

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

AD-A149 456

3

MRC Technical Summary Report # 2751

THE LIKELIHOOD DISPLACEMENT: A UNIFYING PRINCIPLE FOR INFLUENCE MEASURES

R. Dennis Cook, Daniel Pena and Sanford Weisberg

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

September 1984

(Received August 24, 1984)

DTIC FILE COPY

Approved for public release
Distribution unlimited

DTIC
ELECTE
JAN 22 1985
S D D

Sponsored by

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

85 01 16 C44

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

THE LIKELIHOOD DISPLACEMENT: A UNIFYING
PRINCIPLE FOR INFLUENCE MEASURES

R. Dennis Cook*, Daniel Pena** and Sanford Weisberg*

Technical Summary Report #2751
September 1984

ABSTRACT

The young field of statistical diagnostics has produced an array of competing statistics for measuring the influence of individual cases. Two of the most popular measures for linear regression are Cook's (1977) D_i and Belsley, Kuh and Welsch's (1980) $DFITS_i$. Using the likelihood displacement (Cook and Weisberg, 1982) as a unifying concept, these two measures are compared.

next page

AMS(MOS) Subject Classification: 62J05, 62F35, 62J99.

Key Words: Influential observations, Cook's distance, $DFITS$, Likelihood displacement.

Work Unit Number 4 - Statistics and Probability.

* Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108.

** Escuela de Ingenieros Industriales, Universidad Politécica de Madrid, Madrid, Spain.

SIGNIFICANCE AND EXPLANATION

The identification of influential cases seems generally accepted as an important part of linear regression analysis. Although there are many diagnostic methods available for this, two specific diagnostic statistics-- D_i as proposed by Cook (1977), and $DFITS_i$ as proposed by Belsley, Kuh and Welsh (1980)--appear to be used most frequently since they are available in many widely distributed regression packages. For further progress and a deeper understanding of available methodology, larger perspectives seem necessary. We have found the likelihood displacement to be particularly well-suited for this study .

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A/1	



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

THE LIKELIHOOD DISPLACEMENT: A UNIFYING
PRINCIPLE FOR INFLUENCE MEASURES

R. Dennis Cook*, Daniel Pena** and Sanford Weisberg*

1. INTRODUCTION

The identification of influential cases seems generally accepted as an important part of linear regression analysis. Although there are many diagnostic methods available for this, two specific diagnostic statistics-- D_1 as proposed by Cook (1977), and $DFITS_1$ as proposed by Belsley, Kuh and Welsch (1980)--appear to be used most frequently since they are available in many widely distributed regression packages.

A number of authors, including Atkinson (1981), Belsley, Kuh and Welsch (1980), Cook and Weisberg (1982), Hoaglin and Welsch (1978) and Welsch (1982), use special pleading to justify the use of D_1 or $DFITS_1$, generally concentrating on isolated characteristics of these statistics. Although useful, such narrow arguments are not likely to resolve important differences or even allow bilateral recognition of alternative views. One way to further understand this is to cast both diagnostics into a common framework so that they can be judged in a larger perspective. Such a framework is provided by the likelihood displacement (distance) as developed by Cook and Weisberg (1982, p. 182).

In section 2 we review the likelihood displacement and the central results for linear regression. In section 3 we show that both D_1 and $DFITS_1$ fit conveniently into this framework, and address some of the specific arguments alluded to above. Section 4 contains our concluding comments.

* Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108.

** Escuela de Ingenieros Industriales, Universidad Polit3cnica de Madrid, Madrid, Spain.

2. LIKELIHOOD DISPLACEMENT

Let θ be a $p \times 1$ parameter vector partitioned as $\theta^T = (\theta_1^T, \theta_2^T)$, where θ_1 is $p_1 \times 1$, and let $L(\theta; Z) = L(\theta_1, \theta_2; Z)$ denote the log likelihood function for θ based on data Z . To help with later ideas, Figure 1 illustrates the contours of $L(\theta; Z)$ when $p=2$. The maximum likelihood estimate (mle) $\hat{\theta}^T = (\hat{\theta}_1^T, \hat{\theta}_2^T)$ is indicated in Figure 1 by the point F.

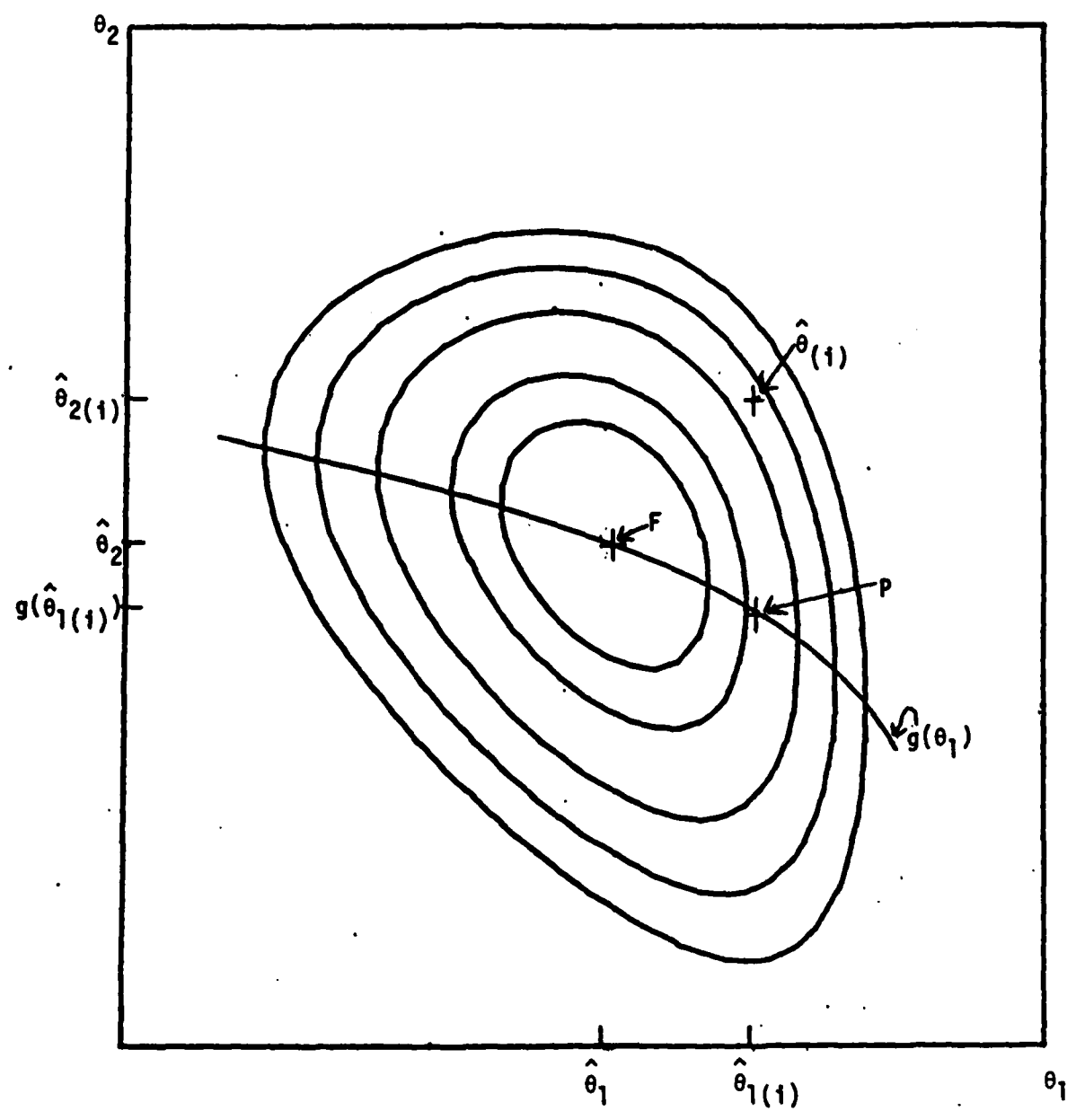
In influence analysis we often wish to compare the full data mle $\hat{\theta}$ to the mle $\hat{\theta}_{(1)}^T = (\hat{\theta}_{1(1)}^T, \hat{\theta}_{2(1)}^T)$ obtained from the log likelihood $L(\theta; Z_{(1)})$ where the subscript "(1)" means "without case 1". One useful and general method for comparing $\hat{\theta}$ and $\hat{\theta}_{(1)}$ is based on the likelihood displacement

$$LD_1(\theta) = 2[L(\hat{\theta}; Z) - L(\hat{\theta}_{(1)}; Z)] \quad (1)$$

In Figure 1, this displacement corresponds to computing twice the difference in the heights of the full data log likelihood at $\hat{\theta}$ and at $\hat{\theta}_{(1)}$. If this difference in heights is large, case 1 is called influential since deleting it may cause a substantial change in important conclusions. The likelihood displacement judges all cases falling on the same contour of L to be equally influential. If desirable, this displacement can be transformed to a more familiar scale by comparing it to percentiles of a chi-squared distribution with p degrees of freedom. This comparison gives the level of the smallest likelihood region for θ that contains $\hat{\theta}_{(1)}$ (Cox and Hinkley, 1974, Chapter 9).

In many problems, a subset of θ can be regarded as nuisance parameters so that only the remaining parameters are of interest. Suppose that θ_1 is of interest while θ_2 represents the nuisance parameters. Define the implicit function $g(\theta_1)$, such that, for fixed θ_1 , $L(\theta_1, g(\theta_1); Z)$ is maximized; $g(\theta_1)$ is

Figure 1. Contours of a log likelihood function $L(\theta_1, \theta_2; Z)$



given as a curved line in Figure 1. The likelihood displacement for θ_1 ignoring θ_2 can now be defined as

$$LD_1(\theta_1|\theta_2) = 2\{L(\hat{\theta};Z) - L[\hat{\theta}_{1(1)},g(\hat{\theta}_{1(1)});Z]\} \quad (2)$$

In Figure 1, the point P is obtained by moving the point $\hat{\theta}_{(1)}$ parallel to the θ_2 axis until it reaches the curve g. Then $LD_1(\theta_1|\theta_2)$ is just twice the difference in height of the point F and the point P. Again, $LD_1(\theta_1|\theta_2)$ may be calibrated by comparison to the percentiles of a chi-squared distribution, now with p_1 degrees of freedom.

It is fairly straightforward to apply the general results (1) and (2) to the standard linear regression model

$$Y = X\beta + \epsilon \quad (3)$$

where $Y = (y_1)$ is an $n \times 1$ vector of observable responses, the $n \times p$ matrix X is known and has full rank, β is a $p \times 1$ vector of unknown parameters and the $n \times 1$ vector of unobservable errors ϵ is at least tentatively assumed to follow a multivariate normal distribution with mean 0 and variance $\sigma^2 I$. Let $\hat{\beta}$ and $\hat{\sigma}^2$ denote the maximum likelihood estimators of β and σ^2 , respectively, and let $H = X(X^T X)^{-1} X^T$ so that the fitted values \hat{Y} and the residuals e can be written $\hat{Y} = Hy$ and $e = (I-H)Y$. The diagonal elements of H will be denoted by h_1 .

Cook and Weisberg (1982) show that

$$LD_1(\beta|\sigma^2) = n \log \left[\frac{p}{n-p} D_1 + 1 \right] \quad (4)$$

where D_1 is the statistic proposed by Cook (1977):

$$\begin{aligned} D_1 &= (\hat{\beta} - \hat{\beta}_{(1)})^T X^T X (\hat{\beta} - \hat{\beta}_{(1)}) / ps^2 \\ &= ||\hat{Y} - \hat{Y}_{(1)}||^2 / ps^2 \\ &= \frac{h_1}{1-h_1} \cdot \frac{r_1^2}{p} \end{aligned} \quad (5)$$

where $s^2 = e^T e / (n-p)$, and $r_1 = e_1 / s(1-h_1)^{1/2}$ is the 1-th internally Studentized residual. Since $LD_1(\beta|\sigma^2)$ is a monotonic function of D_1 , it is equivalent to D_1 for the purpose of ordering cases based on influence. When σ^2 is known, $LD_1(\beta)$ is equal to D_1 with ps^2 replaced by σ^2 .

3. LD_1 and $DFFITS_1$

All of the statistics considered here depend on the leverages h_1 and the residuals e_1 . For later convenience, define

$$b_1 = \frac{e_1^2}{e^T e (1-h_1)} \quad i=1,2,\dots,n \quad (6)$$

Under model (3) b_1 has a beta distribution with parameters $1/2$ and $(n-p-1)/2$. Using (4), (5) and (6) it is immediate that

$$D_1 = \frac{(n-p)}{p} b_1 \frac{h_1}{1-h_1}$$

and thus

$$LD_1(\beta|\sigma^2) = n \log \left\{ \frac{b_1 h_1}{1-h_1} + 1 \right\} \quad (7)$$

We now turn to the statistic $DFITS_1^2$ which is defined as (Belsley, Kuh and Welsch 1980)

$$DFITS_1^2 = \frac{e_1^2 h_1}{s_{(1)}^2 (1-h_1)^2}$$

Using the relationship (Cook and Weisberg, 1982, eq. (2.2.8))

$$\frac{\hat{\sigma}_{(1)}^2}{\hat{\sigma}^2} = \frac{n}{n-1} (1-b_1) \quad (8)$$

it follows easily that $DFITS_1$ can be expressed in the form

$$DFITS_1^2 = (n-p-1) \left(\frac{b_1}{1-b_1} \right) \left(\frac{h_1}{1-h_1} \right) \quad (9)$$

We shall also require expressions for $LD_1(\beta, \sigma^2)$ and $LD_1(\sigma^2|\beta)$; these are derived in the Appendix to be

$$LD_1(\beta, \sigma^2) = n \log \left(\frac{n}{n-1} \right) + n \log(1-b_1) + \left(\frac{b_1}{1-b_1} \right) \left(\frac{n-1}{1-h_1} \right) - 1 \quad (10)$$

and

$$LD_1(\sigma^2|\beta) = n \log\left(\frac{n}{n-1}\right) + n \log(1-b_1) + \frac{nb_1^{-1}}{1-b_1} \quad (11)$$

Equation (11) depends only on b_1 and not the leverage h_1 . Since b_1 is a monotonic transformation of the usual test statistic for a mean shift outlier, the study of the likelihood displacement for σ^2 ignoring β is equivalent to the study of mean shift outliers.

The full likelihood displacement $LD_1(\beta, \sigma^2)$ is monotonically increasing in h_1 , as is clear from an inspection of (10). In general, $h_1 \geq 0$ and for models with a constant $h_1 \geq n^{-1}$. A sufficient condition for (10) to be monotonic in b_1 is $h_1 \geq n^{-1}$. Interestingly, $LD_1(\beta, \sigma^2)$ reduces to $LD_1(\sigma^2|\beta)$ when h_1 is replaced with its minimum value $h_1 = 0$. In other words, when $h_1 = 0$, $LD_1(\beta|\sigma^2) = 0$ and $LD_1(\beta, \sigma^2) = LD_1(\sigma^2|\beta)$.

We now relate $DFFITS_1$ to the likelihood displacement by subtracting $LD_1(\sigma^2|\beta)$ from $LD_1(\beta, \sigma^2)$,

$$\begin{aligned} LD_1(\beta, \sigma^2) - LD_1(\sigma^2|\beta) &= \frac{b_1}{1-b_1} \frac{(n-1)}{1-h_1} - 1 - \frac{nb_1^{-1}}{1-b_1} \\ &= (n-1) \frac{h_1}{1-h_1} \cdot \frac{b_1}{1-b_1} \quad (12) \end{aligned}$$

Comparing (12) and (9) we see that

$$LD_1(\beta, \sigma^2) - LD_1(\sigma^2|\beta) = \frac{n-1}{n-p-1} DFFITS_1^2 \quad (1) \quad (13)$$

The factor $(n-1)/(n-p-1)$ appears in this fundamental relationship since the

likelihood displacement is based on the maximum likelihood estimator of σ^2 while $DFFITS_1$ is based on the usual bias adjusted estimator of σ^2 .

3.1 A Simple Illustration

For illustration, we consider simple regression through the origin so that complete contour plots can be drawn. The log likelihood for (β, σ^2) , given data $Z = (X, Y)$, is

$$L(\beta, \sigma^2; Z) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (y_i - \beta x_i)^2 \quad (14)$$

and the value of L at the mle is

$$L(\hat{\beta}, \hat{\sigma}^2; Z) = -\frac{n}{2} [\log(2\pi\hat{\sigma}^2) + 1] \quad (15)$$

To compute $LD_1(\beta | \sigma^2)$ we need to find the function $g_1(\beta)$ such that $L(\beta, g_1(\beta); Z)$ is maximized for each β . Differentiating (14) with respect to σ^2 and setting the result to zero gives

$$g_1(\beta) = \frac{1}{n} \sum (y_i - \beta x_i)^2 \quad (16)$$

Similarly, the function $g_2(\sigma^2)$ that maximizes $L(\beta, \sigma^2; Z)$ for each σ^2 is given by

$$g_2(\sigma^2) = \frac{\sum x_1 y_1}{\sum x_1^2} \quad (17)$$

We see that $g_2(\sigma^2)$ does not depend on σ^2 .

As a special case of this problem, we take $n=4$ and $(x_1, y_1) = (0,0), (.2, .2), (.2, -.2), (\sqrt{.92}, \sqrt{.92})$. For these data $||X|| = ||Y|| = 1$, and all points but the third fall on a common line. The all-but-one-point-on-a-line problem is mentioned by Dempster and Green (1981), and promoted by Welsch (1982) as a reason for the use of $DFITs_1$ over D_1 . The general idea is that $DFITs_1$ will always find the point that lies off the line to be most influential since $\hat{\sigma}_{(1)}^2 = 0$, while D_1 may identify a point on the line as most influential, a circumstance that is evidently counter to Welsch's (1982) intuition. Although this example is relatively simple, its essential characteristics are perfectly general.

Table 1 lists the maximum likelihood estimates $(\hat{\beta}, \hat{\sigma}^2)$ and $(\hat{\beta}_{(1)}, \hat{\sigma}_{(1)}^2)$, $i=1,2,3,4$. Figure 2 gives a contour plot of $L(\beta, \sigma^2; Z)$ as defined in (14). In addition, $g_1(\beta)$, equation (16), is indicated by the short dashes, and $g_2(\sigma^2)$, equation (17), is indicated by the long dashes. The peak of $L(\hat{\beta}, \hat{\sigma}^2)$ is indicated by "F" and has value given by (15). The points $(\hat{\beta}_{(1)}, \hat{\sigma}_{(1)}^2)$ are marked by $i=1,2,3,4$.

The four influence measures given in (7), (9), (10) and (11) correspond to the differences in heights between various points in Figure 2. Consider case 4, for example. The full likelihood displacement $LD_4(\beta, \sigma^2)$ is simply twice the difference in the heights of the points located at "F" and "4". For the measure $LD_4(\beta | \sigma^2)$, the point "4" is moved parallel to the ordinate until it falls on the curve $g_1(\beta)$; the final position is indicated by "4A" in Figure 2. Now $LD_4(\beta | \sigma^2)$ is just twice the difference in the heights of the points at "F"

Table 1

Maximum likelihood estimates for simple regression through the origin

<u>Index</u>	<u>Case Deleted</u>	<u>$\hat{\beta}$</u>	<u>$\hat{\sigma}^2$</u>
F	none	.920	.0382
1	(0,0)	.920	.0509
2	(.2,.2)	.917	.0511
3	(.2,-.2)	1	0
4	($\sqrt{.92}$, $\sqrt{.92}$)	0	.0266

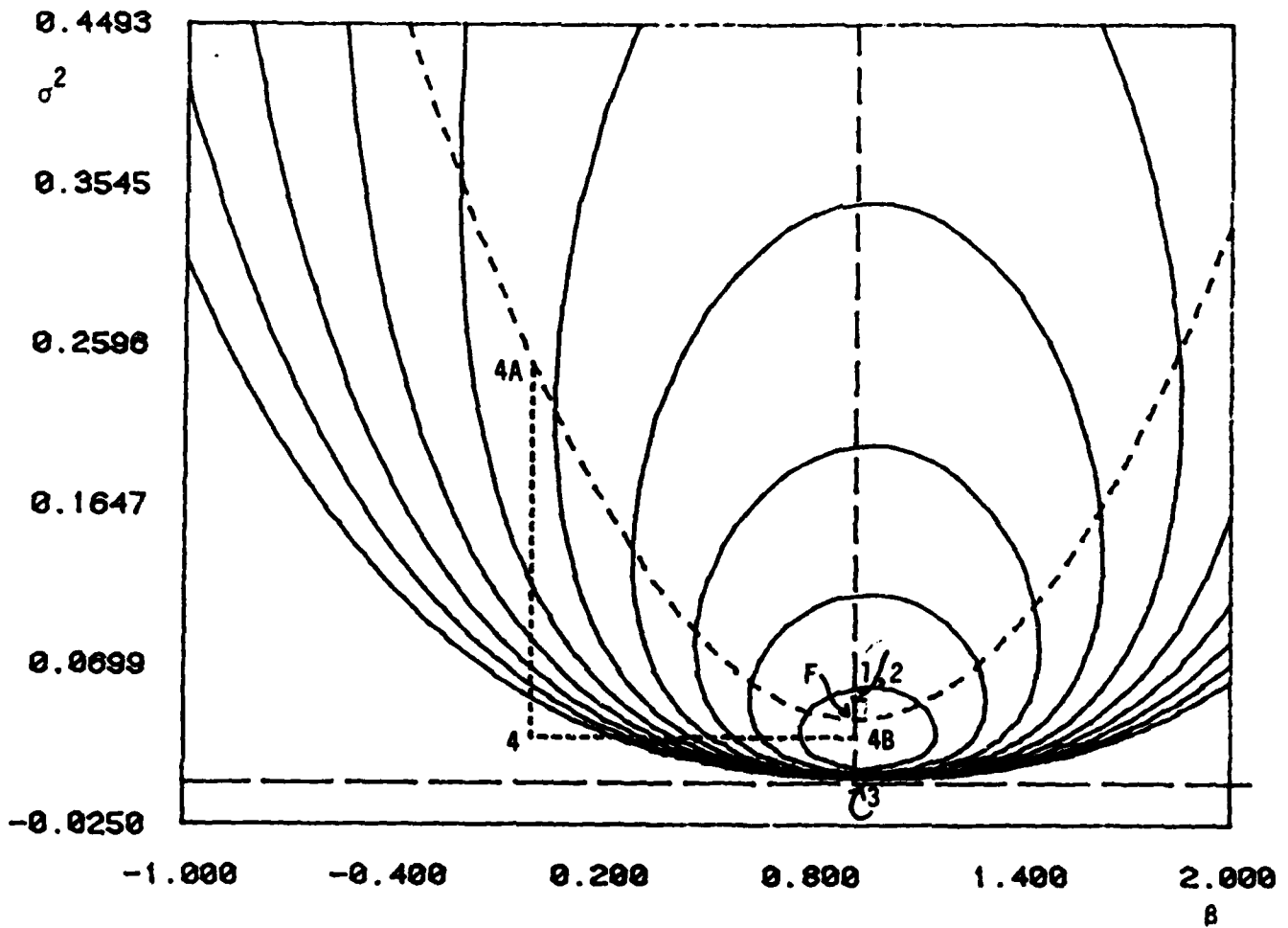


Figure 2. Contour plot of the log likelihood function $L(\beta, \sigma^2)$ for regression through the origin.

and "4A". Similarly, $LD_4(\sigma^2|\beta)$ is obtained by using the heights at "F" and "4B".

Each of the measures $LD_4(\beta, \sigma^2)$, $LD_4(\beta|\sigma^2)$ and $LD_4(\sigma^2|\beta)$ uses the maximum of L as a reference for assessing influence. In contrast, $DFFITS_4^2$ assesses influence by using the heights of points "4" and "4B", both of which lie on the side of L. If $DFFITS_4^2$ is useful then surely the analogous measure obtained by using point "4" and "4A" is useful also.

An inspection of Figure 2 yields the following qualitative conclusions. First, cases 1 and 2 are relatively uninfluential. Second, case 4 is influential for (β, σ^2) and β , but not for σ^2 alone. Finally, case 3 is influential for (β, σ^2) and σ^2 , but not for β alone. Notice that "3" falls just to the right of the vertical line (17) at $\beta = \hat{\beta} = .92$ where $L = -\infty$.

Returning to the all-points-but-one-on-a-line problem, we now see that $LD_1(\beta|\sigma^2)$ will not identify case 3 to be the most influential since "3" will be moved from $-\infty$ to the $g_1(\beta)$ curve prior to the computation of $LD_1(\beta|\sigma^2)$. This movement loses all information on changes in σ^2 , but is essential if we are to isolate changes in location as $LD_1(\beta|\sigma^2)$ is designed to do.

3.2 Contour Comparisons

Further insights can be obtained by comparing the contours of the four measures in the (b_1, h_1) plane. The contours for $LD_1(\beta, \sigma^2)$, $LD_1(\beta|\sigma^2)$ and $DFFITS_1^2$ are given in Figures 3-5, respectively. Recall that $LD_1(\sigma^2|\beta) = LD_1(\beta, \sigma^2)$ when $h_1 = 0$; thus the contours for $LD_1(\sigma^2|\beta)$ are parallel to the x-axis and they intersect the y-axis at the same points as the contours of $LD_1(\beta, \sigma^2)$ in Figure 3.

According to Welsch (1982), $DFFITS_1$ is designed to measure changes in location and scale simultaneously. With this in mind, we first compare

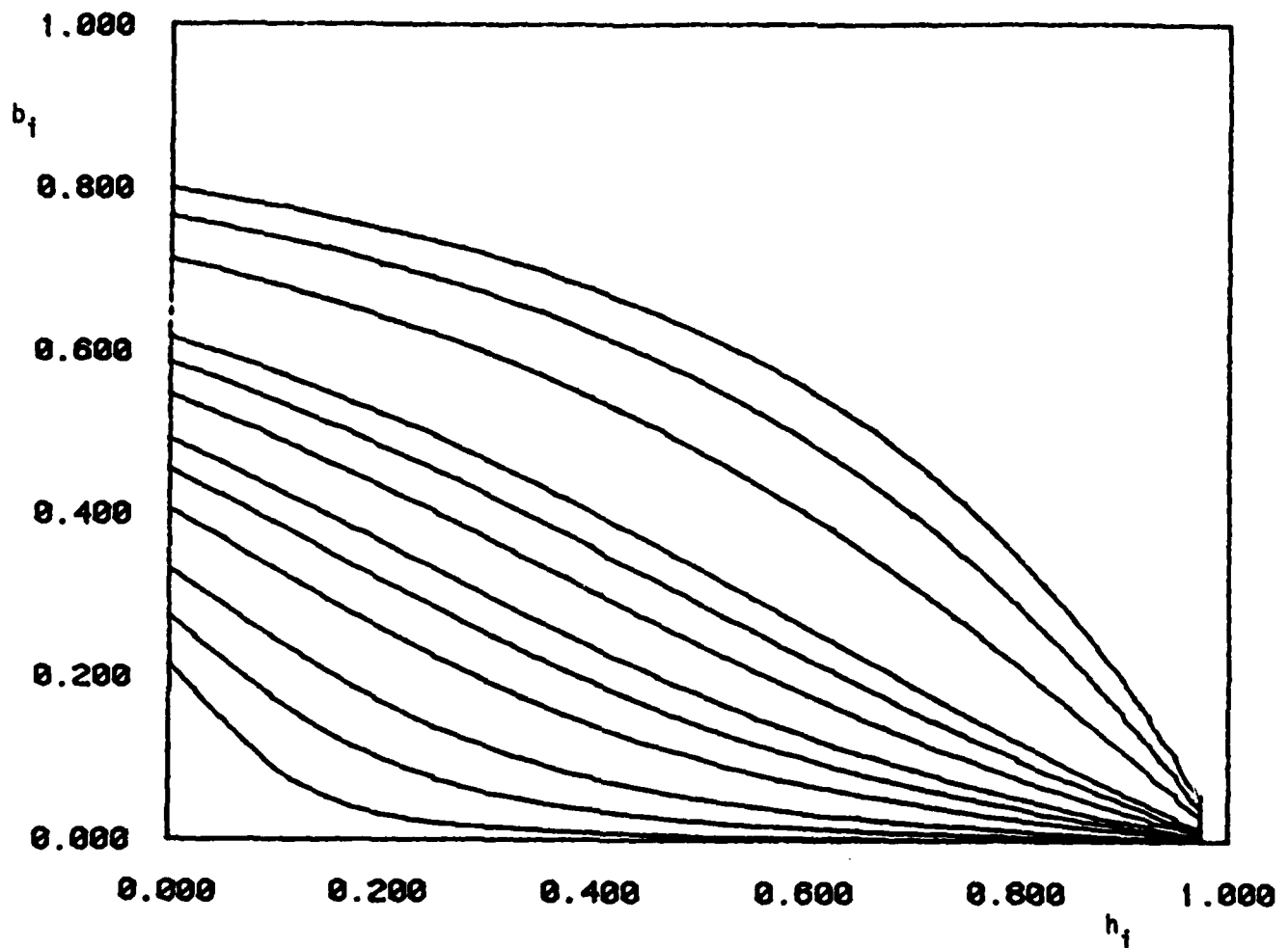


Figure 3. $LD_i(\beta, \sigma^2)$ as a function of (h_i, b_i) . Contours are drawn at .1, .25, .5, 1, 1.5, 2, 3, 4, 5, 10, 15, 20, 50, 100.

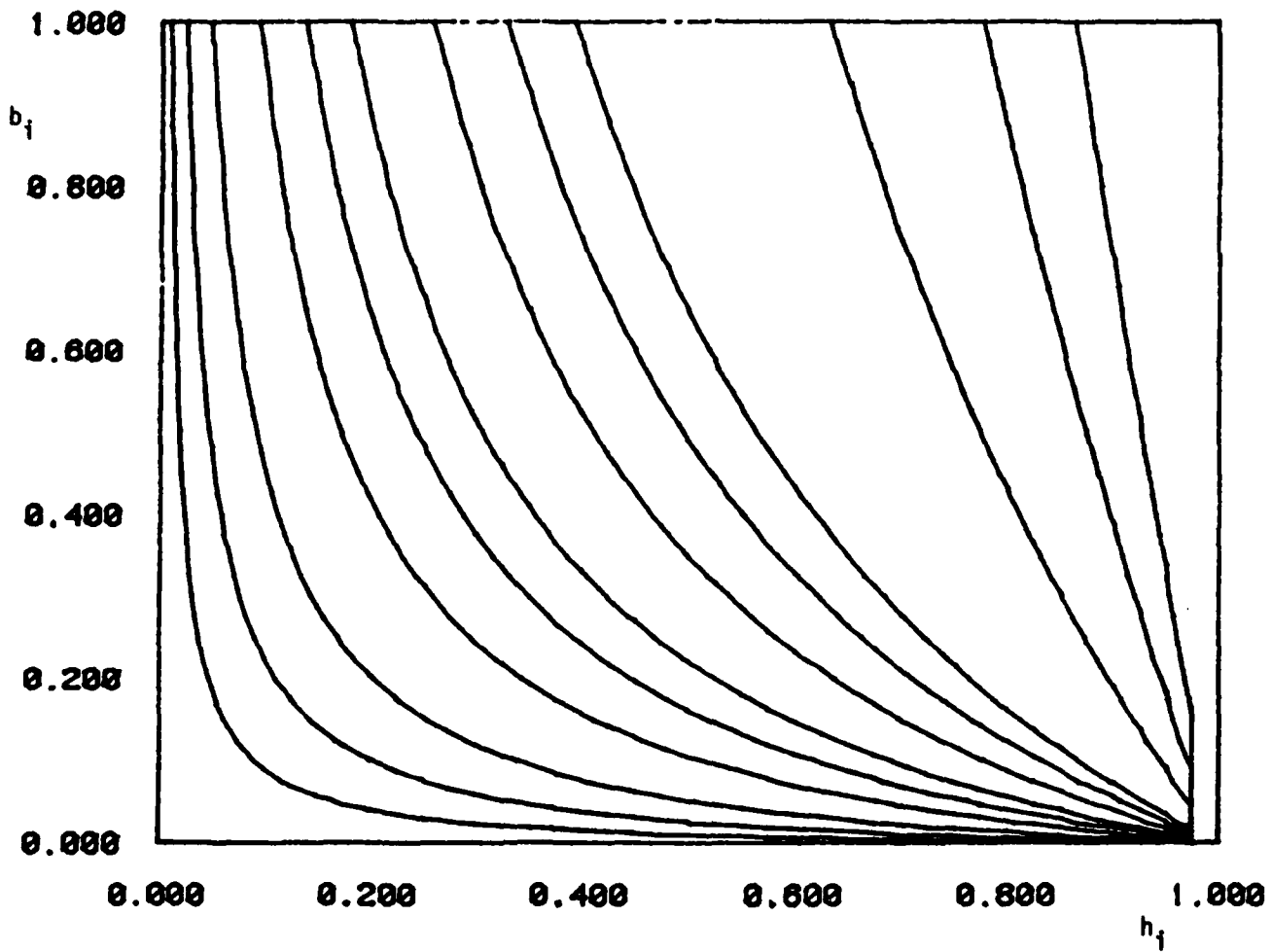


Figure 4. $LD_1(\beta|\sigma^2)$ as a function of (h_i, b_i) . Contours are as given in Figure 3.

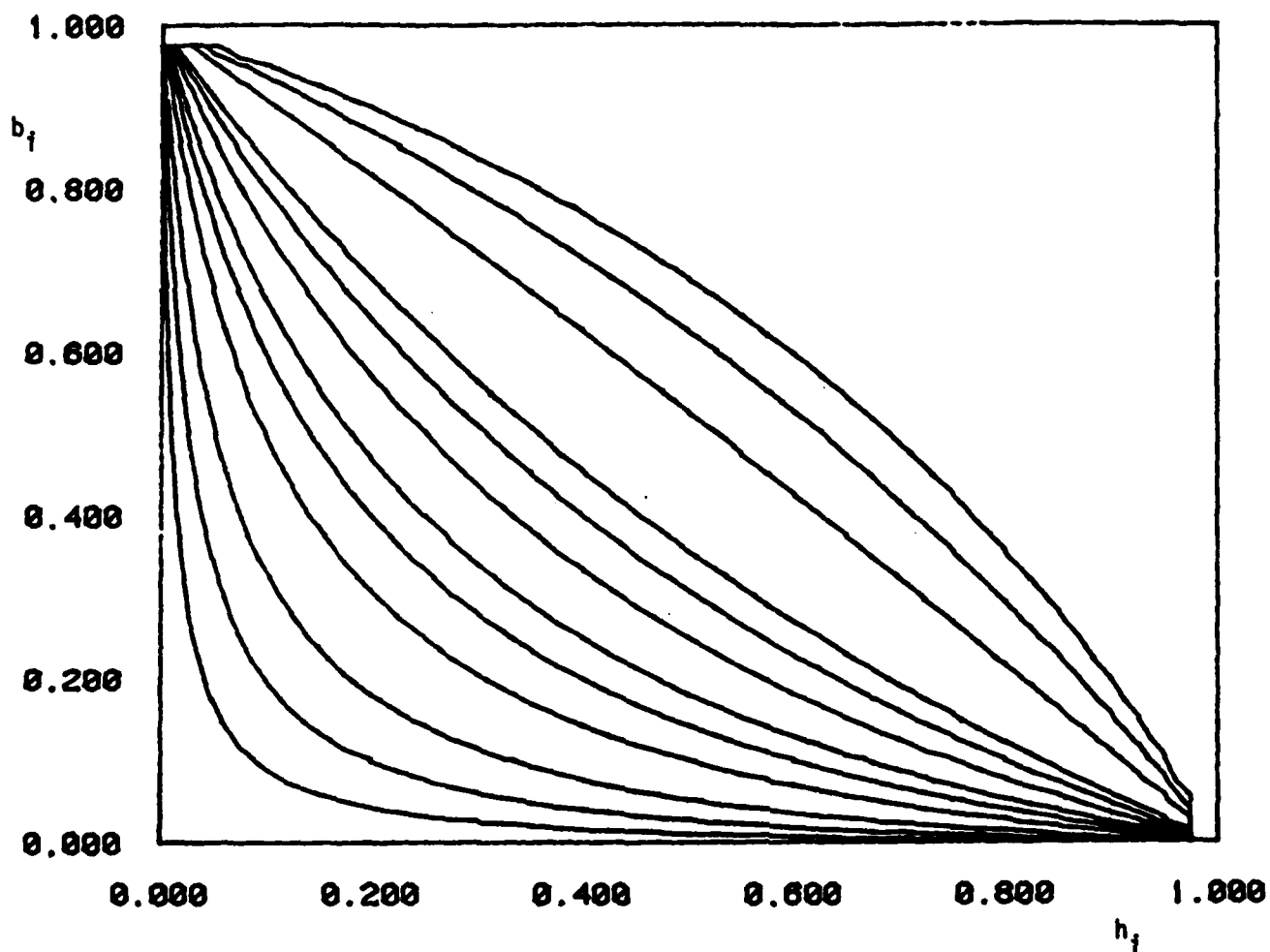


Figure 5. $LD_1(\beta, \sigma^2) - LD_1(\sigma^2 | \beta)$ as a function of (h_1, b_1) . Contours are as given in Figure 3.

Figures 3 and 5. The contours in these two figures are remarkably similar when $b_1 < h_1$; when this condition holds we can expect $\text{DFFITS}_1^2 \approx \text{LD}_1(\beta, \sigma^2)$. When $b_1 > h_1$, the two sets of contours are quite different and $\text{LD}_1(\beta, \sigma^2)$ is considerably more sensitive to increases in b_1 . Evidently, DFFITS_1 is not sufficiently sensitive to changes in scale. Numerical illustrations of this insensitivity are easily constructed. Suppose, for example, that $b_1 = .99$ so that from (8) $\hat{\sigma}_{(1)}^2 = .01 \hat{\sigma}^2$. With b_1 fixed at .99, DFFITS_1^2 can be made arbitrarily small by letting $h_1 \rightarrow 0$. Under these same conditions, however, $\text{LD}_1(\beta, \sigma^2) \rightarrow \text{LD}_1(\sigma^2 | \beta)$. This example can be used to formulate a more realistic all-points-but-one-nearly-on-a-line problem in which DFFITS_1 may fail to find the point that is far from the line.

A variety of other useful insights can be obtained by comparing Figures 3-5. For example, $\text{LD}_1(\beta | \sigma^2)$ responds primarily to h_1 while $\text{LD}_1(\sigma^2 | \beta)$ is independent of h_1 . Clearly, leverage is more important for changes in coefficients while outliers (as reflected by b_1) are important for changes in scale. When examining Figures 3-5 it should be remembered that only DFFITS_1^2 and $\text{LD}_1(\beta, \sigma^2)$ are directly comparable since the other measures concentrate on selected aspects of the problem.

Atkinson (1981) indicates a preference for measures like DFFITS_1 since they emphasize outliers more than D_1 . Relative to the likelihood displacement, such emphasis is insufficient if both β and σ^2 are of interest and is oversufficient if interest centers on β alone. Generally, Figures 3-5 show that DFFITS_1^2 lies between $\text{LD}_1(\beta, \sigma^2)$ and $\text{LD}_1(\beta | \sigma^2)$ when $b_1 > h_1$.

Welsh (1982) favors yet another measure of influence that can be written

$$w_i = \frac{(n-1)}{1-h_i} \cdot \text{DFITS}_i^2 \quad (14)$$

This measure is intended to reflect the influence of cases on location, scale and the shape of the covariance matrix. From the above discussion it seems clear that the shape information is coming at the substantial expense of information on coefficients and scale. Perhaps it is unwise to expect so much information from a single number.

DISCUSSION

Many of the initial developments in the area of influence assessment are based on ad hoc reasoning, as often happens during the infancy of any new methodology. For further progress and a deeper understanding of available methodology, larger perspectives seem necessary. We have found the likelihood displacement to be particularly well-suited for the study of influence, although other reasonable frameworks are possible, of course. For example, Johnson and Geisser (1983) adopt a predictivist view.

Within the likelihood framework, we conclude that $LD_1(\beta, \sigma^2)$ is the most useful one-number summary of influence in the absence of more specific concerns. This conclusion follows from two observations. First, $LD_1(\beta | \sigma^2)$ and $LD_1(\sigma^2 | \beta)$ are bounded above by $LD_1(\beta, \sigma^2)$. Cases that are uninformative for (β, σ^2) must therefore be uninformative for β and σ^2 considered separately. The specific concerns reflected by $LD_1(\beta | \sigma^2)$ and $LD_1(\sigma^2 | \beta)$ need to be addressed only when $LD_1(\beta, \sigma^2)$ is sufficiently large. Second, DFITS_i and related measures like Atkinson's (1981, 1982) modified Cook statistic will be essentially equivalent to $LD_1(\beta, \sigma^2)$ when $h_i > b_i$; otherwise these measures are not sufficiently sensitive to changes in scale.

Since coefficients are often a major concern in linear regression, $LD_1(\beta|\sigma^2)$ or, equivalently, D_1 can be added to give a useful two-number summary of influence. If a subset β_1 of $\beta^T = (\beta_1^T, \beta_2^T)$ is of special interest, $LD_1(\beta|\sigma^2)$ can be refined further by using the general form given in (2).

Since the three likelihood displacements considered here depend only on n , b_1 and h_1 , other summaries might include various combinations or transformations (e.g., to Studentized residuals) of these quantities. Such mixed summaries require different scales for interpretation and are therefore somewhat more difficult to comprehend than constant scale summaries. Of course, b_1 and h_1 might be useful for purposes other than an assessment of influence.

Finally, equations (12) shows one way to generalize DFFITS beyond linear models.

APPENDIX

Derivation of Equations (10) and (11)

By definition,

$$LD_1(\beta, \sigma^2) = 2[L(\hat{\beta}, \hat{\sigma}^2) - L(\hat{\beta}_{(1)}, \hat{\sigma}_{(1)}^2)]$$

where

$$L(\hat{\beta}, \hat{\sigma}^2) = -\frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2} - \frac{n}{2} \log 2\pi$$

$$L(\hat{\beta}_{(1)}, \hat{\sigma}_{(1)}^2) = -\frac{n}{2} \log \hat{\sigma}_{(1)}^2 + \frac{1}{2} \sum_{j=1}^n \frac{(y_j - x_j^T \hat{\beta}_{(1)})^2}{\hat{\sigma}_{(1)}^2} - \frac{n}{2} \log 2\pi$$

Since

$$\sum_j \frac{(y_j - x_j^T \hat{\beta}_{(1)})^2}{\hat{\sigma}_{(1)}^2} = \frac{(n-1) \hat{\sigma}_{(1)}^2 + e_{(1)}^2}{\hat{\sigma}_{(1)}^2}$$

it follows that

$$LD_1(\beta, \sigma^2) = n \log \frac{\hat{\sigma}_{(1)}^2}{\hat{\sigma}^2} + \frac{e_{(1)}^2}{\hat{\sigma}_{(1)}^2} - 1 .$$

Now, using (8)

$$\begin{aligned}
 LD_1(\beta, \sigma^2) &= n \log \frac{n}{n-1} + n \log (1-b_1) + \frac{e_{(1)}^2 (n-1)}{\hat{\sigma}_{(1)}^2 n (1-b_1)} - 1 \\
 &= n \log \frac{n}{n-1} + n \log (1-b_1) + \frac{b_1 (n-1)}{(1-b_1)(1-h_1)} - 1
 \end{aligned}$$

as given by (10).

To derive (11), by definition,

$$LD_1(\sigma^2 | \beta) = 2[L(\hat{\beta}, \hat{\sigma}^2) - L(g(\hat{\sigma}_{(1)}^2), \hat{\sigma}_{(1)}^2)]$$

Since the maximum likelihood estimator of $\hat{\beta}$ does not depend on σ^2 , $g(\hat{\sigma}_{(1)}^2) = \hat{\beta}$ and thus

$$L(\hat{\beta}, \hat{\sigma}_{(1)}^2) = -\frac{n}{2} \log \hat{\sigma}_{(1)}^2 - \frac{n \hat{\sigma}_{(1)}^2}{2 \hat{\sigma}_{(1)}^2} - \frac{n}{2} \log 2\pi .$$

Then, we obtain

$$LD_1(\sigma^2 | \beta) = n \log \frac{\hat{\sigma}_{(1)}^2}{\sigma^2} + n \left(\frac{\hat{\sigma}_{(1)}^2}{\sigma^2} - 1 \right) .$$

Equation (11) now follows from this and equation (8).

REFERENCES

- Atkinson, A.C. (1981). "Robustness, transformations and two graphical displays for outlying and influential observations in regression," Biometrika 68, 13-20.
- Atkinson, A.C. (1982). "Regression diagnostics, transformations and constructed variables (with discussion)," Journal of the Royal Statistical Society, Ser. B, 44, 1-35.
- 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. and Weisberg, S. (1982). Residuals and Influence in Regression, New York and London: Chapman-Hall.
- Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. London: Chapman-Hall.
- Dempster, A.P. and Green, M. (1981). "New tools for residual analysis," Annals of Statistics 9, 945-959.
- Hcaglin, D.C. and Welsch, R (1978). "The hat matrix in regression and ANOVA," American Statistician 32, 17-22.
- 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.
- Welsch, R.E. (1982). "Influence functions and regression diagnostics," in Launer, R.L. and Siegel, A.F., Modern Data Analysis, New York: Academic Press, pp. 149-170.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2751	2. GOVT ACCESSION NO. AD A149456	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Likelihood Displacement: A Unifying Principle for Influence Measures		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. Dennis Cook, Daniel Pena and Sanford Weisberg		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 and 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 September 1984
		13. NUMBER OF PAGES 21
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) Influential observations, Cook's distance, DFFITS, Likelihood displacement.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The young field of statistical diagnostics has produced an array of competing statistics for measuring the influence of individual cases. Two of the most popular measures for linear regression are Cook's (1977) D_i and Belsley, Kuh and Welsch's (1980) DFFITS _i . Using the likelihood displacement (Cook and Weisberg, 1982) as a unifying concept, these two measures are compared.		

END

FILMED

2-85

DTIC