



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

Tesis de Licenciatura

**Estimadores para el modelo lineal altamente robustos y
eficientes para muestras finitas**

Ezequiel Smucler

Director: Dr. Víctor Yohai

Fecha de Presentación: Diciembre de 2013

Índice general

1	Introducción	2
2	Regresión Lineal	4
2.1	Modelo Lineal	4
2.2	Mínimos Cuadrados	5
2.3	Funcionales de regresión	6
2.4	Robustez	7
2.4.1	Punto de ruptura finito	7
2.4.2	Máximo sesgo y punto de ruptura asintóticos	8
3	Estimadores robustos de regresión	10
3.1	M-estimadores	10
3.1.1	Propiedades asintóticas	13
3.2	S-estimadores	14
3.2.1	Máximo sesgo asintótico	15
3.2.2	Punto de ruptura finito	17
3.2.3	Propiedades asintóticas	17
3.3	τ -estimadores	18
3.4	MM-estimadores	21
3.4.1	Robustez	22
3.4.2	Propiedades asintóticas	23
3.5	REWLS	24
4	Propuesta	26
4.1	Robustez	28
4.2	Propiedades asintóticas	29
4.3	Simulación	30
4.3.1	Eficiencia	32
4.3.2	Robustez	33
5	Apéndice	35
	Bibliografía	45

1 Introducción

Todos los modelos estadísticos se basan en asumir hipótesis sobre los datos observados. Estas hipótesis ayudan a hacer a los modelos más simples y manejables desde un punto vista teórico y también computacional. Una de las simplificaciones más utilizada consiste en suponer que los datos observados siguen una distribución exactamente normal. La razón principal en la que se basa esta suposición es que la distribución normal da una buena aproximación para muchos conjuntos de datos y además permite obtener fórmulas explícitas para estadísticos óptimos (por ejemplo de menor varianza uniformemente) y conocer su distribución. Este enfoque, al que llamaremos clásico, ha dominado la Estadística de los últimos doscientos años.

En la práctica suele suceder que la mayoría de los datos observados sigue el modelo normal propuesto, pero algunas observaciones siguen un patrón diferente o directamente ningún patrón. Llamamos outliers o datos atípicos a las observaciones que no siguen el patrón de la mayoría de las observaciones. Intuitivamente se podría esperar que si la hipótesis de normalidad es cierta, aunque sea aproximadamente, los resultados de la teoría clásica también valgan aproximadamente. Lamentablemente, este no es el caso, la gran mayoría de los métodos clásicos pierden completamente sus propiedades de optimalidad ante pequeñas desviaciones del modelo asumido.

La Estadística Robusta busca dar métodos para el modelado estadístico y el análisis de datos que produzcan estimaciones confiables aún cuando los datos observados no siguen la distribución asumida exactamente, sino que aproximadamente. El objetivo de buena parte de los procedimientos robustos es dar estimaciones que sean "casi" tan buenas como la de los estimadores clásicos cuando la distribución de los datos observados es exactamente la asumida y que en la presencia de outliers estas estimaciones sean parecidas a las que se hubieran obtenido por métodos clásicos aplicados sólo a los datos no atípicos.

En este trabajo, estudiaremos métodos robustos para el modelo lineal, que describimos en la primer sección. Una de las primeras propuestas de estimación robusta para el modelo lineal fue dada por Huber en su trabajo de 1973 en el que presenta los M-estimadores. Estos estimadores son un intento por hacer más robusto el enfoque clásico, que es el de mínimos cuadrados. Lamentablemente, los M-estimadores como fueron propuestos por Huber son muy sensibles a observaciones atípicas en las variables regresoras y por lo tanto, su uso se encuentra limitado a modelos lineales con regresores no aleatorios.

En la década del 80, surgieron muchos estimadores robustos para el modelo lineal con regresores aleatorios como el LMS de Rousseeuw (1984), el LTS de Rousseeuw (1984) o los S-estimadores de Rousseeuw y Yohai (1984). Cuando decimos que estos estimadores son robustos nos referimos a que tienen un punto de ruptura alto. Definiremos con precisión este concepto en la primer sección de este trabajo, pero intuitivamente podemos pensar que el punto de ruptura nos indica la proporción de contaminación en una muestra que un estimador puede soportar sin verse completamente arruinado. En este contexto, la eficiencia de un estimador suele medirse como el cociente entre la varianza asintótica

del estimador de máxima verosimilitud y el estimador en cuestión. Todos los estimadores que mencionamos previamente son muy ineficientes bajo normalidad de los errores, es decir que asintóticamente tienen muchas más variabilidad en sus estimaciones que el estimador de mínimos cuadrados, que coincide con el estimador de máxima verosimilitud cuando la distribución de los errores del modelo es exactamente normal.

La primera propuesta de estimación robusta y altamente eficiente es la de Yohai (1987) que introduce los MM-estimadores. Estos estimadores combinan robustez con una eficiencia asintótica que podemos hacer arbitrariamente alta. En los años siguientes surgieron muchos otros estimadores que combinan estas propiedades, pero ninguno de ellos es completamente eficiente. Recién en Gervini y Yohai (2002) se proponen los estimadores REWLS que combinan robustez con una varianza asintótica igual a la de mínimos cuadrados bajo normalidad e independencia de los errores.

Sin embargo, si el tamaño de la muestra es pequeño la eficiencia asintótica de estos estimadores puede ser mucho mayor que la eficiencia para muestras finitas, definida como el cociente entre el error cuadrático medio del estimador de máxima verosimilitud y del estimador en cuestión. Es claro además que para las aplicaciones prácticas la eficiencia para muestras finitas es la medida relevante.

En este trabajo, motivados por los resultados de Bondell y Stefanski (2013), proponemos una familia de estimadores con un punto de ruptura alto y altamente eficientes para muestras finitas.

El presente trabajo está organizado de la siguiente manera:

En la primera sección, presentamos el modelo lineal, el estimador de mínimos cuadrados y definimos algunas de las medidas más utilizadas para medir la robustez de un estimador.

En la segunda sección, damos las propiedades más importantes de algunos de los estimadores robustos para el modelo lineal más conocidos: M-estimadores, S-estimadores, tau-estimadores, MM-estimadores y los estimadores REWLS.

En la tercera sección, presentamos nuestra propuesta, damos las definiciones pertinentes e incluimos un estudio de simulación de Montecarlo que compara el rendimiento de algunos de los estimadores descriptos.

Finalmente, en el apéndice damos las demostraciones de todos los resultados originales de este trabajo: Se prueba la consistencia y normalidad asintótica de los estimadores propuestos, se da su punto de ruptura finito y probamos una generalización de un Teorema de Martin, Yohai y Zamar (1989) que caracteriza el máximo sesgo asintótico de S-estimadores bajo ciertas hipótesis.

2 Regresión Lineal

2.1 Modelo Lineal

Consideremos un conjunto de n observaciones $(x_{i1}, \dots, x_{ip}, y_i)$, donde x_{i1}, \dots, x_{ip} son variables explicativas e y_i es la variable dependiente o de respuesta. Decimos que estos datos siguen un modelo lineal si

$$y_i = \sum_{j=1}^p x_{ij} \beta_{0,j} + u_i \text{ para } i = 1, \dots, n,$$

donde $\beta_{0,1}, \dots, \beta_{0,p}$ son parámetros desconocidos a estimar y las u_i son variables aleatorias (los errores). En un experimento diseñado las x_{ij} son no aleatorias, es decir que están determinadas antes del experimento. Cuando los datos son observacionales, las x_{ij} son variables aleatorias. Si llamamos \mathbf{x}_i y $\boldsymbol{\beta}_0$ a los vectores p -dimensionales con coordenadas $(x_{i1}, \dots, x_{ip})'$ y $(\beta_{0,1}, \dots, \beta_{0,p})'$, podemos expresar el modelo de manera más compacta como

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

Cuando el modelo contiene un término constante, la primer coordenada de cada \mathbf{x}_i es 1 y el modelo puede escribirse como

$$y_i = \beta_{0,0} + \mathbf{x}_i' \boldsymbol{\beta}_{0,1} + u_i \text{ para } i = 1, \dots, n \quad (2.2)$$

donde $\mathbf{x}_i = (x_{i1}, \dots, x_{i(p-1)})'$, $\boldsymbol{\beta}_{0,1} \in \mathbb{R}^{p-1}$ y

$$\mathbf{x}_i = \begin{bmatrix} 1 \\ \mathbf{x}_i \end{bmatrix}, \boldsymbol{\beta}_0 = \begin{bmatrix} \beta_{0,0} \\ \boldsymbol{\beta}_{0,1} \end{bmatrix}.$$

Llamamos a $\beta_{0,0}$ la ordenada al origen y a los elementos de $\boldsymbol{\beta}_{0,1}$ las pendientes. A lo largo de este trabajo, vamos a suponer que las x_{ij} son aleatorias y que nuestro modelo no contiene una ordenada al origen.

Llamemos ahora $\mathbf{X} \in \mathbb{R}^{n \times p}$ a la matriz con elementos x_{ij} y sean \mathbf{y} y \mathbf{u} los vectores con elementos y_i y u_i para $i = 1, \dots, n$. Entonces podemos expresar el modelo lineal como

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta}_0 + \mathbf{u}.$$

Vamos a decir que la matriz \mathbf{X} tiene rango completo si sus columnas son linealmente independientes.

Llamamos $\hat{y}_i(\boldsymbol{\beta})$ a los valores ajustados y $r_i(\boldsymbol{\beta})$ a los residuos que corresponden al vector $\boldsymbol{\beta}$, definidos por

$$\hat{y}_i(\boldsymbol{\beta}) = \mathbf{x}_i' \boldsymbol{\beta} \text{ y } r_i(\boldsymbol{\beta}) = y_i - \hat{y}_i(\boldsymbol{\beta}) \text{ para } i = 1, \dots, n.$$

Llamemos F_0 a la función de distribución de u y G_0 a la función de distribución de \mathbf{x} . Entonces la función de distribución de (\mathbf{x}, y) que llamamos H_0 está dada por

$$H_0(\mathbf{x}, y) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} F_0(y - \boldsymbol{\beta}_0' \mathbf{s}) dG_0(\mathbf{s}). \quad (2.3)$$

Para probar ciertos resultados necesitaremos además suponer que F_0 tiene una densidad f_0 que cumple que:

- (i) f_0 es una función par. (2.4)
- (ii) $f_0(|u|)$ es monótona decreciente.
- (iii) f_0 es estrictamente decreciente en un entorno del cero.
- (iv) f_0 es continua.

A lo largo de todo este trabajo vamos a trabajar bajo la hipótesis de que

$$\text{las } u_i \text{ son i.i.d e independientes de las } \mathbf{x}_i. \quad (2.5)$$

2.2 Mínimos Cuadrados

Supongamos que (\mathbf{x}'_i, y_i) para $i = 1, \dots, n$ siguen el modelo (2.1) con (2.5) y que $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 1$. Esta hipótesis se cumple por ejemplo, cuando \mathbf{x} tiene densidad. Esto implica que la probabilidad de que \mathbf{X} tenga rango completo tiene a uno cuando $n \rightarrow \infty$.

Definimos entonces el estimador de mínimos cuadrados de β_0 como

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

Derivando respecto de β obtenemos

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

que es equivalente a las llamadas ecuaciones normales

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

Si \mathbf{X} tiene rango completo, $\hat{\beta}_{MC}$ queda unívocamente determinado por

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

Por otro lado $\hat{\beta}_{MC}$ cumple

$$E(\hat{\beta}_{MC} | \mathbf{X}) = \beta \text{ y } \mathbf{Var}(\hat{\beta}_{MC} | \mathbf{X}) = \sigma_0^2 (\mathbf{X}' \mathbf{X})^{-1},$$

donde $\sigma_0^2 = \text{Var}(u)$.

Si \mathbf{X} no tiene rango completo, las ecuaciones normales tienen infinitas soluciones pero todas dan lugar a los mismos residuos y valores ajustados.

Si la distribución de las u_i es $N(0, \sigma_0^2)$, tenemos que la distribución de $\hat{\beta}_{MC}$ condicional a \mathbf{X} es normal multivariada y que $\hat{\beta}_{MC}$ es el estimador de Máxima Verosimilitud de β_0 . Además la matriz de covarianza de $\hat{\beta}_{MC}$ alcanza la cota de Rao-Cramér multivariada.

Si las u_i no son normales, pero existe $\mathbf{V}_{\mathbf{x}} = E_{G_0}(\mathbf{x}\mathbf{x}')$ y es inversible, como veremos en la Sección 3.1.1 se puede probar que

$$\sqrt{n}(\hat{\beta}_{MC} - \beta_0) \rightarrow^d N_p(\mathbf{0}, \sigma_0^2 \mathbf{V}_{\mathbf{x}}^{-1}), \quad (2.6)$$

donde $\sigma_0^2 = \text{Var}(u)$.

Sin embargo, es bien sabido que $\hat{\beta}_{MC}$ no es un estimador robusto, es muy sensible a pequeñas desviaciones del modelo de regresión con errores normales. Además, una pequeña proporción de observaciones malas, incluso una sola, pueden arruinar completamente el estimador.

Esta falta de estabilidad es un serio problema en las aplicaciones del modelo lineal.

Para ejemplos de este fenómeno consultar las Secciones 1.4.2 y 5.1 de Maronna, Martin, Yohai (2006).

2.3 Funcionales de regresión

Sea \mathbf{T} un funcional que toma valores en \mathbb{R}^p definido sobre un subconjunto del espacio de funciones de distribución en \mathbb{R}^{p+1} . Supongamos que el subconjunto incluye todas las funciones de distribución empíricas H_n correspondientes a muestras (\mathbf{x}'_i, y_i) de tamaño n con distribución H . Llamaremos a \mathbf{T} funcional de regresión. Se sigue entonces que $\mathbf{T}_n = \mathbf{T}(H_n)$ es un estimador de β_0 , y por lo tanto podemos pensar a \mathbf{T} también como un estimador de regresión.

Sea (\mathbf{x}', y) un vector aleatorio con distribución H . Consideremos $\tilde{y} = y + \mathbf{x}'\mathbf{b}$ para cierto $\mathbf{b} \in \mathbb{R}^p$, $\tilde{y} = cy$ para cierto $c \in \mathbb{R}$, y $\tilde{\mathbf{x}} = \mathbf{A}'\mathbf{x}$ para cierta $\mathbf{A} \in \mathbb{R}^{p \times p}$ inversible. Llamemos \tilde{H}_1 , \tilde{H}_2 y \tilde{H}_3 a las distribuciones de (\mathbf{x}, \tilde{y}) , (\mathbf{x}, \tilde{y}) y $(\tilde{\mathbf{x}}, y)$ respectivamente.

Las siguientes propiedades para un funcional de regresión \mathbf{T} son llamadas equivariancia por regresión, por escala y afín respectivamente:

- $\mathbf{T}(\tilde{H}_1) = \mathbf{T}(H) + \mathbf{b} \ \forall \mathbf{b} \in \mathbb{R}^p$.
- $\mathbf{T}(\tilde{H}_2) = c\mathbf{T}(H) \ \forall c \in \mathbb{R}$.
- $\mathbf{T}(\tilde{H}_3) = \mathbf{A}^{-1}\mathbf{T}(H) \ \forall \mathbf{A} \in \mathbb{R}^{p \times p}$ inversible.

De manera análoga definimos estas propiedades para estimadores de regresión. Diremos que un funcional o estimador de regresión es equivariante si cumple con estas tres propiedades. Las anteriores son propiedades deseables para cualquier funcional o estimador de regresión ya que nos permiten conocer como el estimador cambia bajo estas transformaciones en los datos. Además, estas propiedades facilitan enormemente la simulación.

Es fácil verificar que el estimador $\hat{\beta}_{MC}$ que definimos antes es equivariante por transformaciones de regresión, de escala y afines.

A continuación introducimos la noción de Fisher-consistencia y continuidad de funcionales, a las que haremos referencia más adelante.

Definición 2.1 Diremos que un funcional \mathbf{T} es Fisher-consistente si para cada familia de distribuciones F_{β} en \mathbb{R}^p se tiene que $\mathbf{T}(F_{\beta}) = \beta$.

Definición 2.2 Dada una distribución H en \mathbb{R}^p diremos que un funcional \mathbf{T} es continuo en H si se cumple que

$$H_n \rightarrow^d H \Rightarrow \mathbf{T}(H_n) \rightarrow \mathbf{T}(H).$$

2.4 Robustez

En esta sección definimos algunas medidas de robustez para estimadores de regresión.

2.4.1 Punto de ruptura finito

El concepto de punto de ruptura fue introducido en Hampel (1971) como una manera de medir que tan resistente es un estimador a contaminaciones en las observaciones. A continuación damos una definición de una versión de este concepto introducida en Donoho y Huber (1983).

Definición 2.3 Llamemos $\mathbf{z}_i = (\mathbf{x}'_i, y_i)$ para $i = 1, \dots, n$ a observaciones que siguen el modelo (2.1) y $\hat{\beta}(\mathbf{Z})$ al estimador de regresión basado en $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$. Definimos entonces el punto de ruptura finito por reemplazo de $\hat{\beta}$ como

$$\mathbf{RBP}(\hat{\beta}) = \frac{m^*}{n}, \quad (2.7)$$

donde

$$m^* = \max\{m \geq 0 : \hat{\beta}(\mathbf{Z}_m) \text{ está acotado } \forall \mathbf{Z}_m \in Z_m\}$$

y Z_m es la familia de conjuntos de vectores con al menos $n - m$ elementos en común con \mathbf{Z} .

Intuitivamente, dada una muestra de tamaño n , $\mathbf{RBP}(\hat{\beta})$ nos indica la proporción de las observaciones que puede ser sujeta a modificaciones sin que el estimador $\hat{\beta}$ quede completamente arruinado.

Se puede probar que si $\hat{\beta}$ es un estimador de regresión equivariante

$$\mathbf{RBP}(\hat{\beta}) \leq \frac{1}{n} \left\lceil \frac{n - k^* - 1}{2} \right\rceil \quad (2.8)$$

donde k^* está definido como

$$k^*(\mathbf{X}) = \max \{ \#(\beta \mathbf{x}_i = 0) : \beta \in \mathbb{R}^p, \beta \neq \mathbf{0} \}. \quad (2.9)$$

Este resultado está demostrado en la Sección 4.9.3 de Maronna, Martin, Yohai (2006).

Existe otra noción de punto de ruptura finito similar al punto de ruptura que definimos en (2.7) que es utilizada a menudo. A continuación la precisamos:

Definición 2.4 Llamemos $\mathbf{z}_i = (\mathbf{x}'_i, y_i)$ para $i = 1, \dots, n$ a observaciones que siguen el modelo (2.1). Sean $\hat{\beta}_n(Z_n)$ un estimador de regresión basado en $Z_n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ y

$$b(m, \hat{\beta}_n, Z_n) = \sup \left\| \hat{\beta}_{n+m}(Z_n \cup W_m) - \hat{\beta}_n(Z_n) \right\|,$$

donde el supremo se toma sobre todas las muestras W_m de tamaño m . Definimos el punto de ruptura finito aditivo de $\hat{\beta}_n$ en la muestra Z_n como

$$\mathbf{AFBP}(\hat{\beta}_n, Z_n) = \inf \left\{ \frac{m}{m+n} : b(m, \hat{\beta}_n, Z_n) = \infty \right\}. \quad (2.10)$$

Intuitivamente podemos interpretar $\mathbf{AFBP}(\hat{\beta}_n, Z_n)$ como la máxima fracción de outliers que podemos agregar a Z_n sin arruinar completamente el estimador.

En Serfling (2009) se dan desigualdades que relacionan el punto de ruptura finito aditivo con el punto de ruptura finito por reemplazo de un estimador. En ese trabajo se prueba que estas dos nociones de punto de ruptura son esencialmente equivalentes y que en el límite siempre coinciden. Según Serfling (2009) este resultado "Indulta a los autores que inadvertidamente cometieron el crimen de comparar el **RBP** de un estimador con el **AFBP** de otro."

2.4.2 Máximo sesgo y punto de ruptura asintóticos

Supongamos que (\mathbf{x}', y) cumple el modelo (2.1) con G_0 la función de distribución de \mathbf{x} . Nuevamente sea \mathbf{T} un funcional que toma valores en \mathbb{R}^p definido sobre un subconjunto del espacio de funciones de distribución H en \mathbb{R}^{p+1} . Supongamos que el subconjunto incluye todas las funciones de distribución empíricas H_n correspondientes a muestras (\mathbf{x}'_i, y_i) de tamaño n con distribución H y que \mathbf{T} es equivariante.

Definición 2.5 Definimos el sesgo asintótico $b(\mathbf{T}, H)$ de \mathbf{T} en H para que sea invariante bajo transformaciones de regresión como

$$b(\mathbf{T}, H) = (\mathbf{T}(H) - \beta_0)' \mathbf{C}(G_0) (\mathbf{T}(H) - \beta_0),$$

donde $\mathbf{C}(G_0)$ es funcional de dispersión equivariante por transformaciones afines y G_0 es la distribución de \mathbf{x} .

Si la función de distribución G_0 es elíptica alrededor del origen con matriz de dispersión \mathbf{A} podemos tomar $\mathbf{C}(G_0) = \mathbf{A}$.

Como nuestra definición de sesgo asintótico es invariante por transformaciones afines y de regresión, podemos suponer sin pérdida de generalidad que $\mathbf{C}(G_0) = \mathbf{I}$ donde \mathbf{I} es la matriz identidad de $\mathbb{R}^{p \times p}$ y que $\beta_0 = \mathbf{0}$. De esta forma

$$b(\mathbf{T}, H) = \|\mathbf{T}(H)\|. \quad (2.11)$$

Si el funcional \mathbf{T} es continuo en H , se sigue en virtud del Teorema de Glivenko-Cantelli que $\mathbf{T}(H)$ es el valor asintótico del estimador cuando la distribución subyacente de la muestra es H . Vamos a suponer que \mathbf{T} es asintóticamente insesgado en H_0 , es decir que

$$\mathbf{T}(H_0) = \beta_0.$$

Definimos ahora entornos de contaminación de H_0 .

Definición 2.6 *Dado $\epsilon > 0$ definimos*

$$V_\epsilon(H_0) = \{H: H = (1 - \epsilon)H_0 + \epsilon H^*\} \quad (2.12)$$

donde H^ es cualquier función de distribución en \mathbb{R}^{p+1} .*

Aunque los entornos de contaminación que definimos no son entornos en un sentido topológico, tienen la atractiva interpretación de que si $H \in V_\epsilon(H_0)$ para cierto $\epsilon > 0$, tendremos que una proporción de $1 - \epsilon$ de las observaciones de H provendrá de H_0 y una proporción de ϵ de H^* para cierta H^* .

Definición 2.7 *Dado $\epsilon > 0$ definimos el máximo sesgo asintótico de \mathbf{T} sobre $V_\epsilon(H_0)$ como*

$$B_\epsilon(\mathbf{T}) = \sup\{\|\mathbf{T}(H)\| : H \in V_\epsilon(H_0)\}. \quad (2.13)$$

Definición 2.8 *Definimos el punto de ruptura asintótico de \mathbf{T} como*

$$\mathbf{ABP}(\mathbf{T}) = \inf\{\epsilon : B_\epsilon(\mathbf{T}) = \infty\}. \quad (2.14)$$

Como se menciona en la Sección 5.4.1 de Maronna, Martín, Yohai (2006), se puede probar que si \mathbf{T} es equivariante

$$\mathbf{ABP}(\mathbf{T}) \leq \frac{1 - \alpha}{2} \quad (2.15)$$

donde $\alpha = \sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0)$.

3 Estimadores robustos de regresión

3.1 M-estimadores

En esta sección, vamos a definir y dar algunas de las propiedades más importantes de los M-estimadores de regresión, que son de vital importancia para la teoría de la regresión lineal robusta. Los M-estimadores, introducidos en Huber (1973), se pueden pensar como un intento por hacer del enfoque clásico, el de mínimos cuadrados, más robusto. Para definir los M-estimadores necesitaremos de algunas definiciones preliminares.

Definición 3.1 Sea ρ una función definida en \mathbb{R} , decimos que ρ es una ρ -función si se cumple:

- $\rho(x)$ es no decreciente como función de $|x|$.
- $\rho(0) = 0$.
- $\rho(x)$ es creciente para los $x > 0$ tales que $\rho(x) < \rho(\infty)$.
- Si ρ es acotada, $\rho(\infty) = 1$.
- ρ tiene a lo sumo finitas discontinuidades.

Definición 3.2 Llamaremos ψ -función a cualquier función que sea la derivada de una ρ -función, esto implica en particular que

- ψ es impar y $\psi(x) \geq 0$ si $x \geq 0$.

Si ψ tiene a cero en infinito, diremos que ψ es redescendiente.

A continuación damos dos ejemplos de ρ -funciones.

La familia de ρ -funciones bicuadráticas de Tukey está dada por

$$\rho_c^B(x) = \left[1 - \left(1 - \left(\frac{x}{c} \right)^2 \right)^3 \right] \mathbf{I}_{|x| \leq c}.$$

La correspondiente familia de ψ -funciones está dada por

$$\psi_c^B(x) = \frac{6}{c^2} x \left[1 - \left(\frac{x}{c} \right)^2 \right]^2 \mathbf{I}_{|x| \leq c}.$$

Notemos que como las bicuadráticas son acotadas, sus correspondientes ψ son redescendentes. Estas serán las ρ -funciones de preferencia a lo largo de este trabajo.

La familia de ρ -funciones de Huber está dada por

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

La correspondiente familia de ψ – *funciones* está dada por

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

ψ_k^H es un claro ejemplo de una ψ – *función* monótona.

En la gran mayoría de los casos de interés $\psi(0) = 0$, $\psi'(0)$ existe y es no nula, lo que implica que ψ es aproximadamente lineal en el cero.

Sea

$$W(x) = \begin{cases} \frac{\psi(x)}{\psi'(0)} & \text{si } x \neq 0 \\ \psi'(0) & \text{si } x = 0 \end{cases}. \quad (3.1)$$

Llamaremos $W(x)$ a la función de pesos asociada a ρ o a ψ .

La función de pesos asociada a ρ_c^B esta dada por

$$W(x) = \left[\left(1 - \left(\frac{x}{c} \right)^2 \right)^2 \right] \mathbf{I}_{|x| \leq c}.$$

Ahora podemos definir los M-funcionales y M-estimadores de escala.

Definición 3.3 Sea $0 < b < 1$, F una función de distribución definida en \mathbb{R} y ρ una ρ – función acotada. Entonces, el M-funcional de escala esta definido por

$$s(F) = \inf\{s > 0 : E_F\left(\rho\left(\frac{u}{s}\right)\right) \leq b\}.$$

Dada una muestra $\mathbf{u} = (u_1, \dots, u_n)$ de F , el correspondiente M-estimador de escala $s_n(\mathbf{u})$ se obtiene reemplazando F por la distribución empírica de la muestra.

Es fácil ver que

$$s(F) > 0 \iff P_F(u = 0) < 1 - b. \quad (3.2)$$

Si esta condición se satisface con ρ continua, podemos reemplazar el menor o igual por un igual en la definición de $s(F)$. En este caso dada una muestra $\mathbf{u} = (u_1, \dots, u_n)$ de F , el correspondiente M-estimador de escala $s_n(\mathbf{u})$ queda definido por

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{u_i}{s_n(\mathbf{u})}\right) = b. \quad (3.3)$$

Como nosotros trabajaremos únicamente con ρ – *funciones* continuas y principalmente con la función bicuadrática, usaremos la definición (3.3). Es fácil probar que $s_n(\mathbf{u})$ dado por (3.3) cumple que

$$s_n(c\mathbf{u}) = |c| s_n(\mathbf{u}) \quad \forall c \in \mathbb{R} \quad (3.4)$$

y

$$|u_i| \leq |v_i| \quad \text{para } i = 1, \dots, n \implies s_n(\mathbf{u}) \leq s_n(\mathbf{v}). \quad (3.5)$$

Se prueba en el trabajo de Yohai y Zamar (1988) que el punto de ruptura de $s_n(\mathbf{u})$ dado por (3.3) es $\min(b, 1 - b)$.

Para más detalles, consultar la Sección 5.2 de Huber (1981).

Fijado $b = 0.5$ y tomando $\rho = \rho_c^B$ podemos elegir $c = 1.54$ de manera tal que el valor asintótico de $s_n(\mathbf{u})$ coincida con el desvío standard cuando las u_i son normales, manteniendo además un punto de ruptura de 0.5 como se menciona en la Sección 5.5 de Maronna, Martin y Yohai (2006).

Ahora estamos en condiciones de definir los M-funcionales y M-estimadores de regresión.

Definición 3.4 *Supongamos que (\mathbf{x}', y) sigue el modelo (2.1) con (2.5), que H es una función de distribución en \mathbb{R}^{p+1} , y que ρ es una ρ -función. Definimos el M-funcional de regresión basado en H como*

$$\mathbf{T}_M(H) = \arg \min_{\beta \in \mathbb{R}^p} E_H\left(\rho\left(\frac{y - \beta' \mathbf{x}}{s(H)}\right)\right) = \arg \min_{\beta \in \mathbb{R}^p} E_H\left(\rho\left(\frac{r(\beta)}{s(H)}\right)\right) \quad (3.6)$$

donde $s(H)$ es un funcional de escala auxiliar que se necesita para hacer de \mathbf{T}_M un funcional equivariante por escala.

Notemos que si $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ son observaciones i.i.d del modelo (2.1) y H_n es su función de distribución empírica tenemos que

$$\mathbf{T}_M(H_n) = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho\left(\frac{y_i - \beta' \mathbf{x}_i}{s_n}\right) = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho\left(\frac{r_i(\beta)}{s_n}\right) \quad (3.7)$$

donde $s_n = s(H_n)$.

Veremos como elegir la escala $s(H)$ cuando estudiemos el procedimiento de MM-estimación en la Sección 3.4.

Llamamos a $\hat{\beta}_M(\mathbf{X}, \mathbf{y}) := \mathbf{T}_M(H_n)$ el M-estimador de regresión basado en la muestra $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$.

Derivando (3.7) respecto de β obtenemos el análogo a las ecuaciones normales

$$\sum_{i=1}^n \mathbf{x}_i \psi\left(\frac{r_i(\beta)}{s_n}\right) = \mathbf{0}. \quad (3.8)$$

Para resolver esta ecuación podemos utilizar el algoritmo de mínimos cuadrados con pesos iterados, IRWLS por sus siglas en inglés, descrito en la Sección 4.5.2 de Maronna, Martin, Yohai (2006), que converge siempre que $W(x)$ definida en (3.1) sea no creciente para $x > 0$, propiedad que cumple por ejemplo la función de pesos asociada a la función bicuadrática.

Se prueba en el Teorema 10.15 de Maronna, Martin, Yohai (2006) que si ψ es monótona, la solución de la ecuación (3.8) es única y en el capítulo 6 de Huber (1981) se dan condiciones más generales para asegurar la existencia, Fisher-consistencia y continuidad de funcionales definidos por la solución de (3.6).

Para el caso de ψ redescendiente la ecuación (3.8) puede tener raíces "malas", es decir, que no corresponden a óptimos de (3.7) sino a mínimos locales. En este caso, es en general una tarea muy difícil encontrar el mínimo global de (3.7). Por esta razón, a la hora de calcular M-estimadores con una ψ redescendiente es muy importante el estimador inicial que utilizamos para inicializar el algoritmo de IRWLS; estudiaremos más en profundidad este tema cuando veamos los MM-estimadores en la Sección 3.4.

Recordemos que como trabajamos bajo la hipótesis de que los regresores son variables aleatorias, tanto las y_i como las \mathbf{x}_i pueden ser outliers. Si calculamos un M-estimador con una ψ — función monótona, y \mathbf{x}_j es una observación "muy alejada de las demás", para hacer que la suma (3.8) sea cero $\psi(r_j(\hat{\beta}_{\mathbf{M}}))$ debe estar cerca de cero y por lo tanto $\hat{\beta}_{\mathbf{M}}$ está muy influenciado por (\mathbf{x}'_j, y_j) . De hecho, el punto de ruptura finito de cualquier M-estimador monótono es cero cuando las variables regresoras pueden contener outliers, como se prueba en la Sección 5.16.1 de Maronna, Martin, Yohai (2006). Notemos que estimadores clásicos como el de mínimos cuadrados o L1 son M-estimadores con ψ — funciones $\psi(x) = x$ y $\psi(x) = \text{sgn}(x)$ respectivamente. Como ambos están calculados con ψ — funciones monótonas tendrán punto de ruptura finito igual a cero. Esto no sucede cuando la ψ — función utilizada es redescendiente, ya que las observaciones con residuos muy grandes tienen muy poco peso naturalmente. Por esta razón, y por su importancia para la teoría de MM-estimadores, nos enfocaremos principalmente en los M-estimadores calculados con una ρ — función acotada.

3.1.1 Propiedades asintóticas

Supongamos que $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ son observaciones i.i.d que cumplen el modelo (2.1), que se satisface (2.5) y supongamos además que la escala auxiliar s_n converge en probabilidad a cierto σ_1 . Supongamos también que $E_{G_0}(\|\mathbf{x}\|^2) < \infty$. Se prueba en las Secciones 6.2 y 6.3 de Huber (1981) que bajo condiciones muy generales sobre la ρ — función utilizada y sobre H_0 , $\hat{\beta}_{\mathbf{M}}$ definido por (3.7) es fuertemente consistente para β_0 y es asintóticamente normal:

$$\sqrt{n}(\hat{\beta}_{\mathbf{M}} - \beta_0) \rightarrow^d N_p(\mathbf{0}, v \mathbf{V}_{\mathbf{x}}^{-1}) \quad (3.9)$$

con $\mathbf{V}_{\mathbf{x}} = E_{G_0}(\mathbf{x}\mathbf{x}')$ y v definida por

$$v = \sigma_1^2 \frac{E_{F_0}(\psi^2(\frac{u}{\sigma_1}))}{\left[E_{F_0}(\psi'(\frac{u}{\sigma_1}))\right]^2}. \quad (3.10)$$

Notemos que como $\hat{\beta}_{MC}$ es un M-estimador, este resultado prueba (2.6). Además, siempre que $E_{G_0}(\|\mathbf{x}\|^2) < \infty$ la eficiencia de $\hat{\beta}_{\mathbf{M}}$ no dependerá de G_0 .

En particular, si $\sigma_0^2 = \text{Var}(u)$, deducimos que la eficiencia cuando los errores son normales de $\hat{\beta}_{\mathbf{M}}$ está dada por

$$\text{Eff}(\hat{\beta}_{\mathbf{M}}) = \frac{\sigma_0^2}{v}. \quad (3.11)$$

Es fácil ver que $Eff(\widehat{\beta}_{\mathbf{M}})$ no depende de σ_0^2 , siempre que $\sigma_1 = \sigma_0$.

Por otro lado, ρ_c^B cumple con todos los requerimientos enunciados en las Secciones 6.2 y 6.3 de Huber (1981). Como se muestra en la Sección 2.2.4 de Maronna, Martin, Yohai (2006) siempre podemos elegir la constante c de modo de tener una eficiencia asintótica bajo normalidad de los errores tan alta como deseemos.

A continuación transcribimos una tabla con valores de c y la correspondiente eficiencia de $\widehat{\beta}_{\mathbf{M}}$ calculado utilizando ρ_c^B .

$Eff(\widehat{\beta}_{\mathbf{M}})$	0.80	0.85	0.90	0.95
c	3.14	3.44	3.88	4.68

Tabla 1.

3.2 S-estimadores

Los S-estimadores de regresión fueron introducidos en Rousseeuw y Yohai (1984), y se los puede definir como estimadores que minimizan cierta M-escala robusta de los residuos. Como veremos en esta sección, son estimadores muy robustos, en el sentido de que podemos hacer que alcancen las cotas para puntos de ruptura finito y asintótico para estimadores equivariantes dadas en (2.8), (2.15).

A continuación, damos una definición precisa.

Definición 3.5 Sea $(\mathbf{x}', y) \in \mathbb{R}^{p+1}$ un vector aleatorio con cierta distribución H . Para $\beta \in \mathbb{R}^p$ llamemos F_β a la distribución de $r(\beta) = y - \beta' \mathbf{x}$. Sea $s(F)$ un M -funcional de escala calculado con una ρ -función continua y acotada y llamemos $s(F_\beta) = s(\beta, H)$. Diremos entonces que $\mathbf{T}_S(H)$ es un S -funcional de regresión si

$$s(\mathbf{T}_S(H), H) = \min_{\beta \in \mathbb{R}^p} s(\beta, H). \quad (3.12)$$

La existencia de S -funcionales es consecuencia del siguiente lema y del Lema 5.3 que se encuentra en el apéndice.

Lema 3.1 Supongamos que $\sup_{\|\beta\|=1} P_H(\beta' \mathbf{x} = 0) < 1 - b$, donde b corresponde a la definición de $s(F)$, calculado con una ρ -función continua y acotada.

Luego, si $\|\beta_n\| = \infty$, entonces $\lim_n s(\beta_n, H) = \infty$.

Demostración. Supongamos que $\lim_n \|\beta_n\| = \infty$ y sea $\beta_n^* = \frac{\beta_n}{\|\beta_n\|}$. Podemos suponer sin pérdida de generalidad que $\beta_n^* \rightarrow \beta^*$ con $\|\beta^*\| = 1$.

Para probar el lema, basta ver que $\forall s > 0$ $E_H \rho\left(\frac{y - \beta_n^* \mathbf{x}}{s}\right) > b$ para n suficientemente grande.

En efecto,

$$E_H \rho\left(\frac{y - \beta_n^* \mathbf{x}}{s}\right) \geq E_H \rho\left(\frac{y - \|\beta_n\| \beta_n^{*'} \mathbf{x}}{s}\right) \mathbf{I}_{(\beta_n^{*'} \mathbf{x} \neq 0)}.$$

Es inmediato probar que

$$\rho \left(\frac{y - \|\beta_n\| \beta_n^{*\prime} \mathbf{x}}{s} \right) \mathbf{I}_{(\beta_n^{*\prime} \mathbf{x} \neq 0)} \rightarrow \mathbf{I}_{(\beta^{*\prime} \mathbf{x} \neq 0)} \text{ c.s.}$$

Entonces teniendo en cuenta que $\sup_{\|\beta\|=1} P_H(\beta' \mathbf{x} = 0) < 1 - b$ el resultado se sigue del Teorema de Convergencia Mayorada. ■

En el caso de que exista más de un $\mathbf{T}(H)$ que satisfaga (3.12) la elección de $\mathbf{T}_S(H)$ es arbitraria.

Obtenemos un S-estimador de regresión que notaremos con $\hat{\beta}_S$ reemplazando en (3.12) a H por la distribución empírica de una muestra. Es decir que dadas $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ observaciones i.i.d del modelo (2.1) con (2.5) si llamamos H_n a la distribución empírica de la muestra obtenemos

$$\hat{\beta}_S(\mathbf{X}, \mathbf{y}) = \arg \min_{\beta \in \mathbb{R}^p} s_n(r(\beta)).$$

Es inmediato verificar que los S-funcionales son equivariantes.

3.2.1 Máximo sesgo asintótico

A lo largo de toda esta subsección supondremos que G_0 , la función de distribución de \mathbf{x} , es elíptica. En virtud de lo observado en (2.11), supondremos sin pérdida de generalidad que $\beta_0 = 0$ y que G_0 es esférica. El siguiente lema, probado en Martin, Yohai y Zamar (1989), será de fundamental importancia para caracterizar el máximo sesgo asintótico de S-funcionales.

Lema 3.2 *Supongamos el modelo (2.1) con H_0 dada por (2.3). Supongamos que F_0 cumple (2.4) y que $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) = 0$. Sea ρ una ρ -función continua y acotada y definamos*

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

Entonces g es continua, estrictamente creciente respecto de $\|\beta\|$ y estrictamente decreciente respecto de s .

Se sigue inmediatamente de este lema que un S-funcional $\mathbf{T}_S(H)$ es Fisher-consistente en H_0 .

Observación 3.1 *Como corolario de los resultados de Fasano, Maronna, Sued, y Yohai (2012) tenemos que si (\mathbf{x}', y) sigue el modelo (2.1) con (2.5), F_0 satisface (2.4) y $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) = 0$ entonces $\mathbf{T}_S(H)$ es continuo en H_0 .*

Combinando este corolario con el lema previo obtenemos la consistencia de los S-estimadores. Probaremos por otra vía propiedades asintóticas de S-estimadores en la sección que sigue.

Llamemos $g_1^{-1}(\cdot, \|\beta\|)$ y $g_2^{-1}(s, \cdot)$ a las inversas de g respecto de s y de $\|\beta\|$ respectivamente.

El siguiente teorema, que es una generalización del Teorema 3.1 de Martin, Yohai y Zamar (1989), caracteriza el máximo sesgo asintótico y punto de ruptura asintótico de funcionales que son "casi S-funcionales". Su demostración se halla en el Apéndice.

Teorema 3.2 *Supongamos que se cumplen las condiciones del Lema 3.2. Sea $\delta \geq 0$ y supongamos que \mathbf{T} es un funcional de regresión equivariante y Fisher-consistente en H_0 tal que si H es una distribución en \mathbb{R}^{p+1} se cumple que*

$$s(\mathbf{T}(H), H) \leq (1 + \delta) \min_{\beta \in \mathbb{R}^p} s(\beta, H).$$

Llamemos

$$\alpha_1 = g_2^{-1} \left((1 + \delta) g_1^{-1} \left(\frac{b - \epsilon}{1 - \epsilon}, 0 \right), \frac{b}{1 - \epsilon} \right)$$

$$\alpha_2 = g_2^{-1} \left(\frac{g_1^{-1} \left(\frac{b - \epsilon}{1 - \epsilon}, 0 \right)}{(1 + \delta)}, \frac{b}{1 - \epsilon} \right).$$

Entonces sobre un entorno de contaminación V_ϵ de H_0 , $\mathbf{B}_\epsilon(\mathbf{T})$ cumple que

$$\epsilon < \min(b, 1 - b) \implies \alpha_2 \leq \mathbf{B}_\epsilon(\mathbf{T}) \leq \alpha_1$$

y

$$\epsilon \geq \min(b, 1 - b) \implies \mathbf{B}_\epsilon(\mathbf{T}) = \infty.$$

En particular, $\mathbf{ABP}(\mathbf{T}) = \min(b, 1 - b)$.

Tomando $\delta = 0$ en el teorema anterior, caracterizamos el máximo sesgo y punto de ruptura asintótico de S-funcionales. Notemos que en particular todos los funcionales que cumplan con las condiciones del teorema anterior con $b = 0.5$ tendrán punto de ruptura asintótico 0.5 que es el máximo para funcionales equivariantes con $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) = 0$.

Supongamos que estamos en las condiciones del teorema anterior y que (\mathbf{x}, y) tiene distribución $N(\mathbf{0}, \mathbf{I}_{p+1})$. Entonces

$$g(s, \gamma) = h \left(\frac{\sqrt{(1 + \gamma^2)}}{s} \right)$$

donde $h(\lambda) = E(\rho(\lambda u))$ con $u \sim N(0, 1)$.

Luego

$$g_1^{-1}(t, \gamma) = \frac{\sqrt{(1 + \gamma^2)}}{h^{-1}(t)}$$

y

$$g_2^{-1}(s, t) = \sqrt{(s h^{-1}(t))^2 - 1}.$$

De donde obtenemos las siguiente cotas para el sesgo:

$$\left[\frac{h^{-1}(\frac{b}{1 - \epsilon})}{(1 + \delta) h^{-1}(\frac{b - \epsilon}{1 - \epsilon})} \right]^2 - 1 \leq \mathbf{B}_\epsilon(\mathbf{T})^2 \leq \left[\frac{(1 + \delta) h^{-1}(\frac{b}{1 - \epsilon})}{h^{-1}(\frac{b - \epsilon}{1 - \epsilon})} \right]^2 - 1.$$

3.2.2 Punto de ruptura finito

Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) y $\hat{\beta}_{\mathbf{S}}$ un S-estimador basado en esta muestra calculado con una ρ -función continua y acotada. Sean $\delta_n \geq 0$ y $\hat{\beta}$ un estimador equivariante que cumple que

$$s_n(r(\hat{\beta})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{\mathbf{S}})), \quad (3.13)$$

donde $s_n(r(\beta))$ es la escala de los residuos que minimiza $\hat{\beta}_{\mathbf{S}}$.

Veremos en el siguiente teorema, que es una generalización del Teorema 5.1 de Maronna, Martin y Yohai (2006), que $\hat{\beta}$ alcanza la cota para el punto de ruptura finito por reemplazo de estimadores equivariantes (2.8) siempre que en la definición de s_n elijamos $b = \left(\left\lfloor \frac{n-k^*-1}{2} \right\rfloor + \gamma \right) / n$ con $\gamma \in (0, 1)$ y k^* como en (2.9).

En particular, este teorema caracteriza el punto de ruptura finito por reemplazo de S-estimadores calculados con $b = \left(\left\lfloor \frac{n-k^*-1}{2} \right\rfloor + \gamma \right) / n$.

La demostración del Teorema 3.2 se halla en el apéndice.

Teorema 3.3 *Sea $\hat{\beta}$ un estimador equivariante como en (3.13) y $m_{\max}^* = [(n - k^* - 1) / 2]$. Llamemos $m(b)$ al mayor entero menor que nb . Entonces si m^* es como en (2.7), se cumple*

- (i) $m^* \leq nb$.
- (ii) Si $[nb] \leq m_{\max}^*$ entonces $m^* \geq m(b)$.

Se sigue que si nb no es entero y $b \leq m_{\max}^ / n$, entonces $m^* = [nb]$ y por lo tanto tomando $b = \left(\left\lfloor \frac{n-k^*-1}{2} \right\rfloor + \gamma \right) / n$ obtenemos $\mathbf{RBP}(\hat{\beta}) = \left\lfloor \frac{n-k^*-1}{2} \right\rfloor / n$.*

Notemos que si \mathbf{X} tiene rango completo

$$\mathbf{RBP}(\hat{\beta}) = \frac{1}{n} \left\lfloor \frac{n-p}{2} \right\rfloor,$$

que es aproximadamente 0.5 cuando n es grande.

3.2.3 Propiedades asintóticas

Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1), y $\hat{\beta}_{\mathbf{S}}$ un S-estimador basado en esta muestra calculado con una ρ -función continua y acotada. Veamos que $\hat{\beta}_{\mathbf{S}}$ es un M-estimador en el sentido de (3.7), para lo cual probaremos que

$$\sum_{i=1}^n \rho \left(\frac{r_i(\hat{\beta}_{\mathbf{S}})}{s_0} \right) \leq \sum_{i=1}^n \rho \left(\frac{r_i(\beta)}{s_0} \right) \quad \forall \beta \in \mathbb{R}^p.$$

donde $s_0 = s_n(r(\hat{\beta}_{\mathbf{S}}))$ y $s_n(r(\beta))$ es la M-escala de los residuos que minimiza $\hat{\beta}_{\mathbf{S}}$.

En efecto si $\beta \in \mathbb{R}^p$ es tal que

$$\sum_{i=1}^n \rho \left(\frac{r_i(\hat{\beta})}{s_0} \right) > \sum_{i=1}^n \rho \left(\frac{r_i(\beta)}{s_0} \right)$$

como ρ es monótona tendríamos que existe un $s^* < s_0$ tal que

$$\sum_{i=1}^n \rho \left(\frac{r_i(\beta)}{s^*} \right) = nb.$$

Pero esto último implica que $s_n(r(\beta)) < s_n(r(\hat{\beta}_S))$, que es absurdo por (3.12).

Se sigue que las propiedades asintóticas de $\hat{\beta}_S$ pueden estudiarse con los métodos que mencionamos en la Sección 3.1.1. También obtendremos resultados asintóticos para S-estimadores en la siguiente sección. Lamentablemente los S-estimadores calculados con una ρ – función suave no pueden tener simultáneamente un punto de ruptura y una eficiencia altos. Se demuestra en Hossjer (1992) que un S-estimador con punto de ruptura asintótico de 0.5 tiene una eficiencia bajo normalidad de los errores de a lo sumo 0.33, que aún siendo más alta que la eficiencia de otros estimadores muy robustos como el LMS o el LTS de Rousseeuw, es inaceptable teniendo alternativas que combinan ambas propiedades como veremos en lo que resta del trabajo. El rol que jugaran los S-estimadores de aquí en más será el de estimadores iniciales para resolver la ecuación (3.8) con el algoritmo IRWLS. Este procedimiento, llamado MM-estimación, será estudiado en la Sección 3.4.

Para detalles sobre la computación numérica de S-estimadores consultar la Sección 5.7.2 de Maronna, Martín y Yohai (2006) y Salibian-Barrera y Yohai (2005). Solo mencionaremos que los algoritmos disponibles para su cálculo son computacionalmente muy costosos.

A continuación, presentamos los τ -estimadores, que son una generalización de los S-estimadores, y combinan propiedades de robustez con una eficiencia normal tan alta como deseamos.

3.3 τ -estimadores

Los τ -estimadores fueron introducidos en Yohai y Zamar (1988) y están definidos por minimizar un nuevo estimador de la escala de los residuos. Como mencionamos previamente, estos estimadores combinan un punto de ruptura alto con una eficiencia normal tan alta como deseamos. A continuación damos una definición precisa de estos nuevos estimadores de escala.

Definición 3.6 Sean ρ_1 y ρ_2 dos ρ – funciones acotadas. Supongamos además que ρ_2 es continuamente diferenciable y se cumple que $2\rho_2(x) - \psi_2(x)x \geq 0$. Sea H una distribución en \mathbb{R}^{p+1} . Sea $s(\beta, H)$ un M-funcional de escala calculado utilizando a ρ_1 como ρ – función. Definimos $\tau(\beta, H)$ como

$$\tau(\beta, H)^2 = s(\beta, H)^2 E_H(\rho_2 \left(\frac{y - \beta' \mathbf{x}}{s(\beta, H)} \right)). \quad (3.14)$$

Dadas $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1), si llamamos H_n a la distribución empírica de la muestra, reemplazando en la definición anterior obtenemos el estimador de escala de los residuos $\tau_n(r(\boldsymbol{\beta}))$ dado por

$$\tau_n(r(\boldsymbol{\beta}))^2 = s_n(r(\boldsymbol{\beta}))^2 \frac{1}{n} \sum_{i=1}^n \rho_2 \left(\frac{r_i(\boldsymbol{\beta})}{s_n(r(\boldsymbol{\beta}))} \right).$$

Es claro que τ_n es equivariante por transformaciones de escala. Notemos que si tomamos $\rho_1 = \rho_2$, tenemos que $\tau_n(r(\boldsymbol{\beta})) = \sqrt{b} s_n(r(\boldsymbol{\beta}))$, y si $\rho_2(u) = u^2$, $\tau_n(r(\boldsymbol{\beta}))$ es el desvío standard muestral de los residuos.

Ahora podemos definir los τ -funcionales y τ -estimadores.

Definición 3.7 En las condiciones de la definición 3.6, dada H una distribución en \mathbb{R}^{p+1} definimos el τ -funcional de regresión $\mathbf{T}_\tau(H)$ como

$$\mathbf{T}_\tau(H) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \tau(\boldsymbol{\beta}, H). \quad (3.15)$$

Dadas $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que siguen el modelo (2.1), si llamamos H_n a la distribución empírica de la muestra, reemplazando en la definición anterior obtenemos el τ -estimador de regresión $\hat{\boldsymbol{\beta}}_\tau$ dado por

$$\hat{\boldsymbol{\beta}}_\tau(\mathbf{X}, \mathbf{y}) = \mathbf{T}_\tau(H_n) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \tau_n(r(\boldsymbol{\beta})). \quad (3.16)$$

Por lo tanto si tomamos $\rho_1 = \rho_2$ recuperamos los S-estimadores y si tomamos $\rho_2(u) = u^2$ recuperamos el estimador de mínimos cuadrados.

Se prueba en la Sección 2 de Yohai y Zamar (1988) que los τ -estimadores pueden pensarse como un tipo de M-estimador. De hecho veremos en el Teorema 3.4 que coinciden asintóticamente con un M-estimador calculado con

$$\psi_0(u) = W_0 \psi_1\left(\frac{u}{\sigma_0}\right) + \psi_2\left(\frac{u}{\sigma_0}\right) \quad (3.17)$$

donde

$$W_0 = \frac{2E_{F_0}(\rho_2(\frac{u}{\sigma_0})) - E_{F_0}(\psi_2(\frac{u}{\sigma_0})\frac{u}{\sigma_0})}{E_{F_0}(\psi_1(\frac{u}{\sigma_0})\frac{u}{\sigma_0})},$$

y σ_0 es tal que

$$E_{F_0}(\rho_1(\frac{u}{\sigma_0})) = b,$$

con b el lado derecho de la ecuación que define a s_n .

El siguiente lema, que aparece en Yohai y Zamar (1988), prueba la Fisher-consistencia de τ -funcionales.

Lema 3.3 Sean ρ_1 y ρ_2 dos ρ -funciones acotadas. Supongamos además que ρ_2 es continuamente diferenciable y se cumple que $2\rho_2(x) - \psi_2(x)x \geq 0$. Sea H_0 como en (2.3), con F_0 que cumple (2.4) y supongamos que

$$\sup_{\|\boldsymbol{\beta}\|=1} P_{G_0}(\boldsymbol{\beta}'\mathbf{x}=0) < 1.$$

Sea $s(\boldsymbol{\beta}, H_0)$ un M-funcional de escala calculado utilizando a ρ_1 como ρ -función. Entonces $\tau(\boldsymbol{\beta}, H_0)$ tiene un único mínimo en $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

Para detalles sobre las propiedades de robustez de los τ -estimadores consultar la Sección 3 de Yohai y Zamar (1988), solo mencionaremos que dependen únicamente de la función ρ_1 y que eligiendo correctamente el estimador inicial podemos obtener un estimador con un alto punto de ruptura finito aditivo (2.10).

El siguiente teorema, que aparece en Yohai y Zamar (1988), prueba la consistencia de los τ -estimadores.

Teorema 3.4 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d del modelo (2.1) con (2.5). Sea H_0 como en (2.3), F_0 que satisface (2.4) y G_0 tal que

$$\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 0.5.$$

Sean ρ_1 y ρ_2 dos ρ -funciones acotadas. Supongamos además que ρ_2 es continuamente diferenciable y se cumple que $2\rho_2(x) - \psi_2(x)x \geq 0$. Sea $(\hat{\beta}_{0,n})_n$ una sucesión de estimadores que converge casi seguramente a β_0 , y $(\hat{\beta}_{1,n})_n$ otra sucesión de estimadores tales que

$$\tau_n(r(\hat{\beta}_{1,n})) \leq \tau_n(r(\hat{\beta}_{0,n})) \text{ casi seguro.}$$

Entonces $\hat{\beta}_{1,n}$ converge casi seguro a β_0 .

Notemos que según este teorema no precisamos tener un mínimo absoluto de τ_n para tener consistencia. Alcanza con tener un mínimo local que mejora la escala τ_n de los residuos de un estimador consistente. El algoritmo para computar numéricamente τ -estimadores descrito en la Sección 6 de Yohai y Zamar (1988) hace precisamente esto. Lamentablemente, este algoritmo es computacionalmente tan costoso como el algoritmo para calcular S-estimadores.

El siguiente teorema, que aparece en Yohai y Zamar (1988), prueba la normalidad asintótica de τ -estimadores.

Teorema 3.5 Supongamos que se cumplen las condiciones del Teorema 3.3. Supongamos además que ρ_1 es continuamente diferenciable y tanto ρ_1 como ρ_2 son eventualmente constantes. Si $E_{G_0}(\|\mathbf{x}\|^2) < \infty$ y $\mathbf{V}_{\mathbf{x}} = E_{G_0}(\mathbf{x}\mathbf{x}')$ es no singular tenemos que

$$\sqrt{n}(\hat{\beta}_{1,n} - \beta_0) \rightarrow^d N(\mathbf{0}, \sigma_0^2 \frac{E_{F_0}(\psi_0(u)^2)}{E_{F_0}(\psi'_0(u))^2} \mathbf{V}_{\mathbf{x}}^{-1})$$

con ψ_0 como en (3.17) y $\hat{\beta}_{1,n}$ como en el Teorema 3.3.

Como las propiedades de robustez solo dependen de ρ_1 , podemos elegir ρ_2 para tener una eficiencia asintótica tan alta como deseemos cuando F_0 es normal. En la página 16 de Yohai y Zamar (1988) se muestra que si empezamos con un estimador $\hat{\beta}_{0,n}$ con punto de ruptura finito que tienda a 0.5, podemos elegir por ejemplo $\rho_1 = \rho_{c_1}^B$, $\rho_2 = \rho_{c_2}^B$ con $c_1 = 1.54$, $c_2 = 6.08$ y $b = 0.203$ para obtener un τ -estimador con punto de ruptura finito que tienda a 0.5 y eficiencia normal 0.95.

Lamentablemente, a medida que aumenta la eficiencia del estimador también aumenta el sesgo, como se muestra en la Sección 7 de Yohai y Zamar (1988).

Para cerrar esta sección mencionamos que si en los Teoremas 3.3 y 3.4 tomamos $\rho_1 = \rho_2$ y $\hat{\beta}_{0,n} = \beta_0 \forall n$ obtenemos condiciones para la consistencia y normalidad asintótica de S-estimadores.

3.4 MM-estimadores

Los MM-estimadores fueron introducidos en Yohai (1987) y fueron la primera clase de estimadores que logran un punto de ruptura alto junto con una eficiencia tan alta como deseamos bajo normalidad de los errores. La idea es que para obtener un M-estimador robusto y altamente eficiente no precisamos calcular el mínimo global de (3.7), si no que alcanza con encontrar un mínimo local a partir de un estimador inicial robusto, utilizando una escala robusta del mismo para asegurar la equivariancia del estimador.

A continuación damos una definición precisa de los MM-funcionales y MM-estimadores de regresión.

Definición 3.8 Sea (\mathbf{x}', y) un vector aleatorio que sigue el modelo (2.1) con (2.5) con distribución H . Sean $\rho_0 \geq \rho_1$ dos ρ -funciones continuas y acotadas y sea $\mathbf{T}_S(H)$ un S-funcional calculado con ρ_0 . Definimos el MM-funcional de regresión como

$$\mathbf{T}_{MM}(H) = \arg \min_{\beta \in \mathbb{R}^p} E_H \rho_1 \left(\frac{y - \beta' \mathbf{x}}{s(\mathbf{T}_S(H), H)} \right). \quad (3.18)$$

Dadas $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5) definimos el MM estimador de regresión $\hat{\beta}_{MM}$ en tres etapas de la siguiente manera:

- Calcular un S-estimador $\hat{\beta}_S$ de β_0 con punto de ruptura alto, posiblemente 0.5.
- Calcular $s_0 = s_n(r(\hat{\beta}_S))$ donde s_n es la escala que minimiza $\hat{\beta}_S$.
- Sea ρ_1 otra ρ -función tal que $\rho_0 \geq \rho_1$. Definimos finalmente $\hat{\beta}_{MM}$ como cualquier solución de

$$\sum_{i=1}^n \mathbf{x}_i \psi_1 \left(\frac{r_i(\beta)}{s_0} \right) = \mathbf{0} \quad (3.19)$$

que cumpla que

$$L(\hat{\beta}_{MM}) \leq L(\hat{\beta}_S) \quad (3.20)$$

donde

$$L(\beta) = \sum_{i=1}^n \rho_1 \left(\frac{r_i(\beta)}{s_0} \right)$$

y se define $\rho_1\left(\frac{0}{0}\right) = 0$.

El lema que damos a continuación implica que el mínimo global de $L(\beta)$ siempre existe, pero la definición que damos de $\hat{\beta}_{MM}$ nos permitirá obtener un estimador robusto y altamente eficiente bajo normalidad de los errores encontrando simplemente un mínimo local de $L(\beta)$ que cumpla (3.20). El algoritmo propuesto en Yohai (1985) para calcular MM-estimadores es una modificación de IRWLS que garantiza que $L(\beta)$ disminuye en cada iteración, de manera que asegura (3.20).

La realidad es que en la etapa (i) puede utilizarse cualquier estimador consistente con punto de ruptura alto que no requiera de una escala previa para ser equivariante, pero como en la práctica se utilizan en general S-estimadores, los definimos así.

Es inmediato verificar que los MM-estimadores son equivariantes.

El siguiente lema nos garantiza la Fisher-consistencia de \mathbf{T}_{MM} , la prueba se puede ver en la página 34 de Yohai (1985).

Lema 3.4 *Sea (\mathbf{x}', y) un vector aleatorio que sigue el modelo (2.1) con (2.5). Supongamos que F_0 satisface (2.4) y que $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 1$. Sea ρ una ρ -función acotada. Entonces*

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

tiene un único mínimo en $\beta = \beta_0$.

3.4.1 Robustez

A continuación damos un teorema de Yohai (1985) que garantiza la continuidad de \mathbf{T}_{MM} en la distribución nominal H_0 siempre que \mathbf{T}_S lo sea. Ver la Observación 3.1 para obtener condiciones para la continuidad de \mathbf{T}_S en H_0 .

Teorema 3.6 *Sea (\mathbf{x}', y) un vector aleatorio que sigue el modelo (2.1) con (2.5) y distribución H_0 dada por (2.3). Supongamos que F_0 satisface (2.4) y que $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 0.5$. Entonces si \mathbf{T}_S es continuo en H_0 , \mathbf{T}_{MM} también lo es.*

El siguiente teorema de Yohai (1985) da una cota para el punto de ruptura finito aditivo (2.10) de MM-estimadores.

Teorema 3.7 *Sean $\mathbf{z}_1 = (\mathbf{x}'_1, y_1), \dots, \mathbf{z}_n = (\mathbf{x}'_n, y_n)$ observaciones i.i.d del modelo (2.1) donde los errores cumplen (2.5). Sean $Z_n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ y $k^*(\mathbf{X})$ como en (2.9) y supongamos que $c_n = k^*(\mathbf{X})/n < 0.5$. Sean $\hat{\beta}_{MM,n}$ y $\hat{\beta}_{S,n}$ los MM y S-estimadores basados en Z_n respectivamente. Entonces*

$$\mathbf{AFBP}(\hat{\beta}_{MM,n}, Z_n) \geq \max(\mathbf{AFBP}(\hat{\beta}_{S,n}, Z_n), \frac{1 - 2c_n}{2 - 2c_n}).$$

Notemos que si \mathbf{X} tiene rango completo tendremos que $c_n = (p-1)/n$, luego si $\mathbf{AFBP}(\hat{\beta}_{S,n}, Z_n)$ es aproximadamente 0.5 cuando n es grande, $\mathbf{AFBP}(\hat{\beta}_{MM,n}, Z_n)$ también.

3.4.2 Propiedades asintóticas

El siguiente teorema de Yohai (1985) da condiciones para garantizar la consistencia de $\hat{\beta}_{MM}$,

Teorema 3.8 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5) y distribución H_0 dada por (2.3). Supongamos que F_0 satisface (2.4) y que

$$\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 0.5.$$

Supongamos además que la sucesión de S -estimadores basados en la muestra anterior $(\hat{\beta}_{S,n})_n$ es fuertemente consistente para β_0 . Entonces la sucesión de MM-estimadores basados en la muestra anterior $(\hat{\beta}_{MM,n})_n$ es fuertemente consistente para β_0 .

El siguiente teorema de Yohai (1985) da condiciones para garantizar la normalidad asintótica de $\hat{\beta}_{MM}$.

Teorema 3.9 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) donde los errores cumplen (2.5), y distribución H_0 dada por (2.3). Supongamos que F_0 satisface (2.4) y que G_0 es tal que $E_{G_0}(\|\mathbf{x}\|^2) < \infty$, $\mathbf{V}_{\mathbf{x}} = E_{G_0}(\mathbf{x}\mathbf{x}')$ es no singular y

$$\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 0.5.$$

Supongamos además que ρ_1 es continuamente diferenciable y eventualmente constante. Entonces si la sucesión de S -estimadores basados en la muestra anterior $(\hat{\beta}_{S,n})_n$ es fuertemente consistente para β_0 se cumple que

$$\sqrt{n}(\hat{\beta}_{MM,n} - \beta_0) \rightarrow^d N(\mathbf{0}, \sigma_0^2 \frac{E_{F_0}(\psi_1(\frac{u}{\sigma_0})^2)}{E_{F_0}(\psi_1'(\frac{u}{\sigma_0}))^2} \mathbf{V}_{\mathbf{x}}^{-1}),$$

donde $\sigma_0 = s(\beta_0, H_0) = \lim s_n(r(\hat{\beta}_{S,n}))$ c.s.

Notemos que entonces la eficiencia de $\hat{\beta}_{MM}$ no depende de G_0 y que está dada por (3.11).

En particular, podemos elegir $\rho_0 = \rho_{c_0}^B$ con $c_0 = 1.54$ para garantizar un punto de ruptura asintótico de 0.5 para la escala y que σ_0 coincida con el desvío standard de los errores bajo normalidad. De esta manera, si tomamos $\rho_1 = \rho_{c_1}^B$ con $c_1 \geq c_0$, podemos elegir c_1 tan grande como deseemos para obtener una eficiencia bajo normalidad tan alta como queramos. Recordemos que dimos valores de c_1 para alcanzar una eficiencia bajo normalidad predeterminada en la Tabla 1.

En Henning (1995) se dan cotas para el máximo sesgo asintótico de MM-estimadores. Lamentablemente, a medida que aumentamos c_1 también aumenta el máximo sesgo asintótico, como se muestra en la Sección 5.9 de Maronna, Martin y Yohai (2006). Por lo tanto, es importante elegir la eficiencia de manera tal de tener controlado el sesgo. En la Sección 5.5 de Maronna, Martin y Yohai (2006) se recomienda elegir una eficiencia de 0.85.

3.5 REWLS

Ninguno de los estimadores descriptos hasta ahora puede alcanzar eficiencia 1 bajo normalidad de los errores y al mismo tiempo mantener un punto de ruptura alto y máximo sesgo chico. A continuación, describimos brevemente un estimador presentado en Gervini y Yohai (2002) que cumple con estas tres condiciones.

El estimador, llamado REWLS (robust and efficient weighted least squares), es un estimador de mínimos cuadrados con pesos computado a partir de un estimador inicial $\hat{\beta}_0$ con punto de ruptura alto, pero en vez de eliminar los valores con residuos más grandes que cierto k fijo como en el WLSE de Rousseeuw y Leroy (1987), el procedimiento guarda un número N de observaciones (\mathbf{x}'_i, y_i) correspondientes a los valores más pequeños de $t_i = r_i(\hat{\beta}_0)/\hat{\sigma}$, donde $\hat{\sigma}$ es una escala auxiliar y N depende de los datos como detallamos a continuación. Este N tiene la propiedad de que bajo normalidad de los errores $N/n \rightarrow^n 1$, de manera que la cantidad de observaciones eliminadas tiende a cero, obteniendo así eficiencia 1.

Llamemos F_0 a la distribución de u_i/σ_0 bajo el modelo (2.1) donde los errores cumplen (2.5). Sean $t_{(i)}$ con $i = 1, \dots, n$ los estadísticos de orden de t_i . Sea $\eta = F_0^{-1}(\gamma)$ donde γ es un valor alto, por ejemplo $\gamma = 0.95$. Definamos

$$\begin{aligned} i_0 &= \min(i : t_{(i)} \geq \eta). \\ q &= \min_{i \geq i_0} \left(\frac{i-1}{F_0(t_{(i)})} \right). \\ N &= [q]. \end{aligned}$$

Entonces el estimador REWLS es el estimador de mínimos cuadrados calculado con las observaciones que corresponden a $t_{(i)}$ con $i \leq N$.

La idea intuitiva es considerar como potenciales outliers a las observaciones cuyo $t_{(i)}$ es no sólo más grande que un valor predeterminado, sino también lo suficientemente más grande que el correspondiente estadístico de orden de una muestra de F_0 . Notemos que si los datos contienen outliers, en un Q-Q plot normal de los $t_{(i)}$ contra los respectivos cuantiles de F_0 algunos $t_{(i)}$ grandes van a aparecer arriba de la línea de la identidad. Lo que hace el procedimiento es eliminar observaciones con $t_{(i)}$ grande hasta que el Q-Q plot de los restantes quede por debajo de la línea de la identidad, al menos para valores grandes de $|t_{(i)}|$. Como nos interesan sólo las colas de la distribución, consideramos solo los valores más grandes que cierto η .

Más precisamente, sea para $N \leq n$ F_N la distribución empírica de $t_{(i)}$ con $i = 1, \dots, N$

$$F_N(t) = \frac{\#\{t_{(i)} \leq t\}}{N}.$$

Se sigue entonces que

$$F_N(t) = \frac{i-1}{N} \text{ para } t_{(i-1)} \leq t < t_{(i)}$$

y por lo tanto, cada $t \in [t_{(i-1)}, t_{(i)})$ es un α_i -cuantil de F_N con

$$\alpha_i = \frac{i-1}{N}.$$

Por otro lado el α_i -cuantil de F_0 es $F_0^{-1}(\alpha_i)$. Buscamos entonces un N tal que para $i_0 \leq i \leq N$ el α_i -cuantil de F_N no sea mayor que el de F_0 , es decir

$$\text{para } i \in [i_0, N] \quad t \in [t_{(i-1)}, t_{(i)}) \implies t \leq F_0^{-1}\left(\frac{i-1}{N}\right) \iff F_0(t) \leq \frac{i-1}{N}.$$

Como F es continua, esto último implica que

$$F_0(t_{(i)}) \leq \frac{i-1}{N} \quad \text{para } i \in [i_0, N]. \quad (3.21)$$

Además como

$$i > N \implies \frac{i-1}{N} \geq 1 > F_0(t) \quad \forall t,$$

la restricción $i \leq N$ puede ser eliminada en (3.21).

Por lo tanto obtenemos que (3.21) es equivalente a

$$N \leq \frac{i-1}{F_0(t_{(i)})} \quad \text{para } i \geq i_0 \iff N \leq q.$$

Como queremos N lo más grande posible con $N \leq q$, tomamos $N = [q]$.

Gervini y Yohai (2002) muestran en su trabajo que bajo condiciones muy generales sobre $\hat{\beta}_0$ y $\hat{\sigma}$ y sin importar la velocidad de convergencia de $\hat{\beta}_0$, estos estimadores alcanzan el punto de ruptura máximo para estimadores equivariantes y eficiencia asintótica igual a 1 cuando los errores tienen distribución normal.

4 Propuesta

En esta sección, presentamos una nueva familia de estimadores para la regresión lineal que llamaremos estimadores de escala robusta acotada. La idea intuitiva es la siguiente:

Dadas $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5), calculamos $\hat{\beta}_{\mathbf{S},n}$ un S-estimador basado en la muestra anterior. Si la escala robusta de los residuos de $\hat{\beta}_{MC,n}$ está cerca de la escala robusta que minimiza $\hat{\beta}_{\mathbf{S},n}$, utilizamos $\hat{\beta}_{MC,n}$. Caso contrario utilizamos un MM-estimador con una eficiencia tan alta como podamos, manteniendo la escala de sus residuos cerca de la de $\hat{\beta}_{\mathbf{S},n}$. De esta forma la escala de $\hat{\beta}_{\mathbf{S},n}$ funciona como un control de calidad de la muestra; cuando la muestra no contiene outliers se utilizará en general $\hat{\beta}_{MC,n}$.

A continuación precisamos estas ideas.

Definición 4.1 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5). Sea $\hat{\beta}_{\mathbf{S},n}$ el S-estimador basado en la muestra anterior que minimiza la escala s_n calculada utilizando como ρ – función $\rho_{c_0}^B$ con cierto c_0 y b en el lado derecho de (3.3). Llamemos $\hat{\beta}_{c,n}$ al MM-estimador calculado utilizando $\rho_1 = \rho_c^B$. Sean $\delta_n \geq 0$ y $c_1 \geq c_0$. Definimos entonces

$$c^*(c_0, c_1, b, \delta_n) = \sup\{c_0 \leq c \leq c_1 : s_n(r(\hat{\beta}_{c,n})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{\mathbf{S},n}))\}.$$

Notemos que como $\hat{\beta}_{c_0,n} = \hat{\beta}_{\mathbf{S},n}$, $c^*(c_0, c_1, b, \delta_n)$ está bien definido. Cuando esté claro, o no sea importante la elección de c_0 , c_1 , b y δ_n , nos referiremos a $c^*(c_0, c_1, b, \delta_n)$ simplemente como c^* .

Ahora estamos en condiciones de definir los nuevos estimadores.

Definición 4.2 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5). Sea $c^* = c^*(c_0, c_1, b, \delta_n)$ como en la definición 4.1. Definimos entonces el estimador de escala robusta acotada basado en la muestra anterior como:

- $\hat{\beta}_{MC,n}$ si $s_n(r(\hat{\beta}_{MC,n})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{\mathbf{S},n}))$.
- $\hat{\beta}_{c^*,n}$ caso contrario.

En cualquier caso, llamaremos al estimador de escala robusta acotada $\hat{\beta}_{ERA,n}$.

Es inmediato verificar, utilizando la equivariancia de $\hat{\beta}_{c,n}$, $\hat{\beta}_{MC,n}$ y de $\hat{\beta}_{\mathbf{S},n}$, que $\hat{\beta}_{ERA,n}$ es equivariante.

El siguiente lema nos da cierta intuición de por qué la familia de estimadores que proponemos será altamente eficiente.

Lema 4.1 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5). Sea H_n la distribución empírica de la muestra y supongamos

que $\sup_{\|\beta\|=1} P_{H_n}(\beta' \mathbf{x} = 0) < 1 - b$. En las condiciones de la definición 4.1 tomando $c_1 = \infty$ tenemos que

$$c^* = \infty \implies s_n(r(\hat{\beta}_{MC,n})) \leq (1 + \delta_n) s_n(r(\hat{\beta}_{S,n})).$$

Demostración. Para agilizar la notación llamemos $s_0 = s_n(r(\hat{\beta}_{S,n}))$.

Como $c^* = \infty$ existe $(c_k)_{k \geq 1} \subset \mathbb{R}_{>0}$ tal que $c_k \rightarrow \infty$ y se cumple que

$$s_n(r(\hat{\beta}_{c_k,n})) \leq (1 + \delta_n) s_0 \quad \forall k \in \mathbb{N}.$$

Entonces en virtud del Lema 3.1 podemos suponer que existe $M > 0$ tal que

$$\|\hat{\beta}_{c_k,n}\| \leq M \quad \forall k \in \mathbb{N}. \quad (4.1)$$

Por lo tanto, podemos suponer sin pérdida de generalidad que existe $\hat{\beta} \in \mathbb{R}^p$ tal que

$$\hat{\beta}_{c_k,n} \rightarrow \hat{\beta}.$$

Veamos ahora que $\hat{\beta} = \hat{\beta}_{MC,n}$.

Por la definición de $\hat{\beta}_{c_k,n}$ tenemos que

$$\sum_{i=1}^N \mathbf{x}_i \frac{c_k^2}{6} \psi_{c_k}^B \left(\frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right) = 0.$$

Por lo tanto

$$\left\| \sum_{i=1}^N \mathbf{x}_i \left(\frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right) \right\| = \left\| \sum_{i=1}^N \mathbf{x}_i \left[\left(\frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right) - \frac{c_k^2}{6} \psi_{c_k}^B \left(\frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right) \right] \right\|$$

Ahora, para cada $i = 1, \dots, n$

$$\begin{aligned} \left| \frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} - \frac{c_k^2}{6} \psi_{c_k}^B \left(\frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right) \right| &= \\ \left| \frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right| \left| 1 - \left(1 - \left(\frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0 c_k} \right)^2 \right)^2 \mathbf{I}_{\left| \frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right| \leq c_k} \right|. \end{aligned} \quad (4.2)$$

Por (4.1) existe un $L > 0$ tal que

$$\left| \frac{y_i - \hat{\beta}'_{c_k,n} \mathbf{x}_i}{s_0} \right| \leq L$$

para todo k y para todo i .

Como $c_k \rightarrow \infty$ podemos encontrar k_0 tal que $k \geq k_0 \implies c_k > L$. Entonces si $k \geq k_0$ (4.2) es menor o igual que

$$L \left[1 - \left(1 - \left(\frac{L}{c_k} \right)^2 \right)^2 \right] \rightarrow^k 0.$$

Luego

$$\lim_k \sum_{i=1}^N \mathbf{x}_i \left(\frac{y_i - \hat{\beta}'_{c_k, n} \mathbf{x}_i}{s_0} \right) = \sum_{i=1}^N \mathbf{x}_i \left(\frac{y_i - \hat{\beta}' \mathbf{x}_i}{s_0} \right) = 0.$$

Se sigue que $\hat{\beta}$ cumple con las ecuaciones normales y por lo tanto $\hat{\beta} = \hat{\beta}_{MC, n}$. Sea ahora $0 < \epsilon < 1$. Como $\hat{\beta}_{c_k, n} \rightarrow^k \hat{\beta}_{MC, n}$, tenemos que

$$\left| r_i(\hat{\beta}_{c_k, n}) \right| \rightarrow^k \left| r_i(\hat{\beta}_{MC, n}) \right| \text{ para todo } i = 1, \dots, n.$$

Entonces existe k_1 tal que $k \geq k_1$ implica que

$$(1 - \epsilon) \left| r_i(\hat{\beta}_{MC, n}) \right| \leq \left| r_i(\hat{\beta}_{c_k, n}) \right| \text{ para todo } i = 1, \dots, n.$$

Aplicando (3.4) y (3.5) obtenemos que

$$(1 - \epsilon) s_n(r(\hat{\beta}_{MC, n})) \leq s_n(r(\hat{\beta}_{c_k, n})) \leq (1 + \delta_n) s_0 \text{ si } k \geq k_1.$$

Pero entonces

$$(1 - \epsilon) s_n(r(\hat{\beta}_{MC, n})) \leq (1 + \delta_n) s_0 \quad \forall 0 < \epsilon < 1.$$

Se sigue que

$$s_n(r(\hat{\beta}_{MC, n})) \leq (1 + \delta_n) s_0$$

que es lo que queríamos probar. ■

Es fácil ver, haciendo un razonamiento parecido al anterior que si c^* es finito, entonces

$$s_n(r(\hat{\beta}_{c^*, n})) \leq (1 + \delta_n) s_0.$$

4.1 Robustez

Como consecuencia inmediata del Teorema 3.2 tendremos que sin importar como elijamos (δ_n) , $\mathbf{RBP}(\hat{\beta}_{ERA, n})$ alcanza la cota dada en (2.8) siempre que elijamos

$b = \left(\left\lceil \frac{n - k^* - 1}{2} \right\rceil + \gamma \right) / n$ en la definición 4.1.

4.2 Propiedades asintóticas

Todas las demostraciones de los resultados de esta Sección se encuentran en el apéndice.

El siguiente teorema tiene como corolario la consistencia de $\hat{\beta}_{ERA,n}$.

Teorema 4.1 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5) y distribución H_0 dada por (2.3). Supongamos que F_0 cumple (2.4) y que $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 1 - b$. Sea $(\hat{\beta}_{1,n})_n$ una sucesión cualquiera de estimadores equivariantes basados en la muestra. Supongamos que se cumple que

$$s_n(r(\hat{\beta}_{1,n})) \rightarrow s(\beta_0, H_0) \text{ c.s.}$$

donde $s(\beta_0, H_0)$ es tal que

$$E_{H_0}(\rho \left(\frac{y - \beta'_0 \mathbf{x}}{s(\beta_0, H_0)} \right)) = b$$

y $s_n(r(\beta))$ es la escala que minimizan los S -estimadores, calculada con ρ una ρ -función continuamente diferenciable y eventualmente constante y b en el lado derecho de (3.3).

Entonces $(\hat{\beta}_{1,n})_n$ es fuertemente consistente para β_0 .

Corolario 4.2 En las condiciones del Teorema 4.1, supongamos que $(\hat{\beta}_{S,n})_n$, la sucesión de S -estimadores que minimizan $s_n(r(\beta))$, es fuertemente consistente para β_0 . Supongamos además que en la definición de $(\hat{\beta}_{ERA,n})_n$, tomamos $\delta_n \rightarrow 0$. Entonces $(\hat{\beta}_{ERA,n})_n$ es fuertemente consistente para β_0 .

Notemos que el resultado de consistencia que obtuvimos asume que el modelo lineal que cumplen las observaciones no tiene una ordenada al origen. De cualquier manera, en el estudio de simulación incluimos algunos casos con ordenada al origen, ya que en las aplicaciones prácticas los modelos suelen incluirla.

El siguiente teorema prueba que existe una manera de elegir δ_n de manera tal que $\hat{\beta}_{ERA,n}$ coincide asintóticamente con $\hat{\beta}_{MC,n}$.

Teorema 4.3 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d que cumplen el modelo (2.1) con (2.5) y distribución H_0 dada por (2.3). Sea $(\hat{\beta}_S)_n$ la sucesión de S -estimadores que minimizan s_n . Supongamos que $(\hat{\beta}_{MC,n})_n$ y $(\hat{\beta}_S)_n$ son fuertemente consistentes para β_0 y que $s_n(r(\hat{\beta}_{MC,n}))$ y $s_n(r(\hat{\beta}_S))$ son asintóticamente normales. Sea $(\delta_n) \subset \mathbb{R}_{\geq 0}$ tal que $\delta_n \sqrt{n} \rightarrow \infty$.

Entonces

$$P(\hat{\beta}_{ERA,n} = \hat{\beta}_{MC,n}) \rightarrow 1.$$

y además $\hat{\beta}_{ERA,n}$ tiene la misma distribución asintótica que $\hat{\beta}_{MC,n}$.

Notemos que si estamos en las condiciones del teorema anterior, existe $\mathbf{V}_x = E_{G_0}(\mathbf{x}\mathbf{x}')$ y es inversible, por (2.6) $\hat{\beta}_{ERA,n}$ será asintóticamente normal. Además

si F_0 es normal $\hat{\beta}_{ERA,n}$ tendrá eficiencia asintótica 1. Sin embargo si F_0 tiene colas más pesadas, por ejemplo con distribución t de Student, el hecho de que $\hat{\beta}_{ERA,n}$ coincida asintóticamente con $\hat{\beta}_{MC,n}$ nos indica que $\hat{\beta}_{ERA,n}$ no será eficiente respecto del estimador de máxima verosimilitud, ya que $\hat{\beta}_{MC,n}$ no lo es. Las simulaciones nos mostraron que no parece ser posible calibrar los parámetros de un $\hat{\beta}_{ERA,n}$ cuando $(\delta_n) \subset \mathbb{R}_{\geq 0}$ es tal que $\delta_n \sqrt{n} \rightarrow \infty$ de manera tal de tener un eficiencia por encima del 70% cuando los regresores tienen distribución t de Student.

Para obtener condiciones para garantizar la normalidad asintótica de $s_n(r(\hat{\beta}_{MC,n}))$ y de $s_n(r(\hat{\beta}_{S,n}))$, consultar la Sección 6.3 de Huber (1981) y el Teorema 6 de Fasano, Maronna, Sued y Yohai (2012).

4.3 Simulación

A la hora de calcular un estimador de escala robusta acotada utilizamos el paquete de Estadística Robusta de R para calcular los S-estimadores y MM-estimadores necesarios.

Fijados los parámetros c_0, c_1, b , y δ_n , para calcular $\hat{\beta}_{ERA,n}$ de acuerdo a la definición 4.2 primero verificamos si $\hat{\beta}_{MC,n}$ cumple que

$$s_n(r(\hat{\beta}_{MC,n})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{S,n})),$$

y de ser así tomamos $\hat{\beta}_{ERA,n} = \hat{\beta}_{MC,n}$.

Si $\hat{\beta}_{MC,n}$ no cumple con esta condición debemos estimar $c^*(c_0, c_1, b, \delta_n)$.

En un principio probamos hacer una búsqueda en una grilla entre c_0 y c_1 con cierto paso $h > 0$. Notamos entonces que para que las estimaciones resultantes sean buenas se debe tomar $h \leq 0.1$ y que en ese caso la búsqueda resulta muy costosa computacionalmente, ya que cada paso de la misma requiere calcular un MM-estimador y su escala.

Decidimos entonces utilizar el siguiente procedimiento:

- Primero verificamos si

$$s_n(r(\hat{\beta}_{c_1,n})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{S,n})),$$

en cuyo caso $c^*(c_0, c_1, b, \delta_n) = c_1$.

- Caso contrario calculamos $c' = c'(c_0, c_1, b, \delta_n)$ definido por

$$s_n(r(\hat{\beta}_{c',n})) = (1 + \delta_n)s_n(r(\hat{\beta}_{S,n})) \text{ con } c_0 \leq c' \leq c_1$$

utilizando el algoritmo de Brent para resolver ecuaciones no lineales y utilizamos $\hat{\beta}_{c',n}$ como estimador.

La experimentación con los dos métodos nos mostró que para los casos de interés para nuestro estudio de simulación el algoritmo de Brent converge rápidamente a $c'(c_0, c_1, b, \delta_n)$ y que $\hat{\beta}_{c',n}$ coincide con $\hat{\beta}_{c^*,n}$, donde $c^* = c^*(c_0, c_1, b, \delta_n)$ se estima haciendo una búsqueda en una grilla entre c_0 y c_1 con $h = 0.1$.

De cualquier manera, todos los resultados teóricos que obtuvimos para los estimadores de escala robusta acotada se satisfacen reemplazando en la definición 4.2. $\hat{\beta}_{c^*,n}$ por $\hat{\beta}_{c,n}$ donde c es tal que

$$s_n(r(\hat{\beta}_{c,n})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{\mathbf{S},n})).$$

Para el estudio de simulación asumiremos el modelo (2.1) con (2.5), y en algunos casos incluiremos una ordenada al origen. Como todos los estimadores que estudiamos son equivariantes, podemos suponer sin pérdida de generalidad que $\beta_0 = \mathbf{0}$. Compararemos la robustez y eficiencia para muestras finitas de los siguientes estimadores: MM con 85% de eficiencia asintótica que es el estimador recomendado en Maronna, Martin y Yohai (2006), S-estimador con punto de ruptura por reemplazo máximo, y un estimador de escala robusta acotada que llamaremos ERA calculado con punto de ruptura por reemplazo máximo, tomando $\delta_n = 0.6 \left(\frac{n}{p}\right)^{-0.6}$ donde p es el tamaño del vector de regresores, $c_0 = 1.54$ y $c_1 = 7$. Elegimos estos parámetros para el estimador de escala robusta acotada porque las simulaciones nos mostraron que brindan el mejor balance entre robustez y eficiencia. Notemos que este estimador no está en las condiciones del Teorema 4.2 y por lo tanto no conocemos su distribución asintótica.

Para estudiar la eficiencia de los estimadores cuando los errores tienen distribución normal, consideraremos los siguientes ocho escenarios:

- En los primeros tres, todas las coordenadas de \mathbf{x} son independientes con distribución $N(0, 1)$, $U(0, 1)$ y t de Student con cuatro grados de libertad respectivamente.
- En las siguientes tres, incluimos una ordenada al origen: (x_2, \dots, x_p) es como antes y $x_1 = 1$.
- En los últimos dos incluimos un término cuadrático: (x_1, \dots, x_{p-2}) es como antes, $x_{p-1} = 1$ y $x_p = x_1^2$, con distribución $N(0, 1)$ y $U(0, 1)$.

Tomamos $p = 5, 10$ y 20 y $n = Kp$ con $K = 5, 10$ y 20 .

Para cada n, p y escenario para las variables regresoras calculamos la eficiencia para muestras finitas bajo errores normales standard. Además, calculamos la eficiencia para muestras finitas bajo errores con distribución de Student con tres y cinco grados de libertad y regresores independientes con distribución normal standard.

Para medir la robustez de los estimadores contaminamos los datos de la siguiente manera. Para una tasa de contaminación $\varepsilon \in (0, 1)$ sea $m = \lfloor n\varepsilon \rfloor$. Entonces para $i \leq n-m$, generamos (\mathbf{x}'_i, y_i) de acuerdo al modelo (2.1) con (2.5), errores normales standard y regresores independientes normales standard. Para $i > n-m$ tomamos $\mathbf{x}_i = (x_0, 0, \dots, 0)'$ y $y_i = x_0 k$. El efecto de esta contaminación será entonces arrastrar a los estimadores hacia $(k, 0, \dots, 0)$. Tomamos $x_0 = 5$, k en un grilla de 0.5 a 10 con paso 0.2 y $\varepsilon = 0.1$. Para cada caso damos una tabla en la que se muestra el máximo y mínimo sesgo alcanzado por cada estimador.

El numero total de replicaciones de Monte Carlo es $N_{rep} = 500$ en todos los casos. Dado un estimador $\hat{\beta}$, llamemos $\hat{\beta}_k$ a la k -ésima replica de Montecarlo. Medimos entonces el rendimiento del estimador $\hat{\beta}$ utilizando el error cuadrático medio:

$$ECM(\hat{\beta}) = \frac{\sum_{i=1}^{500} \|\hat{\beta}_k\|^2}{500}.$$

4.3.1 Eficiencia

Con el fin de resumir los resultados del estudio de la eficiencia para muestras finitas de los estimadores con errores normales, para cada combinación (p, n) y para cada estimador tomamos la máxima y mínima eficiencia sobre los ocho escenarios con respecto al estimador de máxima verosimilitud.

Los resultados se presentan en la Tabla 2.

p	n	Máxima Eficiencia			Mínima Eficiencia		
		$MM85$	S	ERA	$MM85$	S	ERA
5	25	0.711	0.328	0.82	0.64	0.253	0.742
	50	0.795	0.303	0.951	0.7	0.273	0.88
	100	0.837	0.276	0.98	0.762	0.258	0.943
10	50	0.742	0.285	0.781	0.63	0.261	0.74
	100	0.814	0.274	0.938	0.77	0.253	0.914
	200	0.833	0.363	1	0.806	0.26	0.971
20	100	0.732	0.285	0.776	0.66	0.26	0.743
	200	0.798	0.258	0.948	0.79	0.246	0.937
	400	0.847	0.255	1	0.765	0.237	0.993

Tabla 2: Eficiencias de los estimadores con errores normales.

Notemos que en algunos casos, por ejemplo $p = 5$ y $n = 25$ o $p = 10$ y $n = 50$, la eficiencia mínima para muestras finitas del MM-estimador esta bien por debajo de su eficiencia asintótica. Los dos casos que mencionamos corresponden para $p = 5$ y $n = 25$ a regresores con distribución t de Student y regresores normales con término cuadrático en el caso $p = 10$ y $n = 50$.

Por otro lado, ERA tiene una eficiencia mínima del 74%.

La Tabla 3 muestra las eficiencias de los estimadores bajo errores con distribución t de Student respecto del estimador de máxima verosimilitud.

$g.l.$	p	n	$MM85$	ERA	LS	S
3	5	25	0.827	0.898	0.625	0.413
		50	0.889	0.847	0.615	0.419
		100	0.952	0.801	0.573	0.458
	10	50	0.848	0.867	0.562	0.368
		100	0.942	0.844	0.502	0.426
		200	0.943	0.8	0.520	0.457
	20	100	0.861	0.936	0.624	0.394
		200	0.941	0.875	0.554	0.429
		400	0.960	0.755	0.524	0.472
5	5	25	0.812	0.837	0.803	0.419
		50	0.903	0.904	0.83	0.439
		100	0.913	0.902	0.84	0.469
	10	50	0.834	0.882	0.878	0.375
		100	0.885	0.936	0.836	0.437
		200	0.927	0.862	0.820	0.38
	20	100	0.856	0.882	0.865	0.349
		200	0.901	0.930	0.848	0.338
		400	0.934	0.865	0.837	0.362

Tabla 3: Eficiencia de los estimadores para errores con distribución de Student con tres y cinco grados de libertad

Estos resultados nos muestran que para errores con distribución de Student, el estimador ERA es más eficiente que el MM para tamaños de muestra pequeños, y lo opuesto sucede para tamaños de muestra grandes. De cualquier manera, vemos que en este caso ambos estimadores son altamente eficientes.

4.3.2 Robustez

La Tabla 4 muestra el máximo y mínimo error cuadrático medio de los estimadores estudiados.

ε	p	n	Máximo ECM			Mínimo ECM		
			$MM85$	S	ERA	$MM85$	S	ERA
0.1	5	25	0.868	1.41	0.860	0.36	0.71	0.33
		50	0.576	0.85	0.61	0.147	0.27	0.137
		100	0.432	0.706	0.444	0.068	0.16	0.063
	10	50	1.35	2.28	1.36	0.341	0.815	0.325
		100	0.621	1.17	0.657	0.143	0.371	0.135
		200	0.489	0.757	0.514	0.068	0.173	0.064
	20	100	1.33	2.36	1.44	0.326	0.838	0.32
		200	0.747	1.23	0.811	0.143	0.39	0.136
		400	0.488	0.659	0.516	0.062	0.172	0.061

Tabla 4: Errores cuadráticos medios de los estimadores para regresores normales, con datos contaminados.

Vemos entonces que el precio que pagamos por el claro aumento en la eficiencia que muestra la Tabla 2, es un aumento de a lo sumo 10% en el máximo error cuadrático medio respecto del MM-estimador.

La Figura 1 compara el ECM de los tres estimadores para $p = 5$ y $n = 25$. En este caso, el comportamiento del MM85 y del ERA es similar, alcanzando el ERA un máximo menor.

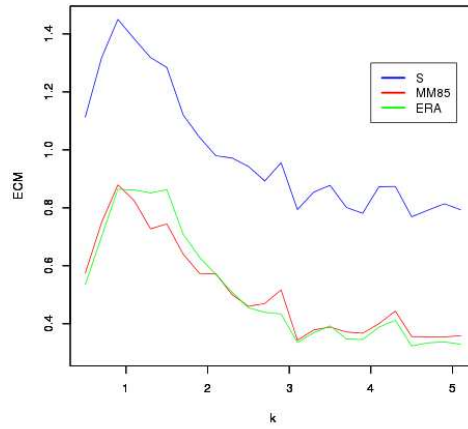


Figura 1: ECM de los tres estimadores, en el caso $p = 5$ y $n = 25$.

5 Apéndice

A continuación damos todas las demostraciones pendientes.

Demostración del Teorema 3.1

Para agilizar la notación llamemos

$$\begin{aligned} s_1 &= g_1^{-1} \left(\frac{b-\epsilon}{1-\epsilon}, 0 \right) \\ \tilde{s}_1 &= (1+\delta)s_1 = (1+\delta)g_1^{-1} \left(\frac{b-\epsilon}{1-\epsilon}, 0 \right) \\ \hat{s}_1 &= \frac{s_1}{(1+\delta)} = \frac{g_1^{-1} \left(\frac{b-\epsilon}{1-\epsilon}, 0 \right)}{(1+\delta)}. \end{aligned}$$

Sea $\epsilon < \min(b, 1-b)$, y $V_\epsilon(H_0)$ un entorno de contaminación de H_0 . Veamos primero que $\mathbf{B}_\epsilon(\mathbf{T}) \leq \alpha_1$. Para esto basta ver que si $H \in V_\epsilon(H_0)$ se cumple que

$$\|\beta\| > \alpha_1 \implies s(\beta, H) > (1+\delta)s(0, H).$$

Sea entonces β con $\|\beta\| > \alpha_1$. Como g es estrictamente creciente respecto de $\|\beta\|$ tenemos que

$$g(\tilde{s}_1, \|\beta\|) > g(\tilde{s}_1, \alpha_1) = \frac{b}{1-\epsilon}$$

y como g es continua y estrictamente decreciente respecto de s podemos encontrar un $s_2 > \tilde{s}_1$ tal que

$$g(s_2, \|\beta\|) > \frac{b}{1-\epsilon}.$$

Entonces

$$E_H(\rho \left(\frac{y - \beta' \mathbf{x}}{s_2} \right)) \geq (1-\epsilon)g(s_2, \|\beta\|) > b,$$

lo que implica que $s_2 \leq s(\beta, H)$.

Por otro lado si $s > s_1$

$$E_H(\rho \left(\frac{y}{s} \right)) \leq (1-\epsilon)g(s, 0) + \epsilon \leq (1-\epsilon)g(s_1, 0) + \epsilon = b.$$

Luego $s > s(\mathbf{0}, H) \forall s > s_1$, de donde obtenemos $s_1 \geq s(\mathbf{0}, H)$.

Juntando todo, tenemos que

$$s(\beta, H) \geq s_2 > \tilde{s}_1 = (1+\delta)s_1 \geq (1+\delta)s(\mathbf{0}, H).$$

Por lo tanto

$$\mathbf{B}_\epsilon(\mathbf{T}) \leq \alpha_1.$$

Veamos a continuación que $\alpha_2 \leq \mathbf{B}_\epsilon(\mathbf{T})$.

Sea $0 < c < \alpha_2$ y sea β^* con $\|\beta^*\| = c$. Consideremos $(\lambda_n)_n \subset \mathbb{R}$ tal que $\lambda_n \rightarrow \infty$ y definamos $\mathbf{x}_n = \lambda_n \beta^*$, $y_n = \beta^{*'} \mathbf{x}_n$. Sea H_n^* la distribución concentrada en (\mathbf{x}_n, y_n) y $H_n = (1 - \epsilon)H_0 + \epsilon H_n^*$.

Luego para probar que $\alpha_2 \leq \mathbf{B}_\epsilon(\mathbf{T})$ nos basta ver que

$$\sup_n \|\mathbf{T}(H_n)\| \geq c.$$

Supongamos por el absurdo que esto no sucede. Entonces eventualmente pasando a una subsucesión, que por conveniencia seguiremos llamando H_n , tendremos que $\mathbf{T}(H_n) = \tilde{\beta}_n$ con

$$\lim_n \tilde{\beta}_n = \tilde{\beta}$$

y

$$\|\tilde{\beta}\| < \|\beta^*\| = c.$$

Se sigue entonces que

$$\lim_n |y_n - \tilde{\beta}' \mathbf{x}_n| = \lim_n \lambda_n (\|\beta^*\|^2 - \beta^* \tilde{\beta}) = \infty.$$

Por otro lado

$$E_{H_n} \rho \left(\frac{y - \tilde{\beta}' \mathbf{x}}{s} \right) = (1 - \epsilon)g(s, \|\tilde{\beta}_n\|) + \epsilon \rho \left(\frac{y_n - \tilde{\beta}' \mathbf{x}_n}{s} \right).$$

Tomemos ahora $s < s_1$, luego

$$\lim_n E_{H_n} \rho \left(\frac{y - \tilde{\beta}' \mathbf{x}}{s} \right) \geq (1 - \epsilon)g(s, 0) + \epsilon > (1 - \epsilon)g(s_1, 0) + \epsilon = b.$$

Por lo tanto

$$\lim_n s(\tilde{\beta}_n, H_n) \geq s \quad \forall s < s_1.$$

Como g es continua y estrictamente decreciente respecto de s tenemos que

$$(1 - \epsilon)g(\hat{s}_1, c) < (1 - \epsilon)g(\hat{s}_1, \alpha_2) = b$$

y que existe $s_3 < \hat{s}_1$ tal que

$$(1 - \epsilon)g(s_3, c) < b.$$

Entonces

$$E_{H_n} \left(\rho \left(\frac{y - \beta^{*'} \mathbf{x}}{s_3} \right) \right) = (1 - \epsilon)g(s_3, c) < b$$

de donde deducimos que

$$s(\beta^*, H_n) \leq s_3 < \hat{s}_1 = \frac{s_1}{(1 + \delta)} \leq \frac{\lim_n s(\tilde{\beta}_n, H_n)}{(1 + \delta)}$$

que es absurdo, ya que $\mathbf{T}(H_n) = \beta_n$ es tal que $s(\tilde{\beta}_n, H_n)/(1 + \delta)$ es mínimo. Por lo tanto

$$\alpha_2 \leq \mathbf{B}_\epsilon(\mathbf{T}).$$

Para completar la prueba basta ver que

$$\epsilon \uparrow \min(b, 1 - b) \implies \alpha_2 = g_2^{-1} \left(\frac{g_1^{-1} \left(\frac{b - \epsilon}{1 - \epsilon}, 0 \right)}{(1 + \delta)}, \frac{b}{1 - \epsilon} \right) \longrightarrow \infty.$$

Tomemos $b \leq 0.5$, para que $\min(b, 1 - b) = b$. Tenemos entonces que

$$\lim_{\epsilon \uparrow b} g_1^{-1} \left(\frac{b - \epsilon}{1 - \epsilon}, 0 \right) = \lim_{r \rightarrow 0} g_1^{-1}(r, 0) = \infty$$

de donde se sigue que

$$\lim_{\epsilon \uparrow b} g_2^{-1} \left(\frac{g_1^{-1} \left(\frac{b - \epsilon}{1 - \epsilon}, 0 \right)}{(1 + \delta)}, \frac{b}{1 - \epsilon} \right) = \lim_{r \rightarrow \infty} g_2^{-1} \left(r, \frac{b}{1 - \epsilon} \right) = \infty.$$

■

Para probar el Teorema 3.2 necesitaremos el siguiente lema, que aparece en Maronna, Martin y Yohai (2006).

Lema 5.1 *Sea $\mathbf{r}_N = (r_{N,1}, \dots, r_{N,n})$ un sucesión. Sean $C = \{i : |r_{N,i}| \rightarrow \infty\}$ y $D = \{i : |r_{N,i}| \text{ está acotada}\}$. Sea $s_n(r(\beta))$ un M -estimador de escala calculado con una ρ -función continua y acotada y sea b el lado derecho de (3.3). Entonces*

- Si $\#C > nb, s_n(\mathbf{r}_N) \rightarrow \infty$
- Si $\#D > n - nb, s_n(\mathbf{r}_N)$ está acotada

Demostración. (i) Supongamos que $s_n(\mathbf{r}_N)$ esta acotada. Entonces la definición de $s(\mathbf{r}_N)$ implica que

$$nb \geq \lim_N \sum_{i \in C} \rho \left(\frac{r_{N,i}}{s_n(\mathbf{r}_N)} \right) = \#C > nb$$

que es absurdo.

(ii) Supongamos que $s_n(\mathbf{r}_N) \rightarrow \infty$. Entonces si $i \in D$ $\frac{r_{N,i}}{s_n(\mathbf{r}_N)} \rightarrow 0$, lo que implica que

$$nb = \lim_N \sum_{i=1}^n \rho \left(\frac{r_{N,i}}{s_n(\mathbf{r}_N)} \right) = \lim_N \sum_{i \notin D} \rho \left(\frac{r_{N,i}}{s_n(\mathbf{r}_N)} \right) \leq n - \#D < nb$$

que es absurdo. ■

Demostración del Teorema 3.2

(i) Aunque suponemos que nuestro modelo no tiene una ordenada al origen, se puede hacer una demostración enteramente análoga cuando si la tiene. Vamos a probar que $m^* \leq nb$. Sea $m > nb$. Sea $C \subset \{1, \dots, n\}$ con $\#C = m$. Sea $\mathbf{x}_0 \in \mathbb{R}^p$ con $\|\mathbf{x}_0\| = 1$. Dada una sucesión $(\mathbf{X}_n, \mathbf{y}_n)$, definamos para $\beta \in \mathbb{R}^{p+1}$

$$\mathbf{r}_N(\beta) = \mathbf{y}_n - \mathbf{X}_n \beta.$$

Tomemos $(\mathbf{X}_n, \mathbf{y}_n)$ de la siguiente manera:

$$(\mathbf{x}_{N,i}, y_{N,i}) = \begin{cases} (N\mathbf{x}_0, N^2) & \text{si } i \in C \\ (\mathbf{x}_i, y_i) & \text{si } i \notin C \end{cases}$$

Veamos que $\hat{\beta}_N$, el estimador basado en $(\mathbf{X}_n, \mathbf{y}_n)$, no puede estar acotado. Supongamos por el absurdo que $\hat{\beta}_N$ está acotado. Entonces $|r_{N,i}| \rightarrow \infty$ cuando $i \in C$. Luego, la parte (i) del Lema 5.1 implica que $s_n(\mathbf{r}_N(\hat{\beta}_N)) \rightarrow \infty$. Como $\frac{nb}{m} < 1 = \rho(\infty)$, tenemos que existe un único γ tal que

$$\rho\left(\frac{1}{\gamma}\right) = \frac{nb}{m}.$$

Vamos a probar que

$$\frac{1}{N^2} s_n(\mathbf{r}_N(\hat{\beta}_N)) \rightarrow^N \gamma.$$

De hecho

$$nb = \sum_{i \notin C} \rho \left(\frac{y_i - \hat{\beta}'_N \mathbf{x}_i}{s_n(\mathbf{r}_N(\hat{\beta}_N))} \right) + \sum_{i \in C} \rho \left(\frac{N^2 - N\hat{\beta}'_N \mathbf{x}_0}{s_n(\mathbf{r}_N(\hat{\beta}_N))} \right).$$

El primer sumando tiende a cero pues $s_n(\mathbf{r}_N(\hat{\beta}_N)) \rightarrow \infty$, y el segundo es

$$m\rho \left(\frac{1 - N^{-1}\hat{\beta}'_N \mathbf{x}_0}{N^{-2}s_n(\mathbf{r}_N(\hat{\beta}_N))} \right).$$

El numerador de la fracción tiene a uno. Si una subsucesión de $N^{-2}s_n(\mathbf{r}_N(\hat{\beta}_N))$ tiene un límite t (posiblemente infinito), debe cumplir $\rho(\frac{1}{t}) = \frac{nb}{m}$ que es lo que queríamos probar.

Llamemos $\eta = (1 + 2\delta_n) / (2 + 2\delta_n)$ y definamos $\beta_N^* = \mathbf{x}_0 N \eta$. Entonces

$$r_{N,i}(\beta_N^*) = N^2(1 - \eta) \text{ si } i \in C$$

y

$$r_{N,i}(\beta_N^*) = y_i - \mathbf{x}'_0 \mathbf{x}_i N \eta \text{ si } i \notin C.$$

Como $\#\{i : r_{N,i}(\beta_N^*) \rightarrow \infty\} = n$, la parte (i) del Lema 5.1 implica que $s_n(\mathbf{r}_N(\beta_N^*)) \rightarrow \infty$. Procediendo como antes vemos que

$$\frac{1}{N^2} s_n(\mathbf{r}_N(\beta_N^*)) \rightarrow^N \frac{\gamma}{2(1 + \delta_n)}$$

lo que implica que

$$s_n(\mathbf{r}_N(\beta_N^*)) < \frac{s_n(\mathbf{r}_N(\hat{\beta}_N))}{(1 + \delta_n)}$$

para N suficientemente grande, que es absurdo por (3.13).

(ii) La demostración es idéntica a la que aparece en la página 165 de Maronna, Martin, Yohai (2006). ■

Para probar el Teorema 4.1 necesitaremos de los siguientes siguientes lemas.

Lema 5.2 *Sea ρ una ρ -función continuamente diferenciable. Sea H_0 dada por (2.3) con F_0 que satisface (2.4) y $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = \mathbf{0}) < 1 - b$. Sea $s(\beta, H_0)$ definido para $\beta \in \mathbb{R}^p$ por*

$$E_{H_0}(\rho\left(\frac{y - \beta \mathbf{x}}{s(\beta, H_0)}\right)) = b. \quad (5.1)$$

Entonces $s(\beta, H_0)$ tiene un único mínimo en $\beta = \beta_0$.

Se puede encontrar una demostración del lema anterior en Yohai y Zamar (1988).

Probamos ahora una leve generalización de un lema de Yohai (1985).

Lema 5.3 *Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d del modelo (2.1) con (2.5) y distribución H_0 dada por (2.3). Supongamos que $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = \mathbf{0}) < 1 - b$. Supongamos que ρ es una ρ -función continua y acotada y que $(\hat{\beta}_{1,n})_n$ es una sucesión de estimadores que converge casi seguro a $\beta^* \in \mathbb{R}^p$. Entonces $s_n(r(\hat{\beta}_{1,n}))$ definida por (3.3) converge casi seguro a $s(\beta^*, H_0)$ definido como en (5.1).*

Demostración. Sea $0 < \epsilon < s(\beta^*, H_0)$ afirmamos que entonces podemos encontrar un $\delta > 0$ tal que

$$\begin{aligned} E_{H_0}\left(\inf_{\|\beta - \beta^*\| \leq \delta} \rho\left(\frac{y - \beta' \mathbf{x}}{s(\beta^*, H_0) - \epsilon}\right)\right) &\geq b + \delta \\ y \\ E_{H_0}\left(\sup_{\|\beta - \beta^*\| \leq \delta} \rho\left(\frac{y - \beta' \mathbf{x}}{s(\beta^*, H_0) + \epsilon}\right)\right) &\leq b - \delta. \end{aligned} \quad (5.2)$$

Supongamos por el absurdo que no. Entonces existe un $\tilde{\epsilon}$ tal que $0 < \tilde{\epsilon} < s(\boldsymbol{\beta}^*, H_0)$ y para todo $\delta > 0$

$$E_{H_0}\left(\inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leq \delta} \rho\left(\frac{y - \boldsymbol{\beta}'\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right)\right) < b + \delta \quad (5.3)$$

o

$$E_{H_0}\left(\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leq \delta} \rho\left(\frac{y - \boldsymbol{\beta}'\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) + \tilde{\epsilon}}\right)\right) > b - \delta. \quad (5.4)$$

Tomemos $(\delta_n) \subset \mathbb{R}_{>0}$ tal que $\delta_n \rightarrow 0$. Entonces para cada $n \in \mathbb{N}$ se cumple (5.3) o (5.4), pero alguna de las dos se cumple infinitas veces. Supongamos que se cumple (5.3) infinitas veces y tomemos una subsucesión de (δ_n) , que por comodidad llamaremos igual, tal que se cumpla (5.3) para todo n . Si se cumple (5.4) infinitas veces el argumento es enteramente análogo.

Ahora, como ρ es una función continua y $\{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leq \delta_n\}$ es compacto para todo n , tendremos que para cada n existe un $\boldsymbol{\theta}_n$ tal que

$$\inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leq \delta_n} \rho\left(\frac{y - \boldsymbol{\beta}'\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right) = \rho\left(\frac{y - \boldsymbol{\theta}_n'\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right).$$

Además como $\boldsymbol{\theta}_n \rightarrow \boldsymbol{\beta}^*$, se sigue que

$$\rho\left(\frac{y - \boldsymbol{\theta}_n'\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right) \rightarrow \rho\left(\frac{y - \boldsymbol{\beta}^*\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right).$$

Recordando que ρ es acotada y aplicando el Teorema de Convergencia Mayorada obtenemos que

$$\lim E_{H_0}\left(\inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leq \delta_n} \rho\left(\frac{y - \boldsymbol{\beta}'\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right)\right) = E_{H_0}\left(\rho\left(\frac{y - \boldsymbol{\beta}^*\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right)\right) \leq b.$$

Como ρ es monótona no decreciente y $s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon} < s(\boldsymbol{\beta}^*, H_0)$ finalmente llegamos a que

$$b = E_{H_0}\left(\rho\left(\frac{y - \boldsymbol{\beta}^*\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0)}\right)\right) \leq E_{H_0}\left(\rho\left(\frac{y - \boldsymbol{\beta}^*\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right)\right) \leq b.$$

Pero entonces

$$E_{H_0}\left(\rho\left(\frac{y - \boldsymbol{\beta}^*\mathbf{x}}{s(\boldsymbol{\beta}^*, H_0) - \tilde{\epsilon}}\right)\right) = b$$

que es absurdo ya que por definición $s(\boldsymbol{\beta}^*, H_0)$ es el ínfimo de los $s > 0$ tales que

$$E_{H_0}\left(\rho\left(\frac{y - \boldsymbol{\beta}^*\mathbf{x}}{s}\right)\right) = b$$

Tenemos probado entonces (5.2).

Se sigue entonces de la Ley de los Grandes Números que

$$\lim_n \frac{1}{n} \sum_{i=1}^n \inf_{\|\beta - \beta^*\| \leq \delta} \rho \left(\frac{y_i - \beta' \mathbf{x}_i}{s(\beta^*, H_0) - \epsilon} \right) \geq b + \delta \text{ c.s.}$$

$$\lim_n \frac{1}{n} \sum_{i=1}^n \sup_{\|\beta - \beta^*\| \leq \delta} \rho \left(\frac{y_i - \beta' \mathbf{x}_i}{s(\beta^*, H_0) + \epsilon} \right) \leq b - \delta \text{ c.s.}$$

Entonces como

$$\lim_n \hat{\beta}_{1,n} = \beta^* \text{ c.s.}$$

tenemos que

$$\lim_n \frac{1}{n} \sum_{i=1}^n \rho \left(\frac{y_i - \hat{\beta}_{1,n}' \mathbf{x}_i}{s(\beta^*, H_0) - \epsilon} \right) \geq b + \delta \text{ c.s.}$$

$$\lim_n \frac{1}{n} \sum_{i=1}^n \rho \left(\frac{y_i - \hat{\beta}_{1,n}' \mathbf{x}_i}{s(\beta^*, H_0) + \epsilon} \right) \leq b - \delta \text{ c.s.}$$

Luego por la monotonía de ρ , con probabilidad uno existe un n_0 tal que $s(\beta^*, H_0) - \epsilon \leq s_n(r(\hat{\beta}_{1,n})) \leq s(\beta^*, H_0) + \epsilon$, que es lo que queríamos probar. ■

El siguiente lema está probado en la página 36 de Yohai y Zamar (1988).

Lema 5.4 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ observaciones i.i.d del modelo (2.1) con (2.5) y distribución H_0 dada por (2.3) con $\beta_0 = \mathbf{0}$. Supongamos que F_0 cumple (2.4) y que $\sup_{\|\beta\|=1} P_{G_0}(\beta' \mathbf{x} = 0) < 1 - b$. Sea $s_n(r(\beta))$ como en (3.3), con ρ una ρ - función continuamente diferenciable y eventualmente constante. Entonces

$$\forall k_2 \exists k_1 \text{ tal que } \lim_n \inf_{\|\beta\| \geq k_1} s_n(r(\beta)) \geq k_2 \text{ c.s.}$$

Demostración del Teorema 4.1

En virtud de la equivariancia de los estimadores involucrados, podemos suponer sin pérdida de generalidad que $\beta_0 = \mathbf{0}$.

Llamemos (Ω, \mathcal{F}, P) al espacio de probabilidad en el que estan definidas nuestras observaciones.

Veamos que con probabilidad uno $(\hat{\beta}_{1,n})_n$ es acotada.

Como

$$s_n(r(\hat{\beta}_{1,n})) \rightarrow s(\mathbf{0}, H_0) \text{ c.s.}$$

tenemos que existe $N_0 \in F$ tal que $P(N_0) = 0$ y si $\omega \notin N_0$ se cumple

$$s_n(r(\hat{\beta}_{1,n}))(\omega) \rightarrow s(\mathbf{0}, H_0).$$

Por el Lema 5.4 tenemos que para cada $m \in \mathbb{N}$ existe un $k(m) \in \mathbb{N}$ y $N_m \in \mathcal{F}$ tal que $P(N_m) = 0$ y si $\omega \notin N_m$ se cumple que

$$\lim_n \inf_{\|\beta\| \geq k(m)} s_n(r(\beta))(\omega) \geq m.$$

Sea $N = \bigcup_{m=0}^{\infty} N_m$. Es claro entonces que $N \in \mathcal{F}$ y $P(N) = 0$.

Sea $\omega_0 \notin N$, veamos que entonces $(\hat{\beta}_{1,n}(\omega_0))_n$ es acotada.

Supongamos por el absurdo que no.

Notemos que como $\omega_0 \notin N$ vale que

$$s_n(r(\hat{\beta}_{1,n}))(\omega_0) \rightarrow s(\mathbf{0}, H_0) \quad (5.5)$$

y para todo $m \in \mathbb{N}$ se tiene

$$\lim_n \inf_{\|\beta\| \geq k(m)} s_n(r(\beta))(\omega_0) \geq m. \quad (5.6)$$

De (5.5) se sigue que existe un $m_0 \in \mathbb{N}$ tal que para todo $n \in \mathbb{N}$

$$s_n(r(\hat{\beta}_{1,n}))(\omega_0) < \frac{m_0}{2}.$$

Como $(\hat{\beta}_{1,n}(\omega_0))_n$ no esta acotada, tiene que existir una subsucesión $(\hat{\beta}_{1,n_j}(\omega_0))_j$ tal que para todo $j \in \mathbb{N}$

$$\|\hat{\beta}_{1,n_j}(\omega_0)\| \geq k(m_0).$$

Pero entonces para todo $j \in \mathbb{N}$

$$\inf_{\|\beta\| \geq k(m_0)} s_{n_j}(r(\beta))(\omega_0) \leq s_{n_j}(r(\hat{\beta}_{1,n_j}))(\omega_0) < \frac{m_0}{2}.$$

Se sigue de (5.6) que

$$m_0 \leq \lim \inf_{\|\beta\| \geq k(m_0)} s_{n_j}(r(\beta))(\omega_0) \leq \frac{m_0}{2}$$

que es absurdo. Concluimos que con probabilidad uno $(\hat{\beta}_{1,n})_n$ es acotada.

Veamos finalmente que $\hat{\beta}_{1,n} \rightarrow^{c.s} \mathbf{0}$. Para esto, basta probar que con probabilidad uno toda subsucesión $(\hat{\beta}_{1,n_j})_j$ de $(\hat{\beta}_{1,n})_n$ tiene una subsucesión $(\hat{\beta}_{1,n_{j_k}})_k$ que converge a $\mathbf{0}$.

Sea entonces $(\hat{\beta}_{1,n_j})_j$ una subsucesión de $(\hat{\beta}_{1,n})_n$. Como $(\hat{\beta}_{1,n})_n$ esta acotada con probabilidad uno, $(\hat{\beta}_{1,n_j})_j$ también. Por lo tanto con probabilidad uno existe $(\hat{\beta}_{1,n_{j_k}})_k$ una subsucesión de $(\hat{\beta}_{1,n_j})_j$ convergente, llamemos $\tilde{\beta} \in \mathbb{R}^p$ a su límite. Tenemos en virtud del Lema 5.3 que

$$s_{n_{j_k}}(r(\hat{\beta}_{1,n_{j_k}})) \rightarrow^{c.s} s(\tilde{\beta}, H_0).$$

Ademas como

$$s_n(r(\hat{\beta}_{1,n})) \rightarrow^{c.s.} s(\mathbf{0}, H_0)$$

tenemos que $s(\tilde{\beta}, H_0) = s(\mathbf{0}, H_0)$. Por el Lema 5.2 sabemos que $s(\beta, H_0)$ tiene un único mínimo en $\beta = \mathbf{0}$, por lo tanto debe ser que $\tilde{\beta} = \mathbf{0}$ y esto termina la demostración. ■

Demostración del Corolario

Por el Teorema 4.1, nos basta probar que

$$s_n(r(\hat{\beta}_{ERA,n})) \rightarrow s(\beta_0, H_0) \text{ c.s.}$$

Tenemos que para todo n

$$s_n(r(\hat{\beta}_{S,n})) \leq s_n(r(\hat{\beta}_{ERA,n})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{S,n})).$$

Como por hipótesis

$$\hat{\beta}_{S,n} \rightarrow \beta_0 \text{ c.s.},$$

obtenemos del Lema 5.3 que

$$s_n(r(\hat{\beta}_{S,n})) \rightarrow s(\beta_0, H_0) \text{ c.s.}$$

y por lo tanto

$$s_n(r(\hat{\beta}_{ERA,n})) \rightarrow s(\beta_0, H_0) \text{ c.s.}$$

■

Demostración Teorema 4.2

Para probar que

$$P(\hat{\beta}_{ERA,n} = \hat{\beta}_{MC,n}) \rightarrow 1 \tag{5.7}$$

nos basta con mostrar que

$$P(s_n(r(\hat{\beta}_{MC,n})) \leq (1 + \delta_n)s_n(r(\hat{\beta}_{S,n}))) \rightarrow 1. \tag{5.8}$$

Para probar esto, notemos que como

$$\sqrt{n}(s_n(r(\hat{\beta}_{MC,n})) - s(\beta_0, H_0)) = O_P(1)$$

y

$$\sqrt{n}(s_n(r(\hat{\beta}_{S,n})) - s(\beta_0, H_0)) = O_P(1),$$

restando estas dos igualdades y multiplicando por $n^{-1/2}$ obtenemos que

$$s_n(r(\hat{\beta}_{MC,n})) - s_n(r(\hat{\beta}_{S,n})) = n^{-1/2}O_P(1).$$

Por lo tanto

$$s_n(r(\hat{\beta}_{MC,n})) - (1 + \delta_n)s_n(r(\hat{\beta}_{S,n})) = n^{-1/2}O_P(1) - \delta_n s_n(r(\hat{\beta}_{S,n})),$$

de donde deducimos que

$$n^{1/2}(s_n(r(\hat{\beta}_{MC,n})) - (1 + \delta_n)s_n(r(\hat{\beta}_{S,n}))) = O_P(1) - \delta_n n^{1/2}s_n(r(\hat{\beta}_{S,n})). \quad (5.9)$$

Finalmente (5.9) implica que

$$n^{1/2}(s_n(r(\hat{\beta}_{MC,n})) - (1 + \delta_n)s_n(r(\hat{\beta}_{S,n}))) \xrightarrow{P} -\infty$$

que a su vez implica (5.8), que es lo que queríamos probar.

Probemos ahora que

$$n^{1/2}(\hat{\beta}_{ERA,n} - \hat{\beta}_{MC,n}) = o_p(1). \quad (5.10)$$

Para probar (5.10) tenemos que mostrar que dado $\varepsilon > 0$ y $\delta > 0$ existe un n_0 tal que

$$P(\|n^{1/2}(\hat{\beta}_{ERA,n} - \hat{\beta}_{MC,n})\| > \delta) \leq \varepsilon \quad (5.11)$$

para todo $n \geq n_0$.

Sean entonces $\varepsilon > 0$ y $\delta > 0$. Tomemos n_0 tal que

$$P(\hat{\beta}_{ERA,n} = \hat{\beta}_{MC,n}) > 1 - \varepsilon \text{ para todo } n \geq n_0.$$

Entonces si $n \geq n_0$ tenemos que

$$P(\|n^{1/2}(\hat{\beta}_{ERA,n} - \hat{\beta}_{MC,n})\| > \delta) \leq P(\hat{\beta}_{ERA,n} \neq \hat{\beta}_{MC,n}) \leq \varepsilon$$

que es lo que queríamos probar.

Finalmente notemos que

$$n^{1/2}(\hat{\beta}_{ERA,n} - \beta_0) = n^{1/2}(\hat{\beta}_{ERA,n} - \hat{\beta}_{MC,n}) + n^{1/2}(\hat{\beta}_{MC,n} - \beta_0) = o_p(1) + n^{1/2}(\hat{\beta}_{MC,n} - \beta_0),$$

que implica que $\hat{\beta}_{ERA,n}$ y $\hat{\beta}_{MC,n}$ tienen la misma distribución asintótica.

■

Bibliografía

- [1] Bondell, H.D. and Stefanski, L.A. (2013). Efficient Robust Regression via Two-Stage Generalized Empirical Likelihood. *J.Amer.Statist.Assoc.* **108** 644-655.
- [2] Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. *Festschrift for Erich L. Lehmann* 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*, **18** 4 1289-1309.
- [4] Hampel, F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887-1896.
- [5] Hennig, C. (1995). Efficient high-breakdown-point estimator in robust regression: Which function to choose?. *Statistics and Decisions*. **13** 221-241.
- [6] Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.
- [7] Huber, P.J. (1973). Robust regression. *Ann. Statist.* **1** 799-821
- [8] Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*. Wiley, New York.
- [9] Maronna, R.A. and Yohai, V.J. (1981). Asymptotic behaviour of general M-estimates of regression and scale with random carriers. *Z. Wahrsch. verw. Gebiete*. **58** 7-20.
- [10] Martin, R.D., Yohai, V.J. and Zamar, R.H. (1989). A min-max bias robust regression. *Ann. Statist.* **17** 1608-1630.
- [11] Rousseuw, P.J. (1984). Least median of squares regression. *J.Amer.Statist.Assoc.* **79** 871-880.
- [12] Rousseuw, P.J. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
- [13] Rousseuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. *Lecture Notes in Statist.* **26** 256-272. Springer, New York.
- [14] Salibián-Barrera, M., and Yohai, V.J. (2005). A fast algorithm for S-regression estimates. *Journal of Computational and Graphical Statistics*. **15** 414-427.
- [15] Serfling, R. (2009). Inequalities relating addition and replacement type finite sample breakdown points. *International Journal of Statistical Sciences*. **9** 71-83
- [16] Yohai, V.J. (1985). High breakdown point and high efficiency robust estimates for regression. Technical Report No. 66, Dept. Statistics, Univ. Washington, Seattle.

- [17] Yohai, V.J. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15** 642-656.
- [18] Yohai, V.J. and Zamar, R. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *J.Amer.Statist.Assoc.* **83** 406-413.