

## Tesis de Maestría

# Algoritmos rápidos para computar estimadores robustos

Ambrosio, Beatriz

2004

Tesis presentada para obtener el grado de Magister de la Universidad de Buenos Aires en el área de Estadística Matemática de la Universidad de Buenos Aires

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the Master's and Doctoral Theses Collection of the Central Library Dr. Luis Federico Leloir, available in [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Ambrosio, Beatriz. (2004). Algoritmos rápidos para computar estimadores robustos. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.  
[http://hdl.handle.net/20.500.12110/tesis\\_n3765\\_Ambrosio](http://hdl.handle.net/20.500.12110/tesis_n3765_Ambrosio)

Cita tipo Chicago:

Ambrosio, Beatriz. "Algoritmos rápidos para computar estimadores robustos". Tesis de Magister. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2004.  
[http://hdl.handle.net/20.500.12110/tesis\\_n3765\\_Ambrosio](http://hdl.handle.net/20.500.12110/tesis_n3765_Ambrosio)

**EXACTAS** UBA

Facultad de Ciencias Exactas y Naturales



**UBA**

Universidad de Buenos Aires



UNIVERSIDAD NACIONAL DE BUENOS AIRES  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

ALGORITMOS RAPIDOS PARA  
COMPUTAR  
ESTIMADORES ROBUSTOS

AUTOR: BEATRIZ AMBROSIO

LUGAR DE TRABAJO: UNIVERSIDAD DE BUENOS AIRES

DIRECTOR: DR. VICTOR YOHAI

TESIS PARA OPTAR AL TITULO DE  
MAGISTER EN ESTADISTICA MATEMATICA

3765

Noviembre de 2004

## INDICE

RESUMEN.....	1
ABSTRACT.....	3
INTRODUCCION.....	5
1. ESTIMADOR DE MINIMOS CUADRADOS.....	7
2. MEDIDAS DE ROBUSTEZ.....	10
3. M-ESTIMADORES.....	13
4.1. PROPIEDADES ASINTOTICAS DE LOS M-ESTIMADORES.....	15
5. ESTIMADOR DE MINIMA MEDIANA DE CUADRADOS.....	21
6. ESTIMADORES DE MINIMOS CUADRADOS PODADOS.....	22
7. S-ESTIMADORES.....	24
7.1. PRPIEDADES ASINTOTICAS DEL S-ESTIMADOR.....	29
7.2. PUNTO DE RUPTURA DE LOS S-ESTIMADORES.....	31
8. MM-ESTIMADORES.....	40
8.1. PROPIEDADES ASINTOTICAS DE LOS MM-ESTIMADORES.....	42
9. ALGORITMOS PARA COMPUTAR ESTIMADORES ROBUSTOS.....	43
9.1. FAST-LTS.....	43
9.2. FAST-S.....	46
10. EJEMPLO.....	50
11. SIMULACIONES.....	53
12. TABLAS.....	56
13. REFERENCIAS.....	58

## RESUMEN

El método más comunmente usado para estimar los coeficientes de una regresión lineal es el de mínimos cuadrados. Este método que es óptimo en el caso de errores distribuidos normalmente, es muy sensible a la presencia de outliers. Para remediar ese problema se han desarrollado otros métodos de estimación llamados métodos robustos, los cuales se ven poco afectados por la presencia de datos atípicos.

Una medida de la robustez de un estimador es su punto de ruptura. Entre los estimadores de regresión robustos se encuentran los LTS, LMS y los S estimadores. Estos estimadores son equivariantes por transformaciones afines, de regresión, y de escala y además tienen un alto punto de ruptura. El inconveniente que presenta su calculo es que requiere muchas horas de computadora.

Rouseeuw y Van Driessen desarrollaron un algoritmo, llamado Fast-LTS, que mejora la velocidad de cálculo del estimador de mínimos cuadrados podados. El objetivo de este trabajo es desarrollar un nuevo algoritmo, análogo al Fast-LTS, para computar S estimadores. Del mismo modo que el algoritmo desarrollado por Rousseeuw, este nuevo algoritmo está basado en el mejoramiento local de los estimadores iniciales. Esto permite una significativa reduccion del número de candidatos requeridos para obtener una buena aproximacion de la solución óptima.

Se ha realizado un estudio de simulación que ha mostrado que los S estimadores calculados con el algoritmo Fast-S, son comparativamente mejores a los estimadores LTS calculados con el algoritmo Fast-LTS. Algunas de las ventajas del nuevo algoritmo son: 1) Menor porcentaje de muestras afectadas por los datos atípicos. 2) Menor error cuadrático medio. 3) Menor tiempo de cómputo.

## ABSTRACT

Regression analysis is an important statistical tool that is applied in most sciences. The purpose of regression analysis is to fit equations to observed variables. The most commonly regression technique is the least squares method, generally adopted because of tradition and easier computation. However these method is very sensitive to the presence of atypical points in the sample. An observation is an atypical point or outlier if it does not follow the model. To remedy this problem the robust methods have been developed that are not so easily affected by outliers.

One mesure of robustness of an estimate is its breakdown point. Heuristically, the breakdown point is the minimum fraction of arbitrary outliers that can take the estimate beyond any limit. The breakdown point as an asymptotic concept has been introduced by Hampel (1971). Donoho and Huber (1983) gave the corresponding finite sample notion.

A desirable property for regression estimates is that the estimate be equivariant with respect to affine, regression and scale transformations. Estimates with this property are the LTS-estimate and S-estimate. This estimates are computationally expensive, and the corresponding algorithms become unfeasible for moder-

ately large number of regressors.

In this thesis we propose an algorithm for computing S-estimates, analogous to the algorithm of Rousseeuw and Van Driessen (fast-LTS) to improve the computational speed of the LTS-estimate . The new algorithm, that we call "fast-S", is based on a "local improvement"-step of the re-sampling initial candidates. This allows a substantial reduction of the number of candidates required to obtain a good approximation to the optimal solution. We performed a simulation study wich shows that S-estimators computed with the fast-S algorithm compare favourably to the LTS estimator computed with the fast-LTS algorithm.

## INTRODUCCION

El análisis de regresión es una herramienta estadística de importancia, con numerosas aplicaciones en múltiples campos del conocimiento humano, tales como la medicina, la economía, la sociología, la psicología y la industria, entre otros. El análisis de regresión se utiliza para modelar la relación existente entre un conjunto de variables. Uno de los posibles modelos dentro del análisis de regresión es el modelo lineal. Este modelo supone que hay una relación *lineal* entre una variable dependiente  $Y$ , llamada variable respuesta y un conjunto  $X_1, X_2, \dots, X_p$  de variables explicativas o regresoras. Más precisamente, supondremos:

$$Y = \theta_1 X_1 + \dots + \theta_p X_p + U \quad (1.1)$$

donde  $U$  es el error que no puede ser explicado por las variables  $X_1, X_2, \dots, X_p$ . El vector  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$  es el vector de los coeficientes de la regresión. El método más frecuentemente utilizado para estimar estos coeficientes es el de *mínimos cuadrados* (MC). Este método ha sido usado tradicionalmente debido a que es óptimo en el caso de que los errores tengan distribución normal y además es simple de computar. Sin embargo tiene el inconveniente de ser muy sensible a la presencia de outliers (observaciones atípicas). Para superar este problema se han



desarrollado los métodos de estimación denominados *robustos*. Estos métodos se ven poco afectados por la presencia de valores atípicos.

El objetivo de este trabajo es desarrollar un nuevo algoritmo para computar un estimador robusto (S-estimador) basado en muestreo de conjuntos elementales y su comparación con otros algoritmos.

## 2. ESTIMADOR DE MINIMOS CUADRADOS (MC)

Sea

$$Y = \theta_1 X_1 + \dots + \theta_p X_p + U$$

un modelo de regresión lineal, y sea  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$  un estimador de  $\theta = (\theta_1, \dots, \theta_p)$  basado en  $n$  observaciones  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ , donde  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ ,  $\mathbf{y} = (y_1, \dots, y_n)$ ,  $\mathbf{U} = (u_1, \dots, u_n)'$ . La mayoría de las aplicaciones de regresión lineal incluyen una constante, en ese caso se toma  $x_{i1} = 1$  para todo  $i$ . Como el modelo se satisface para todas las observaciones, podemos escribir:

$$y_i = \theta_1 x_{i1} + \dots + \theta_p x_{ip} + u_i, \quad 1 \leq i \leq n. \quad (2.1)$$

Denotaremos  $\hat{y}_i(\theta)$  al valor predicho de  $y_i$  cuando los coeficientes de regresión están dados por  $\theta$

$$\hat{y}_i(\theta) = \theta_1 x_{i1} + \dots + \theta_p x_{ip}, \quad (2.2)$$

y denominaremos residuos  $r_i(\theta)$ ,  $1 \leq i \leq n$ , a la diferencia entre el valor observado y el valor predicho, es decir

$$r_i(\boldsymbol{\theta}) = y_i - \hat{y}_i(\boldsymbol{\theta}), 1 \leq i \leq n.$$

El estimador de mínimos cuadrados (MC) se define como el valor de  $\boldsymbol{\theta}$  que minimiza la suma de los cuadrados de los residuos. Luego, se define como

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^n r_i^2(\boldsymbol{\theta}). \quad (2.3)$$

Este estimador fue propuesto por Gauss (1801). El sistema (2.1) se puede expresar en forma matricial como

$$\mathbf{y} = X\boldsymbol{\theta} + \mathbf{u},$$

donde  $X$  es la matriz de  $n \times p$ , cuyas filas son  $\mathbf{x}'_i = (x_{i1}, \dots, x_{ip})$ ,  $\mathbf{y} = (y_1, \dots, y_n)'$  y  $\mathbf{u} = (u_1, \dots, u_n)'$ . Entonces si el rango de la matriz  $X$  es  $p$ , el estimador MC está dado por

$$\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'\mathbf{y},$$

siendo  $X'$  la matriz transpuesta de  $X$ .

Si los errores  $u_i$  son todos independientes y tienen distribución normal con media cero y la misma varianza, entonces el estimador MC tiene la propiedad de ser el estimador insesgado de mínima varianza. Sin embargo el estimador MC es muy

sensible a pequeñas desviaciones de normalidad de los errores, como también a la presencia en la muestra de algunas pocas observaciones atípicas. A las observaciones atípicas se las llama *outliers*. Dos outliers con el mismo residuo pueden tener diferente influencia sobre los estimadores MC, ya que esa influencia depende también de cuán alejado esté el valor  $x$  del outlier, del vector de medias (leverage).

### 3. MEDIDAS DE ROBUSTEZ

Los estimadores robustos son aquellos que tienen la propiedad de ser poco sensibles a la presencia de outliers en la muestra. Por lo tanto el estimador MC no es robusto, ya que un solo outlier puede tener sobre el mismo una influencia tan grande como se quiera.

Para estudiar la robustez de un estimador será conveniente definir el  $\varepsilon$ -entorno de la distribución teórica  $F_0$  del vector  $(y, x_1, \dots, x_p)$ . Este modelo asume que una proporción  $(1 - \varepsilon)$  de las observaciones proviene de la distribución  $F_0$  y el  $\varepsilon$  por ciento restante corresponde a una distribución desconocida  $F_1$ . Luego el entorno de contaminación de  $F_0$  de tamaño  $\varepsilon$  está dado por

$$V_{F,\varepsilon} = \{F \mid (1 - \varepsilon)F_0 + \varepsilon F_1, F_1 \text{ arbitraria}\}.$$

Se han propuesto distintos indicadores para medir la robustez de un estimador.

Se pueden clasificar en:

1. **Medidas para determinar la robustez local.** Determinan la robustez de un estimador frente a la presencia de una proporción infinitesimal de outliers. (Curva de influencia, Sensibilidad a errores groseros, medidas introducidas por Hampel (1971)).

2. **Medidas de robustez global.** Determinan la robustez de un estimador frente a una proporción no infinitesimal de outliers. Una medida de robustez global es el punto de ruptura. Esta medida fue introducida por Hampel (1971) con carácter asintótico, y luego Donoho y Huber (1983) definieron una versión basada en muestras finitas.

Vamos a definir el punto de ruptura basado en muestras finitas de un estimador  $\hat{\theta}$ . Consideramos una muestra

$$Z = \{(y_1, x_{11}, \dots, x_{1p}), \dots, (y_n, x_{n1}, \dots, x_{np})\}$$

y la familia  $\mathcal{Z}$  de todas las muestras que se obtienen al reemplazar  $m$  puntos originales de  $Z$ , por valores arbitrarios (outliers). Se define máximo sesgo producido por  $m$  outliers como

$$S_m(Z) = \sup_{Z^* \in \mathcal{Z}} \|\hat{\theta}(Z^*) - \hat{\theta}(Z)\|.$$

Por lo tanto  $S_m(Z)$  es la máxima variación que puede experimentar el estimador  $\hat{\theta}$  cuando la muestra se contamina con  $m$  outliers.

Sea

$$m^* = \min\{m : S_m(Z) = \infty\},$$

es decir  $m^*$  es la mínima cantidad de outliers que puede provocar que el sesgo del estimador sea mayor que cualquier límite. Entonces el punto de ruptura para una muestra finita se define por

$$\varepsilon_n^* = \frac{m^*}{n}.$$

Por lo tanto si la proporción de contaminación  $\varepsilon$  cumple que  $\varepsilon < \varepsilon_n^*$ , el estimador es informativo y deja de serlo si  $\varepsilon \geq \varepsilon_n^*$ . En general el punto de ruptura de un estimador es menor o igual que  $\frac{1}{2}$ , ya que si la muestra tiene más de un 50 % de outliers, es imposible saber cuáles son las observaciones "buenas" y cuáles los outliers.

## 4. M-ESTIMADORES

Los M-estimadores de regresión fueron propuestos por Huber (1973). Se definen reemplazando la función cuadrática utilizada en (2.4) por otra función  $\rho$  que satisfaga las siguientes propiedades:

**P1.** Si  $0 \leq u \leq v \Rightarrow \rho(u) \leq \rho(v)$  ( $\rho$  es una función monótona no decreciente para los valores positivos de la variable).

**P2.**  $\rho(u) = \rho(-u)$  (es una función par).

**P3.**  $\rho(0) = 0$ .

**P4.** La función  $\rho$  es derivable (notaremos  $\rho' = \Psi$ ).

Para que el M-estimador resulte robusto frente a cualquier tipo de outliers, se necesita agregar la siguiente propiedad a la función  $\rho$ .

**P5**  $\rho$  debe ser acotada.

Diremos entonces que  $\hat{\theta}$  es un M-estimador si minimiza:

$$S(\theta) = \sum_{i=1}^n \rho \left( \frac{r_i(\theta)}{s} \right), \quad (4.1)$$

donde  $s$  es un estimador de la escala de los errores. Es decir:

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \rho \left( \frac{r_i(\theta)}{s} \right).$$



Diferenciando (4.1) con respecto a cada  $\theta_i$  se tiene el siguiente sistema de  $p$  ecuaciones

$$\sum_{i=1}^n \psi \left( \frac{r_i(\boldsymbol{\theta})}{s} \right) \mathbf{x}_i = \mathbf{0}, \quad (4.2)$$

siendo  $\psi = \rho'$ ,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ , y  $\mathbf{0} = (0, \dots, 0)$ .

La escala,  $s$ , se puede estimar separada o simultáneamente con  $\boldsymbol{\theta}$ . En este último caso se puede agregar una ecuación de la forma

$$\frac{1}{n} \sum_{i=1}^n \chi \left( \frac{r_i(\boldsymbol{\theta})}{s} \right) = b,$$

donde la función  $\chi$  cumple las propiedades **P1a** **P5** enunciadas anteriormente. Si se quiere que el estimador de escala estime la desviación típica cuando los errores son normales el número  $b$  debe ser elegido como

$$b = E\phi(\chi(u))$$

siendo  $\phi$  la distribución  $\mathcal{N}(0, 1)$  (distribución normal con media 0 y varianza 1).

Se pueden distinguir dos tipos distintos de funciones  $\rho$ : funciones  $\rho$  acotadas y funciones  $\rho$  convexas. Los M-estimadores con una función  $\rho$  convexa, son robustos solamente frente a outliers que no tienen alto leverage, en cambio los M-

estimadores con una función  $\rho$  acotada son robustos frente a cualquier tipo de outliers, pero como contrapartida su computación es mucho más compleja.

Para obtener algún grado de robustez, es necesario que la función  $\rho$  crezca más lentamente que la función cuadrática, de manera de dar menos peso a las observaciones atípicas.

Una de las posibles elecciones de una función  $\rho$  convexa es considerar la familia de funciones propuesta por Huber (1964), definida por

$$\rho_c^H = \begin{cases} -cu - \frac{c^2}{2} & \text{si } u < -c \\ \frac{u^2}{2} & \text{si } |u| \leq c \\ cu - \frac{c^2}{2} & \text{si } u > c. \end{cases}$$

El valor de  $c$  se puede elegir de modo que se obtenga la eficiencia deseada para el caso que los errores son normales. La derivada de  $\rho_c^H$  está definida por

$$\psi_c^H = \text{sgn}(u) \min\{|u|, c\}.$$

#### 4.1. PROPIEDADES ASINTÓTICAS DE LOS M-ESTIMADORES

El propósito de la teoría asintótica es estudiar la validez aproximada de los procedimientos de inferencia estadística cuando el número de observaciones es grande.

Las tres propiedades más importantes que estudia la teoría asintótica de una sucesión de estimadores  $\{\hat{\theta}_n\}$  de  $\theta$  son las siguientes:

### 1.- Consistencia

(i) Una sucesión  $\{\hat{\theta}_n\}$  de estimadores de  $\theta$  es fuertemente consistente si converge en casi todo punto a  $\theta$   $P_{\theta}\{\hat{\theta}_n \rightarrow \theta\} = 1$

(ii) Una sucesión  $\{\hat{\theta}_n\}$  de estimadores de  $\theta$  es débilmente consistente si converge en probabilidad a  $\theta$ . O sea si  $\lim_{n \rightarrow \infty} P_{\theta}(|\hat{\theta}_n - \theta| > \varepsilon) = 0$  para todo  $\varepsilon > 0$ .

### 2.- Normalidad asintótica.

Sea  $a_n$  una sucesión de números reales positivos. Se dice que una sucesión  $\hat{\theta}_n, n \geq 1$  es asintóticamente normal con media  $\theta$ , matriz de covarianza  $V(\theta)$  y orden  $a_n$  si

$$a_n(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{V}(\theta)),$$

donde  $\mathcal{N}_p(\mu, \Sigma)$  representa la distribución normal multivariada con media  $\mu$  y matriz de covarianza  $\Sigma$ .

### 3.- Eficiencia relativa

Sean dos sucesiones  $\{\hat{\theta}_n^1\}$  y  $\{\hat{\theta}_n^2\}$  de estimadores asintóticamente normales con el mismo orden  $a_n, n \geq 1$ , misma media  $\theta$ , y matrices de covarianza asintóticas

$v_1(\theta)V(\theta)$  y  $v_2(\theta)V(\theta)$  respectivamente, donde  $v_1(\theta)$  y  $v_2(\theta)$  son escalares. Luego la eficiencia relativa del primero respecto del segundo como estimadores de  $\theta$  está dada por  $v_2(\theta)/v_1(\theta)$ .

Las siguientes hipótesis son necesarias para la derivación de los teoremas de normalidad asintótica en el caso de regresores fijos.

**H1.**  $\mathbf{z}_1 = (y_1, \mathbf{x}_1), \dots, \mathbf{z}_n = (y_n, \mathbf{x}_n)$  satisfacen (2.1).

**H2.** Los errores  $u_1, \dots, u_n$  son independientes y tienen todos la misma distribución  $F$ .

**H3.** La distribución de los errores  $F$  es simétrica..

**H4.** La distribución  $F$  tiene una densidad  $f = F'$  unimodal.

**H5.**  $\psi$  tiene dos derivadas continuas y acotadas

**H6.** Existe una sucesión  $a_n$ , y una matriz  $\Sigma$  definida positiva tal que  $a_n \rightarrow \infty$

y tal que

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'}{a_n^2} = \Sigma.$$

Además se cumple la siguiente condición de Lindenberg

$$\lim_{n \rightarrow \infty} \frac{\max_{1 \leq i \leq n} x_{i,j}^2}{\sum_{i=1}^n x_{i,j}^2} = 0, 1 \leq j \leq p.$$

Sea  $\hat{\boldsymbol{\theta}}$  un M-estimador consistente de  $\boldsymbol{\theta}$  tal que satisface

$$\sum_{i=1}^n \psi \left( \frac{r_i(\boldsymbol{\theta})}{s} \right) \mathbf{x}_i = \mathbf{0}, \quad (4.3)$$

donde  $s$  converge en probabilidad a  $\sigma$ . Luego si se satisfacen **H1-H6** se cumplirá

$$a_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, V(\psi, F, \sigma) \Sigma^{-1}). \quad (4.4)$$

donde  $\xrightarrow{D}$  significa convergencia en distribución y  $V(\psi, F, \sigma)$  es un escalar definido por

$$V(\psi, F, \sigma) = \sigma^2 \frac{E_F[\psi^2(u/\sigma)]}{E^2[\psi'(u/\sigma)]}. \quad (4.5)$$

En el caso de  $\mathbf{x}_i$  aleatorios **H3** debe ser reemplazada por

**H3\*** : Los vectores  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$  son independientes e idénticamente distribuidos con distribución  $F$  y  $u_i$  es independiente de  $\mathbf{x}_i$ .

Además **H6** se debe reemplazar por :

**H6\*** : El vector  $\mathbf{x}$  tiene segundos momentos. La matriz de segundos momentos se simbolizará por  $\Sigma = \mathbf{E}(\mathbf{x}_i \mathbf{x}_i')$ .

En este caso se tendrá

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, V(\psi, F, \sigma^2) \Sigma^{-1})$$

En el caso de  $\rho$  convexa la hipótesis **H4** no es necesaria. Solamente se requiere cuando  $\rho$  es acotada.

La matriz de covarianza asintótica de un M-estimador, tanto en el caso de  $\mathbf{x}_i$  fijas como aleatorias, puede ser estimada por

$$\hat{\Sigma}_{\hat{\boldsymbol{\theta}}} = V(\psi, V(\psi, F_n, s^2)(\hat{\Sigma})^{-1}.$$

donde  $\hat{\Sigma} = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' / a_n$  y  $F_n$  es la distribución empírica de los residuos  $\hat{r}_i = y_i - \hat{\boldsymbol{\theta}} \mathbf{x}_i$  con  $1 \leq i \leq n$  siendo

$$V(\psi, F_n, s^2) = s^2 \frac{\sum_{i=1}^n \psi^2(r_i(\hat{\boldsymbol{\theta}})) / n}{\left(\sum_{i=1}^n \psi'(r_i(\hat{\boldsymbol{\theta}})) / n\right)^2}. \quad (4.6)$$

Utilizando el estimador  $\hat{\Sigma}_{\hat{\boldsymbol{\theta}}}$  se pueden encontrar intervalos de confianza aproximados para cualquier combinación lineal del vector  $\boldsymbol{\theta}$ , como así también obtener tests de hipótesis aproximados para cualquier hipótesis lineal.

Teoremas de consistencia para el caso de M-estimadores con función  $\rho$  convexa se pueden encontrar en Huber (1973) y Yohai (1973). Teorema de consistencia para M-estimadores con función  $\rho$  acotada pueden encontrarse en Klein y Yohai (1981), Bickel (1975), Yohai y Maronna (1979) y Yohai (1987). Un estimador más robusto de la matriz de covarianza asintótica de un M-estimador se puede encontrar en Stahel, Yohai y Zamar (1990)

## 5. ESTIMADOR DE MINIMA MEDIANA DE CUADRADOS

El estimador de mínima mediana de cuadrados (LMS- estimador) fue propuesto por Rousseeuw (1984).

Se define el LMS-estimador como

$$\arg \min_{\theta} \text{median} \{r_i^2(\theta) : 1 \leq i \leq n\}$$

El estimador LMS tiene la propiedad de ser regresión, afín y escala equivariante.

Un estimador  $T$  es regresión equivariante si

$T(\{\mathbf{x}_i, y_i + \mathbf{x}_i v\}; i = 1, \dots, n) = T(\{\mathbf{x}_i, y_i\}; i = 1, \dots, n) + \mathbf{v}$ , donde  $\mathbf{v}$  es un vector columna cualquiera.

Un estimador  $T$  es escala equivariante si

$$T(\{\mathbf{x}_i, cy_i\}; i = 1, \dots, n) = cT(\{\mathbf{x}_i, y_i\}; i = 1, \dots, n).$$

**Definición:** Se dice que los vectores  $\mathbf{x}_1, \dots, \mathbf{x}_n$  correspondientes a las variables explicativas de las observaciones, están en "posición general" si cualquier hiperplano en  $\mathbb{R}^p$  de dimensión  $p - 1$  contiene a lo sumo  $p$  de estos vectores. Es decir si  $\mathbf{a} \in \mathbb{R}^p$  el número de  $\mathbf{x}_i$  que satisfacen la ecuación  $\mathbf{a}'\mathbf{x}_i = 0$ , es a lo sumo  $p$ .



Se puede demostrar que si los puntos están en posición general, el LMS-estimador tiene punto de ruptura para muestras finitas igual a

$$\frac{\lfloor \frac{n}{2} \rfloor - p + 2}{n}$$

Un inconveniente del LMS-estimador es que es muy ineficiente. En efecto, su tasa de convergencia es de orden  $n^{\frac{1}{3}}$  en vez de la usual  $n^{\frac{1}{2}}$ .

## 6. ESTIMADORES DE MINIMOS CUADRADOS PODADOS

Los estimadores de mínimos cuadrados podados (LTS-estimadores) fueron propuestos por Rousseeuw (1984). El estimador LTS es el que minimiza la suma de los cuadrados de los  $h$  residuos con valores absolutos más pequeños, donde  $h$  es un valor que está entre  $n/2$  y  $n$ . Más precisamente, sean  $r_{(1)}^2(\boldsymbol{\theta}) \leq \dots \leq r_{(n)}^2(\boldsymbol{\theta})$  los valores que resultan de ordenar en forma creciente los residuos  $r_i^2(\boldsymbol{\theta})$ ,  $1 \leq i \leq n$ , y sea

$$Q_h(\boldsymbol{\theta}) = \sum_{i=1}^h r_{(i)}^2(\boldsymbol{\theta})$$

luego el LTS se define por:

$$\hat{\theta} = \arg \min_{\theta} Q_h(\theta)$$

Esta definición es muy parecida a la del estimador MC, la diferencia radica en que en esta sumatoria no se toman todos los residuos al cuadrado sino solamente los  $h$  más pequeños, lo que hace que el estimador sea poco influenciado por los outliers.

Al igual que el LMS, este estimador es regresión, afín y escala equivariante.

Si los puntos están en posición general, alcanza el máximo punto de ruptura para muestras finitas cuando

$$h = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{p+1}{2} \right\rfloor$$

y en este caso  $\varepsilon_n^* = \frac{\left\lfloor \frac{n-p}{2} \right\rfloor + 1}{n}$

El estimador LTS es también asintóticamente normal:

$$n^{1/2}(\hat{\theta} - \theta) \rightarrow \mathcal{N}(0, V(LTS, F))$$

donde  $V(LTS, F)$  es su varianza asintótica.

## 7. S-ESTIMADORES

Los estimadores basados en una M-escala (S-estimadores) fueron introducidos por Rousseeuw y Yohai (1984).

Dada una muestra  $u_1, \dots, u_n$  un estimador de escala  $s(u_1, \dots, u_n)$  mide el grandor en valor absoluto de las observaciones. Debe cumplir las siguientes propiedades

1.  $s(u_1, \dots, u_n) \geq 0$ .
2. Invarianza por cambio de escala. Es decir para todo  $\lambda$ .

$$s(\lambda u_1, \dots, \lambda u_n) = |\lambda|s(u_1, \dots, u_n).$$

Por ejemplo un estimador de escala es

$$s_0 [u_1, \dots, u_n] = \left[ \frac{1}{n} \sum_{i=1}^n u_i^2 \right]^{\frac{1}{2}}$$

pero no es un estimador robusto. Como minimizar  $\left[ \frac{1}{n} \sum_{i=1}^n r_i^2(\boldsymbol{\theta}) \right]$  es lo mismo que minimizar  $\left[ \frac{1}{n} \sum_{i=1}^n r_i^2(\boldsymbol{\theta}) \right]^{\frac{1}{2}}$  el estimador MC se puede definir como

$$\arg \min_{\boldsymbol{\theta}} \{s_0(r_1(\boldsymbol{\theta}), \dots, r_n(\boldsymbol{\theta}))\}.$$

Un estimador de escala robusto es por ejemplo, la mediana de los valores absolutos de los residuos normalizada (MAD)

$$s_1(u_1, \dots, u_n) = \frac{\pi}{2} \text{median}(|u_1|, \dots, |u_n|).$$

(se multiplica por la constante  $\pi/2$  para que  $s$  estime la desviación típica cuando la distribución es normal con media cero.).

En general la clase de estimadores basados en una escala (S-estimadores) se define utilizando una escala arbitraria. Es decir la ecuación de los S-estimadores de regresión viene dada por:

$$\arg \min_{\theta} \{s(r_1(\theta), \dots, r_n(\theta))\} \tag{7.1}$$

utilizando algún estimador de escala  $s(u_1, \dots, u_n)$ .

Una importante clase de estimadores de escala robusta es la clase de M-estimadores.

Dada una función  $\rho$  satisfaciendo propiedades **P1-P5** enunciadas en la Sección 4, el M-estimador de escala calculado en una muestra  $u_1, u_2, \dots, u_n$ , se define como

el valor  $s$  que se obtiene resolviendo

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{u_i}{s}\right) = b, \quad (7.2)$$

Si se quiere que el estimador de escala estime la desviación típica cuando los errores tienen distribución normal,  $b$  se debe elegir como

$$b = E\phi[\rho(x)],$$

siendo  $\phi$  la distribución normal  $N(0, 1)$ . En efecto, si se supone que  $u_1, \dots, u_n$  es una muestra aleatoria de una distribución normal estándar, entonces, cuando  $n \rightarrow \infty$ , entonces

$$\frac{1}{n} \sum_{i=1}^n \rho(u_i) \rightarrow E\phi(\rho(u)). \quad (7.3)$$

Un ejemplo de familia de funciones  $\rho$  que cumple las propiedades **P1** a **P5** es la familia de las funciones bicuadradas, propuesta por Tukey

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4} & |x| \leq c \\ \frac{c^2}{6} & |x| > c \end{cases} \quad (7.4)$$

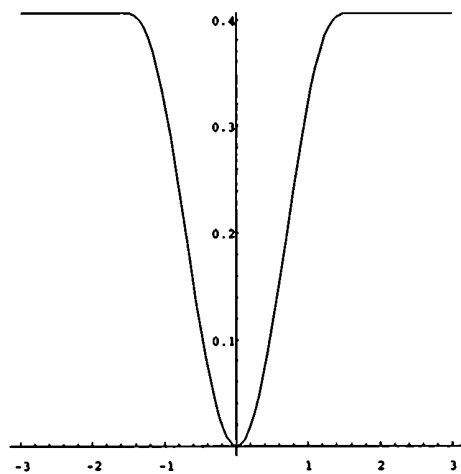
cuya función derivada es:

$$\psi(x) = \begin{cases} x\left(1 - \left(\frac{x}{c}\right)^2\right)^2 & |x| \leq c \\ 0 & |x| > c. \end{cases} \quad (7.5)$$

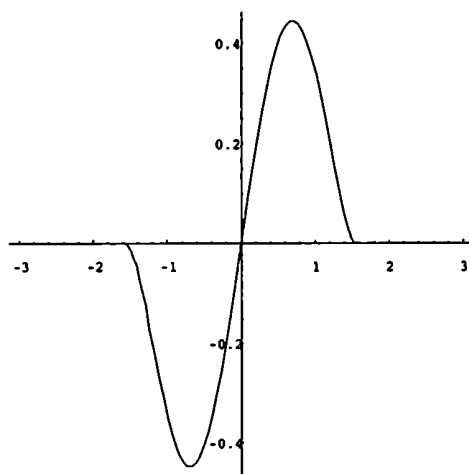
Los S-estimadores de regresión se definen como el valor de  $\theta$  que minimiza un M-estimador de escala de los residuos. Es decir

$$\hat{\theta} = \arg \min_{\theta} s(r_1(\theta), \dots, r_n(\theta)), \quad (7.6)$$

donde  $s$  es un M-estimador de escala. Las figuras de la página 28 muestran las funciones  $\rho$  y  $\psi$  respectivamente, tomando como valor  $c = 1.56$ .



Función  $\rho(x)$



Función  $\psi(x)$

## 7.1. PROPIEDADES ASINTÓTICAS DEL S-ESTIMADOR

Los resultados asintóticos de este tipo de estimadores se basan en el hecho de que los S-estimadores de regresión satisfacen las ecuaciones para M-estimadores simultáneos de regresión y escala . Es decir suponiendo que  $\hat{\theta}$  satisface la ecuación (7.1) con  $s(u_1, \dots, u_n)$  dada por la solución de

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{u_i}{s} \right) = b$$

llamando

$$\hat{\sigma} = s(r_1(\hat{\theta}), \dots, r_n(\hat{\theta})),$$

se puede demostrar que se cumple

$$\hat{\theta} = \arg \min_{\theta \in \mathcal{R}^p} \sum_{i=1}^n \rho \left( \frac{r_i(\theta)}{\hat{\sigma}} \right) \quad (7.7)$$

Entonces si  $\rho$  es diferenciable con  $\psi = \rho'$ ,  $\hat{\theta}$  y  $\hat{\sigma}$  son soluciones del sistema



$$\begin{cases} \sum_{i=1}^n \psi\left(\frac{r_i(\boldsymbol{\theta})}{\hat{\sigma}}\right) \mathbf{x}_i = 0 \\ \frac{1}{n} \sum_{i=1}^n \rho\left(\frac{r_i(\boldsymbol{\theta})}{\hat{\sigma}}\right) = b \end{cases} \quad (7.8)$$

y pueden ser pensados como M-estimadores acotados simultáneos de regresión y escala. Esto implica que se puede aplicar la teoría asintótica desarrollada para M-estimadores. Por lo tanto la distribución asintótica estará dada para el caso de  $\mathbf{x}_i$  fijas por

$$a_n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, V(\psi, F, \sigma^2)\Sigma^{-1}) \text{ donde } \Sigma = \lim_{n \rightarrow \infty} \frac{X'X}{a_n} \quad (7.9)$$

Para el caso de  $\mathbf{x}_i$  aleatorias se pide que éstos tengan momentos de segundo orden.

Sea  $\Sigma = E(\mathbf{x}_i \mathbf{x}_i')$  entonces

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, V(\psi, F, \sigma^2)\Sigma^{-1}), \quad (7.10)$$

donde el escalar  $V(\psi, F, \sigma^2)$  se define como en (4.6).

Como el estimador MC es en particular un M-estimador con  $\psi(u) = u$ , las fórmulas (7.9) y (7.10) también dan su distribución asintótica para el caso de  $\mathbf{x}_i$  fijas o

aleatorias respectivamente. En este caso  $V(\psi, F, \sigma^2) = \text{var}_F(u) = E_F(u^2)$ . Luego la eficiencia relativa de un S-estimador con respecto al estimador MC está dada por

$$E_F(\psi, F, \sigma) = \frac{\text{var}_F(u)}{V(\psi, F, \sigma^2)}. \quad (7.11)$$

Los Teoremas de comportamiento asintótico de los S-estimadores pueden verse en Rousseeuw y Yohai (1984).

## 7.2. PUNTO DE RUPTURA DE LOS S-ESTIMADORES

En esta subsección se estudia el punto de ruptura para muestras finitas de los S-estimadores de regresión.

Sea  $x_1, \dots, x_n$  una muestra aleatoria y  $S_n$  un estimador de escala definido por:

$$S_n = \inf \left\{ s : \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i}{s} \right) \leq b \right\} \quad (7.12)$$

donde  $\rho$  satisface las siguientes propiedades:

- 1)  $\rho(0) = 0$  y continua en 0
- 2)  $\rho(-u) = \rho(u)$
- 3) Si  $0 \leq u \leq v$ , then  $\rho(u) \leq \rho(v)$

4) Sea  $a = \sup_u \rho(u)$ ,  $a < \infty$

Comenzamos demostrando el siguiente

**Lema:** Sea  $m$  el menor entero tal que  $m > \frac{nb}{a}$  y  $m - 1 < \frac{nb}{a}$ , entonces

(i) Dado cualquier  $K > 0$  existe  $K'$  tal que si  $\#\{i : |x_i| > K'\} \geq m$ , entonces  $s(x_1, \dots, x_n) > K$ .

(ii) Dado cualquier  $M > 0$  existe  $M'$  tal que si  $\#\{i : |x_i| > M'\} < m$ , entonces  $s(x_1, \dots, x_n) < M'$ .

(iii) Dado cualquier  $\delta > 0$  existe  $\delta'$  tal que si  $\#\{i : |x_i| < \delta'\} > n - m$ , entonces  $s(x_1, \dots, x_n) < \delta$ . Esto implica que si  $\#\{i : |x_i| = 0\} > n - m$ , entonces  $s(x_1, \dots, x_n) = 0$ .

(iv) Dado cualquier  $\eta > 0$  existe  $\eta'$  tal que si  $\#\{i : |x_i| < \eta'\} \leq n - m$ , entonces  $s(x_1, \dots, x_n) > \eta'$ .

Demostración

(i) Sea  $K > 0$ . Dado que  $am > nb$ , y por definición de supremo, existe  $A$  tal que

$$\rho(A)m > nb, \tag{7.13}$$

y se define  $K' = 2AK$ . Sea  $x_1, \dots, x_n$  una muestra tal que si  $C = \{i : \|x_i\| > K'\}$ ,

luego  $\#C \geq m$ . Entonces

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{2K}\right) \geq \frac{1}{n} \sum_{i \in C} \rho\left(\frac{x_i}{2K}\right) \geq \frac{m}{n} \rho\left(\frac{2AK}{2K}\right) = \frac{m\rho(A)}{n} > b.$$

Luego  $S_n(x_1, \dots, x_n) \geq 2K > K$  y por lo tanto (a) queda probado.

(ii) Sea  $M > 0$ . Como  $(m-1)a < nb$  entonces

$$\alpha = \frac{nb - (m-1)a}{n} > 0. \quad (7.14)$$

Tomemos  $A$  tal que

$$\rho(A) < \alpha \quad (7.15)$$

y sea  $M_1 = M/A$ . Sea  $x_1, \dots, x_n$  una muestra, y sea  $C = \{i : |x_i| > M\}$ .

Supongamos que  $\#C \leq (m-1)$ , entonces

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{M_1}\right) = \frac{1}{n} \left[ \sum_{i \in C} \rho\left(\frac{x_i}{M_1}\right) + \sum_{i \notin C} \rho\left(\frac{x_i}{M_1}\right) \right] \leq$$

$$\leq \frac{(m-1)a}{n} + \rho\left(\frac{A}{M/A}\right) < b - \alpha + \alpha = b$$

Por lo tanto  $S_n(x_1, \dots, x_n) \leq M_1$ , con lo cual (ii) se cumple tomando  $M' = 2M_1$ .

(iii) Sea  $\delta > 0$ . Sea  $\alpha$  como en (7.14),  $A$  como en (7.15),  $\delta' = \delta A/2$  y  $C = \{i : |x_i| < \delta'\}$ . Supongamos que  $\#C \geq n - (m - 1)$  entonces

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{\delta/2}\right) &= \frac{1}{n} \left[ \sum_{i \in C} \rho\left(\frac{x_i}{\frac{1}{2}\delta}\right) + \sum_{i \notin C} \rho\left(\frac{x_i}{\delta/2}\right) \right] \\ &\leq \rho\left(\frac{\delta A}{\delta/2}\right) + \frac{(m-1)a}{n} < \alpha + (b - \alpha) = b. \end{aligned}$$

Por lo tanto  $S_n(x_1, \dots, x_n) \leq \delta/2 < \delta$  con lo cual queda probado (iii).

(iv) Sea  $\eta > 0$ ,  $A$  como en (7.15). Sea  $\eta_1 = \eta/A$ , sea

$C = \{i : |x_i| < \eta\}$  y supongamos  $\#C \leq n - m$ . Entonces

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{\eta_1}\right) \geq \frac{1}{n} \sum_{i \notin C} \rho\left(\frac{x_i}{\eta_1}\right) \geq \frac{m}{n} \rho\left(\frac{\eta}{\eta/A}\right) = \frac{m\rho(A)}{n} > b,$$

y entonces  $S_n(x_1, \dots, x_n) \geq \eta_1$ . Por lo tanto poniendo  $\eta' = \eta_1/2$  se cumple (iv).

**Teorema.** Sea  $S_n(x_1, \dots, x_n)$  un estimador de escala basado en una función  $\rho$  que cumpla las propiedades 1) a 4) definidas anteriormente. Sea  $Z = (\mathbf{z}_1, \dots, \mathbf{z}_n)$  con  $\mathbf{z}_i = (y_i, \mathbf{x}_i)$  y sea  $\mathbf{x}_i \in \mathbb{R}^p$  una muestra para el modelo de regresión lineal.

Consideremos el estimador de escala de la regresión definido por:

$$T_n(\mathbf{Z}) = \arg \min S(r_1(\boldsymbol{\theta}), \dots, r_n(\boldsymbol{\theta}))$$

donde  $r_i(\boldsymbol{\theta}) = y_i - \boldsymbol{\theta}'\mathbf{x}_i$ . Supongamos que el rango de  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  es  $p$ , sea  $q =$

$\max_{\|\boldsymbol{\theta}\|=1} \#\{i : \boldsymbol{\theta}'\mathbf{x}_i = 0\}$ , y sea  $m$  como en el lema anterior. Entonces

$$\varepsilon_n^*(T_n, \mathbf{Z}) \geq \frac{r_1}{n}, \quad (7.16)$$

donde  $r_1 = \min(n - m - q + 1, m)$  y

$$\varepsilon_n^*(T_n, \mathbf{Z}) \leq \frac{r_2}{n}, \quad (7.17)$$

donde  $r_2 = \min(n - m - p + 2, m)$ .

#### Demostración

Supongamos que (7.16) no es cierta. Entonces existe una sucesión de muestras  $\mathbf{Z}^{(j)} = (\mathbf{z}_1^{(j)}, \dots, \mathbf{z}_n^{(j)})$ ,  $1 \leq j < \infty$ ,  $\mathbf{z}_i^{(j)} = (y_i^{(j)}, \mathbf{x}_i^{(j)})$  tal que cada  $\mathbf{Z}^{(j)}$  difiere de  $\mathbf{Z}$  en menos de  $r_1$  observaciones y tal que

$$\lim_{n \rightarrow \infty} \|T_n(\mathbf{Z}^{(j)})\| = \infty. \quad (7.18)$$

Sin pérdida de generalidad podemos suponer que

$$\lim_{n \rightarrow \infty} \frac{T_n(\mathbf{Z}^{(j)})}{\|T_n(\mathbf{Z}^{(j)})\|} = \lambda \quad (7.19)$$

Sea  $M > \sup_{1 \leq i \leq n} |y_i|$ , entonces existe  $M'$  tal que se cumple el inciso (ii) del lema anterior. Por lo tanto todas las muestras  $\mathbf{Z}^{(j)}$  tienen menos de  $m$  observaciones  $\mathbf{z}_i^{(j)} = (y_i^{(j)}, \mathbf{x}_i^{(j)})$  con  $|y_i^{(j)}| > M$ . Entonces

$$S_n(y_1^{(j)}, \dots, y_n^{(j)}) < M' \quad \forall j. \quad (7.20)$$

Sea  $\delta_i = |\lambda' \mathbf{x}_i|$ ,  $1 \leq i \leq n$ , y sea  $\delta = \min\{\delta_i > 0\}$ . Por lo tanto todas las muestras  $\mathbf{Z}^{(j)}$  tienen por lo menos  $n - (r_1 - 1) \geq m$  observaciones  $\mathbf{z}_i^{(j)} = (y_i^{(j)}, \mathbf{x}_i^{(j)})$  de la muestra original  $\mathbf{Z}$  con  $|\lambda' \mathbf{x}_i^{(j)}| > \delta$ . Entonces existe  $K'$  tal que se cumple (i) del lema anterior con  $K = M'$  y por (7.18) y (7.19), podemos encontrar  $j_0$  tal que para todo  $j > j_0$  todas las muestras contienen por lo menos  $m$  puntos tales que  $|y_i^{(j)} - \mathbf{T}_n(\mathbf{Z}^{(j)}) \mathbf{x}_i^{(j)}| > K'$ . Entonces

$$S_n(y_1^{(j)} - T_n(\mathbf{Z}^{(j)})' \mathbf{x}_1^{(j)}, \dots, y_n^{(j)} - T_n(\mathbf{Z}^{(j)})' \mathbf{x}_n^{(j)}) > M', \quad \forall j > j_0 \quad (7.21)$$

Pero (7.20) y (7.21) contradicen la definición de  $T_n(\mathbf{Z}^{(j)})$  y entonces (7.16) se

cumple.

Ahora mostraremos que se cumple (7.17)

Sea  $\mathbf{x}_0$  y  $\lambda_0$  tal que  $y_0 = \lambda_0' \mathbf{x}_0 > 0$  y sea  $\mathbf{w}_j = (j^2 y_0, j \mathbf{x}_0)$ . Sea  $m^* \geq m$ , y sea  $\mathbf{Z}^{(j)}$  una muestra obtenida reemplazando  $m^*$  elementos de  $\mathbf{Z}$  por los elementos  $\mathbf{w}_j$ . Vamos a mostrar que

$$\lim_{j \rightarrow \infty} \|\mathbf{T}_n^{(j)}\| = \infty. \quad (7.22)$$

Supongamos que (7.22) no se cumple. Entonces  $\sup \|\mathbf{T}_n^{(j)}\| < \infty$  y por lo tanto

$$\lim_{j \rightarrow \infty} \frac{1}{j} (j^2 y_0 - \mathbf{T}_n^{(j)'} j \mathbf{x}_0) = \infty.$$

Por lo tanto de acuerdo con (i) y a la equivarianza de  $S_n$

$$\lim_{j \rightarrow \infty} \frac{1}{j} s(y_1^{(j)} - \mathbf{T}_n(\mathbf{Z}^{(j)})' \mathbf{x}_1^{(j)}, \dots, y_n^{(j)} - \mathbf{T}_n(\mathbf{Z}^{(j)})' \mathbf{x}_n^{(j)}) = \infty. \quad (7.23)$$

Llamando  $\theta^{(j)} = j \lambda_0$  entonces  $j^2 y_0 - \theta^{(j)} j \mathbf{x}_i^{(j)} = 0$  y si  $M = \sup_i (|y_i| + |\lambda_0' \mathbf{x}_i|)$

entonces

$$\sup_i \frac{1}{j} |y_i - \theta^{(j)} j \mathbf{x}_i^{(j)}| \leq M \quad 1 \leq i \leq n.$$



Entonces por (ii) y la equivarianza de  $S_n$  existe  $M'$  tal que

$$\frac{1}{j} S_n(y_1^{(j)} - j\theta^{(j)'} \mathbf{x}_1^{(j)}, \dots, y_n^{(j)} - j\theta^{(j)'} \mathbf{x}_n^{(j)}) < M'.$$

(7.23) y (7.24) contradicen nuevamente la definición de  $\mathbf{T}_n$ . Entonces (7.22) se cumple y

$$\varepsilon_n^*(\mathbf{T}_n, Z) \leq \frac{m}{n}. \quad (7.24)$$

Supongamos sin pérdida de generalidad que  $\mathbf{x}_1, \dots, \mathbf{x}_p$  son linealmente independientes. Sea  $\lambda_0$  tal que  $\lambda_0' \mathbf{x}_i = y_i$ ,  $1 \leq i \leq p$ . También podemos encontrar  $\lambda_1 \neq 0$  tal que  $\lambda_1' \mathbf{x}_i = 0$ ,  $1 \leq i \leq p-1$ . Para todo  $j$  llamemos  $\theta^{(j)} = \lambda_0 + j\lambda_1$  y  $\mathbf{w}_j = (\theta^{(j)'} \mathbf{x}_0, \mathbf{x}_0)$ ,  $\mathbf{x}_0$  arbitrario. Sea  $m^* > n - m - p + 1$ , y sea  $\mathbf{Z}^{(j)}$  la muestra obtenida reemplazando  $m^*$  observaciones  $\mathbf{z}_i$ ,  $i > p-1$  de  $\mathbf{Z}$  por  $\mathbf{w}_j$ . Por lo tanto  $y_i - \theta^{(j)'} \mathbf{x}_i = 0$ ,  $i = 1, 2, \dots, p-1$  y los residuos correspondientes a  $\mathbf{w}_j$  usando como parámetros de la regresión al vector  $\theta^{(j)}$  son también son 0. Luego la muestra  $\mathbf{Z}^{(j)}$  tiene más de  $n - m$  elementos tales que  $y_i^{(j)} - \theta^{(j)'} \mathbf{x}_i^{(j)} = 0$ , y entonces de acuerdo a (iii)

$$S_n(y_1^{(j)} - \theta^{(j)'} \mathbf{x}_1^{(j)}, \dots, y_n^{(j)} - \theta^{(j)'} \mathbf{x}_n^{(j)}) = 0.$$

Entonces  $\mathbf{T}_n(\mathbf{Z}^{(j)}) = \theta^{(j)}$ , y por lo tanto  $\lim_{j \rightarrow \infty} \|\mathbf{T}_n(\mathbf{Z}^{(j)})\| = \infty$ . Esto prueba

que

$$\varepsilon_n^*(\mathbf{T}_n, \mathbf{Z}) \leq \frac{n - m - p + 2}{n}. \quad (7.25)$$

(7.25) y (7.26) implican (7.17).

**Observación.** Si  $q = p - 1$ , es decir, si la muestra está en posición general, entonces  $\varepsilon_n^*(\mathbf{T}_n, \mathbf{Z}) = r_2/n$ . y para  $n$  grande esto es aproximadamente  $\min\left(\frac{b}{a}, 1 - \frac{b}{a}\right)$ .

Los teoremas sobre punto de ruptura de los S-estimadores pueden verse en Rousseeuw y Yohai (1984) y Yohai (1987).

## 8. MM-ESTIMADORES

Los MM-estimadores fueron introducidos por Yohai (1987). Estos estimadores combinan un alto punto de ruptura con una alta eficiencia si los errores tienen distribución normal. Se definen de la siguiente manera

1. Se parte de un estimador inicial  $\hat{\theta}_0$  con alto punto de ruptura . Una posibilidad es usar un S-estimador.
2. Se calculan los residuos  $r_i(\hat{\theta}_0) = y_i - \hat{y}_i(\hat{\theta}_0)$ ,  $1 \leq i \leq n$  y un M-estimador de escala de los mismos,  $\hat{\sigma}$ , de la forma

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{r_i(\hat{\theta}_0)}{\hat{\sigma}} \right) = b,$$

donde  $\rho_0$  satisface las propiedades **P1** a **P5** enunciadas en sección 4 y

$$b = \frac{\sup \rho_0(t)}{2}.$$

3. Usando otra función  $\rho_1$  que satisfaga también las propiedades **P1** a **P5** enunciadas en sección 4 y que además para todo  $t$

$$\rho_1(t) \leq \rho_0(t)$$

se define  $\hat{\theta}_1$  como un mínimo local de

$$S(\theta) = \sum_{i=1}^n \rho_1 \left( \frac{r_i(\theta)}{\hat{\sigma}} \right)$$

tal que

$$S(\hat{\theta}_1) \leq S(\hat{\theta}_0).$$

Generalmente  $\hat{\theta}_1$  se calcula a través de un procedimiento iterativo de mínimos cuadrados pesados usando como estimador inicial a  $\hat{\theta}_0$ . Se puede probar que el punto de ruptura de  $\hat{\theta}_1$  es mayor o igual que el de  $\hat{\theta}_0$  y que la eficiencia de  $\hat{\theta}_1$  depende sólo de  $\rho_1$ . Por lo tanto  $\rho_1$  se puede elegir de manera tal que bajo errores normales la eficiencia sea tan alta como se quiera.

## 8.1. PROPIEDADES ASINTÓTICAS DE LOS MM-ESTIMADORES

Los MM-estimadores pueden pensarse como un caso particular de M-estimadores con la escala estimada separadamente . Por lo tanto se puede aplicar la teoría asintótica descrita en la sección 4.1 para M-estimadores. Yohai (1987) prueba la consistencia y normalidad asintótica de los MM-estimadores para el caso de  $x_i$  aleatorias.

## 9. ALGORITMOS RAPIDOS PARA COMPUTAR ESTIMADORES ROBUSTOS.

### 9.1. FAST-LTS

El algoritmo FAST-LTS fue propuesto por Rousseeuw (1984) para computar LTS-estimadores.

La clave de este nuevo algoritmo radica en el hecho de que empezando con cualquier aproximación a los estimadores LTS de los coeficientes de la regresión, es posible computar otra aproximación con menor función objetivo.

Se basa en la siguiente propiedad

Sea  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  un conjunto de datos y sea  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$  un vector cualquiera de dimensión  $p$  y  $h$  un número natural comprendido entre  $\lfloor n/2 \rfloor$  y  $n$ . Definamos  $n_i$ ,  $1 \leq i \leq n$  por  $r_{(i)}^2(\boldsymbol{\theta}) = r_{n_i}^2(\boldsymbol{\theta})$  y  $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_p^*)$  el estimador de mínimos cuadrados usando únicamente las observaciones  $(\mathbf{x}_{n_i}, y_{n_i})$ ,  $1 \leq i \leq h$ . Luego

$$Q_h(\boldsymbol{\theta}^*) \leq Q_h(\boldsymbol{\theta}).$$

En efecto, como  $\boldsymbol{\theta}^*$  es el estimador de mínimos cuadrados de las  $h$  observaciones  $(\mathbf{x}_{n_i}, y_{n_i})$ ,  $1 \leq i \leq h$ , se tiene

$$Q_h(\boldsymbol{\theta}^*) = \sum_{i=1}^h r_{n_i}^2(\boldsymbol{\theta}^*) \leq \sum_{i=1}^h r_{n_i}^2(\boldsymbol{\theta}) = Q(\boldsymbol{\theta})$$

y por lo tanto la propiedad queda demostrada.

Basado en esta propiedad, Rousseeuw propone el siguiente step de concentración

(C-step) para mejorar un valor  $\hat{\boldsymbol{\theta}}_0$

1. Se computan los residuos  $r_i(\hat{\boldsymbol{\theta}}_0)$  para  $i = 1, \dots, n$ .
2. Se consideran las  $h$  observaciones correspondientes a los menores valores absolutos de estos residuos
3. Se computa el estimador de mínimos cuadrados  $\hat{\boldsymbol{\theta}}_1$  basado en estas observaciones

Repitiendo el C-step se obtiene un proceso iterativo obteniéndose valores  $\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_k$  que satisfacen

$$Q_h(\hat{\boldsymbol{\theta}}_0) \geq Q_h(\hat{\boldsymbol{\theta}}_1) \geq \dots \geq Q_h(\hat{\boldsymbol{\theta}}_k).$$

El proceso se acaba cuando  $Q_h(\hat{\boldsymbol{\theta}}_j) = Q_h(\hat{\boldsymbol{\theta}}_{j+1})$ , en cuyo caso se tendrá un mínimo local. Sin embargo ésto no es suficiente para que sea el mínimo global de  $Q_h$ .

Rousseeuw propone el siguiente algoritmo para calcular el LTS:

1. Generar  $M$  valores iniciales  $\hat{\theta}_0^{(1)}, \dots, \hat{\theta}_0^{(M)}$ . Cada uno de estos valores es obtenido tomando al azar una submuestra de  $p$  elementos y ajustando un hiperplano a los mismos.
2. A partir de cada  $\hat{\theta}_0^{(i)}$  se obtiene un valor  $\hat{\theta}^{(i)}$  haciendo dos C-steps y se calcula el valor  $Q_h(\hat{\theta}^{(i)})$ .
3. Se guardan los diez  $\hat{\theta}^{(i)}$  con menor valor  $Q_h(\hat{\theta}^{(i)})$ , que denominaremos  $\hat{\theta}_1^*, \dots, \hat{\theta}_{10}^*$ .
4. A partir de cada  $\hat{\theta}_i^*$  se aplica el C-step hasta convergencia. Llamemos  $\hat{\theta}_i^{**}$  al último valor.
5. Se define el estimador final  $\hat{\theta}^*$  como aquel  $\hat{\theta}_i^{**}$  que minimiza  $Q_h(\hat{\theta}_i^{**})$ .



## 9.2. FAST-S

El objetivo de este trabajo es desarrollar un algoritmo rápido para computar el S-estimador desarrollado en la Sección 7. El algoritmo propuesto utiliza un C-step similar al descrito para el LTS, mediante el cual partiendo de un estimador cualquiera se obtiene en general otro estimador con menor función objetivo.

El C-step en este caso se basa en la siguiente propiedad.

Supongamos que se tiene un estimador  $\hat{\theta}_0$ . Calculemos los residuos  $r_i(\hat{\theta}_0)$  para  $1 \leq i \leq n$  y el estimador de escala  $s_0 = s(r_1(\hat{\theta}_0), \dots, r_n(\hat{\theta}_0))$ . Sean los pesos

$$w_i = \frac{\psi(r_i(\hat{\theta}_0)/s_0)}{r_i(\hat{\theta}_0)/s_0},$$

y definamos  $\hat{\theta}_1$  como el estimador de mínimos cuadrados ponderados con estos pesos.

Se puede demostrar que si la función  $\rho$  que define la escala satisface las propiedades **P1** a **P5** enunciadas en sección 4, entonces

$$s(r_1(\hat{\theta}_1), \dots, r_n(\hat{\theta}_1)) \leq s(r_1(\hat{\theta}_0), \dots, r_n(\hat{\theta}_0)). \quad (9.1)$$

Luego, si se aplica en forma iterativa este paso se obtiene una sucesión de

estimadores que converge a un mínimo local.

Como el cómputo de  $s_0$  es lento, se lo puede reemplazar por  $s_0^*$ , el resultado de aplicar una sola iteración en el algoritmo de cálculo de  $s_0$  partiendo del MAD.

Este algoritmo puede ser el de Newton-Raphson o el basado en la recursión

$$s^{(j+1)} = s^{(j)} \left( \frac{1}{nb} \sum_{i=1}^n \rho \left( \frac{r_i(\hat{\theta}_0)}{s^{(j)}} \right) \right)^{1/2}$$

En este caso no se puede demostrar (8.1), aunque en la mayoría de los casos va a valer debido a que  $s_0^*$  es generalmente cercano a  $s_0$ .

Luego partiendo de  $\hat{\theta}_0$ , definimos el C-step para el cálculo del S-estimador como sigue:

1. Se calculan los residuos  $r_i(\hat{\theta}_0)$ ,  $1 \leq i \leq n$  y el estimador de escala de un paso  $s_0^*$  partiendo del MAD.
2. Se calculan los pesos  $w_i = \frac{\psi(r_i(\hat{\theta}_0)/s_0^*)}{r_i(\hat{\theta}_0)/s_0^*}$ .
3. Definamos  $\hat{\theta}_1$  como el estimador de mínimos cuadrados ponderados con estos pesos.

El algoritmo propuesto depende de 3 parámetros:  $N$ , el número de submuestras,  $h$ , el número de veces que se aplica el C-step a cada estimador obtenido por ajuste

exacto para cada submuestra y  $m$ , el número de estimadores que se guardan para la comparación final. El algoritmo consiste en los siguientes pasos:

1. Generar  $N$  valores iniciales  $\hat{\theta}_0^{(1)}, \dots, \hat{\theta}_0^{(N)}$ . Cada uno de estos valores se obtiene tomando al azar una submuestra de  $p$  elementos de la muestra original y ajustando un hiperplano a los mismos.
2. A partir de cada  $\hat{\theta}_0^{(i)}$  se obtiene, aplicando  $h$  veces el C-step otro estimador  $\hat{\theta}_1^{(i)}$ .
3. Se determina si este estimador se incluye entre los  $m$  de menor escala. Para eso supongamos que  $s_1, \dots, s_m$  son las  $m$  escalas más pequeñas entre los primeros  $(i - 1)$  valores  $s(\hat{\theta}_1^{(j)})$  y que corresponden a  $\theta_1^{(j_1)}, \dots, \theta_1^{(j_m)}$ . Para determinar si se incluye  $\hat{\theta}_1^{(i)}$  entre los  $m$  mejores se computa

$$s^* = \max_{1 \leq j \leq m} \{s_j\}$$

y se determina si  $\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{r_i(\hat{\theta}_1^{(i)})}{s^*} \right) < b$ . Si ésto es cierto se incluye  $\hat{\theta}_1^{(i)}$  entre los  $m$  mejores, y se calcula  $s(\hat{\theta}_1^{(i)})$ . De esta manera se obtienen los  $m$  estimadores con menor función objetivo, los llamaremos  $\hat{\theta}_1^*, \dots, \hat{\theta}_m^*$ . A cada  $\hat{\theta}_j^*$   $1 \leq j \leq m$  se le aplica el C-step hasta convergencia. Sea  $\hat{\theta}_j^{**}$  el

último valor. Se define el estimador final  $\hat{\theta}^*$  por  $\hat{\theta}^* = \arg \min_{1 \leq j \leq m} s(\hat{\theta}_j^{**})$ .

## 10. EJEMPLO

1. El siguiente ejemplo con datos reales y con un gran número de outliers, puede verse en Rousseeuw y Yohai (1984) y demuestra la utilidad del S-estimador. En el "Belgian Statistical Survey" se encuentra un conjunto de datos conteniendo el número total de llamadas internacionales realizadas desde Bélgica. Estos datos están dados en la siguiente tabla donde  $x$  representa las dos últimas cifras del año e  $y$  el número de llamadas en decenas de millones.

$x$	50	51	52	53	54	55	56	57
$y$	0.44	0.47	0.47	0.59	0.66	0.73	0.81	0.88
$x$	58	59	60	61	62	63	64	65
$y$	1.06	1.20	1.35	1.49	1.61	2.12	11.9	12.4
$x$	66	67	68	69	70	71	72	73
$y$	14.2	15.9	18.2	21.2	4.30	2.40	2.70	2.90

La Figura 10.1 parece mostrar una tendencia creciente a lo largo de los años. Sin embargo esta serie de tiempo contiene varios outliers entre los años 1964 y 1969. Ello se debe a que durante ese período se utilizó otro sistema de registro. En lugar

del número de llamadas, se registró el número total de minutos de las llamadas. Los años 1963 y 1970 también se ven afectados parcialmente, porque la transición de un sistema a otro no se hizo el primer día del año, así que se agregó al número de llamadas, el número de minutos registrados. Esto causa una gran cantidad de outliers en la dirección del eje  $y$ .

La regresión por mínimos cuadrados para estos datos viene dada por la recta  $\hat{y} = 0.504x - 26.01$ , la cual está muy afectada por los valores de  $y$  asociados al período 1964 a 1969, y es por eso que tiene una pendiente grande y no ajusta bien ni los datos buenos ni los datos malos.

Aplicando el algoritmo FAST-LTS para calcular un LTS-estimador se obtiene la recta de ecuación  $\hat{y} = 0.116x - 5.61$

Aplicando el algoritmo FAST-S para calcular un S-estimador se obtiene la recta de ecuación  $\hat{y} = 0.113x - 5.4$

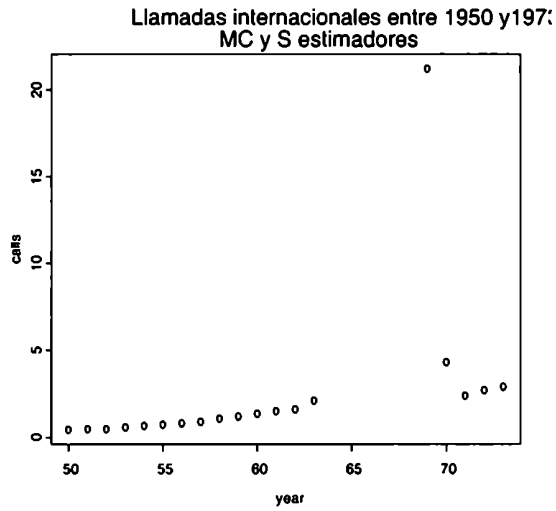


Figure 10.1:

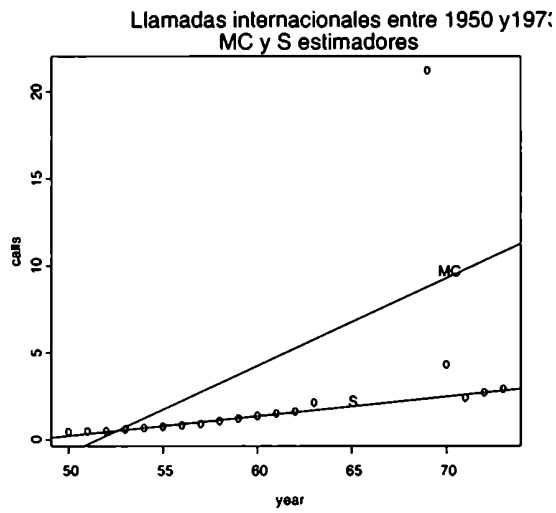


Figure 10.2:

## 11. SIMULACIONES

Se evaluó la performance del algoritmo FAST-S, utilizando como función  $\rho$  la función bicuadrada definida en (7.4) con  $c = 1.56$  y  $b = 0.5$  y se la comparó con la del algoritmo FAST-LTS. Se procedió de la siguiente manera:

Se tomaron muestras de tamaño  $n = 400$  con  $p = 35$  variables explicativas, considerando el modelo

$$y_i = \theta_1 x_{i1} + \dots + \theta_j x_{ij} \dots + \theta_{35} x_{i,35} + \theta_{36} + u_i, \quad 1 \leq i \leq n.$$

con  $\theta_j = 0$  para  $j = 1, 2, \dots, 35$ . Para el 90% de las observaciones, los vectores  $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$  se obtuvieron generando variables aleatorias normales independientes con media cero y varianza 1. El 10% restante de las observaciones son los outliers que son todos iguales al mismo punto  $(100, 0, \dots, 0, M)$ . Se eligieron valores de  $M$  entre 90 y 200 con paso de 10.

Es decir que, para cada algoritmo y para cada valor de  $k$  se realizaron 12 simulaciones, una con cada uno de los valores de  $M$  considerados. El número de replicaciones fue 500. Se aplicó en cada caso el algoritmo FAST-S para  $k = 0$  y 1 y el algoritmo FAST-LTS para  $k = 0, 1$  y 2 ( $k$  representa el número de veces que se aplica el C-step). Se eligió el número  $N$  de submuestras para los distintos



valores de  $k$  a los efectos de que el tiempo de cómputo sea aproximadamente el mismo.

La primer columna de las Tablas 1 y 2 de las páginas 51, 52 , indica el tipo de estimador usado. La segunda columna, el número de veces que se aplica el C-step.

La tercer columna indica el número de submuestras elegidas.

En la Tabla 1, las columnas desde la cuarta hasta la decimoquinta, indican el porcentaje de muestras para cada una de las pendientes indicadas en la primer fila de la tabla, en las cuales el estimador es afectado por los outliers. Que un estimador se ve afectado por la presencia de outliers quiere decir que la estimación no es buena, ya que el valor absoluto de la primer componente del estimador  $\hat{\theta}$  obtenido es mayor que  $0.8 * \frac{M}{100}$ , lo que indica que el estimador converge a la pendiente de los puntos atípicos y no al verdadero valor del parámetro (0 en este caso). Se puede observar que el porcentaje de muestras en las cuales el estimador falla es menor en el caso de aplicar el algoritmo FAST-S.

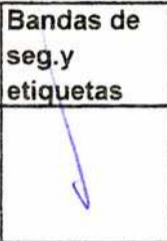
La Tabla 2 muestra el Error Cuadrático Medio para cada uno de los estimadores y cada una de las pendientes. Se observa que para  $k = 1$  es menor el error cuadrático medio al aplicar el FAST-S a partir de una pendiente 1.1.

En conclusión, la performance del algoritmo FAST-S es superior a la del algoritmo FAST-LTS y tanto en el caso del estimador S como el LTS, los algoritmos rápidos

son superiores que los simples correspondientes a  $k = 0$

## Servicios Técnicos

### Proceso técnico de los documentos

Ingreso a S.T.	Verif. list y datos de gestión	Nro. Inventari o	Sig. topogr. MAE MAT	Ingreso base PRE	Bandas de seg. y etiquetas	Exporta ción	Base Biblo	Fin del Proceso
	17/12/04	3765	3765	fig. 23 17/12/04				
Descrip.								
Pal. Clave								

## SIMULACIONES CON 500 REPETICIONES

### PORCENTAJE DE MUESTRAS EN LAS QUE EL ESTIMADOR FALLA

EST	k	N	pendiente											
			0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
S	0	2000	100	99	95	84	72	61	49	39	32	23	18	11
S	1	400	98	91	67	36	16	5	1	1	0	0	0	0
LTS	0	3900	100	98	97	92	84	73	59	52	39	31	23	16
LTS	1	950	100	97	85	67	64	46	27	10	1	0	0	0
LTS	2	400	98	97	94	82	64	43	24	12	4	2	0	0

*TABLA 1*

# SIMULACIONES CON 500 REPETICIONES

## ERROR CUADRATICO MEDIO

EST	k	N	pendiente											
			0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
S	0	2000	1.35	1.46	1.53	1.52	1.47	1.42	1.33	1.24	1.17	1.06	0.99	0.87
S	1	400	1.32	1.37	1.24	1.02	0.83	0.72	0.69	0.67	0.67	0.67	0.67	0.67
LTS	0	3900	1.69	1.80	1.89	1.95	1.97	1.92	1.80	1.77	1.61	1.50	1.36	1.24
LTS	1	950	1.25	1.33	1.33	1.22	1.35	1.17	0.97	0.77	0.67	0.66	0.66	0.66
LTS	2	400	1.25	1.34	1.41	1.39	1.29	1.12	0.93	0.81	0.72	0.69	0.67	0.66

*TABLA 2*

## 12. REFERENCIAS

1. Bickel, P. y Daksum, K (2001) "*Mathematical Statistics: Basic ideas and selected topics.*" Vol I, Prentice Hall.
2. Davies, P. (1987) "*Asimptotic behavior of S-estimates of multivariate location parameters and dispersion matrices.*" The Annals of Statistics. Vol 15, N°3, 1269-1292.
3. Donoho, D y Huber, P. (1983) "*The notion of breakdown-point*" Wadsworth, Belmont, CA.
4. Hampel, F. (1971) "*A general qualitative definition of robustness*", Ann.Math. Stat. 42, 1887-1896.
5. Huber, P.(1973) "*Robust regression: Asymptotic, conjetures and Montecarlo*", Ann. Stat., 1, 799-821.
6. Huber, P.(1980) "*Robust Statistics*" Wiley.
7. Klein R. y Yohai V. "*Asimptotic behavior of M-estimates of location*". Boletín de la Sociedad Brasileira de Matematicas.
8. Rousseeuw, P.(1984) "*Last median of squares regression*" Journal of the

American Statistical Association, 79, 871-880.

9. Rousseeuw, P. y Yohai, V. (1984) "*Robust regression by means of S-estimator*".  
Lecture Notes in Statistics N°26, Springer Verlag, N.Y. 256-272.
10. Rousseeuw, P. y Leroy, A. (1987) "*Robust regression and outlier detection*"  
, Wiley.
11. Rousseeuw, P. y Van Driessen, K. (1998) "*Computing LTS regression for  
large data sets*"
12. Ruppert, D. (1993) "*Computing S-estimators for regression and multivariate  
location/dispersion*" Journal of the American Statistical Association, vol 1,  
N°3, 253-270
13. Yohai V.(1973) "*Asymptotically optimal Bayes designs of experiments for  
estimation*" The Annals of Statistics, vol.2 822-837.
14. Yohai V. y Zamar R. (1986) "*High breakdown point estimates of regression  
by means of minimization of an efficient scale*". Technical Report N° 84
15. Yohai, V. (1987) "*High breakdown-point and high efficiency robust estimates  
for regression*". Annals of Statistics. vol 15, No 2, pp 642-656.