In: DIRECTIONS IN ROBUST STATISTICS AND DIAGNOSTICS, PART II,
eds. W. Stahel and S. Weisberg. The IMA Volumes in Mathematics
and its Applications, Vol. 34, Springer-Verlag, New York (1991).

ROBUSTNESS OF THE p-SUBSET ALGORITHM FOR REGRESSION WITH HIGH BREAKDOWN POINT

PETER J. ROUSSEEUW* AND GILBERT W. BASSETT JR.†

Abstract. Regression techniques with high breakdown point can withstand a substantial amount of outliers in the data. One such method is the least trimmed squares estimator. Unfortunately, its exact computation is quite difficult because the objective function may have a large number of local minima. Therefore, we have been using an approximate algorithm based on p-subsets. In this paper we prove that the algorithm shares the equivariance and good breakdown properties of the exact estimator. The same result is also valid for other high-breakdown-point estimators. Finally, the special case of one-dimensional location is discussed separately because of unexpected results concerning half samples.

1. Multiple Regression with High Breakdown Point.

In the linear regression model

$$y_i = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + e_i \qquad \text{for } i = 1, \ldots, n \tag{1.1}$$

which may also be written as

$$y_i = x_i\beta + e_i \qquad \text{for } i = 1, \ldots, n \tag{1.2}$$

where $x_i = (x_{i1}, \ldots, x_{ip})$, there exist many procedures to estimate the unknown parameter vector $\beta = (\beta_1, \ldots, \beta_p)^t$. The commonly used least squares estimate is not robust at all, because it may be changed an arbitrarily large amount due to a single outlier (x_k, y_k). We say that (x_k, y_k) is a **regression outlier** if it does not agree with the linear pattern set by the majority of the data. Moreover, (x_k, y_k) is called a **leverage point** if x_k sticks out of the p-variate point cloud formed by the majority of the x_i. Note that not all regression outliers are leverage points, and that not all leverage points are regression outliers (see, e.g., Chapter 1 of Rousseeuw and Leroy 1987). However, the worst outliers are those that are also leverage points, especially if several of them form a cluster, in which case they cannot be identified by most robust estimators and outlier diagnostics. This unpleasant phenomenon is called the masking effect.

To prevent masking, one needs a method that can cope with a large fraction of outliers. We will focus our discussion on one such method, and generalize our results in Section 3. The **least trimmed squares estimator** (LTS) proposed by Rousseeuw (1983) is defined as the vector $T = (\beta_1, \ldots, \beta_p)^t$ minimizing the following objective:

$$\underset{\beta}{\text{minimize}} \sum_{i=1}^{h} (r^2(\beta))_{i:n}. \tag{1.3}$$

*U.I.A., Vesaliuslaan 24, B-2650 Edegem, Belgium

†Department of Economics (M/C 144), University of Illinois at Chicago, Box 4348, Chicago, Illinois 60680

Here, the residuals are $r_i(\beta) = y_i - x_{i1}\beta_1 - \cdots - x_{ip}\beta_p$. The residuals are first squared and then ordered, and $(r^2)_{i:n}$ denotes the i-th of these ordered values, hence $0 \leq (r^2)_{1:n} \leq (r^2)_{2:n} \leq \cdots \leq (r^2)_{n:n}$. Note that the sum contains only h terms rather than n. Therefore the largest residuals are not included in the criterion, which allows the fit to stay away from the outliers.

Clearly h has to be larger than p, because we can make any p residuals equal to zero by fitting p points exactly. In the remainder of the paper we will always assume that $n > 2p$.

If we choose h appropriately (about $n/2$, as we will see below) then the LTS estimator has a breakdown point of roughly 50%, which means that it remains bounded when up to half of the data points (x_i, y_i) are replaced by arbitrary values. Because of this desirable breakdown property, the LTS can be used as a method for detecting discrepant points (Rousseeuw 1984) and as a first step towards robust estimators with a high asymptotic efficiency (Rousseeuw and Yohai 1984, and Yohai 1987).

2. The p-subset Algorithm and its Breakdown Point. The objective function of the LTS estimator is not convex and has numerous local minima, so it is difficult to find the exact minimum of (1.3). For this reason, Rousseeuw and Leroy (1987) implemented the following approximate algorithm. Let J be any collection of p distinct indices from $\{1, \ldots, n\}$. The corresponding **p-subset estimate** β_J is the parameter vector which fits these p observations exactly:

$$y_i = x_i\beta_J \qquad \text{for all } i \text{ in } J. \tag{2.1}$$

This β_J exists and is unique whenever the $p \times p$ matrix formed by these x_i is of full rank. The objective function

$$\sum_{i=1}^{h} (r^2(\beta_J))_{i:n} \tag{2.2}$$

is then minimized over all such p-subsets J. As the total number C_n^p of p-subsets is finite, there always exists a solution $T^* = \beta_J$ to

$$\underset{J}{\text{minimize}} \sum_{i=1}^{h} (r^2(\beta_J))_{i:n}. \tag{2.3}$$

The estimator T^* does not necessarily coincide with the actual LTS because the minimization is not over all possible β. However, the restriction to the β_J is quite natural, and p-subset estimates already appear in the literature on L_1 regression (see Bassett 1988 for a survey), regression quantiles (Koenker and Bassett 1978), the median slope estimator of Theil (1950), and the repeated median (Siegel 1982). It is interesting to note that the p-subset estimates β_J occur under quite different names: in Rousseeuw and Leroy (1987) they are called **trial estimates**, whereas Hawkins, Bradu and Kass (1984) refer to them as **elemental regressions** in a diagnostics

setting. In both situations one restricts attention to a collection of random subsets J when C_n^p is too large.

As T^* is an approximation to the LTS, the question naturally arises whether it retains the good properties of the latter. By this we mean whether T^* **considered as an estimator** possesses the same equivariance and breakdown point as the LTS itself. The purpose of our paper is to show that this is indeed the case. Concerning equivariance, it is trivial to adapt the reasoning on page 117 of Rousseeuw and Leroy (1987) to show that T^* is **regression equivariant**

$$T^*(\{x_i, y_i + x_i v); i = 1, \ldots, n\}) = T^*(\{x_i, y_i); i = 1, \ldots, n\}) + v$$

for any column vector v, as well as **scale equivariant**

$$T^*(\{x_i, cy_i); i = 1, \ldots, n\}) = cT^*(\{x_i, y_i); i = 1, \ldots, n\})$$

for any constant factor c, and **affine equivariant**

$$T^*(\{x_i A, y_i); i = 1, \ldots, n\}) = A^{-1}T^*(\{x_i, y_i); i = 1, \ldots, n\})$$

for any nonsingular $p \times p$ matrix A.

To investigate the breakdown behavior of T^* we use a finite-sample version of the breakdown point as introduced by Donoho and Huber (1983). Take any data set Z of n points $(x_1, y_1), \ldots, (x_n, y_n)$. Let $\text{bias}(m; T^*, Z)$ be the supremum of all $\|T^*(Z') - T^*(Z)\|$ where Z' ranges over all contaminated data sets obtained by replacing any m points of Z by arbitrary values. Then the breakdown point of T^* at Z is

$$\varepsilon_n^*(T^*, Z) = \min\left\{\frac{m}{n}; \text{bias}(m; T^*, Z) = \infty\right\}. \tag{2.4}$$

We prefer replacing observations to adding observations, which some authors do, because in our opinion replacement contamination is simple, realistic, and generally applicable. From an intuitive point of view, outliers are not some faulty values that are added to the sample, but they treacherously hide themselves by replacing some of the data that should have been observed. Replacement contamination also avoids formal problems because the contaminated sample Z' has the same size as the original Z, so one only has to consider one estimator T_n instead of several T_{n+m}. Therefore, replacement still applies to many structured situations where adding observations would be difficult (e.g., one cannot just add cells to a two-way table).

In what follows we shall assume that the original sample Z is in general position, which is equivalent to saying that any p-subset J yields a unique estimate β_J. For instance, when $p = 2$ this means that any pair of points (x_{i1}, x_{i2}, y_i) and (x_{j1}, x_{j2}, y_j) determines a unique nonvertical plane through the origin, and hence $(0, 0, 0)$, (x_{i1}, x_{i2}, y_i), and (x_{j1}, x_{j2}, y_j) cannot be collinear. This assumption is quite important, because when Z is far from general position (e.g. when many

points coincide, or when the design matrix is very sparse) it is impossible for any equivariant estimator to achieve a high breakdown point.

The LTS attains its best possible breakdown point when h (the number of terms) is taken equal to

$$h = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right] \tag{2.5}$$

where $[\ldots]$ denotes the integer part. In this way the LTS has breakdown point $([(n-p)/2]+1)/n$, as does Siegel's (1982) repeated median. It was proved in (Rousseeuw 1984) that this is the maximal breakdown point for all regression equivariant estimators. We shall now show that the approximation T^* has exactly the same breakdown point as the LTS itself, under the same conditions.

THEOREM 1. *If $p > 1$ and the data set Z is in general position, then the breakdown point of T^* with h given by (2.5) is*

$$\varepsilon_n^*(T^*, Z) = \left(\left[\frac{n-p}{2}\right] + 1\right)/n. \tag{2.6}$$

The proof of this result can be found in the Appendix. The following corollary tells us that the exact fit property of the LTS also holds for T^*.

COROLLARY ("EXACT FIT PROPERTY"). *If $p > 1$ and there exists some β such that strictly more than $\frac{1}{2}(n+p-1)$ of the observations satisfy $y_i = x_i\beta$ and are in general position, then T^* equals β, whatever the other observations are.*

For instance, in the case of simple regression it follows that whenever 11 out of 20 points lie on one line, this line will be obtained. Some remarks on the relation between breakdown and the exact fit property can be found in Donoho et. al. (1985) and Rousseeuw and Leroy (1987, page 123).

3. Extensions. Some people reject the idea of an estimator like LTS because they intuitively dislike its high breakdown point, saying something like "I don't want to throw away half of my data". Actually, the LTS does not throw away anything, but rather it identifies the linear pattern formed by the majority of the data, and allows discrepant data points (of which there may be just a few) to show up. For a one-dimensional batch of numbers the sample median does something similar, and we never heard anyone complain about the breakdown point of the sample median being too high (it is 50% as well).

If one does want a lower breakdown point (perhaps to obtain a better efficiency in case the data are outlier-free) this is quite easy to achieve by changing the value of h. Setting h equal to

$$h = [(1-\alpha)n]$$

(for any $0 < \alpha < 1/2$, such as $\alpha = 25\%$) will yield a breakdown point tending to α when n becomes large. It takes only a small adaptation of the proof of Theorem 1

to show that the p-subset approximation T^* with the same h will then have exactly the same breakdown point.

Similar results hold for the regression estimator defined by

$$\underset{\beta}{\text{minimize}}(r^2(\beta))_{h:n} \tag{3.1}$$

which is a variant of least median of squares (LMS). In (3.1) we minimize the h-th ordered squared residual and achieve the same breakdown point as before, if we again choose h as in (2.5). The LMS can also be approximated by minimizing the objective function over the set of all p-subset trial estimates β_J as in (2.3). By Theorem 1, the resulting T^* inherits the breakdown point of the exact LMS estimator.

The p-subset algorithm for the LTS takes somewhat more computation time than that for the LMS. Indeed, for the LTS the computation of the objective function at each β_J takes $0(n\log(n))$ steps because the squared residuals have to be ranked (sorted), whereas the objective function in (3.1) needs only $0(n)$ because the median of an array of n numbers can be found in $0(n)$ steps. The resulting difference in speed was investigated by van Zomeren (1987), who ran both methods and found that the LTS consumed just a little more computation time than the LMS.

On the other hand, the LTS objective function turns out to be more stable, and unlike the LMS it yields the usual convergence rate of $n^{1/2}$, so we currently prefer the LTS to the LMS. The extra computation time is not important relative to modern standards, and the present generation of desktop computers is a lot faster than those of 1983, when we wrote our original p-subset program (Rousseeuw and Leroy 1984). On the other hand, the computation of regression S-estimators (Rousseeuw and Yohai 1984) does take substantially more time.

Already in Rousseeuw and Leroy (1984) the p-subset algorithm was refined by computing the intercept separately, by means of an **exact** algorithm for the one-dimensional case. This we do as follows: rather than using β_J as it stands, we keep the slope parameters and merely replace the intercept by running a one-dimensional algorithm on the residuals, thereby making the objective function still smaller. Such exact one-dimensional algorithms exist for both LTS and LMS (see Rousseeuw and Leroy 1987, pages 169, 171, and 201). As described in that book, the adjustment can be made for each trial estimate β_J, or just at the end if one wants to save computation time. Note that Theorem 1 holds for all these variants.

Finally, it should be noted that the p-subset estimator can also be used as a starting point for estimators with higher efficiency such as a one-step M-estimator or a one-step GM-estimator, as proposed by Hampel et al. (1986, page 330).

4. Special Results in the Location Model. By considering the linear model (1.1) with $p = 1$ and all $x_{i1} = 1$ we obtain the location model

$$y_i = \beta + e_i \qquad \text{for } i = 1, \ldots, n \tag{4.1}$$

in which the data is just a one-dimensional sample $\{y_1, \ldots, y_n\}$ and we want to estimate β by some estimator T. For this we can again use the LTS estimator (1.3),

keeping in mind that the residuals become $r_i = y_i - T$. Because $p = 1$, the number of terms is now

$$h = \left[\frac{n}{2}\right] + 1 \tag{4.2}$$

We can also extend the LMS of (3.1) to this model, by using the same h as in (4.2). In this framework both the LTS and the LMS have breakdown point

$$\varepsilon_n^* = \left[\frac{n+1}{2}\right]/n \tag{4.3}$$

(Rousseeuw and Leroy 1987, pages 183–185) which is the upper bound for all translation equivariant estimators.

In contrast with the general regression situation, the LMS and the LTS can now be computed explicitly, as described by Rousseeuw and Leroy (1987, pages 169 and 171). For the LMS one first has to order the data to $y_1 \leq y_2 \leq \cdots \leq y_n$. Then one determines the smallest of the differences

$$y_h - y_1, y_{h+1} - y_2, \ldots, y_n - y_{n-h+1}.$$

The corresponding subset containing h observations is then called the **shortest half sample**, because of all the possible subsets with h elements it possesses the shortest range. The LMS then simply equals the midpoint of this shortest interval. (This is how PROGRESS computes the LMS for one-dimensional data.) The LTS is determined in a similar way.

Although in this case the LMS and the LTS can be computed exactly, we would still like to know how the p-subset algorithm behaves. Because $p = 1$, the p-subset estimates correspond to the sample values y_i. Therefore, T^* minimizes the objective function of the LMS (or the LTS) over the set of sample values. It should be noted that we are **not** advocating to use T^* in actual applications, because this computation is of a higher complexity than the exact algorithm based on the "shortest half" approach. Instead, we want to investigate how these estimators differ.

Since the ordinary LMS (which we will from now on denote by T) is the midpoint of a half sample, and since a midpoint will not generally be identical with any given sample value, it follows that T and T^* are not the same estimator.

In the general linear model we have proven that T and T^* have the same breakdown point, by Theorem 1 above. Does T^* still have the same breakdown point as T in the location model? We cannot use the proof of Theorem 1, which needed the assumption $p > 1$. However, it can be shown directly that the breakdown point of T^* equals (4.3). The following short proof was given by Werner Stahel (personal communication). When fewer than $[(n+1)/2]$ points are replaced, the contaminated sample still contains h of the original values, which we denote by $y_1 \leq y_2 \leq \cdots \leq y_h$. Therefore, the median squared residual at $\beta = y_1$ is at most equal to $(y_h - y_1)^2$. From this it follows that T^*, at which the median squared residual cannot be larger,

must belong to the interval $[y_1 - (y_k - y_1),\ y_k + (y_k - y_1)]$ which has at most three times the range of the original data set.

While T and T^* are not identical, it might be expected that they would both lie in the shortest half. This however is not the case, as shown by the following.

Example. Consider the sample $\{4, 5, 23, 24, 39, 45, 64\}$ with $n = 7$. The shortest half (containing four observations) is the interval $[4, 24]$ hence the LMS equals $T = 14$, the midpoint of that interval. For the corresponding p-subset estimator we evaluate the median squared residual at each sample point. This is shown in Table 1, where T^* is found to be 39. Therefore T^* lies outside the shortest half! In Table 1 we see that the p-subset version of the LTS also equals 39, and hence lies outside the shortest half as well.

Table 1. Computations to obtain the p-subset version of LMS and LTS for the one-dimensional sample $\{4, 5, 23, 24, 39, 45, 64\}$.

y_i	Residuals	Median of Squared Residuals	Trimmed Sum of Squared Residuals
4	0,1,19,20,35,41,60	$(20)^2$	762
5	-1,0,18,19,34,40,59	$(19)^2$	686
23	-19,-18,0,1,16,22,41	$(18)^2$	581
24	-20,-19,-1,0,15,21,40	$(19)^2$	587
39	-35,-34,-16,-15,0,6,25	$(16)^2$	517
45	-41,-40,-22,-21,-6,0,19	$(21)^2$	838
64	-60,-59,-41,-40,-25,19,0	$(40)^2$	2586

Although T^* does not always lie inside the shortest half, it will be shown to be within a bounded distance of the shortest half. Let us denote the ordered data by $y_1 \leq y_2 \leq \cdots \leq y_n$. The LMS objective function will be denoted by

$$m(\beta) = \{(y_i - \beta)^2\}_{h:n}.$$

The actual LMS estimator T is then defined as the value of β which minimizes $m(\beta)$. The LMS can also be defined in terms of half samples. A half sample is any subset of H of $\{y_1, y_2, \ldots, y_n\}$ containing h elements, and its length is

$$\begin{aligned} \text{len}(H) &= \max\{|y_i - y_j|; \quad y_i \text{ and } y_j \text{ belong to } H\} \\ &= \max_H y_j - \min_H y_i \end{aligned}$$

Let $H_{\mathbf{min}}$ denote the half sample which minimizes len(H). The LMS can then be written as the midpoint of $H_{\mathbf{min}}$:

$$T = \left(\min_{H_{\mathbf{min}}} y_i + \max_{H_{\mathbf{min}}} y_j \right) /2$$

The p-subset estimator T^* minimizes $m(\beta)$ on the domain of the sample values (that is, for $\beta = y_i, \quad i = 1, \ldots, n$).

THEOREM 2.

$$|T - T^*| \leq 3\sqrt{m(T^*)} = \frac{3}{2} len(H_{\mathbf{min}}).$$

The proof of this result is given in the Appendix. Note that neither T or T^* need to be unique, because the theorem holds for any pair of solutions T and T^*. Also note that this theorem gives a second proof of the fact that T and T^* have the same breakdown point.

Remark. The inequality in Theorem 2 cannot be improved. This follows from the example $\{-1, -1, 1, 1, 3-\varepsilon, 3.1, 5\}$ for which $T = 0$, $T^* = 3-\varepsilon$, $m(T) = 1$, and len$(H_{\mathbf{min}}) = 2$. The inequality becomes an equality when ε tends to zero.

5. Appendix: Proof of Theorems 1 and 2.

Proof of Theorem 1. (a) Let us first show that $\varepsilon_n^*(T^*, Z) \geq ([(n-p)/2] + 1)/n$. We consider the set of p-subset estimates

$$\{\beta_J; \quad J \text{ is a } p\text{-subset of } 1, \ldots, n\} \tag{5.1}$$

obtained from Z, which contains a solution β to (2.3). Without loss of generality, we may assume $\beta = 0$ (using regression equivariance). Consider the largest absolute residual of any point of Z with respect to any trial estimate:

$$M: = \max_J \max_{1 \leq i \leq n} |r_i(\beta_J)| < \infty.$$

Since $\beta = 0$, we have $M \geq |y_i|$ for all i. Let us now consider the $(p+1)-$dimensional space containing the observarions (x_i, y_i), and its horizontal hyperplane through the origin, denoted by $(y = 0)$. Put

$$\delta = \frac{1}{2} \inf\{ \tau > 0; \quad \text{there exists a } (p-1)-\text{dimensional subspace } V \text{ of } (y = 0), \text{ going through the origin, such that } V^\tau \text{ contains at least } p \text{ of the } x_i\}$$

where V^τ is the set of all x with distance to V not larger than τ. Because Z is in general position, it holds that $\delta > 0$.

We now have to show that T^* remains bounded whenever at most $[(n-p)/2]$ points are changed. For this purpose, construct a contaminated sample $Z' = \{(x_i, y_i); \quad i = 1, \ldots, n\}$ by retaining $n - [(n-p)/2] = [(n+p+1)/2]$ points of

Z, which will be called the "good" points, and by replacing the others by arbitrary values. The remainder of the proof will be devoted to showing that

$$||\beta'|| < (1 + \sqrt{h})M/\delta$$

where β' corresponds to Z', which is sufficient because the right hand member is a finite constant. Suppose the opposite is true, hence $\beta' \neq 0$. Denote by H the hyperplane given by the equation $y = x\beta'$. It follows that $H \neq (y = 0)$ so their intersection $H \cap (y = 0)$ has dimension $p-1$. Therefore, at most $p-1$ of the "good" x_i can lie in $(H \cap (y = 0))^\delta$. Define A as the set of remaining good observations, containing at least $[(n + p + 1)/2] - (p - 1)$ points. Now consider any (x_a, y_a) belonging to A, and put $r_a(\beta') = y_a - x_a\beta'$ so

$$|r_a(\beta')| = |x_a\beta' - y_a| \geq | \, |x_a\beta'| - |y_a| \, |.$$

As in Rousseeuw (1984, page 878) it follows from $||\beta'|| \geq (1 + \sqrt{h})M/\delta$ that

$$|x_a\beta'| > \delta||\beta'|| \geq (1 + \sqrt{h})M > |y_a|$$

so

$$|r_a(\beta')| \geq |x_a\beta'| - |y_a| > (1 + \sqrt{h})M - M = M\sqrt{h}$$

hence

$$r_a^2(\beta') > M^2 h.$$

Since $n - |A| \leq h - 1$, any set of h points of Z' must contain at least one of the (x_a, y_a), so

$$\sum_{i=1}^{h} ((y_i' - x_i'\beta')^2)_{i:n} \geq r_a^2(\beta') > M^2 h. \tag{5.2}$$

However, the set of "good" observations in Z' contains at least p points because $[(n + p + 1)/2] \geq p$. Therefore, the p-subset solution β_J through these p points belongs to the original set (5.1), and at least $[(n+p+1)/2] \geq h$ of its residuals are the same as before, hence

$$\sum_{i=1}^{h} ((y_i' - x_i'\beta_J)^2)_{i:n} \leq M^2 h,$$

contradicting (5.2). Note that nowhere in this proof we have used uniqueness of either β or β', so it is generally true that $||\beta - \beta'||$ remains bounded for any pair of solutions β and β'.

(b) The opposite inequality $\varepsilon_n^*(T^*, Z) \leq ([(n - p)/2] + 1)/n$ holds for any regression equivariant estimator by Theorem 4 on page 125 of Rousseeuw and Leroy (1987). □

Proof of Theorem 2. We first establish

$$|T - T^*| \leq \sqrt{m(T)} + \sqrt{m(T^*)}. \tag{5.3}$$

The subset $\{y_i\,; (y_i - T^*)^2 \leq m(T^*)\}$ contains at least h points, so it has at least one point in common with $H_{\min}$. This point we call y_k, and it satisfies $|y_k - T| \leq \sqrt{m(T)}$ and $|y_k - T^*| \leq \sqrt{m(T^*)}$. Adding and using the triangle inequality gives (5.3).

Next, since T^* minimizes $m(\beta)$ on the set of sample values, we know that $m(T^*) \leq m(y_j)$ for any y_j in $H_{\min}$. Because $H_{\min}$ contains h points it follows that $m(y_j) \leq \max\{(y_i - y_j)^2;\ y_i \text{ and } y_j \text{ belong to } H_{min}\} = (\text{len}(H_{\min}))^2$. Therefore

$$\sqrt{m(T^*)} \leq \text{len}(H_{\min}) = 2\sqrt{m(T)} \tag{5.4}$$

Substituting (5.4) into (5.3) proves Theorem 2. □

Acknowledgement. The question about the robustness of the p-subset algorithm arose in a conversation with Peter Huber. The example where T^* does not lie in the shortest half was constructed in response to a conjecture of Roger Koenker. We are grateful to Werner Stahel for helpful comments improving the presentation and for proposing to combine our results in a single paper.

REFERENCES

BASSETT, G.W., *A p-Subset Property of L_1 and Regression Quantile Estimates*, Computational Statistics and Data Analysis 6 (1988), pp. 297–304.

DONOHO, D.L., AND HUBER, P.J., *The Notion of Breakdown Point, in A Festschrift for Erich Lehmann*, edited by P. Bickel, K. Doksum, and J.L. Hodges, Jr., Belmont: Wadsworth (1983).

DONOHO, D.L., JOHNSTONE, I., ROUSSEEUW, P.J., AND STAHEL, W., *Discussion on "Projection Pursuit"* by P. Huber, Annals of Statistics, 13 (1985), pp. 496–500.

HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J., AND STAHEL, W.A., *Robust Statistics*, New York: John Wiley (1986).

HAWKINS, D.M., BRADU, D., AND KASS, G.V., *Location of Several Outliers in Multiple Regression Data using Elemental Sets*, Technometrics, 26 (1984), pp. 197–208.

KOENKER, R., AND BASSETT, G.W., *Regression Quantiles*, Econometrica, 46 (1978), pp. 33–50.

ROUSSEEUW, P.J., *Regression Techniques with High Breakdown Point*, The IMS Bulletin, 12 (1983), p. 155.

ROUSSEEUW, P.J., *Least Median of Squares Regression*, Journal of the American Statistical Association, 79 (1984), pp. 871–880.

ROUSSEEUW, P.J., AND LEROY, A., *PROGRESS: A Program for Robust Regression Analysis*, Technical Report 201, Centrum STOO, Vrije Universiteit Brussel, Belgium (1984).

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

ROUSSEEUW, P.J., AND YOHAI, V.J., *Robust Regression by Means of S-estimators, in Robust and Nonlinear Time Series Analysis* edited by J. Franke, W. Härdle, and R.D. Martin, Lecture Notes in Statistics No. 26, New York: Springer Verlag (1984), pp. 256–272.

SIEGEL, A.F., *Robust Regression using Repeated Medians*, Biometrika, 69 (1982), pp. 242–244.

THEIL, H., *A Rank-Invariant Method of Linear and Polynomial Regression Analysis (Parts 1-3)*, Ned.Akad. Wetensch. Proc. Ser. A, 53 (1950), pp. 386–392, 521–525, 1397–1412.

VAN ZOMEREN, B.C., *A Comparison of Some High-Breakdown Regression Estimators*, Computational Statistics Quarterly, 4 (1987), pp. 207–218.

YOHAI, V.J., *High Breakdown-Point and High Efficiency Robust Estimates for Regression*, Annals of Statistics, 15 (1987), pp. 642–656.