



**UNIVERSIDAD DE BUENOS AIRES**  
**Facultad de Ciencias Exactas y Naturales**  
**Departamento de Matemática**

**Tesis de Licenciatura**

**Estimación Robusta de Posición con Datos Faltantes: una Nueva Propuesta**

**María Florencia Statti**

**Director:** Dra. Mariela Sued.

**Co-director:** Dr. Víctor Yohai

Diciembre 2012



# Índice general

<b>Introducción</b>	<b>5</b>
<b>1. Estimadores de Posición y Escala</b>	<b>7</b>
1.1. El Modelo de Posición . . . . .	7
1.1.1. M-estimadores de Posición . . . . .	8
1.2. M-estimadores de Escala . . . . .	10
1.3. M-estimadores de Posición con dispersión desconocida . . . . .	10
1.4. Algunas $\rho$ -funciones importantes . . . . .	12
1.5. Estimadores como funcionales . . . . .	14
1.5.1. Consistencia de Fisher . . . . .	16
1.5.2. Continuidad de los Funcionales de Posición . . . . .	18
1.6. Punto de Ruptura . . . . .	19
1.6.1. Punto de Ruptura Muestral . . . . .	19
<b>2. Regresión Lineal</b>	<b>21</b>
2.1. Variables Explicativas Fijas . . . . .	21
2.2. Variables Explicativas Aleatorias . . . . .	24
2.2.1. Punto de Ruptura de MM-estimadores . . . . .	25
2.3. MM-funcionales . . . . .	26
2.3.1. Consistencia de Fisher y Continuidad del MM-funcional de Regresión . . . . .	27
<b>3. Estimación de Posición con Datos Faltantes</b>	<b>29</b>
3.1. Estimadores de Posición con Datos Faltantes . . . . .	29
3.2. Nueva Propuesta de Estimación . . . . .	30
3.3. Consistencia . . . . .	32
3.4. Estimación de los Parámetros del Modelo de Regresión . . . . .	33
3.5. El funcional de Posición . . . . .	34
3.6. Punto de Ruptura . . . . .	34
3.7. Resultados de un estudio de simulación por el Método de Monte Carlo . . . . .	37
<b>4. Apéndice</b>	<b>43</b>
<b>Bibliografía</b>	<b>49</b>



# Agradecimientos

A mi Mamá, mi Papá, Vane, Marian y Lucho por acompañarme siempre, en particular en esta hermosa carrera que elegí, disfrutando de mis alegrías, tratando de apaciguar mis angustias, alentando mis ilusiones. Vane, además, merece una mención especial por ser mi Gran Amiga y no sólo mi hermana.

A Marcos, mi Amor, por ser incondicional y una persona maravillosa con la que me encanta compartir todo. También por su aporte ingenioso en esta tesis, ayudándome a adaptar los gráficos para poder visualizarlos bien.

A mis amigos de la facu: Paulet, Lu, Ale, Romi, Flor, Ema, Marian, Lucio, Lucas, Carlita, Marie y Pablo. Con gran parte de ellos conformamos un hermoso grupo del que me alegra formar parte.

A mis amigas del cole: Lucci, Jime y Coks.

A la facultad, como institución, porque siempre me sentí muy cómoda y bien tratada.

Al Dr. Víctor Yohai por sus acertadas sugerencias y por el conocimiento brindado.

A Mariela por tener vocación docente y ser excelente en el ejercicio de la docencia, por ser generosa y dispuesta a ayudar - en particular porque mucho antes de proponerle ser mi directora, supo escucharme y aconsejarme en la vida académica - por ser una persona comprometida y porque la confección de esta tesis resultó ser mucho más amena bajo su dirección.



# Introducción

En los últimos años, estimar la media de cierta variable utilizando una muestra con datos faltantes es un problema que ha despertado mucho interés. Pese a que la variable respuesta  $y_i$  no es observada en todos los individuos de la muestra, se dispone de un vector de covariables  $\mathbf{x}_i$  observado para cada  $i$ . Además, para que  $E(y_i)$  pueda ser identificado a partir de la distribución de las variables observadas, se asume que el mecanismo de pérdida de las respuestas es condicionalmente independiente de éstas, dado el vector de covariables (missing at random (MAR)). Mas específicamente, si  $a_i = 1$  cuando  $y_i$  es observada, MAR establece que  $y_i$  es independiente de  $a_i$  dado  $\mathbf{x}_i$ . En este contexto, varias son las propuestas consideradas para estimar  $E(y_i)$ . Robins et. al. (2007) presentan varias de ellas. A los efectos de esta tesis, queremos destacar el siguiente estimador, llamado de regresión. Bajo MAR obtenemos que  $E(y_i) = E(E(y_i|\mathbf{x}_i)) = E(E(y_i|\mathbf{x}_i, a_i = 1))$ . Esto sugiere estimar  $E(y_i)$  basado en un modelo de regresión con  $E(y_i|\mathbf{x}_i) = g(\mathbf{x}_i, \beta_0)$ , ajustado con  $\hat{\beta}$  utilizando sólo los pares  $(\mathbf{x}_i, y_i)$  para los cuales la respuesta  $y_i$  es observada. El estimador de regresión para  $E(y_i)$  bajo el modelo lineal propuesto está dado por  $\sum_{i=1}^n g(\mathbf{x}_i, \hat{\beta})/n$ .

Recientemente Sued y Yohai (2012) consideran una propuesta robusta para tratar este problema. Reformulan el problema de interés estadístico procurando construir estimadores de medidas de posición robustas asociadas a la variable respuesta. A modo de ejemplo, podemos pensar en estimar la mediana de  $y_i$  en lugar de su media. Asumiendo MAR y un modelo de regresión consiguen un estimador de la distribución de la variable respuesta y con él estiman cualquier funcional continuo. Es decir, estiman  $F_y$  con  $\hat{F}_n$  y luego proponen  $T(\hat{F}_n)$  para estimar  $T(F_y)$ . La continuidad del funcional  $T$  en  $F_y$  garantiza la consistencia del procedimiento.

La propuesta hecha considera una muestra ficticia combinando valores predichos con residuos, obtenidos de un ajuste robusto del modelo lineal, utilizando los pares para los cuales la respuesta es observada. Sin embargo, a la hora de construir la muestra ficticia, las respuestas observadas no son utilizadas. Una forma natural de intentar mejorar el punto de ruptura del estimador consiste en utilizar las respuestas observadas y predecir solamente las faltantes. Esa modificación es la que se considera en esta tesis. Demostramos la consistencia de esta nueva propuesta y mostramos que obtenemos una cota para el punto de ruptura que supera a la obtenida por Sued y Yohai.

El aporte original de este trabajo se presenta en el Capítulo 3. Sin embargo hemos considerado conveniente resumir en los Capítulos 1 y 2 algunos conceptos sobre estimadores robustos de posición y regresión lineal.

En el Capítulo 1 presentamos el modelo de posición y escala, y diferentes estimadores robustos para los parámetros de interés. Entre ellos, la mediana, las medias  $\alpha$ -podadas, los S-estimadores y, en particular, los M-estimadores de posición. Presentamos también la versión funcional de los estimadores, estudiando condiciones que garanticen continuidad y Fisher-consistencia. El enfoque funcional será de suma utilidad en el Capítulo 3.

Seguiremos con el modelo de Regresión Lineal en el Capítulo 2, distinguiendo los casos de

covariables determinísticas y aleatorias. Introduciendo estimadores de regresión robustos a partir de la modificación de estimadores clásicos, para luego, al final de este capítulo, hacer hincapié en los MM-estimadores, los cuales jugarán un papel crucial en los estimadores propuestos en esta tesis.

Finalmente, en el Capítulo 3 revisamos las propuestas clásicas y robustas para la estimación de medidas de posición en el contexto de datos faltantes y, valiéndonos de todos los contenidos presentados en los Capítulos 1 y 2, definiremos al nuevo estimador. Los resultados más importantes del trabajo son

- Consistencia de los estimadores propuestos.
- Obtención de una cota inferior para el punto de ruptura de los estimadores propuestos. Se demuestra que esta cota es mayor que la correspondiente a los estimadores de Sued y Yohai (2012) para el mismo problema.
- Finalmente se realiza un estudio de simulación por el método de Monte Carlo. Los resultados de este estudio confirman que los estimadores propuestos tienen un mayor grado de robustez que los propuestos en Sued y Yohai (2012).

# Capítulo 1

## Estimadores de Posición y Escala

En este capítulo nos concentraremos en los modelos de posición y escala. Comenzamos presentando el modelo y diferentes propuestas robustas para estimar cada uno de los parámetros involucrados. Luego, presentamos la versión funcional de los estimadores y estudiamos condiciones que garanticen continuidad y Fisher-consistencia. Estas propiedades permiten deducir la consistencia de los estimadores y resultarán fundamentales para demostrar la consistencia de los estimadores propuestos en el contexto de datos faltantes.

### 1.1. El Modelo de Posición

En muchas situaciones donde se tienen observaciones  $x_i$  de una variable cuantitativa, se asume que la misma se obtiene al observar un verdadero valor  $\mu_0$  desconocido, que es el parámetro que queremos estimar, combinado con un error aleatorio no sistemático. El caso más simple es suponer que el error resulta aditivo, en cuyo caso se plantea el modelo:

$$x_i = \mu_0 + u_i \quad i = 1, \dots, n, \quad (1.1)$$

donde los errores  $u_1, \dots, u_n$  son variables aleatorias. A este modelo se lo denomina el *modelo de posición*.

Si las observaciones son réplicas independientes de un mismo experimento bajo iguales condiciones, se puede asumir que  $u_1, \dots, u_n$  son variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d) con función de distribución  $F_0$ . De esta manera,  $x_1, \dots, x_n$  son v.a.i.i.d con función de distribución

$$F(x) = F_0(x - \mu_0). \quad (1.2)$$

Para modelar la ausencia de errores sistemáticos en las mediciones, asumiremos que las variables  $u_i$  tienen distribución simétrica alrededor del origen:  $u_i \sim -u_i$ .

Recordemos que un estimador  $\hat{\mu}$  es una función de la muestra obtenida:  $\hat{\mu} = \hat{\mu}(x_1, \dots, x_n) = \hat{\mu}(\mathbf{x})$ . En el caso del modelo de posición, se busca que el estimador aproxime con alta probabilidad al verdadero valor  $\mu_0$ . La bondad de la estimación se puede medir con el *error cuadrático medio* (ECM), que se define mediante la expresión:

$$\text{ECM}(\hat{\mu}) = E[(\hat{\mu} - \mu_0)^2] = \text{Var}(\hat{\mu}) + [\text{sesgo}(\hat{\mu})]^2$$

con  $\text{sesgo}(\hat{\mu}) = E(\hat{\mu}) - \mu_0$ , siendo “ $E$ ” la esperanza y “ $\text{Var}$ ” la varianza.

Es importante que un estimador de posición no dependa de la unidad con la que se ha elegido medirlo para que sea apto a una modificación de la misma. Para eso es necesario que cumpla las siguientes dos propiedades:

- equivariancia por posición:  $\hat{\mu}(x_1 + c, \dots, x_n + c) = \hat{\mu}(x_1, \dots, x_n) + c$
- equivariancia por escala:  $\hat{\mu}(cx_1, \dots, cx_n) = c\hat{\mu}(x_1, \dots, x_n)$

Los más conocidos estimadores de posición, la media muestral y la mediana muestral, cumplen efectivamente estas dos propiedades que son muy sencillas de probar. También la media  $\alpha$ -podada, estimador de posición robusto, cumple estas propiedades. Recordemos que la media  $\alpha$ -podada se define de la siguiente manera: dado  $\alpha \in [0, \frac{1}{2})$  y sea  $m = [(n - 1)\alpha]$ , donde  $[\cdot]$  es la parte entera de un número, la media  $\alpha$ -podada es

$$\bar{x}_\alpha = \frac{1}{n - 2m} \sum_{i=m+1}^{n-m} x_{(i)}$$

con  $x_{(i)}$  denotando el  $i$ -ésimo estadístico de orden.

En diferentes aplicaciones, es usual asumir que la distribución  $F_0$  es normal con media 0 y varianza desconocida  $\sigma^2$ , y por consiguiente las observaciones se distribuyen según  $F = \mathcal{N}(\mu_0, \sigma^2)$ . En tal caso, la media muestral es un estimador “óptimo” ya que es el estimador de máxima verosimilitud (EMV) de  $\mu_0$  y minimiza el ECM para estimadores insesgados y, también, como se puede revisar en los trabajos de Lehmann E.L. y Casella G. (1998) y Bickel P.J. y Doksum K.A. (2001), para los equivariantes. Sin embargo, puede ocurrir que una proporción central de la muestra se pueda describir mediante la distribución normal, pero no toda. En tal caso, podemos pensar a la distribución  $F$  como aproximadamente normal, en el siguiente sentido: consideramos que una proporción  $1 - \varepsilon$  de las observaciones es generada por el modelo normal, mientras que la proporción  $\varepsilon$  restante es generada por un mecanismo desconocido. En el caso de los experimentos, este modelo puede estar ilustrando la toma de medidas de una cierta magnitud donde el  $(1 - \varepsilon) \cdot 100\%$  de las veces la medición es correcta y el  $\varepsilon \cdot 100\%$  de las veces el aparato que mide falla o el experimentador transcribe mal la medición. Así podemos escribir a la función de distribución  $F$ , denominada *distribución normal contaminada*, como:

$$F = (1 - \varepsilon)G + \varepsilon H, \tag{1.3}$$

donde  $G = \mathcal{N}(\mu_0, \sigma^2)$  y  $H$  es una distribución cualquiera.

No obstante, existen casos donde este tipo de distribución está representando el comportamiento de una subpoblación dentro de la muestra, que se pensaba como una población única, o un cambio desconocido en las condiciones en que se hace la medición. En estos casos los datos atípicos no se corresponden con errores en la medición si no con un modelo que no resulta el apropiado o problemas en la recolección de los datos.

### 1.1.1. M-estimadores de Posición

Asumiendo que  $F_0$  tiene densidad  $f_0 = F'_0$ , obtenemos la densidad conjunta de las observaciones

$$L(x_1, \dots, x_n; \mu) = \prod_{i=1}^n f_0(x_i - \mu).$$

El estimador de máxima verosimilitud (EMV) de  $\mu_0$  es el valor  $\hat{\mu}$  que maximiza  $L(x_1, \dots, x_n; \mu)$ :

$$\hat{\mu} = \arg. \max_{\mu} L(x_1, \dots, x_n; \mu).$$

Conociendo exactamente a  $F_0$ , el EMV es “óptimo” en el sentido de poseer la varianza asintótica más baja respecto de una clase de estimadores razonables. Sin embargo, al considerar distribuciones contaminadas,  $F_0$  no es del todo conocida. En tal caso, el propósito es encontrar estimadores que resulten:

- (a) “cercanamente óptimos” cuando  $F_0$  es exactamente normal.
- (b) “cercanamente óptimos” cuando  $F_0$  es aproximadamente normal.

Tomando a  $f_0$  en su soporte y sabiendo que la función logaritmo es estrictamente creciente, podemos escribir

$$\hat{\mu} = \arg. \min_{\mu} \sum_{i=1}^n \rho(x_i - \mu) \quad (1.4)$$

donde  $\rho = -\ln(f_0)$ . Si  $\rho$  es diferenciable, diferenciando (1.4) con respecto a  $\mu$  obtenemos que  $\hat{\mu}$  satisface

$$\sum_{i=1}^n \psi(x_i - \hat{\mu}) = 0 \quad (1.5)$$

con  $\psi = \rho'$ . Para  $\psi$  discontinua pueden no existir soluciones de (1.5) y en tal caso, consideramos como solución a cualquier punto donde el lado izquierdo de la ecuación (1.5) cambie de signo.

Soluciones particulares de (1.4) son la media muestral y la mediana, con  $F_0 = \mathcal{N}(0, 1)$  y  $f_0(x) = \frac{1}{2}e^{-|x|}$  respectivamente.

Considerando que los estimadores de máxima verosimilitud resuelven (1.4) o (1.5), utilizaremos este tipo de procedimiento para definir nuevos estimadores, utilizando funciones  $\rho$  que no necesariamente provengan de un modelo. Para ello, consideremos la siguiente definición:

**Definición 1.1.1.** Una  $\rho$ -función denota a una función  $\rho$  cumpliendo las siguientes condiciones:

**r1**  $\rho(x)$  es una función no decreciente de  $|x|$ , es decir,  $\rho$  es una función par.

**r2**  $\rho(0) = 0$ .

**r3**  $\rho(x)$  es creciente para  $x > 0$  tal que  $\rho(x) < \rho(\infty)$ .

**r4** Si  $\rho$  es acotada, supondremos que  $\rho(\infty) = 1$ .

**Definición 1.1.2.** Una  $\psi$ -función denota una función  $\psi$  la cual es la derivada de una  $\rho$ -función.

**Observación 1.1.3.**  $\psi$  es una función impar y  $\psi(x) \geq 0$  para todo  $x \geq 0$ .

Huber (1964) introduce la siguiente definición:

**Definición 1.1.4.** Dada una  $\rho$ -función, un M-estimador de posición es una solución de (1.4).

Así podemos estudiar estimadores que no necesariamente sean EMV de alguna distribución. La función  $\rho$  se elegirá de manera que el estimador se encuadre en los ítems (a) y (b) antes mencionados. Es sencillo de probar que los M-estimadores tal cual se definieron son equivariantes por posición pero no así por escala.

## 1.2. M-estimadores de Escala

Consideremos observaciones  $x_i$  satisfaciendo el *modelo de escala*

$$x_i = \sigma_0 u_i \quad i = 1, \dots, n \quad (1.6)$$

donde las  $u_i$  son v.a.i.i.d con función de densidad  $f_0$  y  $\sigma_0 > 0$  es el parámetro desconocido que queremos estimar.

La distribución de las variables  $x_i$  tiene densidad  $f_0(x/\sigma_0)/\sigma_0$ . Ejemplos de familias con este tipo de distribuciones son la exponencial y la normal con media 0.

El EMV de  $\sigma_0$  para el modelo (1.6) es

$$\hat{\sigma} = \arg \cdot \max_{\sigma > 0} \frac{1}{\sigma^n} \prod_{i=1}^n f_0 \left( \frac{x_i}{\sigma} \right).$$

Tomando ln y derivando respecto de  $\sigma$  queda la expresión

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i}{\hat{\sigma}} \right) = 1$$

siendo  $\rho(t) = -t \frac{f'_0(t)}{f_0(t)}$ .

Inspirados en esta construcción, definimos los M-estimadores de escala.

**Definición 1.2.1.** Sea  $\hat{\sigma}$  un estimador de  $\sigma_0$  bajo el modelo (1.6). Se dice que  $\hat{\sigma}$  es un M-estimador de escala si es solución de la identidad

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{x_i}{\hat{\sigma}} \right) = \delta \quad (1.7)$$

siendo  $\rho$  una  $\rho$ -función y  $\delta > 0$  una constante.

Para que esté bien definido el estimador,  $\delta$  debe cumplir:  $0 < \delta < \rho(\infty)$ . Es más, si  $\rho$  es acotada se puede asumir sin pérdida de generalidad que  $\rho(\infty) = 1$  y entonces  $\delta \in (0, 1)$ .

Se puede verificar que los M-estimadores de escala satisfacen la equivariancia por escala:

$$\hat{\sigma}(cx_1, \dots, cx_n) = c \hat{\sigma}(x_1, \dots, x_n), \quad \text{para todo } c > 0. \quad (1.8)$$

## 1.3. M-estimadores de Posición con dispersión desconocida

Los estimadores dados por (1.4) no son equivariantes por escala, lo que implica, como ya hemos mencionado, que dependen fuertemente de las unidades elegidas para modelar. Por ejemplo, si queremos estimar  $\mu_0$  en el modelo (1.1) donde  $F$  está dada por (1.3) con  $G = \mathcal{N}(\mu_0, \sigma^2)$ , si  $\sigma$  es conocido, resulta intuitivo dividir en (1.1) por  $\sigma$  para reducir el problema al caso  $\sigma = 1$ . En tal caso, el estimador de  $\mu_0$  estaría dado por

$$\hat{\mu} = \arg \cdot \min_{\mu} \sum_{i=1}^n \rho \left( \frac{x_i - \mu}{\sigma} \right). \quad (1.9)$$

Una solución natural para poder obtener un M-estimador de posición que sea equivariante por escala, es estimar a  $\sigma$  previamente por un estimador de dispersión. Este tipo de estimadores cumplen esencialmente con las siguientes dos propiedades:

- invariancia por posición:  $\hat{\sigma}(x_1 + c, \dots, x_n + c) = \hat{\sigma}(x_1, \dots, x_n)$
- equivariancia por escala:  $\hat{\sigma}(cx_1, \dots, cx_n) = |c| \hat{\sigma}(x_1, \dots, x_n)$

Con la modificación introducida, efectivamente la solución  $\hat{\mu}$  de (1.9), con  $\sigma = \hat{\sigma}$  un estimador de dispersión, es equivariante por escala. En adelante, consideraremos M-estimadores  $\hat{\mu}$  equivariantes por posición y escala siendo

$$\hat{\mu} = \arg \cdot \min_{\mu} \sum_{i=1}^n \rho \left( \frac{x_i - \mu}{\hat{\sigma}} \right), \quad (1.10)$$

con  $\hat{\sigma}$  un estimador de dispersión. Cuando  $\psi = \rho'$  es monótona, el estimador es la única solución de

$$\sum_{i=1}^n \psi \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) = 0. \quad (1.11)$$

Un estimador de dispersión no robusto muy utilizado es la desviación estándar muestral:

$$SD(\mathbf{x}) = \left[ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}.$$

Una alternativa robusta dentro de esta clase de estimadores es

$$MAD(\mathbf{x}) = \text{med}_i(|x_i - \text{med}(\mathbf{x})|),$$

la mediana de las desviaciones absolutas (en inglés: median absolute deviations).

Nosotros utilizaremos S-estimadores de dispersión, definidos de la siguiente manera. Dada  $\rho_0$ , una  $\rho$ -función acotada y  $\delta \in (0, 1)$ , para cada  $t \in \mathbb{R}$  consideramos  $\hat{s}(t)$  satisfaciendo

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{x_i - t}{\hat{s}(t)} \right) = \delta. \quad (1.12)$$

Luego, el S-estimador de dispersión está dado por  $\hat{\sigma}$ , siendo

$$\hat{\sigma} = \min_t \hat{s}(t). \quad (1.13)$$

Rousseeuw y Yohai (1984) definen los S-estimadores de posición mediante

$$\hat{\mu}_S = \arg \cdot \min_{t \in \mathbb{R}} \hat{s}(t). \quad (1.14)$$

Los S-estimadores de posición resultan M-estimadores como los definidos mediante la ecuación (1.10), tomando  $\rho = \rho_0$  y poniendo el S-estimador de dispersión definido en (1.13) en lugar de  $\hat{\sigma}$ .

Los M-estimadores, utilizando S-estimadores de dispersión con  $\rho \leq \rho_0$ , fueron propuestos por Yohai (1987) y se dicen MM estimadores.

Otra forma de pensar este problema es considerando un nuevo modelo, el *modelo de posición y dispersión* con dos parámetros desconocidos

$$x_i = \mu_0 + \sigma_0 u_i \quad i = 1, \dots, n \quad (1.15)$$

donde  $u_i$  tiene densidad  $f_0$ . Notemos que el parámetro  $\sigma_0$  es de escala para las  $u_i$  pero de dispersión para las  $x_i$ . Conociendo la densidad  $f_0$ , podemos calcular los estimadores de máxima verosimilitud para los parámetros  $(\mu_0, \sigma_0)$ . Inspirados en el sistema de ecuaciones que estos resuelven, consideramos una propuesta robusta para estimar en forma simultánea a los parámetros, mediante la resolución del sistema

$$\frac{1}{n} \sum_{i=1}^n \psi \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) = 0 \quad (1.16)$$

$$\frac{1}{n} \sum_{i=1}^n \rho_{escala} \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) = \delta \quad (1.17)$$

donde  $\psi$  es una  $\psi$ -función y  $\rho_{escala}$  es una  $\rho$ -función.

#### 1.4. Algunas $\rho$ -funciones importantes

La familia de *funciones de Huber* son un tipo de  $\rho$ - y  $\psi$ -funciones con importantes propiedades. Para cada  $k > 0$ , se define:

$$\rho_k^H(x) = \begin{cases} x^2 & \text{si } |x| \leq k \\ 2k|x| - k^2 & \text{si } |x| > k \end{cases} \quad (1.18)$$

con derivada

$$\psi_k^H(x) = \begin{cases} 2x & \text{si } |x| \leq k \\ 2k \operatorname{sgn}(x) & \text{si } |x| > k \end{cases} \quad (1.19)$$

Los correspondientes M-estimadores para los casos límites  $k \rightarrow +\infty$  y  $k \rightarrow 0$  son la media y la mediana, considerando a  $\psi_0(x) = \operatorname{sgn}(x)$ .

Otra familia de funciones interesantes y populares es la de las *funciones bicuadradas*, introducida por Tukey. Fijado  $k > 0$ ,  $\rho_k^B$  está dada por:

$$\rho_k^B(x) = \begin{cases} 1 - [1 - (\frac{x}{k})^2]^3 & \text{si } |x| \leq k \\ 1 & \text{si } |x| > k \end{cases} \quad (1.20)$$

con derivada

$$\psi_k^B(x) = \frac{6x}{k^2} \left[ 1 - \left( \frac{x}{k} \right)^2 \right]^2 I_{\{|x| \leq k\}} \quad (1.21)$$

siendo  $I_A$  la indicadora del conjunto  $A$ . Los M-estimadores definidos mediante una función  $\psi$  soportada en un intervalo, no son EMV de alguna distribución.

Frecuentemente, suele considerarse el S-estimador de dispersión presentado en (1.13) con  $\rho = \rho_1^B$  y  $\delta = 0,5$ .

Una visión intuitiva de los M-estimadores de posición nos permite pensarlos como una media pesada. En la mayoría de los casos interesantes,  $\psi(0) = 0$  y  $\exists \psi'(0)$ , entonces  $\psi$  es aproximadamente lineal en el origen. Sea

$$W(x) = \begin{cases} \psi(x)/x & \text{si } x \neq 0 \\ \psi'(0) & \text{si } x = 0 \end{cases}$$

Luego podemos reescribir a (1.5) como

$$\sum_{i=1}^n W(x_i - \hat{\mu})(x_i - \hat{\mu}) = 0,$$

o equivalentemente

$$\hat{\mu} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad \text{con } w_i = W(x_i - \hat{\mu}),$$

quedando el estimador expresado como una media ponderada.

Particularmente, se tiene la función de pesos de la familia de *funciones bicuadradas* dada de la siguiente manera:

$$W_k^B(x) = \left[1 - \left(\frac{x}{k}\right)^2\right]^2 I_{\{|x| \leq k\}}. \quad (1.22)$$

De esta manera podemos observar que cuando utilizamos  $\rho$ -funciones en la familia bicuadrada, observaciones muy grandes en valor absoluto reciben peso nulo, en cuyo caso podemos considerar que son ignoradas a la hora de calcular el estimador.

Si  $\rho$  es diferenciable en todo punto y  $\psi$  es monótona, se tiene que las expresiones (1.4) y (1.5) son equivalentes. Para  $\psi$  redescendiente, la función objetivo a minimizar puede tener mínimos locales, dando lugar a soluciones de (1.5), que pueden no corresponder al mínimo absoluto del criterio que define al M-estimador.

Estimadores definidos como soluciones de (1.5) o (1.11) con  $\psi$  monótona se denominarán M-estimadores monótonos, mientras que aquellos definidos por (1.4) o (1.10) cuando  $\psi$  no es monótona se denotarán M-estimadores redescendentes.

Como hemos mencionado, los procedimientos robustos permitirán protegernos ante la presencia de datos atípicos pero, generalmente, esta propiedad repercutirá en la eficiencia del procedimiento. Sabemos que los estimadores de máxima verosimilitud resultan eficientes y entonces para medir la eficiencia asintótica de un nuevo estimador se mide cuán cerca está este del estimador óptimo. Tenemos entonces la siguiente definición:

**Definición 1.4.1.** La eficiencia asintótica de  $\hat{\mu}$  mide cuán cerca está  $\hat{\mu}$  del estimador óptimo y está dada por

$$Ef(\hat{\mu}) = \frac{v_0}{v} \quad (1.23)$$

donde  $v_0$  es la varianza asintótica del EMV y  $v$  es la varianza asintótica de  $\hat{\mu}$ .

Si consideramos el modelo de posición presentado en (1.1) con  $u_i \sim \mathcal{N}(0, 1)$ , tenemos que  $x_i \sim \mathcal{N}(\mu_0, 1)$ . Utilizando M-estimadores definidos mediante  $\rho$ -funciones pertenecientes a la familia bicuadrada, la eficiencia asintótica de estos estimadores varía con el valor de  $k$  como se observa en el siguiente cuadro, extraído de la página 30 del libro de Maronna R.A., Martin R.D. y Yohai V.J. (2006).

Cuadro 1.1: Eficiencia de las funciones bicuadradas bajo el modelo normal

<b>k</b>	3,14	3,44	3,88	4,68
<b>Ef</b>	0,80	0,85	0,90	0,95

Se puede probar que, para un  $n$  grande, la distribución de la media  $\alpha$ -podada bajo el modelo (1.1) es aproximadamente normal, y cuando se tienen  $u_i$  con distribución simétrica,  $\bar{x}_\alpha$  tiene una distribución aproximadamente  $\mathcal{N}(\mu_0, \frac{v}{n})$  donde la varianza asintótica  $v$  es aquella de un M-estimador con función de Huber  $\psi_k^H$  siendo  $k$  el  $(1 - \alpha)$ -cuantil de  $u$ :

$$v = \frac{E[\psi_k^2(u)]}{(1 - 2\alpha)^2}.$$

## 1.5. Estimadores como funcionales

Tras haber presentado diferentes estimadores de posición, en este trabajo restringiremos nuestra atención a (i) la mediana, (ii) las medias  $\alpha$ -podadas y (iii) los M-estimadores definidos mediante (1.10), utilizando un S-estimador de dispersión, definido mediante (1.13). Cada uno de estos estimadores puede ser considerado como cierto funcional  $T$  (definido sobre un espacio de distribuciones) evaluado en  $\widehat{F}_n$ , la distribución empírica asociada a la muestra, definida por

**Definición 1.5.1.** Dada  $y_1, \dots, y_n$  una muestra, se define la función de distribución empírica asociada a la muestra como

$$\widehat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n I_{\{y_i \leq y\}}. \quad (1.24)$$

Notemos que para toda función  $g$  medible se verifica la siguiente igualdad:

$$E_{\widehat{F}_n}(g(y)) = \frac{1}{n} \sum_{i=1}^n g(y_i). \quad (1.25)$$

Si definimos  $T$  como el funcional que a cada distribución que tiene esperanza finita le asigna su esperanza, tenemos  $T(F) = E_F(y)$  con  $y \sim F$ . Por lo tanto, si  $\widehat{F}_n$  es como en (1.24),  $T(\widehat{F}_n)$  resulta ser la media muestral:  $T(\widehat{F}_n) = \bar{y}$ .

Veremos en esta sección que tanto la mediana, los M-estimadores de posición presentados como las medias  $\alpha$ -podadas resultan  $\widehat{\mu} = T(\widehat{F}_n)$ , para un apropiado funcional  $T$ . Considerando que  $\widehat{F}_n$  converge débilmente a  $F$  con probabilidad uno ( $\mathbb{P}(\widehat{F}_n \rightarrow F) = 1$ ), el enfoque funcional permite deducir la consistencia del estimador  $T(\widehat{F}_n)$  a partir de la continuidad débil del funcional  $T$ , en la medida que  $T(F)$  coincida con el parámetro que queremos estimar. En tal caso,  $T$  se dice Fisher-consistente para el parámetro de interés. En el presente contexto, la Fisher-consistencia para el parámetro de posición bajo el modelo (1.1) requiere que  $T(F) = \mu_0$ .

El enfoque funcional resulta de suma importancia en el presente trabajo para demostrar la consistencia de la propuesta de estimación que haremos en el contexto de datos faltantes.

Vamos ahora a definir los funcionales asociados a los diferentes estimadores considerados. Empecemos por introducir al funcional de posición asociado a una de las medidas de posición más populares: la mediana. El funcional  $T_{med}$  está dado por

$$T_{med}(F) = \arg \cdot \min_{t \in \mathbb{R}} E_F[|y - t|]. \quad (1.26)$$

Cuando existe más de un valor que realiza el mínimo, el funcional es definido eligiendo alguno de ellos. Luego, dada la muestra  $y_1, \dots, y_n$ ,  $T_{med}(\widehat{F}_n)$  es la mediana muestral.

Para poder establecer el funcional de posición asociado a la media  $\alpha$ -podada, presentaremos los L-funcionales, del que son un caso particular. Los L-funcionales se expresan de la siguiente manera:

$$T(F) = \int_0^1 F^{-1}(v)W(v) dv, \quad (1.27)$$

donde  $F^{-1}(u) = \inf\{y : F(y) \geq u\}$ ,  $W : [0, 1] \rightarrow \mathbb{R}_{\geq 0}$  es una función simétrica alrededor del  $1/2$ , no creciente para  $v \geq 1/2$ , y satisfaciendo  $\int_0^1 W(v) dv = 1$ . Sean  $y_{(i)}$ ,  $i = 1, \dots, n$ , los estadísticos de orden de la muestra. Entonces

$$T(\widehat{F}_n) = \sum_{i=1}^n w_{i,n} y_{(i)}$$

con  $w_{i,n} = \int_{(i-1)/n}^{i/n} W(v) dv$ . Sea

$$\alpha_0 = \inf\{\alpha \in [0, 1] : W(\alpha) > 0\}. \quad (1.28)$$

La media  $\alpha$ -podada,  $0 \leq \alpha < 1/2$ , es un L-funcional con

$$W(u) = \frac{I_{[\alpha, 1-\alpha]}(u)}{1 - 2\alpha}$$

y  $\alpha_0 = \alpha$ . Así,  $T(\widehat{F}_n)$  resulta la media  $\alpha$ -podada de la muestra.

Por último, presentamos los funcionales  $T_S$ ,  $T_B$  y  $T_M$ , asociados a los estimadores dados en (1.14), (1.10) y (1.11), respectivamente. Sean  $\rho_0$  y  $\rho_1$  dos  $\rho$ -funciones acotadas. Podemos pensar a  $\rho_0$  y  $\rho_1$  en la familia bicuadrada. Consideremos  $S^*(F, t)$  la solución de

$$E_F \left[ \rho_0 \left( \frac{y - t}{S^*(F, t)} \right) \right] = \delta \quad (1.29)$$

donde  $\delta \in (0, 1)$ . Luego el funcional de dispersión es

$$S(F) = \min_{t \in \mathbb{R}} S^*(F, t), \quad (1.30)$$

mientras que el S-funcional de posición está dado por

$$T_S(F) = \arg \cdot \min_{t \in \mathbb{R}} S^*(F, t). \quad (1.31)$$

**Definición 1.5.2.** Un funcional  $S$  definido sobre un espacio de distribuciones se dice de dispersión si satisface  $S(F) > 0$  y  $S(F_{\sigma z + \mu}) = |\sigma|S(F_z)$  donde  $F_z$  es la función de distribución de una variable aleatoria  $z$ .

Consideremos el M-funcional de posición correspondiente a  $\rho_1$ , dado por

$$T_B(F) = \arg \cdot \min_{t \in \mathbb{R}} E_F \left[ \rho_1 \left( \frac{y - t}{S(F)} \right) \right]. \quad (1.32)$$

Consideremos ahora una nueva  $\rho$ -función  $\rho$  con derivada  $\psi$  monótona. A modo de ejemplo,  $\rho$  puede pertenecer a la familia de Huber. El M-funcional de posición correspondiente a  $\rho$  es el único valor  $T_M(F)$  satisfaciendo

$$E_F \left[ \psi \left( \frac{y - T_M(F)}{S(F)} \right) \right] = 0. \quad (1.33)$$

Dada una muestra  $y_1, \dots, y_n$ , notemos que el estimador  $\widehat{\sigma}$  definido en (1.13) satisface  $\widehat{\sigma} = S(\widehat{F}_n)$ , mientras que  $T_S(\widehat{F}_n) = \widehat{\mu}_S$ , definido en (1.14).

Además,  $T_B(\widehat{F}_n)$  verifica

$$T_B(\widehat{F}_n) = \arg \cdot \min_{t \in \mathbb{R}} E_{\widehat{F}_n} \left( \rho_1 \left( \frac{y - t}{S(\widehat{F}_n)} \right) \right) = \arg \cdot \min_{t \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{y_i - t}{S(\widehat{F}_n)} \right)$$

y  $T_M(\widehat{F}_n)$  satisface

$$0 = E_{\widehat{F}_n} \left( \psi \left( \frac{y - T_M(\widehat{F}_n)}{S(\widehat{F}_n)} \right) \right) = \frac{1}{n} \sum_{i=1}^n \psi \left( \frac{y_i - T_M(\widehat{F}_n)}{S(\widehat{F}_n)} \right).$$

Es decir, recuperamos los estimadores definidos por (1.14), (1.10) y (1.11), usando como  $\widehat{\sigma}$  el  $S$ -funcional de dispersión evaluado en  $\widehat{F}_n$ .

Es fácil verificar que los funcionales que acabamos de definir resultan funcionales de posición, definidos de la siguiente manera:

**Definición 1.5.3.** Sea  $\mathcal{G}$  una clase de distribuciones univariadas. Un funcional de posición  $T : \mathcal{G} \rightarrow \mathbb{R}$  satisface  $T(F_{ay+b}) = aT(F_y) + b$  donde  $F_y$  denota la función de distribución de la variable aleatoria  $y$ .

Como mencionamos, la consistencia de los estimadores definidos mediante  $T(\widehat{F}_n)$  puede ser deducida de la Fisher-consistencia y la continuidad de  $T$  en  $F$ . Estudiaremos ahora condiciones que garanticen estas dos propiedades.

### 1.5.1. Consistencia de Fisher

Sea el modelo de posición dado por

$$y = \mu_0 + u \tag{1.34}$$

con  $F_0$  la función de distribución de  $u$ . Recordando lo mencionado al principio de este Capítulo tenemos que si  $F_{\mu_0}$  es la función de distribución de  $y$ ,  $F_{\mu_0}(y) = F_0(y - \mu_0)$ . Acabamos de ver que los estimadores presentados en este capítulo pueden ser vistos como cierto funcional  $T$  evaluado en la distribución empírica  $\widehat{F}_n$  asociada a la muestra. Lo que nos interesa ver ahora es bajo qué condiciones los funcionales de posición  $T$  que hemos definido satisfacen  $T(F_{\mu_0}) = \mu_0$ . Más generalmente, daremos la siguiente definición.

**Definición 1.5.4.** Un funcional  $T$  es *Fisher-consistente* si para la familia de distribuciones  $F_\theta$ ,  $T(F_\theta) = \theta$ .

Para estudiar la Fisher-consistencia de los estimadores de posición definidos bajo el modelo de posición, estudiaremos por separado el caso redescendiente y el caso monótono.

#### Estimadores de Posición Redescendientes

Para poder demostrar la Fisher-consistencia, precisamos efectuar las siguientes hipótesis sobre  $F_0$  y  $\rho$ .

- (B1)  $f_0 = F_0'$  es una función par,  $f_0(u)$  es monótona no creciente en  $|u|$  y es estrictamente decreciente en  $|u|$  en un entorno de 0.
- (B2)  $\rho(0) = 0$ ,  $\rho$  es una función par y derivable,  $0 \leq u < v$  implica  $\rho(u) \leq \rho(v)$ ,  $a = \sup_{x \in \mathbb{R}} \rho(x)$  es un número real positivo, y si  $u > 0$  y  $\rho(u) < a$  entonces  $\psi(u) = \rho'(u) > 0$ .

La demostración del siguiente resultado se encuentra en el Apéndice.

**Lema 1.5.5.** Sea  $F_0$  una distribución que satisface (B1) y sea  $\rho$  una función que satisface (B2). Si

$$g(\lambda) = E_{F_0}(\rho(u - \lambda)), \tag{1.35}$$

entonces (i)  $g(\lambda)$  es par, (ii) si  $0 \leq \lambda_1 < \lambda_2$  entonces  $g(\lambda_1) \leq g(\lambda_2)$  y (iii)  $g(0) < g(\lambda)$  para todo  $\lambda \neq 0$ .

**Corolario 1.5.6.** *Consideremos el modelo de posición (1.34), con  $u \sim F_0$  satisfaciendo las condición (B1) y  $\rho$  verificando (B2). Tenemos entonces que  $T_B(F_{\mu_0})$  está unívocamente definido y satisface  $T_B(F_{\mu_0}) = \mu_0$ ; lo mismo ocurre con  $T_S(F_{\mu_0})$ .*

*Demostración.* Recordemos la definición de  $T_B$  dada en (1.32). Definiendo

$$\rho_S(y) := \rho_1 \left( \frac{y}{S(F_{\mu_0})} \right),$$

tenemos que  $\rho_S$  también satisface la condición (B2). Sea  $g^*(\mu) = E_{F_{\mu_0}}(\rho_S(y - \mu))$ . Para  $\mu \neq \mu_0$  sea  $\lambda = \mu - \mu_0 \neq 0$ . Tenemos entonces que

$$g^*(\mu) = E_{F_{\mu_0}}(\rho_S(y - \mu)) = E_{F_0}(\rho_S(u + \mu_0 - \mu)) = E_{F_0}(\rho_S(u - \lambda)) =: g_S(\lambda),$$

siendo  $g_S$  definida como en (1.35) con  $\rho_S$ . Es claro que  $g^*(\mu_0) = g_S(0)$ . Por lo tanto tenemos que

$$g^*(\mu) - g^*(\mu_0) = g_S(\lambda) - g_S(0).$$

Usando el ítem (iii) del Lema 1.5.5, resulta que

$$g^*(\mu) - g^*(\mu_0) > 0 \quad \text{para todo } \mu \neq \mu_0.$$

En conclusión,  $\mu_0$  es el mínimo de  $g^*$  y por consiguiente,  $T_B(F_{\mu_0}) = \mu_0$ , como queríamos demostrar.

Notemos que  $T_S(F)$  satisface

$$T_S(F) = \arg \cdot \min_{t \in \mathbb{R}} E_F \left[ \rho_0 \left( \frac{y - t}{S(F)} \right) \right],$$

como  $\rho_0$  y  $\rho_1$  satisfacen las mismas propiedades, concluimos que el mismo resultado vale para  $T_S$ .  $\square$

### Estimadores de Posición Monótonos

Para establecer la Fisher-Consistencia de los funcionales de posición monótonos  $T_M$  definidos en (1.33), necesitamos que  $F_0$  y  $\psi$  cumplan las siguientes hipótesis.

(M1)  $f_0 = F_0'$  es una función par y existe  $a > 0$  tal que si  $|u| < a$  entonces  $f_0(u) > 0$ .

(M2)  $\psi$  es una función impar y continua,  $u < v$  implica  $\psi(u) \leq \psi(v)$  y existe  $a > 0$  tal que  $0 \leq u < v \leq a$  implica  $\psi(u) < \psi(v)$ .

**Lema 1.5.7.** *Sea  $F_0$  una distribución que satisface (M1) y sea  $\psi$  una función que satisface (M2). Si*

$$g(\lambda) = E_{F_0}(\psi(u - \lambda)),$$

*entonces (i)  $g(\lambda)$  es impar, (ii) si  $\lambda_1 < \lambda_2$  entonces  $g(\lambda_1) \geq g(\lambda_2)$  y (iii)  $g(\lambda) < 0$  para todo  $\lambda > 0$ .*

La demostración de este lema se encuentra en el Apéndice.

**Corolario 1.5.8.** *Consideremos el modelo de posición (1.34), con  $u \sim F_0$  satisfaciendo la condición (M1) y  $\psi$  verificando (M2). Tenemos entonces que  $T_M(F_{\mu_0}) = \mu_0$ .*

*Demostración.* Recordemos la definición de  $T_M$  dada en (1.33). Definiendo

$$\psi_S(y) := \psi \left( \frac{y}{S(F_{\mu_0})} \right),$$

tenemos que  $\psi_S$  también satisface la condición (M2). Sea  $g^*(\mu) = E_{F_{\mu_0}}(\psi_S(y - \mu))$ . Tenemos entonces que

$$g^*(\mu_0) = E_{F_{\mu_0}}(\psi_S(y - \mu_0)) = E_{F_0}(\psi_S(u)) = 0$$

ya que  $u$  es una v.a. simétrica.

Sea  $\mu \neq \mu_0$  y sea  $\lambda = \mu - \mu_0 \neq 0$ . Suponiendo que  $g^*(\mu) = 0$ , llegaremos a un absurdo. En tal caso, se tiene que

$$\lambda g^*(\mu) = 0. \tag{1.36}$$

Por otro lado,  $\lambda g^*(\mu) = E_{F_{\mu_0}}(\psi_S(y - \mu)\lambda) = E_{F_0}(\psi_S(u + \mu_0 - \mu)\lambda) = E_{F_0}(\psi_S(u - \lambda)\lambda) \neq 0$  por el lema anterior, lo cual contradice (1.36).  $\square$

### 1.5.2. Continuidad de los Funcionales de Posición

Comencemos recordando la definición de convergencia débil, para poder definir funcionales débilmente continuos.

**Definición 1.5.9.** Sea  $\{F_n\}_{n \in \mathbb{N}}$  una sucesión de funciones de distribución y sea  $F$  una función de distribución. Se dice que  $F_n$  converge débilmente a  $F$  ( $F_n \rightarrow_w F$ ) si, para todo  $x$  punto de continuidad de  $F$ , tenemos que

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

**Definición 1.5.10.** Un funcional  $T$  es débilmente continuo en  $F$  si para toda sucesión de funciones de distribución  $\{F_n\}$  que converge débilmente a  $F$ ,  $T(F_n) \rightarrow T(F)$ .

Sued y Yohai (2012) establecen condiciones que garantizan la continuidad débil de los L-funcionales, y en particular, de los asociados a las medias  $\alpha$ -podadas. A continuación enunciamos un teorema, extraído del mismo trabajo, que nos da condiciones que asegurarán la continuidad débil en  $F_{\mu_0}$  de  $T_B$  y  $T_M$ .

**Teorema 1.5.11.** Sean  $\rho_0$  una  $\rho$ -función acotada y  $\rho_1$  otra  $\rho$ -función acotada o convexa. Sean  $T_B$  y  $T_M$  los  $M$ -funcionales de posición definidos por (1.32) y (1.33), respectivamente, según  $\rho_1$  sea acotada o convexa, con  $S(F) > 0$  dado por (1.30). Si  $T_B(F)$  y  $T_S(F)$  ( $T_M(F)$  y  $T_S(F)$ ) están unívocamente determinados, entonces  $T_B$  ( $T_M$ ) es un funcional débilmente continuo en  $F$ .

Para finalizar esta sección, enunciamos el Teorema de Glivenko-Cantelli. El mismo garantiza la convergencia débil, con probabilidad total, de las distribuciones empíricas. Este resultado, combinado con la Fisher-consistencia y la continuidad de los funcionales, permiten deducir la consistencia de los estimadores dados por  $T(\hat{F}_n)$ .

**Teorema 1.5.12.** (de Glivenko-Cantelli) Sea  $x_1, \dots, x_n$  una muestra aleatoria de una distribución  $F$ , y sea  $F_n$  su distribución empírica, definida en (1.24). Entonces

$$\lim_{n \rightarrow \infty} \sup_x |F_n(x) - F(x)| = 0 \quad \text{c.t.p.}$$

## 1.6. Punto de Ruptura

Una forma de cuantificar la robustez de un estimador es mediante su punto de ruptura. Informalmente hablando, el punto de ruptura indica qué proporción de puntos atípicos se pueden reemplazar de la muestra para que el estimador siga brindando información sobre el parámetro que se desea estimar. Siguiendo con el enfoque funcional definido en este Capítulo, comenzaremos hablando del punto de ruptura de un funcional, para luego hablar del punto de ruptura de un estimador.

**Definición 1.6.1.** Consideremos un funcional  $T$  definido en un espacio de distribuciones. Su punto de ruptura en una distribución  $F$ , denotado con  $\varepsilon^*(T, F)$  es el mayor  $\varepsilon^* \in (0, 1)$  tal que para todo  $\varepsilon < \varepsilon^*$ ,  $T((1 - \varepsilon)F + \varepsilon H)$ , visto como función de  $H$ , permanece acotado y a distancia positiva del espacio donde toma valores el funcional. Cuando consideramos estimadores de la forma  $\hat{\theta} = T(\hat{F}_n)$ , decimos también que  $\varepsilon^*(T, F)$  es el punto de ruptura asintótico del estimador  $\hat{\theta}$  y el punto de ruptura asintótico de  $T$ .

La mayoría de los funcionales tiene punto de ruptura menor o igual a  $1/2$ . La demostración del siguiente resultado se encuentra en el Apéndice.

**Proposición 1.6.2.** Sea  $\rho$  con derivada  $\rho' = \psi$  monótona cumpliendo que  $k_1 = \lim_{x \rightarrow -\infty} -\psi(x)$  y  $k_2 = \lim_{x \rightarrow +\infty} \psi(x)$  son finitos. Consideremos el funcional  $T(F)$ , la solución de

$$E_F(\psi(y - T(F))) = 0. \quad (1.37)$$

Entonces

$$\varepsilon^*(T, F) = \frac{\min\{k_1, k_2\}}{k_1 + k_2}. \quad (1.38)$$

Vamos ahora a considerar M-funcionales de escala,  $R(F)$ , resolviendo

$$E_F\left(\rho\left(\frac{x}{R(F)}\right)\right) = \delta.$$

Podemos transformar a un M-funcional de escala en un M-funcional de posición con escala logarítmica, obteniendo que el punto de ruptura asintótico  $\varepsilon^*(R, F) = \min\{\delta, 1 - \delta\}$  si  $\rho$  es acotada, con  $\rho(\infty) = 1$ .

El punto de ruptura de  $T_B$  y  $T_M$  también puede ser calculado, pero dejamos estos resultados para el Capítulo 3, donde damos una nueva definición de punto de ruptura para funcionales de posición, que es la que utilizaremos en este trabajo.

### 1.6.1. Punto de Ruptura Muestral

Sea ahora  $\hat{\theta}_n = \hat{\theta}(\mathbf{x})$  un estimador de  $\theta \in \Theta$ , definido para muestras  $\mathbf{x} = \{x_1, \dots, x_n\}$ . El punto de ruptura muestral (PRM) de  $\hat{\theta}_n$  en  $\mathbf{x}$  es la mayor proporción  $\varepsilon_n^*(\hat{\theta}_n; \mathbf{x})$  de datos que pueden ser reemplazados arbitrariamente por outliers sin que  $\hat{\theta}_n$  deje de estar acotado ni tenga puntos de acumulación en la frontera de  $\Theta$ , como se define en el trabajo realizado por Donoho D.L. y Huber P.J. (1983). Simbólicamente, llamando

$$X_m = \{\mathbf{y} \in \mathbb{R}^n : \#(\mathbf{x} \cap \mathbf{y}) = n - m\}.$$

Sea

$$\hat{\theta}(X_m) = \{\theta(\mathbf{y}) : \mathbf{y} \in X_m\}$$

y consideremos  $m^*$  el máximo  $m$  de forma tal que  $\widehat{\theta}(X_m)$  permanece acotado sin puntos de acumulación en la frontera de  $\Theta$ . Luego

$$\varepsilon_n^*(\widehat{\theta}_n; \mathbf{x}) = \frac{m^*}{n}.$$

En la mayoría de los casos interesantes,  $\varepsilon_n^*$  no depende de  $\mathbf{x}$  y tiende al punto de ruptura asintótico, cuando  $n$  tiende a infinito. Se puede probar que para estimadores de posición equivariantes

$$\varepsilon_n^* \leq \frac{1}{n} \left[ \frac{n-1}{2} \right],$$

y esta cota sirve para M-estimadores con  $\psi$  impar y acotada. Para la media  $\alpha$ -podada,  $m^* = [n\alpha]$  entonces  $\varepsilon_n^* \approx \alpha$  para  $n$  grande.

## Capítulo 2

# Regresión Lineal

Sean  $x_{i1}, \dots, x_{ip}$  variables explicativas y sea  $y_i$  la variable respuesta o dependiente, de las cuales se tienen  $n$  observaciones,  $i = 1, \dots, n$ . Estas variables se relacionan mediante el modelo lineal:

$$y_i = \sum_{j=1}^p x_{ij}\beta_{0j} + u_i, \quad i = 1, \dots, n \quad (2.1)$$

donde  $\beta_{01}, \dots, \beta_{0p}$  son los parámetros desconocidos que se quieren estimar y los errores  $u_1, \dots, u_n$  son variables aleatorias. También, podemos expresar al modelo lineal matricialmente de la siguiente manera:  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u}$  siendo  $\mathbf{X} \in \mathbb{R}^{n \times p}$  la matriz que tiene de columnas a las variables explicativas,  $\boldsymbol{\beta}_0$  el vector columna  $p$ -dimensional de los parámetros desconocidos,  $\mathbf{y} = (y_1, \dots, y_n)'$  y  $\mathbf{u} = (u_1, \dots, u_n)'$  los vectores columna de la variable respuesta y del error, respectivamente.

En el caso que el modelo posea un término constante, el parámetro desconocido asociado se denomina ordenada al origen y la matriz  $\mathbf{X}$  resulta con una columna de 1's.

Los valores ajustados  $\hat{y}_i$  y los residuos  $r_i$  del modelo para un vector  $\boldsymbol{\beta}$  son  $\hat{y}_i(\boldsymbol{\beta}) = \mathbf{x}'_i\boldsymbol{\beta}$  y  $r_i(\boldsymbol{\beta}) = y_i - \hat{y}_i(\boldsymbol{\beta})$ .

Las variables explicativas o independientes pueden ser fijas o aleatorias. Vamos a estudiar los dos casos.

### 2.1. Variables Explicativas Fijas

Recordemos que el estimador de mínimos cuadrados (EMC) bajo un modelo lineal está dado por

$$\hat{\boldsymbol{\beta}}_{MC} = \arg \cdot \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n r_i^2(\boldsymbol{\beta}).$$

Diferenciando con respecto a  $\boldsymbol{\beta}$  tenemos

$$\sum_{i=1}^n r_i(\hat{\boldsymbol{\beta}}_{MC})\mathbf{x}_i = \mathbf{0}, \quad (2.2)$$

que es equivalente a resolver las denominadas ecuaciones normales, dadas por

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{MC} = \mathbf{X}'\mathbf{y}.$$

Cuando  $\mathbf{X}$  es de rango completo (columnas linealmente independientes asumiendo  $n > p$ ), garantiza la no singularidad de  $\mathbf{X}'\mathbf{X}$ , y de esta manera, una forma natural de resolver el sistema de ecuaciones es multiplicando a izquierda por la inversa de  $\mathbf{X}'\mathbf{X}$ . En tal caso, el EMC es único:

$$\widehat{\boldsymbol{\beta}}_{MC} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Cuando  $\mathbf{X}$  no es de rango completo, decimos que hay colinealidad entre las variables explicativas. Esto produce la no *identificabilidad* de los parámetros, en el sentido que existen  $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$  tales que  $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$ , lo que implica que se tengan infinitos estimadores, todos con los mismos valores ajustados y residuos.

El EMC cumple con las siguientes tres propiedades, que son deseables para un estimador de regresión:

- equivariancia por regresión:  $\widehat{\boldsymbol{\beta}}_{MC}(\mathbf{X}, \mathbf{y} + \mathbf{X}\boldsymbol{\gamma}) = \widehat{\boldsymbol{\beta}}_{MC}(\mathbf{X}, \mathbf{y}) + \boldsymbol{\gamma}$ ,  $\boldsymbol{\gamma} \in \mathbb{R}^p$
- equivariancia por escala:  $\widehat{\boldsymbol{\beta}}_{MC}(\mathbf{X}, \lambda\mathbf{y}) = \lambda\widehat{\boldsymbol{\beta}}_{MC}(\mathbf{X}, \mathbf{y})$ ,  $\lambda \in \mathbb{R}$
- equivariancia afín:  $\widehat{\boldsymbol{\beta}}_{MC}(\mathbf{X}\mathbf{A}, \mathbf{y}) = \mathbf{A}^{-1}\widehat{\boldsymbol{\beta}}_{MC}(\mathbf{X}, \mathbf{y})$ ,  $\mathbf{A} \in \mathbb{R}^{p \times p}$  no singular

Asumiendo que las  $u_i$ 's son v.a.i.i.d con  $E(u_i) = 0$  y  $Var(u_i) = \sigma^2$  y que  $\mathbf{X}$  es de rango completo, bajo el modelo lineal (2.1)  $\widehat{\boldsymbol{\beta}}_{MC}$  es un estimador insesgado, es decir  $E(\widehat{\boldsymbol{\beta}}_{MC}) = \boldsymbol{\beta}_0$ , y su matriz de covarianzas es  $Var(\widehat{\boldsymbol{\beta}}_{MC}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ . En el caso que las  $u_i$ 's son variables aleatorias normales y  $\mathbf{X}$  es de rango completo,  $\widehat{\boldsymbol{\beta}}_{MC}$  es un vector aleatorio normal multivariado

$$\widehat{\boldsymbol{\beta}}_{MC} \sim \mathcal{N}_p(\boldsymbol{\beta}_0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}).$$

Además, recordando que  $r_i(\boldsymbol{\beta}_0) \sim u_i$ , el EMC de  $\boldsymbol{\beta}_0$  coincide con el estimador de Máxima Verosimilitud (EMV) puesto que al maximizar

$$L(r_1, \dots, r_n; \boldsymbol{\beta}, \sigma) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-r_i^2(\boldsymbol{\beta})/2\sigma^2} = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n r_i^2(\boldsymbol{\beta})},$$

y aplicar  $\ln$ , buscamos encontrar  $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})$  cumpliendo

$$(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}) = \arg. \min_{(\boldsymbol{\beta}, \sigma)} \left[ \frac{1}{n} \sum_{i=1}^n \left( \frac{r_i(\boldsymbol{\beta})}{\sigma} \right)^2 + \ln(\sigma) \right]. \quad (2.3)$$

Nótese que  $\widehat{\boldsymbol{\beta}}$  se obtiene minimizando el primer término, independientemente del valor  $\sigma$ , resultando así ser el estimador de mínimos cuadrados. Además, diferenciando respecto de  $\boldsymbol{\beta}$  y  $\sigma$ , deducimos el siguiente sistema de ecuaciones:

$$\sum_{i=1}^n r_i(\widehat{\boldsymbol{\beta}}_{MV}) \mathbf{x}_i = \mathbf{0} \quad (2.4)$$

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{r_i(\widehat{\boldsymbol{\beta}}_{MV})}{\widehat{\sigma}_{MV}} \right)^2 = \frac{1}{2} \quad (2.5)$$

por lo que podemos corroborar que  $\widehat{\boldsymbol{\beta}}_{MV} = \widehat{\boldsymbol{\beta}}_{MC}$ .

En el caso que  $E(u_i) \neq 0$ , el EMC será un estimador sesgado, pero lo notable es que si el modelo contiene ordenada al origen, el sesgo lo afectará sólo a esta, es decir que si

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_O \\ \boldsymbol{\beta}_P \end{pmatrix}$$

donde  $\beta_O \in \mathbb{R}$  y  $\boldsymbol{\beta}_P \in \mathbb{R}^{p-1}$ , entonces  $E(\widehat{\boldsymbol{\beta}}_{P,MC}) = \boldsymbol{\beta}_P$ .

Sea  $p^*$  el rango de  $\mathbf{X}$ . Si  $p^* < p$ ,  $\mathbf{X}$  no es de rango completo, pero como todas las soluciones de (2.2) tienen los mismos residuos aunque  $\widehat{\boldsymbol{\beta}}_{MC}$  no está definido de manera única, un estimador insesgado de  $\sigma^2$  es

$$s^2 = \frac{1}{n - p^*} \sum_{i=1}^n r_i^2(\widehat{\boldsymbol{\beta}}).$$

Podemos entonces estimar la matriz de covarianzas de  $\widehat{\boldsymbol{\beta}}_{MC}$  mediante  $s^2(\mathbf{X}'\mathbf{X})^{-1}$ .

Ahora estudiemos estimadores de regresión lineal que combinen robustez y eficiencia.

Asumamos el modelo de regresión lineal donde  $u_i$  tiene densidad  $f_0(u/\sigma)/\sigma$ , siendo  $\sigma$  el parámetro de escala. Las  $y_i$ 's son variables aleatorias independientes pero no están idénticamente distribuidas, de hecho tienen densidad:

$$\frac{1}{\sigma} f_0\left(\frac{y - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma}\right).$$

La función de verosimilitud para  $\boldsymbol{\beta}$  y  $\sigma$  es

$$L(\boldsymbol{\beta}, \sigma) = \frac{1}{\sigma^n} \prod_{i=1}^n f_0\left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma}\right),$$

por lo cual maximizarla equivale a

$$(\widehat{\boldsymbol{\beta}}_{MV}, \widehat{\sigma}_{MV}) = \arg \cdot \min_{(\boldsymbol{\beta}, \sigma)} \left[ \frac{1}{n} \sum_{i=1}^n \rho_0\left(\frac{r_i(\boldsymbol{\beta})}{\sigma}\right) + \ln(\sigma) \right]$$

donde  $\rho_0 = -\ln(f_0)$ . En particular, tenemos que  $\widehat{\boldsymbol{\beta}}_{MV}$  satisface

$$\widehat{\boldsymbol{\beta}}_{MV} = \arg \cdot \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho_0\left(\frac{r_i(\boldsymbol{\beta})}{\widehat{\sigma}_{MV}}\right). \quad (2.6)$$

Por otra parte, diferenciando con respecto a  $\boldsymbol{\beta}$  y  $\sigma$  se obtiene el siguiente sistema de ecuaciones:

$$\sum_{i=1}^n \psi_0\left(\frac{r_i(\widehat{\boldsymbol{\beta}}_{MV})}{\widehat{\sigma}_{MV}}\right) \mathbf{x}_i = \mathbf{0} \quad (2.7)$$

$$\frac{1}{n} \sum_{i=1}^n \gamma_0\left(\frac{r_i(\widehat{\boldsymbol{\beta}}_{MV})}{\widehat{\sigma}_{MV}}\right) = 1 \quad (2.8)$$

siendo  $\gamma_0(x) = x \psi_0(x)$ .

La condición (2.6) motiva la siguiente definición para los M-estimadores de regresión

$$\widehat{\boldsymbol{\beta}} = \arg \cdot \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho\left(\frac{r_i(\boldsymbol{\beta})}{\widehat{\sigma}}\right) \quad (2.9)$$

donde  $\rho$  es una  $\rho$ -función y  $\hat{\sigma}$  puede ser estimado previamente o simultáneamente por

$$\frac{1}{n} \sum_{i=1}^n \rho_{escala} \left( \frac{r_i(\hat{\beta})}{\hat{\sigma}} \right) = \delta,$$

con  $\delta \in (0, 1)$ . Además, se obtiene el análogo de las ecuaciones normales, diferenciando respecto de  $\beta$ :

$$\sum_{i=1}^n \psi \left( \frac{r_i(\hat{\beta})}{\hat{\sigma}} \right) \mathbf{x}_i = \mathbf{0} \quad (2.10)$$

con  $\psi = \rho'$ .

Estudiemos ahora el punto de ruptura de los estimadores presentados.

Supongamos que la matriz  $\mathbf{X}$  tiene rango completo. Como  $\mathbf{X}$  está fijo pero  $\mathbf{y}$  es un vector aleatorio, el PRM para la regresión viene dado por

$$\varepsilon^* = \frac{m^*}{n},$$

con  $m^* = \max\{m \geq 0 : \hat{\beta}(\mathbf{X}, \mathbf{y}_m) \text{ está acotado para todo } \mathbf{y}_m \in \mathcal{Y}_m\}$  donde  $\mathcal{Y}_m$  es el conjunto de vectores en  $\mathbb{R}^n$  con al menos  $n - m$  elementos en común con  $\mathbf{y}$ . En particular, el EMC tiene punto de ruptura  $\varepsilon^* = 0$ .

Denotemos  $k^* = k^*(\mathbf{X})$  a la máxima cantidad de  $\mathbf{x}_i$  que se encuentran en el mismo subespacio de dimensión  $< p$ :

$$k^* = \max\{\#(\boldsymbol{\theta}'\mathbf{x}_i = 0) : \boldsymbol{\theta} \in \mathbb{R}^p, \boldsymbol{\theta} \neq \mathbf{0}\}. \quad (2.11)$$

En general, cuando la matriz  $\mathbf{X}$  no tiene rango completo, se redefinen  $m^*$  y  $k^*$  como

$$m^* = \max\{m \geq 0 : \mathbf{X}\hat{\beta}(\mathbf{X}, \mathbf{y}_m) \text{ está acotado para todo } \mathbf{y}_m \in \mathcal{Y}_m\},$$

$$k^* = \max\{\#(\boldsymbol{\theta}'\mathbf{x}_i = 0) : \boldsymbol{\theta} \in \mathbb{R}^p, \mathbf{X}\boldsymbol{\theta} \neq \mathbf{0}\}.$$

En la sección 4.9.3 del libro de Maronna R.A., Martin R.D. y Yohai V.J. (2006) está demostrado que para estimadores equivariantes por regresión

$$\varepsilon^* \leq \varepsilon_{max}^* = \frac{m_{max}^*}{n}$$

donde  $m_{max}^* = \lfloor \frac{n-k^*-1}{2} \rfloor \leq \lfloor \frac{n-p}{2} \rfloor$  pues  $k^* \geq p - 1$  siempre.

## 2.2. Variables Explicativas Aleatorias

A continuación presentaremos estimadores y propiedades de los mismos que serán de suma importancia a lo largo de este trabajo.

En el caso de diseño de experimentos, los  $\mathbf{x}_i$  no resultan fijos si no que son vectores aleatorios. Por lo tanto bajo el modelo de regresión lineal, las  $u_i$ 's son v.a.i.i.d independientes de los  $\mathbf{x}_i$ 's. El análogo de asumir  $\mathbf{X}$  de rango completo en el modelo donde están fijas, es suponer que la distribución de  $\mathbf{x}_i$  no se concentra en ningún hiperplano, es decir  $\mathbb{P}(\mathbf{a}'\mathbf{x}_i = 0) < 1$  para todo  $\mathbf{a} \neq \mathbf{0}$ .

Para el EMC obtenemos  $E(\widehat{\beta}_{MC}|\mathbf{X}) = \beta_0$  y  $Var(\widehat{\beta}_{MC}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$  la matriz de covarianzas, cuando  $Var(u_i) = \sigma^2$ . También se cumple que condicional a  $\mathbf{X}$  el EMC resulte normal multivariado con los mismos parámetros definidos con la  $\mathbf{X}$  fija siempre que se supongan las  $u_i$ 's normales.

Para obtener estimadores de regresión robustos donde tanto las  $\mathbf{x}_i$  como las  $y_i$  puedan ser outliers, es necesario usar M-estimadores  $\widehat{\beta}$  que sean solución de (2.9) con  $\rho$  acotada y un alto punto de ruptura de la escala preliminar  $\widehat{\sigma}$ . En tal caso  $\psi = \rho'$  resultará redescendiente dado que es simple verificar que una función  $\rho$  con  $\psi$  monótona no será acotada.

Vamos ahora a introducir los S y MM estimadores de regresión. Sea  $\mathbf{r}(\beta) = (r_1(\beta), \dots, r_n(\beta))$  un vector de residuos. Para cada  $\beta$  sea  $\widehat{\sigma}(\beta)$  un M-estimador de escala para  $\mathbf{r}(\beta)$ , definido por

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{r_i(\beta)}{\widehat{\sigma}(\beta)} \right) = \delta$$

con  $\rho$  una  $\rho$ -función acotada y  $\delta \in (0, 1)$ . El estimador de regresión resultante

$$\widehat{\beta}_S = \arg \cdot \min_{\beta \in \mathbb{R}^p} \widehat{\sigma}(\beta) \quad (2.12)$$

se denomina un **S-estimador**. También se define la S-escala de los residuos con

$$\widehat{\sigma}_0 = \min_{\beta \in \mathbb{R}^p} \widehat{\sigma}(\beta) \quad (2.13)$$

Lamentablemente, los S-estimadores con  $\rho$  suave no pueden tener simultáneamente alto punto de ruptura asintótico y alta eficiencia. De hecho en el trabajo de Hössjer (1992) se demuestra que la eficiencia asintótica (bajo normalidad de los errores) de un S-estimador con punto de ruptura igual a 1/2 es 0,33.

Yohai (1987) logra superar esta limitación, introduciendo los MM estimadores, que combinan un alto punto de ruptura y alta eficiencia bajo errores normales. Para obtener el estimador propuesto por Yohai, consideremos  $\rho_0$  y  $\rho_1$  dos  $\rho$ -funciones acotadas tales que  $\rho_0 \geq \rho_1$ . Sea  $\widehat{\sigma}_0$  la S-escala definida en (2.13) utilizando  $\rho_0$ . El MM-estimador es el M-estimador construido con  $\rho_1$  que utiliza  $\widehat{\sigma}_0$  por escala. Es decir,  $\widehat{\beta}_{MM}$  es aquel que minimiza

$$L(\beta) = \sum_{i=1}^n \rho_1 \left( \frac{r_i(\beta)}{\widehat{\sigma}_0} \right). \quad (2.14)$$

Equivalentemente, podemos escribir

$$\widehat{\beta}_{MM} = \arg \cdot \min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{r_i(\beta)}{\widehat{\sigma}_0} \right). \quad (2.15)$$

### 2.2.1. Punto de Ruptura de MM-estimadores

Focalizaremos nuestra atención en el PRM de  $\widehat{\beta}$ . Consideremos  $\mathbf{z}_i = (\mathbf{x}_i, y_i)$  y escribamos al estimador dependiendo de los datos, o sea  $\widehat{\beta}(\mathbf{Z})$  con  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ . Por lo tanto resulta

$$\varepsilon^* = \frac{m^*}{n} \quad \text{donde} \quad m^* = \max\{m \geq 0 : \widehat{\beta}(\mathbf{Z}_m) \text{ está acotado para todo } \mathbf{Z}_m \in \mathcal{Z}_m\},$$

y  $\mathcal{Z}_m$  es el conjunto de vectores en  $\mathbb{R}^n$  con al menos  $n - m$  elementos en común con  $\mathbf{Z}$ . Como nos encontramos en la situación donde varían tanto  $\mathbf{x}_i$  como  $y_i$ , el  $m^*$  recién introducido resulta ser menor o igual al presentado antes, en la sección 2.1.

Además se puede probar que, para estimadores equivariantes por regresión, vale la misma cota para  $\varepsilon^*$  con  $k^*$  como se define en (2.11). Es más,

$$\varepsilon_{max}^* = \frac{1 - c}{2} \quad \text{siendo} \quad c = \max_{\boldsymbol{\theta} \neq \mathbf{0}} \mathbb{P}(\boldsymbol{\theta}' \mathbf{x} = 0),$$

según se afirma en la página 130 del libro de Maronna R.A., Martin R.D. y Yohai V.J. (2006).

### 2.3. MM-funcionales

Supongamos que tenemos el siguiente modelo de regresión lineal

$$y_i = \boldsymbol{\beta}'_0 \mathbf{x}_i + u_i, \quad i = 1, \dots, n, \quad (2.16)$$

con  $y_i, u_i \in \mathbb{R}, \mathbf{x}_i \in \mathbb{R}^p, \boldsymbol{\beta}_0 \in \mathbb{R}^p, u_i$  independiente de  $\mathbf{x}_i$ .

En este Capítulo, hemos definido a los MM-estimadores en su versión muestral. Ahora, es conveniente escribirlos en la forma funcional. Incluimos también los S-funcionales de regresión, pese a que en el próximo capítulo sólo utilizaremos los MM.

Consideremos un conjunto  $\mathcal{G}$  de distribuciones en  $\mathbb{R}^{p+1}$  para  $(\mathbf{x}, y)$ . Para determinar el MM-funcional de regresión  $T_{MM}(G)$ , para  $G \in \mathcal{G}$ , consideremos  $\rho_0 \geq \rho_1$ , dos  $\rho$ -funciones acotadas. La función  $\rho_0$  es aquella que define a  $S(G)$  el funcional de dispersión de la distribución de los errores, como se detalla a continuación.

Sea  $S^*(G, \boldsymbol{\beta})$  con  $G \in \mathcal{G}$ , la solución de

$$E_G \left[ \rho_0 \left( \frac{y - \boldsymbol{\beta}' \mathbf{x}}{S^*(G, \boldsymbol{\beta})} \right) \right] = \delta \quad (2.17)$$

donde  $\delta \in (0, 1)$ . Luego el funcional de dispersión  $S(G)$  está dado por

$$S(G) = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} S^*(G, \boldsymbol{\beta}) \quad (2.18)$$

mientras que el S-funcional de regresión está dado por

$$T_S(G) = \arg. \min_{\boldsymbol{\beta} \in \mathbb{R}^p} S^*(G, \boldsymbol{\beta}). \quad (2.19)$$

Finalmente, el MM-funcional de regresión se define siendo

$$T_{MM}(G) = \arg. \min_{\boldsymbol{\beta} \in \mathbb{R}^p} E_G \left[ \rho_1 \left( \frac{y - \boldsymbol{\beta}' \mathbf{x}}{S(G)} \right) \right]. \quad (2.20)$$

Notemos que, teniendo una muestra  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ , el MM-estimador de regresión  $\hat{\boldsymbol{\beta}}_{MM}$  definido en (2.15) coincide con  $T_{MM}(\hat{G}_n)$ , donde  $\hat{G}_n$  es la distribución empírica de la muestra, y lo mismo ocurre con el S estimador de regresión. Es decir, tenemos que  $\hat{\boldsymbol{\beta}}_{MM} = T_{MM}(\hat{G}_n)$  y  $\hat{\boldsymbol{\beta}}_S = T_S(\hat{G}_n)$ . La Fisher-consistencia de estos funcionales bajo el modelo de regresión junto con la continuidad débil de los mismos permiten deducir la consistencia de los estimadores, siendo que  $\mathbb{P}(\hat{G}_n \rightarrow_w G) = 1$  (Glivenko-Cantelli).

El punto de ruptura asintótico de  $T_{MM}$  es

$$\varepsilon^* = \min\{\delta, 1 - \delta - C(G)\}$$

donde

$$C(G) = \sup_{\gamma \neq 0} \mathbb{P}_G(\gamma' \mathbf{x} = 0), \quad (2.21)$$

similar a la igualdad mencionada al final de la subsección anterior. El máximo punto de ruptura se alcanza cuando  $\delta = (1 - C(G))/2$  y su valor es  $(1 - C(G))/2$ , que es el máximo punto de ruptura para funcionales equivariantes por regresión.

Los MM-estimadores para regresiones lineales combinan el punto de ruptura más alto posible con eficiencia bajo normalidad tan alta como se desea, lo cual está demostrado en el trabajo de Yohai (1987).

### 2.3.1. Consistencia de Fisher y Continuidad del MM-funcional de Regresión

El siguiente resultado permite garantizar la Fisher-consistencia de los funcionales de regresión MM y S.

**Lema 2.3.1.** *Consideremos  $(\mathbf{x}_i, y_i)$  con distribución  $G$  satisfaciendo el modelo lineal (2.16). Denotemos con  $f_0$  a la función de densidad de  $u_i$ . Supongamos que  $f_0$  satisface la condición (B1) y  $\rho$  satisface la condición (B2) del Lema 1.5.5. Si  $S(G) > 0$  y para todo  $\gamma \neq \mathbf{0}$   $\mathbb{P}(\gamma' \mathbf{x}_i = 0) < 1$ , tenemos que*

$$L(\boldsymbol{\beta}) = E_G \left( \rho \left( \frac{y_i - \boldsymbol{\beta}' \mathbf{x}_i}{S(G)} \right) \right) \quad (2.22)$$

tiene un único mínimo en  $\boldsymbol{\beta}_0$ .

*Demostración.* Sea  $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$  y llamemos  $\boldsymbol{\gamma} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$ .

$$\begin{aligned} L(\boldsymbol{\beta}) &= E \left( \rho \left( \frac{y_i - \boldsymbol{\beta}' \mathbf{x}_i}{S(G)} \right) \right) = E \left( \rho \left( \frac{u_i + \boldsymbol{\beta}'_0 \mathbf{x}_i - \boldsymbol{\beta}' \mathbf{x}_i}{S(G)} \right) \right) \\ &= E \left( \rho \left( \frac{u_i - \boldsymbol{\gamma}' \mathbf{x}_i}{S(G)} \right) \right) = E \left[ E \left( \rho \left( \frac{u_i - \boldsymbol{\gamma}' \mathbf{x}_i}{S(G)} \right) \middle| \mathbf{x}_i \right) \right]. \end{aligned}$$

Como

$$E \left( \rho \left( \frac{u_i - \boldsymbol{\gamma}' \mathbf{x}_i}{S(G)} \right) \middle| \mathbf{x}_i \right) = g(\boldsymbol{\gamma}' \mathbf{x}_i)$$

resulta que

$$L(\boldsymbol{\beta}) = E(g(\boldsymbol{\gamma}' \mathbf{x}_i)).$$

Además es claro que  $L(\boldsymbol{\beta}_0) = g(0)$ . Luego

$$L(\boldsymbol{\beta}) - L(\boldsymbol{\beta}_0) = E(g(\boldsymbol{\gamma}' \mathbf{x}_i) - g(0)).$$

Por lo tanto, usando el Lema 1.5.5 tenemos que por el ítem (ii)  $g(\boldsymbol{\gamma}' \mathbf{x}_i) - g(0) \geq 0$ , y por el ítem (iii),  $\mathbb{P}(g(\boldsymbol{\gamma}' \mathbf{x}_i) - g(0) > 0) > 0$  dado que  $\mathbb{P}(\boldsymbol{\gamma}' \mathbf{x}_i = 0) < 1$ . Entonces podemos denotar por  $W$  a la variable aleatoria  $W = g(\boldsymbol{\gamma}' \mathbf{x}_i) - g(0)$  que cumple:  $W \geq 0$  y  $\mathbb{P}(W > 0) > 0$ . En conclusión

$$L(\boldsymbol{\beta}) - L(\boldsymbol{\beta}_0) = E(W) > 0$$

lo cual afirma que  $\boldsymbol{\beta}_0$  es mínimo de  $L$ . □

**Corolario 2.3.2.** Consideremos  $(\mathbf{x}_i, y_i)$  con distribución  $G$  satisfaciendo el modelo lineal (2.16). Denotemos con  $f_0$  a la función de densidad de  $u_i$ . Supongamos que  $f_0$  satisface la condición (B1) y  $S(G) > 0$ .

(i) Si  $\rho_1$  satisface la condición (B2) del Lema 1.5.5, el MM-funcional de regresión está unívocamente definido y satisface que  $T_{MM}(G) = \beta_0$ .

(ii) Si  $\rho_0$  satisface la condición (B2) del Lema 1.5.5, el S-funcional de regresión está unívocamente definido y satisface que  $T_S(G) = \beta_0$ .

*Demostración.* El ítem (i) se deduce inmediatamente de la definición del MM-funcional, combinado con el Lema 2.3.1. Para mostrar la Fisher-consistencia del S-funcional de regresión, notemos que puede ser visto como

$$T_S(G) = \arg . \min_{\beta \in \mathbb{R}^p} E_G \left[ \rho_0 \left( \frac{y_i - \beta' \mathbf{x}_i}{S(G)} \right) \right].$$

□

Por último, mencionamos el resultado de continuidad de los S y MM funcionales que se deducen de los resultados presentados en el trabajo de Fasano et. al (2012).

**Lema 2.3.3.** Consideremos  $(\mathbf{x}_i, y_i)$  con distribución  $G$  satisfaciendo el modelo lineal (2.16). Denotemos con  $f_0$  a la función de densidad de  $u_i$ . Supongamos que  $f_0$  satisface la condición (B1) y  $S(G) > 0$ . Si  $\delta < 1 - c(G_0)$  with  $c(G)$  definido en (2.21), entonces  $S(G)$  y  $T_S$  son débilmente continuos en  $G$ . Si además

$$E_G \left[ \rho \left( \frac{y_i - T_{MM}(G)' \mathbf{x}_i}{S(G)} \right) \right] < 1 - c(G),$$

entonces  $T_{MM}$  es débilmente continuo en  $G$ .

Terminamos el capítulo mencionando la consistencia de los S y MM estimadores, como consecuencia de la Fisher-consistencia, la continuidad de los funcionales y el Lema de Glivenko-Cantelli.

## Capítulo 3

# Estimación de Posición con Datos Faltantes

En este capítulo presentaremos una nueva propuesta para estimar medidas de posición de cierta variable respuesta en el contexto de datos faltantes. Estudiaremos, además, algunas propiedades de esta nueva propuesta que es el resultado de modificar levemente la introducida en el trabajo de Sued y Yohai (2012).

### 3.1. Estimadores de Posición con Datos Faltantes

Supongamos que se tienen observaciones  $y_i$  de una variable respuesta, que por alguna circunstancia no son observadas en algunos individuos. Un problema clásico consiste en estimar  $E(y_i)$ , la media de la variable respuesta  $y_i$ . Suponiendo que se tienen variables explicativas asociadas a esta variable respuesta  $\mathbf{x}_i$  que son observadas para todo índice  $i$ , para identificar el parámetro de interés con la distribución de los datos observados se asume Missing at Random (MAR). Este supuesto establece que la distribución condicional de las respuestas dado el vector de variables explicativas permanece igual sin importar si la respuesta también es observada. Para ser más específicos, consideremos la variable dicotómica  $a_i$  tal que

$$a_i = \begin{cases} 1 & \text{si la } i\text{-ésima variable respuesta es observada} \\ 0 & \text{si no} \end{cases} \quad (3.1)$$

La condición MAR afirma que

$$\mathbb{P}(a_i = 1 | \mathbf{x}_i, y_i) = \mathbb{P}(a_i = 1 | \mathbf{x}_i),$$

o equivalentemente

$$y_i | \mathbf{x}_i \sim y_i | \mathbf{x}_i, a_i = 1.$$

Así obtenemos que

$$E(y_i) = E(E(y_i | \mathbf{x}_i)) = E(E(y_i | \mathbf{x}_i, a_i = 1)).$$

Esto sugiere estimar  $E(y_i)$  basado en un modelo de regresión con  $E(y_i | \mathbf{x}_i) = g(\mathbf{x}_i, \boldsymbol{\beta}_0)$ , ajustando  $\hat{\boldsymbol{\beta}}$  de manera de utilizar sólo los pares  $(\mathbf{x}_i, y_i)$  para los cuales la respuesta  $y_i$  es observada. Luego estimamos  $E(y_i)$  con  $\sum_{i=1}^n g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})/n$ .

$E(y_i)$  es la medida de posición clásica, sin embargo no es robusta. Por esta razón, es necesario cambiar el objeto estadístico de interés: estudiaremos estimadores consistentes de

$$\mu_0 = T(F) \quad (3.2)$$

siendo  $T$  un funcional de posición débilmente continuo en  $F$ , la distribución de  $y_i$ . La propuesta asumirá que las covariables  $\mathbf{x}_i$  se relacionan con la respuesta mediante el modelo lineal

$$y_i = \beta_0' \mathbf{x}_i + u_i, \quad (3.3)$$

con  $y_i, u_i \in \mathbb{R}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$ ,  $\beta_0 \in \mathbb{R}^p$ ,  $u_i$  independiente de  $\mathbf{x}_i$ . Es más, para garantizar el cumplimiento de la condición MAR, asumimos que  $u_i$  es independiente de  $(\mathbf{x}_i, a_i)$  y que su distribución satisface la condición (B1) enunciada en el Capítulo 1.

Nótese que siendo  $T$  continuo en  $F$ , para estimar  $T(F)$  basta considerar  $T(F_n)$ , siendo  $F_n$  una sucesión de distribuciones que converge a  $F$  débilmente.

El procedimiento propuesto por Sued y Yohai (2012) consiste en los siguientes pasos.

- Consideran  $\hat{\beta}_n$ , un estimador de  $\beta_0$ , que se obtiene al ajustar de manera robusta el modelo lineal, utilizando sólo los pares  $(\mathbf{x}_i, y_i)$  tales que  $a_i = 1$ . Dicha estimación se puede realizar dado que  $u_i$  y  $a_i$  son independientes.
- Luego consideran la siguiente muestra ficticia

$$\hat{y}_{ij}^{SY} = \hat{u}_i + \hat{\beta}_n' \mathbf{x}_j, \quad i, j \in \{1, \dots, n\}, a_i = 1,$$

donde, para  $i$  con  $a_i = 1$ ,

$$\hat{u}_i = y_i - \hat{\beta}_n' \mathbf{x}_i,$$

y, denominando  $\delta_s$  a la función de distribución del punto masa  $s$ , estiman  $F$  con

$$\hat{H}_n^* = \frac{1}{n \sum_{i=1}^n a_i} \sum_{i=1}^n \sum_{j=1}^n a_i \delta_{\hat{y}_{ij}^{SY}}. \quad (3.4)$$

es decir

$$\hat{H}_n^*(y) = \frac{1}{n \sum_{i=1}^n a_i} \sum_{i=1}^n \sum_{j=1}^n a_i I_{\{\hat{y}_{ij}^{SY} \leq y\}}.$$

- Finalmente, estiman  $T(F)$  con  $T(\hat{H}_n^*)$ .

Los autores prueban la consistencia del estimador, encuentran su distribución asintótica y obtienen una cota inferior para el punto de ruptura del procedimiento.

En este trabajo presentamos una nueva propuesta para estimar  $F$ . La diferencia radica en que cuando  $a_i = 1$  el valor  $\hat{y}_{ij}^{SY}$  se reemplaza por  $y_i$ . De esta forma esperamos mitigar el efecto de la presencia de outliers, obteniendo un estimador con mayor punto de ruptura.

## 3.2. Nueva Propuesta de Estimación

Sea

$$A = \{i : i = 1, \dots, n, a_i = 1\}$$

de cardinal  $\mathcal{C}(A) = m$ , por lo cual

$$\sum_{j=1}^n a_j = m.$$

Sea  $\widehat{\beta}_n$  el estimador de  $\beta_0$  ya introducido. Para cada  $i \in A$ , llamemos  $\widehat{u}_i = y_i - \widehat{\beta}'_n \mathbf{x}_i$ . A partir de la muestra obtenida, consideramos la siguiente muestra ficticia de tamaño  $nm$

$$\widehat{y}_{ij} = \begin{cases} \widehat{u}_i + \widehat{\beta}'_n \mathbf{x}_j & \text{si } i \in A, j \in A^C \\ y_i & \text{si } i, j \in A \end{cases} \quad (3.5)$$

Definimos un nuevo estimador  $\widehat{F}_n$  dado por la función de distribución empírica en  $\widehat{y}_{ij}$ :

$$\widehat{F}_n = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^n (\delta_{y_i} a_i a_j + \delta_{\widehat{y}_{ij}} a_i (1 - a_j)) = \frac{1}{n} \sum_{i=1}^n \delta_{y_i} a_i + \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^n \delta_{\widehat{y}_{ij}} a_i (1 - a_j). \quad (3.6)$$

Luego, estimaremos  $\mu_0$  por

$$\widehat{\mu}_n = T(\widehat{F}_n) \quad (3.7)$$

Nótese que podemos reescribir a  $\widehat{F}_n$  como combinación convexa de funciones de distribución empíricas de la siguiente manera

$$\widehat{F}_n = \frac{m}{n} \frac{1}{m} \sum_{i=1}^n \delta_{y_i} a_i + \left(1 - \frac{m}{n}\right) \frac{1}{(n-m)m} \sum_{i=1}^n \sum_{j=1}^n \delta_{\widehat{y}_{ij}} a_i (1 - a_j).$$

Más aún, si definimos

$$\widehat{R}_n = \frac{1}{\sum_{j=1}^n (1 - a_j)} \sum_{j=1}^n (1 - a_j) \delta_{\widehat{\beta}'_n \mathbf{x}_j} \quad (3.8)$$

y

$$\widehat{K}_n = \frac{1}{\sum_{i=1}^n a_i} \sum_{i=1}^n a_i \delta_{\widehat{u}_i}. \quad (3.9)$$

Ahora necesitamos utilizar la definición de la convolución entre dos funciones de distribución: sean  $\widetilde{U} \sim F_1$  y  $\widetilde{V} \sim F_2$

$$\begin{aligned} (F_1 * F_2)(s) &:= \int_{-\infty}^{+\infty} F_1(s-x) dF_2(x) = E_{F_2}(F_1(s-\widetilde{U})) \\ &= \int_{-\infty}^{+\infty} F_2(s-x) dF_1(x) = E_{F_1}(F_2(s-\widetilde{V})) \end{aligned}$$

donde  $*$  denota la convolución entre dos funciones de distribución, y entonces tenemos que

$$\widehat{F}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i} a_i + \left(1 - \frac{m}{n}\right) \widehat{R}_n * \widehat{K}_n. \quad (3.10)$$

Notemos que, a diferencia de la propuesta hecha en Sued y Yohai (2012), nosotros aprovechamos las respuestas observadas y no las predecimos. De esta manera, obtenemos un procedimiento con mejor cota inferior para el punto de ruptura.

### 3.3. Consistencia

Enunciaremos a continuación uno de los principales resultados de este trabajo.

**Teorema 3.3.1.** *Sea  $\widehat{F}_n$  definida en (3.6), siendo  $\widehat{\beta}_n$  un estimador fuertemente consistente para  $\beta_0$ . Entonces:*

- (a)  $\{\widehat{F}_n\}$  converge débilmente a  $F$  c.t.p., es decir  $\mathbb{P}(\widehat{F}_n \rightarrow_w F) = 1$ .  
 (b) Asumiendo que  $T$  es un funcional débilmente continuo en  $F$ , tenemos que  $\widehat{\mu}_n = T(\widehat{F}_n)$  converge c.t.p. a  $\mu_0 = T(F)$ .

La demostración del Teorema 3.3.1 requiere de los siguientes resultados auxiliares. La demostración del próximo lema se encuentra en el Apéndice mientras que los dos sucesivos resultados que le siguen se encuentran demostrados en Sued y Yohai (2012).

**Lema 3.3.2.** *Sean  $\{z_i\}$  una sucesión de v.a.i.i.d. en  $\mathbb{R}^k$ ,  $h : \mathbb{R}^k \times \mathbb{R}^p \rightarrow \mathbb{R}^q$  una función continua y  $\{c_i\}$  una sucesión de variables Bernoulli i.i.d. Supongamos que  $\{\widehat{\beta}_n\}_{n \in \mathbb{N}}$  es una sucesión de estimadores fuertemente consistente de  $\beta_0 \in \mathbb{R}^p$ . Consideremos*

$$\begin{aligned} \widetilde{H}_n &= \frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n c_i \delta_{h(z_i, \widehat{\beta}_n)} \\ \widetilde{H}_0 &\sim h(z_1, \beta_0) | c_1 = 1. \end{aligned}$$

Entonces  $\widetilde{H}_n$  converge débilmente a  $\widetilde{H}_0$  c.t.p., es decir

$$\mathbb{P}(\widetilde{H}_n \rightarrow_w \widetilde{H}_0) = 1.$$

**Lema 3.3.3.** *Sean  $\{K_n\}_{n \in \mathbb{N}}$  y  $\{R_n\}_{n \in \mathbb{N}}$  dos sucesiones de funciones de distribución y sean  $K_0$  y  $R_0$  dos funciones de distribución. Si  $K_n \rightarrow_w K_0$  y  $R_n \rightarrow_w R_0$ , entonces  $K_n * R_n \rightarrow_w K_0 * R_0$ .*

**Lema 3.3.4.** *Sea  $\{(a_i, z_i)\}$  una sucesión de vectores aleatorios independientes e idénticamente distribuidos,  $a_i$  con distribución Bernoulli y  $z_i \in \mathbb{R}^h$ . Entonces*

$$\sup_{z \in \mathbb{R}^h} \left| \frac{1}{n} \sum_{i=1}^n a_i I_{\{z_i \leq z\}} - E(a_1 I_{\{z_1 \leq z\}}) \right| \rightarrow 0 \quad \text{c.t.p.}$$

**Demostración del Teorema 3.3.1** Recordemos que podemos escribir a  $\widehat{F}_n$  como

$$\widehat{F}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i} a_i + \left(1 - \frac{m}{n}\right) \widehat{R}_n * \widehat{K}_n,$$

con  $\widehat{R}_n$  y  $\widehat{K}_n$  definidos en (3.8) y (3.9), respectivamente. Del Lema 3.3.4 podemos deducir que

$$\sup_{z \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n a_i \delta_{y_i}(z) - P(a=1) F_{|a_1=1}(z) \right| \rightarrow 0 \quad \text{c.t.p.}$$

Sea  $R_0$  la función de distribución de  $\beta'_0 \mathbf{x}_i$  para los  $i$  tales que  $y_i$  es faltante, o sea para los  $a_i = 0$ :  $R_0 \sim \beta'_0 \mathbf{x}_i |_{a_i=0}$ . Sea  $K_0$  la función de distribución de  $u_i$ . Veamos que  $F_{|a_1=0} = R_0 * K_0$ .

Sabemos que

$$F_{|a_1=0}(s) = \mathbb{P}(\beta'_0 \mathbf{x} + u \leq s | a_1=0) = \frac{\mathbb{P}(\beta'_0 \mathbf{x} + u \leq s, a_1=0)}{\mathbb{P}(a_1=0)}.$$

En particular,

$$\begin{aligned}\mathbb{P}(\beta'_0 \mathbf{x} + u \leq s, a_1 = 0) &= E(I_{(-\infty, s] \times \{0\}}(\beta'_0 \mathbf{x} + u, a_1)) \\ &= E[E(I_{(-\infty, s] \times \{0\}}(\beta'_0 \mathbf{x} + u, a_1) | u)].\end{aligned}\quad (3.11)$$

Usando la independencia existente entre las variables  $u$  y  $\mathbf{x}$ , y  $u$  y  $a_1$ , tenemos que

$$\begin{aligned}E(I_{(-\infty, s] \times \{0\}}(\beta'_0 \mathbf{x} + u, a_1) | u=t) &= E(I_{(-\infty, s] \times \{0\}}(\beta'_0 \mathbf{x} + t, a_1) | u=t) \\ &= E(I_{(-\infty, s] \times \{0\}}(\beta'_0 \mathbf{x} + t, a_1)) \\ &= \mathbb{P}(\beta'_0 \mathbf{x} + t \leq s, a_1 = 0) = \mathbb{P}(\beta'_0 \mathbf{x} \leq s - t, a_1 = 0) \\ &= \mathbb{P}(\beta'_0 \mathbf{x} \leq s - t | a_1=0) \mathbb{P}(a_1 = 0) = R_0(s - t) \mathbb{P}(a_1 = 0).\end{aligned}$$

Por lo cual (3.11) resulta ser

$$\mathbb{P}(\beta'_0 \mathbf{x} + u \leq s, a_1 = 0) = E(R_0(s - U)) \mathbb{P}(a_1 = 0)$$

y entonces

$$F|_{a_1=0}(s) = E(R_0(s - U)) = (R_0 * K_0)(s).$$

Nótese que  $m/n \rightarrow \mathbb{P}(a_1 = 1)$  y en consecuencia  $1 - m/n \rightarrow \mathbb{P}(a_1 = 0)$ . Por el Lema 3.3.2 se observa que  $\mathbb{P}(\widehat{R}_n \rightarrow_w R_0) = 1$  y  $\mathbb{P}(\widehat{K}_n \rightarrow_w K_0) = 1$ . Además, dado que  $F$  se puede escribir como la siguiente combinación convexa de funciones de distribución

$$F(z) = F|_{a_1=1}(z) \mathbb{P}(a_1 = 1) + F|_{a_1=0}(z) (1 - \mathbb{P}(a_1 = 1)),$$

resulta que si  $z$  es un punto de continuidad de  $F$ , también es punto de continuidad de  $F|_{a_1=1}$  y de  $F|_{a_1=0}$ . En conclusión, utilizando el Lema 3.3.3, queda probado el ítem (a). Para (b) sólo se usa la continuidad débil de  $T$ .  $\square$

### 3.4. Estimación de los Parámetros del Modelo de Regresión

Sean  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  satisfaciendo el modelo lineal (3.3). Denominemos  $\{F_n^*\}$  a la sucesión de distribuciones empíricas asociadas a los pares  $(\mathbf{x}_i, y_i)$  observados, o sea  $i \in A$ . Por lo tanto,

$$F_n^* = \frac{1}{\sum_{i=1}^n a_i} \sum_{i=1}^n a_i \delta_{(\mathbf{x}_i, y_i)}.$$

Así podemos estimar a  $\beta_0$  por

$$\widehat{\beta}_n = T_{MM}(F_n^*).$$

Utilizaremos un MM-funcional usando  $\rho_0$  y  $\rho_1$  en la familia bicuadrada. Para poder afirmar que  $\widehat{\beta}_n$  resultará un estimador fuertemente consistente de  $\beta_0$ , nos remitiremos a los resultados de Fisher-consistencia y continuidad enunciados en la última sección del capítulo anterior.

**Lema 3.4.1.** *Consideremos  $(\mathbf{x}_i, y_i)$  con distribución  $F$  satisfaciendo el modelo (3.3), y sea  $F^* \sim (\mathbf{x}_i, y_i) |_{a_i=1}$ . Asumamos que  $\mathbb{P}_{F^*}(\mathbf{a}'\mathbf{x}_i = 0) < 1$  para todo  $\mathbf{a} \neq \mathbf{0}$ . Denotemos con  $f_0$  a la densidad de  $u_i$  y supongamos que  $f_0$  satisface la condición (B1) del Lema 1.5.5. Supongamos que  $S(F^*) > 0$ ,  $\delta < 1 - c(F^*)$  y  $E_{F^*} \left( \rho_1 \left( \frac{y_i - \beta'_0 \mathbf{x}_i}{S(F^*)} \right) \right) < 1 - c(F^*)$ . Entonces,*

$$\mathbb{P}(\widehat{\beta}_n \rightarrow \beta_0) = 1.$$

Las funciones  $\rho_0$  y  $\rho_1$  están determinadas por los valores  $k_0$  y  $k_1$ , en el libro de Maronna R.A., Martin R.D. y Yohai V.J. (2006) se recomienda tomar  $k_1 = 3.44$  puesto que efectúa un buen compromiso entre robustez y eficiencia. Más aún, corresponde a la eficiencia asintótica bajo normalidad del 85 % respecto del estimador de mínimos cuadrados. Grandes valores de  $k_1$  permiten alta eficiencia sacrificando robustez.

### 3.5. El funcional de Posición

En la Sección 3.7 realizamos un estudio de Monte Carlo comparando el estimador presentado en Sued y Yohai (2012) con nuestra nueva propuesta. Para ello, consideramos los siguientes funcionales de posición: la mediana, la media  $\alpha$ -podada, y los funcionales  $T_M$  y  $T_B$  definidos en el Capítulo 1. En el Capítulo 1 se estudian condiciones que garantizan la continuidad de estos funcionales.

Cabe mencionar que, como se demuestra en el trabajo de Fasano et. al. (2012), para que el funcional  $T_B$  esté unívocamente determinado en  $F$  basta con que la densidad sea fuertemente unimodal, relajando la condición de simetría requerida en el Corolario 1.5.6.

### 3.6. Punto de Ruptura

Sea

$$\mathbf{W} = \{(\mathbf{x}_1, y_1, a_1), \dots, (\mathbf{x}_n, y_n, a_n)\} \quad (3.12)$$

el conjunto de todas las observaciones e indicadores de pérdidas. Llamemos  $W_{ts}$  al conjunto de muestras obtenidas de  $\mathbf{W}$  reemplazando no más de  $t$  puntos de las observaciones donde las  $y_i$  son faltantes por outliers (solo puede ser en  $\mathbf{x}$ ) y a lo sumo  $s$  reemplazos correspondientes a los pares  $(\mathbf{x}_i, y_i)$  con  $y_i$  observada:

Luego si  $\mathbf{W}^* = \{(\mathbf{x}_1^*, y_1^*, a_1), \dots, (\mathbf{x}_n^*, y_n^*, a_n)\} \in W_{ts}$ , se tiene

$$\sum_{i \in A^C} \mathbb{I}_{\{\mathbf{x}_i^* \neq \mathbf{x}_i\}} \leq t$$

y

$$\sum_{i \in A} \mathbb{I}_{\{(\mathbf{x}_i^*, y_i^*) \neq (\mathbf{x}_i, y_i)\}} \leq s.$$

Dado un estimador  $\hat{\mu}_n$  de  $\mu_0$ , denotemos

$$M_{ts} = \sup_{\mathbf{W}^* \in W_{ts}} |\hat{\mu}_n(\mathbf{W}^*)|$$

y  $\kappa(t, s) = \max\left(\frac{t}{n-m}, \frac{s}{m}\right)$ . El *Punto de Ruptura Muestral* (PRM) de un estimador  $\hat{\mu}_n$  en  $\mathbf{W}$  es

$$\varepsilon^* = \min\{\kappa(t, s) : M_{ts} = \infty\}.$$

En forma coloquial lo que representa el PRM es la menor fracción de datos atípicos en el conjunto de las observaciones faltantes o en el de las observaciones vistas, necesaria para que el estimador no permanezca acotado.

Para obtener una cota inferior del PRM de  $\hat{\mu}_n$  el estimador de posición introducido en (3.7), precisamos definir el concepto de *Punto de Ruptura asintótico uniforme* de un funcional  $T$ , introducido en el trabajo de Sued y Yohai (2012).

**Definición 3.6.1.** Sea  $T$  un funcional. El Punto de Ruptura asintótico uniforme (PRU) de  $T$  es el supremo de todos los  $\varepsilon > 0$  satisfaciendo la siguiente propiedad:

para todo  $M > 0$  existe  $K > 0$  que depende de  $M$  y tal que si

$$\mathbb{P}_F(|y| \leq M) > 1 - \varepsilon \quad \Rightarrow \quad |T(F)| < K. \quad (3.13)$$

En el trabajo de Sued y Yohai (2012), los autores demuestran que, bajo determinadas condiciones, el PRU  $\varepsilon_U^*$  de un M-funcional de posición satisface:

- (i)  $\varepsilon_U^* \geq \min\{0, 5, \delta\}$  si  $\rho_1$  es convexa y  $\psi_1$  es acotada.
- (ii)  $\varepsilon_U^* \geq \min\{1 - \delta, \delta\}$  si  $\rho_1$  es acotada con  $\rho_1 \leq \rho_0$ .

También demuestran que el punto de ruptura uniforme de la media  $\alpha$ -podada es  $\alpha$ .

Una cota inferior del PRM de  $\hat{\mu}_n$  dado por (3.7) la brinda el siguiente teorema:

**Teorema 3.6.2.** Sea  $\mathbf{W}$  dado por (3.12) y sea  $\mathbf{Z} = \{(\mathbf{x}_i, y_i) : i \in A\}$ . Supongamos que  $\hat{\beta}_n = \tilde{\beta}_m(\mathbf{Z})$ , donde  $\tilde{\beta}_m$  es un estimador de regresión para muestras de tamaño  $m$ . Sea  $\varepsilon_1 > 0$  una cota inferior del PRM en  $\mathbf{Z}$  de  $\tilde{\beta}_m$  y sea  $\varepsilon_2 > 0$  una cota inferior del PRU de  $T$ . Entonces el PRM  $\varepsilon^*$  del estimador  $\hat{\mu}_n$  en  $\mathbf{W}$  cumple la siguiente desigualdad:

$$\varepsilon^* \geq \varepsilon_3 = \min \left( \varepsilon_1, 1 - \frac{\sqrt{\tilde{p}^2 + 4(1 - \tilde{p})(1 - \varepsilon_2)} - \tilde{p}}{2(1 - \tilde{p})} \right) \quad (3.14)$$

con  $\tilde{p} = \frac{m}{n}$ .

*Demostración.* Sean  $t < (n - m)\varepsilon_3$  y  $s < m\varepsilon_3$ . Definamos

$$Z_s = \left\{ \mathbf{Z}^* = \{(\mathbf{x}_i^*, y_i^*) : i \in A\} : \sum_{i \in A} \mathbb{I}_{\{(\mathbf{x}_i^*, y_i^*) \neq (\mathbf{x}_i, y_i)\}} \leq s \right\}.$$

Como  $\frac{s}{m} < \varepsilon_1$  por definición del PRM en  $\mathbf{Z}$ , podemos hallar  $M_1$  tal que

$$\sup_{\mathbf{Z}^* \in Z_s} \|\tilde{\beta}_m(\mathbf{Z}^*)\| \leq M_1,$$

y por lo tanto podemos encontrar  $M_2$  tal que

$$\sup_{j \in A^C} \sup_{\|\beta\| \leq M_1} |\beta' \mathbf{x}_j| \leq \frac{M_2}{2} \quad (3.15)$$

y

$$\sup_{i \in A} \sup_{\|\beta\| \leq M_1} |y_i - \beta' \mathbf{x}_i| \leq \frac{M_2}{2}. \quad (3.16)$$

Además, existe  $M_3 > 0$  tal que

$$\max_{i \in A} |y_i| \leq M_3 \quad (3.17)$$

Como consecuencia del modelo con datos faltantes, dado  $\mathbf{W}^* \in W_{ts}$  llamamos  $\hat{\beta}_n^* = \tilde{\beta}_m(\mathbf{Z}^*)$ , con  $\mathbf{Z}^* \in Z_s$ . Considerando  $B = \{j \in A^C : \mathbf{x}_j = \mathbf{x}_j^*\}$  y  $C = \{i \in A : (\mathbf{x}_i, y_i) = (\mathbf{x}_i^*, y_i^*)\}$ , tenemos que  $\mathcal{C}(B) > n - m - t > (n - m)(1 - \varepsilon_3)$  y que  $\mathcal{C}(C) > m - s > m - m\varepsilon_3 = m(1 - \varepsilon_3)$ .

Llamemos

$$\hat{y}_{ij}^* = \begin{cases} \hat{\beta}_n^*{}' \mathbf{x}_j^* + y_i^* - \hat{\beta}_n^*{}' \mathbf{x}_i^* & \text{si } i \in A, j \in A^C \\ y_i^* & \text{si } i, j \in A \end{cases} \quad (3.18)$$

Entonces tomando  $M = \max(M_2, M_3)$  y usando las desigualdades (3.15), (3.16) y (3.17) obtenemos que para  $i \in C$  y  $j \in B$  vale  $|\widehat{y}_{ij}^*| \leq M$ , y por lo tanto

$$\mathcal{C}(\{(i, j) : |\widehat{y}_{ij}^*| \leq M\}) > \mathcal{C}(C)\mathcal{C}(B) + \mathcal{C}(C)m > m(n-m)(1-\varepsilon_3)^2 + m^2(1-\varepsilon_3).$$

Luego usando que

$$1 - \varepsilon_3 \geq \frac{\sqrt{\widetilde{p}^2 + 4(1-\widetilde{p})(1-\varepsilon_2)} - \widetilde{p}}{2(1-\widetilde{p})} = \frac{\sqrt{m^2 + 4n(n-m)(1-\varepsilon_2)} - m}{2(n-m)},$$

resulta que

$$\mathcal{C}(\{(i, j) : |\widehat{y}_{ij}^*| \leq M\}) > nm(1-\varepsilon_2).$$

Como hay  $nm$  pares  $(i, j)$  subindexando  $\widehat{y}_{ij}^*$ , se cumple

$$\mathbb{P}_{\widehat{F}_n^*}(|y| \leq M) > (1-\varepsilon_2),$$

y como consecuencia de que  $\varepsilon_2$  es una cota inferior del PRU, se concluye que  $|T(\widehat{F}_n^*)| \leq K$  y entonces el PRM del estimador definido por (3.7) es mayor o igual a  $\varepsilon_3$ .  $\square$

Afortunadamente, este estimador brinda un PRM que es mayor o igual que el estimador que se presenta en Sued y Yohai (2012). La diferencia entre los estimadores radica en que el estimador que se propone en esta tesis utiliza los  $y_i$  observados y sólo se genera una “muestra ficticia” para los  $y_i$  faltantes. En cambio, en Sued y Yohai (2012) se genera la muestra ficticia tanto para los  $y_i$  observados como para los no observados.

Si bien Sued y Yohai (2012) dan otra definición de punto de ruptura para muestras con datos faltantes, podemos verificar que con nuestra definición obtenemos la misma cota obtenida por ellos. Es más, la cota inferior para el PRM de la propuesta hecha por Sued y Yohai (2012) es

$$\varepsilon_3^{SY} = \min(\varepsilon_1, 1 - \sqrt{1 - \varepsilon_2}). \quad (3.19)$$

**Lema 3.6.3.**  $\varepsilon_3^{SY} \leq \varepsilon_3$ .

*Demostración.* Para todo  $\widetilde{p} \geq 0$  y  $0 \leq \varepsilon_2 \leq 1$  vale que

$$\widetilde{p}(\sqrt{1-\varepsilon_2} - (1-\varepsilon_2)) \geq 0,$$

que es equivalente a

$$(2(1-\widetilde{p})\sqrt{1-\varepsilon_2} + \widetilde{p})^2 \geq \widetilde{p}^2 + 4(1-\widetilde{p})(1-\varepsilon_2).$$

Luego aplicando la raíz cuadrada y operaciones de resta y producto, asumiendo que  $\widetilde{p} \neq 1$ , obtenemos

$$\sqrt{1-\varepsilon_2} \geq \frac{\sqrt{\widetilde{p}^2 + 4(1-\widetilde{p})(1-\varepsilon_2)} - \widetilde{p}}{2(1-\widetilde{p})},$$

que finalmente es equivalente a

$$1 - \sqrt{1-\varepsilon_2} \leq 1 - \frac{\sqrt{\widetilde{p}^2 + 4(1-\widetilde{p})(1-\varepsilon_2)} - \widetilde{p}}{2(1-\widetilde{p})},$$

lo que permite deducir la desigualdad deseada.  $\square$

A continuación se presentan los cuadros correspondientes a las cotas inferiores del punto de ruptura de  $\widehat{\mu}_n = T(\widehat{F}_n)$ , siendo  $\varepsilon_2$  el PRU del funcional débilmente continuo  $T$ , para las distintas propuestas de estimación.

En el cuadro 3.1 se puede observar una cota inferior del punto de ruptura de los estimadores propuestos en este trabajo para distintas proporciones  $\widetilde{p}$  y distintos  $\varepsilon_2$ , y en el cuadro 3.2 figuran cotas inferiores del punto de ruptura de los estimadores del trabajo de Sued y Yohai (2012) para distintos valores de  $\varepsilon_2$ , asumiendo en ambos casos que  $\varepsilon_1 = 0.5$ .

Cuadro 3.1: Cota inferior del Punto de Ruptura de la propuesta de estimación de este trabajo

$\tilde{p} \setminus \varepsilon_2$	0.1	0.2	0.3	0.4	0.5
0.05	0.053	0.108	0.168	0.231	0.300
0.1	0.054	0.111	0.172	0.237	0.308
0.25	0.059	0.121	0.186	0.257	0.333
0.5	0.068	0.140	0.215	0.296	0.382
0.75	0.081	0.165	0.253	0.344	0.438
0.9	0.092	0.185	0.280	0.377	0.475
0.95	0.096	0.192	0.290	0.388	0.488

Cuadro 3.2: Cota inferior del Punto de Ruptura de la propuesta de estimación de Sued y Yohai

$\varepsilon_2$	0,1	0,2	0,3	0,4	0,5
$\varepsilon_3^{SY}$	0.051	0.106	0.163	0.225	0.293

### 3.7. Resultados de un estudio de simulación por el Método de Monte Carlo

Para comparar el comportamiento de la propuesta de estimación de este trabajo y la que se define en Sued y Yohai (2012) se realizó un estudio de Monte Carlo considerando dos modelos de regresión lineal:

**Modelo 1** La variable  $y$  es generada por  $y = 5x_1 + x_2 + x_3 + 4v + 9$  donde  $x_1, x_2, x_3$  y  $v$  son v.a.i.i.d. con distribución  $\mathcal{N}(0, 1)$ .

**Modelo 2** La variable  $y$  es generada por  $y = 5x_1 + x_2 + x_3 + 4v + 4$  donde  $x_1, x_2$  y  $x_3$  son v.a.i.i.d  $\chi_1^2$  y  $v$  se distribuye según la distribución  $\mathcal{N}(0, 1)$  y es independiente de las  $x_i$ 's.

Podemos observar que en el Modelo 1 tanto  $y$  como el error de regresión  $v$  poseen distribuciones simétricas, mientras que en el Modelo 2 la variable  $y$  es asimétrica y el error de regresión se mantiene simétrico.

Para ambos modelos, el mecanismo que genera la variable  $a$ , que indica cuando el  $y$  es observado, está dado por

$$\ln \left( \frac{\mathbb{P}(a = 1 | x_1, x_2, x_3)}{1 - \mathbb{P}(a = 1 | x_1, x_2, x_3)} \right) = 0.15(x_1 + x_2 + x_3).$$

Despejamos  $\mathbb{P}(a = 1 | x_1, x_2, x_3)$  de la identidad anterior:

$$\mathbb{P}(a = 1 | x_1, x_2, x_3) = \frac{e^{0.15(x_1+x_2+x_3)}}{1 + e^{0.15(x_1+x_2+x_3)}},$$

así

$$\mathbb{P}(a = 1) = E(I_{\{a=1\}}) = E(E(I_{\{a=1\}} | (x_1, x_2, x_3))) = E \left( \frac{e^{0.15(x_1+x_2+x_3)}}{1 + e^{0.15(x_1+x_2+x_3)}} \right).$$

Numéricamente se obtuvo que  $\mathbb{P}(a = 1) = 0.5$  para el modelo 1 y  $\mathbb{P}(a = 1) = 0.606$  para el modelo 2.

Se estudiaron 62 casos para cada modelo. El primero corresponde al modelo central sin contaminación de outliers. Los otros 61 casos reemplazan al 10% de las observaciones  $(\mathbf{x}, y)$ 's por el mismo valor  $(\mathbf{x}^*, y^*)$  con  $\mathbf{x}^* = (2, 0, 0)$  e  $y^*$  variando en una grilla de 61 valores equiespaciados en el intervalo  $[-20, 40]$ . Usando muestras de tamaño 100, se realizaron 1000 replicaciones para cada uno de los 62 escenarios.

Consideramos 5 funcionales de posición: la media, la mediana, el M-funcional de posición con  $\rho_1$  en la familia bicuadrada con  $k = 3.44$  (BI), el M-funcional de posición con  $\rho_1$  en la familia de Huber con  $k = 1.37$  (HU), y la media 0.1-podada (MP10). Para ambos M-funcionales de posición, utilizamos el funcional de dispersión  $S(F)$  definido por (2.18) con  $\rho_0$  en la familia de funciones bicuadradas con  $k = 1.57$  y  $\delta = 0.5$ . Dado que el modelo 1 ofrece una distribución de  $y$  simétrica, los valores de los 5 funcionales son el mismo y coinciden con el centro de simetría. Por el contrario, la distribución de  $y$  en el modelo 2 es asimétrica y entonces los valores aproximados de los funcionales se calcularon por simulación usando una muestra de tamaño 100000. El cuadro 3.3 muestra los valores exactos de los 5 funcionales de posición para el modelo 1 y valores aproximados para el modelo 2 que fueron extraídos del trabajo de Sued y Yohai (2012).

Cuadro 3.3: Valores de los Funcionales de Posición

Modelo	Media	Mediana	BI	HU	MP10
1	9	9	9	9	9
2	11	9.47	9.29	10.02	10.05

Se utilizó  $\hat{\beta}_n$  el estimador de Mínimos Cuadrados para estimar el parámetro de la regresión cuando consideramos la media como funcional de posición, recuperando así la propuesta clásica descrita al inicio del presente capítulo. No obstante, para los 4 funcionales de posición robustos, se utilizó, justamente, un estimador robusto:  $\hat{\beta}_n$  el MM-estimador con  $\rho$  funciones en la familia bicuadrada. Más específicamente, se eligió  $k = 1.57$  y  $k = 3.44$  para  $\rho_0$  y  $\rho_1$ , respectivamente, y  $\delta = 0.5$

Para cada uno de los 4 funcionales de posición robustos, consideramos dos estimadores: el propuesto en Sued y Yohai (2012), que está dado por  $T(\hat{H}_n^*)$ , con  $\hat{H}_n^*$  definido en (3.4), y el propuesto en (3.7). Los resultados de las simulaciones para los modelos 1 y 2 se presentan en los cuadros 3.4 y 3.5, respectivamente. En los mismos, el subíndice 1 denotará a los estimadores correspondientes de Sued y Yohai, y el subíndice 2 a los de este trabajo. Las abreviaturas usadas en los cuadros son como se describen a continuación:

**ECM<sub>0</sub>** el error cuadrático medio cuando no se tiene contaminación de outliers.

**Sesgo<sub>0</sub>** el sesgo del estimador considerado sin outliers.

**ECM<sub>max</sub>** el máximo error cuadrático medio con outliers, para los 61 valores de  $y^*$ .

*y<sub>max</sub>* el valor de  $y^*$  para el cual **ECM<sub>max</sub>** se alcanza.

La primer línea del cuadro 3.4 muestra que en ausencia de outliers el estimador clásico es el más eficiente para el modelo 1, como se espera de variables normales. Bajo el modelo 2, los estimadores robustos resultan ser más eficientes que el estimador de la media, como se puede observar en el cuadro 3.5. Esto se debe a que la distribución de  $y$  tiene colas pesadas. Los resultados que se observan en los casos donde hay contaminación con outliers muestran que, para ambos modelos, el estimador que se propone en este trabajo produce un **ECM<sub>max</sub>** menor que el correspondiente a los otros estimadores. Los estimadores que resultan ser más robustos, según la propuesta introducida

Cuadro 3.4: Resultados del método de Monte Carlo para el Modelo 1

Estim.	Media	Mediana <sub>1</sub>	BI <sub>1</sub>	HU <sub>1</sub>	MP10 <sub>1</sub>	Mediana <sub>2</sub>	BI <sub>2</sub>	HU <sub>2</sub>	MP10 <sub>2</sub>
ECM <sub>0</sub>	0.60	0.65	0.65	0.62	0.62	0.77	0.67	0.63	0.63
Sesgo <sub>0</sub> × 10 <sup>3</sup>	22.8	20.7	19.0	21.7	21.6	6.32	14.0	19.3	20.1
ECM <sub>max</sub>	9.23	4.63	5.02	6.56	6.68	2.98	2.65	4.26	4.62
<i>y</i> max	40	35	33	40	40	36	25	40	40

Cuadro 3.5: Resultados del método de Monte Carlo para el Modelo 2

Estim.	Media	Mediana <sub>1</sub>	BI <sub>1</sub>	HU <sub>1</sub>	MP10 <sub>1</sub>	Mediana <sub>2</sub>	BI <sub>2</sub>	HU <sub>2</sub>	MP10 <sub>2</sub>
ECM <sub>0</sub>	0.81	0.57	0.60	0.62	0.63	0.68	0.63	0.62	0.64
Sesgo <sub>0</sub> × 10 <sup>3</sup>	49.4	20.0	30.1	46.2	61.8	31.9	41.4	50.6	72.8
ECM <sub>max</sub>	11.93	3.30	3.97	4.47	6.31	2.49	3.40	3.75	6.03
<i>y</i> max	-20	25	29	37	40	25	25	37	40

en esta tesis, para el modelo 1 son BI<sub>2</sub>, Mediana<sub>2</sub> y HU<sub>2</sub>, y para el modelo 2 son Mediana<sub>2</sub>, BI<sub>2</sub> y HU<sub>2</sub>. Notemos que para las construcciones que estamos realizando, la Media 0.1-podada, no resulta robusta. La razón es que, siendo 0.5 el punto de ruptura del estimador de regresión, de acuerdo con (3.14) el punto de ruptura del estimador MP10<sub>2</sub> es  $< 0, 1$ , y usando (3.19) el punto de ruptura del estimador MP10<sub>1</sub> es 0.0523.

En las figuras 3.1 y 3.2 se comparan los comportamientos de las 2 distintas propuestas de estimación para cada funcional de posición robusto en un entorno de contaminación de datos atípicos. Para tal propósito se graficó el error cuadrático medio (ECM) en función del valor outlier  $y^*$ . Estos gráficos muestran que para el M-funcional de la familia de funciones bicuadradas (BI), la propuesta aquí presentada parece ser preferible respecto de la de Sued y Yohai (2012) para ambos modelos. También se puede observar que a medida que crece el outlier, el ECM de la media 0.1-podada también.

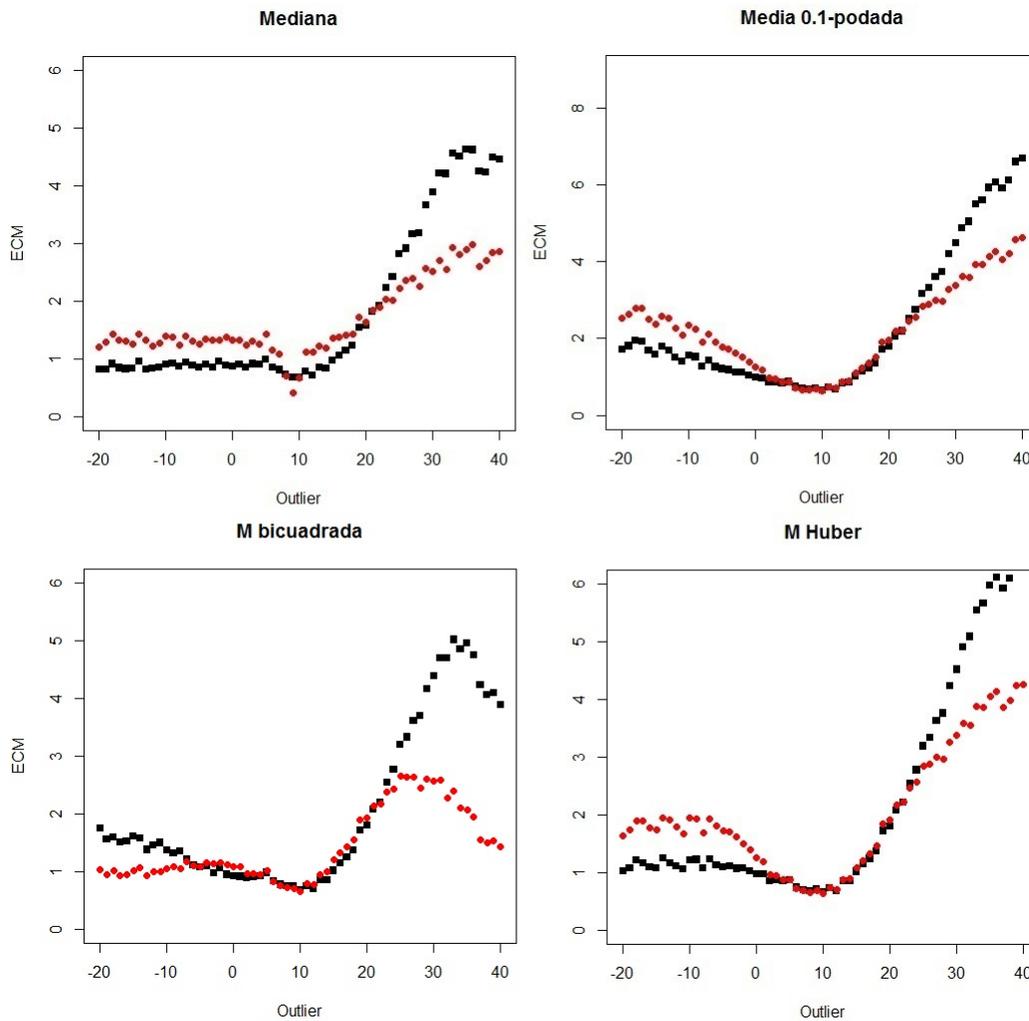


Figura 3.1: Comparación de estimadores bajo contaminación de outliers para el Modelo 1. El color negro identifica a la metodología de estimación introducida en Sued y Yohai (2012), y el rojo a la propuesta en este trabajo.

### 3.7. RESULTADOS DE UN ESTUDIO DE SIMULACIÓN POR EL MÉTODO DE MONTE CARLO41

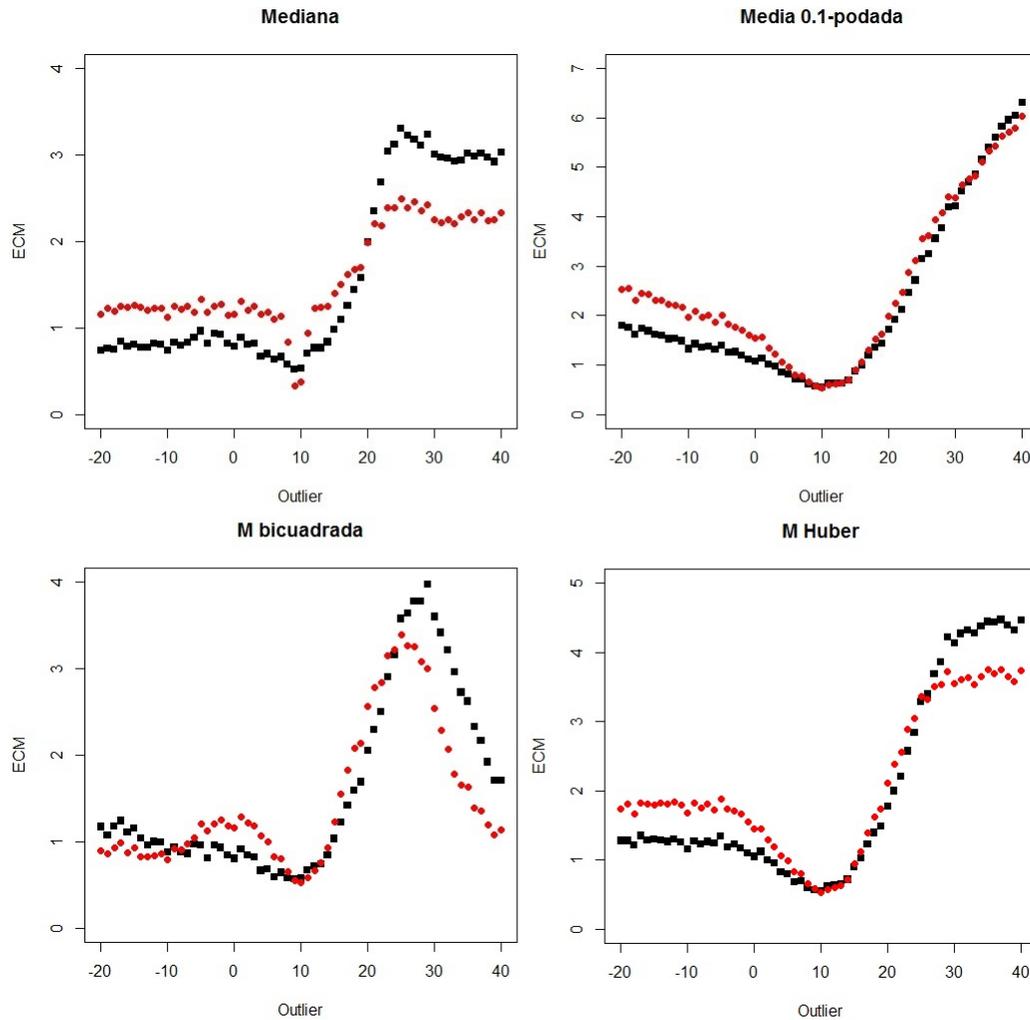


Figura 3.2: Comparación de estimadores bajo contaminación de outliers para el Modelo 2. El color negro identifica a la metodología de estimación introducida en Sued y Yohai (2012), y el rojo a la propuesta en este trabajo.



# Capítulo 4

## Apéndice

Presentamos en este capítulo las demostraciones pendientes.

### **Demostración del Lema 1.5.5**

Sea  $\lambda > 0$ . Como  $u$  es una v.a. simétrica,  $\rho(u - \lambda)$  tiene la misma distribución que  $\rho(-u - \lambda) = \rho(u + \lambda)$ . Por lo tanto vale que

$$g(\lambda) = g(-\lambda), \quad (4.1)$$

y obtenemos (i).

Para probar (ii) tomemos  $\lambda \geq 0$  y consideremos la función de distribución de  $v = |u - \lambda|$

$$R_\lambda(v) = P(|u - \lambda| \leq v) = P(\lambda - v \leq u \leq \lambda + v).$$

Luego para  $v \geq 0$

$$R_\lambda(v) = \int_{\lambda-v}^{\lambda+v} f_0(u) du$$

y entonces como  $f_0(u)$  es monótona no creciente en  $|u|$ ,

$$\frac{\partial R_\lambda(v)}{\partial \lambda} = f_0(\lambda + v) - f_0(\lambda - v) \leq 0, \quad (4.2)$$

y por consiguiente,

$$R_{\lambda_1}(v) \geq R_{\lambda_2}(v), \text{ para todo } v \geq 0, \text{ para todo } 0 \leq \lambda_1 \leq \lambda_2. \quad (4.3)$$

Sea  $\lambda_0$  tal que para  $0 \leq u \leq 2\lambda_0$ , (a)  $f_0(\cdot)$  es estrictamente decreciente en  $[0, \lambda_0]$  y (b)  $\psi(u) > 0$  en  $[0, \lambda_0]$ . Si  $0 \leq \lambda \leq \lambda_0, v \leq \lambda_0$ , tenemos que

$$\frac{\partial R_\lambda(v)}{\partial \lambda} < 0,$$

de donde concluimos que

$$R_\lambda(v) < R_0(v), \text{ si } 0 < \lambda \leq \lambda_0, 0 < v \leq \lambda_0. \quad (4.4)$$

Llamando  $r_\lambda(u) = R'_\lambda(u)$  e integrando por partes, resulta

$$g(\lambda) = E_{F_0}(\rho(u - \lambda)) = \int_0^{+\infty} \rho(v) r_\lambda(v) dv = R_\lambda(v) \rho(v) \Big|_0^{v \rightarrow +\infty} - \int_0^{+\infty} \rho'(v) R_\lambda(v) dv = a - \int_0^{+\infty} \psi(v) R_\lambda(v) dv$$

y si  $0 \leq \lambda_1 \leq \lambda_2$  usando (4.3)

$$g(\lambda_2) - g(\lambda_1) = \int_0^{+\infty} \psi(v)(R_{\lambda_1}(v) - R_{\lambda_2}(v))dv \geq 0 \quad (4.5)$$

queda probado (ii). También tenemos para  $0 < \lambda \leq \lambda_0$

$$g(\lambda) - g(0) = \int_0^{\lambda_0} \psi(v)(R_0(v) - R_\lambda(v))dv + \int_{\lambda_0}^{+\infty} \psi(v)(R_0(v) - R_\lambda(v))dv$$

y como para  $0 < v \leq \lambda_0$ , por (4.4) y por  $\psi(v) > 0$  se tiene

$$\int_0^{\lambda_0} \psi(v)(R_0(v) - R_\lambda(v))dv > 0$$

y

$$\int_{\lambda_0}^{+\infty} \psi(v)(R_0(v) - R_\lambda(v))dv \geq 0.$$

Luego

$$g(\lambda) > g(0) \text{ si } 0 < \lambda \leq \lambda_0$$

y recordando (4.5), resulta que

$$g(\lambda) > g(0) \text{ si } 0 < \lambda < \infty.$$

Finalmente utilizando (4.1) llegamos a que

$$g(\lambda) > g(0) \text{ si } \lambda \neq 0$$

o sea (iii).  $\square$

### **Demostración del Lema 1.5.7**

Como  $u \sim -u$  y  $\psi$  es una función impar, entonces  $\psi(u - \lambda)$  tiene la misma distribución que  $\psi(-u - \lambda) = -\psi(u + \lambda)$ . Por lo tanto vale que

$$E_{F_0}(\psi(u - \lambda)) = -E_{F_0}(\psi(u + \lambda)),$$

o sea que

$$g(-\lambda) = -g(\lambda).$$

Así queda probado (i).

Sean  $\lambda_1 < \lambda_2$ .

$$g(\lambda_1) - g(\lambda_2) = E_{F_0}(\psi(u - \lambda_1) - \psi(u - \lambda_2)). \quad (4.6)$$

$\psi$  es una función monótona no decreciente y  $u - \lambda_1 > u - \lambda_2$ , por lo que resulta

$$\psi(u - \lambda_1) - \psi(u - \lambda_2) \geq 0$$

y entonces

$$g(\lambda_1) - g(\lambda_2) \geq 0,$$

quedando demostrado (ii).

Por último, para probar el ítem (iii), usando las hipótesis (1) y (2), existe  $a > 0$  tal que si  $|u| \leq a$ ,  $\psi$  es estrictamente creciente y  $f_0(u) > 0$ . Tomemos  $0 < \lambda \leq a$ . Apliquemos (4.6) a  $\lambda_1 = 0$  y  $\lambda_2 = \lambda$ . Al ser  $g$  una función impar,  $g(0) = 0$ . Por lo cual

$$-g(\lambda) = E_{F_0}(\psi(u) - \psi(u - \lambda)).$$

Luego

$$g(\lambda) = E_{F_0}(\psi(u - \lambda) - \psi(u)) = \int_{-\infty}^{+\infty} (\psi(u - \lambda) - \psi(u))f_0(u) du. \quad (4.7)$$

Sea  $R(u, \lambda) = (\psi(u - \lambda) - \psi(u))f_0(u)$ . Es claro que

$$R(u, \lambda) \leq 0 \quad \text{para todo } u. \quad (4.8)$$

Por otro lado, si  $\lambda > 0$  y  $0 \leq u \leq \lambda \leq a$  tenemos que  $|u - \lambda| \leq a$ , y entonces

$$R(u, \lambda) < 0. \quad (4.9)$$

Podemos reescribir (4.7) de la siguiente manera

$$g(\lambda) = \int_{-\infty}^0 R(u, \lambda) du + \int_0^{\lambda} R(u, \lambda) du + \int_{\lambda}^{+\infty} R(u, \lambda) du = I_1 + I_2 + I_3.$$

$I_j \leq 0, j = 1, 2, 3$  por (4.8). Además, usando (4.9),  $I_2 < 0$ . Por lo tanto, si  $0 < \lambda \leq a$ ,  $g(\lambda) < 0$ , y si  $\lambda > a$ , por el ítem (ii)  $g(\lambda) \leq g(a) < 0$ .  $\square$

**Demostración de la Proposición 1.6.2** Para una  $G$  dada, sea

$$F_\varepsilon = (1 - \varepsilon)F + \varepsilon G \quad (4.10)$$

y sea  $\mu_\varepsilon = \widehat{\mu}(F_\varepsilon)$ . Por definición,  $\mu_\varepsilon$  es solución de  $E_{F_\varepsilon}(\psi(x - \mu_\varepsilon)) = 0$ . Entonces aplicando el operador esperanza en (4.10) tenemos

$$(1 - \varepsilon)E_F(\psi(x - \mu_\varepsilon)) + \varepsilon E_G(\psi(x - \mu_\varepsilon)) = E_{F_\varepsilon}(\psi(x - \mu_\varepsilon)) = 0.$$

Sea  $\varepsilon < \varepsilon^*$ , el punto de ruptura asintótico del M-funcional. Luego, existe  $c > 0$  tal que  $|\mu_\varepsilon| \leq c$  para todo  $G$ . Tomando  $G = \delta_{x_0}$ , la función de distribución del punto masa  $x_0$  tenemos que  $(1 - \varepsilon)E_F(\psi(x - \mu_\varepsilon)) + \varepsilon\psi(x_0 - \mu_\varepsilon) = 0$ . Sabiendo que

$$\lim_{x_0 \rightarrow +\infty} \psi(x_0 - \mu_\varepsilon) = k_2$$

y asumiendo que  $\psi$  es no decreciente,  $\psi \geq -k_1$ , entonces

$$0 = (1 - \varepsilon)E_F(\psi(x - \mu_\varepsilon)) + \varepsilon\psi(x_0 - \mu_\varepsilon) \geq (1 - \varepsilon)(-k_1) + \varepsilon\psi(x_0 - \mu_\varepsilon).$$

Luego  $0 \geq -k_1(1 - \varepsilon) + \varepsilon k_2$ , lo que implica que  $\varepsilon \leq \frac{k_1}{k_1 + k_2} := \varepsilon_1^*$ . Tomando  $x_0 \rightarrow -\infty$  se obtiene análogamente  $\varepsilon \leq \frac{k_2}{k_1 + k_2} := \varepsilon_2^*$ . Por lo cual

$$\varepsilon^* \leq \frac{\min\{k_1, k_2\}}{k_1 + k_2}.$$

Ahora queremos probar la desigualdad opuesta. Sea  $\varepsilon > \varepsilon^*$ . Por definición de punto de ruptura, sabemos que existe  $\{G_n\}_{n \in \mathbb{N}}$  tal que  $\mu_{\varepsilon, n} = \widehat{\mu}((1 - \varepsilon)F + \varepsilon G_n)$  no está acotada. Sea, entonces  $\{\mu_{\varepsilon, n_j}\}$  una subsucesión que, por el momento, supondremos converge a  $+\infty$ . Por lo tanto, para cada  $x$ ,  $x - \mu_{\varepsilon, n_j} \rightarrow -\infty$  y como  $\psi \leq k_2$ , nos queda que

$$0 \leq (1 - \varepsilon) \lim_{j \rightarrow +\infty} E_F(\psi(x - \mu_{\varepsilon, n_j})) + \varepsilon k_2.$$

Usando el Teorema de Convergencia Monótona:

$$\lim_{j \rightarrow +\infty} E_F(\psi(x - \mu_{\varepsilon, n_j})) = E_F(\lim_{j \rightarrow +\infty} \psi((x - \mu_{\varepsilon, n_j}))) = \psi(-\infty) = -k_1$$

obtenemos  $0 \leq -k_1(1 - \varepsilon) + \varepsilon k_2$  o sea que  $\varepsilon \geq \varepsilon_1^*$ . Si  $\mu_{\varepsilon, n_j} \rightarrow -\infty$  se tiene  $\varepsilon \geq \varepsilon_2^*$ .  $\square$

### Demostración del Lema 3.3.2

La convergencia débil puede ser caracterizada por la siguiente propiedad:

$$H_n \rightarrow_w H \Leftrightarrow \lim_{n \rightarrow \infty} \int f dH_n = \int f dH \text{ para toda } f \in \mathcal{C}_0(\mathbb{R})$$

donde  $\mathcal{C}_0(\mathbb{R})$  es el conjunto de funciones continuas y acotadas. Consideremos

$$\widehat{H}_n = \frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n c_i \delta_{h(\mathbf{z}_i, \beta_0)}.$$

Veamos que  $\widehat{H}_n$  converge uniformemente a  $\widetilde{H}_0$  c.t.p. Para ello, utilizaremos el Lema 3.3.4, que garantiza que

$$\sup_{\mathbf{z}} \left| \frac{1}{n} \sum_{i=1}^n c_i I_{\{h(\mathbf{z}_i, \beta_0) \leq \mathbf{z}\}} - E(c_1 I_{\{h(\mathbf{z}_1, \beta_0) \leq \mathbf{z}\}}) \right| \rightarrow 0 \quad \text{c.t.p.}$$

Luego, notando que  $\sum_{i=1}^n c_i/n$  converge a  $P(c_1 = 1)$  c.t.p., podemos concluir la convergencia uniforme deseada. Resta encontrar un conjunto de probabilidad 1 que verifique

$$\lim_{n \rightarrow +\infty} \left| \int f d\widetilde{H}_n - \int f d\widehat{H}_n \right| = 0 \quad \text{para toda } f \in \mathcal{C}_0(\mathbb{R}).$$

Observemos que

$$\int f d\widetilde{H}_n = \frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n c_i f(h(\mathbf{z}_i, \widehat{\beta}_n)) \text{ y } \int f d\widehat{H}_n = \frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n c_i f(h(\mathbf{z}_i, \beta_0)),$$

por lo cual usando que  $c_i \leq 1$

$$\left| \int f d\widetilde{H}_n - \int f d\widehat{H}_n \right| \leq \frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n \left| f(h(\mathbf{z}_i, \widehat{\beta}_n)) - f(h(\mathbf{z}_i, \beta_0)) \right| I_{\{\|\mathbf{z}_i\| \leq K\}} + 2\|f\|_{\infty} \frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n I_{\{\|\mathbf{z}_i\| > K\}}.$$

Llamando

$$C_K = \{(\mathbf{z}, \beta) : \|\mathbf{z}\| \leq K, \|\beta - \beta_0\| \leq 1\},$$

conjunto compacto, tenemos que  $f \circ h : C_K \rightarrow \mathbb{R}$  es uniformemente continua, o sea dado  $\varepsilon > 0$ , existe  $\delta > 0$  tal que si  $(\mathbf{z}_i, \beta_i) \in C_K$  y  $\|(\mathbf{z}_1, \beta_1) - (\mathbf{z}_2, \beta_2)\| < \delta$ , entonces  $|f(h(\mathbf{z}_1, \beta_1)) -$

$f(h(\mathbf{z}_2, \boldsymbol{\beta}_2))| < \varepsilon$ . Por la consistencia fuerte de  $\widehat{\boldsymbol{\beta}}_n$  y sabiendo por la Ley de los Grandes Números que

$$\frac{n}{\sum_{i=1}^n c_i} \rightarrow \frac{1}{E(c_1)},$$

con probabilidad 1 existe  $n_0$  entero aleatorio tal que  $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| < \delta$  para toda  $n \geq n_0$ . Luego

$$\left| \int f d\widehat{H}_n - \int f d\widetilde{H}_n \right| \leq \frac{n}{\sum_{i=1}^n c_i} \varepsilon + 2\|f\|_\infty \frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n I_{\{\|\mathbf{z}_i\| > K\}} \quad \text{para toda } n \geq n_0.$$

Usando otra vez la Ley de los Grandes Números,

$$\frac{1}{\sum_{i=1}^n c_i} \sum_{i=1}^n I_{\{\|\mathbf{z}_i\| > K\}} \rightarrow \frac{1}{E(c_1)} \mathbb{P}(\|\mathbf{z}_i\| > K) \text{ c.t.p. para todo } K.$$

Finalmente, con probabilidad 1

$$\lim_{n \rightarrow \infty} \left| \int f d\widehat{H}_n - \int f d\widetilde{H}_n \right| \leq \frac{\varepsilon}{E(c_1)} + \frac{2\|f\|_\infty}{E(c_1)} \mathbb{P}(\|\mathbf{z}_i\| > K) \quad \text{para todo } K \text{ y para todo } \varepsilon > 0.$$

El resultado buscado sale de hacer tender  $\varepsilon \rightarrow 0$  y  $K \rightarrow +\infty$ .  $\square$



# Bibliografía

- [1] Bickel P.J. y Doksum K.A. (2001) *Mathematical Statistics: Basic Ideas and Selected Topics*. Vol. I, 2nd edition, London, Prentice Hall.
- [2] Donoho D.L. y Huber P.J. (1983) *The Notation of Breakdown Point*. A Festschrift for E. L. Lehmann, Bickel P.J., Doksum, K.A., & Hodges, J.L., editors. Wadsworth, Belmont, CA. 157-184.
- [3] Fasano M.V., Maronna R.A., Sued R.M. y Yohai V.J. (2012) *Continuity and differentiability of regression M functionals*. Bernoulli, vol. 18, n° 4 1289-1309.
- [4] Huber P.J. (1964) *Robust estimation of a location parameter*. The Annals of Mathematical Statistics, vol. 35, 73-101.
- [5] Hössjer O. (1992) *On the optimality of S-estimators*. Statistics and Probability Letters, vol. 14, 413-419.
- [6] Lehmann E.L. y Casella G. (1998) *Theory of Point Estimation*. 2nd edition, Springer Texts in Statistics, New York, Springer.
- [7] Maronna R.A., Martin R.D. y Yohai V.J. (2006) *Robust Statistics: Theory and Methods*. John Wiley & Sons, Ltd.
- [8] Robins J., Sued, M., Lei-Gomez, Q., y Rotnitzky, A. (2007). *Comment: Performance of double-robust estimators when inverse probability weights are highly variable*. Statistical Science, vol. 22, 544-559.
- [9] Rousseeuw P.J. y Yohai V.J. (1984) *Robust Regression by means of S-estimators*. Robust and Nonlinear Time Series, vol. 26, 256-272. Franke, J., Hardle, W. & Martin, R.D. editors. Lecture Notes in Statistics, Springer, New York.
- [10] Sued R.M. y Yohai V.J. *Robust location estimation with missing data*. The Canadian Journal of Statistics. Aceptado.
- [11] Yohai V.J. (1987) *High Breakdown-Point and High Efficiency Robust Estimates for Regression*. The Annals of Statistics, vol. 15, 642-656.