Alternatives to the Median Absolute Deviation

Peter J. Rousseeuw and Christophe Croux*

In robust estimation one frequently needs an initial or auxiliary estimate of scale. For this one usually takes the median absolute deviation $\mathrm{MAD}_n = 1.4826\ \mathrm{med}_i\{|x_i - \mathrm{med}_j x_j|\}$, because it has a simple explicit formula, needs little computation time, and is very robust as witnessed by its bounded influence function and its 50% breakdown point. But there is still room for improvement in two areas: the fact that MAD_n is aimed at symmetric distributions and its low (37%) Gaussian efficiency. In this article we set out to construct explicit and 50% breakdown scale estimators that are more efficient. We consider the estimator $S_n = 1.1926\ \mathrm{med}_i\{\mathrm{med}_j|x_i - x_j|\}$ and the estimator Q_n given by the .25 quantile of the distances $\{|x_i - x_j|; i < j\}$. Note that S_n and Q_n do not need any location estimate. Both S_n and Q_n can be computed using $O(n \log n)$ time and $O(n)$ storage. The Gaussian efficiency of S_n is 58%, whereas Q_n attains 82%. We study S_n and Q_n by means of their influence functions, their bias curves (for implosion as well as explosion), and their finite-sample performance. Their behavior is also compared at non-Gaussian models, including the negative exponential model where S_n has a lower gross-error sensitivity than the MAD.

KEY WORDS: Bias curve; Breakdown point; Influence function; Robustness; Scale estimation.

1. INTRODUCTION

Although many robust estimators of location exist, the sample median is still the most widely known. If $\{x_1, \ldots, x_n\}$ is a batch of numbers, we will denote its sample median by

$$\mathrm{med}_i x_i$$

which is simply the middle order statistic when n is odd. When n is even, we shall use the average of the order statistics with ranks $(n/2)$ and $(n/2) + 1$. The median has a breakdown point of 50% (which is the highest possible), because the estimate remains bounded when fewer than 50% of the data points are replaced by arbitrary numbers. Its influence function is also bounded, with the sharpest bound for any location estimator (see Hampel, Ronchetti, Rousseeuw, and Stahel 1986). These are but a few results attesting to the median's robustness.

Robust estimation of scale appears to have gained somewhat less acceptance among general users of statistical methods. The only robust scale estimator to be found in most statistical packages is the *interquartile range,* which has a breakdown point of 25% (which is rather good, although it can be improved on). Some people erroneously consider the *average deviation*

$$\mathrm{ave}_i|x_i - \mathrm{ave}_j x_j| \tag{1.1}$$

(where ave stands for "average") to be a robust estimator, although its breakdown point is 0 and its influence function is unbounded. If in (1.1) one of the averages is replaced by a median, we obtain the "median deviation about the average" and the "average deviation about the median," both of which suffer from a breakdown point of 0 as well.

A very robust scale estimator is the *median absolute deviation about the median,* given by

$$\mathrm{MAD}_n = b\ \mathrm{med}_i|x_i - \mathrm{med}_j x_j|. \tag{1.2}$$

This estimator is also known under the shorter name of median absolute deviation (MAD) or even median deviation. The MAD has the best possible breakdown point (50%, twice as much as the interquartile range), and its influence function is bounded, with the sharpest possible bound among all scale estimators. The MAD was first promoted by Hampel (1974), who attributed it to Gauss. The constant b in (1.2) is needed to make the estimator consistent for the parameter of interest. In the case of the usual parameter σ at Gaussian distributions, we need to set $b = 1.4826$. (In the same situation, the average deviation in (1.1) needs to be multiplied by 1.2533 to become consistent.)

The sample median and the MAD are simple and easy to compute, but nevertheless very useful. Their extreme sturdiness makes them ideal for screening the data for outliers in a quick way, by computing

$$\frac{|x_i - \mathrm{med}_j x_j|}{\mathrm{MAD}_n} \tag{1.3}$$

for each x_i and flagging those x_i as spurious for which this statistic exceeds a certain cutoff (say, 2.5 or 3.0). Also, the median and the MAD are often used as initial values for the computation of more efficient robust estimators. It was confirmed by simulation (Andrews et al. 1972) that it is very important to have robust starting values for the computation of M-estimators, and that it won't do to start from the average and the standard deviation. Also note that location M-estimators need an ancillary estimate of scale to make them equivariant. It has turned out that the MAD's high breakdown property makes it a better ancillary scale estimator than the interquartile range, also in regression problems. This led Huber (1981, p. 107) to conclude that "the MAD has emerged as the single most useful ancillary estimate of scale."

In spite of all these advantages, the MAD also has some drawbacks. First, its efficiency at Gaussian distributions is very low; whereas the location median's asymptotic efficiency is still 64%, the MAD is only 37% efficient. Second, the MAD takes a symmetric view on dispersion, because one first estimates a central value (the median) and then attaches equal importance to positive and negative deviations from it. Actually, the MAD corresponds to finding the symmetric in-

* Peter J. Rousseeuw is Professor and Christophe Croux is Assistant, Department of Mathematics and Computer Science, Universitaire Instelling Antwerpen (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium. The authors thank the associate editor and referees for helpful comments.

terval (around the median) that contains 50% of the data (or 50% of the probability), which does not seem to be a natural approach at asymmetric distributions. The interquartile range does not have this problem, as the quartiles need not be equally far away from the center. In fact, Huber (1981, p. 114) presented the MAD as the *symmetrized version* of the interquartile range. This implicit reliance on symmetry is at odds with the general theory of M-estimators, in which the symmetry assumption is not needed. Of course, there is nothing to stop us from using the MAD at highly skewed distributions, but it may be rather inefficient and artificial to do so.

2. THE ESTIMATOR S_n

The purpose of this article is to search for alternatives to the MAD that can be used as initial or ancillary scale estimates in the same way but that are more efficient and not slanted towards symmetric distributions. Two such estimators will be proposed and investigated; the first is

$$S_n = c \operatorname{med}_i \{\operatorname{med}_j |x_i - x_j|\}. \tag{2.1}$$

This should be read as follows. For each i we compute the median of $\{|x_i - x_j|; j = 1, \ldots, n\}$. This yields n numbers, the median of which gives our final estimate S_n. (The factor c is again for consistency, and its default value is 1.1926, as we will see later.) In our implementation of (2.1) the outer median is a low median, which is the order statistic of rank $[(n + 1)/2]$, and the inner median is a high median, which is the order statistic of rank $h = [n/2] + 1$. The estimator S_n can be seen as an analog of Gini's average difference (Gini 1912), which one would obtain when replacing the medians by averages in (2.1). Note that (2.1) is very similar to formula (1.2) for the MAD, the only difference being that the med_j operation was moved outside the absolute value. The idea to apply medians repeatedly was introduced by Tukey (1977) for estimation in two-way tables and by Siegel (1982) for estimating regression coefficients. Rousseeuw and Bassett (1990) used recursive medians for estimating location in large data sets.

Note that (2.1) is an explicit formula; hence S_n is always uniquely defined. We also see immediately that S_n does behave like a scale estimator, in the sense that transforming the observations x_i to $ax_i + b$ will multiply S_n by $|a|$. We will refer to this property as *affine equivariance.*

Like the MAD, the new estimator S_n is a simple combination of medians and absolute values. Instead of the absolute values, we could also use squares and then take a square root at the end, which would yield exactly the same estimate. (If we would replace the medians by averages in that formula, we would recover the standard deviation.)

On the other hand, S_n is unlike the MAD in that it does not need any location estimate of the data. Instead of measuring how far away the observations are from a central value, S_n looks at a typical distance between observations, which is still valid at asymmetric distributions.

A straightforward algorithm for computing (2.1) would need $O(n^2)$ computation time. However, Croux and Rousseeuw (1992b) have constructed an $O(n \log n)$-time algorithm for S_n. Its source code can be obtained from the authors (rousse@wins.uia.ac.be), and it has been incorporated in Statistical Calculator (T. Dusoir, fbgj23@ujvax.ulster.ac.uk) and in Statlib (statlib@stat.cmu.edu).

Table 1. Average Estimated Value of MAD_n, S_n, Q_n, and SD_n at Gaussian Data

	Average estimated value			
n	MAD_n	S_n	Q_n	SD_n
10	.911	.992	1.392	.973
20	.959	.999	1.193	.990
40	.978	.999	1.093	.993
60	.987	1.001	1.064	.996
80	.991	1.002	1.048	.997
100	.992	.997	1.038	.997
200	.996	1.000	1.019	.999
∞	1.000	1.000	1.000	1.000

NOTE: Based on 10,000 samples for each n.

To check whether the correction factor $c = 1.1926$ (obtained through an asymptotic argument) succeeds in making S_n approximately unbiased for finite samples, we performed a modest simulation study. The estimators in Table 1 are the MAD_n, S_n, the estimator Q_n (which will be described in Section 3), and the sample standard deviation SD_n. Each table entry is the average scale estimate on 10,000 batches of Gaussian observations. We see that S_n behaves better than MAD_n in this experiment. Moreover, Croux and Rousseeuw (1992b) have derived finite-sample correction factors that render MAD_n, S_n, and Q_n almost exactly unbiased.

In the following theorem we prove that the finite sample breakdown point of S_n is the highest possible. We use the replacement version of the breakdown point: For any sample $X = \{x_1, \ldots, x_n\}$, the breakdown point of S_n is defined by

$$\varepsilon_n^*(S_n, X) = \min\{\varepsilon_n^+(S_n, X), \varepsilon_n^-(S_n, X)\}, \tag{2.2}$$

where

$$\varepsilon_n^+(S_n, X) = \min\left\{\frac{m}{n}; \sup_{X'} S_n(X') = \infty\right\}$$

$$\text{and} \quad \varepsilon_n^-(S_n, X) = \min\left\{\frac{m}{n}; \inf_{X'} S_n(X') = 0\right\}$$

and X' is obtained by replacing any m observations of X by arbitrary values. The quantities ε_n^+ and ε_n^- are called the explosion breakdown point and the implosion breakdown point.

Theorem 1. At any sample $X = \{x_1, \ldots, x_n\}$ in which no two points coincide, we have

$$\varepsilon_n^+(S_n, X) = [(n + 1)/2]/n \quad \text{and} \quad \varepsilon_n^-(S_n, X) = [n/2]/n.$$

The breakdown point of the scale estimator S_n thus is given by

$$\varepsilon_n^*(S_n, X) = [n/2]/n,$$

which is the highest possible value for any affine equivariant scale estimator. (The proof is given in the Appendix.)

We now turn to the asymptotic version of our estimator. Let X and Y be independent stochastic variables with dis-

tribution function G. If we denote $g_G(x) = \text{med}_Y |x - Y|$, then

$$S(G) = c \operatorname*{med}_X g_G(X) \tag{2.3}$$

is the desired functional. If G_n is the empirical distribution function, then $S(G_n) = S_n$.

Let us consider a location-scale model $F_{\theta,\sigma}(x) = F((x-\theta)/\sigma)$, where F is called the model distribution. We want our estimator to be Fisher-consistent, which means that $S(F_{\theta,\sigma}) = \sigma$ for all θ and all $\sigma > 0$. It is easily verified that S is Fisher-consistent if we take

$$c = 1/(\operatorname*{med}_X g_F(X)), \tag{2.4}$$

where X is distributed according to the model distribution F. The following theorem gives the value of c in case $F = \Phi$, where $\Phi(x)$ is the standard Gaussian distribution function.

Theorem 2. For $F = \Phi$, the constant c satisfies the equation

$$\Phi(\Phi^{-1}(3/4) + c^{-1}) - \Phi(\Phi^{-1}(3/4) - c^{-1}) = 1/2.$$

Equivalently, c^{-1} is the square root of the median of the noncentral chi-squared distribution with 1 degree of freedom and noncentrality parameter $\Phi^{-1}(3/4)$. Numerical calculation yields $c = 1.1926$. (If we want to work with another model distribution, we must take a different c, as in the examples in Sec. 4.)

The influence function of the functional S at the distribution F is defind by

$$IF(x; S, F) = \lim_{\varepsilon \downarrow 0} \frac{S((1-\varepsilon)F + \varepsilon\Delta_x) - S(F)}{\varepsilon}, \tag{2.5}$$

where Δ_x has all its mass in x (Hampel 1974). Theorem 3 gives the expression of $IF(x; S, F)$.

Theorem 3. Assume that the following conditions hold:

1. There exists an x_0 such that $g_F(x)$ is increasing on $[x_0, \infty[$ and decreasing on $]-\infty, x_0]$.
2. There exist $q_1 < x_0$ and $q_2 > x_0$ for which $g_F(q_1) = c^{-1} = g_F(q_2)$. Assume that F has a density f in neighborhoods of q_1 and q_2, with $f(q_1) > 0$ and $f(q_2) > 0$. Assume further that g_F has a strictly positive derivative in a neighborhood of q_2 and a strictly negative derivative in a neighborhood of q_1.
3. f exists in neighborhoods of the points $q_1 \pm c^{-1}$ and $q_2 \pm c^{-1}$.

Then the influence function of S is given by

$$\begin{aligned} IF(x; S, F) = c\Bigg\{ & F(q_2) - F(q_1) - I(q_1 < x < q_2) \\ & - \frac{f(q_2)\,\text{sgn}(|q_2 - x| - c^{-1})}{2(f(q_2 + c^{-1}) - f(q_2 - c^{-1}))} \\ & + \frac{f(q_1)\,\text{sgn}(|q_1 - x| - c^{-1})}{2(f(q_1 + c^{-1}) - f(q_1 - c^{-1}))} \Bigg\} \\ & \times \left\{ \frac{f(q_2)}{g'_F(q_2)} - \frac{f(q_1)}{g'_F(q_1)} \right\}^{-1}. \end{aligned} \tag{2.6}$$

(A more general "chain rule" for influence functions was given by Rousseeuw and Croux 1992.) Let us consider the special case where $F = \Phi$. We will use the notation $q = \Phi^{-1}(3/4)$. From the proof of Theorem 2, it follows that $q_1 = -q$, $q_2 = q$, and $g'_\Phi(q) = -g'_\Phi(-q)$. Because we have

$$\Phi(x + g_\Phi(x)) - \Phi(x - g_\Phi(x)) = .5$$

for all x, it follows that

$$g'_\Phi(q) = \frac{\phi(q - c^{-1}) - \phi(q + c^{-1})}{\phi(q + c^{-1}) + \phi(q - c^{-1})}.$$

This yields the expression

$$\begin{aligned} IF(x; S, \Phi) = c\Bigg\{ & \frac{g'_\Phi(q)\,\text{sgn}(|x| - q)}{4\phi(q)} \\ & + \frac{\text{sgn}(|q - x| - c^{-1}) + \text{sgn}(|-q - x| - c^{-1})}{4(\phi(q + c^{-1}) + \phi(q - c^{-1}))} \Bigg\}. \end{aligned} \tag{2.7}$$

The influence function of S is plotted in Figure 1. We see that it is a step function that takes on four values (which are symmetric about 0), unlike $IF(x; \text{MAD}, \Phi)$, which takes on only two values. Let us also consider the gross-error sensitivity

$$\gamma^*(S, \Phi) = \sup_x |IF(x; S, \Phi)| = 1.625, \tag{2.8}$$

which is rather small, indicating that S is quite robust against outliers. In contrast, $\gamma^*(\text{SD}, \Phi) = \infty$, whereas $\gamma^*(\text{MAD}, \Phi) = 1.167$ is the smallest value that we can obtain for any scale estimator at the Gaussian distribution.

The influence function of the classical standard deviation is given by $IF(x; \text{SD}, \Phi) = (x^2 - 1)/2$. Note that $IF(x; S, \Phi)$ looks more like that U-shaped curve than does $IF(x; \text{MAD}, \Phi)$, indicating that S will be more efficient than the MAD. Indeed, the asymptotic variance of S_n is given by

$$V(S, \Phi) = \int IF(x; S, \Phi)^2 \, d\Phi(x) = .8573. \tag{2.9}$$

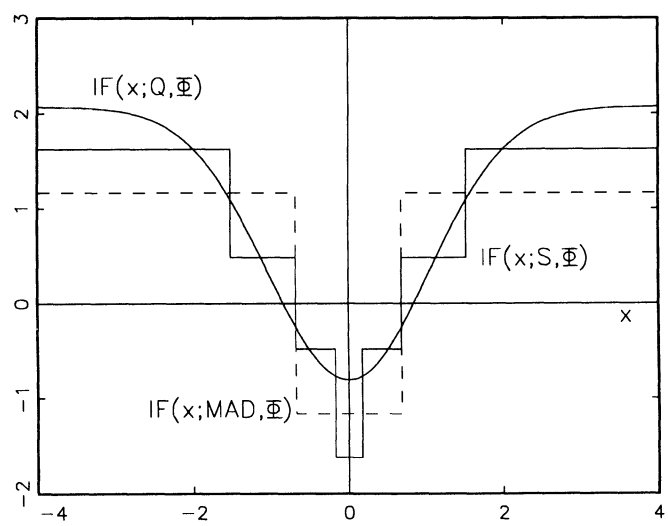

Figure 1. Influence Functions of the MAD, the Estimator S, and the Estimator Q When the Model Distribution is Gaussian.

Table 2. Standardized Variance of MAD_n, S_n, Q_n, and SD_n at Gaussian Data

	Standardized variance			
n	MAD_n	S_n	Q_n	SD_n
10	1.361	1.125	.910	.584
20	1.368	.984	.773	.532
40	1.338	.890	.701	.513
60	1.381	.893	.679	.515
80	1.342	.878	.652	.511
100	1.377	.869	.650	.507
200	1.361	.873	.636	.499
∞	1.361	.857	.608	.500

NOTE: Based on 10,000 samples for each n.

(The asymptotic normality of S_n was proved in Hössjer, Croux, and Rousseeuw 1993.) This yields an efficiency of 58.23%, which is a marked improvement relative to the MAD whose efficiency at Gaussian distributions is 36.74%. We pay for this by a slight increase in the gross-error sensitivity and in the required computation time.

We carried out a simulation to verify this efficiency gain at finite samples. For each n in Table 2, we computed the variance $\text{var}_m(S_n)$ of the scale estimator S_n over $m = 10{,}000$ samples. Table 2 lists the standardized variances

$$n \text{ var}_m(S_n)/(\text{ave}_m(S_n))^2, \tag{2.10}$$

where $\text{ave}_m(S_n)$ is the average estimated value as given in Table 1. (It was argued by Bickel and Lehmann (1976) that the denominator of (2.10) is needed to obtain a natural measure of accuracy for scale estimators.) The results show that the asymptotic variance provides a good approximation for (not too small) finite samples, and that S_n is more efficient than MAD_n even for small n.

Whereas the influence function describes how the estimator reacts to a single outlier, the *bias curve* tells us how much the estimator can change (in the worst case) when a fraction ε of the data is contaminated. Bias curves were briefly mentioned by Hampel et al. (1986, p. 177), but their full potential was not realized until the work of Martin and Zamar (1989, 1991), Martin, Yohai and Zamar (1989), and He and Simpson (1993).

In the case of scale estimators, we distinguish between an increase of the estimate (explosion) and a decrease (implosion). For $\varepsilon > 0$, we define $\mathcal{F}_\varepsilon = \{G; G = (1-\varepsilon)F + \varepsilon H\}$, where H ranges over all distributions. Then the *explosion bias curve* of S is defined as

$$B^+(\varepsilon, F) = \sup_{G \in \mathcal{F}_\varepsilon} S(G)$$

(plotted as a function of ε), and the *implosion bias curve* of S is given by

$$B^-(\varepsilon, F) = \inf_{G \in \mathcal{F}_\varepsilon} S(G).$$

Theorem 4. For a distribution F that is symmetric around the origin, is unimodal, and has a strictly positive density, we have

$$B^+(\varepsilon, F) = cg^+\left(F^{-1}\left(\frac{3-2\varepsilon}{4(1-\varepsilon)}\right)\right), \tag{2.11}$$

where g^+ is defined implicitly by

$$F(x + g^+(x)) - F(x - g^+(x)) = \frac{1}{2(1-\varepsilon)}. \tag{2.12}$$

The implosion bias curve is given by

$$B^-(\varepsilon, F) = cg^-\left(F^{-1}\left(\frac{3-4\varepsilon}{4(1-\varepsilon)}\right)\right), \tag{2.13}$$

where g^- is defined implicitly by

$$F(x + g^-(x)) - F(x - g^-(x)) = \frac{1-2\varepsilon}{1-\varepsilon}. \tag{2.14}$$

Figure 2 displays both bias curves at $F = \Phi$. We see that the explosion bias curve of S is nearly as good as that of the MAD, and for ε close to the breakdown point it is even better. For the implosion bias curve the MAD performs slightly better than S overall. Note that $[\partial B^+(\varepsilon, \Phi)/\partial\varepsilon]|_{\varepsilon=0} = \gamma^*(S, \Phi)$ and $[\partial B^-(\varepsilon, \Phi)/\partial\varepsilon]|_{\varepsilon=0} = -\gamma^*(S, \Phi)$. Moreover, at the breakdown point ($\varepsilon \to \frac{1}{2}$) we find $B^+(\varepsilon, \Phi) \to \infty$ and $B^-(\varepsilon, \Phi) \to 0$.

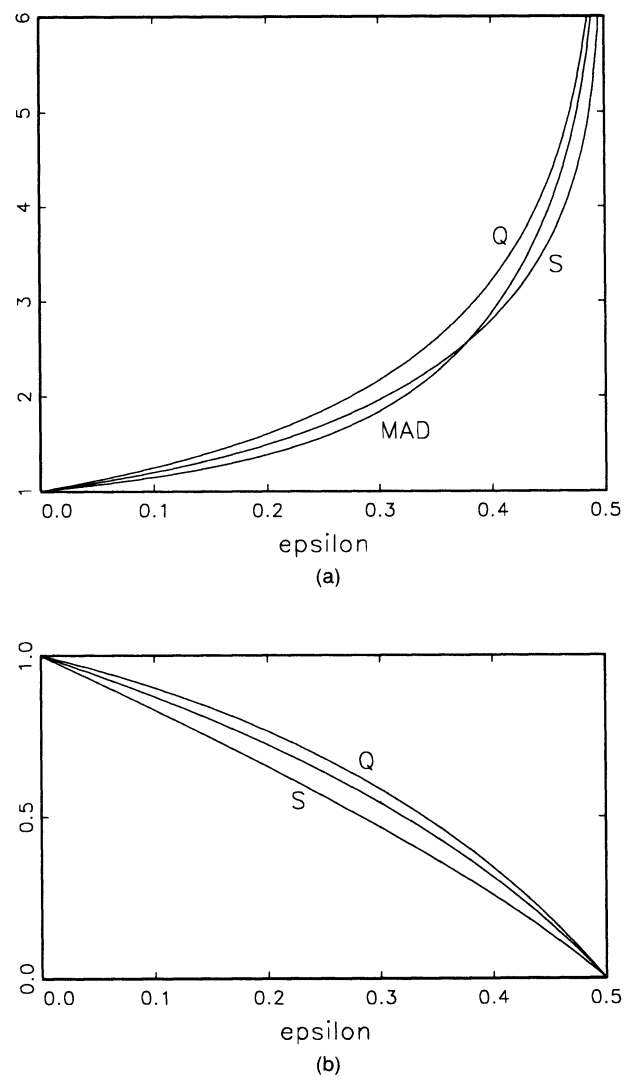

Figure 2. Bias Curves of the MAD, S, and Q as a Function of the Fraction of Contamination: (a) for Explosion; (b) for Implosion.

Note that the influence function and the bias curve are asymptotic concepts. For finite samples, we have confirmed the above results by computing the corresponding sensitivity curves and empirical bias curves.

3. THE ESTIMATOR Q_n

A drawback of MAD_n and S_n is that their influence functions have discontinuities. In the location framework, the sample median has the same drawback, unlike the Hodges–Lehmann (1963) estimator

$$\text{med}\left\{\frac{x_i + x_j}{2}\,;\, i < j\right\}, \tag{3.1}$$

which possesses a smooth influence function (see, for example, Hampel et al. 1986, p. 112). Therefore, the Hodges–Lehmann estimator might be viewed as a "smooth version" of the median.

An analogous scale estimator, mentioned by Shamos (1976, p. 260) and Bickel and Lehmann (1979, p. 38), is obtained by replacing the pairwise averages by pairwise distances, yielding

$$\text{med}\{|x_i - x_j|;\, i < j\}. \tag{3.2}$$

This resembles definition (2.1) of S_n, where separate medians were taken over i and j, whereas (3.2) uses an overall median over $\binom{n}{2}$ pairs. We propose to multiply (3.2) by 1.0483 to achieve consistency for the parameter σ of Gaussian distributions. Like the Hodges–Lehmann estimator, this scale estimator has only a 29% breakdown point, but a rather high Gaussian efficiency (about 86%).

But we want an estimator with a 50% breakdown point like the MAD. We found that, somewhat surprisingly, this goal can be attained by replacing the median in (3.2) by a different order statistic. Therefore, we propose the estimator

$$Q_n = d\{|x_i - x_j|;\, i < j\}_{(k)}, \tag{3.3}$$

where d is a constant factor and $k = \binom{h}{2} \approx \binom{n}{2}/4$, where $h = [n/2] + 1$ is roughly half the number of observations. That is, we take the kth order statistic of the $\binom{n}{2}$ interpoint distances. This bears some resemblance to S_n, because the double median of the interpoint distances is at least as large as their .25 quantile.

The estimator Q_n shares the attractive properties of S_n: a simple and explicit formula, a definition that is equally suitable for asymmetric distributions, and a 50% breakdown point. In addition, we will see that its influence function is smooth, and that its efficiency at Gaussian distributions is very high (about 82%). At first sight, these advantages are offset by a larger computational complexity, because the naive algorithm (which begins by computing and storing all $\binom{n}{2}$ pairwise distances) needs $O(n^2)$ space and $O(n^2)$ time. But Croux and Rousseeuw (1992b) have constructed an algorithm for computing Q_n with $O(n)$ space and $O(n \log n)$ time. Due to the availability of fast algorithms for S_n and Q_n, a financial company is now using these estimators on a daily basis in analyses of the behavior of stocks, with $n \approx 8000$.

We first show that the finite sample breakdown point of Q_n attains the optimal value.

Theorem 5. At any sample $X = \{x_1, \ldots, x_n\}$ in which no two points coincide, we have $\varepsilon_n^+(Q_n, X) = [(n+1)/2]/n$ and $\varepsilon_n^-(Q_n, X) = [n/2]/n$; thus the breakdown point of Q_n is

$$\varepsilon_n^*(Q_n, X) = [n/2]/n.$$

We now give the asymptotic version of our estimator. Let X and Y be independent random variables with distribution G. Then we define

$$Q(G) = dH_G^{-1}\left(\frac{1}{4}\right) \quad \text{with} \quad H_G = \mathcal{L}(|X - Y|). \tag{3.4}$$

Note that $Q(G_n)$ is not exactly the same as Q_n (where we take an order statistic among $\binom{n}{2}$ elements instead of n^2), but asymptotically this makes no difference. Representation (3.4) belongs to the class of generalized L-statistics, as introduced by Serfling (1984). Because $X - Y$ is symmetric, we could also write

$$Q(G) = dK_G^{-1}\left(\frac{5}{8}\right) \quad \text{with} \quad K_G = \mathcal{L}(X - Y) \tag{3.5}$$

instead of (3.4). Finally, we obtain

$$Q(G) = \inf\left\{s > 0;\, \int G(t + d^{-1}s)\, dG(t) \geq 5/8\right\}. \tag{3.6}$$

In the parametric model $F_{\theta,\sigma}(x) = F((x - \theta)/\sigma)$, the functional Q is Fisher-consistent for σ if we choose the constant d according to

$$d = 1/H_F^{-1}\left(\frac{1}{4}\right) = 1/K_F^{-1}\left(\frac{5}{8}\right). \tag{3.7}$$

For symmetric F this reduces to $d = 1/((F^{*2})^{-1}(5/8))$, where F^{*2} denotes the convolution $F * F$. In the case $F = \Phi$, we obtain $K_F = \mathcal{L}(X - Y) = \mathcal{L}(X + Y) = \mathcal{L}(\sqrt{2}X)$; hence

$$d = 1/(\sqrt{2}\,\Phi^{-1}(5/8)) = 2.2219.$$

In Table 1 we see that with this constant the estimator Q_n has a considerable small-sample bias, but we can use the correction factors derived by Croux and Rousseeuw (1992b).

By means of some standard calculations (or using expression (2.12) in Serfling 1984), one obtains the following formula for the influence function.

Theorem 6. If K_F has a positive derivative at $1/d$, then

$$IF(x; Q, F) = d\,\frac{\frac{1}{4} - P\left(|x - Y| \leq \frac{1}{d}\right)}{K_F'\left(\frac{1}{d}\right)}. \tag{3.8}$$

If F has a density f, then (3.8) becomes

$$IF(x; Q, F) = d\,\frac{\frac{1}{4} - F(x + d^{-1}) + F(x - d^{-1})}{\int f(y + d^{-1}) f(y)\, dy}. \tag{3.9}$$

In Figure 1 we plotted the function $IF(x; Q, \Phi)$. It turns out that the gross-error sensitivity $\gamma^*(Q, \Phi) = 2.069$ is larger than that of the MAD and S. Note that this influence function is unbalanced: The influence of an inlier at the center of the distribution is smaller (in absolute value) than the influence of an outlier at infinity.

With the aid of numerical integration, Equations (2.9) and (3.9) yield $V(Q, \Phi) = .6077$. By applying Theorem 3.1 in Serfling (1984), we obtain a rigorous proof of the asymptotics of Q_n under the same conditions as in our Theorem 6. The resulting Gaussian efficiency of 82.27% is surprisingly high.

In Table 2 we see that Q_n is considerably more efficient than either MAD_n or S_n. For instance, Q_{40} has about the same precision as MAD_{80}. But Q_n looses some of its efficiency at small sample sizes.

Theorem 7. If F is symmetric about 0 and possesses a density, then

$$B^+(\varepsilon, F) = d(F^{*2})^{-1}\left(\frac{5 - 8\varepsilon + 4\varepsilon^2}{8(1 - \varepsilon)^2}\right), \qquad (3.10)$$

and the implosion bias is given by the solution of the equation

$$(1 - \varepsilon)^2 F^{*2}(d^{-1}B^-(\varepsilon, F)) + 2\varepsilon(1 - \varepsilon)F(B^-(\varepsilon, F)) + \varepsilon^2 = 5/8. \qquad (3.11)$$

Figure 2 displays the bias curves of Q (at $F = \Phi$) along with those of the MAD and S. Note that the MAD is slightly more robust than Q regarding their explosion bias curves, whereas Q is more robust than MAD for implosion bias. It is also interesting to note (see the derivation in the Appendix) that the supremum bias of Q does not correspond to contamination by a point mass moving to infinity, but rather to a contaminating distribution of which both location *and* scale tend to infinity. This explains why the slope of the explosion bias curve at $\varepsilon = 0$ is not equal to the supremum of the influence function in this case.

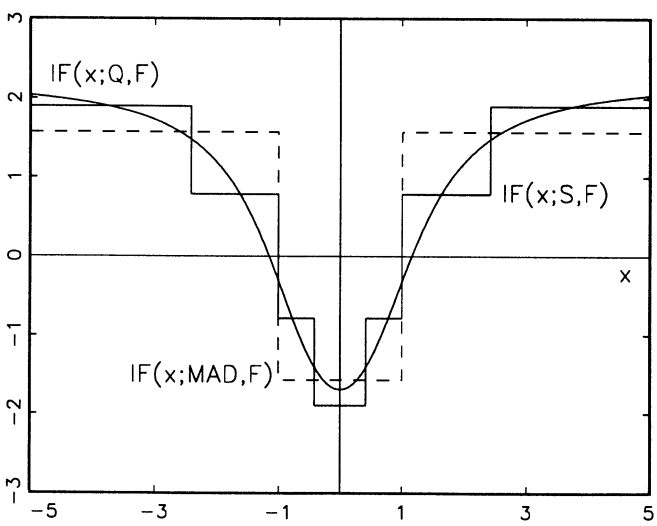

Figure 3. Influence Functions of the MAD, S, and Q at the Cauchy Distribution.

Table 3. Simulation Results for Cauchy Data, Based on 10,000 Samples for Each n

	Average value			Standardized variance		
n	MAD_n	S_n	Q_n	MAD_n	S_n	Q_n
10	1.097	1.126	1.629	3.150	3.194	3.363
20	1.045	1.044	1.284	2.819	2.554	2.606
40	1.025	1.021	1.136	2.569	2.280	2.299
60	1.014	1.010	1.085	2.571	2.202	2.172
80	1.009	1.007	1.062	2.586	2.240	2.215
100	1.011	1.010	1.055	2.497	2.164	2.119
200	1.005	1.004	1.027	2.454	2.130	2.058
∞	1.000	1.000	1.000	2.467	2.106	2.044

4. ROBUSTNESS AT NON-GAUSSIAN MODELS

Quite often one uses a parametric model $F_{\theta,\sigma}(x) = F((x - \theta)/\sigma)$ in which the model distribution F is itself non-Gaussian. In this section we investigate the behavior of S_n and Q_n for a heavy-tailed model distribution and for an asymmetric model distribution.

Let us first consider the Cauchy distribution

$$F(x) = \frac{\arctan(x)}{\pi} + \frac{1}{2}.$$

At this heavy-tailed distribution, the sample standard deviation is not a useful estimator of σ because the second moment of $F_{\theta,\sigma}$ does not exist. But we can still use MAD_n, S_n, and Q_n, whose consistency factors become $b = 1$, $c = .7071$, and $d = 1.2071$ in this model. Figure 3 shows the influence functions of these three estimators, which were computed in roughly the same way as in the Gaussian model. The shape of these functions is quite similar to Figure 1, except for $IF(x; Q, F)$, which is more balanced than in the Gaussian case. The gross-error sensitivity $\gamma^*(Q, F) = 2.2214$ is attained for x going to infinity, but $IF(x; Q, F)$ approaches this limit very slowly.

For the asymptotic variances we obtain $V(\text{MAD}, F) = 2.4674$, $V(S, F) = 2.1060$, and $V(Q, F) = 2.0438$. These should be compared to the theoretical lower bound (i.e., the inverse of the Fisher information) which equals 2 in this case. Therefore, the absolute asymptotic efficiencies become $e(\text{MAD}) = 81\%$, $e(S) = 95\%$, and $e(Q) = 98\%$. At this heavy-tailed distribution, all three robust estimators are more efficient than at the Gaussian distribution.

Table 3 summarizes a simulation (analogous to Tables 1 and 2) confirming that MAD_n and S_n are approximately unbiased for finite samples. For not too small n, the variability of the estimators is described reasonably well by their asymptotic variance. Note that S_n performs somewhat better than Q_n at small samples.

As an example of an asymmetric model distribution, we take the negative exponential

$$F(x) = (1 - \exp(-x))I(x \geq 0),$$

which generates the location-scale family

$$F_{\theta,\sigma}(x) = (1 - \exp(-(x - \theta)/\sigma))I(x \geq \theta),$$

which we will call the *shifted exponential model.* Note that the location parameter θ is not at the center of the distribution, but rather at its left boundary. To estimate σ by MAD_n, S_n, or Q_n, we need to set $b = 2.0781$, $c = 1.6982$, and $d = 3.4760$. Figure 4 displays the influence functions of these estimators at F. Note that the influence functions were also computed at negative x although F has no mass there, because outliers can occur to the left of the unknown θ.

The influence function of the MAD looks very different from those in Figures 1 and 3, because this time F is not symmetric. Note that $IF(x; \text{MAD}, F) = 0$ for negative x (due to the compensating effect of the outlier at x on the location median) and that it has three jumps for positive x (at $\ln(2) - b^{-1}$, at $\ln(2)$, and at $\ln(2) + b^{-1}$). The influence functions of S and Q are roughly U-shaped, as in Figures 1 and 3. Surprisingly, the gross-error sensitivity of S is smaller than that of the MAD. Indeed,

$$\gamma^*(S, F) = 1.8447 < \gamma^*(\text{MAD}, F) = 1.8587.$$

This would seem to disprove a theorem stating that the MAD has the lowest possible gross-error sensitivity (see Hampel et al. 1986, p. 142), but there is no contradiction because that theorem was for symmetric F.

For the asymptotic variances we have obtained $V(\text{MAD}, F) = 2.1352$, $V(S, F) = 1.8217$, and $V(Q, F) = 1.3433$. As in the Gaussian and the Cauchy models, we again see that Q is more efficient than S, which in turn is more efficient than the MAD. (This also holds for finite samples, as confirmed by Table 4.) Note that the Fisher information for this model does not exist, so the absolute efficiency of the estimators is not defined. But we can still compute relative efficiencies between them, such as $\text{ARE}(\text{MAD}, Q) = 1.3433/2.1351 = 63\%$.

The maximum likelihood estimator (MLE) of σ in this model is $\hat{\sigma}_n = \text{ave}_i x_i - \hat{\theta}_n$ where $\hat{\theta}_n = \min_j x_j$ is the MLE of θ (see Johnson and Kotz 1970, p. 211). It turns out that $\hat{\theta}_n$ converges at the unusually fast rate of n^{-1}, whereas $\hat{\sigma}_n$ converges at the standard rate of $n^{-1/2}$ with asymptotic variance

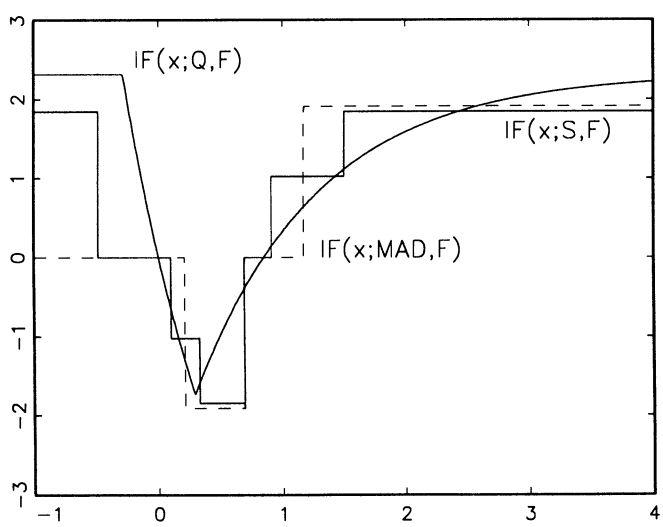

Figure 4. Influence Functions of the MAD, S, and Q at the Exponential Distribution.

Table 4. Simulation Results for Exponentially Distributed Data, Based on 10,000 Samples for Each n

	Average value			Standardized variance		
n	MAD_n	S_n	Q_n	MAD_n	S_n	Q_n
10	0.925	1.002	1.536	2.152	2.121	1.677
20	0.959	0.981	1.246	2.182	2.065	1.521
40	0.983	0.991	1.120	2.167	2.006	1.481
60	0.986	0.993	1.078	2.172	1.989	1.420
80	0.988	0.993	1.056	2.153	1.924	1.395
100	0.993	0.996	1.046	2.122	1.911	1.373
200	0.996	0.998	1.024	2.173	1.919	1.381
∞	1.000	1.000	1.000	2.135	1.822	1.343

$V(\hat{\sigma}, F) = 1$. But the extreme efficiency of $\hat{\theta}_n$ is offset by its extreme sensitivity to any outlier $x_i < \theta$. One possibility for constructing a more robust estimator of θ would be to use

$$\tilde{\theta}_n = \text{med}_i x_i - \tilde{\sigma}_n \ln 2,$$

where $\tilde{\sigma}_n$ is a robust scale estimator such as MAD_n, S_n, or Q_n.

5. DISCUSSION

Let us compare S_n and Q_n to another explicit scale estimator with a 50% breakdown point, which is based on the length of the shortest half sample:

$$\text{LMS}_n = c' \min_i |x_{(i+h-1)} - x_{(i)}|, \tag{5.1}$$

where again $h = [n/2] + 1$ and $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ are the ordered data. (Note that the minimum length is always unique, even if the half sample itself is not.) The default value of c' is .7413, which achieves consistency at Gaussian distributions. The estimator (5.1) first occurred as the scale part of the least median of squares (LMS) regression estimator (Rousseeuw 1984) in the special case of one-dimensional data. As we can see from (5.1), the estimator LMS_n is simple, has the optimal breakdown point, and needs only $O(n \log n)$ computation time and $O(n)$ storage. Its influence function is the same as that of the MAD (Rousseeuw and Leroy 1988), and its efficiency equals that of the MAD as well (Grübel 1988). Therefore, LMS_n is less efficient than either S_n or Q_n. Note that if we use the p-subset algorithm of Rousseeuw and Leroy (1987), we obtain the estimator

$$c' \min_i \operatorname*{med}_j |x_i - x_j|, \tag{5.2}$$

which is asymptotically equivalent to LMS_n and also has a 50% breakdown point.

The estimators MAD_n, LMS_n, and S_n are all based on half samples. MAD_n looks for the shortest half that is symmetric about the location median, whereas LMS_n looks for the shortest half without such a constraint (hence LMS_n can be used at asymmetric distributions as well). Similarly, S_n can be said to reflect the "typical" length of half samples in the data set. The following theorem shows that MAD_n, LMS_n, and S_n are strongly interrelated.

Theorem 8. At every data set $\{x_1, \ldots, x_n\}$, it holds that

$$\frac{1}{b}\,\mathrm{MAD}_n \le \frac{1}{c}\,S_n \le \frac{2}{b}\,\mathrm{MAD}_n$$

$$\frac{1}{2c'}\,\mathrm{LMS}_n \le \frac{1}{c}\,S_n \le \frac{1}{c'}\,\mathrm{LMS}_n.$$

If we define the collection of interpoint distances by $D = \{|x_i - x_j|;\ 1 \le i, j \le n\}$, then S_n is the remedian with base n (see Rousseeuw and Bassett 1990) of the set D, assuming that the elements of D are read row by row. This implies that S_n corresponds to an element between the .25 and .75 quantiles of D. On the other hand, Q_n is asymptotically equivalent to the .25 quantile of D. Roughly speaking, and ignoring the constants involved, it follows that S_n is larger than Q_n. (Note that $S_n = Q_n$ for $n \le 4$.)

We have seen that S_n and Q_n have a better efficiency than MAD_n, both asymptotically and for finite samples, while their gross-error sensitivities and bias curves are almost as good. Another advantage is that they do not presuppose a symmetric model distribution, but can be considered as nonparametric measures of spread. (The estimator LMS_n shares this advantage, but is less efficient.) The only price we pay is a small increase of the computation time, which is $O(n \log n)$ for S_n and Q_n as compared to $O(n)$ for MAD_n, but today's computers can easily afford this.

One could also compare S_n and Q_n with the class of M-estimators. The latter can be more efficient, and their computational complexity is similar (or even lower, if we restrict their algorithm to a fixed number of iteration steps). But M-estimators are defined implicitly, whereas S_n and Q_n have an explicit formula, which guarantees uniqueness of the estimate and is more intuitive. Moreover, there is no need to select a function χ or to choose tuning constants.

Another class is that of one-step M-estimators, which have an explicit formula and can also attain a high efficiency and a high breakdown point. But like fully iterated M-estimators, they are not location-free because they need an ancillary location estimator. And, also like fully iterated M-estimators, their computation requires an initial high-breakdown scale estimator such as MAD_n, S_n, or Q_n in the first place. Finally, it should be noted that although k-step M-estimators inherit the breakdown point of the initial estimator, it turns out that their actual contamination bias does increase (Rousseeuw and Croux 1993).

Choosing between S_n and Q_n comes down to a tradeoff between advantages that are hard to compare. Although Q_n is more efficient, in most applications we would prefer S_n because it is very robust, as witnessed by its low gross-error sensitivity. Another advantage of S_n is that its simplicity makes it easier to compute.

We expect the application potential of the new robust scale estimators to be mainly in the following three uses: (1) as a data analytic tool by itself; (2) as an ancillary scale estimate to make M-estimators affine equivariant and as a starting value for the iterative computation of M-estimators; and (3) as an objective function for regression analysis, as will be described in the next section.

6. EXTENSIONS AND OUTLOOK

Our estimators S_n and Q_n are special cases ($v = 2$) of the following formulas:

$$S_n^{(v)} = c_v \mathrm{med}_{i_1}(\mathrm{med}_{i_2}(\ldots \mathrm{med}_{i_v}\mathrm{SD}(x_{i_1}, x_{i_2}, \ldots, x_{i_v})\ldots)) \tag{6.1}$$

and

$$Q_n^{(v)} = d_v\{\mathrm{SD}(x_{i_1}, \ldots, x_{i_v});\ i_1 < i_2 < \cdots < i_v\}_{\binom{h}{v}}, \tag{6.2}$$

where $\mathrm{SD}(x_{i_1}, \ldots, x_{i_v})$ denotes the standard deviation of the v observations $\{x_{i_1}, \ldots, x_{i_v}\}$ and $h = [n/2] + 1$. These estimators all have 50% breakdown points. We expect the efficiency to increase with the order v. But the estimators (6.1) and (6.2) demand a high computational effort. Also, for increasing v their gross-error sensitivity may become arbitrarily large. Therefore, using (6.1) or (6.2) with a large v is not recommended.

If we also allow scale estimators that measure dispersion around a location estimate, possible alternatives would be

$$\mathrm{med}_i\left(\mathrm{med}_j\left(\left|\frac{x_i + x_j}{2} - \mathrm{med}_x x_k\right|\right)\right) \tag{6.3}$$

and

$$\left\{\left|\frac{x_i + x_j}{2} - \mathrm{med}_k x_k\right|;\ i < j\right\}_{\binom{h}{2}}, \tag{6.4}$$

which also have a 50% breakdown point. At symmetric distributions, (6.3) and (6.4) are asymptotically equivalent to S_n and Q_n; however, simulations have indicated that at finite samples these estimators do not perform better than S_n and Q_n, so we see no reason to switch to either (6.3) or (6.4). A possible application of (6.3) would be to divide it by S_n to obtain a test for symmetry (analogously, (6.4) can be divided by Q_n). This is similar to a proposal of Boos (1982) using Gini's average deviation.

Recently, Rousseeuw and Croux (1992) proposed several other explicit scale estimators with high breakdown point. A promising estimator is

$$T_n = 1.3800\,\frac{1}{h}\sum_{k=1}^{h}\{\mathrm{med}_{j \ne i}\,|x_i - x_j|\}_{(k)}. \tag{6.5}$$

It was proved that T_n has a 50% breakdown point, a continuous influence function, an efficiency of 52%, and a gross-error sensitivity of 1.4688.

In cluster analysis one often needs a measure of dissimilarity between two groups A and B, which is based on the interobject dissimilarities. (For a discussion of such measures see Kaufman and Rousseeuw 1990, sec. 5.5.) Our estimators S_n and Q_n can be extended to this situation, yielding

$$d_S(A, B) = \mathrm{med}_{i \in A}\{\mathrm{med}_{j \in B} d(i, j)\} \tag{6.6}$$

and

$$d_Q(A, B) = \{d(i, j);\ i \in A, j \in B\}_{(k)}, \tag{6.7}$$

where k is approximately $(\#A)(\#B)/4$. (Both measures are equivariant for monotone transformations on the $d(i, j)$;

hence they can still be applied when the dissimilarities are on an ordinal scale.) Note that d_Q is symmetric in the sense that $d_Q(A, B) = d_Q(B, A)$, whereas d_S is not, but this can be repaired by working with $d'_S(A, B) = \min\{d_S(A, B), d_S(B, A)\}$.

Both d'_S and d_Q can be seen as robustifying existing methods. Indeed, if the medians in (6.6) were replaced by averages, or if the order statistic in (6.7) was replaced by an average, then we would recover the well-known average linkage criterion. On the other hand, replacing the medians by minima (or equivalently, setting k in (6.7) equal to 1) produces the single linkage criterion. Moreover, replacing the medians by maxima (or setting $k = (\#A)(\#B)$ in (6.7)) yields complete linkage. We can thus think of d'_S and d_Q as intermediate between single linkage and complete linkage, which are often considered to be too extreme. For instance, the latter criteria do not have an asymptotic limit (because usually the smallest dissimilarity tends to 0 as $n \to \infty$, and the largest dissimilarity tends to ∞), whereas (6.6) and (6.7) do.

Another application is to regression. The approach put forward in (Rousseeuw 1984) was to construct affine equivariant high-breakdown regression estimates by minimizing a high-breakdown scale estimate of the residuals $r_1, \ldots, r_n$. For instance, minimizing the scale estimator (5.1) yields the LMS regression estimator. In the same vein, S-estimators (Rousseeuw and Yohai 1984) were obtained by minimization of an M-estimator of scale based on a smooth function ρ. The high-breakdown scale estimators introduced in this article could be used in the same way. It seems that Q_n is preferable to S_n in this regard, because of its smooth influence function. Therefore, we may consider the regression estimator

$$\hat{\theta}_n = \operatorname{argmin}_\theta Q_n(r_1, \ldots, r_n). \tag{6.8}$$

We will refer to $\hat{\theta}_n$ as the *least quartile difference* (LQD) estimator. The LQD belongs to the larger class of *generalized S-estimators* (Croux, Rousseeuw, and Hössjer 1993). The asymptotic efficiency of the LQD turns out to be 67.1%, which is much higher than that of S-estimators.

A final extension is to scale estimation in a linear model. Existing estimators of the error dispersion are based on the residuals from some regression fit. But it is possible to generalize the location-free estimators S_n and Q_n to scale estimators that are *regression-free,* in the sense that they do not depend on any previous estimate of the regression parameters. In the simple linear model, we may consider triangles formed by data points and compute their vertical height $h(i, j, k)$. This leads to several scale estimators (Rousseeuw and Hubert 1993), obtained from quantiles or repeated medians of the $h(i, j, k)$. When attention is restricted to adjacent triangles, this also yields a test for linearity.

APPENDIX: PROOFS

Due to lack of space, some proofs and computations have been omitted and referred to a technical report (Rousseeuw and Croux 1991).

Proof of Theorem 1. Let X be a sample in general position and denote $\varepsilon_n^+ = \varepsilon_n^+(S_n, X)$ and $\varepsilon_n^- = \varepsilon_n^-(S_n, X)$, where $S_n = c\ \text{lomed}_i \text{himed}_j |x_i - x_j|$. We first show that $\varepsilon_n^- \le [n/2]/n$. Construct a contaminated sample X' by replacing the observations $x_2, \ldots, x_{[n/2]+1}$ by x_1. Then we have $\text{himed}_j |x'_i - x'_j| = 0$ for $[n/2] + 1$ observations x'_i, hence $S_n(X') = 0$. Moreover, $\varepsilon_n^- \ge [n/2]/n$. Indeed, take any sample X' where fewer than $[n/2]$ observations are replaced. Because X was in general position, $\text{himed}_j |x'_i - x'_j| \ge \min_{i<j} |x_i - x_j|/2 = \delta > 0$ for all i; hence $S_n(X') > c\delta$.

Further, $\varepsilon_n^+ \le [(n+1)/2]/n$. Construct a sample X' by replacing x_1 by $x_{(n)} + L$, x_2 by $x_{(n)} + 2L, \ldots,$ and $x_{[(n+1)/2]}$ by $x_{(n)} + [(n+1)/2]L$, with $L > 0$. Then $\text{himed}_j |x'_i - x'_j| \ge L$ for all i. Letting L tend to infinity will then inflate the estimator S_n beyond all bounds. On the other hand, $\varepsilon_n^+ \ge [(n+1)/2]/n$. Take any sample X' where fewer than $[(n+1)/2]$ observations are replaced. If x'_i belongs to the original sample, then $\text{himed}_j |x'_i - x'_j| \le |x_{(n)} - x_{(1)}|$. Because this holds for at least half the points, we find that $S_n(X') \le c|x_{(n)} - x_{(1)}| < \infty$.

Remark. This proof also implies that if we replace the outer median by an order statistic of rank $k \le h = [n/2] + 1$, then the estimator will keep the same breakdown points.

The fact that $\varepsilon_n^*(S_n, X) \le [n/2]/n$ for any affine equivariant scale estimator was proved by Croux and Rousseeuw (1992a).

Proof of Theorem 3. We use the notations

$$g_1 = ((g_F))_{|]-\infty, x_0]})^{-1} \quad \text{and} \quad g_2 = ((g_F)_{|[x_0, \infty[})^{-1}.$$

Condition 1 ensures that these functions are well defined. With $G_F(u) = P(g_F(Y) \le u)$, we have $S(F) = cG_F^{-1}(1/2) = 1$, using (2.3). Let us denote $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$. Differentiating the expression $G_{F_\varepsilon}(c^{-1}S(F_\varepsilon)) = 1/2$ yields

$$IF(x; S, F) = \left.\frac{\partial S(F_\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = c\,\frac{-\left.\dfrac{\partial G_{F_\varepsilon}(c^{-1})}{\partial \varepsilon}\right|_{\varepsilon=0}}{G'_F(c^{-1})}. \tag{A.1}$$

Using Conditions 1 and 2, we obtain for u in a neighborhood of c^{-1} that

$$G_F(u) = P(g_1(u) \le Y \le g_2(u)) = F(g_2(u)) - F(g_1(u)^-).$$

Condition 2 yields $g_1(c^{-1}) = q_1$ and $g_2(c^{-1}) = q_2$; hence

$$G'_F(c^{-1}) = \frac{f(q_2)}{g'_F(q_2)} - \frac{f(q_1)}{g'_F(q_1)}. \tag{A.2}$$

Condition 3 allows us to assume that x is different from the points $q_1 \pm c^{-1}$ and $q_2 \pm c^{-1}$ (in which the influence function will not be defined). For ε sufficiently small, we may write

$$G_{F_\varepsilon}(u) = F_\varepsilon(g_{2,\varepsilon}(u)) - F_\varepsilon(g_{1,\varepsilon}(u)) \tag{A.3}$$

for all u in a small neighborhood of c^{-1}. We choose a neighborhood such that g_{F_ε} stays decreasing in a neighborhood of q_1 and increasing near q_2. We may also assume that $g_{1,\varepsilon}(c^{-1})$ and $g_{2,\varepsilon}(c^{-1})$ are lying in that neighborhood. Differentiation of (A.3) yields

$$\left.\frac{\partial G_{F_\varepsilon}(u)}{\partial \varepsilon}\right|_{\varepsilon=0} = f(g_2(u)) \left.\frac{\partial g_{2,\varepsilon}(u)}{\partial \varepsilon}\right|_{\varepsilon=0} + \left.\frac{\partial F_\varepsilon(g_2(u))}{\partial \varepsilon}\right|_{\varepsilon=0} - f(g_1(u)) \left.\frac{\partial g_{1,\varepsilon}(u)}{\partial \varepsilon}\right|_{\varepsilon=0} - \left.\frac{\partial F_\varepsilon(g_1(u))}{\partial \varepsilon}\right|_{\varepsilon=0}.$$

Evaluating in c^{-1} gives

$$\left.\frac{\partial G_{F_\varepsilon}(c^{-1})}{\partial \varepsilon}\right|_{\varepsilon=0} = f(q_2) \left.\frac{\partial g_{2,\varepsilon}(c^{-1})}{\partial \varepsilon}\right|_{\varepsilon=0} - F(q_2) + I(x < q_2) - f(q_1) \left.\frac{\partial g_{1,\varepsilon}(c^{-1})}{\partial \varepsilon}\right|_{\varepsilon=0} + F(q_1) - I(x < q_1). \tag{A.4}$$

By definition of g_{F_ε}, we have

$$(1-\varepsilon)\{F(c^{-1}+g_{2,\varepsilon}(c^{-1}))-F(g_{2,\varepsilon}(c^{-1})-c^{-1})\}+\varepsilon\{\Delta_x(c^{-1}+g_{2,\varepsilon}(c^{-1}))-\Delta_x(g_{2,\varepsilon}(c^{-1})-c^{-1})\}=1/2.$$

Differentiating yields

$$\{f(g_{2,\varepsilon}(c^{-1})+c^{-1})-f(g_{2,\varepsilon}(c^{-1})-c^{-1})\}\left.\frac{\partial g_{2,\varepsilon}(c^{-1})}{\partial\varepsilon}\right|_{\varepsilon=0}-\{F(c^{-1}+g_{2,\varepsilon}(c^{-1}))-F(g_{2,\varepsilon}(c^{-1})-c^{-1})\}+\{I(x\le c^{-1}+g_{2,\varepsilon}(c^{-1}))-I(x\le g_{2,\varepsilon}(c^{-1})-c^{-1})\}=0.$$

Using the relation $F(q_2+c^{-1})-F(q_2-c^{-1})=1/2$, we obtain

$$\left.\frac{\partial g_{2,\varepsilon}(c^{-1})}{\partial\varepsilon}\right|_{\varepsilon=0}=\frac{\operatorname{sgn}(|q_2-x|-c^{-1})}{2(f(q_2+c^{-1})-f(q_2-c^{-1}))}. \tag{A.5}$$

Analogously, we also find

$$\left.\frac{\partial g_{1,\varepsilon}(c^{-1})}{\partial\varepsilon}\right|_{\varepsilon=0}=\frac{\operatorname{sgn}(|q_1-x|-c^{-1})}{2(f(q_1+c^{-1})-f(q_1-c^{-1}))}. \tag{A.6}$$

Combining (A.1), (A.2), (A.4), (A.5), and (A.6) gives the desired equation (2.6).

Proof of Theorem 4. Let ε be any value in $]0, \frac{1}{2}[$. We obtain $B^+(\varepsilon, F)$ as the limit of $S(G_n)$, where $G_n=(1-\varepsilon)F+\varepsilon\Delta_{x_n}$ and x_n goes to infinity. Let Y be distributed according to G_n. Then $g_{G_n}(x)$ is the smallest positive value for which

$$G_n(x+g_{G_n}(x))-G_n(x-g_{G_n}(x))+P(Y=x-g_{G_n}(x))\ge .5.$$

Substituting G_n yields

$$(1-\varepsilon)\{F(x+g_{G_n}(x))-F(x-g_{G_n}(x))\}+\varepsilon\{\Delta_{x_n}(x+g_{G_n}(x))-\Delta_{x_n}(x-g_{G_n}(x))\}+\varepsilon I(x-g_{G_n}(x)=x_n)\ge .5.$$

Because $\varepsilon<.5$, for each $M>0$ we can choose n_0 such that for all $n>n_0$

$$\sup_{|x|<M}(x+g_{G_n}(x))<x_n \tag{A.7}$$

and

$$\inf_{|x|\ge M} g_{G_n}(x)\ge\sup_{|x|<M} g_{G_n}(x), \tag{A.8}$$

making use of the properties of F. Using (A.7) yields $g^+(x)=g_{G_n}(x)$ for all $|x|<M$ and for all $n>n_0$. If we choose M such that

$$\operatorname{med}|Y|=F^{-1}\left(\frac{3-2\varepsilon}{4(1-\varepsilon)}\right)<M,$$

then, using (A.8) and the symmetry and monotonicity of $g_{G_n}(x)$ on $]-M, M[$ (from Eq. (2.12)), we obtain

$$S(G_n)=cg_{G_n}(\operatorname{med}|Y|),$$

which does not depend on n (provided that $n>n_0$). Therefore, Equations (2.11) and (2.12) together determine $B^+(\varepsilon, F)$.

For the implosion bias curve, we have $B^-(\varepsilon, F)=S(G_0)$, with $G_0=(1-\varepsilon)F+\varepsilon\Delta_0$. Now $g_{G_0}(x)$ is the smallest positive solution of

$$(1-\varepsilon)\{F(x+g_{G_0}(x))-F(x-g_{G_0}(x))\}+\varepsilon\{\Delta_0(x+g_{G_0}(x))-\Delta_0(x-g_{G_0}(x))\}+\varepsilon I(x-g_{G_0}(x)=0)\ge .5. \tag{7.9}$$

It follows that $x\le g_{G_0}(x)$ and that

$$x=g_{G_0}(x)\quad\text{iff } x\ge F^{-1}\left(\frac{2-3\varepsilon}{2(1-\varepsilon)}\right)\Big/2.$$

When $x<g_{G_0}(x)$, Equation (A.9) entails $g_{G_0}(x)=g^-(x)$. Therefore, $g_{G_0}(x)$ is symmetric and increasing on the positive numbers. This implies that

$$B^-(\varepsilon, F)=S(G_0)=cg_{G_0}(\operatorname{med}|Y|)=cg_{G_0}\left(F^{-1}\left(\frac{3-4\varepsilon}{4(1-\varepsilon)}\right)\right), \tag{A.10}$$

where Y is distributed according to G_0. Because $F^{-1}\{(3-4\varepsilon)/[4(1-\varepsilon)]\}$ is smaller than $F^{-1}\{(2-3\varepsilon)/[2(1-\varepsilon)]\}/2$, Equations (2.13) and (2.14) provide an implicit determination of $B^-(\varepsilon, F)$.

Proof of Theorem 7. Let the model distribution F be symmetric about 0 and have a density. At $G=(1-\varepsilon)F+\varepsilon H$, the functional $Q(G)$ is the smallest positive solution of

$$\int G(y+d^{-1}Q(G))\,dG(y)\ge 5/8, \tag{A.11}$$

where d is the constant defined by (3.7). If X is distributed according to F and Y_1, Y_2 according to G, then we can rewrite (A.11) as

$$(1-\varepsilon)^2F^{*2}(d^{-1}Q(G))+\varepsilon(1-\varepsilon)\{1+P(|X-Y_1|\le d^{-1}Q(G))\}+\varepsilon^2P((Y_1-Y_2)\le d^{-1}Q(G))\ge 5/8. \tag{A.12}$$

Note that each term in (A.12) is increasing in $Q(G)$. To maximize $Q(G)$, we have to minimize $P(|X-Y_1|\le d^{-1}Q(G))$ and $P((Y_1-Y_2)\le d^{-1}Q(G))$. These terms approach their lower bounds (0 and $\frac{1}{2}$) when both location and scale of H tend to infinity. (For instance, consider a sequence of Gaussian distributions with location n and standard deviation n.) Therefore, (A.12) yields formula (3.10).

For the implosion bias curve, we have to maximize $P(|X-Y_1|\le d^{-1}Q(G))$ and $P((Y_1-Y_2)\le d^{-1}Q(G))$. When we choose $H=\Delta_0$, we attain the maximal values $2F(Q(G))-1$ and 1. Combining this with (A.12), we obtain Equation (3.11), which determines $B^-(\varepsilon, F)$.

[Received September 1991. Revised June 1992.]

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), *Robust Estimates of Location: Survey and Advances,* Princeton, NJ: Princeton University Press.

Bickel, P. J., and Lehmann, E. L. (1976), "Descriptive Statistics for Nonparametric Models III: Dispersion," *The Annals of Statistics,* 4, 1139–1158.

——— (1979), "Descriptive Statistics for Nonparametric Models IV: Spread," in *Contributions to Statistics, Hájek Memorial Volume,* ed. J. Jurečková, Prague: Academia, pp. 33–40.

Boos, D. (1982), "A Test for Asymmetry Associated With the Hodges–Lehmann Estimator," *Journal of the American Statistical Association,* 77, 647–651.

Croux, C., and Rousseeuw, P. J. (1992a), "A Class of High-Breakdown Scale Estimators Based on Subranges," *Communications in Statistics, Theory and Methods,* 21, 1935–1951.

——— (1992b), "Time-Efficient Algorithms for Two Highly Robust Estimators of Scale," in *Computational Statistics, Volume 1,* eds. Y. Dodge and J. Whittaker, Heidelberg: Physika-Verlag, 411–428.

Croux, C., Rousseeuw, P. J., and Hössjer, O. (1993), "Generalized *S*-Estimators," technical report 93-03, Universitaire Instelling Antwerpen, Belgium.

Gini, C. (1912), "Variabilità é mutabilita, contributo allo studio delle distribuzioni e delle relazioni statistiche," *Studi Economico—Giuridici della R. Università di Cagliari,* 3, 3–159.

Grübel, R. (1988), "The Length of the Shorth," *The Annals of Statistics,* 16, 619–628.

Hampel, F. R. (1974), "The Influence Curve and its Role in Robust Estimation," *Journal of the American Statistical Association,* 69, 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions,* New York: John Wiley.

He, X., and Simpson, D. G. (1993), "Lower Bounds for Contamination Bias: Globally Minimax versus Locally Linear Estimation," *The Annals of Statistics,* 21, 314–337.

Hodges, J. L., Jr., and Lehmann, E. L. (1963), "Estimates of Location Based on Rank Tests," *Annals of Mathematical Statistics,* 34, 598–611.

Hössjer, O., Croux, C., and Rousseeuw, P. J. (1993), "Asymptotics of a Very Robust Scale Estimator," technical report 93-10, Universitaire Instelling Antwerpen, Belgium.

Huber, P. J. (1981), *Robust Statistics,* New York: John Wiley.

Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions-I,* New York: John Wiley.

Kaufman, L., and Rousseeuw, P. J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis,* New York: John Wiley.

Martin, R. D., Yohai, V. J., and Zamar, R. H. (1989), "Min-Max Bias Robust Regression," *The Annals of Statistics,* 17, 1608–1630.

Martin, R. D., and Zamar, R. H. (1989), "Asymptotically Min-Max Bias Robust M-estimates of Scale for Positive Random Variables," *Journal of the American Statistical Association,* 84, 494–501.

——— (1991), "Bias Robust Estimation of Scale when Location is Unknown," submitted to *The Annals of Statistics.*

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

Rousseeuw, P. J., and Bassett, G. W., Jr. (1990), "The Remedian: A Robust Averaging Method for Large Data Sets," *Journal of the American Statistical Association,* 85, 97–104.

Rousseeuw, P. J., and Croux, C. (1991), "Alternatives to the Median Absolute Deviation," Technical Report 91-43, Universitaire Instelling Antwerpen, Belgium.

——— (1992), "Explicit Scale Estimators with High Breakdown Point," in L_1-*Statistical Analysis and Related Methods,* ed. Y. Dodge, Amsterdam: North-Holland, pp. 77–92.

——— (1993), "The Bias of k-Step M-estimators," submitted to *Statistics and Probability Letters.*

Rousseeuw, P. J., and Hubert, M. (1993), "Regression-Free and Robust Estimation of Scale," technical report 93-24, Universitaire Instelling Antwerpen, Belgium.

Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection,* New York: John Wiley.

——— (1988), "A Robust Scale Estimator Based on the Shortest Half," *Statistica Neerlandica,* 42, 103–116.

Rousseeuw, P. J., and Yohai, V. J. (1984), "Robust Regression by Means of S-Estimators," in *Robust and Nonlinear Time Series Analysis,* (Lecture Notes in Statistics 26), eds. J. Franke, W. Härdle, and R. D. Martin, New York: Springer-Verlag, pp. 256–272.

Serfling, R. J. (1984), "Generalized L-, M-, and R-Statistics," *The Annals of Statistics,* 12, 76–86.

Shamos, M. I. (1976), "Geometry and Statistics: Problems at the Interface," in *New Directions and Recent Results in Algorithms and Complexity,* ed. J. F. Traub, New York: Academic Press, pp. 251–280.

Siegel, A. F. (1982), "Robust Regression Using Repeated Medians," *Biometrika,* 69, 242–244.

Tukey, J. W. (1977), *Exploratory Data Analysis,* Reading, MA: Addison-Wesley.