Comment

A. H. Welsh

This is a stimulating and timely paper written from a practical data-analytic viewpoint. The introduction contains a careful discussion of strategies for data analysis which should be read critically by all statisticians. The difference in perspective makes this discussion a most useful supplement to the related discussions in Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986). The remainder of the paper provides an easily accessible exposition of two rank-based approaches to analyzing linear models. I would like to discuss several points which arise from the paper.

1. INFERENTIAL STRATEGIES

Draper's discussion of the widely used but naive "do-nothing" approach to dealing with violations of the classical assumptions highlights the failings of the approach. Although the more sophisticated data-analytic approach offers a potential improvement over the do-nothing approach, it is not as widely appreciated as it should be that the data-analytic approach has serious pitfalls. Essentially, problems arise from the fact that in linear model (and more complicated model) problems, much of the data analysis is based on the residuals from some preliminary parameter estimate (which is usually the least squares estimate) rather than on the observations themselves and any subsequent analysis should take this fact into account. This is particularly true of methods for outlier detection. It is a popular misconception, for example, that we can proceed to delete outliers from the sample and then apply classical least squares techniques to the reduced sample. Ruppert and Carroll (1980) showed that if we delete observations with extreme residuals and apply least squares to the reduced data set, the initial estimator (from which the residuals were calculated) has a persistent effect which does not vanish asymptotically. Consequently, the second-stage estimator can be no more efficient or robust than the initial estimator and the standard errors obtained from standard least squares formulae will be too small so that confidence intervals and tests will be misleading. The problem is that the second stage of the analysis (which is a naive least squares analysis) incorrectly ignores the effects of the preliminary analysis. The problem is not solved by basing the preliminary analysis on an initial estimator other than least squares. However, the problem can be overcome by constructing a trimmed estimator which takes the preliminary analysis into account (see Welsh, 1987a).

Incidentally, I cannot agree with Draper's view that recent developments in nonparametric regression are a part of robustness work. Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986) have argued convincingly that robustness is concerned with underlying parametric models. Although some extension of the theory away from this strict viewpoint may be possible, without the specification of a precise underlying model, the whole concept of a deviation from such a model loses its foundation. Indeed, in nonparametric regression all data points are treated as equally good and consequently receive equal weight in the analysis.

2. ASYMMETRY

The role of symmetry in the robust analysis of linear models has caused persistent confusion. Draper makes some comments on the role of symmetry which deserve clarification. In most applications of linear models, the important inferential questions involve the slope parameters. The slope parameters are identifiable (and hence can be estimated) when the errors have an asymmetric distribution. L- and M-estimators, for example, estimate the slopes when the errors have an asymmetric distribution (see Carroll and Welsh (1987) for references) and the same ought to be true of R-estimators. If symmetry has any role to play, it is in the estimation of the intercept. However, as Draper notes, in practice we can either use the ordinary median of the residuals (as advocated by Aubuchon and Hettmansperger, 1984b) as the intercept estimator or we can try to determine a more appropriate estimator based on the nature of the problem at hand. Hence, the presence of asymmetry in the errors means that we should think carefully about the choice of intercept estimator; it certainly does not mean that a robust analysis is inappropriate.

3. VARIANCE ESTIMATION

The problem of estimating the variance of R-estimators is interesting and important but still seems to have been relatively little studied. A number of questions immediately arise.

A. H. Welsh is Lecturer, Department of Statistics, The Faculties at The Australian National University, G. P. O. Box 4, Canberra A.C.T. 2601, Australia.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science.* www.jstor.org

At a basic level, there is a question as to whether, in Draper's notation, we should regard $1/\theta$ or θ as the parameter to be estimated. I have a personal preference for $1/\theta$ because this seems more natural to me and because I feel that the bias enters in a simpler way than when we take the reciprocal of an estimator of θ.

For the L_1-estimator (which is related to the R-estimator with sign scores; see (3.10) in Draper), we can construct a kernel estimator of $1/\theta$ directly (Welsh, 1987c). What is interesting about this estimator is that the shape of the kernel or window function does seem to matter as a poor choice can lead to an estimator with excessive bias. This is in conflict with the usual advice (reported by Draper) that in estimating a density, the choice of kernel is unimportant.

In evaluating competing estimates of the variance of an R-estimator, we should evaluate their properties as studentizing factors rather than as estimates of the variance per se. Although this is quite often done in simulation studies, it is not often done in theoretical investigations. However, recently Hall and Sheather (1988) derived an Edgeworth expansion for the sample median studentized by a particular variance estimator and showed that the optimal choice of smoothing parameter is different from that obtained from mean squared error considerations. In fact, their result indicated that it is important to decrease the bias more than one would if the variance was a parameter of interest. In other words, the bias/variance tradeoff is different when the density is a nuisance parameter than when it is a parameter of interest. These results are in agreement with the practical experience reported by Draper that the bias is more important than the variance in estimating $1/\theta$ (or θ).

4. *L*-, *M*- OR *R*-ESTIMATORS?

In advocating the use of R-estimators over M-estimators, Draper notes only that they often have simple, closed-form expressions. He does not mention that perhaps a more serious objection to M-estimation is that scale equivariance is usually achieved through the use of a concomitant scale estimator which may have subtle effects on the properties of the M-estimator and on the resulting inference. Now L-estimators (Welsh, 1987b; Koenker and Portnoy, 1987) have been developed further since Draper's work and they share the advantages of R-estimators. However, they have one further advantage: if the weight function is chosen to be smooth, the asymptotic variance of the resulting L-estimator is straightforward to estimate. That is, the complete analysis (including inference) is easier for L-estimators than for R-estimators. Consequently, I welcome Draper's paper for the indirect support it provides for the use of L-estimators in the linear model problem.

ADDITIONAL REFERENCES

CARROLL, R. J. and WELSH, A. H. (1987). A note on asymmetry and robustness in linear regression. *Amer. Statist.* To appear.

HALL, P. and SHEATHER, S. J. (1988). On the distribution of a studentised quantile. *J. Roy. Statist. Soc. Ser. B.* To appear.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

KOENKER, R. and PORTNOY, S. (1987). L-estimation for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

WELSH, A. H. (1987b). One-step L-estimators for the linear model. *Ann. Statist.* **15** 626–641.

WELSH, A. H. (1987c). Kernel estimates of the sparsity function. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (Y. Dodge, ed.) 369–378. North-Holland, Amsterdam.

Comment

Roger Koenker and Stephen Portnoy

David Draper's survey of rank-based robust methods for estimation and inference in linear models vividly illustrates the vitality of the R approach. The emphasis on inference is, in our view, particularly welcome, because despite the rapid growth of the foundations of robust *estimation* for linear models, the framework for robust *inference* has languished in a state of benign neglect. Certainly in applied fields like econometrics, unless we are able to suggest simple, yet reliable, robust methods of computing "those little numbers in parentheses," robust methods in general will continue to be a curiosity of the "theorists" with little impact on empirical research.

On Draper's three *desiderata* for a successful robust method: (i) intuitive appeal, (ii) unified theory and (iii) computability, we would like to offer some highly

Roger Koenker is Professor of Economics, Department of Economics and Stephen Portnoy is Professor of Statistics, Department of Statistics, University of Illinois, Champaign, Illinois 61820.