The Annals of Statistics
1993, Vol. 21, No. 2, 991–1017

BIAS ROBUST ESTIMATION OF SCALE[1]

BY R. D. MARTIN AND RUBEN H. ZAMAR

University of Washington and University of British Columbia

In this paper we consider the problem of robust estimation of the scale of the location residuals when the underlying distribution of the data belongs to a contamination neighborhood of a parametric location-scale family. We define the class of M-estimates of scale with general location, and show that under certain regularity assumptions, these scale estimates converge to their asymptotic functionals uniformly with respect to the underlying distribution, and with respect to the M-estimate defining score function χ. We establish expressions for the maximum asymptotic bias of M-estimates of scale over the contamination neighborhood as a function of the fraction of contamination. Using these expressions we construct asymptotically *min-max bias robust* estimates of scale. In particular, we show that a scaled version of the Madm (median of absolute residuals about the median) is approximately min-max bias-robust within the class of Huber's Proposal 2 joint estimates of location and scale. We also 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 min-max bias robust in this class. Finally, we present the results of a Monte Carlo study showing that the Shorth has attractive finite sample size mean squared 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., ε-contamination or Levy neighborhood) of a parametric model. In this context one tries to obtain estimates which optimize some appealing criterion, for example, minimize the maximum asymptotic variance over a given neighborhood. Huber (1964) is the earliest example of this approach, with 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 ε-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 ε-contamination neighborhoods of a nominal Gaussian model: Among all translation equivariant estimates, the median minimizes the maximum asymptotic bias over such

Received July 1989; revised June 1992.
[1]Research partially supported by the Office of Naval Research, contracts N00014-88-K-0265 and N00014-91-J-1074, and also a Canadian NSERC grant.
AMS 1991 *subject classifications*. Primary 62F35, secondary 62G35.
Key words and phrases. Bias robustness, M-estimates, scale, location, Madm, Shorth.

neighborhoods. The relevance of this result seems considerable in view of the needed realism of allowing asymmetric contamination.

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

In this paper we obtain minimax bias robust estimates of scale for contamination models with a nominal distribution which is symmetric about an unknown location parameter. More precisely, we assume the following:

Assumption 0. F_0 is a specified distribution function with an even and unimodal density f_0.

The distribution F for independent and identically distributed observations $X_1, \dots, X_n$ belongs to the ε-contaminated family

$$\mathscr{F}_\varepsilon = \left\{ F(x) \colon F(x) = (1-\varepsilon) F_0\left(\frac{x-\mu_0}{s_0}\right) + \varepsilon H(x), \quad x \in R,\ \varepsilon \text{ fixed in } (0, 0.5) \right\}, \tag{1}$$

where μ_0 is the unknown location parameter, s_0 is the unknown scale parameter and H is an arbitrary (and unspecified) distribution.

The first step in obtaining a minimax estimate is to derive the maximal asymptotic bias $B_T(\varepsilon)$ of an estimate T over the family $\mathscr{F}_\varepsilon$. From $B_T(\varepsilon)$ one may construct a *maximum bias curve*, namely a plot of $B_T(\varepsilon)$ versus ε. The maximum bias curve includes the *gross error sensitivity* GES_T, namely the slope of $B_T(\varepsilon)$ at $\varepsilon = 0$, and also the *breakdown-point* ε_T^*, which is the location of the singularity where $B_T(\varepsilon)$ goes to infinity. While the two-number summary consisting of the GES_T and ε_T^* provides considerable information, one naturally would like to have the entire curve $B_T(\varepsilon)$ if possible. Not only do such curves allow one to check the range of accuracy of the GES_T as a linear approximation, they may also lead to different preference ordering of competing estimates that one might make on the basis of GES_T and ε_T^* [e.g., see Section 5 and also Martin, Yohai and Zamar (1989), who find min-max bias robust regression estimates with $\mathrm{GES}_T = \infty$].

Figure 1 displays the maximum bias curves for three proposed robust estimates of scale: $H95$, a Huber proposal 2 estimate of scale, adjusted for 95% efficiency at the Gaussian model [Huber (1964)]; the median of absolute deviations about the median (Madm); and the "shortest half" of the data (Shorth). Observe that $\varepsilon^*_{\mathrm{Shorth}} = \varepsilon^*_{\mathrm{Madm}} = 0.5$, the largest possible value of ε^* and $\varepsilon^*_{H95} = 0.17$. The breakdown point of a classical Gaussian maximum likelihood estimate is typically zero. The GES_T lines provide local linear

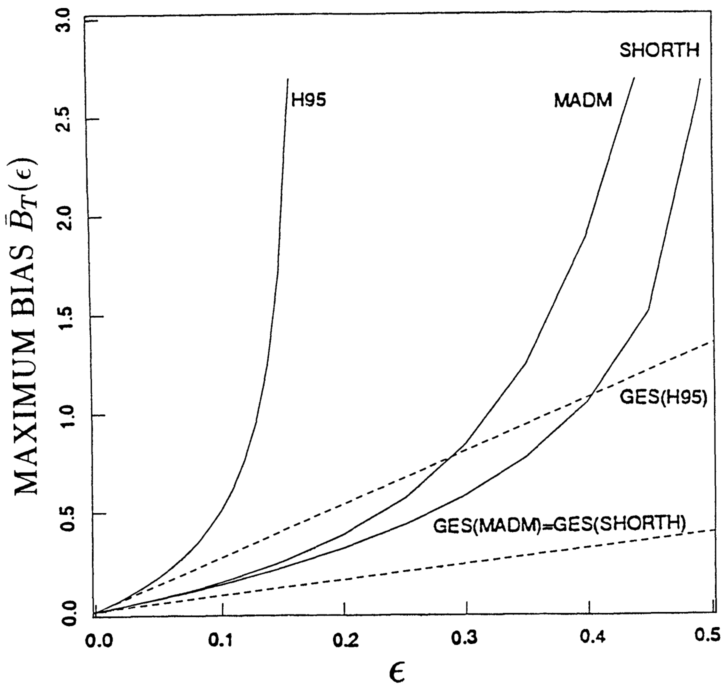

FIG. 1. *Maximum bias curves and GES linear approximations for the* 95% *efficient Huber M-estimate of scale and for the* 68% *efficient Madm and Shorth.*

approximations to the maximum bias, which are reasonable for not too large values of ε [just how large the reader can judge for himself; see the rule of thumb in Hampel, Ronchetti, Rousseeuw and Stahel (1986)].

The remainder of the paper is organized as follows. Section 2 introduces the class of M-estimates of scale with general location. This class includes the well known Huber (Proposal 2) M-estimates of location and scale, and also the class of *scale estimates* called *S-estimates*, which are associated with so-called S-estimates of regression [Rousseeuw and Yohai (1984)]. Section 2 also shows that, under certain regularity conditions, the finite sample value and the asymptotic value of M-estimates of scale are uniformly close, as F ranges over the family $\mathscr{F}_\varepsilon$. Moreover, prior results in Martin and Zamar (1989) indicate that the bias is a significant component of the mean-squared error for rather small to moderate sample sizes, depending on the value of ε. Section 3 gives a class of generalized bias functions to deal with the intrinsic asymmetry of the bias of scale estimates. Section 4 constructs minimax bias-robust estimates for the class of Huber (Proposal 2) M-estimates of location and scale, and shows that the bias robust estimates are well approximated by the Madm. Section 5 constructs minimax bias-robust S-estimates of scale, which are shown to be min-max in the larger class of M-estimates of scale with general location introduced in Section 2. Section 5 also shows that these estimates are reasonably well approximated by the Shorth. Section 6 briefly discusses the difference between bias-robust Huber estimates and S-estimates. Sections 7 and 8 give some encouraging finite sample results. Finally, Section 9 closes with a brief discussion of the GES linear approximation to the maximum bias curve. Proofs of lemmas and theorems are given in Section 10.

Our results on the Shorth complement recent results of Rousseeuw and Leroy (1988), who propose the Shorth as a robust scale estimate. They derive the influence function, the finite sample breakdown-point, and a correction factor to achieve approximate finite sample size unbiasedness at the normal distribution. Another interesting recent work on the Shorth is that of Grübel (1988), who establishes asymptotic normality.

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 subset F of distribution functions $\mathscr{F}$, which is assumed to include all the empirical distribution functions F_n and the ε-contamination family (1). The scale estimate $\hat{s}_n$ is then obtained by evaluating the functional $S(F)$ at F_n: $\hat{s}_n = S(F_n)$.

Suppose the following:

ASSUMPTION 1. χ is even, nondecreasing on $[0, \infty)$, bounded, with at most a finite number of jumps, and that $\chi(\infty) = 1$.

Let $b(\chi) = E_{F_0}\chi(X)$ and for each $t \in R$, let $S(F, t)$ be the M-estimate of scale of $X - t$ defined by

$$S(F, t) = \sup\{s\colon E_F\chi[(X - t)/s] > b(\chi)\}. \tag{2}$$

ASSUMPTION 2. $\varepsilon < b(\chi) < 1 - \varepsilon$, for a fixed value of $\varepsilon \in (0, 0.5)$.

In view of (2) and Assumptions 1 and 2, there is no loss of generality in the assumption that $\chi(\infty) = 1$.

The definition (2) is needed to insure uniqueness and to handle possible discontinuities of F and χ. If χ (or F) is continuous, then $S(F, t)$ satisfies

$$E_F\chi[(X - t)/S(F, t)] = b(\chi). \tag{3}$$

Since the location parameter μ_0 in (1) is unknown, it must be estimated along with s_0. Let $T(F)$ be a location and scale equivalent functional, that is,

$$T[F((x - t)/s)] = sT[F(x)] + t, \qquad \forall\, t \in R, \forall\, s > 0.$$

The *M-estimate of scale with general location* is now defined as

$$S(F) = S[F, T(F)].$$

Some particular cases are:

Huber Proposal 2. In this case $T(F)$ and $S[F, T(F)]$ simultaneously satisfy (3), with $t = T(F)$, and

$$E_F \psi[(X - T(F))/S(F, T(F))] = 0, \tag{4}$$

where

ASSUMPTION 3. $\psi(x)$ is odd, nondecreasing, bounded, with at most a finite number of jumps.

In particular the Madm is obtained when χ is the jump function,

$$\chi_a(x) = \begin{cases} 0, & \text{if } |x| \le a, \\ 1, & \text{if } |x| > a, \end{cases} \tag{5}$$

with $a = F_0^{-1}(3/4)$, and ψ is the "sign" function

$$\psi_0(x) = \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases} \tag{6}$$

In this case $T(F) = F^{-1}(1/2)$ is the median of F.

S-Estimate of scale. In this case the location estimate $T(F)$ is a minimizer of $S(F, t)$, that is,

$$S(F) = \inf_{t \in R} S(F, t). \tag{7}$$

It is not difficult to see that $S(F)$ and $T(F)$ satisfies (3) and (4) with $\psi(x) = \chi'(x)$. Since $\chi(x)$ is bounded, $\psi(x)$ tends to zero as x tends to infinity, that is, $\psi(x)$ is redescending. In particular, the Shorth is obtained when χ is given by (5). Observe that the Madm and the Shorth have both the same chi-function (6) but different centering functionals.

The following lemma shows that under mild assumptions the breakdown point of the functional $S(F, t)$ is larger than ε. The proof is straightforward and therefore omitted.

LEMMA 1. *Let $K > 0$ be given and suppose that Assumptions* 0–2 *hold. Then, there exist $0 < s_1 < s_2 < \infty$ such that $s_1 \le S(F, t) \le s_2$ for all $|t| < K$ and $F \in \mathscr{F}_\varepsilon$.*

Theorem 1 below shows that, under some regularity conditions which include the continuity of χ, $S(F_n) \to S(F)$ a.s. $[F]$ as $n \to \infty$, uniformly over $\mathscr{F}_\varepsilon \times \mathscr{C}$, where $\mathscr{C}$ is a certain class of χ-functions. Unfortunately, the case of χ-functions of the jump type given by (5) is not covered by Theorem 1. However, Theorem 6 and the Monte Carlo results presented in Sections 7 and 8 support the finite sample relevance of the asymptotic minimax-bias theory for this important special case.

The following definitions are needed for stating Theorem 1:

$$(8) \qquad g_\chi(s,t) = E_{F_0}\chi[(X-t)/s], \qquad h_\chi(s,t) = (\partial/\partial s)g_\chi(s,t).$$

THEOREM 1. *Suppose that Assumptions* 0–2 *hold. Assume also that* χ *and* $h_\chi(s,t)$ *are continuous and that* $h_\chi(s,t) < 0$ *for all* $s > 0$, $t \in R$. *Let* $K > 0$ *be fixed. Then, for all* $\delta > 0$:

(a) $\lim_{m\to\infty} \sup_{F\in\mathscr{F}_\varepsilon} P_F\{\sup_{n\ge m} \sup_{|t|\le K} |S(F_n,t) - S(F,t)| > \delta\} = 0$.

(b) *If* $S(F)$ *is given by* (7), *then* $\lim_{m\to\infty} \sup_{F\in\mathscr{F}_\varepsilon} P_F\{\sup_{n\ge m}|S(F_n) - S(F)| > \delta\} = 0$.

(c) *If* $\sup_{F\in\mathscr{F}_\varepsilon} |T(F)| < \infty$ *and*

$$\lim_{m\to\infty} \sup_{F\in\mathscr{F}_\varepsilon} P_F\left\{\sup_{n\ge m} |T(F_n) - T(F)| > \delta\right\} = 0,$$

then

$$\lim_{m\to\infty} \sup_{F\in\mathscr{F}_\varepsilon} P_F\left\{\sup_{n\ge m} |S[F_n, T(F_n)] - S[F, T(F)]| > \delta\right\} = 0.$$

(d) *Let* $x_0 > 0$ *be fixed. The class* $\mathscr{C}$ *is defined as the set of* χ-*functions satisfying all the previous assumptions and* (i) $\chi(x) = 1\ \forall\ |x| \ge x_0$ *and* (ii) *there exists* $h_0(s,t)$ *such that* $h_\chi(s,t) \le h_0(s,t) < 0$, $\forall\ s > 0$, $\forall\ t \in R$.

Then (a), (b) *and* (c) *hold uniformly on* $\mathscr{C}$.

REMARK. Suppose that a certain function χ_0 satisfies Assumptions 1 and 2 and is such that

$$h_{\chi_0}(s,t) = -\frac{1}{s}E_{F_0}\left[\chi_0'\left(\frac{x-t}{s}\right)\frac{x-t}{s}\right] < 0, \qquad \forall\, s > 0, t \in R$$

and

$$\chi_0(x) = 1, \qquad \forall\, |x| > x_0.$$

Then the set

$$\mathscr{C} = \left\{\chi\colon \chi'(x) \le \chi_0'(x), \forall\, x > 0 \text{ and } h_\chi(s,t) = -\frac{1}{s}E_{F_0}\left[\chi'\left(\frac{X-t}{s}\right)\frac{X-t}{s}\right]\right\}$$

satisfies the assumptions of Theorem 1.

3. 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 are in general asymptotically biased for $F \in \mathscr{F}_\varepsilon$. 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, the bias $B_r[S(F)]$ is positive, and when F is an inliers generating distribution, the bias $B_r[S(F)$ is 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

$$B[S(F)] = \begin{cases} L_1[S(F)/s_0], & \text{if } 0 < S(F) \leq s_0, \\ L_2[S(F)/s_0], & \text{if } s_0 < S(F) \leq \infty, \end{cases} \tag{9}$$

where L_1 and L_2 are continuous, nonnegative and monotone, with $L_1(1) = L_2(1) = 0$ and

$$\lim_{t \to 0} L_1(t) = \lim_{t \to \infty} L_2(t) = \infty.$$

We are interested in the maximum generalized bias,

$$\overline{B}(\varepsilon) = \max_{F \in \mathscr{F}_\varepsilon} B[S(F)]. \tag{10}$$

From monotonicity of L_1 and L_2, it follows that

$$\overline{B}(\chi, T) = \max\{L_1[S^-/s_0], L_2[S^+/s_0]\},$$

where S^- and S^+ denote the supremum and the infimum of the functional $S(F)$ as F ranges over $\mathscr{F}_\varepsilon$.

4. Bias robust Huber estimates. In view of the historical importance and high degree of familiarity of Huber (Proposal 2) estimates we first focus on obtaining bias robust estimates in this class. To emphasize the dependence on χ and ψ we use the notation $S(F, \chi, \psi)$, $S^+(\chi, \psi)$, $S^+(\chi, \psi)$ and so on.

The first step toward finding the bias robust Huber estimate is deriving the expressions (16) and (17) for $S^-(\chi, T_\psi)$ and $S^+(\chi, T_\psi)$. Claims which are made below without proof can be easily verified under Assumptions 0–3.

The maximum value $S^+(\chi, \psi)$ of the scale functional $S(F, \chi, \psi)$ is produced by a point mass contamination at infinity δ_∞ and such contamination also produces the maximum value of the location estimate $T_\psi(F)$. The estimating equations in this limit case are

$$(1 - \varepsilon) E_{F_0} \chi[(X - t)/s] + \varepsilon = b(\chi) \tag{11}$$

and

$$(1 - \varepsilon) E_{F_0} \psi[(X - t)/s] + \varepsilon = 0. \tag{12}$$

Let $\gamma_\chi(t)$ be the unique solution of (11) for fixed t and let $r_\psi(s)$ be the unique solution of (12) for fixed $s > 0$. The function $m_{\chi,\psi}(t) = r_\psi[\gamma_\chi(t)]$ is continuous and nondecreasing. Also, the pair (s^*, t^*) simultaneously satisfy (11) and (12) if and only if $t^* = m_{\chi,\psi}(t^*)$ and $s^* = \gamma_\chi(t^*)$.

The following lemma characterizes the maximum asymptotic bias due to outliers of the location and scale Huber M-estimates. This lemma also provides an algorithm for computing these maximum biases. We recall that Huber (1964) has shown that the maximum asymptotic bias of the median (the minimax-bias estimate of location) is

$$t_0 = F_0^{-1}[0.5/(1-\varepsilon)]. \tag{13}$$

LEMMA 2. *Suppose that Assumptions* 0–3 *hold. For each* $n \geq 1$ *let* $t_n = m_{\chi,\psi}(t_{n-1})$, *with* t_0 *given by* (13). *Let* $s_n = \gamma_\chi(t_n)$ *and*

$$t^* = \inf\{t > t_0 \colon m_{\chi,\psi}(t) = t\}.$$

Then, (a) $\lim_{n\to\infty} t_n = t^*$ *and* $\lim_{n\to\infty} s_n = \gamma_\chi(t^*) = s^*$; (b) *the maximum asymptotic bias of the location estimate* $T(F,\chi,\psi)$ *is* t^* *and the maximum value of the scale functional* $S(F,\chi,\psi)$ *is* $S^+(\chi,\psi) = s^*$.

The minimum value of the scale functional $S(F,\chi,\psi)$, $S^-(\chi,\psi)$, is produced by a point mass contamination δ_0 at zero. In this case the estimating equations are

$$(1-\varepsilon)E_{F_0}\chi[(X-t)/s] + \varepsilon\chi(t/s) = b(\chi) \tag{14}$$

and

$$(1-\varepsilon)E_{F_0}\psi[(X-t)/s] + \varepsilon\psi(-t/s) = 0. \tag{15}$$

By monotonicity of ψ, $t = 0$ for all $s > 0$. Let g_t^{-1} be the inverse of $g_\chi(s,t)$ with respect to s, for fixed t. Then, from (14) with $t = 0$ it follows that

$$S^-(\chi,\psi) = g_0^{-1}[b(\chi)/(1-\varepsilon)]. \tag{16}$$

Optimal centering. The choice of ψ has an effect on the maximum asymptotic bias of the scale estimate by virtue of affecting the bias t^* of the location estimate. Observe that since $S^-(\chi,\psi)$ does not depend on ψ [see (16)], the optimal choice of ψ must be based on $S^+(\chi,\psi)$ alone.

It follows from Lemma 2 and (11), with $t = t^*$, that

$$S^+(\chi,\psi) = g_{t^*}^{-1}[(b-\varepsilon)/(1-\varepsilon)]. \tag{17}$$

For all $0 < \alpha < 1$ the function $g_t^{-1}(\alpha)$ is nondecreasing in t. Therefore, by the Huber (1964) minimax-bias result, $r_\psi(s) \geq t_0 = r_{\psi_0}(s)$, for all $s > 0$ and ψ, where ψ_0 is the "sign" function (6). Thus we have the following result:

THEOREM 2. *For each fixed* χ *satisfying Assumption* 1 *the median centering functional minimizes the maximum asymptotic bias of both location and scale among Huber estimates with* ψ *satisfying Assumption* 3.

More generally, it is not difficult to show that Theorem 2 holds for the class of all M-estimates of scale with centering functional $T(F)$ having the *monotonicity* property

$$T(F) \leq T[(1-\varepsilon)F_0 + \varepsilon\delta_\infty], \qquad \forall\, F \in \mathscr{F}_\varepsilon. \tag{18}$$

The minimax-bias Huber estimate of scale. By Theorem 2 it suffices to consider $S^+(\chi,\psi_0)$ and $S^-(\chi,\psi_0)$ and the function ψ can be dropped from the notations. It will be shown that under certain conditions the maximum generalized bias $\overline{B}(\chi)$ [see (10)] is minimized by a jump function χ_{a^*} [see (6)].

For each $a > 0$, let $\overline{B}(a) = \overline{B}(\chi_a)$ and

$$b(a) = b(\chi_a) = 2[1 - F_0(a)]. \tag{19}$$

We begin by showing that given $0 < \varepsilon < 0.5$, F_0, L_1 and L_2 there exists a^* such that

$$\overline{B}(a^*) \le \overline{B}(\chi_a), \qquad \forall\, a > 0. \tag{20}$$

Let $a_0 = F_0^{-1}[(1+\varepsilon)/2]$ and $a_1 = F_0^{-1}[(2-\varepsilon)/2]$. From (19), the corresponding values of b are $b_0 = b(a_0) = 1 - \varepsilon$ and $b_1 = b(a_1) = \varepsilon$. Hence, letting $S^-(a) = S^-(\chi_a)$ and $S^+(a) = S^+(\chi_a)$, we have

$$\begin{aligned} \lim_{a\to a_0} S^-(a) &= \lim_{a\to a_0} g_0^{-1}\left[\frac{b(a)}{1-\varepsilon}\right] \\ &= \lim_{a\to a_0} \frac{1}{a} F_0^{-1}\left[1 - \frac{b(a)}{2(1-\varepsilon)}\right] = \frac{1}{a_0} F_0^{-1}(0.5) = 0 \end{aligned} \tag{21}$$

and

$$\begin{aligned} \lim_{a\to a_1} S^+(a) &= \lim_{a\to a_1} g_{t^*}^{-1}\left[\frac{b(a)-\varepsilon}{1-\varepsilon}\right] \ge \lim_{a\to a_1} g_0^{-1}\left[\frac{b(a)-\varepsilon}{1-\varepsilon}\right] \\ &= \lim_{a\to a_1} \frac{1}{a} F_0^{-1}\left[1 - \frac{b(a)-\varepsilon}{2(1-\varepsilon)}\right] = \frac{1}{a_1} F_0^{-1}(1) = +\infty. \end{aligned} \tag{22}$$

Therefore, by the assumptions on L_1 and L_2, $\overline{B}(a) \to +\infty$ when either $a \to a_0$ or $a \to a_1$. By continuity of $\overline{B}(a)$ there exists $a_0 < a^* < a_1$ such that

$$\overline{B}(a^*) \le \overline{B}(a), \qquad \forall\, a_0 < a^* < a_1. \tag{23}$$

Thus, the jump function χ_{a^*} is bias robust among all jump functions χ_a. The following theorem gives conditions under which $S(F, \chi_{a^*}, \psi_0)$ is bias robust among all Huber estimates of scale.

THEOREM 3. *Let $s^* = S^+(a^*)$, where a^* is given by* (23), *and let t_0 be as in* (13). *Suppose that in addition to Assumptions* 0 *and* 1 *the following conditions hold*:

(a) $f_0(x) > 0$, $\forall\, x \in R$ *and* $f_0(sx)/f_0(x)$ *is increasing in* $|x|$, $\forall\, 0 < s < 1$.

(b) *The function* $k_0(x) = [f_0(s^*x - t_0) + f_0(s^*x + t_0)]/f_0(x)$ *is decreasing in* $|x|$.

(c) $S^-(a)$ *and* $S^+(a)$ *are both strictly monotone at* $a = a^*$.

Then $\overline{B}(a^*) \le \overline{B}(\chi, \psi)$ *for all pairs* (χ, ψ) *satisfying Assumptions* 1–3.

It can be shown that the conditions of Theorem 3 hold, for example, when F_0 is the standard normal distribution and $\varepsilon < 0.35$ [see Martin and Zamar (1987)].

TABLE 1
Bias-robust Huber Proposal 2 estimates of scale when F_0 = standard normal. Logarithmic loss function

ε	a^*	$b(a^*)$	$\bar{B}(a^*)$	$\bar{B}$(Madm)
0.05	0.650	0.516	0.062	0.063
0.10	0.674	0.500	0.135	0.135
0.15	0.673	0.501	0.221	0.221
0.20	0.673	0.501	0.324	0.324
0.30	0.676	0.499	0.609	0.612
0.40	0.695	0.487	1.072	1.166
0.45	0.713	0.476	1.440	1.779

Near optimality of the Madm. Let b^* be the value of $b(\chi_{a^*}) = E_{F_0}\chi_{a^*}(x)$. Since the bias robust estimate of Theorem 3 is based on χ_{a^*}, using the median for centering, it follows that the bias robust Huber estimate is the $n - [nb^*]$ order statistic of the absolute value of the residuals about the median (scaled by $1/a^*$), where $a_0(\varepsilon) < a^* < a_1(\varepsilon)$. Since both, $a_0(\varepsilon)$ and $a_1(\varepsilon)$ tend to $F_0^{-1}(0.75)$ as $\varepsilon \to 0.5$, so does a^*. Thus, as $\varepsilon \to 0.5$, the bias robust Huber estimate is the well known Madm, whose breakdown-point is equal to 0.5.

It came as a pleasant surprise that for a broad range of ε, the maximum bias of the bias robust estimate is very close to the Madm for the leading case of the nominal Gaussian distribution and the logarithmic loss function $L_2(t) = -L_1(t) = \log(t)$. Table 1 shows the values of $a^*, b^* = b(a^*)$, the minimaxbias $\bar{B}(a^*)$ and the maximum bias $\bar{B}$(Madm) of the Madm for some values of ε. 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 estimates. Note in particular that even when we choose ε small, for example, $\varepsilon = 0.05$, the breakdown-point of the minimax-bias scale estimate is very close to 0.5.

5. Bias robust *S*-estimates. One naturally wonders whether the greater bias robustness can be obtained by enlarging the class of estimates over which one searches for a minimax solution. In particular one may consider the entire class of M-estimates of scale with general location. This larger class of course includes joint M-estimates of location and scale with *redescending* as well as monotone ψ for the location estimate.

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

The following notation is needed for stating Theorem 4. Set $S^+(\chi)$ and $S^-(\chi)$ denote the maximum and minimum asymptotic values of the S-estimate of scale based on χ [see (7)]. Let $S^+(\chi, T)$ and $S^-(\chi, T)$ denote the maximum and minimum asymptotic values of the M-estimate of scale $S_\chi[F, T(F)]$ with general location, based on χ and the location estimate $T(F)$.

THEOREM 4. *Suppose that Assumptions* 0–2 *hold and let* $\gamma_\chi(s) = g_\chi(s, 0)$, *where* $g_\chi(s, t)$ *is given by* (8). *Let* T *be any location-scale equivariant estimate satisfying the condition*

$$T[(1-\varepsilon)F_0 + \varepsilon\delta_0] = 0.$$

Then,

(a) $S^+(\chi) = \gamma_\chi^{-1}[(1-b)/(1-\varepsilon)] \leq S^+(\chi, T)$.
(b) $S^-(\chi) = \gamma_\chi^{-1}[b/(1-\varepsilon)] = S^-(\chi, T)$.

This paves the way for the following main result.

THEOREM 5. *Suppose that Assumptions* 0–2 *hold. Then there exists a jump function* χ_{a^*} *such that the S-estimate based on* χ_{a^*} *has the minimax asymptotic bias over the class of all M-estimates of scale with general location.*

Therefore, the minimax estimate is an S-estimate based on the jump function χ_{a^*}.

Near optimality of the Shorth. As in Section 4, $a^* \to F_0^{-1}(0.75)$ as $\varepsilon \to 0.5$. Thus the minimax estimate of scale with general location tends to the Shorth as $\varepsilon \to 0.5$. Table 2 shows the values of $a^*, b^* = b(a^*)$, the minimax bias $\overline{B}(a^*)$ and the maximum bias $\overline{B}$(Shorth) of the Shorth for some values of ε. These results show that the minimax estimate is reasonably well approximated by the Shorth in terms of maximum bias, the approximation being less good for larger values of ε. One again finds that a breakdown point reasonably close to 0.5 is obtained by the minimax estimate for a wide range of values of ε.

It should be remarked that the S-estimate of location associated with the Shorth, namely the midpoint of the shortest half of the data, has a slow rate of convergence [Andrews, Bickel, Hampel, Huber, Rogers and Tukey, (1972)]. However, the Shorth estimate of scale has the usual rate of convergence [Grübel (1988)].

TABLE 2
Bias-robust M-estimates of scale with general location when F_0 = *standard normal*
Logarithmic loss function

ε	a^*	$b(a^*)$	$\overline{B}(a^*)$	$\overline{B}$(Shorth)
0.05	0.650	0.516	0.060	0.060
0.10	0.700	0.484	0.127	0.135
0.15	0.716	0.474	0.201	0.220
0.20	0.726	0.468	0.284	0.322
0.30	0.751	0.453	0.495	0.612
0.40	0.763	0.445	0.845	1.166
0.45	0.746	0.456	1.236	1.779

6. Huber estimates versus S-estimates of scale: Madm versus Shorth. The class of Huber estimates of scale considered in Section 3 excludes centering functionals which are M-estimates of location with redescending ψ. These location estimates are of course allowed in the larger class considered in Section 5. We now show that the S-estimate of location $T_\chi(F)$ is in fact an M-estimate of location with redescending psi-function $\psi(x) = \chi'(x)$: Let $t^* = \arg\min_t S(F, t)$. The monotonicity of $\chi(x)$ on $[0, \infty)$ and the definition of the S-estimate of scale $S(F)$ [see (7)] implies that

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

So, t^* minimizes the function $l(t) = E_F \chi[X - t/S(F)]$ and therefore satisfies the equation $l'(t^*) = 0$, that is, t^* satisfies the location M-estimate equation

$$l^*(t) = E_F \chi'[X - t^*/S(F)] = 0.$$

Figure 2(a) and 2(b) display the maximum bias curves of the minimax Huber and S-estimates of scale (for the case of logarithmic loss function) for

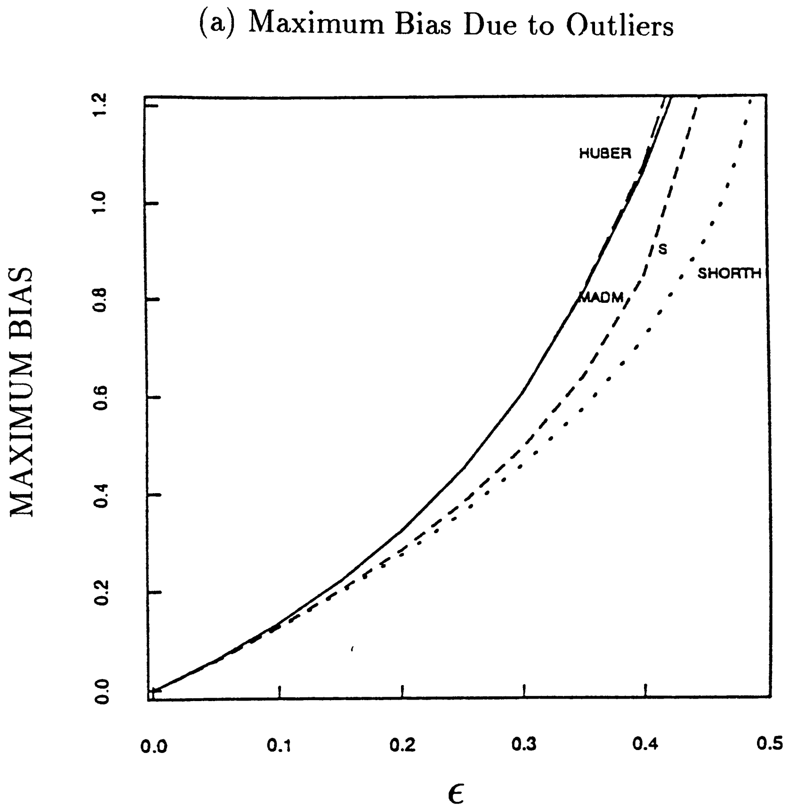

FIG. 2. *Maximum bias of the optimal S-estimate of scale with general location, the optimal Huber Proposal 2 M-estimate of scale. The Madm and the Shorth.* $F_0 = N(0, 1)$, *logarithmic loss.*

(b) Maximum Bias Due to Inliers

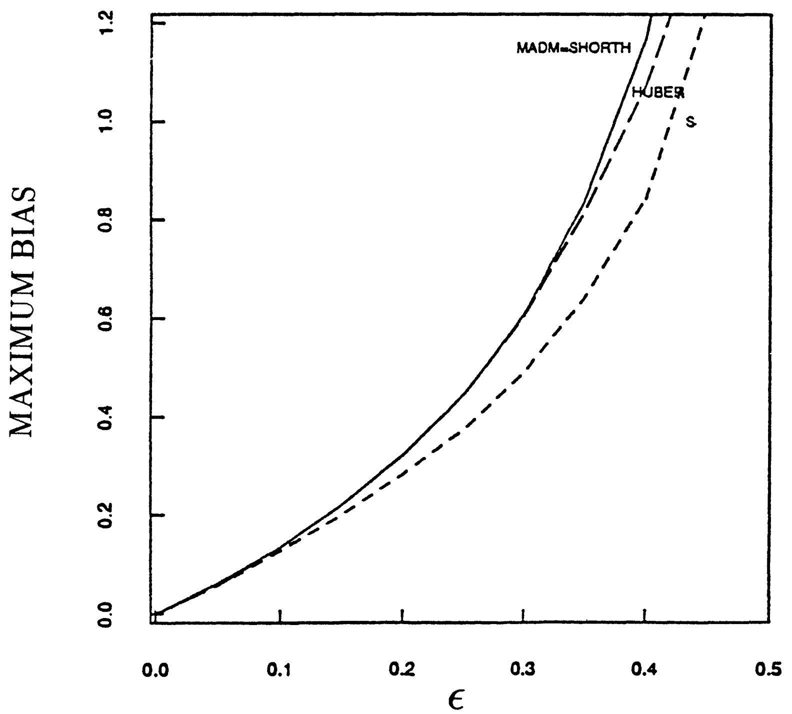

FIG. 2. (*Continued*)

outliers and inliers, respectively. The logarithmic bias for the Madm and the Shorth are also shown. Figure 2 reveals uniformly smaller bias for the minimax S-estimate than for the minimax Huber estimate.

We notice that in Figure 2(a) the maximum bias curve for the Shorth is uniformly smaller than that of the minimax S-estimate, whereas the opposite is true in Figure 2(b). This is a consequence of the relative way in which the logarithmic loss function penalizes positive and negative bias. It is worth noticing that if one is concerned only about outliers, then the Shorth is the best choice with respect to bias. The better performance of S-estimates relative to the Huber M-estimates in the case of outliers is a consequence of the S-estimate of location being an M-estimate with redescending ψ, which suffers no bias for gross outliers.

Also referring back to Figure 1 we would remark that the price paid for using a high efficiency Huber estimate is in terms of maximum bias and breakdown point. Table 3 presents mean-squared-error relative efficiencies of the Shorth relative to Madm for finite sample sizes $n = 20, 40, 100$, computed by Monte Carlo simulation. These results indicate considerable superiority of the Shorth for outliers, and moderate superiority of Madm for inliers.

TABLE 3
Mean-squared-error relative efficiencies of Shorth and Madm

ε	n = 20		n = 40		n = 100	
	Outliers	**Inliers**	**Outliers**	**Inliers**	**Outliers**	**Inliers**
0.00	1.09	1.09	1.10	1.10	1.18	1.18
0.05	1.11	1.11	1.06	1.12	1.06	1.00
0.10	1.08	1.07	1.00	1.02	1.07	0.90
0.15	1.10	0.98	1.04	0.91	1.12	0.83
0.20	1.10	0.89	1.14	0.82	1.27	0.77
0.25	1.23	0.85	1.27	0.76	1.39	0.75
0.30	1.43	0.83	1.46	0.77	1.61	0.78
0.35	1.62	0.82	1.66	0.77	1.83	0.77
0.40	1.74	0.85	1.86	0.80	2.01	0.78
0.45	1.87	0.92	2.00	0.85	2.16	0.82

7. Finite sample relevance of asymptotic bias robustness. Unfortunately, the functions χ_a are discontinuous and so Theorem 1 cannot be invoked to claim finite sample relevance for the asymptotic minimax theory. However we can prove the following result, which is relevant to the finite sample size situation.

THEOREM 6. *Let* $0 < a < \infty$. *For each* $\Delta > 0$,

$$\lim_{m \to \infty} \sup_{F \in \mathscr{F}_\varepsilon} P_F\left\{S^-(a) - \Delta \le S_{\chi_a}(F_n) \ge S^+(a) + \Delta, \forall\, n \ge m\right\} = 1.$$

So $S^-(a) - \Delta$ and $S^+(a) + \Delta$ are almost sure uniform lower and upper bounds for the maximum and the minimum values of the S-estimate of scale $S_{\chi_a}(F_n)$ for m large enough.

The Monte Carlo results summarized in Figures 3 and 4 suggest that the required values of m are moderately small. These figures display the finite sample bias (logarithmic loss) for several contamination models for the Shorth and for the Madm, as well as the corresponding maximum bias curves. Observe that for both cases, for outliers and for inliers, the asymptotic maximum bias curves tend to be rather close to the finite sample bias curves.

8. Finite sample comparison with other estimates. A Monte Carlo simulation was carried out to compare the bias and the mean-squared-error performance of the following scale estimates: the Madm, the Shorth, the rejection-plus-standard-deviation (RPSD) estimate [Simonoff (1987)] and the scale A-estimate [Lax (1985)].

To define the RPSD estimate let $B_0 = \{x_1, \dots, x_n\}$ be the complete sample and, for $i \le (n-1)/2$, let B_i be the subset obtained from B_{i-1} by deleting the observation farthest from its mean. Such observation is denoted by $x^{(i-1)}$.

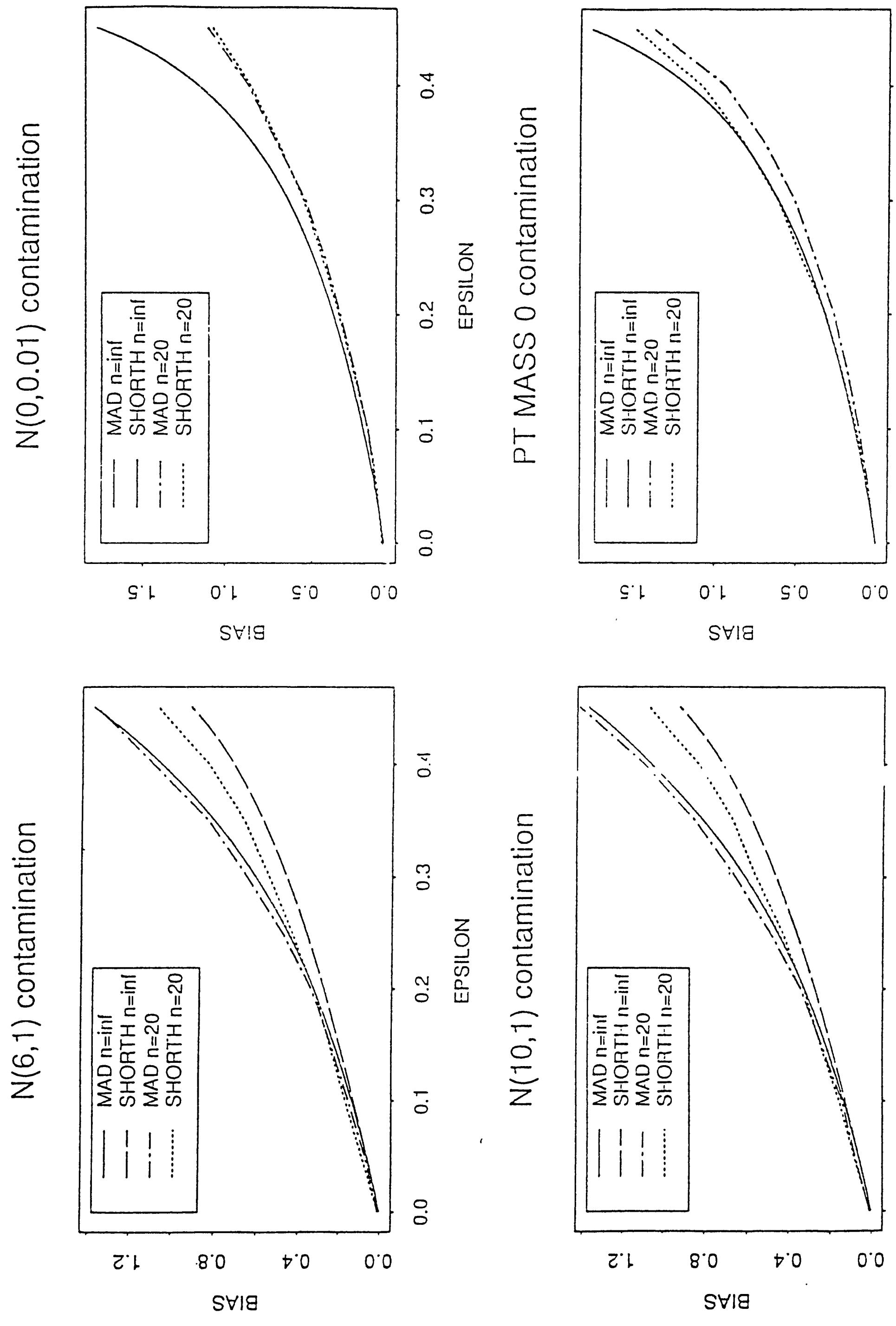

FIG. 3. *Finite sample* ($n = 20$) *and asymptotic bias curve for Madm and Shorth.* $F_0 = N(0,1)$, *logarithmic loss.*

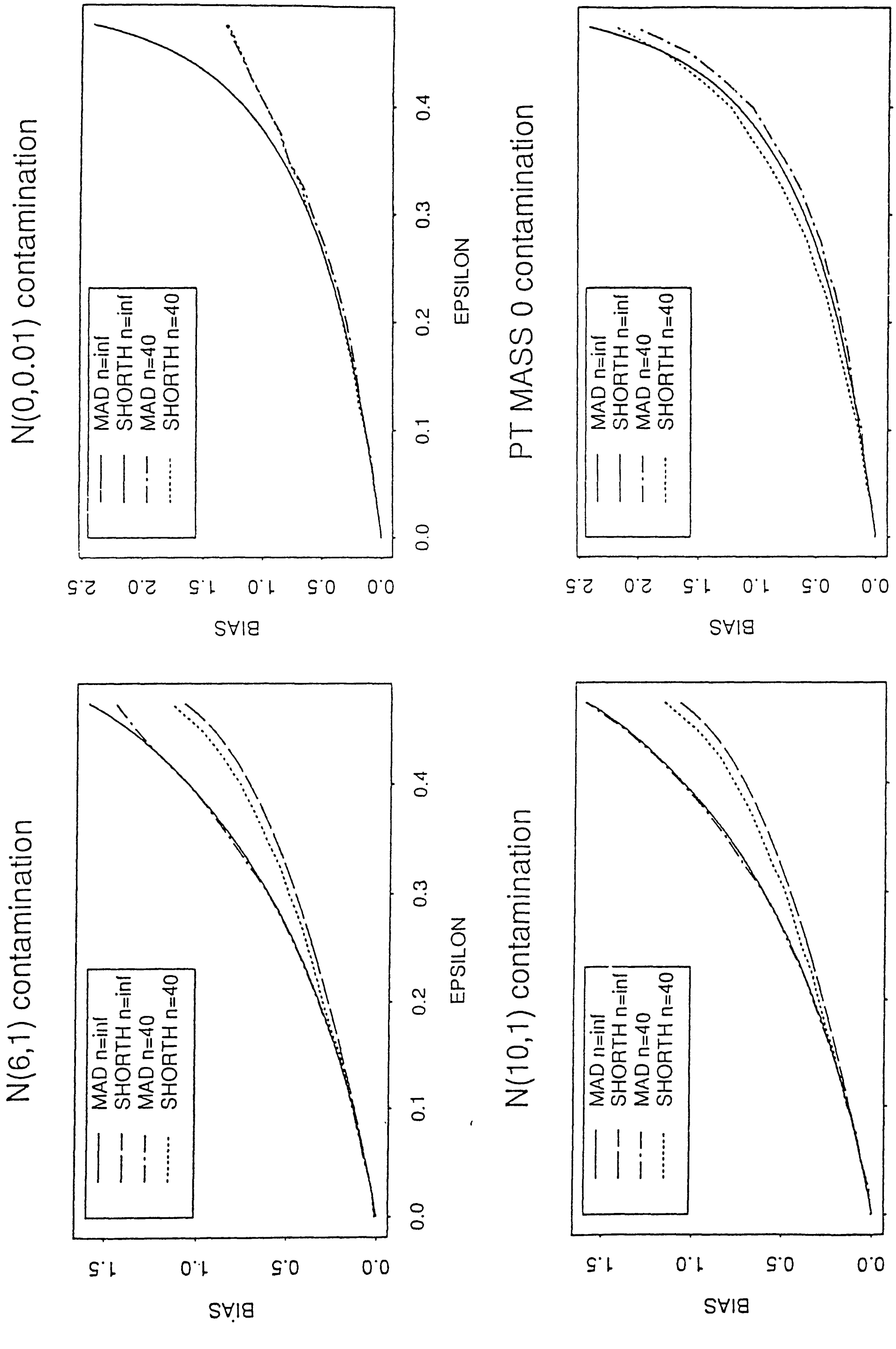

FIG. 4. *Finite sample* ($n = 40$) *and asymptotic bias curve for Madm and Shorth.* $F_0 = N(0,1)$, *logarithmic loss.*

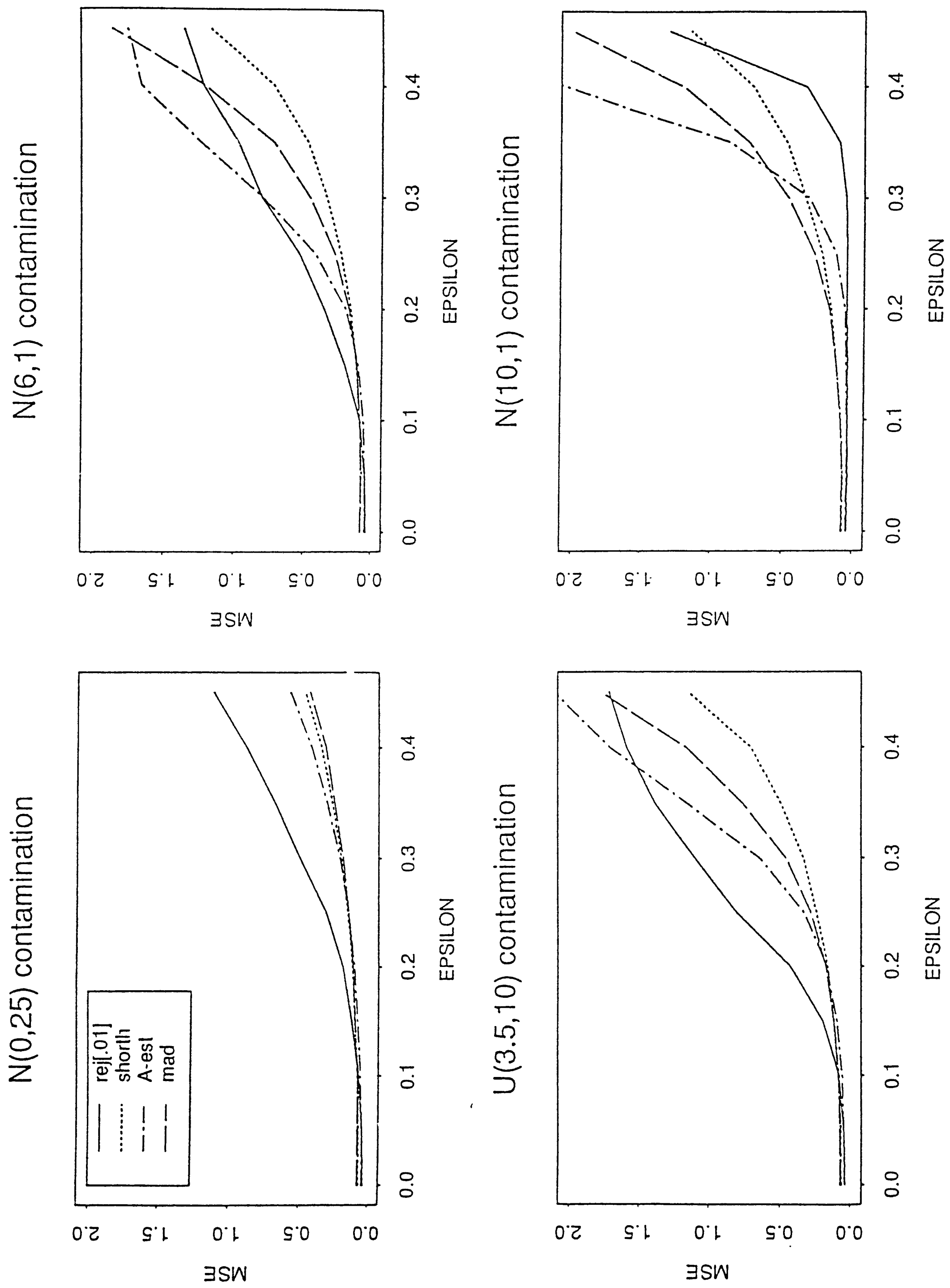

FIG. 5. *Mean-squared-error curves for* $n = 20$, $F_0 = N(0, 1)$ *and logarithmic loss.*

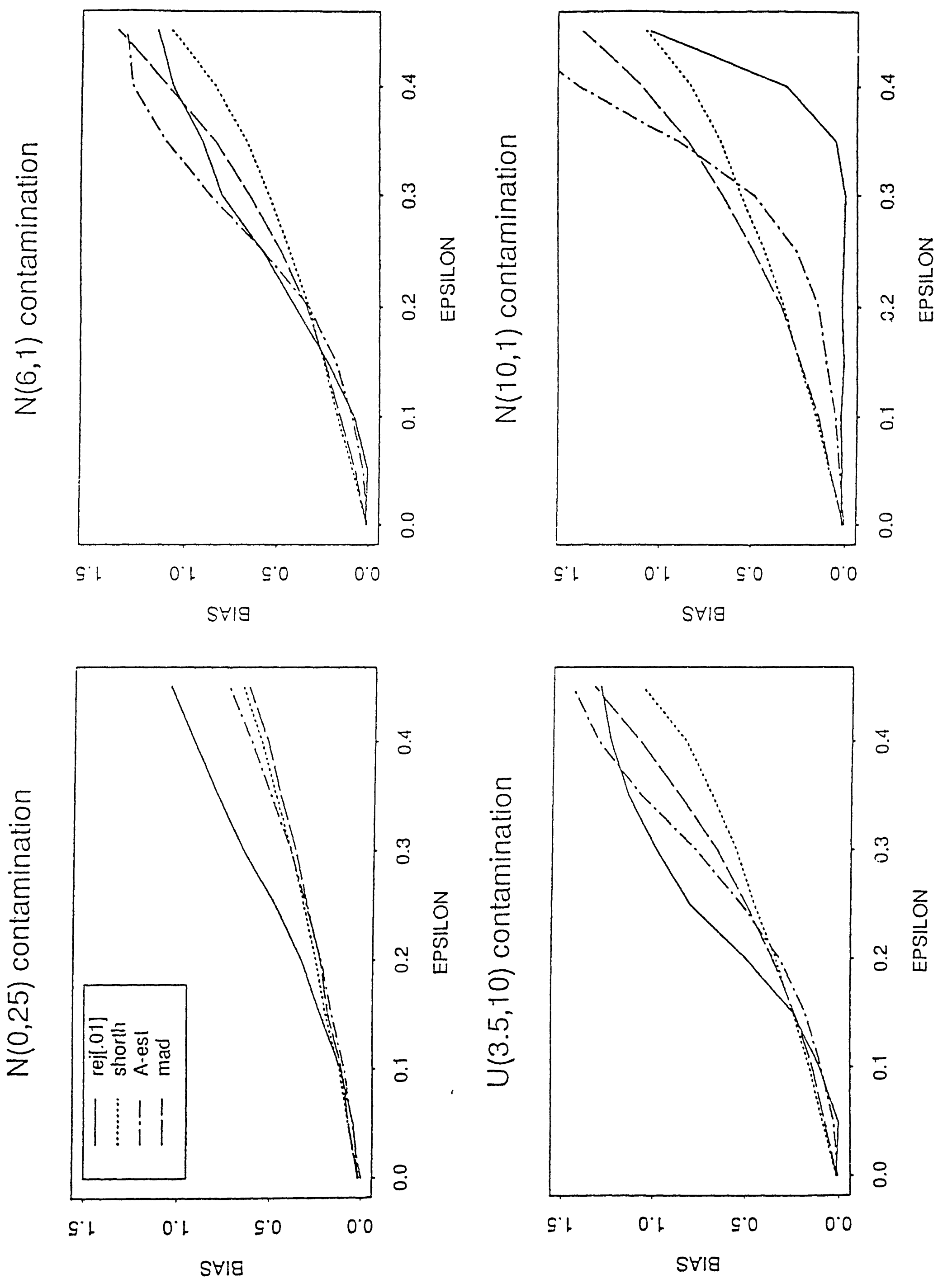

FIG. 6. *Bias curves for* $n = 20$, $F_0 = N(\ ,1)$ *and logarithmic loss.*

Let $\bar{x}(B_{i-1})$ and $sd(B_{i-1})$ be the mean and the standard deviation of B_{i-1},

$$ESD_i = \frac{\left|x^{(i-1)} - \bar{x}(B_{i-1})\right|}{sd(B_{i-1})}$$

and

$$i_0 = \max\left\{i \le \frac{n-1}{2}: ESD_i > d^*\right\}.$$

The constant d^* is chosen so that $P(ESD_1 > d^*) = 0.01$ for Gaussian data. The RPSD estimate is now defined as $sd(B_{i_0+1})$, the standard deviation of the "clean" sample.

To define the A-estimate S_A, let T_0 and S_0 be some initial estimates of location and scale and set $u_i = (x_i - T_0)/S_0$, $i = 1, \dots, n$. The A-estimate S_A is now defined as

$$S_A = S_0 \sqrt{\frac{n}{n-1} \frac{\sum^n \psi^2(u_i)/n}{\left[\sum^n \psi'(u_i)/n\right]^2}},$$

where ψ is a location score function. Notice that if $\psi(u) = u$ and $T_0 = \bar{x}$ then S_A reduces to the classical standard deviation. In our Monte Carlo simulation T_0 = Median, S_0 = Madm and ψ is the Tukey's bisquare score function with $c = 4.7$.

Some results for sample size $n = 20$ are presented on Figure 5 (MSE) and Figure 6 (bias), for the case $F_0 = N(0, 1)$ and logarithmic loss. Each simulated sample contains exactly $\varepsilon 20$ outliers generated from the four different distributions indicated at the tops of the figures. Similar results (not presented here) were obtained for $n = 40$, $n = 100$ and for other type of contaminating distributions. The main conclusions are:

1. When $\varepsilon < 0.10$ the four estimates perform equally well.
2. For larger fractions of outliers the Shorth and the Madm usually outperform the other two estimates, with the Shorth being somewhat better.
3. When the outliers are large and well separated from the rest of the data, for example, generated from a $N(10, 1)$, the rejection-plus-standard-deviation estimate performs better than the other three estimates.

9. The GES approximation. Hampel, Ronchetti, Rousseeuw and Stahel (1986) established that based on the gross-error sensitivity, the Madm is the *most bias robust* M-estimate of scale for vanishingly small fractions of contamination ε. In fact the Shorth has the same influence function and hence the same gross-error-sensitivity as the Madm, namely 0.787 [see Rousseeuw and Leroy (1988)]. However, this leaves unanswered the question of optimality for each $\varepsilon \in (0, 0.5)$, and our results show that the Shorth is a better estimate than the Madm from the global (i.e., $\varepsilon > 0$) point of view.

On the other hand, it must be noted that the gross-error-sensitivity approximation is remarkably good for $\varepsilon < 0.05$, with the approximation being better the more bias-robustness the estimate possesses. This provides substantial reconfirmation of the utility of the influence curve and the gross-error-sensitivity as a measure of maximal bias.

At the same time one should be aware that the gross-error-sensitivity 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 of the Madm does not reflects the impact of the bias of estimation of the nuisance location parameter. Since the maximum asymptotic bias of the Shorth is unaffected by the asymptotic bias of the location estimate, the GES approximation is better in this case.

10. Proofs of lemmas and theorems. The following lemma is needed to prove Theorem 1.

LEMMA 3. *Let $0 < s_1 < s_2 < \infty$ be as in Lemma 1. Suppose that Assumptions 0–2 hold and also assume that χ and $h_\chi(s,t)$ are continuous and $h_\chi(s,t) < 0$, $\forall\, s > 0$, $t \in R$. Then, for all $K > 0$, we have*:

(a) *$\chi[(x-t)/s]$ is uniformly continuous on $(s,x,t) \in R \times [s_1, s_2] \times [-K, K]$.*

(b) *$S(F,t)$ is uniformly continuous on $t \in R$, uniformly on $F \in \mathscr{F}_\varepsilon$.*

(c) *$\chi[(x-t)/S(F,t)]$ is uniformly continuous on $(x,t) \in R \times [-K,K]$, uniformly on $F \in \mathscr{F}_\varepsilon$.*

PROOF. Let $\delta > 0$ be fixed and let $B = [s_1, s_2] \times [-K, K]$. Since $\chi(x)$ is continuous, even, monotone on $[0, \infty)$ and bounded, it is uniformly continuous. Let $\Delta_1 > 0$ be such that $|x - x'| < \Delta_1$ implies $|\chi(x) - \chi(x')| < \delta$. Also, since $\lim_{|x| \to \infty} \chi(x) = 1$, there exists $x_0 > 0$ such that $|x| > x_0$ and $|x'| > x_0$ imply $|\chi(x) - \chi(x')| < \delta$. Let $x_1 > 0$ be such that if $|x| > x_1$, then $|(x-t)/s| > x_0$ for all $(s,t) \in B$. So $\chi[(x-t)/s]$ is uniformly continuous on $\{x\colon |x| > x_1\} \times B$. If $|x| \le x_1$, then $|(x-t)/s - (x-t')/s'| \le x_1 |1/s - 1/s'| + |t'/s'| - |t/s|$ and (a) follows. To prove (b) notice that the assumptions on $h_\chi(s,t)$ imply that

$$\min_{(s,t)\in B} |h_\chi(s,t)| > 0,$$

and so $\delta_0 = \delta(1-\varepsilon)\min_{(s,t)\in B} |h_\chi(s,t)| > 0$. By (a) there exists $\Delta > 0$ such that $|t - t'| < \Delta$, $|t| \le K$, $|t'| \le K$ imply

$$\left| E_F \chi\left[\frac{X - t'}{S(F,t) - \delta}\right] - E_F \chi\left[\frac{X - t}{S(F,t) - \delta}\right]\right| \le \frac{\delta_0}{4}.$$

and so, using the mean value theorem and

$$E_F\chi[(X-t)/s] \geq (1-\varepsilon)E_{F_0}\chi[(X-t)/s], \forall F \in \mathscr{F}_\varepsilon,$$

$$\begin{aligned} E_F\chi\left[\frac{X-t'}{S(F,t)-\delta}\right] - b(\chi) &\geq E_F\chi\left[\frac{X-t}{S(F,t)-\delta}\right] - b(\chi) - \frac{\delta_0}{4} \\ &\geq (1-\varepsilon)\left\{E_{F_0}\chi\left[\frac{X-t}{S(F,t)-\delta}\right] - E_{F_0}\chi\left[\frac{X-t}{S(F,t)}\right]\right\} - \frac{\delta_0}{4} \\ &\geq \delta(1-\varepsilon)\min_{(s,t)\in B}|h_\chi(s,t)| - \frac{\delta}{4} \geq \frac{3\delta_0}{4} > 0. \end{aligned}$$

Thus, $S(F,t) \geq S(F,t') - \delta$ and (b) holds. Finally, (c) follows directly from (a) and (b). □

Proof of Theorem 1. Let $\delta > 0$ be fixed. It can be shown, as in the proof of Lemma 3(b), that there exists $0 < \delta_0 < 1$ such that

$$E_F\chi\left[\frac{X-t}{S(F,t)-\delta}\right] - b(\chi) \geq \delta_0, \quad \forall\, |t| \leq K, \forall F \in \mathscr{F}_\varepsilon. \tag{24}$$

For all $\gamma \geq 0$, let $B_n(t,\gamma) = \{(1/n)\Sigma\chi[(X_i - t)/S((F,t)-\delta)] - b(\chi) > \gamma\}$. By Lemma 3(c) there exist $-K = t_1 < t_2 < \cdots < t_m = K$ such that

$$\bigcap_{|t|\leq K} B_n(t,0) \supseteq \bigcap_{j=1}^{m} B_n(t_j, \delta_0/2), \tag{25}$$

for all n. By (25) and Bernstein's inequality, for each $j = 1,\dots,m$ and for all $F \in \mathscr{F}_\varepsilon$,

$$\begin{aligned} P_F\left\{B_n^c\left(t_j, \frac{\delta_0}{2}\right)\right\} &\leq P_F\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\chi\left[\frac{X_i - t_j}{S(F,t_j)-\delta}\right] - E_F\chi\left[\frac{X-t_j}{S(F,t_j)-\delta}\right]\right| > \frac{\delta_0}{2}\right\} \\ &\leq e^{-(n\delta_0^2/12)}. \end{aligned} \tag{26}$$

By (25) and (26),

$$\begin{aligned} P_F\left\{\inf_{|t|\leq K}[S(F_n,t) - S(F,t)] > -\delta\right\} &\geq P_F\left\{\bigcap_{|t|\leq K} B_n(t,0)\right\} \\ &\geq 1 - \sum_{j=1}^{m} P_F\{B_n^c(t_j,\delta_0/2)\} \\ &\geq 1 - me^{-n\delta_0^2/12}, \end{aligned}$$

for all $F \in \mathscr{F}_\varepsilon$. Analogously, we can show that

$$P_F\Big\{\sup_{|t| \le K}[S(F_n,t) - S(F,t)] < \delta\Big\} \ge 1 - me^{-n\delta_0^2/12}, \qquad \forall\, F \in \mathscr{F}_\varepsilon.$$

Therefore,

$$\sum_{n=1}^{\infty} P_F\Big\{\sup_{|t| \le K}|S(F_n,t) - S(F,t)| > \delta\Big\} \le \frac{2m}{1 - e^{-\delta_0^2/12}}, \qquad \forall\, F \in \mathscr{F}_\varepsilon$$

and (a) follows by the Borel–Cantelli lemma. To prove (b) first notice that, since $b(\chi) < (1-\varepsilon)$, there exists $K_1 > 0$ such that for all $|t| > K_1$,

$$E_F\chi\left[\frac{X - t}{S(F,0)}\right] \ge E_F\chi\left(\frac{X-t}{s_1}\right) \ge (1-\varepsilon)E_{F_0}\chi\left(\frac{X-t}{s_1}\right) > b(\chi),$$

where s_1 is as in Lemma 1. Notice that by the dominated convergence theorem

$$\lim_{t \to \infty} E_{F_0}\chi[(X - t)/s_1] = 1.$$

Hence, $S(F,t) > S(F,0)$, $\forall\ |t| > K_1$, $\forall\ F \in \mathscr{F}_\varepsilon$ and so

$$S(F) = \inf_{t \in R} S(F,t) = \inf_{|t| \le K_1} S(F,t).$$

On the other hand, let K_3 and $\delta_1 > 0$ be such that $(1-\varepsilon)P_{F_0}(|X| \le K_3) > b(\chi) + \delta_1$. Observe now that

$$\lim_{K_2 \to \infty}\ \inf_{|t| > K_2} \chi\left(\frac{x-t}{s_1 + \delta}\right) I(|x| < K_3) = I(|x| < K_3), \qquad \forall\, x \in R,$$

where $I(|x| < K_3) = 1$ if $|x| < K_3$ and equal to zero otherwise. Hence, by the dominated convergence theorem

$$\lim_{K_2 \to \infty} E_{F_0}\left\{\inf_{|t| > K_2} \chi\left(\frac{x-t}{s_1+\delta}\right) I(|X| < K_3)\right\} = (1-\varepsilon)P_{F_0}(|X| \le K_3) > b(\chi) + \delta_1.$$

Therefore, there exist $K_2 \ge K_1$ such that

$$\begin{aligned} & E_F\left\{\inf_{|t| > K_2} \xi\left[\frac{X-t}{S(F,0)+\delta}\right]\right\} \\ \text{(27)} \qquad & \ge (1-\varepsilon)E_{F_0}\left\{\inf_{|t| > K_2} \chi\left[\frac{X-t}{S(F,0)+\delta}\right] I(|X| \le K_3)\right\} \\ & \ge b(\chi) + \delta_1. \end{aligned}$$

Let $\delta_2 = \min\{\delta_0^2, \delta_1^2\}/12$. By (a), (27) and the Bernstein inequality,

$$\begin{aligned}
P_F&\Big\{S(F_n) = \inf_{|t|\le K_2} S(F_n,t)\Big\}\\
&\ge P_F\Big\{\inf_{|t|>K_2} S(F_n,t) > S(F_n,0)\Big\}\\
&\ge P_F\Big\{\inf_{|t|>K_2} \frac{1}{n}\sum\chi\Big[\frac{X_i-t}{S(F,0)+\delta}\Big] > b(\chi), |S(F_n,0)-S(F,0)|<\delta\Big\}\\
&\ge 1 - P_F\Big\{\frac{1}{n}\sum \inf_{|t|>K_2}\chi\Big[\frac{X_i-t}{S(F,0)+\delta}\Big] \le b(\chi)\Big\}\\
&\quad - P_F\{|S(F_n,0)-S(F,0)|\ge\delta\}\\
&\ge 1-2e^{-n\delta_2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
P_F\{|S(F_n)-S(F)|<\delta\} &\ge P_F\Big\{\sup_{|t|\le K_2}|S(F_n,t)-S(F,t)|<\delta\Big\}\\
&\quad + P_F\Big\{S(F_n) \ne \inf_{|t|\le K_2} S(F,t)\Big\}\\
&\ge 1-e^{-n\gamma}, \qquad \forall F\in\mathscr{F}_\varepsilon,
\end{aligned}$$

for some $\gamma > 0$ and (b) follows. Since

$$\begin{aligned}
P_F&\{|S[F_n,T(F_n)] - S[F,T(F)] > \delta\}\\
&\le P_F\Big\{\sup_{|t|\le 2K}|S(F_n,t)-S(F,t)| > \frac{\delta}{2}\Big\}\\
&\quad + P_F\Big\{|S[F,T(F_n)] - S[F,T(F)]| > \frac{\delta}{2}\Big\},
\end{aligned}$$

(c) follows from (a) and Lemma 3(c).

Finally, (d) follows by noticing that, under the given assumptions, all the statements made in the proof of (a), (b) and (c) hold uniformly for all χ in $\mathscr{C}$. □

Proof of Lemma 2. Since the median minimizes the maximum asymptotic bias among location equivariant estimates [Huber (1964)] and since t_0 and t_1 are the maximum asymptotic biases of the median and a location equivariant estimate, $t_0 \le t_1$. Thus, $t_1 = m(t_0) \le m(t_1) = t_2$ and in general, $t_n \le t_{n+1}$. Let $t^{**} = \lim_{n\to\infty} t_n$. Since $t^{**} = \lim_{n\to\infty} t_{n+1} = \lim_{n\to\infty} m(t_n) = m(\lim_{n\to\infty} t_n) = m(t^{**})$, we have $t^{**} \ge t^*$. On the other hand, if t satisfies $t = m(t) > t_0$, then $t = m(t) \ge t_n$ for all n. Therefore $t \ge t^{**}$ and so $t^{**} \le t^*$. The second part of (a) follows directly from the continuity of $\gamma(t)$. To prove (b) 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 $t_{n+1} = m(t_n)$, starting from the median. □

For each $b \in (\varepsilon, 1 - \varepsilon)$, let $\mathbf{C}_b$ be the class of χ-functions satisfying Assumption 1 and $b(\chi) = b$. Also let $\mathbf{C}$ be the class of functions satisfying Assumptions 1 and 2, that is, $\mathbf{C} = \bigcup_{\varepsilon < b < 1-\varepsilon} \mathbf{C}_b$. The following lemma is needed to prove Theorem 3.

LEMMA 4. *Fix $b \in (\varepsilon, 1 - \varepsilon)$ and let $a = F_0^{-1}[1 - (b/2)]$. Under the assumptions of Theorem 3 we have*:

(a) $S^-(\chi_a) \geq S^-(\chi)$ *for all* $\chi \in \mathbf{C}_b$.
(b) $g_\chi(s^*, t_0) \geq g_{\chi_a}(s^*, t_0)$ *for all* $\chi \in \mathbf{C}_b$.

PROOF. Part (a) follows directly from (16) and Lemma A3 in Martin and Zamar (1989). To prove (b), notice that for all $\chi \in \mathbf{C}_b$ we have $\int_{-a}^{a} \chi(x) f_0(x)\, dx = 2\int_a^\infty [1 - \chi(x)] f_0(x)\, dx$. Thus,

$$s^* \int_{-a}^{a} \chi(x)[f_0(s^*x - t_0) + f_0(s^*x + t_0)]\, dx$$

$$= s^* \int_{-a}^{a} \chi(x) f_0(x) k_0(x)\, dx \geq s^* k_0(a) \int_{-a}^{a} \chi(x) f_0(x)\, dx$$

$$= 2s^* k_0(a) \int_a^\infty [1 - \chi(x)] f_0(x)\, dx$$

$$= s^* k_0(a) \left[\int_{-\infty}^{-a} [1 - \chi(x)] f_0(x)\, dx + \int_a^\infty [1 - \chi(x)] f_0(x)\, dx \right]$$

$$\geq s^* \int_{-\infty}^{-a} [1 - \chi(x)] k_0(x)\, dx + s^* \int_a^\infty [1 - \chi(x)] k_0(x) f_0(x)\, dx$$

and (b) follows. □

PROOF OF THEOREM 3. First of all notice that since $S^+(a)$ and $S^-(a)$ are increasing at a^* and L_1 and L_2 are strictly monotone, we have $L_1[S^+(a)] = L_2[S^-(a)] = \overline{B}(a^*)$. Let $\chi \in \mathbf{C}$ be fixed and set $b = b(\chi)$. Let $a = F_0^{-1}[1 - (b/2)]$ so that $b(\chi_a) = b$. If $g_\chi(s^*, t_0) \geq (b - \varepsilon)/(1 - \varepsilon)$, then $S^+(\chi) \geq s^*$. So $\overline{B}(\chi) \geq L_2[S^+(\chi)] \geq L_2(s^*) = \overline{B}(a^*)$. On the other hand, suppose that $g_\chi(s^*, t_0) < (b - \varepsilon)/(1 - \varepsilon)$, that is, $S^+(\chi) < s^*$. Since $\chi \in \mathbf{C}_b$, by Lemma 4(b) we have $g_{\chi_a}(s^*, t_0) \leq g_\chi(s^*, t_0) < (b - \varepsilon)/(1 - \varepsilon)$. Hence $S^+(a) < s^*$, too. In view of the optimality of χ_{a^*} among jump functions we have $\overline{B}(a) \geq \overline{B}(a^*)$ and so $L_1[S^-(a)] \geq L_1[S^-(a^*)]$. For the particular b in question, by Lemma 4(a), $S^-(\chi_a) \geq S^-(\chi)$. Therefore,

$$\overline{B}(\chi) \geq L_1[S^-(\chi)] \geq L_1[S^-(a)] \geq L_1[S^-(a^*)] = \overline{B}(a^*),$$

and the theorem follows. □

PROOF OF THEOREM 4. Let $F_\infty = (1-\varepsilon)F_0 + \varepsilon\delta_\infty$, $t_\infty = T(F_\infty)$ and $s_\infty = S_\chi(F_\infty)$. First notice that

$$h^{-1}[(b-\varepsilon)/(1-\varepsilon)] = \sup_{F\in\mathscr{F}_\varepsilon} S_\chi(F,0) = S_\chi(F_\infty,0), \tag{28}$$

where $S_\chi(F,0)$ is the S-scale functional based on χ and the *true location* 0. By definition of the S-estimate of scale, for all $F \in \mathscr{F}_\varepsilon$, $S_\chi(F) = \inf_t S_\chi(F,t) \le \inf_t S_\chi(F_\infty, t) = S_\chi(F_\infty, 0)$, and so $S^+(\chi) \le S_\chi(F_\infty, 0)$. Assume first that $s_\infty < \infty$ and so $t_\infty < \infty$. By monotonicity of $g_\chi(s,t)$, $b = E_{F_\infty}\chi[(X-t_\infty)/s_\infty] = (1-\varepsilon)g_\chi(s_\infty, t_\infty) + \varepsilon \ge (1-\varepsilon)g_\chi(s_\infty, 0) + \varepsilon = E_{F_\infty}\chi(X/s_\infty)$. Therefore,

$$S_\chi(F_\infty, 0) \le s_\infty \le S^+(\chi). \tag{29}$$

Observe that, if $s_\infty = \infty$, then (29) trivially holds. Now, (28) and (29) imply that $S^+(\chi) \le \gamma_\chi^{-1}[(b-\varepsilon)/(1-\varepsilon)] \le S^+(\chi, T)$ and (a) follows by taking $T(F) = \arg\min S_\chi(F,t)$.

To prove (b) write $\gamma_\chi^{-1}[(b/(1-\varepsilon)] = \inf_{F\in\mathscr{F}_\varepsilon} S_\chi(F,0) = S_\chi(F_0^*, 0)$, where $F_0^* = (1-\varepsilon)F_0 + \varepsilon\delta_0$. For all $F \in \mathscr{F}_\varepsilon$, $t \in R$ and $s > 0$, $E_F\chi[(X-t)/s] \ge (1-\varepsilon)E_{F_0}\chi[(X-t)/s] \ge (1-\varepsilon)E_{F_0}\chi(X/s) = E_{F_0^*}\chi(X/s)$. Therefore, for all M-estimates of scale based on the given χ and for all T satisfying the assumptions of this theorem we have, $S^-(\chi, T) = \gamma_\chi^{-1}[(b/(1-\varepsilon)]$. □

PROOF OF THEOREM 5. Follows directly from Theorem 4 and Theorem 2 in Martin and Zamar (1989). □

PROOF THEOREM 6. Let $0 < a < \infty$. It suffices to show that, for each $\Delta > 0$,

$$\lim_{m\to\infty} \sup_{F\in\mathscr{F}_\varepsilon} P_F\left\{ \sup_{n\ge m} S_{\chi_a}(F_n) \ge S^+(a) + \Delta \right\} = 0 \tag{30}$$

and

$$\lim_{m\to\infty} \sup_{F\in\mathscr{F}_\varepsilon} P_F\left\{ \inf_{n\le m} S_{\chi_a}(F_n) \le S^-(a) - \Delta \right\} = 0. \tag{31}$$

For each $\delta > 0$ the approximating (even) function $\bar\rho_\delta(x)$ is defined as

$$\bar\rho_\delta(x) = \begin{cases} 0, & \text{if } 0 \le x \le a-\delta, \\ 1 - [(a-x)/\delta], & \text{if } a-\delta \le x \le a, \\ 1, & \text{if } x \ge a. \end{cases} \tag{32}$$

Notice that $\bar\rho_\delta(x)$ is continuous and that $\bar\rho_\delta(x) \ge \chi_a(x)$ for all x. For each $t \in R$ and all F, let

$$\bar S_\delta(F,t) = \sup\{s\colon E_F\bar\rho_\delta[(x-t)/s] \ge b(\chi_a)\}. \tag{33}$$

Clearly, for all t and all F (including the empirical c.d.f. F_n) we have $\bar S_{\chi_a}(F,t) \le \bar S_\delta(F,t)$ and so

$$S_{\chi_a}(F) \le \bar S_\delta(F) = \inf_t \bar S_\delta(F,t). \tag{34}$$

It is not difficult to verify that, for all given $\Delta > 0$, there exists $\delta_0 > 0$ such that

$$\bar{S}^+_{\delta_0} = \sup_{F \in \mathscr{F}_\varepsilon} \bar{S}_{\delta_0}(F) \le S^+(a) + (\Delta/2). \tag{35}$$

By Theorem 1,

$$\lim_{m \to \infty} \sup_{F \in \mathscr{F}_\varepsilon} P_F\left\{ \sup_{n \ge m} \bar{S}_{\delta_0}(F_n) > \bar{S}^+_{\delta_0} \ge S^+(a) + (\Delta/2) \right\} = 0. \tag{36}$$

Now (30) follows from (33), (34) and (35). Finally, (31) can be proved in a similar way, using the approximating function

$$\rho_\delta(x) = \begin{cases} 0, & \text{if } 0 \le x \le a, \\ (x-a)/\delta, & \text{if } a \le x \le a + \delta, \\ 1, & \text{if } x \ge a + \delta, \end{cases} \tag{37}$$

and the approximating scale functional

$$S_\delta(F, t) = \inf\{s \colon E_F \rho_\delta[(x-t)/s] \le b(\chi_a)\}. \tag{38}$$ □

Acknowledgments. We wish to thank Ricardo Fraiman for some useful comments for establishing the uniform consistency of M-estimates of scale, Rick White for helping us with the Monte Carlo simulations and the referees for a number of very helpful comments and suggestions.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton Univ. Press.

DONOHO, D. L. and LIU, R. C. (1988). The automatic robustness of minimum distance functionals. *Ann. Statist.* **16** 552–586.

GRÜBEL, R. (1988). The length of the shorth. *Ann. Statist.* **16** 619–628.

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Univ. California, Berkeley.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

LAX, D. A. (1985). Robust estimators of scale: finite sample performance in long-tailed symmetric distributions. *J. Amer. Statist. Assoc.* **80** 736–741.

MARTIN, R. D., YOHAI, V. J. and ZAMAR, R. H. (1989). Min-max bias robust regression. *Ann. Statist.* **17** 1608–1630.

MARTIN, R. D. and ZAMAR, R. H. (1987). Minimax bias robust M-estimates of scale. Technical Report 72, Dept. Statistics, Univ. Washington, Seattle.

MARTIN, R. D. and ZAMAR, R. H. (1989). Asymptotically min-max bias robust M-estimates of scale for positive random variables. *J. Amer. Statist. Assoc.* **84** 494–501.

ROUSSEEUW, P. J. and LEROY, A. M. (1988). A robust scale estimator based on the shortest half. *Statist. Neerlandica* **42** 103–116.

ROUSSEEUW, P. J. and YOHAI, V. J. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 265–272. Springer, New York.

SIMONOFF, J. S. (1987). Outlier detection and robust estimation of scale. *J. Statist. Comput. Simulation* **27** 79–92.

YOHAI, V. J. and ZAMAR, R. H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83** 406–413.

DEPARTMENT OF STATISTICS, GN-22
ROOM B-314
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
2021 WEST MALL
VANCOUVER, BRITISH COLUMBIA
CANADA V6T 1W5