used to set the directions in (9) which can then be used to construct plots, similar to those in Figure 1, of the lifted line $\alpha(\omega(a))$. Similarly, the eigenvectors associated with intermediate eigenvalues can be used to investigate the behavior of $\alpha(\omega)$ in directions corresponding to less extreme curvatures.

Another option is the average curvature $\bar{C}$ obtained by averaging (19) with respect to a uniform distribution on the surface of the unit sphere in q dimensions: Let S_q denote the surface area of a q-dimensional unit sphere and assume that $\ddot{F}$ is negative semidefinite. Then

$$\begin{aligned}\bar{C} &= S_q^{-1} \int_{\|\ell\|=1} C_\ell \, dS \\ &= 2(q(q+2))^{-1}[\mathrm{tr}^2(\ddot{F}) + 2\mathrm{tr}(\ddot{F}^2)]\end{aligned} \tag{20}$$

where $\ddot{F}$ is as defined in (18).

Finally, a relationship between the curvature C_ℓ and LD can be obtained by expanding $LD(\omega(a)) = LD(\omega_0 + a\ell)$ as a function of a:

$$LD(\omega_0 + a\ell) = a^2 C_\ell/2 + o(a^2)$$

where $\|\ell\| = 1$. This representation provides a useful alternative interpretation of C_ℓ.

3.2 Other Influence Graphs

In this subsection we investigate the influence graphs $\alpha_1(\omega)$ and $\alpha'(\omega)$ defined in equations (7) and (8), respectively.

By replacing α with α' in the development that led to (10) and (11) and using the chain rule for differentiation, it is not difficult to verify that the velocity and acceleration vectors at ω_0 for α' are the same as those for α. It follows that α and α' have identical curvatures at ω_0, although the two influence graphs can differe considerably in global behavior. Since we are primarily interested in assessing local influence around ω_0, α

and the analogous graph α_1 for subsets will be used in the remainder of this paper.

To develop the curvatures for $\alpha_1(\omega)$, we first note that the development leading to (13) is valid with $L(\hat{\theta}_\omega)$ replaced by $L[\gamma(\hat{\theta}_{1\omega})]$ where $\gamma^T = (\hat{\theta}^T_{1\omega}, g^T(\hat{\theta}_{1\omega}))$ and g is defined following (7). It follows that (13) can be adapted for α_1 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, (14) 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

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

where $\ddot{L}$ is as defined following (15) 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

$$K = \binom{I}{K_2} K_1 \quad . \tag{22}$$

Note that K_1 is just the matrix consisting of the first p_1 rows of J defined in (17). To evaluate K_2 we make use of the fact that

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

where g_i is the i-th 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

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

Finally, combining (14), (21) and (24) with the form of K_1 mentioned above, we obtain the normal curvature for subsets,

$$C_\ell = 2|\ell^T\Delta^T(\ddot{L}^{-1} - B_{22})\Delta\,\ell|/\ell^T\ell \tag{26}$$

where

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

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

4. APPLICATIONS

In this section we describe several possible applications of the ideas described in the previous sections. Our intent is to illustrate the range of possible use rather than to develop any particular application in full detail. The refinements of the individual applications and adaptations for applications not discussed here should be straightforward.

4.1 Case Weights in Normal Linear Regression

Let $\boldsymbol{\omega}$ denote the $n \times 1$ vector of case weights for the regression model (1) and again assume that σ^2 is known. The modified log likelihood is

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

where ω_i and y_i are the i-th components of $\boldsymbol{\omega}$ and $\mathbf{Y}$, respectively, and x_i^T is the i-th row of $\mathbf{X}$. Differentiating (27) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$, and evaluating at $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\omega}_0 = 1$, we find

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

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

$$\begin{aligned} C_\ell &= 2|\boldsymbol{\ell}^T\Delta^T(\ddot{L})^{-1}\Delta\,\boldsymbol{\ell}|/\boldsymbol{\ell}^T\boldsymbol{\ell} \\ &= 2\,\boldsymbol{\ell}^T D(\mathbf{e})P_x D(\mathbf{e})\boldsymbol{\ell}/\boldsymbol{\ell}^T\boldsymbol{\ell}\,\sigma^2 \end{aligned} \tag{29}$$

where $P_x = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the projection operator for the column space of $\mathbf{X}$.

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

$$\Delta = \begin{pmatrix} \mathbf{X}^T D(\mathbf{e})/\hat{\sigma}^2 \\ \mathbf{e}_2^T/2\hat{\sigma}^4 \end{pmatrix} \tag{30}$$

where $\hat{\sigma}^2$ is the maximum likelihood estimator of σ^2 and $\mathbf{e}_2 = (e_i^2)$. Since

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

we have the analogous result for θ,

$$C_\ell = 2\ell^T[D(e)P_x D(e) + e_2 e_2^T/2n\hat{\sigma}^2]\ell/\ell^T\ell\hat{\sigma}^2 \quad . \tag{31}$$

If only β is of interest, the above results in combination with (26) show that the curvature is given by (29) with σ^2 replaced with $\hat{\sigma}^2$. The following three special cases should furnish some insight into the behavior of the curvature in the situation.

First, for a simple random sample $\ddot{F}$ has only one nonzero eigenvalue C_{max} with corresponding eigenvector $\ell_{max} = e/\|e\|$. Thus, the local changes in $\hat{\beta}$ will be zero when $\omega_0 = 1$ is perturbed in any direction ℓ' that is orthogonal to e. This is easily confirmed by direct calculation: $\hat{\beta}_\omega = \hat{\beta}$ for $\omega = 1 + a\ell'$. 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. Experience has shown that curvatures smaller than 2 can generally be neglected while curvatures much larger than 2 suggest further investigation. 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 is indicated by a relatively large element of ℓ_{max}. 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. This idea will be illustrated further in later examples.

Second, the curvature for simple linear regression through the origin with $X = (x_i)$ is

$$C_{max} = 2\Sigma(x_i e_i)^2/\hat{\sigma}^2\Sigma x_i^2$$

which occurs in the direction $\ell = (x_i e_i)$. This curvature is bounded above by $2n$ and will tend to be large when the residuals attached to remote x_i's are relatively large.

Finally, the curvature for the influence graph obtained by modifying the weight attached to a single case, say the i-th, is

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

where h_{ij} is the (i,j)-th element of P_x. Form (32) was used to construct Figure 1. For case A, $h_{ii} = .95$ and for case B $h_{ii} = .01$. 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. 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 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 Figure 1, the variance of case A might be set at 10 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.

We use the drill data as given in Cook and Weisberg (1982, p. 148) for a numerical illustration of the use of (29) with $\sigma^2 = \hat{\sigma}^2$. The data consist of $n = 31$ observations on the axial load on a drill bit under condition set by three design variables. We use the full second-order response surface model so that there are 10 location parameters.

The maximum curvature $C_{max} = 7.42$ occurs in approximately the direction $\ell = (\ell_i)$ with $\ell_5 = .61$, $\ell_6 = -.16$, $\ell_9 = -1$, $\ell_{26} = .38$, $\ell_{28} = -.15$, $\ell_{31} = .21$ and $\ell_i = 0$ otherwise. Five of the six cases with nonnegligible ℓ_i's correspond to the cases with the five largest ordinary residuals. Using ℓ as given here and (9) with $\omega^* = 1$, we have displayed $LD(\omega(a))$ in the direction of the maximum curvature in Figure 2. Clearly, appropriately

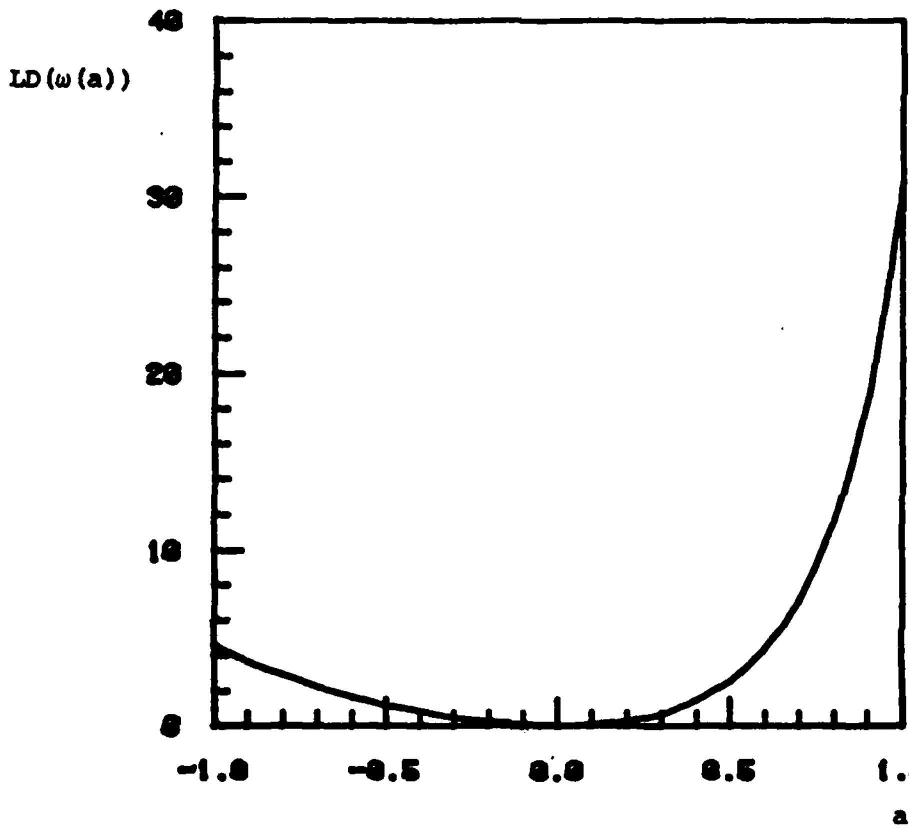


Figure 2

Plot of the likelihood difference LD in the direction of maximum curvature for the drill data.

modifying a few selected weights can substantially change $\hat{\beta}$ as measured using L. For example, at $a = .9$ the approximate weights are $\omega_5 = 1.55$, $\omega_6 = .86$, $\omega_9 = .1$, $\omega_{26} = 1.34$, $\omega_{28} = .86$, $\omega_{31} = 1.19$, and $\omega_i = 1$ otherwise. For these weights $LD(\boldsymbol{\omega}) = 18.7$ so that $\hat{\beta}_\omega$ will lie on the edge of a 96% confidence for β.

Further information on $\alpha(\boldsymbol{\omega})$ could 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.

4.2 Correlations in Normal Linear Regression

In linear regression the assumption of uncorrelated errors is often difficult to justify. In such situations it may be important to ask if the analysis is sensitive to deviations from this assumption.

Let $\boldsymbol{\omega}$ now denote an $n(n-1)/2 \times 1$ vector of error correlations indexed by (i,j), $i < j$, and let $\Sigma_\omega = \mathrm{var}(\varepsilon)$. The (i,j)-th element of Σ is $\omega_{ij}\sigma^2$ for $i < j$ and σ^2 for $i = j$. For convenience we assume σ^2 to be known.

The log likelihood for the perturbed model can now be written as

$$L(\beta|\boldsymbol{\omega}) = -\frac{1}{2}\log|\Sigma_\omega| - \frac{1}{2}(Y-X\beta)^T\Sigma_\omega^{-1}(Y-X\beta) \quad . \tag{33}$$

Differentiating (33) with respect to β and $\boldsymbol{\omega}_{ij}$, and evaluating at $\hat{\beta}$ and $\boldsymbol{\omega}_0 = 0$, it is not difficult to verify that the (i,j)-th column of Δ is $-(x_ie_j + x_je_i)/\sigma^2$ where x_i^T is the i-th row of X. It follows that the $((i,j),(k,m))$-th element of $\ddot{F}$ is

$$h_{ki}e_me_j + h_{mi}e_ke_j + h_{kj}e_me_i + h_{mj}e_ke_i \tag{34}$$

where h_{ij} is defined following (32).

In this application $\ddot{F}$ is an $n(n-1)/2 \times n(n-1)/2$ matrix. The eigenvalues of $\ddot{F}$ can be determined by replacing $(X^TX)^{-1}$ with $(X^TX)^{-1/2}(X^TX)^{-1/2}$ and using the fact that the eigenvalues of A^TA are the same as those of AA^T which will be a manageable $p \times p$ matrix in this case. The eigenvectors of $\ddot{F}$ may be more of a problem but we expect that by using the structure of the experiment ω can be restricted to a subspace in many applications.

When only a single correlation is considered, $\ddot{F}$ becomes a scalar and the corresponding curvature is

$$2(h_{ii}e_j^2 + 2h_{ij}e_ie_j + h_{jj}e_i^2)/\sigma^2 \tag{35}$$

where (i,j) indexes the perturbed correlation.

4.3 Explanatory Variables in Normal Linear Regression

It is well known that perturbations, within the limits of measurement error, of the explanatory variables in linear regression can seriously influence the results of a least squares analysis, particularly when collinearity is present. To handle this situation in the present context, let s_j denote the standard deviation of the measurement error associated with the j-th explanatory variable. 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 $\hat{\sigma}^2$.

The perturbed log likelihood $L(\beta|\omega)$ is constructed from (1) with X replaced by

$$X_\omega = X + W\,S \tag{36}$$

where $W = (\omega_{ij})$ is an $n \times p$ matrix of perturbations and $S = \text{diag}(s_1,\ldots,s_p)$. 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 $\partial^2L(\theta|\omega)/\partial\theta_i\partial\omega_{jk}$, $i = 1,2,\ldots,p$, $j = 1,2,\ldots,n$. Then

$$A_k = s_k(b_k e^T - \hat{\beta}_k X^T)/\sigma^2 \tag{37}$$

where b_k is the k-th standard basis vector for R^p.

In this application, $\ddot{F}$ is a potentially large $np \times np$ matrix and determining the eigenvalues of $\ddot{F}$ may be an unpleasant task. However, using the method described following (34), it can be shown that the nonzero eigenvalues of $\ddot{F}$ are

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

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

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

When only the k-th column of X is perturbed, $s_j = 0$ for $j \neq k$ and (39) can be written as

$$C_{max} = 2s_k^2(e^T e/RSS_k + \hat{\beta}_k^2)/\sigma^2 \tag{40}$$

where RSS_k is the residual sum of squares from the regression of the k-th column of X on the remaining columns.

For a first numerical illustration we use the perturbation scheme for the Longley data that is described in Weisberg (1980, p. 70-72). For this setup, which consists essentially of using the s_k's to represent round-off errors in the last digit of the explanatory variables, evaluating (39) with $\sigma^2 = \hat{\sigma}^2$ gives $C_{max} = .18$. 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 such variation does not reflect important changes in the estimates when judged against the log likelihood.

For a second numerical illustration we use the rat data from Weisberg (1980, p. 110-113). This data set consists of 19 cases and 4 explanatory variables, X_0 = constant, X_1 = body weight, X_2 = liver weight and X_3 = relative dose. The perturbation schemes we consider are characterized by S =

$\text{diag}(s_0,s_1,s_2,s_3) = \text{diag}(0,1,0,s_3)$. For $s_3 = .01(.01).04$ 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 $s_3 = .01$ is relatively small while the curvatures for the remaining s_3's indicate that a minor perturbation within the limits of these measurement errors may lead to drastic changes in $\hat{\beta}$. 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 Figure 3 for $s_3 = .03$. Interestingly, the largest element in the eigenvector for C_{max} always corresponds to the relative dose for case 3 which is the anomolous care identified by Weisberg (1980). The scale on the x-axis in Figure 3 is the amount that the relative dose for case 3 is perturbed in units of s_3. Thus, for example, $a = .5$ indicates $as_3 = .015$ was added to the relative dose for case 3. Clearly, the influence of perturbations for $s_3 = .03$ is very strong. In particular, the value of LD at $a = -.5$ shows that $\hat{\beta}_\omega$ will be moved outside of a 95% confidence region for $\hat{\beta}$ when perturbing each element of X by an amount that is no greater than 1/2 of the respective standard deviations s_j. The nonmonotonic behavior in Figure 3 arises since globally the lifted line $\alpha(\omega(a))$ need not correspond to a path of monotonic descent.

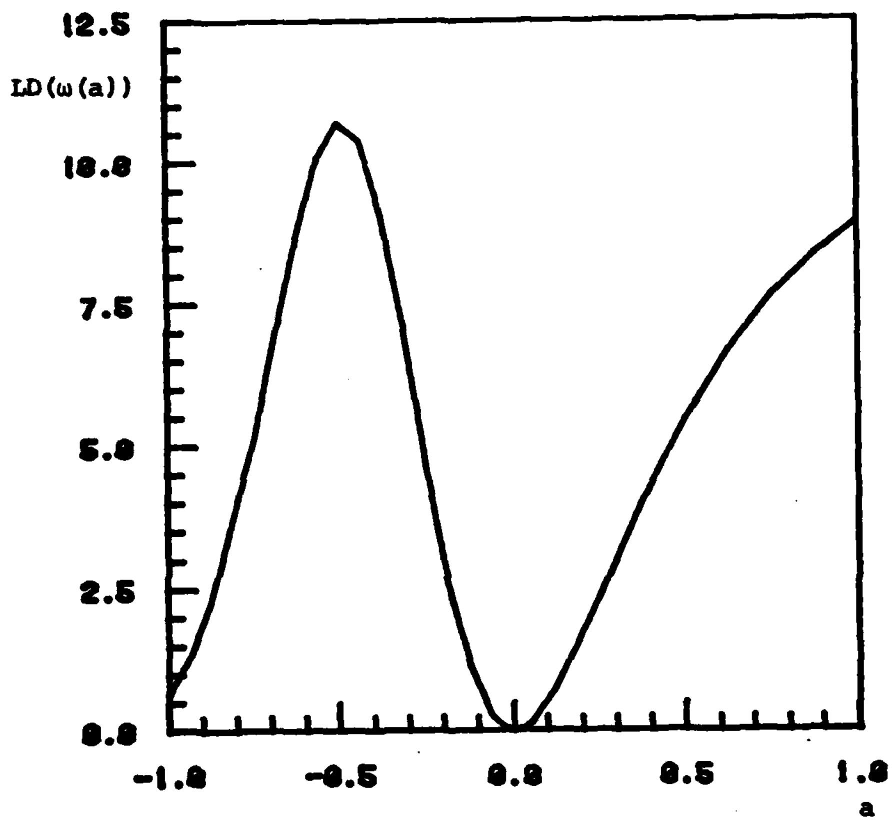
Kelly (1984) investigates various extensions of D_i for use in the errors in variables problems.

4.4 Case Weights in Curved Exponential Families

In this and the following subsection, 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 representation

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



<u>Figure 3</u>

Plot of the likelihood difference LD in the direction of maximum curvature for the rat data.

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

$$L(\theta) = \sum_i (y_i\eta_i - \psi_i(\eta_i)) \quad . \tag{41}$$

Next, the log likelihood obtained by attaching a weight ω_i to the i-th case case be written as simply

$$L(\theta|\omega) = \sum_i \omega_i(y_i\eta_i - \psi_i(\eta_i)) \quad . \tag{42}$$

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

Let
$$\eta = (\eta_i) \quad ,$$
$$\dot{\eta} = \partial\eta/\partial\theta \qquad (n \times p)$$
$$\eta_i = \partial^2\eta_i(\theta)/\partial\theta^2 \qquad (p \times p)$$

and

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

where all derivatives are evaluated at $\hat{\theta}$, the maximum likelihood estimator of θ. Using the results of section 3.1 it is not difficult to verify that

$$\ddot{F} = D_r\dot{\eta}[\sum_i r_i\eta_i - \dot{\eta}^T\Psi\,\dot{\eta}]^{-1}\dot{\eta}^T D_r \tag{43}$$

where D_r is an $n \times n$ diagonal matrix with the score residuals

$$r_i = (y_i - \partial\psi_i/\partial\eta_i) \tag{44}$$

as the diagonal entries.

Many generalized linear models are special cases of (41) with $\eta_i(\theta) = K(x_i^T\theta)$ where K is the link function. Further,

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

and

$$\ddot{\eta} = \ddot{K}_i\; x_i x_i^T \quad ,$$

where $\dot{K}_i$ and $\ddot{K}_i$ are the first and second derivatives of K evaluated at

$x_i^T\hat{\theta}$, respectively. In particular, $\ddot{\eta} = 0$ when the cononical link is used.

4.5 Explanatory Variables in Generalized Linear Models

Consider the log likelihood (41) with $\eta_i = K(x_i^T\theta)$. The log likelihood $L(\theta|\omega)$ obtained after the explanatory variables have been perturbed by an amount ω can be constructed by replacing x_i^T with $x_{i\omega}^T$, the i-th row of X_ω defined in (36). From this it can be verified that Δ has the same structure as described in section 4.3 and that

$$\Delta_k = s_k\{b_k r^T \operatorname{diag}(\dot{K}_i) + \hat{\theta}_k X^T \operatorname{diag}(r_i\ddot{K}_i - \ddot{\psi}_i\dot{K}_i^2)\} \tag{45}$$

where b_k is defined following (37) and $r = (r_i)$. Further, the observed information matrix is

$$-\ddot{L} = -X^T \operatorname{diag}(r_i\ddot{K}_i - \ddot{\psi}_i\dot{K}_i^2)X \quad . \tag{46}$$

For a concrete illustration we use the leukemia data as reported in Cook and Weisberg (1982, p. 179). Here, a patients survival time in weeks y_i, $i = 1,2,\ldots,17$, is assumed to follows 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 i-th patient.

The log likelihood for the original data is of the form given in (41) 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

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

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

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

$$\ddot{\psi}_i = \dot{\eta}_i^{-2} = \exp[2(\hat{\theta}_1 + \hat{\theta}_2 x_i)] = \widehat{\operatorname{var}}(Y_i)$$

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

and thus that

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

These calculations along with (45) and (46) can now be used to construct $\ddot{F}$ as given in (18).

To assess the influence of measurement errors associated with WBC we perturb $x = \log_{10}(WBC)$ rather than WBC itself. This implies that the measurement errors associated with WBC are multiplicative rather than additive and that the standard deviation of WBC_i is proportional to $E(WBC_i)$. Both implications seem reasonable.

The maximum curvature for this perturbation scheme is $C_{max} = 17.014s_x^2$ where s_x is the standard deviation of the measurement error associated with $x = \log_{10}WBC$. Clearly, the measurement error must be substantial for the local influence to be large. The eigenvector associated with C_{max} lies substantially in the direction of case 17: The largest element of this vector corresponds to case 17 and is about 7 times larger than the second largest element. Thus, although the local 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.

5. DISCUSSION

For a complete understanding of the influence of a particular perturbation scheme it is probably necessary to know the full behavior of the selected influence graph. We have found the central methodology discussed in this paper to be a useful and relatively simple way of characterizing the local behavior of an influence graph around ω_0. The maximum curvature C_{max} seems to be a reliable indicator of extreme local behavior, and the plot of the corresponding lifted line provides a reasonably easy way to confine such indications. Also, the methodology can be easily adapted to handle loss functions other than LD or LD_1. In a Bayesian analysis, for example, LD might be replaced with a loss function that reflects the sensitivity of the analysis to perturbations in the prior parameters.

As demonstrated in Figure 1, gross errors can have a substantial influence on an analysis even when the curvatures are small. To understand the consequences of gross errors it is necessary to characterize the behavior of an influence graph near the boundaries of Ω, as in case deletion diagnostics attempt to do. Generally, this might be done by simply evaluating an influence graph at various points near the boundary of Ω. However, our experience has shown that a plot of the lifted line associated with C_{max} may indicate the seriousness of gross errors, even when C_{max} is small. This happens, for example, in the numerical illustrations of sections 4.3 and 4.5, and may be expected whenever the influence graph is strongly quadratic.

ACKNOWLEDGEMENT

The author wishes to thank D. Pena and S. Weisberg for their comments on an earlier version of this report.

REFERENCES

Andrews, D. F. and Pregibon, D. (1978), "Finding the Outliers that Matter," Journal of the Royal Statistical Society, Ser. B, 40, 85-93.

Atkinson, A. C. (1982), "Regression Diagnostics, Transformations and Constructed Variables," Journal of the Royal Statistical Society, Ser. B, 44, 1-36.

Barnard, G. A. (1980), "Discussion of Professor Box's Paper," Journal of the Royal Statistical Society, Ser. B, 404-406.

Bates, D. M. and Watts, D. G. (1980), "Relative Curvature Measures of Non-linearity," Journal of the Royal Statistical Society, Ser. B, 42, 1-25.

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

Cook, R. D. (1977), "Detection of Influential Observations in Linear Regression," Technometrics, 19, 15-18.

Cook, R. D. (1979), "Influential Observations in Linear Regression," Journal of the American Statistical Association, 74, 169-174.

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

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

Johnson W. and Geisser, S. (1983), "A Predictive View of the Detection and Characterization of Influential Observations in Regression Analysis," Journal of the American Statistical Association, 78, 137-144.

Kelly, G. (1984), "The Influence Function in the Errors in Variables Problems," Annals of Statistics, 12, 87-100.

Lustbader, E. D. and Moolgavkar, S. H. (1984), A Diagnostic Statistic for the Score Test," Journal of the American Statistical Association, to appear.

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," Annals of Statistics, to appear.

Oman, S. D. (1984), "Analyzing Residuals in Calibration Problems," Technometics, to appear.

Pregibon, D. (1981), "Logistic Regression Diagnostics," Annals of Statistics, 9, 705-724.

Welsch, R. E. (1982), "Influence Functions and Regression Diagnostics," in Launer, R. L. and Siegel, A. F. (Eds), Modern Data Analysis, New York: Academic Press.

RDC/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER # 2694	2. GOVT ACCESSION NO. AD-A143002
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Assessment of Local Influence	5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. Dennis Cook	8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709	12. REPORT DATE May 1984
	13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Collinearity, curvature, diagnostics, influence graphs, influential observations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

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 a particular class of models, and it seems to provide a relatively simple, unified approach for handling a variety of problems.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

8 84

DTIC