Stability under contamination of robust regression estimators based on differences of residuals

José R. Berrendero* and Juan Romo†

Departamento de Estadística y Econometría

Universidad Carlos III de Madrid

Abstract

A reasonable approach to robust regression estimation is minimizing a robust scale estimator of the pairwise differences of residuals. We introduce a large class of estimators based on this strategy extending ideas of Yohai and Zamar (1993) and Croux, Rousseeuw and Hössjer (1994). The asymptotic robustness properties of the estimators in this class are addressed using the maxbias curve. We provide a general principle to compute this curve and present explicit formulae for several particular cases including generalized versions of S-, R- and τ-estimators. Finally, the most stable estimator in the class, that is, the estimator with the minimum maxbias curve, is shown to be the set of coefficients that minimizes an appropriate quantile of the distribution of the absolute pairwise differences of residuals.

AMS 1991 subject classification: 62F35.

KEY WORDS AND PHRASES: Robust regression, maxbias curve, S-estimators, GS-estimators, minimax bias.

*Research partially supported by DGICYT PB93-0232

†Research partially supported by DGICYT PB93-0232

1 Introduction

Let $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$, $\mathbf{x}_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$ be independent observations from the linear model

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

and define the residuals corresponding to $\boldsymbol{\theta}$ as $r_i = r_i(\boldsymbol{\theta}) = y_i - \boldsymbol{\theta}' \mathbf{x}_i$.

In model (1), we denote by F_0 the (nominal) distribution of the errors u_i which are assumed to be independent of the carriers $\mathbf{x}_i$. Let G_0 be the (nominal) distribution of these carriers and suppose that there exists $E_{G_0} \mathbf{x}\mathbf{x}' = A$. From F_0, G_0 and the independence assumption, it is possible to compute the (nominal) joint distribution of $(y_i, \mathbf{x}_i)$, denoted by H_0. To allow for the presence of a proportion ϵ of outliers in the sample, we will assume that the true joint distribution of the data lies in the contamination neighborhood

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

The least squares method is the classical procedure to estimate the vector of parameters $\boldsymbol{\theta}_0$. However, as it is well-known, least squares estimators does not behave well when there are outliers in the sample. In fact, only one outlier may cause inferences to be highly unreliable. More stable methods consist in choosing the vector of parameters that minimizes a robust estimator of the scale of the residuals. For instance, regression S-estimators (see Rousseeuw and Yohai, 1984) are defined as

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}), \tag{2}$$

where $S_n(\boldsymbol{\theta})$ is a scale M-estimator computed from the residuals $r_i(\boldsymbol{\theta})$, that is,

$$S_n(\boldsymbol{\theta}) = \inf\{s > 0 : n^{-1} \sum_{i=1}^{n} \chi(r_i/s) < 0\},$$

for an appropriate score function χ. Another possibility is to define

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \sum_{p=1}^{[n(1-\alpha)]} r(\boldsymbol{\theta})^2_{(p)}, \tag{3}$$

where the subscript (p) stands for the p-th order statistic. This is the least trimmed squares (LTS) estimator proposed by Rousseeuw (1984). Other regression estimators included in this setup are τ-estimators (Yohai and Zamar, 1988) or R-estimators (Hössjer, 1994).

More recently, Croux, Rousseeuw and Hössjer (1994) have introduced the class of GS-estimators which is defined as in equation (2) but taking $S_n(\boldsymbol{\theta})$ to be a scale M-estimator of the pairwise differences $\{|r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta})| : \; i < j\}$ instead of a scale M-estimator of the residuals $\{r_i(\boldsymbol{\theta})\}$. This approach has several advantages: (1) efficiency is higher than in the case of simple S-estimators, (2) robustness properties are greatly preserved, (3) computation is not very expensive for some important particular cases, and (4) the objective function of GS-estimators does not depend on the intercept term of the model, which can be estimated as a second stage. (See Croux, *et al.* (1994) for a detailed discussion.)

The basic idea of using $\{|r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta})| : \; i < j\}$ rather than the residuals to compute the scale is also promising when applied to regression methods other than S-estimators. For example, if $h_n = n(n-1)/2$, a reasonable regression procedure could be

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \sum_{p=1}^{[h_n(1-\alpha)]} \{(r_i - r_j)^2 : \; i < j\}_{(p)}, \tag{4}$$

which is a generalized version of Rousseeuw's LTS-estimator. It is also possible to define generalized versions of τ- and R-estimators (see Section 4).

In this paper we provide a general method to study the robustness, under contamination of the sample, of regression estimators based on differences of residuals, such as the examples above. To measure the asymptotic stability, we consider the *maxbias curve.* This curve contains most of the asymptotic robustness properties of an estimator. It was initially defined by Huber (1964) and extensively exploited as a gauge of robustness by Martin and Zamar (1989) for scale M-estimators, and Martin, Yohai and Zamar (1989) for regression M-estimators with general scale.

Let $\mathbf{T}$ be a functional taking values on $\mathbb{R}^{p+1}$ and defined on a set of distributions on $\mathbb{R}^{p+1}$ that is large enough to contain both the empirical distribution functions and the neighborhood V_ϵ. The asymptotic bias of $\mathbf{T}$ at H, $b_A(\mathbf{T}, H)$, is defined so that it is invariant under regression equivariant transformations,

$$b_A(\mathbf{T}, H) = \{[\mathbf{T}(H) - \boldsymbol{\theta}_0]' A [\mathbf{T}(H) - \boldsymbol{\theta}_0]\}^{1/2}.$$

As we will only consider regression and affine equivariant estimators, we can assume without loss of generality that A is the identity matrix I and $\boldsymbol{\theta}_0 = \mathbf{0}$. Therefore, $b_A(\mathbf{T}, H) = b(\mathbf{T}, H) = \|\mathbf{T}(H)\|$. The maxbias curve of $\mathbf{T}$ is defined as

$$B_{\mathbf{T}}(\epsilon) = \sup_{H \in V_\epsilon} b(\mathbf{T}, H) = \sup_{H \in V_\epsilon} \|\mathbf{T}(H)\|, \tag{5}$$

that is, the maximum conceivable discrepancy between the value of the functional $\mathbf{T}$ at the nominal central distribution of the data and $\mathbf{T}(H)$, when H ranges over the neighborhood V_ϵ of all the possible true distributions of the data.

Several widely known one-figure summaries of the robustness of $\mathbf{T}$ can be computed from $B_\mathbf{T}(\epsilon)$. For instance, the *gross-error-sensitivity* introduced by Hampel (1974) equals, under regularity conditions, the derivative of the maxbias curve at zero, which was called the *contamination sensitivity* by He and Simpson (1993). On the other hand, the *breakdown point* of $\mathbf{T}$, also introduced by Hampel (1974), can be defined as

$$\epsilon^* = \sup\{\epsilon : \ B_\mathbf{T}(\epsilon) < \infty\}.$$

Therefore, the maxbias curve helps us to understand the robustness properties of an estimator both for small and large fractions of contamination. Unfortunately, it is sometimes a function difficult to derive and often each estimator requires a somewhat specialized method to compute it. So, it is useful to have general principles to perform the maxbias curve analysis. Section 3 points out one of these general principles.

The broad set of estimators our method deals with arises as a modification of the class of residual admissible regression estimators defined by Yohai and Zamar (1993). Roughly speaking, this class consists of estimators for which the empirical distribution of the absolute residuals cannot be uniformly improved –in the sense of stochastic dominance– by using any other set of regression coefficients. It can be shown that S-, R- and τ-estimators have this property. When we compute a residual admissible estimator from the pairwise differences of residuals, we obtain what we call a *generalized residual admissible estimator*. These are the estimators our method accounts for. Once we have computed the maxbias curves of the estimators in the general class, we will solve the problem of finding the most stable one.

The rest of the paper is organized as follows. Section 2 is devoted to introduce the required background on residual admissible estimators. Also, a formal definition is given for the class of generalized residual admissible estimators. Section 3 contains the main result of the paper. Some numerical examples show the applicability of this result in Section 4. In Section 5, we find the minimax bias estimator within the general class. The proofs of all the results are relegated to a final appendix.

2 Generalized residual admissible estimators

Let $F_{H,\boldsymbol{\theta}}$ denote the distribution function of the residuals $r_i(\boldsymbol{\theta})$ corresponding to $\boldsymbol{\theta}$ when H is the distribution of $(y_i, \mathbf{x}_i)$. If X and Y are independent random variables distributed as F, let F^* denote the distribution of the random variable $|X - Y|$. So, $F^*_{H,\boldsymbol{\theta}}$ is the distribution of $|r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta})|$ when $i \neq j$. Finally, let $F_{H_1 \times H_2,\boldsymbol{\theta}}$ denote the distribution of $|r_1(\boldsymbol{\theta}) - r_2(\boldsymbol{\theta})|$ when $(y_1, \mathbf{x}_1)$ is distributed as H_1 and $(y_2, \mathbf{x}_2)$ is distributed as H_2.

We start from estimators whose functional form is defined as

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} J(F_{H,\boldsymbol{\theta}}), \tag{6}$$

where $J(F)$ is a functional that measures the scale of F. Both the estimators defined in (2) and (3) are instances of these functionals when applied to the empirical distribution function of the data.

We will suppose that J satisfies

Assumption 1 *(a) If F and G are two distribution functions on $[0,\infty)$ such that $F(u) \leq G(u)$ for every $u \in \mathbb{R}$, then $J(F) \geq J(G)$.*

(b) (η-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 substochastic and continuous on $(0,\infty)$, with $G(\infty) \geq 1 - \eta$ and

$$G(u) \geq F(u), \text{ for every } u > 0, \tag{7}$$

then

$$\lim_{n\to\infty} J(F_n) \geq \lim_{n\to\infty} J(G_n). \tag{8}$$

Moreover, if (7) holds strictly, then (8) also holds strictly.

Assumption 1(a) is a monotonicity condition which implies that if the absolute residuals corresponding to $\boldsymbol{\theta}_1$ are stochastically smaller than the absolute residuals corresponding to $\boldsymbol{\theta}_2$, then $\boldsymbol{\theta}_2$ will never be the only solution of (6). Assumption 1(b) of η-monotonicity was introduced by Yohai and Zamar (1993). Notice that if we take $F_n = F$ and $G_n = G$ for each n, then η-monotonicity implies that if F and G are distribution functions on $[0,\infty)$, continuous on $(0,\infty)$ and such that $G(u) > F(u)$ for $u > 0$, then $J(F) > J(G)$. Therefore, we can view η-monotonicity as a strict monotonicity condition for certain especial distributions.

An estimator defined as in (6), where J satisfies Assumption 1, is called a *residual admissible estimator*. This definition is slightly more restrictive than that of Yohai and Zamar (1993) but it is suitable to cover all the relevant examples.

Following the ideas sketched in the introduction, we define the generalized residual admissible estimators as those that come up from functionals defined as

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} J(F^*_{H,\boldsymbol{\theta}}), \tag{9}$$

where J satisfies Assumption 1. The only difference with respect to equation (6) is that now we apply the scale functional J to the distribution of the absolute pairwise differences. In the following section we find out the maxbias curve of any generalized residual admissible estimator. Sometimes it will be convenient to use the notation $J^*(F) \equiv J(F^*)$.

3 A general method to compute maxbias curves

First, we list the assumptions required to prove our results. The nominal distributions F_0 and G_0 must verify the following hypothesis:

Assumption 2 *The distribution F_0 of the errors has a symmetric and strictly unimodal density f_0. The distribution G_0 of the carriers is such that $\boldsymbol{\theta}'\mathbf{x}$ has a symmetric and strictly unimodal density for each $\boldsymbol{\theta} \neq \mathbf{0}$.*

This hypothesis is fulfilled, for instance, when F_0 is the standard normal distribution and G_0 is the spherical multivariate normal distribution. As we will see, the applications of Theorem 1 become simpler when the regressors are spherical. However, this condition is not strictly necessary.

The following assumption is a regularity condition to be imposed on the scale functional J. It can be easily checked for the most important examples.

Assumption 3 *Let V_n be the uniform distribution on the interval $[n, 2n]$. Then*

$$\lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n] \geq J^*[F_{H,0}],$$

for each distribution $H \in V_\epsilon$.

This assumption implies that, if the scale is measured using pairwise differences of observations, the most harmful contamination occurs when both the location and the dispersion move away to infinity.

Next, we state the main result of the paper:

Theorem 1 *Let* $\mathbf{T}$ *be a regression functional defined as in equation (9). Then, under Assumption 1, for* $\eta = \epsilon(2-\epsilon)$*, and Assumptions 2 and 3,* $B_{\mathbf{T}}(\epsilon) = t^*$*, where* $t^* \in \mathbb{R}$ *is such that*

$$m(t^*) \equiv \inf_{\|\boldsymbol{\theta}\|=t^*} J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0] = \lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n], \tag{10}$$

V_n *is the uniform distribution on* $[n, 2n]$*, and* δ_0 *is the degenerated distribution giving probability one to zero.*

When G_0 is spherical, it is easy to prove that $F^*_{H_0,\boldsymbol{\theta}}$ only depends on $\boldsymbol{\theta}$ through the value of $\|\boldsymbol{\theta}\|$. Therefore, the infimum in equation (10) is no longer needed as each direction of $\boldsymbol{\theta}$ gives the same value of $J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0]$. We obtain the following result:

Corollary 1 *Under the notation and assumptions of Theorem 1 and assuming further that* G_0 *is spherical,* $B_{\mathbf{T}}(\epsilon) = \|\boldsymbol{\theta}^{\square}\|$*, where* $\boldsymbol{\theta}^{\square} \in \mathbb{R}^p$ *satisfies*

$$J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^{\square}} + \epsilon\delta_0] = \lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n]. \tag{11}$$

It is possible to give an intuitive interpretation of Corollary 1. Suppose there is a proportion ϵ of outliers placed at $(\boldsymbol{\theta}^{\square\prime}\mathbf{x}_n, \mathbf{x}_n)$, where $\mathbf{x}_n$ is uniformly distributed on the interval $(n\boldsymbol{\theta}^{\square}, 2n\boldsymbol{\theta}^{\square})$. Then $J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^{\square}} + \epsilon\delta_0]$ is the scale of the absolute residuals obtained when these outliers are perfectly fitted. On the other hand, $\lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n]$ is the scale obtained when the outliers are completely ignored and $n \to \infty$. Corollary 1 says that the maxbias curve is the value of $\|\boldsymbol{\theta}^{\square}\|$ such that both scales coincide.

4 Examples and numerical computations

For the sake of simplicity, we assume in this section that F_0 is the standard normal distribution, hereafter denoted by Φ, and G_0 is the spherical multivariate normal distribution. We will apply Corollary 1 to generalized versions of S-, R- and τ regression estimators. Assumptions of Corollary 1 can be checked under mild conditions in each case (see Lemmas 4 and 5 in the Appendix). We highlight here the wide applicability of the method.

4.1 GS-estimators

Consider the functional $\mathbf{T}_S(H) = \arg\min_{\boldsymbol{\theta}} S(F^*_{H,\boldsymbol{\theta}})$, where $S(F)$ is defined as

$$S(F) = \inf\{s : E_F\chi(y/s) < b\}, \tag{12}$$

and χ is a score function satisfying the assumptions of Lemma 4 in the appendix.

To illustrate the application of Corollary 1, define the following functions

$$g(s) = E_\Phi\chi(y/s),$$

and

$$h(s,\epsilon) = (1-\epsilon)^2 g(2^{-1/2}s) + 2\epsilon(1-\epsilon)g(s).$$

Let $S_1 = S^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^{\square}} + \epsilon\delta_0]$. We have that

$$(1-\epsilon)^2 E_{H_0\times H_0}\chi\left(\frac{r_1(\boldsymbol{\theta}^{\square}) - r_2(\boldsymbol{\theta}^{\square})}{S_1}\right) + 2\epsilon(1-\epsilon)E_{H_0}\chi\left(\frac{r(\boldsymbol{\theta}^{\square})}{S_1}\right) = b,$$

that is, $S_1 = (1 + \|\boldsymbol{\theta}^{\square}\|^2)^{1/2} h^{-1}(b,\epsilon)$.

On the other hand, let $S_2 = \lim_{n\to\infty} S^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n]$. Then, it is easy to show that

$$S_2 = 2^{1/2} g^{-1}\left(\frac{b - 2\epsilon + \epsilon^2}{(1-\epsilon)^2}\right).$$

Imposing the condition $S_1 = S_2$ and solving for $\|\boldsymbol{\theta}^{\square}\|$, we get

$$B_S(\epsilon) = \left[2\left(\frac{g^{-1}\left(\frac{b-2\epsilon+\epsilon^2}{(1-\epsilon)^2}\right)}{h^{-1}(b,\epsilon)}\right)^2 - 1\right]^{1/2}, \tag{13}$$

which amounts to the expression for the maxbias curve found by Croux *et al.* (1994).

Observe that $B_S(\epsilon)$ goes to infinity when the numerator of the fraction in the formula above goes to infinity or when the denominator goes to zero. Therefore, the asymptotic breakdown point of a GS-estimator is $\epsilon^* = \min\{1-(1-b)^{1/2}, (1-b)^{1/2}\}$ which equals 1/2 when $b = 0.75$. If we want the corresponding scale estimator to be consistent we must impose the Fisher-consistency condition $b = E_{F_0\times F_0}\chi(y_1 - y_2)$. This condition determines, through the value of b, the breakdown point of the regression estimator.

In Table 1 we present some numerical results. First, we have considered a jump score function $\chi_a(y) = I\{|y| > a\}$ where $a > 0$ is chosen so that $b = 0.75$. The corresponding estimator $\hat{\boldsymbol{\theta}}_S$ is the least quartile of differences (LQD) estimator. Another widely used option is provided by the biweight Tukey score function $\chi(y) = \min\{3y^2/c^2 - 3y^4/c^4 + y^6/c^6, 1\}$. When the tuning constant is $c = 0.9958$ we have $\epsilon^* = 1/2$. Call this estimator TUKEYGS. We have applied (13) to obtain the maxbias curves of both estimates for several values of the proportion of contamination. The obtained values are similar; so the corresponding robustness properties are not very different. Also, it can be proved that the efficiencies are alike; hence, what makes LQD preferable to TUKEYGS is its easier computability.

TABLE 1 ABOUT HERE

4.2 GR-estimators

Generalized estimators based on signed ranks (GR-estimators) originate from the functional $\mathbf{T}_R(H) = \arg\min_{\boldsymbol{\theta}} R(F^*_{H,\boldsymbol{\theta}})$, where

$$R(F) = \int_0^\infty a[F(u)]u^k dF(u), \quad a(u) \geq 0, \tag{14}$$

for some positive integer k.

These estimators select the vector of parameters that minimizes a weighted average of powers of the absolute pairwise differences of the residuals. The weights are given by a function $a(u)$ applied to the signed ranks of these absolute differences.

An interesting particular case is the generalized α-least trimmed absolute value (α-GLTAV) estimator which is defined by taking $k = 1$ and

$$a(u) = \begin{cases} 1, & |u| \leq 1 - \alpha \\ 0, & |u| > 1 - \alpha \end{cases} . \tag{15}$$

The corresponding estimator is

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \sum_{p=1}^{[h_n(1-\alpha)]} \{|r_i - r_j| : \ i < j\}_{(p)}, \tag{16}$$

where $h_n = n(n-1)/2$. The α-GLTS-estimator defined in equation (4) is obtained when $k = 2$ and $a(u)$ is as defined in (15).

To apply Corollary 1 we have to solve for $\|\boldsymbol{\theta}^\square\|$ the equation

$$R^*(F_{\boldsymbol{\theta}^\square}) = \lim_{n\to\infty} R^*(F_n),$$

where $F_{\boldsymbol{\theta}^{\square}} = (1-\epsilon)F_{H_0,\boldsymbol{\theta}^{\square}} + \epsilon\delta_0$ and $F_n = (1-\epsilon)F_{H_0,0} + \epsilon V_n$.

Denote $I_k(c) = \int_0^c u^k \varphi(u)du$, where φ is the standard normal density function. Let $\alpha \geq 1/2$. If $\epsilon < (1-\alpha)^{1/2} < 1-(1-\alpha)^{1/2}$, there exist $c_1 > 0$ and $c_2 > 0$ such that $F_{\boldsymbol{\theta}^{\square}}(\gamma c_1) = 1-\alpha$, and $(1-\epsilon)^2 F^*_{H_0,0}(2^{1/2}c_2) = 1-\alpha$, where $\gamma = (1+\|\boldsymbol{\theta}^{\square}\|^2)^{1/2}$.

We have that

$$\begin{aligned} R^*(F_{\boldsymbol{\theta}^{\square}}) &= \int_0^\infty a[F^*_{\boldsymbol{\theta}^{\square}}(u)]u^k dF^*_{\boldsymbol{\theta}^{\square}}(u) = \int_0^{\gamma c_1} u^k dF^*_{\boldsymbol{\theta}^{\square}}(u) \\ &= (1-\epsilon)^2 \int_0^{\gamma c_1} u^k dF^*_{H_0,\boldsymbol{\theta}^{\square}}(u) + 2\epsilon(1-\epsilon)\int_0^{\gamma c_1} u^k dF_{H_0,\boldsymbol{\theta}^{\square}}(u) \\ &= \gamma^k(1-\epsilon)[(1-\epsilon)2^{\frac{2+k}{2}} I_k(2^{-1/2}c_1) + 4\epsilon I_k(c_1)]. \end{aligned} \tag{17}$$

On the other hand, it can be shown applying Lemma 1 that

$$\lim_{n\to\infty} R^*(F_n) = (1-\epsilon)^2 \int_0^{2^{1/2}c_2} u^k dF^*_{H_0,0}(u) = (1-\epsilon)^2 2^{\frac{2+k}{2}} I_k(c_2). \tag{18}$$

From (17) and (18) we obtain, for $\epsilon < (1-\alpha)^{1/2}$,

$$[1+B_R^2(\epsilon)]^{k/2} = \frac{(1-\epsilon)I_k(c_2)}{(1-\epsilon)I_k(2^{-1/2}c_1) + 2^{\frac{2-k}{2}}\epsilon I_k(c_1)}. \tag{19}$$

Since $B_R(\epsilon)$ goes to infinity as $c_1 \to 0$, what in turn occurs whenever $\epsilon \to (1-\alpha)^{1/2}$, it follows that the breakdown point of $\mathbf{T}_R$ is $\epsilon^* = (1-\alpha)^{1/2}$.

Numerical results for both the 0.75-GLTAV and the 0.75-GLTS estimators can also be found in Table 1. These estimators are less robust than the GS-estimators studied above. The difference seems to be greater when the proportion of contamination is larger. Notice also that the 0.75-GLTS estimator is more robust than the 0.75-GLTAV estimator. In general, as the value of k increases, the robustness properties of the corresponding estimator are better. This is not surprising since, as $k \to \infty$, the sequence approaches the LQD estimator, which is fairly robust.

Plots of the maxbias curves of the four estimators we have studied can be found in Figure 1.

FIGURE 1 ABOUT HERE

4.3 Gτ-estimators

Let $S(F)$ be a functional based on a score function χ_1 that defines an S-estimator. Let χ_2 be another score function and suppose that both χ_1 and χ_2 satisfy the assumptions in

Lemma 4. Define

$$\tau(F) = S(F)\left[E_{F\times F}\chi_2\left(\frac{y_1 - y_2}{S^*(F)}\right)\right]^{1/2}.$$

Then $\mathbf{T}_\tau(H) = \arg\min_{\boldsymbol{\theta}} \tau(F^*_{H,\boldsymbol{\theta}})$ is called a generalized τ-estimator. For finite samples, we have

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} S_n \left[\sum_{i<j} \chi_2\left(\frac{r_i - r_j}{S_n}\right)\right]^{1/2},$$

where $S_n = S_n(\boldsymbol{\theta})$ is the scale M-estimator based on χ_1 of the pairwise differences $\{|r_i - r_j| : i < j\}$.

Yohai and Zamar (1993) proved that $\tau(F)$ is η-monotone for each $\eta > 0$. Furthermore, if it is assumed that χ_2 is such that $f(s) = s^2 E_{F\times F}\chi_2[(y_1 - y_2)/s]$ is not decreasing for each distribution F, then both Assumptions 1(a) and 3 are fulfilled.

As in the examples above, we can apply Corollary 1 to compute the maxbias curve of the Gτ-estimator based on χ_1 and χ_2, that will be denoted by $B_\tau(\epsilon)$. Let $B_S(\epsilon)$ be the maxbias curve of the GS-estimate defined by χ_1. Some manipulations analogous to those corresponding to GS-estimators allow us to show that $B_\tau(\epsilon)$ satisfies

$$1 + B_\tau^2(\epsilon) = [1 + B_S^2(\epsilon)]H(\epsilon), \tag{20}$$

where

$$H(\epsilon) = \frac{(1-\epsilon)^2 g_2\left[g_1^{-1}\left(\frac{b - 2\epsilon + \epsilon^2}{(1-\epsilon)^2}\right)\right] + 2\epsilon - \epsilon^2}{h_2\left[h_1^{-1}(b, \epsilon), \epsilon\right]},$$

with $g_i(s) = E_\Phi \chi_i(y/s)$ and

$$h_i(s, \epsilon) = (1-\epsilon)^2 g_i(2^{-1/2}s) + 2\epsilon(1-\epsilon)g_i(s) \quad \text{for } i = 1, 2.$$

Since $0 < H(\epsilon) < \infty$, for $\epsilon < \min\{(1-b)^{1/2}, 1 - (1-b)^{1/2}\}$, where $b = E_{\Phi\times\Phi}\chi_1(y_1 - y_2)$, it follows that the breakdown point of a Gτ-estimator is solely characterized by χ_1 irrespective of χ_2. This second score function may be chosen in order to reach higher efficiency.

5 Minimax bias theory

Classical robustness theory deals with minimax problems. In his pioneering paper, Huber (1964) proved that the median is minimax bias (it has the minimum maxbias curve) within the set of all the affine equivariant location estimators, for any fraction of contamination. In this section we obtain a similar result for the class of generalized residual admissible regression estimators, although in this case the minimax solution will be slightly different depending on the proportion of contamination.

Assume in this section that G_0 is a spherical distribution. It will be shown that, for each $0 < \epsilon < 1/2$, there exists $0 < \alpha^* < 1$ such that the maxbias curve at ϵ of the functional

$$\mathbf{T}_{\alpha^*}(H) = \arg\min_{\boldsymbol{\theta}} F^{-1}_{H\times H,\boldsymbol{\theta}}(\alpha^*) \tag{21}$$

is less than the maxbias curve at ϵ of any other generalized residual admissible regression estimator. Notice that $\mathbf{T}_{\alpha^*}(H)$ is the set of regression coefficients that minimizes the α^*-quantile of the distribution of the absolute pairwise differences of residuals. The value of the optimal quantile to be minimized for each proportion of contamination ϵ is a by-product of the proof of Theorem 2 and will be computed later on. The proofs in this section follow closely those by Yohai and Zamar (1993); so, the reader may want to go through that paper for some of the details. The announced result, proved after some lemmas in the appendix, is the following:

Theorem 2 *Let* $\mathbf{T}$ *be a generalized residual admissible regression estimator based on a scale fuctional* J *satisfying Assumption 1 with* $\eta = \epsilon(2-\epsilon)$ *and Assumption 3. Suppose that* G_0 *is spherical and that Assumption 2 holds. For each* $0 < \epsilon < 1/2$*, there exists* $0 < \alpha^* < 1$ *such that* $B_{\mathbf{T}_{\alpha^*}}(\epsilon) \leq B_{\mathbf{T}}(\epsilon)$*, where* $\mathbf{T}_{\alpha^*}$ *was defined in equation (21).*

Table 2 gives the values of the optimal quantiles to be minimized for the gaussian central model. Also, the maxbias of the minimax estimator (the minimax bias) has been computed and compared with the minimax bias for the class of the (non-generalized) residual admissible estimators, as reported by Yohai and Zamar (1993). Generalized optimal estimates are less robust, the difference being larger as ϵ increases.

TABLE 2 ABOUT HERE

On the other hand, notice that the optimal quantile does not change much with the value of ϵ. In Figure 2, both the minimax bias and the maxbias curve of the LQD have been plotted. Note that the robustness properties of the LQD are quite close to the optimum for the whole class *for any fraction of contamination.* The existence of such a quasi-optimal estimator for any amount of outliers is not surprising as both the LMS-estimator, for (non-generalized) admissible residual estimators, and an appropriatedly scaled median, for scale M-estimators, play a similar role (see Yohai and Zamar, 1993, and Martin and Zamar, 1989).

FIGURE 2 ABOUT HERE

Appendix. Proofs

We need three auxiliary lemmas before proving Theorem 1.

Lemma 1 *Let* $\{c_n\}$ *be a sequence of real numbers such that* $\lim_{n\to\infty} c_n = \infty$. *Let* $\{\boldsymbol{\theta}_n\}$ *be a convergent sequence in* $\mathbb{R}^p$. *Finally, let* V_n *be a distribution of* $(y_n, \mathbf{x}_n)$ *such that the residuals* $r(\boldsymbol{\theta}_n)$, *under* V_n, *are uniformly distributed on the interval* $(c_n, 2c_n)$. *If Assumption 2 holds, then*

(a) $\lim_{n\to\infty} F_{V_n\times V_n,\boldsymbol{\theta}_n}(u) = 0,$ *for each* $u > 0$.

(b) $\lim_{n\to\infty} F_{H_0\times V_n,\boldsymbol{\theta}_n}(u) = 0,$ *for each* $u > 0$.

Proof:

For proving part (a), notice that

$$F_{V_n\times V_n,\boldsymbol{\theta}_n}(u) = \mathrm{P}_{V_n\times V_n}\{|r_1(\boldsymbol{\theta}_n) - r_2(\boldsymbol{\theta}_n)| \leq u\} = 2\int_0^u \frac{c_n - x}{c_n^2} I\{0 < x < c_n\}dx.$$

If n is large enough,

$$F_{V_n\times V_n,\boldsymbol{\theta}_n}(u) = 2\int_0^u \frac{c_n - x}{c_n^2}dx = \frac{2}{c_n^2}\left[c_n u - \frac{u^2}{2}\right],$$

that converges to zero as $n \to \infty$, for each $u > 0$.

To show part (b), we write

$$\begin{aligned} F_{H_0\times V_n,\boldsymbol{\theta}_n}(u) &= \mathrm{P}_{H_0\times V_n}\{|r_1(\boldsymbol{\theta}_n) - r_2(\boldsymbol{\theta}_n)| \leq u\} = \mathrm{P}_{H_0\times V_n}\{r_2 - u \leq r_1 \leq r_2 + u\} \\ &= \frac{1}{c_n}\int_{c_n}^{2c_n} P_{H_0}\{r_2 - u \leq r_1 \leq r_2 + u\}dr_2 \\ &= F_{H_0,\boldsymbol{\theta}_n}(\xi_n + u) - F_{H_0,\boldsymbol{\theta}_n}(\xi_n - u), \end{aligned}$$

where $\xi_n \in (c_n, 2c_n)$ (by the mean value theorem for integrals of positive continuous functions). Since $\{\boldsymbol{\theta}_n\}$ converges,

$$\lim_{n\to\infty} F_{H_0,\boldsymbol{\theta}_n}(\xi_n + u) - F_{H_0,\boldsymbol{\theta}_n}(\xi_n - u) = 0,$$

for each $u > 0$.

Lemma 2 *Let Δ_0 be the degenerated distribution giving probability one to the vector $\mathbf{0} \in \mathbb{R}^{p+1}$. If Assumption 2 holds,*

$$F_{H_0\times\tilde{H},\boldsymbol{\theta}}(u) \leq F_{H_0,\boldsymbol{\theta}}(u) = F_{H_0\times\Delta_0,\boldsymbol{\theta}}(u)$$

for each $u \geq 0$, $\boldsymbol{\theta} \in \mathbb{R}^p$ and for each distribution $\tilde{H}$.

Proof:

By Assumption 2, the distribution of the residuals $r_i(\boldsymbol{\theta})$ is symmetric and strictly unimodal. From this fact and given that r_1 and r_2 are independent,

$$\begin{aligned} F_{H_0\times\tilde{H},\boldsymbol{\theta}}(u) &= \int \mathrm{P}_{H_0}\{r_2 - u \leq r_1 \leq r_2 + u\} dF_{\tilde{H},\boldsymbol{\theta}}(r_2) \\ &\leq \int \mathrm{P}_{H_0}\{-u \leq r_1 \leq u\} dF_{\tilde{H},\boldsymbol{\theta}}(r_2) = F_{H_0,\boldsymbol{\theta}}(u). \end{aligned}$$

Lemma 3 *Let J be a functional satisfying Assumption 1(b) for some $\eta > 0$. Define for each $t > 0$,*

$$m(t) = \inf_{\|\boldsymbol{\theta}\|=t} J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0].$$

If Assumption 2 holds,

(a) There exists $\boldsymbol{\theta}_t \in \mathbb{R}^p$ such that $\|\boldsymbol{\theta}_t\| = t$, and $m(t) = J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_t} + \epsilon\delta_0]$.

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

Proof:

Since $\{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| = t\}$ is a compact set, to prove part (a) it is enough to show that the function $f(\boldsymbol{\theta}) = J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0]$ is continuous. Let $\{\boldsymbol{\theta}_n\}$ be a sequence in $\mathbb{R}^p$ such that $\lim_{n\to\infty} \boldsymbol{\theta}_n = \boldsymbol{\theta}$. We have that

$$\begin{aligned} J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_n} + \epsilon\delta_0] &= J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_n} + \epsilon F_{\Delta_0,\boldsymbol{\theta}}] \\ &= J[(1-\epsilon)^2 F^*_{H_0,\boldsymbol{\theta}_n} + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\boldsymbol{\theta}_n} + \epsilon^2 F^*_{\Delta_0,\boldsymbol{\theta}_n}], \end{aligned}$$

where Δ_0 stands for the distribution that gives probability one to the vector $\mathbf{0}$. By Assumption 1(b),

$$\begin{aligned} \lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_n} + \epsilon\delta_0] &= J[(1-\epsilon)^2 F^*_{H_0,\boldsymbol{\theta}} + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\boldsymbol{\theta}} + \epsilon^2 F^*_{\Delta_0,\boldsymbol{\theta}}] \\ &= J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0]. \end{aligned}$$

To prove part (b), consider $t_1 > t_2 > 0$. By part (a), there exists $\boldsymbol{\theta}_1 \in \mathbb{R}^p$ such that $m(t_1) = J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_1} + \epsilon\delta_0]$. Under Assumption 2, Lemma A.3 of Yohai and Zamar (1993) holds. Therefore, both $F_{H_0,\lambda\boldsymbol{\theta}}(u)$ and $F^*_{H_0,\lambda\boldsymbol{\theta}}(u)$ are strictly increasing as functions of λ. Then, for each $u > 0$,

$$\begin{aligned} & (1-\epsilon)^2 F^*_{H_0,\boldsymbol{\theta}_1}(u) + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\boldsymbol{\theta}_1}(u) + \epsilon^2 F^*_{\Delta_0,\boldsymbol{\theta}_1}(u) \\ < \; & (1-\epsilon)^2 F^*_{H_0,(t_2/t_1)\boldsymbol{\theta}_1}(u) + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,(t_2/t_1)\boldsymbol{\theta}_1}(u) + \epsilon^2 F^*_{\Delta_0,(t_2/t_1)\boldsymbol{\theta}_1}(u). \end{aligned}$$

Applying this inequality and Assumption 1(b),

$$m(t_1) > J^*[(1-\epsilon)F_{H_0,(t_2/t_1)\boldsymbol{\theta}_1} + \epsilon\delta_0] \geq m(t_2),$$

but $J^*[(1-\epsilon)F_{H_0,(t_2/t_1)\boldsymbol{\theta}_1} + \epsilon\delta_0] \geq m(t_2)$ by definition of $m(t)$. It follows that $m(t)$ is strictly increasing.

Proof of Theorem 1

First, we prove that $B_{\mathbf{T}}(\epsilon) \leq t^*$. Let $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ be such that $\|\tilde{\boldsymbol{\theta}}\| = t > t^*$. It is enough to show that there is no $H \in V_\epsilon$ such that $\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J^*(F_{H,\boldsymbol{\theta}})$. We will actually show that for every $H \in V_\epsilon$, $J^*(F_{H,\tilde{\boldsymbol{\theta}}}) > J^*(F_{H,0})$.

Applying Lemma 2, we have that for each $H \in V_\epsilon$, and $u > 0$,

$$F^*_{H,\tilde{\boldsymbol{\theta}}}(u) \leq (1-\epsilon)^2 F_{H_0\times H_0,\tilde{\boldsymbol{\theta}}}(u) + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\tilde{\boldsymbol{\theta}}} + \epsilon^2\delta_0(u). \tag{22}$$

Therefore, by inequality (22) and Assumption 1(a),

$$J^*(F_{H,\tilde{\boldsymbol{\theta}}}) \geq J^*[(1-\epsilon)F_{H_0,\tilde{\boldsymbol{\theta}}} + \epsilon\delta_0]. \tag{23}$$

By the definition of the function $m(t)$, Lemma 3(b), and condition (10), given that $t > t^*$,

$$J^*[(1-\epsilon)F_{H_0,\tilde{\boldsymbol{\theta}}} + \epsilon\delta_0] \geq m(t) > m(t^*) = \lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n]. \tag{24}$$

Finally, by Assumption 3,

$$\lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n] \geq J^*[F_{H,0}]. \tag{25}$$

Putting together the inequalities (23), (24), and (25) it follows that

$$J^*(F_{H,\tilde{\boldsymbol{\theta}}}) > J^*(F_{H,0}), \quad \text{for each } H \in V_\epsilon.$$

Now, we establish the converse inequality, $B_{\mathbf{T}}(\epsilon) \geq t^*$. Given $0 < t < t^*$, by Lemma 3 there exists $\boldsymbol{\theta}_t \in \mathrm{I\!R}^p$ such that $\|\boldsymbol{\theta}_t\| = t$ and

$$m(t) = J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_t} + \epsilon\delta_0].$$

Define the following sequence of contaminating distributions: $\tilde{H}_n = \delta_{(y_n,\mathbf{x}_n)}$, where $y_n = \boldsymbol{\theta}_t'\mathbf{x}_n$ and $\mathbf{x}_n$ is uniformly distributed on the interval $(n\boldsymbol{\theta}_t, 2n\boldsymbol{\theta}_t)$. Given $\boldsymbol{\beta} \in \mathrm{I\!R}^p$, the following distributional equality,

$$(\boldsymbol{\theta}_t - \boldsymbol{\beta})'\mathbf{x}_n \equiv U[n(t^2 - \boldsymbol{\beta}'\boldsymbol{\theta}_t), 2n(t^2 - \boldsymbol{\beta}'\boldsymbol{\theta}_t)], \tag{26}$$

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

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

Using Lemma 1, it is not difficult to show the following two facts:

$$\lim_{n\to\infty} F_{H_0\times\tilde{H},\mathbf{T}_n}(u) = 0, \tag{27}$$

and

$$\lim_{n\to\infty} F_{\tilde{H}\times\tilde{H},\mathbf{T}_n}(u) = 0, \tag{28}$$

for each $u \geq 0$. For instance, to show (27), we apply (26) to deduce that $r_2(\mathbf{T}_n) = (\boldsymbol{\theta}_t - \mathbf{T}_n)'\mathbf{x}_n$ is uniformly distributed on the interval $(c_n, 2c_n)$, where $c_n = n(t^2 - \mathbf{T}_n'\boldsymbol{\theta}_t)$ goes to infinity since $\lim_{n\to\infty} \mathbf{T}_n = \tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}'\boldsymbol{\theta}_t \leq \tilde{t}t < t^2$. We can apply Lemma 1(b) with $V_n = \tilde{H}_n$ y $\boldsymbol{\theta}_n = \mathbf{T}_n$ to obtain (27). The equality (28) is proved similarly.

By (27) and (28),

$$\lim_{n\to\infty} F_{H_n \times H_n, \mathbf{T}_n}(u) = (1-\epsilon)^2 F_{H_0\times H_0, \tilde{\boldsymbol{\theta}}}(u), \quad \text{for each } u \geq 0, \tag{29}$$

and

$$\lim_{n\to\infty} F_{H_n\times H_n,0}(u) = (1-\epsilon)^2 F_{H_0\times H_0,0}(u), \quad \text{for each } u \geq 0. \tag{30}$$

By (29), (30) and given that J is η-monotone for $\eta = \epsilon(1-\epsilon)$,

$$\lim_{n\to\infty} J^*(F_{H_n,\mathbf{T}_n}) \geq \lim_{n\to\infty} J^*(F_{H_n,0}). \tag{31}$$

Since $t < t^*$, and using condition (10),

$$\begin{aligned} \lim_{n\to\infty} J^*(F_{H_n,0}) &= \lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n] = m(t^*) \\ &> m(t) = J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_t} + \epsilon\delta_0], \end{aligned} \tag{32}$$

as we chose $\boldsymbol{\theta}_t$ in the required direction to get the last equality.

Observe that, under $\tilde{H}_n$, we have that $r(\boldsymbol{\theta}_t) = \boldsymbol{\theta}_t'\mathbf{x}_n - \boldsymbol{\theta}_t'\mathbf{x}_n = 0$ and therefore

$$\lim_{n\to\infty} F_{H_n\times H_n,\boldsymbol{\theta}_t}(u) = (1-\epsilon)^2 F^*_{H_0,\boldsymbol{\theta}_t}(u) + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\boldsymbol{\theta}_t}(u) + \epsilon^2 F^*_{\Delta_0,\boldsymbol{\theta}_t}(u).$$

Again, as J is η-monotone for $\eta = \epsilon(1-\epsilon)$,

$$\lim_{n\to\infty} J^*(F_{H_n,\boldsymbol{\theta}_t}) = J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_t} + \epsilon F_{\Delta_0,\boldsymbol{\theta}_t}] = J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_t} + \epsilon\delta_0]. \tag{33}$$

Joining (31), (32) and (33), it follows that

$$\lim_{n\to\infty} J^*(F_{H_n,\mathbf{T}_n}) > \lim_{n\to\infty} J^*(F_{H_n,\boldsymbol{\theta}_t}).$$

Therefore, if n is large enough, $J^*(F_{H_n,\mathbf{T}_n}) > J^*(F_{H_n,\boldsymbol{\theta}_t})$. This last inequality is a contradiction since $\mathbf{T}_n = \arg\min_{\boldsymbol{\theta}} J^*(F_{H_n,\boldsymbol{\theta}})$.

For every $t < t^*$ we have found a sequence of distributions $\{H_n\}$ in the neighborhood V_ϵ such that $\sup_n \|\mathbf{T}(H_n)\| \geq t$. The second part of the result follows immediately from this fact.

Both Lemma 4 and Lemma 5 are devoted to check that GS- and GR-estimators satisfy the assumptions of Theorem 1.

Lemma 4 *If the score function χ is even, monotone on $[0, \infty)$, bounded, continuous at 0 with $0 = \chi(0) < \chi(\infty) = 1$ and with at most a finite number of discontinuities, then $S(F)$ satisfies Assumption 1, for each $\eta > 0$, and Assumption 3.*

Proof:

Let F and G be a pair of distribution functions on $[0, \infty)$ such that $F(u) \le G(u)$ for each $u \ge 0$. There exist two random variables X and Y distributed as F and G respectively such that $X \ge Y$ and, hence, $\chi(X/s) \ge \chi(Y/s)$ for each $s > 0$. It follows that

$$S(F) = \inf\{s > 0 : E\chi(X/s) \le b\} \ge \inf\{s > 0 : E\chi(Y/s) \le b\} = S(G).$$

Assumption 1(a) holds.

Yohai and Zamar (1993), Lemma 5.1, proved that $S(H)$ is η-monotone for each $\eta > 0$. Therefore Assumption 1(b) also holds.

Finally, to check Assumption 3, notice that $s_\infty = \lim_{n\to\infty} S^*[(1-\epsilon)F + \epsilon V_n]$ satisfies

$$(1-\epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_\infty}\right) dF(y) + 2\epsilon(1-\epsilon) + \epsilon^2 \le b. \tag{34}$$

On the other hand, if $s_0 = S^*[(1-\epsilon)F + \epsilon G]$ was such that

$$(1-\epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_0}\right) dF(y) + 2\epsilon(1-\epsilon) + \epsilon^2 < b,$$

then there would exist $s < s_0$ such that

$$(1-\epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s}\right) dF(y) + 2\epsilon(1-\epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{u}{s_n}\right) dF_{H_0\times G,0}(u)$$
$$+\epsilon^2 \int_{-\infty}^{\infty} \chi\left(\frac{u}{s_n}\right) dF_{G\times G,0}(u) < b,$$

which contradicts the definition of s_0. It follows that

$$(1-\epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_0}\right) dF(y) + 2\epsilon(1-\epsilon) + \epsilon^2 \ge b.$$

By inequality (34), we must have $s_\infty \ge s_0$ and therefore Assumption 3 also holds.

Lemma 5 *Let $\alpha \ge 1/2$. Suppose that $a(u)$ is continuous on $[0, 1-\alpha]$, $a(u) = 0$ if $1-\alpha < u \le 1$, and $a(u) > 0$ if $0 < u < 1-\alpha$. Then $R(F)$ satisfies Assumption 1, for $\eta < \alpha$, and Assumption 3. 3*

Proof:

Assumption 1(a) holds since it amounts to the conclusion of Lemma A.4 of Yohai and Zamar (1993).

Yohai and Zamar (1993), Theorem 5.2, proved that $R(F)$ is η-monotone for $\eta < \alpha$. As it is required that $\eta = \epsilon(1-\epsilon)$, we will be able to compute the maxbias curve for $\epsilon < 1-(1-\alpha)^{1/2}$. This is not a real restriction as the breakdown point of these estimators is always less than or equal to $1-(1-\alpha)^{1/2}$.

Finally, Assumption 3 follows straightforwardly from the fact that Assumption 1(a) holds even for substochastic distributions and applying η-monotonicity.

The following two lemmas are needed to prove the minimax result of Section 5.

Lemma 6 *Let $\tilde{H} = (1-\epsilon)H_0 + \epsilon\Delta_0$. Define*

$$\|\boldsymbol{\theta}^{\square}\| = \sup\{\|\boldsymbol{\theta}\| : \; F^*_{\tilde{H},\boldsymbol{\theta}}(u) \geq (1-\epsilon)^2 F^*_{H_0,0}(u), \quad \textit{for each } u \geq 0\}. \tag{35}$$

Then, under Assumption 2,

(a) $0 < \|\boldsymbol{\theta}^{\square}\| < \infty$, *for each* $0 < \epsilon < 1/2$.

(b) There exists $u^* > 0$ *such that,*

$$F^*_{\tilde{H},\boldsymbol{\theta}^{\square}}(u^*) = (1-\epsilon)^2 F^*_{H_0,0}(u^*). \tag{36}$$

Proof:

It can be obtained by followin the proofs of Lemmas A.2 and A.3 of Yohai and Zamar (1993), rewritting the details when necessary.

The following lemma is perhaps interesting by itself. It gives a general lower bound for the maxbias curve of any generalized residual admissible estimator.

Lemma 7 *Let $\|\boldsymbol{\theta}^{\square}\|$ be as defined in equation (35). Let $\mathbf{T}$ be a regression estimator defined as in (9). Under Assumption 1(b), for $\eta = \epsilon(1-\epsilon)$, and Assumptions 2 and 3, and assuming further that G_0 is spherical, $B_{\mathbf{T}}(\epsilon) \geq \|\boldsymbol{\theta}^{\square}\|$.*

Proof:

By Corollary 1, $B_{\mathbf{T}}(\epsilon) = \|\boldsymbol{\theta}\|$, where $\boldsymbol{\theta}$ is such that

$$\begin{aligned} J[F^*_{\tilde{H},\boldsymbol{\theta}}] &= J^*[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^\square} + \epsilon\delta_0] = \lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n] \\ &= \lim_{n\to\infty} J[(1-\epsilon)^2 F^*_{H_0,0} + 2\epsilon(1-\epsilon)F_{H_0\times V_n,0} + \epsilon^2 F_{V_n\times V_n,0}]. \end{aligned}$$

This fact implies that there exists $\tilde{u} > 0$ such that

$$F^*_{\tilde{H},\boldsymbol{\theta}}(\tilde{u}) = (1-\epsilon)^2 F^*_{H_0,0}(\tilde{u}), \tag{37}$$

since if $F^*_{\tilde{H},\boldsymbol{\theta}}(u) = (1-\epsilon)^2 F^*_{H_0,0}(u)$ for each $u > 0$, then $J^*[F^*_{\tilde{H},\boldsymbol{\theta}}] < \lim_{n\to\infty} J^*[(1-\epsilon)F_{H_0,0} + \epsilon V_n]$ by Assumption 1(b). Notice that by Lemma 1(a) and (b),

$$\lim_{n\to\infty}(1-\epsilon)^2 F^*_{H_0,0}(u) + 2\epsilon(1-\epsilon)F_{H_0\times V_n,0}(u) + \epsilon^2 F_{V_n\times V_n,0}(u) = (1-\epsilon)^2 F^*_{H_0,0}(u).$$

Lets assume that $\|\boldsymbol{\theta}\| < \|\boldsymbol{\theta}^\square\|$. By the definition of $\boldsymbol{\theta}^\square$, equation (35),

$$F^*_{\tilde{H},\boldsymbol{\theta}}(\tilde{u}) > F^*_{\tilde{H},\boldsymbol{\theta}^\square}(\tilde{u}) \geq (1-\epsilon)^2 F_{H_0,0}(\tilde{u}),$$

which is a contradiction with equation (37). As $\|\boldsymbol{\theta}\| < \|\boldsymbol{\theta}^\square\|$ leads to a contradiction, $\|\boldsymbol{\theta}\| \geq \|\boldsymbol{\theta}^\square\|$ holds.

Proof of Theorem 2:

Let $\alpha^* = (1-\epsilon)^2 F_{H_0\times H_0,0}(u^*)$, where $u^* > 0$ was defined in Lemma 6(b). By Lemma 7, it is enough to show that $\|\mathbf{T}_{\alpha^*}(H)\| \leq \|\boldsymbol{\theta}^\square\|$ for each $H \in V_\epsilon$. On the contrary, suppose that there exists $H = (1-\epsilon)H_0 + \epsilon H^*$ such that $\boldsymbol{\theta} = \mathbf{T}_{\alpha^*}(H)$ and $\|\boldsymbol{\theta}\| > \|\boldsymbol{\theta}^\square\|$. Let $\lambda = \|\boldsymbol{\theta}\|/\|\boldsymbol{\theta}^\square\| > 1$ and define $\tilde{\boldsymbol{\theta}} = \lambda\boldsymbol{\theta}^\square$. Since $\|\tilde{\boldsymbol{\theta}}\| = \|\boldsymbol{\theta}\|$, applying Lemmas 2 and 6(b) we have that,

$$\begin{aligned} F^*_{H,\boldsymbol{\theta}}(u^*) &= (1-\epsilon)^2 F^*_{H_0,\boldsymbol{\theta}}(u^*) + 2\epsilon(1-\epsilon)F_{H_0\times H^*,\boldsymbol{\theta}}(u^*) + \epsilon^2 F^*_{H^*\times H^*,\boldsymbol{\theta}}(u^*) \\ &\leq (1-\epsilon)^2 F^*_{H_0,\boldsymbol{\theta}}(u^*) + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\boldsymbol{\theta}}(u^*) + \epsilon^2 \\ &= (1-\epsilon)^2 F^*_{H_0,\tilde{\boldsymbol{\theta}}}(u^*) + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\tilde{\boldsymbol{\theta}}}(u^*) + \epsilon^2 \\ &< (1-\epsilon)^2 F^*_{H_0,\boldsymbol{\theta}^\square}(u^*) + 2\epsilon(1-\epsilon)F_{H_0\times\Delta_0,\boldsymbol{\theta}^\square}(u^*) + \epsilon^2 \\ &= (1-\epsilon)^2 F_{H_0\times H_0,0}(u^*) = \alpha^*. \end{aligned}$$

Therefore,

$$F^{-1}_{H\times H,\boldsymbol{\theta}}(\alpha^*) > u^*. \tag{38}$$

On the other hand, $F^*_{H\times H,0}(u^*) \geq (1-\epsilon)^2 F_{H_0\times H_0,0}(u^*) = \alpha^*$ and, hence,

$$F^{-1}_{H\times H,0}(\alpha^*) \leq u^*. \tag{39}$$

From (38) and (39),

$$F^{-1}_{H\times H,\boldsymbol{\theta}}(\alpha^*) > F^{-1}_{H\times H,0}(\alpha^*),$$

which is a contradiction with $\boldsymbol{\theta} = \mathbf{T}_{\alpha^*}(H)$. Therefore $\|\mathbf{T}_{\alpha^*}(H)\| \leq \|\boldsymbol{\theta}^\square\|$.

References

Croux, C., P.J. Rousseeuw and O. Hössjer (1994) Generalized S-estimators. *J. Amer. Statist. Assoc.* **89**, 1271–1281.

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

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

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.

Martin, R.D., V.J. Yohai and R.H. Zamar (1989). Min-Max bias robust regression. *Ann. Statist.* **17**, 1608–1630.

Martin, R.D. and R.H. Zamar (1989). Asymptotically Min-Max bias robust M-estimates of scale for positive random variables. *J. Amer. Statist. Assoc.* **17**, 494–501.

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

Rousseeuw, P.J. and V.J. Yohai (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle, R.D. Martin., Eds., *Robust and Nonlinear Time Series Analysis*, Lecture Notes in Statistics 26, Springer, New-York, 256–272.

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

epsilon	0.05	0.10	0.15	0.20	0.25	0.30
LQD	0.59	0.95	1.34	1.83	2.51	3.56
TUKEYGS	0.60	0.96	1.36	1.88	2.60	3.73
α-GLTS	0.60	0.98	1.42	2.02	2.90	4.37
α-GLTAV	0.61	1.02	1.52	2.24	3.38	5.49

Table 1: Maxbias curves of several robust regression estimates based on pairwise differences of residuals.

epsilon	0.05	0.10	0.15	0.20
Quantile	0.27	0.32	0.33	0.34
$B^*(\epsilon)$	0.59	0.94	1.32	1.78
$B(\epsilon)$	0.49	0.77	1.05	1.37

Table 2: Optimal quantiles and minimax bias values in the class of generalized residual admissible estimators, $B^*(\epsilon)$. For comparison purposes, the minimax bias values for the class of (non-generalized) residual admissible estimator, $B(\epsilon)$, are also included.