Leverage, local influence and curvature in nonlinear regression

BY ROY T. ST. LAURENT
Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

AND R. DENNIS COOK
Department of Applied Statistics, School of Statistics, University of Minnesota, St. Paul, Minnesota 55108, U.S.A.

SUMMARY

The connections between measures of leverage, statistical curvature, and the local influence of a mean-shift perturbation are explored in nonlinear regression models. The circumstances under which the Jacobian leverage reduces to a commonly-used linear approximation are found to be connected to the effective residual curvature of the nonlinear model. The local influence of a mean-shift perturbation in the model function is seen to be closely related to the Jacobian leverage.

Some key words: Diagnostics; Influential observation; Jacobian leverage; Mean-shift perturbation.

1. Introduction

In assessing the importance of individual cases in regression models, one is typically interested in identifying observations which have a greater-than-average impact on the estimation of model parameters and fitted values. These influential points often are identified through the use of case deletion diagnostics. Recently, local influence methods for examining the effect of small perturbations in the model or data have been developed.

Leverage is one of the key components of influence in linear models (Chatterjee & Hadi, 1986; Cook & Weisberg, 1982, Ch. 2). Several authors (del Rio, 1988; Hoaglin & Welsch, 1978; Huber, 1983) have argued that in linear models leverage is a measure of influence in its own right. One might suspect that, suitably defined, the concept of leverage would be important in measuring influence in more complex models as well. Leverage has been generalized to nonlinear regression models via the tangent plane leverage (Belsley, Kuh & Welsch, 1980, p. 271; Ross, 1987) and the Jacobian leverage (Emerson, Hoaglin & Kempthorne, 1984; St. Laurent & Cook, 1992).

Here we show how the nonlinearity of the model may affect the leverage, and we explore the relationship between leverage measures and statistical curvature. We discuss the connection between leverage and local influence (Cook, 1986) in a nonlinear regression model, emphasizing the relation between the Jacobian leverage and local influence on the model fitted values.

In § 2 we briefly describe the tangent plane and Jacobian leverage measures. We investigate the relation between these measures, and their relationship to curvature measures of nonlinearity in § 3. In § 4, we connect the development of leverage to the local influence of a mean-shift perturbation. Examples are available in § 5.
2. Definitions of leverage

2.1. Tangent plane leverage

A nonlinear regression model consists of \( n \) observations with the structure

\[
y_i = \eta_i(\theta) + \varepsilon_i, \quad (i = 1, \ldots, n),
\]

where \( y_i \) is the \( i \)th observed response, and \( \eta_i \) is a known function with domain \( \Theta \), an open subset of \( \mathbb{R}^p \), and range a convex subset of \( \mathbb{R} \). The \( p \)-vector \( \theta \in \Theta \) is unknown. The error terms \( \varepsilon_i \) are independent with mean 0 and variance \( \sigma^2 \).

Let \( \eta(\theta) \) denote the \( n \times 1 \) vector with \( i \)th element \( \eta_i(\theta) \), and let \( Y \) denote the \( n \times 1 \) vector of responses. Given the response vector \( Y \), the least squares estimate of \( \theta \) is denoted by \( \hat{\theta} \). Quantities that depend upon \( \hat{\theta} \) will be 'hatted' to indicate evaluation at \( \hat{\theta} \). For example, \( V = V(\hat{\theta}) \) is the \( n \times p \) matrix with \( i \)th row \( \partial \eta_i(\theta)/\partial \theta^T \) and \( \hat{V} = V(\hat{\theta}) \).

Often a tangent plane approximation to the expectation surface

\[
M = \{ \eta(\theta): \theta \in \Theta \} \subset \mathbb{R}^n
\]
at \( \hat{\theta} \) is used to make inference about \( \theta \) through the derived linear model \( \nu(\theta) = \eta(\hat{\theta}) + \hat{V}(\hat{\theta} - \theta) \). Based on this approximation, the tangent plane leverage matrix, denoted by \( \hat{H} \), may be written

\[
\hat{H} = \hat{V}(\hat{V}^T\hat{V})^{-1}\hat{V}^T.
\]

The diagonal elements \( \hat{h}_{ii} \) of \( \hat{H} \) are frequently used to measure leverage in nonlinear regression models (Ross, 1987).

2.2. Jacobian leverage

Let \( W_i = W_i(\theta) \) (\( i = 1, \ldots, n \)) denote the \( p \times p \) Hessian matrix of \( \eta_i(\theta) \). The \( n \times p \times p \) three dimensional array \( W \) is constructed from the \( W_i \)'s: the \( i \)th face of \( W \) is given by \( W_i \) (Bates & Watts, 1988, Ch. 7). It will also be convenient to consider an \( n \times p \times p \) array as a collection of \( p^2 \) \( n \)-vectors, denoted by \( W_{ab} \) (\( a, b = 1, \ldots, p \)). We shall make use of column multiplication of three-dimensional arrays, indicated by \([.]\,[.]\), as defined by Bates & Watts (1980) and illustrated in (3) below.

Emerson et al. (1984) define the Jacobian leverage \( \hat{j}_{ik} \) as the instantaneous rate of change in the \( i \)th fitted value with respect to the \( k \)th response. Formally \( \hat{j}_{ik} \) is the limit as \( b \to 0 \) of \( \{ \eta_i(\hat{\theta}) - \eta_i(\hat{\theta}(k, b)) \}/b \), where \( \hat{\theta}(k, b) \) is the least squares estimate of \( \theta \) when \( y_k \) is replaced by \( y_k + b \). Collecting the \( \hat{j}_{ik} \) in an \( n \times n \) matrix gives the Jacobian leverage matrix

\[
\hat{J} = \hat{V}(\hat{V}^T\hat{V}[\hat{\varepsilon}^T\hat{W}])^{-1}\hat{V}^T.
\]

Here \( \hat{\varepsilon} \) is the \( n \times 1 \) vector of residuals from (1) with elements \( \hat{\varepsilon}_i = y_i - \eta_i(\hat{\theta}) \), and \( [\hat{\varepsilon}^T\hat{W}] = \sum \hat{\varepsilon}_i\hat{W}_i \) where the sum is taken over \( i = 1, \ldots, n \).

In the linear model \( \eta(\theta) = X\theta \), where \( X \) is a known \( n \times p \) matrix, the leverage matrix \( H \) is the projection matrix onto the column-space of \( X \), denoted by \( \mathcal{X}(X) \). The diagonal elements of \( H \), denoted by \( h_{ii} \), measure the rate of change in \( \hat{y}_i \) as \( y_i \) changes. As \( 0 \leq h_{ii} \leq 1 \), the rate of increase in \( \hat{y}_i \) is constrained to be no greater than the rate of increase in \( y_i \).

While \( \hat{J} \) is positive definite and hence \( \hat{j}_{ii} \geq 0 \), it is not necessarily the case that \( \hat{j}_{ii} \) is bounded above by one. St. Laurent & Cook (1992) show that in a nonlinear regression model the rate of change in the predicted response \( \hat{y}_i \) may be greater than the rate of
change in $y_i$; in other words it is possible to have $\hat{f}_n > 1$. Cases with $\hat{f}_n > 1$ are said to exhibit superleverage.

3. LEVERAGE MEASURES AND CURVATURE

Taking $\hat{J}$ as our definition of leverage for a nonlinear model, we may consider $\hat{H}$ as an approximation to it. The appropriateness of this approximation will depend upon the adequacy of the tangent plane approximation to $M$ at $\hat{\eta}$.

Insight into the difference between $\hat{H}$ and $\hat{J}$ is provided by examining these measures in a standard form. Let the $QR$ decomposition of $\hat{V}$ be given by $\hat{V} = \hat{Q}\hat{R}$, where $\hat{Q}$ is an $n \times p$ orthonormal matrix with $\mathcal{E}(\hat{Q}) = \mathcal{E}(\hat{V})$, and $\hat{R}$ is a $p \times p$ upper triangular matrix. The leverage matrices can be written as $\hat{H} = \hat{Q}\hat{Q}^T$ and $\hat{J} = \hat{Q}(I - \hat{B})^{-1}\hat{Q}^T$, where the $p \times p$ matrix $\hat{B} = [e^T][\hat{U}]$ is symmetric and $\hat{U} = \hat{R}^{-T}\hat{W}\hat{R}^{-1}$ is $n \times p \times p$. The $p^2$ $n$-vectors comprising $\hat{U}$, $\hat{U}^{ab}$ ($a, b = 1, \ldots, p$) are the second partial derivatives with respect to the elements of $\tau$ for the reparameterized model $\delta(\tau) = \eta(\theta)$, where $\tau = \hat{Q}^T[\eta(\theta) - \hat{\eta}]$.

Hamilton, Watts & Bates (1982) call $\hat{B}$ the effective residual curvature matrix since its elements $B_{ab}$ give the effective normal curvatures in the direction of the residual vector $e$. The scalar $B_{ab}$ is proportional to the coordinate of the projection of the column vector $\hat{U}^{ab}$ onto the residual vector, and the signs of the eigenvalues of $\hat{B}$ depend upon the orientation of the vectors $\hat{U}^{ab}$ ($a, b = 1, \ldots, n$) relative to $e$. In effect, $\hat{B}$ measures whether the expectation surface $M$ is curving toward or away from the residual vector. As $e$ is orthogonal to $\mathcal{E}(\hat{V})$, it follows that $\hat{B}$ is a function of only the intrinsic curvature portion of the second derivative array of Bates & Watts (1980). The leverage matrix $\hat{J}$ takes into account the normal curvatures of $M$ at $\hat{\eta}$ relative to the residual vector, while $\hat{H}$ does not. Parameter-effects curvature does not play a role here, as both $\hat{J}$ and $\hat{H}$ are invariant under reparameterization.

The Jacobian and tangent plane leverage matrices are identical if and only if $\hat{B} = 0$. Two special cases in which this occurs are when the model provides an exact fit to the data ($e = 0$) and when the model is intrinsically linear ($\hat{U}^{ab} = 0, a, b = 1, \ldots, p$). Measures of the extent to which $\hat{J}$ and $\hat{H}$ differ may be constructed from the eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$ of $\hat{B}$, since $\hat{B} = 0$ if and only if $\lambda_1 = \lambda_p = 0$. The smallest and largest nonzero eigenvalues of $\hat{J}$ are $\psi_1 = (1 - \lambda_1)^{-1}$ and $\psi_p = (1 - \lambda_p)^{-1}$ respectively, a fact we make use of in § 4. A succinct summary of the information about $\hat{J}$ provided by $(\lambda_1, \lambda_p)$ is given by $K = \max \{|\lambda_1|, |\lambda_p|\}$, and may be interpreted as the maximum effective residual curvature (Cook & Tsai, 1990). Then $K = 0$ if and only if $\hat{J} = \hat{H}$.

4. LEVERAGE AND LOCAL INFLUENCE

In addition to the model assumptions near (1), suppose that $e_i$ is a Gaussian random variable. In this section we show that the local influence of a perturbation in the mean of the nonlinear model is just a rescaling of the Jacobian leverage.

Consider perturbing the response vector $Y$ by adding the vector $\omega$ to $Y$. Here perturbation of the data is equivalent to perturbation of the model, since its effect on the likelihood is measured as a function of $Y$ and $\eta(\theta)$ only through their difference, $e(\theta) = Y - \eta(\theta)$. This duality leads to a close connection between the Jacobian leverage and the local influence of an additive, or mean-shift, perturbation in the model.

Cook (1986) considers the influence of a perturbation in a model on the maximum likelihood estimate of the parameter vector $\xi$ through its effect on $L$, the log likelihood
for the unperturbed model. Indexing the perturbation scheme by a \( q \)-vector \( \omega \in \Omega \), he uses the \((q + 1)\)-dimensional influence graph, \( \alpha(\omega) = (\omega^T, LD(\omega))^T \) to display the effect of perturbation of the model by \( \omega \). Here \( LD(\omega) \) is the likelihood displacement \( LD(\omega) = 2(L(\hat{\xi}) - L(\hat{\xi}_\omega)) \), where \( \hat{\xi}_\omega \) and \( \hat{\xi} \) are estimates of \( \xi \) based on the perturbed and unperturbed models, respectively. Cook (1986) defines analogous likelihood displacements and influence graphs when only a subset of \( \xi \) is of interest.

To capture the local effect of perturbing the model in a neighbourhood of \( \omega = \omega_0 \), Cook examines the normal curvature \( C_l \) of a lifted line on the influence graph generated by the line \( \omega(a) = \omega_0 + al \) in \( \Omega \). Here \( l \) is a \( q \)-vector, \( \|l\| = 1 \) and \( a \in \mathbb{R} \). In what follows, we take \( q = n \).

For an additive perturbation of the response \( Y \) in the direction \( l \), the normal curvature measure of the local influence on estimation of model location and scale \( \xi = (\theta^T, \sigma^2)^T \) at \( \hat{\xi} \), namely \( C_l(\eta, \sigma^2) \), can be decomposed into the sum of two orthogonal quadratic forms which themselves are normal curvatures:

\[
C_l(\eta) = \frac{2}{\sigma^2} l^T \hat{\xi} \hat{J} l
\]

measures the effect of the perturbation on the estimation of \( \eta(\theta) \); and

\[
C_l(\sigma^2) = \frac{4}{\sigma^2} l^T \hat{\sigma}^2 l
\]

measures the effect of the perturbation on the estimation of \( \sigma^2 \). Here \( \hat{\sigma}^2 = \hat{\sigma}^2 / (\hat{\sigma}^T \hat{\sigma}) \), and \( \hat{\sigma}^2 \) is the maximum likelihood estimate of \( \sigma^2 \). Thus \( C_l(\eta, \sigma^2) = C_l(\eta) + C_l(\sigma^2) \). The proof of this result is outlined in the Appendix. Schwarzmann (1991) gives the special case of \( C_l(\eta, \sigma^2) \) when \( \eta(\theta) \) is linear in \( \theta \).

For this perturbation scheme, the normal curvatures are invariant under reparameterization. Thus we use \( \eta \) to designate the influence effects on the estimation of the model mean, rather than using the parameterization-specific designator \( \theta \).

The orthogonality of the quadratic forms comprising \( C_l(\eta, \sigma^2) \) implies that influence on estimation of the mean function \( \eta(\theta) \) is caused by perturbations in directions \( l \in \mathcal{C}(\hat{V}) \); while influence on estimation of the variance \( \sigma^2 \) is caused by perturbations in the direction \( l = \hat{\sigma} / \|\hat{\sigma}\| \); and vectors \( l \) orthogonal to \( \mathcal{C}(\hat{V}, \hat{\sigma}) \) are directions of no influence.

Cook (1986) suggests that it is useful to find \( l_{\text{max}} \), the direction yielding the maximum normal curvature at \( \hat{\xi} \), to assess the local influence of the worst-case perturbation. From the decomposition above, it follows that

\[
C_{\text{max}}(\eta, \sigma^2) = \max \{ C_{\text{max}}(\eta), C_{\text{max}}(\sigma^2) \}.
\]

Taking into account the relationship between the eigenvalues of the matrices \( \hat{\mathcal{B}} \) and \( \hat{J} = \hat{Q}(I - \hat{\mathcal{B}})^{-1}\hat{Q}^T \), referred to in § 3, it can be shown that \( C_{\text{max}}(\eta) = 2\psi / \hat{\sigma}^2 \). The only nonzero eigenvalue of \( \hat{\mathcal{B}} \) is 1, corresponding to the eigenvector \( \hat{\sigma} / \|\hat{\sigma}\| \), hence \( C_{\text{max}}(\sigma^2) = 4 / \hat{\sigma}^2 \) and therefore \( C_{\text{max}}(\eta, \sigma^2) = \max \{ 2\psi / \hat{\sigma}^2, 4 / \hat{\sigma}^2 \} \).

As \( I - \hat{\mathcal{B}} \) is positive definite (Hamilton et al., 1982), it follows that \( \lambda_p < 1 \). If \( \lambda_p < \frac{1}{2} \), then \( C_{\text{max}}(\eta, \sigma^2) = C_{\text{max}}(\sigma^2) \) and \( l_{\text{max}} = \hat{\sigma} / \|\hat{\sigma}\| \), that is the mean-shift perturbation that yields the maximum local influence on estimation is a perturbation in the direction of the residual vector, and that perturbation affects only the estimation of the scale parameter. A common situation in which this occurs is when \( \eta(\theta) \) is intrinsically linear, for then \( \lambda_p = 0 \).

When \( \hat{J} = \hat{H} \) and as long as \( p > 1 \), any \( l \in \mathcal{C}(\hat{V}) \) will give \( C_l(\eta) = C_{\text{max}}(\eta) = 2 / \hat{\sigma}^2 \), since \( \hat{H} \) is the projection matrix onto that space. Thus for intrinsically linear models and other
model/data sets where \( \hat{J} = \hat{H} \), the entire \( p \)-dimensional linear subspace \( \mathcal{C}(\hat{V}) \) is associated with worst-case local influence on the mean response.

When \( l \) is the \( i \)-th standard basis vector in \( \mathbb{R}^p \), the normal curvature \( C_l(\eta) \) is proportional to \( \hat{j}_{ll} \). Hence when interest is in assessing which single case has the largest local effect on estimation of \( \eta \), the observation which satisfies this criterion is the one with the largest Jacobian leverage. If in addition to the mean \( \eta \), scale is taken into account, then the observation which yields the largest local influence is the one for which \( \hat{j}_{ll} + 2\hat{\sigma}^2/(n\hat{\sigma}^2) \) is maximized.

### 5. Examples

**Example 1.** Consider the class of partially nonlinear regression models,

\[
Y = X\alpha + \beta g(\gamma) + \varepsilon, \tag{6}
\]

where \( X \) is a known \( n \times (p-2) \) full rank matrix, \( \alpha \) is a \((p-2) \times 1\) vector of unknown parameters, \( \beta \) and \( \gamma \) are unknown scalars and \( g \) is a known vector-valued function. For this model it can be shown that \( K = |\tau| \), where

\[
\tau = \hat{\beta}^{T} g''(\hat{\gamma}), \quad c = d_{p}^{T}(\hat{V}^{T}\hat{V})^{-1}d_{p}, \quad \hat{V} = (X, g(\hat{\gamma}), \hat{\beta}g'(\hat{\gamma})),
\]

\( d_{p} \) is the \( p \)-th standard basis vector in \( \mathbb{R}^p \), and \( g' \) and \( g'' \) are the first and second derivatives of \( g \) with respect to \( \gamma \). The sign of \( \tau \) indicates whether the leverage is, on average, greater \((\tau > 0)\) or less \((\tau < 0)\) than what one would get in linear regression.

For a numerical illustration we use the model and contrived data from J. R. Hill and C. L. Tsai in a personal communication:

\[
y_i = \alpha x_{i1} + \exp(\gamma x_{i2}) + \varepsilon, \quad (i = 1, \ldots, 9).
\]

For their data given in Table 1, \( \hat{\alpha} = 2 \), \( \hat{\gamma} = 0 \), \( \hat{\sigma}^2 = 2.314 \) and \( K = 96 \). Operationally, the results for model (6) apply here by deleting the column \( g(-\gamma) \) from \( \hat{V} \) and setting \( \hat{\beta} = 1 \).

Given the large value of \( K \) we expect to see notable differences between the Jacobian and tangent plane leverage measures for these data. Since \( \tau = -32 \) the leverages for this problem should be less on the average than predicted by the linear regression analogy. Figure 1 gives an index plot of \( \hat{j}_{ll} \) and \( \hat{h}_{ll} \) which shows the individual downward adjustments of the leverage values.

Here the largest eigenvalue of \( \hat{B} \) is \( \lambda_2 = 0 \), indicating that the direction of perturbation in the data that yields the greatest local influence on model fit is \( L_{max} = \hat{\varepsilon}/\|\hat{\varepsilon}\| \). It follows

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( y_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.0</td>
<td>-2.5</td>
<td>-10.0</td>
</tr>
<tr>
<td>-3.0</td>
<td>-2.0</td>
<td>-5.0</td>
</tr>
<tr>
<td>-2.0</td>
<td>-1.5</td>
<td>-2.0</td>
</tr>
<tr>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>1.5</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>4.0</td>
</tr>
<tr>
<td>2.0</td>
<td>1.5</td>
<td>5.0</td>
</tr>
<tr>
<td>3.0</td>
<td>2.0</td>
<td>6.0</td>
</tr>
<tr>
<td>4.0</td>
<td>2.5</td>
<td>7.0</td>
</tr>
</tbody>
</table>

Table 1. Contrived data for Example 1, from J. R. Hill and C. L. Tsai (personal communication)
that the largest eigenvalue of $\hat{J}$ is $\psi_2 = 1$. The corresponding eigenvector is $x_1/\|x_1\|$. Thus the local influence on predicted values, ignoring scale, is maximized by perturbing the data in the direction $x_1/\|x_1\|$. As the contribution from the second eigenvector is quite small relative to the first ($\psi_1 = 0.01$), $\hat{J}$ is well approximated by the leverage matrix associated with the simple linear regression model in $x_1$ through the origin, namely $x_1x_1^T/(x_1x_1^T)$. Hence the index plot of $j_{\nu}$ in Fig. 1 also gives approximately the square of the elements of the perturbation vector yielding the maximum leverage or local influence on predicted values.

**Example 2.** Bates & Watts (1988, p. 306) give data relating molar concentration of nifedipine, NIF, to radioactivity counts, $y_{\nu}$, in rat heart tissue radioactively tagged with nitrendipine. They propose fitting the four-parameter logistic model

$$
\eta_i(\theta) = \theta_1 + \frac{\theta_2}{1 + \exp \{\theta_4(x_i - \theta_3)\}},
$$

where

$$x_i = \log_{10} \text{(concentration NIF)}.
$$

We consider only the data reported for tissue sample 2 consisting of $n = 16$ cases: two replicates at each of eight concentrations. An interesting feature of these data is that for cases 1 and 2 NIF concentration is zero and thus $x_i = -\infty$. For these cases, (7) becomes $\eta_i(\theta) = \theta_1 + \theta_2$ as long as $\theta_4 > 0$. The least squares estimates of the parameters are $\hat{\theta}_1 = 1923.52$, $\hat{\theta}_2 = 3194.92$, $\hat{\theta}_3 = -8.3214$, $\hat{\theta}_4 = 1.2687$ and $\hat{\sigma} = 535.6$. There is no clear evidence for lack of fit. The maximum effective residual curvature is $K = 0.65$, a value sufficiently different from zero to cast suspicion on the usefulness of $H$ as an approximation to $\hat{J}$.

In Fig. 2 is a plot of the data with the fitted response curve. The two zero concentrations have been plotted at abscissa $x = -27$. We might expect that these two cases would have a great deal of influence on model fit as the asymptote of the curve, $\theta_1 + \theta_2$, would be ill-determined were these points deleted. For the most part however, the value of the asymptote is of little importance in determining the predicted values for any but these two cases and to a lesser extent the observations at $x = -11$. Therefore, the local influence of these points when the data is subject to perturbation in the response should not be unduly large.
Leverage, local influence and curvature

Fig. 2. Nifedipine data and fitted response curve, \( \eta_i(\theta) = \theta_1 + \theta_2/[1 + \exp(\theta_3(x_i - \theta_3))] \). The two points plotted at abscissa -27 correspond to log concentrations of \(-\infty\).

Fig. 3. Index plots for the nifedipine data and the model \( \eta_i(\theta) = \theta_1 + \theta_2/[1 + \exp(\theta_3(x_i - \theta_3))] \) of:

(a) \( \hat{j}_{i\bar{u}} \), shown by solid circles, and \( \hat{h}_{i\bar{u}} \), triangles; and (b) the elements of \( L_{\max}^{(\eta)} \).

The index plot of \( \hat{j}_{i\bar{u}} \) and \( \hat{h}_{i\bar{u}} \) in Fig. 3(a) shows fairly striking disagreement between the two measures. Here and in Fig. 3(b), cases are ordered by x-value, from smallest to largest. The tangent plane leverage is greatest for cases 1 and 2, the cases for which NIF concentration is zero. The Jacobian measure clearly identifies the replicate pairs 9-10 at \( x = -8 \), and 15-16 at \( x = -5 \) as large leverage points relative to the remaining cases, while the tangent plane measure does not. The influence of these replicate pairs is also clearly illustrated in Fig. 3(b) which contains the index plot of the elements of \( L_{\max}^{(\eta)} \) corresponding to \( C_{\max}(\eta) = 3.38/\hat{\sigma}^2 \), compared to \( 2/\hat{\sigma}^2 \) for any \( l \in \mathcal{C}(\hat{H}) \). By perturbing the data in this direction, the resulting fitted curve will drop more steeply since \( \hat{\theta}_4 \) will become larger, and hence the predicted values will change most rapidly.

Acknowledgement

This work was supported in part by grants from the National Institutes of Health and the National Science Foundation.
Derivation of $C_t(\eta, \sigma^2)$

For $\xi = (\theta^T, \sigma^2)^T$, the log likelihood for the perturbed model is

$$L(\xi | \omega) = -(n \log \sigma^2) + \| Y + \omega - \eta(\theta) \|^2 / \sigma^2 / 2.$$  

Cook (1986, eqn (16)) gives the formula for the normal curvature

$$C_t = 2||\Delta^T \Phi^{-1} \Delta||,$$  

where $\Delta$ is a $(p + 1) \times n$ matrix with elements the second mixed partial derivatives of $L(\xi | \omega)$ with respect to the elements of $\xi$ and $\omega$, and $-\Phi$ is the $(p + 1) \times (p + 1)$ observed information matrix for the unperturbed model. For a mean shift perturbation in $T$, we obtain

$$\Delta = \frac{\partial^2 L(\xi | \omega)}{\partial \xi \partial \omega} \bigg|_{\xi = \hat{\xi}, \omega = \hat{\omega}} = \frac{1}{\sigma^4} \begin{pmatrix} \hat{\sigma}^2 \hat{v}^T \\ \hat{e}^T \end{pmatrix}, \quad -\Phi = \frac{1}{2\sigma^4} \begin{pmatrix} 2\hat{\sigma}^2 \hat{A} & 0 \\ 0 & n \end{pmatrix},$$

where $\hat{A} = \hat{V}^T \hat{Y} - [\hat{e}^T][\hat{W}]$. Substitution into (A1) gives the result $C_t(\eta, \sigma^2)$ just below (5).

The results (4) and (5) are based on similar calculations using equations (16), (25) and (26) of Cook (1986). The orthogonality of the quadratic forms $C_t(\eta)$ and $C_t(\sigma^2)$ follows from the normal equations: $\hat{V}^T \hat{e} = 0$ and hence $\hat{J} \hat{P} e = 0$.

REFERENCES


[Received January 1992. Revised July 1992]