BIAS ROBUST ESTIMATION OF SCALE WHERE LOCATION IS UNKNOWN

by

Douglas R. Martin
Ruben H. Zamar

TECHNICAL REPORT No. 184
October, 1989

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA
Bias Robust Estimation of Scale When Location Is Unknown

R. Douglas Martin
University of Washington
and
Ruben H. Zamar
University of British Columbia

ABSTRACT

In this paper we consider the problem of robust estimation of the scale of the location residuals when the "true" underlying distribution of the data belongs to a contamination neighborhood of a parametric location-scale family.

First we show that a scaled version of the MADAM (median of absolute residuals about the median) is approximately most bias-robust within the class of Huber's proposal II joint estimates of location and scale. Then we consider the larger class of M-estimates of scale with general location and show that a scaled version of the SHORTH (the shortest half of the data) is approximately most bias-robust in this case. The exact min-max asymptotic bias estimate is a scaled order statistic of the residuals about a certain location estimate. The exact order, scaling and location depend on the fraction of contamination, the loss function and the central parametric model.

Finally, we present the results of a Monte Carlo simulation study showing that the scaled SHORTH has attractive finite sample mean square error properties for contaminated Gaussian data.
1. INTRODUCTION

A main theoretical approach to robustness has consisted of studying the asymptotic behavior of an estimate when the underlying distribution of the data belongs to some neighborhood (e.g., $\varepsilon$-contamination or Levy neighborhood) of a parametric model. In this context one tries to obtain estimates which optimize some appealing criterion, e.g., find estimates which minimize the maximum asymptotic variance over a given neighborhood. Huber (1964) is the earliest example of this approach, with a focus on M-estimates of location. The best-known part of Huber (1964) is the result that a particular M-estimate of location, namely the one with psi-function $\psi(x) = \min\{c, \max(x, -c)\}$, minimizes the maximum asymptotic variance over symmetric $\varepsilon$-contamination neighborhoods of a Gaussian model.

A considerably less well known part of Huber (1964) is that concerned with asymptotic bias of location estimates for unrestricted asymmetric $\varepsilon$-contamination neighborhoods of a nominal Gaussian model: Among all translation equivariant estimates, the median minimizes the maximum asymptotic bias over such neighborhoods. The relevance of this result seems considerable in view of the needed realism of allowing asymmetric contamination distributions.

Aside from the above early approach of constructing bias-robust estimates, by which we mean estimates which minimize the maximum asymptotic bias, the main approaches to the bias problem until quite recently have been through the influence curve, the breakdown point, and shrinking $\sqrt{n}$-neighborhoods.

The supremum of the (suitably) normed influence curve of an estimate $T$ is called the gross-error sensitivity, $\text{GES}_T$, a term introduced by Hampel (1974). See also Hampel et al. (1986). $\text{GES}_T$ is an infinitesimal quantity which provides a local linear approximation to the maximal bias of an estimate $T$ for small fractions of contamination. The breakdown
point $\varepsilon_T^*$ of an estimate $T$ is a global measure, namely, the largest fraction of contamination the estimate can tolerate before its bias can be arbitrarily large. The combination of $\text{GES}_T$ and $\varepsilon_T^*$ provides a convenient two-number summary which often provides a good indication of the likely maximum bias behavior of an estimate for all fractions of contamination $\varepsilon$ between $\varepsilon = 0$ and $\varepsilon_T^*$. In this regard, see Hampel (1974), Hampel (1968), Hampel et al. (1986).

The latter approach, which puts asymptotic variance and squared bias on the same footing, and minimizes the maximum mean-squared error, was introduced by Jaeckel (1972), and pursued by a number of authors (e.g., Beran, 1981; Bickel, 1984).

Recently, there has been a renewed interest in bias-robustness. In particular Donoho and Liu (1988a, 1988b) have shown that minimum distance estimates have desirable bias robustness properties. Martin, Yohai and Zamar (1989) have obtained asymptotically min-max bias regression estimates, and Martin and Zamar (1989) have obtained min-max bias estimates of scale for positive random variables.

In this paper we obtain min-max bias robust estimates of scale for contamination models with a nominal distribution which is symmetric about an unknown location parameter. Specifically we shall assume that the distribution $F$ for independent and identically distributed observations $X_1, X_2, \ldots, X_n$ belongs to the $\varepsilon$-contaminated family

$$F_\varepsilon = \left\{ F(x): F(x) = (1-\varepsilon)F_0 \left[ \frac{x - \mu_0}{s_0} \right] + \varepsilon H(x), \ x \in \mathbb{R} \right\}$$

where $F_0$ is symmetric, $\mu_0$ is the unknown location parameter, $s_0$ is the unknown scale parameter, and $\varepsilon$ is fixed in the interval $[0, .5)$.

The first step in obtaining a min-max bias estimate will be to derive the maximal asymptotic bias $B_T(\varepsilon)$ of an estimate $T$ over the family $F_\varepsilon$. From $B_T(\varepsilon)$ one may construct a maximal bias curve, namely a plot of $B_T(\varepsilon)$ versus $\varepsilon$. Such a curve provides a complete view of the maximum bias of an estimate $T$. The curve includes at the origin the
gross error sensitivity $GES_T$, namely the slope of $B_T(\varepsilon)$ at $\varepsilon = 0$, and also the breakdown point $\varepsilon^*_T$, which is the location of the singularity where $B_T(\varepsilon)$ goes to infinity (or for problems with compact parameter space, where $B_T(\varepsilon)$ corresponds to taking $T$ to the boundary of the parameter space). While the two-number summary consisting of $GES_T$ and $\varepsilon^*_T$ provides considerable information, one naturally would like to have the entire curve $B_T(\varepsilon)$ if possible. Not only would such curves allow one to check the range of accuracy of $GES_T$ as a linear approximation, but also such knowledge may lead to different preference orderings of competing estimates than one might make on the basis of $GES_T$ and $\varepsilon^*_T$ alone (e.g., Martin, Yohai and Zamar, 1989, find quite attractive regression estimates for which $GES_T = \infty$).

A pleasant surprise of our work has been that maximal bias curves admit to analytic expressions, which may be evaluated (by numerical methods, if necessary) for a wide range of problems.

Figure 1 displays maximal bias curves associated with outliers for three proposed robust estimates of scale: $H_{95}$, a Huber proposal 2 estimate of scale, adjusted for 95% efficiency at the Guassian model (Huber, 1964); MADM, the median absolute deviation about the median; and SHORTH, the "shortest half" of the data. Observe that $\varepsilon^*_{\text{SHORT}} = \varepsilon^*_{\text{MADM}} = 0.5$, the largest possible value of $\varepsilon^*$ and $\varepsilon^*_{H_{95}} = .17$. The breakdown point of a classical Gaussian maximum likelihood estimator is typically 0. The figure also displays the $GES_T$ slopes, for each of the estimates. These lines provide local linear approximation to the maximum bias which are reasonable for not too large values of $\varepsilon$ (just how large, the reader can judge for himself – see the rule of thumb in Hampel et al., 1986). For asymptotic bias protection against outliers, one clearly chooses the SHORTH, all other things being equal. Of course we need to consider bias due to inliers as well, and we shall do so subsequently.
Bias, maximal bias and breakdown point are all highly transparent concepts, which we have found can easily be communicated to scientists and engineers in many fields. Since a min-max bias robust estimate is one which produces the "smallest" maximal bias curve, this notion is also rather transparent. Thus the value of maximal bias curves as a useful qualitative and quantitative display seems considerable. The emergence of the maximal bias curve as an important tool is due to Hampel et al. (1986).

It may be noted that in general the solution of a min-max bias estimation problem will result in a different estimate, $T^*_\varepsilon$, for each $\varepsilon$. In this regard Huber's (1964) result for location was unusual in that the sample median solved the problem for every $\varepsilon < .5$. On the other hand it has been our experience to date that neither the actual form of a min-max bias estimate $T^*_\varepsilon$ nor its maximal bias curve $B^*(\varepsilon)$ change very much as $\varepsilon$ ranges over $(0,.5)$.

Of course, there is always the question of whether or not an asymptotic theory is relevant for finite sample sizes. Prior results in Martin and Zamar (1989), indicate that: (a) Bias will be a significant component of mean-squared error for rather small to moderate sample sizes, depending on the size of $\varepsilon$, one's notion of "small", and the application at hand; (b) Simulations presented herein show that the finite-sample size mean-squared errors of the bias-robust estimates are smaller than those of previously proposed robust estimates of scale even when both $\varepsilon$ and the sample size are quite small.

The paper is organized as follows. Section 2 introduces M-estimates of scale with very general location estimate. This class includes well known Huber (Proposal 2) M-estimates of location and scale, and also a new class of scale estimates called S-estimates, which are associated with so-called S-estimates of location introduced by Rousseeuw and Yohai (1984) in the regression context. Section 3 mentions the asymmetry of the bias issue for scale estimates, with regard to inliers and outliers, and introduces a class of generalized bias function for dealing with this aspect of the problem. Section 4 constructs min-max bias
robust estimates for the class of Huber (Proposal 2) estimates of location and scale, and shows that the bias robust estimates are well approximated by the well-known median absolute deviation about the median (MADM) estimate. Section 5 constructs min-max bias robust S-estimates of scale, which are shown to be bias robust in the entire class of M-estimates of scale with general location, and also shows that these estimates are reasonably well approximated by the shortest-half of the data (the SHORTH). Section 6 gives some encouraging finite sample size Monte Carlo results. Section 7 closes with some brief comments on the difference between bias robust Huber estimates and S-estimates, and between gross-error sensitivity (GES) and maximal bias.

Our results on the SHORTH complement recent results of Rousseeuw and Leroy (1988), who propose the SHORTH as a robust scale estimator. They derive the influence curve, the finite sample size breakdown point, and a correction factor to achieve approximate finite sample size unbiasedness at the normal distribution.
2. M-ESTIMATES OF SCALE WITH GENERAL LOCATION

Estimates of scale are conveniently viewed as translation invariant, scale equivariant functionals $S(F)$ defined over a more or less rich subset $F$ of distribution functions $F$. In this paper we take $F$ to be large enough to include all empirical distribution functions $F_n$, as well as the $\varepsilon$-contamination family (1.1). The finite sample estimate $\hat{s}_n$ is then obtained by evaluating the functional $S(F)$ at the empirical distribution function $F_n$ of the sample: $\hat{s}_n = S(F_n)$.

Since the location parameter $\mu_0$ in (1.1) is unknown, it must be estimated along with the scale parameter, $s_0$. Let $T(F)$ be a location and scale equivariant "centering" functional, namely one which satisfies

$$T\left\{F\left(\frac{x-b}{a}\right)\right\} = aT\{F(x)\} + b$$

for any real $b$ and positive $a$. The centering estimate $\hat{\mu}_n$ is obtained by evaluating the functional $T(F)$ at $F_n$: $\hat{\mu}_n = T(F_n)$. We work only with $T$ which are Fisher consistent, and continuous at $F_0$.

Let $T$ be any such centering functional. Then an M-estimate of scale with general location $T(F)$ is defined as a solution $S(F)$ to the following equation in $s$:

$$\int \chi \left\{ \frac{x - T(F)}{s} \right\} dF(x) = b \quad (2.1)$$

where the function $\chi$ is even, monotone on $[0, \infty)$, bounded and with at most a finite number of discontinuities. If equation (2.1) has no solution or more than one solution, then $S(F)$ is defined as the infimum of the set of values of $s$ for which the left hand side is smaller than or equal to $b$. In the sequel we often write (2.1) with $s$ replaced by $S(F)$ as the defining equation for $S(F)$. The integral will be replaced with the expectation operator when convenient.
The value of \( b \) is selected in the usual way (Huber, 1981; Hampel et al, 1986) to obtain Fisher consistency of \( S(F) \), i.e., so that \( S(F_0) = s_0 \):

\[
b = E_{F_0} \chi \left\{ \frac{X - T(F_0)}{s_0} \right\}.
\]

(2.2)

Because of the invariance and equivariance properties of \( T \) and \( S \), and the invariance property of the generalized measure of bias introduced in Section 3, it will suffice to take \( \mu_0 = 0 \) and \( s_0 = 1 \).

Because (2.1) assumes only quite general properties of the centering functional \( T \), our definition of \( S(F) \) is sufficiently general to include rather different subclasses. We consider in particular two main sub-classes of M-estimates of scale with general location:

**HUBER ESTIMATES OF SCALE**

In this case the centering functional \( T(F) \) is an M-estimate of location computed simultaneously with \( S(F) \), that is \( T(F) \) and \( S(F) \) are joint solutions of (2.1) and

\[
\int \psi \left\{ \frac{x - T(F)}{S(F)} \right\} dF(x) = 0.
\]

(2.3)

Typically \( \psi \) is chosen to be odd, monotone, bounded and continuous. This class of estimates is often referred to as Huber Proposal 2 joint M-estimates of location and scale (see Huber, 1964).

**S-ESTIMATES OF SCALE**

This class of estimates of scale was introduced by Rousseeuw and Yohai (1984) in the regression context, where the emphasis was on obtaining high breakdown point regression estimates with the usual rates of normal convergence. For each fixed \( t \) let \( S(F, t) \) be the smallest solution to (2.1) with \( T(F) \) replaced by \( t \). Then the S-estimate \( S(F) \) is defined as

\[
S(F) = \inf_t S(F, t).
\]

(2.4)
2.3

The minimizing value \( t^* = T(F) \) is the S-estimate of location. Unlike in Rousseeuw and Yohai (1985), our interest here is not primarily in the behavior of \( t^* \) but in the behavior of \( S(F) \).

Examples 2–3 below are special types of Huber estimates of scale while Examples 4–5 are special types of S-estimates of scale.

**Example 1. The Standard Deviation:** When \( \chi(t) = t^2 \) and \( b = 1 \), we obtain the standard deviation as an S-estimate, and also as a Huber estimate with \( \psi(t) = t \).

**Example 2. Order Statistic of Absolute Residuals about the Median:** Consider the Huber estimate of scale with \( \psi(t) = \text{sgn}(t) \), so that \( T(F) \) is the median, and \( \chi(t) = \chi_a(t) \), where

\[
\chi_a(t) = \begin{cases} 
0 & |t| \leq a \\
1 & |t| > a 
\end{cases}
\]

(2.5)

is a jump function with jump locations \( \pm a \). Then using symmetry of \( F_0 \) gives \( b = 2 \left[ 1 - F_0(a) \right] \). For a sample \( x_1, \ldots, x_n \), let \( r_i = x_i - \text{med} \{ x_i \} \), and let \( |r|_i \) be the \( i \)th order-statistic of the absolute residuals \( |r_i| \). If \( [x] \) denotes the largest integer no greater than \( x \), then

\[
\hat{s}_n = \frac{|r|_{n-[nb]}}{a}.
\]

(2.6)

The min-max bias robust estimate derived in Section 4 is of this form.
Example 3. The Median Absolute Deviation about the Median (MADM): This well-known estimate is a special case of (2.6) obtained when \( a = F_0^{-1} (.75) \), and correspondingly \( b = 1/2 \).

Example 4. S-Estimators for Jump Functions \( \chi_a \): When the function \( \chi \) is of the jump type (2.5), the S-estimate has an interesting simple form. Let \( X_1, \ldots, X_n \) be a sample and \( X_{(1)}, \ldots, X_{(n)} \) be the corresponding order statistics. For convenience let \( a \) be of the form

\[
a = F_0^{-1} \left( \frac{1}{2} + \frac{1}{2} \frac{m}{n} \right) \tag{2.7}
\]

where \( m \) is an integer with \( 1 < m < n \). Then \( b = 2 \left( 1 - F_0(a) \right) = 1 - \frac{m}{n} \). In this case the equation defining \( s_n(t) \) for a sample \( x_1, \ldots, x_n \) is

\[
\frac{1}{n} \sum_{i=1}^{n} \chi_a \left( \frac{x_i - t}{s_n(t)} \right) = 1 - \frac{m}{n}
\]

which is equivalent to

\[
\# \left\{ |x_i - t| \leq a s_n(t) \right\} = m.
\]

From this it follows that the S-estimate \( \hat{s}_n = \min_{i} s_n(t) \) is given by

\[
\hat{s}_n = \frac{Y_{(1)}}{2F_0^{-1} \left( \frac{1}{2} + \frac{1}{2} \frac{m}{n} \right)} \tag{2.8}
\]

where \( Y_j = X_{(m-1+j)} - X_{(j)} \) for \( j = 1, \ldots, n - m + 1 \). Up to a scale factor, \( \hat{s}_n \) is the smallest spacing (of order statistics) of "order" \( m - 1 \). The min-max bias robust estimate derived in Section 5 is of this form.
Example 5. The Shortest Half: When $m = n/2$, (2.8) becomes a scaled version of the "shortest half" of the data. We call this estimate the SHORTH, pointing out to the reader that the SHORTH originally referred to the location estimate obtained as the midpoint of the shortest half of the data. The name SHORTH seems more appropriately applied to the scale estimate. See for example Andrews et al. (1972).
3. Generalized Bias and Technical Assumptions

Generalized Bias

Although the M-estimates of scale with general location introduced in Section 2 are Fisher consistent at the nominal distribution $F_0$, they will in general be asymptotically biased for $F$ different from $F_0$ in (1.1). Furthermore, the "raw" asymptotic bias

$$B_r \{ S(F) \} = S(F) - s_0$$

can be of two distinct kinds: When $F$ is an outliers generating distribution function, the bias $B_r \{ S(F) \}$ will be positive, and when $F$ is an inliers generating distribution, the bias $B_r \{ S(F) \}$ will be negative.

As in Martin and Zamar (1989), we consider generalized bias functions which are scale invariant and flexible. Penalization of positive and negative bias is independently chosen, by allowing the user to put positive and negative bias on different scales. Specifically we define the generalized bias $B \{ S(F) \}$ as

$$B \{ S(F) \} = \begin{cases} g_1 \{ S(F)/s_0 \} & \text{if } 0 < S(F) \leq s_0 \\ g_2 \{ S(F)/s_0 \} & \text{if } s_0 < S(F) < \infty \end{cases}$$

(3.1)

where $g_1$ and $g_2$ are continuous, non-negative and monotone, with $g_1(1) = g_2(1) = 0$ and $\lim_{t \to 0} g_1(t) = \lim_{t \to \infty} g_2(t) = \infty$.

To stress the dependence of the scale functional $S(F)$ on $\chi$ and $T$, we write $S(F) = S(F, \chi, T)$. In addition, $S^+(\chi, T)$ and $S^-(\chi, T)$ will denote the supremum and the infimum of $S(F, \chi, T)$ over $F$. We are interested in the maximum generalized bias

$$\bar{B}(\chi, T) = \max_{F \in F} B \{ S(F, \chi, T) \}.$$ 

(3.2)

From monotonicity of $g_1$ and $g_2$, it follows that

$$\bar{B}(\chi, T) = \max [ g_1 \{ S^-(\chi, T)/s_0 \}, g_2 \{ S^+(\chi, T)/s_0 \} ].$$

(3.3)
Technical Assumptions

In the sections to follow we use the following assumptions:

A0) $F_0$ has an even and unimodal density $f_0$.

A1) $\chi$ is even, monotone nondecreasing on $[0, \infty)$, and bounded, with at most a finite number of discontinuities. In view of (2.1) and (2.2) we can assume without loss of generality that $\chi(\infty) = 1$.

A2) $\chi$ is such that $\varepsilon < E_{F_0} \chi(X) < 1 - \varepsilon$.

Breakdown Point Considerations

The reason for assumption A2 is to prevent the "breakdown" of the scale estimate. From (3.5) and the assumptions on $g_1$ and $g_2$ it follows that in this case the breakdown point is given by the smallest fraction of contamination $\varepsilon$ which will take the scale functional to either zero or infinity. It was shown by Huber (1981) that the breakdown point $BP$ of the "pure" M-scale functional obtained from (2.1) by setting $T(F) = 0$, is given by

$$BP = \min \{ b, 1 - b \}$$

where $b = E_{F_0} \chi(X)$. The case $b < 1 - b$ corresponds to breakdown due to a fraction $b$ of outliers at infinity, and correspondingly the value of the pure scale functional is $\infty$. The case $1 - b < b$ corresponds to the breakdown due to a fraction $b$ of outliers at the origin (often called "implosion"), and correspondingly the value of the pure scale functional is 0.

These comments hold for the M-estimate of scale with general location as well, provided the breakdown point of the latter is at least as large as $\min \{ b, 1 - b \}$. In this case $\varepsilon = b$ results in $S^+(\chi, T) = \infty$, and $\varepsilon = 1 - b$ gives $S^-(\chi, T) = 0$. 
4. BIAS ROBUST HUBER ESTIMATES

In view of the historical importance and high degree of familiarity of Huber estimates of scale we first focus on obtaining bias robust estimates in this class with the restriction that \( \psi \) is monotone. The monotonicity restriction is made in order to have estimates which are asymptotically uniquely defined and thereby make the problem tractable. The degree to which monotonicity restricts the bias robustness of our solution will become clear in Sections 5 and 6.

For the class of Huber estimates we replace \( T \) by \( \psi \) and write \( S^+(\chi, \psi) \) and \( S^-(\chi, \psi) \) for the maximum and minimum of the scale functional \( S(F, \chi, \psi) \). The first step toward finding the bias robust Huber estimate is obtaining expressions (4.8) and (4.10) below for \( S^+(\chi, \psi) \) and \( S^-(\chi, \psi) \), respectively. These expressions are useful in their own right for calculating the maximal generalized bias \( B(\chi, \psi) \) of any Huber scale estimate. The second step is showing that the median, already known to be the bias-robust estimate of location, is also the bias-robust centering functional with regard to the Huber scale estimate. This result is stated in Theorem 1. The third step is to show that there exists a bias-robust jump function \( \chi_{a^*} \) among all jump functions. This is shown by the arguments leading to (4.4) below. Finally, Theorem 2 establishes the main result that \( \chi_{a^*} \) is bias robust among a broad class of \( \chi \) functions. It then follows from Example 2 of Section 2, that the Huber bias-robust estimate of a scale is a scaled order statistic of the absolute deviations about the median.

In this section we use the following assumption in addition to A0–A2 given in the previous section:

A3) \( \psi \) is odd, monotone and bounded, with at most a finite number of discontinuities.

In view of (A3) and equation (2.3) we can assume without losing generality that \( \psi(\infty) = 1 \).

Claims which are made below without proof can be easily verified under A0–A3.
Bias Due to Outliers

The maximum value $S^+(\chi, \Psi)$ of the scale functional $S(\chi, \Psi, F)$ is produced by a point mass contamination at infinity, and such contamination also produces the maximum value for the location functional $T(F)$. Thus both the maximum positive bias for scale and the maximum absolute bias for location are produced by gross outliers. (Note that for location, the maximum and minimum biases are equal in magnitude and of opposite signs.) The estimating equations in this limit case are

\begin{equation}
(1-\varepsilon) E_{F_0} \chi \left\{ \frac{(X-t)}{s} \right\} + \varepsilon = b
\end{equation}

and

\begin{equation}
(1-\varepsilon) E_{F_0} \Psi \left\{ \frac{(X-t)}{s} \right\} + \varepsilon = 0 .
\end{equation}

For all $s > 0$ and $-\infty < t < \infty$ let

\begin{equation}
g_\chi (s, t) = E_{F_0} \chi \left\{ \frac{(X-t)}{s} \right\} .
\end{equation}

Clearly $g_\chi (s, t)$ is continuous, decreasing in $s$. Using the fact that $F_0$ is symmetric and $\chi$ is even, it is not difficult to verify that $g_\chi (s, t)$ is increasing in $t$.

Let $\gamma(t)$ be the unique solution of (4.1) for fixed $t$, and let $r(s)$ be the unique solution of (4.2) for fixed $s > 0$. The function

\begin{equation}
m(t) = r \{ \gamma(t) \}
\end{equation}

is continuous and non-decreasing. Also, $s^*, t^*$ simultaneously satisfy (4.1) and (4.2) if and only if $t^* = m(t^*)$ and $s^* = \gamma(t^*)$.

The median and its maximum asymptotic bias play a role in the remainder of this section. It can be shown (see Huber, 1981, p. 74) that the maximum asymptotic bias of the median is

\begin{equation}
t_0 = F_0^{-1} \{ .5 / (1-\varepsilon) \} .
\end{equation}
The following lemma characterizes the maximum asymptotic biases of location and scale for Huber estimates. It also provides an algorithm for computing these maximum biases.

**LEMMA 1.** Suppose assumptions A0–A3 hold. For each $n \geq 1$ let

$$t_n = m \{ t_{n-1} \}$$

(4.4)

with initial condition $t_0$ given by (4.3'). Let

$$s_n = \gamma(t_n)$$

(4.5)

and

$$t^* = \inf \{ t \geq t_0 : m(t) = t \}.$$

Then

(a) $\lim t_n = t^*$ and $\lim s_n = \gamma(t^*) = s^*$.

(b) The maximum asymptotic bias of $T$ is $t^*$.

(c) $S^+(\chi, \psi) = s^*$.

**Bias Due to Inliers**

The minimum value $S^-(\chi, \psi)$ of the scale functional $S(\chi, \psi, F)$ is produced by a point mass contamination at zero. Thus the maximum negative bias is produced by inliers. In this case the estimating equations are

$$(1-\varepsilon)E_{\nu} \chi\left( (X-t)/s \right) + \varepsilon \chi(t/s) = b$$

(4.6)

and

$$(1-\varepsilon)E_{\nu} \psi\left( (X-t)/s \right) + \varepsilon \psi(-t/s) = 0.$$

(4.7)

Monotonicity of $\psi$ and (4.7) imply $t = 0$ for all $s > 0$. 
Let \( g_t^{-1} \) be the inverse of \( g_\chi(\cdot, t) \) for fixed \( t \), where \( g_\chi(s, t) \) is given by (4.3). Then, from (4.6) with \( t = 0 \), it follows that

\[
S^-(\chi, \psi) = g_0^{-1} \left\{ b / (1 - \varepsilon) \right\} .
\] (4.8)

Observe that this expression is independent of \( \psi \).

**Optimal Centering**

The choice of \( \psi \) will have an effect on the maximum asymptotic bias of the scale estimate by virtue of affecting the bias \( \hat{t}^* \) of the location estimate. The bias-robust choice of \( \psi \) would minimize the asymptotic bias of both the location and scale estimate with respect to the choice of monotone \( \psi \). In fact rather brief arguments show that the median, obtained when \( \psi(t) = \text{sign}(t) \), is the optimal choice.

Since, as noticed before, \( S^-(\chi, \psi) \) doesn't depend on \( \psi \), the optimal choice of centering must be based on \( S^+(\chi, \psi) \) alone. It follows from Lemma 1 and (4.1) with \( t = t^* \) that

\[
S^+(\chi, \psi) = g_{t^*}^{-1} \left\{ (b - \varepsilon) / (1 - \varepsilon) \right\} .
\] (4.10)

By Huber (1964), and 4.2, \( t^* = t^*(\chi, \psi) \geq t_0 \) for all \( \psi \) and \( t^*(\chi, \text{sign}) = t_0 \). Since in addition for all \( 0 < \alpha < 1 \), \( g_t^{-1}(\alpha) \) is a non-decreasing function of \( t \), we have the following result:

**THEOREM 1.** For each fixed \( \chi \) satisfying A1, the median centering functional minimizes the maximum bias of both location and scale among all Huber estimates with \( \psi \) satisfying A3.
More generally, it is not difficult to show that Theorem 1 holds for the class of all M-estimates of scale with centering functionals $T$ having the "monotone" property

$$T(F) \leq T\left\{ (1 - \varepsilon) F_0 + \varepsilon \delta_{\infty} \right\} \quad \forall \quad F \in F_{\varepsilon}$$  \hspace{1cm} (4.11)

where $\delta_{\infty}$ is a point mass at infinity.

The Min-Max Bias Huber Estimate of Scale

Let $\psi^*$ denote the sign function, $\psi^*(t) = \text{sgn}(t)$. By Theorem 1 it will suffice to consider the maximum and minimum values $S^+(\chi, \psi^*)$ and $S^-(\chi, \psi^*)$ of the scale functional when the median is used as centering. It will be shown that under certain conditions the maximum generalized bias

$$\overline{B}(\chi, \psi^*) = \max\left\{ g_1 \{ S^-(\chi, \psi^*) \}, g_2 \{ S^+(\chi, \psi^*) \} \right\}$$  \hspace{1cm} (4.12)

is minimized by a jump function $\chi_a$ with a particular value $a = a^*$ which depends upon $\varepsilon$ (and also on $F_0, g_1, g_2$).

We begin by showing that among all jump functions $\chi_a$ there is an optimal jump function, i.e., one which minimizes

$$\overline{B}(a) = \overline{B}(\chi_a, \psi^*), \quad 0 < a < \infty.$$  \hspace{1cm} (4.13)

Let $a_0 = F_0^{-1} \{ (1 + \varepsilon)/2 \}$ and $a_1 = F_0^{-1} \{ (2 - \varepsilon)/2 \}$, with $0 < \varepsilon < .5$. For the jump functions $\chi_{a_0}, \chi_{a_1}$, the corresponding values of $b$ are $b_0 = 1 - \varepsilon, b_1 = \varepsilon$, respectively. From our discussion of breakdown points in Section 3 we see that as $a \uparrow a_1$, $b \downarrow b_1 = \varepsilon$ and $S^+(\chi_a, \psi^*) \to \infty$. Similarly, as $a \downarrow a_0$, $b \uparrow b_0 = 1 - \varepsilon$ and $S^-(\chi_a, \psi^*) \to 0$. Because of the properties of $g_1, g_2$ we have $\overline{B}(a) \to \infty$ in either case, i.e., as $a \downarrow a_0$ or $a \uparrow a_1$. Now, by continuity of $\overline{B}(a)$ there exists an $a^* \in (a_0, a_1)$ such that

$$\overline{B}(a^*) \leq \overline{B}(a) \quad \forall \quad a_0 < a < a_1.$$  \hspace{1cm} (4.14)

Thus, the jump function $\chi_{a^*}$ is bias-robust among all jump functions $\chi_a$. 
The following theorem gives conditions under which $\chi_{a^*}$ is bias-robust among all Huber estimates.

**THEOREM 2:** Let $s^+(a) = S^+(\chi_a, \psi^*)$, $s^-(a) = S^-(\chi_a, \psi^*)$ and $s^* = s^+(a^*)$. As before $\psi^*(t) = \text{sgn}(t)$. Suppose that in addition to A0–A3, the following conditions hold:

1. $f_0$ is positive and for all $0 < s < 1$, $k_0(x) = f_0(sx)/f_0(x)$ is increasing in $|x|$.
2. $k_0(x) = \{f_0(s^*x - t_0) + f_0(s^*x + t_0)\}/f_0(x)$ is decreasing in $|x|$.
3. $s^-(a)$ and $s^+(a)$ are both strictly monotone at $a = a^*$.

Then

$$\bar{B}(a^*) \leq \bar{B}(\chi, \psi) \quad (4.15)$$

for all $\chi$.

It can be shown that the conditions of the above theorem hold, for example, when $F_0$ is the standard normal distribution and $\varepsilon \leq .35$ (see Martin and Zamar, 1987).

**Near Optimality of MADM**

Since the bias robust estimate of Theorem 2 is based on $\chi_{a^*}$, using the median for centering, it follows from Example 2 that the bias robust Huber estimate is an order statistic of the absolute residuals about the median, scale by $(a^*)^{-1}$: Namely, the $n - \lfloor nb^* \rfloor$ order statistic, where $b^* = 2(1 - F_0(a^*))$ and $a^*$, $b^*$, depend upon $\varepsilon$ (and $F_0, g_1, g_2$).

From the discussion preceding Theorem 2, one sees that $a^* \rightarrow F_0^{-1}(.75)$ as $\varepsilon \rightarrow .5$. That is $a^*$ goes to the upper quartile of the nominal symmetric distribution $F_0$, and correspondingly $b^* \rightarrow .5$. Thus as $\varepsilon \rightarrow .5$, the bias robust estimate is the well known
median absolute deviation about the median (MADM), which has breakdown point $BP = .5$.

It came as a pleasant surprise that for a broad range of $\varepsilon \in (0,.5)$ the bias robust estimate is rather close to MADM with regard to the bias curve and to the value of $b^*$ and the min-max bias, for the leading case of a nominal Gaussian distribution $F_0 = N(0,1)$, and the choice of logarithmic loss $g_1(t) = -\log(t)$, $g_2(t) = \log(t)$. Table 1 shows values of $b^* = b^*(\varepsilon)$, $a^* = a^*(\varepsilon)$ and the min-max bias $\bar{B}(a^*)$, for a range of $\varepsilon$ values.

The value of $a$ for the MADM is 0.674. Therefore in this case there is no appreciable difference between the MADM and the bias-robust estimate.
5. BIAS ROBUST S-ESTIMATES

One naturally wonders whether greater bias robustness can be obtained by enlarging the class of estimates over which one searches for a min-max bias solution. In particular one may consider the entire class of M-estimates of scale with very general location, by letting \( T \) be an arbitrary location-scale equivariant estimate of location. This larger class will of course include Huber estimates of scale with redescending as well as monotone \( \psi \).

As a first step in dealing with this problem, we show that it suffices to work with the smaller class of S-estimates of scale.

**THEOREM 3:** Suppose that \( F_0 \) satisfies A0 and \( \chi \) satisfies A1 and A2. Let \( h_\chi(s) = g_\chi(s, 0) \), where \( g_\chi(s, t) \) is given by (4.3). Let \( T_\chi \) be the S-estimate of location based on \( \chi \), and let \( T \) be any location-scale equivariant estimate satisfying the quite reasonable condition that

\[
T \{ (1 - \epsilon) F_0 + \epsilon \delta_0 \} = 0
\]

where \( \delta_0 \) is a point mass distribution at 0. Then,

(a) \( S^+(\chi, T_\chi) = h_\chi^{-1} \{ (b - \epsilon) / (1 - \epsilon) \} \leq S^+(\chi, T) \)

and

(b) \( S^-(\chi, T_\chi) = h_\chi^{-1} \{ b / (1 - \epsilon) \} = S^-(\chi, T) \).

This result paves the way for obtaining our main result.

**THEOREM 4:** Suppose that \( F_0 \) satisfies A0 and \( \chi \) satisfies A1 and A2. Then there exists a jump function \( \chi_{a^*} \), such that the S-estimate of scale based on \( \chi_{a^*} \) has min-max asymptotic bias over the class of all M-estimates of scale with general location.

It follows that the bias robust S-estimate of scale is given by equation (2.8) with

\[
m^* = \left\lceil \frac{2 F_0(a^*) - 1}{n} \right\rceil n
\]
where \([x]\) denotes the largest integer less than or equal to \(x\). The value of \(a^*\) depends on \(\varepsilon\) (and also on \(F_0, g_1, g_2\)). As \(\varepsilon \to 0.5\), \(a^* \to F_0^{-1}(0.75)\) \(m^* \to [n/2]\), and the bias-robust estimate tends to the SHORT estimate of scale.

We again had the pleasant surprise that, for \(F_0 = N(0, 1)\) and a wide range of values of \(\varepsilon\), the bias-robust estimate is well approximated by the SHORT scale estimate when the generalized bias is logarithmic. Table 2 shows the values of the jump point \(a^* = a^*(\varepsilon)\), the breakdown point \(BP^* = BP^*(\varepsilon)\), the min-max bias \(\bar{B}(a^*)\), and the maximum bias of the MADM \(\bar{B}(MADM)\), for a range of \(\varepsilon\) values. When \(F_0 = N(0, 1)\) and the generalized bias is logarithmic, the bias-robust scale estimate is well approximated by the SHORT.

Note that once again high values of \(BP^*(\varepsilon)\) are obtained even when the estimate is designed for small \(\varepsilon\).

It should be remarked that the S-estimate of location associated with the name SHORT, namely the midpoint of the shortest half of the data, has a "slow" rate of convergence (see Andrews et al., 1972). However, the SHORT estimate of scale has the usual rate of normal convergence (see Grubel, 1988).
6. FINITE-SAMPLE SIZE PERFORMANCE

To get some idea about the relevance of our asymptotic results for finite sample sizes, we carried out some Monte Carlo studies to estimate the bias and mean-squared error of the MADM and SHORTH estimates of scale for the case of a nominal Gaussian model \( F_0 = N(0, 1) \). Since the structure and asymptotic biases of MADM are extremely close to those of the optimal Huber estimate, while those of the SHORTH are moderately close to those of the optimal SS-estimate (cf. relevant discussion at the ends of Sections 4 and 5), the results of such Monte Carlo studies are expected to reflect quite well the behavior of the min-max bias robust estimates.

Bias

Figures 3(a)–(d) display the maximum bias due to outliers, on the log scale, of MADM and SHORTH for sample sizes \( n = 20, 40, 100, \infty \). The corresponding results for inliers are displayed in Figures 3(e)-(h). For each sample size \( n \), the corresponding sample contains exactly \( n \) “bad” data points all equal to the same value \( x \). For the inlier cases \( x = 0 \) and for the outlier cases \( x = 4 \). The value \( x = 4 \) was empirically determined to be already large enough to produce maximum biases. The number of Monte Carlo replications was 1000, and the values displayed are

\[
\frac{1}{1000} \sum_{m=1}^{100} \log \hat{s}_m
\]

were \( \hat{s}_m \) is the value of the MADM or SHORTH estimate for the \( m \)th Monte Carlo replication.

The general conclusions are as follows:

(1) For both outliers and inliers the asymptotic bias tends to be rather close to the finite sample size bias.
(b) The SHORTH is substantially superior to the MADM for outliers when the fraction of contamination is larger than .2 to .3, while below $\varepsilon = .2$, their performance is nearly identical.

(c) The MADM is slightly superior to the SHORTH for inliers when the fraction of contamination is larger than .1, while their performance is nearly identical for $\varepsilon < .1$.

Mean-Squared Errors

For two scale estimates $\hat{s}^1$ and $\hat{s}^2$ at sample size $n$ we define Monte Carlo estimates of mean-squared-error relative efficiency as

$$\text{EFF}_n(\hat{s}^1, \hat{s}^2) = \frac{\sum_{m=1}^{M} (\log \hat{s}^2_m)^2}{\sum_{m=1}^{M} (\log \hat{s}^1_m)^2}$$

where $M$ is the number of Monte Carlo replicates, $\hat{s}^1_m, \hat{s}^2_m$ are the $m$th replicates for each of the two estimates at sample size $n$, and it is assumed that the true scale is $s_0 = 1$.

Table 3 displays the Monte Carlo values of EEF (SHORTH, MADM) for several sampling situations. The main conclusions from Table 3 are:

1. For small values of epsilon ($\varepsilon \leq 0.10$) SHORTH is slightly better than MADM (between 5 and 10% better) for both inliers and outliers and all considered sample sizes.

2. For larger values of epsilon ($\varepsilon > 0.10$) SHORTH is better than MADM in the presence of outliers and MADM is better than SHORTH in the presence of inliers.
(3) Except for $\varepsilon = 0$, the difference between the mean-squared-error performance of MADM and SHORTH increases somewhat with increasing $n$, the more so the larger $\varepsilon$ is.
7. HUBER ESTIMATES VERSUS S-ESTIMATES: MADM VERSUS SHORTH

As we commented in the beginning of Section 5, by working with the class of S-estimates, one is able to obtain a bias robust estimate of scale for the entire class of M-estimates with very general location, and the latter class includes Huber estimates with redescending \( \psi \) functions. In fact it is not difficult to see that the S-estimate of location functional \( t^* = T(F) \) associated with the S-scale functional \( S(F) \) (see (2.4)) is in fact an M-estimate of location with redescending psi-function \( \psi = \chi' \):

For each real \( t \), monotonicity of \( \chi \) gives

\[
E_F \chi \left( \frac{X - t}{S(F)} \right) \geq E_F \chi \left( \frac{X - t}{S(F,t)} \right) = E_F \chi \left( \frac{X - t^*}{S(F)} \right) = b.
\]

So \( t^* \) minimizes the loss function \( L(t) = E_F \chi \left[ \frac{X - t}{S(F)} \right] \). And hence \( t^* \) satisfies the location M-estimate equation

\[
E_F \chi' \left( \frac{X - t^*}{S(F,t)} \right) = 0.
\]

Figures 2a and 2b display the maximal generalized bias curves of the bias robust Huber and S-estimates of scale, for the case of outliers and inliers respectively. The choice of generalized bias is the "logarithmic" bias obtained with \( g_1(t) = -\log t \), \( g_2(t) = \log t \). Figures 2a and 2b also give the maximal "logarithmic" bias for the MADM and SHORTH which as we already have noted are limiting cases of the bias robust Huber and S-estimates, respectively, as \( \varepsilon \to .5 \). Figure 2 reveals uniformly smaller bias for the min-max S-estimate than for the min-max Huber-estimate, and similarly for the SHORTH versus the MADM, in the case of Figure 2a. In the outliers case this is evidently a consequence of the S-estimate location estimate being an M-estimate with redescending \( \psi \), which suffers no bias for gross outliers.
It is apparent that the optimal S-estimate gives up some bias robustness toward outliers relative to the SHORTH, in order to achieve greater bias robustness toward inliers than the SHORTH. This, of course, is a consequence of the relative way in which the generalized logarithmic bias penalizes positive and negative bias. It is worth noting that if one is concerned only about outliers, then the SHORTH is the estimate of choice with regard to bias. Also, referring back to Figure 1 we would remark that the price paid for using a high-efficiency (e.g., 95%) Huber estimate is exceedingly high in terms of maximal bias and breakdown point.

Hampel et al. (1986) established that based on the gross-error sensitivity, the MADM is the most bias robust estimate of scale for (vanishingly) small fractions of contamination $\varepsilon$. In fact the SHORTH has the same influence curve and hence the same GES as the MADM, 0.787. However, this leaves unanswered the question of optimality for each $\varepsilon \in (0,.5)$, and it should not be too surprising that there exists a better estimate than MADM from the global (i.e., $\varepsilon > 0$) point of view.

It must be noted that the GES approximation is remarkable good for $\varepsilon < .2$, with the approximation being better the more bias robustness the estimate possesses. This provides substantial reconfirmation of the utility of the influence curve and GES as a measure of maximal bias.

At the same time one should be aware that the GES linear approximation may be less accurate for problems with nuisance parameters. For example, in the present context the GES approximation to the maximal bias curve for the MADM in the case of outliers does not reflect the impact of the bias of the location estimate for positive $\varepsilon$. This, of course, is not a problem for the SHORTH scale estimate with outliers, which helps explain why the GES approximation is better is this case.
APPENDIX

Proof of Lemma 1

Since the median minimizes the maximum asymptotic bias among translation equivariant estimates (Huber, 1964), and since $t_0$ and $t_1$ are the maximum asymptotic biases of the median and a translation equivariant estimate of location, we have $t_0 \leq t_1$. By definition of $t^*$, $t^* \leq t_0$ and by monotonicity of $m(t)$, for all $n$, $t^* \leq t_n \leq t_{n+1} \leq \lim_{n \to \infty} t_n = t^{**}$.

On the other hand, by continuity of $m(t)$, $t^{**} = \lim_{n \to \infty} t_n = \lim_{n \to \infty} m(t_n) = m(t^{**})$. This implies $t^{**} \leq t^*$, proving the first part of (a). The second part follows from the first and continuity of $\gamma$.

To prove (b) and (c) observe that $t^*$ is a lower bound for the maximum bias of the Huber estimate of location. This lower bound is achieved if the estimate is computed by the recursion formula (4.4).

For each fixed $b \in (\epsilon, 1-\epsilon)$, let
\[
C_b = \{ \chi : E_{F_0} \chi(X) = b, \chi \text{ satisfies } (A1) \}.
\]

Also, let
\[
C = \{ \chi : \chi \text{satisfies } (A1) \} = \bigcup_{\epsilon < b < 1-\epsilon} C_b.
\]

The following lemma is needed in the proof of Theorem 2.

LEMMA 3: Fix $b \in (\epsilon, 1-\epsilon)$ and let $a = F_0^{-1} \{ (b/2) - 1 \}$. Let $\psi^*(t) = \text{sgn}(t)$.

Under the assumptions of Theorem 2 we have
\[
(a) \quad S^-\chi_a, \psi^* \geq S^-\chi, \psi^*, \forall \chi \in C_b
\]
\[
(b) \quad g_\chi(s^*, t_0) \geq g_\chi(s^*, t_0), \forall \chi \in C_b
\]
Proof: Part (a) directly follows from Lemma A3 in Martin and Zamar (1988).

To prove (b), notice that for all $\chi \in C_b$ we have \[ \int_{-a}^{a} \chi(x) f_0(x) dx = 2\int_{a}^{\infty} (1 - \chi(x)) f_0(x) dx. \] Thus,

\[ s^* \int_{-a}^{a} \chi(x) \left\{ f_0(s^*x - t_0) + f_0(s^*x + t_0) \right\} dx = \]

\[ = s^* \int_{-a}^{a} \chi(x) f_0(x) k_1(x) dx \geq s^* k_1(a) \int_{-a}^{a} \chi(x) f_0(x) dx \]

\[ = s^* k_1(a) 2\int_{a}^{\infty} \{ 1 - \chi(x) \} f_0(x) dx \]

\[ = k_1(a) s^* \left[ \int_{-\infty}^{-a} \{ 1 - \chi(x) \} f_0(x) dx + \int_{a}^{\infty} \{ 1 + \chi(x) \} f_0(x) dx \right] \]

\[ \geq s^* \int_{-\infty}^{-a} \{ 1 - \chi(x) \} f_0(x) k_1(x) dx + s^* \int_{a}^{\infty} \{ 1 - \chi(x) \} f_0(x) k_1(x) dx. \]

Therefore,

\[ \int_{-\infty}^{\infty} \chi \left( \frac{x - t_0}{s^*} \right) f_0(x) dx \geq \int_{-\infty}^{-a} k_1(x) f_0(x) dx + \int_{a}^{\infty} k_1(x) f_0(x) dx \]

\[ = \int_{-\infty}^{\infty} \chi_a \left( \frac{x - t_0}{s^*} \right) f_0(x) dx. \]

Proof of Theorem 2

First of all we notice that since $s^+(a)$ and $s^-(a)$ are increasing at $a^*$ and $g_1$ and $g_2$ (see (3.5)) are strictly monotone, we have $g_1(s^+(a^*)) = g_2(s^-(a^*)) = \overline{B}(a^*)$.

Let $\chi \in C$ with $b = E_{F_0} \chi(X)$. Let $a = F_0^{-1} \left( 1 - \frac{b}{2} \right)$ so that $b = E_{F_0} \chi_a(X)$. If $g_\chi(s^*, t_0) \geq (b - \varepsilon)/(1 - \varepsilon)$ then $S^+(\chi, \psi^*) \geq s^*$. So $\overline{B}(\chi, \psi^*) \geq g_2(S^+(\chi, \psi^*)) \geq g_2(s^*) = \overline{B}(a^*)$. 

\[ \square \]
On the other hand, suppose that \( g_\chi(s^*, t_0) < (b - \varepsilon)/(1 - \varepsilon) \), that is \( S^+(\chi, \psi^*) < s^* \).

Since \( \chi \in C_b \), by Lemma 3(b) we have \( g_\chi(s^*, t_0) \leq g_\chi(s^*, t_0) < (b - \varepsilon)/(1 - \varepsilon) \).

Hence \( s^+(a) < s^* \), too. In view of the optimality of \( \chi_a^* \) among jump functions we have \( \overline{B}(a) \geq \overline{B}(a^*) \) and so \( g_1 \{ s^-(a) \} \geq g_1 \{ s^-(a^*) \} \). For the particular \( b \) in question, by Lemma 3(a), \( S^-(\chi_a^*, \psi^*) \leq S^-(\chi, \psi^*) \). Therefore \( \overline{B}(\chi, \psi^*) \geq g_1 \{ S(\chi, \psi^*) \} \geq g_1 \{ s^-(a) \} \geq g_1 \{ s^-(a^*) \} \geq \overline{B}(a^*) \), and the theorem follows.

**Proof of Theorem 3.**

Let \( F_\infty = (1 - \varepsilon)F_0 + \varepsilon \delta_\infty \), \( t_\infty = T(F_\infty) \) and \( s_\infty = S \{ F_\infty, \chi, T(F_\infty) \} \).

First notice that

\[
    h^{-1} \left\{ \frac{(b - \varepsilon)}{(1 - \varepsilon)} \right\} = \sup_{F \in \mathcal{F}_\varepsilon} S(F, \chi, 0) = S(F_\infty, \chi, 0) \tag{*}
\]

where \( S(F, \chi, 0) \) is the M-scale functional based on \( \chi \) and the true location \( 0 \).

By definition of the SS-estimate of scale, \( S(F, \chi) \), for all \( F \in \mathcal{F}_\varepsilon \),

\[
    S(F, \chi) = \inf_t \gamma(t, F) \leq \inf_t \gamma(t, F_\infty) = S(F_\infty, \chi, 0),
\]

and so

\[
    S^+(\chi, T_\chi) \leq S(F_\infty, \chi, 0)
\]

Assume first that \( s_\infty < \infty \) and so \( |t_\infty| < \infty \). By monotonicity of \( g(s, t) \)

\[
    b = E_{F_\infty} \left[ \chi \{ (X - t_\infty)/s_\infty \} \right] = (1 - \varepsilon)g(s_\infty, t_\infty) + \varepsilon
\]

\[
    \geq (1 - \varepsilon)g(s_\infty, 0) + \varepsilon = E_{F_\infty} \{ \chi(X/s_\infty) \}.
\]

Therefore,

\[
    S(F_\infty, \chi, 0) \leq s_\infty \leq S^+(\chi, T) \tag{**}
\]
If $s_\infty = \infty$, (***) trivially holds.

Now (*) and (**) imply

$$S^+(\chi, T) \leq h^{-1} \{ (b-\varepsilon)/(1-\varepsilon) \} \leq S^+(\chi, T)$$

and (a) follows by taking $T = T_\chi$.

To prove (b) write

$$h^{-1} \{ b/(1-\varepsilon) \} = \inf_{F \in F_\varepsilon} S(F, \chi, 0) = S\{ F_0^*, \chi, 0 \}$$

where $F_0^* = (1-\varepsilon)F_0 + \varepsilon \delta_0$. For all $F \in F_\varepsilon$, $t \in \mathbb{R}$ and $s > 0$

$$E_F[\chi((X-t)/s)] \geq (1-\varepsilon)E_{F_0}[\chi((X-t)/s)] \geq (1-\varepsilon)E_{F_0}[\chi(X/s)]$$

$$= E_{F_0}[\chi(X/s)]$$

Therefore, for all M-estimate of scale based on the given $\chi$ and for all $T$ satisfying the assumption of this theorem we have

$$S^-(\chi, T) = h^{-1} \{ b/(1-\varepsilon) \}.$$  \quad \square

Proof of Theorem 4

Follows directly from Theorem 3 and Theorem 2 in Martin and Zamar (1987).
<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>$a^*$</th>
<th>$BP^*$</th>
<th>$\bar{B}(a^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.674</td>
<td>0.50</td>
<td>0.063</td>
</tr>
<tr>
<td>0.10</td>
<td>0.673</td>
<td>0.50</td>
<td>0.135</td>
</tr>
<tr>
<td>0.20</td>
<td>0.673</td>
<td>0.50</td>
<td>0.324</td>
</tr>
<tr>
<td>0.30</td>
<td>0.676</td>
<td>0.50</td>
<td>0.609</td>
</tr>
<tr>
<td>0.35</td>
<td>0.682</td>
<td>0.50</td>
<td>0.812</td>
</tr>
</tbody>
</table>

Table 1  Bias-robust Huber Estimates of Scale
When $F_0$ = Standard Normal.
Logarithmic Scale
<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$a^*$</th>
<th>$BP^*$</th>
<th>$\bar{B}(a^*)$</th>
<th>$B(\text{MADM})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.700</td>
<td>0.48</td>
<td>0.127</td>
<td>0.135</td>
</tr>
<tr>
<td>0.20</td>
<td>0.726</td>
<td>0.47</td>
<td>0.284</td>
<td>0.322</td>
</tr>
<tr>
<td>0.30</td>
<td>0.750</td>
<td>0.45</td>
<td>0.494</td>
<td>0.612</td>
</tr>
<tr>
<td>0.40</td>
<td>0.762</td>
<td>0.45</td>
<td>0.844</td>
<td>1.166</td>
</tr>
<tr>
<td>0.45</td>
<td>0.746</td>
<td>0.46</td>
<td>1.236</td>
<td>1.780</td>
</tr>
</tbody>
</table>

Table 2  Bias-robust M-estimates of Scale
When $F_0 = \text{Standard Normal}$.
Logarithmic Scale
<table>
<thead>
<tr>
<th>eps</th>
<th>Outliers</th>
<th>Inliers</th>
<th>Outliers</th>
<th>Inliers</th>
<th>Outliers</th>
<th>Inliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.09</td>
<td>1.09</td>
<td>1.10</td>
<td>1.10</td>
<td>1.18</td>
<td>1.18</td>
</tr>
<tr>
<td>0.05</td>
<td>1.11</td>
<td>1.11</td>
<td>1.06</td>
<td>1.12</td>
<td>1.06</td>
<td>1.00</td>
</tr>
<tr>
<td>0.10</td>
<td>1.08</td>
<td>1.07</td>
<td>1.00</td>
<td>1.02</td>
<td>1.07</td>
<td>0.90</td>
</tr>
<tr>
<td>0.15</td>
<td>1.10</td>
<td>0.98</td>
<td>1.04</td>
<td>0.91</td>
<td>1.12</td>
<td>0.83</td>
</tr>
<tr>
<td>0.20</td>
<td>1.10</td>
<td>0.89</td>
<td>1.14</td>
<td>0.82</td>
<td>1.27</td>
<td>0.77</td>
</tr>
<tr>
<td>0.25</td>
<td>1.23</td>
<td>0.85</td>
<td>1.27</td>
<td>0.76</td>
<td>1.39</td>
<td>0.75</td>
</tr>
<tr>
<td>0.30</td>
<td>1.43</td>
<td>0.83</td>
<td>1.46</td>
<td>0.77</td>
<td>1.61</td>
<td>0.78</td>
</tr>
<tr>
<td>0.35</td>
<td>1.62</td>
<td>0.82</td>
<td>1.66</td>
<td>0.77</td>
<td>1.83</td>
<td>0.77</td>
</tr>
<tr>
<td>0.40</td>
<td>1.74</td>
<td>0.85</td>
<td>1.86</td>
<td>0.80</td>
<td>2.01</td>
<td>0.78</td>
</tr>
<tr>
<td>0.45</td>
<td>1.87</td>
<td>0.92</td>
<td>2.00</td>
<td>0.85</td>
<td>2.16</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 3. Mean-squared-error relative efficiencies of SHORTH and MADM.
FIGURE 2: MAXIMUM BIAS OF OPTIMAL SS, OPTIMAL HUBER, MADM AND SHORTH. LOG SCALE

(a) Maximum Bias Due to Outliers

(b) Absolute Maximum Bias Due to Inliers
Figure 3: Finite sample maximum biases for Shorth (solid line) and MADM. Log scale.
References


In this paper we consider the problem of robust estimation of the scale of the location residuals when the "true" underlying distribution of the data belongs to a contamination neighborhood of a parametric location-scale family.

First we show that a scaled version of the MADAM (median of absolute residuals about the median) is approximately most bias-robust
within the class of Huber's proposal II joint estimates of location and scale. Then we consider the larger class of M-estimates of scale with general location and show that a scaled version of the SHORTH (the shortest half of the data) is approximately most bias-robust in this case. The exact min-max asymptotic bias estimate is a scaled order statistic of the residuals about a certain location estimate. The exact order, scaling and location depend on the fraction of contamination, the loss function and the central parametric model.

Finally, we present the results of a Monte Carlo simulation study showing that the scaled SHORTH has attractive finite sample mean square error properties for contaminated Gaussian data.