



Facultad de Ciencias Exactas y Naturales  
Departamento de Matemática

**ESTIMADORES DE REGRESIÓN DE TIPO M  
COMBINANDO ALTA EFICIENCIA Y BAJO MÁXIMO  
SESGO**

Tesis presentada para optar al título de Doctor de la Universidad de  
Buenos Aires en el área Ciencias Matemáticas

**Marcela Svarc**

Director de tesis: Dr. Victor J. Yohai

Lugar de trabajo: Universidad de San Andrés.

Buenos Aires, 19 de octubre del 2006.

# Contenido

<b>Contenido</b>	<b>1</b>
<b>1 Introducción</b>	<b>2</b>
<b>2 Capítulo 1</b>	<b>7</b>
2.1 Notación y definiciones básicas . . . . .	7
2.2 Algunas definiciones generales y medidas de robustez . . . . .	8
2.3 Propiedades de equivarianza de los estimadores . . . . .	11
2.4 Algunas propuestas de estimadores robustos de regresión . . .	12
2.4.1 Estimador $L_1$ de regresión . . . . .	12
2.4.2 Estimadores de tipo M de regresión . . . . .	13
2.4.3 Estimadores de tipo S de regresión . . . . .	17
2.4.4 Estimador de mínima mediana de cuadrados . . . . .	21
2.4.5 Estimadores MM de regresión . . . . .	21
2.4.6 Estimadores de regresión basados en proyecciones . . .	22
<b>3 Capítulo 2</b>	<b>26</b>
3.1 Estimadores de tipo M de regresión combinando robustez y eficiencia . . . . .	26
3.2 El problema de optimalidad de Hampel para el modelo de posición . . . . .	26
3.3 El problema de Hampel para el modelo de regresión . . . . .	27
3.4 Estimadores de tipo M de regresión óptimos . . . . .	30
3.5 Algunos resultados numéricos . . . . .	34
<b>4 Capítulo 3</b>	<b>41</b>
4.1 Elección del estimador inicial para un estimador de tipo M de regresión . . . . .	41
4.2 Comparación por Monte Carlo de los estimadores MM de re- gresión con diferentes estimadores iniciales . . . . .	42
4.3 Cálculo del máximo sesgo asintótico para los estimadores PMM	57
<b>5 Bibliografía</b>	<b>62</b>

# 1 Introducción

Uno de los objetivos más importantes de la estadística robusta es el desarrollo de estimadores poco sensibles a pequeñas desviaciones de la distribución de las observaciones y a la presencia de observaciones atípicas, y que simultáneamente son eficientes en el caso que el modelo supuesto es correcto. Esto se puede lograr resolviendo el problema de encontrar estimadores que minimicen la varianza asintótica bajo un modelo central sujeto a que el máximo sesgo asintótico sea menor que una cota dada. Este enfoque fue propuesto primero por Hampel (1974) quien encontró una solución aproximada para un modelo con un solo parámetro. La solución de Hampel simplifica el problema reemplazando el máximo sesgo asintótico por su aproximación lineal infinitesimal.

En diversos trabajos posteriores se resolvieron problemas similares para modelos más complejos, incluido el modelo de regresión. En esta tesis se encuentra el estimador de tipo M de regresión que minimiza la varianza asintótica bajo el modelo central sujeto a que el máximo sesgo asintótico (no su aproximación lineal infinitesimal) sea menor o igual a una cota dada.

Este trabajo se encuentra organizado del siguiente modo.

En el Capítulo 1 presentamos el modelo de regresión y hacemos una breve reseña de los conceptos más importantes para medir el grado de robustez y eficiencia de un estimador. También introducimos varias clases de estimadores que se utilizarán en esta tesis enunciando sus propiedades más importantes.

En el Capítulo 2 planteamos un problema tipo Hampel para los estimadores de tipo M en el modelo de regresión. Hallamos la función de scores que corresponde al estimador de tipo M que minimiza la varianza asintótica bajo el modelo central sujeta a una cota en el máximo sesgo asintótico. La solución que resuelve este problema es comparada con la solución aproximada obtenida por Yohai y Zamar (1997) basada en un concepto de sesgo infinitesimal.

Por último, en el Capítulo 3, obtenemos un estimador de regresión que simultáneamente es altamente robusto desde el punto de vista del sesgo, y que es eficiente cuando los errores tienen distribución normal. Ese estimador está definido en dos etapas. En la primera de ellas calculamos un estimador basado en proyecciones de regresión. Esta elección se basa en el hecho de que este estimador es en cuanto a sesgo el más robusto de los estimadores conocidos. En la segunda etapa calculamos un estimador MM con alta eficiencia mediante el algoritmo de mínimos cuadrados pesados iterados usando como estimador inicial el estimador basado en proyecciones calculado en la primera etapa. Para estudiar la robustez y eficiencia de este estimador se realizó un

estudio de Monte Carlo que muestra que el estimador propuesto se compara favorablemente con un estimador MM que usa como inicial un estimador de tipo S.

*Palabras claves y frases: estimadores de tipo M, regresión robusta, función de máximo sesgo, problema de Hampel, sesgo minimax.*

# M-estimates of regression combining high efficiency and small maximum bias

One of the main goals of Robust Statistics is to develop estimates which simultaneously are not very sensitive to small departures of the distribution of the observations or to outliers observations and highly efficient when the model is correct. This can be achieved by solving the problem of finding estimates which minimizes the asymptotic variance subject to the constrain that the maximum bias is smaller that a given bound. This approach was first proposed by Hampel (1974) who found an approximate solution for one-parameters models. Hampel's solution simplifies the problem by replacing the maximum asymptotic bias by a linear infinitesimal approximation.

After Hampel's work, several authors extended this infinitesimal approach to more complex models, including regression. In this thesis we find the regression M-estimate that minimizes the asymptotic variance subject to the constrain that the maximum asymptotic bias (not its infinitesimal linear approximation) be smaller or equal than a given bound.

This thesis is organized as follows:

In Chapter 1 we describe the regression model and present the most important concepts to measure the degree of robustness and efficiency of an estimate. We also introduce several classes of robust estimates that will be used in this thesis and state its main properties.

In Chapter 2, we solve a Hampel type problem for the regression model. We find the score function corresponding to the M-estimate which minimizes the maximum asymptotic bias subject to a bound on the asymptotic variance. The solution of this problem is compared with the approximate solution obtained by Yohai and Zamar (1997) based on an infinitesimal biased concept.

Finally, in Chapter 3 we obtain a regression estimate which is simultaneously highly bias-robust and highly efficient under normal errors. This estimate is defined in two steps. In the first step we compute a projection estimate. This choice is due to the fact that this estimate is the most bias-robust among all the known estimates. In a second stage we compute a highly efficient MM-estimate by means of the iterated weighted least squares algorithm using as starting point the P-estimate computed in the first stage.

*Key words and phrases:* M-estimates, robust regression, maximum bias function, Hampel 's problem.

## Agradecimientos

Agradezco, en primer lugar, al Dr Victor Yohai por haber ayudado a introducirme en el mundo de la investigación en estadística a través de sus enseñanzas y su orientación en éste mi primer trabajo de investigación.

En segundo término, le agradezco al Departamento de Matemática y Ciencias de la Universidad de San Andrés donde encontré un excelente ámbito de trabajo y un estímulo constante en esta difícil etapa. En particular quiero agradecerle a Ricardo Fraiman por su constante confianza y optimismo.

En un ámbito más personal, quiero agradecerle a mis padres y hermanos que siempre estuvieron a mi lado a lo largo de este camino, por creer en mí y por recordarme constantemente cuáles son las cosas importantes en la vida.

Por último, quiero agradecerle a Neco por estar a mi lado y esforzarse, continuamente, por comprenderme.

*Para Ana y Federico*

## 2 Capítulo 1

### 2.1 Notación y definiciones básicas

Consideramos el modelo usual de regresión lineal con intercept y regresores aleatorios; i.e. sean  $\mathbf{z}_i = (y_i, \mathbf{x}_i')$   $i = 1, \dots, n$  el vector de dimensión  $p + 1$  de observaciones que satisfacen

$$y_i = \alpha_0 + \boldsymbol{\beta}'_0 \mathbf{x}_i + u_i, \quad i = 1, \dots, n. \quad (2.1)$$

El vector  $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{0p})'$  está compuesto por las pendientes y  $\alpha_0$  es el intercept; estos son los parámetros a estimar. Los  $u_i$  son los errores que son independientes de las variables regresoras  $\mathbf{x}_i$ . Asumimos que los  $u_i$  son independientes entre si tienen media cero y varianza  $\sigma^2$ .

En este trabajo los vectores serán vectores columna y notaremos por  $'$  a la traspuesta.

Llamaremos  $G_0(\mathbf{x})$  a la distribución de  $\mathbf{x}_i$ ,  $F_0(u)$  a la distribución de  $u_i$  y  $H_{\alpha_0, \boldsymbol{\beta}_0}(y, \mathbf{x})$  la distribución de  $\mathbf{z}$  que está dada por

$$H_{\alpha_0, \boldsymbol{\beta}_0}(\mathbf{x}, y) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} F_0(y - \alpha - \boldsymbol{\beta}'\mathbf{s}) dG_0(\mathbf{s}). \quad (2.2)$$

El estimador más usual para resolver este problema es el de mínimos cuadrados, se basa en minimizar la suma del cuadrado de los residuos, es decir,

$$\left( \widehat{\alpha}, \widehat{\boldsymbol{\beta}}' \right) = \arg \min_{(\alpha, \boldsymbol{\beta})} \sum_{i=1}^n r_i^2(\alpha, \boldsymbol{\beta}), \quad (2.3)$$

donde los residuos son

$$r_i(\alpha, \boldsymbol{\beta}) = y_i - \alpha - \boldsymbol{\beta}'\mathbf{x}_i. \quad (2.4)$$

Este estimador fue propuesto a principio de siglo XIX simultáneamente por Gauss y Legendre (ver Stigler, 1981). Bajo ciertas condiciones de regularidad este estimador presenta propiedades optimales. Si la distribución de los errores es tal que  $E(u) = 0$ , y  $\text{cov}(\mathbf{u}) = \sigma^2 I$ , donde  $\mathbf{u} = (u_1, \dots, u_n)$  este estimador es el único de mínima varianza dentro de los estimadores lineales insesgados de  $\boldsymbol{\beta}'\mathbf{x}$  (Teorema de Gauss-Markov). Por otra parte, si además la distribución de los errores es normal, entonces este estimador coincide con el de máxima verosimilitud y su matriz de covarianza alcanza la cota de Rao-Cramer. Otra característica importante es que se lo puede calcular de manera explícita.



Sin embargo, el método de mínimos cuadrados deja de ser óptimo cuando se producen desviaciones del modelo normal. Por ejemplo, cuando la distribución de los errores presenta colas pesadas o en presencia de datos atípicos. De hecho, basta un outlier para modificar drásticamente la estimación. Esta problemática da lugar a la definición de estimadores robustos.

## 2.2 Algunas definiciones generales y medidas de robustez

En esta sección definiremos a los estimadores como funcionales en el espacio de las distribuciones, esto facilita el estudio de las propiedades que se desea que un estimador tenga.

Consideramos estimadores para el modelo de regresión dado en (2.1) basados en funcionales que actúan sobre las distribuciones empíricas. Luego se tendrá

$$\left(\widehat{\alpha}_n, \widehat{\beta}'_n\right) = \left(\widehat{\alpha}(H_n), \widehat{\beta}'(H_n)\right),$$

donde  $H_n$  es la función de distribución empírica de  $\{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  y  $(\widehat{\alpha}, \widehat{\beta}')$  es un funcional  $(\widehat{\alpha}, \widehat{\beta}') : \mathcal{F}_n \rightarrow \mathbb{R}^{p+1}$ , donde  $\mathcal{F}_n$  denota el conjunto de todas las posibles distribuciones en  $\mathbb{R}^{p+1}$ .

Asumiremos que el estimador es consistente en el sentido de Fisher, es decir, que bajo la distribución verdadera  $H_0$  (2.2) toma los valores poblacionales de los parámetros en (2.1).

$$\begin{aligned}\widehat{\alpha}(H_{\alpha_0, \beta'_0}) &\equiv \alpha_0 \\ \widehat{\beta}'(H_{\alpha_0, \beta'_0}) &\equiv \beta_0\end{aligned}$$

Nuestro propósito es estudiar estimadores para los parámetros del modelo de regresión dado en (2.1), cuando se producen alejamientos del modelo central descrito en la sección 2.1. Con ese propósito definiremos los entornos de contaminación.

**Definición 2.1 (Entorno de contaminación)** *Sea  $\varepsilon > 0$ , luego el entorno de contaminación del modelo central  $H_0$  de tamaño  $\varepsilon$ , que denotaremos por  $\nu_{H_0, \varepsilon}$  se define por*

$$\nu_{H_0, \varepsilon} = \{H = (1 - \varepsilon) H_0 + \varepsilon H^* \text{ donde } H^* \text{ es una distribución arbitraria}\}. \quad (2.5)$$

*De este modo se asume que la contaminación se puede dar tanto en  $G_0$  como en  $F_0$ .*

**Observación 2.2** *En algunos casos resulta complicado estudiar el comportamiento del estimador en un entorno de contaminación tan general y por lo tanto, se lo estudia en un subconjunto de  $\nu_{H_0, \varepsilon}$  por ejemplo, considerando solo contaminaciones puntuales,*

$$\nu_{H_0, \varepsilon}^* = \{H = (1 - \varepsilon) H_0 + \varepsilon \delta_{y_0, \mathbf{x}_0}\},$$

donde  $\delta_{y_0, \mathbf{x}_0}$  es la masa puntual en el punto  $(y_0, \mathbf{x}_0)$  arbitrario.

Intuitivamente un estimador robusto es aquel que no es muy influenciado por un pequeño número de observaciones atípicas. Estos estimadores no buscan ajustar todos los datos sino solamente aquellos que siguen la distribución del modelo central. Una medida importante que permite cuantificar la robustez de un estimador es el máximo sesgo asintótico.

**Definición 2.3 (Sesgo asintótico)** *Dado un funcional de estimación de regresión  $(\hat{\alpha}, \hat{\beta}')$  y  $H \in \nu_{H_0, \varepsilon}$ , los sesgos de  $\hat{\alpha}$  y  $\hat{\beta}$  se definen por*

$$b(\hat{\beta}, H) = \left( \left( \hat{\beta}(H) - \beta_0 \right)' A(G_0) \left( \hat{\beta}(H) - \beta_0 \right) \right)^{1/2}, \quad (2.6)$$

$$b(\hat{\alpha}, H) = \frac{|\hat{\alpha}(H) - \alpha_0|}{s(F_0)} \quad (2.7)$$

donde  $A(G)$  es un funcional de covarianza afín equivariante, i.e. si  $\mathbf{x}$  tiene distribución  $G$  y llamamos  $\tilde{\mathbf{x}} = B\mathbf{x} + \mathbf{c} \sim \tilde{G}$  para alguna matriz  $B$  de  $p \times p$  no singular y un vector  $\mathbf{c} \in \mathbb{R}^p$ , luego  $A(\tilde{G}) = BA(G)B'$  y donde  $s(F)$  es un funcional de escala. El sesgo total de  $(\hat{\alpha}, \hat{\beta}')$  está dado por

$$b(\hat{\alpha}, \hat{\beta}, H) = \left( b^2(\hat{\beta}, H) + b^2(\hat{\alpha}, H) \right)^{1/2}.$$

**Observación 2.4** *En el caso particular en el que  $A(G_0) = I$ ,  $s(F_0) = 1$ ,  $\alpha = 0$  y  $\beta = \mathbf{0}$  tenemos que  $b(\hat{\beta}', H) = \|\hat{\beta}(H)\|$  y  $b(\hat{\alpha}, H) = |\hat{\alpha}(H)|$ .*

**Definición 2.5 (Máximo sesgo asintótico)** *Dado un funcional de regresión  $(\hat{\alpha}, \hat{\beta}')$ , los máximos sesgos de  $\hat{\alpha}$ ,  $\hat{\beta}$  y  $(\hat{\alpha}, \hat{\beta}')$  se definen respectivamente por*

$$B(\hat{\alpha}, \varepsilon, H_0) = \sup_{H \in \nu_{H_0, \varepsilon}} b(\hat{\alpha}, H)$$

$$B(\hat{\beta}, \varepsilon, H_0) = \sup_{H \in \nu_{H_0, \varepsilon}} b(\hat{\beta}, H)$$

$$B(\widehat{\alpha}, \widehat{\beta}, \varepsilon, H_0) = \sup_{H \in \nu_{H_0, \varepsilon}} b(\widehat{\alpha}, \widehat{\beta}, H)$$

Si bien el máximo sesgo asintótico es una medida conveniente para medir la robustez de un estimador, en muchos casos resulta muy difícil calcularlo. Por esta razón se lo aproxima por una función lineal de  $\varepsilon$ . Para esto, Hampel (1974) introduce la función de influencia que mide el sesgo producido por una contaminación infinitesimal.

**Definición 2.6 (Función de Influencia)** *Dado un funcional de estimación de regresión  $(\widehat{\alpha}, \widehat{\beta}')$ , las funciones de influencia de los estimadores  $\widehat{\alpha}, \widehat{\beta}$  en una distribución  $H$  están dados respectivamente por*

$$IF(\mathbf{z}; \widehat{\alpha}, H) = \lim_{\varepsilon \rightarrow 0} \frac{\widehat{\alpha}((1 - \varepsilon)H + \varepsilon\delta_{\mathbf{z}}) - \widehat{\alpha}(H)}{\varepsilon}$$

$$IF(\mathbf{z}; \widehat{\beta}, H) = \lim_{\varepsilon \rightarrow 0} \frac{\widehat{\beta}((1 - \varepsilon)H + \varepsilon\delta_{\mathbf{z}}) - \widehat{\beta}(H)}{\varepsilon}$$

A partir de la función de influencia se puede definir la *sensibilidad a errores groseros* y el punto de ruptura.

**Definición 2.7 (Sensibilidad a errores groseros)** *Dado un funcional de estimación de regresión  $(\widehat{\alpha}, \widehat{\beta}')$  la sensibilidad a errores groseros de  $\widehat{\alpha}, \widehat{\beta}$  y  $(\widehat{\alpha}, \widehat{\beta}')$  se definen respectivamente por*

$$\gamma(\widehat{\alpha}, H_0) = \sup_{\mathbf{z}} \frac{|IF(\mathbf{z}; \widehat{\alpha}, H_0)|}{s(F_0)}, \quad (2.8)$$

$$\gamma(\widehat{\beta}, H_0) = \sup_{\mathbf{z}} \left( IF(\mathbf{z}; \widehat{\beta}, H_0)' A(G_0) IF(\mathbf{z}; \widehat{\beta}, H_0) \right)^{1/2}, \quad (2.9)$$

y

$$\gamma(\widehat{\alpha}, \widehat{\beta}, H_0) = \sup_{\mathbf{z}} \left( IF(\mathbf{z}; \widehat{\beta}, H_0)' A(G_0) IF(\mathbf{z}; \widehat{\beta}, H_0) + \frac{IF^2(\mathbf{z}; \widehat{\alpha}, H_0)}{s^2(F_0)} \right)^{1/2}. \quad (2.10)$$

Bajo condiciones de regularidad y cuando la sensibilidad sea finita, He y Simpson (1993) probaron que la sensibilidad a errores groseros aproxima linealmente a máximo sesgo para valores suficientemente pequeños de  $\varepsilon$ . Es decir se tiene que

$$B(\widehat{\alpha}, \varepsilon, H_0) = \gamma(\widehat{\alpha}, H_0) \varepsilon + o(\varepsilon).$$

$$B(\widehat{\beta}, \varepsilon, H_0) = \gamma(\widehat{\beta}, H_0) \varepsilon + o(\varepsilon).$$

$$B(\widehat{\alpha}, \widehat{\beta}, \varepsilon, H_0) = \gamma(\widehat{\alpha}, \widehat{\beta}, H_0) \varepsilon + o(\varepsilon).$$

Estos no se puede aplicar al caso de los estimadores de tipo M (en particular los estimadores de tipo MM) ya que estos tiene sensibilidad a errores groseros infinita.

El punto de ruptura  $\varepsilon^*$ , intuitivamente representa el menor porcentaje de contaminación que puede causar que el estimador tome valores arbitrariamente grandes. Por lo tanto cuando la fracción de contaminación es mayor que  $\varepsilon^*$  el estimador puede ser totalmente no informativo.

**Definición 2.8 (Punto de ruptura asintótico)** *Consideremos el estimador de regresión  $(\hat{\alpha}, \hat{\beta}')$  y sea  $B(\hat{\alpha}, \hat{\beta}, \varepsilon, H_0)$  el máximo sesgo asintótico para una contaminación de tamaño  $\varepsilon$  llamamos punto de ruptura,  $\varepsilon^*$*

$$\varepsilon^* := \sup \left\{ \varepsilon \in [0, 1] \text{ tal que } B(\hat{\alpha}, \hat{\beta}, \varepsilon, H_0) < \infty \right\}. \quad (2.11)$$

Donoho y Huber (1983) introdujeron el concepto de punto de ruptura para muestras finitas. Sea  $\mathbf{Z}_n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  una muestra de tamaño  $n$ , y sea  $(\hat{\alpha}_n, \hat{\beta}'_n)$  un estimador. Luego consideremos el conjunto  $\mathcal{Z}_m$  de muestras que se pueden obtener reemplazando  $m$  ( $m < n$ ) observaciones en  $\mathbf{Z}_n$

**Definición 2.9 (Punto de ruptura para muestras finitas)** *El punto de ruptura de un estimador  $(\hat{\alpha}, \hat{\beta}')$  de regresión está dado por*

$$\varepsilon_n^* = \min \left\{ \frac{m}{n} \left| \sup_{\mathbf{z}_n^*} \left\| (\hat{\alpha}, \hat{\beta}')(\mathbf{Z}_n) - (\hat{\alpha}, \hat{\beta}')(\mathbf{Z}_n^*) \right\| = \infty \right. \right\}.$$

## 2.3 Propiedades de equivarianza de los estimadores

Sea  $(\hat{\alpha}_n, \hat{\beta}'_n)$  un estimador de regresión, basado en una muestra de tamaño  $n$ . Una propiedad deseable es que sea equivariante respecto de las transformaciones que dejan el problema invariante. En regresión se pueden definir tres formas de equivarianza.

**Definición 2.10 (Estimadores equivariantes por regresión)** *Un estimador de regresión  $(\hat{\alpha}_n, \hat{\beta}'_n)$  es equivariante por regresión si para todo  $\mathbf{v} \in \mathbb{R}^p$*

$$\begin{aligned} \hat{\beta}'_n((\mathbf{x}_1, y_1 - \mathbf{x}'_1 \mathbf{v}), \dots, (\mathbf{x}_n, y_n - \mathbf{x}'_n \mathbf{v})) &= \hat{\beta}'_n((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)) + \mathbf{v}. \\ \hat{\alpha}_n((\mathbf{x}_1, y_1 - \mathbf{x}'_1 \mathbf{v}), \dots, (\mathbf{x}_n, y_n - \mathbf{x}'_n \mathbf{v})) &= \hat{\alpha}_n((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)). \end{aligned}$$

En el modelo de posición, esta propiedad es equivalente a ser invariante por traslaciones.

**Definición 2.11 (Estimadores equivariantes por escala)** *Un estimador de regresión  $(\hat{\alpha}_n, \hat{\beta}'_n)$  es equivariante por escala si para cualquier constante  $c$*

$$\begin{aligned}\hat{\alpha}_n((\mathbf{x}_1, cy_1), \dots, (\mathbf{x}_n, cy_n)) &= c\hat{\alpha}_n((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)). \\ \hat{\beta}_n((\mathbf{x}_1, cy_1), \dots, (\mathbf{x}_n, cy_n)) &= c\hat{\beta}_n((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)).\end{aligned}$$

Si un estimador es equivariante por escala el ajuste no dependerá de la unidad en que se mide la variable de respuesta.

**Definición 2.12 (Estimadores afín equivariantes)** *Un estimador de regresión  $(\hat{\alpha}_n, \hat{\beta}'_n)$  es afín equivariante si para cualquier matriz  $\mathbf{A}$  no singular se tiene*

$$\begin{aligned}\hat{\beta}_n((\mathbf{x}_1\mathbf{A}, y_1), \dots, (\mathbf{x}_n\mathbf{A}, y_n)) &= \mathbf{A}^{-1}\hat{\beta}_n((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)). \\ \hat{\alpha}_n((\mathbf{x}_1\mathbf{A}, y_1), \dots, (\mathbf{x}_n\mathbf{A}, y_n)) &= \hat{\alpha}_n((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)).\end{aligned}$$

## 2.4 Algunas propuestas de estimadores robustos de regresión

Se han presentado diferentes propuestas para estimadores robustos para el modelo lineal. Alcanzar el objetivo ideal sería encontrar estimadores resistente a la presencia de outliers, es decir, que tenga bajo punto de ruptura y sesgo asintótico y simultáneamente sean altamente eficientes cuando la distribución de los errores es normal. Por otra parte, es importante que el estimador sea factible de ser calculado numéricamente y que cumpla las propiedades de equivarianza. También es conveniente que sea asintóticamente normal de manera de facilitar la inferencia.

A continuación haremos una recopilación de las propuestas de estimación robustas para modelo lineal que nos resultan de especial interés y con las cuales trabajaremos en los capítulos siguientes.

### 2.4.1 Estimador $L_1$ de regresión

Antes de describir este estimador creemos que es conveniente realizar la siguiente observación. Se puede pensar el modelo de posición como un caso particular del modelo de regresión, si consideramos el modelo lineal con intercept y  $p = 0$  (2.1), i.e.,

$$y_i = \alpha + u_i \quad i = 1, \dots, n. \quad (2.12)$$

Por lo tanto, los residuos en este caso son  $r_i(\alpha) = y_i - \alpha$ .

Si no hubiera datos atípicos y la distribución de los errores fuese normal centrada en cero y con varianza  $\sigma^2$ , resultaría fácil ver que el estimador óptimo para el parámetro  $\alpha$  es la media muestral. Sin embargo, al estar ante la presencia de datos atípicos este estimador puede tomar valores arbitrariamente grandes pues tiene punto de ruptura nulo. Si en lugar de minimizar la suma del cuadrado de los residuos se minimiza la suma de los valores absolutos de los residuos el estimador de  $\alpha$  es la mediana muestral, que tiene punto de ruptura  $1/2$ . Huber (1964) mostró que la mediana tiene mínimo sesgo asintótico cuando la distribución de los errores es simétrica y unimodal. Por lo tanto, se puede considerar que resuelve satisfactoriamente el problema de robustez si la muestra es muy grande y no estamos interesados en la eficiencia.

Esto nos sugiere a definir un estimador para el modelo lineal dado en (2.1) por la minimización de los valores absolutos de los residuos dados por (2.4). Luego podemos dar la siguiente definición.

**Definición 2.13 (Estimador  $L_1$ )** *El estimador  $L_1$  de regresión se define por*

$$\left(\hat{\alpha}, \hat{\beta}'\right) = \arg \min_{\alpha, \beta} \sum_{i=1}^n |r_i(\alpha, \beta)|.$$

Este estimador fue propuesto originariamente por Laplace y Boscovich y es la primera propuesta de estimador robusto (ver Edgeworth, 1887).

Si bien este método es robusto frente a outliers de bajo leverage, no es robusto a outliers con alto leverage. Un solo outlier con alto leverage puede llevar al estimador fuera de todo límite y por lo tanto su punto de ruptura es 0.

En lo que respecta al cálculo numérico de estos estimadores se puede decir que si bien no existen ecuaciones explícitas se han hallado métodos numéricos que hacen que su cálculo sea factible (Barrodale y Roberts, 1973) y (Portnoy y Koenker, 1997) aún cuando se tenga un número grande de observaciones y de variables.

#### 2.4.2 Estimadores de tipo M de regresión

Sean  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  una muestra de observaciones  $p + 1$  dimensionales independientes e idénticamente distribuidas que satisfacen el modelo 2.1. Huber (1973) introduce la familia de los estimadores de tipo M de regresión en forma similar al estimador de mínimos cuadrados, pero reemplazando la función cuadrado por otra que crezca más lentamente cuando el residuo se aleje de cero.

**Definición 2.14 (Estimadores de tipo M de regresión)** *Los estimadores de tipo M de regresión están definidos por*

$$\left(\widehat{\alpha}, \widehat{\beta}'\right) = \arg \min_{\beta} \sum_{i=1}^n \rho(r_i(\alpha, \beta)). \quad (2.13)$$

donde  $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$ , tiene las siguientes propiedades: es par  $\rho(u) = \rho(-u)$ , es monótona no decreciente en  $|u|$ .

Esta familia de estimadores contiene como casos particulares el estimador de mínimos cuadrados ( $\rho(u) = u^2$ ) y el estimador  $L_1$  ( $\rho(u) = |u|$ ).

Huber (1964), trata el problema de posición y propone elegir la función  $\rho$  dentro de la siguiente familia (son todas funciones intermedias entre las funciones cuadrática y valor absoluto)

$$\rho_c^H(u) = \begin{cases} u^2 & \text{si } |u| \leq c \\ 2c|u| - c^2 & \text{si } |u| > c \end{cases}, \quad (2.14)$$

donde  $c$  es una constante que determina la eficiencia y el grado de robustez del estimador. Esta función es continua y derivable aun en los puntos  $c$  y en  $-c$ . Sin embargo, como no son acotadas los estimadores correspondientes resultan no robustos frente a puntos atípicos de alto leverage.

Para obtener un estimador que sea resistente a todo tipo de outliers se debe usar una función  $\rho$  acotada. En resumen, para obtener estimadores robustos frente a todo tipo de observaciones atípicas debemos considerar funciones  $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$ , satisfaciendo las siguientes propiedades.

**A1** La función  $\rho$  debe ser par, es decir  $\rho(u) = \rho(-u)$  para todo  $u$ .

**A2** Se debe cumplir que  $\rho(0) = 0$  y  $\lim_{u \rightarrow \infty} \rho(u) = c_\rho < \infty$ .

**A3** La función debe ser no decreciente en  $[0, \infty)$ .

Una familia de funciones  $\rho$  que cumple con estas propiedades es la bicuadrada introducida por Tukey, Esta familia está dada por

$$\rho_c^B(u) = \begin{cases} 3(u/c)^2 - 3(u/c)^4 + (u/c)^6 & \text{si } |u| \leq c \\ 1 & \text{si } |u| > c \end{cases}, \quad (2.15)$$

La constante  $c$  determina la eficiencia asintótica del estimador como así también el grado de robustez.

Sin embargo, los estimadores de tipo M de regresión definidos por (2.13) no son equivariante por escala. Con el objetivo de conseguir un estimador

que tenga esta propiedad se deben estandarizar los residuos,  $r_i(\alpha, \boldsymbol{\beta})$ , por un estimador de escala de los errores, es decir los estimadores de tipo M de regresión se deben definir por

$$\left(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}'\right) = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho \left( \frac{r_i(\alpha, \boldsymbol{\beta})}{\widehat{\sigma}} \right), \quad (2.16)$$

donde  $\widehat{\sigma}$  es un estimador de escala de los  $u_i$ 's.

La función de influencia para los estimadores de tipo M de regresión (Hampel et al., 1986) está dada por

$$IF(\mathbf{x}, y; (\alpha, \boldsymbol{\beta}), H_0) = \psi \left( \frac{r(\alpha, \boldsymbol{\beta})}{\sigma} \right) M^{-1}(1, \mathbf{x}), \quad (2.17)$$

donde  $\psi = \rho'$ ,  $M = E_{H_0}(\psi') E_{H_0}[(1, \mathbf{x}')'(1, \mathbf{x}')]^{-1}$  y  $\sigma$  es el valor límite de  $\widehat{\sigma}$ . Como se ve por esta expresión, la función de influencia de los estimadores de tipo M de regresión no es acotada, por lo tanto la sensibilidad a errores groseros para los estimadores de tipo M de regresión (Hampel et al., 1986) es infinita.

Bajo condiciones generales de regularidad de la función  $\rho$  y si  $E(\|\mathbf{x}\|^2)$  es finita los estimadores de tipo M de regresión tienen distribución asintótica normal (Maronna, Martin y Yohai, 2006), más precisamente se tiene que

$$\sqrt{n}((\alpha_n, \boldsymbol{\beta}'_n) - (\alpha_0, \boldsymbol{\beta}'_0)) \rightarrow^D N(0, AV(\psi, H_0)), \quad (2.18)$$

donde

$$AV(\psi, F_0, G_0) = \sigma^2 \frac{E_{F_0}(\psi^2(u))}{[E_{F_0}(\psi'(u))]^2} E_{G_0} \left[ ((1, \mathbf{x}')'(1, \mathbf{x}'))^{-1} \right]. \quad (2.19)$$

### Algoritmos para el cálculo numérico de los estimadores de tipo M.

Consideraremos el caso de estimadores de tipo M de regresión de regresión con escala desconocida. En virtud de **A3**, derivando respecto de  $\alpha$ ,  $\boldsymbol{\beta}$  y  $\sigma$  e igualando a cero; se obtienen las siguientes ecuaciones que son análogas a las ecuaciones normales en el caso de los estimadores de mínimos cuadrados,

$$\sum_{i=1}^n \psi \left( \frac{r_i(\widehat{\alpha}, \widehat{\boldsymbol{\beta}})}{\widehat{\sigma}} \right) (1, \mathbf{x}'_i)' = 0. \quad (2.20)$$

En (2.20) queda un sistema de  $p + 1$  ecuaciones que en general no tiene solución única. A continuación describiremos un método numérico iterativo



para hallar los estimadores suponiendo conocido  $\hat{\sigma}$  y un estimador inicial  $(\hat{\alpha}_0, \hat{\beta}'_0)$ . Este estimador inicial puede ser un estimador  $L_1$ , aunque en este caso el estimador resultante no será robusto si se presentan outliers con alto leverage. Más adelante describiremos un algoritmo que permite hallar estimadores iniciales robustos contra cualquier tipo de outliers.

El estimador de escala  $\hat{\sigma}$  puede ser cualquier escala robusta de los residuos correspondientes a  $(\hat{\alpha}_0, \hat{\beta}'_0)$ . Por ejemplo puede tomarse

$$\hat{\sigma} = \frac{1}{0.675} \text{mediana}_{1 \leq i \leq n} \left( r_i \left( \hat{\alpha}_0, \hat{\beta}'_0 \right) \right).$$

Podemos reescribir la ecuación (2.20) como

$$\sum_{i=1}^n \psi \left( \frac{r_i \left( \hat{\alpha}, \hat{\beta} \right)}{\hat{\sigma}} \right) (1, \mathbf{x}'_i)' = \sum_{i=1}^n \frac{\psi \left( r_i \left( \hat{\alpha}, \hat{\beta} \right) / \hat{\sigma} \right)}{r_i \left( \hat{\alpha}, \hat{\beta} \right) / \hat{\sigma}} \frac{r_i \left( \hat{\alpha}, \hat{\beta} \right)}{\hat{\sigma}} (1, \mathbf{x}'_i)' = 0. \quad (2.21)$$

Luego, si llamamos

$$W(r) = \begin{cases} \psi(r)/r & \text{si } r \neq 0 \\ \psi'(0) & \text{si } r = 0 \end{cases}, \quad (2.22)$$

(2.21) se obtiene

$$\sum_{i=1}^n W \left( \frac{r_i \left( \hat{\alpha}, \hat{\beta} \right)}{\hat{\sigma}} \right) r_i \left( \hat{\alpha}, \hat{\beta} \right) (1, \mathbf{x}'_i)' = 0.$$

Esto sugiere el siguiente algoritmo iterativo para calcular los estimadores de tipo M de regresión.

Se parte del estimador inicial  $(\hat{\alpha}_0, \hat{\beta}'_0)$ . Suponiendo que ya se ha calculado  $(\hat{\alpha}_j, \hat{\beta}'_j)$ , el siguiente valor  $(\hat{\alpha}_{j+1}, \hat{\beta}'_{j+1})$  se obtiene resolviendo

$$\sum_{i=1}^n w_i^{(j)} r_i \left( \hat{\alpha}, \hat{\beta} \right) (1, \mathbf{x}'_i) = 0,$$

donde los pesos  $w_i^{(j)}$  vienen dados por

$$w_i^{(j)} = W \left( \frac{r_i \left( \hat{\alpha}_j, \hat{\beta}'_j \right)}{\hat{\sigma}} \right), \quad (2.23)$$

lo cual equivale a mínimos cuadrados pesados donde la observación  $i$  recibe peso  $w_i^N$ .

El algoritmo se continua mientras que

$$\frac{\left\| \left( \widehat{\alpha}_N, \widehat{\beta}'_N \right) - \left( \widehat{\alpha}_{N-1}, \widehat{\beta}'_{N-1} \right) \right\|}{\left\| \left( \widehat{\alpha}_{N-1}, \widehat{\beta}'_{N-1} \right) \right\|} > \delta,$$

donde  $\delta$  es una constante de tolerancia. Este procedimiento se denomina algoritmo de mínimos cuadrados pesados iterados.

Si  $\psi$  es no decreciente ( $\rho$  convexa) la ecuación (2.20) tiene solución que es esencialmente única y por lo tanto, el punto de partida sólo afecta al número de iteraciones necesarios para que el algoritmo converja. En cambio, si  $\psi$  corresponde a una  $\rho$  acotada, la solución puede no ser única y por lo tanto el algoritmo descrito puede converger a soluciones distintas de acuerdo al valor inicial. En estos casos resulta crucial elegir de manera robusta el estimador inicial.

### 2.4.3 Estimadores de tipo S de regresión

Consideramos  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  una muestra de observaciones  $p+1$  dimensionales independientes e idénticamente distribuidas que satisfacen el modelo (2.1) y sean sus residuos,  $r_i(\alpha, \beta) = y_i - \alpha - \beta' \mathbf{x}_i$ .

**Definición 2.15 (Escala de tipo M)** Sea  $u_1, \dots, u_n$  una muestra de tamaño  $n$ . Dada una función  $\rho$  que satisfaga **A1–A3** y  $b \in [0, < \max_t \rho(t)]$  se define una escala  $S_n(u_1, \dots, u_n)$  de tipo M como la solución de la ecuación

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{u_i}{s}\right) = b. \quad (2.24)$$

En general se elige  $b = E_{\Phi}(\rho)$ , donde  $\Phi$  es la distribución normal estándar. Con esa definición se puede obtener un estimador de escala consistente al desviación estándar cuando los  $u_i$  constituyen una muestra aleatoria de una distribución normal.

Si  $\rho(t)$  es la función salto

$$\rho(t) = I_{(|t|>c)},$$

la escala será la mediana de las observaciones  $S_n(u_1, \dots, u_n) = (\text{mediana}_{1 \leq i \leq n} u_i)/c$ .

Rousseeuw y Yohai (1984) definieron los estimadores de tipo S de regresión que se basan en minimizar una M-escala robusta de los residuos (2.4).

**Definición 2.16 (Estimadores de tipo S de regresión)** *Los estimadores de tipo S de regresión se definen por,*

$$\left(\hat{\alpha}, \hat{\beta}'\right) = \arg \min_{\alpha, \beta} S_n(r_1(\alpha, \beta), \dots, r_n(\alpha, \beta)), \quad (2.25)$$

donde,  $S_n$  es una escala de tipo M definida como en (2.24).

Sea  $\mathbf{r}(\alpha, \beta) = (r_1(\alpha, \beta), \dots, r_n(\alpha, \beta))$ . Si la ecuación (2.25) tiene más de una solución entonces definiremos  $S_n(\mathbf{r}(\alpha, \beta))$  como el supremo de todas ellas, es decir,

$$S_n(\mathbf{r}(\alpha, \beta)) = \sup \left\{ \sigma; \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{r_i(\alpha, \beta)}{\sigma} \right) = b \right\}.$$

Si el conjunto de soluciones es vacío tomamos  $S_n(\mathbf{r}(\alpha, \beta)) = 0$ . Como se puede observar los estimadores de tipo S de regresión no dependen de un estimador de escala previo.

El punto de ruptura asintótico de un estimador de tipo S de regresión (Rousseeuw y Yohai 1984) está dado por

$$\varepsilon^* = \min \{b, 1 - b\}.$$

Una observación importante es que los estimadores de tipo S se pueden considerar también estimadores de tipo M. En efecto ellos están dados por (2.16) donde  $\hat{\sigma}$  está dada por  $\hat{\sigma} = S_n(\mathbf{r}(\hat{\alpha}, \hat{\beta}))$ . En efecto, sea  $(\alpha, \beta')$  un vector arbitrario de parámetros. Por definición,

$$S_n(\mathbf{r}(\alpha, \beta)) \geq S_n(\mathbf{r}(\hat{\alpha}, \hat{\beta})),$$

notaremos  $\hat{\sigma} = S_n(\mathbf{r}(\hat{\alpha}, \hat{\beta}))$ , luego como  $S_n(\mathbf{r}(\alpha, \beta))$  satisface

$$\sum_{i=1}^n \rho \left( \frac{r_i(\alpha, \beta)}{S_n(\mathbf{r}(\alpha, \beta))} \right) = b,$$

y  $\rho(u)$  es no decreciente en  $|u|$ , tenemos que

$$\sum_{i=1}^n \rho \left( \frac{r_i(\alpha, \beta)}{\hat{\sigma}} \right) \geq b,$$

y que la igualdad se alcanza en  $(\hat{\alpha}, \hat{\beta}) = (\alpha, \beta)$ . Por lo tanto,

$$\left(\hat{\alpha}, \hat{\beta}'\right) = \arg \min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{r_i(\alpha, \beta)}{\hat{\sigma}} \right). \quad (2.26)$$

De acuerdo a los resultados para estimadores de regresión de tipo M bajo condiciones generales de regularidad de  $\rho$  y de  $F_0$  se tendrá que

$$\sqrt{n} \left( (\hat{\alpha}_n, \hat{\beta}'_n) - (\alpha_0, \beta'_0) \right) \rightarrow^D N(0, AV(\psi, F_0, G_0)), \quad (2.27)$$

donde la matriz de covarianza asintótica es

$$AV(\psi, F_0, G_0) = \sigma^2 \frac{E_{F_0}(\psi^2(y))}{[E_{F_0}(\psi'(u))]^2} E_{G_0} \left[ ((1, \mathbf{x}')'(1, \mathbf{x}'))^{-1} \right]. \quad (2.28)$$

En el caso de que usemos  $\rho$  en la familia biacuadrada de Tukey, la constante  $c$  puede elegirse de manera que su eficiencia asintótica

$$e = \frac{[E_{F_0}(\psi')]^2}{E_{F_0}(\psi^2)}, \quad (2.29)$$

tome un valor dado, por ejemplo 0.95. La principal desventaja de este estimador es que no puede ser simultáneamente altamente robusto y eficiente. Más adelante presentaremos un estimador que tenga ambas propiedades.

### Algoritmo para calcular los estimadores de tipo S de regresión

Como un estimador de tipo S de regresión es también un estimador de tipo M, puede calcularse usando un algoritmo iterativo de mínimos cuadrados ponderados similar al propuesto en la sección (2.4.2). La única diferencia es que la escala  $\hat{\sigma}$  se estima conjuntamente y por lo tanto debe actualizarse en cada iteración.

Por lo tanto los pesos  $w_i^{(j)}$  que se utilizan para calcular  $(\hat{\alpha}_j, \hat{\beta}'_j)$  se obtienen usando la siguiente fórmula recursiva

$$\hat{\sigma}^{(j)} = S_n(\mathbf{r}(\hat{\alpha}_j, \hat{\beta}_j)),$$

$$w_i^{(j)} = W \left( \frac{r_i(\hat{\alpha}_j, \hat{\beta}_j)}{\hat{\sigma}^{(j)}} \right).$$

En cuanto al estimador inicial se obtiene de la siguiente manera

En el caso de los estimadores de tipo S la función objetivo que se minimiza es la escala de los residuos. Una manera de aproximar el valor que realiza el mínimo de la función objetivo es obtener un conjunto finito, de  $N$  candidatos al cual restringiremos la búsqueda del mínimo de la función objetivo.

Cada candidato se obtiene en dos pasos. En el primer paso se obtiene una submuestra aleatoria de tamaño  $p + 1$  y se obtiene el hiperplano que los

contiene. En un segundo paso esta solución se la modifica de tal manera de disminuir la función objetivo.

El número de candidatos debe ser lo suficientemente grande como poder garantizar con probabilidad alta que llamaremos  $\eta$ , por ejemplo 0.99, de que entre ellos se encuentra un valor cercano al valor verdadero. Para que esto ocurra una condición necesaria pero no suficiente es que con probabilidad  $1 - \eta$  se pueda garantizar que al menos una muestra no contenga outliers.

Como veremos lamentablemente, el número de submuestras  $N$  que hay que tomar para que con probabilidad alta,  $(1 - \eta)$ , haya alguna sin observaciones atípicas crece exponencialmente a medida que aumenta la cantidad de covariables. Sea  $\varepsilon$  la proporción de datos atípicos en la muestra (este porcentaje tiene que ser menor que el punto de ruptura del estimador). La probabilidad de que una submuestra no tenga ningún outlier es  $(1 - \varepsilon)^{p+1}$ , por ende, la probabilidad de que haya alguna submuestra que no este contaminada entre  $N$  submuestras es

$$1 - \left[1 - (1 - \varepsilon)^{(p+1)}\right]^N.$$

Por lo tanto, para tener una submuestra sin outliers  $N$  debe ser tal que,

$$1 - \left[1 - (1 - \varepsilon)^{(p+1)}\right]^N > (1 - \eta),$$

despejando obtenemos un tamaño mínimo de submuestras está dado por

$$N > \frac{\ln(\eta)}{\ln\left(1 - (1 - \varepsilon)^{(p+1)}\right)}.$$

En la Tabla 2.1 se encuentran los tamaños mínimos de muestra si se considera  $\eta = 0.01$ .

Tabla 2.1 Cantidad mínima de submuestras

$p$	$\varepsilon$	
	<b>0.10</b>	<b>0.20</b>
<b>3</b>	5	9
<b>5</b>	7	16
<b>10</b>	13	52
<b>20</b>	41	497
<b>50</b>	991	403324

Por otra parte, el hecho de que una submuestra no esté contaminada no garantiza que el hiperplano que pase por esos puntos sea una buena solución inicial ya que la matriz determinada por esos puntos podría estar mal condicionada. Por lo tanto, siguiendo una idea de Rousseeuw y Van Driessen

(1999 y 2006) las soluciones obtenidas usando submuestras serán modificadas de tal manera que la función objetivos correspondiente disminuya.

Sean  $(\alpha_1, \beta'_1), \dots, (\alpha_N, \beta'_N)$  los  $N$  candidatos obtenidos ajustando hiperplanos a las submuestras. Para cada  $(\alpha_i, \beta'_i)$  llevaremos a cabo el siguiente procedimiento.

Paso 1 Calculamos la escala robusta de tipo M que obtenemos de la ecuación dada por (2.24) y que la notaremos por  $\tilde{s}_i = s \left( r \left( \tilde{\alpha}_i, \tilde{\beta}'_i \right) \right)$ .

Paso 2 Definimos los pesos  $w_i = w \left( s \left( r \left( \tilde{\alpha}_i, \tilde{\beta}'_i \right) \right) / \tilde{s}_i \right)$  donde  $w(u) = \psi(u)/u$  igual que en (2.22).

Paso 3 Utilizando estos pesos hallamos el estimador de mínimos cuadrados pesados obteniendo un nuevo candidato que llamaremos  $(\alpha_i^0, \beta_i^0)$ .

#### 2.4.4 Estimador de mínima mediana de cuadrados

Rousseeuw (1984) definió el estimador de mínima mediana de cuadrados (LMS) de regresión por

$$\left( \hat{\alpha}, \hat{\beta}' \right) = \arg \min_{\alpha, \beta} \text{mediana}_{1 \leq i \leq n} r_i^2(\alpha, \beta). \quad (2.30)$$

Las propiedades más importantes de este estimador es que es equivariante por regresión, escala y transformaciones afines y además que su punto de ruptura asintótico es 0.5 si se cumple que  $P(\gamma' \mathbf{x} = c) = 0$ , para todo  $\gamma \in \mathbb{R}^p$  y todo  $c \in \mathbb{R}$ . Su mayor desventaja es la falta de eficiencia ya que su tasa de convergencia es  $n^{1/3}$  en lugar de la habitual  $n^{1/2}$ .

**Observación 2.17** *El estimador LMS puede ser visto como un estimador de tipo S si se considera como función  $\rho$  de pérdida la función de salto dada por,*

$$\rho_c^S(u) = \begin{cases} 0 & \text{si } |u| < 1 \\ 1 & \text{si } |u| \geq 1 \end{cases}. \quad (2.31)$$

#### 2.4.5 Estimadores MM de regresión

Estos estimadores propuestos por Yohai (1987) tienen punto de ruptura 0.5 y su eficiencia bajo errores normales puede hacerse tan alta como se quiera. Por otra parte, los estimadores MM son asintóticamente normales.

Para obtener un estimador MM hay que seguir el siguiente procedimiento.

Etapa 1 Se obtiene un estimador robusto de regresión  $(\hat{\alpha}_0, \hat{\beta}'_0)$  con punto de ruptura 0.5. Este estimador puede ser ineficiente, por ejemplo un estimador de tipo S con punto de ruptura 0.5.

Etapa 2 Se computa una escala de tipo M dada por  $\hat{\sigma} = S_n(\mathbf{r}(\hat{\alpha}_0, \hat{\beta}'_0))$   $1 \leq i \leq n$  con punto de ruptura 0.5. Llamaremos  $\rho_1$  a la función en la que está basada esta escala.

Etapa 3 Se calcula un estimador de tipo M eficiente basado en una función  $\rho_2 \leq \rho_1$  usando como estimador inicial  $(\hat{\alpha}_0, \hat{\beta}'_0)$  y como escala  $\hat{\sigma}$ . Más precisamente se encuentra  $(\hat{\alpha}, \hat{\beta}')$  como una solución de

$$\sum_{i=1}^n \psi_2 \left( \frac{r_i(\hat{\alpha}, \hat{\beta})}{\hat{\sigma}} \right) (1, \mathbf{x}'_i) = 0, \quad (2.32)$$

donde  $\psi_2 = \rho'_2$  que satisface

$$\sum_{i=1}^n \rho_2 \left( \frac{r_i(\hat{\alpha}, \hat{\beta})}{\hat{\sigma}} \right) \leq \sum_{i=1}^n \rho_1 \left( \frac{r_i(\hat{\alpha}, \hat{\beta})}{\hat{\sigma}} \right).$$

Esta solución  $(\hat{\alpha}, \hat{\beta}')$  puede ser encontrada usando el algoritmo de mínimos cuadrados pesados iterados descrito en la Sección 2.4.2.

Se puede demostrar que el estimador de tipo M  $(\hat{\alpha}, \hat{\beta}')$  así obtenido conserva las propiedades de punto de ruptura igual a 0.5 que tiene el estimador  $(\hat{\alpha}_0, \hat{\beta}'_0)$  y simultáneamente es altamente eficiente.

#### 2.4.6 Estimadores de regresión basados en proyecciones

Maronna y Yohai (1993) definieron los estimadores de regresión basados en proyecciones, también llamados estimadores de tipo P. Para comenzar daremos algunas definiciones que serán de utilidad para definir estos estimadores.

Sea  $y$  en  $\mathbb{R}$  la variable de respuesta y  $\mathbf{x}$  en  $\mathbb{R}^p$  las variables regresoras. Si  $(\lambda_0, \boldsymbol{\lambda}')$  es un vector  $(p+1)$ -dimensional, diremos que  $(1, \mathbf{x})'(\lambda_0, \boldsymbol{\lambda}')$  es la proyección de  $(1, \mathbf{x})$  en la dirección  $(\lambda_0, \boldsymbol{\lambda}')$ .

Por otra parte, definiremos un estimador robusto de escala, consideremos una muestra aleatoria  $x_1, \dots, x_n$  llamaremos *mediana de los residuos absolutos* (mad),

$$\text{mad}(x_1, \dots, x_n) = \text{mediana}_{1 \leq i \leq n} |x_i - \text{mediana}_{1 \leq j \leq n}(x_j)|. \quad (2.33)$$

Si consideramos el modelo lineal dado por (2.1) con los mismos supuestos que en la sección 2.1. Llamemos  $A(\alpha, \beta, \lambda_0, \lambda)$  a la mediana del cociente de los residuos de,  $r_i(\alpha, \beta) = y_i - \mathbf{x}'_i \beta - \alpha$ , y las proyecciones de las variables regresoras en la dirección  $(\lambda_0, \lambda)$ .

$$A(\alpha, \beta, \lambda_0, \lambda) = \text{mediana}_{1 \leq i \leq n} \left( \frac{r_i(\alpha, \beta)}{(1, \mathbf{x}'_i)(\lambda_0, \lambda)'} \right). \quad (2.34)$$

Si  $(\alpha_0, \beta'_0)$  son los valores verdaderos de los parámetros entonces se tiene que  $A(\alpha_0, \beta'_0, \lambda_0, \lambda)$  se hace asintóticamente cero para cualquier  $(\lambda_0, \lambda')$  perteneciente a  $\mathbb{R}^{p+1}$ . Luego, el objetivo ideal sería hallar  $(\hat{\alpha}, \hat{\beta}')$  para los cuales

$$A(\hat{\alpha}, \hat{\beta}', \lambda_0, \lambda) = 0,$$

para cualquier  $(\lambda_0, \lambda')$  en  $\mathbb{R}^{p+1}$ . Sin embargo como esta ecuación puede no tener solución una forma posible de definir el estimador sería

$$(\hat{\alpha}, \hat{\beta}') = \arg \min_{(\alpha, \beta')} \max_{\|(\lambda_0, \lambda')\|=1} |A(\alpha, \beta, \lambda_0, \lambda)|.$$

Como estos estimadores no serían equivariantes, para solucionar este problema se los estimadores de tipo P de regresión se definen como

$$(\hat{\alpha}, \hat{\beta}') = \arg \min_{(\alpha, \beta')} \max_{\|(\lambda_0, \lambda')\|=1} |A(\alpha, \beta, (\lambda_0, \lambda))| \text{mad}((1, \mathbf{x}')(\lambda_0, \lambda)). \quad (2.35)$$

La propiedad más importante de estos estimadores es que son altamente robustos desde el punto de vista del sesgo. Se lo puede ver como una extensión del estimador de regresión univariado de mínimo máximo sesgo asintótico que está dado por

$$\text{mediana}_{1 \leq i \leq n} \left( \frac{y_i}{x_i} \right).$$

En efecto, Martin, Yohai y Zamar (1989) probaron que si la distribución de  $F_0$  es simétrica y unimodal luego este estimador es minimax en la clase de todos los estimadores equivariantes por regresión.

Es importante destacar si bien el orden de convergencia de estos estimadores es 0.5 su distribución asintótica es desconocida y se sabe que no es normal.

### **Algoritmo para el cálculo de estimadores de tipo P de regresión.**

Desde el punto de vista numérico el cálculo de este estimador es costoso. A continuación describiremos un algoritmo basado en submuestreo que permite encontrar una solución aproximada a este problema.



Sea  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  una muestra aleatoria de observaciones  $p + 1$ -dimensionales del modelo lineal dado por (2.1). En primer lugar elegimos in conjunto de  $N$  candidatos como en el caso de los estimadores de tipo S de regresión descrito en (2.4.3). De este modo obtenemos el conjunto de candidatos para  $(\hat{\alpha}, \hat{\beta}')$

$$\mathbf{B} = \left\{ \left( \hat{\alpha}_1, \hat{\beta}'_1 \right), \dots, \left( \hat{\alpha}_N, \hat{\beta}'_N \right) \right\}. \quad (2.36)$$

El número de candidatos también se puede determinar como en el caso de los estimadores de tipo S.

Las direcciones unitarias que se utilizan cuando se evalúa el candidato  $(\hat{\alpha}_i, \hat{\beta}'_i)$ ,  $1 \leq i \leq n$  son las del conjunto

$$\mathbf{L}_i = \left\{ \hat{\lambda}_{i,j} = \frac{\left( \hat{\alpha}_i, \hat{\beta}'_i \right) - \left( \hat{\alpha}_j, \hat{\beta}'_j \right)}{\left\| \left( \hat{\alpha}_i, \hat{\beta}'_i \right) - \left( \hat{\alpha}_j, \hat{\beta}'_j \right) \right\|} : j \neq i \right\}. \quad (2.37)$$

Luego el estimador aproximado de proyección que calcularemos será

$$\left( \hat{\alpha}^*, \hat{\beta}^{*'} \right) = \arg \min_{(\alpha, \beta') \in B} \max_{(\lambda_0, \lambda) \in L_i} |A(\alpha, \beta, \lambda_0, \lambda)| \text{mad} \left( (1, \mathbf{x}') (\lambda_0, \lambda)' \right).$$

Llamaremos

$$\gamma_{i,j} = \left| A \left( \hat{\alpha}_i, \hat{\beta}'_i, \hat{\lambda}_{i,j} \right) \right| \text{mad} \left( (1, \mathbf{x}') \hat{\lambda}_{i,j} \right).$$

y definimos

$$\gamma_i^* = \max_{j \neq i} \gamma_{i,j}.$$

y sea

$$i_0 = \arg_i \min \gamma_i^*.$$

Obsérvese que

$$\left( \hat{\alpha}^*, \hat{\beta}^{*'} \right) = \left( \hat{\alpha}_{i_0}, \hat{\beta}'_{i_0} \right).$$

Sea

$$\eta_i = \min_{k \leq i} \gamma_k^*$$

y

$$k_i = \min \{ k : \eta_i = \gamma_k^* \}.$$

Para calcular  $(\hat{\alpha}_{i_0}, \hat{\beta}'_{i_0})$  procedemos a calcular inductivamente  $\eta_i$  y  $k_i$  del siguiente modo.

1. Calculamos

$$\gamma_1^* = \max_{j \neq 1} \gamma_{1,j}$$

y

$$\eta_1 = \gamma_1^*.$$

2. Supongamos que ya hemos calculado  $\eta_i$  y  $k_i$ . Para calcular  $\eta_{i+1}$  y  $k_{i+1}$  procedemos a calcular  $\gamma_{i+1,j}$  comenzando desde  $j = 1$  en adelante. Si encontramos un  $j$  tal que  $\gamma_{i+1,j} \geq \eta_i$  podemos asegurar que  $\gamma_{i+1}^* \geq \eta_i$ . Por lo tanto paramos de calcular los  $\gamma_{i+1,j}$  y hacemos

$$\eta_{i+1} = \gamma_i^*, \quad k_{i+1} = k_i.$$

En caso contrario calculamos todos los  $\gamma_{i+1,j}$  con  $j \neq i + 1$  y luego  $\gamma_{i+1}^* = \max_{j \neq i+1} \gamma_{i+1,j}$ . En este caso hacemos

$$\eta_{i+1} = \gamma_{i+1}^*, \quad k_{i+1} = i + 1.$$

3. Luego se tendrá  $i_0 = k_N$  y el estimador aproximado de proyección será  $(\hat{\alpha}^*, \hat{\beta}^{*'}) = (\hat{\alpha}_{i_0}, \hat{\beta}'_{i_0})$ .

Obsérvese que de acuerdo a la descripción del paso 2, no es necesario calcular todos los  $\gamma_{i,j}$ , lo que resulta en un aumento muy importante en la velocidad del algoritmo.

## 3 Capítulo 2

### 3.1 Estimadores de tipo M de regresión combinando robustez y eficiencia

En este capítulo trataremos el problema de como hallar una función  $\rho$ , a partir de la cual se obtengan estimadores de tipo M de regresión que sean robustos en un entorno de contaminación y eficientes bajo el modelo central.

Notaremos por  $AV(\psi, H)$  a la varianza asintótica de un estimador de tipo M basado en la función  $\psi$  cuando la distribución de  $(y, \mathbf{x})$  es  $H$  y por  $B(T, \varepsilon, H)$ , al máximo sesgo asintótico del estimador  $T$  cuando la distribución central es  $H$  y hay una contaminación de tamaño  $\varepsilon$ .

### 3.2 El problema de optimalidad de Hampel para el modelo de posición

Hampel (1974) con el objetivo de encontrar un estimador que equilibrase el sesgo y la eficiencia consideró el siguiente problema para el modelo de posición. Supongamos que se tiene una muestra aleatoria  $y_1, \dots, y_n$  de

$$y = \alpha_0 + u_i,$$

donde  $u_i$  tiene distribución  $F_0$  simétrica. Este es un caso particular del modelo de regresión donde no hay variables explicativas ( $p = 0$ ) los estimadores de tipo M cuando la escala es conocida están definidos por

$$\hat{\alpha} = \arg \min_{\alpha} \sum_{i=1}^n \rho \left( \frac{y_i - \alpha}{\sigma} \right).$$

Hampel planteó el problema de encontrar el estimador de tipo M de menor varianza asintótica sujeto a una cota superior en la sensibilidad a errores groseros (2.8). Sea  $AV(\psi, F_0)$  es la varianza asintótica del estimador de tipo M con función de scores  $\psi$  cuando  $F_0$  es la distribución de los  $u_i$ . Luego se tiene que

$$AV(\psi, H_0) = \frac{\sigma^2 E_{F_0} \left( \psi^2 \left( \frac{y-\mu}{\sigma} \right) \right)}{\left[ E_{F_0} \left( \psi' \left( \frac{y-\mu}{\sigma} \right) \right) \right]^2}. \quad (3.1)$$

Por otro lado sensibilidad a errores groseros de un estimador de tipo M está dada por

$$\gamma(\psi, H_0) = \frac{\max_y |\psi(y)|}{\left| E_{F_0} \left( \psi' \left( \frac{y-\mu}{\sigma} \right) \right) \right|}. \quad (3.2)$$

Luego el problema de Hampel consiste en buscar  $\psi^*$  tal que

$$\psi^* = \arg \min_{\gamma(\psi, F_0) \leq K} AV(\psi, F_0).$$

Hampel probó que si  $F_0$  es simétrica  $\psi_c^*$  es de la siguiente forma

$$\psi_c^*(y) = \psi_c^H \left( \psi_0 \left( \frac{y}{\sigma} \right) \right), \quad (3.3)$$

donde  $\psi_c^H$  es la función de scores propuesta por Huber (2.14) y

$$\psi_0(y) = -\frac{f'_0(y)}{f_0(y)},$$

donde  $f_0$  es la densidad de  $F_0$ . El valor de  $K$  depende  $c$ . La elección de  $c$  (y luego de  $K$ ) debe hacerse como un compromiso ente eficiencia y robustez. Cuando menor es el  $c$  mayor es la varianza del estimador. Por otro lado hay un valor mínimo posible de  $c$  que se obtiene cuando se toma  $K$  tendiendo a 0. Por ejemplo si  $F_0$  es normal con media cero y varianza uno entonces el valor mínimo de  $K = 1/(2\varphi(0))$ ; donde  $\varphi(u)$  es la densidad de la normal estándar. Este valor se alcanza cuando se utiliza la mediana que es el estimador de tipo M correspondiente a tomar como  $\psi$  la función signo.

Es importante destacar que Huber resolvió el problema de hallar el estimador con mínima varianza asintótica en un entorno de contaminación en el que sólo considero contaminaciones simétricas (al considerar este tipo de contaminaciones el problema del sesgo desaparece) y la  $\psi$  óptima es también de la forma dada por (3.3).

El problema dual de Hampel consiste en encontrar la función  $\psi$  que minimiza la sensibilidad a errores groseros sujeta a una cota inferior en la varianza asintótica. Es fácil mostrar que la solución a este problema es también de la forma dada en (3.3).

### 3.3 El problema de Hampel para el modelo de regresión

En esta sección trataremos de plantear un problema análogo al de la sección anterior pero para el modelo de regresión. El primer problema que se presenta en este caso es que los estimadores de tipo M de regresión tienen función de influencia no acotada y por lo tanto, la sensibilidad a errores groseros tampoco es acotada, ya que es el supremo de esta última.

He y Simpson (1993) introdujeron el concepto de *sensibilidad a la contaminación* que se desprende del concepto de máximo sesgo.

Dado un funcional de estimación de regresión  $(\hat{\alpha}, \hat{\beta}')$ , la *sensibilidad a la contaminación* de  $\hat{\alpha}$  y de  $\hat{\beta}$  se definen respectivamente por

$$\gamma^*(\hat{\alpha}, F_0, G_0) = \frac{\partial}{\partial \varepsilon} B(\hat{\alpha}, \varepsilon, F_0, G_0)|_{\varepsilon=0}. \quad (3.4)$$

$$\gamma^*(\hat{\beta}, F_0, G_0) = \frac{\partial}{\partial \varepsilon} B(\hat{\beta}, \varepsilon, F_0, G_0)|_{\varepsilon=0}.$$

Este concepto permite aproximar el máximo sesgo para valores suficientemente pequeños de  $\varepsilon$

$$B(\hat{\alpha}, \varepsilon, F_0, G_0) = \gamma^*(\hat{\alpha}, F_0, G_0)\varepsilon + o(\varepsilon),$$

y

$$B(\hat{\beta}, \varepsilon, F_0, G_0) = \gamma^*(\hat{\beta}, F_0, G_0)\varepsilon + o(\varepsilon),$$

Hay dos diferencias entre  $\gamma$  y  $\gamma^*$ . La primera es el orden en que se toma el supremo y el límite cuando  $\varepsilon$  tiende a 0 y la segunda es que en el caso de  $\gamma^*$  se toma el supremo en un entorno de contaminación general,  $\nu_{H_0, \varepsilon}$ , mientras que en el caso (2.8) se restringe a contaminaciones puntuales. He y Simpson (1993) probaron que en general  $\gamma(T, H) \geq \gamma_T^*$ , y que bajo condiciones de regularidad la igualdad se alcanza. He (1991) probó que para una clase de estimadores llamada *estimadores de residuos admisibles*, dentro de los que se encuentran los estimadores de tipo M, de tipo S y el estimador LMS entre otros estimadores,  $B(\hat{\alpha}, \varepsilon, F_0, G_0)$  y  $B(\hat{\beta}, \varepsilon, F_0, G_0)$  son de orden mayor o igual que  $\sqrt{\varepsilon}$  en un entorno del cero y que por lo tanto,  $\gamma_T^* = \infty$ . Esto motivó que Yohai y Zamar (1997) extendieran la definición de *sensibilidad a la contaminación*.

**Definición 3.1 (Sensibilidad a la contaminación)** *La sensibilidad a la contaminación de orden  $\delta$ ,  $\gamma_T^\delta$ , definida por,*

$$\gamma^{(\delta)}(\hat{\alpha}, F_0, G_0) = \lim_{\varepsilon \rightarrow 0} \frac{B(\hat{\alpha}, \varepsilon, F_0, G_0)}{\varepsilon^\delta}, \quad (3.5)$$

$$\gamma^{(\delta)}(\hat{\beta}, F_0, G_0) = \lim_{\varepsilon \rightarrow 0} \frac{B(\hat{\beta}, \varepsilon, F_0, G_0)}{\varepsilon^\delta},$$

*cuando estos límites existen.*

En particular si  $\delta = 1$  se tiene la *sensibilidad a la contaminación*. Cuando  $\delta = 0.5$  podemos obtener la siguiente aproximación a los máximos sesgos de  $\hat{\alpha}$  y  $\hat{\beta}$ ,

$$B(\hat{\alpha}, \varepsilon, F_0, G_0) = \gamma^{(\delta)}(\hat{\alpha}, F_0, G_0)\varepsilon^\delta + o(\varepsilon^\delta),$$

y

$$B(\widehat{\beta}, \varepsilon, F_0, G_0) = \gamma^{(\delta)}(\widehat{\beta}, F_0, G_0)\varepsilon^\delta + o(\varepsilon^\delta).$$

Una forma natural de combinar robustez con eficiencia sería resolver el problema de encontrar el estimador de tipo M con escala general (2.16) que minimice la traza de la varianza asintótica bajo el modelo central sujeta a una cota en el máximo sesgo. Este problema parece muy difícil de tratar y solo será resuelto en el caso en que la distribución sea elíptica y la varianza sea conocida.

Martin y Zamar (1993) encontraron el estimador que minimiza la varianza asintótica bajo  $H_0$  sujeta a cota en el máximo sesgo para estimadores de tipo M de posición y encontraron que la solución era muy próxima a la del problema de Hampel. Esto sugiere que algo similar puede ocurrir en el caso del modelo de regresión.

Estas consideraciones motivaron que Yohai y Zamar (1997) resolvieran una generalización del problema de Hampel para estimadores de tipo M de regresión.

Hallar la función  $\psi$  tal que

$$\text{traza}(AV(\psi, F_0, G_0)) = \text{mínimo}, \quad (3.6)$$

sujeto a

$$\gamma^{(1/2)}(\widehat{\beta}, F_0, G_0) \leq k. \quad (3.7)$$

Como los estimadores son escala equivariantes podemos considerar que  $\sigma = 1$ . En ese caso en Yohai (1987) se demuestra que

$$AV(\psi, F_0, G_0) = \frac{E_{F_0}(\psi^2(u))}{\lambda^2(F_0)} E_{G_0}(\mathbf{xx}')^{-1},$$

donde

$$\lambda(F_0) = \int_{-\infty}^{\infty} \psi(y) f'_0(y) dy.$$

Por otro lado Yohai y Zamar (1997) probaron que

$$\gamma^{(1/2)}(\widehat{\beta}, F_0, G_0) = \sqrt{2\rho(\infty)/\lambda(F_0)}.$$

y que la solución del problema dado por (3.6) y (3.7) está dada por

$$\psi_\beta^*(u) = \text{signo}(y) \frac{q_\beta^+(|u|)}{\langle q_\beta^+, f'_0/f_0 \rangle} \quad \text{para } 0 \leq \beta \leq \beta_0,$$

donde  $f_0 = F'_0$ ,

$$q_\beta(u) = -\frac{f'_0(u) + \beta}{f_0(u)},$$

$q^+$  denota la parte positiva de  $q$ , y  $\langle \psi_1, \psi_2 \rangle = \int_0^\infty \psi_1(u) \psi_2(u) du$ .

### 3.4 Estimadores de tipo M de regresión óptimos

Consideremos el modelo dado en (2.1). Encontraremos el estimador  $(\hat{\alpha}, \hat{\beta}')$  de tipo M de  $(\alpha_0, \beta'_0)$  tal que el correspondiente estimador de las pendientes  $\hat{\beta}$  tenga máxima eficiencia asintótica sujeta a una cota en el máximo sesgo asintótico. No consideraremos el problema de encontrar un estimador de tipo M óptimo para  $\hat{\alpha}$  ya que como muestran Berrendero y Zamar (2001) en lugar de utilizar un estimador de tipo M para el intercept es mejor desde el punto de vista del sesgo calcular primero el estimador de tipo M y luego estimar el intercept mediante los residuos de  $y - \hat{\beta}'\mathbf{x}$  donde  $\hat{\beta}$  es el estimador de tipo M de la pendiente.

Haremos los siguientes supuestos:

- (i) El parámetro de escala  $\sigma_0$  es conocido y sin pérdida de generalidad lo podemos hacer igual a 1.
- (ii) Berrendero y Zamar (2001) obtuvieron el máximo sesgo asintótico para un estimador de regresión de tipo M con escala fija. Para establecer los resultados requirieron las siguientes hipótesis sobre la distribución de las variables regresoras  $G_0$ .

**B.1** La distribución de  $\mathbf{x}$  es elíptica y su densidad es de la forma

$$g(\mathbf{x}) = \det(\Sigma)^{-1/2} g_0((\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})).$$

Para simplificar el problema en el resto del Capítulo supondremos que el parámetro de escala  $\sigma_0$  es conocido y sin pérdida de generalidad lo podemos hacer igual a 1.

Sin pérdida de generalidad, (teniendo en cuenta las propiedades de equivarianza), en lo que sigue del capítulo supongamos que  $\Sigma = I$  y que  $\boldsymbol{\mu} = \mathbf{0}$ . Luego la condición **B1** implica que la distribución de  $\beta'\mathbf{x}$  depende sólo de  $\|\beta\|$  y por lo tanto lo mismo ocurre con la distribución de  $y - \beta'\mathbf{x}$ . Luego podemos definir

$$k_\rho(\|\beta\|) = E_{H_0}(\rho(y - \beta'\mathbf{x})).$$

En todos los resultados de este capítulo supondremos que se cumple **A1-A3** y **B1**

El siguiente lema probado en Martín Yohai y Zamar (1989) será de gran utilidad a lo largo del trabajo.

**Lema 3.2** Sea  $H_0$  es absolutamente continua con densidad  $f_0$  simétrica, continua monótona decreciente para valores positivos. Sea

$$g(s, \|\beta\|) = E_{H_0} \left( \rho \left( \frac{y - \beta' \mathbf{x}}{s} \right) \right),$$

entonces  $g$  es continua, estrictamente creciente como función de  $\|\beta\|$  y estrictamente decreciente como función de  $s$ . En particular  $g(1, \|\beta\|) = k_\rho(\|\beta\|)$  y por lo tanto resulta  $k_\rho$  monótona creciente.

El siguiente Teorema se encuentra demostrado en Barrendero y Zamar (2001).

**Teorema 3.3** Sea  $c_\rho = \max \rho$ , y  $\hat{\beta}$  el funcional de estimación de tipo  $M$  correspondiente a la función  $\rho$ . Luego se tiene

$$B(\hat{\beta}, \varepsilon, H_0) = k_\rho^{-1} \left( k_\rho(0) + c_\rho \frac{\varepsilon}{1 - \varepsilon} \right).$$

donde  $c_\rho = \max_u \rho(u)$ .

Sea

$$\Delta_0(u) = -\frac{f_0'(u)}{f_0(u)}, \quad \Delta_1(u, t) = \frac{2M_0(u, 0) - 2M_0(u, t) - \varepsilon/(1 - \varepsilon)}{f_0(u)},$$

donde  $M_0(\cdot, t)$  es la distribución de  $u - t'x_1$ . Consideremos la familia de funciones

$$q_{\delta b}(u) = \Delta_0(u) + \delta \Delta_1(u, b), \quad (3.8)$$

con  $\delta > 0$  y definamos

$$\psi_{\delta b}(u) = q_{\delta b}^+(u) \text{sign}(u), \quad (3.9)$$

donde  $q^+$  es la parte positiva de  $q$ .

Para el próximo lema necesitamos la siguiente hipótesis adicional.

**B3**  $f_0(y)$  es simétrica y unimodal.

**Lema 3.4** Sea  $\hat{\beta}$  el estimador de tipo  $M$  basado en una función  $\psi = \rho'$ , definamos

$$C_\psi(b) = \int_0^\infty \psi(u) \Delta_1(u, b) f_0(u) du,$$

y supongamos que  $B(\hat{\beta}, \varepsilon, H_0) = b$ . Entonces,



$$(i) \quad C_\psi(b) = 0. \quad (3.10)$$

$$(ii) \quad \int_{-\infty}^{\infty} \psi'(u) f_0(u) du = 2 \int_0^{\infty} \psi(u) \Delta_0(u) f_0(u) du.$$

(iii) Si además se cumple **B3**, entonces  $C_\psi$  es no decreciente.

**Demostración.** Sea  $m_0(u, t) = (\partial/\partial u) M(u, t)$  y

$$D_\psi(t) = \int_{-\infty}^{\infty} \rho(u) m_0(u, t) du - \int_{-\infty}^{\infty} \rho(u) m_0(u, 0) du - \frac{\varepsilon}{1-\varepsilon} \int_0^{\infty} \psi(u) du.$$

Como  $B_{\hat{\beta}}(\varepsilon, H_0) = b$ , por el Teorema 3.3 tenemos que

$$D_\psi(b) = 0. \quad (3.11)$$

Usando que  $\rho(u)$  y que  $m_0(u, t)$  son funciones pares, tenemos que

$$D_\psi(t) = 2 \int_0^{\infty} \rho(u) m_0(u, t) du - 2 \int_0^{\infty} \rho(u) m_0(u, 0) du - \frac{\varepsilon}{1-\varepsilon} \int_0^{\infty} \psi(u) du,$$

y luego integrando por partes resulta

$$\begin{aligned} D_\psi(t) &= -2 \int_0^{\infty} \psi(u) M_0(u, t) du + 2 \int_0^{\infty} \psi(u) M_0(u, 0) du - \\ &\quad - \frac{\varepsilon}{1-\varepsilon} \int_0^{\infty} \psi(u) du \\ &= \int_0^{\infty} \psi(u) \left[ \frac{2M_0(u, 0) - 2M_0(u, t) - \varepsilon/(1-\varepsilon)}{f_0(u)} \right] f_0(u) du \\ &= \int_0^{\infty} \psi(u) \Delta_1(u, t) f_0(u) du = C_\psi(t). \end{aligned} \quad (3.12)$$

Luego la parte (i) sale de (3.11). La parte (ii) sale de aplicando integración por partes. Teniendo en cuenta que  $C_\psi(t) = D_\psi(t)$ , para probar la parte (iii) basta ver que

$$\int_{-\infty}^{\infty} \rho(u) m_0(u, t) du = E_{H_0} \rho(u - tx_1)$$

es no decreciente en  $t$ . Esto está probado en el Lemma 3.1 de Martin et al. (1989). ■

En el siguiente teorema mostramos que la familia de funciones de scores (3.9) son dominantes con respecto a la varianza entre todas las funciones de scores con el mismo sesgo.

**Teorema 3.5** *Supongamos que se cumple **B3**. Sean  $\delta > 0$  y  $b > 0$  tales que*

(i) *Existen constantes  $0 < c_1(\delta, b) < c_2(\delta, b) < \infty$  tales que*

$$q_{\delta b}(u) \begin{cases} < 0 & \text{si } 0 < u < c_1(\delta, b) \\ \geq 0 & \text{si } c_1(\delta, b) < u < c_2(\delta, b) \\ < 0 & \text{si } c_2(\delta, b) < u. \end{cases} .$$

(ii) *Si  $\widehat{\beta}_{\delta b}$  es el funcional de estimación de tipo M de la pendiente basado en  $\psi_{\delta b}$ , entonces se tiene  $B(\widehat{\beta}_{\delta b}, \varepsilon, H_0) = b$ .*

*Luego, para cualquier estimador funcional de tipo M de la pendiente  $\widehat{\beta}$ , que satisfaga  $B(\widehat{\beta}, \varepsilon, H_0) \leq b$  se tiene que*

$$V(\psi_{\delta b}, F_0) \leq V(\psi, F_0).$$

**Demostración.** Consideramos un estimador  $\widehat{\beta}$  de tipo M basado  $\psi$  en tal que  $B(\widehat{\beta}, \varepsilon, H_0) \leq b$ . Como  $\psi$  y  $\lambda\psi$  definen el mismo estimador para cualquier escalar  $\lambda$ , sin pérdida de generalidad podemos asumir que

$$\int_{-\infty}^{\infty} \psi'(u) f_0(u) du = \int_{-\infty}^{\infty} \psi'_{\delta b}(u) f_0(u) du,$$

y por lo tanto por el Lema 3.4 (ii) resulta

$$\int_0^{\infty} \psi(u) \Delta_0(u) f_0(u) du = \int_0^{\infty} \psi_{\delta b}(u) \Delta_0(u) f_0(u) du. \quad (3.13)$$

Luego, para probar el teorema es suficiente ver que

$$\int_0^{\infty} \psi^2(u) f_0(u) du \geq \int_0^{\infty} \psi_{\delta b}^2(u) f_0(u) du. \quad (3.14)$$

Por el Lema 3.4 (i) y (iii) y el hecho de que  $B(\widehat{\beta}, \varepsilon, H_0) \leq b$  se obtiene

$$C_{\psi}(b) = \int_0^{\infty} \psi(u) \Delta_1(u, b) f_0(u) du \leq 0. \quad (3.15)$$

Además se tiene que

$$\begin{aligned} \int_0^{\infty} \psi^2(u) f_0(u) du &= \int_0^{\infty} (\psi(u) - q_{\delta b}(u))^2 f_0(u) du \\ &+ \int_0^{\infty} q_{\delta b}^2(u) f_0(u) du - 2 \int_0^{\infty} \psi(u) q_{\delta b}(u) f_0(u) du. \end{aligned} \quad (3.16)$$

Luego de (3.8) se deriva que

$$\begin{aligned} \int_0^{\infty} \psi(u) q_{\delta b}(u) f_0(u) du &= \int_0^{\infty} \psi(u) \Delta_0(u) f_0(u) du \\ &+ \delta \int_0^{\infty} \psi(u) \Delta_1(u, b) f_0(u) du, \end{aligned} \quad (3.17)$$

por el Lema 3.4 y (3.15) resulta que

$$\int_0^{\infty} \psi(u) \Delta_1(u, b) f_0(u) du \leq 0 = \int_0^{\infty} \psi_{\delta b}(u) \Delta_1(u, b) f_0(u) du. \quad (3.18)$$

Por lo tanto, de (3.13), (3.15), (3.16); (3.17) y (3.18) se deduce (3.14). ■

**Observación 3.6** Si  $H_0$  es normal multivariada entonces  $M_0(u, t) = \Phi(u/\sqrt{1+t})$ .

### 3.5 Algunos resultados numéricos

En esta sección compararemos las tres familias de estimadores de tipo M basados en máximo sesgo

(a) la función bicuadrada de Tukey

$$\psi_c^B(u) = u \left(1 - \frac{u^2}{c^2}\right)^2 I_{[0,c]}(u),$$

que es la derivada de (2.15).

(b) la familia de las aproximaciones a la función óptima, estudiadas por Yohai y Zamar (1997).

$$\psi_c^{AO}(u) = \text{sign}(u) r_c^+(|u|)$$

donde

$$r_c(u) = -\frac{(f_0'(u) + c)}{f_0(u)}.$$

(c) la familia de funciones óptimas presentadas en este trabajo (3.9).

En el último caso a diferencia de los primeros la función óptima para cada eficiencia fija depende del porcentaje de contaminación. La función de pérdida óptima  $\psi_{\delta b}$  depende de  $\delta$  y de  $b$ , que están relacionados por medio de la ecuación,

$$C_{\psi_{\delta b}}(b) = 0.$$

Las funciones  $\psi_{\delta b}$  estrictamente positivas en el intervalo  $(c_0(\delta, b), c_1(\delta, b))$ , por lo tanto, la función  $\rho_{\delta b}$  en ese intervalo es estrictamente positiva.

Para cada una de estas familias consideramos estimadores de eficiencia asintótica (medida por  $AV(\psi, H_0)^{-1}$ ) 0.80; 0.90 y 0.95. Consideramos  $G_0$  normal multivariada y sin pérdida de generalidad su media es  $\mathbf{0}$  y su matriz de covarianza la identidad. Por otro lado tomamos  $F_0$  normal con media 0 y varianza 1.

En las Tablas 3.1, 3.2 Y 3.3 se puede ver que la eficiencia para la función aproximada y para la función óptima son muy parecidas y que son mejor que para la función de Tukey.

Tabla 3.1. Sesgo máximo para estimadores con eficiencia 0.80

función $\psi$	$\varepsilon$					
	<b>0.05</b>	<b>0.10</b>	<b>0.15</b>	<b>0.20</b>	<b>0.25</b>	<b>0.30</b>
<b>Tukey</b> ( $c = 3.14$ )	0.59	0.91	1.24	1.63	2.16	2.99
<b>Apr. Optima</b> ( $c = 0.06$ )	0.57	0.87	1.18	1.54	2.02	2.76
<b>Optima</b>	0.57	0.87	1.17	1.52	1.97	2.62

Tabla 3.2. Sesgo máximo para estimadores con eficiencia 0.90

función $\psi$	$\varepsilon$					
	<b>0.05</b>	<b>0.10</b>	<b>0.15</b>	<b>0.20</b>	<b>0.25</b>	<b>0.30</b>
<b>Tukey</b> ( $c = 3.89$ )	0.66	1.00	1.35	1.76	2.28	3.06
<b>Apr. Optima</b> ( $c = 0.028$ )	0.61	0.93	1.25	1.62	2.10	2.82
<b>Optima</b>	0.61	0.93	1.25	1.61	2.08	2.75

Tabla 3.3. Sesgo máximo para estimadores con eficiencia 0.95

función $\psi$	$\varepsilon$					
	<b>0.05</b>	<b>0.10</b>	<b>0.15</b>	<b>0.20</b>	<b>0.25</b>	<b>0.30</b>
<b>Tukey</b> ( $c = 4.685$ )	0.74	1.12	1.54	1.94	2.49	3.28
<b>Apr. Optima</b> ( $c = 0.0127$ )	0.66	1.00	1.34	1.72	2.21	2.92
<b>Optima</b>	0.66	1.00	1.33	1.72	2.19	2.89

En la Tabla 3.4 calculamos los valores de la constante  $\delta$  para diferentes porcentajes de contaminación y eficiencia. Estos parámetros son importantes ya que a partir de ellos y de la ecuación (3.10) se calculan  $c_i(\delta, b)$   $i = 1, 2$ .

Tabla 3.4. Constante  $\delta$  para la función  $\psi$  óptima

Eficiencia	$\varepsilon$					
	<b>0.05</b>	<b>0.10</b>	<b>0.15</b>	<b>0.20</b>	<b>0.25</b>	<b>0.30</b>
<b>0.80</b>	2.0675	1.1458	0.9390	0.8367	0.8293	0.8401
<b>0.90</b>	0.7656	0.4312	0.3298	0.3066	0.3085	0.3377
<b>0.95</b>	0.3083	0.2045	0.1437	0.1224	0.1307	0.14

En el la Figura 1 se encuentran las funciones  $\psi$  para estimadores de tipo M de regresión con una eficiencia del 95%. En el caso de la función óptima se realizó el gráfico para una contaminación del 10%. Se puede ver la aproximación de la función óptima es muy próxima a la misma, es decir, que es una muy buena aproximación. Por otro lado, si bien las tres funciones alcanzan sus extremos aproximadamente en el mismo lugar la óptima y su aproximación

luego del extremo descienden a cero más abruptamente que la bicuadrada.

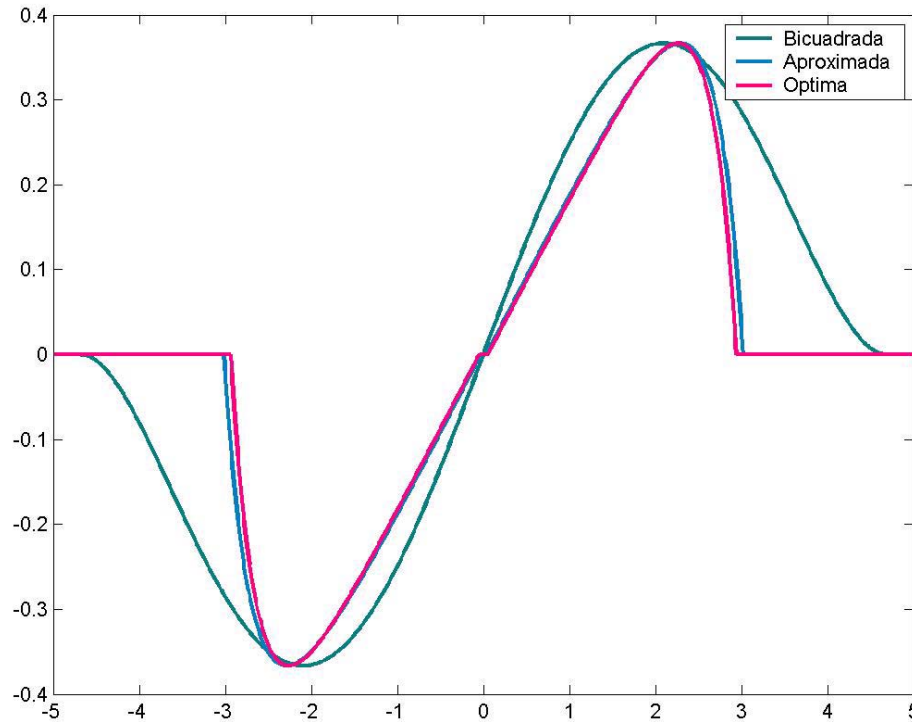


Figura 1: Funciones de escores para M-estimadores con 95% de eficiencia.

Por otro lado graficamos las funciones de escores óptimas para contaminaciones entre el 10% y el 30%, cuando las variables regresoras tienen colas más pesadas que la normal: Cauchy y  $t$  con 3 grados de libertad; y cuando tienen colas menos pesadas que la normal: uniforme en el intervalo  $[0, 1]$  (ver Figuras 2, 3 y 4). En estos gráficos se puede ver que la función  $\psi$  óptima varía muy poco cuando varía la distribución de las variables regresoras y que por lo tanto, la función aproximada sigue siendo una buena aproximación. Sin embargo, a medida que el porcentaje de contaminación aumenta las funciones óptimas para diferentes distribuciones en las variables regresoras son más fáciles de distinguir.

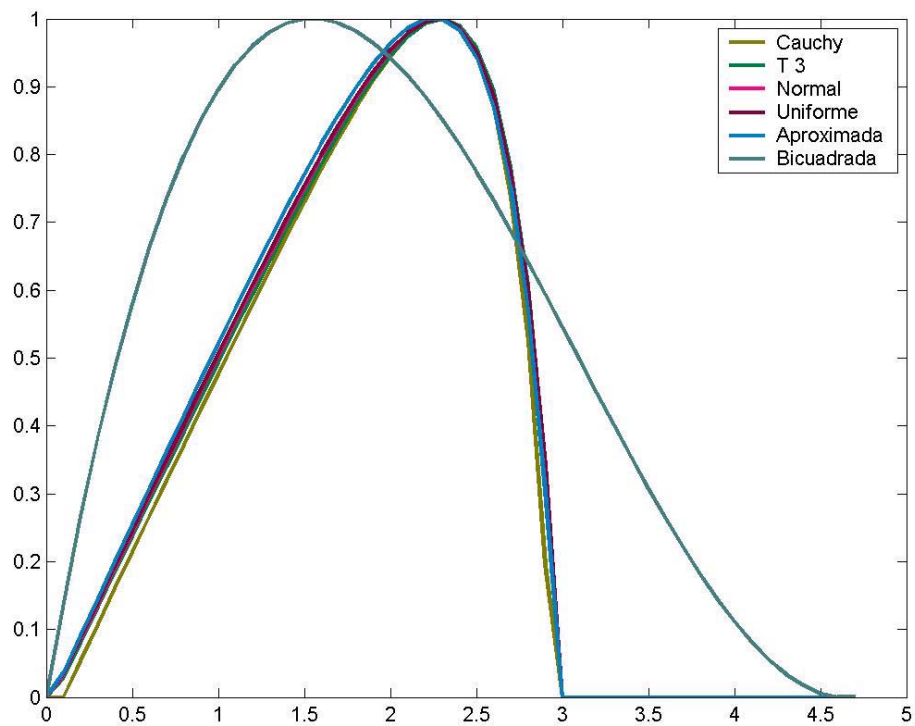


Figura 2: Funciones de escores para M-estimadores considerando diferentes distribuciones en las variables regresoras para una contaminación del 10% y una eficiencia del 95%.

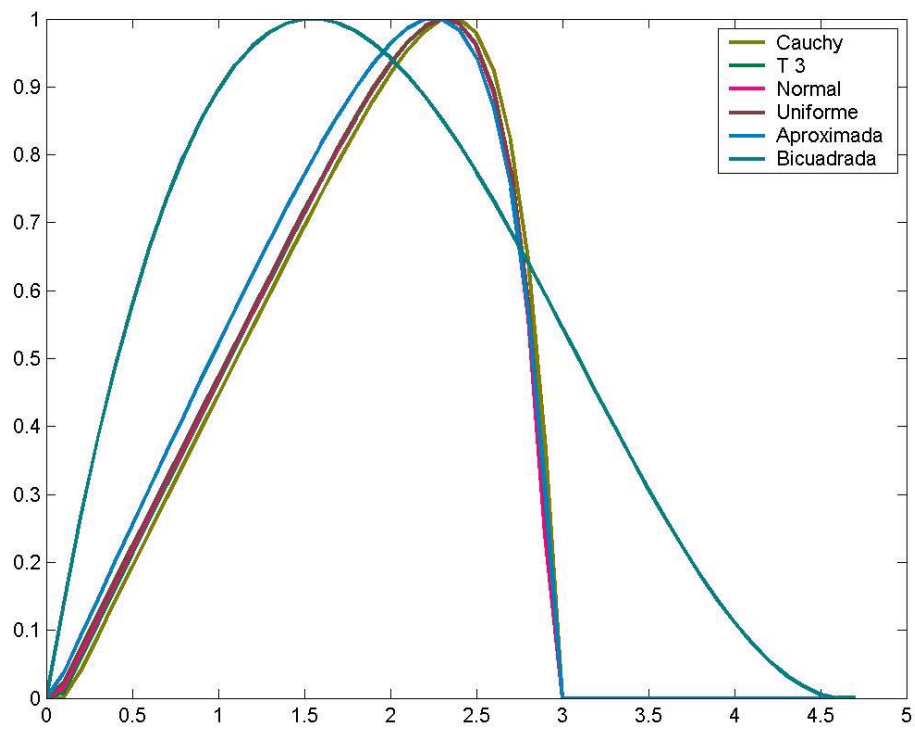


Figure 1: Figura 3: Funciones de escores para M-estimadores considerando diferentes distribuciones en las variables regresoras para una contaminación del 20% y una eficiencia del 95%.



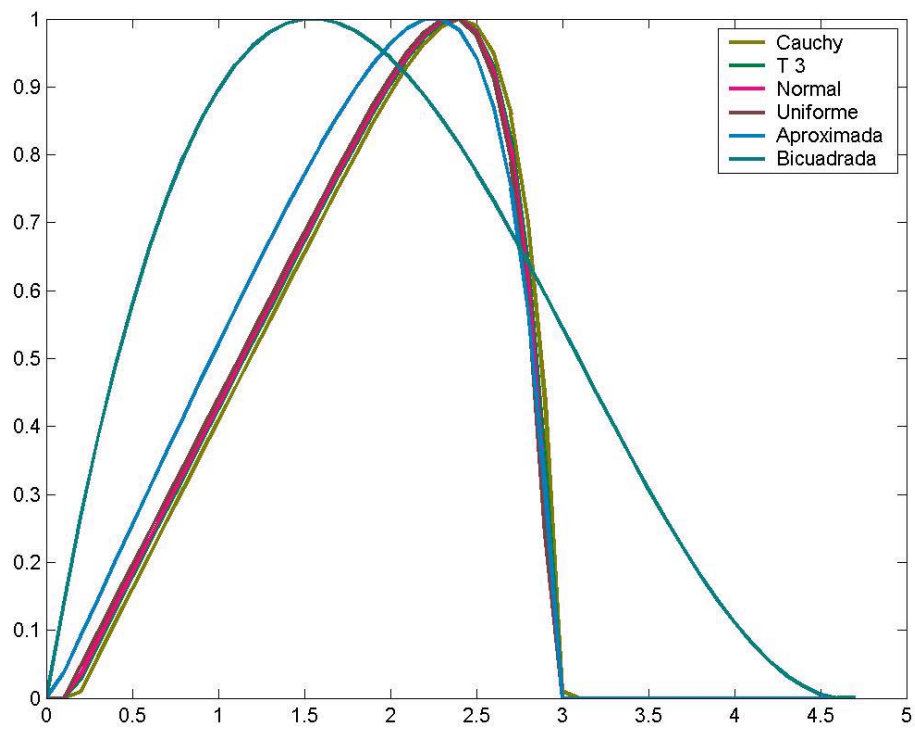


Figure 2: Figura 4: Funciones de escores para M-estimadores considerando diferentes distribuciones en las variables regresoras para una contaminación del 30% y una eficiencia del 95%.

## 4 Capítulo 3

### 4.1 Elección del estimador inicial para un estimador de tipo M de regresión

El objetivo de este capítulo es obtener un estimador de regresión que sea altamente robusto desde el punto de vista del sesgo, y que sea eficiente cuando los errores tienen distribución normal.

Las primeras propuestas de estimadores robustos no alcanzaban estas propiedades de manera simultánea. Esto pasa, por ejemplo, en el caso de los estimadores de tipo S que no pueden tener simultáneamente alto punto de ruptura y alta eficiencia bajo errores normales. Recién a finales de la década del 80 se hicieron dos propuestas de estimadores de regresión que tienen simultáneamente con alto punto de ruptura y alta eficiencia. Estos son los estimadores de tipo  $\tau$  definidos por Yohai y Zamar (1988) y los estimadores MM propuestos por Yohai (1987) que describimos en la Sección 2.4.5.

Como veremos el grado de robustez de los estimadores MM depende fuertemente del estimador inicial utilizado. En la literatura se recomienda considerar como estimador inicial un estimador de tipo S (ver 2.4.3) con punto de ruptura 0.5. Sin embargo, en este trabajo mostraremos que si elegimos como estimador inicial los estimadores de proyección introducidos por Maronna y Yohai (1993) (ver 2.4.6), los estimadores MM resultantes son más robustos. En este capítulo estudiaremos y comparemos la performance de los estimadores MM usando ambos estimadores iniciales: los estimadores de proyección y los estimadores de tipo S.

Con el objetivo de reducir el sesgo del intercept en el estimador de proyección haremos una pequeña modificación en su definición. La intercept será siempre estimada como la mediana de los residuos. Luego si  $\alpha(\boldsymbol{\beta}) = \text{mediana}(y_i - \mathbf{x}'_i \boldsymbol{\beta})$ , los estimadores de proyección se definirán por

$$\widehat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \max_{\|(\lambda_0, \boldsymbol{\lambda}')\|=1} \left| \text{mediana} \left( \frac{y_i - \boldsymbol{\beta}' \mathbf{x}_i - \alpha(\boldsymbol{\beta})}{\lambda_0 + \boldsymbol{\lambda}' \mathbf{x}_i} \right) \right| \text{mad}(\boldsymbol{\lambda}' \mathbf{x}_i), \quad (4.1)$$

y

$$\widehat{\alpha} = \text{mediana} \left( y_i - \widehat{\boldsymbol{\beta}}' \mathbf{x}_i \right).$$

Mediante un estudio de simulación compararemos el error cuadrático medio (ECM) de un estimador MM (2.32) considerando como estimadores iniciales un estimador de tipo S (2.25) y un estimador de tipo P, estimando el intercept de acuerdo con (4.1).

## 4.2 Comparación por Monte Carlo de los estimadores MM de regresión con diferentes estimadores iniciales

En este Capítulo comparamos la performance de los estimadores MM utilizando como estimadores iniciales los estimadores de tipo S y los estimadores de tipo P. Las comparaciones se harán en ambas situaciones, cuando la muestra no contiene puntos atípicos y cuando los contiene.

Tanto los estimadores de tipo S como los estimadores MM están basados en una función  $\rho$  de la familia de funciones de scores bicuadrada definida en (2.15). Para obtener un estimador de tipo S con punto de ruptura 0.5, las constantes que utilizamos fueron  $c = 1.56$  y  $b = 0.5$  que se obtienen a partir de

$$\frac{E_{\Phi}(\rho)}{\rho(c)} = 0.5,$$

(ver Rousseeuw y Leroy, 1987) donde  $\Phi$  es la distribución normal estandar. Consideramos estimadores MM con eficiencia asintótica del 85%. Para esto se requiere utilizar como constante de la familia bicuadrada  $c = 3.44$ . Este valor se obtiene utilizando (2.29).

Para calcular los estimadores iniciales de tipo S y P tomamos  $N = 500$  submuestras y aplicamos los algoritmos descritos en las Secciones (2.4.3) y (2.4.6) respectivamente.

Consideramos el modelo

$$y_i = \alpha + \beta' \mathbf{x}_i + u_i \quad i = 1, \dots, 100;$$

como valores de  $p$  se tomaron 3, 5 y 10. En las muestras donde la proporción  $(1 - \varepsilon)$  de las observaciones  $(y, \mathbf{x})$  son obtenidas de una distribución normal multivariada, que sin pérdida de generalidad debido a la equivarianza se eligió con media cero y matriz de covarianza identidad. Por lo tanto, el valor de los parámetros es  $(\alpha_0, \beta'_0) = \mathbf{0}'$ . Las restantes observaciones son outliers iguales a un valor  $(y_0, \mathbf{x}_0)$ . Nuevamente por equivarianza sin pérdida de generalidad se eligió  $(\mathbf{x}_0) = (x_0, 0, \dots, 0)$  e  $y_0 = mx_0$ .

Consideramos los casos donde  $\varepsilon = 0, 0.10$  y  $0.20$ . En el caso de  $\varepsilon = 0.10$  el valor de  $m$  varía entre 0.18 y 1.28 ya que 1.24 es el máximo sesgo posible para un estimador MM (Martin, Yohai y Zamar, 2006). De manera análoga, para  $\varepsilon = 0.20$  el valor de  $m$  varía entre 0.18 y 2.48. La variable  $x_0$  toma valores naturales hasta 10 para el primer caso y hasta 20 para el segundo caso. Llevamos a cabo un estudio de simulación tipo Monte Carlo con 100 repeticiones para cada una de las contaminaciones.

Para medir la bondad del comportamiento de los estimadores utilizamos el ECM.

**Definición 4.1 (Error Cuadrático Medio)** Sea  $\widehat{\beta}$  en  $\mathbb{R}^p$  el estimador de un parámetro  $\beta$  definimos el error cuadrático medio como

$$ECM(\widehat{\beta}) = E\left(\|\widehat{\beta} - \beta_0\|^2\right). \quad (4.2)$$

Luego dadas  $m$  replicasiones  $\widehat{\beta}_1, \dots, \widehat{\beta}_m$  de un estimador  $\widehat{\beta}$  de  $\beta_0$  estimamos su ECM por

$$\widehat{ECM}(\widehat{\beta}) = \frac{\sum_{i=1}^m \|\widehat{\beta}_i - \beta_0\|^2}{m}. \quad (4.3)$$

Similarmente dadas  $m$  replicasiones  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_m$  de un estimador  $\widehat{\alpha}$  de  $\alpha_0$  estimamos su ECM por

$$\widehat{ECM}(\widehat{\alpha}) = \frac{\sum_{i=1}^m (\widehat{\alpha}_i - \alpha_0)^2}{m}.$$

Dado un estimador  $(\widehat{\alpha}, \widehat{\beta}')$  de  $(\alpha_0, \beta'_0)$ , su error cuadrático global se define como

$$ECMG((\widehat{\alpha}, \widehat{\beta}')) = ECM(\widehat{\alpha}) + ECM(\widehat{\beta})$$

y es estimado por

$$\widehat{ECMG}((\widehat{\alpha}, \widehat{\beta}')) = \widehat{ECM}(\widehat{\alpha}) + \widehat{ECM}(\widehat{\beta})$$

Consideramos los siguiente estimadores:

1. Estimador de tipo P (P).
2. Estimador de tipo S (S).
3. Estimador MM usando el estimador de tipo S como inicial (SMM).
4. Estimador MM usando el estimador de tipo P como inicial (PMM).
5. Estimador MM utilizando como valor inicial el verdadero valor de los parámetros (VMM). Este no es un verdadero estimador, pero dado que parte del valor verdadero correspondería al mejor comportamiento que podemos aspirar de un estimador MM calculado por el algoritmo de mínimos cuadrados pesados iterados.

En las tablas que se presentan a continuación se encuentran los máximos ECM que se obtienen variando  $x_0$  y  $m$ .

En la Tabla 4.1 se puede ver que bajo el modelo central y con pocas covariables el máximo ECM global para todos los estimadores MM tiene un comportamiento similar pero que a medida que el número de covariables aumenta se hace notorio que el máximo ECM para el estimador de tipo S es más grande. Por otra parte bajo contaminación, la performance del estimador PMM es muy superior a la del SMM y esa diferencia aumenta a medida que crece el porcentaje de contaminación y el número de covariables. Esto muestra lo importante que es el estimador inicial utilizado en la primera etapa para los estimadores MM.

Tabla 4.1 Máximo ECM global

Estimadores	$\varepsilon = 0$			$\varepsilon = 0.10$			$\varepsilon = 0.20$		
	$p$			$p$			$p$		
	<b>3</b>	<b>5</b>	<b>10</b>	<b>3</b>	<b>5</b>	<b>10</b>	<b>3</b>	<b>5</b>	<b>10</b>
<b>P</b>	0.065	0.101	0.184	0.17	0.24	0.36	1.33	1.53	2.39
<b>S</b>	0.145	0.265	0.543	0.57	0.82	1.41	2.14	3.07	4.71
<b>PMM</b>	0.073	0.099	0.178	0.34	0.41	0.50	1.74	1.94	2.75
<b>SMM</b>	0.079	0.111	0.238	0.48	0.62	1.17	2.25	3.06	10.92
<b>VMM</b>	0.069	0.092	0.161	0.08	0.33	0.43	0.80	0.88	1.22

En las Figuras 5 a 8 se muestran las curvas del ECM en función de la pendiente de contaminación ( $m$ ), para el valor de  $x_0$  para los cuales el ECM del estimador PMM alcanza su máximo.

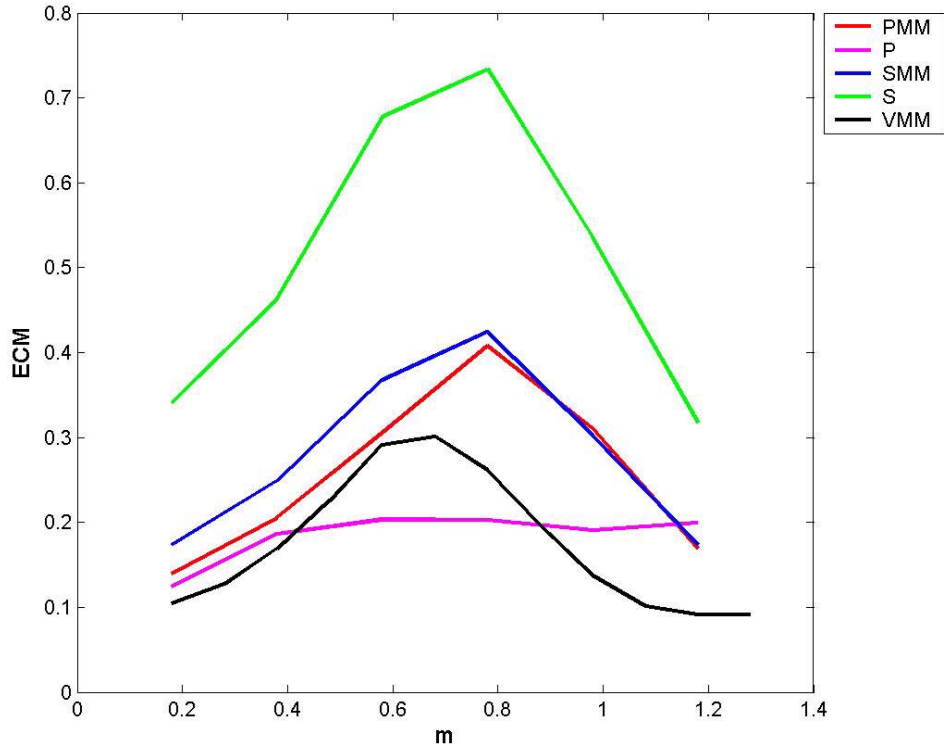


Figura 5: ECMG en el caso de 10% de contaminación y de 5 covariables para una contaminación puntual de  $x_0 = 5$

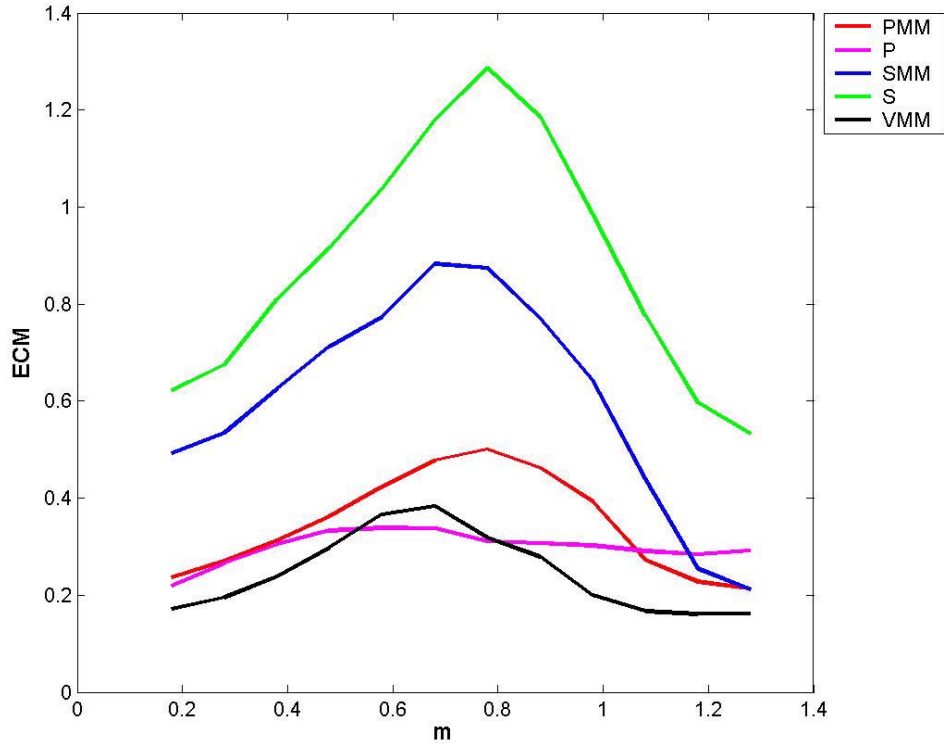


Figura 6: ECMG para el caso de 10% de contaminación y de 10 covariables para una contaminación puntual de  $x_0 = 5$

Las Figuras 5 y 6 muestran las curvas de ECM en el caso de 10% de contaminación puntual. Es claro que aunque se consideren las contaminaciones puntuales menos favorables para el estimador PMM su sesgo es globalmente menor que el del estimador SMM

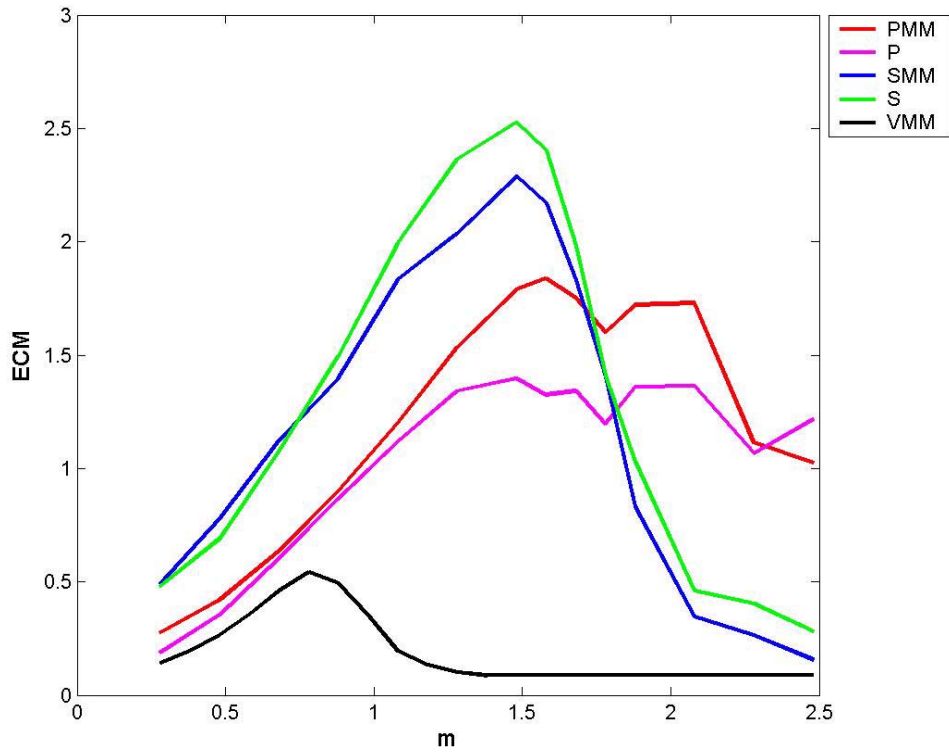


Figura 7: ECMG para el caso de 20% de contaminación y de 5 covariables para una contaminación puntual de  $x_0 = 5$



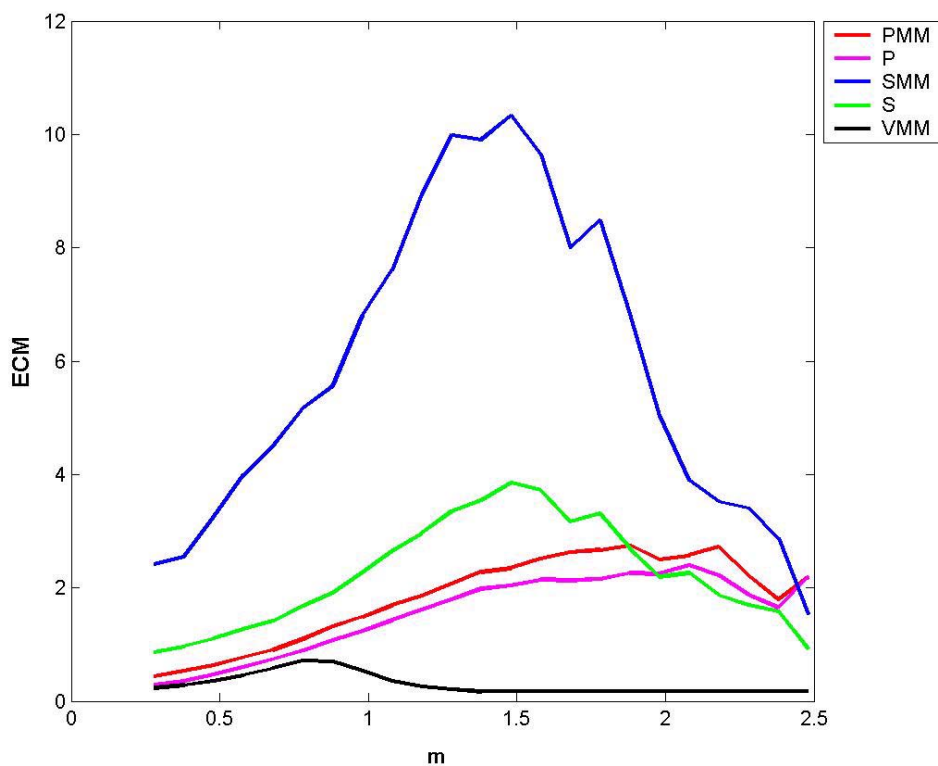


Figura 8: ECMG para el caso de 20% de contaminación y de 10 covariables para una contaminación puntual de  $x_0 = 5$

Las figuras 7 y 8 corresponden a un porcentaje de contaminación en los datos del 20%. Al aumentar el porcentaje de contaminación las diferencias entre los estimadores SMM y el PMM se vuelven más notorias (y favorables para el último).

A continuación buscamos hacer un estudio más detallado del comportamiento de los estimadores analizando en forma separada el ECM de la pendiente  $\beta$  y del intercept  $\alpha$ .

En la tabla 4.2 se encuentra el ECM para el intercept.

Tabla 4.2 Máximo *ECM* de los estimadores de  $\alpha_0$

Estimadores	$\varepsilon = 0$			$\varepsilon = 0.10$			$\varepsilon = 0.20$		
	$p$			$p$			$p$		
	<b>3</b>	<b>5</b>	<b>10</b>	<b>3</b>	<b>5</b>	<b>10</b>	<b>3</b>	<b>5</b>	<b>10</b>
<b>P</b>	0.013	0.019	0.019	0.04	0.04	0.04	0.20	0.24	0.23
<b>S</b>	0.034	0.046	0.056	0.13	0.13	0.14	0.35	0.52	0.49
<b>PMM</b>	0.033	0.036	0.040	0.05	0.05	0.06	0.11	0.15	0.33
<b>SMM</b>	0.038	0.046	0.076	0.07	0.10	0.28	0.41	0.63	4.69
<b>VMM</b>	0.030	0.032	0.031	0.04	0.03	0.03	0.09	0.10	0.12

Las características más salientes de la Tabla 4.3 son que en el caso de estimadores calculados en el modelo sin contaminación los ECM para los estimadores MM son similares independientemente de cual sea el estimador inicial y que siempre son superiores a los de los estimadores de tipo P y menores que los de los estimadores de tipo S . Para el caso del los estimadores considerados bajo el modelo contaminado podemos ver que si bien el ECM de los estimadores MM siempre se encuentra entre el de los de tipo P y el de los de tipo S, las diferencias entre los estimadores SMM y los PMM se vuelve mucho más notoria. Es importante destacar que la performace del estimador PMM es muy superior a la del SMM y que esto se debe principalmente a que el intercept está siendo estimado como mediana de los residuos y que la mediana es el estimador óptimo desde el punto de vista del sesgo para el modelo de posición (2.12).

En las Figuras 9 y 10 se encuentran graficadas las curvas de ECM para el intercept para las contaminaciones menos favorables para los estimadores PMM.

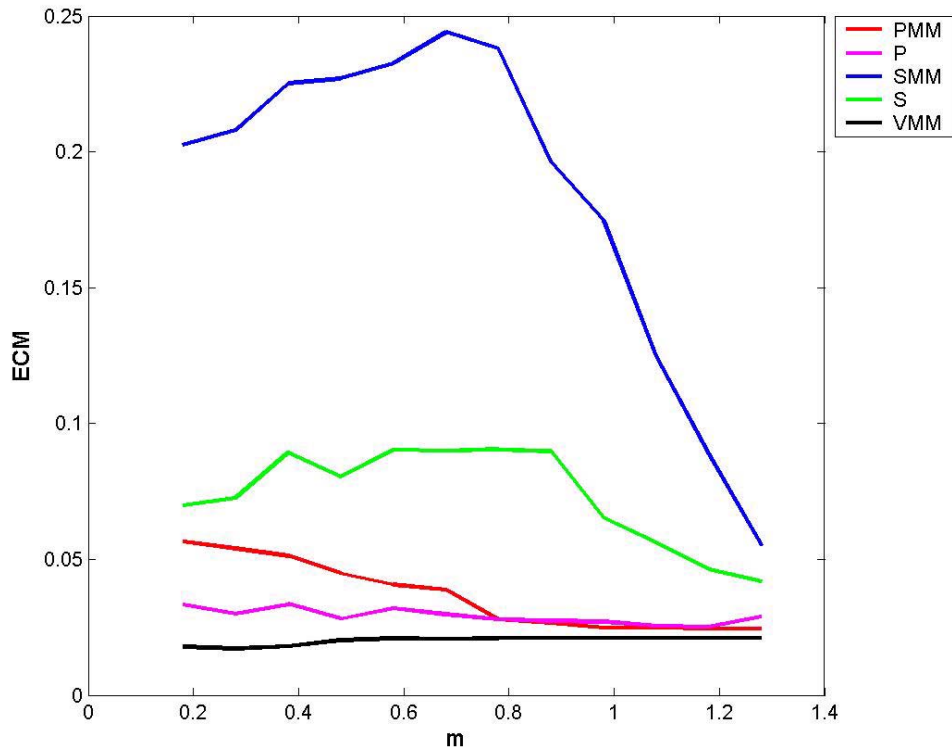


Figura 9: ECM para los estimadores de  $\alpha_0$  en el caso de 10% de contaminación y de 10 covariables para una contaminación puntual de  $x_0 = 9$ .

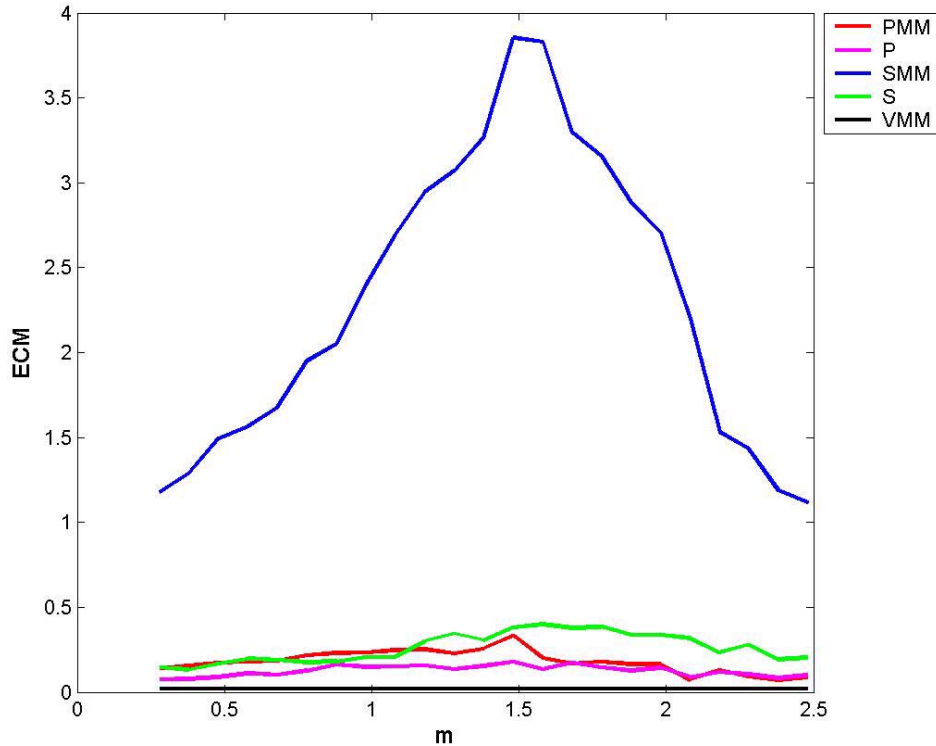


Figura 10: ECM para los estimadores de  $\alpha_0$  en el caso de 20% de contaminación y de 10 covariables para una contaminación puntual de  $x_0 = 11$ .

Una diferencia importante que presentan las Figuras 9 y 10 si se las compara con las gráficas del ECM global es que del estimador SMM tiene un desempeño peor que el del S mientras que en el caso global el ECM del estimador S era superior al del SMM. Al analizar el resto de las curvas se confirma el análisis hecho en base a los máximos ECM para el intercept alcanzado por los estimadores.

Por último, estudiamos el ECM de las estimaciones de la pendiente.

Tabla 4.3 Máximo ECM de los estimadores de  $\beta_0$

Estimadores	$\varepsilon = 0$			$\varepsilon = 0.10$			$\varepsilon = 0.20$		
	$p$			$p$			$p$		
	<b>3</b>	<b>5</b>	<b>10</b>	<b>3</b>	<b>5</b>	<b>10</b>	<b>3</b>	<b>5</b>	<b>10</b>
<b>P</b>	0.05	0.082	0.16	0.14	0.21	0.33	1.13	1.30	2.17
<b>S</b>	0.11	0.212	0.49	0.52	0.73	1.31	1.98	2.77	4.26
<b>PMM</b>	0.039	0.062	0.138	0.32	0.38	0.47	1.70	1.88	2.62
<b>SMM</b>	0.041	0.065	0.162	0.42	0.53	0.90	1.87	2.46	6.34
<b>VMM</b>	0.038	0.060	0.13	0.05	0.32	0.42	0.76	0.84	1.18

Analizando la Tabla 4.3 se puede ver que el ECM en el caso en que las observaciones no están contaminadas es análogo al realizado anteriormente. Para los casos en que estamos en presencia de contaminación se puede ver nuevamente que las propiedades de robustez del estimador PMM son superiores a las del SMM. Además en este caso nuevamente el máximo ECM del estimador de tipo S es mayor que el de el SMM. Todas estas características se acentúan a medida que aumenta el porcentaje de contaminación y el número de covariables.

En las Figuras 11 y 12 graficamos las curvas de ECM para la pendiente de manera análoga a lo hecho anteriormente.

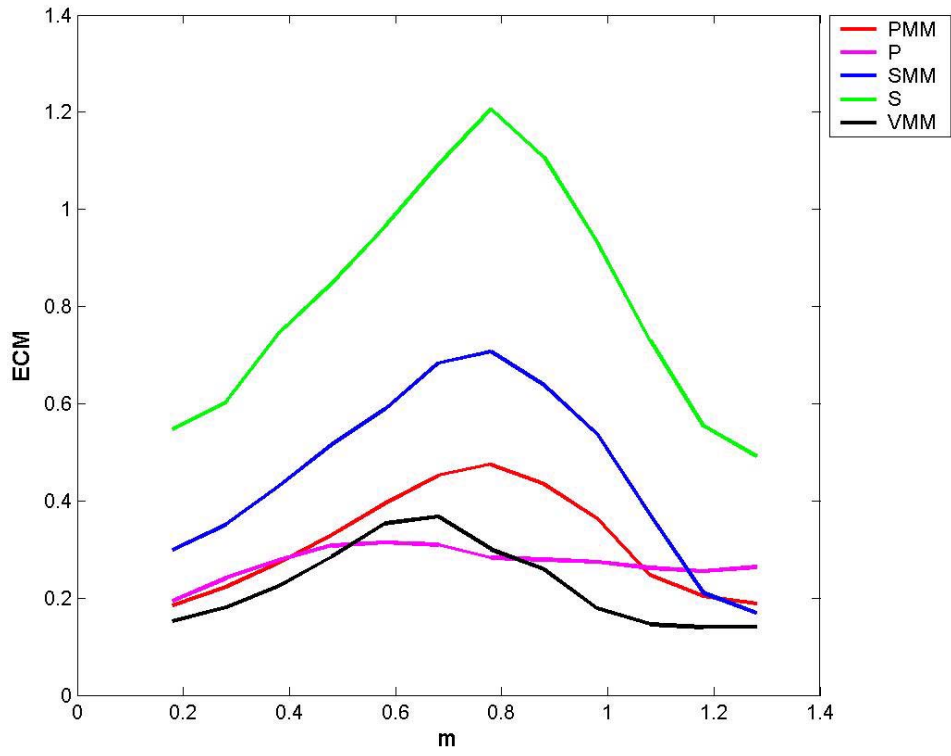


Figura 11: ECM para  $\beta$  en el caso de 10% de contaminación y de 10 covariables para una contaminación puntual de  $x_0 = 5$ .

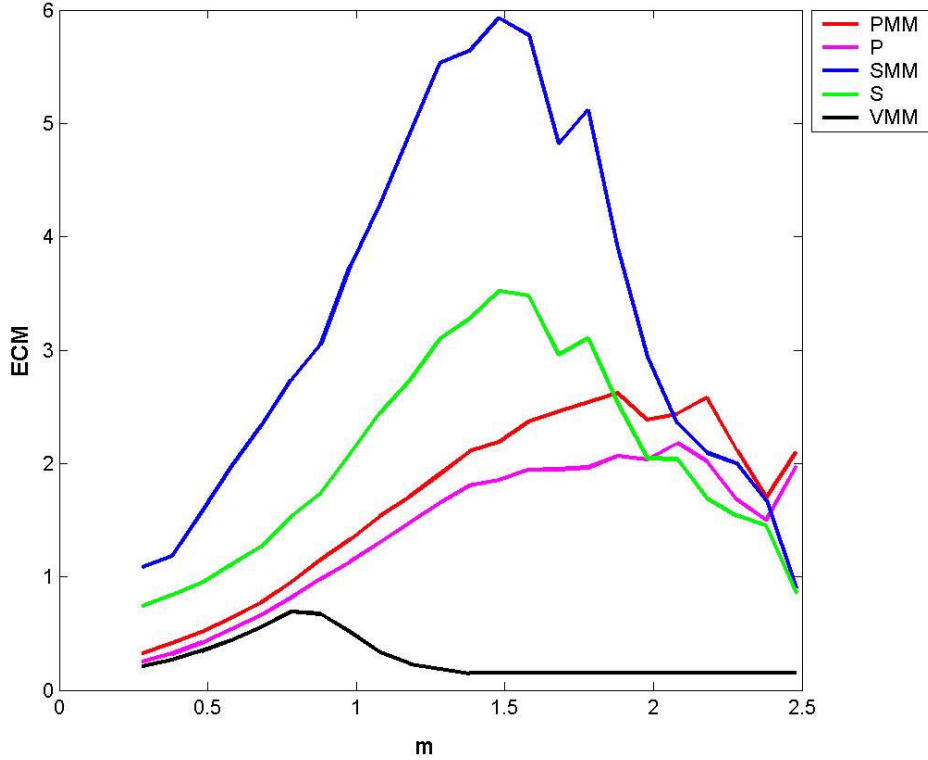


Figura 12: ECM para  $\beta$  en el caso de 20% de contaminación y de 10 covariables para una contaminación puntual de  $x_0 = 2$ .

Con el objetivo de tener un conocimiento más preciso del comportamiento de los estimadores, para cada una de las contaminaciones puntuales consideradas calculamos los percentiles 5, 10,  $\dots$ , 95 de los cuadrados de las normas de los errores. Es decir sean  $(\hat{\alpha}_1, \hat{\beta}_1), \dots, (\hat{\alpha}_m, \hat{\beta}_m)$  las distintas replicas de un estimador para una contaminación dada. Entonces calculamos los errores totales

$$E_i = (\hat{\alpha}_i - \alpha_0)^2 + \|\hat{\beta}_i - \beta_0\|^2,$$

y representamos los percentiles de los  $E_i$ . Luego hallamos el máximo para cada uno de estos percentiles cuando cambiamos la contaminación.

En los Figuras 13 y 14 se encuentran las curvas de máximos percentiles para 10 covariables y contaminaciones de 10% y 20% respectivamente. Es importante destacar que al considerar los diferentes estimadores las contaminaciones que hacen máximos los percentiles no necesariamente son las mismas, por lo tanto en estos gráficos se consideran las contaminaciones más desfavorables para cada uno de ellos.

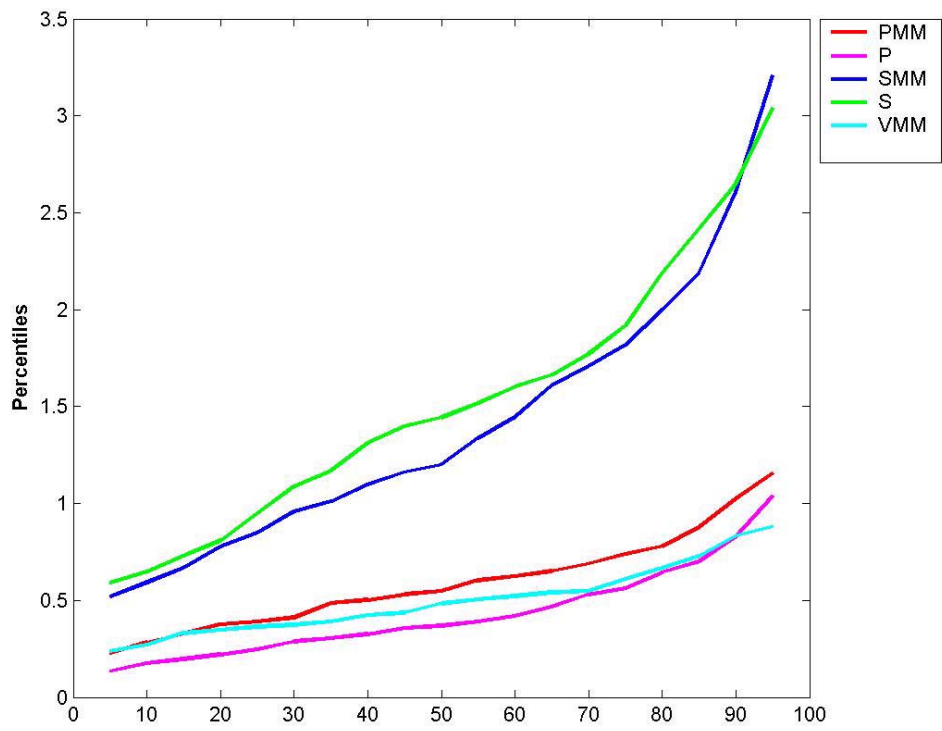


Figura 13: Máximos percentiles para 10% de contaminación y 10 covariables.



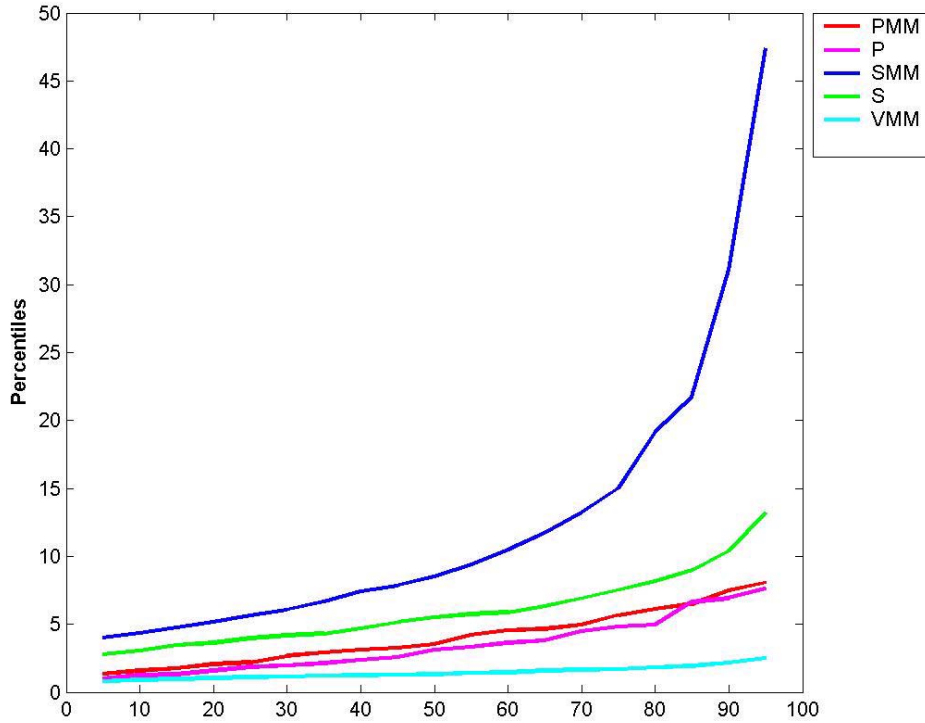


Figura 14: Máximos percentiles para 20% de contaminación y 10 covariables.

En los Figuras 13 y 14 se puede ver que la performance de estimador PMM es superior a la del SMM, y esta superioridad se acentúa para los percentiles más altos. En el caso particular de 20% de contaminación se puede observar que para percentiles superiores al 60% los valores correspondientes al estimador SMM por lo menos duplica al resto de los estimadores.

### 4.3 Cálculo del máximo sesgo asintótico para los estimadores PMM

Se calculó numéricamente el máximo sesgo asintótico para el estimador PMM cuando la fracción de outliers  $\varepsilon$  es igual a 0.10 y 0.20 y cuando la distribución de  $(y, \mathbf{x})$  es normal multivariada.

En primer lugar, mostraremos que si las variables regresoras  $\mathbf{x}$  siguen una distribución normal multivariada el máximo sesgo para la pendiente del estimador definido en (2.35) no depende de la dimensión, si  $p \geq 2$ .

**Teorema 4.2** *Consideramos el modelo lineal de regresión con intercept dado por (2.1), donde la distribución  $G_0$  es normal multivariada. Luego el máximo sesgo para la pendiente y la intercept de los estimadores de tipo P cuando  $H = (1 - \varepsilon)H_0 + \varepsilon\delta_{(y_0, \mathbf{x}_0)}$  donde  $\delta_{(y_0, \mathbf{x}_0)}$  es la masa puntual en el punto  $(y_0, \mathbf{x}_0)$  es independiente de  $p$  si  $p \geq 2$ .*

**Demostración.** Por equivarianza podemos suponer que  $G_0 = N(\mathbf{0}, I)$  y que  $\mathbf{x}_0 = x_0\mathbf{e}_1$ , donde  $\mathbf{e}_1 = (1, 0, \dots, 0)$ .

Luego bastará probar que si  $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$  es el funcional que define el estimador de proyección y  $H_{y_0, x_0} = (1 - \varepsilon)H_0 + \varepsilon\delta_{(y_0, x_0\mathbf{e}_1)}$ , entonces  $\|\hat{\boldsymbol{\beta}}(H_{y_0, x_0})\|$  y  $\hat{\alpha}(H_{y_0, x_0})$  no dependen de  $p$ .

Observemos que la distribución conjunta  $y - \boldsymbol{\beta}'\mathbf{x}$  y de  $\boldsymbol{\lambda}'\mathbf{x}$  es normal bivariada con medias iguales a 0 y con varianzas  $1 + \|\boldsymbol{\beta}\|^2$  y  $\|\boldsymbol{\lambda}\|$  y con covarianza  $\boldsymbol{\beta}'\boldsymbol{\lambda}$ . Luego depende de  $\|\boldsymbol{\beta}\|$ ,  $\|\boldsymbol{\lambda}\|$  y  $\rho = \boldsymbol{\beta}'\boldsymbol{\lambda}/\|\boldsymbol{\beta}\|\|\boldsymbol{\lambda}\|$  y no de la dimensión  $p$ . Obsérvese que también tenemos

$$\frac{y_0 - x_0\boldsymbol{\beta}'\mathbf{e}_1 - \alpha}{\lambda_0 + x_0\boldsymbol{\lambda}'\mathbf{e}} = \frac{y_0 - x_0\beta_1 - \alpha}{\lambda_0 + x_0\lambda_1}$$

Luego para  $x_0$  y  $y_0$  fijos resulta que

$$\left| \text{mediana} \left( (1 - \varepsilon) \frac{y - \boldsymbol{\beta}'\mathbf{x} - \alpha}{\lambda_0 + \boldsymbol{\lambda}'\mathbf{x}} + \varepsilon \frac{y_0 - x_0\boldsymbol{\beta}'\mathbf{e}_1 - \alpha}{\lambda_0 + x_0\boldsymbol{\lambda}'\mathbf{e}} \right) \right|$$

depende solo de  $\|\boldsymbol{\beta}\|$ ,  $\|\boldsymbol{\lambda}\|$ ,  $\rho$ ,  $\beta_1$ ,  $\alpha$ ,  $\lambda_1\lambda_0$ , y no de la dimensión  $p$ . Llamémosle  $D(\|\boldsymbol{\beta}\|, \|\boldsymbol{\lambda}\|, \rho, \beta_1, \alpha, \lambda_1\lambda_0)$ . También tenemos que

$$\text{mad}((1 - \varepsilon)(\lambda_0 + \boldsymbol{\lambda}'\mathbf{x}) + \varepsilon(\lambda_0 + x_0\boldsymbol{\lambda}'\mathbf{e}))$$

depende solo de  $\|\boldsymbol{\lambda}\|$ ,  $\boldsymbol{\lambda}_0$ ,  $\lambda_1$ . Llamémosle  $F(\|\boldsymbol{\lambda}\|, \boldsymbol{\lambda}_0, \lambda_1)$ .

Luego

$$\sup_{\lambda_0^2 + \|\boldsymbol{\lambda}\|^2 = 1, \rho} D(\|\boldsymbol{\beta}\|, \|\boldsymbol{\lambda}\|, \rho, \beta_1, \alpha, \lambda_1\lambda_0) F(\|\boldsymbol{\lambda}\|, \boldsymbol{\lambda}_0, \lambda_1)$$

depende solo de  $\|\beta\|, \beta_1, \alpha$  y no depende de  $p$ . Llamémosle  $J(\|\beta\|, \beta_1, \alpha)$ . Como  $(\hat{\alpha}(H_{y_0, x_0}), \hat{\beta}(H_{y_0, x_0})')$  se obtienen minimizando  $J(\|\beta\|, \beta_1, \alpha)$  obtenemos que  $\|T_\beta(H_{y_0, x_0})\|$  y  $T_\alpha$  no dependen de  $p$ . ■

Como consecuencia del Teorema 4.2. podemos restringir el estudio al modelo

$$y_i = \alpha + x_{1i}\beta_1 + x_{2i}\beta_2 + u_i \quad i = 1, \dots, n;$$

donde  $\alpha, \beta_j \in \mathbb{R}$ ; con  $j = 1, 2$  y  $n = 10000$

Tomamos una muestra de tamaño  $n = 10000$ . Consideramos que una proporción  $(1 - \varepsilon)$  de las observaciones  $(\mathbf{x}, y)$  son generadas bajo el modelo central, que en nuestro caso es  $N_3(\mathbf{0}, I)$ . Para generar una muestra que tenga aproximadamente esta distribución tomamos tres vectores que contienen los cuantiles  $(i + 0.5)/n, 1 \leq i \leq n$  de la distribución normal estándar y los permutamos aleatoriamente. Las primeras  $(1 - \varepsilon) \times n$  observaciones son reemplazadas el vector  $(y_0, x_0, 0) = (mx_0, x_0, 0)$ , ya que por equivarianza podemos asumir que toda la contaminación de las covariables está en la primer coordenada. Los valores de las contaminaciones puntuales que utilizamos son los mismos que en la Sección 4.2 En primer lugar calculamos el estimador de tipo P utilizando la función `fminmax` del programa MATLAB que lleva a cabo el proceso de encontrar un valor minimax. Los inputs de esta función son un valor inicial de  $\beta$  para comenzar a resolver el problema de minimización y un conjunto de valores para  $\lambda$  sobre los cuales se resuelve el problema de maximización. Siguiendo el Lema A.1 probado en (Maronna y Yohai, 1993), el valor inicial para  $\beta$  lo asignamos del siguiente modo

$$(\beta_{10}, \beta_{20}) = \begin{cases} (0, 0) & \text{si } m \leq k_0 \\ (m, 0) & \text{si } k_0 < m \leq 2k_0 \end{cases},$$

donde  $k_0$  es 0.18 (respectivamente 0.41) para una proporción de 0.10 (respectivamente 0.20) datos contaminados. Donde  $k_0 = 0.18$  (respectivamente 0.41) estos valores los obtuvimos en (Maronna y Yohai, 1993).

El subconjunto de vectores unitarios  $(\lambda_0, \lambda_1)$  que utilizamos para resolver el problema de maximización fue el siguiente. La primer coordenada  $\lambda_0$  toma valores entre  $-1$  y  $1$  con una distancia de 0.025 y la segunda coordenada  $\lambda_2$  satisface

$$\lambda_1 = \sqrt{1 - \lambda_0^2}.$$

Una vez obtenido el estimador de tipo P (4.1) lo utilizamos como punto inicial de un estimador MM del mismo modo que hicimos en la Sección 4.2. A partir de este estimador calculamos numéricamente el máximo sesgo asintótico (2.6) para contaminaciones puntuales.

Además, como patrón de referencia calculamos el estimador VMM utilizando como estimador inicial  $(\beta_{10}, \beta_{20}) = (0, 0)$ .

El máximo sesgo asintótico para el estimador SMM se encuentra en Maronna, Martin y Yohai (2006), considerando como estimador inicial un estimador de tipo S con punto de ruptura 0.5 y luego un estimador de tipo M con eficiencia 0.85.

En lo que sigue, analizamos los resultados obtenidos y los contrastamos con los obtenidos en (Maronna, Martin y Yohai, 2006) y en (Maronna y Yohai, 1993).

Tabla 4.4 Máximo sesgo asintótico

Estimadores	$\varepsilon = 0.10$	$\varepsilon = 0.20$
P	0.36	0.82
VMM	0.61	0.67
PMM	0.58	1.16
SMM	0.68	1.65

Como muestra la Tabla 4.4 se produce una importante disminución en máximo sesgo asintótico de los estimadores MM cuando se considera un estimador de tipo P como estimador inicial, la disminución en el sesgo alcanza un 35%. (respectivamente 30%) cuando la proporción de datos contaminados es de 0.10 (respectivamente 0.20).

En las Figuras 15 y 16 presentamos las curvas de máximo sesgo asintótico global para 10% y 20% de contaminación respectivamente. En ellas se compara el sesgo del estimador PMM con el sesgo del VMM estimador y se confirma el análisis realizado en base a los valores máximos. Se puede ver que en el caso de una contaminación del 10% ambas curvas se encuentran muy próximas. Sin embargo, el estimador PMM se ve mucho más afectado al aumentar el porcentaje de contaminación.

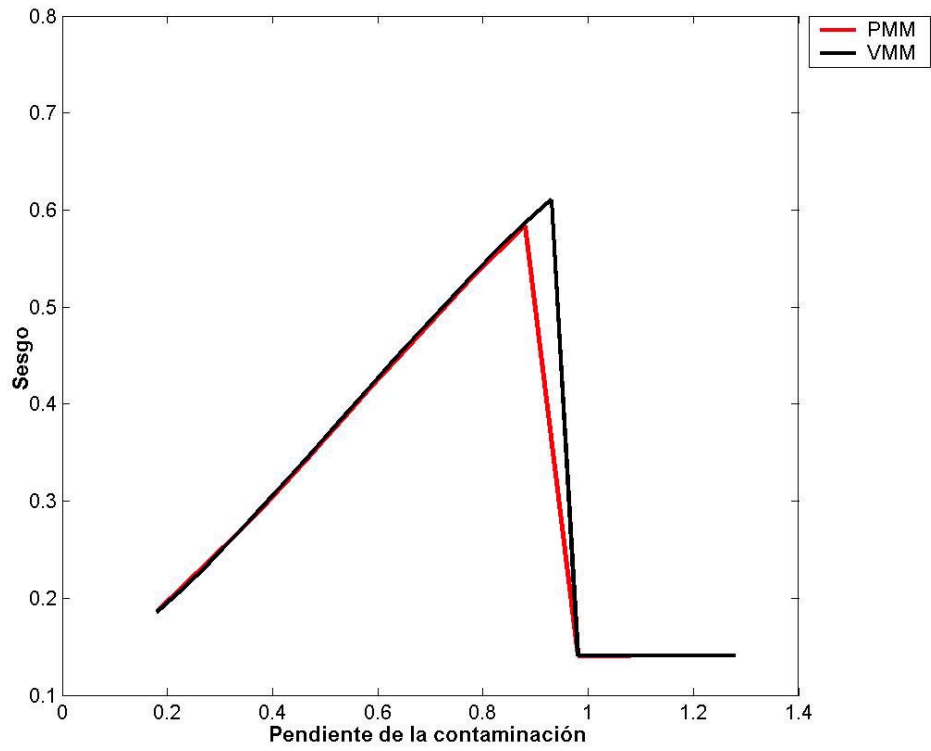


Figura 15: Curvas de máximo sesgo asintótico para una contaminación el 10%

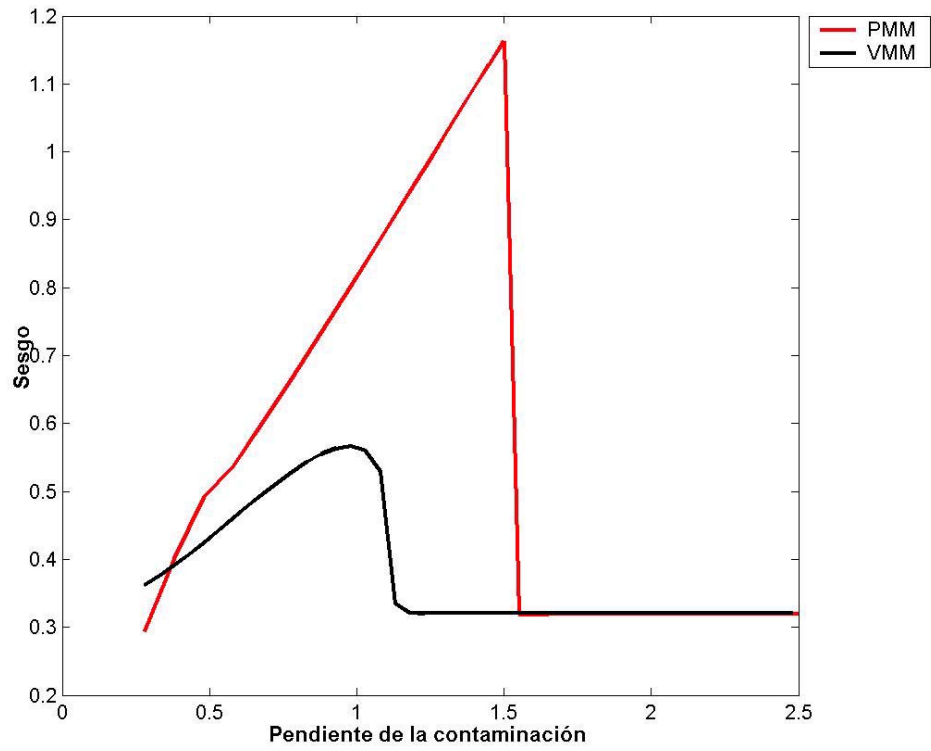


Figura 16: Curvas de máximo sesgo asintótico para una contaminación de 20%

## 5 Bibliografía

1. Barrodale, I. y Roberts, F.D.K. (1973) An improved algorithm for discrete  $L_1$  linear approximation, *SIAM Journal of Numerical Analysis*, Vol. **10**, 839-848.
2. Berrendero, J. R. y Zamar, R. H. (2001) Maximun Bias Curves for Robust Regression with Non-elliptical Regresors, *Ann. Statist.*, Vol **29** No. **1** , 224-251.
3. Donoho, D. L. y Huber, P. J. (1983). The notion of breakdown point. *In a Festschrift for Erich L. Lehmann*, P. J. Bickel, k. A. Donksun, J. L. Jr. Hodges (eds.). Wadsworth, Belmont, California.
4. Edgeworth, F. Y. (1887) On observations relating to several quantities, *Hermathena*, **6**, 279-285.
5. Hampel, F. R. (1971). A general qualitative definition of rubustness, *The Annals of Mathematical Statistics*, Vol. **42**, No. **6**, 1887-1896.
6. Hampel, F. R. (1974). The influence curve and its role in Robust Estimation, *Journal of the American Statistical Association.*, **62**, 1179-1186.
7. Hampel, F. R., Ronchetti, E, M., Rousseeuw, P.J. y Stahel, W.A. (1986). Robust Statistics The Approach Based on Influence Functions. *John Wiley and Sons, Inc.*
8. He, X. (1991) A Local Breakdown Property of Robust Tests in Linear Regression, *J. Multivariate Anal.*, **38**, 294-305.
9. He, X. y Simpson, D. G. (1993) Lower Bounds for Contamination Bias: Globally Minimax Versus Locally Linear Estimation, *Ann. Statist.*, Vol **21** No. **1** , 314-337.
10. Hodges, J. L. Jr. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, University of California Press, Berkeley, California, 163-186.
11. Huber, P. J. (1964) Robust Estimation of a Location Parameter, *Ann. Math. Statist*, **35**, 73-101.
12. Huber, P. J. (1973) Robust Regression: Asymptotics, conjectures and Monte Carlo, *Ann. Statist*, **1**, 799-821.

13. Huber, P. J. (1981) Robust Statistics. *John Wiley and Sons, Inc.*
14. Maronna, R.A, Yohai, V. J. (1993) Bias-Robust Estimates of Regression Based on Projections, *Ann. Statist.*, Vol **21** No. **2** , 965-990.
15. Maronna, R.A, Yohai, V. J (2000) Robust Regression with both Continuous and Categorical Predictors, *Journal of Statistical Planning and Inference*, Vol. **89**, 197-214.
16. Maronna, R.A; Martin, R. D. y Yohai, V. J.(2006) Robust Statistics: Theory and Methods, **John Wiley & Sons Canada, Ltd**, 1 Ed.
17. Martin, R. D; Yohai, V. J. y Zamar, R. H. (1989) Minimax Robust Regression, *Ann. Statist.*, Vol **17** No. **4** , 1608-1630.
18. Martin, R. D. y Zamar, R. H. (1993), Efficiency-constrain Bias-robust Estimation of Location, *Ann. Statist.*, Vol **21** No. **1** , 338-354.
19. Portnoy, S. y Koenker, R. (1997), The Gaussian hare and the Laplacian tortoise: Computability of square error versus absolute-error estimators, *Statistical Science*, Vol. **12**, 279-296.
20. Rousseeuw, P. J (1984) Least median of squares regression, *Journal of the American Statistical Association*. Vol. **79**, No. **388**, 871-880.
21. Rousseeuw, P. J y Leroy, A. M. (1987), Robust Regression and Outlier Detection, *John Wiley and Sons, Inc.*
22. Rousseeuw, P. J y Van Driessen, K. (1999), A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics*, **41**, 212-223
23. Rousseeuw, P. J y Van Driessen, K. (2006), Computing LTS Regression for Large Data Sets, *Data Mining and Knowledge Discovery*, **12**, 29-45.
24. Rousseeuw, P. J. y Yohai, V. J. (1984) Robust Regression by mean of S-estimators, in *Robust and Nonlinear Time Series Analysis*, edited by J. Franke, W. Härdle, and R. D. Martin, Lectures Notes in Statistics, **26**, Springer Verlag, New York , 256-272.
25. Stigler, S. M. (1981). Gauss and the invention of least squares. *Ann. Statist.*, **9**, 465-474.
26. Yohai, V. J. (1987), High Breakdown point and High Efficiency Robust Estimates for Regression, *Ann. Statist.*, Vol **15** No. **2** , 642-656.



27. Yohai, V.J. and Zamar, R.H. (1988), High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal American Statistical Association*, **83**, 406-413.
28. Yohai, V. J. y Zamar, R. H. (1993) A Minimax-Bias Property of the Least  $\alpha$ -Quantile Estimates, *Ann. Statist.*, Vol **21** No. **4** , 1824-1842.
29. Yohai, V. J. y Zamar, R. H. (1997) Optimal Locally Robust M-estimate of Regression, *Journal of Statistical Planning and Inference*, Vol. **57**, 73-92.