Working Paper 99-60
Statistics and Econometrics Series 21
July 1999

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax 34 - 91- 624.9849

LOCALLY AND GLOBALLY ROBUST ESTIMATORS IN REGRESSION

Sonia Hernández and Victor J. Yohai*

Abstract

A new class of estimates for the linear model is introduced. These estimators, that we call *C*-estimators, are defined as a linear convex combination of the Rousseeuw's least median squares (LMS-) estimator and any other estimate, $\mathbf{T}_2$. We prove that *C*-estimators retain the high breakdown point of the LMS-estimator, but inherit the asymptotic properties and the behaviour in terms of local robustness of $\mathbf{T}_2$. In particular, a *C*-estimators will have $\sqrt{n}$-asymptotics and bounded contamination sensitivity if $\mathbf{T}_2$ does. In addition, efficiency, local robustness properties and the maximum bias curve of *C*-estimators are investigated for different choices of $\mathbf{T}_2$.

Key Words

Linear regression, robust estimates; maximum bias function, high breakdown point, contamination sensivity; high efficiency.

*Hernández, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain, e-mail: sha@est-econ.uc3m.es; Yohai, Universidad de Buenos Aires.

1. Introduction. The linear regression model with random regressors assumes

$$y = \theta'\mathbf{x} + u, \tag{1}$$

where y is the response variable, $\mathbf{x}$ is a p-dimensional vector of explanatory random variables, θ is the vector of p unknown true regression parameters and the error u is a random variable stochastically independent of $\mathbf{x}$.

The classical aproach assumes model (1) to hold exactly for all the points of the observed sample $Z = \{\mathbf{z}_1, ..., \mathbf{z}_n\} = \{(y_1, \mathbf{x}_1')', ..., (y_n, \mathbf{x}_n')'\}$. Nevertheless, it is often the case in practice that the sample contains outliers, that is, observations that do not follow the distribution of most of the data. Robust regression methods are based on the idea that it is more realistic to suppose that the model is only valid for most of the data, and thus they try to devise estimators that are not strongly affected by outliers. A robust method is one that is still valid for partially contaminated samples. The local robustness of an estimator refers to its stability as the amount of contamination in the data approaches zero, whereas the concept of global robustness concerns the behavior of the estimator when the sample contains a large fraction of outliers.

A very informative and natural measure of an estimator robustness is its maximum bias curve, which states the maximum variation caused by a fraction ϵ of outlier contamination. Naturally, such quantity will increase with ϵ, and eventually will become infinite, but we would like to make it, in some sense, as small as possible. This curve combines information on local and global robustness features. In order to measure the local robustness of an estimator we may look at the rate of convergence to zero of its bias curve. On the other hand, to measure global stability we can use the breakdown point (Hampel, 1971), which indicates the minimum fraction of contamination that may yield a completely uninformative value of the estimate.

In recent years, several authors have proposed regression estimators which have a bias curve with both the optimum rate of convergence to zero and the highest breakdown point. Included among these estimators are the one-step GM-estimators (Simpson, Rupert and Carroll, 1992), the projection estimators proposed by Maronna and Yohai (1993) and the class of generalized τ (Gτ)-estimators introduced by Ferretti, Kelmansky, Yohai and Zamar (1994).

However, the breakdown point and the rate of convergence of the bias curve may not suffice to adequately describe the bias. Note that for fractions of contamination smaller than the breakdown point, the maximum bias will be bounded, but it may still be very large. On the other hand, the rate of convergence to zero gives information about the behavior of the estimate only for very small fractions of contamination. Hence it is important to have a more complete account of the bias.

The behavior of projection estimates in terms of their maximum bias curve is very good. However, they are not asymptotically normal, which severely hinders the construction of confidence intervals and hypothesis testing, and hence they are not very adecuate for inference. Furthermore, He and Simpson (1993) prove that they are not locally linear. The locally linearity property esentially means that the estimate has an influence function with finite second moments.

On the other hand, one-step GM-estimates and $G\tau$-estimates also have good local and global propierties, but numerical computations show that the behavior of their maximum bias curve is not very good, specially for ϵ near their breakdown point and large p.

The propose of this article is to present an alternative class of estimators being simultaneously locally and globally robust. These estimators, which we will call C-estimators, are defined as a linear convex combination of a high breakdown point estimator, $\mathbf{T}_1$, and any other estimator $\mathbf{T}_2$. C-estimators retain the breakdown point of $\mathbf{T}_1$ but inherit the asymptotic propierties and the behaviour in terms of local robustness of $\mathbf{T}_2$. Therefore, if we choose a locally robust estimator as $\mathbf{T}_2$, then the resulting C-estimator will be locally and globally robust. On the other hand, choosing the least squares estimator as $\mathbf{T}_2$, we will obtain a C-estimator with high breakdown point and as efficient as the optimal under normal errors.

In the next section we give the basic definitions and notation. In Section 3 we define C-estimators. In Section 4 we stablish that C-estimators have high breakdown point, independently of the choice of $\mathbf{T}_2$. In Section 5 we prove that C-estimators have the same rate of convergence and asymptotic distribution as $\mathbf{T}_2$. In the final two sections we propose choices of $\mathbf{T}_2$ that provide C-estimates which are locally robust or efficient under normal errors respectivily.

2. Basic definitions and notation. Consider the linear model in (1), where the error u is independent of $\mathbf{x}$. Let F_0 be the distribution function of u and let G_0 be the distribution function of $\mathbf{x}$. Then, for each θ, the joint distribution function of $(y, \mathbf{x}')'$ under model (1), H_θ, is given by

$$H_\theta(y, \mathbf{x}')' = \int_{-\infty}^{x_1} \ldots\ldots \int_{-\infty}^{x_p} F_0(y - \theta'\mathbf{s}) dG_0(\mathbf{s}). \tag{2}$$

In particular, for $\theta = 0$, namely when y is independent of $\mathbf{x}$, the joint distribution function H_0 of $(y, \mathbf{x}')'$ is

$$H_0(y, \mathbf{x}')' = F_0(y)\, G_0(\mathbf{x}). \tag{3}$$

Let $\mathcal{H}$ be the set of distributions on $\mathbb{R}^{p+1}$. To allow for a fraction ϵ of data points not following model (1), we consider ϵ-contamination neighborhoods of H_θ of the form

$$V_\epsilon(H_\theta) = \{H : H = (1-\epsilon)H_\theta + \epsilon H^*, \quad H^* \in \mathcal{H}\}.$$

Most estimators of θ can be defined by functionals. Let $\mathbf{T}$ be an $\mathbb{R}^p$ valued functional defined on a broad subset of $\mathcal{H}$ which includes all the empirical distribution functions and the contamination neighborhoods $V_\epsilon(H_\theta)$ for $0 < \epsilon < 0.5$. If H_n is the distribution function corresponding to the observed sample Z, then the estimator of θ associated with $\mathbf{T}$ will be $\hat{\theta}_n(Z) = \mathbf{T}(H_n)$.

Suppose that $(y, \mathbf{x}')'$ satisfies $y = \theta'\mathbf{x} + u$; let $\tilde{\mathbf{x}} = A\mathbf{x}$ and $\tilde{y} = ay + \gamma'\mathbf{x}$, where A is a $p \times p$ non singular matrix, $\gamma \in \mathbb{R}^p$ and $a \in \mathbb{R}$; then $(\tilde{y}, \tilde{\mathbf{x}})$ satisfies $\tilde{y} = \beta'\tilde{\mathbf{x}} + v$, where $\beta = A'^{-1}(a\theta + \gamma)$ and $v = au$. So it will be natural to consider only equivariant estimators, i.e., $\mathbf{T}(\tilde{H}) = A'^{-1}(a\mathbf{T}(H) + \gamma)$, where $\tilde{H}$ is the distribution of $(\tilde{y}, \tilde{\mathbf{x}})$.

The maximum asymptotic bias function of an estimator $\mathbf{T}$ is given by

$$B_{\mathbf{T}}(\epsilon) = \sup_{H \in V_\epsilon(H_\theta)} \left[(\mathbf{T}(H) - \theta)' C_0(G_0)(\mathbf{T}(H) - \theta)\right]^{1/2},$$

where C_0 is a positive definite affine equivariant scatter functional. The affine equivariance of C_0 means that, if G is the distribution of $\mathbf{x}$ and $\tilde{G}$ is the distribution of $\tilde{\mathbf{x}} = A\mathbf{x}$, where A is a $p \times p$ matrix, then

$$C_0(\tilde{G}) = AC_0(G)A'. \tag{4}$$

If C_0 verifies (4) and $\mathbf{T}$ is equivariant, then $B_{\mathbf{T}}(\epsilon)$ does not depend on θ.

In order to measure local robustness of an estimating functional we can use the contamination sensitivity of order q defined by Yohai and Zamar (1992) as

$$\gamma_{\mathbf{T}}^{(q)} = \lim_{\epsilon \downarrow 0} \frac{B_{\mathbf{T}}(\epsilon)}{\epsilon^q}$$

for $q \geq 0$. $\mathbf{T}$ is said to be locally stable of order q if $\gamma_{\mathbf{T}}{}^{(q)} < \infty$. He and Simpson (1993) prove that the optimal rate is $q = 1$, and therefore an estimating functional $\mathbf{T}$ will be considered locally robust if it is locally stable of order 1, namely, if $B_{\mathbf{T}} = O(\epsilon)$.

One measure of global stability is the asymptotic breakdown point, which is given by

$$\epsilon_{\mathbf{T}}^* = \inf\{\epsilon > 0 : B_{\mathbf{T}}(\epsilon) = \infty\}.$$

Equivariant functionals always have $\epsilon_{\mathbf{T}}^* \leq 1/2$ (He and Simpson, 1993) and therefore, in this paper, an estimating functional $\mathbf{T}$ will be considered globally robust if it asymptotic breakdown point is $\epsilon_{\mathbf{T}}^* \leq 1/2$.

Donoho and Huber (1983) give a finite sample version of the breakdown point, closely related to the asymptotic one. Given a sample $Z = \{\mathbf{z}_1, ..., \mathbf{z}_n\}$ of size n and $m < n$, consider the set of all possible contaminated samples that are obtained by replacing any m of the original data points by arbitrary values

$$\Upsilon_n(Z, m) = \left\{\tilde{Z} = \{\tilde{\mathbf{z}}_1, ..., \tilde{\mathbf{z}}_n\} : \sharp\{i \in \{1, ..., n\} : \tilde{\mathbf{z}}_i \neq \mathbf{z}_i\} \leq m\right\}.$$

The maximum bias of the estimate $\hat{\theta}$ of θ at the sample Z when is contamined with m outliers is given by

$$B_n\left(\hat{\theta}, Z, m\right) = \sup_{\widetilde{Z}\in\Upsilon_n(Z,m)} \left\|\hat{\theta}(\widetilde{Z}) - \hat{\theta}(Z)\right\|,$$

where $\| \ \|$ denotes the Euclidean norm. Then, the breakdown point of $\hat{\theta}$ at Z is given by

$$\epsilon_n^*\left(\hat{\theta}, Z\right) = \frac{\min\left\{m : B_n(\hat{\theta}, Z, m) = \infty\right\}}{n}.$$

3. Definition of C-estimators. Given $\theta \in \mathbb{R}^p$, and $(y, \mathbf{x}')'$ with joint distribution H, let $F_{H,\theta}$ be the distribution function of $|y - \theta'\mathbf{x}|$. For each $H \in \mathcal{H}$ and $\theta \in \mathbb{R}^p$ let

$$S(H, \theta) = F_{H,\theta}^{-1}(0.5) = \text{median}_H|y - \theta'\mathbf{x}|.$$

The least-median of squares (LMS)-estimate (Rousseeuw, 1984) is defined by the functional

$$\mathbf{T}(H) \ = \ \arg\min_{\theta\in\mathbb{R}^p} S(H, \theta). \tag{5}$$

Let S^p be defined as $S^p = \{\lambda \in \mathbb{R}^p : \|\lambda\| = 1\}$. A sample $Z = \{(y_1, \mathbf{x}_1')', ..., (y_n, \mathbf{x}_n')'\}$ is said to be in general position if for all $\lambda \in S^p$ verifies

$$\sharp\{i \in \{1, ..., n\} : \lambda'\mathbf{x}_i = 0\} \leq p - 1.$$

For any sample in general position, the finite-sample breakdown point of the LMS-estimator is

$$\epsilon_n^*(Z) = \frac{\left[\frac{n}{2}\right] - p + 2}{n},$$

and therefore its asymptotic breakdown point is $\epsilon^* = 0.5$. However, the LMS-estimator is inefficient, since its rate of convergence is $n^{1/3}$; furthermore, its asymptotic distribution is not normal (Davies, 1990).

We now introduce a new class of estimators which, although are based on the LMS, may not suffer from its disadvantages. Let $\mathbf{T}_1$ be the LMS-estimator defined in (5) and $\mathbf{T}_2$ any other functional. For each $H \in \mathcal{H}$ define

$$S_1(H) = S(H, \mathbf{T}_1(H)),$$

$$S_2(H) = S(H, \mathbf{T}_2(H)),$$

and

$$d(H) = \frac{S_1(H)}{S_2(H)}.$$

The definition of $\mathbf{T}_1$ implies that $0 \leq d(H) \leq 1$ for any distribution H.

Fix constants $0 < c_1 \leq c_2 < 1$. Let $h : \mathbb{R} \rightarrow [0,1]$ be a nondecreasing function which satisfies $h(t) = 1$ for $t \leq c_1$ and $h(t) = 0$ for $t \geq c_2$, and call $\alpha(H) = h(d(H))$. We now define the C-estimator associated with $\mathbf{T}_2(H)$ by

$$\mathbf{T}_3(H) = \alpha(H)\, \mathbf{T}_1(H) + (1 - \alpha(H))\, \mathbf{T}_2(H).$$

In particular, the functional $\mathbf{T}_3$ will be the LMS-estimator when $d(H) < c_1$, that is, when the median of the absolute values of the residuals of $\mathbf{T}_2$ is significantly larger than the minimum median attained by the LMS-estimator. If $d(H) > c_2$, $\mathbf{T}_3$ will be $\mathbf{T}_2$, and for the intermediate values of $d(H)$ $\mathbf{T}_3$ will be a linear combination of both estimators. So the idea is to choose $\mathbf{T}_1$ only if this estimator is significantly "better" than $\mathbf{T}_2$ in terms of the scale that yields the LMS-estimator.

The choice of the LMS-estimator is not essential. The estimator $\mathbf{T}_1$ may be replaced by any S-estimator whith high breakdown point. In this more general case, the criterium for deciding between $\mathbf{T}_1$ and $\mathbf{T}_2$ will be based on the scale which define the S-estimator. The results we obtain here are easily generalized to different choices of $\mathbf{T}_1$.

4. Global robustness of C-estimators. In this section we will prove that any C-estimator is globally robust.

Since we will deal with the finite sample breakdown point, we need to introduce the finite sample version of C-estimators. For each $\theta \in \mathbb{R}^p$ and each sample $Z = \{(y_1, \mathbf{x}_1')', ..., (y_n, \mathbf{x}_n')'\}$ of size n, denote

$$s_n(Z, \theta) = \text{median}\left\{|(y_1 - \theta'\mathbf{x}_1|, ..., |(y_n - \theta'\mathbf{x}_n|\right\}.$$

Let $\hat{\theta}_{1n}$ be the sample LMS-estimator, which is defined for each sample Z by

$$\hat{\theta}_{1n}(Z) = \arg\min_{\theta \in \mathbb{R}^p} s_n(Z, \theta),$$

and let $\hat{\theta}_{2n}$ be any other estimator. Let

$$s_{1n}(Z) = s_n(Z, \hat{\theta}_{1n}(Z)),$$

$$s_{2n}(Z) = s_n(Z, \hat{\theta}_{2n}(Z)),$$

and

$$d_n(Z) = \frac{s_{1n}(Z)}{s_{2n}(Z)}.$$

Call $\alpha_n(Z) = h(d_n(Z))$. The finite sample C-estimator associated with $\hat{\theta}_{2n}$ is defined by

$$\hat{\theta}_{3n}(Z) = \alpha_n(Z)\ \hat{\theta}_{1n}(Z) + (1 - \alpha_n(Z))\ \hat{\theta}_{2n}(Z).$$

In order to show that C-estimators retain the high breakdown point of the LMS-estimator, the following lemma is needed.

LEMMA 4.1. *If $p > 1$ and Z is in general position, then*

$$\sup_{\widetilde{Z} \in \Upsilon_n\left(Z, \left[\frac{n}{2}\right] - p + 1\right)} s_{1n}(\widetilde{Z}) < \infty.$$

PROOF. Let $Z = \{\mathbf{z}_1, ..., \mathbf{z}_n\} = \{(y_1, \mathbf{x}_1')', ..., (y_n, \mathbf{x}_n')'\}$ be a sample in general position and suppose that

$$\sup_{\widetilde{Z} \in \Upsilon_n\left(Z, \left[\frac{n}{2}\right] - p + 1\right)} s_{1n}(\widetilde{Z}) = \infty.$$

Then there exists a sequence

$$\left\{\widetilde{Z}^k\right\}_{k \in \mathbf{N}} = \left\{\left(\widetilde{\mathbf{z}}_1^k, ..., \widetilde{\mathbf{z}}_n^k\right)\right\} = \left\{\left(\widetilde{y}_1^k, \widetilde{\mathbf{x}}_1^k\right), ..., \left(\widetilde{y}_n^k, \widetilde{\mathbf{x}}_n^k\right)\right\} \in \Upsilon_n\left(Z, \left[\frac{n}{2}\right] - p + 1\right),$$

with $\lim_{k \to \infty} s_{1n}(\widetilde{Z}^k) = \infty$.

Let $M = \max\ \{\mid y_1 \mid, ..., \mid y_n \mid\}$ and for each k define the set

$$A_k = \{i \in \{1, ..., n\} : \left|\widetilde{y}_i{}^k\right| > M\}.$$

Consider the sequence $s(\widetilde{Z}^k, 0) = \text{median}\left\{\left|\widetilde{y}_1^k\right|, ..., \left|\widetilde{y}_n^k\right|\right\}$. The definition of $\hat{\theta}_{1n}$ implies that $s(\widetilde{Z}^k, 0) \geq s_{1n}(\widetilde{Z}^k)$ for all k, and hence

$$\lim_{k \to \infty} s(\widetilde{Z}^k, 0) \geq \lim_{k \to \infty} s_{1n}(\widetilde{Z}^k) = \infty,$$

so we can find k_0 such that $k \geq k_0$ implies $s(\widetilde{Z}^k, 0) > M$, and hence

$$\sharp A_k \geq \left[\frac{n+1}{2}\right]. \tag{6}$$

Let $B_k = \left\{i \in \{1, ..., n\} : \widetilde{\mathbf{z}}_i{}^k \neq \mathbf{z}_i\right\}$. Since $|y_i| \leq M$ for all i, it is clear that for every k, $A_k \subset B_k$, and so

$$\sharp A_k \leq \sharp B_k \leq \left[\frac{n}{2}\right] - p + 1 < \left[\frac{n+1}{2}\right],$$

which contradicts (6). □

The following result yields a very large class of estimators with asymptotic breakdown point $\epsilon^* = 1/2$.

THEOREM 4.1. *If $p > 1$ and Z is a sample in general position, then the finite sample breakdown point of any C-estimator satisfies*

$$\epsilon_n^* \left(\hat{\theta}_{3n}, Z\right) \geq \frac{\left[\frac{n}{2}\right] - p + 2}{n}.$$

PROOF. Suppose that $\epsilon_n^* \left(\hat{\theta}_{3n}, Z\right) < \left(\left[\frac{n}{2}\right] - p + 2\right)/n$ for some sample Z in general position. Then there exists a sequence,

$$\left\{\widetilde{Z}^{\,k}\right\}_{k\in\mathbb{N}} \in \Upsilon_n\left(Z, \left[\frac{n}{2}\right] - p + 1\right),$$

such that $\lim_{k\to\infty} \left\|\hat{\theta}_{3n}\left(\widetilde{Z}^{\,k}\right)\right\| = \infty$. Since the breakdown point of the LMS-estimator for any sample in general position is $\epsilon_n^*(\hat{\theta}_{1n}, Z) = \left(\left[\frac{n}{2}\right] - p + 2\right)/n$, then $\limsup_{k\to\infty} \left\|\hat{\theta}_{1n}\left(\widetilde{Z}^{\,k}\right)\right\| < \infty$, so it has to be

$$\lim_{k\to\infty} \left\|\hat{\theta}_{3n}\left(\widetilde{Z}^{\,k}\right)\right\| = \lim_{k\to\infty} \left\|\hat{\theta}_{2n}\left(\widetilde{Z}^{\,k}\right)\right\| = \infty,$$

which implies

$$\liminf_{k\to\infty} d_n\left(\widetilde{Z}^{\,k}\right) > c_1. \tag{7}$$

For each k call $\lambda_k = \hat{\theta}_{2n}(\widetilde{Z}^{\,k}) / \left\|\hat{\theta}_{2n}(\widetilde{Z}^{\,k})\right\|$. Without loss of generality we can assume that $\lim_{k\to\infty} \lambda_k = \lambda$ for some $\lambda \in S^p$. Let $D = \{i \in \{1, ..., n\} : \lambda' \mathbf{x}_i \neq 0\}$. For all $i \in D$ we have

$$\lim_{k\to\infty} \left| y_i - \hat{\theta}_{2n}(\widetilde{Z}^{\,k})' \mathbf{x}_i \right| = \lim_{k\to\infty} \left| y_i - \frac{\hat{\theta}_{2n}(\widetilde{Z}^{\,k})' \mathbf{x}_i}{\left\|\hat{\theta}_{2n}(\widetilde{Z}^{\,k})\right\|} \left\|\hat{\theta}_{2n}(\widetilde{Z}^{\,k})\right\| \right| =$$

$$= \left| y_i - \lambda' \mathbf{x}_i \lim_{k\to\infty} \left\|\hat{\theta}_{2n}(\widetilde{Z}^{\,k})\right\| \right| = \infty,$$

so for each $L > 0$ we can find k_L such that $k \geq k_L$ implies $\left|y_i - \hat{\theta}_{2n}(\widetilde{Z}^{\,k})' \mathbf{x}_i\right| > L$ for all $i \in D$.

For each $k \in \mathbb{N}$, let $C_k = \left\{i \in \{1, ..., n\} : \widetilde{\mathbf{z}}_i^{\,k} = \mathbf{z}_i\right\}$. Fix $L > 0$ and $k \geq k_L$. For every $i \in (C_k \cap D)$ we have

$$\left| \widetilde{y}_i^{\,k} - \hat{\theta}_{2n}(\widetilde{Z}^{\,k})' \widetilde{\mathbf{x}}_i^{\,k} \right| = \left| y_i - \hat{\theta}_{2n}(\widetilde{Z}^{\,k})' \mathbf{x}_i \right| > L.$$

Since Z is in general position, $\sharp D^c \leq p-1$; on the other hand, it is clear that for all k, $\sharp C_k^c \leq [\frac{n}{2}] - p + 1$, and hence

$$\sharp(C_k \cap D) = n - \sharp(C_k^c \cup D^c) \geq n - \sharp C_k^c - \sharp D^c \geq n - \left(\left[\frac{n}{2}\right] - p + 1\right) - (p-1) = \left[\frac{n+1}{2}\right].$$

Thus for $k \geq k_L$,

$$\sharp\left\{i \in \{1, ..., n\} : \left|\tilde{y}_i^{\,k} - \hat{\theta}_{2n}(\tilde{Z}^{\,k})'\tilde{\mathbf{x}}_i^{\,k}\right| > L\right\} \geq \left[\frac{n+1}{2}\right],$$

and hence $s_{2n}(\tilde{Z}^{\,k}) > L$. This yields $\lim_{k\to\infty} s_{2n}(\tilde{Z}^{\,k}) = \infty$, and due to the previous lemma, it follows that

$$\lim_{k\to\infty} d_n(\tilde{Z}^{\,k}) = \frac{\lim_{k\to\infty} s_{1n}(\tilde{Z}^{\,k})}{\lim_{k\to\infty} s_{2n}(\tilde{Z}^{\,k})} = 0,$$

which contradits (7). $\square$

5. Asymptotic Distribution. C-estimators inherit the asymptotic distribution of $\hat{\theta}_{2n}$ under the following assumption

(H1) The median of $|u|$ under F_0 is unique.

Call s_0 this median, namely $s_0 = F_{H_\theta,\theta}^{-1}(0.5) = F_{H_0,0}^{-1}(0.5)$, here and throughout.

Along this section we assume that the data comes from an uncontaminated distribution, and hence, H_n will be the empirical distribution function corresponding to a sample of size n coming from a distribution H_θ given by (2). We will use the following notation:

$$\hat{\theta}_{1n} = \mathbf{T}_1(H_n), \quad \hat{\theta}_{2n} = \mathbf{T}_2(H_n), \quad \hat{\theta}_{3n} = \mathbf{T}_3(H_n);$$

$$s_{1n} = S_1(H_n), \qquad s_{2n} = S_2(H_n);$$

$$d_n = d(H_n) = \frac{s_{1n}}{s_{2n}};$$

$$\alpha_n = \alpha(H_n) = h(d_n).$$

Before stating the asymptotic distribution of C-estimators we will prove the following two lemmas.

LEMMA 5.1. *Let $(y_1, \mathbf{x}_1')', ..., (y_n, \mathbf{x}_n')'$ be i.i.d. observations with distribution H_θ verifying* (2). *Assume that F_0 satisfies* (H1) *and let $\hat{\theta}_n$ be any sequence of consistent estimators of θ. Then, the sequence s_n* = median$\{|y_1 - \hat{\theta}_n'\mathbf{x}_1|, ..., |y_n - \hat{\theta}_n'\mathbf{x}_n|\}$ *is consistent for s_0.*

PROOF. We can assume without loss of generality that the true parameter is $\theta = 0$, i.e., data come from H_0 given by (3).

Fix $\epsilon > 0$ and $\delta > 0$. In orther to prove the lemma we must show that $P_{H_0}[s_n > s_0+\epsilon] < \delta$ and $P_{H_0}[s_n < s_0 - \epsilon] < \delta$ for sufficiently large n.

Let

$$r = \frac{P_{F_0}[\,|u| \le s_0 + \frac{\epsilon}{2}] - \frac{1}{2}}{2}.$$

It is clear that (H1) implies $r > 0$. For every n we have

$$P_{H_0}[\,s_n > s_0 + \epsilon\,] \le P_{H_0}\left[\sharp\left\{i \in \{1, ..., n\} : |u_i - \hat{\theta}_n' \mathbf{x}_i| > s_0 + \epsilon\right\} \ge \frac{n}{2}\right] \le$$

$$\le P_{H_0}\left[\sharp\left\{i \in \{1, ..., n\} : |u_i| + \|\hat{\theta}_n\| \, \|\mathbf{x}_i\| > s_0 + \epsilon\right\} \ge \frac{n}{2}\right] \le$$

$$\le P_{H_0}\left[\sharp\left\{i \in \{1, ..., n\} : \left(|u_i| > s_0 + \frac{\epsilon}{2}\right) \text{ or } \left(\|\hat{\theta}_n\| \, \|\mathbf{x}_i\| > \frac{\epsilon}{2}\right)\right\} \ge \frac{n}{2}\right] \le$$

$$\le P_{F_0}\left[\sharp\left\{i \in \{1, ..., n\} : |u_i| > s_0 + \frac{\epsilon}{2}\right\} \ge n\left(\frac{1}{2} - r\right)\right] +$$

$$+ P_{H_0}\left[\sharp\left\{i \in \{1, ..., n\} : \left(\|\hat{\theta}_n\| \, \|\mathbf{x}_i\| > \frac{\epsilon}{2}\right)\right\} \ge nr\right].$$

For each i, define the random variable

$$N_i = \begin{cases} 0 & \text{if } |u_i| \le s_0 + \frac{\epsilon}{2}, \\ 1 & \text{if } |u_i| > s_0 + \frac{\epsilon}{2}. \end{cases}$$

Since the u_i's are i.i.d., also are the N_i's, and hence we have

$$\sum_{i=1}^{n} N_i/n \xrightarrow{d} E(N_1) = P_{F_0}\left[\,|u_1| > s_0 + \frac{\epsilon}{2}\right] = \frac{1}{2} - 2r \; < \; \frac{1}{2} - r,$$

where $\xrightarrow{d}$ denotes convergence in distribution, so we can find n_1 such that $n \ge n_1$ implies

$$P_{F_0}\left[\sharp\left\{i \in \{1, ..., n\} : |u_i| > s_0 + \frac{\epsilon}{2}\right\} \ge n\left(\frac{1}{2} - r\right)\right] =$$

$$= P_{F_0}\left[\sum_{i=1}^{n} N_i \ge n\left(\frac{1}{2} - r\right)\right] = P_{F_0}\left[\sum_{i=1}^{n} N_i/n \ge \frac{1}{2} - r\right] < \frac{\delta}{2}.$$

Let K be a constant such that $P_{G_0}[\,\|\mathbf{x}\| > K\,] \le r/2$. For each i, define

$$M_i = \begin{cases} 0 & \text{if } \|\mathbf{x}_i\| \le K, \\ 1 & \text{if } \|\mathbf{x}_i\| > K. \end{cases}$$

Since the M_i's are i.i.d., we have

$$\sum_{i=1}^{n} M_i/n \xrightarrow{d} E(M_1) = P_{G_0}[\,\|\mathbf{x}_1\| > K\,] \le \frac{r}{2} < r,$$

and hence there exists n_2 such that $n \ge n_2$ implies

$$P_{G_0}[\,\sharp\{i \in \{1, ..., n\} : \|\mathbf{x}_i\| > K\} \ge nr\,] = P_{G_0}\left[\sum_{i=1}^{n} M_i \ge nr\right] = P_{G_0}\left[\sum_{i=1}^{n} M_i/n \ge r\right] < \frac{\delta}{4}.$$

Since $\hat{\theta}_n \xrightarrow{d} 0$, we can find another number, n_3, such that $n \ge n_3$ implies

$$P_{H_0}\left[\|\,\hat{\theta}_n\| > \frac{\epsilon}{2K}\right] < \frac{\delta}{4}.$$

For $n \ge \max\{\, n_2, n_3 \,\}$ we have

$$P_{H_0}\left[\sharp\left\{i \in \{1, ..., n\} : \|\hat{\theta}_n\|\,\|\mathbf{x}_i\| > \frac{\epsilon}{2}\right\} > nr\right] =$$

$$= P_{H_0}\left[\sharp\left\{i \in \{1, ..., n\} : \|\hat{\theta}_n\|\,\|\mathbf{x}_i\| > \frac{\epsilon}{2}\right\} > nr \;\middle|\; \|\hat{\theta}_n\| \le \frac{\epsilon}{2K}\right] P_{H_0}\left[\|\hat{\theta}_n\| \le \frac{\epsilon}{2K}\right] +$$

$$+ P_{H_0}\left[\sharp\left\{i \in \{1, ..., n\} : \|\hat{\theta}_n\|\,\|\mathbf{x}_i\| > \frac{\epsilon}{2}\right\} > nr \;\middle|\; \|\hat{\theta}_n\| > \frac{\epsilon}{2K}\right] P_{H_0}\left[\|\hat{\theta}_n\| > \frac{\epsilon}{2K}\right] \le$$

$$\le P_{G_0}[\,\sharp\{i \in \{1, ..., n\} : \|\mathbf{x}_i\| > K\} > nr\,] + P_{H_0}\left[\,|\hat{\theta}_n\| > \frac{\epsilon}{2K}\right] < \frac{\delta}{2}.$$

Let $n_0 = \max\{\, n_1, n_2, n_3 \,\}$. Clearly, $P_{H_0}[s_n > s_0 + \epsilon] < \delta$ for $n \ge n_0$.

Proving that $P_{H_0}[S_n < s_0 - \epsilon] < \delta$ for sufficiently large n is similar. □

LEMMA 5.2. *Let $(y_1, \mathbf{x}_1')', ..., (y_n, \mathbf{x}_n')'$ be i.i.d. random variables with distribution H_θ given by* (2). *Assume that F_0 satisfies* (H1) *and suposse that $\hat{\theta}_{2n}$ is a sequence of consistent estimates of θ. Then $n^k\,\alpha_n \xrightarrow{d} 0$ for all $k \in \mathbb{R}$.*

PROOF. By the previous lemma we have $s_{1n} \xrightarrow{d} s_0$ and $s_{2n} \xrightarrow{d} s_0$, and therefore

$$d_n = \frac{s_{1n}}{s_{2n}} \xrightarrow{d} 1,$$

which implies $\lim_{n\to\infty} P_{H_\theta}[\, d_n \le c\,] = 0$ for any $c < 1$. Hence, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P_{H_\theta}[\, |n^k\alpha_n| > \epsilon\,] \le \lim_{n\to\infty} P_{H_\theta}[\, \alpha_n \ne 0\,] \le \lim_{n\to\infty} P_{H_\theta}[\, d_n \le c_1\,] = 0. \quad \square$$

The next theorem establishes that a C-estimator has the same asymptotic distribution as $\mathbf{T}_2$.

THEOREM 5.1. *Let $(y_1, \mathbf{x}_1')', ..., (y_n, \mathbf{x}_n')'$ be i.i.d. with distribution H_θ given by* (2) *where F_0 satisfies* (H1). *If the estimator $\hat{\theta}_{2n}$ verifies $n^k(\hat{\theta}_{2n} - \theta) \xrightarrow{d} Z$ for some $k \ge 0$ and some distribution Z, then the resulting C-estimator also verifies $n^k(\hat{\theta}_{3n} - \theta) \xrightarrow{d} Z$.*

PROOF. Note that

$$n^k(\hat{\theta}_{3n} - \hat{\theta}_{2n}) = n^k(\alpha_n\, \hat{\theta}_{1n} + (1 - \alpha_n)\, \hat{\theta}_{2n} - \hat{\theta}_{2n}) =$$

$$= n^k\alpha_n(\hat{\theta}_{1n} - \hat{\theta}_{2n}) = n^{(k-1/3)}\, \alpha_n\, n^{1/3}(\hat{\theta}_{1n} - \theta) - \alpha_n\, n^k(\hat{\theta}_{2n} - \theta).$$

By LEMMA 5.2 , $n^{(k-1/3)}\, \alpha_n \xrightarrow{d} 0$ and $\alpha_n \xrightarrow{d} 0$; furthermore $n^k(\hat{\theta}_{2n} - \theta) \xrightarrow{d} Z$ and $n^{1/3}(\hat{\theta}_{1n} - \theta)$ converges in distribution to a certain random variable C_τ (Rousseeuw and Leroy, 1987). Hence $n^k(\hat{\theta}_{3n} - \hat{\theta}_{2n}) \xrightarrow{d} 0$ which implies $n^k(\hat{\theta}_{3n} - \theta)$ has the same limit distribution as $n^k(\hat{\theta}_{2n} - \theta)$. $\square$

6. Locally robustness of C-estimators. In this section we establish the local robustness of the C-estimators that combines the LMS with a GM-estimator.

GM-estimators are the best known class of estimators which are locally stable of order 1. For each $H \in \mathcal{H}$, a GM-estimate of regression is obtained by solving for θ an equation of the form

$$E_H\left[\psi\left(y - \theta'\mathbf{x}, \|\mathbf{x}\|_G\right)\, \mathbf{x}\right] = 0, \tag{8}$$

where $\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ satisfies the following properties:

A1. $\psi(u, v)$ is continuous except for a finite number of points.

A2. $\psi(u, v)$ is odd with respect to u for all v.

A3. For all v, the function $\psi(u, v)$ is nondecreasing with respect to u for $u > 0$ and strictly increasing at $u = 0$.

Moreover $\|\mathbf{x}\|_G = (\mathbf{x}'C_0(G)\mathbf{x})^{1/2}$, where C_0 is a robust scatter matrix satisfying the affine equivariance propierty (4).

$\gamma_{\mathbf{T}}^{(1)}$ is finite for the GM-estimator defined by (8) if and only if its function ψ verifies

$$\sup_{u,v} |\, \psi(u,v)\ v| < \infty. \tag{9}$$

However, Maronna, Bustos and Yohai (1989) proved that the breakdown point of GM-estimators tends to zero as p increases, and therefore global stability of these estimators is satisfactory only for small p.

In this section we prove that taking a GM-estimator as $\mathbf{T}_2$, the resulting C-estimator, $\mathbf{T}_3$, will also verify $\gamma_{\mathbf{T}_3}^{(1)} < \infty$. We will require some additional assumptions.

(H2) F_0 is absolutely continuous, with density f_0 which is symmetric, continuous and strictly decreasing for $u \geq 0$.

(H3) $\sup_{\lambda \in S^p} P_{G_0}[\lambda'\mathbf{x} = 0] = 0$.

In order to prove that the C-estimators based on GM are locally stable of order one, we will use the following five lemmas.

LEMMA 6.1. *Let the functions $l, h : \mathbb{R}^2 \to [0,1]$ and $s : \mathbb{R} \to \mathbb{R}^+$ defined by*

$$l(t,k) = \inf_{\lambda \in S^p} P_{H_0}\left[\, |y - t\lambda'\mathbf{x}| \leq k\right],$$

$$h(t,k) = \sup_{\lambda \in S^p} P_{H_0}\left[\, |y - t\lambda'\mathbf{x}| \leq k\right],$$

and

$$s(t) = \inf_{\lambda \in S^p} S(H_0, t\lambda).$$

Assume that F_0 satisfies (H2) *and G_0 satisfies* (H3). *Then*

(i) *$l(t,k)$ is continuous in t for all $k > 0$.*

(ii) *$h(t,k)$ is strictly decreasing with respect to $|t|$ for every $k > 0$.*

(iii) *$s(t)$ is a strictly increasing function of $|t|$.*

PROOF.

(i) Fix $k > 0$ and let t_n be any sequence of real numbers with $\lim_{n\to\infty} t_n = t_0$. Since the function $p(\lambda) = P_{H_0}[\,|y - t\lambda'\mathbf{x}| \leq k]$ is continuous with respect to λ, and by compactness of S^p, we can find $\lambda_0 \in S^p$ such $l(t_0,k) = P_{H_0}[\,|y - t_0\lambda_0'\mathbf{x}| \leq k]$ and a sequence $\{\lambda_n\}_{n\in\mathbb{N}} \subset S^p$ such that for each n, $l(t_n,k) = P_{H_0}[\,|y - t_n\lambda_n'\mathbf{x}| \leq k]$.

Assume without loss of generality that $\lim_{n\to\infty} \lambda_n = \tilde{\lambda}$ for some $\tilde{\lambda} \in S^p$. Since

$$P_{H_0}[\,|y - t_n\lambda_n'\mathbf{x}| \leq k] \leq P_{H_0}[\,|y - t_n\lambda_0'\mathbf{x}| \leq k]$$

for all n, we have

$$P_{H_0}\left[\,|y - t_0\widetilde{\lambda}'\mathbf{x}| \le k\right] \le P_{H_0}\left[\,|y - t_0\lambda_0'\mathbf{x}| \le k\right],$$

which yields $\lim_{n\to\infty} l(t_n, k) = l(t_0, k)$ $\square$

(ii) From (H2) it is straightforward that for each $k > 0$ the function $P_{F_0}[\,|u - m| \le k]$ is strictly decreasing with respect to $|m|$, and hence, the conditional probability

$$\widetilde{h}(t, \lambda, k, \mathbf{x}) = P_{H_0}[\,|u - t\lambda'\mathbf{x}| \le k \mid \mathbf{x}]$$

is a strictly decreasing function of $|t|$ for every $k > 0$, every $\lambda \in S^p$ and every $\mathbf{x} \in \mathbb{R}^p$ such that $\lambda'\mathbf{x} \neq 0$.

Fix $k > 0$ and $|t_1| < |t_2|$; for any $\lambda \in S^p$ we have

$$\widetilde{h}(t_1, \lambda, k, \mathbf{x}) - \widetilde{h}(t_2, \lambda, k, \mathbf{x}) \ge 0,$$

for all $\mathbf{x} \in \mathbb{R}$, and equality holds only if $\lambda'\mathbf{x} = 0$, which has null probability under G_0 because of (H3); therefore,

$$P_{H_0}[\,|y - t_1\lambda'\mathbf{x} \le k] > P_{H_0}[\,|y - t_2\lambda'\mathbf{x} \le k].$$

Take $\lambda_1, \lambda_2 \in S^p$ such that $h(t_1, k) = P_{H_0}\left[\,|y - t_1\lambda_1'\mathbf{x}| \le k\right]$ and $h(t_2, k) = P_{H_0}\left[\,|y - t_2\lambda_2'\mathbf{x}| \le k\right]$. We have

$$h(t_1, k) = P_{H_0}\left[\,|y - t_1\lambda_1'\mathbf{x}| \le k\right] \ge P_{H_0}\left[\,|y - t_1\lambda_2'\mathbf{x}| \le k\right] > P_{H_0}\left[\,|y - t_2\lambda_2'\mathbf{x}| \le k\right] = h(t_2, k),$$

and thus $h(t, k)$ is strictly decreasing with respect to $|t|$. $\square$

(iii) Let $|t_1| < |t_2|$; from (ii) and (H2) we have that

$$\sup_{\lambda \in S^p} P_{H_0}\left[\,|y - t_2\lambda'\mathbf{x}| \le s(t_1)\right] < \sup_{\lambda \in S^p} P_{H_0}\left[\,|y - t_1\lambda'\mathbf{x}| \le s(t_1)\right] = \frac{1}{2},$$

and hence $s(t_1) < s(t_2)$. $\square$

LEMMA 6.2. *Let* $\mathbf{T}_1$ *be the LMS-estimator defined in* (5). *Under the assumptions in Lemma* 6.1,

$$\lim_{\epsilon \downarrow 0} \sup_{H \in V_\epsilon(H_0)} \|\,\mathbf{T}_1(H)\| = 0.$$

PROOF. Fix $\delta > 0$; we have to prove that there exists $\epsilon_0 > 0$ such that $\|\mathbf{T}_1(H)\| < \delta$ for all $H \in V_{\epsilon_0}(H_0)$. Let

$$k = \frac{s(\delta) - s_0}{2},$$

$$w = P_{F_0}[\,|u| \le s_0 + k] - \frac{1}{2},$$

and

$$r = \frac{1}{2} - h(\delta, s_0 + k).$$

The previous lemma implies that $w > 0$ and $r > 0$. Put

$$\epsilon_1 = \frac{w}{\frac{1}{2} + w}$$

and

$$\epsilon_2 = \frac{r}{\frac{1}{2} + r}.$$

For any $H \in V_{\epsilon_1}(H_0)$ we have

$$P_H[\,|y| \leq s_0 + k] \geq (1 - \epsilon_1)P_{F_0}[\,|u| \leq s_0 + k] = (1 - \epsilon_1)\left(\frac{1}{2} + w\right) = \frac{1}{2},$$

and therefore $S(H, 0) \leq s_0 + k$. On the other hand, LEMMA 6.1 also implies that for each $H \in V_{\epsilon_2}(H_0)$ and each $\theta \in \mathbb{R}^p$ with $\|\theta\| > \delta$,

$$P_H[\,|y - \theta'\mathbf{x}| \leq s_0 + k] \leq (1 - \epsilon_2)P_{H_0}[\,|y - \theta'\mathbf{x}| \leq s_0 + k] + \epsilon_2 \leq (1 - \epsilon_2)\, h\left(\,\|\theta\|, s_0 + k\right) + \epsilon_2 <$$

$$< (1 - \epsilon_2)\, h(\delta, s_0 + k) + \epsilon_2 = (1 - \epsilon_2)\left(\frac{1}{2} - r\right) + \epsilon_2 = \frac{1}{2},$$

and hence $S(H, \theta) > s_0 + k$.

Let $\epsilon_0 = \min\{\epsilon_1, \epsilon_2\}$. It is clear that for all $H \in V_{\epsilon_0}(H_0)$ we have $S(H, 0) < S(H, \theta)$ for any $\theta \in \mathbb{R}^p$ with $\|\theta\| > \delta$, so it has to be $\|\mathbf{T}_1(H)\| < \delta$. □

LEMMA 6.3. *Define the functions*

$$g(t) = \inf_{\lambda \in S^p} |E_{H_0}\left[\,\psi(y - t\lambda'\mathbf{x}, \|\mathbf{x}\|_G)\, \lambda'\mathbf{x}\right]|$$

and

$$n(t) = \inf_{\lambda \in S^p} \|E_{H_0}\left[\,\psi(y - t\lambda'\mathbf{x}, \|\mathbf{x}\|_G)\, \mathbf{x}\right]\|\,.$$

Suppose that ψ satisfies A1, A2 *and* A3, *F_0 satisfies* (H2) *and G_0 satisfies* (H3). *Then*

(i) *$g(t)$ is a nondecreasing function of $|t|$ and verifies $g(t) > 0$ for all $t \neq 0$.*

(ii) *$n(t) \geq g(t)$ for all $t \in \mathbb{R}$.*

PROOF.

(i) For every $\lambda \in S^p$ and $\mathbf{x} \in \mathbb{R}^p$, define the function

$$q_{\lambda,\mathbf{x}}(t) = E_{F_0}\left[\,\psi(y - t\lambda'\mathbf{x}, \|\mathbf{x}\|_G)\right].$$

It is easy to show that (H2) and A2 imply that $q_{\lambda,\mathbf{x}}(t)$ is odd with respect to t; moreover A3 implies that $q_{\lambda,\mathbf{x}}(t)$ is a nonincreasing function of t if $\lambda'\mathbf{x} > 0$ and a nondecreasing function if $\lambda'\mathbf{x} < 0$. Therefore, $q_{\lambda,\mathbf{x}}(t)\lambda'\mathbf{x}$ is an odd and nonincreasing function of t, which implies that the function

$$\tilde{g}(\lambda, t) = |\, E_{H_0}\left[\, \psi\left(y - t\lambda'\mathbf{x}, \|\mathbf{x}\|_G\right) \lambda'\mathbf{x}\,\right] | = \left| \int_{\mathbb{R}^p} q_{\lambda'\mathbf{x}}(t)\, \lambda'\mathbf{x}\, dG_0(\mathbf{x}) \right|$$

is even and nondecreasing with respect to $|t|$.

By compactness of S^p and continuity of g, for each $t \in \mathbb{R}$ we can find $\lambda_t \in S^p$ such that $g(t) = \tilde{g}(\lambda_t, t)$. Then, for $|t_1| < |t_2|$ we have

$$g(t_1) = \tilde{g}(\lambda_{t_1}, t_1) \leq \tilde{g}(\lambda_{t_2}, t_1) \leq \tilde{g}(\lambda_{t_2}, t_2) = g(t_2)$$

and hence $g(t)$ is a nondecreasing function of $|t|$.

Fix now $t_0 \neq 0$. By A3 and (H2) we get that $q_{\lambda,\mathbf{x}}(t_0)\lambda'\mathbf{x} = 0$ if and only if $\lambda'\mathbf{x} = 0$, which has nulle probability under G_0. Thereby, $\tilde{g}(\lambda, t_0) > 0$ for any $\lambda \in S^p$ and, in particular, $g(t_0) = \tilde{g}(\lambda_{t_0}, t_0) > 0$. □

(ii) Using the Cauchy-Schwartz inequality we get

$$\|E_{H_0}\left[\psi(y - t\lambda'\mathbf{x}, \|\mathbf{x}\|_G)\mathbf{x}\right]\| = \|\lambda\|\ \|E_{H_0}\left[\psi(y - t\lambda'\mathbf{x}, \|\mathbf{x}\|_G)\mathbf{x}\right]\| \geq$$

$$\geq |\lambda' E_{H_0}\left[\psi(y - t\lambda'\mathbf{x}, \|\mathbf{x}\|_G)\mathbf{x}\right]| = \tilde{g}(\lambda, t)$$

for all $t \in \mathbb{R}$ and $\lambda \in S^p$. Since the function $\|E_{H_0}\left[\psi(y - t\lambda'\mathbf{x}\|\mathbf{x}\|_G)\,\mathbf{x}\right]\|$ is continuous with respect to λ, a straighforward argument show that $n(t) \geq g(t)$ □

LEMMA 6.4. *Assume that ψ satisfies* A1, A2, A3 and (9), F_0 *satisfies* (H2) *and* G_0 *satisfies* (H3). Let $\mathbf{T}_2$ be the GM-estimator based on ψ. *Then*

$$\lim_{\epsilon \downarrow 0} \sup_{H \in V_\epsilon(H_0)} \|\mathbf{T}_2(H)\| = 0.$$

PROOF. Let $A = \sup_{u,v} |\psi\,(u, v)\, v| + 1$. Given $\delta > 0$, put

$$\epsilon_0 = \frac{g(\delta)}{g(\delta) + A}\,.$$

LEMMA 6.3 implies that $\epsilon_0 > 0$. We will prove that $\|\mathbf{T}_2(H)\| < \delta$ for any $H \in V_{\epsilon_0}(H_0)$.

For each $H \in V_{\epsilon_0}(H_0)$ we have that $\mathbf{T}_2(H)$ verifies the equality

$$(1 - \epsilon_0)\, E_{H_0}[\, \psi\,(y - \mathbf{T}_2(H)'\mathbf{x}, \|\mathbf{x}\|_G)\, \mathbf{x}] + \epsilon_0\, E_{H^*}[\, \psi\,(y - \mathbf{T}_2(H)'\mathbf{x}, \|\mathbf{x}\|_G)\, \mathbf{x}\,] = 0$$

for some arbitrary distribution H^*, and hence verifies

$$\|E_{H_0}[\psi\,(y-\mathbf{T}_2(H)'\mathbf{x},\|\mathbf{x}\|_G)\,\mathbf{x}]\,\| = \frac{\epsilon_0}{1-\epsilon_0}\ \|E_{H^*}[\psi\,(y-\mathbf{T}_2(H)'\mathbf{x},\|\mathbf{x}\|_G)\,\mathbf{x}]\,\| <$$

$$< \frac{\epsilon_0}{1-\epsilon_0}\,A = g\,(\delta). \tag{10}$$

Suppose that $\|\mathbf{T}_2(H)\| \geq \delta$. Then, by LEMMA 6.3 we would have

$$\|E_{H_0}[\psi\,(y-\mathbf{T}_2(H)'\mathbf{x},\|\mathbf{x}\|_G)\,\mathbf{x}]\,\| \geq n(\|\mathbf{T}_2(H)\|) \geq g(\|\mathbf{T}_2(H)\|) \geq g(\delta),$$

which contradicts (10). □

LEMMA 6.5. *Suppose that all the assumptions in Lemma 6.4 hold. Let $\mathbf{T}_1$ be the LMS-estimator and let $\mathbf{T}_2$ be the GM-estimator based on ψ. Then, for any $\delta > 0$ there exists $\epsilon_0 > 0$ such that $H \in V_{\epsilon_0}(H_0)$ implies $s_0 - \delta < S_1(H) \leq S_2(H) < s_0 + \delta$.*

PROOF. Given $\delta > 0$, put

$$r = \frac{\frac{1}{2} - P_{F_0}[\,|u| \leq s_0 - \delta]}{2}$$

and

$$w = \frac{P_{F_0}[\,|u| \leq s_0 + \delta] - \frac{1}{2}}{2}.$$

Clearly (H2) implies that $w > 0$ and $r > 0$. Let

$$\epsilon_1 = \frac{r}{\frac{1}{2}+r}.$$

LEMMA 6.1 implies that $P_{H_0}[\,|y-\theta'\mathbf{x}| \leq s_0 - \delta] \leq 1/2 - 2r$ for all $\theta \in \mathbb{R}^p$, whereby, for any distribution $H \in V_{\epsilon_1}(H_0)$,

$$P_H[\,|y-\theta'\mathbf{x}| \leq s_0-\delta] \leq (1-\epsilon_1)P_{H_0}[\,|y-\theta'\mathbf{x}| \leq s_0-\delta] + \epsilon_1 \leq (1-\epsilon_1)\left(\frac{1}{2}-2r\right) + \epsilon_1 < \frac{1}{2},$$

so $S(H,\theta) > s_0 - \delta$, and in particular $S_1(H) = S(H, \mathbf{T}_1(H)) > s_0 - \delta$.

On the other hand, LEMMA 6.1 also implies that

$$\lim_{t\to 0} l(t, s_0+k) = l(0, s_0+k) = \frac{1}{2} + 2w,$$

so we can find $t_o > 0$ such that $|t| < t_o$ implies

$$l(t, s_0+k) > \frac{1}{2} + w.$$

Furthermore, by LEMMA 6.4, we can find $\epsilon_2 > 0$ such that $H \in V_{\epsilon_2}(H_0)$ implies $\|\mathbf{T}_2(H)\| < t_0$. Put

$$\epsilon_3 = \min\left\{\epsilon_2, \frac{w}{\frac{1}{2}+w}\right\}.$$

For each $H \in V_{\epsilon_3}(H_0)$ we have

$$P_H[\,|y - \mathbf{T}_2(H)'\mathbf{x}| < s_0 + \delta] \geq (1-\epsilon_3)P_{H_0}[\,|y - \mathbf{T}_2(H)'\mathbf{x}| < s_0 + \delta] \geq$$

$$\geq (1-\epsilon_3)l(\|\mathbf{T}_2(H)\|, s_0 + k) > (1-\epsilon_3)\left(\frac{1}{2}+w\right) \geq \frac{1}{2},$$

and hence $S_2(H) < s_0 + \delta$.

Now it is enough to take $\epsilon_0 = \min\{\epsilon_1\,,\ \epsilon_3\,\}$. □

The next theorem establish the local robustness of the C-estimators considered in this section.

THEOREM 6.1. *Suppose that all the assumptions in Lemma 6.4 hold. Let $\mathbf{T}_1$ be the LMS-estimator, $\mathbf{T}_2$ be the GM-estimator based on ψ and let $\mathbf{T}_3$ be the resulting C-estimator. Then $\mathbf{T}_3$ is locally stable of order one.*

PROOF. We can suppose without loss of generality that the true parameter is $\theta = 0$. Let

$$\delta = \frac{(1-c_2)s_0}{1+c_2}.$$

(H2) implies that $\delta > 0$, and then, by LEMMA 6.5, we can find $\epsilon_0 > 0$ such that for all $H \in V_{\epsilon_0}(H_0)$ is $S_1(H) > s_0 - \delta$ and $S_2(H) < s_0 + \delta$, and hence

$$d(H) = \frac{S_1(H)}{S_2(H)} > \frac{s_0 - \delta}{s_0 + \delta} = c_2,$$

so $\mathbf{T}_3(H) = \mathbf{T}_2(H)$. This yields

$$\lim_{\epsilon \downarrow 0} \frac{B_{\mathbf{T}_3}(\epsilon)}{\epsilon^q} = \lim_{\epsilon \downarrow 0} \frac{B_{\mathbf{T}_2}(\epsilon)}{\epsilon^q}$$

for any $q > 0$, and then (9) implies that

$$\gamma^{(1)}_{\mathbf{T}_3} = \gamma^{(1)}_{\mathbf{T}_2} < \infty. \quad \square$$

7. Numerical evaluation of C-estimators. The main goal of this section is to assess the performance of locally and globally robust regression estimates and to investigate the behavior of hybrid estimates proposed in section 6, which are simultaneously locally and globally robust. To this effect we have computed the maximum asymptotics bias of these C-estimators and conducted a Monte Carlo study.

From the results in previous sections, it follows that C-estimators constructed with a GM-estimator as $\mathbf{T}_2$ have the following properties:

- Their asymptotic breakdown point is $\epsilon^*_{\mathbf{T}_3} = 0.5$.

- They are locally stable of order 1, that is, $\gamma^{(1)}_{\mathbf{T}_3} < \infty$.

- Under the uncontaminated model they are $n^{1/2}$ consistent and asymptotically normal.

Hence they are asymptotically normal estimators with the best possible breakdown point, the best order of convergence to zero of the maximum bias function and the best order of consistency. However, as we explained in section 1, these properties are not sufficient to guarantee a complete robust behavior.

7.1. Asymptotic bias. A more complete understanding of an estimator's behavior is obtained by computing the maximum bias for mass point contaminated distribution, which is defined by

$$B^*_{\mathbf{T}}(\epsilon) = \sup_{(y,\mathbf{x}')'} \left[(\mathbf{T}((1-\epsilon)H_\theta + \epsilon\delta_{y,\mathbf{x}}) - \theta)' C_0(G_0)(\mathbf{T}((1-\epsilon)H_\theta + \epsilon\delta_{y,\mathbf{x}}) - \theta)\right]^{1/2},$$

where $\delta_{y,\mathbf{x}}$ is the point mass distribution that gives probability 1 to $(y,\mathbf{x}')'$, and C_0 is a functional that satisfies the equivariance condition in (4).

We have computed numerically the maximum bias under mass point contamination for C-estimators in the case that H_θ is $N_{p+1}(0,I)$. By equivariance, there is no loss of generality in standardizing the distribution.

We have taken the estimator $\mathbf{T}_2$ to be that with minimax bias within the class GM. Martin, Yohai y Zamar (1989) prove that this estimator corresponds to the function $\psi(u,v) = \mathrm{sgn}(u)/v$, and therefore is defined by

$$E_{H_0}\left[\mathrm{sgn}(y - \mathbf{T}_2(H)'\mathbf{x})\frac{\mathbf{x}}{||\mathbf{x}||}\right] = 0,$$

or equivalently

$$\mathbf{T}_2(H) = \arg\min_{\theta\in\mathbf{R}^p} E_H\left(\frac{|y-\theta'\mathbf{x}|}{||\mathbf{x}||}\right), \tag{11}$$

with $||\mathbf{x}|| = (\mathbf{x}'\Sigma\mathbf{x})^{-1}$, where Σ is the covariance matrix of $\mathbf{x}$. Since Σ is unknown it has been replaced by the robust covariance estimator based on projections proposed by Maronna, Stahel y Yohai (1992).

In order to simplify the calculations we have taken $c_1 = c_2 = c$, so that the functional $\mathbf{T}_3$ is of the form

$$\mathbf{T}_3(H) = \begin{cases} \mathbf{T}_1(H) & \text{if } \frac{S_1(H)}{S_2(H)} < c, \\ \\ \mathbf{T}_2(H) & \text{if } \frac{S_1(H)}{S_2(H)} \geq c. \end{cases}$$

Since GM-estimators depend on the number of regressors, we have considered different values of p. Tables 7.1 to 7.5 present the values of $B^*_{\mathbf{T}_3}$ for several values of p, ϵ and c.

Tables also include the values of B^* for the LMS-estimator and the GM-estimator in (11) and for two more clases of estimators having both high breakdown point and bounded contamination sensitivity of order 1: the generalized τ-estimators, which we will denote here by $G\tau$, and the one-step GM-estimators, denoted as GM1. The $G\tau$-estimate considered here is the same as in section 3 of Ferreti et al. (1994). GM1 is a one–step Newton–Raphson version of the Ryan type GM–estimate with ψ–function in the family

$$\psi_c(x) = c \tanh(x/k),$$

and weight function $w(x) = 1/|x|$. Simpson and Yohai (1997) show that when $k \to 0$ the corresponding gross error sensitivity tends to the lower bound for locally linear functionals. This value is attained by the minimax GM–estimate found by Martin et al. (1989) and given in (11). The values of k for each p are chosen to match the gross error sensitivity of $G\tau$ and may be found in Table 4.2. of Ferreti et al. (1994).

The robust covariance matrix estimate used in GM1 and $G\tau$ is in the Donoho–Stahel family (see Donoho, 1982, Stahel, 1981, and Maronna and Yohai, 1995). The specific definition of this estimate requires the choice of another weight function w^*. Based on the results of Maronna and Yohai (1995), we took $w^*(u) = min(1, l/u)$ where l is the square root of the 95% percentile of a χ-squared distribution with p degrees of freedom.

Table 7.1: Maximum bias of several C-estimators that combines LMS and GM, of $G\tau$-estimator and of GM1-estimator for $p = 2$

	p=2	$\epsilon = .05$	$\epsilon = .10$	$\epsilon = .15$	$\epsilon = .20$
	$B_{\mathbf{T}_{LMS}}(\epsilon)$	0.53	0.83	1.14	1.52
	$B_{\mathbf{T}_{GM}}(\epsilon)$	0.11	0.26	0.47	0.83
$B^*_{\mathbf{T}_3}$	$c = .9$	0.22	0.66	1.09	1.51
	$c = .8$	0.11	0.38	0.86	1.47
	$c = .7$	0.11	0.26	0.57	1.19
	$\mathbf{c = .6}$	0.11	0.26	0.47	0.83
	$B^*_{G\tau}$	0.25	0.43	0.76	1.32
	B^*_{GM1}	0.15	0.31	0.51	0.75

Table 7.2: Maximum bias of several C-estimators that combines LMS and GM, of $G\tau$-estimator and of GM1-estimator for $p = 5$

p = 5		$\epsilon = .05$	$\epsilon = .10$	$\epsilon = .15$	$\epsilon = .20$
	$B_{\mathbf{T}_{LMS}}(\epsilon)$	0.53	0.83	1.14	1.52
	$B_{\mathbf{T}_{GM}}(\epsilon)$	0.18	0.47	1.23	∞
$B^*_{\mathbf{T}_3}$	$c = .9$	0.25	0.77	1.14	1.75
	c = .8	0.18	0.51	1.14	2.04
	$c = .7$	0.18	0.47	1.23	2.39
	$B^*_{G\tau}(\epsilon)$	0.25	0.56	1.50	2.15
	$B^*_{GM1}(\epsilon)$	0.18	0.42	0.74	2.29

Table 7.3: Maximum bias of several C-estimators that combines LMS and GM, of $G\tau$-estimator and of GM1-estimator for $p = 10$

p = 10		$\epsilon = .05$	$\epsilon = .10$	$\epsilon = .15$	$\epsilon = .20$
	$B_{\mathbf{T}_{LMS}}(\epsilon)$	0.53	0.83	1.14	1.52
	$B_{\mathbf{T}_{GM}}(\epsilon)$	0.26	0.82	∞	∞
$B^*_{\mathbf{T}_3}$	c = .9	0.31	0.83	1.14	1.75
	$c = .8$	0.26	0.83	1.61	2.04
	$c = .7$	0.26	0.82	1.92	2.39
	$B^*_{G\tau}(\epsilon)$	0.32	0.95	1.60	2.15
	$B^*_{GM1}(\epsilon)$	0.37	0.75	2.03	5.99

Table 7.4: Maximum bias of several C-estimators that combines LMS and GM, of $G\tau$-estimator and of GM1-estimator for $p = 15$

p = 15		$\epsilon = .05$	$\epsilon = .10$	$\epsilon = .15$	$\epsilon = .20$
	$B_{\mathbf{T}_{LMS}}(\epsilon)$	0.53	0.83	1.14	1.52
	$B_{\mathbf{T}_{GM}}(\epsilon)$	0.33	1.29	∞	∞
$B^*_{\mathbf{T}_3}$	c = .9	0.34	0.98	1.26	1.75
	$c = .8$	0.33	1.28	1.61	2.04
	$c = .7$	0.33	1.28	1.92	2.39
	$B^*_{G\tau}(\epsilon)$	0.36	1.12	1.60	2.15
	$B^*_{GM1}(\epsilon)$	0.42	1.02	3.07	9.68

Table 7.5: Maximum bias of several C-estimators that combines LMS and GM, of $G\tau$-estimator and of GM1-estimator for $p = 20$

	p = 20	$\epsilon = .05$	$\epsilon = .10$	$\epsilon = .15$	$\epsilon = .20$
	$B_{\mathbf{T}_{LMS}}(\epsilon)$	0.53	0.83	1.14	1.52
	$B_{\mathbf{T}_{GM}}(\epsilon)$	0.39	2.36	∞	∞
	c = .9	0.39	1.04	1.36	1.75
$B^*_{\mathbf{T}_3}$	$c = .8$	0.39	1.28	1.61	2.04
	$c = .7$	0.39	1.28	1.92	2.39
	$B^*_{G\tau}(\epsilon)$	0.42	1.12	1.60	2.15
	$B^*_{GM1}(\epsilon)$	0.42	1.27	3.35	8.44

Since the values of ϵ considered are smaller than the asymptotic breakdown point of $\mathbf{T}_3$, $\epsilon^*_{\mathbf{T}_3} = 0.5$, then $B^*_{\mathbf{T}_3}$ will be finite in all cases. In addition, $B^*_{\mathbf{T}_3}$ is bounded between $B^*_{\mathbf{T}_{LMS}}$ and $B^*_{\mathbf{T}_{GM}}$, and approaches $B^*_{\mathbf{T}_{LMS}}$ as c increases.

Note that, for all p, there exists c such that the maximum bias for mass point contamination of $\mathbf{T}_3$ is less than that of the $G\tau$-estimator for all the values of ϵ. This value of c is highlighted in the tables. For large values of p, the bias of the estimator $\mathbf{T}_3$ corresponding to this choice of c is also less than $\mathbf{T}_{GM1}$, for all ϵ. When the number of regressors is small, $\mathbf{T}_3$ behaves somewhat worse than $\mathbf{T}_{GM1}$ for some values of ϵ, but it is much better for most values of ϵ.

7.2. Monte Carlo finite sample size results. In order to understand and compare the behavior of the C-estimators for finite sample size, we have run a simulation experiment. The estimates includes in our study are,

1. The least median of squares estimate (LMS).
2. The generalized τ estimates ($G\tau$) proposed Ferreti et al. (1994).
3. The one-step Newton-Raphson GM-estimates (GM1) introduced by Simpson et al. (1992).
4. The C-estimator of the form introduced in section 6 (C).

The $G\tau$ and GM1-estimators considered here are the same as in subsection 7.1.

As in the previous subsection, we consither the C-estimator that combines the LMS with the GM-estimator given by (11). LMS is computed by resampling with 1000 subsamples. The values of the constants are $c_1 = c_2 = 0.6$ for $p = 2$, $c_1 = 0.6$ and $c_2 = 0.8$ for $p = 5$, and $c_1 = 0.7$ and $c_2 = 0.9$ for $p = 10$, for all the sample sizes. We have analysed differents values of c_1 and c_2, but for shortness we only include the values that provide the best results.

Because of efficiency considerations, we also included the one-step reweighted least squares versions of the four estimates listed above (see Rousseeuw and Leroy, 1987). Given the initial estimate $\mathbf{T}$, the corresponding one-step reweighted estimate is a weighted LS with weights $w_i = w((y_i - \mathbf{T}'\mathbf{x}_i)/\hat{\sigma})$, where

$$\hat{\sigma} = 1.481 \text{ median } |y_i - \mathbf{T}'\mathbf{x}_i|.$$

Following Rousseeuw and Leroy. we use the "hard rejection" weight function $w(t) = I(|t| \leq a)$, with $a = 2.5$. The goal of this one-step reweighted least squares estimates is to gain efficiency under normal errors while preserving the robustness properties of the initial estimates. The reweighted estimates will be denoted by RLMS, RGM1 , RG$\mathcal{T}$ and RC.

We consider again a regression model without intercept. And we analize three different values of p (2, 5 and 10) and three sample sizes (n = 60, 100 and 140). The number of replications is 500. Each sample contains $n(1-\epsilon)$ observations $(y, \mathbf{x}')'$ from a $\mathrm{N}_{p+1}(0, I)$ and $n\epsilon$ identical observations equal to $(\tilde{y}, (10, 0, ..., 0))$ (ϵ ranges from 0 to .50 with increments of .05).

For each estimate and sampling situation, we compute the total mean squared error

$$MSE = \frac{1}{m}\left(\sum_{i=1}^{m}\sum_{j=1}^{p} T_{ij}^2\right),$$

where $\mathbf{T}_i = (T_{i1}, \ldots, T_{ip})$, $(i = 1, \ldots, m)$. The value of the "contaminating" slope, $sl = \tilde{y}/10$ is changed with increments of 0.05 in searching for the maximum value of the MSE. For shortness, we only report the overall maximum MSE for each estimate, p and ϵ. The results are given in Tables 7.6, 7.7 and 7.8.

From these tables we can see that there is not and overall best robust estimate.

In general, reweighting improves the performance of the estimates. Therefore, the following comments focus on the reweighted estimates RG$\mathcal{T}$, RGM1 and RC.

For $p = 2$ the best estimator is RGM1 for $\epsilon \leq 0.15$ if $n = 60$ and for $\epsilon \leq 0.10$ if $n = 100$ or $n = 140$. In the rest of the cases, the performance of the RC-estimator is better.

For $p = 5$, the RC-estimator is uniformily better than RG$\mathcal{T}$ and RGM1.

For $p = 10$, RG$\mathcal{T}$ has the best performance for $\epsilon \geq 0.10$ if $n = 60$, for $\epsilon \geq 0.15$ if $n = 100$, for $\epsilon = 0.20$ if $n = 140$, and RC is the best estimator in all the other situations.

Table 7.6: Simulation Results for $n = 60$. Maximum Mean Squared Errors

n	p	ϵ	LS	LMS	$G_\mathcal{T}$	GM1	C	RLMS	$RG_\mathcal{T}$	RGM1	RC
60	2	0.00	0.036	0.186	0.075	0.074	0.087	0.060	0.043	0.042	0.044
		0.05		0.257	0.097	0.101	0.107	0.149	0.086	0.077	0.105
		0.10		0.458	0.195	0.198	0.232	0.344	0.185	0.132	0.210
		0.15		0.825	0.502	0.488	0.467	0.712	0.477	0.316	0.448
		0.20		1.580	2.232	1.305	0.957	1.424	2.010	1.175	0.972
	5	0.00	0.097	0.380	0.157	0.148	0.119	0.183	0.117	0.119	0.117
		0.05		0.535	0.245	0.243	0.207	0.350	0.207	0.215	0.205
		0.10		1.030	0.798	0.753	0.435	0.844	0.705	0.706	0.414
		0.15		2.298	2.057	2.359	1.186	2.036	1.916	1.942	1.216
		0.20		5.193	4.602	9.105	2.916	4.926	4.283	6.942	3.356
	10	0.00	0.205	0.649	0.349	0.382	0.306	0.493	0.316	0.295	0.301
		0.05		1.007	0.648	0.716	0.503	0.853	0.606	0.562	0.511
		0.10		2.178	1.495	3.271	1.433	2.031	1.452	1.817	1.711
		0.15		5.432	3.598	21.592	30972	5.241	3.467	9.122	6.458
		0.20		16.034	9.830	145.585	17.841	15.835	9.679	73.092	16.451

Table 7.7: Simulation Results for $n = 100$. Maximum Mean Squared Errors

n	p	ϵ	LS	LMS	$G_\mathcal{T}$	GM1	C	RLMS	$RG_\mathcal{T}$	RGM1	RC
100	2	0.00	0.022	0.126	0.043	0.044	0.052	0.031	0.025	0.025	0.028
		0.05		0.220	0.067	0.068	0.091	0.123	0.076	0.060	0.091
		0.10		0.452	0.170	0.160	0.224	0.340	0.189	0.122	0.199
		0.15		0.851	0.492	0.436	0.471	0.712	0.524	0.320	0.304
		0.20		1.628	2.368	1.185	0.947	1.440	2.256	1.270	0.914
	5	0.00	0.055	0.370	0.084	0.083	1.012	0.089	0.065	0.065	0.062
		0.05		0.525	0.153	0.156	0.139	0.232	0.149	0.157	0.134
		0.10		0.982	0.649	0.551	0.518	0.596	0.591	0.537	0.433
		0.15		1.837	1.797	1.782	1.852	1.323	1.713	1.452	1.391
		0.20		3.600	3.642	6.682	3.775	2.885	3.380	5.315	3.144
	10	0.00	0.118	0.878	0.165	0.337	0.193	0.208	0.146	0.142	0.152
		0.05		1.226	0.388	0.560	0.562	0.445	0.362	0.369	0.309
		0.10		2.152	0.996	2.431	2.251	1.092	0.971	1.468	0.823
		0.15		4.153	2.091	14.112	5.316	2.691	1.986	7.754	2.368
		0.20		9.368	4.636	89.347	12.323	7.044	4.298	46.765	8.072

Table 7.8: Simulation Results for $n = 140$. Maximum Mean Squared Errors

n	p	ϵ	LS	LMS	$G_{\mathcal{T}}$	GM1	C	RLMS	R$G_{\mathcal{T}}$	RGM1	RC
140	2	0.00	0.015	0.087	0.029	0.026	0.032	0.020	0.017	0.017	0.017
		0.05		0.182	0.055	0.051	0.051	0.113	0.069	0.055	0.059
		0.10		0.412	0.147	0.135	0.124	0.349	0.176	0.115	0.141
		0.15		0.829	0.458	0.379	0.317	0.762	0.526	0.330	0.306
		0.20		1.650	2.362	1.087	0.852	1.538	2.225	1.336	0.862
	5	0.00	0.037	0.100	0.058	0.055	0.059	0.048	0.043	0.043	0.043
		0.05		0.229	0.123	0.116	0.116	0.171	0.131	0.144	0.121
		0.10		0.619	0.614	0.510	0.427	0.553	0.576	0.531	0.479
		0.15		1.238	1.817	1.579	1.396	1.160	1.747	1.419	1.431
		0.20		2.628	3.606	5.955	3.144	2.484	3.368	4.890	2.892
	10	0.00	0.078	0.124	0.106	0.111	0.125	0.098	0.092	0.093	0.093
		0.05		0.281	0.306	0.289	0.273	0.243	0.296	0.264	0.245
		0.10		0.756	0.913	1.636	0.896	0.712	0.893	1.322	0.789
		0.15		1.601	1.831	8.359	1.814	1.564	1.770	6.043	1.678
		0.20		3.412	3.635	39.204	3.612	3.349	3.348	29.426	3.406

Therefore, the performance of C-estimators is very good for $p = 5$ and for other values of p becomes better than $G_{\mathcal{T}}$ and GM1 as ϵ increase.

8. A C-estimator as efficient as the LSE. In this section we propose an estimator that, although is not locally robust, has very interesting properties.

The most widely used estimator in regression is the Least Squares Estimator (LSE), which is defined by the functional

$$\mathrm{T}_{LSE}(H) = \arg\min_{\theta \in \mathbf{R}^p} E_H\left[(y - \theta'\mathrm{x})^2\right]. \tag{12}$$

It is known that when the distribution H is normal the LSE is efficient, since its covariance matrix attains the Rao-Cramer bound matrix.

However $\mathbf{T}_{LS}$ is neither globally nor locally robust. In fact, $B_{\mathbf{T}}(\epsilon) = \infty$ for all $\epsilon > 0$, and therefore its breakdown point is zero and it is not locally stable for any $q > 0$.

We now analyze the properties of the C-estimator with $\mathbf{T}_2$ as the LSE. By results in Section 5, it follows that the asymptotic covariance matrix of $\mathbf{T}_3$ coincides with that of $\mathbf{T}_2$, and then, the assymptotic efficiency of this $C-$estimator under normality is optimal.

Consequently, the C-estimator constructed with the LSE has the following properties:

- Its asymptotic break down point is $\epsilon^* = 0.5$.

- Under the uncontaminated model, it is $n^{1/2}$ consistent and asymptotically normal.

- If data are normally distributed, it has the same efficiency as LSE.

In addition, this C-estimator can be computed in S-Plus, since this program calculates the LMS and LS-estimators. So the C estimator that we propose in this section is very easy to compute.

Obviously, this estimator will not be locally robust, given that the LSE is not locally stable of any order $q > 0$. But it has maximum breakdown point and its efficiency is optimal under normal errors.

We have calculated the maximum bias for mass point contaminated distribution of this estimator for several values of ϵ and $c_1 = c_2 = c$. In this case, the bias does not depend on the number of regressors. Since $B_{\mathbf{T}_2}(\epsilon) = 0$ for all $\epsilon > 0$, it is clear that will reduce the maximum bias taking c close to one. Note that for large c, $B_{\mathbf{T}_3}(\epsilon)$ is very close to $B_{\mathbf{T}_{LMS}}(\epsilon)$.

The calculations are in Table 8.1.

Table 8.1: Maximum bias of the C-estimator that combines LMS and LS

	Contaminación	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.15$	$\epsilon = 0.20$
	$B_{\mathbf{T}_{LMS}}(\epsilon)$	0.53	0.83	1.14	1.52
$B^*_{\mathbf{T}_C}(\epsilon)$	$c = 0.99$	0.55	0.85	1.16	1.54
	$c = 0.95$	0.64	0.93	1.24	1.63
	$c = 0.9$	0.76	1.04	1.36	1.75
	$c = 0.8$	0.99	1.28	1.61	2.04
	$c = 0.7$	1.27	1.56	1.92	2.39
	$c = 0.6$	1.59	1.92	2.32	2.85
	$c = 0.5$	2.03	2.39	2.86	3.49
	$c = 0.4$	2.64	3.09	3.66	4.43
	$c = 0.3$	3.63	4.21	4.95	5.97

References

Davies P.J. (1990). The asmyptotics of S-estimates in the linear regression model, *Ann. Statist.*,**18**, 1651-1675.

Donoho, D.L. (1982). Breakdown properties of multivariate location estimators. Unpublished manuscript. Harvard University. Ph. D. qualifying paper.

Donoho, D.L. and Huber P.J.(1983). The notion of breakdown point. *Festschrift for Erich Lehmann.* P.J. Bickel, K.Docksum y J.L.Hodges, Jr.,eds.,Wadsworth, Belmont, 157-184.

Ferretti, N., Kelmansky, D., Yohai, V. J. and Zamar, R. H.(1994). Generalized τ–estimates of regression. Notas de Matemática No. 55, Departamento de Matematicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata.

Hampel, F. R. (1971). A general qualitative definition of robustness, *Ann. Math. Statist.*, **42**, 1887–1896.

He, X. y Simpson, D.G. (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *The Annals of Statistics,* **21**, 314–337.

Maronna, R.A., Bustos, O.H. and Yohai, V.J. (1979). Bias– and efficiency–robustness of general M–estimators for regression with random carriers. In *Smoothing Techniques for Curve Estimation,* T. Gasser and M. Rosenblatt (eds.). Lecture Notes in Mathematics **757**, 91–116, Springer–Verlag, New York.

Maronna, R.A., Stahel, W.A. and Yohai, V.J. (1992). Bias–robust estimators of multivariate scatter based on projections. *Journal of Multivariate Analysis,* **42**, 141–161.

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

Martin, R.D., Yohai, V.J. and Zamar, R.H. (1989). Min–max bias robust regression. *The Annals of Statistics,* **17**, 1608–1630.

Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association,* **79**, 871–880.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*, Wiley, New York.

Simpson, D. G., Ruppert, D. y Carroll, R.J. (1992). On one–step GM–estimates and stability of inferences in linear regression. *Journal of the American Statistical Association,* **87**, 439–450.

Simpson, D. G. and Yohai, V. J. (1993). Functional stability of one–step GM–estimators in linear regression. Technical Report #71, Department of Statistics, University of Illinois, Urbana-Champaign.

Stahel, W. A. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen. Ph. D. dissertation, ETH, Zürich.

Yohai, V. y Zamar, R. H (1992). Optimally bounding the GES of unbounded influence estimates of regression. Working paper 92–44. Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain.