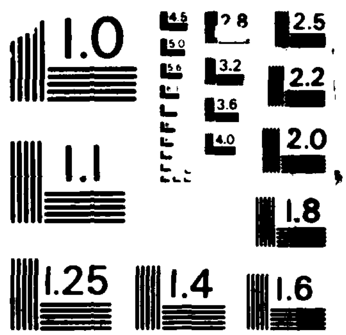
AD-A179 255 MIN-MAX BIAS ROBUST M-ESTIMATES OF SCALE(U) WASHINGTON 1/1
UNIV SEATTLE DEPT OF STATISTICS R D MARTIN ET AL.
DEC 86 TR-72 N00014-84-C-0169

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

MIN–MAX BIAS ROBUST M–ESTIMATES OF SCALE

*R. Douglas Martin**
*Ruben H. Zamar**

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA

ABSTRACT

Min-max bias robust M-estimates of scale are obtained for positive random variables which have ϵ-contaminated distributions. Any such estimate is a scaled order statistic, with the order statistic determined by ϵ. As $\epsilon \to 0.5$ the min-max bias robust estimate becomes a scaled sample median, which thereby enjoys both high breakdown point of 0.5 and min-max bias robustness. Furthermore, for a wide range of ϵ, min-max estimate is quite close to the scaled median in terms of both structure and min-max bias behavior. Results are also obtained for random variables whose distribution is $F = (1-\epsilon)F_0 + \epsilon H$ with F_0 symmetric about an unknown location parameter. In particular we show that when F_0 is normal and $0 \leq \epsilon \leq .35$, the min-max bias M-estimate of scale is a scaled order statistic applied to the absolute value of centered data, with the median as the centering estimate. This estimate is extremely close to a scaled median absolute deviation about the median, in terms of both structure and bias behavior.

February 18, 1987

---

* Research supported by ONR Contract N00014-84-C-0169

1. INTRODUCTION

In a somewhat neglected section of his by now classic paper, Huber (1964) considered the problem of minimizing the maximum asymptotic bias of translation equivariant location estimates over ε-contaminated models with asymmetric contamination. Huber showed that the sample median solves this problem. He also downplayed the significance of the result and concentrated on the "deeper theory" of min-max variance robustness over symmetric neighborhoods (see Section 4.3, 4.4 of Huber, 1981).

In spite of substantial awareness of the importance of controlling bias due to contamination, and other deviations from the nominal model, the literature is relatively devoid of global results on this robustness problem. Hampel's optimal bounded influence estimates (see Hampel et al., 1986, and references therein) control bias only for vanishingly small fractions of contamination, and the shrinking $\sqrt{n}$-neighborhood approach (see Jaeckel, 1971, Bickel, 1984 and references in the latter) is also local in nature. The main global approach to robustness has been the construction of high breakdown point estimates for multi-dimensional problems. Interesting recent results along these lines may be found in Donoho (1982), Donoho and Huber (1983), Rousseeuw (1984), Rousseeuw and Yohai (1984), Yohai (1986), Hampel et al. (1986), Yohai and Zamar (1986).

It should be noted that the problem of bias robustness and the desirability of optimal bias robust estimators, namely min-max bias estimates, is clearly recognized in Hampel et al., (1986, Table 2A, p.176) where the term "most B-robust" is used. We are also aware of very recent work on bias robustness by Donoho and Liu (1986), who demonstrate attractive properties of minimum distance estimators in this regard.

In the present paper we take a small step beyond Huber's (1964) early results by constructing min-max bias robust estimates of scale using M-estimates and ε-contaminated distributions. An M-estimate $\hat{s}_n$ of scale for positive random variables X_i, is defined as

$$\hat{s}_n = \inf\left\{ s : \frac{1}{n}\sum_1^n \chi\left[\frac{X_i}{s}\right] \le b \right\}. \tag{1.1}$$

for an appropriate function χ. We shall work with monotone χ which satisfy A1–A3 stated in Section 2.1. The general definition (1.1) is needed to handle χ which are discontinuous. If χ is continuous, then $\hat{s}_n$ will satisfy

$$\sum_1^n \chi\left[\frac{X_i}{\hat{s}_n}\right] = b \tag{1.1'}$$

and $\hat{s}_n$ will be unique when χ is strictly increasing on $\{ t : \chi(t) \le b \}$.

The positive random variables X_i are assumed to be independent and identically distributed, with common distribution F belonging to the ε-contaminated family

$$\mathbf{F} = \{F : F = (1-\varepsilon)F_0 + \varepsilon H\} \tag{1.2}$$

where H is any distribution function and $0 < \varepsilon < 1$, ε fixed. In a later section of the paper, this setup is modified somewhat to deal with the case of random variables X_i which have a nominal distribution which is symmetric about an unknown location parameter.

Estimation of scale is an asymmetric problem with regard to bias. Outliers give rise to asymptotic positive bias whose maximum positive value is denoted by $B^+(\chi)$, whereas inliers result in a negative bias whose maximum absolute value is denoted by $B^-(\chi)$. These biases are related to the minimum and maximum asymptotic values of the scale estimate, $s^+(\chi)$ and $s^-(\chi)$, by $s^+(\chi) = 1 + B^+(\chi)$ and $s^-(\chi) = 1 - B^-(\chi)$, respectively.

In general $B^+(\chi) \ne B^-(\chi)$, and this asymmetry leads one to introduce a loss function $l(s^-(\chi), s^+(\chi)) = \bar{l}(B^-(\chi), B^+(\chi))$ to be minimized with respect to the choice of χ, thereby resulting in a min-max bias estimate. We shall call any such estimate a *bias-robust* estimate.

It came as a pleasant surprise to find that the maximal biases $B^{+}(\chi)$ and $B^{-}(\chi)$, or equivalently, the minimal and maximal scales $s^{+}(\chi)=1+B^{+}(\chi)$, $s^{-}(\chi)=1-B^{-}(\chi)$, can be obtained in a manageable analytic forms. It is this manageability which allowed us to minimize $l\,(s^{-}(\chi), s^{+}(\chi))$, and thereby obtain bias robust estimates for certain classes of loss functions.

Section 2 of the paper presents the basic theory for the contamination model (1.2) for positive random variables. In Section 2.1 we show that it suffices to work with point mass contaminations. In Section 2.2 we establish the maximal and minimal scale expressions (2.10)–(2.11) for $s^{+}(\chi)$, $s^{-}(\chi)$, which are of value in their own right for calculating the maximal biases of any scale M-estimate. Section 2.3 proves the pleasant result that $s^{+}(\chi)$ is minimized and $s^{-}(\chi)$ is simultaneously maximized by a χ of 0–1 jump function type, χ_a, with jump location a. The resulting estimate is a scaled order statistic. This result is then used in Section 2.4 to establish the optimal choice of a for loss functions l of a fairly general form. In Section 2.5, the breakdown point of the bias robust estimates is exhibited. It is shown that as $\varepsilon \rightarrow 0.5$, the min-max bias robust estimate is a scaled median, which has breakdown point 0.5. Thus the scaled median enjoys two global robustness properties simultaneously.

Section 3 provides some explicit bias, breakdown point and efficiency calculations for the cases of exponential and positive normal F_0. It turns out that for a wide range of contamination fractions, the bias robust estimate is quite close to the scaled median, both in form and with regard to actual bias!

Section 4 briefly treats maximal bias curves, and also the issue of the finite-sample size relevance of our asymptotic theory.

Finally, Section 5 treats the case of estimating the scale of distributions which are symmetric about an unknown location parameter. The problem is now considerably more complicated, but we nonetheless show that jump-function type M-estimates are again optimal under certain conditions. For the case of normal F_0 we are able to establish bias optimality for the range $0 \leq \varepsilon \leq .35$. The bias optimal estimate uses the median for centering, and is very close to the well-known scaled (for Fisher consistency) median-absolute deviation about the median (MADM) estimate, both in form and with regard to bias performance. Thus the scaled MADM is very nearly optimal with regard to both breakdown point and bias.

Proofs are provided in the body of the paper for the key results, whereas proofs of lesser results are relegated to Section 6.

2. THE GENERAL SCALE PROBLEM

2.1 M-estimate Scale Functionals for Contamination Models

The asymptotic value of $\hat{s}_n$, defined by (1.1), is

$$s(\chi, F) = \inf \left\{ s : \int \chi\left(\frac{x}{s}\right) dF(x) \leq b \right\}, \quad F \in \mathbf{F} . \tag{2.1}$$

The constant b will always be determined by F_0 and χ as

$$b = E_{F_0} \chi(X) . \tag{2.2}$$

which insures Fisher consistency: $s(F_0, \chi) = 1$.

We shall work with the class $\mathbf{C}$ of χ functions which satisfy the following conditions:

(A1) χ is nondecreasing and bounded on $[0,\infty)$, with at most a finite number of discontinuities, and $\chi(0) = 0$.

(A2) χ is such that $b = b(\chi)$ satisfies the inequalities $\varepsilon < b < 1 - \varepsilon$

In view of (A1) and (2.1)–(2.2), there is no loss of generality in making the assumption

(A3) $\sup_x \chi(x) = 1$.

Lemma 2.1: Under (A1)-(A3) the asymptotic equation (2.1) has a unique solution $s = s(\chi, F) \in (0, \infty)$ for any $F \in \mathbf{F}$, including the case where H is a point mass δ_∞ at $+\infty$ or δ_0 at 0.

REMARK 2.1: It is easy to see that for any unbounded χ, $s(\chi, F) = \infty$, when $H = \delta_\infty$. Thus, boundedness of χ is a necessary (but not sufficient) condition to obtain a bounded bias over $\mathbf{F}$.

REMARK 2.2: The proof of Lemma 2.1 shows that if $b = 1-\varepsilon$, then (2.1) will have only the solution $s = 0$ when H is concentrated at the origin. This constitutes (implosive) *breakdown* of the estimate due to inliers. The proof also reveals that if $b = \varepsilon$ and H is a point mass at infinity, then $s = \infty$ is the only solution and we have (explosive) *breakdown* due to outliers. See Hampel (1971), Huber (1981) and Donoho and Huber (1983) for various definitions of breakdown point.

Consider the family of point-mass contaminations

$$\mathbf{F}_1 = \{F : F = (1-\varepsilon)F_0 + \varepsilon\delta_y ,\ y \in \bar{R}^+\}$$

were δ_y is a point mass at y and $\bar{R}^+$ is the set of extended nonnegative real numbers. The following lemma shows that the maximum and minimum biases over $\mathbf{F}$ are achieved over the smaller subfamily $\mathbf{F}_1$.

<u>*Lemma 2.2:*</u> Suppose (A1) and (A2) hold. Then

$$\inf_{F \in \mathbf{F}_1} s(\chi, F) = \inf_{F \in \mathbf{F}} s(\chi, F)$$

and

$$\sup_{F \in \mathbf{F}_1} s(\chi, F) = \sup_{f \in \mathbf{F}} s(\chi, F)$$

2.2 Minimum and Maximum Scale Functionals

Since it suffices to work with F in $\mathbf{F}_1$, it is convenient to replace $s(\chi, F)$ by either $s(\chi, y)$, or more simply s_y, where y is the location of the contamination point mass δ_y. Let

$$s^+(\chi) = \sup_{y \geq 0} s(\chi, y) \tag{2.3}$$

denote the *maximal scale*, and let

$$s^{-}(\chi) = \inf_{y \geq 0} s(\chi, y) . \tag{2.4}$$

denote the *minimal scale*. Then the maximum positive bias of $s(\chi, F)$ over $\mathbf{F}$ is

$$B^{+} = s^{+}(\chi) - 1 \tag{2.5}$$

and the maximum of the absolute values of the negative bias is

$$B^{-} = 1 - s^{-}(\chi) . \tag{2.6}$$

Let

$$h_{\chi}(s) = \int_0^{\infty} \chi(\frac{x}{s}) \, dF_0(x) \quad , \qquad 0 < s < \infty \tag{2.7}$$

and note that for $F \in \mathbf{F}_1$, and any $s > s(\chi, y)$ we have

$$(1-\varepsilon) h_{\chi}(s) + \varepsilon \chi(\frac{y}{s}) - b \leq 0 . \tag{2.8}$$

where $b = h_{\chi}(1)$. It is easy to see that h_{χ} is continuous, non-negative and strictly decreasing, with $\lim_{s \to 0} h_{\chi}(s) = 1$, $\lim_{s \to \infty} h_{\chi}(s) = 0$. When χ is continuous, so that s_y solves (2.8) with equality, the following result is almost immediate.

Theorem 2.1: Suppose that (A1) and (A2) hold. Then

$$s^{+}(\chi) = h_{\chi}^{-1}(\frac{b-\varepsilon}{1-\varepsilon}) \tag{2.9}$$

and

$$s^{-}(\chi) = h_{\chi}^{-1}(\frac{b}{1-\varepsilon}) \tag{2.10}$$

with $s^{+}(\chi)$ achieved by point mass contamination δ_y with $y \to \infty$, and $s^{-}(\chi)$ achieved by point mass contamination δ_0 at the origin.

2.3 Optimality of Jump Function χ's

We now suppose that F_0 has a density f_0. Consider the jump function type χ given by

$$\chi_a(x) = \begin{cases} 1 & |x| > a \\ 0 & |x| \leq a \end{cases} . \tag{2.11}$$

When χ in (1.1) is a jump function χ_a, the estimate s_n is a scaled order statistic:

$$s_n = \frac{1}{a} X_{(n-[nb])} .$$

Fix $b \in (\varepsilon, 1-\varepsilon)$, and let

$$\mathbf{C}_b = \{\chi : \chi \in \mathbf{C} \quad \text{and} \quad \int_0^\infty \chi(x) f_0(x)\, dx = b\} \tag{2.12}$$

The following lemma shows that if a is appropriately chosen, then for a large class of densities f_0, χ_a minimizes the maximal scale $s^+ = s^+(\chi)$ over $\mathbf{C}_b$, and simultaneously maximizes the minimal scale $s^- = s^-(\chi)$!

Lemma 2.3: Let a satisfy

$$1 - F_0(a) = b \tag{2.13}$$

and suppose that F_0 has a positive density f_0 with the property that $\frac{f_0(s\,x)}{f_0(x)}$ is decreasing in x for $s > 1$ and increasing in x for $s < 1$. Then for every $s \geq 1$

$$h_{\chi_a}(s) \leq h_\chi(s) \qquad \forall\ \chi \in \mathbf{C}_b . \tag{2.14}$$

and for every $s < 1$

$$h_{\chi_a}(s) \geq h_\chi(s) \qquad \forall\ \chi \in \mathbf{C}_b . \tag{2.15}$$

Proof: χ_a is a member of C_b by virtue of (2.13). For every $\chi \in C_b$ we have

$$\int_0^a \chi(x) f_0(x)\,dx = \int_a^\infty [1-\chi(x)] f_0(x)\,dx \quad .$$

Then for $s \geq 1$ we have

$$s\int_0^a \chi(x)\, f_0(sx)\,dx = s\int_0^a \chi(x) f_0(x)\,\frac{f_0(sx)}{f_0(x)}\,dx$$

$$\geq s\,\frac{f_0(sa)}{f_0(a)}\int_0^a \chi(x)\, f_0(x)\,dx$$

$$= s\,\frac{f_0(sa)}{f_0(a)}\int_a^\infty [1-\chi(x)] f_0(x)\,dx$$

$$\geq s\int_a^\infty [1-\chi(x)] f_0(sx)\,dx \; .$$

Then for every $\chi \in C_b$ and for every $s \geq 1$ we have

$$h_\chi(s) = s\int_0^\infty \chi(x) f_0(xs)\,dx \geq s\int_0^\infty \chi_a(x) f_0(sx)\,dx \; . = h_{\chi_a}(s) \; .$$

For $s < 1$, the above inequalities are reversed. □

<u>*Lemma 2.4:*</u> Suppose that $f_0(x)$ is non-increasing, differentiable, and that $-\log f_0(x)$ is convex. Then the hypotheses of Lemma 2.3 are satisfied.

<u>*REMARK 2.3:*</u> The exponential and positive normal distributions treated in Section 3 satisfy Lemma 2.4.

<u>*REMARK 2.4:*</u> The M-estimate of scale based on χ_a is a maximum-likelihood estimate (M.L.E.) of scale based on a density having the form

$$f^*(x) = \begin{cases} \tau & 0 < x \le a \\ \dfrac{\tau^3}{x^2} & x \ge a \end{cases} .$$

where $\tau > 0$ satisfies $\tau^3 + \tau^2 a^2 - a = 0$. Of course bounded χ's can only arise from M.L.E.'s for densities having tails at least as heavy as the Cauchy tails of f^*. While the distribution F^* corresponding to f^* is not a member of $\mathbf{F}$, there is no particular reason that it should be for the problem we are treating.

Set

$$\underline{a} = F_0^{-1}(\varepsilon) \tag{2.16}$$

and

$$\bar{a} = F_0^{-1}(1-\varepsilon) . \tag{2.17}$$

Then in view of (2.13) and (A2), the jump point a of the jump function χ_a must satisfy

$$\underline{a} < a < \bar{a} . \tag{2.18}$$

For $\chi = \chi_a$, let $s^+(a) = s^+(\chi_a)$ and $s^-(a) = s^-(\chi_a)$, and note that $h_{\chi_a}(s) = 1 - F_0(sa)$. Thus (2.9) and (2.10) give

$$s^+(a) = \frac{1}{a} F_0^{-1} \left[\frac{F_0(a)}{1-\varepsilon} \right] \tag{2.19}$$

$$s^-(a) = \frac{1}{a} F_0^{-1} \left[\frac{F_0(a)-\varepsilon}{1-\varepsilon} \right] . \tag{2.19'}$$

Note that as $a \uparrow \bar{a}$, $s^+(a) \to \infty$ and breakdown due to outliers occurs. On the other hand, when $a \downarrow \underline{a}$, $s^-(a) \to 0$ and breakdown due to inliers occurs. For $a \in (\underline{a}, \bar{a})$ we clearly have $0 < s^-(a) \le 1$ and $1 \le s^+(a) < \infty$.

For some distributions F_0, it may turn out that $s^+(a)$ and $s^-(a)$ are both monotone increasing functions of a (this is the case for example when F_0 is the exponential distribution treated in Section 3). In such cases we wish to choose a small in order to make the maximal scale $s^+(a)$ and the maximum positive bias $s^+(a)-1$ small. At the same time one wishes to choose a large to make the minimal scale $s^-(a)$ large, thereby making maximum negative bias $1-s^-(a)$ small. The precise choice of a will depend upon the loss function used to assess bias.

2.4 Min-Max Bias Results for Certain Loss Functions

Bias in scale estimation is a fundamentally asymmetric problem, and one may wish to consider any of a variety of loss functions $l = l(s^-, s^+)$. Most loss functions of interest will assign infinite loss to $s^+ = \infty$ or $s^- = 0$. Some possibilities are, for $0 < s^- \le 1 \le s^+ < \infty$:

(i) $l(s^-, s^+) = (s^-)^{-1} + s^+ - 2$

(ii) $l(s^-, s^+) = \log s^+ - \log s^-$

(iii) $l(s^-, s^+) = \max\{(s^-)^{-1}, s^+\}$

(iv) $l(s^-, s^+) = \max\{-\log s^-, \log s^+\}$

Since several considerations lead one to the logarithmic transformation when estimating scale, the choices (ii) and (iv) may be particularly appealing.

The above possibilities suggest consideration of two distinct classes of loss functions.

L_1: All continuous functions $l(u_1, u_2)$ on $(0, 1] \times [1, \infty)$ which are decreasing in u_1, increasing in u_2, and such that $\lim_{u_1 \to 0} l(u_1, u_2) = \infty$ for each real u_2, and $\lim_{u_2 \to \infty} l(u_1, u_2) = \infty$ for each real u_1.

L_2: All functions on $(0, 1] \times [1, \infty)$ of the form $l(u_1, u_2) = \max\{g_1(u_1), g_2(u_2)\}$ where g_1 and g_2 are continuous with

g_1 strictly monotone decreasing, g_2 strictly monotone increasing, and

$\lim_{u_1 \to 0} g_1(u_1) = \lim_{u_2 \to \infty} g_2(u_2) = \infty$.

While $\mathbf{L}_2$ is a special case of $\mathbf{L}_1$, $\mathbf{L}_2$ has enough additional structure to warrant separate treatment. Note that for l in $\mathbf{L}_1$ or $\mathbf{L}_2$, $l(s^-(\chi), s^+(\chi))$ represents the maximum loss, accounting for both positive and negative bias, due to use of a particular χ. Let

$$L(a) = l(s^-(a), s^+(a)) = l(s^-(\chi_a), s^+(\chi_a))$$

denote the value of the loss for a jump function χ_a.

Theorem 2.2: Suppose that $l \in \mathbf{L}_i$, $i = 1$, or $i = 2$. Then under the assumptions of Lemma 2.3, for $i = 1, 2$

$$a_0 = \underset{a \in (\underline{a}, \bar{a})}{argmin} L(a)$$

exists, and

$$L(a_0) = \min_{\chi \in \mathbf{C}} l(s^-(\chi), s^+(\chi)) .$$

Proof: Since F_0 is not only absolutely continuous, but also strictly increasing, F_0^{-1} exists and is continuous. Thus $s^-(a)$ and $s^+(a)$ are continuous functions of a. Furthermore, $s^-(a) \to 0$ as $a \downarrow \underline{a} = F_0^{-1}(\varepsilon)$ and $s^+(a) \to \infty$ as $a \uparrow \bar{a} = F_0^{-1}(1-\varepsilon)$. Thus $L(a) \to \infty$ as $a \downarrow \underline{a}$ or $a \uparrow \bar{a}$ and since $L(a)$ is continuous, its minimum a_0 is attained on a closed subset of $(\underline{a}, \bar{a})$. It follows from Lemma 2.3 and the properties of L that for each $b \in (\varepsilon, 1-\varepsilon)$, and correspondingly each $a \in (\underline{a}, \bar{a})$, $L(s^-(\chi), s^+(\chi))$ is minimized by the jump function χ_a, with a determined by (2.13). $\square$

Corollary 2.2: Let $l \in \mathbf{L}_2$, and let a_0 be the optimal a of the theorem.

(i) Either a_0 satisfies the equation

$$g_1(s^-(a_0)) = g_2(s^+(a_0)) \ . \tag{2.20}$$

or a_0 is a local minimum of at least one of the functions $s^-(a)$, $s^+(a)$.

(ii) If $s^-(a)$ and $s^+(a)$ are both strictly monotone increasing, then a_0 uniquely satisfies (2.20).

Proof: (i) Suppose that a_0 does not satisfy (2.20). Then there is a neighborhood $(a_0-\delta, a_0+\delta)$ such that either $l(a)=g_1(s^-(a))$ or $l(a)=g_2(s^+(a))$ on $(a_0-\delta, a_0+\delta)$. Then by definition of a_0 it must be a local minimum of either $s^-(a)$ or $s^+(a)$.

(ii) $g_1(s^-(a))$ is a strictly decreasing function of a, which tends to $+\infty$ as $a \downarrow \underline{a}$, while $g_2(s^+(a))$ is a monotone increasing function of a strictly tends to $+\infty$ as $a \uparrow \bar{a}$. Thus the equation $g_1(s^-(a)) = g_2(s^+(a))$ has a unique solution $a_0 \in (\underline{a}, \bar{a})$. □

2.5 Breakdown and Bias Optimality of the Median

For discussions of the breakdown point of scale estimates, including M-estimates, see Hampel et al (1986) and Huber (1981). Here we comment only on an M-estimate of scale based on jump functions χ_a with $a \in (\underline{a}, \bar{a})$.

Let ε^+ denote the supremum of all ε for which $s(\chi, y)$, given by (2.1) with $f \in \mathbf{F}_1$ is finite when $y = \infty$, and let ε^- denote the supremum of all ε for which this $s(\chi, y)$ is positive when $y = 0$. Then take the breakdown point (BP) of the estimate to be $BP = \min\{\varepsilon^-, \varepsilon^+\}$. For the case of a jump function χ_a, it follows from (2.19)–(2.19') that

$$\varepsilon^- = F_0(a) \tag{2.21}$$

$$\varepsilon^+ = 1 - F_0(a) \tag{2.21'}$$

and so

$$BP = \min\{F_0(a), 1 - F_0(a)\} \ .$$

Clearly $BP \leq 0.5$ and $BP = 0.5$ only if $F_0(a) = 0.5$. In this case $a = a_{med}$ is the median of F_0, (2.2) gives $b = b_{med} = 0.5$, and the solution of (2.1) is

$$s(a_{med}, F) = \frac{F^{-1}(\frac{1}{2})}{F_0^{-1}(\frac{1}{2})} \tag{2.23}$$

assuming $F^{-1}(\frac{1}{2})$ exists. In words: the only jump function type M-estimate of scale with breakdown point $\frac{1}{2}$ is the median functional, standardized by the median of F_0 to insure Fisher consistency.

Now a bias optimal estimate of the type given by Theorem 2.2 or Corollary 2.2 is of jump function type, with jump location a_0 in $(\underline{a}, \bar{a})$. For $\varepsilon < \frac{1}{2}$ we will seldom, if ever, have $BP = \frac{1}{2}$ for the optimal a_0. Note however that as $\varepsilon \uparrow \frac{1}{2}$, $\underline{a} = F_0^{-1}(\varepsilon) \rightarrow F_0^{-1}(\frac{1}{2})$ and $\bar{a} = F_0^{-1}(1-\varepsilon) \rightarrow a_0 = F_0^{-1}(\frac{1}{2})$. Thus any bias optimal estimate given by Theorem 2.2 or Corollary 2.2 tends to the median estimate (2.23) as $\varepsilon \rightarrow \frac{1}{2}$. *Thus the median, as an estimate of scale of a positive random variable with a contamination distribution, simultaneously enjoys two global robustness properties: a maximal breakdown point of one-half and min-max bias optimality.*

3. THE EXPONENTIAL AND POSITIVE NORMAL DISTRIBUTIONS

In this section the loss function $l(s^{-}, s^{+}) = \max\{\log s^{-}, \log s^{+}\}$ is used.

3.1 Exponential Distribution

Suppose that our target distribution is exponential with scale parameter s, that is

$$f(x,s) = \frac{1}{s} f_0\left(\frac{x}{s}\right), \qquad f_0(x) = e^{-x} \tag{3.1}$$

In this particular case $s^{+}(a)$ and $s^{-}(a)$ (See Section 2) reduce to

$$s^{+}(a) = -\frac{1}{a} \log \left[\frac{e^{-a} - \varepsilon}{1 - \varepsilon} \right] \tag{3.2}$$

$$s^{-}(a) = 1 + \frac{\log(1-\varepsilon)}{a} \tag{3.3}$$

for $a \in (\bar{a}, \underline{a})$.

We notice immediately that $s^{-}(a)$ is strictly increasing with $\lim_{a \to \bar{a}} s^{-}(a) = 1$, $\lim_{a \to \underline{a}} s^{-}(a) = 0$. It is also easy to check that the derivative of $s^{+}(a)$ is positive (see Section 5 for details), so that $s^{+}(a)$ is strictly increasing, with $\lim_{a \to \bar{a}} s^{+}(a) = \infty$ and $\lim_{a \to \underline{a}} s^{-}(a) = \log(1-2\varepsilon)/\log(1-\varepsilon)$. Therefore Corollary 2.2 is in force.

In Table 1a we show values of the optimal a_0, along with $\log s^{+}(a_0) = -\log {}^{-}(a_0)$, $1 - s^{-}(a_0)$, and $s^{+}(a_0) - 1$ for several values of ε. Breakdown points and efficiences at the nominal model are also given. We note that the breakdown points of the optimal estimators are fairly high (at least 0.46), even when they are designed for contamination models with small fractions of contamination!

For the sake of comparison with the optimal estimator, Table 1b displays corresponding values of $-\log s^{-}$, $\log s^{+}$, $1 - s^{-}$ and $s^{+} - 1$ for the scaled median estimator

Table 1 Min-Max and Median Estimates for Exponential Distribution

(a) Min-Max Estimate

ε	a_0	$\log s^+(a_0)$	$1-s^-(a_0)$	$s^+(a_0)-1$	Breakdown Point	Efficiency
0.10	0.716	0.159	0.147	0.173	0.49	0.49
0.20	0.739	0.359	0.302	0.432	0.48	0.50
0.30	0.761	0.632	0.469	0.882	0.47	0.51
0.40	0.773	1.080	0.661	1.946	0.46	0.51
0.45	0.763	1.530	0.783	3.618	0.47	0.48

(b) Scaled Median Estimate ($a = .693$)

ε	$-\log s^-$	$\log s^+$	$1-s^-$	s^+-1	Breakdown Point	Efficiency
0.10	0.165	0.157	0.152	0.170	0.50	0.48
0.20	0.389	0.347	0.322	0.415	0.50	0.48
0.30	0.723	0.592	0.515	0.807	0.50	0.48
0.40	1.335	0.950	0.737	1.585	0.50	0.48
0.45	1.984	1.241	0.862	2.459	0.50	0.48

$s_n = med/.693$. While the scaled median is actually less sensitive to outliers than the optimal estimate, it is sufficiently more sensitive to inliers that it is somewhat dominated by the optimal estimate using the loss $\max\{-\log s^-, \log s^+\}$. As one might expect, both estimates are more sensitive to outliers than inliers on the raw scale. Note also that the asymptotic efficiency of .48 for the scaled median at the nominal model is only slightly smaller than that of the min-max bias estimates displayed in Table 1a. In summary we can state: *the scaled median is a nearly optimal estimate of scale in terms of asymptotic bias.*

Table 2 displays similar results for the positive normal distribution, whose density is $f_{PN}(x) = 2\phi(x)$, $x \geq 0$, where $\phi(x)$ is the standard normal density. The results are rather similar to those for the exponential distribution. Again: *the scaled median (in this case* *med* / .6751) *is a rather good approximation to the optimal bias-robust estimate.*

Table 2 Min-Max and Median Estimates for Positive Normal Distribution

(a) Min-Max Estimate

ε	a_0	$\log s^+(a_0)$	$1-s^-(a_0)$	$s^+(a_0)-1$	Breakdown Point	Efficiency
0.10	0.700	0.127	0.120	0.136	0.48	0.38
0.20	0.726	0.284	0.247	0.329	0.47	0.40
0.30	0.750	0.494	0.390	0.639	0.45	0.41
0.40	0.762	0.844	0.570	1.325	0.45	0.42
0.45	0.746	1.236	0.710	2.441	0.46	0.41

(b) Scaled Median Estimate (a = .675)

ε	$-\log s^-$	$\log s^+$	$1-s^-$	s^+-1	Breakdown Point	Efficiency
0.10	0.135	0.126	0.126	0.134	0.50	0.37
0.20	0.322	0.274	0.276	0.315	0.50	0.37
0.30	0.612	0.459	0.458	0.583	0.50	0.37
0.40	1.166	0.718	0.688	1.050	0.50	0.37
0.45	1.780	0.918	0.831	1.505	0.50	0.37

4. MAXIMAL BIAS CURVES AND FINITE SAMPLE SIZE RELEVANCE

It should be noted that, min-max bias optimality results aside, the calculation of maximal bias curves (i.e., plots of maximal bias of a given estimate for each fraction ε of contamination, versus ε) for contamination models for any proposed robust estimate is a very worthwhile goal (see Sec. 2.7 of Hampel et al, 1986). It was a pleasant surprise to discover that for a wide range of estimation problems (in addition to estimation of scale), and types of estimate, one can usually obtain an analytic expression from which one can calculate the maximal bias curves.

For example, (2.9) and (2.10) of Theorem 2.1 lead to the maximal biases (2.5) and (2.6), for outliers and inliers, respectively. It is not always possible to express the inverse h_χ^{-1} in explicit analytic form (e.g., it is possible for the case of exponential F_0, but not for the positive normal F_0 in Section 3). Nonetheless, the expression (2.7) for $h_\chi(s)$ is nice enough (as are analogous expressions for other problems such as location, regression, etc.), that one can always solve numerically for $s^+(\chi)$ and $s^-(\chi)$ by computing $h_\chi(s)$ for a grid of s values, and interpolating to match with $\frac{b-\varepsilon}{1-\varepsilon}$ and $\frac{b}{1-\varepsilon}$, respectively.

We have done this for the following two estimators in the case of the positive normal F_0: (i) The median; and (ii) Huber's M-estimate of scale based on

$$\chi(t) = \psi_{H,\tau}^2(t) = \begin{cases} t^2, & 0 \le t \le \tau \\ t^2, & t > \tau \ . \end{cases}$$

The results are shown in Figure 1 for $\tau = 1.5$, which results in approximately 50% efficiency for the Huber estimate, at the nominal positive normal distribution (where the scaled median has 37% efficiency). The Huber estimate is much more sensitive to the outliers than the median, as reflected in the fact that $B_{hub}^+ > B_{med}^+$ for all ε. Furthermore, the Huber estimate has a breakdown point toward outliers of only .29, (whereas that of the scaled median is .5). On the other hand the Huber estimate is less sensitive than the median to

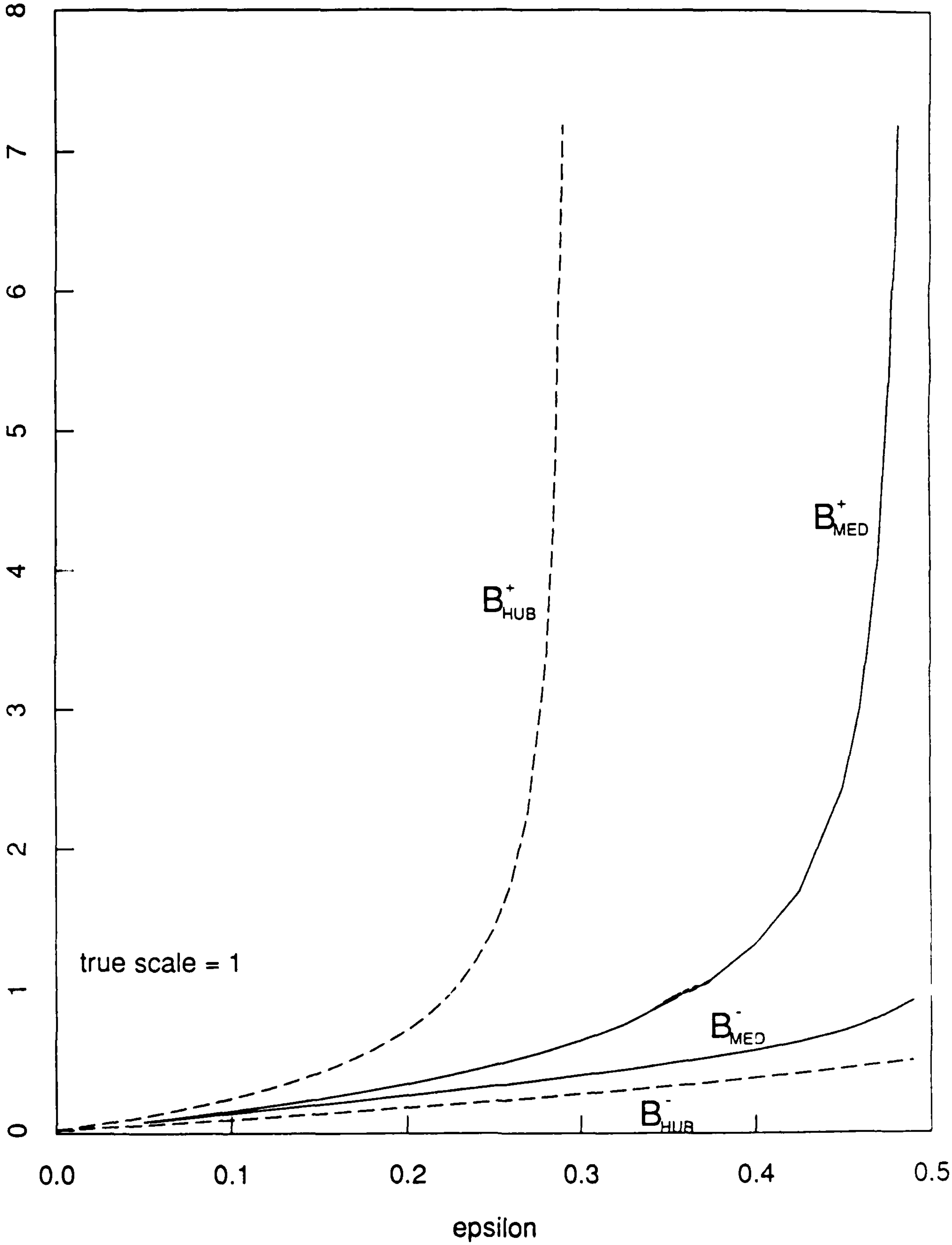


FIGURE 1. Maximal Biases B^{+}_{med} , B^{-}_{med} for Median, and B^{+}_{hub} , B^{-}_{hub} for Huber M-estimate of Scale. F_0 = positive normal distribution.

inliers, and accordingly $B_{hub}^{-} < B_{med}^{-}$ for all ε. Also, the breakdown point of the Huber estimate to inliers is .71 (as compared with .5 for the scaled median).

Finally, one should be concerned about the relevance of an asymptotic min-max bias theory for finite sample situations. Here one is naturally interested in the sample size $n(\varepsilon)$ at which squared bias is equal to the variance. One can get a rough idea of what $n(\varepsilon)$ will be by acting as if the asymptotic bias and variance projects well back to finite sample sizes.

For example, consider the positive bias $B^{+}(\varepsilon) = s^{+}(\varepsilon) - 1$ of the mix-max scale estimate for the exponential F_0 (see Section 3). This bias is achieved by point mass contamination δ_∞ at infinity. The corresponding asymptotic variance is easily shown to be

$$V^{+}(\varepsilon) = \frac{b(a_0)[1-b(a_0)]}{(1-\varepsilon)^2 a_0^2 \varepsilon^{-2s^{+}a_0}}$$

where $s^{+} = s^{+}(\varepsilon)$, $a_0 = a_0(\varepsilon)$. Some values of $n^{+}(\varepsilon) = [V^{+}(\varepsilon)]^{\frac{1}{2}} / B^{+}(\varepsilon)$ are shown in Table 3.

Table 3: Sample Size $n(\varepsilon) = [V^{+}(\varepsilon)]^{\frac{1}{2}}/B^{+}(\varepsilon)$

for Matching Variance and Squared Bias

(Min-Max Scale Estimate for Exponential F_0)

ε	.02	.05	.1	.15	.2
$n^{+}(\varepsilon)$	50	20	10	7	6

Evidently, the asymptotic theory is relevant down to quite small sample sizes unless the fraction of contamination is exceedingly small. More precise finite-sample size calculations should be carried out to confirm this.

5. SCALE FOR SYMMETRIC DISTRIBUTIONS WITH NUISANCE LOCATION PARAMETER

Suppose now that our observations X_i are not restricted to being positive, and have a location-scale family distribution $F_{\mu,s}(x) = F\left[\frac{x-\mu}{s}\right]$. We now assume that

$$\mathbf{F} = \{ F : F = (1-\varepsilon)F_0 + \varepsilon H, F_0 \text{ symmetric} \} \tag{5.0}$$

and H is again any distribution function. One now has to estimate the nuisance location parameter μ, and so let $\hat{\mu}_n$ be any location and scale equivariant estimate of location. Then the M-estimate of scale $\hat{s}_n$ is given by

$$\hat{s}_n = \inf \left\{ s : \sum_1^n \chi\left[\frac{|X_i - \hat{\mu}_n|}{s}\right] \le b \right\} \tag{5.1}$$

where χ is as in Section 2. The corresponding asymptotic functional is

$$s(\chi, F) = \inf \left\{ s : \int_{-\infty}^{\infty} \chi\left[\frac{|X - \mu(F)|}{s}\right] dF(x) \le b \ . \right\} \tag{5.2}$$

Here $\hat{\mu}_n$ is assumed to have an asymptotic functional representation $\mu(F)$, and b is given by (2.2). Existence of a unique solution $s(\chi, F) \in (0, \infty)$ follows under (A1)–(A3) and finiteness of $\mu(F)$, just as in Lemma 2.1. Without loss of generality we take $\mu(F_0) = 0$ and $s(\chi, F_0) = 1$. Then $\mu(F)$ is the bias in estimating μ and $s(\chi, F) - 1$ is the bias in estimating s.

Let X be a random variable with distribution F_X, and let F_{-X} denote the distribution of $-X$. Then most $\mu(F)$ of interest will satisfy the symmetry property

(L1) $\quad \mu(F_{-X}) = -\mu(F_X)$.

Suppose that $s(\chi, F_X)$ is given by (5.2), with $x \sim F_X$ resulting in $\mu(F_X) < 0$. Then in view of (5.0) and (L1), there exists F_{-X} in $\mathbf{F}$ with $\mu(F_{-X}) = -\mu(F_X) > 0$ and $s(\chi, F_{-X}) = s(\chi, F_X)$. Thus, it suffices under L1 to restrict attention to the sub-family

$$\mathbf{F}^+ = \{F \in \mathbf{F} : \mu(F) \geq 0\} .$$

In addition we will use the assumption

(L2) For all $F \in \mathbf{F}^+$, $\mu(F) \leq \bar{\mu} < \infty$ where $\bar{\mu} = \mu((1-\varepsilon)F_0 + \varepsilon\delta_\infty)$.

REMARK 5.1: Any location functional $\mu(F)$ with breakdown point greater than ε, such that $\mu(\delta_\infty) = \sup_{F \in \mathbf{F}^+} \mu(F)$ will satisfy (L2). For example location M-estimates based on bounded, monotone psi-functions will satisfy (L2).

Let

$$h(s,t) = \int_{-\infty}^{\infty} \chi\left[\frac{|x-t|}{s}\right] f_0(x)\,dx . \tag{5.3}$$

By change of variable we have

$$h(s,t) = s\int_0^{\infty} \chi(x)\ \{f_0(xs-t) + f_0(xs+t)\}\,dx . \tag{5.4}$$

Lemma 5.1: Let χ satisfy (A1). In addition suppose that the density f_0 is not only symmetric, but also decreasing on $[0,\infty)$. Then for each fixed $s > 0$, $h(s,t)$ is increasing in t on $[0,\infty)$.

Let

$$\lambda_\mu(s,F) = (1-\varepsilon)h(s,\mu(F)) + \varepsilon\int\chi\left[\frac{x-\mu(F)}{s}\right]dH(x) \tag{5.5}$$

and let $s_\mu^+(\chi)$ and $s_\mu^-(\chi)$ denote the largest and smallest values of $s(\chi,F)$, as F

ranges over $\mathbf{F}^+$. Here the subscript μ denotes dependence on the form of the location functional $\mu = \mu(\cdot)$. $\lambda_\mu(s, F)$ is a decreasing function of s, and so $s_\mu^+(\chi)$ is achieved by a possibly substochastic contamination distribution H which makes $\lambda_\mu(s, F)$ as large as possible for $s > 1$, and $s_\mu^-(\chi)$ is achieved by an H which makes $\lambda_\mu(s, F)$ as small as possible for $s < 1$. Even though $\mu(F)$ is now involved, it follows from Lemma 5.1 that point mass contamination at infinity achieves $s_\mu^+(\chi)$, and point mass contamination at the origin achieves $s_\mu^+(\chi)$.

Let $h_{\bar{\mu}}^{-1}(t)$ denote the inverse of $h(s, \bar{\mu})$ with respect to the first argument, with the second argument fixed at $\bar{\mu} = \mu((1-\varepsilon)F_0 + \varepsilon\delta_\infty)$, and define $h_0^{-1}(t)$ similarly in terms of $h(s, 0)$.

Theorem 5.1: Suppose that (A1)–(A3) and (L1)–(L2) hold. Then

$$s_\mu^+(\chi) = h_{\bar{\mu}}^{-1}\left[\frac{b-\varepsilon}{1-\varepsilon}\right] \tag{5.7}$$

$$s_\mu^-(\chi) = h_0^{-1}\left[\frac{b}{1-\varepsilon}\right] \tag{5.8}$$

REMARK 5.2: The maximal scale $s_\mu^+(\chi)$ depends on the location functional $\mu = \mu(\cdot)$ through $\bar{\mu}$. On the other hand the minimal scale $s^-(\chi)$ is independent of the choice of the location functional $\mu(F)$ which is why we denote it simply $s^-(\chi)$.

Since the median minimizes the maximum bias $\bar{\mu}$ of any equivariant location estimate over $\mathbf{F}^+$ (Huber, 1964), Lemma 5.1 gives the following important corollary.

Corollary 5.1: The median minimizes $s_\mu^+(\chi)$ over all location functionals $\mu = \mu(\cdot)$ which satisfy L1 and L2.

The results of Section 2.4 hinge on the key Lemma 2.3, the proof of which relies on the hypothesis that $k_s(x) = f_0(sx)/f_0(x)$ is decreasing for $s > 1$, and increasing for $s < 1$. In the present context we still have $k_s(x) = f_0(sx)/f_0(x)$ for $s < 1$, but for $s > 1$ $k_s(x)$ is replaced by

$$k_{s,\bar{\mu}_{med}}(x) = \frac{f_0(sx - \bar{\mu}_{med}) + f_0(sx + \bar{\mu}_{med})}{f_0(x)} \tag{5.9}$$

where $\bar{\mu}_{med}$ is now the maximal bias of the median. Unfortunately, $k_{s,\bar{\mu}_{med}}(x)$ is not decreasing for all s and all values of $\bar{\mu}_{med}$, and this complicates matters somewhat. However, we can proceed as follows.

First, note that the same type of argument used in the proof of Theorem 2.2 shows that there is an optimal jump function χ_{a_*} among the class of all jump functions χ_a. That is, for $l \in \mathbf{L}_i$, $i = 1$ or $i = 2$, there is an $a_* \in (0, \infty)$ such that

$$L(a_*) \le l(s^-(\chi_a), s^+(\chi_a)), \qquad a \in (0,\infty) . \tag{5.10}$$

Let $s_*^+ = s^+(a_*)$, $s_*^- = s^-(a_*)$, and set

$$k_\varepsilon(x) = \frac{f_0(xs_*^+ - \bar{\mu}_{med}) + f_0(xs_*^+ + \bar{\mu}_{med})}{f_0(x)} \tag{5.11}$$

where the subscript now indicates the dependence on ε through s_*^+ and $\bar{\mu}_{med}$, each of which depend on ε. Note that the conclusion of Lemma 2.3 holds when $k_\varepsilon(x)$ is decreasing (just replace $f_0(sx)/f_0(x)$ by $k_\varepsilon(x)$ in the proof). In applications there will always be an interval $[0, \varepsilon_0)$ for which $k_\varepsilon(x)$ is decreasing for $\varepsilon \in [0, \varepsilon_0)$. The normal distribution provides a concrete example of this, as we see shortly.

The next theorem gives conditions under which the optimality χ_{a_*} over the class of all jump functions extends to $\mathbf{C}$.

Theorem 5.2: Let a_* be as in (5.10), so that χ_{a_*} is optimal over all jump functions. Suppose that $k_\varepsilon(x)$ is decreasing and let $l \in \mathrm{L}_2$. If $s^+(a)$, $s^-(a)$ are both strictly increasing, or if the derivatives of $s^+(a)$, $s^-(a)$ are non-vanishing at a_*, then

$$L(a_*) = \min_{\chi \in \mathrm{C}} l\,(s^-(\chi), s^+(\chi)) \ .$$

Proof: Choose any $\chi \in \mathrm{C}$, and fix b corresponding to this χ, namely $b = E_{F_0}\chi(X)$. Let χ_a be the unique jump function which corresponds to this particular b. Let $h_\chi(s_*^+) = h\,(s_*^+, \bar{\mu}_{med})$ for the given χ. If $h_\chi(s_*^+) \geq \frac{b-\varepsilon}{1-\varepsilon}$, then from (5.7) for both χ and χ_{a_*}, and the monotonicity of $h\,(\cdot, \bar{\mu}_{med})$, χ is not better than χ_{a_*}, i.e., $s^+(\chi) \geq s^+(\chi_{a_*})$. So, suppose that $h_\chi(s_*^+) < \frac{b-\varepsilon}{1-\varepsilon}$, i.e., suppose χ is strictly better than χ_{a_*} for s^+, i.e., $s^+(\chi) < s_*^+ = s^+(a_*)$. We will show that then χ must be worse than χ_a for s^-, and hence worse than χ_{a_*} for s^-. Since $k_\varepsilon(x)$ is decreasing, the conclusion of Lemma 2.3 applies *at the point* $s = s_*^+$. Thus, the jump function χ_a is optimal *at the point* $s = s_*^+$, and so $s^+(a) \leq s^+(\chi) < s_*^+$. In view of Corollary 2.2, and the optimality of χ_{a_*} we have

$$g_1(s_*^-) = g_2(s_*^+) \leq \max\{g_1(s^-(a)), g_2(s^+(a))\}, \quad a \in (0, \infty)$$

and it follows that $s^-(a) \leq s_*^-$. For the particular b in question the jump function χ_a is optimal for all $s < 1$, and so we have

$$s^-(\chi) \leq s^-(a) \leq s_*^- \ .$$

Thus χ_{a_*} is as good as any $\chi \in \mathrm{C}$. $\square$

The Normal Distribution

We now apply this general result to the case of the normal distribution and the loss function $l(s^-, s^+) = \max\{-\log s^-, \log s^+\}$.

The optimal value of a_* is obtained for each ε by minimizing $\max\{-\log s^-(a), \log s^+(a)\}$ for $a \in (\underline{a}, \bar{a})$, where $s^+(a)$ is given by (5.7) with $\chi = \chi_a$ and $\bar{\mu} = \bar{\mu}_{med}$, and $s^-(a)$ is given by (5.8). This cannot be done analytically, since $h(\cdot, t)$ given by (5.4) does not admit a closed form inverse when f_0 is normal. We determined $a_* = a_*(\varepsilon)$ by numerical search for $\varepsilon = .05, .5(.05)$, evaluated the corresponding values $s^+(a_*), s^-(a_*)$ and $\bar{\mu}_{med} = \bar{\mu}_{med}(\varepsilon)$. These are displayed in Table 5.1a for $\varepsilon = .05, .35(.05)$. With the help of Lemma 6.1 (see end of Section 6) we have numerically verified that the conditions of Theorem 5.2 hold, and therefore χ_{a_*} is optimal for these value of ε.

Based on further numerical checks, we conjecture that χ_{a_*} is optimal for all $\varepsilon \in (0, .35)$. On the other hand a calculation showed that $k_\varepsilon(x)$ is not monotone for $\varepsilon = .4$, and thus we cannot conclude that χ_{a_*} is optimal for all values of $\varepsilon \in (0, .5)$, at least not by the current method of proof. The question of optimality of χ_{a_*} for the full range of ε is $(0, .5)$ remains an open problem.

The min-max bias values $\log s^+(a_*) = -\log s^-(a_*)$ are shown in Table 4a, as are the negative and positive biases $1 - s^-(a_*)$ and $s^+(a_*) - 1$ on the raw scale. Breakdown points and efficiences are also displayed.

As in the case of the solutions for exponential and positive normal distributions in Section 3, we find that a_* is quite insensitive to the value of ε, and not too different from those for the positive normal distribution. However, the actual min-max biases $\log s^+(a_*)$ (and also the corresponding row biases) are substantially higher, as one might expect, due to the nontrivial bias of the median as centering functional.

Table 4b displays similar quantities for the scaled median-absolute deviation about the median (MADM) estimate. The comparison of Tables 4a and 4b shows that even though the scaled MADM is not optimal for the given loss function and values of ε, it is strikingly close to optimal for the range of ε values considered (the performance of the scaled MADM is much nearer to that of the optimal estimate than was that of the scaled median for the positive normal distribution treated in Section 3.2). Except for $\varepsilon = .35$, the common entries of Tables 4a and 4b agree to at least two decimal places, and at $\varepsilon = .35$ the differences are only about .01. The breakdown point of the min-max estimates round to .500 for all ε values shown. In effect a single Table would suffice, and we are led to assert: *For all practical purposes the scaled MADM is the min-max bias estimate for* $\varepsilon = .05, .35\,(.05)$, *(and probably for all* $\varepsilon \in (0, .35)$*).*

Table 4 Normal Distribution with Location Unknown

(a) Min-Max Estimate

ε	$\bar{\mu}_{med}$	a_*	$\log s^+(a_*)$	$1-s^-(a_*)$	$s^+(a_*)-1$	Breakdown Point	EFF
0.05	0.066	0.674	0.063	0.061	0.065	0.50	0.37
0.10	0.139	0.673	0.135	0.126	0.145	0.50	0.37
0.15	0.223	0.673	0.221	0.198	0.247	0.50	0.37
0.20	0.318	0.673	0.324	0.276	0.382	0.50	0.37
0.25	0.430	0.673	0.450	0.362	0.568	0.50	0.37
0.30	0.566	0.676	0.609	0.456	0.839	0.50	0.37
0.35	0.736	0.682	0.812	0.556	1.252	0.50	0.37

(b) Scaled MADM Estimate (a = .675)

ε	$-\log s^-$	$\log s^+$	$1-s^-$	s^+-1	Breakdown Point	Efficiency
0.05	0.063	0.063	0.061	0.065	0.50	0.37
0.10	0.135	0.135	0.126	0.145	0.50	0.37
0.15	0.220	0.221	0.197	0.247	0.50	0.37
0.20	0.322	0.324	0.276	0.383	0.50	0.37
0.25	0.449	0.450	0.362	0.569	0.50	0.37
0.30	0.612	0.608	0.458	0.838	0.50	0.37
0.35	0.833	0.808	0.565	1.243	0.50	0.37

6. PROOFS OF RESULTS

Proof of Lemma 2.1

Proof: For $F \in \mathbf{F}$, we have

$$\lambda(s) = \int \chi(\frac{x}{s})\, dF(x) = (1-\varepsilon)\int \chi(\frac{x}{s}) dF_0(x) + \varepsilon \int \chi(\frac{x}{s})\, dH(x)$$

Since for any distribution function H, including the point mass at infinity, we have $\lim_{s \to \infty} \lambda(s) \le \varepsilon < b$, it suffices to show that $\lim_{s \to 0^+} \lambda(s) > b$. Since $\lambda(s) \ge (1-\varepsilon) \int_0^\infty \chi(\frac{x}{s}) dF_0(x)$ for any H, with equality when H is concentrated at the origin,

$$\lim_{s \to 0^+} \lambda(s) \ge (1-\varepsilon) \lim_{s \to 0^+} \int_0^\infty \chi(\frac{x}{s}) dF_0(x) = (1-\varepsilon) > b .$$

Thus $s(F, \chi) \in (0, \infty)$ for all H. □

Proof of Lemma 2.2

Proof: We'll show that for all $F \in \mathbf{F}$ there exist $F_1, F_2, \in \mathbf{F}_1$ such that $s(\chi, F_1) \le s(\chi, F) \le s(\chi, F_2)$. Let $\chi \in \mathbf{C}$ and $s = s(\chi, F)$ be defined as in (2.1), with $F = (1-\varepsilon) F_0 + \varepsilon H$.

If $E_H\{\chi(X/\bar{s})\} = 1$ for some $\bar{s} < s$ then $\chi(x/s) = 1$ a.s. $[H]$ and there exists $y \in \mathbf{R}$ such that $\chi(y/s) = 1$. Let $F_2 = (1-G) F_0 + \varepsilon \delta y$; for all $\bar{s} \le s$ we have

$$E_{F_2}\{\chi(X/\bar{s})\} = (1-\varepsilon) h_\chi(\bar{s}) + \varepsilon = E_F\{\chi(X/\bar{s})\} > b ,$$

therefore $s(F_2, \chi) \ge s$. If $E_H\{\chi(X/\bar{s})\} < 1$ for some $0 < \bar{s} < s$ (and then for all $\bar{s} \le s' \le s$), let $y \in \mathbf{R}$ be such that $\chi(y/s) > E_H\{\chi(X/\bar{s})\}$ and let F_2 be as defined above;

for some $\bar{s} < s' \leq s$, by continuity of h_χ we have

$$E_{F_2}\{\chi(X/s)\} = (1-\varepsilon)h_\chi(s) + \varepsilon\chi(y/s) > (1-\varepsilon)h_\chi(s') + \varepsilon E_H\{\chi(X/s')\} > b$$

Therefore $s(x, F_2) \geq s$.

On the other hand, let $F_1 = (1-\varepsilon)F_0 + \varepsilon\delta_0$. For all $\bar{s} > s$ we have

$$E_{F_1}\{\chi(X/\bar{s})\} = (1-\varepsilon)h_\chi(\bar{s}) \leq (1-\varepsilon)h_\chi(\bar{s}) + \varepsilon E_H\{\chi(X/\bar{s})\} = E_F\{\chi(X/\bar{s})\} \leq b,$$

hence $s(x, F_1) = \inf\{\bar{s} : E_{F_1}\{\chi(X/\bar{s})\}\} \leq s$ • $\square$

Proof of Theorem 2.1

For any y we have, by definition of s_y

$$s < s_y \Rightarrow (1-\varepsilon)h(s) + \varepsilon - b \geq (1-\varepsilon)h(s) + \varepsilon\chi(\frac{y}{s}) - b \geq 0$$

Thus, by continuity of $h(s)$ we have

$$h(s_y) \geq \frac{b-\varepsilon}{1-\varepsilon}. \tag{6.1}$$

On the other hand, again by definition of s_y, we have

$$(1-\varepsilon)h(s_y) \leq (1-\varepsilon)h(s_y) + \varepsilon\chi(\frac{y}{s_y}) \leq b \tag{6.2}$$

from (6.1) and (6.2) follows

$$\frac{b-\varepsilon}{1-\varepsilon} \leq h(s_y) \leq \frac{b}{1-\varepsilon}.$$

or

$$h^{-1}\left(\frac{b}{1-\varepsilon}\right) \le s_y \le h^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}\right).$$

For $y=0$, the left-hand inequality becomes an equality. Now note that s_y satisfies the monotonicity property $s_{y_1} \ge s_{y_2}$ when $y_1 \ge y_2$. Thus if y_n is a sequence of values tending to infinity, the limit $\bar{s} = \lim_{n\to\infty} s_{y_n}$ exists. We claim that for $s_y = \bar{s}$, the right hand inequality in (2.11) becomes an equality. For suppose

$$\bar{s} < h^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}\right).$$

Then

$$h(\bar{s}) > \frac{b-\varepsilon}{1-\varepsilon}$$

and we can choose $\delta > 0$ and y_0 sufficiently large that

$$(1-\varepsilon)h(\bar{s}) + \varepsilon\chi\left[\frac{y_0}{\bar{s}}\right] - b \ge (1-\varepsilon)h(\bar{s}) + \varepsilon(1-\delta) - b > 0.$$

Contradicting $s_y \le \bar{s}$. □

Proof of Lemma 2.4

Let $s > 1$. For all $x \ge 0$ we have

$$\frac{d}{dx}\left[\frac{f_0(sx)}{f_0(x)}\right] \le 0$$

if and only if

$$s\left[-\frac{f_0'(sx)}{f_0(sx)}\right] \le -\frac{f_0'(x)}{f_0(x)}.$$

The last inequality follows since $-f_0'(x)/f_0(x)$ is non-negative and increasing. For $s < 1$

all the reverse inequalities hold. □

Proof that $\frac{d}{da} s^+(a) \geq 0$

$$\frac{d}{da} s^+(a) = \left[\frac{a\, e^{-a}}{e^{-a} - \varepsilon} + \log\left(\frac{e^{-1} - \varepsilon}{1 - \varepsilon}\right) \right] / a^2 \geq 0$$

if and only if

$$\log(a - \varepsilon) \leq \gamma(a) \overset{\Delta}{=} \frac{1\, e^{-a}}{e^{-a} - \varepsilon} + \log(e^{-a} - \varepsilon)$$

The last inequality follows because $\gamma(0) = \log(1 - \varepsilon)$ and

$$\frac{d}{da} \gamma(a) = a \frac{\varepsilon^{-a}}{e^{-a} - \varepsilon} \left[\frac{e^{-a}}{e^{-a} - \varepsilon} - 1 \right] \geq 0$$

Proof of Lemma 5.1

Proof: It suffices to consider the case $s = 1$. Extend the domain of χ to $\mathbf{R}$ by setting $\chi(-t) = \chi(t), t \leq 0$. Set $\Delta = \frac{t_1 + t_2}{2}, 0 \leq t_1 \leq t_2 < \infty$. Then

$$\begin{aligned}
&\int_{-\infty}^{\infty} [\chi(x - t_2) - \chi(x - t_1)] f_0(x)\, dx \\
&= \int_{-\infty}^{\Delta} [\chi(x - t_2) - \chi(x - t_1)] f_0(x)\, dx + \int_{\Delta}^{\infty} [\chi(x - t_2) - \chi(x - t_1)] f_0(x)\, dx \\
&= \int_{\Delta}^{\infty} [\chi(t_1 - y) - \chi(t_2 - y)] f_0(t_1 + t_2 - y)\, dy + \int_{\Delta}^{\infty} [\chi(y - t_2) - \chi(y - t_1)] f_0(y)\, dy \\
&= \int_{\Delta}^{\infty} [\chi(y - t_2) - \chi(y - t_1)] [f_0(y) - f_0(y - 2\Delta)]\, dy \\
&\geq 0 .
\end{aligned}$$

□

Proof of Theorem 5.1

Proof: By (L1) and the Lemma 5.1 we have for all $s > 0$

$$h(s, \mu(F)) \leq h(s, \bar{\mu}) \qquad \forall\, F \in \mathbf{F}^+$$

and so

$$\lambda_\mu(s, F) \leq (1-\varepsilon) h(s, \bar{\mu}) + \varepsilon$$

with the right-hand side achieved when $H = \delta_\infty$, the point mass at infinity. Equating the right-hand side to b yields 5.7. On the other hand for all $s > 0$ we also have

$$\lambda(s, F) \geq (1-\varepsilon) h(s, 0) + \quad F \in \mathbf{F}^+$$

with the right-hand side achieved when $H = \delta_0$, the point mass at the origin. Now, equating the right-hand side to b gives (5.8). □

Statement and Proof of Lemma 6.1

The following notation is needed for Lemma 6.1, below:

Let $t = su$ and set

$$A(s, t, b_1, b_2) = (s^2 - 1)[1 + e^{-2tb_2} - 2t\, b_2 e^{-tb_1}] - 2t^2 e^{-tb_1}$$

Let

$$A_0(s,t) = A(s, t, 0, \tfrac{1}{8})$$

$$A_1(s,t) = A(s, t, \tfrac{1}{8}, \tfrac{1}{4},)$$

$$A_2(s,t) = A(s, t, \tfrac{1}{4}, \tfrac{1}{2},)$$

$$A_3(s,t) = A(s, t, \tfrac{1}{2}, \tfrac{3}{4},)$$

$$A_4(s,t) = A\,(s,t,\tfrac{3}{4},1)$$

Lemma 6.1: Suppose that

$$s\,(s-u) > 1$$

and

$$\min_{0 \le j \le 4} A_j(s,t) > 0$$

Then

$$h\,(x,s,u) = \frac{\phi(sx+u)+\phi(sx-u)}{\phi(x)}$$

is a decreasing function of x .

Proof:

$$h\,(x,s,u) = e^{-\frac{1}{2}((s^2-1)x^2+u^2)}(e^{sxu}+e^{-sxu})$$

and

$$\frac{\partial}{\partial x} h\,(x,s,u) = e^{-\frac{1}{2}((s^2-1)+u^2)}[-x\,(s^2-1)(e^{-sxu}+e^{sxu}) - su\,e^{-sxu} + su\,e^{sxu}]$$

Then

$$\frac{\partial}{\partial x} h\,(x,x,u) \le 0 \iff su \le x\,(s^2-1)(1+e^{-2sux}) + su\,e^{-2sux} \tag{6.1}$$

The last inequality holds for $x \ge 1$ since, by hypothesis

$$x\,(s^2-1)(1+e^{-2sux}) + su\,e^{-2sux} \ge (s^2-1) \ge su\ .$$

If $x = 0$ then equality holds. Therefore, we only need to prove that right-hand side of (6.1) holds when $0 < x < 1$. Set $\delta(x) = x\,(s^2-1)(1+e^{-2tx}) + t\,e^{-2tx}$;

$$\delta'(x) = (s^2 - 1)[\,1 + e^{-2tx} - x\,2t\,e^{-2tx}\,] - 2t^2 e^{-2tx}$$

$$\delta'(x) \geq \min_{0 \leq j \leq 4} A_j > 0$$

and the lemma follows. $\square$

References

Bickel, P.J. (1984). "Robust regression based on infinitestimal neighborhoods." *Ann. Statist.* **12,** 1347–1368.

Donoho, D.L. (1982). "Breakdown properties of multivariate location estimators." Unpublished manuscript, Harvard University Ph.D. qualifying paper.

Donoho, D.L. and Huber, P.J. (1983). "The notion of breakdown point." In *Festschrift fur Erich L. Lehman,* edited by P.J. Bickel, K. Doksum and J.L. Hodges. Wadsworth, Belmont.

Donoho, D.L., and Liu, R.C. (1986). "The automatic robustness of minimum distance functionals." Technical Rept., Dept. of Statistics, University of California.

Hampel, F.R. (1971). "A general qualitative definition of robustness." *Annals Math. Stat.* **42,** 1887–1896.

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

Huber, P.J. (1964). "Robust estimation of a location parameter." *Annals Math. Stat.* **35,** 73-101.

Huber, P.J. (1981). *Robust Statistics.* Wiley.

Jaeckel, L.A. (1971). "Robust estimates of location: Symmetry and asymmetric contamination," *Ann. Math. Statist.* **43** 1449–1458

Rousseeuw, P.J. (1984). "Least median of squares regression," *J. Ann. Statist. Assoc.* **79,** 871–880.

Rousseeuw, P.J. and Yohai, V.J. (1984). "Robust regression by means of S-estimators." In *Robust and Nonlinear Time Series Analysis.* J. Franke, W. Härdle and R.D. Martin (eds.). Lecture Notes in Statistics 26, 256–272. Springer.

Yohai, V.J. (1986). "High breakdown point and high efficiency robust estimates for regression." To appear in *Annals of Statistics.*

Yohai, V.J. and Zamar, R. (1986). "High breakdown-point estimates of Regression by means of the minimization of an efficient scale." Submitted to *J. Ann. Statist. Assoc.*

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 72	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER A179255
4. TITLE (and Subtitle) Min-max Bias Robust M-estimates of Scale		5. TYPE OF REPORT & PERIOD COVERED TR 12/1/86-11/30/87
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. Douglas Martin & Ruben H. Zamar		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0169
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, GN-2 University of Washington Seattle, Washington 98195		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-611-003
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N63374 1107 NE 45th St Seattle, WA 98105		12. REPORT DATE December 1986
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robustness, M-estimates of Scale, Min-Max Bias, Maximal Bias, Breakdown Point.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END
5_87
DTIC