The Annals of Statistics
2001, Vol. 29, No. 1, 224–251

MAXIMUM BIAS CURVES FOR ROBUST REGRESSION WITH NON-ELLIPTICAL REGRESSORS

BY JOSÉ R. BERRENDERO[1] AND RUBEN H. ZAMAR[2]

Universidad Autónoma de Madrid and University of British Columbia

Maximum bias curves for some regression estimates were previously derived assuming that (i) the intercept term is known and/or (ii) the regressors have an elliptical distribution. We present a single method to obtain the maximum bias curves for a large class of regression estimates. Our results are derived under very mild conditions and, in particular, do not require the restrictive assumptions (i) and (ii) above. Using these results it is shown that the maximum bias curves heavily depend on the shape of the regressors' distribution which we call the **x**-configuration. Despite this big effect, the relative performance of different estimates remains unchanged under different **x**-configurations. We also explore the links between maxbias curves and bias bounds. Finally, we compare the robustness properties of some estimates for the intercept parameter.

1. Introduction. The concept of maximum asymptotic bias was introduced by Huber (1964) for the simple location model. Extensions to scale and location-dispersion models were later obtained by Martin and Zamar (1989, 1993). Adrover (1998) derived the maximum asymptotic bias of dispersion matrices within the class of M-estimates. Martin Yohai and Zamar (1989), Croux, Rousseeuw and Hössjer (1994, 1996), Hennig (1995) and Berrendero and Romo (1998) considered the linear regression model. See also Donoho and Liu (1988), He and Simpson (1993) and Zamar (1992).

We will focus on the regression model where the maximum bias theory has two main possible applications:

(i) the comparison of competing robust regression estimates in terms of their bias behavior and
(ii) the estimation of bias bounds for robust estimates in practical situations.

Given two estimates, the one with smaller maxbias curve is obviously more robust. In particular, its maxbias curve will have a smaller derivative at zero (called gross-error-sensitivity) and a larger asymptote (called breakdown point, BP). The comparison of maxbias functions naturally leads to the minimax bias theory also initiated by Huber's seminal 1964 work. The minimax bias theory seeks estimates which minimize the maximum asymptotic bias

Received June 1999; revised October 2000.
[1]Supported in part by DGESIC Spanish Grant PB97-0021.
[2]Supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).
AMS 2000 *subject classification*. 62F35.
Key words and phrases. Robust regression, maxbias curve, S-estimates, τ-estimates, R-estimates.

in a certain class of estimates. Huber (1964) showed that the median minimizes the maxbias curve among translation equivariant location estimates. Martin Yohai and Zamar (1989) showed that the least median of square estimate (LMS) introduced by Rousseeuw (1984) is nearly minimax in the class of regression M-estimates with general scale. Yohai and Zamar (1993) extend this result to the larger class of residual admissible estimates [see He (1990)].

A main criticism of the maxbias theory when applied to the comparison of regression estimates is that its results apply only to regression-through-the-origin models and/or to neighborhoods of a central regression model with elliptically distributed regressors. We address these criticisms by considering regression models with general regressors and including the intercept. Our results will also help to unify the existing theory because they yield a general methodology applicable to the large class of regression estimates satisfying (3) below. This class includes most regression and affine equivariant estimates for which maxbias curves were known.

Regarding the second application, we notice that, in general, estimates face two sources of uncertainty: sampling variability and bias. In the case of robust estimates, the sampling variability can be assessed by their standard errors (estimated using their asymptotic variances). On the other hand, the bias caused by outliers and other departures from symmetry can be assessed using their maximum asymptotic bias. To fix ideas, consider the location case and the median functional, $M(F)$. Suppose we have a large sample from a distribution F containing at most a fraction $\varepsilon 100\%$ of contamination. Suppose we wish to bound the absolute difference $D(F) = |M(F) - M(F_0)|$ between the median $M(F)$ of the contaminated distribution and the median $M(F_0)$ of the core (uncontaminated) distribution. Huber (1964) showed that

$$\left|\frac{M(F) - M(F_0)}{\sigma_0}\right| \le F_0^{-1}\left(\frac{1}{2(1-\varepsilon)}\right) \doteq B(\varepsilon) \tag{1}$$

and therefore $D(F)$ is bounded by $\sigma_0 B(\varepsilon)$. In practice σ_0 is seldom known and must be estimated by a robust scale functional $S(F)$, for example, the median of absolute deviations about the median. Unfortunately the quantity $S(F)B(\varepsilon)$ is not an upper bound for $D(F)$ because $S(F)$ may underestimate σ_0. For instance, if $F = 0.90N(0,1) + 0.10\delta_{0.15}$, then $|M(F) - M(F_0)| = 0.1397 > MAD(F)B(0.10) = 0.8818 \times 0.1397 = 0.1232$. A quantity, $K(\varepsilon)$ such that $S(F)K(\varepsilon)$ is a bound for $D(F)$ will be called *bias bound*. In Section 6 we study the relation between maxbias curves and bias bounds for location and regression estimates.

The bias bound is a new theoretical concept which highlights the practical potential of maxbias curves. In the case of regression estimates we face challenging problems because maxbias curves for regression estimates are derived using a normalized distance (quadratic form) between the asymptotic and the true values of the regression coefficients. The normalization is based on a certain *unknown* scatter matrix of the regressors. Moreover, the maximum bias curve depends on the joint distribution of the regressors and the available formulas relied on unrealistic assumptions (elliptical regressors and regression-through-the-origin model). Using the results of Theorem 1 we are able to derive satisfactory (but possibly not optimal) bias bounds for robust regression estimates satisfying (3) below. Similar results for other classes of robust estimates, for example, one-step Newton-Raphson estimates [Simpson, Ruppert and Carroll (1992)], projection estimates [Maronna and Yohai (1993)] and maximum depth estimates [Rousseeuw and Hubert (1999)] would be desirable.

The rest of the paper is organized as follows. In Section 2 we present some notation and technical background. In Section 3 we lay down the theoretical ground for our method. In Section 4 we illustrate our method with elliptical regressors. In Section 5 we consider several cases of non-elliptical regressors. In Section 6 we show how bias bounds can be obtained. In Section 7 we compare the robustness properties of some robust estimates for the intercept parameter. In Section 8 we give some concluding remarks. Finally, we collect all the proofs in the Appendix.

2. Notation and technical background. Consider the linear regression model with p-dimensional regressors and intercept parameter

$$y_i = \alpha_0 + \boldsymbol{\theta}_0'\mathbf{x}_i + \sigma_0 u_i, \qquad 1 \le i \le n$$

where the independent errors, u_i, have distribution F_0 and are independent of the $\mathbf{x}_i$. We assume that the regressors $\mathbf{x}_i$ are independent random vectors with common distribution G_0. The joint distribution of $(y_i, \mathbf{x}_i)$ under this model is denoted H_0. To allow for a fraction ε of contamination in the data we assume that the actual true distribution H of $(y_i, \mathbf{x}_i)$ belongs to the contamination neighborhood

$$V_\varepsilon(H_0) = \Big\{H:\ H = (1-\varepsilon)H_0 + \varepsilon\tilde{H},\ \tilde{H}\text{ arbitrary distribution}\Big\}.$$

Let $\mathbf{T}$ be an $\mathbb{R}^p$ valued regression affine equivariant functional for the estimation of $\boldsymbol{\theta}_0$, defined on a subset of distribution functions H on $\mathbb{R}^{p+1}$, which includes all H in $V_\varepsilon(H_0)$ and all the empirical distributions H_n. A natural invariant measure of the robustness of $\mathbf{T}$ is given by the maximum bias function (maxbias function) which gives the maximum bias caused by a fraction ε of contamination,

$$B_{\mathbf{T}}(\varepsilon) = \sup_{H\in V_\varepsilon(H_0)} \{[\mathbf{T}(H) - \boldsymbol{\theta}_0]'\Sigma_0[\mathbf{T}(H) - \boldsymbol{\theta}_0]\}^{1/2}/\sigma_0. \tag{2}$$

The matrix Σ_0 is an affine equivariant scatter matrix of the regressors under G_0. In view of the equivariance of $\mathbf{T}$ and the invariance of $B_{\mathbf{T}}(\varepsilon)$, we can assume without loss of generality that $\boldsymbol{\theta}_0 = \mathbf{0}$, $\sigma_0 = 1$ and $\Sigma_0 = I$, and so $B_{\mathbf{T}}(\varepsilon) = \sup_{H\in V_\varepsilon} \|\mathbf{T}(H)\|$. We will also assume without loss of generality that $\alpha_0 = 0$.

So far, derivations of $B_{\mathbf{T}}(\varepsilon)$ for regression estimates were done assuming (i) that the regression is through the origin and/or (ii) that G_0 is elliptical. Moreover, the methods and formulas used were rather specific for each particular type of estimates.

We present a method to compute the maxbias curve of regression estimates which is general in several senses. First, it does not assume that the intercept term is known. Second, it does not require the assumption of ellipticity of G_0. Finally, it applies to a broad class of robust regression estimates, namely, those defined as

$$[T_0(H), \mathbf{T}(H)] = \arg\min_{\alpha,\boldsymbol{\theta}} J(F_{H,\alpha,\boldsymbol{\theta}}), \tag{3}$$

where $J(\cdot)$ is a robust loss functional and $F_{H,\alpha,\boldsymbol{\theta}}$ is the distribution of the absolute residuals, $r_i(\alpha, \boldsymbol{\theta}) = |y_i - \alpha - \boldsymbol{\theta}'\mathbf{x}_i|$, under H. This definition includes, among others, S-estimates [Rousseeuw and Yohai (1984)], τ-estimates [Yohai and Zamar (1988)] and R-estimates [Hössjer (1994)]. Some examples of estimates which are not residual admissible are *GM*-estimates [see, e.g., Hampel et al. (1986)], *GS*-estimates [Croux, Rousseeuw and Hössjer (1994)] and P-estimates [Maronna and Yohai (1993)].

3. The main result. We will assume that the robust loss functional satisfies:

A1. (a) *If F and G are two distribution functions on $[0, \infty)$ such that $F(u) \leq G(u)$ for every $u \geq 0$, then $J(F) \geq J(G)$.*

(b) (ε-*monotonicity*). *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 sub-stochastic and continuous on $(0, \infty)$, with $G(\infty) \geq 1 - \varepsilon$ and*

$$G(u) \geq F(u) \quad \text{for every } u > 0, \tag{4}$$

then

$$\lim_{n\to\infty} J(F_n) \geq \lim_{n\to\infty} J(G_n). \tag{5}$$

Moreover, if (4) holds strictly, then (5) also holds strictly.

(c) *If F and G are two distribution functions on $[0, \infty)$, with F continuous, then*

$$J[(1-\varepsilon)F + \varepsilon\delta_\infty] \doteq \lim_{n\to\infty} J[(1-\varepsilon)F + \varepsilon U_n] \geq J[(1-\varepsilon)F + \varepsilon G],$$

where U_n stands for the uniform distribution function on $[n-(1/n), n+(1/n)]$.

A1(a) is a monotonicity condition that can be easily checked in all the examples of Sections 4 and 5. A1(b) was introduced by Yohai and Zamar (1993) who show that when J is ε-monotone, the corresponding estimate belongs to the general class of residual admissible regression estimates [He (1990)]. The definition of residual admissible estimate is rather involved but, loosely speaking, one can say that the distribution of the absolute residuals produced by a residual admissible estimate cannot be improved (in the sense of stochastic dominance) by using any other set of parameters. Finally, A1(c) is a condition that specifies the form of the contaminations that cause the greatest loss.

We will assume that the errors and regressors satisfy:

A2. *F_0 has an even and strictly unimodal density f_0 with $f_0(u) > 0$ for every $u \in \mathbb{R}$, and $P_{G_0}(\boldsymbol{\theta}'\mathbf{x} = c) < 1$, for each $\boldsymbol{\theta} \neq \mathbf{0}$, $c \in \mathbb{R}$.*

Assumption A2 allows for great generality. In particular, it does not require ellipticity nor continuity of the regressors distribution.

The following is our main result.

THEOREM 1. *Let $\mathbf{T}$ be a regression estimate defined by* (3). *Let $c = J[(1 - \varepsilon)F_{H_0,0,\mathbf{0}} + \varepsilon\delta_\infty]$ and*

$$m(t) = \inf_{\|\boldsymbol{\theta}\|=t} \inf_{\alpha\in\mathbb{R}} J\big[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0\big], \tag{6}$$

where δ_0 and δ_∞ are point mass distributions at zero and infinity, respectively. Then, under A1 *and* A2, *the maxbias curve for $\mathbf{T}$ is given by*

$$B_{\mathbf{T}}(\varepsilon) = m^{-1}(c). \tag{7}$$

Some general properties of the contamination sensitivity and BP of residual admissible estimates with unknown intercept can now be obtained. These properties are stated in the following corollary.

COROLLARY 1. *Let $\mathbf{T}$ be a regression estimate defined by* (3). *With the same notation and assumptions of Theorem* 1:

(a) *The slope of $B_{\mathbf{T}}(\varepsilon)$ at zero is infinity*: $\lim_{\varepsilon\to 0} B_{\mathbf{T}}(\varepsilon)/\varepsilon = \infty$

(b) *The BP of $\mathbf{T}$, ε^*, is given by $\varepsilon^* = \inf\{\varepsilon > 0 : m(t) < c$, for all $t > 0\}$.*

Notice that part (a) extends the results by He (1990) and Yohai and Zamar (1993, 1997) to the unknown intercept case. Moreover, part (b) shows how the BP of residual admissible estimates can be characterized through the loss function J.

In the next section we illustrate the application of Theorem 1 in the Gaussian case. Non-elliptical distributions are considered in Section 5.

4. Gaussian and elliptical regressors. The following theorem shows that under strong symmetry assumptions on the regressors distribution the function $m(t)$, defined in (6), can be substantially simplified.

THEOREM 2. *Assume* A1 *and* A2, *and that the distribution of $\boldsymbol{\theta}'\mathbf{x}$ is symmetric, unimodal and only depends on $\|\boldsymbol{\theta}\|$ for all $\boldsymbol{\theta} \neq \mathbf{0}$. Then, it holds that*

$$\inf_{\alpha\in\mathbb{R}} J\left[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0\right] = J\left[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}} + \varepsilon\delta_0\right] = m(\|\boldsymbol{\theta}\|).$$

As a consequence the infima in (6) are no longer needed and the maxbias function $B_{\mathbf{T}}(\varepsilon)$ satisfies

$$m\left(B_{\mathbf{T}}(\varepsilon)\right) = J\left[(1-\varepsilon)F_{H_0,0,\mathbf{0}} + \varepsilon\delta_\infty\right]. \tag{8}$$

The symmetry assumption of Theorem 2 is clearly satisfied by Gaussian regressors, and more generally, by elliptically symmetric and unimodal regressors. Moreover, in the Gaussian case the distribution of $y - \boldsymbol{\theta}'\mathbf{x}$ is normal with mean 0 and variance $1 + \|\boldsymbol{\theta}\|^2$ and the function $m(t)$ then becomes particularly simple.

4.1. *Maxbias curves for S-estimates.* Regression S-estimates were defined by Rousseeuw and Yohai (1984) by the property of minimizing a scale M-estimate [see Huber (1964, 1981)]. It can be easily verified that in this case the functional J in (3) is given by $J(F_{H,\alpha,\boldsymbol{\theta}}) = S(F_{H,\alpha,\boldsymbol{\theta}})$, where

$$S(F) = \inf\left\{s > 0 : E_F\chi(u/s) \le b\right\}. \tag{9}$$

We assume that the score function χ satisfies:

A3. *The function χ is even, bounded, monotone on $[0,\infty)$, continuous at 0 with $0 = \chi(0) < \chi(\infty) = 1$ and with at most a finite number of discontinuities.*

It can be easily checked that A1(a) and A1(c) hold under A3. Moreover, Yohai and Zamar (1993) showed that $S(F)$ is ε-monotone for all $\varepsilon > 0$.

The method given by Theorems 1 and 2 can now be used to strengthen the results in Martin Yohai and Zamar (1989). First, it is easy to see that the formula (3.18) in that paper derived for S-estimators of regression through the origin is also valid for the general regression model. Second, Martin Yohai and Zamar (1989) and Yohai and Zamar (1993) established a minimax bias theory for the regression model. According to this theory, the least median of squares (LMS) estimate [Rousseeuw (1984)] minimizes the maximum bias among residual admissible estimates of regression through the origin. It is now immediate from Theorem 2 that the minimaxity of LMS extends to models including the intercept.

4.2. *Maxbias curves for τ-estimates.* The loss functional J in (3) for the case of τ-estimates is given by $J(F_{H,\alpha,\boldsymbol{\theta}}) = \tau^2(F_{H,\alpha,\boldsymbol{\theta}})$ with

$$\tau^2(F) = S^2(F)E_F\chi_2\left(\frac{u}{S(F)}\right),$$

and $S(F)$ is defined as in (9) with $\chi = \chi_1$. If χ_1 and χ_2 satisfy A3 and χ_2 is differentiable with $2\chi_2(u) - \chi_2'(u)u \ge 0$, then A1(a) and A1(c) hold. Moreover, Yohai and Zamar (1993) showed that $\tau(F)$ is ε -monotone for all $\varepsilon > 0$.

As shown by Hössjer (1992), efficient S-estimates have a very low BP. Yohai and Zamar (1988) showed that τ-estimates inherit the BP of the initial S-estimate defined by χ_1 whereas its efficiency is mainly determined by χ_2.

Therefore they can have high efficiency and BP simultaneously. A natural question then is to what extent the τ-estimate inherits the good bias behavior of the initial S-estimate. We can answer this question using Theorem 3 below. Notice that (10) establishes a nice relationship between the maxbias curves of the τ-estimates and their initial S-estimates.

THEOREM 3. *Under the assumptions of Theorem* 1 *and assuming that the errors and the regressors are Gaussian*, *the maxbias curves for* τ-*estimates*, $B_\tau(\varepsilon)$, *are given by*

$$B_\tau(\varepsilon) = \left\{[1 + B_S^2(\varepsilon)]H(\varepsilon) - 1\right\}^{1/2}, \tag{10}$$

where $B_S(\varepsilon)$ *is the maxbias curve of the initial* S-*estimate based on* χ_1 *and* $H(\varepsilon)$ *is defined as*

$$H(\varepsilon) = \left[\overline{g}\left(\frac{b-\varepsilon}{1-\varepsilon}\right) + \frac{\varepsilon}{1-\varepsilon}\right] \Big/ \overline{g}\left(\frac{b}{1-\varepsilon}\right),$$

with $\overline{g}(s) = g_2[g_1^{-1}(s)]$ *and* $g_i(s) = E_\Phi \chi_i(u/s)$, *for* $i = 1, 2$. *Here*, Φ *is the standard normal distribution function*.

A similar theorem can be stated for elliptical regressors.

The maxbias curves for 95% efficient τ-estimates (solid line) and their initial S-estimates (dashed line) are given for the Huber loss functions $\chi(y) = \min\{(y/c)^2, 1\}$ in Figure 1(a) and for the Tukey loss functions $\chi(y) = \min\{3(y/c)^2 - 3(y/c)^4 + (y/c)^6, 1\}$ in Figure 1(b). The maxbias curves for the corresponding 95% efficient S-estimates are also displayed (dotted-dashed line). The breakdown points are 0.13 (Huber) and 0.12 (Tukey). The question posed above is answered then by these figures. The maxbias curves of the efficient τ-estimates are closer to those of the initial S-estimates than those

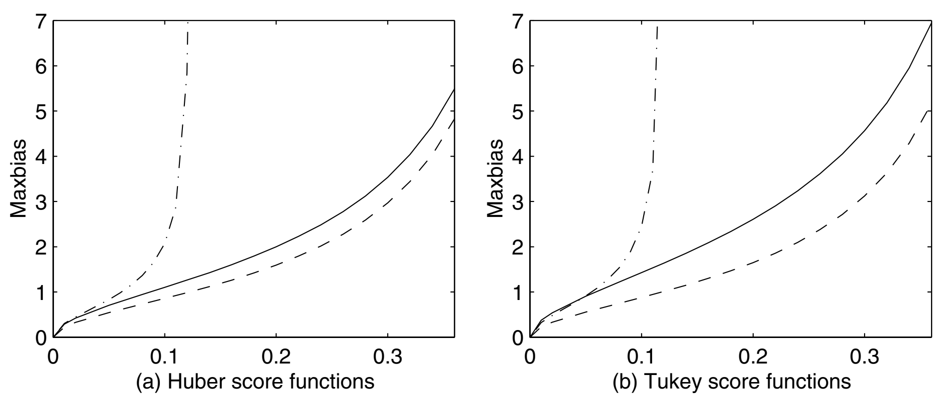

FIG. 1. *Maxbias curves for tau-estimates* (*solid line*), *initial S-estimates* (*dashed line*) *and efficient S-estimates* (*dotted-dashed line*)

of the efficient S-estimates. Clearly, the gain in efficiency is less costly in the Huber case.

REMARK. Our method can also be applied to R-estimates [see Hössjer (1994)]. In this case $J(F_{H,\alpha,\boldsymbol{\theta}}) = R(F_{H,\alpha,\boldsymbol{\theta}})$, where

$$R(F) = \int_0^\infty a[F(u)]u^k dF(u).$$

Berrendero and Zamar (1995) and Croux, Rousseeuw and Hössjer (1996) independently gave formulas for the maxbias curve of R-estimators of regression through the origin. The method given by Theorems 1 and 2 shows that these formulas extend to the general regression model.

5. Non-elliptical regressors. We wish to investigate the effect of non-elliptical regressors on the bias of robust estimates. To that effect, we consider a regression through the origin model and compute the maxbias curve for three different regression S-estimates ($\hat{\boldsymbol{\theta}}_{LMS}$, $\hat{\boldsymbol{\theta}}_H$ and $\hat{\boldsymbol{\theta}}_L$) for several bivariate $\mathbf{x}$-configurations. The corresponding score functions are

$$\chi_{LMS}(y) = \begin{cases} 0, & y \le c_1, \\ 1, & y > c_1, \end{cases}$$

for $\hat{\boldsymbol{\theta}}_{LMS}$,

$$\chi_H(y) = \min\{(y/c_2)^2, 1\},$$

for $\hat{\boldsymbol{\theta}}_H$ and

$$\chi_L(y) = \min\{|y/c_3|, 1\},$$

for $\hat{\boldsymbol{\theta}}_L$. The constants $c_1 = 0.674$, $c_2 = 1.041$ and $c_3 = 1.470$ are chosen so that these estimates have BP equal to 1/2.

The multiple integration needed to evaluate $J[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}} + \varepsilon\delta_0]$ has been carried out using Monte-Carlo methods with 100,000 replicates. For each $t > 0$, the infimum over $\|\boldsymbol{\theta}\| = t$ has been approximated by evaluating $J[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}} + \varepsilon\delta_0]$ at $\boldsymbol{\theta}(\gamma) = (t\cos(\gamma), t\sin(\gamma))'$ with γ ranging over the grid $\{0^o, 1^o, \dots, 180^o\}$. Finally, the equation $m(t) = c$ has been solved by bisection.

5.1. *Kurtosis of the marginal* $\mathbf{x}$*-distributions.* To investigate the effect of the tails of the marginal $\mathbf{x}$-distributions, we compute the maxbias curves for several Student-t bivariate distributions.

We found that, when the regressors x_1 and x_2 are independent Student-t random variables with $n_1 \le n_2$ degrees of freedom, the maxbias curve is solely determined by the minimum of the two numbers, namely n_1.

Figure 2(a), (b) and (c) correspond to $n_1 = 3$, $n_1 = 6$ and $n_1 = 12$ respectively. Figure 2 (d) displays the Gaussian limiting case ($n_1 \to \infty$). As n_1 decreases, the maxbias curves of the three estimates increase, more dramatically when

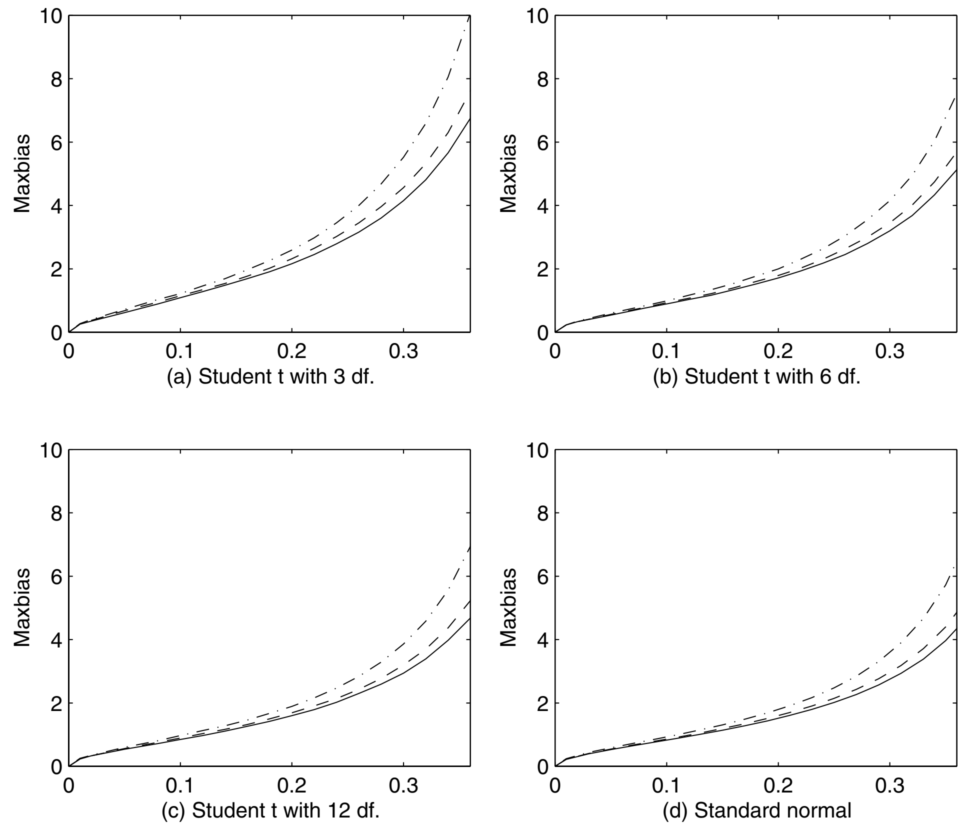

FIG. 2. *Maxbias curves of* $\hat{\boldsymbol{\theta}}_{LMS}$ *(solid line),* $\hat{\boldsymbol{\theta}}_H$ *(dashed line) and* $\hat{\boldsymbol{\theta}}_L$ *(dotted-dashed line) for independent Student-t and Gaussian regressors.*

$n_1 = 3$. The maxbias curves are already very close to the limiting Gaussian case when $n_1 = 12$.

Although the maxbias curves behave differently for each **x**-configuration, the relative maxbias behaviors are remarkably preserved across the considered cases: the maxbias curves of $\hat{\boldsymbol{\theta}}_{LMS}$ and $\hat{\theta}_H$ are quite similar and the maxbias curve of $\hat{\boldsymbol{\theta}}_L$ is somewhat larger than the other two, in all the considered cases.

5.2. *Skewness of the marginal* **x***-distributions.* To investigate the effect of unbalanced **x**-configurations (asymmetric marginal) we compute the maxbias curves for several Chi-square bivariate distributions.

In Figure 3(a), (b) and (c) the maxbias curves for two independent Chi-square regressors (x_1, x_2) with several pairs of degrees of freedom (n_1, n_2) are displayed. As the regressors become more asymmetric the maxbias curves of the three estimates increase. However, when one of the two regressors is

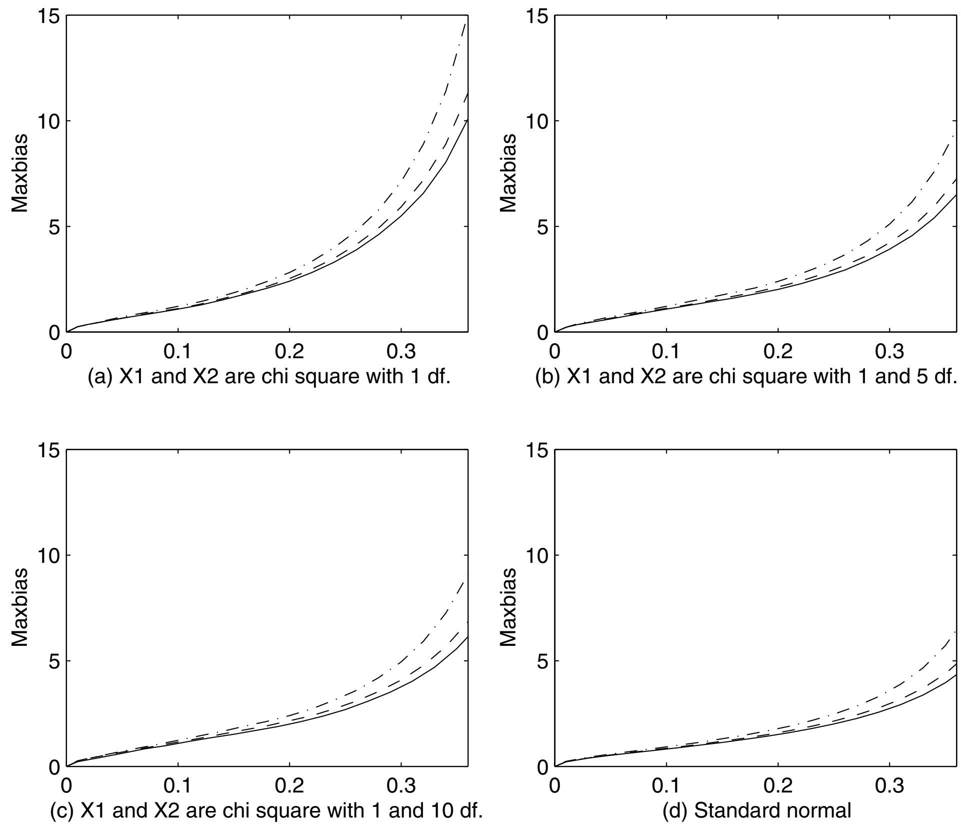

FIG. 3. *Maxbias curves of* $\hat{\boldsymbol{\theta}}_{LMS}$ (*solid line*), $\hat{\boldsymbol{\theta}}_H$ (*dashed line*) *and* $\hat{\boldsymbol{\theta}}_L$ (*dotted-dashed line*) *for independent Chi-square and Gaussian regressors.*

nearly Gaussian [see Figure 3(d)] the maxbias curves approach a Gaussian behavior.

As in the previous case, the relative maxbias behaviors are preserved across the considered cases.

5.3. *Dependent regressors.* To investigate the effect of dependent regressors we compute the maxbias curves for two closely related **x**-configurations. In the two cases x_1 is standard normal and x_2 is Chi-square with one degree of freedom. In the first case $x_2 = x_1^2$ which corresponds to a second degree polynomial regression [see Figure 4(a)]. In the second case [Figure 4(b)] x_1 and x_2 are independent. Surprisingly, the maxbias curves are larger in the independent case although the results for the two cases are strikingly close.

6. Bias bound. To introduce the main ideas we will consider first the simple location model.

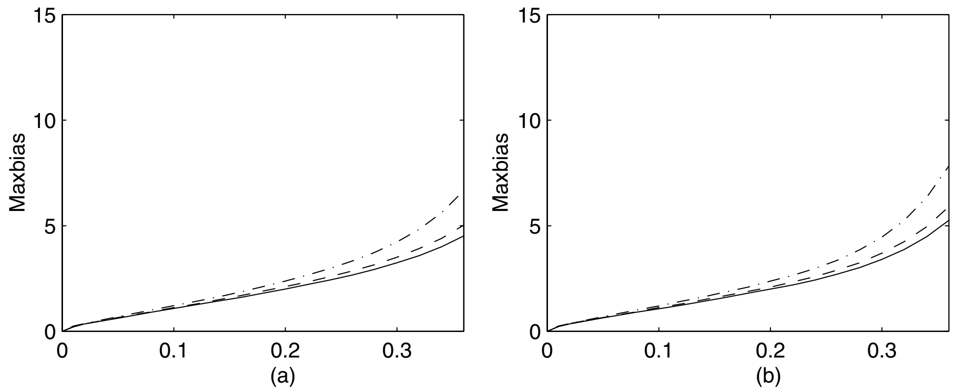

FIG. 4. *Maxbias curves of* $\hat{\boldsymbol{\theta}}_{LMS}$ (*solid line*), $\hat{\boldsymbol{\theta}}_H$ (*dashed line*) *and* $\hat{\boldsymbol{\theta}}_L$ (*dotted-dashed line*) *for* (a) x_1 *standard normal and* $x_2 = x_1^2$ *and* (b) x_1 *standard normal and* x_2 *Chi-square with one degree of freedom and independent of* x_1.

6.1. *Location model.* Let $y_i = \theta_0 + \sigma_0 u_i$, $i = 1, \dots, n$, where the errors, u_i, follow a nominal distribution F_0 satisfying A2. We assume that the actual distribution of the observations $y_1, \dots, y_n$ is $F \in V_\varepsilon(F_0)$, that is, $F(y) = (1-\varepsilon)F_0[(y-\theta_0)/\sigma_0] + \varepsilon\tilde{F}$, for some arbitrary contamination distribution $\tilde{F}$. Let $M(F)$ be a location M-functional to estimate θ_0 and let $S(F)$ be a scale M-functional to estimate σ_0. Martin and Zamar (1993) give formulas for the maxbias of the location functional

$$B_M(\varepsilon) = \sup_{F \in V_\varepsilon(F_0)} \left| \frac{M(F) - \theta_0}{\sigma_0} \right|$$

and for the implosion and explosion maxbiases of the scale functional

$$S_S^+(\varepsilon) = \sup_{F \in V_\varepsilon(F_0)} \frac{S(F)}{\sigma_0} \qquad \text{and} \qquad S_S^-(\varepsilon) = \inf_{F \in V_\varepsilon(F_0)} \frac{S(F)}{\sigma_0}$$

As argued earlier on (in the Introduction) the quantity $S(F)B_M(\varepsilon)$ is not a bound for $D(F) = |M(F) - \theta_0|$. On the other hand,

$$D(F) \le \sigma_0 B_M(\varepsilon) = S(F) B_M(\varepsilon) \frac{\sigma_0}{S(F)} \le S(F) B_M(\varepsilon)/S_S^-(\varepsilon)$$

and so $\tilde{K}(\varepsilon) = B_M(\varepsilon)/S_S^-(\varepsilon)$ is a bias bound. A refinement (of practical value when $\varepsilon > 0.2$) is provided by the following lemma.

LEMMA 1. *Let* $M(F)$ *be an equivariant location functional with maxbias function* $B_M(\varepsilon)$ *and BP* 0.5. *Let* $S(F)$ *be a scale* M*-functional with score function* χ *satisfying* A3 *and such that* $E_{F_0}\chi(X) = 1/2$. *Let* $\gamma(t)$ *be defined as the* *unique solution to* $(1-\varepsilon)E_{F_0}\chi[(X-t)/\gamma(t)] = 1/2$. *Then*

$$K(\varepsilon) = \sup_{|t| \le B_M(\varepsilon)} \frac{|t|}{\gamma(t)},$$

is a bias bound for $M(F)$.

Table 1 gives the values of $B_M(\varepsilon)$, $K(\varepsilon)$ and $\tilde{K}(\varepsilon)$ for the median when $S =$ MAD, for several values of ε and for $F_0 = \Phi$, the standard normal distribution. Notice that $K(\varepsilon)$ and $\tilde{K}(\varepsilon)$ are larger than $B_M(\varepsilon)$ because they take into account the possible underestimation of σ_0.

6.2. *Regression model.* In the regression case one would normally be interested in bounding the difference between the estimated regression coefficients and their true values. Notice that $B_{\mathbf{T}}(\varepsilon)$ gives an upper bound for the invariant quantity $\{[\mathbf{T}(H)-\mathbf{T}(H_0)]'\Sigma_0[\mathbf{T}(H)-\mathbf{T}(H_0)]\}^{1/2}/\sigma_0$ [see (2)] whereas the straight difference $\|\mathbf{T}(H)-\mathbf{T}(H_0)\|$ depends on the units used to measure the response and the regressors. Moreover, as in the location model, to obtain a usable bound we must take into account the possible bias in the estimation of σ_0 and Σ_0.

To fix ideas, we will assume that $\Sigma_0 = \mathrm{Cov}_{G_0}(\mathbf{x})$. Let $\mathbf{a}_1(G_0), \dots, \mathbf{a}_p(G_0)$ be the unit eigenvectors of Σ_0 and $\lambda_1(G_0) \le \cdots \le \lambda_p(G_0)$ be the corresponding eigenvalues. We have the following result:

LEMMA 2. *For all* $H \in V_\varepsilon(H_0)$, $\|\mathbf{T}(H)-\mathbf{T}(H_0)\| \le \sigma_0 B_{\mathbf{T}}(\varepsilon)/\sqrt{\lambda_1(G_0)}$.

In applications, the nuisance parameters σ_0 and $\sqrt{\lambda_1(G_0)}$ must be robustly estimated. The residual scale, σ_0, can be estimated by a robust residual scale estimate $S_0(F)$ applied to the distribution of the regression residuals $F_{H,T_0(H),\mathbf{T}(H)}$, that is,

$$\hat{\sigma}_0(H) \doteq S_0(F_{H,T_0(H),\mathbf{T}(H)}). \tag{11}$$

ESTIMATION OF $\lambda_1(G_0)$. Suppose that the core $\mathbf{x}$-configuration is known except for an affine transformation. More precisely, suppose that G_0 is the distribution function of $\mathbf{x} = A^{-1}\mathbf{x}^* + \boldsymbol{\mu}$, where A is an unknown invertible matrix, $\boldsymbol{\mu}$

TABLE 1
Maxbias and bias bound for the median ($S =$ MAD) *when* F_0 *is the standard normal distribution*

ε	$B_M(\varepsilon)$	$K(\varepsilon)$	$\tilde{K}(\varepsilon)$
0.05	0.066	0.070	0.070
0.10	0.140	0.159	0.160
0.15	0.223	0.271	0.278
0.20	0.319	0.417	0.440
0.25	0.431	0.614	0.675
0.30	0.566	0.889	1.043

is an unknown vector and $\mathbf{x}^*$ has some *specified* canonical distribution function G_0^* with $\mathrm{E}_{G_0^*}(\mathbf{x}^*) = \mathbf{0}$ and $\mathrm{Cov}_{G_0^*}(\mathbf{x}^*) = I$. Some examples are (i) the core distribution of $\mathbf{x}$ is multivariate normal or Student-t and (ii) the non-contaminated entries of $\mathbf{x}$ are linear combinations of some independent random variables (e.g., Student-t random variables with $n_1, n_2, \ldots, n_p$ degrees of freedom).

We will see that in order to be useful for the construction of the bias bound, the functional $\hat{\lambda}_1(G)$ must have the following two properties: (a) it must have a *known* upper bound for its over-estimation bias; and (b) it must satisfy the inequality $\hat{\lambda}_1(G_0) \le \lambda_1(G_0)$. We have found a functional which satisfies these requirements. The definition of this functional resembles the projection pursuit robust principal component methods proposed by Li and Chen (1985).

Let ρ be a continuous loss function satisfying A3 and such that

$$\max_{\|\mathbf{a}\|=1} \inf_t \mathrm{E}_{G_0^*} \rho(\mathbf{a}'\mathbf{x} - t) = 1/2. \tag{12}$$

For each distribution G, each (unitary) vector $\mathbf{a}$, and each real number t, let $S(G, \mathbf{a}, t)$ be the solution to

$$\mathrm{E}_G \rho[(\mathbf{a}'\mathbf{x} - t)/s] = 1/2 \tag{13}$$

and let

$$S(G, \mathbf{a}) = \inf_t \; S(G, \mathbf{a}, t). \tag{14}$$

Define,

$$\sqrt{\hat{\lambda}_1(G)} \doteq \min_{\|\mathbf{a}\|=1} S(G, \mathbf{a}). \tag{15}$$

The following lemma shows that $\hat{\lambda}_1(G)$ satisfies the conditions (a) and (b) above. Statement (c) of the lemma gives a condition under which the estimate is Fisher consistent. Note that this condition holds when G_0 is spherically symmetric. In general, ρ should be chosen so that $\min_{\|\mathbf{a}\|=1} \inf_t \mathrm{E}_{G_0^*} \rho(\mathbf{a}'\mathbf{x} - t)$ is close to 1/2.

LEMMA 3. *Let* $\mathbf{x} \sim G$ *and* $\mathbf{z} = A^{-1}\mathbf{x} + \mathbf{v} \sim G^*$, $\mathbf{b} = A'\mathbf{a}/\|A'\mathbf{a}\|$ *and* $t^* = (t + \mathbf{a}'A\mathbf{v})/\|A'\mathbf{a}\|$. *Suppose that* ρ *satisfies* A3 *and is continuous and strictly monotone at any* u *such that* $\rho(u) < 1$. *Then*:

(a) $\|A'\mathbf{a}\| S(G^*, \mathbf{b}, t^*) = S(G, \mathbf{a}, t)$, *for all* $\|\mathbf{a}\| = 1$ *and for all* t.
(b) $\|A'\mathbf{a}\| \, S(G^*, \mathbf{b}) = S(G, \mathbf{a})$ *for all* $\|\mathbf{a}\| = 1$.
(c) $\hat{\lambda}_1(G_0) \le \lambda_1(G_0)$, *with equality if* $\min_{\|\mathbf{a}\|=1} \inf_t \; E_{G_0^*} \rho(\mathbf{a}'\mathbf{x} - t) = 1/2$.

As in the location case, we wish to find a bias bound for $\mathbf{T}(H)$,that is a quantity $K(\varepsilon)$ such that

$$\|\mathbf{T}(H) - \mathbf{T}(H_0)\| \le \frac{\hat{\sigma}_0(H)}{\sqrt{\hat{\lambda}_1(G)}} K(\varepsilon) \doteq \hat{a}(H) K(\varepsilon) \qquad \text{for all } H \in V_\varepsilon(H_0) \tag{16}$$

where $K(\varepsilon)$ only depends on H_0 and ε, and $\hat{a}(H)$ can be consistently estimated.

We have the following result:

THEOREM 4. *Let* $\hat{\sigma}_0(H)$ *and* $\hat{\lambda}_1(G)$ *be defined by* (11) *and* (15). *Suppose that* ρ *satisfies* A3 *and is continuous and strictly monotone at any* u *such that* $\rho(u) < 1$. *Then* (16) *holds with*

$$K(\varepsilon) = \max_{\|\mathbf{a}\|=1} \frac{S^+(\mathbf{a}, \varepsilon)}{\hat{\sigma}_0^-(\varepsilon)} B_{\mathbf{T}}(\varepsilon), \tag{17}$$

where $\hat{\sigma}_0^-(\varepsilon)$ *is the implosion bias of the scale functional used in* (11) [*see Martin and Zamar* (1993)] *and*

$$S^+(\mathbf{a}, \varepsilon) = \frac{N(\mathbf{a}, \varepsilon)}{S(G_0^*, \mathbf{a})},$$

where $N(\mathbf{a}, \varepsilon) = \inf_t N(\mathbf{a}, \varepsilon, t)$ *and* $N(\mathbf{a}, \varepsilon, t)$ *satisfies*

$$(1 - \varepsilon) E_{G_0^*} \rho \left(\frac{\mathbf{a}'\mathbf{x} - t}{N(\mathbf{a}, \varepsilon, t)} \right) + \varepsilon = \frac{1}{2}.$$

If the core distribution of the regressors, G_0, *is multivariate normal, then* (16) *holds with*

$$K(\varepsilon) = \frac{S^+(\varepsilon)}{\hat{\sigma}_0^-(\varepsilon)} B_{\mathbf{T}}(\varepsilon), \tag{18}$$

where $S^+(\varepsilon)$ *is the explosion bias of the* S*-scale functional* $S(F)$ *and* $\hat{\sigma}_0^-(\varepsilon)$ *is as before*.

Table 2 gives the values of $B_{\mathbf{T}}(\varepsilon)$ and $K(\varepsilon)$ when $\mathbf{T}$ is the least median of squares, G_0 is Gaussian, and ρ is a jump function. As in Table 1, $K(\varepsilon)$ is larger than $B_{\mathbf{T}}(\varepsilon)$ because it takes into account the asymptotic bias in the estimation of $\sigma_0(H)$ and $\lambda_1(G)$. Notice that, because the estimation problems are now more involved, the differences observed in this table are larger than those observed in Table 1. However, it should also be noted that the bias bound given in Theorem 4 is probably not optimal. It is maybe an interesting open problem to devise bias bounds which are optimal so that the differences in Table 2 are minimized.

Often one is interested in linear combinations of the regression coefficients $\phi(H_0) = \mathbf{c}'\boldsymbol{\theta}_0$ (e.g., contrasts or predictions) which can be estimated by the

TABLE 2
Maxbias and bias bound for the least median of squares when G_0 is a Gaussian distribution

ε	$B_{\mathbf{T}}(\varepsilon)$	$K(\varepsilon)$
0.05	0.528	0.597
0.10	0.827	1.073
0.15	1.134	1.728
0.20	1.515	2.751
0.25	2.012	4.522
0.30	2.714	7.999

functional $\phi(H) = \mathbf{c}'\mathbf{T}(H)$. Applying the results above we have

$$\|\phi(H) - \phi(H_0)\| \le \|\mathbf{c}\| K(\varepsilon)\hat{a}(H),$$

and therefore, the bias bound for $\phi(H_n)$ is $\|\mathbf{c}\| K(\varepsilon)\hat{a}(H_n)$.

To use in practice the bias bound given by Theorem 4 we should specify (up to an affine transformation) the core distribution G_0 and the amount of contamination ε and then compute the quantity $K(\varepsilon)$. The ratio $\hat{\sigma}_0(H)/\sqrt{\hat{\lambda}_1(G)}$ depends on the sampling distributions H and G. Although these distributions are unknown, we have observations drawn from them and, therefore, we can replace H and G by the empirical distributions H_n and G_n. For the bias bound to be asymptotically valid, it is necessary that

$$\hat{a}(H_n) = \hat{\sigma}_0(H_n)\Big/\sqrt{\hat{\lambda}_1(G_n)} \to \hat{\sigma}_0(H)\Big/\sqrt{\hat{\lambda}_1(G)} = \hat{a}(H).$$

Since most robust scale estimates are consistent, the main concern is to show $\hat{\lambda}_1(G_n) \to \hat{\lambda}_1(G)$. This is accomplished in the following result:

THEOREM 5. *Let G_n be the empirical distribution corresponding to a sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ drawn from G. Assume that the loss function ρ is continuous and satisfies* A3. *Furthermore, assume that the loss function ρ and the distribution G are such that the functions $E_G\rho(\frac{\mathbf{a}'\mathbf{x}-t}{s})$ and $E_{G_n}\rho(\frac{\mathbf{a}'\mathbf{x}-t}{s})$ are strictly decreasing in $s > 0$ a.s. for each $\mathbf{a} \in A = \{\mathbf{a}\colon \|\mathbf{a}\| = 1\}$, $\mathbf{x} \in \mathbb{R}^p$ and $t \in \mathbb{R}$. Then,*

$$\hat{\lambda}_1(G_n) \to \hat{\lambda}_1(G) \qquad a.s.$$

7. Intercept maxbias. Let $(T_0(H), \mathbf{T}(H))$ be the regression functional defined by (3). Due to invariance considerations we define the intercept bias and maxbias as

$$b(H) = \left|T_0(H) - \alpha_0 + (\mathbf{T}(H) - \boldsymbol{\theta}_0)'\boldsymbol{\mu}_0\right|\Big/\sigma_0 \qquad \text{and} \qquad B_{T_0}(\varepsilon) = \sup_{H \in V_\varepsilon(H_0)} b(H),$$

respectively. Compare with Adrover, Salibian and Zamar (1999). Here, $\boldsymbol{\mu}_0$ is a multivariate location parameter for $\mathbf{x}$ under G_0. We can then assume without loss of generality that $\boldsymbol{\theta}_0 = \boldsymbol{\mu}_0 = \mathbf{0}$, $\alpha_0 = 0$ and $\sigma_0 = 1$ and therefore $B_{T_0}(\varepsilon) = \sup_{H \in V_\varepsilon(H_0)} |T_0(H)|$.

TABLE 3

Upper bounds for the maxbias of $\hat{\alpha}_M$ and lower bounds for the maxbias of several residual admissible intercept estimates: T_0^{LMS} (least median of squares), T_0^H (0.5 BP Huber S-estimate), T_0^T (0.5 BP Tukey S-estimate), $T_0^{\tau,H}$ (95% efficient and 0.5 BP τ-estimate with Huber score function) and $T_0^{\tau,T}$ (95% efficient and 0.5 BP τ-estimate with Tukey score functions)

ε	$\hat{\alpha}_M$	T_0^{LMS}	T_0^H	T_0^T	$T_0^{\tau,H}$	$T_0^{\tau,T}$
0.05	0.07	0.49	0.30	0.15	0.15	0.10
0.10	0.18	0.67	1.21	0.31	0.33	0.21
0.15	0.35	0.91	1.62	0.50	0.54	0.35
0.20	0.60	1.13	1.96	0.70	0.72	0.50
0.25	1.02	1.35	2.24	0.92	0.91	0.75

Although we cannot compute the maxbias for T_0, we are able to obtain (see Theorem 6 below) an upper bound for the maxbias of the intercept estimate $\hat{\alpha}_M$ defined as

$$\hat{\alpha}_M(H) \doteq \operatorname{med}\big(F^*_{H,\mathbf{T}(H)}\big),$$

where $F^*_{H,\mathbf{T}(H)}$ stands for the distribution function of $y - \mathbf{T}(H)'\mathbf{x}$ under H. We will provide some evidence suggesting that for small ε ($\varepsilon \leq 0.15$ in our calculations) the intercept maxbias of T_0 is larger than that of $\hat{\alpha}_M$.

THEOREM 6. *For every $u \in \mathbb{R}$, define*

$$F_1^*(u) \doteq \inf_{0\leq t\leq B_T(\varepsilon)} \inf_{\|\boldsymbol{\theta}\|=t} (1-\varepsilon)F_{H_0,\boldsymbol{\theta}^*(u)}, \tag{19}$$

$$F_2^*(u) \doteq \sup_{0\leq t\leq B_T(\varepsilon)} \sup_{\|\boldsymbol{\theta}\|=t} (1-\varepsilon)F^*_{H_0,\boldsymbol{\theta}}(u) + \varepsilon. \tag{20}$$

Let $B_1^*(\varepsilon) \doteq \inf\{u : F_1^*(u) > 1/2\}$, $B_2^*(\varepsilon) \doteq \inf\{u : F_2^*(u) > 1/2\}$ (and $B_1^*(\varepsilon) \doteq \infty$ and $B_2^*(\varepsilon) \doteq \infty$, respectively, if the conditions are never satisfied). Define

$$B^*_{\hat{\alpha}_M}(\varepsilon) \doteq \max\big\{|B_1^*(\varepsilon)|, |B_2^*(\varepsilon)|\big\}.$$

Then, $B_{\hat{\alpha}_M}(\varepsilon) \leq B^*_{\hat{\alpha}_M}(\varepsilon)$. Moreover, if the core distribution of the data, H_0, is multivariate normal, then $B^*_{\hat{\alpha}_M}(\varepsilon) = [1 + B^2_{\mathbf{T}}(\varepsilon)]^{1/2}\,\Phi^{-1}[1/(2(1-\varepsilon))]$.

The second column of Table 3 gives the values of the maxbias upper bound, $B^*_{\hat{\alpha}_M}(\varepsilon)$, for the Gaussian case. The estimate $\mathbf{T}(H)$ used to compute $\hat{\alpha}_M$ is the Huber S-estimate with BP equal to 1/2. The remaining columns of Table 3 give some maxbias lower bounds for several residual admissible estimates, for Gaussian H_0. More precisely, these are the values of $T_0[(1-\varepsilon)H_0 + \varepsilon\delta_{(y_0,0)}]$ where for each case the value of y_0 has been chosen so that $T_0[(1-\varepsilon)H_0 + \varepsilon\delta_{(y_0,0)}]$ is maximized over the grid $\{0, 0.1, \ldots, 2.5\}$. From Table 3 we conclude that, for $\varepsilon \leq 0.15$, $B_{\hat{\alpha}_M}(\varepsilon) \leq B_{T_0}(\varepsilon)$ for all the considered cases.

8. Concluding remarks. As mentioned in the Introduction, the maximum bias theory has two main possible applications: (i) the comparison of competing robust regression estimates in terms of their bias behavior and (ii) the computation of bias bounds for robust estimates in practical situations.

Regarding the first application, our results confirm the relevance of the existing global robustness theory in two ways. First, they show that, under elliptical configurations, taking the intercept parameter equal to zero corresponds to the least favorable situation and so the maxbias curves for regression-through-the-origin models formally agree with those for the general model. Therefore, the minimaxity of the LMS-estimator established by Martin Yohai and Zamar (1989) and Yohai and Zamar (1993) is proved to be valid for models including the intercept parameter. Second, our calculation of maxbias curves for several robust regression estimates under various **x**-configurations indicates that the "maxbias ranking" obtained using the Gaussian regressors also hold for other **x**-configurations. In summary, the comparison of competing robust estimates done using the Gaussian theory have implications beyond this restrictive setting.

Regarding the second application we have defined bias bounds and discussed their relation and difference with the maxbias curves. Our results show that in this regard the **x**-configuration matters since the maxbias curves changed substantially across the several cases considered.

It would be desirable to obtain bias bounds which are locally of order ε [that is, $K(\varepsilon) = o(\varepsilon)$] In order to do that one should look outside the class of residual admissible estimates. However, it is not yet clear how to do this with the needed generality (unknown nuisance parameters, non-elliptical regressors, model including the intercept, etc.). Natural candidates are *GM*-estimates. However, recent work by Adrover, Salibian and Zamar (1999) show that there are enormous technical difficulties and that the BP of *GM*-estimates may take a big drop when the intercept is in the model. For the $p = 1$ case considered in the given reference one has:

1. BP = 1/2 when the intercept is known.
2. BP = 1/3 when the intercept is unknown and the location of the regressor is known.
3. BP = 1/4 when the intercept and the location of the regressor are unknown.

Another possibility are one-step Newton-Raphson estimates [see Simpson, Ruppert and Carroll (1992)]. The technical difficulties in this case are also considerable, specially when $p > 1$. To achieve high BP and local maxbias of order $o(\varepsilon)$, the initial estimate must have high BP and local maxbias of order $o(\sqrt{\varepsilon})$ [see Simpson and Yohai (1998)]. We believe that knowledge about the maxbias behavior of the initial estimate (a residual admissible estimate in all the given proposals) will be useful, if not essential, for the derivation of bias bounds for these estimates.

Bias bounds take into account both the core distribution G_0 (which is specified by the user) and the data [through the estimation of $\hat{a}(H)$]. Perhaps it may be desirable to obtain completely empirical bias bounds, that is, solely based on data. This could be achieved by replacing G_0, in our bias bounds, by an estimate $\hat{G}_0$ computed from the sample. This approach poses interesting new theoretical and practical problems.

APPENDIX

LEMMA 4. *Let* $M(\boldsymbol{\theta}) = \inf_{\alpha\in\mathbb{R}} J[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0]$. *Under* A1(b) *and* A2, *there exists* $\alpha(\boldsymbol{\theta}) \in \mathbb{R}$ *such that*

$$M(\boldsymbol{\theta}) = J\left[(1-\varepsilon)F_{H_0,\alpha(\boldsymbol{\theta}),\boldsymbol{\theta}} + \varepsilon\delta_0\right].$$

Moreover, for every $t > 0$ *there exists* $K_t > 0$ *such that* $|\alpha(\boldsymbol{\theta})| \le K_t$ *for each* $\boldsymbol{\theta} \in \{\boldsymbol{\theta}\colon \|\boldsymbol{\theta}\| = t\}$.

PROOF. From A1(b) and A2, it follows that $J[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0]$ is a continuous function of α and $\boldsymbol{\theta}$. Since, for each $u > 0$, $\lim_{|\alpha|\to\infty} F_{H_0,\alpha,\boldsymbol{\theta}}(u) < F_{H_0,0,\boldsymbol{\theta}}(u)$, it also follows from A1(b) that

$$\lim_{|\alpha|\to\infty} J\big[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0\big] > J\big[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}} + \varepsilon\delta_0\big].$$

Therefore, for each $\boldsymbol{\theta} \in \mathbb{R}^p$, there exists $K_{\boldsymbol{\theta}} > 0$ such that the infimum must be attained in the compact set $[-K_{\boldsymbol{\theta}}, K_{\boldsymbol{\theta}}]$ and, therefore, the infimum is actually a minimum. Assume that the last assertion of this lemma is not true, it is then possible for some $t > 0$ to find a sequence $\{\boldsymbol{\theta}_n\} \subset \{\boldsymbol{\theta}\colon \|\boldsymbol{\theta}\| = t\}$ such that $\lim_{n\to\infty} |\alpha(\boldsymbol{\theta}_n)| = \infty$. Suppose w.l.o.g. that $\boldsymbol{\theta}_n \to \bar{\boldsymbol{\theta}}$. For each $\alpha \in \mathbb{R}$ and $u \ge 0$ we have that

$$\lim_{n\to\infty}\big[(1-\varepsilon)F_{H_0,\alpha(\boldsymbol{\theta}_n),\boldsymbol{\theta}_n}(u) + \varepsilon\delta_0(u)\big] = \varepsilon \le (1-\varepsilon)F_{H_0,\alpha,\bar{\boldsymbol{\theta}}}(u) + \varepsilon\delta_0(u).$$

Hence,

$$\lim_{n\to\infty} J\left[(1-\varepsilon)F_{H_0,\alpha(\boldsymbol{\theta}_n),\boldsymbol{\theta}_n} + \varepsilon\delta_0\right] \ge J\left[(1-\varepsilon)F_{H_0,\alpha,\bar{\boldsymbol{\theta}}} + \varepsilon\delta_0\right]. \tag{21}$$

On the other hand, the definition of $\alpha(\boldsymbol{\theta})$ implies that, for each $\alpha \in \mathbb{R}$,

$$\lim_{n\to\infty} J\left[(1-\varepsilon)F_{H_0,\alpha(\boldsymbol{\theta}_n),\boldsymbol{\theta}_n} + \varepsilon\delta_0\right] \le J\left[(1-\varepsilon)F_{H_0,\alpha,\bar{\boldsymbol{\theta}}} + \varepsilon\delta_0\right]. \tag{22}$$

It follows from (21) and (22), that the value of $J[(1-\varepsilon)F_{H_0,\alpha,\bar{\boldsymbol{\theta}}} + \varepsilon\delta_0]$ does not depend on α, but this is a contradiction since $\lim_{|\alpha|\to\infty} J[(1-\varepsilon)F_{H_0,\alpha,\bar{\boldsymbol{\theta}}} + \varepsilon\delta_0] < J[(1-\varepsilon)F_{H_0,0,\bar{\boldsymbol{\theta}}} + \varepsilon\delta_0]$. □

LEMMA 5. *Under* A2, *for all* $\|\boldsymbol{\theta}\| = 1$, $\lambda > 0$ *and* $u > 0$, $F_{H_0,\lambda\alpha,\lambda\boldsymbol{\theta}}(u)$ *is strictly decreasing in* λ.

PROOF. The reader can check this result by following the lines of the proof of Lemma A.1 in Yohai and Zamar (1993) rewriting the details when necessary. □

The following lemma shows that the function $m^{-1}(\cdot)$ is well defined.

LEMMA 6. *Let $m(t)$ be as in equation* (6). *Under* A1(b) *and* A2:

(a) *There exists $\boldsymbol{\theta}_t \in \mathbb{R}^p$ and $\alpha(\boldsymbol{\theta}_t) \in \mathbb{R}$ such that $\|\boldsymbol{\theta}_t\| = t$ and*

$$m(t) = J\left[(1-\varepsilon)F_{H_0,\alpha(\boldsymbol{\theta}_t),\boldsymbol{\theta}_t} + \varepsilon\delta_0\right].$$

(b) *$m(t)$ is strictly increasing.*

PROOF. With the notation of Lemma 4, notice that

$$m(t) = \inf_{\|\boldsymbol{\theta}\|=t} M(\boldsymbol{\theta}) = \inf_{\|\boldsymbol{\theta}\|=t}\ \inf_{[-K_t,K_t]} J\big[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0\big],$$

where $J[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0]$ is uniformly continuous on the compact set $\{\boldsymbol{\theta}\colon \|\boldsymbol{\theta}\| = t\} \times [-K_t, K_t]$. Therefore, $M(\boldsymbol{\theta})$ is continuous on the compact set $\boldsymbol{\theta}\colon \|\boldsymbol{\theta}\| = t\}$, as it is the infimum of an equicontinuous family of functions evaluated at $\boldsymbol{\theta}$. Part (a) follows from this fact.

Let t_1 and t_2 be such that $t_1 > t_2$. Define $\lambda = t_2/t_1 < 1$. Applying part (a), there exists $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that $m(t_1) = M(\boldsymbol{\theta}_1)$ and $m(t_2) = M(\boldsymbol{\theta}_2)$. Since, by Lemma 5, $F_{H_0,\alpha(\boldsymbol{\theta}_1),\boldsymbol{\theta}_1}(u) < F_{H_0,\lambda\alpha(\boldsymbol{\theta}_1),\lambda\boldsymbol{\theta}_1}(u)$, it follows from A1(b) and the definition of $\alpha(\boldsymbol{\theta})$ that

$$m(t_1) > J\Big[(1-\varepsilon)F_{H_0,\lambda\alpha(\boldsymbol{\theta}_1),\lambda\boldsymbol{\theta}_1} + \varepsilon\delta_0\Big] \geq J\left[(1-\varepsilon)F_{H_0,\alpha(\lambda\boldsymbol{\theta}_1),\lambda\boldsymbol{\theta}_1} + \varepsilon\delta_0\right].$$

But, by the definition of $m(t)$ and given that $\|\lambda\boldsymbol{\theta}_1\| = t_2$,

$$m(t_2) \leq M(\lambda\boldsymbol{\theta}_1) = J\left[(1-\varepsilon)F_{H_0,\alpha(\lambda\boldsymbol{\theta}_1),\lambda\boldsymbol{\theta}_1} + \varepsilon\delta_0\right].$$

The last two inequalities prove part (b). □

PROOF OF THEOREM 1. Let t^* be such that $c = m(t^*)$. First, we prove that $B_{\mathbf{T}}(\varepsilon) \leq t^*$. Let $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ be such that $\|\tilde{\boldsymbol{\theta}}\| = t > t^*$. It is enough to show that for every $H \in V_\varepsilon(H_0)$ and every $\alpha \in \mathbb{R}$, $J(F_{H,\alpha,\boldsymbol{\theta}}) > J(F_{H,0,\mathbf{0}})$.

It is clear that for each $H \in V_\varepsilon(H_0)$, $\alpha \in \mathbb{R}$ and $u > 0$,

$$F_{H,\alpha,\tilde{\boldsymbol{\theta}}}(u) \leq (1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}}(u) + \varepsilon\delta_0(u). \tag{23}$$

Inequality (23), A1(a), the definition of the function $m(t)$ and Lemma 6(b) imply that, for each $H \in V_\varepsilon(H_0)$,

$$J\left(F_{H,\alpha,\tilde{\boldsymbol{\theta}}}\right) \geq J\Big[(1-\varepsilon)F_{H_0,\alpha,\tilde{\boldsymbol{\theta}}} + \varepsilon\delta_0\Big] \geq m(t) > m(t^*). \tag{24}$$

The condition $c = m(t^*)$ and A1(c) imply

$$m(t^*) = \lim_{n\to\infty} J\left[(1-\varepsilon)F_{H_0,0,\mathbf{0}} + \varepsilon U_n\right] \geq J(F_{H,0,\mathbf{0}}). \tag{25}$$

Finally, inequalities (24) and (25) yield the first part of the result.

Now, we show the inequality $B_{\mathbf{T}}(\varepsilon) \geq t^*$. Let $t \in \mathbb{R}$ be such that $t < t^*$. The idea of the proof is to find a distribution $H \in V_\varepsilon(H_0)$ such that $\|\mathbf{T}(H)\| > t$. The inequality must hold if we can exhibit such a distribution for every $t < t^*$.

By Lemma 6(a), there exist $\boldsymbol{\theta}_t$ and α_t such that $m(t) = J[(1-\varepsilon)F_{H_0,\alpha_t,\boldsymbol{\theta}_t} + \varepsilon\delta_0]$. Define the following sequence of contaminated distributions: $\tilde{H}_n = \delta_{(y_n,\mathbf{x}_n)}$ where $\mathbf{x}_n = n\boldsymbol{\theta}_t$ and y_n is uniformly distributed on the interval $[\alpha_t + nt^2 - (1/n), \alpha_t + nt^2 + (1/n)]$. If F_n is the uniform distribution function on $[-1/n, 1/n]$, then for each $\boldsymbol{\beta} \in \mathbb{R}^p$, $u > 0$ and $\alpha \in \mathbb{R}$,

$$F_{\tilde{H}_n,\alpha,\boldsymbol{\beta}}(u) = F_n\left[u + \alpha - \alpha_t - n\left(t^2 - \boldsymbol{\beta}'\boldsymbol{\theta}_t\right)\right] - F_n\left[-u + \alpha - \alpha_t - n\left(t^2 - \boldsymbol{\beta}'\boldsymbol{\theta}_t\right)\right]. \tag{26}$$

Let $H_n = (1-\varepsilon)H_0 + \varepsilon\tilde{H}_n$. Suppose that $\sup_n \|\mathbf{T}(H_n)\| < t$ in order to find a contradiction. Under this assumption, there exists a convergent subsequence, denoted by $\{\mathbf{T}(H_n)\}$, such that

$$\lim_{n\to\infty} \mathbf{T}(H_n) \doteq \lim_{n\to\infty} \boldsymbol{\theta}_n = \tilde{\boldsymbol{\theta}} \qquad \text{where } \|\tilde{\boldsymbol{\theta}}\| = \tilde{t} < t.$$

Since $t^2 - |\boldsymbol{\theta}_t'\boldsymbol{\theta}_t| = 0$, it follows from (26) that

$$\lim_{n\to\infty} F_{\tilde{H}_n,\alpha_t,\boldsymbol{\theta}_t}(u) = 1 \quad \text{for } u > 0. \tag{27}$$

Next, we show that we can also assume w.l.o.g that the subsequence of intercepts corresponding to $\boldsymbol{\theta}_n$, denoted by $T_0(H_n) \doteq \alpha_n$, converges to a limit $\tilde{\alpha}$. Assume for a moment that $\lim_{n\to\infty} |\alpha_n| = \infty$. Then, for each $u > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} F_{H_n,\alpha_n,\boldsymbol{\theta}_n}(u) \\
&\quad = \varepsilon \lim_{n\to\infty} F_{\tilde{H}_n,\alpha_n,\boldsymbol{\theta}_n}(u) < (1-\varepsilon)F_{H_0,\alpha_t,\boldsymbol{\theta}_t}(u) + \varepsilon\delta_0(u) \\
&\quad = \lim_{n\to\infty} F_{H_n,\alpha_t,\boldsymbol{\theta}_t}(u).
\end{aligned} \tag{28}$$

For the last equality, we apply (27). The strict inequality follows from A2 since

$$F_{H_0,\alpha_t,\boldsymbol{\theta}_t}(u) = \int \mathrm{P}_{F_0}\{|y - \alpha_t - \boldsymbol{\theta}_t'\mathbf{x}| \le u|\mathbf{x}\}\, dG_0(\mathbf{x}) > 0 \quad \text{for } u > 0.$$

From (28) and assumption A1(b) we have that $J(F_{H_n,\alpha_n,\boldsymbol{\theta}_n}) > J(F_{H_n,\alpha_t,\boldsymbol{\theta}_t})$ for large enough n,and this fact contradicts the definition of $(\alpha_n, \boldsymbol{\theta}_n)$. Notice that $0 \le |\boldsymbol{\theta}_t'\tilde{\boldsymbol{\theta}}| \le \|\boldsymbol{\theta}_t\|\,\|\tilde{\boldsymbol{\theta}}\| = t\tilde{t} < t^2$. Hence, $t^2 - |\boldsymbol{\theta}'\tilde{\boldsymbol{\theta}}| > 0$. It follows from (26) that

$$\lim_{n\to\infty} F_{\tilde{H}_n,\alpha_n,\mathbf{T}_n}(u) = 0 \qquad \text{for } u > 0. \tag{29}$$

From (29) and Lemma 5, we have that for each $u > 0$,

$$\begin{aligned}
\lim_{n\to\infty} F_{H_n,\alpha_n,\mathbf{T}_n}(u) &= (1-\varepsilon)F_{H_0,\tilde{\alpha},\tilde{\boldsymbol{\theta}}}(u) \\
&\le (1-\varepsilon)F_{H_0,0,\mathbf{0}}(u) \\
&= \lim_{n\to\infty}\left[(1-\varepsilon)F_{H_0,0,\mathbf{0}}(u) + \varepsilon U_n(u)\right].
\end{aligned} \tag{30}$$

Applying A1(b), A1(c) and inequality (30),

$$\lim_{n\to\infty} J\left(F_{H_n,\alpha_n,\boldsymbol{\theta}_n}\right) \ge \lim_{n\to\infty} J\left[(1-\varepsilon)F_{H_0,0,\mathbf{0}} + \varepsilon U_n\right] = c = m(t^*). \tag{31}$$

From (27), we have that for each $u > 0$,

$$\lim_{n\to\infty} F_{H_n,\alpha_t,\boldsymbol{\theta}_t}(u) = (1-\varepsilon)F_{H_0,\alpha_t,\boldsymbol{\theta}_t}(u) + \varepsilon\delta_0(u). \tag{32}$$

From A1(b), equation (32) and Lemma 6(b),

$$\lim_{n\to\infty} J\left[F_{H_n,\alpha_t,\boldsymbol{\theta}_t}\right] = J\left[(1-\varepsilon)F_{H_0,\alpha_t,\boldsymbol{\theta}_t} + \varepsilon\delta_0\right] = m(t) < m(t^*). \tag{33}$$

Applying (31) and (33), we have that for large enough n,

$$J\big(F_{H_n,\alpha_n,\boldsymbol{\theta}_n}\big) > J\big(F_{H_n,\alpha_t,\boldsymbol{\theta}_t}\big).$$

This last inequality is a contradiction since $(\alpha_n, \boldsymbol{\theta}_n) = \arg\min_{\alpha,\boldsymbol{\theta}} J(F_{H_n,\alpha,\boldsymbol{\theta}})$. For every $t < t^*$ we have found a sequence of distributions $\{H_n\}$ in the neighborhood $V_\varepsilon(H_0)$ such that $\sup_n \|\mathbf{T}(H_n)\| \geq t$. □

PROOF OF COROLLARY 1. From Theorem 1 and Lemma 6, there exist $\boldsymbol{\theta}_t \in \mathbb{R}^p$ and α_t such that $J[(1-\varepsilon)F_{H_0,\alpha_t,\boldsymbol{\theta}_t} + \varepsilon\delta_0] = c$, and $B_{\mathbf{T}}(\varepsilon) = \|\boldsymbol{\theta}_t\| = t$. Moreover, since $\alpha_t = \alpha(\boldsymbol{\theta}_t)$,

$$\begin{aligned} J\left[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}_t} + \varepsilon\delta_0\right] &\geq J\left[(1-\varepsilon)F_{H_0,\alpha_t,\boldsymbol{\theta}_t} + \varepsilon\delta_0\right] \\ &= c = J\left[(1-\varepsilon)F_{H_0,0,\mathbf{0}} + \varepsilon\delta_\infty\right]. \end{aligned}$$

Therefore, by A1(b), there exists $u^* \geq 0$ such that $(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}_t}(u^*) + \varepsilon \leq (1-\varepsilon)F_{H_0,0,\mathbf{0}}(u^*)$. This implies that $|F_{H_0,0,\boldsymbol{\theta}_t}(u^*) - F_{H_0,0,\mathbf{0}}(u^*)| \geq \varepsilon/(1-\varepsilon)$. At this point, we obtain the conclusion of part (a) by reproducing the proof of Theorem 2.1 in Yohai and Zamar (1997).

To prove part (b), notice that, from Theorem 1, $\varepsilon < \varepsilon^*$ implies that $B_{\mathbf{T}}(\varepsilon) < \infty$. On the other hand, $\varepsilon > \varepsilon^*$ implies that $m(t) < c$ for all $t > 0$. Therefore, following exactly the same lines of the second part of the proof of Theorem 1, we deduce that $B_{\mathbf{T}}(\varepsilon) > t$ for all $t > 0$ and, hence, $B_{\mathbf{T}}(\varepsilon) = \infty$. □

PROOF OF THEOREM 2. It is easy to check that, for all $u > 0$,

$$F_{H_0,\alpha,\boldsymbol{\theta}}(u) = \mathrm{P}\big\{-u + \alpha \leq y - \boldsymbol{\theta}'\mathbf{x} \leq u + \alpha\big\}.$$

By the symmetry and unimodality assumptions on F_0 and G_0, we have that, for all $\alpha \in \mathbb{R}$,

$$F_{H_0,\alpha,\boldsymbol{\theta}}(u) \leq \mathrm{P}\big\{-u \leq y - \boldsymbol{\theta}'\mathbf{x} \leq u\big\} = F_{H_0,0,\boldsymbol{\theta}}(u),$$

and therefore, using A1(a),

$$J\big[(1-\varepsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \varepsilon\delta_0\big] \geq J\big[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}} + \varepsilon\delta_0\big],$$

for all $\alpha \in \mathbb{R}$. Finally, notice that, under the assumptions of the theorem, $F_{H_0,0,\boldsymbol{\theta}}$ only depends on $\boldsymbol{\theta}$ through the value of $\|\boldsymbol{\theta}\|$. □

PROOF OF THEOREM 3. We will apply (7) and (8) to obtain (10). First, notice that

$$c = \tau^2\big[(1-\varepsilon)F_{H_0,0,\mathbf{0}} + \varepsilon\delta_\infty\big] = \left[g_1^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}\right)\right]^2\left[(1-\varepsilon)\overline{g}\left(\frac{b-\varepsilon}{1-\varepsilon}\right)+\varepsilon\right]. \tag{34}$$

Define $m_S(\|\boldsymbol{\theta}\|) = S[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}} + \varepsilon\delta_\infty]$, then

$$\begin{aligned} m_\tau(\|\boldsymbol{\theta}\|) &= \tau^2\big[(1-\varepsilon)F_{H_0,0,\boldsymbol{\theta}} + \varepsilon\delta_\infty\big] \\ &= m_S^2(\|\boldsymbol{\theta}\|)(1-\varepsilon)E_{H_0}\chi_2\left(\frac{y-\boldsymbol{\theta}'\mathbf{x}}{m_S(\|\boldsymbol{\theta}\|)}\right) \\ &= \big(1+\|\boldsymbol{\theta}\|^2\big)\left[g_1^{-1}\left(\frac{b}{1-\varepsilon}\right)\right]^2(1-\varepsilon)\overline{g}\left(\frac{b}{1-\varepsilon}\right) \end{aligned} \tag{35}$$

and, therefore, from (7), (34) and (35), the condition $c = m_\tau(\|\boldsymbol{\theta}\|)$ implies [see also formula (3.18) in Martin, Yohai and Zamar (1989)]

$$1 + B_\tau^2(\varepsilon) = 1 + \|\boldsymbol{\theta}\|^2 = \left[\frac{g_1^{-1}\left(\frac{b-\varepsilon}{1-\varepsilon}\right)}{g_1^{-1}\left(\frac{b}{1-\varepsilon}\right)}\right]^2 H(\varepsilon) = B_S^2(\varepsilon)H(\varepsilon). \qquad \square$$

PROOF OF LEMMA 1. Since M is location equivariant, we can assume w.l.o.g. that $\theta_0 = 0$ and $\sigma_0 = 1$. Let $V_{\varepsilon,t}(F_0) = \{F \in V_\varepsilon(F_0)\colon M(F) = t\}$. Notice that $|M(F)| \le B_M(\varepsilon)$ for all $F \in V_\varepsilon(F_0)$ and therefore $V_\varepsilon(F_0) = \bigcup_{|t|\le B(\varepsilon)} V_{\varepsilon,t}(F_0)$. We have that

$$\sup_{F\in V_\varepsilon(F_0)} |M(F)/S(F)| = \sup_{|t|\le B_M(\varepsilon)}\ \sup_{F\in V_{\varepsilon,t}} |t|/S(F) \tag{36}$$

Now, for each $F \in V_{\varepsilon,t}(F_0)$, $F = (1-\varepsilon)F_0 + \varepsilon\tilde{F}$, it holds that

$$S(F) = \inf\left\{s > 0\colon (1-\varepsilon)E_{F_0}\chi\left(\frac{X-t}{S(F)}\right) + \varepsilon E_{\tilde{F}}\chi\left(\frac{X-t}{S(F)}\right) \le 1/2\right\},$$

and therefore, $S(F) \ge \gamma(t)$. This fact, together with (36), proves the result. $\square$

PROOF OF LEMMA 2. There exist coefficients β_i, $i = 1,\dots,p$, such that $\mathbf{T}(H) - \mathbf{T}(H_0) = \sum_{i=1}^p \beta_i \mathbf{a}_i(G_0)$. Moreover, for all $H \in V_\varepsilon(H_0)$,

$$\begin{aligned} B_{\mathbf{T}}^2(\varepsilon) &\ge [\mathbf{T}(H) - \mathbf{T}(H_0)]'\Sigma_0[\mathbf{T}(H) - \mathbf{T}(H_0)]/\sigma_0^2 \\ &= \sum_{i=1}^p \beta_i^2\lambda_i(G_0)/\sigma_0^2 \ge \big[\lambda_1(G_0)/\sigma_0^2\big]\sum_{i=1}^p \beta_i^2 \\ &= \big[\lambda_1(G_0)/\sigma_0^2\big]\|\mathbf{T}(H) - \mathbf{T}(H_0)\|^2. \end{aligned}$$

Therefore, $\sup_{H\in V_\varepsilon(H_0)} \|\mathbf{T}(H) - \mathbf{T}(H_0)\| \le [\sigma_0/\sqrt{\lambda_1(G_0)}]B_{\mathbf{T}}(\varepsilon)$. $\square$

PROOF OF LEMMA 3. Part (a) follows directly from the fact that

$$\mathrm{E}_G\rho[(\mathbf{a}'\mathbf{x}-t)/s] = \mathrm{E}_{G^*}\rho\big[(\mathbf{b}'\mathbf{z}-t^*)/(s/\|A'\mathbf{a}\|)\big].$$

Part (b) follows because $S(G,\mathbf{a}) = \inf_t S(G,\mathbf{a},t) = \|A'\mathbf{a}\| \inf_{t^*} S(G^*,\mathbf{b},t^*) = \|A'\mathbf{a}\|\, S(G^*,\mathbf{b})$. To prove (c) first notice that

$$1/2 \le \inf_t \ \mathrm{E}_{G_0^*}\rho\big[(\mathbf{a}'\mathbf{x}-t)\big/S(G_0^*,\mathbf{a})\big] \tag{37}$$

because otherwise there exists t_1 such that $\mathrm{E}_{G_0^*}\rho[(\mathbf{a}'\mathbf{x}-t_1)/S(G_0^*,\mathbf{a})] < 1/2$ and $S(G_0^*,\mathbf{a},t_1) < S(G_0^*,\mathbf{a})$. Also, since $\lim_{|t|\to\infty} \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t) = 1$, there exists $0 < K < \infty$ and $-K \le t_0 \le K$ such that

$$\inf_t \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t) = \min_{-K\le t\le K} \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t) = \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t_0). \tag{38}$$

Suppose now that $S(G_0^*,\mathbf{a}) > 1$. Then,

$$\inf_t \mathrm{E}_{G_0^*}\rho\big[(\mathbf{a}'\mathbf{x}-t)/S(G_0^*,\mathbf{a})\big] < \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t_0). \tag{39}$$

If (39) doesn't hold we arrive at a contradiction:

$$\inf_t \mathrm{E}_{G_0^*}\rho\big[(\mathbf{a}'\mathbf{x}-t)/S(G_0^*,\mathbf{a})\big] \ge \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t_0) > \mathrm{E}_{G_0^*}\rho\big[(\mathbf{a}'\mathbf{x}-t_0)/S(G_0^*,\mathbf{a})\big],$$

where the strict inequality holds because $\mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t_0) \le 1/2 < 1$ together with the monotonicity assumption on ρ imply that the function (in s) $\mathrm{E}_{G_0^*}\rho[(\mathbf{a}'\mathbf{x}-t_0)/s]$ is strictly decreasing at $s=1$. By (12), (37), (38) and (39),

$$\begin{aligned} 1/2 &\le \inf_t \ \mathrm{E}_{G_0^*}\rho\big[(\mathbf{a}'\mathbf{x}-t)\big/S(G_0^*,\mathbf{a})\big] < \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t_0) \\ &= \inf_t \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t) \le \max_{\|\mathbf{a}\|=1} \inf_t \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t) = 1/2. \end{aligned} \tag{40}$$

Therefore, $S(G_0^*,\mathbf{a}) \le 1$. Finally, if $\mathrm{Cov}_{G_0}(\mathbf{x}) = AA'$, by (b), $\sqrt{\hat\lambda_1(G_0)} = \min_{\|\mathbf{a}\|=1} S(G_0,\mathbf{a}) = \min_{\|\mathbf{a}\|=1} S(G_0^*,\mathbf{b})\|A'\mathbf{a}\| \le \min_{\|\mathbf{a}\|=1} \|A'\mathbf{a}\| = \sqrt{\mathbf{a}_1'AA'\mathbf{a}_1} = \sqrt{\lambda_1(G_0)}$. Finally, if $\min_{\|\mathbf{a}\|=1} \inf_t \mathrm{E}_{G_0^*}\rho(\mathbf{a}'\mathbf{x}-t) = 1/2$, then $S(G_0^*,\mathbf{b}) = 1$ for all $\|\mathbf{b}\| = 1$ and the equality holds. □

PROOF OF THEOREM 4. Let $V_\varepsilon(G_0^*)$ be the ε-contamination neighborhood for G_0^* and let $V_\varepsilon(G_0)$ be the corresponding ε-contamination neighborhood for G_0. First of all notice that, given $\mathbf{x} \sim G$ and $\mathbf{z} = A^{-1}\mathbf{x}+\mathbf{v} \sim G^*$, by Lemma 3(b),

$$S(G,\mathbf{a}) = \|A'\mathbf{a}\|\, S(G^*,\mathbf{b}), \tag{41}$$

and so

$$\sup_{G\in V_\varepsilon(G_0)} \frac{S(G,\mathbf{a})}{S(G_0,\mathbf{a})} = \sup_{G^*\in V_\varepsilon(G_0^*)} \frac{S(G^*,\mathbf{b})\|A'\mathbf{a}\|}{S(G_0^*,\mathbf{b})\|A'\mathbf{a}\|} = \sup_{G^*\in V_\varepsilon(G_0^*)} \frac{S(G^*,\mathbf{b})}{S(G_0^*,\mathbf{b})}. \tag{42}$$

Moreover, by Lemma 3(c),

$$\frac{\sigma_0}{\sqrt{\lambda_1(G_0)}} = \frac{\sigma_0}{\sqrt{\hat\lambda_1(G_0)}} \frac{\sqrt{\hat\lambda_1(G_0)}}{\sqrt{\lambda_1(G_0)}} \le \frac{\sigma_0}{\sqrt{\hat\lambda_1(G_0)}} \tag{43}$$

and

$$\frac{\sigma_0}{\sqrt{\hat{\lambda}_1(G_0)}} = \frac{\hat{\sigma}_0(H)}{\sqrt{\hat{\lambda}_1(G)}} \frac{\sigma_0}{\hat{\sigma}_0(H)} \frac{\sqrt{\hat{\lambda}_1(G)}}{\sqrt{\hat{\lambda}_1(G_0)}} \tag{44}$$
$$\leq \frac{\hat{\sigma}_0(H)}{\sqrt{\hat{\lambda}_1(G)}} \frac{1}{\hat{\sigma}_0^-(\varepsilon)} \frac{\sqrt{\hat{\lambda}_1(G)}}{\sqrt{\hat{\lambda}_1(G_0)}}.$$

By (43) and (44)

$$\frac{\sigma_0}{\sqrt{\lambda_1(G_0)}} \leq \frac{\hat{\sigma}_0(H)}{\sqrt{\hat{\lambda}_1(G)}} \frac{1}{\sigma_0^-(\varepsilon)} \sup_{G \in V_\varepsilon(G_0)} \frac{\min_{\|\mathbf{a}\|=1} S(G,\mathbf{a})}{\min_{\|\mathbf{a}\|=1} S(G_0,\mathbf{a})}.$$

In addition, by (42),

$$\begin{aligned}
&\sup_{G \in V_\varepsilon(G_0)} \frac{\min_{\|\mathbf{a}\|=1} S(G,\mathbf{a})}{\min_{\|\mathbf{a}\|=1} S(G_0,\mathbf{a})} \\
&\quad = \sup_{G \in V_\varepsilon(G_0)} \frac{\min_{\|\mathbf{a}\|=1}[S(G,\mathbf{a})/S(G_0,\mathbf{a})]S(G_0,\mathbf{a})}{\min_{\|\mathbf{a}\|=1} S(G_0,\mathbf{a})} \\
&\quad \leq \min_{\|\mathbf{a}\|=1} \frac{\sup_{G \in V_\varepsilon(G_0)}[S(G,\mathbf{a})/S(G_0,\mathbf{a})]S(G_0,\mathbf{a})}{\min_{\|\mathbf{a}\|=1} S(G_0,\mathbf{a})} \\
&\quad = \min_{\|\mathbf{a}\|=1} \frac{\sup_{G^* \in V_\varepsilon(G_0^*)}[S(G^*,\mathbf{a})/S(G_0^*,\mathbf{a})]S(G_0,\mathbf{a})}{\min_{\|\mathbf{a}\|=1} S(G_0,\mathbf{a})} \\
&\quad \leq \min_{\|\mathbf{a}\|=1} \frac{S^+(\mathbf{a},\varepsilon)S(G_0,\mathbf{a})}{\min_{\|\mathbf{a}\|=1} S(G_0,\mathbf{a})} \leq \max_{\|\mathbf{a}\|=1} S^+(\mathbf{a},\varepsilon).
\end{aligned} \tag{45}$$

To obtain the next to the last inequality notice that, for each $\|\mathbf{a}\| = 1$ and t, $S(G^*, \mathbf{a}, t) \leq N(\mathbf{a}, \varepsilon, t)$.

In the Gaussian case, we have that $S(G_0, \mathbf{a}) = \sigma_{\mathbf{a}} \doteq \sqrt{\mathbf{a}'\Sigma_0\mathbf{a}}$. Moreover, $S(G, \mathbf{a}) \leq \bar{s}$, where $\bar{s}$ satisfies

$$(1-\varepsilon)E_{G_0}\rho\left(\frac{\mathbf{a}'\mathbf{x}}{\bar{s}}\right) + \varepsilon = 1/2.$$

From Theorem 4 in Martin and Zamar (1993) we have that $\bar{s} = \sigma_a S^+(\varepsilon)$, and therefore,

$$\sup_{G \in V_\varepsilon(G_0)} \sqrt{\hat{\lambda}_1(G)} = \sup_{G \in V_\varepsilon(G_0)} \min_{\|\mathbf{a}\|=1} S(G,\mathbf{a}) \leq \min_{\|\mathbf{a}\|=1} \sup_{G \in V_\varepsilon(G_0)} S(G,\mathbf{a})$$
$$\leq \min_{\|\mathbf{a}\|=1} \sigma_{\mathbf{a}} S^+(\varepsilon) = \sqrt{\lambda_1(G_0)}S^+(\varepsilon).$$

The final assertion of the theorem follows from this inequality. □

The following lemma is needed to prove Theorem 5:

LEMMA 7. *Under the assumptions of Theorem* 5, *for each* $c > 0$:

(a) $S(G, \mathbf{a}, t)$ *and* $S(G_n, \mathbf{a}, t)$ *are uniformly continuous a.s. on* $(\mathbf{a}, t) \in A \times [-c, c]$.

(b) $\rho\,(\frac{\mathbf{a}'\mathbf{x}-t}{S(G,\mathbf{a},t)})$ *is uniformly continuous on* $(\mathbf{a}, t) \in A \times [-c, c]$.

(c) *There exists* $C > 0$ *such that* $S(G_n, \mathbf{a}) = \min_{|t|\le C} S(G_n, \mathbf{a}, t)$ *and* $S(G, \mathbf{a}) = \min_{|t|\le C} S(G, \mathbf{a}, t)$.

(d) $S(G_n, \mathbf{a})$ *and* $S(G, \mathbf{a})$ *are uniformly continuous on* $\mathbf{a} \in A$.

PROOF. From Lemma 3(a) in Martin and Zamar (1993) we know that for $c > 0$ and $0 < s_0 < s_\infty < \infty$, the function $\rho[(x-t)/s]$ is uniformly continuous on $(x, t, s) \in \mathbb{R}\times[-c, c]\times[s_0, s_\infty]$. Let $(\mathbf{a}_1, t_1, s_1), (\mathbf{a}_2, t_2, s_2) \in A\times[-c, c]\times[s_0, s_\infty]$. Since

$$\left|E_G\rho\left(\frac{\mathbf{a}_1'\mathbf{x}-t_1}{s_1}\right) - E_G\rho\left(\frac{\mathbf{a}_2'\mathbf{x}-t_2}{s_2}\right)\right| \le E_G\left|\rho\left(\frac{\mathbf{a}_1'\mathbf{x}-t_1}{s_1}\right) - \rho\left(\frac{\mathbf{a}_2'\mathbf{x}-t_2}{s_2}\right)\right|,$$

then $E_G\rho\,(\frac{\mathbf{a}'\mathbf{x}-t}{s})$ is uniformly continuous on $(\mathbf{a}, t, s) \in A \times [-c, c] \times [s_0, s_\infty]$. Let $\delta > 0$ and let $(\mathbf{a}_1, t_1), (\mathbf{a}_2, t_2) \in A \times [-c, c]$. Define $s_1 = S(G, \mathbf{a}_1, t_1)$ and $s_2 = S(G, \mathbf{a}_2, t_2)$. Since $E_G\rho[(\mathbf{a}'\mathbf{x} - t)/s]$ is strictly decreasing in $s > 0$, $\delta_0 \doteq E_G\rho\,(\frac{\mathbf{a}_1'\mathbf{x}-t_1}{s_1-\delta}) - \frac{1}{2} > 0$. There exists $\delta_1 > 0$ such that if $\|\mathbf{a}_1 - \mathbf{a}_2\| < \delta_1$ and $|t_1 - t_2| < \delta_1$, then

$$\left|E_G\rho\left(\frac{\mathbf{a}_1'\mathbf{x}-t_1}{s_1-\delta}\right) - E_G\rho\left(\frac{\mathbf{a}_2'\mathbf{x}-t_2}{s_1-\delta}\right)\right| < \delta_0/2,$$

which implies that $E_G\rho\,(\frac{\mathbf{a}_2'\mathbf{x}-t_2}{s_1-\delta}) > 1/2$ and, therefore, $s_2 > s_1-\delta$. Analogously, it can be shown that there exists $\delta_2 > 0$ such that if $\|\mathbf{a}_1 - \mathbf{a}_2\| < \delta_2$ and $|t_1-t_2| < \delta_2$, then $s_1 > s_2-\delta$. Considering $\min\{\delta_1, \delta_2\}$, we obtain $|s_1-s_2| < \delta$ and part (a) follows. (The result for G_n can be proved analogously.)

Part (b) follows straightforwardly from Lemma 3(a) in Martin and Zamar (1993) and part (a).

To prove part (c), notice that $S(G, \mathbf{a}, 0) < \infty$ and $\lim_{|t|\to\infty} S(G, \mathbf{a}, t) = \infty$. Therefore, for all $\mathbf{a} \in A$ there exists $c(\mathbf{a}) > 0$ such that $S(G, \mathbf{a}) = \min_{|t|\le c(\mathbf{a})} S(G, \mathbf{a}, t) \doteq S(G, \mathbf{a}, t(\mathbf{a}))$. Moreover, there exists $C > 0$ such that $t(\mathbf{a}) \le C$ for all $\mathbf{a} \in A$ because, if this was not the case, there would exist a sequence $\{\mathbf{a}_k\} \subset A$ (we assume $\mathbf{a}_k \to \mathbf{a}^*$ without loss of generality) such that $|t(\mathbf{a}_k)| \to \infty$. Then,

$$\lim_{k\to\infty} S(G, \mathbf{a}_k, t(\mathbf{a}_k)) = \infty. \tag{46}$$

On the other hand, from the definition of $t(\mathbf{a})$ and part (a) we have that

$$\lim_{k\to\infty} S(G, \mathbf{a}_k, t(\mathbf{a}_k)) \le \lim_{k\to\infty} S(G, \mathbf{a}_k, 0) = S(G, \mathbf{a}^*, 0) < \infty. \tag{47}$$

Equation (47) contradicts equation (46). Therefore, $S(G, \mathbf{a}) = \min_{|t|\le C} S(G, \mathbf{a}, t)$. Analogously, we can prove that $S(G_n, \mathbf{a}) = \min_{|t|\le C} S(G_n, \mathbf{a}, t)$. (We can assume the same value for C without loss of generality.)

Finally, both $S(G, \mathbf{a})$ and $S(G_n, \mathbf{a})$ are continuous since they are defined as the minimum on the compact set $[-C, C]$ of a family of uniformly continuous functions. Since A is compact, both $S(G, \mathbf{a})$ and $S(G_n, \mathbf{a})$ are uniformly continuous on A and therefore (d) follows. □

PROOF OF THEOREM 5. Notice that it is enough to prove

$$\max_{\mathbf{a}\in A} |S(G_n, \mathbf{a}) - S(G, \mathbf{a})| \to 0 \quad \text{a.s.} \tag{48}$$

Since A is a compact set, this fact, together with Lemma 7 (d) implies the conclusion of the theorem. Let $\delta > 0$, $\mathbf{a} \in A$ and $t \in [-C, C]$. Since $E_G\rho\,(\frac{\mathbf{a}'\mathbf{x}-t}{s})$ is strictly decreasing, there exists δ_0 such that $E_G\rho\,(\frac{\mathbf{a}'\mathbf{x}-t}{S(G,\mathbf{a},t)-\delta}) - 1/2 \ge \delta_0 > 0$. For any $\gamma \ge 0$, define

$$B_n(\mathbf{a}, t, \gamma) = \left\{\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{\mathbf{a}'\mathbf{x}_i - t}{S(G, \mathbf{a}, t) - \delta}\right) - \frac{1}{2} \ge \gamma\right\}.$$

By Lemma 7 (b), there exist $t_1, \dots, t_m$ with $t_j \in [-C, C]$ such that

$$\bigcap_{j=1}^{m} B_n\left(\mathbf{a}, t_j, \frac{\delta_0}{2}\right) \subset \bigcap_{|t|\le C} B_n(\mathbf{a}, t, 0) \quad \text{a.s.}$$

Now, from Bernstein's inequality, there exists $r > 0$ such that

$$\begin{aligned}
&\mathrm{P}_G\{B_n^c(\mathbf{a}, t_j, \delta_0/2)\}\\
&\quad\le \mathrm{P}_G\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{\mathbf{a}'\mathbf{x}_i - t_j}{S(G, \mathbf{a}, t_j) - \delta}\right) - E_G\rho\left(\frac{\mathbf{a}'\mathbf{x}_i - t_j}{S(G, \mathbf{a}, t_j) - \delta}\right)\right| > \frac{\delta_0}{2}\right\}\\
&\quad\le 2\exp\{-rn\delta_0^2\}.
\end{aligned}$$

For all $\mathbf{a} \in A$, we have

$$\begin{aligned}
&\mathrm{P}_G\left\{\min_{|t|\le C}[S(G_n, \mathbf{a}, t) - S(G, \mathbf{a}, t)] > -\delta\right\}\\
&\qquad\ge \mathrm{P}_G\left\{\bigcap_{|t|\le C} B_n(\mathbf{a}, t, 0)\right\} \ge \mathrm{P}_G\left\{\bigcap_{j=1}^{m} B_n\left(\mathbf{a}, t_j, \frac{\delta_0}{2}\right)\right\}\\
&\qquad= 1 - \mathrm{P}_G\left\{\bigcup_{j=1}^{m} B_n^c\left(\mathbf{a}, t_j, \frac{\delta_0}{2}\right)\right\} \ge 1 - 2m\exp\{-rn\delta_0^2\}.
\end{aligned}$$

Analogously, it can be proved that, for all $\mathbf{a} \in A$,

$$\mathrm{P}_G\left\{\max_{|t|\le C}[S(G_n, \mathbf{a}, t) - S(G, \mathbf{a}, t)] < \delta\right\} \ge 1 - 2m\exp\{-rn\delta_0^2\}.$$

As a consequence, for all $\mathbf{a} \in A$,

$$\mathrm{P}_G\left\{\max_{|t|\le C} |S(G_n, \mathbf{a}, t) - S(G, \mathbf{a}, t)| > \delta\right\} \le 2m \exp\{-rn\delta_0^2\}.$$

Therefore,

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathrm{P}_G\left\{\max_{\mathbf{a}\in A} |S(G_n, \mathbf{a}) - S(G, \mathbf{a})| > \delta\right\} \\
&\quad \doteq \sum_{n=1}^{\infty} \mathrm{P}_G\left\{|S(G_n, \mathbf{a}_n) - S(G, \mathbf{a}_n)| > \delta\right\} \\
&\quad \le \sum_{n=1}^{\infty} \mathrm{P}_G\left\{\max_{|t|\le C} |S(G_n, \mathbf{a}_n, t) - S(G, \mathbf{a}_n, t)| > \delta/2\right\} \\
&\quad < \infty.
\end{aligned}$$

This inequality implies (48). □

PROOF OF THEOREM 6. Let $H \in V_\varepsilon(H_0)$ and notice that $\|\mathbf{T}(H)\| \le B_{\mathbf{T}}(\varepsilon)$. Since for some arbitrary distribution $\tilde{H}$, $F^*_{H,\mathbf{T}(H)} = (1-\varepsilon)F^*_{H_0,\mathbf{T}}(H) + \varepsilon F^*_{\tilde{H},\mathbf{T}}(H)$, we have $F_1^*(u) \le F^*_{H,\mathbf{T}(H)}(u) \le F_2^*(u)$ for all $u \in \mathbb{R}$. Therefore, $B_2^*(\varepsilon) \le \hat{\alpha}_M(H) \le B_1^*(\varepsilon)$, and the first part of the result follows. When G_0 is multivariate normal the inner infimum and supremum in (19) and (20) are not needed because $F^*_{H_0,\boldsymbol{\theta}}(u)$ only depends on the length of $\boldsymbol{\theta}$. Indeed, for all $\|\boldsymbol{\theta}\| = t$, we have

$$F^*_{H_0,\boldsymbol{\theta}}(u) = P_{H_0}\{y - \boldsymbol{\theta}'\mathbf{x} \le u\} = \Phi\left(\frac{u}{\sqrt{1+t^2}}\right).$$

As a consequence, $F_1^*(u) = (1-\varepsilon)\Phi[u/(1+B^2_{\mathbf{T}}(\varepsilon))]$ and $F_2^*(u) = (1-\varepsilon)\Phi(u) + \varepsilon$. Hence, $B_1^*(\varepsilon) = [1+B^2_{\mathbf{T}}(\varepsilon)]^{1/2}\,\Phi^{-1}[1/(2(1-\varepsilon))]$ and $B_2^*(\varepsilon) = -\Phi^{-1}[1/(2(1-\varepsilon))]$. Therefore, $B^*_{\hat{\alpha}_M}(\varepsilon) = [1 + B^2_{\mathbf{T}}(\varepsilon)]^{1/2}\ \Phi^{-1}[1/(2(1-\varepsilon))]$. □

REFERENCES

ADROVER, J. G. (1998). Minimax bias robust estimation of the dispersion matrix of a multivariate distribution. *Ann. Statist.* **26** 2301–2320.

ADROVER, J. G., SALIBIAN, M. and ZAMAR, R. H. (1999). Robust inference for the simple linear regression model. Unpublished manuscript. Available at ftp://hajek.stat.ubc.ca/pub/ruben/rislrm.

BERRENDERO, J. R. and ZAMAR, R. H. (1995). On the maxbias curve of residual admissible robust regression estimates. Working Paper 95–62, Univ. Carlos III de Madrid.

BERRENDERO, J. R. and ROMO, J. (1998) Stability under contamination of robust regression estimates based on differences of residuals. *J. Statist. Plann. Inference* **70** 149–165.

CROUX, C., ROUSSEEUW, P. J. and HÖSSJER, O. (1994). Generalized S-estimators. *J. Amer. Statist. Assoc.* **89** 1271–1281.

CROUX, C., ROUSSEEUW, P. J. and VAN BAEL, A. (1996). Positive-breakdown regression by minimizing nested scale estimators. *J. Statist. Plann. Inference* **53** 197–235.

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

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.

HE, X. (1990). A local breakdown property of robust tests in linear regression. *J. Multivariate Anal.* **38** 293–305.

HE, X. and SIMPSON, D. G. (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *Ann. Statist.* **21** 314–337.

HENNIG, C. (1995). Efficient high breakdown point estimators in robust regression: which function to choose? *Statist. Decisions* **13** 221–241.

HÖSSJER, O. (1992). On the optimality of S-estimators. *Statist. Probab. Lett.* **14** 413–419.

HÖSSJER, O. (1994). Rank-based estimates in the linear model with high breakdown point. *J. Amer. Statist. Assoc.* **89** 149–158.

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.

LI, G. and CHEN, Z. (1985). Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo. *J. Amer. Statist. Assoc.* **80** 759–766.

MARONNA, R. A. and YOHAI, V. J. (1993). Bias-robust estimates of regression based on projections. *Ann. Statist.* **21** 965–990.

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

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

MARTIN, R. D. and ZAMAR, R. (1993) Bias robust estimation of scale. *Ann. Statist.* **21** 991–1017.

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

ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth. *J. Amer. Statist. Assoc.* **94** 388–402.

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,

SIMPSON, D. G., RUPPERT, D. and CARROLL, R. J. (1992). On one-step *GM*–estimates and stability of inferences in linear regression model. *J. Amer. Statist. Assoc.* **87** 439–450.

SIMPSON, D. G. and YOHAI, V. J. (1998). Functional stability of one-step *GM*-estimators in linear regression. *Ann. Statist.* **26** 1147–1169.

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.

YOHAI, V. J. and ZAMAR, R. H. (1993). A minimax-bias property of the least α-quantile estimates. *Ann. Statist.* **21** 1824–1842.

YOHAI, V. J. and ZAMAR, R. H. (1997). Optimal locally robust M-estimates of regression. *J. Statist. Plann. Inference* **64** 309–323.

ZAMAR, R. H. (1992). Bias robust estimation in orthogonal regression. *Ann. Statist.* **20** 1875–1888.

DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD AUTÓNOMA DE MADRID
CAMPUS DE CANTOBLANCO
28049 MADRID
SPAIN
E-MAIL: joser.berrendero@uam.es

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
6356 AGRICULTURAL ROAD
ROOM 333
VANCOUVER, B.C.
CANADA V6T 1Z2
E-MAIL: ruben@stat.ubc.ca