Massachusetts Institute of Technology indicates that good subset diagnostics can be constructed in this way. If the computing is overly burdensome in a particular situation, sampling of $p$ from $n$ or simulated annealing can provide useful approximations.

Why do this at all? Some have argued that there are few real examples where clumps occur. Yet, it is easy to construct simulated examples. The reason we have few real examples is because we have no microscope with enough resolution to see the problem. When we do and still find no real examples, then we can go on to other things.

## CONCLUDING REMARKS

Perhaps the most difficult task I have undertaken (and by no means completed) in recent years is to develop regression analysis strategies for the guided-computing project at the Massachusetts Institute of Technology (Oldford and Peters, 1985). Even with a vast arsonal of diagnostics, it is very hard to write down rules that can be used to guide a data analysis. So much is really subjective and subtle. Guided computing forces us to consider chance as a possible cause in any diagnostic exploration. It is perhaps a form of controlled magical thinking (Diaconis, 1985). A great deal of what we teach in applied statistics is *not* written down, let alone in a form suitable for formal encoding. It is just simply "lore."

Progress can be made for very restricted problems detected by diagnostics. However, as soon as we try to attack influential data and collinearity, or influential data and model selection, or influential data and transformations, etc. it gets much harder. Multiplicity and simultaneity (of problems and analyses) are additional important words for statisticans to remember. They provide an incredible challenge for the future of diagnostics and statistics.

## ACKNOWLEDGMENT

This research was supported in part by National Science Foundation Grant DCR-8116778 and Army Research Office Contract DAAG29-84-K-0207.

## ADDITIONAL REFERENCES

ATKINSON, A. C. (1985). *Plots, Transformations, and Regression.* University Press, Oxford.

DIACONIS, P. (1985). Theories of data analysis: From magical thinking through classical statistics. In *Exploring Data Tables, Trends, and Shapes* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds.). Wiley, New York.

HAWKINS, D. M., BRADU, D. and KASS, G. V. (1984). Location of several outliers in multiple-regression data using elemental sets. *Technometrics* **26** 197–208.

KEMPTHORNE, P. J. (1985). Identifying rank-influential groups of observations in linear regression modeling. Memorandum NS-539, Dept. Statistics, Harvard Univ.

KEMPTHORNE, P. J. (1986). Identifying derivative-influential groups of observations in regression. Memorandum NS-540, Dept. Statistics, Harvard Univ.

KRASKER, W. S. and WELSCH, R. D. (1983). The use of bounded-influence regression in data analysis: Theory, computation and graphics. In *Computer Science and Statistics: Fourteenth Symposium on the Interface* (K. W. Heiner, R. S. Sacher, and J. W. Wilkinson, eds.) 45–51. Springer, New York.

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, New York.

MYERS, R. H. (1986). *Classical and Modern Regression with Applications.* Duxbury, Boston.

OLDFORD, R. W. and PETERS, S. C. (1985). DINDE: Towards more statistically sophisticated software. Technical Report 55, Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge.

ROUSSEEUW, P. J. (1985). A regression diagnostic for multiple outliers and leverage points. Abstract 85t-74. *Institute Math. Statist. Bull.* **14** 399.

SAMAROV, A. and WELSCH, R. E. (1982). Computational procedures for bounded-influence regression. In *COMPSTAT 1982: Proceedings in Computational Statistics* (H. Caussinus, P. Ettinger and R. Tomassone, eds.) 412–418. Physica Verlag, Wien.

# Comment

Rollin Brant

I strongly agree with the authors' characterization of the number of proposals that have been made regarding outliers and influential points as "bewildering." However, these proposals themselves comprise only a part of a much larger number of methods put forward as useful adjuncts to the criticism of regression models. As a consequence, the conscientious and up to date investigator finds that model validation can be both time-consuming and difficult, requiring the consideration of a myriad of diagnostic quantities and plots. Most difficult of all is the integration of the often fragmentary evidence provided by these procedures into a coherent set of recommendations and/or conclusions. Any attempts at cutting through any portion of this tangled web must be welcomed by all.

*Rollin Brant is Assistant Professor, Department of Applied Statistics, University of Minnesota, St. Paul, Minnesota 55108.*

The authors surely deserve credit for providing a convenient and compact summary of the currently available measures. Where their efforts fall short, however, is in failing to provide guidance in the practical application of these measures. Two issues are key in this regard. The first, which is partially addressed by the authors, is the need for a distillation of the currently chaotic mass of diagnostics down to a compact and integrated set of procedures. The second, which has received little direct attention, concerns the practical implications of diagnostic findings.

With regard to the first issue, the authors propose that "three measures are sufficient to display the major characteristics of a data set with reference to its leverage, influence, and lack of fit." My concern here is not with the number, or indeed the particular choice, but with the lack of adequate substantiation for this claim, the sole basis for which appears to be "experience with several data sets." While I do not wish to denigrate the role of practical experience in the shaping of methodology, one wonders whether the experience referred to has been sufficiently wide ranging, or even in any sense practical. If their rationale is somewhat unclear, the authors can perhaps be forgiven; current statistical theory provides little guidance in this area, for a number of reasons discussed below.

A characteristic shared by all diagnostics is their nonspecificity with regard to the ills they potentially signify. For instance, a large outlier can stem from causes ranging from the need to transform variables through misspecification of the error distribution to possible defects in data collection and reporting. A second and associated feature of most measures is the difficulty of calibrating and/or comparing procedures. For most statistical procedures, such considerations are based on operating characteristics under varying model assumptions. However, the omnibus scope of diagnostics makes it difficult to represent their aims in terms of models and loss functions. Moreover, important anomalies often arise from the behavior of the predictor variables, regarding which the usual models have little to say. Consequently, most diagnostics can provide only relative measures, i.e., rankings of cases and not measurements on scales with more general relevance. Additionally, there is little formal basis for comparing competing measures.

Since traditional theory cannot provide the sought after logical basis for assessment, we must turn to practical issues to provide such guidance. In particular, we must consider the role of diagnostics in suggesting remedial measures and/or substantial conclusions in practical situations. Guidance in these matters is distributed rather sparsely throughout the current literature. The failure to relate the development of diagnostics to the more substantial aims of investigators has had some unfortunate effects. For instance, some naive users have come to regard case deletion as the customary cure for the ills apparently indicated by diagnostic measures.

Of course, most authors have provided appropriate cautions regarding the need for further investigation into the underlying conditions which give rise to diagnostic indications before taking such definite actions. In the same places, however, such illustrative examples as are provided are usually furnished merely to illustrate the more mechanical aspects of procedures, and are rarely accompanied by examples of the type of investigation that should follow the diagnostic phase. Of course many illustrative examples are obtained "second hand," and further investigation is often hampered, although not entirely impossible.

Unfortunately, the authors' presentation makes little progress toward correcting this deficiency. One of the few substantial recommendations they make in this connection is to "collect more data." While it is certain that in many instances the potential problems uncovered by diagnostics are not amenable to statistical remedies, it is also certain that statisticians should have more to offer. Indeed, the mere investigation of diagnostic measures on their own is seldom as enlightening as it should be. Such measures are most usefully taken as pointers toward potentially interesting features of the data. Owing to the variety of forms these features may take, they are most appropriately examined through graphical exposition, which has been largely overlooked in the authors' presentation.

In the illustrative example, all measures seem to point to case 17. What can be said regarding this case, aside from recommending further investigation into the substantial background of the associated observations? The plot in Figure 1 of $x_4$ versus $x_3$ clearly exposes the nature of the peculiarity of case 17, revealing that the remaining cases are concentrated along the indicated hyperplane. Aside from suggesting the possibility of some sort of anomaly, the plot makes clear that any inference regarding the form of the regression relationship away from this region is determined strongly by this case. In short, this plot communicates much more usable information than any purely numerical measure.

This plot is further revealing when one considers the joint influence of case 17 together with its near neighbor, case 13. In Table 1, values of the generalized Cook distance measure (see Cook and Weisberg, 1982, page 136) for sets of cases are given for the 10 most influential subsets of sizes 2 and 3, revealing the overwhelming influence of the rather sparse data in this region. The authors only touch on the subject of jointly influential subsets, and summarily dismiss the issue as being primarily computational. The extent of the above tabulation, however, points out that merely calculating influence measures over a catalog of

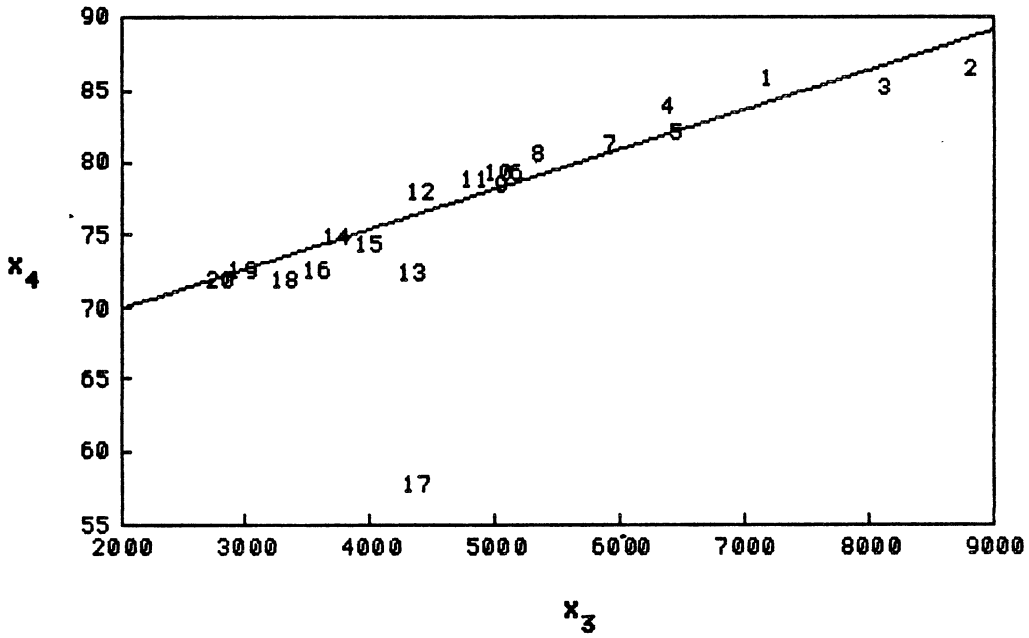

FIG. 1. *Plot of $x_4$ versus $x_3$ for Moore's data. Numbers plotted are case numbers.*

TABLE 1
*Influence measures for influential sets of size 2 and 3*

| Set | Cases | | | Generalized Cook's distance |
|---|---|---|---|---|
| a. Sets of size 2 | | | | |
| 1. | 3 | 17 | | 5.42 |
| 2. | 4 | 17 | | 2.97 |
| 3. | 6 | 17 | | 2.68 |
| 4. | 8 | 17 | | 2.48 |
| 5. | 10 | 17 | | 2.44 |
| 6. | 13 | 17 | | 2.30 |
| 7. | 14 | 17 | | 2.28 |
| 8. | 16 | 17 | | 1.99 |
| 9. | 17 | 18 | | 1.90 |
| 10. | 17 | 19 | | 1.88 |
| b. Sets of size 3 | | | | |
| 1. | 3 | 13 | 17 | 11.68 |
| 2. | 4 | 13 | 17 | 9.21 |
| 3. | 6 | 13 | 17 | 8.40 |
| 4. | 8 | 13 | 17 | 7.58 |
| 5. | 10 | 13 | 17 | 7.23 |
| 6. | 12 | 13 | 17 | 7.22 |
| 7. | 13 | 14 | 17 | 6.83 |
| 8. | 13 | 16 | 17 | 6.10 |
| 9. | 13 | 17 | 18 | 5.86 |
| 10. | 13 | 17 | 19 | 5.66 |

subsets is not of itself informative, for not all apparently influential subsets will correspond to interesting behavior in the data. Additionally, the benefits to be derived from consideration of joint influence go beyond defeating the masking effect. Such considerations can, in addition, provide clues as to the substantial relevance of influence, by helping to relate similarly influential observations. These issues are considered at greater length in Brant (1986).

Although the literature concerning diagnostics is extensive, it has yet to fully address certain vital issues. Perhaps most important is the need for integration of this and the many other varied aspects of regression and model fitting into a coherent whole. Regression methodology has undergone almost explosive growth during the past 10 years, owing in part to the ever increasing availability of computing equipment. One potentially important development that now looms on the horizon are "expert systems" for regression, which will implement diagnostic, and possibly remedial, procedures in more or less automated computer packages.

While some may argue the desirability of such systems, we can also be sure that they will appear. The heart of any such system will be a "regression strategy" for combining the various relevant methods, and it is the development of sensible strategies that is a major research problem. Fortunately, current advances in software development systems will facilitate experimentation with a variety of proposed strategies. Thus technological change, while posing new challenges, at the same time affords statisticians with exciting new opportunities. It is imperative that statisticians take up this challenge, for if we fail to do so the job will only fall to the inexpert.

## ADDITIONAL REFERENCE

BRANT, R. (1986). Finding and understanding influential sets in regression. Technical Report 466, School of Statistics, Univ. Minnesota.