The Annals of Statistics
1989, Vol. 17, No. 4, 1608–1630

MIN-MAX BIAS ROBUST REGRESSION

By R. D. Martin,[1] V. J. Yohai[1,2] and R. H. Zamar[1,3]

University of Washington, University of Buenos Aires and University of British Columbia

This paper considers the problem of minimizing the maximum asymptotic bias of regression estimates over ε-contamination neighborhoods for the joint distribution of the response and carriers. Two classes of estimates are treated: (i) M-estimates with bounded function ρ applied to the scaled residuals, using a very general class of scale estimates, and (ii) bounded influence function type generalized M-estimates. Estimates in the first class are obtained as the solution of a minimization problem, while estimates in the second class are specified by an estimating equation. The first class of M-estimates is sufficiently general to include both Huber Proposal 2 simultaneous estimates of regression coefficients and residuals scale, and Rousseeuw–Yohai S-estimates of regression. It is shown than an S-estimate based on a jump-function type ρ solves the min-max bias problem for the class of M-estimates with very general scale. This estimate is obtained by the minimization of the α-quantile of the squared residuals, where $\alpha = \alpha(\varepsilon)$ depends on the fraction of contamination ε. When $\varepsilon \to 0.5$, $\alpha(\varepsilon) \to 0.5$ and the min-max estimator approaches the least median of squared residuals estimator introduced by Rousseeuw. For the bounded influence class of GM-estimates, it is shown the "sign" type nonlinearity yields the min-max estimate. This estimate coincides with the minimum gross-error sensitivity GM-estimate. For $p = 1$, the optimal GM-estimate is optimal among the class of all equivariant regression estimates. The min-max S-estimator has a breakdown point which is independent of the number of carriers p and tends to 0.5 as ε increases to 0.5, but has a slow rate of convergence. The min-max GM-estimate has the usual rate of convergence, but a breakdown point which decreases to zero with increasing p. Finally, we compare the min-max biases for both types of estimates, for the case where the nominal model is multivariate normal.

1. Introduction. In spite of the considerable existing literature on robustness, there is relatively little published work on *global* robustness. Huber's (1964) min-max variance approach is based on neighborhoods which are not global by virtue of excluding asymmetric distributions. The shrinking neighborhood approach introduced by Jaeckel (1971) and used also by Bickel (1984) and Beran (1977a, b), among others, attempts to deal with asymmetry by putting bias on the same asymptotic footing as variance. But the shrinking neighborhood approach could hardly be called global. Approaches based on the influence curve, such as optimal bounded influence regression [Hampel (1974), Krasker (1980), Krasker and Welsch (1982) and Huber (1983)] inherit the local or infinitesimal aspect of the influence curve itself.

Received November 1987; revised August 1988.
[1]Research supported by ONR Contract N00014-84-C-0169.
[2]Research supported by a J. S. Guggenheim Memorial Foundation fellowship.
[3]Research supported by NSERC Operating Grant A9276.
AMS 1980 *subject classifications*. Primary 62J02; secondary 62J05.
Key words and phrases. Min-max bias, regression, robust estimates.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

An important measure of the global robustness of an estimate is given by its maximum asymptotic bias over an ε-contamination neighborhood of the target model. Naturally, such quantity will increase with ε and eventually will become infinite. The smallest values of ε, ε^*, for which the maximum asymptotic bias is infinite is called the breakdown point of the estimate. More formal definitions of these concepts are given in Section 2.1.

It seems that the main global approach to robustness in recent years has been centered around the construction of high breakdown point estimates, particularly for multivariate problems where this approach presents real challenges. See, for example, Donoho (1982), Donoho and Huber (1983), Stahel (1981), Rousseeuw (1984), Rousseeuw and Yohai (1984), Yohai (1987) and Yohai and Zamar (1988). In the last two papers, the authors construct regression estimators which have both high breakdown points and high efficiency.

The breakdown point approach is highly attractive for a number of reasons, not the least of which is the transparency of the concept and the ease with which it can be communicated to applied statisticians and scientists. On the other hand, one nonetheless wishes to have global optimality theory of robustness which emphasizes bias control for fractions of contamination smaller than the breakdown point. Furthermore, bias is itself a very transparent concept.

Along these lines we recall that Huber (1964) established the following result in his by now classic paper: The sample median minimizes the maximum asymptotic bias among all translation equivariant estimators of location, the maximum being over epsilon contaminated distributions (and also Lévy neighborhoods). It seems that this approach to global robustness, namely the construction of min-max bias robust estimators has been essentially neglected until quite recently, and this problem is quite clearly articulated in Hampel, Ronchetti, Rousseeuw and Stahel (1986) (see lower left entry of Table 2, page 176). Among the recent work in this area, we know of the following: Donoho and Liu (1988), who establish attractive bias robustness properties of minimum distance estimators, Martin and Zamar (1987a), who obtain min-max bias robust estimates of scale, and Martin and Zamar (1987b), who construct min-max bias robust estimates of location, subject to an efficiency constraint at the nominal model. See also, Zamar (1985) for min-max bias orthogonal regression M-estimates.

In this paper, we construct min-max bias robust regression estimates for two different classes of estimates: (i) M-estimates based on bounded ρ functions and general scale (i.e., general scale estimate for residuals) and (ii) GM-estimates having bounded influence curves. In the first case, the estimates are defined by a minimization problem, whereas in the second case the estimates are defined by an estimating equation.

It turns out that *S-estimators* introduced by Rousseeuw and Yohai (1984) can be regarded as special cases of M-estimates with general scale, as can Huber Proposal 2 M-estimates for regression and residuals scale. In fact, our min-max bias M-estimate is just that, an S-estimate.

The paper is organized in the following way. Section 2 introduces the ε-contaminated model for regression, M-estimates of scale based on bounded, symmetric functions ρ and the related S-estimates for regression. Section 3 establishes an expression for the maximum bias of an S-estimate. We also display the special form this expression takes for nominal multivariate normal models and also the special form obtained for jump functions ρ_c, which take on the values 0–1, with jumps at $\pm c$. Section 4 introduces the class of M-estimates with general scale, constructs a lower bound A_ρ for the maximum bias for fixed ρ and a lower bound A^* for A_ρ as ρ ranges over a broad class of loss functions. It is then shown that an S-estimate achieves A^*. Section 5 constructs min-max bias GM-estimates. These estimates are based on a "sign" function type nonlinearity in the estimating equations, which corresponds to a weighted L_1 regression, with weights inversely proportional to the norm of the vector of carriers. Throughout Sections 2–5, we have, for simplicity, considered the case where the intercept is known. In Section 6 we indicate how our results may be extended to the case when the intercept is unknown and must be estimated along with the slope parameters. Finally, Section 7 provides a comparison of the biases of min-max S-estimates and GM-estimates for the case where the nominal model is multivariate normal.

2. General setup and S-estimates.

2.1. *The target model and maximum asymptotic bias.* We assume the target model is the linear model

$$y = \mathbf{x}'\boldsymbol{\theta}_0 + u,$$

where $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ is a random vector in $\mathbb{R}^p$, $\boldsymbol{\theta}_0 = (\theta_{10}, \dots, \theta_{p0})'$ are the true regression parameters and the error u is a random variable independent of $\mathbf{x}$. Let F_0 be the nominal distribution function of u and G_0 the nominal distribution function of $\mathbf{x}$. Then the nominal distribution function H_0 of $(y, \mathbf{x})$ is

$$H_0(y, \mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} F_0(y - \boldsymbol{\theta}_0'\mathbf{s})\, dG_0(\mathbf{s}). \tag{2.1}$$

We assume that G_0 is elliptical about the origin, with scatter matrix $\mathbf{A}$. Correspondingly, we work with zero intercept until Section 6, which discusses how our results can be extended to deal with an intercept.

Let $\mathbf{T}$ be an $\mathbb{R}^p$ valued functional defined on a ("large") subset of the space of distribution functions H on $\mathbb{R}^{p+1}$. This subset is assumed to include all empirical distribution functions H_n corresponding to a sample $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ of size n from H. Then $\mathbf{T}_n = \mathbf{T}(H_n)$ is an estimate of $\boldsymbol{\theta}_0$.

It is further assumed that $\mathbf{T}$ is regression equivariant, i.e., if $\tilde{y} = y + \mathbf{x}'\mathbf{b}$ and $\tilde{\mathbf{x}} = \mathbf{C}^T\mathbf{x}$ for some full rank $p \times p$ matrix $\mathbf{C}$, then $\mathbf{T}(\tilde{H}) = \mathbf{C}^{-1}[\mathbf{T}(H) + \mathbf{b}]$, where $\tilde{H}$ is the distribution of $(\tilde{y}, \tilde{\mathbf{x}})$. Correspondingly, the transformed model parameter is $\tilde{\boldsymbol{\theta}}_0 = \mathbf{C}^{-1}[\boldsymbol{\theta}_0 + \mathbf{b}]$.

We define the asymptotic bias $b_A = b_A(\mathbf{T}, H)$ of $\mathbf{T}$ at H so that it is invariant under regression equivariant transformations,

$$b_A(\mathbf{T}, H) = (\mathbf{T}(H) - \boldsymbol{\theta}_0)'\mathbf{A}(\mathbf{T}(H) - \boldsymbol{\theta}_0). \tag{2.2}$$

Therefore, we can assume without loss of generality, that G_0 is spherical, i.e., $\mathbf{A}$ is the identity matrix, and that $\boldsymbol{\theta}_0 = \mathbf{0}$. Accordingly, the nominal model (2.1) becomes

$$H_0(y, \mathbf{x}) = \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_p} F_0(y)\, dG_0(\|s\|) \tag{2.3}$$

and, correspondingly, the asymptotic bias of $\mathbf{T}$ at H is given by the Euclidean norm of $\mathbf{T}$,

$$b(\mathbf{T}, H) = \|\mathbf{T}(H)\|. \tag{2.4}$$

If the operator $\mathbf{T}$ is continuous at H, then $\mathbf{T}(H)$ is the asymptotic value of the estimate when the underlying distribution of the sample is H. It is assumed that $\mathbf{T}$ is asymptotically unbiased at the nominal model H_0,

$$\mathbf{T}(H_0) = \mathbf{0}. \tag{2.5}$$

We will work the ε-contamination neighborhood of the fixed nominal distribution H_0,

$$V_\varepsilon = \{H\colon H = (1-\varepsilon)H_0 + \varepsilon H^*\}, \tag{2.6}$$

where H^* is any arbitrary distribution on $\mathbb{R}^{p+1}$. The maximum asymptotic bias of $\mathbf{T}$ over V_ε is

$$B_\varepsilon(\mathbf{T}) = \sup\{\|\mathbf{T}(H)\|\colon H \in V_\varepsilon\}. \tag{2.7}$$

Finally the asymptotic breakdown point of $\mathbf{T}$ [see Hampel (1971)] is defined as

$$\varepsilon^* = \inf\{\varepsilon\colon B_\varepsilon(\mathbf{T}) = \infty\}.$$

2.2. *M-estimates of scale.* Let ρ be a real-valued function on $\mathbb{R}^1$ satisfying the following assumptions:

A1. (i) Symmetric and nondecreasing on $[0, \infty)$, with $\rho(0) = 0$.
(ii) Bounded, with $\lim_{u\to\infty} \rho(u) = 1$.
(iii) ρ has only a finite number of discontinuities.

Let $0 < b < 1$. Then given a distribution function F, the M-scale functional is defined [see Huber (1981)] as

$$s(F) = \inf\left\{s > 0\colon E_F\rho\left(\frac{u}{s}\right) \le b\right\}. \tag{2.8}$$

Given a sample $\mathbf{u} = (u_1, \dots, u_n)$ from F, the corresponding M-estimate of scale is obtained from (2.8) by replacing F by the empirical distribution F_n of $\mathbf{u}$.

It is easy to prove that

$$s(F) > 0 \quad \text{iff} \quad P_F(u = 0) < 1 - b. \tag{2.9}$$

If this condition is satisfied with $s(F)$ finite and ρ is continuous, we can replace the inequality by equality in (2.8).

It can be shown that the *breakdown point* due to implosion, i.e., due to contamination at the origin which results in $s(F) = 0$, is $1 - b$ and the breakdown point due to explosion, i.e., due to contamination tending to infinity which results in $s(F) = \infty$, is b. The overall breakdown point is then $\varepsilon^* = \min\{b, 1 - b\}$. For details, see Huber (1981).

In the case where one is interested in estimating scale for its own sake, one usually forces consistency at a nominal model F_0 by setting $b = E_{F_0}\rho(u)$. This issue turns out to be irrelevant for our present purposes, since as we see in the next section, we will only be interested in obtaining a smallest M-estimate of scale with respect to the regression parameter θ in a particular parametrization of the scale functional. The choice of b will therefore remain at our disposal in obtaining a min-max bias regression estimate.

2.3. *S-estimators of regression for general H.* Let $(y, \mathbf{x}) \in \mathbb{R}^{p+1}$ be a random vector with arbitrary distribution function H, e.g., H could be the empirical distribution function for $(y, \mathbf{x})$. For any $\theta \in \mathbb{R}$ let F^θ be the distribution of the residuals

$$r(\theta) = y - \mathbf{x}'\theta. \tag{2.10}$$

Let $s(F)$ be any M-estimate of scale as defined in Section 2.2 and, to emphasize the independent roles of θ and H in determining F^θ, let $s(\theta, H) = s(F^\theta)$.

A functional $\mathbf{T}(H)$ is said to be an S-estimate functional of regression [see Rousseeuw and Yohai (1984)] if there exists a sequence $\theta_n \in \mathbb{R}^p$ such that

$$\lim_{n \to \infty} \theta_n = \mathbf{T}(H) \tag{2.11}$$

and

$$\lim_{n \to \infty} s(\theta_n, H) = \inf_{\theta \in \mathbb{R}^p} s(\theta, H). \tag{2.12}$$

With regard to the existence of such a sequence, we assert:

If ρ satisfies A1 and H satisfies

$$\sup_{\|\theta\| = 1} P_H(\mathbf{x}'\theta = 0) < 1 - b, \tag{2.13}$$

then there exist some sequence θ_n and $\mathbf{T}(H)$ satisfying (2.11) and (2.12).

This is a consequence of the following lemma.

LEMMA 2.1. *Suppose that* ρ *satisfies* A1 *and* H *satisfies* (2.13). *Then* $\|\theta_n\| \to \infty$ *implies* $\lim_{n \to \infty} s(\theta_n, H) = \infty$.

PROOF. Suppose that $\|\theta_n\| \to \infty$ and let $\theta_n^* = \theta_n / \|\theta_n\|$. Without loss of generality we can assume that $\theta_n^* \to \theta^*$ with $\|\theta^*\| = 1$. To prove the lemma it is enough to show that for all $s > 0$,

$$E_H\{\rho([y - \mathbf{x}'\theta_n]/s)\} > b,$$

for sufficiently large n. Indeed, we can write

$$E_H\{\rho([y - \mathbf{x}'\theta_n]/s)\} \ge E_H\{\rho([y - \|\theta_n\|\mathbf{x}'\theta_n^*]/s)\mathbf{I}_{(\mathbf{x}'\theta_n^* \neq 0)}\},$$

where $\mathbf{I}_A$ is the indicator of the set A. Since it is immediate to prove that $\rho((y - \|\theta_n\|\mathbf{x}'\theta_n^*)/s)\mathbf{I}_{(\mathbf{x}'\theta_n^* \neq 0)} \to \mathbf{I}_{(\mathbf{x}'\theta^* \neq 0)}$ a.s. H_0, the lemma follows from the dominated convergence theorem and (2.13). □

It is easy to prove that if A1 is satisfied and ρ is continuous, then (2.11) and (2.13) will imply

$$s(\mathbf{T}(H), H) = \min\{s(\theta, H) \colon \theta \in \mathbb{R}^p\}. \tag{2.14}$$

However, in general, (2.14) may not be true.

Observe that there may be more than one value $\mathbf{T}(H)$ satisfying (2.11) and (2.12). In that case, the choice of $\mathbf{T}(H)$ is arbitrary.

It is easy to verify that S-estimates of regression are regression equivariant, as defined in Section 2.1. Furthermore, under very general conditions, these estimates are consistent and asymptotically normal [see Yohai and Zamar (1988)].

3. Maximum bias of S-estimates.

3.1. *Maximum bias of general S-estimates.* Assume now the target model H_0 is given by (2.3). We will need the following assumptions.

A2. F_0 is absolutely continuous with density f_0 which is symmetric, continuous and strictly decreasing for $u \geq 0$.

A3. G_0 is spherical and $P_{G_0}(\mathbf{x}'\theta = 0) = 0\ \forall\ \theta \in \mathbb{R}^p$ with $\theta \neq 0$.

Under A3, it is easy to see that the distribution of $\mathbf{x}'\theta$ depends only on $\|\theta\|$. Thus we set

$$g(s, \|\theta\|) = E_{H_0}\rho\left(\frac{y - \mathbf{x}'\theta}{s}\right). \tag{3.1}$$

The following lemma is a key result.

Lemma 3.1. *Assume that ρ satisfies* A1, *F_0 satisfies* A2 *and G_0 satisfies* A3. *Then g is continuous, strictly increasing with respect to $\|\theta\|$ and strictly decreasing in s for $s > 0$.*

Proof. Continuity of g follows from A1(iii) and A2: Since ρ is continuous a.s. $[F_0]$, the expectation of $\rho(y - \mathbf{x}'\theta)$ with respect to F_0 is a continuous function of $\mathbf{x}'\theta$ [see for example Billingsley (1968), page 181]. Let $s_2 > s_1$. Since

$$\rho((y - \mathbf{x}'\theta)/s_1) \geq \rho((y - \mathbf{x}'\theta)/s_2) \text{ a.s. } [H_0],$$

we have

$$E_{H_0}\rho((y - \mathbf{x}'\theta)/s_1) \geq E_{H_0}\rho((y - \mathbf{x}'\theta)/s_2).$$

In addition, we have strict inequality unless $(y - \mathbf{x}'\theta)/s_1 = (y - \mathbf{x}'\theta)/s_2$ a.s. $[H_0]$, that is unless $y - \mathbf{x}'\theta = 0$ a.s. $[H_0]$. The last is impossible because of independence of y and $\mathbf{x}$. By A3, the distribution of $\mathbf{x}'\mathbf{a}$ is the same for any unit vector $\mathbf{a}$. Thus the distribution of $\mathbf{x}'\boldsymbol{\theta}$ is the same as that of $\|\boldsymbol{\theta}\|z$, where z is a random variable distributed as $\mathbf{x}'\mathbf{a}$, $\|\mathbf{a}\| = 1$. Assume without loss of generality that $s = 1$ and let $t_2 > t_1 \geq 0$. Since y is symmetric about 0 and independent of z, the conditional expectation $\tilde{g}(t, z) = E[\rho(y - tz)|z]$ is a nondecreasing function of $|t|$. Hence

$$E\{\tilde{g}(t_2, z) - \tilde{g}(t_1, z)\} \geq 0$$

and equality holds only if $t_1 z = t_2 z$ a.s., that is only if $z = 0$ a.s. The last is impossible because of A3. □

From Lemma 3.1 it is immediate that an S-estimate $\mathbf{T}(H)$ is Fisher consistent at the target model H_0.

Let $g_1^{-1}(\cdot, \|\boldsymbol{\theta}\|)$ be the inverse of g with respect to s and $g_2^{-1}(s, \cdot)$ the inverse of g with respect to $\|\boldsymbol{\theta}\|$. The following theorem gives the maximum bias of an S-estimate.

THEOREM 3.1. *Under the same assumptions as in Lemma* 3.1, *the maximum bias* $B_\varepsilon(\mathbf{T})$ *of an S-estimate* $\mathbf{T}$ *over the contamination neighborhood* V_ε *is given by*

$$(3.2) \qquad B_\varepsilon(\mathbf{T}) = \begin{cases} g_2^{-1}\left(g_1^{-1}\left(\dfrac{b-\varepsilon}{1-\varepsilon}, 0\right), \dfrac{b}{1-\varepsilon}\right) & \text{if } \varepsilon < \min(b, 1-b), \\ \infty & \text{if } \varepsilon \geq \min(b, 1-b). \end{cases}$$

Therefore, the asymptotic breakdown point of $\mathbf{T}$ *is* $\varepsilon = \min(b, 1 - b)$.

PROOF. Let

$$c = g_2^{-1}\left(g_1^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}, 0\right), \frac{b}{1-\varepsilon}\right)$$

and suppose that $\varepsilon < \min(b, 1 - b)$. To prove that

$$(3.3) \qquad B_\varepsilon(\mathbf{T}) \leq c,$$

it is enough to show that for any H of the form $H = (1 - \varepsilon)H_0 + \varepsilon H^*$, $\|\boldsymbol{\theta}\| > c$ implies

$$(3.4) \qquad s(\boldsymbol{\theta}, H) > s(\mathbf{0}, H).$$

Put $s_1 = g_1^{-1}((b - \varepsilon)/(1 - \varepsilon), 0)$. Then by Lemma 3.1, $\|\boldsymbol{\theta}\| > c$ implies that

$$(3.5) \qquad g(s_1, \|\boldsymbol{\theta}\|) > \frac{b}{1-\varepsilon}.$$

Also by Lemma 3.1, there exists $s_2 > s_1$ such that

$$(3.6) \qquad g(s_2, \|\boldsymbol{\theta}\|) > \frac{b}{1-\varepsilon}.$$

Then

$$E_H\rho\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}}{s_2}\right) \geq (1-\varepsilon)g(s_2, \|\boldsymbol{\theta}\|) > b$$

and, therefore, by definition of $s(\boldsymbol{\theta}, H) = s(F^{\theta})$ [see (2.8)], we have

$$s_2 \leq s(\boldsymbol{\theta}, H). \tag{3.7}$$

On the other hand,

$$g(s_1, 0) = \frac{b-\varepsilon}{1-\varepsilon}. \tag{3.8}$$

Combining (3.8) and Lemma 3.1, we have for any $H = (1-\varepsilon)H_0 + \varepsilon H^*$ and any $s > s_1$,

$$E_H\rho\left(\frac{y}{s}\right) \leq (1-\varepsilon)g(s,0) + \varepsilon \leq (1-\varepsilon)g(s_1,0) + \varepsilon = b.$$

Therefore, $s \geq s(\mathbf{0}, H)$ for all $s > s_1$ and so

$$s_1 \geq s(\mathbf{0}, H). \tag{3.9}$$

Thus (3.4) follows from (3.7) and (3.9), and so (3.3) holds. Now we will prove that

$$B_\varepsilon(\mathbf{T}) \geq c. \tag{3.10}$$

Let c_1 be any positive number smaller than c and let $\|\boldsymbol{\theta}^*\| = c_1$. Let H_n^* be the distribution concentrated at the point mass $(y_n, \mathbf{x}_n)$ where $\mathbf{x}_n = \lambda_n \boldsymbol{\theta}^*$, $\lambda_n \to \infty$ and $y_n = \mathbf{x}_n'\boldsymbol{\theta}^*$. Set $H_n = (1-\varepsilon)H_0 + \varepsilon H_n^*$. In order to prove (3.10), it is enough to show that

$$\sup_n \|\mathbf{T}(H_n)\| \geq c_1. \tag{3.11}$$

Suppose (3.11) is not true. Then by passing to a subsequence, which we continue to label H_n, we have $\mathbf{T}(H_n) = \boldsymbol{\theta}_n$, with

$$\lim_{n\to\infty} \boldsymbol{\theta}_n = \tilde{\boldsymbol{\theta}} \tag{3.12}$$

and

$$\|\tilde{\boldsymbol{\theta}}\| < \|\boldsymbol{\theta}^*\| = c_1. \tag{3.13}$$

It follows that

$$\lim_{n\to\infty} |y_n - \mathbf{x}_n'\boldsymbol{\theta}| = \lim_{n\to\infty} \lambda_n\left(\|\boldsymbol{\theta}^*\|^2 - \boldsymbol{\theta}^{*\prime}\tilde{\boldsymbol{\theta}}\right) = \infty.$$

Then since

$$E_{H_n}\rho\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}_n}{s}\right) = (1-\varepsilon)g(s, \|\boldsymbol{\theta}_n\|) + \varepsilon\rho\left(\frac{y_n - \mathbf{x}_n'\boldsymbol{\theta}_n}{s}\right), \tag{3.14}$$

letting $s < s_1 = g_1^{-1}((b-\varepsilon)/(1-\varepsilon), 0)$ and using Lemma 3.1 gives

$$\begin{aligned}\lim_{n\to\infty} E_{H_n}\rho\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}_n}{s}\right) &\geq (1-\varepsilon)g(s,0) + \varepsilon > (1-\varepsilon)g(s_1,0) + \varepsilon \\ &= (1-\varepsilon)\frac{b-\varepsilon}{1-\varepsilon} + \varepsilon = b.\end{aligned}$$

This implies that

$$\lim_{n\to\infty} s(\theta_n, H_n) \geq s \qquad \forall\, s < s_1$$

and so we have

$$\lim_{n\to\infty} s(\theta_n, H_n) \geq s_1. \tag{3.15}$$

On the other hand,

$$(1-\varepsilon)g(s_1, c_1) < (1-\varepsilon)g(s_1, c) = b$$

and by Lemma 3.1 we can find $s_2 < s_1$ such that

$$(1-\varepsilon)g(s_2, c_1) < b.$$

This gives

$$E_{H_n}\rho\left(\frac{y - \mathbf{x}'\theta^*}{s_2}\right) = (1-\varepsilon)g(s_2, c_1) < b$$

and

$$s(\theta^*, H_n) \leq s_2. \tag{3.16}$$

Since (3.15) and (3.16) contradict the fact that $\mathbf{T}(H_n) = \theta_n$ minimizes $s(\cdot, H_n)$ for each n, we have established (3.10). In order to complete the proof it is enough to show that if $\varepsilon \uparrow \min(b, 1-b)$, then

$$g_2^{-1}\left(g_1^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}, 0\right)\frac{b}{1-\varepsilon}\right) \to \infty. \tag{3.17}$$

Let $b \leq 0.5$, so that $\min(b, 1-b) = b$. Then we have

$$\lim_{\varepsilon\uparrow b} g_1^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}, 0\right) = \lim_{\delta\to 0} g_1^{-1}(\delta, 0) = \infty$$

and so

$$\lim_{\varepsilon\uparrow b} g_2^{-1}\left(g_1^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}, 0\right), \frac{b}{1-\varepsilon}\right) = \lim_{s\uparrow\infty} g_2^{-1}\left(s, \frac{b}{1-b}\right) = \infty. \qquad \square$$

3.2. *Maximum bias of S-estimates for* $(y, \mathbf{x})$ *multivariate normal.* If $\mathbf{z} = (y, \mathbf{x}) \sim N(\mathbf{0}, \mathbf{I}_{p+1})$, then

$$g(s, \gamma) = h\left((1+\gamma^2)^{1/2}/s\right),$$

where $h(\lambda) = E\rho(\lambda u)$, with $u \sim N(0,1)$. Then

$$g_1^{-1}(t, \gamma) = \frac{(1+\gamma^2)^{1/2}}{h^{-1}(t)}$$

and

$$g_2^{-1}(s, t) = \left(\left[sh^{-1}(t)\right]^2 - 1\right)^{1/2}.$$

This gives the expression for squared bias,

$$B_\varepsilon^2(\mathbf{T}) = \left(\frac{h^{-1}(b/(1-\varepsilon))}{h^{-1}((b-\varepsilon)/(1-\varepsilon))}\right)^2 - 1. \tag{3.18}$$

3.3. *Maximum bias of S-estimates when ρ is a jump function.* Consider the special family of *jump functions* ρ_c (which satisfy A1),

$$\rho_c(u) = \begin{cases} 0 & \text{if } |u| < c, \\ 1 & \text{if } |u| \geq c. \end{cases} \tag{3.19}$$

Given a sample $\mathbf{u} = (u_1, \dots, u_n)$, the corresponding M-estimate of scale is given by

$$s_n(\mathbf{u}) = \frac{1}{c}|u|_{(n-[nb])},$$

where $|u|_{(1)}, \dots, |u|_{(n)}$ are the order statistics for the absolute values $|u_1|, \dots, |u_n|$.

For the choice ρ_c, the corresponding regression S-estimate minimizes the absolute value of the (approximate) $1 - b$ quantile of the absolute values $|y_i - \mathbf{x}_i'\boldsymbol{\theta}|$ of the residuals. Note that this regression S-estimate does not depend on the choice of c and so we henceforth set $c = 1$.

When $b = 0.5$, $s_n(\mathbf{u}) = |u|_{([n/2])}$ is the median absolute value (MAV) estimate of scale. The corresponding S-estimate is identical to Rousseeuw's (1984) least median of squared residuals (LMS) regression estimate. (Minimization with respect to $\boldsymbol{\theta}$ of a quantile of any monotone transformation of the absolute values $|y_i - \mathbf{x}_i'\boldsymbol{\theta}|$ results in the same estimate.)

The following lemma gives the maximum bias of an S-estimate when $\rho = \rho_1$.

LEMMA 3.2. *Let $\mathbf{T}_b$ be the S-estimate with jump function ρ_1 and right-hand side b. Assume F_0 satisfies* A2 *and G_0 satisfies* A3. *Then*

(i)
$$B_\varepsilon(\mathbf{T}_b) = G^{-1}_{F_0^{-1}(1-(b-\varepsilon)/[2(1-\varepsilon)])}\left(\frac{b}{1-\varepsilon}\right),$$

where

$$G_t(\|\boldsymbol{\theta}\|) = 1 - E_{G_0}F_0(t + \mathbf{x}'\boldsymbol{\theta}) + E_{G_0}F_0(-t + \mathbf{x}'\boldsymbol{\theta}). \tag{3.20}$$

(ii) *The following equation is valid*:

$$\inf_{\varepsilon < b < 1-\varepsilon} B_\varepsilon(\mathbf{T}_b) = \inf_{F_0^{-1}(1/[2(1-\varepsilon)]) < t < \infty} G_t^{-1}\left(2(1 - F_0(t)) + \frac{\varepsilon}{1-\varepsilon}\right). \tag{3.21}$$

PROOF. In this case we have

$$g(s, \|\boldsymbol{\theta}\|) = P(|y - \mathbf{x}'\boldsymbol{\theta}| \geq s) = G_s(\|\boldsymbol{\theta}\|)$$

and so

$$g_2^{-1}(s, \lambda) = G_s^{-1}(\lambda). \tag{3.22}$$

We also have

$$g(s,0) = 2(1 - F_0(s)). \tag{3.23}$$

Using (3.20) and (3.23) in (3.2) gives (i). The result (ii) is obtained by substituting $t = F_0^{-1}(1 - (b - \varepsilon)/[2(1 - \varepsilon)])$ in (i). $\square$

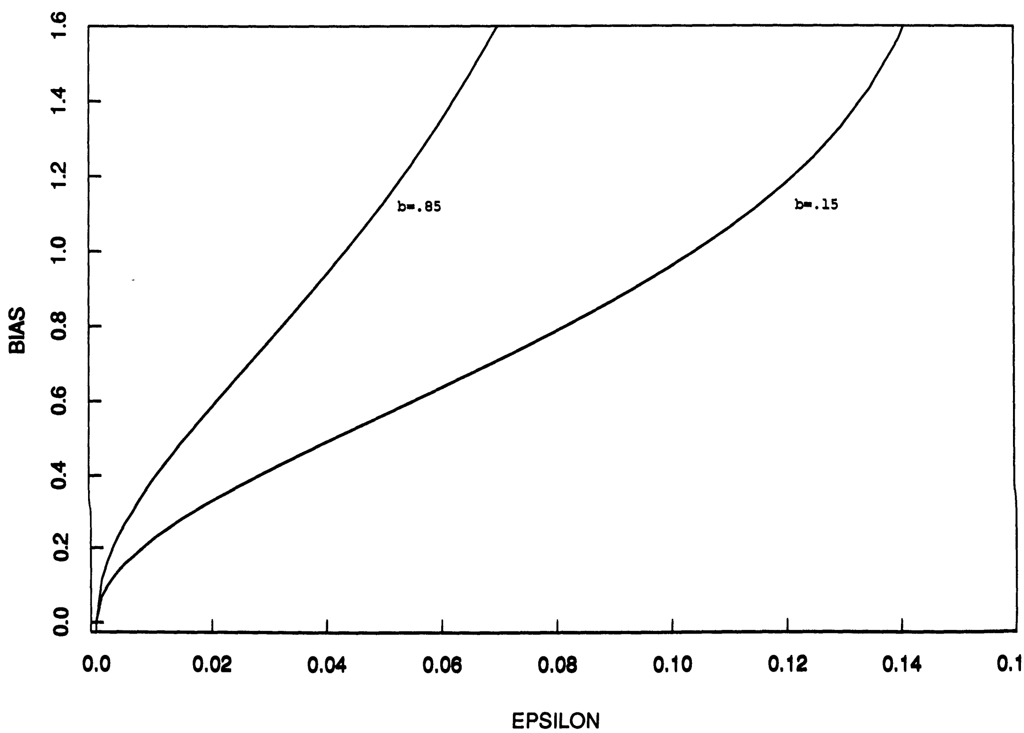

FIG. 1. *Maximal biases of T_b for $b = 0.85$ and $b = 0.15$, with corresponding breakdown point* $\varepsilon^* = 0.15$.

In the case that $(y, \mathbf{x})$ is multivariate normal, using (3.18) and the fact that for $\rho = \rho_c$,

$$h(\lambda) = 2\left(1 - \Phi\left(\frac{1}{\lambda}\right)\right),$$

we get

$$B_\varepsilon^2(\mathbf{T}_b) = \left(\frac{\Phi^{-1}(1 - (b - \varepsilon)/[2(1 - \varepsilon)])}{\Phi^{-1}(1 - b/[2(1 - \varepsilon)])}\right)^2 - 1, \tag{3.24}$$

where Φ is the $N(0,1)$ distribution function.

It is interesting to note from (3.2) that two distinct values of b give rise to any specified breakdown point $\varepsilon^* \in (0, 0.5)$, namely, $b = \varepsilon^*$ and $b = 1 - \varepsilon^*$. The estimates $\mathbf{T}_b$ for two such values of b have different maximal bias curves [i.e., plots of $B_\varepsilon(\mathbf{T}_b)$ versus ε], both of which explode at ε^*. In Figure 1 we display two such curves, with bias as a function of ε given by (3.24) for the values $b = 0.15$ and $b = 0.85$, which corresponds to a breakdown point $\varepsilon^* = 0.15$. The breakdown at $\varepsilon^* = 0.15$ is due to *implosion* for $b = 0.85$ and due to *explosion* for $b = 0.15$ (cf. comments in Section 2.2).

4. *M*-estimates with general scale.

4.1. *Definition of M-estimates with general scale.* Let ρ be a function satisfying A1 and let $s(H)$ be a (very) general estimate of the residuals scale. For example, the general scale functional $s(H)$ may be determined simultaneously with $\boldsymbol{\theta}$ or independently of $\boldsymbol{\theta}$. It is assumed that $s(H)$ is *regression invariant* (i.e., invariant under regression transformations $\tilde{y} = y + \mathbf{x}'\mathbf{b}$ and $\tilde{\mathbf{x}} = \mathbf{C}^T\mathbf{x}$) and *residuals scale equivariant* (i.e., equivariant under residuals scale change $\tilde{u} = au$). Furthermore, we will assume that $s(H)$ has a breakdown point greater than ε, namely,

A4.
$$s_1 = \inf\{s(H)\colon H = (1-\varepsilon)H_0 + \varepsilon H^*\} > 0,$$
$$s_2 = \sup\{s(H)\colon H = (1-\varepsilon)H_0 + \varepsilon H^*\} < \infty.$$

Then an M-estimator $\mathbf{T}(H)$ of regression, *with general scale,* is determined by solving the minimization problem

$$\inf_{\boldsymbol{\theta}} E_H \rho\left(\frac{y - \mathbf{x}'\boldsymbol{\theta}}{s(H)}\right). \tag{4.1}$$

Under the assumptions on $s(H)$, $\mathbf{T}(H)$ is clearly regression equivariant.

If the infimum in (4.1) is attained, then it defines $\mathbf{T}(H)$, with the choice of $\mathbf{T}(H)$ arbitrary in the case of nonuniqueness. If a value $\boldsymbol{\theta}$ which attains (4.1) does not exist, then $\mathbf{T}(H)$ is defined by

$$\mathbf{T}(H) = \lim_{n\to\infty} \boldsymbol{\theta}_n, \tag{4.2}$$

where $\boldsymbol{\theta}_n$ satisfy

$$\lim_{n\to\infty} E_H \rho\left(\frac{y - \mathbf{x}'\boldsymbol{\theta}_n}{s(H)}\right) = \inf_{\boldsymbol{\theta}\in\mathbb{R}^p} E_H \rho\left(\frac{y - \mathbf{x}'\boldsymbol{\theta}}{s(H)}\right). \tag{4.3}$$

Again, in the case of nonuniqueness, the choice of $\mathbf{T}(H)$ is arbitrary. It is easy to check that S-estimates are special types of M-estimates with general scale [see Rousseeuw and Yohai (1984)], as are Huber (1981) Proposal 2 simultaneous M-estimates of regression and scale.

4.2. *Lower bound for the minimax bias of M-estimators.* Let $g(s, \|\boldsymbol{\theta}\|)$ be as in (3.1) and put

$$A_\rho(s) = g_2^{-1}\left(s, g(s,0) + \frac{\varepsilon}{1-\varepsilon}\right), \tag{4.4}$$

$$A_\rho = \inf_{s\in[s_1,s_2]} A_\rho(s). \tag{4.5}$$

The following lemma shows that A_ρ is in fact a lower bound for the maximum bias over V_ε of an M-estimate with general scale.

LEMMA 4.1. *This proof follows closely the second part of the proof of Theorem* 3.1. *Let* $\mathbf{T}$ *be an M-estimate with general scale. Assume* ρ *satisfies* A1,

F_0 *satisfies* A2, G_0 *satisfies* A3 *and the scale* $s(H)$ *satisfies* A4. *Then*

$$B_\varepsilon(\mathbf{T}) \geq A_\rho.$$

PROOF. Let $B = B_\varepsilon(\mathbf{T})$, suppose that $B < A_\rho$ and take $\gamma > 0$ such that

$$B \leq A_\rho - \gamma. \tag{4.6}$$

Also take $\tilde{\theta}$ such that

$$A_\rho - \frac{\gamma}{2} \leq \|\tilde{\theta}\| \leq A_\rho - \frac{\gamma}{4}. \tag{4.7}$$

Let H_n^* be the distribution corresponding to a point mass at $(y_n, \mathbf{x}_n)$ where $y_n - \mathbf{x}_n'\tilde{\theta} = 0$ and $\mathbf{x}_n = \tilde{\theta}\lambda_n$ with $\lambda_n \to \infty$. Put $H_n = (1-\varepsilon)H_0 + \varepsilon H_n^*$ and

$$\theta_n^* = \mathbf{T}(H_n). \tag{4.8}$$

If θ_n^* is unbounded, (4.6) is contradicted and the theorem is proved. Assume θ_n^* is bounded and then we may also assume that $\theta_n^* \to \theta^*$. By A4 we may assume that $s_n = s(H_n) \to s > 0$. According to (4.6) we have

$$\|\theta'\| \leq A_\rho - \gamma. \tag{4.9}$$

Let

$$L_n(\theta) = E_{H_n}\rho\left(\frac{y - \mathbf{x}'\theta}{s_n}\right).$$

Then by Lemma 3.1 we have

$$L_n(\theta_n^*) \geq (1-\varepsilon)E_{H_0}\rho\left(\frac{\gamma}{s_n}\right) + \varepsilon\rho\left(\lambda_n \frac{\|\tilde{\theta}\|^2 - \theta_n^{*\prime}\tilde{\theta}}{s_n}\right).$$

Since (4.7) and (4.9) imply

$$\frac{\left|\lambda_n\left(\|\tilde{\theta}\|^2 - \theta_n^{*\prime}\tilde{\theta}\right)\right|}{s_n} \to \infty,$$

we have

$$\begin{aligned} \lim L_n(\theta_n^*) &\geq (1-\varepsilon)E_{H_0}\rho\left(\frac{y}{s}\right) + \varepsilon \\ &= (1-\varepsilon)g(s, \mathbf{0}) + \varepsilon. \end{aligned}$$

We also have

$$L_n(\tilde{\theta}) = (1-\varepsilon)E_{H_0}\rho\left(\frac{y - \mathbf{x}'\tilde{\theta}}{s_n}\right) = (1-\varepsilon)g\left(s_n, \tilde{\theta}\right)$$

and then by Lemma 3.1 we have

$$\lim L_n(\tilde{\theta}) \geq (1-\varepsilon)g(s, \tilde{\theta}).$$

Since $L_n(\theta_n^*) \leq L_n(\tilde{\theta})$ we also have

$$(1-\varepsilon)g(s, 0) + \varepsilon \leq (1-\varepsilon)g(s, \tilde{\theta}).$$

Therefore, by Lemma 3.1 we have

$$\|\tilde{\theta}\| \geq g_2^{-1}\left(s, g(s,0) + \frac{\varepsilon}{1-\varepsilon}\right)$$
$$= A_\rho(s) \geq A_\rho$$

and this contradicts (4.7). □

4.3. *Optimality of S-estimates with jump function* ρ. From now on it will be convenient to show explicitly that g depends on ρ and so we will write $g_\rho(s, \|\theta\|)$. For $t \in \mathbb{R}$ and $s > 0$ define

$$h_\rho(s,t) = E_{F_0}\rho\left(\frac{y-t}{s}\right).$$

We will need the following assumption.

A2*. F_0 has a density f_0 satisfying A2 and, for $t > 0$ and $y > 0$,

$$a(y) = \frac{f_0(y+t) + f_0(y-t)}{f_0(y)}$$

is a nondecreasing function of y.

A2* is satisfied, for example, in the important case where F_0 is the Gaussian distribution $N(0, \sigma^2)$. This follows because in the Gaussian case we have

$$a(y) = \frac{f_0(y+t) + f_0(y-t)}{f_0(y)} = 2e^{-t^2/\sigma^2}\cosh\left(\frac{ty}{\sigma}\right)$$

and

$$a'(y) = 2\frac{t}{\sigma}e^{-t^2/\sigma^2}\sinh\left(\frac{ty}{\sigma}\right) \geq 0 \quad \text{if } t > 0 \text{ and } y > 0.$$

A2* evidently holds in a number of other interesting situations: For example, it is easy to verify A2* when F_0 is double exponential.

The following lemmas will show that A_ρ is minimized when ρ is a jump function. This will enable us to compute the minimum of A_ρ.

LEMMA 4.2. *Assume* ρ *satisfies* A1 *and* F_0 *satisfies* A2*. *Let* $s > 0$ *and suppose that the jump function* ρ_c *satisfies*

$$h_{\rho_c}(s,0) = h_\rho(s,0). \tag{4.10}$$

Then

$$h_{\rho_c}(s,t) \geq h_\rho(s,t) \qquad \forall\, t \in \mathbb{R}.$$

PROOF.

$$h_{\rho_c}(s,t) - h_\rho(s,t) = -s\int_0^c \rho(y)(f_0(sy+t) + f_0(sy-t))\,dy$$
$$+s\int_c^\infty (1-\rho(y))(f_0(sy+t) + f_0(sy-t))\,dy$$
$$= -I_1 + I_2.$$

With

$$k = \frac{f_0(sc+t) + f_0(sc-t)}{f_0(sc)},$$

A2* gives

$$I_1 \leq sk \int_0^c \rho(y) f_0(sy)\, dy,$$

$$I_2 \geq sk \int_0^\infty \rho(y) f_0(sy)\, dy.$$

Thus (4.10) gives

$$h_{\rho_c}(s,\lambda) - h_\rho(s,\lambda) \geq k\left(-s\int_0^c \rho(y) f_0(sy)\, dy + s\int_c^\infty (1-\rho(y)) f_0(sy)\, dy\right) \geq 0.$$

□

LEMMA 4.3. *Assume* ρ *satisfies* A1, F_0 *satisfies* A2* *and* G_0 *satisfies* A3. *Then for any* $s > 0$ *there exists a jump function* ρ_c *such that*:

(i) $g_{\rho_c}(s,0) = g_\rho(s,0)$.
(ii) $g_{\rho_c}(s,t) \geq g_\rho(s,t)\ \forall\ t \in \mathbb{R}$.

PROOF. Follows from Lemma 4.2 conditioning on $\mathbf{x}$. □

LEMMA 4.4. *Assume* ρ *satisfies* A1, F_0 *satisfies* A2* *and* G_0 *satisfies* A3. *Then*:

(i) $A_\rho(s) \geq \inf_c A_{\rho_c}$.
(ii) $A_{\rho_c}(s) = G_{sc}^{-1}(2(1 - F_0(sc)) + \varepsilon/(1-\varepsilon))$,

where $G_t(\lambda)$ *is defined in* (3.20).

PROOF. (i) Follows immediately from Lemma 4.3.
(ii) Follows from the definition of $A_\rho(s)$, (3.22) and (3.23). □

The following theorem gives a lower bound for the maximum bias of an M-estimate for each fixed ε.

THEOREM 4.1. *Let* $\mathbf{T}$ *be an* M*-estimate and assume* A1, A2*, A3 *and* A4. *Then*

$$B_\varepsilon(\mathbf{T}) \geq \inf_{F_0^{-1}(1/[2(1-\varepsilon)]) < t < \infty} G_t^{-1}\left(2(1 - F_0(t)) + \frac{\varepsilon}{1-\varepsilon}\right). \tag{4.11}$$

PROOF. The theorem follows from Lemma 4.4 since $G_t^{-1}(2(1 - F_0(t)) + \varepsilon/(1-\varepsilon))$ is only defined when $2[1 - F_0(t)] + \varepsilon/(1-\varepsilon) < 1$ and this is equivalent to

$$t > F_0^{-1}(1/[2(1-\varepsilon)]).$$

□

TABLE 1
Min-max biases of optimal GM-estimates with estimated covariance matrix and optimal S-estimates

		GM-estimates		
p	$\varepsilon = 0.05$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$
1	0.083	0.18	0.28	0.41
2	0.11 (0.11)[a]	0.25 (0.23)	0.42 (0.37)	0.68 (0.55)
3	0.12 (0.11)	0.29 (0.25)	0.60 (0.44)	1.39 (0.70)
4	0.15 (0.14)	0.39 (0.31)	0.95 (0.52)	∞ (0.82)
5	0.19 (0.17)	0.49 (0.36)	2.85 (0.59)	∞ (1.00)
10	0.31 (0.23)	∞ (0.50)	∞ (0.97)	∞ (∞)
15	0.62 (0.29	∞ (0.68)	∞ (1.71)	∞ (∞)
		S-estimates		
$b(\varepsilon)$	0.33	0.34	0.35	0.36
S^*	0.49	0.77	1.05	1.37
LMS	0.53	0.83	1.07	1.52

[a] Numbers in parentheses are biases with covariance known (i.e., they correspond to points on the curves in Figure 2).

The following theorem shows that for a proper choice of b, which depends on ε, the S-estimate $\mathbf{T}_b$ as defined in Lemma 3.2, minimizes the maximum bias over the class of M-estimates with general scale.

THEOREM 4.2. *Assume that* F_0 *satisfies* A2* *and* G_0 *satisfies* A3. *Given* $\varepsilon > 0$, *there exists* $b = b(\varepsilon)$ *such that*

$$B_\varepsilon(\mathbf{T}_b) \leq B_\varepsilon(\mathbf{T})$$

for all M-estimates with general scale $\mathbf{T}$ *with* ρ*-function satisfying* A1 *and scale* s *satisfying* A4.

PROOF. Follows immediately from Lemma 3.2(ii) and Theorem 4.1. □

The value $b(\varepsilon)$ is the minimizing value of b in (3.21). For the Gaussian situation $b(\varepsilon)$ may be obtained by minimizing (3.24). For this case values of $b(\varepsilon)$ for $\varepsilon = 0.05, 0.1, 0.15, 0.2$ are given in Table 1.

5. *GM*-estimates.

5.1. *Characterizing the bias of GM-estimates.* We now consider *GM*-estimates of regression $\mathbf{T} = \mathbf{T}(H)$ obtained by solving

$$E_H \eta(y - \mathbf{x}'\boldsymbol{\theta}, \|\mathbf{x}\|) \frac{\mathbf{x}}{\|\mathbf{x}\|} = \mathbf{0} \tag{5.1}$$

for $\boldsymbol{\theta}$. The following assumptions will be used.

A5. $\eta(u, \nu)$ is:

(i) Continuous.
(ii) Odd and monotone nondecreasing in u.
(iii) Bounded, with $\sup_{u,\nu} \eta(u, \nu) = 1$.

Observe the optimal bounded influence estimates obtained by Krasker (1980) and Krasker and Welsch (1982) of this form with $\eta(u, \nu) = \psi_c(u\nu)$ in the Huber family

$$\psi_c(u) = \operatorname{sign}(u)\min(c, |u|). \tag{5.2}$$

The following lemma characterizes the possible biases of GM-estimates when $H \in V_\varepsilon$.

LEMMA 5.1. *Assume that* η *satisfies* A5 *and* F_0 *satisfies* A2. *Let* $\mathbf{T}(H)$ *be the GM-estimator defined by* (5.1). *Then there exists* $H_n = (1-\varepsilon)H_0 + \varepsilon H_n^*$ *such that* $\mathbf{T}(\tilde{H}_n) \to \tilde{\boldsymbol{\theta}}$ *if and only if*

$$\left\| E_{H_0}\eta(y - \mathbf{x}'\tilde{\boldsymbol{\theta}}, \|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}\right\| \le \frac{\varepsilon}{1-\varepsilon}. \tag{5.3}$$

PROOF. If there exists an $H \in V_\varepsilon$ such that $\mathbf{T}(H) = \tilde{\boldsymbol{\theta}}$, then (5.3) follows immediately from (5.1). Suppose now that (5.3) is satisfied with strict inequality. We will show that in this case there exists an $H \in V_\varepsilon(H_0)$ such that $\mathbf{T}(H) = \tilde{\boldsymbol{\theta}}$. In this case, there exist u, ν with $\nu > 0$ such that

$$\eta(u, \nu) = \|\mathbf{w}\|\frac{1-\varepsilon}{\varepsilon},$$

where $\mathbf{w} = E_{H_0}\eta(y - \mathbf{x}'\tilde{\boldsymbol{\theta}}, \|\mathbf{x}\|)\mathbf{x}/\|\mathbf{x}\|$. Take as H^* the distribution with point mass at $\mathbf{x}_0 = -\nu\|\mathbf{w}\|$, $y_0 = u + \mathbf{x}'\tilde{\boldsymbol{\theta}}_0$. Then if $H = (1-\varepsilon)H_0 + \varepsilon H^*$, we have

$$E_H\eta(y - \mathbf{x}'\boldsymbol{\theta}, \|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|} = (1-\varepsilon)\mathbf{w} + \varepsilon\|\mathbf{w}\|\frac{1-\varepsilon}{\varepsilon}\left(\frac{-\nu\mathbf{w}}{\nu\|\mathbf{w}\|}\right) = \mathbf{0}. \qquad \square$$

5.2. *Optimality of the sign function* η_0. Consider the GM-estimate based on the sign function $\eta_0(u, \nu) = \operatorname{sgn}(u)$,

$$E\operatorname{sgn}(y - \mathbf{x}'\boldsymbol{\theta})\frac{\mathbf{x}}{\|\mathbf{x}\|} = 0. \tag{5.4}$$

The solution $\boldsymbol{\theta}(H)$ of (5.4) minimizes

$$E\frac{1}{\|\mathbf{x}\|}|y - \mathbf{x}'\boldsymbol{\theta}|. \tag{5.5}$$

Thus the estimate is a weighted L_1 estimate with weights $\|\mathbf{x}_i\|^{-1}$ for a finite sample $(y_i, \mathbf{x}_i)$, $i = 1, \dots, n$. In the case of $p = 1$ it is easy to see that the estimate is the median of the slopes,

$$\hat{\theta}_{\mathrm{ms}} = \operatorname{med}\left\{\frac{y_i}{x_i}\right\}. \tag{5.6}$$

We shall now show that the choice η_0 minimizes the maximum bias over V_ε. We need the following lemma.

LEMMA 5.2. *Assume* $\psi: \mathbb{R} \to \mathbb{R}$ *is* (a) *odd*, (b) *monotone nondecreasing and* (c) $\sup \psi = 1$. *It follows that*:

(i) $q_\psi(t) = E_{F_0}\psi(y+t)$ *is monotone nondecreasing.*
(ii) *If* F_0 *is symmetric, then* $q_\psi(t)t \ge 0$ *and* $q_\psi(-t) = -q_\psi(t)$.
(iii) *If* F_0 *satisfies* A2, *then* $|q_\psi(t)| \le |q_{\psi_0}(t)|$ *where* $\psi_0(u) = \mathrm{sgn}(u)$.

PROOF. (i) Let $t_2 > t_1$. Then

$$q_\psi(t_2) - q_\psi(t_1) = \int_{-\infty}^{\infty} [\psi(y+t_2) - \psi(y+t_1)]\, dF_0(y) \ge 0$$

by property (b) of ψ.

(ii) Since $q_\psi(0) = 0$, (i) gives $q_\psi(t)t \ge 0$. On the other hand,

$$q_\psi(-t) = E_{F_0}\psi(y-t) = E_{F_0}\psi(-y-t)$$
$$= -E_{F_0}\psi(y+t) = -q_\psi(t).$$

(iii) By (ii) we can assume $t > 0$ and, therefore,

$$q_\psi(t) = \int_0^{\infty} \psi(y)[f(y-t) - f(y+t)]\, dy.$$

Since $\psi(y) \le 1$ and $f(y-t) \ge f(y+t)$ for $y \ge 0$, we have

$$q_\psi(t) \le \int_0^{\infty} [f(y-t) - f(y+t)]\, dy$$
$$= q_{\psi_0}(t).$$ □

Now we can prove that η_0 is optimal.

THEOREM 5.1. *Suppose that* η *satisfies* A5, F_0 *satisfies* A2 *and* G_0 *satisfies* A3. *Let* $\mathbf{T}$ *be the GM-estimate based on* η *and* $\mathbf{T}_0$ *be the GM-estimate based on* η_0. *Then*

$$B_\varepsilon(\mathbf{T}) \ge B_\varepsilon(\mathbf{T}_0).$$

PROOF. Let

$$(5.7)\qquad t_\eta(\|\boldsymbol\theta\|) = \left\| E_{H_0}\eta(y - \mathbf{x}'\boldsymbol\theta, \|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}\right\|.$$

A3 implies that the right-hand side expectation depends only on $\|\boldsymbol\theta\|$. Then according to Lemma 5.1, it is enough to show that

$$(5.8)\qquad t_\eta(\|\boldsymbol\theta\|) \le t_{\eta_0}(\|\boldsymbol\theta\|).$$

Setting $\boldsymbol\theta = \lambda(1,0,\dots,0)'$ for $\lambda \ge 0$ without loss of generality, we have

$$(5.9)\qquad t_\eta(\lambda) = E_{H_0}\eta(y - \lambda x_1, \|\mathbf{x}\|)\frac{x_1}{\|\mathbf{x}\|}.$$

Taking conditional expectation with respect to $\mathbf{x}$ in (5.9) we get

$$(5.10)\quad t_\eta^*(\lambda,\mathbf{x}) \triangleq E_{H_0}\left(\eta(y-\lambda x_1, \|\mathbf{x}\|)\frac{x_1}{\|\mathbf{x}\|}\middle|\mathbf{x}\right) = E_{F_0}\eta(y-\lambda x_1, \|\mathbf{x}\|)\frac{x_1}{\|\mathbf{x}\|}$$

and, therefore, by Lemma 5.2, putting $\psi(y) = \eta(y, \|\mathbf{x}\|)$ and $t = -\lambda x_1$, we get

$$(5.11)\qquad E_{H_0}\eta(y-\lambda x_1, \|\mathbf{x}\|)\frac{x_1}{\|\mathbf{x}\|} \le E_{H_0}\eta_0(y-\lambda x, \|\mathbf{x}\|)\frac{x_1}{\|\mathbf{x}\|}.$$

Then (5.9)–(5.11) yield (5.8). □

5.3. *Optimality of η_0 among all equivariant estimates for $p = 1$.* So far we have obtained min-max bias robust estimates over two specific classes of equivariant regression estimates. It would, of course, be highly desirable to obtain a min-max bias solution over the class of *all* equivariant regression estimates. Although it is not yet clear how to obtain such an estimate for general p, we have the following solution for the special case $p = 1$.

THEOREM 5.2. *For the model* (2.6) *with $p = 1$, the median of the slopes estimate $\hat{\theta}_{\mathrm{ms}}$ given by* (5.6) *minimizes the maximum bias among all regression equivariant estimates.*

PROOF. The proof follows lines quite analogous to Huber's (1964) proof of the min-max bias property of the median among all translation equivariant estimates. □

5.4. *Computing the maximum bias.*

LEMMA 5.3. *Assume η satisfies* A5, *F_0 satisfies* A2 *and G_0 satisfies* A3. *Then if* $\mathbf{T}$ *is the GM-estimate corresponding to η, we have*:

(i) *$t_\eta(\lambda)$ is monotone nondecreasing in λ.*
(ii) $B_\varepsilon(\mathbf{T}) = t_\eta^{-1}(\varepsilon/(1-\varepsilon))$.

PROOF. According to Lemma 5.2, $t_\eta^*(\lambda,\mathbf{x})$ defined in (5.7) is monotone nondecreasing in $|\lambda|$ for all $\mathbf{x}$. Then (i) follows. Use of (i) and Lemma 5.1 gives (ii). □

We will compute now $t_{\eta_0}(\lambda)$, when y and $\mathbf{x}$ are normal. From (5.6) we have for $p > 1$,

$$t_{\eta_0}(\lambda) = \left| E\,\mathrm{sign}(y-\lambda x)\frac{x}{(x^2+\nu)^{1/2}}\right|,$$

where y and x are $N(0,1)$ and ν is chi-squared with $p-1$ degrees of freedom. (χ^2_{p-1}), y, x and ν are independent. Then

$$t_{\eta_0}(\lambda) = \left| E(2\phi(\lambda x)-1)\frac{x}{(x^2+\nu)^{1/2}}\right|.$$

In the case that $p = 1$,

$$t_{\eta_0}(\lambda) = |E\,\text{sign}(y - \lambda x)\text{sign}\,x| = \left|E\,\text{sign}\left(\frac{y}{x} - \lambda\right)\right|,$$

where y, x are independent $N(0,1)$. In this case $\nu = y/x$ is Cauchy and then

$$t_{\eta_0}(\lambda) = \left(1 - 2\left(\frac{\tan^{-1}(\lambda)}{\pi} + \frac{1}{2}\right)\right) = \frac{2\tan^{-1}(\lambda)}{\pi}.$$

Therefore, in this case we have

$$B_\varepsilon(\mathbf{T}_0) = \tan\left(\frac{\pi\varepsilon}{2(1-\varepsilon)}\right).$$

6. Including the intercept. The results so far do not cover the case of a regression model with an intercept. This is because they were obtained under the assumptions that the contamination affects all the coordinates of $\mathbf{x}$. Nevertheless, all our results for the regression parameter remain unchanged for the regression model with intercept

$$y = \alpha + \mathbf{x}'\boldsymbol{\theta} + u, \tag{6.1}$$

where y, $\mathbf{x}$, $\boldsymbol{\theta}$ and u are as before and α is the intercept parameter.

Consider the following class of S-estimates of $(\alpha, \boldsymbol{\theta})$: Let T^* be any location functional defined on the class of distribution functions on $\mathbb{R}$. Given a ρ function as in Section 2.1 and a distribution function H on $\mathbb{R}^{p+1}$, we define an S-estimate $\mathbf{T}(H)$ of the regression parameter as the vector $\boldsymbol{\theta}$ which minimizes the scale functional $s(H_{\boldsymbol{\theta}}^*)$, where $H_{\boldsymbol{\theta}}^*$ is the distribution function of $y - \mathbf{x}'\boldsymbol{\theta} - T^*(H_\theta)$ and where H_θ is the distribution function of $y - \mathbf{x}'\boldsymbol{\theta}$. Now one naturally takes the final location estimate to be $T^*(H_{\mathbf{T}(H)})$, i.e., the location estimate T^* applied to the "residuals" $y - \mathbf{x}'\mathbf{T}(H)$. This class contains as a particular case the usual S-estimate of the regression and intercept parameters, simply by taking T^* equal to the corresponding S-estimate of location. Similar extensions are possible for M- and GM-estimates.

Assume now that T^* is Fisher consistent, i.e., for any symmetric distribution F on $\mathbb{R}$, $T^*(F) = 0$ and has breakdown point at least ε. Then it can be shown that the results of Theorems 3.1, 4.1 and 5.1 still hold for estimating $\boldsymbol{\theta}$ in the model (6.1).

It remains to find $(\mathbf{T}, T^*)$, with $\mathbf{T}$ an M-estimate with general scale (or a GM-estimate) and T^* a location estimate, such that the maximum bias of the intercept is minimized. We conjecture that choosing T^* to be the median and $\mathbf{T}$ the corresponding min-max bias estimate for $\boldsymbol{\theta}$ will solve this problem.

7. Comparing min-max bias estimates. The result of solving the min-max bias problem over the class of regression M-estimates with general scale and bounded ρ, yields the discontinuous jump function ρ_c. Consequently, the S-estimate which achieves the min-max bias does not have an influence curve and it has a slower rate of convergence than usual, namely $n^{-1/3}$, the same rate

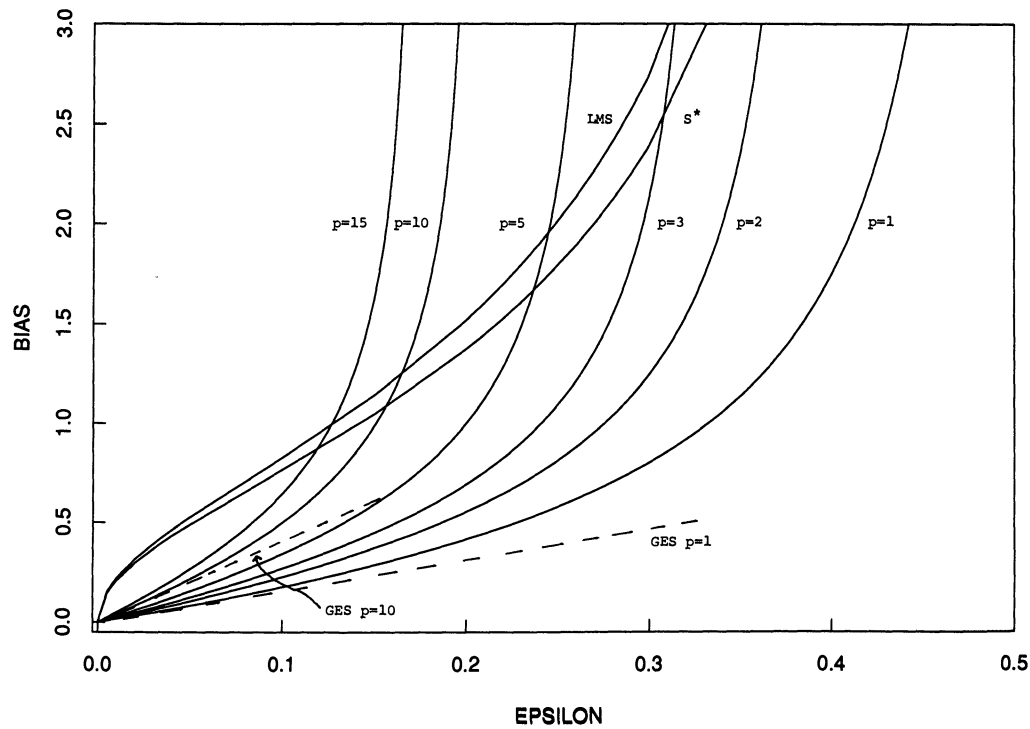

FIG. 2. *Bias curves for min-max bias S-estimate* (S^*) *and min-max bias GM-estimates* ($p = 1, 2, 3, 5, 10, 15$), *and maximal bias curve for least median of squared residuals* (LMS) *estimate.*

of convergence as Rousseeuw's (1984) least median squared residuals (LMS) estimate. This is evidently the price one has to pay when one wishes to control bias over the class of M-estimates with bounded ρ. On the other hand, the min-max bias is independent of the number of carriers p.

The min-max bias GM-estimate of Section 5 does have a bounded influence curve [see Hampel, Ronchetti, Rousseeuw and Stahel (1986)] and enjoys the usual rate of convergence under regularity conditions. However, its bias and breakdown point depend on the dimensionality p of the carrier space [see Maronna, Bustos and Yohai (1979) and Maronna and Yohai (1987a)]. Furthermore, it is necessary to robustly estimate the covariance matrix to implement the GM-estimate, and this is not necessary for the S-estimate.

Nonetheless one wonders how the two min-max estimates compare for fractions of contamination smaller than their breakdown points. First some computations were carried out under the unrealistic assumption that the covariance matrix for the carriers is known. Figure 2 displays the resulting bias curves of the min-max GM-estimate $p = 1, 2, 3, 5, 10, 15$ carriers, along with the bias curves of the min-max S-estimate S^* and the maximal bias curve of the LMS estimate (these latter biases being independent of the number of carriers ρ). Several observations are immediate: For each $p \geq 2$ the optimal GM-estimate has significantly smaller bias than the optimal S-estimate for fractions of contamination not too close to the GM-estimate breakdown point. Of course, as ε approaches the breakdown point of a GM-estimate for any given p, the S-estimate will strongly dominate the GM-estimate. Also, it is interesting to note that the performance of LMS $= T_{0.5}$ [which is the limiting form of $S^* = T_{b(\varepsilon)}$ as $\varepsilon \to 0.5$ and correspondingly $b(\varepsilon) \to 0.5$] is sufficiently close to that of the min-max bias solution $S^* = T_{b(\varepsilon)}$, which uses an optimal $b = b(\varepsilon)$ for each ε (see Theorem 4.2), to regard it as an "excellent" approximation. [This is very similar to the results of Martin and Zamar (1987a), who show that an appropriately scaled median is an excellent approximation to the min-max bias scale estimate for a positive random variable.]

By Theorem 5.2 the optimal GM-estimate $\hat{\theta}_{ms}$ for $p = 1$ is min-max bias optimal among all regression equivariant estimates with model intercept zero and also has breakdown point 0.5. This global optimality of the GM-estimate and its actual degree of dominance over the optimal S-estimate at $p = 1$ begs the following important question: Does there exist a min-max bias regression estimate among the class of all regression equivariant estimates?

We also made some calculations to reveal how estimation of the covariance matrix inflates the min-max biases of the GM-estimates. In order to do so we made use of recent results on the maximal bias of covariance estimates due to Maronna and Yohai (1987b). The results are displayed in Table 1 for the case of the covariance matrix estimate studied by Tyler (1987). Clearly, the price of estimating covariance can be high, even when the fraction of contamination is far from the breakdown point of the GM-estimate with known covariance. See, for example, the $\varepsilon = 0.05$, $p = 15$ and $\varepsilon = 0.2$, $p = 3$ cases. Of course, the smaller breakdown points of the covariance matrix estimates result in smaller breakdown points for the GM-estimates with estimated covariance.

The *gross-error-sensitivity* (GES) is the supremum of the norm of the influence curve and it is a measure of the maximal bias caused by a vanishingly small fraction of contamination. The GES is the derivative of the maximal bias curve at $\varepsilon = 0$, for well-behaved estimators having an influence curve (which LMS and S^* do not). In Figure 2, we display GES-based linear approximations to maximal bias for the optimal GM-estimates for $p = 1$ and $p = 10$. The GES approximation seems rather good for values of ε up to, say, 40 or 50% of the breakdown point. This is in agreement with Hampel's rule of thumb [see Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 178].

REFERENCES

BERAN, R. (1977a). Robust location estimates. *Ann. Statist.* **5** 431–444.

BERAN, R. (1977b). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BICKEL, P. J. (1984). Robust regression based on infinitesimal neighborhoods. *Ann. Statist.* **12** 1349–1368.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimates. Ph.D. qualifying paper, Dept. Statistics, Harvard Univ.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, Calif.

DONOHO, D. L. and LIU, R. C. (1988). The "automatic" robustness of minimum distance functionals. *Ann. Statist.* **16** 587–608.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1895.

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.

HUBER, P. J. (1983). Minimax aspects of bounded-influence regression (with discussion). *J. Amer. Statist. Assoc.* **78** 66–80.

JAECKEL, L. A. (1971). Robust estimates of location: Symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1020–1034.

KRASKER, W. A. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48** 1333–1346.

KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.* **77** 595–604.

MARONNA, R. and YOHAI, V. J. (1987a). The breakdown point of simultaneous general M-estimates of regression scale. Unpublished.

MARONNA, R. and YOHAI, V. J. (1987b). The maximum bias of robust covariances. Unpublished.

MARONNA, R., BUSTOS, O. H. and YOHAI, V. J. (1979). Bias and efficiency—robustness of general M-estimators for regression with random carriers. *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* **757** 91–116. Springer, New York.

MARTIN, R. D. and ZAMAR, R. H. (1987a). Min-max bias robust M-estimates of scale. Technical Report No. 72, Dept. Statistics, Univ. Washington.

MARTIN, R. D. and ZAMAR, R. H. (1987b). Min-max bias robust M-estimates of location. Unpublished.

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

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** 256–272. Springer, New York.

STAHEL, W. A. (1981). Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, ETH, Zurich.

TYLER, D. E. (1987). A distribution free M-estimate of multivariate scatter. *Ann. Statist.* **15** 234–251.

YOHAI, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15** 642–656.

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

ZAMAR, R. H. (1985). Robust estimation for the errors-in-variables model. Ph.D. dissertation, Dept. Statistics, Univ. Washington.

DEPARTMENT OF STATISTICS GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DEPARTAMENTO DE MATEMATICA
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD DE BUENOS AIRES
PABELLON 1
1428 BUENOS AIRES
ARGENTINA

R. H. ZAMAR
DEPARTMENT OF STATISTICS
WEST MALL 2021
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA
CANADA V6T 1W5