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.

subjective rankings of the four predominant robustness schools: M, R, L, and S. The tripartite division of the realm of robustness into M (*M*aximum likelihood), R (*R*ank-based) and L (*L*inear functions of order statistics) is attributable to Huber. S-estimators, a term recently coined by Rousseeuw and Yohai (1984) include the least median of squares estimator and, generically, minimize an estimate of *S*cale. We should hasten to add at this point that the rankings are motivated, in large part, by the desire to dispel the widespread belief that, as Draper puts it, "the L methods have historically been the most awkward of the three [M, R, L] in generalizing to the linear model" (Table 1).

Intuition is of course in the eye of the beholder. But it is a common complaint in robustness circles, that it is difficult "to sell" robust methods for applied work, because the methods are difficult to explain. Here a strong analogy with simple methods in the location model is helpful, and in this respect the L-estimators, particularly analogues of the trimmed means, have a strong advantage. The intuitively appealing fact that the empirical regression quantile function is estimating such a natural quantity as the population regression quantile function cannot be over emphasized. Most S-estimators, especially Rousseeuw's least median of squares (the infamous shorth of the Princeton Study of location estimators) are also intuitively appealing. The close connection between the Wilcoxon-Jaeckel-Adichie estimator for the bivariate linear model (Draper, Example 4) and the Hodges-Lehmann estimator is very appealing, but one might conclude it to be intuitively obscure, given the special treatment received by the intercept parameter in the R approach. Least appealing, in our view, from an intuitive (can-this-be-explained-successfully-to-practitioners) point of view are the M-estimators. Immediately, even in the location model, we are faced with the problem of solving nonlinear equations. Furthermore, there is the question of scale invariance. It is extremely difficult to explain to those accustomed to the natural equivariance properties of the least squares estimator why certain robust methods fail to satisfy these conditions. The contortions required to achieve scale invariance for most M-estimators are a serious drawback of these methods relative to R, L and S alternatives.

TABLE 1
A subjective ranking of the robustness sweepstakes

Method	Intuition	Theory	Computation
M	4	1	2
R	3	2	3
L	1	2	1
S	2	2	4

The classical asymptotic theory of M-, R- and L-estimators for the location model available, for example, in Serfling (1980) or Huber (1981), has been successfully extended to the linear model. The M-theory elaborated by Huber (1973) has been refined and extended in various directions and is undoubtedly best developed. The R-theory, already quite complete in Jurečková (1971) and Jaeckel (1972), has been considerably expanded, especially in the inference dimension. However, most of the attention, as Draper notes, has been focused on the Wilcoxon score function. The rudimentary L-theory in Koenker and Bassett (1978) for finite linear combinations of regression quantiles has recently been substantially generalized to cover smooth linear functionals of the regression quantile process in Koenker and Portnoy (1987) and to adaptive estimation for iid linear models in Portnoy and Koenker (1987). Adaptive estimation is especially important for the "slope" parameters in linear models when errors are asymmetric or lack a smooth density. The close interrelations among the M, R and L theories has been explored in a series of papers by Jurečková.

One serious theoretical defect shared by all three approaches is their sensitivity to influential design points. Considerable effort has been devoted to modifications of the M approach to reduce this sensitivity. Recently, Antoch and Jurečková (1985) and de Jongh, de Wet and Welsh (1987) have suggested bounded influence versions of the L approach. It is rather surprising that this issue is not raised regarding the R approach by Draper. Of course, it is on just this point that the S-estimator theory shines most brightly. However, in exchange for spectacular breakdown performance, one must bear the burden of rather intractable computation and often of rather low efficiency at the normal model.

On the practical ground of computability the L approach seems very attractive. The basic minimization problem,

$$\min_{b \in R^p} \sum \rho_\theta(y_i - x_i b)$$

with $\rho_\theta(u) = u(\theta - I(u < 0))$, is efficiently solved for all $\theta \in [0, 1]$ with elementary parametric linear programming techniques (see Koenker and d'Orey, 1987, for details). Portnoy (1988) shows that the computation of the regression quantile function requires only $O(n \log n)$ simplex pivots from the initial solution in probability, and the algorithm completely avoids iterative procedures which can fail for multimodal objective functions. This yields a p-dimensional, piecewise constant regression quantile process $\hat{\beta}_n(\theta)$ the mass points of which play the role of order statistics in the linear model L approach. General

L-estimators of the form

$$\hat{\beta}_n[J, w] = \int_0^1 J(\theta)\hat{\beta}_n(\theta)\, d\theta + \sum_{i=1}^{m} w_i\hat{\beta}_n(\theta_i)$$

can be readily computed once $\hat{\beta}_n(\cdot)$ is available. This is particularly valuable in the adaptive estimation problem where $\hat{\beta}_n(\theta)$ is used initially to construct an estimator, $\hat{J}_n(\theta)$, of the optimal L-score function, and $\hat{\beta}_n[\hat{J}_n, 0]$ may then be computed directly without further iteration. In contrast, a change in the M kernel $\psi(\cdot)$ or the R-score function, $a(\cdot)$, yields a completely new optimization problem which must be solved *de novo*. This is particularly problematic in the case of redescending $\psi(\cdot)$, and $a(\cdot)$ where the objective function is nonconvex, and further complicated in the case of M-estimators by the problem of scale estimation. A valuable discussion of computational issues in R-estimation is available in Osborne (1987), where the combinatorial complexity of the R-problem is observed to involve $n!$, because every permutation of $1, 2, \cdots, n$ is a possible ranking of the residuals for some parameter vector b. This contrasts sharply with the relatively simple (l_1-type) problems of computing the regression quantiles. Computation of S-estimators is notoriously difficult, and although some recent progress has been made, see Rousseeuw and Leroy (1987) and Souvaine and Steele (1987), the task is still daunting.

To conclude, we would like to draw attention to a intriguing parallel between L and R methods for the linear model developed in a recent thesis by Guttenbrunner (1987). For each $\theta \in [0, 1]$, the dual variables in the Koenker-Bassett formulation of regression quantiles, $\hat{\alpha} = \hat{\alpha}_{ni}(\theta)$ are defined as the solution to the linear programming problem: maximize $Y'\hat{\alpha}$ over $\hat{\alpha} \in [0, 1]^n$ subject to $X'\hat{\alpha} = (1 - \theta)X'1$. It can be shown that $\hat{\alpha}_{ni}(\theta)$ equals one if $Y_i > x_i'\hat{\beta}(\theta)$, equals zero if the inequality is reversed and is between zero and one for equality. Thus, $\hat{\alpha}$ can be used to define rank scores: let a be a measure on $[0, 1]$ and define the rank scores

$$\hat{a}_{ni} = \int_0^1 \hat{\alpha}_{ni}(\theta)\, da(\theta) + a(0).$$

Rank statistics can then be defined by

$$V_n^{a,d} = \frac{1}{n}\sum_{i=1}^{n} \hat{a}_{ni}\, d_{ni}$$

where d_{ni} are given weights. Gutenbrunner shows that the statistics $V_n^{a,d}$ are asymptotically equivalent to rank statistics of the form

$$S_n^{a,d} = \frac{1}{n}\sum_{i=1}^{n} \tilde{a}_{n\hat{R}_{ni}} d_{ni}$$

where $\hat{R}_{ni}$ is the rank of the residuals $e_i(\hat{\beta})$ based on the Jaeckel form of the regression parameter estimates (see Draper, (3.4)), and

$$\tilde{a}_{ni} = n\int_{(i-1)/n}^{i/n} da(\theta).$$

The statistics $V_n^{a,d}$ appear to be quite natural, and they are relatively easy to compute. These ideas deserve substantially more study and analysis, and may lead to a better understanding of the application of rank concepts in linear models.

ADDITIONAL REFERENCES

ANTOCH, J. and JUREČKOVÁ, J. (1985). Trimmed least squares estimator resistant to leverage points. *Comput. Statist. Q.* **2** 329–339.

DE JONGH, P. J., DE WET, T. and WELSH, A. H. (1987). Mallows-type bounded influence regression trimmed means. Institute of Statistics Mimeo Series 1719, Univ. North Carolina.

GUTTENBRUNNER, C. (1987). Zur Asymptotik von Regressionquantilprozessen und daraus abgeleiten Statistiken. Ph.D. thesis, Faculty of Mathematics, Univ. Freiburg.

KOENKER, R. and D'OREY, V. (1987). Computing regression quantiles. *Appl. Statist.* **36** 383–393.

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

OSBORNE, M. (1987). The reduced gradient algorithm. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (Y. Dodge, ed.). North-Holland, Amsterdam.

PORTNOY, S. (1988). Asymptotic behavior of the number of regression quantile breakpoints. Unpublished manuscript.

PORTNOY, S. and KOENKER, R. (1987). Adaptive L-estimation for linear models. *Ann. Statist.* To appear.

ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

ROUSSEEUW, P. and YOHAI, V. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SOUVAINE, D. L. and STEELE, J. M. (1987). Time- and space-efficient algorithms for least median of squares regression. *J. Amer. Statist. Assoc.* **82** 794–802.