The Annals of Statistics
1993, Vol. 21, No. 4, 1824–1842

A MINIMAX-BIAS PROPERTY OF THE LEAST α-QUANTILE ESTIMATES

BY VICTOR J. YOHAI[1] AND RUBEN H. ZAMAR[2]

University of San Andrés, University of Buenos Aires and CONICET and University of British Columbia

A natural measure of the degree of robustness of an estimate $\mathbf{T}$ is the maximum asymptotic bias $B_{\mathbf{T}}(\varepsilon)$ over an ε-contamination neighborhood. Martin, Yohai and Zamar have shown that the class of least α-quantile regression estimates is minimax bias in the class of M-estimates, that is, they minimize $B_{\mathbf{T}}(\varepsilon)$, with α depending on ε. In this paper we generalize this result, proving that the least α-quantile estimates are minimax bias in a much broader class of estimates which we call *residual admissible* and which includes most of the known robust estimates defined as a function of the regression residuals (e.g., least median of squares, least trimmed of squares, S-estimates, τ-estimates, M-estimates, signed R-estimates, etc.). The minimax results obtained here, likewise the results obtained by Martin, Yohai and Zamar, require that the carriers have elliptical distribution under the central model.

1. Introduction. It is well known that the least squares estimates of the regression coefficients are very sensitive to the presence of outliers in the sample. In fact, the outliers can severely bias these estimates, especially when they are grouped and associated with high leverage points. This fact motivated the development of robust methods which are not so sensitive to outliers.

Several robust regression estimates have been proposed in recent years; therefore, there is a need for quantitative measures to assess the degree of robustness of these estimates. One such measure is the maximum bias $B_{\mathbf{T}}(\varepsilon)$ caused by a fraction ε of outliers. The function $B_{\mathbf{T}}(\varepsilon)$ was first introduced by Huber (1964) for the location model; a precise definition of $B_{\mathbf{T}}(\varepsilon)$ for the regression model is given in Section 2. Other robustness measures closely related to the maximum bias $B_{\mathbf{T}}(\varepsilon)$ are the gross-error sensitivity, GES [Hampel (1974)] and the breakdown point, BP [Hampel (1971)]. It turns out that, under some regularity conditions, GES $= B'_{\mathbf{T}}(0)$ and, therefore, the GES gives a linear approximation for $B_{\mathbf{T}}(\varepsilon)$ near zero. The BP is the smallest ε for which $B_{\mathbf{T}}(\varepsilon) = \infty$. Although the GES and the BP carry much information about $B_{\mathbf{T}}(\varepsilon)$, especially for ε near zero and the BP, the curve $B_{\mathbf{T}}(\varepsilon)$ constitutes a more complete description of the robustness properties of the estimate.

Given the class $\mathscr{T}$ of estimates and the fixed fraction ε of contamination, the most robust estimate in $\mathscr{T}$ is naturally defined by the property of

Received January 1991; revised June 1992.
[1]Supported by ONR contract N000014-84-C-0169.
[2]Supported by Natural Sciences and Engineering Research Council of Canada Grant A9276.
AMS 1991 *subject classification*. 62F35.
Key words and phrases. Minimax bias, regression, robust estimates.

minimizing $B_T(\varepsilon)$ over $\mathscr{T}$; such an estimate is called *minimax bias*. Huber (1964) showed that the median is minimax-bias for all $0 < \varepsilon < 0.5$ among translation equivariant location estimates. He considered neighborhoods of a symmetric and unimodal central distribution. Location models with unimodal central distributions were also studied by Rychlik and Zielinski (1987), Rychlik (1987) and Zielinski (1985, 1987, 1988). These authors derived minimax-bias estimates using a modified definition of maximum bias for both gross-error and Lévy–Kolmogorov-type neighborhoods. Riedel [(1987), (1989a, b)] derived the minimax-bias estimate among all equivariant location estimates, for several types of contamination neighborhoods and arbitrary central distributions.

Martin, Yohai and Zamar (1989) found the minimax-bias M- and GM-estimates of regression, using gross-error neighborhoods of a central model with elliptically distributed carriers. The minimax-bias M-estimate minimizes a certain $\alpha(\varepsilon)$-quantile of the absolute value of the error's distribution. Since $\lim_{\varepsilon \to 0.5} \alpha(\varepsilon) = \frac{1}{2}$, the minimax-bias M-estimate tends to Rousseeuw's (1984) least median of squares, LMS. The minimax-bias GM-estimate is a weighted L_1-estimate, with weights inversely proportional to the leverage of the carriers. In the case of only one carrier, this estimate (which reduces to the median of the ratios y_i/x_i) is also minimax bias among all equivariant estimates.

The interesting work by Riedel (1991) was brought to our attention by a referee. This paper studies the minimax-bias problem for general models when the parameter space is endowed with a group and metric structure. Riedel considers arbitrary families of contamination neighborhoods and proves the existence of the minimax-bias estimate among all equivariant and Fisher-consistent estimates. He also finds an expression for the minimax-bias function which, in the particular case of the regression model with only one random carrier, no intercept and gross-error contamination neighborhoods, agrees with the maximum bias of the minimax-bias GM-estimate described previously [see Theorem 5.3 of Riedel (1991)]. This expression only gives a lower bound for the maximum bias when $p > 1$.

Unfortunately, it does not seem possible to use the general results of Riedel (1991) to derive a computable version of the minimax-bias estimate and the minimax-bias function in the case of multiple regression.

Other related results can be found in Donoho and Liu (1988), He, Simpson and Portnoy (1990) and Maronna and Yohai (1989). Donoho and Liu (1988) study the minimax-bias properties of minimum distance estimates. He, Simpson and Portnoy (1990) introduce the concept of breakdown point of a test using an approach closely connected to the maximum bias curve. Finally, Maronna and Yohai (1989) define the class of projection estimates of regression and show that these estimates have bounded influence and good maximum bias performance.

In this paper we show that the least α-quantile estimates are minimax bias in a very large class of estimates which we call *residual admissible*. Roughly speaking, a residual admissible estimate is one for which the empirical distribution of the absolute value of its regression residuals cannot be uniformly improved by using any other set of regression coefficients. We will show that many robust estimates defined as a function of the regression residuals are residual admissible.

As in Martin, Yohai and Zamar (1989), the minimax results obtained here require that the carriers have elliptical distribution under the central model.

In Section 2 we give the basic definitions and notation. In Section 3 we define residual admissible and least α-quantile estimates. In Section 4 we show that the least α-quantile estimates are minimax bias for the class of residual admissible estimates. In Section 5 we show that the class of regression admissible estimates contains M-, S-, τ-, the LMS-, the LTS- and some R-estimates. All the proofs are given in the Appendix.

2. Basic definitions and notation. Let $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, $\mathbf{x}_i \in R^p$, $y_i \in R$ be independent observations satisfying the linear regression model

$$y_i = \boldsymbol{\theta}_0' \mathbf{x}_i + u_i, \qquad 1 \le i \le n, \tag{1}$$

where the u_i have a common distribution, F_0, and are independent of the $\mathbf{x}_i$. We assume, for simplicity, the the carriers $\mathbf{x}_i$ are independent random vectors with common distribution G_0. Then

$$H_0(\mathbf{x}, y) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} F_0(y - \boldsymbol{\theta}_0'\mathbf{s})\, dG_0(\mathbf{s}) \tag{2}$$

is the distribution function of $(\mathbf{x}_i, y_i)$ under model (1). To allow for a certain fraction ε of data points which violates the "target" model (1), we consider the contamination neighborhood

$$\mathscr{H}_\varepsilon = \{H: (1-\varepsilon)H_0 + \varepsilon H^*\}, \tag{3}$$

where $0 < \varepsilon < 0.5$ and H^* is an arbitrary distribution on R^{p+1}.

Let $\mathbf{T}$ be an R^p-valued functional defined on a "large" subset of distribution functions H on R^{p+1}, which includes $\mathscr{H}_\varepsilon$ and all the empirical distribution functions H_n.

We assume that $\mathbf{T}$ is regression and affine equivariant, that is, if $\tilde{y} = y + \mathbf{x}'\mathbf{b}$ and $\tilde{\mathbf{x}} = C'\mathbf{b}$ for some full rank $p \times p$ matrix C and $\tilde{H}$ is the distribution of $(\tilde{\mathbf{x}}, \tilde{y})$, then $\mathbf{T}(\tilde{H}) = C^{-1}(\mathbf{T}(H) + \mathbf{b})$.

The asymptotic bias of $\mathbf{T}$ at $H \in \mathscr{H}_\varepsilon$ is defined as

$$b_A(\mathbf{T}, H) = (\mathbf{T}(H) - \boldsymbol{\theta}_0)' A (\mathbf{T}(H) - \boldsymbol{\theta}_0), \tag{4}$$

where $A = A(G_0)$ is an affine equivariant covariance functional, that is, if $\mathbf{x} \sim G_0$ and $\tilde{\mathbf{x}} = B\mathbf{x}$ for some nonsingular $p \times p$ matrix B, then $A(\tilde{G}_0) = BA(G_0)B'$.

If $\mathbf{T}(H)$ is not uniquely defined [i.e., $\mathbf{T}(H)$ is set valued], the definition (4) is modified as follows:

$$b_A(\mathbf{T}, H) = \sup_{\boldsymbol{\theta} \in \mathbf{T}(H)} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' A (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{5}$$

Since we only work with regression and affine equivariant estimates and (4) and (5) are invariant under regression and affine equivariant transformations, we can assume without loss of generality that $A = I$ and $\boldsymbol{\theta}_0 = 0$. Therefore,

$$b(\mathbf{T}, H) = \|\mathbf{T}(H)\|^2. \tag{6}$$

If the functional $\mathbf{T}$ 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, that is, $\mathbf{T}(H_0) = 0$. The maximum asymptotic bias of $\mathbf{T}$ over $\mathscr{H}_\varepsilon$ is defined by

$$B_{\mathbf{T}}(\varepsilon) = \sup_{H \in \mathscr{H}_\varepsilon} b(\mathbf{T}, H) = \sup_{H \in \mathscr{H}_\varepsilon} \|\mathbf{T}(H)\|^2. \tag{7}$$

Finally, given the class $\mathscr{T}$, the regression estimate $\mathbf{T}_0 \in \mathscr{T}$ is called minimax bias in $\mathscr{T}$ if

$$B_{\mathbf{T}}(\varepsilon) \geq B_{\mathbf{T}_0}(\varepsilon), \qquad \forall\, \mathbf{T} \in \mathscr{T}. \tag{8}$$

3. Residual admissible estimates. Given $\boldsymbol{\theta} \in R^p$ and $(\mathbf{x}, y) \in R^{p+1}$ with joint distribution H, let $F_{H,\boldsymbol{\theta}}(v)$ be the distribution function of $|y - \boldsymbol{\theta}'\mathbf{x}|$. The following definition is central to this paper.

Definition 3.1. The estimating regression functional $\mathbf{T}$ is residual admissible on $\mathscr{H}$ if given two possibly substochastic distributions F_1 and F_2 which are continuous on $(0, \infty)$ and satisfy

$$F_1(v) < F_2(v), \qquad \forall\, v > 0, \tag{9}$$

there are not a sequence $H_n \in \mathscr{H}$ and a vector $\boldsymbol{\theta}^* \in R^p$ such that $F_{H_n, \mathbf{T}(H_n)}(v)$ and $F_{H_n, \boldsymbol{\theta}^*}(v)$ are continuous on $(0, \infty)$ and

$$\lim_{n \to \infty} F_{H_n, \mathbf{T}(H_n)}(v) = F_1(v) \quad \text{and} \quad \lim_{n \to \infty} F_{H_n, \boldsymbol{\theta}^*}(v) = F_2(v), \qquad \forall\, v > 0.$$

The restriction to continuous distributions is needed to include M- and S-estimates with discontinuous ρ-functions, in particular, the minimax-bias estimate $\mathbf{T}_\alpha$ defined below. In any case, by weakening the definition of residual admissible estimates, we enlarge the class over which the minimax result of Section 4 holds.

Definition 3.1 seems natural, especially when $H_n = H$ for all n. If $F_{H, \mathbf{T}(H)}(v) < F_{H, \boldsymbol{\theta}^*}(v)$ for all $v > 0$, then the absolute residuals for $\boldsymbol{\theta}^*$ are stochastically smaller than those for $\mathbf{T}(H)$ and so $\boldsymbol{\theta}^*$ would appear to provide a "better fit" than $\mathbf{T}(H)$.

It will be shown in Section 5 that most of the known estimates which only depend on the regression residuals are residual admissible for an appropriate choice of $\mathscr{H}$. It will also be shown in that section (see Lemma 5.3) that bounded influence estimates which penalize high leverage observations, for example, GM- and projection estimates [see Maronna and Yohai (1989)], are not residual admissible. The minimax theory for GM-estimates can be found in Martin, Yohai and Zamar (1989).

To define the least α-quantile estimates, which are minimax-bias regression admissible estimates, let $0 < \alpha < 1$ and let $\mathbf{t} \in R^p$ be a "tentative" regression estimate. A measure of how well $\mathbf{t}$ fits the distribution of $(\mathbf{x}, y)$ under H may be given by the α-quantile of the distribution of $|y - \mathbf{t}'\mathbf{x}|$, $F_{H,\mathbf{t}}^{-1}(\alpha)$, where as usual, $F^{-1}(\alpha) = \inf\{x: F(x) \geq \alpha\}$.

DEFINITION 3.2. The least α-quantile regression estimate $\mathbf{T}_\alpha$ is defined by the property of minimizing $F_{H,\mathbf{t}}^{-1}(\alpha)$ as $\mathbf{t}$ ranges over R^p, that is,

$$\mathbf{T}_\alpha(H) = \arg\min_{t \in R^p} F_{H,\mathbf{t}}^{-1}(\alpha).$$

The following lemma shows that $\mathbf{T}_\alpha$ exists in the case of finite samples [i.e., when H is an empirical c.d.f.].

LEMMA 3.1. *Let* $(\mathbf{x}_i, y_i)$, $1 \leq i \leq n$, *be a sample with empirical c.d.f.* H_n. *Let* $0 < \alpha < 1$ *be fixed and suppose that* $\#\{i: \mathbf{t}'\mathbf{x}_i = 0\} < n\alpha$, $\forall\ \|\mathbf{t}\| = 1$. *Then there exists* $\hat{\boldsymbol{\theta}}$ (*not necessarily unique*) *such that* $F_{H_n,\hat{\boldsymbol{\theta}}}^{-1}(\alpha) \leq F_{H_n,\mathbf{t}}^{-1}(\alpha)$ *for all* $\mathbf{t}$.

The next lemma shows that under mild regularity assumptions on H_0, $\mathbf{T}_\alpha(H)$ exists for all $H \in \mathscr{H}_\varepsilon$.

LEMMA 3.2. *Suppose that* $H \in \mathscr{H}_\varepsilon$ *and* H_0 *satisfies*: (a) F_0 *has a positive density* f_0; *and* (b) *there exists* $\gamma > 0$ *such that* $\sup_{\|\mathbf{a}\|=1} P_{G_0}(\mathbf{a}'\mathbf{x} = 0) \leq (\alpha - \varepsilon) - \gamma$. *Then there exists* $\hat{\boldsymbol{\theta}}$ (*not necessarily unique*) *such that* $F_{H,\hat{\boldsymbol{\theta}}}^{-1}(\alpha) \leq F_{H,\mathbf{t}}^{-1}(\alpha)$ *for all* $\mathbf{t}$.

4. The minimax result. First, we find a lower bound, $a(\varepsilon) > 0$, for the maximum bias of all the residual admissible estimates, $\mathbf{T}$, that is, we show that

$$B_{\mathbf{T}}(\varepsilon) \geq a(\varepsilon), \tag{10}$$

for all residual admissible estimates $\mathbf{T}$. The lower bound $a(\varepsilon)$ is given by

$$a(\varepsilon) = \sup\left\{\|\boldsymbol{\theta}\|: (1-\varepsilon)F_{H_0,\boldsymbol{\theta}}(v) + \varepsilon \geq (1-\varepsilon)F_{H_0,0}(v), \forall\ v \geq 0\right\}. \tag{11}$$

Suppose that $H = (1-\varepsilon)H_0 + \varepsilon\delta_{(\mathbf{x},\boldsymbol{\theta}'\mathbf{x})}$, where $\delta_{(\mathbf{x},\boldsymbol{\theta}'\mathbf{x})}$ is a point mass distribution at $(\mathbf{x}, \boldsymbol{\theta}'\mathbf{x})$. Then $F_{H,\boldsymbol{\theta}}(v) = (1-\varepsilon)F_{H_0,\boldsymbol{\theta}}(v) + \varepsilon$ and $F_{H,0}(v) = \lim_{|\boldsymbol{\theta}'\mathbf{x}| \to \infty}(1-\varepsilon)F_{H_0,0}(v) + \varepsilon\delta_{|\boldsymbol{\theta}'\mathbf{x}|}(v) = (1-\varepsilon)F_{H_0,0}(v)$. Therefore, the distribution on the left-hand side of the inequality in (11) corresponds to the case when the "outliers" are perfectly fitted; the distribution on the right-hand side corresponds to the case when the "outliers" are completely ignored. So, $a(\varepsilon)$ is the maximum value of $\|\boldsymbol{\theta}\|$ for which the perfect fit of the "outliers" produces a better set of regression residuals than those obtained by completely ignoring them.

THEOREM 4.1. *Suppose that* H_0 *is given by* (2) *and* $f_0(v) = F_0'(v)$ *is even and strictly unimodal. Suppose that* $\mathbf{T}$ *is residual admissible on* $\mathscr{H}_\varepsilon$. *Then the inequality* (10) *holds.*

TABLE 1
Optimal quantile and maximum bias of the minimax estimate for the Gaussian central model

ε	**Quantile**	$B_{\mathbf{T}}(\varepsilon)$
0.05	0.67	0.49
0.10	0.66	0.77
0.15	0.65	1.05
0.20	0.64	1.37

A natural question at this point is whether the lower bound for $B_{\mathbf{T}}(\varepsilon)$ given by Theorem 4.1 can be attained. In the next theorem we show that $\alpha(\varepsilon)$ can be attained when $G_0(\mathbf{x})$ is elliptical. Unfortunately, the problem becomes more complex for general $G_0(\mathbf{x})$ and we were not able to find a solution. Observe that when $G_0(\mathbf{x})$ is nonelliptical $b_A(\mathbf{T}, H)$ is not necessarily a function of $\|\mathbf{T}(H)\|^2$ and thus the minimax estimate, if it exists, may depend on the particular shape of G_0.

THEOREM 4.2. *Suppose that H_0 is given by* (2) *and $f_0(v) = F_0'(v)$ is even and strictly unimodal. Suppose also that $P_{G_0}(\boldsymbol{\theta}'\mathbf{x} = 0) < 1$, $\forall$ $\|\boldsymbol{\theta}\| = 1$, and $G_0(\mathbf{x})$ is elliptical. Then there exists α^* such that the least α^*-quantile estimate $\mathbf{T}_{\alpha^*}$ is minimax bias in the class of residual admissible estimates. The value α^* is explicitly given in the proof.*

Table 1, taken from Martin, Yohai and Zamar (1989), gives the values of the optimal quantiles and the corresponding maximum asymptotic biases of the minimax-bias estimates for several values of ε, when H_0 is Gaussian. Observe that the value of α does not change much with ε.

5. A general class of residual admissible estimates. In this section we show that several classes of robust estimates (including LMS-, M-, S- and τ-estimates) are residual admissible. We also define a class of estimates containing the LTS- and some signed R-estimates and show that these estimates are residual admissible.

Let $F_{H,\boldsymbol{\theta},s}$ be the distribution function of $|(y - \boldsymbol{\theta}'\mathbf{x})/s|$ under H, that is, $F_{H,\boldsymbol{\theta},s}(v) = F_{H,\boldsymbol{\theta}}(sv)$. We consider the class of regression functionals

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} J\big(F_{H,\boldsymbol{\theta},s(H)}\big), \tag{12}$$

where $J(F)$ is defined on a set of distribution functions on $[0, \infty)$ containing the empirical distributions, and $s(H)$ is either an estimating functional of the error scale or simply $s(H) = 1$. Notice that $s(H)$ can be computed separately or simultaneously with the regression coefficients. In the former case (12) is not the operating definition of $\mathbf{T}(H)$ but rather a property satisfied by $\mathbf{T}(H)$ after $\mathbf{T}(H)$ and $s(H)$ have been simultaneously determined.

The following "monotonicity" property of $J(F)$ is used to prove the regression admissibility of an estimate $\mathbf{T}(H)$ satisfying (12).

DEFINITION 5.1. A functional $J(F)$ is ε-monotone if given two sequences of distribution functions on $[0,\infty)$, F_n and G_n, which are continuous on $(0,\infty)$ and such that $F_n(u) \to F(u)$ and $G_n(u) \to G(u)$, where F and G are possibly substochastic and continuous on $(0,\infty)$, with $G(\infty) \geq 1 - \varepsilon$ and

$$G(u) > F(u), \qquad \forall\, u > 0, \tag{13}$$

then $\lim_{n\to\infty} J(F_n) > \lim_{n\to\infty} J(G_n)$.

THEOREM 5.1. *Let* $\mathbf{T}(H)$ *be an estimating functional which satisfies* (12). *Suppose that* J *is* ε*-monotone and* $s(H)$ *is such that* $\sup_{H\in\mathscr{H}_\varepsilon} s(H) < \infty$ *and* $\inf_{H\in\mathscr{H}_\varepsilon} s(H) > 0$. *Then* $\mathbf{T}(H)$ *is residual admissible on* $\mathscr{H}_\varepsilon$.

Many classes of robust regression estimates are residual admissible by virtue of satisfying (12) with an ε-monotone J.

M-estimates. The class of M-estimates with general scale is defined by

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} E_H\left(\rho\left(\frac{y - \mathbf{T}(H)'\mathbf{x}}{s(H)}\right)\right), \tag{14}$$

where ρ is an appropriate loss function and $s(H)$ is a scale estimating functional. The following lemma and Theorem 5.1 show that M-estimates with bounded ρ are residual admissible.

LEMMA 5.1. *Suppose that* $\rho(v)$ *is even, monotone on* $[0,\infty)$, *bounded, continuous at* 0 *and at* ∞ *and* $0 = \rho(0) < \rho(\infty)$. *Then* $J(F) = \int_0^\infty \rho(v)\,dF(v)$ *is* ε*-monotone for all* $\varepsilon > 0$.

Although Lemma 5.1 excludes M-estimates with unbounded ρ (e.g., L_1-estimates), the following lemma shows that the maximum bias of such estimates is infinite for all $\varepsilon > 0$ and so they can be excluded from the bias-robustness point of view.

LEMMA 5.2. *Consider M-estimates* $\mathbf{T}$ *defined by* (14) *with* $0 < a_1 = \inf_{H\in\mathscr{H}_\varepsilon} s(H)$ *and* $a_2 = \sup_{H\in\mathscr{H}_\varepsilon} s(H) < \infty$. *Suppose that* ρ *is as in Lemma* 5.1 *except that* $\lim \rho(u) = \infty$. *If*

$$E_{H_0}\rho(\alpha|y| + \beta\|bx\|) < \infty, \qquad \forall\, \alpha, \beta \in R, \tag{15}$$

then $B_{\mathbf{T}}(\varepsilon) = \infty$, $\forall\, \varepsilon > 0$.

S-estimates. S-estimates are defined by $\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} S(F_{H,\boldsymbol{\theta}})$, with $S(F)$ defined by

$$E_F\left(\rho\left(\frac{u}{S(F)}\right)\right) = b. \tag{16}$$

The function ρ has the same properties as in the case of M-estimates. Rousseeuw and Yohai (1984) show that S-estimates minimize (14) with $s(H) = \min_{\boldsymbol{\theta}} S(F_{H,\boldsymbol{\theta}})$. Therefore, the monotonicity property for S-estimates follows directly from Lemma 5.1.

It can be readily seen that Rousseeuw's LMS is an S-estimate with $\rho(t) = 0$ for $|t| < 1$ and $\rho(t) = 1$ otherwise and $b = 0.5$. Therefore, the LMS also satisfies (12) with a monotone $J(F)$.

τ-estimates. Yohai and Zamar (1988) define τ-estimates as $\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} \tau(F_{H,\boldsymbol{\theta}})$, where

$$\tau^2(F) = s^2(F) E_F\left(\rho_1\left(\frac{u}{S(F)}\right)\right),$$

and $S(F)$ is given by (16). As in the case of M-estimates, it can be proved that if ρ and ρ_1 are even, monotone on $[0,\infty)$, nonconstant and bounded, then $J(F) = \tau(F)$ is ε-monotone for all $\varepsilon > 0$.

LTS-estimates and R-estimates. R-estimates can be based on unsigned or signed ranks. It was observed by Jaeckel (1972) that R-estimates based on unsigned ranks minimize a dispersion measure D of the residuals. Given a distribution F, $D(F) = \int_{-\infty}^{\infty} a(F(u))u\,dF(u)$, where the score function $a(v)$: $[0,1] \to R$ is nondecreasing and $a(1-v) = -a(v)$. Jaeckel (1972) showed that D is a dispersion measure and proposed to define (unsigned) R-estimates as

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} D(F^*_{H,\boldsymbol{\theta}}), \tag{17}$$

where $F^*_{H,\boldsymbol{\theta}}$ is the distribution of $(y - \boldsymbol{\theta}'\mathbf{x})$ under H. Since D is a dispersion measure, $\mathbf{T}(H)$ does not necessarily have small residuals and consequently these R-estimators are not residual admissible.

Hössjer (1991) defined signed R-estimates using the estimating functional

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} J(F_{H,\boldsymbol{\theta}}), \tag{18}$$

where

$$J(F) = \int_0^{\infty} a(F(u))u\,dF(u), \qquad a(u) \geq 0. \tag{19}$$

The finite sample version of this estimate is given by

$$\mathbf{T}(H_n) = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} a(R_i^+(\boldsymbol{\theta})/n)|y_i - \boldsymbol{\theta}'\mathbf{x}_i|, \tag{20}$$

where $R_i^+(\boldsymbol{\theta})$ is the rank of $|y_i - \boldsymbol{\theta}'\mathbf{x}_i|$ among $|y_1 - \boldsymbol{\theta}'\mathbf{x}_1|, \dots, |y_n - \boldsymbol{\theta}'\mathbf{x}_n|$.

Differentiating the sum in (20) we get the estimating equation

$$\sum_{i=1}^{n} a(R_i^+/n)\operatorname{sign}(y_i - \boldsymbol{\theta}'\mathbf{x}_i)\mathbf{x}_i = 0.$$

Estimates related to the solution of this equation were also studied by van Eeden (1972) and Kraft and van Eeden (1972).

It can be shown that if $a(v)$ is monotone, the estimates defined by (18) and (19) are not robust, that is, $B_{\mathbf{T}}(\varepsilon) = \infty$ for all $\varepsilon > 0$. The proof is similar to Lemma 5.2. Notice that the usual score functions [e.g., Wilcoxon, $a(v) = v$, and normal scores, $a(v) = \Phi^{-1}(v)$, where Φ is the $N(0,1)$ distribution function] are monotone.

Hössjer (1991) also observed that if $a(v)$ vanishes outside the interval $[0, 1-\alpha]$, where $0.5 \le \alpha < 1$ the estimate defined by (18) and (19) has breakdown point α, that is, $B_{\mathbf{T}}(\varepsilon) < \infty$ for all $\varepsilon < \alpha$. An interesting estimate of this type which we call the α-least trimmed absolute value (α-LTAV) estimate is defined by taking

$$a(v) = \begin{cases} 1, & \text{if } |v| \le 1-\alpha, \\ 0, & \text{if } |v| > 1-\alpha. \end{cases}$$

The sampling version of this estimate is obtained by minimizing

$$\sum_{i=1}^{[n(1-\alpha)]} |r(\boldsymbol{\theta})|_{(i)}, \tag{21}$$

where $|r(\boldsymbol{\theta})|_{(i)}$ is the ith-order statistic of $|r_1(\boldsymbol{\theta})|, \dots, |r_n(\boldsymbol{\theta})|$.

Observe that Rousseeuw's LTS-estimates are defined as in (21) with absolute value replaced by square. This suggests the definition of a larger class of estimates which includes signed R- and LTS-estimates. This class is defined by (18) with

$$J(F) = \int_0^\infty a(F(u))u^k\,dF(u). \tag{22}$$

The next theorem shows that estimates in this class (including LTS- and LTAV-estimates) are residual admissible.

THEOREM 5.2. *Suppose that* (i) $a(u)$ *is continuous on* $[0, 1-\alpha]$, (ii) $a(u) = 0$ *if* $1-\alpha < u \le 1$ *and* (iii) $a(u) > 0$ *if* $0 < u < 1-\alpha$. *Then* $J(F)$ *defined by* (22) *is* ε*-monotone for all* $\varepsilon < \alpha$.

Bounded influence estimates. Finally, we show that bounded influence estimates of regression are not residual admissible. It is enough to show that the minimax-bias regression admissible estimates $\mathbf{T}_\alpha$ have unbounded influence. This immediately follows from Lemma 5.3.

LEMMA 5.3. *Suppose that G_0 is elliptical, $f_0(u) = F_0'(u)$ is even and strictly unimodal and $f_0'(F_0^{-1}((1+\alpha)/2)) > 0$. Then*

$$0 < \lim_{\varepsilon \to 0} \frac{B_{\mathbf{T}_\alpha}(\varepsilon)}{\sqrt{\varepsilon}} < \frac{1}{f_0'\big(F_0^{-1}((1+\alpha)/2)\big)} < \infty.$$

APPENDIX

PROOF OF LEMMA 3.1. It suffices to show that (i) $F_{H_n,\mathbf{t}}^{-1}(\alpha)$ is a continuous function of $\mathbf{t}$ and (ii) $F_{H_n,\mathbf{t}}^{-1}(\alpha) \to \infty$ as $\|\mathbf{t}\| \to \infty$. To prove (i), let $\mathbf{t}_0$ be fixed, let $0 \le r_1 < r_2 < \cdots < r_k < \infty$ be the discontinuity points of $F_{H_n,\mathbf{t}_0}$ and let $1 \le m_i \le n$, $i = 1,\dots,k$, be such that $F_{H_n,\mathbf{t}_0}(r_i) = m_i/n$, $i = 1,\dots,k$. Put $m_0 = 0$. Since $m_k = n$, there exists $1 \le i \le k$ such that $m_{i-1} < n\alpha \le m_i$ and so $F_{H_n,\mathbf{t}_0}^{-1}(\alpha) = \inf\{x\colon F_{H_n,\mathbf{t}_0}(x) \ge \alpha\} = r_i$. Given $\delta > 0$, there exist $0 < \delta_1 < \delta$ and $0 < \delta_2 < \delta$ such that $r_i - \delta_1$ and $r_i + \delta_2$ are continuity points of $F_{H_n,\mathbf{t}_0}$. Clearly, $F_{H_n,\mathbf{t}_0}(r_i + \delta_2) \ge \alpha$ and $F_{H_n,\mathbf{t}_0}(r_i - \delta_1) < \alpha$. Now let $0 < \Delta < 1/n$ be fixed. Since $F_{H_n,\mathbf{t}}$ converges in law to $F_{H_n,\mathbf{t}_0}$ as $\mathbf{t} \to \mathbf{t}_0$, $r_i - \delta_1$ and $r_i + \delta_2$ are continuity points of $F_{H_n,\mathbf{t}_0}$, there exists $\gamma > 0$ such that $\|\mathbf{t} - \mathbf{t}_0\| < \gamma$ implies $|F_{H_n,\mathbf{t}_0}(r_i - \delta_1) - F_{H_n,\mathbf{t}}(r_i - \delta_1)| \le \Delta$ and $|F_{H_n,\mathbf{t}_0}(r_i + \delta_2) - F_{H_n,\mathbf{t}}(r_i + \delta_2)| \le \Delta$. Since $\Delta < 1/n$ we get $\alpha < F_{H_n,\mathbf{t}_0}(r_i - \delta_1) = F_{H_n,\mathbf{t}}(r_i - \delta_1)$ and $\alpha \ge F_{H_n,\mathbf{t}_0}(r_i + \delta_2) = F_{H_n,\mathbf{t}}(r_i + \delta_2)$. Therefore, $r_i - \delta_1 \le F_{H_n,\mathbf{t}}^{-1}(\alpha) \le r_i + \delta_2$, and (i) follows. To prove (ii), let $d_i(\mathbf{t}) = |\mathbf{t}'\mathbf{x}_i|$ and let $R_1,\dots,R_n$ be defined by $d_{(i)}(\mathbf{t}) = |\mathbf{t}'\mathbf{x}_{R_i}|$, $i = 1,\dots,n$, where $d_{(1)}(\mathbf{t}) \le d_{(2)}(\mathbf{t}) \le \cdots \le d_{(n)}(\mathbf{t})$. Let $1 \le m \le n$ be such that $(m-1) < \alpha n \le m$. By assumption $d_{(m)}(\mathbf{t}) > 0$, $\forall\ \|\mathbf{t}\| = 1$. Therefore, using the continuity of $d_{(m)}(\mathbf{t})$, we have that

$$\min_{\|\mathbf{t}\|=1} d_{(m)}(\mathbf{t}) = d_{(m)}(\mathbf{t}_0^*) = \delta_0 > 0,$$

for some $\mathbf{t}_0^*$ with $\|\mathbf{t}_0^*\| = 1$. Finally, for all $\lambda > 0$ and $\|\mathbf{t}\| = 1$, we have

$$\begin{aligned} |y_{R_i} - \lambda\mathbf{t}'\mathbf{x}_{R_i}| &\ge \lambda|\mathbf{t}'\mathbf{x}_{R_i}| - |y_{R_i}| = \lambda d_{(i)}(\mathbf{t}) - |y_{R_i}| \\ &\ge \lambda\delta_0 - \max_{1 \le i \le n}|y_i|, \forall\ i = m,\dots,n, \end{aligned}$$

and so $F_{H_n,\lambda\mathbf{a}}^{-1}(\alpha) \ge \lambda\delta_0 - \max_{1\le i\le n}|y_i| \to \infty$ as $\lambda \to \infty$. □

PROOF OF LEMMA 3.2. This proof follows along the lines of the proof of Lemma 3.1. The continuity of $F_{H,\mathbf{t}}^{-1}(\alpha)$ follows from the strict monotonicity of $F_{H,\mathbf{t}}(v)$ which in turn follows directly from (a).

To show that $F_{H,\mathbf{t}}^{-1}(\alpha) \to \infty$ as $\|\mathbf{t}\| \to \infty$, it suffices to show that for all $K > 0$, $F_{H_0,\mathbf{t}}(K) < \alpha - \varepsilon$ for $\|\mathbf{t}\|$ sufficiently large. Let $\mathbf{t} = \lambda\mathbf{a}$ with $\|\mathbf{a}\| = 1$. Since $P_{F_0}(|y| > v) \to 0$ as $v \to \infty$, it is enough to show that $\inf_{\|\mathbf{a}\|=1} P_{G_0}(|\mathbf{a}'\mathbf{x}| > \delta_0) > 1 - \alpha - \varepsilon$ for some $\delta_0 > 0$. But this follows directly from (b) using a standard compactness argument. □

Lemmas A.1, A.2 and A.3 are needed to prove Theorems 4.1 and 4.2.

LEMMA A.1. *Suppose that* H_0 *is given by* (2), $f_0(v) = F_0'(v)$ *is even and strictly unimodal and* $P_{G_0}(\boldsymbol{\theta}'\mathbf{x} = 0) < 1$ *for all* $\|\boldsymbol{\theta}\| = 1$. *Then, for all* $\|\boldsymbol{\theta}\| = 1$, $\lambda > 0$ *and* $v > 0$, $F_{H_0, \lambda\boldsymbol{\theta}}(v)$ *is strictly decreasing in* λ.

PROOF. Let $v > 0$ and $\boldsymbol{\theta}$, with $\|\boldsymbol{\theta}\| = 1$, be fixed. If $g(a) = F_0(v + a) - F_0(-v + a)$, then $F_{H_0, \lambda\boldsymbol{\theta}}(v) = P_{H_0}(-v \le y - \lambda\boldsymbol{\theta}'\mathbf{x} \le v) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\lambda\boldsymbol{\theta}'\mathbf{x})\, dG_0(\mathbf{x})$. Notice that

$$(\partial/\partial\lambda) g(\lambda\boldsymbol{\theta}'\mathbf{x}) = \left(f_0(v + \lambda\boldsymbol{\theta}'\mathbf{x}) - f_0(-v + \lambda\boldsymbol{\theta}'\mathbf{x})\right)\boldsymbol{\theta}'\mathbf{x} < 0, \qquad \forall\ |\boldsymbol{\theta}'\mathbf{x}| > 0,$$

and so if $\lambda_1 > \lambda_2$ then $g(\lambda_2\boldsymbol{\theta}'\mathbf{x}) < g(\lambda_1\boldsymbol{\theta}'\mathbf{x})$, $\forall\ |\boldsymbol{\theta}'\mathbf{x}| > 0$. Thus, $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\lambda_2\boldsymbol{\theta}'\mathbf{x})\, dG_0(\mathbf{x}) < \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\lambda_1\boldsymbol{\theta}'\mathbf{x})\, dG_0(\mathbf{x})$, and the lemma follows. □

LEMMA A.2. *Let* $\gamma = \sup_{\|\boldsymbol{\theta}\|=1} P_{G_0}(\mathbf{x}'\boldsymbol{\theta} = 0)$. *Then we have the following*: (a) *For each* $0 < v < \infty$, $\lim_{\|\boldsymbol{\theta}\| \to \infty} F_{H_0, \boldsymbol{\theta}}(v) \le \gamma$; *and* (b) $F_{H_0, \boldsymbol{\theta}}(v)$ *is continuous in* $\boldsymbol{\theta}$, *uniformly on* $v \in [0, \infty)$.

PROOF. Suppose that for some $0 < v < \infty$ there exists a sequence $\boldsymbol{\theta}_n$ such that $\|\boldsymbol{\theta}_n\| \to \infty$ and

$$\lim_{n \to \infty} F_{H_0, \boldsymbol{\theta}_n}(v) > \gamma, \tag{23}$$

for some $v > 0$. Let $\mathbf{t}_n = \boldsymbol{\theta}_n/\|\boldsymbol{\theta}_n\|$ and without loss of generality assume that $\mathbf{t}_n \to \mathbf{b}$ as $n \to \infty$. Clearly, for all $K > v$,

$$\{(\mathbf{x}, y): |y - \boldsymbol{\theta}_n'\mathbf{x}| < v\} \subset \{(\mathbf{x}, y): |\boldsymbol{\theta}'\mathbf{x}| \le K\} \cup \{(\mathbf{x}, y): |y| > K - v\}. \tag{24}$$

For $\delta > 0$ (arbitrary) choose $K_0 > v$ such that

$$P(|y| > K_0 - v) < \delta. \tag{25}$$

Furthermore, by assumption

$$\lim_{n \to \infty} P(|\mathbf{x}'\boldsymbol{\theta}_n| \le K_0) = \lim_{n \to \infty} P\left(|\mathbf{x}'\mathbf{t}_n| \le \frac{K_0}{\|\theta_n\|}\right) = P(|\mathbf{x}'\mathbf{b}| = 0) \le \gamma. \tag{26}$$

Using (24), (25) and (26), we get $\lim_{n \to \infty} F_{H_0, \boldsymbol{\theta}_n}(v) \le \gamma + \delta$ for all $\delta > 0$. The last inequality contradicts (23), proving (a). The proof of (b) is straightforward and not included here. □

LEMMA A.3. *Suppose that* $F_0(v)$ *is continuous and let* γ *be defined as in Lemma* A.2. *Then we have the following*: (a) $0 < a(\varepsilon)$ *for all* $0 < \varepsilon < 0.5$; (b) $a(\varepsilon) < \infty$ *for all* $\varepsilon < (1 - \gamma)/(2 - \gamma)$; *and* (c) *there exists* $\boldsymbol{\theta}^*$ *with* $\|\boldsymbol{\theta}^*\| = a(\varepsilon)$ *and* $v^* > 0$ *such that*

$$(1 - \varepsilon) F_{H_0, \boldsymbol{\theta}^*}(v) + \varepsilon \ge (1 - \varepsilon) F_{H_0, 0}(v), \qquad \forall\ v > 0, \tag{27}$$

and

$$(1 - \varepsilon) F_{H_0, \boldsymbol{\theta}^*}(v^*) + \varepsilon = (1 - \varepsilon) F_{H_0, 0}(v^*). \tag{28}$$

PROOF. Since 0 belongs to the set used in the definition of $a(\varepsilon)$, this quantity is well defined. By Lemma A.2(b), there exists $\delta > 0$ such that $F_{H_0,\boldsymbol{\theta}}(v) \geq F_{H_0,0}(v) - \varepsilon/2$ for all $\|\boldsymbol{\theta}\| < \delta$ and $v \geq 0$. Suppose now that $\|\boldsymbol{\theta}\| < \delta$. Then, for all $v \geq 0$,

$$
\begin{aligned}
(1-\varepsilon)F_{H_0,\boldsymbol{\theta}}(v) + \varepsilon &> (1-\varepsilon)\big(F_{H_0,0}(v) - \varepsilon/2\big) + \varepsilon \\
&= (1-\varepsilon)F_{H_0,0}(v) + \varepsilon - \varepsilon(1-\varepsilon)/2 \geq (1-\varepsilon)F_{H_0,0}(v).
\end{aligned}
$$

Therefore, $a(\varepsilon) > \delta$ and (a) holds. To prove (b), notice that by assumption $(1-\varepsilon)\gamma + \varepsilon < (1-\varepsilon)$. Therefore, there exists $\delta > 0$ such that

$$(1-\varepsilon)\gamma + \varepsilon < (1-\varepsilon) - \delta. \tag{29}$$

Let now $0 < v_0 < \infty$ be such that

$$(1-\varepsilon)F_{H_0,0}(v_0) > (1-\varepsilon) - \delta/2. \tag{30}$$

By Lemma A.2(a) and (29),

$$\lim_{\|\boldsymbol{\theta}\|\to\infty} (1-\varepsilon)F_{H_0,\boldsymbol{\theta}}(v_0) + \varepsilon \leq (1-\varepsilon)\gamma + \varepsilon < (1-\varepsilon) - \delta.$$

Therefore, there exists $0 < K < \infty$ such that $\|\theta\| \geq K$ implies

$$(1-\varepsilon)F_{H_0,\boldsymbol{\theta}}(v_0) + \varepsilon < (1-\varepsilon) - \delta/2. \tag{31}$$

Equations (30) and (31) imply that $a(\varepsilon) < K$ and (b) follows. Finally, to prove (c), notice that by the definition of $a(\varepsilon)$ and (b), we can find a sequence $\boldsymbol{\theta}_n$ such that $\lim_{n\to\infty}\boldsymbol{\theta}_n = \boldsymbol{\theta}^*$, $\|\boldsymbol{\theta}^*\| = a(\varepsilon)$ and

$$(1-\varepsilon)F_{H_0,\boldsymbol{\theta}_n}(v) + \varepsilon \geq (1-\varepsilon)F_{H_0,0}(v), \qquad \forall\, v \geq 0. \tag{32}$$

Since by Lemma A.2(b) $F_{H_0,\boldsymbol{\theta}}(v)$ is continuous in $\boldsymbol{\theta}$, it follows

$$(1-\varepsilon)F_{H_0,\boldsymbol{\theta}^*}(v) + \varepsilon \geq (1-\varepsilon)F_{H_0,0}(v), \qquad \forall\, v \geq 0. \tag{33}$$

To prove the existence of v^* satisfying (28), suppose that (33) holds with strict inequality for all $v \geq 0$. We will show that in such case there exists $\delta > 0$ such that

$$(1-\varepsilon)F_{H_0,\boldsymbol{\theta}^*}(v) + \varepsilon > (1-\varepsilon)F_{H_0,0}(v) + \delta, \qquad \forall\, v \geq 0. \tag{34}$$

To show this, we begin by finding $v_0 > 0$ such that

$$(1-\varepsilon)F_{H_0,\boldsymbol{\theta}^*}(v) + \varepsilon > 1 - \varepsilon/2, \qquad \forall\, v \geq v_0. \tag{35}$$

Since $F_{H_0,\boldsymbol{\theta}^*}(v)$ and $F_{H_0,0}(v)$ are uniformly continuous functions of v on $[0, v_0]$, there exists $\delta' > 0$ such that

$$(1-\varepsilon)F_{H_0,\boldsymbol{\theta}^*}(v) + \varepsilon > (1-\varepsilon)F_{H_0,0}(v) + \delta', \qquad \forall\, 0 \leq v \leq v_0, \tag{36}$$

and (34) follows now from (35) and (36) with $\delta = \min\{\delta', \varepsilon/2\}$.

Finally, using (34) and Lemma A.2(b), we can find $\boldsymbol{\theta}^{**}$ with $\|\boldsymbol{\theta}^{**}\| > \|\boldsymbol{\theta}^{*}\|$ such that

$$(37) \qquad (1-\varepsilon)F_{H_0,\boldsymbol{\theta}^{**}}(v) + \varepsilon \geq (1-\varepsilon)F_{H_0,0}(v), \qquad \forall\, v \geq 0.$$

But this contradicts the definition of $a(\varepsilon)$, proving (c) and the lemma. □

PROOF OF THEOREM 4.1. Let $\mathbf{T}$ be residual admissible, let $b < a(\varepsilon)$ and let $\boldsymbol{\theta}^*$ be as in Lemma A.1. Take $\mathbf{x}_n = n\boldsymbol{\theta}^*/a(\varepsilon)$ and let y_n be uniformly distributed on the interval $[nb - (1/n), nb + (1/n)]$. Let

$$(38) \qquad H_n = (1-\varepsilon)H_0 + \varepsilon \tilde{H}_n,$$

where $\tilde{H}_n$ is the point mass distribution at $(\mathbf{x}_n, y_n)$. Consequently,

$$(39) \qquad F_{H_n,\boldsymbol{\theta}}(v) = (1-\varepsilon)F_{H_0,\boldsymbol{\theta}}(v) + \varepsilon F_{\tilde{H}_n,\boldsymbol{\theta}}(v).$$

We will show now that

$$(40) \qquad \liminf_{n\to\infty} \|\mathbf{T}(H_n)\| \geq b.$$

A straightforward computation shows that

$$(41) \qquad F_{\tilde{H}_n,\boldsymbol{\theta}}(v) = U_n(v - c_n(\boldsymbol{\theta})) - U_n(-v - c_n(\boldsymbol{\theta})), \qquad \forall\, v \geq 0,$$

where U_n denotes the uniform distribution function on $[-(1/n), (1/n)]$ and where

$$(42) \qquad c_n(\boldsymbol{\theta}) = n\left(b - \frac{\boldsymbol{\theta}'\boldsymbol{\theta}^*}{a(\varepsilon)}\right).$$

Observe that by (39) and (41), $F_{H_n,\boldsymbol{\theta}}(v)$ is continuous in v. Suppose now that (40) does not hold and let $\mathbf{T}_n = \mathbf{T}(H_n)$. Passing to a subsequence if necessary, we can write

$$(43) \qquad \lim_{n\to\infty} \mathbf{T}_n = \tilde{\boldsymbol{\theta}}, \qquad \|\tilde{\boldsymbol{\theta}}\| = \tilde{b} < b.$$

Since $\lim_{n\to\infty}|b - \mathbf{T}_n'\boldsymbol{\theta}^*/a(\varepsilon)| \geq b - \tilde{b} > 0$, it follows that

$$(44) \qquad \lim_{n\to\infty} |c_n(\mathbf{T}_n)| = \infty.$$

Using this and (41), we obtain

$$(45) \qquad \lim F_{\tilde{H}_n,\mathbf{T}_n}(v) = 0, \qquad \forall\, v \geq 0,$$

and so, by (39), (43) and (45), we have

$$(46) \qquad \lim_{n\to\infty} F_{H_n,\mathbf{T}_n}(v) = (1-\varepsilon)F_{H_0,\tilde{\boldsymbol{\theta}}}(v), \qquad \forall\, v \geq 0.$$

On the other hand, if $\bar{\boldsymbol{\theta}} = (b/a)\boldsymbol{\theta}^*$, then $c_n(\bar{\boldsymbol{\theta}}) = 0$ for all $n \geq 1$, and thus

$$(47) \qquad \lim_{n\to\infty} F_{\tilde{H}_n,\bar{\boldsymbol{\theta}}}(v) = 1, \qquad \forall\, v \geq 0.$$

Therefore, by this and (39), we get

$$\lim_{n\to\infty} F_{H_n,\tilde{\theta}}(v) = (1-\varepsilon)F_{H_0,\tilde{\theta}}(v) + \varepsilon, \qquad \forall\, v \geq 0. \tag{48}$$

Now, by Lemmas A.1 and A.3(c), we have

$$\begin{aligned}(1-\varepsilon)F_{H_0,\tilde{\theta}}(v) + \varepsilon &> (1-\varepsilon)F_{H_0,\theta^*}(v) - \varepsilon \geq (1-\varepsilon)F_{H_0,0}(v) \\ &> (1-\varepsilon)F_{H_0,\tilde{\theta}}(v), \qquad \forall\, v > 0.\end{aligned} \tag{49}$$

Finally, (46), (48) and (49) contradict the admissibility assumption and thus (40) holds. Since (40) holds for all $b < a(\varepsilon)$, the theorem follows. □

Proof of Theorem 4.2. We will show that Theorem 4.2 holds with

$$\alpha^* = (1-\varepsilon)F_{H_0,0}(v^*), \tag{50}$$

where v^* is defined in Lemma A.3.

Without loss of generality we can assume that $G_0(\mathbf{x})$ is spherical. Moreover, by Theorem 4.1 it is enough to show $B_{\mathbf{T}_{\alpha^*}}(\varepsilon) \leq a(\varepsilon)$ and for this it suffices to show that for all $H = (1-\varepsilon)H_0 + \varepsilon\tilde{H}$ we have

$$\|\mathbf{T}_{\alpha^*}(H)\| \leq a(\varepsilon). \tag{51}$$

Suppose that this does not hold, and so there exists some H such that if $\theta = \mathbf{T}_{\alpha^*}(H)$, then $\|\mathbf{T}_{\alpha^*}(H)\| = \|\theta\| > \|\theta^*\| = a(\varepsilon)$, where $\|\theta^*\|$ is as defined in Lemma A.3.

Let $\lambda = \|\theta\|/\|\theta^*\|$ and $\tilde{\theta} = \lambda\theta^*$. Then $\|\tilde{\theta}\| = \|\theta\| > \|\theta^*\|$. Since G_0 is spherical, the distributions of $\theta'\mathbf{x}$ and $\tilde{\theta}'\mathbf{x}$ are the same. Thus, by Lemma A.1 and (28),

$$\begin{aligned}F_{H,\theta}(v^*) &\leq (1-\varepsilon)F_{H_0,\theta}(v^*) + \varepsilon = (1-\varepsilon)F_{H_0,\tilde{\theta}}(v^*) + \varepsilon \\ &< (1-\varepsilon)F_{H_0,\theta^*}(v^*) + \varepsilon = (1-\varepsilon)F_{H_0,0}(v^*) = \alpha^*.\end{aligned} \tag{52}$$

Therefore,

$$F_{H,\theta}^{-1}(\alpha^*) > v^*. \tag{53}$$

On the other hand, $F_{H,0}(v^*) \geq (1-\varepsilon)F_{H_0,0}(v^*) = \alpha^*$, and so

$$F_{H,0}^{-1}(\alpha^*) \leq v^*. \tag{54}$$

Finally, (53) and (54) show that $\mathbf{T}_{\alpha^*}(H) \neq \theta$, proving the theorem. □

Proof of Theorem 5.1. Suppose that the admissibility condition is violated. Then there exist two possibly substochastic distribution functions F^* and F, continuous on $(0,\infty)$ and satisfying

$$F^*(v) > F(v), \qquad \forall\, v > 0, \tag{55}$$

together with a sequence H_n in $\mathscr{H}_\varepsilon$ and a vector θ^* such that $F_{H_n,\mathbf{T}(H_n)}(v)$ and $F_{H_n,\theta^*}(v)$ are both continuous on $(0,\infty)$ and

$$F_{H_n,\mathbf{T}(H_n)}(v) \to F(v), \qquad \forall\, v > 0, \tag{56}$$

and

$$F_{H_n,\theta^*}(v) \to F^*(v), \qquad \forall\, v > 0. \tag{57}$$

Without loss of generality we can assume that

$$0 < \lim_{n\to\infty} s(H_n) = s < \infty. \tag{58}$$

Using the continuity of the distribution functions together with (55), (57) and (58), it follows that

$$F_{H_n,\mathbf{T}(H_n),\, s(H_n)}(v) \to F_s(v), \qquad \forall\, v > 0, \tag{59}$$

and

$$F_{H_n,\theta^*,\, s(H_n)}(v) \to F_s^*(v), \qquad \forall\, v > 0, \tag{60}$$

where $F_s(v) = F(sv)$ and $F_s^*(v) = F^*(sv)$. Since $F_s^*(\infty) \geq 1 - \varepsilon$, from (55), (59), (60) and the ε-monotonicity of $J(F)$, we have

$$\lim_{n\to\infty} J\big(F_{H_n,\mathbf{T}(H_n),\, s(H_n)}\big) > \lim_{n\to\infty} J\big(F_{H_n,\theta^*,\, s(H_n)}\big),$$

contradicting (13). Therefore, the theorem holds. □

Proof of Lemma 5.1. Let F, G, F_n and G_n be as in Definition 5.1. Observe that the discontinuities of ρ and the distribution functions cannot occur at the same points. So, using by-part integration and the dominated convergence theorem,

$$\lim_{n\to\infty} J(F_n) = \lim_{n\to\infty} \int_0^\infty \rho(v)\, dF_n(v) = \lim_{n\to\infty} \int_0^\infty (1 - F_n(v))\, d\rho(v)$$

$$= \int_0^\infty (1 - F(v))\, d\rho(v) = \int_0^\infty \rho(v)\, dF(v) + (1 - F(\infty))\rho(\infty).$$

Analogously, $\lim_{n\to\infty} J(G_n) = \int_0^\infty \rho(v)\, dG(v) + (1 - G(\infty))\rho(\infty)$. The lemma follows then because

$$\lim_{n\to\infty} J(F_n) - \lim_{n\to\infty} J(G_n) = \int_0^\infty \rho(v)\, d(F - G)(v) + (G(\infty) - F(\infty))\rho(\infty)$$

$$= \int_0^\infty (G(v) - F(v))\, d\rho(v) > 0. \qquad \square$$

Proof of Lemma 5.2. Suppose that

$$B_{\mathbf{T}}(\varepsilon) = b < \infty. \tag{61}$$

By (15), $\mu_n = E_{H_0}\rho((|y| + n|x_1|)/a_1)$ is finite for all $n \geq 1$. Using the unboundedness of ρ, we can choose y_n such that $\rho(y_n/2a_2) > \mu_n/\varepsilon$. Let $\mathbf{v}' = (1, 0, \ldots, 0)$, $\mathbf{x}_n = (y_n/n)\mathbf{v}$ and H_n be given by (38) with $\tilde{H}_n$ equal to a point-mass at $(\mathbf{x}_n, y_n)$. By (61), $\|\mathbf{T}(H_n\| \le b$ and, therefore, for large n,

$$E_{H_n\rho}\left(\frac{y - \mathbf{T}(H_n)'\mathbf{x}}{s(H_n)}\right) \ge \varepsilon\rho\left(\frac{y_n - T_{1y_n}/n}{a_2}\right) \ge \varepsilon\rho\left(\frac{y_n}{a_2}\left(1 - \frac{b}{n}\right)\right) \ge \varepsilon\rho\left(\frac{y_n}{2a_2}\right) > \mu_n. \tag{62}$$

On the other hand, taking $\boldsymbol{\theta}_n = n\mathbf{v}$ gives

$$E_{H_n\rho}\left(\frac{y - \boldsymbol{\theta}_n'\mathbf{x}}{s(H_n)}\right) \le (1-\varepsilon)E_{H_0\rho}\left(\frac{y - nx_1}{a_1}\right) \le E_{H_0\rho}\left(\frac{|y| - n|x_1|}{a_1}\right) = \mu_n.$$

But this, together with (62), contradicts the definition of $\mathbf{T}(H_n)$ given in (14). □

The following lemma is needed to prove Theorem 5.2.

LEMMA A.4. (a) *If F_i, $i = 1, 2$, are distribution functions (possibly substochastic) such that $F_i(\infty) > 1 - \alpha$, $F_i(u) = 0$ for $u < 0$ and*

$$F_2(u) \ge F_1(u), \qquad \forall\, u > 0, \tag{63}$$

then

$$\int_0^\infty a(F_2(u))u^k\, dF_2(u) \le \int_0^\infty a(F_1(u))u^k\, dF_1(u). \tag{64}$$

Moreover, if the inequality (63) *holds strictly for all $u > 0$, then the inequality* (64) *also holds strictly.*

(b) *Let G_n be a sequence of distribution functions such that $G_n(u) = 0$ for $u < 0$ and continuous in $(0, \infty)$, and suppose that $\lim_{n\to\infty} G_n(u) = G(u)$, with G possibly substochastic, continuous in $(0, \infty)$ and $G(\infty) > 1 - \alpha$. Then*

$$\lim_{n\to\infty} J(G_n) = \int_0^\infty a(G(u))u^k\, dG(u). \tag{65}$$

PROOF. First notice that

$$\int_0^\infty a(F_i(u))u^k\, dF_i(u) = \int_0^{1-\alpha} a(v)\left(F_i^{-1}(v)\right)^k dv, \qquad i = 1, 2. \tag{66}$$

Observe that since $F_i(\infty) > 1 - \alpha$, $F_i^{-1}(v)$ is well defined for all $v \le 1 - \alpha$. Then (64) follows from the fact that (63) implies

$$F_2^{-1}(v) \ge F_1^{-1}(v), \tag{67}$$

for all $v > 0$. Moreover, if (63) holds strictly for all $u > 0$, then (67) holds strictly for all $v > 0$, and then (64) holds strictly, too. To prove (b), let $A_n(u) = |a(G(u)) - a(G_n(u))|u^k$, $C(u) = a(G(u))u^k$ and let u_0 be any point such that $G(u_0) > (1 - \alpha)$. Then

$$\begin{aligned}(68)\quad &\left| J(G_n) - \int_0^\infty a(G(u))\, dG(u) \right| \\ &\quad \le \int_0^\infty A_n(u)\, dG_n(u) + \left| \int_0^\infty C(u)\, dG_n(u) - \int_0^{u_0} C(u)\, dG(u) \right|.\end{aligned}$$

Given $\delta > 0$, we can find u_1 and u_2, with $u_2 < u_0$ such that $G(u_1) < 1 - \alpha$, $G(u_2) > 1 - \alpha$ and $G(u_2) - G(u_1) < \delta/(4K)$, where $K = u_0^k \max a(u)$. Then

$$\begin{aligned}(69)\quad \int_0^\infty A_n(u)\, dG_n(u) &\le \int_0^{u_1} A_n(u)\, dG_n(u) + \int_{u_1}^{u_2} A_n(u)\, dG_n(u) \\ &\quad + \int_{u_2}^\infty A_n(u)\, dG_n(u) \\ &= I_{1n} + I_{2n} + I_{3n}.\end{aligned}$$

Since $A_n(u) \to 0$ uniformly in the interval $[0, u_1]$, there exists n_1 such that $I_{n1} \le \delta/2$ for $n > n_1$. We can find n_2 such that if $n > n_2$, then $G_n(u_2) - G_n(u_1) < \delta/4K$ and, therefore, $I_{2n} < \delta/2$. Finally, we can find n_3 such that if $n > n_3$, then $G_n(u_2) > 1 - \alpha$, and by the definition of $a(u)$, $I_{3n} = 0$ for $n > n_3$. Now put $n_0 = \max(n_1, n_2, n_3)$; then $n > n_0$ implies $\int_0^\infty A_n(u)\, dG_n(u) < \delta$. This implies that

$$(70)\quad \int_0^\infty A_n(u)\, dG_n(u) \to 0.$$

Moreover, by Helly's lemma,

$$(71)\quad \int_0^{u_0} C(u)\, dG_n(u) \to \int_0^{u_0} C(u)\, dG(u),$$

and part (b) follows from (68), (70) and (71). □

PROOF OF THEOREM 5.2. By Theorem 5.1 it is enough to show that J given by (22) is ε-monotone. To prove this, take two sequences G_n and F_n satisfying the conditions of Definition 5.1. We shall prove that

$$(72)\quad \lim_{n\to\infty} J(F_n) > \lim_{n\to\infty} J(G_n).$$

If $F(\infty) > 1 - \alpha$, this follows from Lemma A.4. Now suppose that $F(\infty) \le 1 - \alpha$. In this case define $F_n^*(x) = \max(F_n(x), G_n(x) - 1/x)$. It is immediate that the distribution functions F_n^* are continuous on $(0, \infty)$ and such that $F_n^*(x) = 0$ for $x < 0$. Moreover, $\lim_{n\to\infty} F_n^*(x) = F^*(x) = \max(F(x), G(x) - 1/x)$. We also have $F^*(x) < G(x)$ for all $x > 0$ and $F^*(\infty) = G^*(\infty) > 1 - \alpha$. Therefore, by Lemma A.4,

$$(73)\quad J(F_n^*) \le J(F_n)$$

and

$$(74)\quad \lim_{n\to\infty} J(F_n^*) > \lim_{n\to\infty} J(G_n).$$

Clearly, (73) and (74) imply (72). □

PROOF OF LEMMA 5.3. The estimate $\mathbf{T}_\alpha$ can be viewed as an S-estimate with scale $S(F)$ given by (16), where ρ is the jump function given by

$$\chi(u) = \begin{cases} 0, & \text{if } |u| \le 1, \\ 1, & \text{if } |u| > 1, \end{cases}$$

and $b = 1 - \alpha$. Let $g(\mathbf{t}, s) = E_{H_0}\rho((y - \mathbf{t}'\mathbf{x})/s)$ and $g^*(t, s) = E_{H_0}\rho((y - tx_1)/s)$. Observe that if G_0 is spherical $g^*(\|\mathbf{t}\|, s) = g(\mathbf{t}, s)$ and, therefore, (3.2) of Martin, Yohai and Zamar (1989) implies that $B_{\mathbf{T}_\alpha}(\varepsilon)$ satisfies

$$h\left(B_{\mathbf{T}_\alpha}(\varepsilon), \frac{1-\alpha}{1-\varepsilon}\right) - h\left(0, \frac{1-\alpha-\varepsilon}{1-\varepsilon}\right) = 0, \tag{75}$$

where $h(t, \cdot)$ is the inverse of $g^*(t, \cdot)$. Put

$$h_{i,j}(t, \lambda) = \frac{\partial^{i+j} h(t, \lambda)}{\partial t^i \, \partial \lambda^j}, \qquad 0 \le i, j \le 2.$$

Similar notation is used with the function g.

Making a second-order Taylor expansion of (75), we get

$$h_{1,0}(0, 1-\alpha) B_{\mathbf{T}_\alpha}(\varepsilon) + h_{0,1}(0, 1-\alpha)\frac{\varepsilon}{1-\varepsilon} + h_{2,0}(0, 1-\alpha)\frac{B^2_{\mathbf{T}_\alpha}(\varepsilon)}{2}$$
$$+ h_{0,2}(0, 1-\alpha)\frac{\varepsilon^2}{2(1-\varepsilon)^2} + h_{1,1}(0, 1-\alpha) B_{\mathbf{T}_\alpha}(\varepsilon)\frac{\varepsilon}{1-\varepsilon} + o(\varepsilon^2) = 0.$$

Using that $h_{1,0}(0, 1-\alpha) = 0$ and $\lim_{\varepsilon \to 0} B_{\mathbf{T}_\alpha}(\varepsilon) = 0$ and solving for $B_{\mathbf{T}_\alpha}(\varepsilon)$ in the last equation, we get

$$\lim_{\varepsilon \to \infty} B^2_{\mathbf{T}_\alpha}(\varepsilon) = -2\frac{h_{0,1}(0, 1-\alpha)}{h_{2,0}(0, 1-\alpha)}. \tag{76}$$

A straightforward computation shows that

$$-\frac{h_{0,1}(0, 1-\alpha)}{h_{2,0}(0, 1-\alpha)} = \frac{1}{g_{2,0}} = \frac{1}{2 f_0'(F_0^{-1}((1+\alpha)/2))}. \tag{77}$$

From (76) and (77) we get

$$\lim_{\varepsilon \to \infty} B^2_{\mathbf{T}_\alpha}(\varepsilon) = \frac{1}{f_0'(F_0^{-1}((1+\alpha)/2))},$$

proving the lemma. □

REFERENCES

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

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.

HE, X., SIMPSON, D. G. and PORTNOY, S. L. (1990). Breakdown robustness of tests. *J. Amer. Statist. Assoc.* **85** 446–452.

HÖSSJER, O. (1991). Robust linear regression by means of M- and R-statistics. Ph.D. dissertation, Dept. Mathematics, Uppsala Univ.

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

JAECKEL, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. *Ann. Math. Statist.* **43** 1449–1458.

KRAFT, C. H. and VAN EEDEN, C. (1972). Linearized rank estimates and signed-rank estimates for the general linear model. *Ann. Math. Statist.* **43** 42–57.

MARONNA, R. A. and YOHAI, V. J. (1989). A new class of bias-robust regression estimates. Technical Report 177, Dept. Statistics, Univ. Washington.

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

RIEDEL, M. (1987). The asymptotic bias in the deviation of a local model. *Stability Problems for Stochastic Models. Lecture Notes in Math.* **1233** 134–144. Springer, Berlin.

RIEDEL, M. (1989a). On the bias robustness in the location model I. *Statistics* **20** 223–233.

RIEDEL, M. (1989b). On the bias robustness in the location model II. *Statistics* **20** 235–246.

RIEDEL, M. (1991). Bias robustness in parametric models generated by groups. Technical Report, Sektion Mathematik, Karl Marx Univ.

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-estimates. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

RYCHLIK, T. (1987). An asymptotically most bias-stable estimator of location. *Statistics* **18** 563–571.

RYCHLIK, T. and ZIELINSKI, R. (1987). An asymptotically most bias-robust invariant estimator of location. *Stability Problems for Stochastic Models. Lecture Notes in Math.* **1233** 156–171. Springer, Berlin.

VAN EEDEN, C. (1972). An analogue for signed ranked statistics of Jurečková's asymptotic linearity theorem for rank statistics. *Ann. Math. Statist.* **43** 791–802.

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.

ZIELINSKI, R. (1985). A most bias robust estimate of the location parameter: a general existence theorem. *Statistics* **16** 637–640.

ZIELINSKI, R. (1987). Robustness of sample mean and median under restrictions on outliers. *Zastos. Mat.* **19** 239–240.

ZIELINSKI, R. (1988). A distribution-free median-unbiased quantile estimator. *Statistics* **19** 223–227.

UNIVERSITY OF SAN ANDRÉS
VITO DUMAS 284 ESQ. ARIAS
1644 VICTORIA, PCIA. DE BUENOS AIRES
ARGENTINA

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