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

WELSCH, R. E. and KUH, E. (1977). Linear regression diagnostics. Technical Report 923-77, Sloan School of Management, Massachusetts Institute of Technology.

WELSCH, R. E. and PETERS, S. C. (1978). Finding influential subsets of data in regression models. *Proc. Eleventh Interface Symp. Comput. Sci. Statist.* 240–244.

WOOD, F. S. (1973). The use of individual effects and residuals in fitting equations to data. *Technometrics* **15** 677–695.

Comment

R. Dennis Cook

Chatterjee and Hadi present a disturbing account of the disorientation that can result from attempting to sort through the variety of methods that are available for studying influence, leverage, and outliers in linear regression. Their admonition that the goals of an analysis must be used to guide our choice of methodology is entirely appropriate. The question "Influence on what?" is indeed important, particularly when it is asked of a specific method. I find that answers to this question can form a useful guidebook to influence methodology and can thereby remove much of the perceived confusion. With this key question in mind, Chatterjee and Hadi describe several useful distinctions between the various methods, but some confusion evidently remains, as exemplified by the all-but-one-point-on-a-line problem. For further clarity, it is necessary to take a closer look at the appropriate uses of various influence diagnostics. Beginning with a general introduction, the following discussion is intended to emphasize critical distinctions between selected methods and to further illustrate the importance of Chatterjee and Hadi's question. Unless indicated otherwise, notation is the same as that used by Chatterjee and Hadi.

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, considerations of model adequacy are extremely important. The recent paper by Freedman and Navidi (1986) in combination with the discussants' remarks provides a forceful lesson on modeling. Depending on the situation, a universally compelling demonstration of the adequacy of a model may not be possible. But what we can always do is strive for the reassurance that what we have done is sensible in light of the available information, that the data do not contradict the model or vice versa, and that reasonable alternative formulations will not lead to drastically different conclusions. How much reassurance we may need depends on the particular problem. In well studied situations where we have considerable prior information and experience, a little reassurance may be sufficient, while in fresh problems we may require much more. But some reassurance is always necessary.

Many methods are available for gaining necessary reassurance. For example, we may empirically validate a model through continued observation of the process under study or use robust methods to mitigate the impact of questionable aspects of the model. In addition, diagnostic methods should be used to look for contradictory or other relevant information in the observed data. The absence of such information will not prove that the model is accurate, but it can provide the reassurance that the model is not contradicted by available information or unduly influenced by isolated characteristics of the data.

Chatterjee and Hadi describe their experiences with a particular class of diagnostic methods that are intended to aid in assessing the role that individual observations play in determining a fitted model. A fitted model can be viewed as a smoothed representation that captures global and essential features of the data, but this view is not always appropriate. Key features of a fitted model can be dominated by a single observation and conclusions in such situations tend to depend critically on the model. It seems generally recognized that a concern for influential observations should be part of any analysis, and in recent years there has been a proliferation of methods for their detection.

2. t_i AND t_i^*

Chatterjee and Hadi discuss several reasons for preferring t_i^* over t_i, but their discussion seems to lack

R. Dennis Cook is Professor and Chair, Department of Applied Statistics, University of Minnesota, St. Paul, Minnesota 55108.

specificity since the intended use of the eventual choice is not made clear. Why are these statistics being compared? Perhaps they will be used in probability plots to assess distributional assumptions. If the goal is to test for a single outlier based on a normal mean shift alternative (see Cook and Weisberg, 1982, page 20), then t_i and t_i^* are statistically equivalent since t_i^{*2} is a monotonic function of t_i^2. In this case, the choice seems to hinge only on the availability of tables or other mechanisms for generating critical values. Under the hypothesis of no outliers, $t_i^2/(n-p)$ follows a standard beta distribution so that it may be more convenient to use t_i^*.

3. C_i AND WK_i

Judging from widely distributed regression packages, these are two of the most commonly used influence measures and, as the comments of Chatterjee and Hadi make clear, they are often viewed as competitors. However, it can be argued justifiably that they should not be compared since they measure distinctly different aspects of the influence of a single observation.

Originally developed by Cook (1975), C_i measures only the influence of a single observation on the ordinary least squares estimate $\hat{\beta}$ of the coefficient vector β in linear regression. It is the squared length of $\hat{\beta} - \hat{\beta}_{(i)}$ relative to the fixed inner product defined by $X^TX/p\hat{\sigma}^2$, although useful alternative interpretations have been put forth (see Cook and Weisberg, 1982). The comparison of C_i with the probability points of an F distribution is nothing more than a monotonic transformation to a more familiar scale and is certainly not a test of significance. A discussion of this interpretation is available in Cook's (1977b) response to Obenchain (1977). To my knowledge there are only two other influence diagnostics—$LD_i(\beta)$ as discussed below and that resulting from the Bayesian approach of Johnson and Geisser (1985)—that are intended to measure only the influence of a single observation on $\hat{\beta}$.

According to Welsch (1982), WK_i assesses the influence of a single observation on "coefficients and scale." In other words, WK_i should be viewed as a measure of influence for $\hat{\beta}$ and $\hat{\sigma}^2$ simultaneously. The essential algebraic difference between C_i and WK_i is in the use of $\hat{\sigma}^2$ (C_i) or $\hat{\sigma}^2_{(i)}$(WK_i). One consequence of this difference is that if we try to view WK_i as an influence measure for coefficients, then we are forced to the interpretation that WK_i is the (signed) length of $\hat{\beta} - \hat{\beta}_{(i)}$ relative to the variable inner product defined by $X^TX/\hat{\sigma}^2_{(i)}$. As Chatterjee and Hadi indicate, this makes WK_i difficult to interpret as a measure of influence on $\hat{\beta}$ only, since the ruler that we use to compare $\hat{\beta}$ and $\hat{\beta}_{(i)}$ changes with i. In this situation, the difficulty of interpretation reflects the inappropriateness of the interpretation.

Further, it has been argued that $\hat{\sigma}^2_{(i)}$ is preferable since it is robust to gross errors in the ith observation. When a robust estimate of scale is required and only coefficients are of interest, it seems much better to use C_i with $\hat{\sigma}^2$ replaced by a robust estimate of σ^2 that has a high breakdown point, thereby achieving a constant metric and a robust scale. For further discussion see Cook (1982) and Cook and Weisberg (1982).

4. ALL-BUT-ONE-POINT-ON-A-LINE PROBLEM

This problem has been promoted by a number of authors (e.g., Welsch 1982) as a reason for the use of WK_i over C_i. Chatterjee and Hadi use it to argue that using $\hat{\sigma}^2$ as a scale estimate can produce noninformative results. Whether any influence measure is noninformative depends on the desired information and again this leads back to the importance of a clear statement of goals.

An easy resolution of this problem can be obtained by recognizing the appropriate uses of C_i and WK_i. Figure 1 displays four points, three of which fall on a line. Point A will be selected by C_i as the most influential for the coefficients from a simple linear regression model and this is the correct choice. Deleting observation A will change the coefficients considerably, while deleting observation B, the only point falling off the line, will change the coefficients by a relatively small amount. Point B will be selected by WK_i as the most influential for $\hat{\beta}$ and $\hat{\sigma}^2$ simultaneously and again this choice seems to be correct in view of the intended use of WK_i. Although $\hat{\beta}$ changes a lot when point A is deleted, $\hat{\sigma}^2_{(B)} = 0$, clearly a dramatic change in the estimate of σ^2. This problem may be

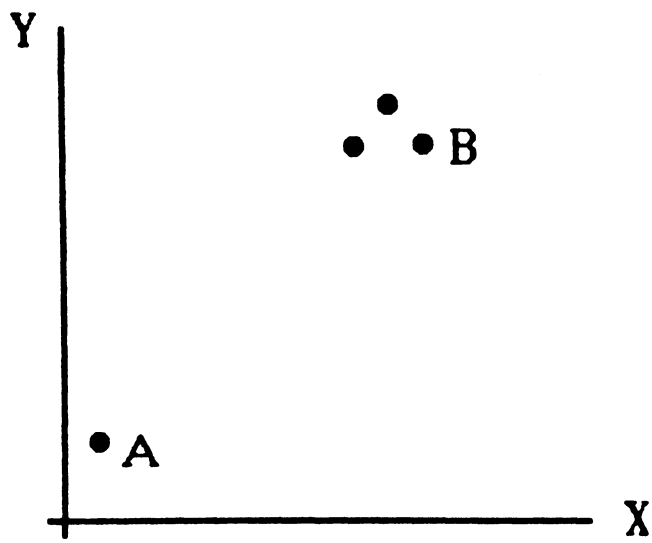

FIG. 1. *Schematic illustration of the all-but-one-point-on-a-line problem.*

used to emphasize the importance of carefully specifying the objectives of an influence analysis (i.e., whether we should emphasize coefficients, scale, a combination thereof, or some other aspect of the problem at hand), but C_i and WK_i seem to behave exactly as expected. However, as described below, WK_i does have serious deficiencies as a measure of influence for $\hat{\beta}$ and $\hat{\sigma}^2$ simultaneously.

Generally, this problem represents situations in which there are two influential observations, a high leverage observation that is in substantial agreement with the model (point A) and a low leverage observation that is not in agreement (point B). Again, judgments about the relative importance of these observations must depend on the goals of the analysis. In many situations, we will wish to identify both observations and this may require using more than one measure or a multiple deletion version of a single measure.

Various other reasons have been advanced for preferring WK_i over C_i. These arguments use little more than special pleading since they nearly always neglect any reference to the different uses underlying these statistics.

5. C_i, WK_i, AND LD_i

Likelihood displacement (a.k.a. likelihood distance) was developed by Cook and Weisberg (1982) as a unifying method for the development of influence measures. Briefly, it assesses the amount that an estimate is displaced when an observation is removed, with displacement being gauged relative to the contours of the appropriate profile log likelihood. The methodology is not restricted to linear regression and can easily be adapted to assess influence on parameter subsets, predictions, etc. Cook and Wang (1983) use this approach when developing influence measures for transformation parameters and Cook (1986) uses it as one ingredient in an extension of the notion of influence beyond the deletion of observations.

Let $b_i = t_i^2/(n - p)$. Under the normal version of the model that Chatterjee and Hadi describe in (1), b_i has a beta $(\frac{1}{2}, (n - p - 1)/2)$ distribution. The likelihood displacements for $\hat{\beta}$ only, $\hat{\sigma}^2$ only, and $\hat{\beta}$ and $\hat{\sigma}^2$ simulatenously can be written as

$$\text{(1)} \quad \begin{aligned} LD_i(\beta) &= N \log[pC_i/(N - p) + 1] \\ &= N \log[b_i h_i/(1 - h_i) + 1], \end{aligned}$$

$$\text{(2)} \quad \begin{aligned} LD_i(\sigma^2) &= N \log[N/(N - 1)] + N \log[1 - b_i] \\ &\quad + b_i(N - 1)/(1 - b_i) - 1, \end{aligned}$$

and

$$\text{(3)} \quad \begin{aligned} LD_i(\beta, \sigma^2) &= N \log[N/(N-1)] + N \log[1 - b_i] \\ &\quad + b_i(N-1)/[(1-b_i)(1-h_i)] - 1, \end{aligned}$$

respectively. Here, h_i is the ith diagonal of the hat matrix. Results (1) and (3) are given in Cook and Weisberg (1982) in different notation, while (2) and the version of $LD_i(\beta, \sigma^2)$ given in (3) are developed in Cook, Pena, and Weisberg (1984).

The discussion of likelihood displacement by Chatterjee and Hadi is seriously misleading. First, likelihood displacement is *not* based on the change in volume of confidence ellipsoids. As far as I know, there is no fundamental connection between volume change and likelihood displacement. Second, in contrast to the statement by Chatterjee and Hadi, equation (23) is *not* the likelihood displacement for measuring the influence of the ith observation on $\hat{\beta}$ only. Rather, equation (23) is the likelihood displacement for measuring the influence of the ith observation on $\hat{\beta}$ and $\hat{\sigma}^2$ simultaneously. Apart from differences in notation, equation (23) is the same as (3) above.

Several useful conclusions can be obtained from an inspection of (1)–(3). First, $LD_i(\beta)$ is a monotonic function of C_i and is therefore equivalent to C_i for the purpose of ordering observations based on influence. Second, $LD_i(\sigma^2)$ does not depend on h_i in the sense that the distribution of b_i is independent of h_i under the normal theory version of (1) in Chatterjee and Hadi. Third, $LD_i(\sigma^2) = LD_i(\beta, \sigma^2)$ when $h_i = 0$. Evidently, leverage is unimportant when assessing the influence of an observation on $\hat{\sigma}^2$ only.

Recall that WK_i measures the influence of the ith observation on $\hat{\beta}$ and $\hat{\sigma}^2$ simultaneously. Thus, WK_i and $LD_i(\beta, \sigma^2)$ may be compared without confusing different aims and for this purpose it is useful to write WK_i in the form (Cook, Pena, and Weisberg, 1984)

$$\text{(4)} \quad WK_i^2 = (n - p - 1)b_i h_i/[(1 - b_i)(1 - h_i)].$$

Comparing (3) and (4), its clear that WK_i and $LD_i(\beta, \sigma^2)$ can respond very differently to b_i and h_i. In particular, consider the data in Figure 2 which we suppose are to be described by using simple regression through the origin. Observation A is surely outlying. Since $h_A = 0$, $\hat{\beta}$ is independent of (x_A, y_A), and thus observation A can influence only $\hat{\sigma}^2$. Further, since WK_i measures influence on $\hat{\beta}$ and $\hat{\sigma}^2$ simultaneously, it is reasonable to expect it to identify observation A. This will not happen, however, since $WK_A = 0$. On the other hand, both $LD_i(\sigma^2)$ and $LD_i(\beta, \sigma^2)$ will identify observation A. The conclusion to be drawn from this example is that WK_i is not sufficiently sensitive to changes in scale. Similar comments apply to Welsch's distance W_i. As an aside, $LD_i(\beta)$ will not identify observation A, but this is entirely appropriate in view of the stated objective: $LD_i(\beta)$ measures influence on $\hat{\beta}$ only, and $\hat{\beta} = \hat{\beta}_{(A)}$.

As Chatterjee and Hadi indicate, Atkinson's C_i^* is proportional to WK_i. Thus, like WK_i, it is not sufficiently sensitive to changes in $\hat{\sigma}^2$ when viewed as a

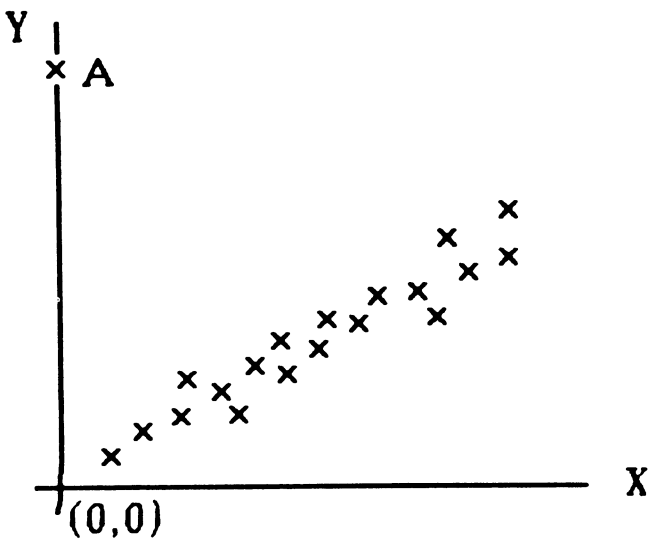

FIG. 2. *Schematic representation of simple linear regression through the origin with an outlier at* $X = 0$.

measure of influence for $(\hat{\beta}, \hat{\sigma}^2)$. The advantages attributed to C_i^* are illusionary. For example, it is true that C_i^* gives more weight to extreme values, but this idea forces the question, "What is the right amount of weight to give to extreme values?" As the previous example demonstrates, the weight used in C_i^* does not seem to be the right amount. The suitability of C_i^* for graphical displays comes from taking the square root in (38), a standard method that can be used with many influence measures, e.g., $\sqrt{C_i}$.

The data from Mickey, Dunn, and Clark (1967) can be used to illustrate the contrasting roles of these influence measures. Chatterjee and Hadi state that the point marked by an "o" in their Figure 1 is an example of an outlier that does not matter. However, this observation is identified as the most influential when using either $LD_i(\beta, \sigma^2)$ or $LD_i(\sigma^2)$, and the numerical values of these two measures indicate that the amount of influence is non-negligible. When using $LD_i(\beta)$, or equivalently C_i, to gauge influence, the observation that Chatterjee and Hadi marked with an "I" is identified as the most influential, but observation "o" is the second most influential and is clearly distinguished from the remainder of the data. From these results, the sense in which observation "o" does not matter is rather elusive. Observation "o" is the most influential when $\hat{\sigma}^2$ is of interest, either solely or in combination with $\hat{\beta}$, and has notable relative influence when interest centers on $\hat{\beta}$ alone. In the latter case, adscititious information may be required for decisions on the importance of this observation.

6. PARTIAL INFLUENCE

For the most part, the above remarks cover partial influence measures. First introduced by Cook (1975), D_{ij} is intended to measure only the influence of the ith observation on the jth coefficient. Similarly, although the intent is a little vague, it seems best to view D_{ij}^* as measure of the influence of the ith observation on $(\hat{\beta}_j, \hat{\sigma}^2)$.

As Chatterjee and Hadi emphasize, a measure that involves all coefficients can be noninformative, particularly when special importance is attached to a subset of β. This need not be the situation, however, even if we rely primarily on a measure that involves all coefficients. It has been repeatedly pointed out in varying degrees of generality (Cook, 1977b; Cook, 1979; Cook and Weisberg, 1980; Cook and Weisberg, 1982) that $D_{ij} \leq pC_i$. The important point is that we need not worry about the effects of the ith observation on the jth coefficient when pC_i is sufficiently small. We may need to compute D_{ij} only when pC_i is large enough to cause concern.

7. CONCLUSIONS

I disagree with the conclusions of Chatterjee and Hadi regarding the routine use of sufficient configurations for the detection of influential observations in linear regression. Although examining $\{WK_i, CW_i, D_{ij}\}$ will provide much useful information, I would not look forward to routinely examining $(p + 2)N$ influence values in addition to diagnostics for heteroscedasticity, outliers, curvature, etc. The recommended configurations are not sufficient to represent all of the specific goals that can arise. They do not allow us to concentrate attention on changes in $\hat{\sigma}^2$, for example. Further, as indicated above, an inspection of C_i will often show that all D_{ij}'s are negligible.

As discussed in Cook and Weisberg (1982), an alternative approach is to select a single influence diagnostic for application in every problem and, if appropriate, supplement this with a parsimonious selection from the remaining diagnostics that reflect the specific objectives in the problem at hand. I personally prefer to routinely examine $LD_i(\beta, \sigma^2)$. Since $LD_i(\beta)$, $LD_i(\sigma^2)$, and $LD_i(\beta_j)$, the likelihood displacement for $\hat{\beta}_j$ only, are bounded above by $LD_i(\beta, \sigma^2)$, I may need to worry about individual coefficients and scale only when $LD_i(\beta, \sigma^2)$ is sufficiently large. Occasionally, it is necessary to compute additional influence diagnostics that reflect more specific aims, but in many problems a single index plot of $LD_i(\beta, \sigma^2)$ is sufficient to provide the necessary reassurance that the analysis is not being dominated by a single observation. Multiple influential observations is a more difficult issue and is not covered by these remarks.

8. THE FUTURE

For the most part, the development of influence methodology for linear regression is based on ad hoc reasoning and this partially accounts for the diversity of recommendations. The value of past work is

substantial and has greatly increased my awareness of the structure of regression problems, particularly with regard to the role of individual and groups of observations. However, for progress beyond linear models and a more complete understanding of past results, ad hoc reasoning no longer seems sufficient. Competing goals must be carefully weighed and influence measures must be formulated with a broader base. Likelihood is the foundation for many analyses and in the long term we should strive for methods that directly reflect the difference between the full sample likelihood and the likelihood obtained after deletion. From a Bayesian perspective, the pioneering work of Johnson and Geisser (1982, 1983, 1985) is relevant.

Broadening the concept of influence to include more than the deletion of observations is a second direction that may prove fruitful. Deletion can be viewed as just one of many ways of perturbing a problem formulation to assess influence. Minor modifications of the values of a selected explanatory variable in linear or nonlinear regression, for example, can uncover relevant structure in the data that would not normally be detected by deletion, and lead to fresh interpretations of certain patterns in added variable plots. These and related issues are addressed in Cook (1986).

ADDITIONAL REFERENCES

COOK, R. D. (1975). Detection of influential observations in linear regression. Technical Report 256, School of Statistics, Univ. Minnesota.

COOK, R. D. (1977b). Letter to the Editor. *Technometrics* **19** 348.

COOK, R. D. (1979). Influential observations in linear regression. *J. Amer. Statist. Assoc.* **74** 169–174.

COOK, R. D. (1982). Discussion of Dr. Atkinson's paper. *J. Roy. Statist. Soc. Ser. B* **44** 28.

COOK, R. D. (1986). Assessment of local influence (with discussion). To appear in *J. Roy. Statist. Soc. Ser. B.*

COOK, R. D., PENA, D. and WEISBERG, S. (1984). The likelihood displacement: a unifying principle for influence measures. MRC Technical Summary Report 2751, Univ. Wisconsin, Madison.

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

FREEDMAN, D. A. and NAVIDI, W. C. (1986). Regression models for adjusting the 1980 census (with discussion). *Statist. Sci.* **1** 3–39.

JOHNSON, W. and GEISSER, S. (1982). Assessing the predictive influence of observations. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, eds.) 343–358. North-Holland, Amsterdam.

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

JOHNSON, W. and GEISSER, S. (1985). Estimative influence measures for the multivariate general linear model. *J. Statist. Plann. Inference* **11** 33–56.

Comment: Aspects of Diagnostic Regression Analysis

A. C. Atkinson

1. INTRODUCTION

The rapidity of acceptance of the group of techniques known as regression diagnostics is remarkable. The methods are already included in many regression packages and there are at least three books devoted to the subject. The emphasis of each book is distinct. Belsley, Kuh, and Welsch (1980) are primarily concerned with applications in economics; Cook and Weisberg (1982) are the most mathematical of the three; Atkinson (1985) includes much material on transformations. In addition, an introduction is given by Weisberg (1985, Chapters 5 and 6). Now we have the present review article by Chatterjee and Hadi. In my comments I shall go beyond the area defined by their title, to describe several recent developments which reflect important aspects of diagnostic regression analysis. An example of the use of these methods is given in Section 5.

A. C. Atkinson is Professor of Statistics, Department of Mathematics, Imperial College, Queen's Gate, London SW7 2BZ, United Kingdom.

Diagnostic procedures are essentially concerned with the detection of disagreements between the model and the data to which it is fitted. As Chatterjee and Hadi suggest, the variety of such procedures can be bewildering. There are, however, some underlying ideas which provide a framework for comparisons. A succinct summary of principles is given by Weisberg (1983). Among other aspects he stresses: 1) the relationship with score tests for parameterized departures from assumptions, 2) the importance of graphical methods, and 3) influence analysis, that is calculation of the effect of individual observations on inferences drawn from the data.

2. GENERALIZATIONS

Chatterjee and Hadi's discussion is almost entirely concerned with the normal theory linear model. Pregibon (1981) gives the extension of diagnostic methods to generalized linear models, although his detailed discussion and examples concentrate on the analysis of binary data. Chapter 12 of McCullagh and Nelder (1983), Model Checking, also describes the extension