The Annals of Statistics
1992, Vol. 20, No. 4, 1875–1888

BIAS ROBUST ESTIMATION IN ORTHOGONAL REGRESSION[1]

BY RUBEN H. ZAMAR

University of British Columbia

Orthogonal regression M-estimates are considered from a bias robust point of view. Their maximum bias over epsilon-contamination neighborhoods is characterized, and maximum bias curves are computed. The most bias robust orthogonal regression M-estimate is derived and shown to be a "mode type" estimate; for instance, in the two-dimensional case this estimate can be computed by locating a strip of fixed width covering the maximum number of data points.

It will be shown that, although orthogonal regression M-estimates with bounded loss function have unbounded influence function, the derivative of their maximum bias curve at zero is finite.

Finally, an implicit formula for an upper bound for the breakdown point of all orthogonal regression M-estimates is found. The upper bound, which depends on the signal-to-noise ratio, is sharp and attained by the most bias robust estimate.

1. Introduction. In this paper we consider the problem of estimation in the structural relationship model when outliers are present in the data. Under the classical Gaussian model the $(p+1)$-dimensional data points Z_i satisfy

$$Z_i = z_i + \varepsilon_i, \qquad \beta_0' z_i = \alpha_0, \tag{1}$$

where $\|\beta_0\| = 1$, z_i and ε_i are independent Gaussian vectors with $E(z_i) = \mu$, $E(\varepsilon_i) = 0$, $\mathrm{Cov}(z_i) = \Sigma$ and $\mathrm{Cov}(\varepsilon_i) = \sigma^2 I$. Notice that these assumptions imply that the ratio of the error variances is known. The eigenvalues of Σ (in increasing order) and corresponding eigenvectors are denoted λ_k and β_k, $k = 0, \dots, p$, respectively. We assume that $\lambda_1 > 0$ and notice that, by (1), $\lambda_0 = 0$. The so-called *signal-to-noise ratio*,

$$\Delta = \frac{\sqrt{\lambda_1}}{\sigma}, \tag{2}$$

plays an important role in the minimax bias theory of Sections 2 and 3.

The distribution of Z_i under the Gaussian model is $F_0 = N(\mu, \Sigma + \sigma^2 I)$. We will assume that the actual common distribution F of the Z_i belongs to the epsilon-contamination neighborhood

$$\mathscr{F}_\epsilon = \{F\colon F = (1-\epsilon)F_0 + \epsilon H\}. \tag{3}$$

Received January 1990; revised December 1991.
[1]Research supported by NSERC and ONR Contract N00014-84-C-0169.
AMS 1980 *subject classifications*. Primary 62J02; secondary 62J05.
Key words and phrases. M-estimates, bias robust, orthogonal regression, structural relationship, functional relationship, errors-in-variables, measurement error model.

Epsilon-contamination neighborhoods were first introduced by Huber (1964) in the location model setup and provide a simple way for modeling data contaminated by outliers. F_0, called the *central distribution*, is a classical parametric model which fits well the bulk of the data; H, called the *contaminating distribution*, is an unspecified distribution which generates the outliers; ϵ, the *fraction of contamination* is a fixed number between 0 and 0.5.

In general, robust estimates offer some degree of bias control by limiting the maximum effect that any individual data point can have on the overall fit. However, even though robust estimates can drastically reduce the bias caused by outliers, they cannot completely eliminate this bias, even when arbitrarily large data sets are available, as shown by the asymptotic theory.

Following Zamar (1989), the asymptotic bias at F of an estimate $\hat{\beta}_n$ (of β_0) is defined as

$$B(F) = 1 - \left|\hat{\beta}(F)'\beta_0\right|, \tag{4}$$

where $\hat{\beta}(F)$ is the asymptotic value of $\hat{\beta}_n$ under F. Notice that $B(F)$ is orthogonal invariant and that $0 \le B(F) \le 1$.

The maximum asymptotic bias of $\hat{\beta}_n$ over $\mathscr{F}_\epsilon$ and its breakdown point are defined as

$$\overline{B}(\epsilon) = \sup_{F \in \mathscr{F}_\epsilon} B(F) \quad \text{and} \quad BP = \sup\{\epsilon: \overline{B}(\epsilon) < 1\}, \tag{5}$$

respectively.

Huber (1964) showed that, in the simple location model, the median minimizes the maximum asymptotic bias among all translation equivariant location estimates. The minimax bias of the median depends on the fraction ϵ of outliers and is given by the formula

$$\overline{B}(\epsilon) = F_0^{-1}\left[\frac{\epsilon}{2(1-\epsilon)}\right].$$

An estimate which like the median minimizes the maximum asymptotic bias over a certain class will be called *bias robust*. Martin and Zamar (1989a, b) and Martin, Yohai and Zamar (1989) derived a bias robust M-estimate of scale and bias robust M-, S- and GM-estimates of regression.

The maximum likelihood estimate under the Gaussian model (1) is called the method of orthogonal regression and consists of minimizing the sum of the squares of the orthogonal residuals, $\Sigma r_i^2(\beta, \alpha)$, where $r_i(\beta, \alpha) = \beta' Z_i - \alpha$. Orthogonal regression (OR) M-estimates defined by Zamar (1989) minimize instead the sum $\Sigma\rho[r_i(\beta, \alpha)/\hat{s}_n]$, where ρ is an even and bounded function designed to downweight the effect of outliers and $\hat{s}_n$ is a robust estimate of the scale of the orthogonal regression residuals, for instance, the orthogonal regression S-estimate of scale defined in Zamar (1989). It will be shown in Section 3 that a particularly important case emerges when $\rho(t)$ is a jump function, that is,

(6) $$\rho(t) = \begin{cases} 0, & \text{if } |t| \le t^*, \\ 1, & \text{otherwise}. \end{cases}$$

In fact, the OR M bias robust estimate has a ρ of this kind. In this case,

$$\sum \rho\left[\frac{r_i(\beta,\alpha)}{\hat{s}_n}\right] = \{i\colon |r_i(\beta,\alpha)| > t^*\hat{s}_n\}.$$

The bias robust fit has then a simple geometric interpretation. It is the "center" of a strip of width $2t^*\hat{s}_n$ located in such a position that it includes the maximum possible number of data points. This estimate can be approximated by using a resampling scheme, along the lines proposed by Rousseeuw and Leroy [(1988), Chapter 5] for classical regression.

In Section 2 the maximum asymptotic bias of orthogonal regression M-estimates is characterized and the maximum bias curve of these estimates is calculated. A simple linear approximation to the maximum bias curve using its derivative at zero is briefly discussed. In Section 3 the bias robust orthogonal regression M-estimate is derived and shown to have a loss function ρ of the jump type. A sharp upper bound for the breakdown point of orthogonal regression estimates is found. In Section 4 some possible extensions are briefly discussed.

2. Bias robust orthogonal regression M-estimates. Assume that σ is known, that the "slope" β_0 is the parameter of interest and that the "intercept" α_0 is a nuisance parameter. The unknown σ case is briefly discussed in Section 4. For $F \in \mathscr{F}_\epsilon$, let $(\hat{\beta}(F), \hat{\alpha}(F))$ be the asymptotic values of the OR M-estimates $(\hat{\beta}_n, \hat{\alpha}_n)$, that is, $\hat{\beta}(F)$ and $\hat{\alpha}(F)$ solve the minimization problem

(7) $$\min_{\|\beta\|=1,\, \alpha \in R} E_F \rho[r(\beta,\alpha)].$$

Zamar (1989) shows that OR M-estimates are Fisher consistent at the Gaussian model (1), that is, $\hat{\alpha}(F_0) = \alpha_0$ and $\hat{\beta}(F_0) = \beta_0$.

Let $\overline{B}_\rho(\epsilon)$ be the maximum asymptotic bias of the OR M-estimate with loss function ρ. It can be shown that if ρ is unbounded then $\overline{B}_\rho(\epsilon) = 1$ for all $\epsilon > 0$, that is, $BP(\rho) = 0$. Therefore we can assume without loss of generality (from the minimax-bias theory point of view) that ρ is bounded and that $\rho(\infty) = 1$.

Notice that β_1, being the eigenvector of Σ associated with the second smallest eigenvalue λ_1 [see the paragraph below (1)] is the eigenvector that could most easily be mistaken for β_0 in the presence of strategically located outliers. Thus β_1 can be considered the "most vulnerable direction" for the given covariance structure. Not surprisingly then, it will be shown that $\overline{B}_\rho(\epsilon)$

is completely determined in terms of the function

$$\beta(\gamma) = (1-\gamma)\beta_0 + \left[1-(1-\gamma)^2\right]^{1/2}\beta_1, \qquad \alpha(\gamma) = \mu'\beta(\gamma). \tag{8}$$

Observe that $\beta(\gamma)$ moves away from β_0 toward β_1 as γ goes from 0 to 1.

The function

$$\begin{aligned} g_\rho(\gamma) &= E_{F_0}\{\rho[\beta(\gamma)'Z - \alpha(\gamma)]\} - E_{F_0}\{\rho(\beta_0'Z - \alpha_0)\} \\ &= E\{\rho[r(\gamma)Y]\} - E\{\rho(\sigma Y)\}, \qquad Y \sim N(0,1), \end{aligned} \tag{9}$$

where $r^2(\gamma) = \sigma^2 + \lambda_1[1-(1-\gamma)^2]$, plays an important role in the minimax theory. It gives the increase in the expected value of the loss function under the central model when $\beta(\gamma)$ and $\alpha(\gamma)$ are used instead of β_0 and α_0. If ρ is almost everywhere continuous, then $g_\rho(\gamma)$ is continuous; since $r(\gamma)$ is strictly increasing, if ρ is even and nondecreasing on $[0,\infty)$, then $g_\rho(\gamma)$ is strictly increasing. Finally, let

$$\mathscr{L} = \{\rho\colon\ \rho \text{ is even, monotone on } [0,\infty) \text{ and has a finite number of discontinuities}\}.$$

Theorem 1, characterizing $\overline{B}_\rho(\epsilon)$, is proved in the Appendix.

THEOREM 1. *Let $0 < \epsilon < 0.5$ be fixed. Suppose that $\rho \in \mathscr{L}$ and assume, without loss of generality, that $\rho(0) = 0$ and $\rho(\infty) = 1$.*

(a) *If $g_\rho(1) > \epsilon/(1-\epsilon)$ (regular case), then $\overline{B}_\rho(\epsilon) = g^{-1}[\epsilon/(1-\epsilon)]$.*
(b) *If $g_\rho(1) \le \epsilon/(1-\epsilon)$ and $\hat{\alpha}(F)$ is bounded, then $\overline{B}_\rho(\epsilon) = 1$.*

REMARK 1. If the first condition of Theorem 1(b) holds then the breakdown point of the corresponding OR M-estimate is smaller than or equal to ϵ. In fact, if $\hat{\alpha}_n$ does not break down for the given ϵ, that is, if $\sup_{F\in\mathscr{F}_\epsilon}|\hat{\alpha}(F)| < \infty$, then the second condition also holds and consequently $\hat{\beta}_n$ breaks down $[\overline{B}_\rho(\epsilon) = 1]$.

REMARK 2. Observe that Theorem 1 and the other results in this paper do not depend on the dimension p.

Linear approximation for $\overline{B}_\rho(\epsilon)$ for ϵ near zero. A linear approximation for $\overline{B}_\rho(\epsilon)$ for ϵ near zero can be obtained using $\overline{B}_\rho'(0)$, the *sensitivity* of the estimate. It can be shown that in many cases including M-estimates of location and M- and S-estimates of scale, the sensitivity and Hampel's *gross-error-sensitivity* [see Hampel, Ronchetti, Rousseeuw and Stahel (1986)] are the same. So, one might expect that in our case

$$l(\epsilon) = \left[\sup_{w\in R^{(p+1)}} \frac{\partial}{\partial\epsilon} B_\rho(F_{\epsilon,w})\bigg|_{\epsilon=0}\right]\epsilon, \qquad F_{\epsilon,w} = (1-\epsilon)F_0 + \epsilon\delta_w,$$

where δ_w is a point-mass distribution at $w \in R^{p+1}$, will provide a good linear approximation for $\overline{B}_\rho(\epsilon)$, for ϵ near zero. However, the order in which the supremum and the derivative are taken has an important effect in the present setup.

For simplicity, consider the case when $p = 1$ and $\alpha_0 = 0$ is known and use the notation $\beta(b) = (1 + b^2)^{-1/2}(1, -b)'$, $\beta_0 = \beta(b_0)$ and $Z = (Y, X)$. The asymptotic value of the OR M-estimate of b_0, $\hat{b}(F)$, solves the minimization problem

$$\min_{b \in R} E_F\{\rho[\beta(b)'Z]\} = \min_{b \in R} E_F\left\{\rho\left(\frac{Y - bX}{\sqrt{1 + b^2}}\right)\right\}.$$

Assume, without loss of generality, that $b_0 = 0$ and so $\beta_0 = (1, 0)'$. It can be easily verified that, for all $w = (y, x)'$, the influence function of $\hat{b}(F)$ is

$$IF(w, \hat{b}, F_0) = \frac{\partial}{\partial \epsilon}\hat{b}(F_{\epsilon, w})\bigg|_{\epsilon = 0} = \frac{\psi(y)x}{\lambda_1 E\{\psi'(Y)\}}, \qquad Y \sim N(0, 1),$$

and so

$$\frac{\partial}{\partial \epsilon}B_\rho(F_{\epsilon, w})\bigg|_{\epsilon = 0} = \frac{\partial}{\partial \epsilon}\left\{1 - \left[1 + \hat{b}^2(F_{\epsilon, w})\right]^{-1/2}\right\}\bigg|_{\epsilon = 0}$$
$$= \frac{\hat{b}(F_0)}{\left[1 + \hat{b}^2(F_0)\right]^{3/2}} IF(w, \hat{b}, F_0) = 0.$$

On the other hand, by Theorem 1(a), the maximum bias of the OR M-estimate with bounded loss function ρ, $\overline{B}_\rho(\epsilon)$, satisfies the equation

$$(1 - \epsilon)g_\rho\left[\overline{B}_\rho(\epsilon)\right] - \rho(\infty)\epsilon = 0, \qquad \forall\, 0 < \epsilon < BP_\rho.$$

Differentiating with respect to ϵ at $\epsilon = 0$ gives

$$-g_\rho(0) + g_\rho'(0)\overline{B}_\rho'(0) - \rho(\infty) = 0.$$

If $\rho'(x) = \psi(x)$ and Δ is given by (2), then

$$g_\rho'(0) = \Delta^2 E\{\psi(\sigma Y)\sigma Y\}, \qquad Y \sim N(0, 1), \qquad g_\rho(0) = 0;$$

so,

$$\overline{B}_\rho'(0) = \frac{\rho(\infty)}{\Delta^2 E\{\psi(\sigma Y)\sigma Y\}}.$$

This formula does not hold when $\rho(x)$ is not differentiable. In the important case of the jump function (6), $g_\rho'(\gamma) = 2\{\phi(t/\sigma) - \phi[t/r(\gamma)]\}$ and

$$\overline{B}_\rho'(0) = \frac{\rho(\infty)}{2\Delta^2\phi(t/\sigma)(t/\sigma)}.$$

Figure 1 gives the maximum bias curves and the corresponding linear approximations for the 95% efficient Tukey's OR M-estimate (dashed line) with loss

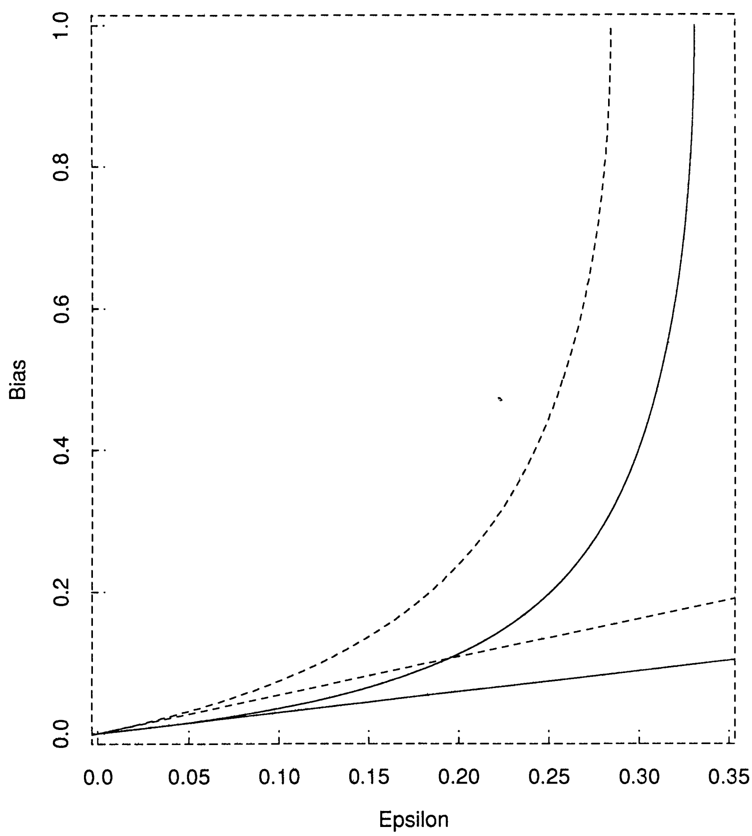

FIG. 1. *Maximum bias curves and sensitivity-based linear approximations for the* 95% *efficient Tukey's OR M-estimate (dashed lines) and for the bias robust OR M-estimate (solid lines).*

function

$$\rho(x) = \begin{cases} \dfrac{3}{c^2}\left(x^2 - \dfrac{x^4}{c^2} + \dfrac{x^6}{3c^4}\right), & \text{if } |x| \le c, \\ 1, & \text{otherwise,} \end{cases} \tag{10}$$

with $c = 4.7$, and for the bias robust OR M-estimate with loss function (6), with $t = 1.372$ (solid line) when $\sigma = 1$ and $\Delta = 3$. Observe that for $\epsilon < 0.10$ the linear approximations are fairly good, especially in the case of the bias robust estimate. As one would expect, the breakdown point of the bias robust estimate (approximately 0.33) is larger than that of Tukey's estimate (approximately 0.27). The corresponding sensitivities are 0.29 and 0.54.

3. Bias robust orthogonal regression *M*-estimates. In this section we solve the following optimization problem.

PROBLEM P. For a fixed $0 < \epsilon < 0.5$, find ρ^* such that $\overline{B}_{\rho^*}(\epsilon) \le \overline{B}_{\rho}(\epsilon)$ for all $\rho \in \mathscr{L}$.

To solve Problem P, consider the following family of auxiliary problems, one for each $0 < \gamma < 1$.

PROBLEM P_γ. Find ρ_γ such that $g_{\rho_\gamma}(\gamma) \ge g_\rho(\gamma)$ for all $\rho \in \mathscr{L}$.

The following theorem is the main result.

THEOREM 2. *Under the assumptions of Theorem 1 the following hold:*

(a) *For each $0 < \gamma < 1$, Problem P_γ has a solution denoted ρ_γ.*
(b) *The function $G(\gamma) = g_{\rho_\gamma}(\gamma)$ is continuous and strictly increasing in $(0, 1)$.*
(c) *Let $0 < \epsilon < 0.5$ be given. If $\lim_{\gamma \to 1} G(\gamma) > \epsilon/(1-\epsilon)$ then there exists a unique $\gamma^* = \gamma^*(\epsilon)$ such that $G(\gamma^*) = \epsilon/(1-\epsilon)$ and the bias robust estimate for the given ϵ has loss function (6), with $t = \gamma^*$. If $\lim_{\gamma \to 1} G(\gamma) \le \epsilon/(1-\epsilon)$, then all orthogonal regression M-estimates, $(\hat{\beta}_n, \hat{\alpha}_n)$, break down, that is, either $\hat{\alpha}_n$ breaks down (its maximum asymptotic bias is unbounded) or $\hat{\beta}_n$ breaks down (its maximum asymptotic bias is equal to 1) for the given ϵ and the given covariance structure.*

PROOF. Let $0 < \gamma < 1$, $b^2 = 1 - (1-\gamma)^2$, $U_\gamma = (Z-u)'\beta(\gamma)$ and $V = (Z-u)'\beta_0$. Then, under F_0, $U_\gamma \sim N[0, \tau^2(\gamma)]$, $\tau^2(\gamma) = \sigma^2 + \lambda_1 b^2$ and $U \sim N(0, \sigma^2)$. For $\rho \in \mathscr{L}$ and $\gamma \in (0,1)$,

$$g_{\rho_\gamma}(\gamma) = E_{F_0}\{\rho(U_\gamma) - \rho(V)\} = 2\int_0^\infty \rho(t)\delta_\gamma(t)\,dt,$$

where

$$\delta_\gamma(t) = \tau(\gamma)^{-1}\phi[t/\tau(\gamma)] - \sigma^{-1}\phi(t/\sigma), \qquad \phi(t) = (2\pi)^{-1}\exp(-t^2/2).$$

Restrict attention to nonnegative t. Notice that $\delta_\gamma(t) < 0$ if t is smaller than

$$t_\gamma = \sigma\left[\frac{1+\Delta^2 b^2}{\Delta^2 b^2}\log(1+\Delta^2 b^2)\right]^{1/2}, \tag{11}$$

and $\delta_\gamma(t) > 0$ if t is larger than t_γ. Since $\rho \in \mathscr{L}$ takes values between 0 and 1, it is clear that taking $\rho_\gamma(t) = 1$ when $t > t_\gamma$ and $\rho_\gamma(t) = 0$ otherwise maximizes $g_\rho(\gamma)$, for the given γ; this proves (a). To prove (b) notice that

$$G(\gamma) = 2\left\{\Phi\left[\left(\frac{1+\Delta^2 b^2}{\Delta^2 b^2}\log(1+\Delta^2 b^2)\right)^{1/2}\right] - \Phi\left[\left(\log(1+\Delta^2 b^2)\right)^{1/2}\right]\right\}, \tag{12}$$

where $\Phi(t)$ is the standard normal distribution function; thus, $G(\gamma)$ is continuous. Finally, for any $0 < \gamma_1 < \gamma_2 < 1$,

$$G(\gamma_1) = g_{\rho_{\gamma_1}}(\gamma_1) < g_{\rho_{\gamma_1}}(\gamma_2) \le g_{\rho_{\gamma_2}}(\gamma_2) = G(\gamma_2).$$

The first inequality follows because, as noticed in the comments below (9), $g_{\rho_{\gamma_1}}(\gamma)$ is an increasing function of γ; the second inequality follows from the optimality of ρ_{γ_2} already proved in (a). To prove (c), let $\rho \in \mathscr{L}$ be fixed. By the optimality of ρ_{γ^*}, if $g_{\rho_\gamma} = \epsilon/(1-\epsilon)$, then $\gamma \ge \gamma^*$. So,

$$\overline{B}_\rho(\epsilon) = g_\rho^{-1}\left(\frac{\epsilon}{1-\epsilon}\right) = \gamma \ge \gamma^* = g_{\rho_{\gamma^*}}^{-1}\left(\frac{\epsilon}{1-\epsilon}\right) = \overline{B}_{\rho_{\gamma^*}}(\epsilon). \qquad \square$$

The bias robust estimate. By the proof of Theorem 2, the bias robust orthogonal regression M-estimate has loss function (6) with $t = t^* = t_{\gamma^*}$ given by (11). When σ is unknown, it must be replaced by a robust estimate of the scale of the orthogonal residuals, for example, the S-estimate defined by Zamar (1989). Given the fraction of contamination ϵ, the next step for determining the value of t^* is to find γ^* by solving the nonlinear equation $G(\gamma) = \epsilon/(1-\epsilon)$. Although the function $G(\gamma)$ (and, consequently, the bias robust estimate) appears to depend on the signal-to-noise ratio Δ^2, the following argument shows that, fortunately, this is not the case. In fact, (11) and (12) only depend on γ and Δ through the product $\Delta^2 b^2$. Furthermore, if $\tilde{G}$ is the function obtained by replacing $\Delta^2 b^2$ by s in (12), then by Theorem 2(b) (with $\Delta^2 = 1$) $\tilde{G}$ is continuous and strictly increasing. It can also be checked that $\lim_{s\to\infty} \tilde{G}(s) = 1$. Given $0 < \epsilon < 0.5$, let $s^* = s^*(\epsilon)$ be the unique solution to the equation $\tilde{G}(s) = \epsilon/(1-\epsilon)$. Replacing $\Delta^2 b^2$ by s^* in (11) gives the desired value of t^* which clearly, then, does not depend on Δ^2. Table 1 gives the values of t^* for several values of ϵ. As shown in Figure 2, the choice $t = 1.37$ (optimal for $\epsilon = 0.25$) has maximum-bias curve (dashed line) almost indistinguishable from the lower bound (lower solid line). The dotted line corresponds to the bias robust estimate when $\epsilon = 0.05$. The remaining solid line is the maximum bias curve of the 95% efficient Tukey's estimate.

Breakdown-point considerations. Although the bias robust estimate itself does not depend on the signal-to-noise ratio Δ^2, the corresponding minimax bias does. Consequently, the breakdown point of the bias robust estimate depends on Δ^2. Let Δ^2 be given and fixed throughout this discussion. For each $0 < \epsilon < 0.5$, there is a unique $s = s(\epsilon)$ such that $\tilde{G}(s) = \epsilon/(1-\epsilon)$. Using (8)

TABLE 1
Optimal values of the jump point t^* for various choices of ϵ

ϵ	$t^*(\epsilon)$	ϵ	$t^*(\epsilon)$
0.05	1.054	0.20	1.272
0.10	1.118	0.25	1.327
0.15	1.189	0.30	1.495

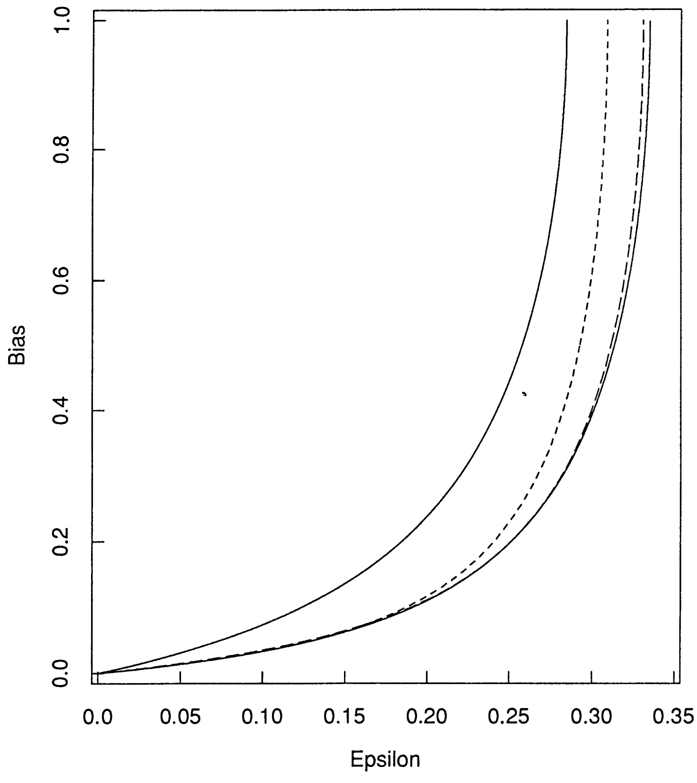

FIG. 2. *Lower bound for the maximum bias of all OR M-estimates (lower solid line) and maximum bias curves for the 95% efficient Tukey's OR M-estimate (upper solid line), for the bias robust OR M-estimate when $\epsilon = 0.05$ (dotted line) and for the bias robust OR M-estimate when $\epsilon = 0.25$ (dashed line).*

and the identity $\Delta^2 b^2 = s(\epsilon)$, one obtains that the maximum bias $\overline{B}^*(\epsilon)$ of the bias robust estimate (i.e., the minimax bias) is given by

$$\overline{B}^*(\epsilon) = 1 - \left[1 - \frac{s(\epsilon)}{\Delta^2}\right]^{1/2},$$

provided that $s(\epsilon) < \Delta^2$; otherwise $\overline{B}^*(\epsilon) = 1$. Thus, the breakdown point of the bias robust estimate (and, consequently, the maximum breakdown-point achievable by any orthogonal regression M-estimate) is given by the expression

$$\epsilon^* = \sup\{\epsilon\colon s(\epsilon) > \Delta^2\}.$$

Table 2 gives the values of ϵ^* for several values of Δ.

TABLE 2
Upper bound for the breakdown point of all ORM-*estimates*

Signal-to-noise ratio (Δ)	ϵ^*
1.0	0.14
1.5	0.22
2.0	0.27
3.0	0.33
10.0	0.44
100.0	0.49

4. Brief discussion of some extensions.

The unknown-σ case. When σ is unknown the orthogonal residuals $r(a, \alpha)$ in (7) must be divided by a robust estimate of scale, $\hat{\sigma}$, to obtain a scale-equivariant and robust OR M-estimate. One can use, for instance, the auxiliary orthogonal regression S-estimate of scale (OR S-estimate) described in Zamar (1989). Notice that this scale estimate is calculated before computing the OR M-estimate. It can be shown that the breakdown point of the OR S-estimate of scale with loss function ρ satisfying the assumptions of Theorem 1 is equal to $\min\{E[\rho(Y)], 1 - E[\rho(Y)]\}$, where $Y \sim N(0,1)$. To achieve a breakdown point of $1/2$, one can use the loss function (10) and $c = 1.08$.

Evidently, the contamination in the data will produce an increase in the value of the scale estimate and this in turn will cause an increase in the value of $\overline{B}_\rho(\epsilon)$. It can be shown, however, that the bias of the OR S-estimate of scale is small for ϵ moderately small, so the increase in $\overline{B}_\rho(\epsilon)$ due to the estimation of σ will also be small.

Let $d(F) = \hat{\sigma}(F)/\sigma$, where $\hat{\sigma}(F)$ is the scale estimate asymptotic functional. Let

$$d^+ = \sup_{F \in \mathscr{F}_\epsilon} d(F), \qquad d^- = \inf_{F \in \mathscr{F}_\epsilon} d(F)$$

and

$$\tilde{g}_\rho(\gamma, d) = E_{F_0}\left\{\rho\left[\frac{a(\gamma)'Z - \alpha(\gamma)}{d}\right]\right\} - E_{F_0}\left\{\rho\left[\frac{\beta_0' Z - \alpha_0}{d}\right]\right\}.$$

Finally, let

$$\bar{g}_\rho(\gamma) = \inf_{d^- \le d \le d^+} \tilde{g}_\rho(\gamma, d) \quad \text{and} \quad \underline{g}_\rho(\gamma) = \tilde{g}_\rho(\gamma, d^+).$$

Under the assumptions of Theorem 1, and using arguments similar to those in the proof of Theorem 1, it can be shown that replacing g_ρ by $\bar{g}_\rho$ gives an upper bound for $\overline{B}_\rho(\epsilon)$, that is,

$$\bar{g}_\rho(1) < \frac{\epsilon}{1-\epsilon} \quad \Rightarrow \quad \overline{B}_\rho(\epsilon) \le \bar{g}_\rho^{-1}\left[\frac{\epsilon}{1-\epsilon}\right].$$

On the other hand, replacing g by $\underline{g}_\rho$ gives a lower bound for $\overline{B}_\rho(\epsilon)$, that is,

$$\overline{B}_\rho(\epsilon) \geq \underline{g}_\rho^{-1}\left[\frac{\epsilon}{1-\epsilon}\right].$$

In particular, when for each fixed γ the function $\tilde{g}_\rho(\gamma, d)$ is decreasing in d, it clearly follows that $\bar{g}_\rho(\gamma) = \underline{g}_\rho(\gamma)$ and $\overline{B}_\rho(\epsilon)$ can be simply characterized in terms of $\underline{g}_\rho(\gamma)$. A simple condition for monotonicity of $\tilde{g}_\rho(\gamma, d)$ is that the covariance between $\rho'(Y)Y$ and Y^2 is nonnegative, that is,

$$E_{F_0}\{\psi(Y)Y(Y^2-1)\} \geq 0, \qquad Y \sim N(0,1).$$

Non-Gaussian F_0. To deal with the non-Gaussian case, the function $g_\rho(\gamma)$ in (9) must be replaced by

$$g_\rho(\gamma) = \inf_{\|\beta\|=\gamma,\ \alpha\in R} E_{F_0}\{\rho[\beta' Z - \alpha]\} - E_{F_0}\{\rho[\beta_0' Z - \alpha_0]\}, \tag{13}$$

and the minimax theory becomes more involved, particularly when $\rho > 1$.

Let G_0 denote the common distribution of the ε_i and let H_0 be the common distribution of the z_i. If we assume that G_0 is spherically symmetric and that there exists $\mu \in R^{p+1}$ such that for all unit vectors β in R^{p+1} the distribution of $\beta' z_i$ is symmetric about $\beta'\mu$, then (13) takes the simpler form

$$g_\rho(\gamma) = \inf_{\|\beta\|=\gamma} E_{F_0}\{\rho[\beta'(Z-\mu)]\} - E_{F_0}\{\rho[\beta_0' Z - \alpha_0]\}, \tag{14}$$

and a result similar to that in Theorem 1 can be obtained. If in addition $p = 1$, then the infimum in (14) is no longer needed and $g_\rho(\gamma)$ takes the simple form

$$g_\rho(\gamma) = E_{F_0}\{\rho[\beta(\gamma)'(Z-\mu)]\} - E_{F_0}\{\rho[\beta_0' Z - \alpha_0]\},$$

where $\beta(\gamma)$ is as in (8). Let $f_\gamma(x)$ be the density of $\beta(\gamma)'(Z-\mu)$ under the central model. If for each $0 < \gamma < 1$ there exists $x_\gamma > 0$ such that $f_\gamma(x) - f_0(x) < 0$ if and only if $|x| < x_\gamma$, then a result similar to Theorem 2 can be obtained.

The functional errors-in-variables model. Suppose that, under the central model, $Z_1, \dots, Z_n$ are independent multivariate normal random vectors with

$$E(Z_i) = m_i, \qquad \mathrm{Cov}(Z_i) = \sigma^2 I \ \text{ and } \ \beta' m_i = \alpha_0, \qquad i = 1, \dots, n.$$

To deal with this model, the function $g_\rho(\gamma)$ in (9) must be replaced by

$$g_\rho(\gamma) = \inf_{\|\beta\|=\gamma,\ \alpha\in R} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left\{E_{F_0}[\rho(\beta' Z_i - \alpha)] - E_{F_0}[\rho(\beta_0' Z_i - \alpha_0)]\right\}. \tag{15}$$

In general, since the Z_i are no longer identically distributed, the results obtained for the structural relationship model cannot easily be extended to the functional errors-in-variables model. However, if one assumes that the empirical distribution function of the μ_i, $G_n(m)$, approaches a limiting distribution $G_0(m)$ as $n \to \infty$, then (15) becomes identical to (13); so, from the minimax bias theory point of view, the functional and the structural setups can be treated in the same way. In particular, if G_0 is Gaussian, then Theorems 1 and 2 hold for the functional errors-in-variables model.

APPENDIX

Proof of Theorem 1. To prove Theorem 1, we need two auxiliary lemmas. Lemma 1 is itself of interest because it shows that in the case of OR M-estimates, the breakdown point of the intercept $\hat{\alpha}$ is larger than or equal to the breakdown point of $\hat{\beta}$. Lemma 2 shows that, for computing $\overline{B}_\rho(\epsilon)$, attention can be restricted to unit vectors $\beta(\gamma)$ defined in (8).

LEMMA 1. *If $g_\rho(1) > \epsilon/(1-\epsilon)$, there exists $K > 0$ such that $|\hat{\alpha} - \alpha_0| \le K$ for all $F \in \mathscr{F}_\epsilon$.*

PROOF. Since $g_\rho(1) > \epsilon/(1-\epsilon)$, there exists $0 < \delta < 1$ such that

$$E_{F_0}\{\rho(\beta_0' Z - \alpha_0)\} < \frac{(1-\delta)(1-\epsilon) - \epsilon}{1-\epsilon}.$$

By the dominated convergence theorem

$$\lim_{\alpha \to \infty} E_{F_0}\{\rho(\beta' Z - \alpha)\} = 1, \qquad \forall\, \|\beta\| = 1.$$

By this and the compactness of the set $\{\beta\colon \|\beta\| = 1\}$, there exists a constant $K > 0$ such that

$$|\alpha - \alpha_0| > K \quad \text{implies} \quad E_{F_0}\{\rho(\beta' Z - \alpha)\} \ge 1 - \delta, \qquad \forall\, \|\beta\| = 1.$$

Therefore, for all $F \in \mathscr{F}_\epsilon$ and all $\|\beta\| = 1$ we have

$$\begin{aligned} E_{F_0}\{\rho(\beta_0' Z - \alpha_0)\} &\le (1-\epsilon) E_{F_0}\{\rho(\beta_0' Z - \alpha_0)\} + \epsilon < (1-\epsilon)(1-\delta) \\ &< (1-\epsilon) E_{F_0}\{\rho(\beta' Z - \alpha)\} \le E_F\{\rho(\beta' Z - \alpha)\}. \end{aligned}$$

LEMMA 2. *Let $\mathscr{C}_\gamma$ be the set of unit vectors β satisfying $1 - |\beta'\beta_0| \ge \gamma$. Then,*

$$E_{F_0}\{\rho[\beta(\gamma)'(Z-\mu)]\} \le E_{F_0}\{\rho(\beta' Z - \alpha)\}, \qquad \forall\, \beta \in \mathscr{C}_\gamma, \forall\, \alpha \in R.$$

PROOF. By the assumptions on ρ and the fact that, under F_0, for all $\alpha \in R$, $(\beta' Z - \alpha)^2$ is stochastically larger than $[\beta'(Z-\mu)]^2$, we have

$$E_{F_0}\{\rho[\beta'(Z-\mu)]\} \le E_{F_0}\{\rho(\beta' Z - \alpha)\}. \tag{16}$$

To prove the lemma, then, it suffices to show that

$$E_{F_0}\{\rho[\beta(\gamma)'(Z-\mu)]\} \le E_{F_0}\{\rho[\beta'(Z-\mu)]\}, \qquad \forall\, \beta \in \mathscr{C}_\gamma.$$

But any vector $\beta \in \mathscr{C}_\gamma$ can be expressed in the form

$$\beta = (1-\delta)\beta_0 + \sum_1^p \delta_i \beta_i, \qquad \gamma \le \delta \le 1, \qquad (1-\delta)^2 + \sum_1^p \delta_i^2 = 1.$$

Thus, $\sum_1^p \delta_i^2 \geq 1 - (1-\gamma)^2$ and

$$\operatorname{Var}(\beta' Z) = \sigma^2 + \sum_1^p \delta_i^2 \lambda_i \geq \sigma^2 + \lambda_1 \sum_1^p \delta_i^2 \geq \lambda_1\left[1 - (1-\gamma)^2\right] = \operatorname{Var}[\beta(\gamma)' Z].$$

So $[\beta'(Z-\mu)]^2$ is stochastically larger than $[\alpha(\gamma)'(Z-\mu)]^2$ and the lemma follows.

PROOF OF THEOREM 1. Suppose first that $g_\rho(1) > \epsilon/(1-\epsilon)$. To prove (a) it suffices to show that, for all $0 < \gamma < 1$, the following hold:

(i) $g_\rho(\gamma) > \epsilon/(1-\epsilon)$ implies $\overline{B}_\rho(\epsilon) < \gamma$;
(ii) $g_\rho(\gamma) < \epsilon/(1-\epsilon)$ implies $\overline{B}_\rho(\epsilon) \geq \gamma$.

To prove (i), let $\beta \in \mathscr{C}_\gamma$ and $\alpha \in R$. By the given assumption and Lemma 2,

$$E_{F_0}\{\rho(\beta' Z - \alpha)\} - E_{F_0}\{\rho(\beta_0' Z - \alpha_0)\} \geq g_\rho(\gamma) > \frac{\epsilon}{1-\epsilon}$$

$$\Rightarrow \quad (1-\epsilon) E_{F_0}\{\rho(\beta' Z - \alpha)\} > (1-\epsilon) E_{F_0}\{\rho(\beta_0' Z - \alpha_0)\} + \epsilon$$

$$\Rightarrow \quad E_F\{\rho(\beta' Z - \alpha)\} > E_F\{\rho(\beta_0' Z - \alpha_0)\}, \qquad \forall\, F \in \mathscr{F}_\epsilon.$$

Hence, $\hat{\beta}(F)$ does not belong to $\mathscr{C}_\gamma$ for all $F \in \mathscr{F}_\epsilon$, and the right-hand side of (i) follows. To prove (ii) it suffices to show that there exists a sequence G_n in $\mathscr{F}_\epsilon$ such that

$$\lim_{n\to\infty}\left[1 - \left|\beta_0' \hat{\beta}(G_n)\right|\right] = \gamma.$$

Let $0 < \delta < 1$ be such that $g_\rho(\gamma) < (1-\delta)\epsilon/(1-\epsilon)$ and let

$$\tilde{\beta}(\gamma) = \left[1 - (1-\gamma)^2\right]^{1/2} \beta_0 - (1-\gamma)\beta_1.$$

Observe that $\{\beta(\gamma), \tilde{\beta}(\gamma), \beta_2, \dots, \beta_p\}$ is an orthonormal basis of R^{p+1}. Let $\mathscr{V}$ be the p-dimensional space spanned by $\{\tilde{\beta}(\gamma), \beta_2, \dots, \beta_p\}$. Any unit vector β can be (uniquely) expressed as

$$\beta = b\tilde{\beta}(\gamma) + (1-b^2)^{1/2} c, \qquad c = b_1 \beta(\gamma) + \sum_2^p b_j \beta_j, \qquad b^2 + \sum_1^p b_j^2 = 1. \tag{17}$$

For $0 < t < 1$, let $\mathscr{V}_t$ be the set of all the unit vectors with $|b| \geq t$. If β is not in $\mathscr{V}_t$, then $1 - |\beta'\beta_0| \geq \gamma - t$. In fact, for any unit vector β [see (17)]

$$\begin{aligned} 1 - |\beta'\beta_0| &\geq 1 - |b|\left|\tilde{\beta}(\gamma)'\beta_0\right| - (1-b^2)^{1/2}|c'\beta_0| \\ &\geq 1 - |b| - |c'\beta_0| = 1 - |b_1|(1-\gamma) - |b| \geq \gamma - |b|. \end{aligned}$$

Let $0 < t_n \leq 1$ be such that $\lim t_n = 0$, let $y_n = \mu + k_n \tilde{\beta}(\gamma)$, where the k_n are chosen so that $\rho(\beta' y_n - \alpha) > 1 - \delta$ for all $|\alpha - \alpha_0| \leq K$ and all $\beta \in \mathscr{V}_{t_n}$ (see Lemma 1 and notice that $|\beta' y_n| \geq k_n t_n - \|\mu\|$).

Let $G_n = (1-\epsilon)F_0 + \epsilon\delta_{y_n}$ and notice that, by Lemma 1, $|\tilde{\alpha}(G_n) - \alpha_0| \le K$. Now, if $|\alpha - \alpha_0| \le K$ and $\beta \in \mathscr{V}_{t_n}$, then

$$\begin{aligned} E_{G_n}\{\rho(\beta'Z - \alpha)\} &= (1-\epsilon)E_{F_0}\{\rho(\beta'Z - \alpha)\} + \epsilon\rho(\beta' y_n - \alpha) \\ &\ge (1-\epsilon)E_{F_0}\{\rho(\beta_0'Z - \alpha_0)\} + \epsilon(1-\delta) \\ &> (1-\epsilon)E_{F_0}\{\rho[\beta(\gamma)'Z - \alpha(\gamma)]\} \\ &= E_{G_n}\{\rho[\beta(\gamma)'Z - \alpha(\gamma)]\}. \end{aligned}$$

Therefore, $\hat{\beta}(G_n)$ is not in $\mathscr{V}_{t_n}$ and so $1 - |\beta_0'\hat{\beta}(G_n)| \ge \gamma - t_n \to \gamma$, proving (ii) and part (a) of the theorem. Finally, if the left-hand side of (b) holds, then the left-hand side of (ii) follows for all $\gamma < 1$. If $\hat{\alpha}$ does not break down (remains bounded), then the argument used to prove (ii) follows and since it is true for all $\gamma < 1$, we conclude that $\overline{B}_\rho(\epsilon) = 1$. This proves part (b) and the theorem. □

Acknowledgments. I thank an Associate Editor and the referees for their useful comments and suggestions.

REFERENCES

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

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.

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

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

MARTIN, D. R. and ZAMAR, R. H. (1989b). Bias robust estimation of scale when location is unknown. Technical report, Dept. Statistics., Univ. Washington.

ROUSSEEUW, P. J. and LEROY, A. M. (1988). *Robust Regression and Outlier Detection*. Wiley, New York.

ZAMAR, R. H. (1989). Robust estimation in the errors in variables models. *Biometrika* **76** 149–160.

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