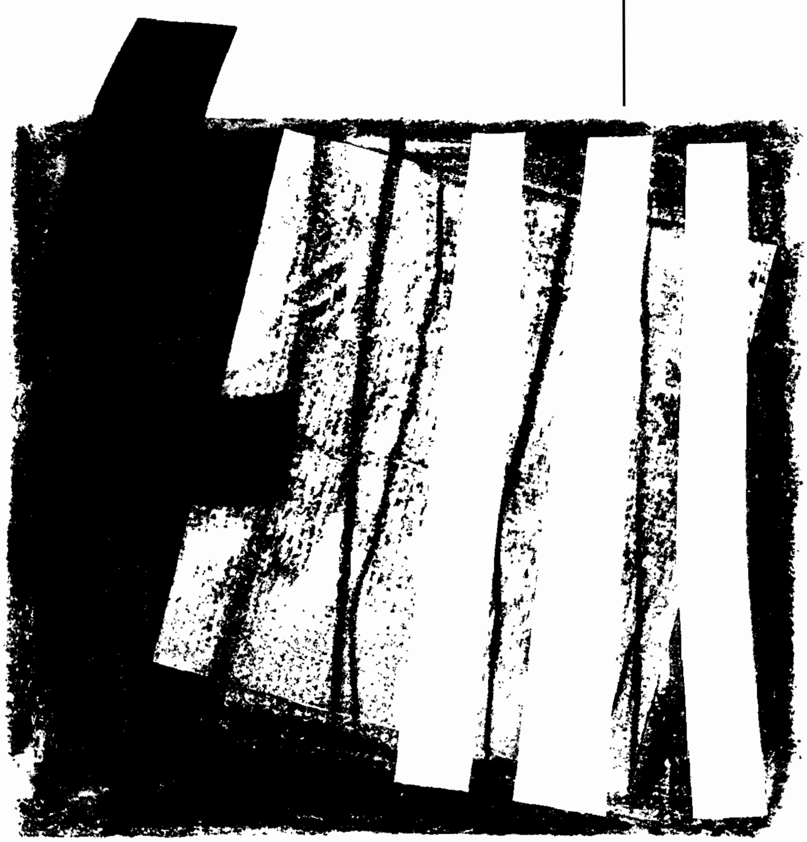
ON THE MAXBIAS CURVE OF RESIDUAL ADMISSIBLE ROBUST REGRESSION ESTIMATES

José R. Berrendero
and Rubén Zamar

95-62



WORKING PAPERS

Working Paper 95-62
Statistics and Econometrics Series 26
December 1995

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe, Madrid (Spain)
Fax (341) 624-9849

ON THE MAXBIAS CURVE OF RESIDUAL ADMISSIBLE ROBUST REGRESSION ESTIMATES

José R. Berrendero and Rubén Zamar [1]

Abstract

The robustness properties of a regression estimate are throughly described by its maxbias curve. However, this function is difficult to compute, especially when the regressors are not elliptically distributed. In this paper, we propose a general method for computing maxbias curves, valid for a large number of robust regression estimates, namely, those estimates defined by residual admissible functionals. Our results are also useful to compute maxbias curves when the regressors are not elliptically distributed. We provide several examples of application of the method which include S-, τ-, and signed R-estimates. MM-estimates are also studied under a related, although slightly different, approach.

Keywords:
Robust regression, Bias robustness, Maxbias curve, S-estimates, tau-estimates, R-estimates, MM-estimates.

[1] Berrendero, Department of Statistics and Econometrics, Universidad Carlos III de Madrid, research partially supported by DGICYT PB93-0232, and Zamar, Department of Statistics, University of British Columbia, research partially supported by the Natural Sciences and Engineering Research Council of Canada.

1 Introduction

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

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

where the errors u_i have a common distribution, F_0, and are independent of the variables $\mathbf{x}_i$. Assume that the regressors $\mathbf{x}_i$ are independent random vectors with common distribution G_0 such that there exists $E_{G_0}\mathbf{x}\mathbf{x}' = A^{-1}$. From F_0, G_0, and the independence between them we can obtain the distribution of $(\mathbf{x}_i, y_i)$ under the model (1), which we denote by H_0. To allow for the existence of a fraction ϵ of outliers, we will suppose that the distribution of the data lies in the contamination neighborhood

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

As it is well known, outliers may alter drastically the conclusions of a regression analysis performed with the usual minimum least squares estimate. In fact, just one outlier, strategically placed, may lead the estimates to take an arbitrary value. For this reason, several alternative robust regression estimates have been developed. First attempts in dealing with this problem can be found in Huber (1981) and Hampel *et al.* (1986). Other important references are Rousseeuw (1984), Rousseeuw and Yohai (1984), Yohai (1987), Yohai and Zamar (1988), Simpson, Rupert and Carroll (1992), Maronna and Yohai (1993), Coakley and Hettmansperger (1993), and Croux, Rousseeuw, and Hössjer (1994).

Most of robust regression estimates can be viewed as a functional, $\mathbf{T}$, with values in $\mathbb{R}^p$ defined on a large set of distribution functions H on $\mathbb{R}^{p+1}$, which includes the neighborhood V_ϵ. We will suppose that $\mathbf{T}$ is regression and affine equivariant, that is, if $y^* = y + \mathbf{x}'\boldsymbol{b}$, $\mathbf{x}^* = C'\mathbf{x}$ for some full rank $p \times p$ matrix C, and H^* is the distribution of $(\mathbf{x}^*, y^*)$, then $\mathbf{T}(H^*) = C^{-1}[\mathbf{T}(H) + \boldsymbol{b}\,]$.

It is important to measure quantitatively the degree of robustness of these functionals in order to compare them. One possible measure is the gross error sensitivity (γ_T^*) which assesses the maximum relative influence of adding individual observations to the sample (see Hampel, 1974). It is convenient to obtain bounded influence estimates, that is, estimates $\mathbf{T}$ such that $\gamma_T^* < \infty$.

Another way of looking at the robustness properties of the estimate is the breakdown point (ϵ^*) or limiting fraction of bad outliers that the estimate can tolerate (see also

Hampel, 1974). A fraction of outliers greater than ϵ^* may lead the estimate to a value totally determined by the outliers and therefore a goal in robust estimation must be to devise methods with high breakdown point.

The gross error sensitivity and the breakdown point can be viewed as part of a global measure of robustness: the maximum asymptotic bias curve (or maxbias curve), $B_{\mathbf{T}}(\epsilon)$, caused by a fraction ϵ of outliers. This function was introduced by Huber (1964) in the location model. Martin and Zamar (1989) and Martin, Yohai and Zamar (1989) computed $B_{\mathbf{T}}(\epsilon)$ for scale M-estimates and regression S-estimates respectively. 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 estimates, 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

$$B_{\mathbf{T}}(\epsilon) = \sup_{H \in V_\epsilon} b(\mathbf{T}, H) = \sup_{H \in V_\epsilon} \|\mathbf{T}(H)\|. \tag{2}$$

Under regularity conditions, $\gamma_T^* = B'_{\mathbf{T}}(0)$. It follows that $\gamma_T^* \epsilon$ is a linear approximation of $B_{\mathbf{T}}(\epsilon)$ for small values of ϵ. On the other hand, the breakdown point can be defined as $\epsilon^* = \inf\{\epsilon : B_{\mathbf{T}}(\epsilon) = \infty\}$. Therefore, ϵ^* contains another relevant feature of $B_{\mathbf{T}}(\epsilon)$, namely, the location of the point at which $B_{\mathbf{T}}(\epsilon)$ explodes.

As we have seen, $B_{\mathbf{T}}(\epsilon)$ sums up different approaches to quantitative robustness. From the knowledge of $B_{\mathbf{T}}(\epsilon)$ we can compute both the gross error sensitivity and the breakdown point of $\mathbf{T}$, and we possess a comprehensive description of the robustness properties of the estimate. Unfortunately, the function $B_{\mathbf{T}}(\epsilon)$ is sometimes difficult to compute. Maxbias curves are known for few robust regression estimates and only when the regressors are elliptically distributed. Moreover, procedures to obtain them are rather specialized.

This paper aims at proposing a general method for computing maxbias curves, valid for a broad class of robust regression estimates. We will consider the class of residual admissible regression estimates (for a precise definition, see Yohai and Zamar, 1993). Roughly speaking, this class consists of estimates for which the empirical distribution of the absolute value of the residuals cannot be uniformly improved by using any other set of regression coefficients. Yohai and Zamar (1993) show that many regression estimates defined as a function of the regression residuals belong to this class. In particular, we apply our results to S-, τ-, MM-, and signed R-estimates. Our results are also valid when the carriers are not elliptically distributed, so we also extend the existing results in this direction. Therefore, we provide a unified bias robustness theory for regression estimates based on residuals.

In Section 2, we give the basic result. Section 3 contains some applications of this result and gives the maxbias curves of several important regression estimates. In Section 4 we deal with the class of M-estimates with general scale which requires a slightly different approach. Proofs and technical details can be found in a final appendix.

2 Main result

Many interesting equivariant regression estimates are defined as the value of $\boldsymbol{\theta}$ that minimizes a functional at the empirical distribution of the absolute value of the residuals, $|y_i - \boldsymbol{\theta}'\mathbf{x}_i|$. If $F_{H,\boldsymbol{\theta}}$ is the distribution function of these absolute values when the distribution of the data $(\mathbf{x}_i, y_i)$ is H, the corresponding functional form of these estimates is

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

where $J(F)$ is defined on a set of distribution functions containing the empirical distribution functions and the neighborhood V_ϵ.

We will suppose that J satisfies the following assumption:

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

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

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

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

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

REMARK 1. Assumption 1(a) is a monotonicity condition which can be easily checked in the important examples that we present in Section 3. Assumption 1(b) of ϵ-monotonicity was introduced by Yohai and Zamar (1993). They show that the ϵ-monotonicity condition implies the residual admissibility of the corresponding estimate and hence it is a slightly stronger assumption. However, all interesting residual admissible estimates can be shown to be also ϵ-monotone and therefore there is not a relevant difference between assuming ϵ-monotonicity or residual admissibility in the applications of our result. Notice that if we take $F_n = F$ and $G_n = G$ for each n, then ϵ-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 also view ϵ-monotonicity as a strict monotonicity condition for certain especial distributions. Finally, Assumption 1(c) is a mild technical condition which can be easily verified in most of important examples.

We will also need the following assumption on the distributions F_0 and G_0.

Assumption 2 *F_0 has an even and strictly unimodal density f_0, and $P_{G_0}(\boldsymbol{\theta}'\mathbf{x} = 0) = 0$, for each $\boldsymbol{\theta} \neq \mathbf{0}$.*

Next, we give a general result useful to compute the maxbias curve (see equation (2)) of any class of robust regression estimates based on residuals.

Theorem 1 *Let $\mathbf{T}$ be a regression estimate defined as in equation (3), let $t^* \in \mathbb{R}$ be such that*

$$m(t^*) \equiv \inf_{\|\boldsymbol{\theta}\|=t^*} J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0] = \lim_{n\to\infty} J[(1-\epsilon)F_{H_0,0} + \epsilon U_n], \tag{6}$$

where U_n is the uniform distribution function on $[n-(1/n), n+(1/n)]$. Then, under Assumptions 1 and 2, $B_{\mathbf{T}}(\epsilon) = t^*$.

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 $\|\boldsymbol{\theta}\|$. We can drop the infimum in equation (6) as each direction gives the same value of $J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}}+\epsilon\delta_0]$. We obtain the following result.

Corollary 1 *Suppose that G_0 is spherical and that there exists $\boldsymbol{\theta}^* \in \mathbb{R}^p$ such that*

$$J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^*}+\epsilon\delta_0] = \lim_{n\to\infty} J[(1-\epsilon)F_{H_0,0}+\epsilon U_n], \tag{7}$$

where U_n is the uniform distribution function on $[n-(1/n), n+(1/n)]$. Then, under Assumptions 1 and 2, $B_{\mathbf{T}}(\epsilon) = \|\boldsymbol{\theta}^*\|$.

REMARK 2. Let δ_n be the distribution that assigns probability one to the point n. In most of examples, the following equation holds:

$$\lim_{n\to\infty} J[(1-\epsilon)F_{H_0,0}+\epsilon U_n] = \lim_{n\to\infty} J[(1-\epsilon)F_{H_0,0}+\epsilon\delta_n] \equiv J[(1-\epsilon)F_{H_0,0}+\epsilon\delta_\infty].$$

Therefore, equality (7) turns out to be

$$J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^*}+\epsilon\delta_0] = J[(1-\epsilon)F_{H_0,0}+\epsilon\delta_\infty].$$

We can use this expression to give an intuitive interpretation of Corollary 1. Suppose that there is a proportion ϵ of outliers at the point $(\boldsymbol{\theta}^{*\prime}\mathbf{x},\mathbf{x})$. Then, $(1-\epsilon)F_{H_0,\boldsymbol{\theta}^*}+\epsilon\delta_0$ is the distribution of the residuals when these outliers are perfectly fitted and hence $J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^*}+\epsilon\delta_0]$ is the value of the target functional in this case. On the other hand, $J[(1-\epsilon)F_{H_0,0}+\epsilon\delta_\infty]$ is the value of the target functional when the outliers are completely ignored and $|\boldsymbol{\theta}^{*\prime}\mathbf{x}| \to \infty$. In Corollary 1 we show that the maximum bias is the value of $\|\boldsymbol{\theta}^*\|$ such that the value of the target functional that we obtain by fitting the outliers perfectly is the same as the value of the target functional that we obtain by ignoring them completely.

3 Maxbias curves of several residual admissible regression estimates

In this section, we give some examples to illustrate the wide applicability of Theorem 1 and Corollary 1. First, we consider the well-known Rousseeuw's (1984) LMS-estimate and derive the maxbias curves for several distributions of the regressors. The purpose of this study is to find out which is the effect in the maxbias curve of deviations from ellipticity in the distribution of the regressors. In the rest of the section we assume that the regressors are spherical and compute the maxbias curves of S-estimates, τ-estimates, and some signed R-estimates including Rousseeuw's LTS- and LTAV-estimates. So far, only the maxbias curves for S-estimates were known.

3.1 LMS-ESTIMATE MAXBIAS CURVE WITH NON-ELLIPTICAL REGRESSORS

Assume that the distribution of the errors, F_0, is standard normal. We consider the linear model (1) with two regressors X_1 and X_2 and the LMS-estimate, that is, the functional defined as

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} F^{-1}_{H,\boldsymbol{\theta}}(1/2).$$

In Table 1 we report the maxbias curves of this estimate when the distributions of the regressors are:

(a) The vector (X_1, X_2) is distributed as a bivariate normal with mean vector $\mathbf{0}$ and covariance matrix I. This situation may be considered as a benchmark to interpret the other results.

(b) For $i = 1, 2$, X_i follows a standardized t-distribution with 3 degrees of freedom and both regressors are independent. This case points at the effect of regressors whose distributions have heavy tails.

(c) For $i = 1, 2$, X_i is distributed as a standardized chi-square with 4 degrees of freedom. Assymetric regressors are considered in this case.

In Table 1 we observe that regressors (b) and (c) give a larger maxbias curve than the elliptical case. Moreover, asymmetric regressors of item (c) yield a better behavior of the maxbias curve than heavy tailed regressors of case (b).

3.2 S-ESTIMATES

For the sake of simplicity we suppose in the rest of this section that F_0 is the standard normal distribution and G_0 is the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix the identity I. We will apply Corollary 1 to different sorts of robust regression estimates.

First, consider an S-estimate, $\mathbf{T}$, based on a function χ. Then,

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

where $S(F_{H,\boldsymbol{\theta}})$ is a scale M-estimate of the absolute value of the residuals, that is, the functional $S(F)$ is defined as the solution of the equation

$$E_F \chi\left(\frac{u}{S(F)}\right) = b. \tag{8}$$

Suppose that the function χ satisfies the following Assumption:

Assumption 3 *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.*

It is not difficult to check Assumptions 1(a) and 1(c). Yohai and Zamar (1993), Lemma 5.1, show that S-estimates with χ functions satisfying Assumption 3 are ϵ-monotone. Therefore we can obtain the maxbias curve applying Corollary 1.

It is easy to show that, for S-estimates, (see Remark 2)

$$\lim_{n\to\infty} S[(1-\epsilon)F_{H_0,0} + \epsilon U_n] = \lim_{n\to\infty} S[(1-\epsilon)F_{H_0,0} + \epsilon\delta_n].$$

To illustrate the use of Corollary 1, define $g(s) = E_\Phi \chi(Y/s)$, where Φ denotes the standard normal distribution. Let $S_1 = S[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^*} + \epsilon\delta_0]$. Since the distribution of $y - \boldsymbol{\theta}^{*\prime}\mathbf{x}$ is normal with mean 0 and variance $1 + \|\boldsymbol{\theta}^*\|^2$,

$$(1-\epsilon)E_{H_0}\chi\left(\frac{y - \boldsymbol{\theta}^{*\prime}\mathbf{x}}{S_1}\right) = (1-\epsilon)g\left(\frac{S_1}{(1+\|\boldsymbol{\theta}^*\|^2)^{1/2}}\right) = b.$$

Therefore

$$S_1 = (1+\|\boldsymbol{\theta}^*\|^2)^{1/2} g^{-1}\left(\frac{b}{1-\epsilon}\right). \tag{9}$$

Let $S_2 = \lim_{n\to\infty} S[(1-\epsilon)F_{H_0,0} + \epsilon\delta_n]$. Therefore, S_2 must satisfy

$$(1-\epsilon)E_{H_0}\chi(Y/S_2) + \epsilon = (1-\epsilon)g(S_2) + \epsilon = b,$$

and it follows that

$$S_2 = g^{-1}\left(\frac{b-\epsilon}{1-\epsilon}\right). \tag{10}$$

To apply Corollary 1, we compute the value of $\|\boldsymbol{\theta}^*\|$ resulting from the equation $S_1 = S_2$ (see equation (7)). Using expressions (9) and (10), we have

$$B_S^2(\epsilon) = \|\boldsymbol{\theta}^*\|^2 = \left[\frac{g^{-1}\left(\frac{b-\epsilon}{1-\epsilon}\right)}{g^{-1}\left(\frac{b}{1-\epsilon}\right)}\right]^2 - 1. \tag{11}$$

Notice that this expression coincides with formula (3.18) in Martin *et al.* (1989).

3.3 τ-ESTIMATES

Let $S(F)$ be an S-estimate based on a function χ_1 as defined in equation (8). Yohai and Zamar (1988) defined the class of regression τ-estimates as $\tau(H) = \arg\min_{\boldsymbol{\theta}} \tau(F_{H,\boldsymbol{\theta}})$, where

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

The idea is to minimize an efficient and robust scale estimate, which is an M-estimate modified by the factor $E_F\chi_2[u/S(F)]$, to reach simultaneously high efficiency and high breakdown point.

Suppose that both χ_1 and χ_2 satisfy Assumption 3 and that χ_2 satisfies

Assumption 4 *χ_2 is differentiable and $2\chi_2(u) - \chi_2'(u)u \geq 0$, for every u.*

Under this hypothesis, it can be shown that the functional $\tau(F)$ satisfies Assumption 1 and therefore Corollary 1 can be applied (see Yohai and Zamar, 1993). As in the case of S-estimates we also have

$$\lim_{n\to\infty} \tau[(1-\epsilon)F_{H_0,0} + \epsilon U_n] = \lim_{n\to\infty} \tau[(1-\epsilon)F_{H_0,0} + \epsilon\delta_n].$$

Define $\tau_1 = \tau^2[(1-\epsilon)F_{H_0,\boldsymbol{\theta}^*} + \epsilon\delta_0]$, $\tau_2 = \lim_{n\to\infty} \tau^2[(1-\epsilon)F_{H_0,0} + \epsilon\delta_n]$, and $g_i(s) = E_\Phi\chi_i(Y/s)$, for $i = 1, 2$. Some tedious manipulations, similar to those for S-estimates, give

$$\tau_1 = (1 + \|\boldsymbol{\theta}^*\|^2)\left[g_1^{-1}\left(\frac{b}{1-\epsilon}\right)\right]^2 (1-\epsilon) g_2\left[g_1^{-1}\left(\frac{b}{1-\epsilon}\right)\right],$$

and

$$\tau_2 = \left[g_1^{-1}\left(\frac{b-\epsilon}{1-\epsilon}\right)\right]^2 \left[(1-\epsilon) g_2\left[g_1^{-1}\left(\frac{b}{1-\epsilon}\right)\right] + \epsilon\right].$$

We impose $\tau_1 = \tau_2$ to obtain

$$B_\tau^2(\epsilon) = \left[\frac{g_1^{-1}\left(\frac{b-\epsilon}{1-\epsilon}\right)}{g_1^{-1}\left(\frac{b}{1-\epsilon}\right)}\right]^2 \left[\frac{g_2\left[g_1^{-1}\left(\frac{b-\epsilon}{1-\epsilon}\right)\right]}{g_2\left[g_1^{-1}\left(\frac{b}{1-\epsilon}\right)\right]} + \frac{\epsilon}{1-\epsilon}\frac{1}{g_2\left[g_1^{-1}\left(\frac{b}{1-\epsilon}\right)\right]}\right] - 1. \tag{12}$$

REMARK 3. Define

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

From (11) and (12),

$$1 + B_\tau^2(\epsilon) = [1 + B_S^2(\epsilon)]H(\epsilon). \tag{13}$$

This is the relationship between the bias of the S-estimate based on χ_1 and the bias of the τ-estimate based on χ_1 and χ_2. Since $H(\epsilon)$ is bounded for $\epsilon < \min\{b, 1-b\}$, the breakdown point of the τ-estimate is $\epsilon^* = \min\{b, 1-b\}$, equal to that of the S-estimate corresponding to χ_1. Therefore, to determine the breakdown point of the regression τ-estimates, the function χ_2 is irrelevant and we can choose it appropriately to attain a high efficiency (see details in Yohai and Zamar, 1988).

In Table 2 some selected values for the maxbias curves of several S- and τ-estimates are displayed. First, we have considered an S-estimate based on a bisquare Tukey function. The tuning constant has been fitted to attain a breakdown point of 0.5. We have combined this S-estimate with another bisquare Tukey function to obtain a τ-estimate with breakdown point of 0.5 and efficiency of 95%. The maximum bias of the S-estimate is considerably better although its efficiency is very low (we recall that Hössjer (1992) has shown that the efficiency of an S-estimate with breakdown point 0.5 is at most 35%).

(Table 2 about here)

(Figure 1 about here)

It is more interesting to compare the bias performance of the Tukey τ-estimate with another τ-estimate. With this purpose we have considered the τ-estimate built by taking Rousseeuw's LMS-estimate as robust S-estimate and a bisquare function as χ_2. (LMS $\tau-$estimate). Since the bias of the LMS-estimate is better than that of Tukey S-estimate, we obtain a lower bias for the second τ-estimate. Both bias curves have been plotted in Figure 1.

3.4 R-ESTIMATES

Hössjer (1994) defined the signed R-estimates which are based on the following estimating functional $T(H) = \arg\min_{\boldsymbol{\theta}} J(F_{H,\boldsymbol{\theta}})$, where

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

These estimates consist in choosing the vector of parameters that minimizes a weighted average of the absolute values of the residuals. The weights are given by a function $a(u)$ of the signed ranks of these residuals. Hössjer (1994) observed that if $a(u)$ vanishes outside the interval $[0, 1-\alpha]$, these estimates have breakdown point $\epsilon^* = \min\{\alpha, 1-\alpha\}$.

An interesting particular case is the α-least trimmed absolute value (α-LTAV) estimate which is defined by taking

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

Observe that Rousseeuw's least trimmed squares estimate is also defined with the function $a(u)$ given in (15) but replacing the functional (14) with

$$J(F) = \int_0^\infty a[F(u)]u^2 dF(u).$$

Yohai and Zamar (1993) defined the general class of functionals

$$J(F) = \int_0^\infty a[F(u)]u^k dF(u), \tag{16}$$

which includes all the estimates defined in this subsection.

Suppose that $a(u)$ satisfies the following assumption,

Assumption 5 *(a)* $a(u)$ *is continuous on* $[0, 1-\alpha]$. *(b)* $a(u) = 0$ *if* $1-\alpha < u \leq 1$. *(c)* $a(u) > 0$ *if* $0 < u < 1-\alpha$.

Yohai and Zamar (1993), Theorem 5.2, proved that under Assumption 5 the functionals defined in equation (16) are ϵ-monotone. Assumptions 1(a) and 1(c) are a consequence of Lemma A.4. in the same work. Therefore we can apply Corollary 1 to obtain the maxbias curve of these estimates.

Let us introduce the following notation:

$$\Delta_1 = \Phi^{-1}\left[1 - \frac{\alpha - \epsilon}{2(1-\epsilon)}\right],\ \Delta_2 = \Phi^{-1}\left[1 - \frac{\alpha}{2(1-\epsilon)}\right], \text{ and}$$

$$l(u) = (1-\epsilon)F_{H_0,0}(u) = (1-\epsilon)[2\Phi(u) - 1].$$

Next, we apply Corollary 1 and use Lemmas 2 and 3 in the appendix to obtain the maxbias curves of the estimates defined in (16) for $\epsilon < \min\{\alpha, 1-\alpha\}$:

$$B^2(\epsilon) = \left[\frac{\int_0^{\Delta_1} a[l(u)]u^k\varphi(u)du}{\int_0^{\Delta_2} a[l(u)+\epsilon]u^k\varphi(u)du}\right]^{2/k} - 1. \tag{17}$$

It is interesting to particularize this formula to the following two especial cases:

1. α-LTAV ESTIMATES. In this case $k = 1$ and $a(u)$ is defined as in the equation (15). It follows from (17) that

$$B_\alpha^2(\epsilon) = \left[\frac{\int_0^{\Delta_1} u\varphi(u)du}{\int_0^{\Delta_2} u\varphi(u)du}\right]^2 - 1 = \left[\frac{\varphi(0) - \varphi(\Delta_1)}{\varphi(0) - \varphi(\Delta_2)}\right]^2 - 1.$$

2. α-LTS ESTIMATES. In this case $k = 2$ and $a(u)$ is defined as in the equation (15). It follows from (17) that

$$B_\alpha^2(\epsilon) = \frac{\int_0^{\Delta_1} u^2\varphi(u)du}{\int_0^{\Delta_2} u^2\varphi(u)du} - 1 = \frac{\Phi(\Delta_1) - 0.5 - \Delta_1\varphi(\Delta_1)}{\Phi(\Delta_2) - 0.5 - \Delta_2\varphi(\Delta_2)} - 1.$$

In Table 2, several values are displayed of the maxbias curves of these two particular cases when $\alpha = 0.5$. Observe that 0.5-LTS performs better than 0.5-LTAV, especially for large amounts of contamination. In general, if we consider the sequence of estimates defined by $a(u)$ as in the equation (15) and $\alpha = 0.5$ when k ranges over the positive integers, it is possible to check numerically that as k increases the bias performance is better uniformly in ϵ. This is not surprising since as $k \to \infty$, these estimates approach the LMS-estimate. Also notice that for small values of ϵ, the bias curves of LTS and LTAV are better than those for τ-estimates. However, we recall that, in general, R-estimates with high breakdown point are not efficient (see Hössjer, 1994). In Figure 2, we plot the maxbias curves of 0.5-LTS and 0.5-LTAV estimates.

(Figure 2 about here)

4 Maxbias curve for M-estimates with general scale

M-estimates of regression with general scale are an important example of regression admissible estimates. Martin *et al.* (1989) defined them as

$$\mathbf{T}(H) = \arg\min_{\boldsymbol{\theta}} E_H \rho \left(\frac{y - \boldsymbol{\theta}'\mathbf{x}}{S(H)} \right), \tag{18}$$

where $S(H)$ is a bias robust estimate of the scale of the residuals and ρ satisfies Assumption 3. Yohai's (1987) MM-estimates are within this class.

Unfortunately, we cannot apply Theorem 1 to compute the maximum bias of M-estimates with general scale because it depends on the bias properties of the scale functional, $S(H)$, which we did not consider in the result of Section 2. Despite of this fact, we obtain in this section an upper bound and a lower one for the maximum bias of M-estimates with general scale. We also show that in some important cases and small values of ϵ, both bounds are identical so that we can give the exact value of the maxbias curve in those particular cases.

Introduce some notation: $s_1 = \inf_{H \in V_\epsilon} S(H)$, $s_2 = \sup_{H \in V_\epsilon} S(H)$, and

$$h(t,s) = \inf_{\|\boldsymbol{\theta}\|=t} E_{H_0} \rho \left(\frac{y - \boldsymbol{\theta}'\mathbf{x}}{s} \right) - E_{H_0} \rho \left(\frac{y}{s} \right).$$

The following two functions play an important role in our bounds:

$$h_1(t) = h(t, s_2), \text{ and } h_2(t) = \inf_{s_1 \le s \le s_2} h(t,s).$$

We will consider as scale estimate, the scale of the residuals obtained from a bias robust S-estimate of regression. That is, we will assume

Assumption 6 χ *is a function satisfying Assumption 3 and* $S(H) = \min_{\boldsymbol{\theta}} S(F_{H,\boldsymbol{\theta}})$, *where* $S(F_{H,\boldsymbol{\theta}})$ *is defined as in equation (8) with* $b = b_\chi$. *The maxbias curve of the corresponding regression S-estimate,* $B_\chi(\epsilon)$, *is such that* $B_\chi(\epsilon) < h_1^{-1}[\epsilon/(1-\epsilon)]$.

Since it is important for $S(H)$ to be bias robust, it is convenient to use jump χ functions due to their minimax-bias properties (see Martin *et al.*, 1989).

Theorem 2 *Let* T *be an M-estimate of regression with general scale as defined in (18), where* ρ *satisfies Assumption 3 and* $S(H)$ *satisfies Assumption 6. Then, under Assumption 2,*

$$h_1^{-1}\left(\frac{\epsilon}{1-\epsilon}\right) \leq B_{\mathbf{T}}(\epsilon) \leq h_2^{-1}\left(\frac{\epsilon}{1-\epsilon}\right). \tag{19}$$

The exact value of the maxbias curve can be found by proving that, for some values of ϵ, the bounds given in Theorem 2 coincide. We will give a sufficient analytical condition for this fact when the regressors are spherically distributed and will check it numerically for some important examples. This fact will prove that our bounds are sharp. Also notice that s_2 is easy to compute and hence we can easily obtain the function h_1. However, computing the function h_2 involves an optimization problem that may be very difficult to solve in general. If both bounds are identical, we will only have to handle the simpler function h_1.

When the regressors are spherically distributed, we can drop the infimum in the definition of $h(t,s)$ and hence we can write $h(\boldsymbol{\theta},s)$, $h_1(\boldsymbol{\theta})$ and $h_2(\boldsymbol{\theta})$. Let $\boldsymbol{\theta}^* = h_1^{-1}[\epsilon/(1-\epsilon)]$. If the function $h(\boldsymbol{\theta}^*,s)$ is decreasing for $s_1 < s < s_2$, then $h_1(\boldsymbol{\theta}^*) = h_2(\boldsymbol{\theta}^*)$. In this case both, the lower bound and the upper one, would take the same value. Therefore, a sufficient condition for this to happen is

$$\frac{\partial h(\boldsymbol{\theta}^*,s)}{\partial s} < 0, \quad \text{for each} \quad s \in (s_1,s_2). \tag{20}$$

From now on, suppose normality and sphericity as in Section 3. Let $\psi = \rho'$ and $\gamma = (1+\|\boldsymbol{\theta}^*\|^2)^{1/2}$. Assuming that we can derive under the integral sign, the condition (20) is equivalent to

$$G(\gamma,s) = E_\Phi[(\gamma Z/s)\psi(\gamma Z/s)] - E_\Phi[(Z/s)\psi(Z/s)] > 0 \quad \text{for each} \quad s \in (s_1,s_2) \tag{21}$$

Since $\gamma \geq 1$ and $G(1,s) = 0$, a sufficient condition for (21) is that $G(r,s)$ is increasing in r for $s \in (s_1,s_2)$ and $1 \leq r \leq \gamma$. Therefore, a new sufficient condition is

$$\begin{aligned} H(r/s) &= r\frac{\partial G(r,s)}{\partial r} \\ &= E_\Phi[(r/s)Z\psi(rZ/s)] + E_\Phi[(r^2Z^2/s^2)\psi'(rZ/s)] \\ &> 0, \quad \text{for each} \quad s \in (s_1,s_2), \quad 1 \leq r \leq \gamma. \end{aligned} \tag{22}$$

In Figure 3, we have plotted the function $H(r/s)$ corresponding to the case when ρ is the Tukey bisquare function. We can see that there is a range of values for r/s such that the function is positive. For $r/s = 2.45$, we have $H(r/s) = 0$, therefore $H(r/s) > 0$ for each $r/s < 2.45$.

(Figure 3 about here)

To verify (22), we can compute $\boldsymbol{\theta}^*$ and s_1 for each fixed amount of contamination ϵ. Since for each $s \in (s_1, s_2)$, we have $\gamma/s_1 > r/s$, then $H(\gamma/s_1) > 0$ implies $H(r/s) > 0$ for each $s \in (s_1, s_2)$ and $1 \leq r \leq \gamma$, and the sufficient condition (22) is satisfied.

In the example of the Tukey bisquare function (where the tuning constant c has been fitted to attain an efficiency of 95%), we report the results in Table 3. We have considered two robust scale S-estimates: the S-estimate based on a jump function and the one based on a Tukey bisquare function where the tuning constant has been fitted to attain breakdown point 0.5. We call these estimates $\mathbf{T}_1$ and $\mathbf{T}_2$ respectively. For $\epsilon < .15$, it can be numerically checked that the condition (22) is satisfied and hence the lower bound and the upper one are identical both for $\mathbf{T}_1$ and $\mathbf{T}_2$. The plot of the maxbias curves is displayed in Figure 4.

(Table 3 about here)

(Figure 4 about here)

5 Further remarks

We finish our work by pointing out some related open problems and further applications of the ideas we have introduced. Croux *et al.* (1994) have proposed to minimize a scale M-estimate of the residual differences $\{|(y_i - \boldsymbol{\theta}'\mathbf{x}_i) - (y_j - \boldsymbol{\theta}'\mathbf{x}_j)| : i < j\}$ instead of a scale M-estimate of the absolute value of the residuals, defining in this way a GS-estimate. Although Assumption 1 does not hold for GS-estimates and we cannot apply Theorem 1, it is possible to use some of the ideas introduced in this paper to avoid the sphericity assumption in the maxbias curve computations. If we redefine $\tilde{h}(\epsilon, s, \|\boldsymbol{\beta}\|)$, see equation (14) in the quoted work, to be

$$\tilde{h}(\epsilon, s, t) = \inf_{\|\boldsymbol{\beta}\|=t} [(1-\epsilon)^2 g(s, \|\boldsymbol{\beta}\|) + 2\epsilon(1-\epsilon)\tilde{g}(s, \|\boldsymbol{\beta}\|)],$$

it is possible to show that Theorem 4 in Croux *et al.* (1994) is valid even for nonspherical distributions. However, we must still assume that the distribution of $\boldsymbol{\theta}'\mathbf{x}$ is symmetric and unimodal for each $\boldsymbol{\theta} \neq 0$. The key point is that in the proof of Theorem 4 we must appropriately choose the direction of the contaminations which is no longer indifferent.

Our method to compute maxbias curves cannot be applied when the residuals are weighted to penalize high leverage observations. Such is the case of, for example, GM-estimates. Only Martin *et al.* (1989) have found out an expression for the maxbias curve of these estimates but assuming sphericity and known regressors covariance matrix. Therefore, the general problem of computing the maxbias curves of the estimates in this class still remains open.

Appendix. Proofs

The following lemma is needed to prove Theorem 1.

Lemma 1 *Define*

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

Then, under Assumptions 1(b) and 2,

(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:

From Assumption 1(b), it is obvious that if F_n and F are distribution functions on $[0,\infty)$, continuous on $(0,\infty)$, such that $F_n(u) \to F(u)$ for each $u > 0$, then

$$\lim_{n\to\infty} J(F_n) = J(F).$$

Suppose that $\{\boldsymbol{\theta}_n\}$ is a sequence such that $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$. Define $F_n = (1-\epsilon)F_{H_0,\boldsymbol{\theta}_n} + \epsilon\delta_0$. It follows that $\lim_{n\to\infty} J(F_n) = J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0]$. That is, the function $f(\boldsymbol{\theta}) = J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}} + \epsilon\delta_0]$ is continuous. Since $\{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| = t\}$ is a compact set for each t, the infimum in the definition of $m(t)$ is actually a minimum. This proves part (a).

Lemma A.1 in Yohai and Zamar (1993) shows that, under Assumption 2, $F_{H_0,\lambda\boldsymbol{\theta}}(u)$ is strictly decreasing as a function of λ. Let t_1 and t_2 be such that $t_1 > t_2$. Applying part (a), there exists $\boldsymbol{\theta}_1$ such that $m(t_1) = J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_1} + \epsilon\delta_0]$. Since $F_{H_0,\boldsymbol{\theta}_1}(u) < F_{H_0,(t_2/t_1)\boldsymbol{\theta}_1}(u)$, it follows from Assumption 1(b) that

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

But, by definition of $m(t)$,

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

The last two inequalities prove part (b). □

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 not any $H \in V_\epsilon$ such that $\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J(F_{H,\boldsymbol{\theta}})$. We will actually show that for every $H \in V_\epsilon$, $J(F_{H,\tilde{\boldsymbol{\theta}}}) > J(F_{H,0})$.

It is clear that for each $H \in V_\epsilon$, and $u > 0$,

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

Inequality (23), Assumption 1(a), the definition of the function $m(t)$, and Lemma 1(b) imply that, for each $H \in V_\epsilon$,

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

Equation (6) and Assumption 1(c) imply

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

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

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

By Lemma 1(a), there exists $\boldsymbol{\theta}_t$ such that $m(t) = J[(1-\epsilon)F_{H_0,\boldsymbol{\theta}_t} + \epsilon\delta_0]$. 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 $[nt^2 - (1/n), nt^2 + (1/n)]$. If F_n is the uniform distribution function on $[-1/n, 1/n]$, then for each $\boldsymbol{\beta} \in \mathbb{R}^p$ and $u > 0$,

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

Let $H_n = (1-\epsilon)H_0 + \epsilon\tilde{H}_n$. 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

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

Notice that $0 \leq |\boldsymbol{\theta}_t'\tilde{\boldsymbol{\theta}}| \leq \|\boldsymbol{\theta}_t\|\|\tilde{\boldsymbol{\theta}}\| = t\tilde{t} < t^2$. Hence, $|t^2 - |\boldsymbol{\theta}'\tilde{\boldsymbol{\theta}}|| > 0$. It follows from (26) that

$$\lim_{n\to\infty} F_{\tilde{H}_n, \mathbf{T}_n}(u) = 0, \quad \text{for } u > 0. \tag{27}$$

On the other hand, $t^2 - \boldsymbol{\theta}_t'\boldsymbol{\theta}_t = 0$. From (26),

$$\lim_{n\to\infty} F_{\tilde{H}_n, \boldsymbol{\theta}_t}(u) = 1, \quad \text{for } u > 0. \tag{28}$$

From (27) and Lemma A.1 in Yohai and Zamar (1993), we have that for each $u > 0$,

$$\begin{aligned} \lim_{n\to\infty} F_{H_n, \mathbf{T}_n}(u) &= (1-\epsilon)F_{H_0, \tilde{\boldsymbol{\theta}}}(u) \\ &\leq (1-\epsilon)F_{H_0,0}(u) \\ &= \lim_{n\to\infty}[(1-\epsilon)F_{H_0,0}(u) + \epsilon U_n(u)]. \end{aligned} \tag{29}$$

Applying Assumption 1(b) and inequality (29),

$$\lim_{n\to\infty} J(F_{H_n, \mathbf{T}_n}) \geq \lim_{n\to\infty} J[(1-\epsilon)F_{H_0,0} + \epsilon U_n] = m(t^*). \tag{30}$$

From (28), we have that for each $u > 0$,

$$\lim_{n\to\infty} F_{H_n, \boldsymbol{\theta}_t}(u) = (1-\epsilon)F_{H_0, \boldsymbol{\theta}_t}(u) + \epsilon\delta_0(u), \quad \text{for } u > 0. \tag{31}$$

From Assumption 1(b), equation (31), and Lemma 1(b),

$$\lim_{n\to\infty} J[F_{H_n, \boldsymbol{\theta}_t}] = J[(1-\epsilon)F_{H_0, \boldsymbol{\theta}_t} + \epsilon\delta_0] = m(t) < m(t^*). \tag{32}$$

Applying (30) and (32), we have that for n 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. □

The following two lemmas are needed to get the expression (17).

Lemma 2 *Let* $F_n = (1-\epsilon)F_{H_0,0} + \epsilon U_n$. *Under Assumption 4 and for* $\epsilon < \alpha$,

$$\lim_{n\to\infty} J(F_n) = 2(1-\epsilon)\int_0^{\Delta_1} a[l(u)]u^k\varphi(u)du, \tag{33}$$

where J *is the functional defined in (16).*

Proof of Lemma 2:

Observe that $F_n(u) \dot{\to} l(u)$ for each $u > 0$. We just have to apply Lemma A.4(b) in Yohai and Zamar (1993) to the sequence $\{F_n\}$. □

The proof of the following lemma is a simple calculation and it will be omitted.

Lemma 3 *Let* $F_{\boldsymbol{\theta}^*} = (1-\epsilon)F_{H_0,\boldsymbol{\theta}^*} + \epsilon\delta_0$. *Under Assumption 4 and for* $\epsilon < \min\{\alpha, 1-\alpha\}$,

$$J_\alpha(F_{\boldsymbol{\theta}^*}) = 2(1-\epsilon)[1+\|\boldsymbol{\theta}^*\|^2]^{k/2}\int_0^{\Delta_2} a[l(u)+\epsilon]u^k\varphi(u)du, \tag{34}$$

where J *is the functional defined in (16).*

Remark 4 To obtain the expression (17) we just have to solve for $\|\boldsymbol{\theta}^*\|$ the equation (see the notation of Lemmas 2 and 3)

$$J(F_{\boldsymbol{\theta}^*}) = \lim_{n\to\infty} J(F_n).$$

Proof of Theorem 2:

Since for every $s > 0$ the functional $J(F) = E_F\rho(y/s)$ satisfies Assumption 1(b), we can prove, following the lines of the proof of Lemma 1, that for every $s > 0$, $h(t,s)$ is a strictly increasing function of t. Moreover, for each $t > 0$, there exists $\boldsymbol{\theta}_t \in \mathbb{R}^p$ such that

$$h_1(t) = E_{H_0}\rho\left(\frac{y-\boldsymbol{\theta}_t'\mathbf{x}}{s_2}\right) - E_{H_0}\rho\left(\frac{y}{s}\right).$$

It follows that $h_1(t)$ is also strictly increasing.

We show first that $B_{\mathbf{T}}(\epsilon) \leq t_2$ where t_2 is such that $h_2(t_2) = \epsilon/(1-\epsilon)$. Let $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ be such that $\tilde{t} = \|\tilde{\boldsymbol{\theta}}\| > t_2$. We shall prove that

$$E_H \rho\left(\frac{y - \tilde{\boldsymbol{\theta}}'\mathbf{x}}{S(H)}\right) > E_H \rho\left(\frac{y}{S(H)}\right), \text{ for each } H \in V_\epsilon. \tag{35}$$

Let $H = (1-\epsilon)H_0 + \epsilon\tilde{H}$, the following inequalities hold:

$$h[\tilde{t}, S(H)] > h[t_2, S(H)] \geq \inf_{s_1 \leq s \leq s_2} h(t_2, s) = h_2(t_2) = \frac{\epsilon}{1-\epsilon}.$$

Therefore, for each $H \in V_\epsilon$,

$$E_{H_0} \rho\left(\frac{y - \tilde{\boldsymbol{\theta}}'\mathbf{x}}{S(H)}\right) - E_{H_0} \rho\left(\frac{y}{S(H)}\right) > \frac{\epsilon}{1-\epsilon}, \quad \text{that is,}$$

$$(1-\epsilon)E_{H_0} \rho\left(\frac{y - \tilde{\boldsymbol{\theta}}'\mathbf{x}}{S(H)}\right) > (1-\epsilon)E_{H_0} \rho\left(\frac{y}{S(H)}\right) + \epsilon.$$

It follows that, for every $H \in V_\epsilon$,

$$\begin{aligned} E_H \rho\left(\frac{y - \tilde{\boldsymbol{\theta}}'\mathbf{x}}{S(H)}\right) &\geq (1-\epsilon)E_{H_0} \rho\left(\frac{y - \tilde{\boldsymbol{\theta}}'\mathbf{x}}{S(H)}\right) \\ &> (1-\epsilon)E_{H_0} \rho\left(\frac{y}{S(H)}\right) + \epsilon \geq E_H \rho\left(\frac{y}{S(H)}\right), \end{aligned}$$

that is, inequality (35) holds.

Next, we show that $B_{\mathbf{T}}(\epsilon) \geq t_1$, where t_1 is such that $h_1(t_1) = \epsilon/(1-\epsilon)$. Let $t > 0$ be such that $B_{\chi}(\epsilon) < t < t_1$, where $B_{\chi}(\epsilon) < t_1$ by Assumption 6. It is enough to show that $B_{\mathbf{T}}(\epsilon) \geq t$.

There exists $\boldsymbol{\theta}_t \in \mathbb{R}^p$ such that

$$h_1(t) = h(t, s_2) = E_{H_0} \rho\left(\frac{y - \boldsymbol{\theta}_t'\mathbf{x}}{s_2}\right) - E_{H_0} \rho\left(\frac{y}{s}\right).$$

Observe that, as h_1 is strictly increasing, $h_1(t) < h_1(t_1)$. It follows that

$$(1-\epsilon)E_{H_0} \rho\left(\frac{y - \boldsymbol{\theta}_t'\mathbf{x}}{s_2}\right) < (1-\epsilon)E_{H_0} \rho\left(\frac{y}{s_2}\right) + \epsilon. \tag{36}$$

Define the following sequence of contamination distributions: $\tilde{H}_n = \delta_{(y_n, \mathbf{x}_n)}$ where $\mathbf{x}_n = n\boldsymbol{\theta}_t$ and $y_n = \mathbf{x}_n'\boldsymbol{\theta}_t = nt^2$. Let $H_n = (1-\epsilon)H_0 + \epsilon\tilde{H}_n$. Suppose that $\sup_n \|\mathbf{T}(H_n)\| < t$ in order to find a contradiction. Under this assumption, a convergent subsequence, denoted by $\{\mathbf{T}(H_n)\}$, exists such that

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

We have that

$$\lim_{n\to\infty} \left| \frac{y_n - \mathbf{T}_n'\mathbf{x}_n}{S(H_n)} \right| = \infty, \quad \text{and} \quad \left| \frac{y_n - \boldsymbol{\theta}_t'\mathbf{x}_n}{S(H_n)} \right| = 0, \quad \text{for each } n. \tag{37}$$

Next, we prove that $\lim_{n\to\infty} S(H_n) = s_2$. From equation (10),

$$(1-\epsilon)E_{H_0}\chi(y/s_2) + \epsilon = b_\chi.$$

Therefore, if the S-estimate of regression, $T_\chi(H)$, satisfies $\boldsymbol{\beta} = \lim_{n\to\infty} T_\chi(H_n) = 0$, then $\lim_{n\to\infty} S(H_n) = s_2$. Since $\|\boldsymbol{\beta}\| \le B_\chi(\epsilon) < b$, then $\lim_{n\to\infty} |y_n - \boldsymbol{\beta}'\mathbf{x}_n| = \infty$. If $l = \lim_{n\to\infty} S(H_n)$, then we obtain that

$$(1-\epsilon)E_{H_0}\chi\left(\frac{y - \boldsymbol{\beta}'\mathbf{x}}{l}\right) + \epsilon = b_\chi. \tag{38}$$

Suppose that $\|\boldsymbol{\beta}\| > 0$. Then,

$$(1-\epsilon)E_{H_0}\chi\left(\frac{y - \boldsymbol{\beta}'\mathbf{x}}{s_2}\right) + \epsilon > (1-\epsilon)E_{H_0}\chi(Y/s_2) + \epsilon = b_\chi. \tag{39}$$

By (38) and (39), $l > s_2$. This is a contradiction since $s_2 = \sup_{H\in V_\epsilon} S(H)$. Therefore $\|\boldsymbol{\beta}\| = 0$ and $\lim_{n\to\infty} S(H_n) = s_2$. We use this fact to obtain equations (40) and (41).

Equations (36) and (37) imply,

$$\begin{aligned} \lim_{n\to\infty} E_{H_n}\rho\left(\frac{y - \mathbf{T}_n'\mathbf{x}}{S(H_n)}\right) &= (1-\epsilon)E_{H_0}\rho\left(\frac{y - \tilde{\boldsymbol{\theta}}'\mathbf{x}}{s_2}\right) + \epsilon \\ &\ge (1-\epsilon)E_{H_0}\rho\left(\frac{y}{s_2}\right) + \epsilon > (1-\epsilon)E_{H_0}\rho\left(\frac{y - \boldsymbol{\theta}_t'\mathbf{x}}{s_2}\right) \end{aligned} \tag{40}$$

On the other hand, applying (37),

$$\lim_{n\to\infty} E_{H_n}\rho\left(\frac{y - \boldsymbol{\theta}_t'\mathbf{x}}{S(H_n)}\right) = (1-\epsilon)E_{H_0}\rho\left(\frac{y - \boldsymbol{\theta}_t'\mathbf{x}}{s_2}\right). \tag{41}$$

Therefore, for n large enough,

$$E_{H_n}\rho\left(\frac{y-\mathbf{T}_n'\mathbf{x}}{S(H_n)}\right) > E_{H_n}\rho\left(\frac{y-\boldsymbol{\theta}_t'\mathbf{x}}{S(H_n)}\right).$$

This last inequality is a contradiction since

$$\mathbf{T}_n = \arg\min_{\boldsymbol{\theta}} E_{H_n}\rho\left(\frac{y-\boldsymbol{\theta}'\mathbf{x}}{S(H_n)}\right).$$

For every $t>0$ such that $B_\chi(\epsilon)<t<t_1$ we have found a sequence of distributions $\{H_n\}$ in the neighborhood V_ϵ such that $\sup_n \|\mathbf{T}(H_n)\| \geq t$. Therefore $B_{\mathbf{T}}(\epsilon) \geq t_1$. □

References

Coakley, C.W. and Hettmansperger, T.P. (1993). A bounded influence, high breakdown, efficient regression estimator. *J. Amer. Statist. Assoc.* **88**, 872–880.

Croux, C., Rousseeuw, P.J., and Hössjer, O. (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. Am. Statist. Assoc.* **69**, 383-393.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics. The approach based on influence functions.* Wiley, New York.

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.

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

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

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

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

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

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

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

Yohai, V.J. and Zamar, R.H. (1988). High breakdown 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.

ϵ	0.05	0.1	0.15	0.2	0.25	0.3
Case(1)	0.52	0.82	1.13	1.51	2.01	2.73
Case(2)	0.62	1.12	1.63	2.21	3.00	4.33
Case(3)	0.52	0.92	1.23	1.71	2.31	3.23

Table 1: LMS-estimate maxbias curve for non-elliptical distributions of the regressors.

ϵ	0.05	0.1	0.15	0.2	0.25	0.3
Tukey S-estimate	0.55	0.87	1.23	1.65	2.17	3.02
Tukey τ-estimate	0.92	1.45	2.00	2.65	3.39	4.49
LMS-estimate	0.52	0.82	1.13	1.51	2.01	2.73
LMS τ-estimate	0.90	1.41	1.92	2.51	3.23	4.18
0.5-LTAV-estimate	0.74	1.34	1.84	2.61	3.98	6.62
0.5-LTS-estimate	0.63	1.02	1.45	2.02	2.85	4.19

Table 2: Maxbias curves for several important robust regression estimates.

ϵ	.01	.03	.05	.07	.09	0.11	0.13	0.15
$B_{\mathbf{T}_1}(\epsilon)$	0.31	0.57	0.77	0.96	1.15	1.34	1.55	1.77
$B_{\mathbf{T}_2}(\epsilon)$	0.31	0.57	0.77	0.95	1.14	1.32	1.52	1.73

Table 3: Maxbias curve computations for the MM-estimates $\mathbf{T}_1$ and $\mathbf{T}_2$.

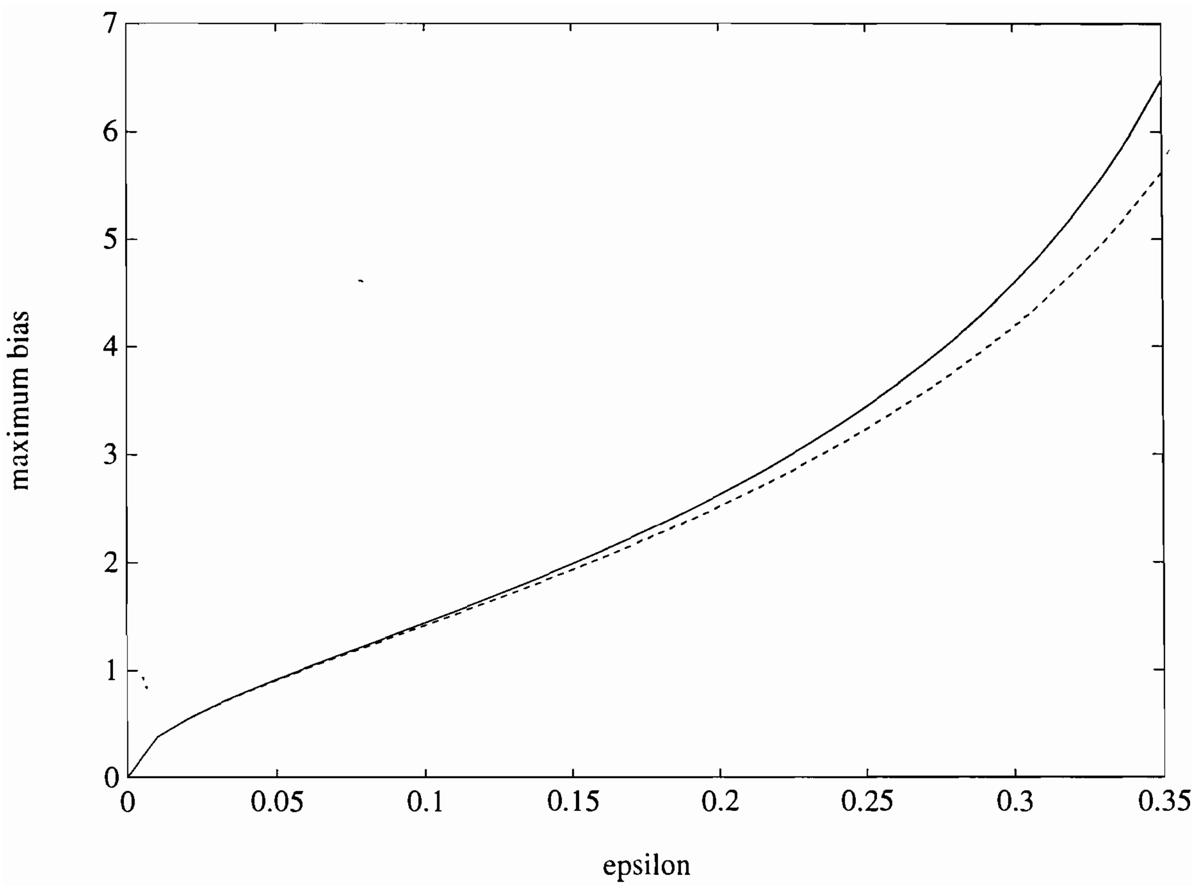


Figure 1: Maxbias curves of Tukey (solid line) and LMS (dashed line) τ-estimates.

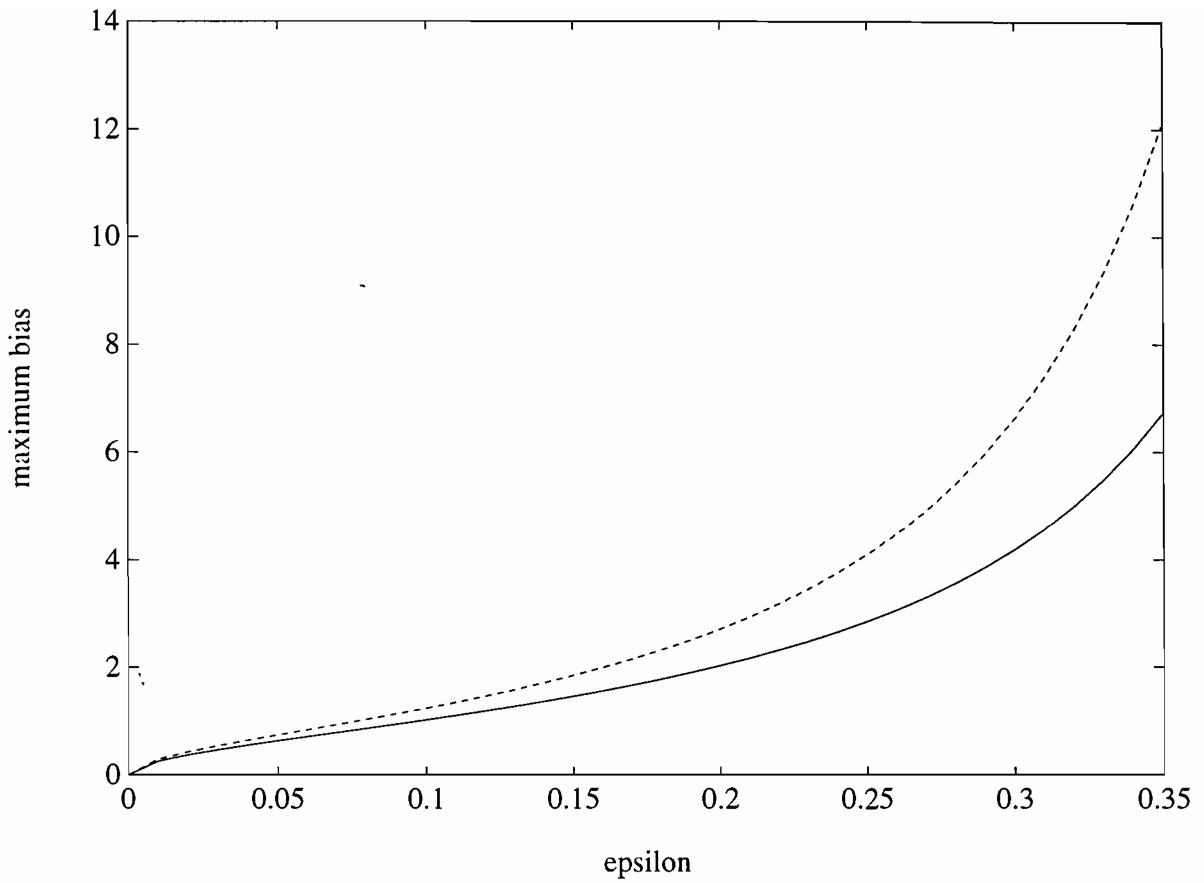


Figure 2: Maxbias curves of 0.5-LTS (solid line) and 0.5-LTAV (dashed line) estimates.

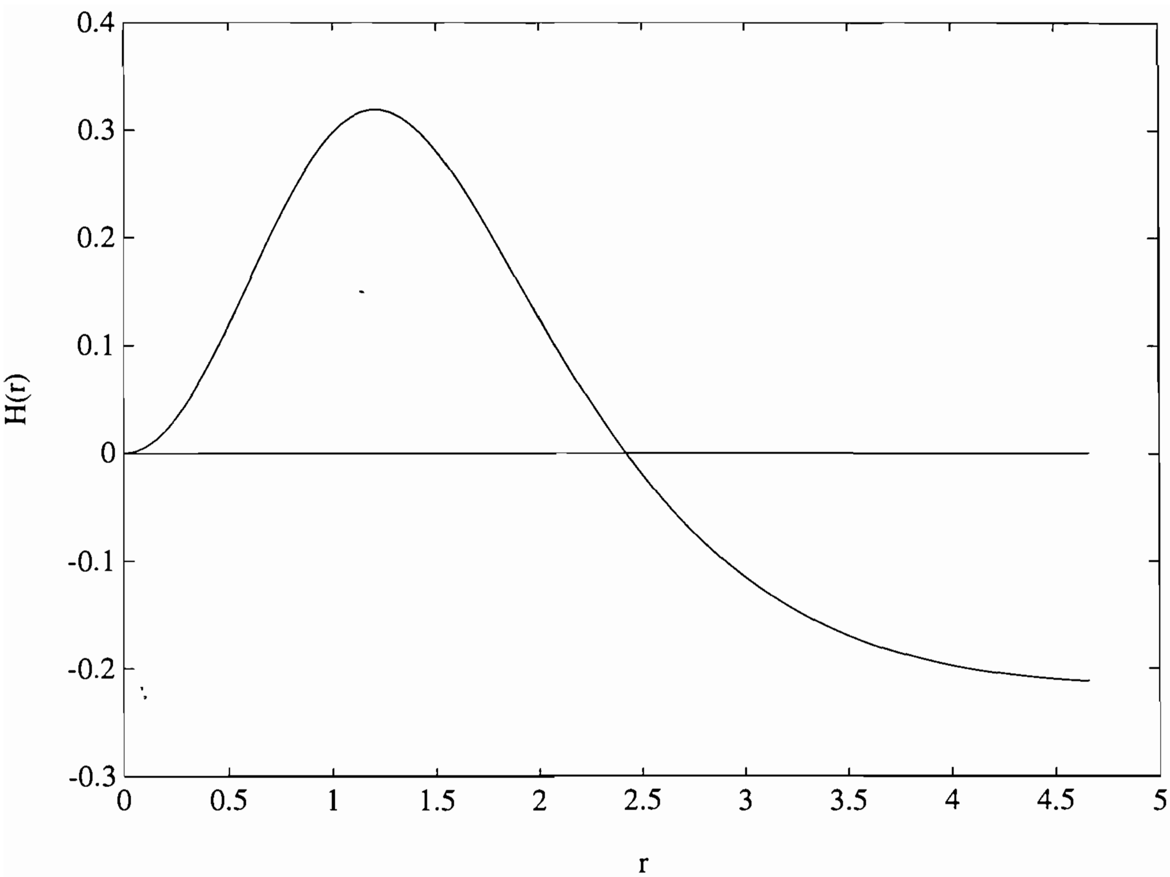


Figure 3: Plot of the function $H(r)$ when ρ is the Tukey bisquare function.

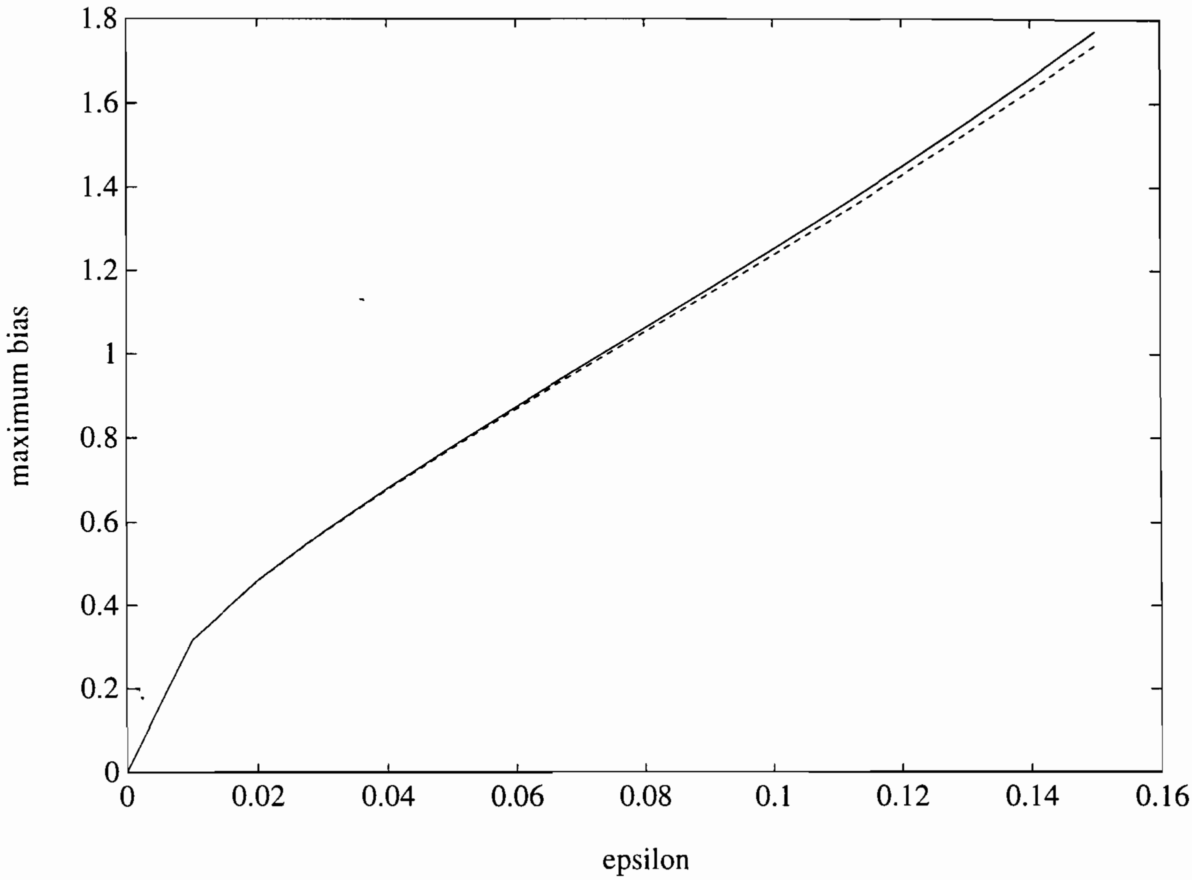


Figure 4: Maxbias curves for the MM-estimates $\mathbf{T}_1$ (dashed line) and $\mathbf{T}_2$ (solid line).