Estimadores robustos y eficientes para el modelo de Regresión Lineal

Gervini, Daniel

1999

Tesis presentada para obtener el grado de Doctor en Ciencias Matemáticas de la Universidad de Buenos Aires

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Estimadores robustos y eficientes para el modelo de Regresión Lineal

Autor: Lic. Daniel Gervini
Director: Dr. Víctor J. Yohai

Lugar de trabajo: Depto. de Matemática, Fac. de Cs. Exactas y Naturales, UBA

Tesis presentada para optar por el título de Doctor de la Universidad de Buenos Aires

1999

RESUMEN

Estimadores robustos y eficientes para el modelo de Regresión Lineal

En esta Tesis presentamos una nueva clase de estimadores (que llamaremos REWLS) para el modelo de Regresión Lineal. Son estimadores de mínimos cuadrados pesados, con pesos que se calculan de manera adaptiva a partir de la distribución empírica de los residuos de un estimador robusto inicial. Se demuestra que el punto de ruptura de los REWLS no es menor que el del estimador inicial, de modo que pueden alcanzar el punto de ruptura máximo 1/2. Para el caso particular del estimador de mínima mediana de cuadrados (LMS) como estimador inicial y pesos "hard rejection", se muestra numéricamente que los sesgos máximos del REWLS para contaminaciones puntuales son prácticamente iguales los del LMS. Pero además, y esto constituye el aporte original de la Tesis, se demuestra que bajo el modelo los REWLS son asintóticamente equivalentes al estimador de mínimos cuadrados y entonces alcanzan la máxima eficiencia asintótica para el modelo de errores normales. En conclusión, los estimadores que proponemos logran alcanzar la *máxima* eficiencia asintótica bajo el modelo sin afectar las cualidades de robustez del estimador inicial.

PALABRAS CLAVE: Robustez; Regresión Lineal; Estimación eficiente; Mínimos Cuadrados Pesados.

ABSTRACT

Robust and efficient estimators for the Linear Regression model

In this Thesis we introduce a new class of estimators (that we will call REWLS) for the Linear Regression model. They are weighted least squares estimators, with weights adaptively computed from the empirical distribution of the residuals of some initial robust estimator. It is shown that the breakdown point of the REWLS is not smaller than the breakdown point of the initial estimator, so that they can attain the maximum 1/2 breakdown point. For the particular case of the least median of squares (LMS) as the initial estimator and hard rejection weights, it is shown that the maximum biases of the REWLS for pointmass contaminations are practically equal to those of the LMS. Moreover –and this is the original contribution of this Thesis– it is shown that the REWLS are asymptotically equivalent to the least squares estimator under the model and hence they attain the maximum asymptotic efficiency for the normal error model. To summarize, the estimators we propose attain the *maximum* asymptotic efficiency under the model with no damage to the robust qualities of the initial estimator.

KEY WORDS: Robustness; Linear Regression; Efficient estimation; Weighted Least Squares.

Indice

1. Introducción

En esta Tesis estudiaremos el problema de estimación puntual en el modelo de Regresión Lineal. En este modelo se tiene una muestra aleatoria $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, donde cada $\mathbf{x}_i$ es un vector con p variables explicativas e y_i es la correspondiente variable de respuesta, relacionada con $\mathbf{x}_i$ mediante la fórmula

$$y_i = \mathbf{x}_i'\boldsymbol{\theta} + u_i, \tag{1.1}$$

donde $\boldsymbol{\theta} \in \mathbb{R}^p$ es el parámetro desconocido que se desea estimar. Los errores $\{u_i\}$ son variables aleatorias i.i.d. no observables, con distribución $F_0(\cdot/\sigma)$ desconocida. El parámetro de escala $\sigma > 0$ también será desconocido pero lo consideraremos un parámetro de ruido, centrándonos en la estimación de $\boldsymbol{\theta}$. A la distribución de los errores absolutos $|u_i|$ la llamaremos $F_0^+(\cdot/\sigma)$. Las variables explicativas $\{\mathbf{x}_i\}$ se suponen no aleatorias; en caso de que fueran una muestra de una distribución G_0 en $\mathbb{R}^p$, supondremos que son estocásticamente independientes de los errores $\{u_i\}$ y todos los resultados asintóticos dados en la Sección 4 serán válidos (usando un argumento condicional).

En lo que sigue haremos las siguientes suposiciones acerca del modelo:

A1. F_0 es continua, estrictamente creciente y simétrica alrededor de cero.

A2. Sea X la matriz de diseño con filas $\mathbf{x}_1', \ldots, \mathbf{x}_n'$ y llamemos $\Sigma_n = (X'X)/n$. Entonces Σ_n es definida positiva y además la secuencia de los menores autovalores de Σ_n no tiende a cero.

En el problema de estimación así planteado, es sabido que el estimador de mínimos cuadrados (EMC) de $\boldsymbol{\theta}$ es el de máxima verosimilitud para $F_0 = \Phi$, la distribución normal estándar, y por lo tanto alcanza la mínima varianza asintótica. Teniendo esta propiedad de optimalidad y siendo quizá el más fácil de calcular, no es de extrañar que el EMC sea el estimador más utilizado en las aplicaciones del modelo lineal. Sin embargo, esta propiedad de optimalidad se ve empañada por la extrema sensibilidad que muestra el EMC ante la presencia de datos atípicos. De hecho, solamente *una* observación colocada suficientemente lejos del resto de los datos puede llevar al EMC más allá de cualquier cota, sin importar cuán grande sea el tamaño muestral n. Por lo tanto, dado que en la práctica ningún conjunto de datos se ajusta perfectamente a un modelo teórico, esta falta de estabilidad del EMC es un problema serio y deben buscarse estimadores alternativos. Estos estimadores deberían ser estables ante la presencia de outliers y a la vez eficientes para datos que se ajusten al modelo.

El primer estimador equivariante capaz de tolerar hasta la mitad de observaciones atípicas en la muestra sin irse a infinito (i.e., en alcanzar el punto de ruptura máximo de 1/2) fue el "least median of squares" (LMS) y fue propuesto por Rousseeuw (1984). Lamentablemente el LMS tiene una grave desventaja: su orden de convergencia bajo el modelo es $n^{-1/3}$ (ver Davies,1990), por lo que su eficiencia relativa al EMC es 0. Para obtener un estimador más eficiente, Rousseeuw y Leroy (1987) sugieren calcular un estimador de mínimos cuadrados pesados (EMCP) a partir del LMS. Su propuesta consiste en hallar primero el LMS y después calcular el EMC con aquellos datos cuyos residuos absolutos estandarizados sea menor que cierta constante, por ejemplo 2.5 si se supone que los errores son normales. Así se obtiene un estimador más eficiente, aunque He y Portnoy (1992) mostraron que el orden de convergencia de este EMCP es aun $n^{-1/3}$.

Los primeros estimadores que alcanzaron el máximo punto de ruptura junto con el orden de convergencia usual de $n^{-1/2}$ fueron los S-estimadores, propuestos por Rousseeuw y Yohai (1984). Sin embargo, estos estimadores no pueden alcanzar simultáneamente un alto punto de ruptura y una alta eficiencia asintótica bajo el modelo. Los primeros estimadores en lograr ambas cosas fueron los MM-estimadores (Yohai, 1987) seguidos por los τ-estimadores (Yohai y Zamar, 1988). Es importante mencionar, sin embargo, que estos estimadores nunca alcanzan la *máxima* eficiencia asintótica bajo el modelo.

El aporte original de esta Tesis consiste, justamente, en presentar una clase de estimadores que no sólo alcanzan el punto de ruptura máximo sino también la máxima eficiencia asintótica bajo el modelo. Estos estimadores son en realidad EMCP y por lo tanto los denominaremos REWLS (Robust and Efficient Weighted Least Squares).

2. Definición del estimador REWLS

Dados dos estimadores robustos iniciales, $\mathbf{T}_{0n}$ de $\boldsymbol{\theta}$ y s_n de σ, consideremos los residuos estandarizados

$$r_i = \left(y_i - \mathbf{x}_i'\mathbf{T}_{0n}\right)/s_n. \tag{2.1}$$

Cuando una observación $(\mathbf{x}_i, y_i)$ tenga un valor $|r_i|$ muy "grande", es natural sospechar que pueda ser un outlier y por lo tanto debe dársele un peso menor que a los demás datos o bien eliminársela completamente. Si $F_0 = \Phi$, un valor "grande" podría ser 2.5, por ejemplo. Con esto en mente, Rousseeuw y Leroy (1987) definen los pesos

$$w_i = \begin{cases} 1 & \text{si } |r_i| < 2.5 \\ 0 & \text{si } |r_i| \geq 2.5 \end{cases}$$

y luego calculan el EMCP $\mathbf{T}_{1n} = \left(X'WX\right)^{-1} X'WY$, donde $W = \operatorname{diag}\left(w_1, \ldots, w_n\right)$ e $Y = \left(y_1, \ldots, y_n\right)'$.

Aunque este estimador mejora la eficiencia asintótica del estimador inicial, es intuitivamente claro que no puede alcanzar la máxima eficiencia relativa, ya que la probabilidad de que un residuo absoluto sea mayor que 2.5 (o que cualquier otra constante) es pequeña pero no es cero; por lo tanto para n grande igual se descartarían algunas observaciones aunque no hubiera outliers en la muestra. De hecho, si $\mathbf{T}_{0n}$ es el LMS entonces T_{1n} sigue teniendo el mismo orden de convergencia $n^{-1/3}$, aunque la varianza asintótica sea menor.

Nuestra propuesta es similar a la de Rousseeuw, pero la diferencia fundamental radica en que usamos un valor de corte *adaptivo* y no fijo. Para calcularlo, definimos primero la función de distribución empírica de los residuos absolutos,

$$F_n^+(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(|r_i| \leq t\right).$$

Cuando los datos se ajustan al modelo (1.1), F_n^+ converge uniformemente a F_0^+ cuando $n \to \infty$. Por lo tanto, para detectar outliers podríamos comparar F_n^+ con F_0^+ (si la conociéramos). Un valor $F_n^+(t) < F_0^+(t)$ estaría indicando que la proporción de residuos cuyo valor absoluto excede t es mayor que la proporción teórica bajo el modelo, lo que podría deberse a la presencia de datos atípicos. Entonces podríamos estimar la proporción de outliers en la muestra mediante $\sup_{t\geq 0}\left\{\max\left\{F_0^+(t) - F_n^+(t), 0\right\}\right\}$. Sin embargo, en la práctica la distribución de los errores F_0 va a ser desconocida, por lo que vamos a tener que usar alguna distribución hipotética F. En la práctica se utilizará $F = \Phi$, aunque en realidad lo que se requerirá de F es que cumpla:

F1. La función de distribución F es continua, simétrica y con varianza finita.

Entonces, como medida de la proporción de outliers en la muestra definimos

$$d_n = \sup_{t\geq 0}\left\{\max\left\{F^+(t) - F_n^+(t), 0\right\}\right\} \tag{2.2}$$

con $F^+(t) = 2F(t) - 1$. Si llamamos $|r|_{(1)} \leq \dots \leq |r|_{(n)}$ a los residuos absolutos ordenados, entonces

$$d_n = \max_{1\leq i\leq n}\left\{\max\left\{F^+\left(|r|_{(i)}\right) - \frac{(i-1)}{n}, 0\right\}\right\}.$$

Ahora debemos darle un peso menor (o eliminar) las $\lfloor nd_n \rfloor$ observaciones con los residuos absolutos más grandes. De manera que un posible valor de corte sería

$$\bar{t}_n = \min\left\{t : F_n^+(t) \geq 1 - d_n\right\} = |r|_{(n-\lfloor nd_n\rfloor)}. \tag{2.3}$$

Sin embargo, estudios mediante simulación mostraron que d_n suele ser relativamente grande para n chico, aunque no haya outliers. Por lo tanto, para obtener una alta eficiencia para muestras chicas definiremos como valor de corte

$$t_n = \max\{\bar{t}_n, \eta\}, \tag{2.4}$$

donde η es el p-ésimo cuantil de F^+ para algún p cerca de uno. Si bien al aumentar el valor de corte de esta manera podríamos estar dejando en la muestra algunos outliers, los outliers más extremos (que son los que afectan seriamente al estimador bajo el modelo de errores normales) serán eliminados de todos modos.

Los pesos asignados a las observaciones tendrán la forma $w_i = w(r_i/t_n)$, donde la función w satisface:

W1. La función de peso $w(u)$ es no negativa, simétrica alrededor de cero, continua a derecha y no creciente para $u \geq 0$, $w(0) = 1$ y $w(u) = 0$ para $|u| \geq 1$.

Una vez calculados los pesos, definimos el REWLS como $\mathbf{T}_{1n} = (X'WX)^{-1} X'WY$. Eligiendo $\eta = 2.5$ en (2.4) obtendremos un estimador que es por lo menos tan eficiente como el EMCP de Rousseeuw y Leroy para muestras chicas.

Notemos que a partir de W1 se deduce que $w_i = 0$ si $|r_i| \geq t_n$, de modo que las observaciones con residuos muy grandes son completamente eliminadas. Esto hace que el REWLS sea robusto. Por otro lado, $|r_i|/t_n \to 0$ cuando $t_n \to \infty$, y como $w_i(0) = 1$ ninguna observación será eliminada asintóticamente. Esto hace que bajo el modelo el REWLS tenga eficiencia asintótica 1 respecto al EMC.

3. Robustez del REWLS

En esta sección estudiaremos el comportamiento del REWLS cuando los datos no se ajustan al modelo (1.1) con exactitud. Para eso supondremos que $\mathbf{x}_1, \ldots, \mathbf{x}_n$ es una muestra aleatoria de cierta distribución G_0 en $\mathbb{R}^p$ tal que

A2'. G_0 tiene segundos momentos finitos y $\Sigma = E_{G_0}\{\mathbf{x}\mathbf{x}'\}$ es definida positiva.

Llamaremos H_0 a la distribución conjunta de $(\mathbf{x}_i, y_i)$ cuando se cumple (1.1), $u_i \sim F_0(\cdot/\sigma)$ y $\mathbf{x}_i \sim G_0$. Dado $\varepsilon \in [0, 0.5)$, consideremos el entorno de H_0 dado por

$$\mathcal{H}_\varepsilon = \left\{H : H = (1-\varepsilon) H_0 + \varepsilon H^* \text{ y } H^* \text{ es una distribución en } \mathbb{R}^{p+1}\right\}. \tag{3.1}$$

Aunque $\mathcal{H}\varepsilon$ no es estrictamente un entorno en el sentido topológico del término, nos permite una interpretación intuitiva muy útil para nuestro problema: una distribución $H \in \mathcal{H}\varepsilon$ genera datos que con probabilidad $1-\varepsilon$ se ajustan al modelo (1.1) y con probabilidad ε son outliers. Por eso tomaremos siempre $\varepsilon < 0.5$ en (3.1), ya que la mayoría de los datos debe provenir de la distribución H_0 para que el término "outlier" tenga sentido.

Los estimadores de $\boldsymbol{\theta}$ que vamos a considerar en esta Tesis pueden escribirse como $\mathbf{T}(H_n)$, donde H_n es la función de distribución empírica de la muestra y $\mathbf{T}$ es un funcional definido sobre (un subconjunto de) el espacio de las funciones de distribución sobre $\mathbb{R}^{p+1}$. Supondremos que $\mathbf{T}$ es consistente Fisher y equivariante. Consistente Fisher significa que $\mathbf{T}(H_0) = \boldsymbol{\theta}$, o sea que $\mathbf{T}(H_n)$ es asintóticamente insesgado bajo el modelo (1.1) si $\mathbf{T}$ es continua. La equivariancia que se pide en el contexto de regresión lineal es que si $(\mathbf{x}, y)$ satisface (1.1) y se definen $y^* = (y + \mathbf{x}'\mathbf{b})/a$ y $\mathbf{x}^* = C'\mathbf{x}$, con $a \neq 0$, $\mathbf{b} \in \mathbb{R}^p$ y $C \in \mathbb{R}^{p \times p}$ no singular, entonces llamando H^* a la distribución de $(\mathbf{x}^*, y^*)$ se tiene que $\mathbf{T}(H^*) = a^{-1}C^{-1}\{\mathbf{T}(H_0) + \mathbf{b}\}$. Este requerimiento de equivariancia es muy natural, ya que $(\mathbf{x}^*, y^*)$ satisface (1.1) con $\boldsymbol{\theta}^* = a^{-1}C^{-1}(\boldsymbol{\theta} + \mathbf{b})$. Al estimador de escala asociado, $S(H_n)$, se le pedirá que sea consistente Fisher, equivariante por cambio de escala e invariante por traslaciones.

Como en general $\mathbf{T}(H) \neq \boldsymbol{\theta}$ para $H \neq H_0$, nos enfrentaremos al problema del sesgo cuando estemos muestreando de una distribución en $\mathcal{H}_\varepsilon$ con $\varepsilon > 0$. El sesgo asintótico de $\mathbf{T}$ en H se puede definir como

$$b_A(\mathbf{T}, H) = \left\{(\mathbf{T}(H) - \boldsymbol{\theta})' A(G_0)(\mathbf{T}(H) - \boldsymbol{\theta})\right\}^{1/2}$$

para un funcional $A(G_0)$ que sea equivariante por transformaciones afines. Tomaremos $A(G_0) = E_{G_0}\{\mathbf{x}\mathbf{x}'\}$ y entonces, dado que nosotros consideramos solamente estimadores equivariantes, a los efectos de estudiar el sesgo asintótico podemos suponer que $\Sigma = \mathrm{I}$ y $\boldsymbol{\theta} = \mathbf{0}$, de modo que $b_A(\mathbf{T}, H) = \|\mathbf{T}(H)\|$. También supondremos, para simplificar, que $\sigma = 1$.

Para estudiar la resistencia de un estimador $\mathbf{T}$ frente a las desviaciones del modelo permitidas en el entorno $\mathcal{H}_\varepsilon$ consideraremos el mayor sesgo posible de $\mathbf{T}(H)$ sobre $\mathcal{H}_\varepsilon$. Con ese fin definimos la *función de sesgo máximo*

$$\mathcal{B}_{\mathbf{T}}(\varepsilon) = \sup\{\|\mathbf{T}(H)\| : H \in \mathcal{H}_\varepsilon\}. \tag{3.2}$$

Un estimador se dirá *robusto* cuando $\lim_{\varepsilon \searrow 0} \mathcal{B}_{\mathbf{T}}(\varepsilon) = 0$, lo que implicará que $\mathbf{T}(H)$ se mantenga acotado sobre todo el entorno $\mathcal{H}_\varepsilon$ para algún $\varepsilon > 0$. Sin embargo, cualquier estimador equivariante explotará para ε suficientemente grande. Definimos entonces el punto de ruptura asintótico

$$\varepsilon^*_{\mathbf{T}} = \inf\{\varepsilon : \mathcal{B}_{\mathbf{T}}(\varepsilon) = \infty\},$$

que será la menor proporción de outliers que el estimador no puede aguantar. He y Simpson (1993) probaron que $\varepsilon^*_{\mathbf{T}} \leq 0.5$ para cualquier estimador equivariante.

Hallemos ahora la forma funcional del REWLS para una distribución $H \in \mathcal{H}\varepsilon$ con $\varepsilon < \varepsilon^*_{\mathbf{T}_0}$. Dados los estimadores iniciales $\mathrm{T}_0(H)$ y $S(H)$ de $\boldsymbol{\theta}$ y σ respectivamente, el residuo estandarizado es

$$r_H(\mathbf{x}, y) = (y - \mathbf{x}'\mathbf{T}_0(H))/S(H)$$

y la función de distribución del residuo estandarizado absoluto será

$$F_H^+(t) = P_H\{|r_H(\mathbf{x}, y)| \leq t\}.$$

Entonces las formas funcionales correspondientes a (2.2), (2.3) y (2.4) son

$$\begin{aligned} d_H &= \sup_{t\geq 0}\left\{\max\left\{F^+(t) - F_H^+(t), 0\right\}\right\} \\ \bar{t}_H &= \min\left\{t : F_H^+(t) \geq 1 - d_H\right\} \\ t_H &= \max\left\{\bar{t}_H, \eta\right\}. \end{aligned} \tag{3.3}$$

Los pesos tienen la forma $w_H(\mathbf{x}, y) = w(r_H(\mathbf{x}, y)/t_H)$ para una función w que satisface W1.

Entonces la forma funcional del REWLS será

$$\mathbf{T}_1(H) = \arg\min_{\mathbf{t}} E_H\left\{w_H(\mathbf{x}, y)(y - \mathbf{x}'\mathbf{t})^2\right\}. \tag{3.4}$$

Obsérvese que si definimos

$$\gamma(H) = \arg\min_{\mathbf{t}} E_H\left\{w_H(\mathbf{x}, y)(y - \mathbf{x}'\mathbf{T}_0(H) - \mathbf{x}'\mathbf{t})^2\right\} \tag{3.5}$$

se tendrá que

$$\mathbf{T}_1(H) = \mathbf{T}_0(H) + \gamma(H).$$

En el Apéndice A se demuestra que $E_H\left\{w_H(\mathbf{x}, y)(y - \mathbf{x}'\mathbf{T}_0(H))^2\right\}$ es finita para cualquier $H \in \mathcal{H}\varepsilon$ con $\varepsilon < \varepsilon^*_{\mathbf{T}_0}$ (ver la desigualdad (A.3)). Además, si $E_H\left\{w_H(\mathbf{x}, y)\|\mathbf{x}\|^2\right\} < \infty$ del Lema A.1 del Apéndice A se deduce que $E_H\{w_H(\mathbf{x}, y)\mathbf{x}\mathbf{x}'\}$ es inversible y entonces se tiene la expresión explícita

$$\gamma(H) = [E_H\{w_H(\mathbf{x}, y)\mathbf{x}\mathbf{x}'\}]^{-1} E_H\{w_H(\mathbf{x}, y)\mathbf{x}(y - \mathbf{x}'\mathbf{T}_0(H))\}. \tag{3.6}$$

Para obtener una expresión explícita de $\gamma(H)$ cuando $E_H\left\{w_H(\mathbf{x}, y)\|\mathbf{x}\|^2\right\} = \infty$, se procede de manera similar a la de Simpson y Yohai (1998, pág. 1155). Sea

$$\mathcal{L}(H) = \left\{\mathbf{v} \in \mathbb{R}^p : E_H\left[w_H(\mathbf{x}, y)|\mathbf{x}'\mathbf{v}|^2\right] < \infty\right\} \tag{3.7}$$

que es un subespacio vectorial de $\mathbb{R}^p$. Llamamos p_1 a la dimensión de $\mathcal{L}(H)$. Cuando $p_1 = 0$, se tiene que

$$E_H\left\{w_H(\mathbf{x}, y)(y - \mathbf{x}'\mathbf{T}_0(H) - \mathbf{x}'\mathbf{t})^2\right\} = \infty \text{ si } \mathbf{t} \neq \mathbf{0}$$

y entonces $\gamma(H) = 0$, es decir $\mathbf{T}_1(H) = \mathbf{T}_0(H)$. Cuando $p_1 > 0$, tomamos una matriz Γ_1 cuyas p_1 columnas formen una base ortogonal de $\mathcal{L}(H)$ y una matriz Γ_2 cuyas $p_2 = p - p_1$ columnas formen una base ortogonal de $\mathcal{L}(H)^{\perp}$. Sea $\Gamma = [\Gamma_1\ \Gamma_2]$, que es una matriz ortogonal de $p \times p$. Definimos $\mathbf{z} = \Gamma'\mathbf{x}$ y llamamos H^* a la distribución de $(\mathbf{z}, y)$. Por equivariancia se tiene que $\mathbf{T}_0(H^*) = \Gamma'\mathbf{T}_0(H)$ y que $w_{H^*}(\mathbf{z}, y) = w_H(\mathbf{x}, y)$. Entonces

$$E_H\left\{w_H(\mathbf{x}, y)\left(y - \mathbf{x}'\mathbf{T}_0(H) - \mathbf{x}'\mathbf{t}\right)^2\right\} = E_{H^*}\left\{w_{H^*}(\mathbf{z}, y)\left(y - \mathbf{z}'\mathbf{T}_0(H^*) - \mathbf{z}'\Gamma'\mathbf{t}\right)^2\right\}$$

de modo que si

$$\gamma(H^*) = \arg\min_{\mathbf{s}} E_{H^*}\left\{w_{H^*}(\mathbf{z}, y)\left(y - \mathbf{z}'\mathbf{T}_0(H^*) - \mathbf{z}'\mathbf{s}\right)^2\right\}$$

se tendrá que $\gamma(H) = \Gamma\gamma(H^*)$ y en consecuencia $\mathbf{T}_1(H) = \mathbf{T}_0(H) + \Gamma\gamma(H^*)$. Ahora bien, si llamamos $\mathbf{z}_1$ al vector de las primeras p_1 coordenadas de $\mathbf{z}$ y $\mathbf{z}_2$ al de las p_2 coordenadas restantes, y hacemos lo mismo con el vector $\mathbf{s}$, se tiene que

$$\begin{aligned} E_{H^*}\left\{w_{H^*}(\mathbf{z}, y)\|\mathbf{z}_1\|^2\right\} &< \infty \\ E_{H^*}\left\{w_{H^*}(\mathbf{z}, y)|\mathbf{s}_2'\mathbf{z}_2|^2\right\} &= \infty\ \forall \mathbf{s}_2 \neq \mathbf{0}. \end{aligned}$$

Por lo tanto

$$E_{H^*}\left\{w_{H^*}(\mathbf{z}, y)\left(y - \mathbf{z}'\mathbf{T}_0(H^*) - \mathbf{z}'\mathbf{s}\right)^2\right\} = \infty \text{ si } \mathbf{s}_2 \neq \mathbf{0}$$

así que tiene que ser

$$\gamma(H^*) = \begin{bmatrix} \gamma_1(H^*) \\ \mathbf{0} \end{bmatrix}$$

donde

$$\gamma_1(H^*) = \arg\min_{\mathbf{s}_1} E_{H^*}\left\{w_{H^*}(\mathbf{z}, y)\left(y - \mathbf{z}'\mathbf{T}_0(H^*) - \mathbf{z}_1'\mathbf{s}_1\right)^2\right\}.$$

Si llamamos $[E_{H^*}\{w_{H^*}(\mathbf{z}, y)\mathbf{z}_1\mathbf{z}_1'\}]^-$ a una pseudoinversa de $E_{H^*}\{w_{H^*}(\mathbf{z}, y)\mathbf{z}_1\mathbf{z}_1'\}$, se tendrá la expresión explícita

$$\begin{aligned} \gamma_1(H^*) &= [E_{H^*}\{w_{H^*}(\mathbf{z}, y)\mathbf{z}_1\mathbf{z}_1'\}]^- E_H\{w_{H^*}(\mathbf{z}, y)\mathbf{z}_1(y - \mathbf{z}'\mathbf{T}_0(H^*))\} \\ &= [E_H\{w_H(\mathbf{x}, y)\Gamma_1'\mathbf{x}\mathbf{x}'\Gamma_1\}]^- E_H\{w_H(\mathbf{x}, y)\Gamma_1'\mathbf{x}(y - \mathbf{x}'\mathbf{T}_0(H))\} \end{aligned}$$

y será

$$\mathbf{T}_1(H) = \mathbf{T}_0(H) + \gamma(H) = \mathbf{T}_0(H) + \Gamma_1\gamma_1(H^*). \tag{3.8}$$

De todas maneras, por definición $\mathbf{T}_1(H)$ cumple que

$$E_H\left\{w_H(\mathbf{x},y)\left(y-\mathbf{x}'\mathbf{T}_1(H)\right)^2\right\} \le E_H\left\{w_H(\mathbf{x},y)\left(y-\mathbf{x}'\mathbf{T}_0(H)\right)^2\right\} \tag{3.9}$$

y es esta desigualdad lo que se usa en la demostración del Teorema 3.1, no las expresiones explícitas del funcional.

Del Teorema 3.1 se deduce inmediatamente que el punto de ruptura del REWLS no es menor que el del estimador inicial. Esto muestra que, al menos desde el punto de vista del punto de ruptura, el ECMP que proponemos no empeora las cualidades de robustez del estimador inicial. El Teorema 3.2 muestra que el REWLS es robusto cuando $F = F_0$. La demostración de ambos teoremas se da en el Apéndice A. Para estos resultados necesitaremos algunas suposiciones:

B1. El estimador inicial es robusto, es decir $\lim_{\varepsilon\searrow 0}\mathcal{B}_{\mathbf{T}_0}(\varepsilon)=0$.

B2. El estimador inicial de escala es robusto, con punto de ruptura no menor que el de $\mathbf{T}_0$. Es decir, para $\varepsilon<\varepsilon^*_{\mathbf{T}_0}$ existen $\sigma_1(\varepsilon)>0$ y $\sigma_2(\varepsilon)<\infty$ tales que $\sigma_1(\varepsilon)\le S(H)\le\sigma_2(\varepsilon)$ para todo $H\in\mathcal{H}\varepsilon$, $\sigma_1(\varepsilon)\le\sigma\le\sigma_2(\varepsilon)$ y $\lim_{\varepsilon\searrow 0}\sigma_1(\varepsilon)=\lim_{\varepsilon\searrow 0}\sigma_2(\varepsilon)=\sigma$.

Teorema 3.1. *Si se cumplen F1, A1, A2', B1, B2 y W1, entonces existe una constante finita* $K=K(\varepsilon,H_0)$ *tal que* $\|\mathbf{T}_1(H)-\mathbf{T}_0(H)\|\le K$ *para toda* $H\in\mathcal{H}\varepsilon$ *con* $\varepsilon<\varepsilon^*_{\mathbf{T}_0}$.

Teorema 3.2. *Si se cumplen F1, A1, A2', B1, B2, W1, y* F_0 *tiene densidad* f_0 *acotada, entonces para el caso* $F=F_0$ *se tendrá que* $\lim_{\varepsilon\searrow 0}\mathcal{B}_{\mathbf{T}_1}(\varepsilon)=0$.

Además de conocer la resistencia de un estimador frente a una gran proporción de outliers, cosa que describe el punto de ruptura, es importante estudiar el comportamiento del estimador cuando $\varepsilon\approx 0$. Una herramienta útil en este sentido es la función de influencia. Dado $(\mathbf{x},y)\in\mathbb{R}^{p+1}$ y $\delta_{(\mathbf{x},y)}$ la distribución concentrada en ese punto, la función de influencia de $\mathbf{T}$ evaluada en $(\mathbf{x},y)$ se define como

$$\mathrm{IF}_{\mathbf{T}}(\mathbf{x},y)=\lim_{\varepsilon\searrow 0}\frac{1}{\varepsilon}\left[\mathbf{T}\left((1-\varepsilon)H_0+\varepsilon\delta_{(\mathbf{x},y)}\right)-\boldsymbol{\theta}\right].$$

Mientras que la función de sesgo máximo mide el peor efecto que una proporción fija de outliers puede tener en el estimador, la función de influencia describe el efecto de una proporción infinitesimal de outliers fijos en un punto. El libro de Hampel et al. (1986) desarrolla una teoría de la robustez basada en la función de influencia y centra su atención en estimadores con influencia acotada. Cabe señalar que, bajo ciertas condiciones de regularidad, los estimadores son asintóticamente normales bajo el modelo y la matriz de covarianza asintótica está dada por $E_{H_0}\left\{\mathrm{IF}_{\mathbf{T}}\mathrm{IF}_{\mathbf{T}}'\right\}$.

El siguiente teorema da la función de influencia del REWLS. Primero, requeriremos que la función de peso w sea suave:

W2. $w(u)$ es continuamente derivable.

Sean

$$\begin{aligned} d^* &= \sup_{t\geq 0}\left\{\max\left\{F^+(t) - F_0^+(t), 0\right\}\right\} \\ \bar{t}^* &= \left(F_0^+\right)^{-1}(1-d^*) \\ t^* &= \max\left\{\bar{t}^*, \eta\right\}. \end{aligned} \tag{3.10}$$

Teorema 3.3. *Si A1, A2', B1, B2, W1 y W2 se satisfacen, entonces la función de influencia del REWLS es*

$$\mathrm{IF}_{\mathbf{T}_1}(\mathbf{x}, y) = \left\{ w\left(\frac{y - \mathbf{x}'\boldsymbol{\theta}}{\sigma t^*}\right) \Sigma^{-1}\mathbf{x}\left(y - \mathbf{x}'\boldsymbol{\theta}\right) - h_2(t^*)\, \mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}, y) \right\} / h_1(t^*) \tag{3.11}$$

donde $\mathrm{IF}_{\mathbf{T}_0}$ *es la función de influencia del estimador inicial y*

$$\begin{aligned} h_1(t^*) &= \int w(u/t^*)\, dF_0(u) \\ h_2(t^*) &= \int w'(u/t^*)(u/t^*)\, dF_0(u). \end{aligned}$$

Notemos que si F^+ es estocásticamente mayor o igual que F_0^+ se tendrá que $t^* = \infty$, y entonces la función de influencia del REWLS coincidirá con la del EMC. Así que tenemos razones para pensar que si usamos $F = \Phi$, el REWLS alcanzará la máxima eficiencia asintótica para el modelo de errores normales. En la Sección 4 se demuestra esto rigurosamente.

Para tener una descripción más completa de las cualidades de robustez del REWLS necesitaríamos poder calcular la función de sesgo máximo para todo ε entre 0 y $\varepsilon^*_{\mathbf{T}_1}$. Para algunos estimadores esto es posible (ver por ejemplo Martin et al., 1989). Lamentablemente, no pudimos hacerlo para el REWLS. Sin embargo, para el caso del LMS como estimador inicial se puede calcular (numéricamente) la función de sesgo máximo para contaminaciones puntuales. De esto se habla en la siguiente sección.

3.1. El LMS como estimador inicial

En esta sección analizamos más detenidamente el caso en que $\mathbf{T}_0$ es el LMS. Este caso es particularmente importante dado que el sesgo máximo del LMS está muy cerca del mínimo posible dentro de la clase de estimadores "residual admissible"

ε	REWLS	LMS
0.01	0.225	0.220
0.05	0.542	0.528
0.10	0.849	0.827
0.15	1.173	1.140
0.20	1.563	1.515

Table 3.1: Sesgos máximos para contaminaciones puntuales

(ver Yohai y Zamar, 1993). Además el LMS es quizá el estimador robusto de regresión más popularmente usado en la práctica y es muy posible que quienes hayan de utilizar el REWLS empleen el LMS como estimador inicial.

Como ya dijimos, no nos fue posible calcular la función de sesgo máximo del REWLS. En su lugar, nos restringimos a un entorno más pequeño de H_0 donde sólo se permiten contaminaciones puntuales, y definimos la función de sesgo máximo puntual

$$\mathcal{B}_{\mathrm{T}}^{*}(\varepsilon)=\sup\left\{\|\mathrm{T}\left((1-\varepsilon)H_0+\varepsilon\delta_{\mathbf{z}}\right)\|:\mathbf{z}\in\mathbb{R}^{p+1}\right\}. \tag{3.12}$$

Para el LMS la función de sesgo máximo y la de sesgo máximo puntual coinciden. Esto se demuestra en Martin et al. (1989), donde además se da la expresión explícita de $\mathcal{B}_{LMS}(\varepsilon)$. Como H_0 tomamos la distribución $N_{p+1}(0,\mathrm{I})$. En el Apéndice C se explica detalladamente cómo se puede calcular 3.12 para el REWLS, tomando $F=\Phi$, $\eta=2.5$ y $w(u)=\mathbb{I}(|u|<1)$. Algunos valores de $\mathcal{B}_{REWLS}^{*}(\varepsilon)$ se muestran en la Tabla 3.1.

Como podemos observar en la Tabla 3.1, el sesgo máximo puntual del REWLS es apenas mayor que el del LMS. También calculamos los sesgos máximos del estimador a un paso de Rousseeuw con constante de corte 2.5, pero coinciden con los del REWLS y por eso no se incluyen en la tabla.

En conclusión, vemos que el REWLS sacrifica muy poco de la robustez del LMS, y esta pérdida se ve ampliamente compensada por la ganancia en eficiencia, según se muestra en la siguiente sección.

4. Distribución asintótica del REWLS

Como se dijo en la Introducción, la contribución original de esta Tesis consiste en un estimador robusto que además de alcanzar el punto de ruptura máximo (cosa que se enunció en la Sección 3), sea asintóticamente equivalente al EMC bajo el modelo central y por lo tanto tenga la máxima eficiencia asintótica en el modelo de errores normales. Esto es lo que veremos ahora.

En primer lugar, probaremos que la proporción de outliers estimada d_n, definida en (2.2), tiende a cero cuando la distribución hipotética F^{+} es estocásticamente mayor o igual que la distribución verdadera F_0^+. En tal caso, la constante de corte t_n tiende a infinito, de manera que asintóticamente no se elimina ninguna observación. El Lema siguiente es en realidad más general, ya que muestra que d_n y $\bar{t}_n$ tienen límite aunque F^+ sea estocásticamente menor que F_0^+. No obstante, tengamos presente que el caso que más nos interesa es cuando $F_0 = \Phi$, pues es sólo ahí donde el EMC alcanza la máxima eficiencia asintótica y por lo tanto también la alcanzará el REWLS si se usa $F = \Phi$, como veremos.

Antes del Lema, agregaremos algunas suposiciones acerca del diseño y de los estimadores iniciales:

A3. Para todo $\varepsilon > 0$, existe un K tal que

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left(\|\mathbf{x}_i\| > K\right) < \varepsilon.$$

A4. Los estimadores iniciales son débilmente consistentes, es decir, $\lim \mathbf{T}_{0n} = \boldsymbol{\theta}$ y $\lim s_n = \sigma$ en probabilidad.

Los S-estimadores, y en particular el LMS, cumplen A4 bajo condiciones muy generales. Ver el Teorema 3 de Davies (1990) para más detalles.

Lema 4.1. *Sean d_n y $\bar{t}_n$ dados por (2.2) y (2.3), y d^* y $\bar{t}^*$ definidos en (3.10). Si se cumplen A1, A3 y A4 y $|f_0|_\infty$ es finita, entonces $\left|F_n^+ - F_0^+\right|_\infty = o_p(1)$, lo que implica que $d_n \to d^*$ y $\bar{t}_n \to \bar{t}^*$ en probabilidad.*

Cabe aclarar que la norma de funciones utilizada en el lema previo y en lo que sigue será la norma del supremo.

Cuando el estimador inicial sea el LMS, necesitaremos un resultado más fuerte, para el cual a su vez necesitamos las siguientes suposiciones:

A5. Para todo $\varepsilon > 0$, existe un K tal que

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x}_i\|^2\, \mathbb{I}\left(\|\mathbf{x}_i\| > K\right) < \varepsilon.$$

A6. Los estimadores iniciales satisfacen $\|\mathbf{T}_{0n} - \boldsymbol{\theta}\| = O_p\left(n^{-\tau}\right)$ y $|s_n - \sigma| = O_p\left(n^{-\tau}\right)$ con $\tau > 1/4$.

Lema 4.2. *Bajo los supuestos A1, A5 y A6, suponiendo además que la densidad f_0 es derivable y que $|f_0|_\infty$, $|f_0'|_\infty$ y $|u^2 f_0'(u)|_\infty$ son finitas, se tiene que $\left|F_n^+ - F_0^+\right|_\infty = O_p\left(n^{-1/2}\right)$.*

El LMS cumple A6 bajo condiciones muy generales. De hecho, los Teoremas 4 y 5 del paper de Davies (1990) muestran que $\|\mathbf{T}_{0n} - \boldsymbol{\theta}\| = O_p\left(n^{-1/3}\right)$ y $|s_n - \sigma| = O_p\left(n^{-1/2}\right)$. La condición sobre el diseño que Davies llama D1 se deduce de nuestra condición A5. Las otras condiciones con respecto a f_0 (tanto las impuestas por Davies como las mencionadas en el Lema 4.2) se cumplen para todas las densidades usadas en la práctica.

Para encontrar la distribución asintótica del REWLS, analizaremos primero el caso en que $\mathbf{T}_{0n}$ es consistente de orden $n^{-1/2}$. Supongamos que:

A7. El estimador de regresión inicial admite el desarrollo asintótico

$$\mathbf{T}_{0n} - \boldsymbol{\theta} = \Gamma_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi\left(\frac{u_i}{\sigma}\right) \mathbf{x}_i + o_p\left(n^{-1/2}\right),$$

donde ψ es una función impar acotada y $\Sigma_n^{1/2}\Gamma_n^{-1}\Sigma_n^{1/2}$ converge en probabilidad a Γ, una matriz simétrica y definida positiva.

Bajo ciertas condiciones de regularidad, A7 se cumple cuando $\mathbf{T}_{0n}$ es solución de la ecuación

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{x}_i'\mathbf{T}_{0n}}{s_n}\right) \mathbf{x}_i = 0. \tag{4.1}$$

Tal es el caso, por ejemplo, de los estimadores MM, S y τ. Además, en esos tres casos se tiene que Γ es simplemente un múltiplo de la identidad; específicamente,

$$\Gamma = \frac{\sigma}{E_{F_0}\left\{\psi'\right\}} \mathrm{I}. \tag{4.2}$$

El siguiente lema da condiciones suficientes para que esto suceda.

Lema 4.3. *Supongamos que el estimador inicial $\mathbf{T}_{0n}$ cumple (4.1) para un estimador de escala s_n y para una ψ impar, acotada, con derivada ψ' continua tal que $|u\psi'(u)|_\infty$ es finita y $E_{F_0}\{\psi'\} \neq 0$. Bajo las condiciones A1, A2, A5, y suponiendo además que $\|\mathbf{T}_{0n} - \boldsymbol{\theta}\|$ y $|s_n - \sigma|$ son $O_p\left(n^{-1/2}\right)$, se cumple A7 con Γ dada por (4.2).*

En general, para que $\mathbf{T}_{0n}$ sea robusto deberá usarse una ψ redescendiente en (4.1), es decir que existe u_0 tal que $\psi(u) = 0$ si $|u| > u_0$, de modo que la condición $|u\psi'(u)|_\infty$ finita que se pide en el Lema 4.3 se cumple. Cuando $\mathbf{T}_{0n}$ es un S estimador y s_n es el estimador de escala asociado, el Teorema 7 y el Lema 5 de Davies (1990) dan el orden $O_p\left(n^{-1/2}\right)$ requerido por el Lema 4.3. Cabe remarcar que las hipótesis requeridas en estos dos resultados de Davies se cumplen bajo las condiciones del Lema 4.3 si se pide además que $f_0(u)$ sea no creciente en $|u|$.

Ahora sí presentamos el resultado que da la distribución asintótica del REWLS.

Teorema 4.4. *Si se cumplen F1, A1, A2, A5, A7, W1 y W2, entonces*

$$\sqrt{n}\,\Sigma_n^{1/2}\left(\mathbf{T}_{1n}-\boldsymbol{\theta}\right)\to_D N_p\left(\mathbf{0},V\left(t^*\right)/h_1^2\left(t^*\right)\right)$$

donde

$$V\left(t^*\right)=\sigma^2E_{F_0}\left\{\left[w\left(\frac{u}{t^*}\right)u\mathrm{I}-h_2\left(t^*\right)\frac{1}{\sigma}\psi\left(u\right)\Gamma\right]^2\right\} \tag{4.3}$$

y las funciones h_1 y h_2 son las definidas en el Teorema 3.3.

En este teorema vemos que cuando $t^*=\infty$, cosa que ocurre sólo si F^+ es estocásticamente mayor que F_0^+, la variancia asintótica del REWLS (estandarizado por $\Sigma_n^{1/2}$) es $\sigma^2\mathrm{I}$, siendo equivalente entonces al EMC, como queríamos probar. En el caso particular en que valga (4.2), la fórmula (4.3) se simplifica a

$$V\left(t^*\right)=\sigma^2E_{F_0}\left\{\left[w\left(\frac{u}{t^*}\right)u-\frac{h_2\left(t^*\right)}{E_{F_0}\left\{\psi'\right\}}\psi\left(u\right)\right]^2\right\}\mathrm{I}.$$

Veamos ahora qué pasa si $\mathbf{T}_{0n}$ es el LMS, en cuyo caso A7 no se cumple. En primer lugar, debemos hacer una suposición más fuerte sobre el diseño:

A8. Para todo $\varepsilon>0$, existe un K tal que

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{x}_i\right\|^3\mathbb{I}\left(\left\|\mathbf{x}_i\right\|>K\right)<\varepsilon.$$

Teorema 4.5. *Definamos $g\left(u\right)=uf_0\left(u\right)$. Supongamos que se cumplen F1, A1, A2, A6, A8, que la densidad f_0 es derivable con $\left|f_0\right|_\infty$, $\left|f_0'\right|_\infty$ y $\left|u^2f_0'\left(u\right)\right|_\infty$ finitas, y que existe $C>0$ tal que $g\left(u\right)\le C\left(1-F_0^+\left(u\right)\right)^{1/2}$ $\forall u\ge0$. Si $w\left(u\right)=\mathbb{I}\left(\left|u\right|<1\right)$ y $t^*=\infty$, entonces*

$$\sqrt{n}\,\Sigma_n^{1/2}\left(\mathbf{T}_{1n}-\boldsymbol{\theta}\right)\to_D N_p\left(\mathbf{0},\sigma^2\mathrm{I}\right).$$

La condición impuesta por el Teorema 4.5 de que $g\left(u\right)\le C\left(1-F_0^+\left(u\right)\right)^{1/2}$ para cierta $C>0$ y para todo $u\ge0$ no es demasiado restrictiva. Lo que dice, esencialmente, es que $u^2f_0^2\left(u\right)$ tiende a cero a mayor velocidad que la probabilidad de las colas cuando $\left|u\right|\to\infty$, y esto pasa en general para las densidades con decrecimiento exponencial. En el caso $F_0=\Phi$ se ve porque la diferencia $2\left(1-\Phi\left(u\right)\right)-u^2\varphi^2\left(u\right)$ tiende a cero cuando $u\to\infty$ y como es estrictamente decreciente a partir de cierto u_0, no puede hacerse negativa en $\left[u_0,\infty\right)$. Así que

$g(u) \leq C\left(1 - F_0^+(u)\right)^{1/2}$ para $u \geq u_0$ y $C \geq 1$, y es claro que esta desigualdad valdrá también para u en $[0, u_0]$ si se toma una C suficientemente grande.

Aunque el resultado del Teorema 4.5 es más débil que el del Teorema 4.4, de todos modos obtenemos la eficiencia asintótica del REWLS para errores normales, como queríamos. En realidad, el Teorema 4.5 demuestra que nuestro EMCP es superior al propuesto por Rousseeuw, en el sentido que el REWLS logra un orden de convergencia de $n^{-1/2}$ mientras que el de Rousseeuw conserva el orden de convergencia $n^{-1/3}$ del LMS.

Finalmente nos resta considerar el caso en que $\mathbf{T}_{0n}$ es el LMS pero $t^* < \infty$. Esto es, asintóticamente, lo mismo que usar una constante de corte fija y por lo tanto el orden de consistencia del REWLS permanecerá $n^{-1/3}$.

Teorema 4.6. *Si se cumplen F1, A1, A2, A5, A6, W1 y W2, entonces*

$$h_1(t^*)\,\Sigma_n^{1/2}(\mathbf{T}_{1n} - \boldsymbol{\theta}) = h_2(t^*)\,\Sigma_n^{1/2}(\mathbf{T}_{0n} - \boldsymbol{\theta}) + O_p\left(n^{-1/2}\right).$$

En la mayoría de los casos prácticos, si $t^* < \infty$ entonces $h_2(t^*) \neq 0$. Si además

$$n^{1/3}\Sigma_n^{1/2}(\mathbf{T}_{0n} - \boldsymbol{\theta}) \to_D \mathcal{Z}$$

para cierta variable aleatoria $\mathcal{Z}$, como es el caso del LMS bajo ciertas condiciones de regularidad (ver el Teorema 6 de Davies, 1990), el Teorema 4.6 implica que

$$n^{1/3}\Sigma_n^{1/2}(\mathbf{T}_{1n} - \boldsymbol{\theta}) \to_D \frac{h_2(t^*)}{h_1(t^*)}\mathcal{Z}.$$

Es decir, el REWLS permanece consistente a un orden $n^{-1/3}$. Sin embargo sigue habiendo ventaja en usar el estimador a un paso, porque $h_2(t^*)/h_1(t^*)$ converge a 0 cuando $t^* \to \infty$ y entonces será usualmente una constante pequeña.

A. Resultados robustos

Esta sección contiene las demostraciones y resultados adicionales correspondientes a la Sección 3. Recordemos que por la equivariancia de los estimadores estamos suponiendo que $\boldsymbol{\theta} = \mathbf{0}$, $\Sigma = \mathrm{I}$ y $\sigma = 1$, de modo que bajo el modelo central se tiene que $\mathbf{x} \sim G_0$, $y \sim F_0$ y son estocásticamente independientes.

Dada $H \in \mathcal{H}\varepsilon$ con $\varepsilon < \varepsilon^*_{\mathrm{T}_0}$ y $E_H \left\{ w_H(\mathbf{x}, y) \|\mathbf{x}\|^2 \right\} < \infty$, llamemos $\lambda_1(H)$ al menor autovalor de $E_H \{ w_H(\mathbf{x}, y) \mathbf{x}\mathbf{x}' \}$. El Lema A.1 implica que $E_H \{ w_H(\mathbf{x}, y) \mathbf{x}\mathbf{x}' \}$ es inversible y por lo tanto la expresión explícita (3.6) para $\gamma(H)$ está bien definida (usando además (A.3)).

Lema A.1. *Si A1, A2', B1, B2 y W1 se satisfacen, existe un* $\lambda_0(\varepsilon) > 0$ *tal que* $\lambda_1(H) \geq (1-\varepsilon)\lambda_0(\varepsilon)$ *para toda* $H \in \mathcal{H}\varepsilon$ *con* $\varepsilon < \varepsilon^*_{\mathrm{T}_0}$ *y* $E_H \left\{ w_H(\mathbf{x}, y) \|\mathbf{x}\|^2 \right\} < \infty$.

Demostración: En primer lugar, obsérvese que por W1 la función de peso $w(u)$ es estrictamente positiva en un entorno de 0 y entonces existe un $c_1 \in (0, 1)$ tal que

$$w(u) \geq \frac{1}{2}\mathbb{I}(|u| \leq c_1).$$

Sea $H \in \mathcal{H}\varepsilon$ con $\varepsilon < \varepsilon^*_{\mathrm{T}_0}$ y $E_H \left\{ w_H(\mathbf{x}, y) \|\mathbf{x}\|^2 \right\} < \infty$. Por simplicidad, llamemos $w_H = w_H(\mathbf{x}, y)$ y $r_H = r_H(\mathbf{x}, y)$. Definamos

$$\tilde{w}_H = \frac{1}{2}\mathbb{I}(|r_H| \leq c_1\eta). \tag{A.1}$$

Entonces $w_H \geq \tilde{w}_H$. También, como

$$\lambda_1(H) = \min_{\|\mathbf{v}\|=1} \mathbf{v}' E_H \{ w_H \mathbf{x}\mathbf{x}' \} \mathbf{v} = E_H \left\{ w_H |\mathbf{x}'\mathbf{v}_H|^2 \right\}$$

para algún $\mathbf{v}_H$ con $\|\mathbf{v}_H\| = 1$, tenemos que

$$\lambda_1(H) \geq (1-\varepsilon) E_{H_0} \left\{ \tilde{w}_H |\mathbf{x}'\mathbf{v}_H|^2 \right\}.$$

Si el $\lambda_0(\varepsilon) > 0$ que buscamos no existiera, habría una sucesión $\{H_n\}$ en $\mathcal{H}\varepsilon$ con $E_{H_n} \left\{ w_{H_n} \|\mathbf{x}\|^2 \right\} < \infty$ tal que

$$\lim_{n\to\infty} \lambda_1(H_n) = 0 = \lim_{n\to\infty} E_{H_0} \left\{ \tilde{w}_{H_n} |\mathbf{x}'\mathbf{v}_{H_n}|^2 \right\}. \tag{A.2}$$

Sin embargo, como $\|\mathbf{v}_{H_n}\| = 1$, $\|\mathrm{T}_0(H_n)\| \leq \mathcal{B}_{\mathrm{T}_0}(\varepsilon)$ y $\sigma_1(\varepsilon) \leq S(H_n) \leq \sigma_2(\varepsilon)$, deben existir una subsucesión $\{H_{n_k}\}$, $\mathbf{v}_0 \in \mathbb{R}^p$, $\mathbf{v}_1 \in \mathbb{R}^p$ con $\|\mathbf{v}_1\| = 1$ y $s_0 > 0$ tales que

$$\mathbf{v}_{H_{n_k}} \to \mathbf{v}_1,\ \mathrm{T}_0(H_{n_k}) \to \mathbf{v}_0 \text{ y } S(H_{n_k}) \to s_0.$$

Por lo tanto

$$\tilde{w}_{H_{n_k}}(\mathrm{x},y) \to \frac{1}{2}\mathbb{I}\left(\frac{|y-\mathrm{x}'\mathrm{v}_0|}{s_0} \leq c_1\eta\right) \triangleq \tilde{w}_0(\mathrm{x},y)$$

puntualmente en (x,y), salvo quizá para aquellos (x,y) tales que $|y-\mathrm{x}'\mathrm{v}_0| = s_0c_1\eta$, conjunto que tiene probabilidad H_0 nula por ser F_0 continua. Como

$$\liminf_{k\to\infty} E_{H_0}\left\{\tilde{w}_{H_{n_k}}\left|\mathrm{x}'\mathrm{v}_{H_{n_k}}\right|^2\right\} \geq E_{H_0}\left\{\liminf_{k\to\infty}\tilde{w}_{H_{n_k}}\left|\mathrm{x}'\mathrm{v}_{H_{n_k}}\right|^2\right\} = E_{H_0}\left\{\tilde{w}_0\left|\mathrm{x}'\mathrm{v}_1\right|^2\right\}$$

de (A.2) se tiene que $E_{H_0}\left\{\tilde{w}_0\left|\mathrm{x}'\mathrm{v}_1\right|^2\right\} = 0$. Es decir, si llamamos $c_2 = s_0c_1\eta$ se tiene que

$$\frac{1}{2}E_{G_0}\left\{\left|\mathrm{x}'\mathrm{v}_1\right|^2\left[F_0(\mathrm{x}'\mathrm{v}_0+c_2)-F_0(\mathrm{x}'\mathrm{v}_0-c_2)\right]\right\} = 0.$$

Pero como $F_0(u)$ es estrictamente creciente, esto implica que

$$P_{G_0}\left\{\left|\mathrm{x}'\mathrm{v}_1\right|^2 = 0\right\} = 1$$

y entonces $\mathrm{v}_1'\Sigma\mathrm{v}_1 = 0$. Siendo Σ definida positiva debe ser $\mathrm{v}_1 = \mathbf{0}$, pero era $\|\mathrm{v}_1\| = 1$, contradicción. ■

Demostración del Teorema 3.1: En primer lugar mostraremos que

$$E_H\left\{w_H(\mathrm{x},y)\left(y-\mathrm{x}'\mathrm{T}_0(H)\right)^2\right\} \leq \left(2\eta^2 + E_F\left\{r^2\right\}\right)\sigma_2^2(\varepsilon) \qquad \text{(A.3)}$$

para toda $H \in \mathcal{H}\varepsilon$ con $\varepsilon < \varepsilon^2_{\mathrm{T}_0}$. Notar que $w(u) \leq \mathbb{I}(|u|<1)$, así que

$$E_H\left\{w_H(\mathrm{x},y)\left(\frac{y-\mathrm{x}'\mathrm{T}_0(H)}{S(H)}\right)^2\right\} \leq \lim_{t\nearrow t_H}\int_0^t r^2dF_H^+(r).$$

Ahora bien,

$$\int_0^t r^2dF_H^+(r) = t^2\left(F_H^+(t)-1\right) + \int_0^t\left(1-F_H^+(r)\right)2rdr.$$

La definición de d_H implica que $F^+(r) \leq F_H^+(r) + d_H$ para todo $r \geq 0$, entonces

$$\int_0^t\left(1-F_H^+(r)\right)2rdr \leq \int_0^t\left(1-F^+(r)\right)2rdr + t^2d_H.$$

Por lo tanto

$$\int_0^t r^2dF_H^+(r) \leq t^2\left(F_H^+(t)-1+d_H\right) + \int_0^t\left(1-F^+(r)\right)2rdr.$$

y como

$$\int_0^\infty \left(1 - F^+(r)\right) 2r dr = E_F\left\{r^2\right\}$$

se tiene que

$$\int_0^t r^2 dF_H^+(r) \le t^2\left(F_H^+(t) - 1 + d_H\right) + E_F\left\{r^2\right\}. \tag{A.4}$$

Para completar la demostración, consideremos las dos posibilidades: $t_H = \eta$ o $t_H = \bar{t}_H$. En el primer caso, es inmediato de (A.4) que

$$E_H\left\{w_H r_H^2\right\} \le 2\eta^2 + E_F\left\{r^2\right\}$$

y entonces se cumple (A.3). En el segundo caso, usando que $F_H^+(t) \le 1 - d_H$ para toda $t < \bar{t}_H$ se sigue de (A.4) que

$$E_H\left\{w_H r_H^2\right\} \le E_F\left\{r^2\right\} \tag{A.5}$$

y por lo tanto también vale la cota (A.3).

Ahora bien, por definición sabemos que $\mathbf{T}_1(H)$ satisface (3.9) y entonces

$$E_H\left\{w_H\left(y - \mathbf{x}'\mathbf{T}_1(H)\right)^2\right\} \le \left(2\eta^2 + E_F\left\{r^2\right\}\right)\sigma_2^2(\varepsilon)$$

para toda $H \in \mathcal{H}\varepsilon$ con $\varepsilon < \varepsilon^*_{\mathbf{T}_0}$. Supongamos que exista un $\tilde{\varepsilon} < \varepsilon^*_{\mathbf{T}_0}$ tal que $\mathcal{B}_{\mathbf{T}_1}(\tilde{\varepsilon}) = \infty$. Entonces existiría una sucesión $\{H_n\}$ de distribuciones en $\mathcal{H}_{\tilde{\varepsilon}}$ tal que $\|\mathbf{T}_1(H_n)\| \to \infty$. Sea $\mathbf{v}_n = \mathbf{T}_1(H_n) / \|\mathbf{T}_1(H_n)\|$. El resto del argumento es similar al de la demostración del Lema A.1. Como $\mathbf{v}_n$, $\mathbf{T}_0(H_n)$ y $S(H_n)$ están acotadas, existe una subsucesión $\{H_{n_k}\}$ tal que

$$\mathbf{v}_{n_k} \to \mathbf{v}_1,\ \mathbf{T}_0(H_{n_k}) \to \mathbf{v}_0 \text{ y } S(H_{n_k}) \to s_0$$

para ciertos vectores $\mathbf{v}_0$ y $\mathbf{v}_1$ de norma 1 y cierta $s_0 > 0$. Como

$$(1-\tilde{\varepsilon}) E_{H_0}\left\{\tilde{w}_{H_{n_k}}\left(\frac{y - \mathbf{x}'\mathbf{T}_1(H_{n_k})}{\|\mathbf{T}_1(H_{n_k})\|}\right)^2\right\}\|\mathbf{T}_1(H_{n_k})\|^2 \le \left(2\eta^2 + E_F\left\{r^2\right\}\right)\sigma_2^2(\tilde{\varepsilon})$$

para $\tilde{w}_H$ como en (A.1), entonces

$$\lim_{k\to\infty} E_{H_0}\left\{\tilde{w}_{H_{n_k}}\left(\frac{y - \mathbf{x}'\mathbf{T}_1(H_{n_k})}{\|\mathbf{T}_1(H_{n_k})\|}\right)^2\right\} = 0.$$

Como

$$\begin{aligned}\liminf_{k\to\infty} E_{H_0}\left\{\tilde{w}_{H_{n_k}}\left(\frac{y-\mathrm{x}'\mathrm{T}_1(H_{n_k})}{\|\mathrm{T}_1(H_{n_k})\|}\right)^2\right\} &\geq E_{H_0}\left\{\liminf_{k\to\infty}\tilde{w}_{H_{n_k}}\left(\frac{y-\mathrm{x}'\mathrm{T}_1(H_{n_k})}{\|\mathrm{T}_1(H_{n_k})\|}\right)^2\right\}\\ &= E_{H_0}\left\{\tilde{w}_0(-\mathrm{x}'\mathrm{v}_1)^2\right\}\end{aligned}$$

si llamamos $c_2 = s_0 c_1 \eta$ se tiene que

$$E_{H_0}\left\{\tilde{w}_0(-\mathrm{x}'\mathrm{v}_1)^2\right\} = \frac{1}{2}E_{G_0}\left\{|\mathrm{x}'\mathrm{v}_1|^2\left[F_0(\mathrm{x}'\mathrm{v}_0+c_2)-F_0(\mathrm{x}'\mathrm{v}_0-c_2)\right]\right\} = 0$$

lo que es una contradicción nuevamente. Así que no puede ser $\mathcal{B}_{\mathrm{T}_1}(\varepsilon) = \infty$ si $\varepsilon < \varepsilon^*_{\mathrm{T}_0}$, por lo que debe ser $\varepsilon^*_{\mathrm{T}_1} \geq \varepsilon^*_{\mathrm{T}_0}$. ∎

Ahora probaremos que el sesgo máximo del REWLS tiende a cero con ε en el caso $F = F_0$, para lo cual necesitaremos primero el siguiente resultado:

Lema A.2. *Supongamos que se cumplen A1, B1 y B2. Además supondremos que F_0 tiene densidad f_0 acotada. Entonces:*
(i) $\lim_{\varepsilon\searrow 0}\sup_{H\in\mathcal{H}_\varepsilon}|d_H - d^*| = 0.$
(ii) $\lim_{\varepsilon\searrow 0}\sup_{H\in\mathcal{H}_\varepsilon}\left|F_0^+(\bar{t}_H) - (1-d^*)\right| = 0.$

Demostración: (i) Es fácil verificar que

$$|d_H - d^*| \leq \left|F_0^+ - F_H^+\right|_\infty .$$

Demostraremos entonces que

$$\lim_{\varepsilon\searrow 0}\sup_{H\in\mathcal{H}_\varepsilon}\left|F_0^+ - F_H^+\right|_\infty = 0. \tag{A.6}$$

Este resultado se usará luego para demostrar la parte (ii) de este lema.

Sean $M_1 = \sup_u f_0(u)$ y $M_2 = \sup_u |uf_0(u)|$. Notemos que $M_1 < \infty$ implica $M_2 < \infty$, porque si $f_0(u)$ es acotada en $\mathbb{R}$ entonces $uf_0(u)$ es acotada en cualquier compacto de $\mathbb{R}$, así que si $uf_0(u)$ no fuera acotada en $\mathbb{R}$ se tendría que $uf_0(u) \to \infty$ cuando (y solamente cuando) $u \to \infty$ y entonces existiría A tal que $uf_0(u) \geq 2$ para todo $u \geq A$, lo que implicaría el absurdo

$$1 \geq \int_A^{2A} f_0(u)\,du = \int_A^{2A}\frac{2}{u}du \geq 2\ln(2).$$

Para cada $t \geq 0$, tenemos que

$$F_H^+(t) \geq (1-\varepsilon)E_{G_0}\left\{F_0^+(t\sigma_1(\varepsilon) - \|\mathrm{x}\|\mathcal{B}_{\mathrm{T}_0}(\varepsilon))\right\}$$

y entonces

$$F_0^+(t) - F_H^+(t) \leq \varepsilon + (1-\varepsilon) E_{G_0} \left\{ F_0^+(t) - F_0^+(t\sigma_1(\varepsilon) - \|\mathbf{x}\| \mathcal{B}_{\mathrm{T}_0}(\varepsilon)) \right\}$$

Llamemos $A\varepsilon = \left\{ \mathbf{x} : \|\mathbf{x}\| \leq (\mathcal{B}_{\mathrm{T}_0}(\varepsilon))^{-1/2} \right\}$. Por el Teorema del Valor Medio (TVM), se tiene que

$$F_0^+(t) - F_0^+(t\sigma_1(\varepsilon)) \leq 2f_0(\xi)\, t\,(1-\sigma_1(\varepsilon))$$

para $t\sigma_1(\varepsilon) < \xi < t$, así que $(t/\xi) < 1/\sigma_1(\varepsilon)$ y entonces

$$F_0^+(t) - F_0^+(t\sigma_1(\varepsilon)) \leq 2M_2 \left(\frac{1-\sigma_1(\varepsilon)}{\sigma_1(\varepsilon)} \right).$$

También, como $\left| f_0^+ \right|_\infty \leq 2M_1$ se tiene

$$F_0^+(t\sigma_1(\varepsilon)) - F_0^+(t\sigma_1(\varepsilon) - \|\mathbf{x}\| \mathcal{B}_{\mathrm{T}_0}(\varepsilon)) \leq 2M_1 \|\mathbf{x}\| \mathcal{B}_{\mathrm{T}_0}(\varepsilon)$$

y entonces para todo $\mathbf{x} \in A\varepsilon$ es

$$F_0^+(t\sigma_1(\varepsilon)) - F_0^+(t\sigma_1(\varepsilon) - \|\mathbf{x}\| \mathcal{B}_{\mathrm{T}_0}(\varepsilon)) \leq 2M_1 (\mathcal{B}_{\mathrm{T}_0}(\varepsilon))^{1/2}.$$

Por lo tanto

$$F_0^+(t) - F_H^+(t) \leq$$

$$\varepsilon + (1-\varepsilon) \left[2M_2 \left(\frac{1-\sigma_1(\varepsilon)}{\sigma_1(\varepsilon)} \right) + 2M_1 (\mathcal{B}_{\mathrm{T}_0}(\varepsilon))^{1/2} + P_{G_0}\left(A_\varepsilon^C\right) \right] \triangleq c_1(\varepsilon). \quad \text{(A.7)}$$

Por otro lado,

$$F_H^+(t) \leq \varepsilon + (1-\varepsilon) E_{G_0} \left\{ F_0^+(t\sigma_2(\varepsilon) + \|\mathbf{x}\| \mathcal{B}_{\mathrm{T}_0}(\varepsilon)) \right\}$$

y entonces podemos obtener una cota inferior

$$F_0^+(t) - F_H^+(t) \geq$$

$$-\varepsilon - (1-\varepsilon) \left[2M_2(\sigma_2(\varepsilon) - 1) + 2M_1 (\mathcal{B}_{\mathrm{T}_0}(\varepsilon))^{1/2} + P_{G_0}\left(A_\varepsilon^C\right) \right] \triangleq -c_2(\varepsilon). \quad \text{(A.8)}$$

Como $c(\varepsilon) \triangleq \max\{c_1(\varepsilon), c_2(\varepsilon)\}$ no depende de t ni de H, de (A.7) y (A.8) deducimos que $\sup_{H\in\mathcal{H}\varepsilon} \left| F_0^+ - F_H^+ \right|_\infty \leq c(\varepsilon)$ y $\lim_{\varepsilon\searrow 0} c(\varepsilon) = 0$.

(ii) Si $\lim_{\varepsilon\searrow 0} \sup_{H\in\mathcal{H}\varepsilon} \left| F_0^+(\bar{t}_H) - (1-d^*) \right| > 0$, habría un $\delta > 0$ y sucesiones $\{\varepsilon_n\}$ y $\{H_n\}$ tales que $\varepsilon_n \to 0$, $H_n \in \mathcal{H}_{\varepsilon_n}$ para cada n, y

$$\lim_{n\to\infty} \left| F_0^+(\bar{t}_{H_n}) - (1-d^*) \right| = \delta. \quad \text{(A.9)}$$

Como $F_{H_n}^{+}(\bar{t}_{H_n}) \geq 1 - d_{H_n}$ y $d_{H_n} \to d^*$ por (i), se tiene que

$$\liminf_{n\to\infty} F_{H_n}^{+}(\bar{t}_{H_n}) \geq 1 - d^*$$

y como

$$\lim_{n\to\infty} \left| F_0^{+}(\bar{t}_{H_n}) - F_{H_n}^{+}(\bar{t}_{H_n}) \right| = 0$$

por (A.6), existe una subsucesión tal que

$$\lim_{k\to\infty} F_0^{+}\left(\bar{t}_{H_{n_k}}\right) \geq 1 - d^*.$$

Esto y (A.9) implican que existe un k_0 tal que

$$F_0^{+}\left(\bar{t}_{H_{n_k}}\right) > 1 - d^* + \frac{\delta}{2} \text{ para todo } k \geq k_0.$$

Entonces, en virtud de la continuidad de F_0^{+} debe existir otra sucesión $\left\{\tilde{t}_k\right\}$ tal que $\tilde{t}_k < \bar{t}_{H_{n_k}}$ para todo k y

$$F_0^{+}\left(\tilde{t}_k\right) > 1 - d^* + \frac{\delta}{2} \quad \text{para todo } k \geq k_0.$$

Pero $F_{H_{n_k}}\left(\tilde{t}_k\right) < 1 - d_{H_{n_k}}$ para todo k. Por lo tanto

$$F_0^{+}\left(\tilde{t}_k\right) - F_{H_{n_k}}^{+}\left(\tilde{t}_k\right) > d_{H_{n_k}} - d^* + \frac{\delta}{2}$$

para todo $k \geq k_0$. Usando nuevamente (A.6) se llega al absurdo $0 \geq \delta/2$. ∎

Demostración del Teorema 3.2: Como $\mathcal{B}_{\mathbf{T}_1}(\varepsilon)$ es decreciente y finita a partir de $\varepsilon^*_{\mathbf{T}_0}$, existe $\delta \triangleq \lim_{\varepsilon \searrow 0} \mathcal{B}_{\mathbf{T}_1}(\varepsilon)$. Si fuera $\delta > 0$, debería haber un $\mathbf{v}_0 \in \mathbb{R}^p$ y sucesiones $\{\varepsilon_n\}$ y $\{H_n\}$, con $\varepsilon_n \to 0$ y $H_n \in \mathcal{H}_{\varepsilon_n}$ para cada n, tales que $\mathbf{T}_1(H_n) \to \mathbf{v}_0$ y $\|\mathbf{v}_0\| = \delta$. Por la definición de $\mathbf{T}_1(H_n)$ sabemos que

$$E_{H_n}\left\{w_{H_n}\left(y - \mathbf{x}'\mathbf{T}_1(H_n)\right)^2\right\} \leq E_{H_n}\left\{w_{H_n} r_{H_n}^2\right\} S^2(H_n).$$

Como $t^* = \infty$ por ser $F = F_0$, podemos usar la desigualdad (A.5) para concluir que

$$\limsup_{n\to\infty} E_{H_n}\left\{w_{H_n}\left(y - \mathbf{x}'\mathbf{T}_1(H_n)\right)^2\right\} \leq E_F\left\{r^2\right\}. \tag{A.10}$$

Por otra parte,

$$E_{H_n}\left\{w_{H_n}\left(y - \mathbf{x}'\mathbf{T}_1(H_n)\right)^2\right\} \geq (1 - \varepsilon_n) E_{H_0}\left\{w_{H_n}\left(y - \mathbf{x}'\mathbf{T}_1(H_n)\right)^2\right\}. \tag{A.11}$$

Dado que $t_{H_n} \to \infty$, $S(H_n) \to 1$ y $\mathbf{T}_0(H_n) \to \mathbf{0}$, se tiene que

$$w_{H_n}(\mathbf{x}, y) \to w(0) = 1$$

porque de W1 se deduce que $w(u)$ es continua en 0. Entonces

$$\liminf_{n\to\infty} E_{H_0}\left\{(1-\varepsilon_n)\, w_{H_n}\left(y - \mathbf{x}'\mathbf{T}_1(H_n)\right)^2\right\} \geq E_{H_0}\left\{(y-\mathbf{x}'\mathbf{v}_0)^2\right\}. \tag{A.12}$$

Juntando las desigualdades (A.10) y (A.11) con (A.12), obtenemos

$$E_{H_0}\left\{(y-\mathbf{x}'\mathbf{v}_0)^2\right\} \leq E_F\left\{r^2\right\} \tag{A.13}$$

pero

$$E_{H_0}\left\{(y-\mathbf{x}'\mathbf{v}_0)^2\right\} = E_{F_0}\left\{y^2\right\} + \mathbf{v}_0'\Sigma\mathbf{v}_0$$

y entonces la desigualdad (A.13) implica que $\mathbf{v}_0 = \mathbf{0}$, una contradicción. Luego, debe ser $\lim_{\varepsilon\searrow 0} \mathcal{B}_{\mathbf{T}_1}(\varepsilon) = 0$. ■

Para evitar confusiones, a la siguiente demostración la haremos para Σ, $\boldsymbol{\theta}$ y σ generales.

Demostración del Teorema 3.3: Dado $0 < \varepsilon < \varepsilon^*_{\mathbf{T}_0}$ y un punto de contaminación $(\mathbf{x}_0, y_0)$, llamemos $H\varepsilon = (1-\varepsilon)\, H_0 + \varepsilon\delta_{(\mathbf{x}_0, y_0)}$. Entonces

$$\mathbf{T}_1(H_\varepsilon) - \boldsymbol{\theta} = \left(E_{H\varepsilon}\left\{w_{H\varepsilon}\mathbf{x}\mathbf{x}'\right\}\right)^{-1} E_{H\varepsilon}\left\{w_{H\varepsilon}\mathbf{x}\left(y-\mathbf{x}'\boldsymbol{\theta}\right)\right\}.$$

Como

$$E_{H\varepsilon}\left\{w_{H\varepsilon}\mathbf{x}\mathbf{x}'\right\} = (1-\varepsilon)\, E_{H_0}\left\{w_{H\varepsilon}\mathbf{x}\mathbf{x}'\right\} + \varepsilon\, w_{H\varepsilon}(\mathbf{x}_0, y_0)\, \mathbf{x}_0\mathbf{x}_0'$$

y el segundo término de esta igualdad tiende a cero con ε, basta usar el TCM para obtener que

$$\lim_{\varepsilon\searrow 0} E_{H\varepsilon}\left\{w_{H\varepsilon}\mathbf{x}\mathbf{x}'\right\} = h_1(t^*)\,\Sigma. \tag{A.14}$$

Consideremos primero el caso $t^* < \infty$. Definamos $\chi(u) = w(u)\, u$. Esta función tiene derivada continua con soporte compacto. Haciendo un desarrollo de Taylor obtenemos que

$$\chi\left(\frac{y-\mathbf{x}'\mathbf{T}_0(H\varepsilon)}{t_{H\varepsilon}S(H\varepsilon)}\right) = \chi\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}}{t_{H\varepsilon}S(H\varepsilon)}\right) - \chi'\left(\xi(\mathbf{x}, y, \varepsilon)\right)\frac{\mathbf{x}'\left(\mathbf{T}_0(H\varepsilon)-\boldsymbol{\theta}\right)}{t_{H\varepsilon}S(H\varepsilon)}$$

donde $\lim_{\varepsilon\searrow 0}\xi(\mathbf{x}, y, \varepsilon) = (y-\mathbf{x}'\boldsymbol{\theta})/(\sigma t^*)$. Usando la imparidad de χ y el TCM (que se puede usar porque χ es acotada y G_0 tiene segundos momentos finitos) se obtiene que

$$\lim_{\varepsilon\searrow 0}\frac{1}{\varepsilon} E_{H_0}\left\{\chi\left(\frac{y-\mathbf{x}'\mathbf{T}_0(H\varepsilon)}{t_{H\varepsilon}S(H\varepsilon)}\right)\mathbf{x}\right\} = -E_{H_0}\left\{\chi'\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}}{\sigma t^*}\right)\mathbf{x}\mathbf{x}'\right\}\frac{1}{\sigma t^*}\,\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0, y_0) \tag{A.15}$$

donde

$$E_{H_0}\left\{\chi'\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}}{\sigma t^*}\right)\mathbf{x}\mathbf{x}'\right\} = \{h_2(t^*)+h_1(t^*)\}\,\Sigma. \tag{A.16}$$

Pero

$$E_{H_0}\left\{\chi\left(\frac{y-\mathbf{x}'\mathbf{T}_0(H\varepsilon)}{t_{H\varepsilon}S(H\varepsilon)}\right)\mathbf{x}\right\} =$$

$$E_{H_0}\{w_{H\varepsilon}\mathbf{x}(y-\mathbf{x}'\boldsymbol{\theta})\}\frac{1}{t_{H\varepsilon}S(H\varepsilon)} - E_{H_0}\{w_{H\varepsilon}\mathbf{x}\mathbf{x}'\}\frac{1}{t_{H\varepsilon}S(H\varepsilon)}(\mathbf{T}_0(H\varepsilon)-\boldsymbol{\theta})$$

y como

$$\lim_{\varepsilon\searrow 0}\frac{1}{\varepsilon}E_{H_0}\{w_{H\varepsilon}\mathbf{x}\mathbf{x}'\}\frac{1}{t_{H\varepsilon}S(H\varepsilon)}(\mathbf{T}_0(H\varepsilon)-\boldsymbol{\theta}) = h_1(t^*)\,\Sigma\frac{1}{\sigma t^*}\,\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0,y_0)$$

de (A.15) y (A.16) se deduce que

$$\lim_{\varepsilon\searrow 0}\frac{1}{\varepsilon}E_{H_0}\{w_{H\varepsilon}\mathbf{x}(y-\mathbf{x}'\boldsymbol{\theta})\} = -h_2(t^*)\,\Sigma\,\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0,y_0).$$

El caso $t^*=\infty$ es más sencillo porque $E_{F_0}\{y^2\}<\infty$. Usando el TVM se tiene que

$$\left|w\left(\frac{y-\mathbf{x}'\mathbf{T}_0(H\varepsilon)}{t_{H\varepsilon}S(H\varepsilon)}\right)-w\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}}{t_{H\varepsilon}S(H\varepsilon)}\right)\right| \le |w'|_\infty\frac{\|\mathbf{x}\|\;\|\mathbf{T}_0(H\varepsilon)-\boldsymbol{\theta}\|}{t_{H\varepsilon}S(H\varepsilon)}.$$

Por lo tanto

$$|E_{H_0}\{w_{H\varepsilon}\mathbf{x}(y-\mathbf{x}'\boldsymbol{\theta})\}| \le |w'|_\infty E_{G_0}\{\|\mathbf{x}\|^2\}E_{H_0}\{|y-\mathbf{x}'\boldsymbol{\theta}|\}\frac{\|\mathbf{T}_0(H\varepsilon)-\boldsymbol{\theta}\|}{t_{H\varepsilon}S(H\varepsilon)}$$

y entonces

$$\lim_{\varepsilon\searrow 0}\frac{1}{\varepsilon}E_{H_0}\{w_{H\varepsilon}\mathbf{x}(y-\mathbf{x}'\boldsymbol{\theta})\}=0.$$

En cualquiera de los dos casos, se tiene que

$$\lim_{\varepsilon\searrow 0}\frac{1}{\varepsilon}E_{H\varepsilon}\{w_{H\varepsilon}\mathbf{x}(y-\mathbf{x}'\boldsymbol{\theta})\} = -h_2(t^*)\,\Sigma\,\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0,y_0)+w\left(\frac{y_0-\mathbf{x}_0'\boldsymbol{\theta}}{\sigma t^*}\right)\mathbf{x}_0(y_0-\mathbf{x}_0'\boldsymbol{\theta}) \tag{A.17}$$

ya que $h_2(\infty)=0$. De (A.14) y (A.17) se obtiene, como queríamos probar, que

$$\lim_{\varepsilon\searrow 0}\frac{1}{\varepsilon}(\mathbf{T}_1(H_\varepsilon)-\boldsymbol{\theta}) = \left\{w\left(\frac{y_0-\mathbf{x}_0'\boldsymbol{\theta}}{\sigma t^*}\right)\Sigma^{-1}\mathbf{x}_0(y_0-\mathbf{x}_0'\boldsymbol{\theta}) - h_2(t^*)\,\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0,y_0)\right\}/h_1(t^*).$$

B. Resultados asintóticos

En esta sección damos las demostraciones y algunos resultados auxiliares necesarios para la Sección 4. La notación será algo diferente a la del resto de la Tesis. Supondremos que los errores $u_i(\omega)$ son v.a.i.i.d. definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{S}, \mathbb{P})$. El operador $\mathbb{E}$ es la esperanza de una v.a. $\mathcal{S}$-medible con respecto a la probabilidad $\mathbb{P}$. Se utilizarán herramientas de procesos estocásticos que pueden hallarse en Pollard (1990). Para facilitar la lectura, se transcriben a continuación las principales definiciones y resultados que se usarán, siguiendo la numeración del libro de Pollard.

Definición 3.3. El *packing number* $D(\varepsilon, T_0)$ para un conjunto T_0 de un espacio métrico se define como el mayor m para el cual existen puntos $t_1, \ldots, t_m$ en T_0 con $d(t_i, t_j) > \varepsilon$ para $i \neq j$. Cuando d sea la distancia euclídea en $\mathbb{R}^n$ usaremos la notación $D_2(\varepsilon, T_0)$.

Definición 4.2. Para cada $\mathbf{t} \in \mathbb{R}^k$ y cada subconjunto J de $\{1, \ldots, k\}$ se define el J-ésimo ortante alrededor de $\mathbf{t}$ como el conjunto de los $\mathbf{x} \in \mathbb{R}^k$ tales que

$$\begin{aligned} x_i &> t_i \text{ si } i \in J \\ x_i &< t_i \text{ si } i \in J^C. \end{aligned}$$

Se dirá que un subconjunto de $\mathbb{R}^k$ ocupa el J-ésimo ortante de $\mathbf{t}$ cuando contenga al menos un punto de ese ortante. Se dirá que un subconjunto de $\mathbb{R}^k$ *rodea* $\mathbf{t}$ si ocupa todos los (2^k) ortantes de $\mathbf{t}$.

Definición 4.3. Dado $\mathbf{x} \in \mathbb{R}^n$, una proyección coordenada de $\mathbf{x}$ es $(x_{i_1}, \ldots, x_{i_k})$, donde $\{i_1, \ldots, i_k\}$ es un conjunto de k subíndices no necesariamente distintos; si son todos distintos, la proyección se dirá propia. Un subconjunto $\mathcal{F} \subset \mathbb{R}^n$ tiene *seudodimensión* a lo sumo V si para cada $\mathbf{t} \in \mathbb{R}^{V+1}$ ninguna proyección coordenada propia de $\mathcal{F}$ puede rodear $\mathbf{t}$.

Corolario 4.10. Si $\mathcal{F}$ es un subconjunto acotado de $\mathbb{R}^n$ con envolvente $\mathbf{F}$ (es decir que para todo $\mathbf{f} \in \mathcal{F}$ se tiene que $f_i \leq F_i \; \forall i = 1, \ldots, n$) y con seudodimensión a lo sumo V, entonces existen constantes A y W que dependen solamente de V y tales que

$$D_2(\varepsilon \|\boldsymbol{\alpha} \odot \mathbf{F}\|, \boldsymbol{\alpha} \odot \mathcal{F}) \leq A\varepsilon^{-W} \quad \forall \varepsilon \in (0, 1]$$

y para todo vector $\boldsymbol{\alpha}$ de coordenadas no negativas. (El operador $\odot$ es el producto coordenada a coordenada.)

Definición 7.9. Un arreglo triangular de procesos estocásticos

$$\{f_{ni}(\omega, t) : t \in T, \; 1 \leq i \leq k_n\}$$

es *manejable* con respecto a las envolventes $\mathbf{F}_n(\omega)$ si existe una función determinística λ (llamada cota de capacidad) tal que

$$\int_0^1 \sqrt{\ln \lambda(x)}dx < \infty$$

y

$$D_2\left(x \left\|\boldsymbol{\alpha} \odot \mathbf{F}_n(\omega)\right\|, \boldsymbol{\alpha} \odot \mathcal{F}_{n\omega}\right) \leq \lambda(x)$$

para todo $x \in (0,1]$, para todo vector $\boldsymbol{\alpha}$ de coordenadas no negativas, para todo $\omega \in \Omega$, y para todo n, siendo

$$\mathcal{F}_{n\omega} = \left\{\left(f_{n1}(\omega, t), \ldots, f_{nk_n}(\omega, t)\right) : t \in T\right\}.$$

En el caso particular en que $\lambda(x) = Ax^{-W}$ como en el Corolario 4.10, el proceso se dice *euclídeo.*

Lema B.1. *Si* $\{f_{ni}(\omega, t)\}$ *es manejable con envolventes* $\mathbf{F}_n(\omega)$, *entonces para* $1 \leq p < \infty$ *se tiene que*

$$\mathbb{E}\left\{\sup_{t\in T}\left|\sum_{i=1}^{k_n}\left[f_{ni}(\omega, t) - \mathbb{E}f_{ni}(\cdot, t)\right]\right|^p\right\} \leq (18C_p\Lambda(1))^p\, \mathbb{E}\left(\|\mathbf{F}_n\|^p\right)$$

donde

$$\Lambda(t) = \int_0^t \sqrt{\ln \lambda(x)}dx$$

y C_p *depende solamente de* p.

Nota: El Lema B.1 es simplemente la desigualdad maximal 7.10 de Pollard (1990). Es importante aclarar que $\Lambda(1)$ depende únicamente de la seudodimensión V. La constante C_p que aparece es la que cumple que $\|Z\|_p \leq C_p \|Z\|_\Psi$, con $\|Z\|_\Psi$ la norma de Orlicz, definida en la página 3 de Pollard, y $\Psi(x) = \exp(x^2)/5$. Concretamente, como se explica en el último renglón de la página 12 de Pollard, C_p debe cumplir que $|x|^p \leq \Psi(C_p x)$. En definitiva, ni C_p ni $\Lambda(1)$ dependen del conjunto T que indexa el proceso.

Lo que vamos a usar nosotros esencialmente es la desigualdad maximal dada en el Lema B.1 y el concepto de manejabilidad de un proceso, así que el resultado que damos a continuación será la base de todas las demostraciones de esta sección.

Lema B.2. *Dadas dos sucesiones no aleatorias,* $\{u_{ni}\}_{i=1}^n$ *en* $\mathbb{R}$ *y* $\{\mathbf{x}_{ni}\}_{i=1}^n$ *en* $\mathbb{R}^p$, *y* $w(u)$ *una función que cumple la condición W1, sea*

$$f_{ni}(\alpha, \boldsymbol{\beta}, \gamma) = w\left(\alpha u_{ni} + \boldsymbol{\beta}'\mathbf{x}_{ni} + \gamma\right).$$

Entonces el subconjunto de $\mathbb{R}^n$

$$\mathcal{F}_n = \left\{\left(f_{n1}(\alpha, \boldsymbol{\beta}, \gamma), \ldots, f_{nn}(\alpha, \boldsymbol{\beta}, \gamma)\right) : (\alpha, \boldsymbol{\beta}, \gamma) \in \mathbb{R}^{p+2}\right\}$$

tiene seudodimensión a lo sumo $p + 2$ *para todo* n.

Demostración: Sea

$$g_{ni}(\alpha,\boldsymbol{\beta},\gamma)=|\alpha u_{ni}+\boldsymbol{\beta}'\mathbf{x}_{ni}+\gamma|$$

y

$$\mathcal{G}_n=\left\{(g_{n1}(\alpha,\boldsymbol{\beta},\gamma),\ldots,g_{nn}(\alpha,\boldsymbol{\beta},\gamma)):(\alpha,\boldsymbol{\beta},\gamma)\in\mathbb{R}^{p+2}\right\}.$$

Es claro que $\mathcal{G}_n$ es un subconjunto de un subespacio vectorial de $\mathbb{R}^n$ de dimensión $p+2$. Dado $\mathbf{t}\in\mathbb{R}^{p+3}$ y un conjunto I de $p+3$ números naturales distintos tomados entre $\{1,\ldots,n\}$, vamos a construir un subconjunto $J\subseteq I$ tal que ningún $\mathbf{f}\in\mathcal{F}_n$ pueda cumplir

$$f_i<t_i\;\forall i\in J \text{ y } f_i>t_i\;\forall i\in I\setminus J \tag{B.1}$$

y entonces $\mathcal{F}_n$ tendrá pseudodimensión a lo sumo $p+2$. Claramente sólo basta considerar el caso $t_i\in(0,1)$ para todo i (porque si algún t_i es 0 o 1 entonces (B.1) sólo podría cumplirse para $J=\emptyset$ o $J=I$ respectivamente). Para $t\in(0,1)$ definamos entonces la inversa generalizada

$$w^{-1}(t)=\min\{u:w(u)\leq t\}.$$

Notar que como $w(u)$ es continua a derecha para $u\geq 0$, se tiene que $w(w^{-1}(t))=t$ y entonces $w(a)>b$ implica que $a<w^{-1}(b)$. Si (B.1) se cumpliera para algún $\mathbf{f}\in\mathcal{F}_n$, entonces debería haber un $\mathbf{g}\in\mathcal{G}_n$ tal que

$$g_i\geq t_i^*\;\forall i\in J \text{ y } g_i<t_i^*\;\forall i\in I\setminus J \tag{B.2}$$

para $t_i^*=w^{-1}(t_i)$. Sin embargo, como $\mathcal{G}_n$ es un subconjunto de un subespacio de dimensión $p+2$, existe un vector no nulo $\mathbf{v}\in\mathbb{R}^{p+3}$, independiente de $\mathbf{t}$, tal que $\sum_{i\in I}v_ig_i=0$ para todo $\mathbf{g}\in\mathcal{G}_n$. Sin pérdida de generalidad podemos suponer que $v_i>0$ para algún $i\in I$. Si $\mathbf{v}'\mathbf{t}^*\leq 0$, definimos

$$J=\{i\in I:v_i\leq 0\}$$

y llegamos a la contradicción

$$0=\sum_{i\in I}v_ig_i<\mathbf{v}'\mathbf{t}^*\leq 0.$$

Si $\mathbf{v}'\mathbf{t}^*>0$ reemplazamos $v_i\leq 0$ por $v_i>0$ en la definición de J y llegamos a otra contradicción. Entonces ningún $\mathbf{f}\in\mathcal{F}_n$ puede cumplir (B.1) para este J, como queríamos demostrar. ■

Demostración del Lema 4.1: Por simplicidad, trabajaremos con el estimador inicial centrado, $\boldsymbol{\beta}_{0n}=\mathbf{T}_{0n}-\boldsymbol{\theta}$. Consideremos la familia de funciones

$$f_i(\omega,\mathbf{v},s)=\mathbb{I}\left(|u_i(\omega)-\mathbf{x}_i'\mathbf{v}|\leq s\right) \text{ para } (\mathbf{v},s)\in T\triangleq\mathbb{R}^p\times\mathbb{R}^+.$$

Para un $\omega \in \Omega$ fijo,

$$\{(f_1(\omega, \mathbf{v}, s), \ldots, f_n(\omega, \mathbf{v}, s)) : (\mathbf{v}, s) \in T\}$$

está incluído en un conjunto de la forma $\mathcal{F}_n$ considerado en el Lema B.2, así que tiene pseudimensión a lo sumo $p+2$. Entonces por el Corolario 4.10 de Pollard el proceso $\{f_i(\omega, \mathbf{v}, s) : (\mathbf{v}, s) \in T\}$ es euclídeo (ver Definición 7.9). Una envolvente para este proceso es sencillamente $\mathbf{F} = \mathbf{1}_n$. Si definimos

$$S_n(\omega, \mathbf{v}, s) = \sum_{i=1}^{n} f_i(\omega, \mathbf{v}, s) \text{ y } M_n(\mathbf{v}, s) = \sum_{i=1}^{n} \mathbb{E} f_i(\cdot, \mathbf{v}, s)$$

es claro que

$$F_n^+(t) = \frac{1}{n} S_n(\omega, \boldsymbol{\beta}_{0n}, s_n t).$$

Usando el Lema B.1 con $p = 2$ obtenemos que

$$\mathbb{E}\left\{\sup_{(\mathbf{v},s)\in T} |S_n(\cdot, \mathbf{v}, s) - M_n(\mathbf{v}, s)|^2\right\} \leq Cn$$

y entonces, aplicando simplemente la desigualdad de Tchebyshev se obtiene que

$$\sup_{t\geq 0} \frac{1}{\sqrt{n}} |S_n(\omega, \boldsymbol{\beta}_{0n}, s_n t) - M_n(\boldsymbol{\beta}_{0n}, s_n t)| = O_p(1).$$

Esto implica que

$$\sup_{t\geq 0} \left| F_n^+(t) - \frac{1}{n} M_n(\boldsymbol{\beta}_{0n}, s_n t) \right| = o_p(1),$$

así que para probar que $|F_n^+ - F_0^+|_\infty = o_p(1)$ sólo falta demostrar que

$$\sup_{t\geq 0} \left| \frac{1}{n} M_n(\boldsymbol{\beta}_{0n}, s_n t) - F_0^+(t) \right| = o_p(1).$$

Para esto basta ver que

$$\sup_{t\geq 0} \left| \frac{1}{n} M_n(\boldsymbol{\beta}_{0n}, s_n t) - F_0^+\left(\frac{s_n t}{\sigma}\right) \right| = o_p(1) \tag{B.3}$$

y que

$$\sup_{t\geq 0} \left| F_0^+\left(\frac{s_n t}{\sigma}\right) - F_0^+(t) \right| = o_p(1). \tag{B.4}$$

Notemos que

$$\frac{1}{n}M_n\left(\boldsymbol{\beta}_{0n}, s_n t\right) = \frac{1}{n}\sum_{i=1}^{n}\left[F_0\left(\frac{s_n t + \mathbf{x}_i'\boldsymbol{\beta}_{0n}}{\sigma}\right) - F_0\left(\frac{-s_n t + \mathbf{x}_i'\boldsymbol{\beta}_{0n}}{\sigma}\right)\right].$$

Haciendo un desarrollo de Taylor de orden 1 se obtiene que

$$F_0\left(\frac{s_n t + \mathbf{x}_i'\boldsymbol{\beta}_{0n}}{\sigma}\right) = F_0\left(\frac{s_n t}{\sigma}\right) + f_0\left(\xi_{in}\right)\frac{\mathbf{x}_i'\boldsymbol{\beta}_{0n}}{\sigma} \tag{B.5}$$

y entonces

$$\left|\frac{1}{n}\sum_{i=1}^{n}F_0\left(\frac{s_n t + \mathbf{x}_i'\boldsymbol{\beta}_{0n}}{\sigma}\right) - F_0\left(\frac{s_n t}{\sigma}\right)\right| \leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\left(\|\mathbf{x}_i\| > K\right) + |f_0|_\infty \frac{K}{\sigma}\|\boldsymbol{\beta}_{0n}\|.$$

Exactamente la misma cota vale para

$$\left|\frac{1}{n}\sum_{i=1}^{n}F_0\left(\frac{-s_n t + \mathbf{x}_i'\boldsymbol{\beta}_{0n}}{\sigma}\right) - F_0\left(\frac{-s_n t}{\sigma}\right)\right|$$

y entonces (B.3) queda probado. Para (B.4) se usa el hecho que F_0^+ es continua, monótona y acotada, por lo que la convergencia puntual en probabilidad de $F_0^+\left(s_n t/\sigma\right)$ a $F_0^+\left(t\right)$ vale uniformemente en t.

Ahora bien, como

$$|d_n - d^*| \leq \left|F_n^+ - F_0^+\right|_\infty$$

tenemos que $d_n \longrightarrow d^*$ en probabilidad, como queríamos.

Para demostrar que $\bar{t}_n \longrightarrow \bar{t}^*$ en probabilidad bastaría probar que $F_0^+\left(\bar{t}_n\right) \longrightarrow 1 - d^*$ en probabilidad. La demostración de esto último es idéntica a la del Lema A.2 parte (ii). ∎

Demostración del Lema 4.2: Consideremos nuevamente $f_i\left(\omega, \mathrm{v}, s\right)$ como en la demostración del Lema 4.1. Ahí probamos que

$$\sqrt{n}\sup_{t\geq 0}\left|F_n^+\left(t\right) - \frac{1}{n}M_n\left(\boldsymbol{\beta}_{0n}, s_n t\right)\right| = O_p\left(1\right),$$

así que solo resta ver que

$$\sup_{t\geq 0}\left|\frac{1}{n}M_n\left(\boldsymbol{\beta}_{0n}, s_n t\right) - F_0^+\left(t\right)\right| = O_p\left(n^{-1/2}\right).$$

Como en el lema anterior, probaremos por separado que

$$\sup_{t\geq 0}\left|\frac{1}{n}M_n\left(\boldsymbol{\beta}_{0n}, s_n t\right) - F_0^+\left(\frac{s_n t}{\sigma}\right)\right| = O_p\left(n^{-1/2}\right) \tag{B.6}$$

y que

$$\sup_{t\geq 0}\left|F_0^+\left(\frac{s_n t}{\sigma}\right)-F_0^+(t)\right|=O_p\left(n^{-1/2}\right). \qquad \text{(B.7)}$$

Haremos desarrollos de Taylor como antes, pero de un orden más. Se tiene entonces

$$F_0\left(\frac{s_n t+x_i'\beta_{0n}}{\sigma}\right)-F_0\left(\frac{s_n t}{\sigma}\right)=f_0\left(\frac{s_n t}{\sigma}\right)\frac{x_i'\beta_{0n}}{\sigma}+f_0'(\xi_{in})\frac{1}{2}\left(\frac{x_i'\beta_{0n}}{\sigma}\right)^2 \qquad \text{(B.8)}$$

y

$$F_0\left(\frac{-s_n t+x_i'\beta_{0n}}{\sigma}\right)-F_0\left(\frac{-s_n t}{\sigma}\right)=f_0\left(\frac{-s_n t}{\sigma}\right)\frac{x_i'\beta_{0n}}{\sigma}+f_0'(\eta_{in})\frac{1}{2}\left(\frac{x_i'\beta_{0n}}{\sigma}\right)^2 \qquad \text{(B.9)}$$

Ahora restamos (B.9) de (B.8) y, usando la simetría de f_0, se obtiene que

$$\frac{1}{n}M_n(\beta_{0n},s_n t)-F_0^+\left(\frac{s_n t}{\sigma}\right)=\frac{1}{n}\sum_{i=1}^n\left[f_0'(\xi_{in})-f_0'(\eta_{in})\right]\frac{1}{2}\left(\frac{x_i'\beta_{0n}}{\sigma}\right)^2$$

y entonces

$$\sup_{t\geq 0}\left|\frac{1}{n}M_n(\beta_{0n},s_n t)-F_0^+\left(\frac{s_n t}{\sigma}\right)\right|\leq|f_0'|_\infty\frac{1}{\sigma^2}\|\beta_{0n}\|^2\frac{1}{n}\sum_{i=1}^n\|x_i\|^2=O_p\left(n^{-2\tau}\right)$$

que implica (B.6) por ser $\tau>1/4$. Para probar (B.7) hacemos otros desarrollos de Taylor similares a (B.8) y (B.9), usamos nuevamente la simetría de f_0 y obtenemos

$$F_0^+\left(\frac{s_n t}{\sigma}\right)=F_0^+(t)+\frac{1}{2}\left[f_0'(\xi_n)-f_0'(\eta_n)\right]\left(\frac{s_n t}{\sigma}-t\right)^2$$

con ξ_n entre $s_n t/\sigma$ y t y η_n entre $-s_n t/\sigma$ y $-t$. Por lo tanto $|t/\xi_n|$ y $|t/\eta_n|$ están entre 1 y σ/s_n, de modo que

$$\sup_{t\geq 0}\left|F_0^+\left(\frac{s_n t}{\sigma}\right)-F_0^+(t)\right|\leq\left|u^2 f_0'(u)\right|_\infty\left(\max\left\{\frac{\sigma}{s_n},1\right\}\right)^2\left|\frac{s_n}{\sigma}-1\right|^2=O_p\left(n^{-2\tau}\right)$$

que implica (B.7) por ser $\tau>1/4$. ∎

Para los siguientes lemas definimos $\nu_n=(s_n t_n)^{-1}$. Notar que $\nu_n=o_p(1)$ cuando $t^*=\infty$.

Lema B.3. *Si se cumplen A1-A5 y W1, entonces*

$$\frac{1}{n}\sum_{i=1}^n\{w_i-h_1(t^*)\}\,x_i x_i'=o_p(1).$$

Demostración: Observemos primero que dado $K > 0$,

$$\left| \frac{1}{n} \sum_{i=1}^{n} \{w_i - h_1(t^*)\} \mathbf{x}_i \mathbf{x}_i' \right| \leq$$

$$\left| \frac{1}{n} \sum_{i=1}^{n} \{w_i - h_1(t^*)\} \mathbf{x}_i \mathbf{x}_i' \mathbb{I}(\|\mathbf{x}_i\| \leq K) \right| + 2\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x}_i\|^2 \mathbb{I}(\|\mathbf{x}_i\| > K). \quad \text{(B.10)}$$

Dado $\varepsilon > 0$ podemos elegir K tal que

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x}_i\|^2 \mathbb{I}(\|\mathbf{x}_i\| > K) < \varepsilon.$$

Así que solamente debemos mostrar que el primer término en (B.10) es $o_p(1)$. Fijemos dos coordenadas j y k y consideremos los procesos

$$f_i(\omega, \mathbf{v}, s) = w\left(s\left(u_i(\omega) - \mathbf{x}_i'\mathbf{v}\right)\right) x_{ij} x_{ik} \mathbb{I}(\|\mathbf{x}_i\| \leq K)$$

para $(\mathbf{v}, s) \in T = \mathbb{R}^p \times \mathbb{R}^+$. Nuevamente podemos aplicar el resultado del Lema B.2 a esta familia de procesos porque el factor $x_{ij} x_{ik} \mathbb{I}(\|\mathbf{x}_i\| \leq K)$ no altera la pseudodimensión. Entonces la sucesión de procesos $\{f_i(\omega, \mathbf{v}, s) : (\mathbf{v}, s) \in T\}$ es manejable para la envolvente $\mathbf{F} = K^2 \mathbf{1}_n$ y podemos aplicar nuevamente la desigualdad maximal del Lema B.1 con $p = 2$ para obtener

$$\mathbb{E}\left\{ \sup_{(\mathbf{v},s)\in T} \left| \sum_{i=1}^{n} f_i(\omega, \mathbf{v}, s) - M_n(\mathbf{v}, s) \right|^2 \right\} \leq CnK^4 \quad \text{(B.11)}$$

donde

$$M_n(\mathbf{v}, s) = \sum_{i=1}^{n} \mathbb{E} f_i(\cdot, \mathbf{v}, s) = \sum_{i=1}^{n} \int w\left(s\left(\sigma u - \mathbf{x}_i'\mathbf{v}\right)\right) f_0(u)\, du\, x_{ij} x_{ik} \mathbb{I}(\|\mathbf{x}_i\| \leq K).$$

De (B.11) se deduce que

$$\frac{1}{n} \left| \sum_{i=1}^{n} w_i x_{ij} x_{ik} \mathbb{I}(\|\mathbf{x}_i\| \leq K) - M_n(\beta_{0n}, \nu_n) \right| = o_p(1) \quad \text{(B.12)}$$

así que sólo resta probar que

$$\frac{1}{n} \left| M_n(\beta_{0n}, \nu_n) - h_1(t^*) \sum_{i=1}^{n} x_{ij} x_{ik} \mathbb{I}(\|\mathbf{x}_i\| \leq K) \right| = o_p(1). \quad \text{(B.13)}$$

Como
$$g(s,t)=\int w(s(\sigma u-t))\,f_0(u)\,du$$
es continua (una consecuencia del TCM), es fácil ver que
$$\left|\frac{1}{n}\sum_{i=1}^{n}\left\{g(\nu_n,\mathbf{x}_i'\boldsymbol{\beta}_{0n})-g\left(\frac{1}{\sigma t^*},0\right)\right\}x_{ij}x_{ik}\mathbb{I}(\|\mathbf{x}_i\|\leq K)\right|=o_p(1)$$
que es otra forma de escribir (B.13). ■

Lema B.4. *Si se cumplen A1-A5 y W1-W2, y llamamos* $\rho(u)=w'(u)\,u$, *entonces*
$$K_n\triangleq\frac{1}{n}\sum_{i=1}^{n}\{\rho(\nu_n u_i)-h_2(t^*)\}\,\mathbf{x}_i\mathbf{x}_i'=o_p(1)\,.$$

Demostración: Podemos escribir $K_n=A_n+B_n$ con
$$\begin{aligned}A_n&=\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\{\rho(\nu_n u_i)-\rho(u_i/(\sigma t^*))\}\,\mathbf{x}_i\mathbf{x}_i'\\B_n&=\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\{\rho(u_i/(\sigma t^*))-h_2(t^*)\}\,\mathbf{x}_i\mathbf{x}_i'.\end{aligned}$$
Veamos primero que $A_n=o_p(1)$. Dado $\varepsilon>0$, por A5 existe $K>0$ tal que
$$\limsup_{n\to\infty}2\,|\rho|_\infty\,\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i\|^2\,\mathbb{I}(\|\mathbf{x}_i\|>K)<\frac{\varepsilon}{2}\tag{B.14}$$
y entonces basta con probar que
$$\lim_{n\to\infty}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\rho(\nu_n u_i)-\rho\left(\frac{u_i}{\sigma t^*}\right)\right|>\frac{\varepsilon}{2K^2}\right)=0.$$
Como existe $c>0$ tal que $\mathbb{P}(|u|>c)<\varepsilon/(8\,|\rho|_\infty\,K^2)$,
$$\lim_{n\to\infty}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\rho(\nu_n u_i)-\rho\left(\frac{u_i}{\sigma t^*}\right)\right|\mathbb{I}(|u_i|>c)>\frac{\varepsilon}{4K^2}\right)=0.$$
Por otra parte, como $\rho(u)$ es continua con soporte $[-1,1]$, es uniformemente continua y entonces existe $\delta=\delta(\varepsilon)$ tal que $|\rho(u)-\rho(v)|<\varepsilon/(8K^2)$ si $|u-v|<\delta$. Por lo tanto
$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\rho(\nu_n u_i)-\rho\left(\frac{u_i}{\sigma t^*}\right)\right|\mathbb{I}(|u_i|\leq c)>\frac{\varepsilon}{4K^2}\right)\leq\mathbb{P}\left(\left|\nu_n-\frac{1}{\sigma t^*}\right|\geq\frac{\delta}{c}\right)$$

que tiende a cero.

Para ver que $B_n = o_p(1)$, nuevamente usamos (B.14) y entonces basta probar que

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\rho\left(\frac{u_i}{\sigma t^*}\right) - h_2(t^*)\right\} x_{ij}x_{ik}\mathbb{I}(\|\mathbf{x}_i\| \le K)\right| > \frac{\varepsilon}{2}\right) = 0 \qquad \text{(B.15)}$$

para cada par de coordenadas j y k. Pero como

$$\mathbb{E}\left[\left\{\rho\left(\frac{u_i}{\sigma t^*}\right) - h_2(t^*)\right\} x_{ij}x_{ik}\mathbb{I}(\|\mathbf{x}_i\| \le K)\right] = 0$$

y

$$\mathbb{V}\left[\left\{\rho\left(\frac{u_i}{\sigma t^*}\right) - h_2(t^*)\right\} x_{ij}x_{ik}\mathbb{I}(\|\mathbf{x}_i\| \le K)\right] \le \left(2\,|\rho|_\infty K^2\right)^2$$

usando la desigualdad de Tchebyschev se tiene (B.15). ■

Por simplicidad, en los teoremas siguientes trabajaremos con las variables explicativas estandarizadas:

$$\mathbf{z}_{ni} = \frac{1}{\sqrt{n}}\Sigma_n^{-1/2}\mathbf{x}_i.$$

Notar que $\sum_{i=1}^{n} \mathbf{z}_{ni}\mathbf{z}_{ni}' = \mathrm{I}$. También, que A2 y A5 implican que

$$\lim_{n\to\infty}\sum_{i=1}^{n}\|\mathbf{z}_{ni}\|^2\,\mathbb{I}(\|\mathbf{z}_{ni}\| > \varepsilon) = 0$$

para todo $\varepsilon > 0$.

Lema B.5. *Consideremos el proceso*

$$\mathbf{W}_n(s) = \sum_{i=1}^{n}\left\{w(su_i)\,u_i\mathrm{I} - h_2(t^*)\,\Gamma\psi\left(\frac{u_i}{\sigma}\right)\right\}\mathbf{z}_{ni}$$

para $s \in T \triangleq \left[(2\sigma t^*)^{-1}, (\sigma t^*)^{-1} + 1\right]$, Γ de $p \times p$ simétrica y ψ impar y acotada. Si se cumplen A1-A5 y W1, entonces

$$\mathbf{W}_n(\nu_n) \to_D N_p(\mathbf{0}, V(t^*))$$

donde $V(t^*)$ está dada por (4.3).

Demostración: Si tomamos $\boldsymbol{\alpha} \in \mathbb{R}^p$ entonces $\boldsymbol{\alpha}'\mathbf{W}_n(s) = \sum_{i=1}^{n} f_{in}(\omega, s)$, donde

$$f_{in}(\omega, s) = w(su_i)\,u_i\boldsymbol{\alpha}'\mathbf{z}_{ni} - h_2(t^*)\,\psi\left(\frac{u_i}{\sigma}\right)\boldsymbol{\alpha}'\Gamma\mathbf{z}_{ni}.$$

Invocamos nuevamente el Lema B.2 para esta familia, ya que el segundo término de f_{in} es fijo con respecto al índice s. El arreglo triangular de procesos $\{f_{in}(\omega, s) : s \in T\}$ es manejable, con envolventes $\mathbf{F}_n$ dadas por

$$F_{ni} = 2\sigma t^* \|\alpha\| \|\mathbf{z}_{ni}\| + |h_2(t^*)| \|\alpha' \Gamma\| |\psi|_\infty \|\mathbf{z}_{ni}\|$$

cuando $t^* < \infty$, o

$$F_{ni}(\omega) = |u_i| \|\alpha\| \|\mathbf{z}_{ni}\|$$

cuando $t^* = \infty$. En este último caso u_i tiene varianza finita, así que las envolventes tienen cuadrado integrable en ambos casos. Notar también que $\mathbb{E} f_{in}(\cdot, s) = 0$. Ahora definamos la covariancia

$$H(s,t) =$$

$$\sigma^2 E_{F_0}\left[\left\{w(s\sigma u) u\alpha' - h_2(t^*)\frac{1}{\sigma}\psi(u)\alpha'\Gamma\right\}\left\{w(t\sigma u) u\alpha' - h_2(t^*)\frac{1}{\sigma}\psi(u)\alpha'\Gamma\right\}'\right]$$

y la seudométrica en T

$$\rho(s,t) = \sigma \|\alpha\| E_{F_0}^{1/2}\left\{(w(s\sigma u) - w(t\sigma u))^2 u^2\right\}.$$

Llamaremos $U\rho(T)$ al subconjunto de funciones reales sobre T que son acotadas y uniformemente continuas respecto a ρ. Para más detalles, ver la Sección 10 de Pollard (1990).

Ahora empleamos el Teorema Central del Límite Funcional 10.6 de Pollard (1990) para obtener que $\alpha' \mathbf{W}_n$ converge en distribución a un proceso W cuya distribución es gaussiana y concentrada en $U\rho(T)$. Las fidis (distribuciones de las proyecciones finito dimensionales) de W están determinadas por las covariancias H. Notar que $H\left((\sigma t^*)^{-1}, (\sigma t^*)^{-1}\right) = \alpha' V(t^*)\alpha$. Si definimos la aplicación $\xi(x,s) = x(s)$, que es continua respecto a la distancia

$$d((x_1, s_1), (x_2, s_2)) = \max\{|x_1 - x_2|_\infty, |s_1 - s_2|\},$$

podemos aplicar el resultado del Ejemplo 10 y el Teorema de la Aplicación Continua 12, páginas 69 y 70 de Pollard (1984), para concluir que $\xi(\alpha' \mathbf{W}_n, \nu_n) \to_D \xi\left(W, (\sigma t^*)^{-1}\right)$, lo que es otra manera de escribir que $\alpha' \mathbf{W}_n(\nu_n) \to_D N(0, \alpha' V(t^*)\alpha)$ y entonces la demostración está completa. ■

Demostración del Teorema 4.4: Sea $\beta_{1n} = \mathbf{T}_{1n} - \theta$. Entonces

$$\sum_{i=1}^{n} w_i \mathbf{x}_i u_i = \sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{x}_i' \beta_{1n}. \tag{B.16}$$

Como

$$\Sigma_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \mathbf{x}_i\mathbf{x}_i'\boldsymbol{\beta}_{1n} =$$

$$\left(\Sigma_n^{-1/2}\left[\frac{1}{n}\sum_{i=1}^{n}\{w_i - h_1(t^*)\}\,\mathbf{x}_i\mathbf{x}_i'\right]\Sigma_n^{-1/2} + h_1(t^*)\,\mathrm{I}\right)\sqrt{n}\Sigma_n^{1/2}\boldsymbol{\beta}_{1n},$$

utilizamos el Lema B.3 en el lado derecho de esta igualdad y sólo nos resta probar que

$$\Sigma_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\mathbf{x}_i u_i = \mathbf{W}_n(\nu_n) + o_p(1)$$

con $\mathbf{W}_n$ como en el Lema B.5.

En primer lugar, llamemos $\chi(u) = w(u)\,u$, que es una función con derivada continua y de soporte compacto, así que $\chi'(u)$ es uniformemente continua. Haciendo un desarrollo de Taylor se tiene que

$$\chi(\nu_n(u_i - \mathbf{x}_i'\boldsymbol{\beta}_{0n})) = \chi(\nu_n u_i) - \chi'(\nu_n u_i)\,\nu_n\mathbf{x}_i'\boldsymbol{\beta}_{0n} + \nu_n R_{ni}$$

donde

$$R_{ni} = [\chi'(\nu_n u_i) - \chi'(\xi_{ni})]\,\mathbf{x}_i'\boldsymbol{\beta}_{0n}$$

con ξ_{ni} tales que $|\nu_n u_i - \xi_{ni}| \leq |\nu_n\mathbf{x}_i'\boldsymbol{\beta}_{0n}|$. Llamemos

$$\mathbf{R}_n = \sum_{i=1}^{n} R_{ni}\mathbf{x}_i.$$

Queremos llegar a que $\mathbf{R}_n = o_p\left(n^{1/2}\right)$. Como

$$\|\mathbf{R}_n\| \leq \|\boldsymbol{\beta}_{0n}\|\sum_{i=1}^{n}|\chi'(\nu_n u_i) - \chi'(\xi_{ni})|\,\|\mathbf{x}_i\|^2$$

y $\|\boldsymbol{\beta}_{0n}\| = O_p\left(n^{-1/2}\right)$, basta con probar que

$$\sum_{i=1}^{n}|\chi'(\nu_n u_i) - \chi'(\xi_{ni})|\,\|\mathbf{x}_i\|^2 = o_p(n). \tag{B.17}$$

Dado $\varepsilon > 0$ existe $K = K(\varepsilon)$ tal que

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i\|^2\,\mathbb{I}(\|\mathbf{x}_i\| > K) < \frac{\varepsilon}{4\|\chi'\|}\quad\forall n.$$

También existe $\delta = \delta(\varepsilon) > 0$ tal que $|\chi'(u) - \chi'(v)| < \varepsilon/(2K^2)$ si $|u - v| < \delta$. Entonces, si $\nu_n \|\boldsymbol{\beta}_{0n}\| < \delta/K$, tendremos que

$$\frac{1}{n}\sum_{i=1}^{n} |\chi'(\nu_n u_i) - \chi'(\xi_{ni})| \, \|\mathbf{x}_i\|^2 < \varepsilon.$$

Por lo tanto

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n} |\chi'(\nu_n u_i) - \chi'(\xi_{ni})| \, \|\mathbf{x}_i\|^2 \geq \varepsilon\right\} \leq \mathbb{P}\left\{\nu_n \|\boldsymbol{\beta}_{0n}\| \geq \frac{\delta}{K}\right\}$$

y el lado derecho de esta desigualdad tiende a cero, de donde se deduce (B.17).

Ahora bien,

$$\sum_{i=1}^{n} w_i \mathbf{x}_i u_i = \frac{1}{\nu_n}\sum_{i=1}^{n} \chi(\nu_n (u_i - \mathbf{x}_i'\boldsymbol{\beta}_{0n})) \mathbf{x}_i + \sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{0n} \tag{B.18}$$

y

$$\frac{1}{\nu_n}\sum_{i=1}^{n} \chi(\nu_n (u_i - \mathbf{x}_i'\boldsymbol{\beta}_{0n})) \mathbf{x}_i = \frac{1}{\nu_n}\sum_{i=1}^{n} \chi(\nu_n u_i) \mathbf{x}_i - \sum_{i=1}^{n} \chi'(\nu_n u_i) \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{0n} + \mathbf{R}_n.$$

Como

$$\frac{1}{\nu_n}\sum_{i=1}^{n} \chi(\nu_n u_i) \mathbf{x}_i = \sum_{i=1}^{n} w(\nu_n u_i) u_i \mathbf{x}_i$$

y

$$\sum_{i=1}^{n} \chi'(\nu_n u_i) \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{0n} = \sum_{i=1}^{n} [w'(\nu_n u_i) \nu_n u_i + w(\nu_n u_i)] \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{0n},$$

de (B.18) obtenemos que

$$\sum_{i=1}^{n} w_i \mathbf{x}_i u_i = \sum_{i=1}^{n} w(\nu_n u_i) u_i \mathbf{x}_i - \sum_{i=1}^{n} w'(\nu_n u_i) \nu_n u_i \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{0n} + \mathbf{R}_n^{(2)} \tag{B.19}$$

con

$$\mathbf{R}_n^{(2)} = \sum_{i=1}^{n} [w(\nu_n (u_i - \mathbf{x}_i'\boldsymbol{\beta}_{0n})) - w(\nu_n u_i)] \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{0n} + \mathbf{R}_n.$$

Como $w(u)$ también es uniformemente continua, igual que antes se puede demostrar que

$$\mathbf{R}_n^{(2)} = o_p\left(n^{1/2}\right).$$

Ahora, usando la notación del Lema B.4,

$$\sum_{i=1}^{n} w'(\nu_n u_i)\, \nu_n u_i \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{0n} = nK_n \boldsymbol{\beta}_{0n} + nh_2(t^*)\, \Sigma_n \boldsymbol{\beta}_{0n},$$

y $nK_n\boldsymbol{\beta}_{0n} = o_p\left(n^{1/2}\right)$. Así que, en definitiva, de (B.19) se obtiene que

$$\sum_{i=1}^{n} w_i \mathbf{x}_i u_i = \sum_{i=1}^{n} w(\nu_n u_i)\, u_i \mathbf{x}_i - nh_2(t^*)\, \Sigma_n \boldsymbol{\beta}_{0n} + o_p\left(n^{1/2}\right)$$

y entonces

$$\Sigma_n^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i u_i = \sum_{i=1}^{n} w(\nu_n u_i)\, u_i \mathbf{z}_{ni} - \sqrt{n} h_2(t^*)\, \Sigma_n^{1/2} \boldsymbol{\beta}_{0n} + o_p(1). \qquad \text{(B.20)}$$

Pero todavía no hemos terminado. Como

$$\boldsymbol{\beta}_{0n} = \Gamma_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi\left(\frac{u_i}{\sigma}\right) \mathbf{x}_i + o_p\left(n^{-1/2}\right)$$

tenemos que

$$\sqrt{n} h_2(t^*)\, \Sigma_n^{1/2} \boldsymbol{\beta}_{0n} = h_2(t^*)\, \Sigma_n^{1/2} \Gamma_n^{-1} \Sigma_n^{1/2} \sum_{i=1}^{n} \psi\left(\frac{u_i}{\sigma}\right) \mathbf{z}_{ni} + o_p(1)$$

donde $\Sigma_n^{1/2}\Gamma_n^{-1}\Sigma_n^{1/2} \to \Gamma$ en probabilidad y

$$\sum_{i=1}^{n} \psi\left(\frac{u_i}{\sigma}\right) \mathbf{z}_{ni} = O_p(1)$$

de modo que

$$\sqrt{n} h_2(t^*)\, \Sigma_n^{1/2} \boldsymbol{\beta}_{0n} = h_2(t^*)\, \Gamma \sum_{i=1}^{n} \psi\left(\frac{u_i}{\sigma}\right) \mathbf{z}_{ni} + o_p(1)$$

y entonces, finalmente, reemplazando en (B.20) obtenemos que

$$\Sigma_n^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i u_i = \mathbf{W}_n(\nu_n) + o_p(1)$$

como queríamos. ■

Antes de pasar a la demostración del Teorema 4.5, veamos la

Demostración del Lema 4.3: Mediante un desarrollo de Taylor tenemos que

$$\psi\left(\frac{u_i - \mathbf{x}_i'\boldsymbol{\beta}_{0n}}{s_n}\right) = \psi\left(\frac{u_i}{s_n}\right) - \psi'\left(\frac{u_i}{s_n}\right)\frac{\mathbf{x}_i'\boldsymbol{\beta}_{0n}}{s_n} + R_{ni}$$

donde

$$R_{ni} = \left[\psi'\left(\frac{u_i}{s_n}\right) - \psi'\left(\xi_{ni}\right)\right]\frac{\mathbf{x}_i'\boldsymbol{\beta}_{0n}}{s_n}$$

para cierto ξ_{ni} tal que

$$\left|\xi_{ni} - \frac{u_i}{s_n}\right| \leq \left|\frac{\mathbf{x}_i'\boldsymbol{\beta}_{0n}}{s_n}\right|.$$

Como el estimador inicial satisface

$$\sum_{i=1}^{n}\psi\left(\frac{u_i - \mathbf{x}_i'\boldsymbol{\beta}_{0n}}{s_n}\right)\mathbf{x}_i = \mathbf{0}$$

se tiene que

$$\Gamma_n\boldsymbol{\beta}_{0n} = \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{u_i}{s_n}\right)\mathbf{x}_i + \mathbf{R}_n$$

donde

$$\Gamma_n = \frac{1}{s_n}\sum_{i=1}^{n}\psi'\left(\frac{u_i}{s_n}\right)\mathbf{x}_i\mathbf{x}_i'$$

y

$$\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{n}R_{ni}\mathbf{x}_i.$$

Con una demostración igual a la del Lema B.4 usando ψ' en lugar de ρ (notar que ψ' es uniformemente continua bajo nuestras hipótesis) se deduce que

$$\frac{1}{n}\sum_{i=1}^{n}\left[\psi'\left(\frac{u_i}{s_n}\right) - E_{F_0}\left\{\psi'\right\}\right]\mathbf{x}_i\mathbf{x}_i' = o_p(1)$$

y entonces

$$\Sigma_n^{1/2}\Gamma_n^{-1}\Sigma_n^{1/2} \to \frac{\sigma}{E_{F_0}\left\{\psi'\right\}}\mathrm{I}$$

en probabilidad, de modo que Γ tiene la forma que queríamos. Para ver que $\mathbf{R}_n = o_p\left(n^{-1/2}\right)$ se procede como en la demostración del Teorema 4.4, usando nuevamente que ψ' es uniformemente continua. Para completar la demostración, usamos otra vez un desarrollo de Taylor de orden 1 y como $|s_n - \sigma| = O_p\left(n^{-1/2}\right)$ y $|u\psi'(u)|$ está acotado se obtiene que

$$\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{u_i}{s_n}\right)\mathbf{x}_i = \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{u_i}{\sigma}\right)\mathbf{x}_i + o_p\left(n^{-1/2}\right).$$

En definitiva,

$$\beta_{0n} = \Gamma_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{u_i}{\sigma}\right)\mathbf{x}_i + o_p\left(n^{-1/2}\right)$$

como queríamos probar. ■

Demostración del Teorema 4.5: Dado que el EMC es asintóticamente normal bajo estas hipótesis, es suficiente demostrar que

$$\sum_{i=1}^{n}(1-w_i)\,\mathbf{x}_i u_i = o_p\left(n^{1/2}\right),$$

ya que con el otro miembro de (B.16) se puede trabajar como en la demostración del Teorema 4.4.

Fijemos una coordenada j y consideremos el proceso

$$f_i(\omega, s, \mathbf{v}) = \mathbb{I}\left(s\left|u_i(\omega) - \mathbf{x}_i'\mathbf{v}\right| \geq 1\right)x_{ij}u_i(\omega) \;\text{ para }\; (s,\mathbf{v}) \in T(\delta)$$

con $T(\delta) = \{(s,\mathbf{v}) : s < \delta, \|\mathbf{v}\| < \delta\}$ para cierto $\delta > 0$ que se especificará después. Nuevamente por el Lema B.2 se tiene que esta sucesión de procesos es euclídea con envolvente $\mathbf{F}$ dada por

$$F_i(\omega) = \left[\mathbb{I}\left(|u_i(\omega)| \geq \frac{1}{2\delta}\right) + \mathbb{I}\left(\|\mathbf{x}_i\| \geq \frac{1}{2\delta^2}\right)\right]\|\mathbf{x}_i\|\,|u_i(\omega)|.$$

Es decir, podemos aplicar el Lema B.1 para obtener que

$$\mathbb{E}\left\{\sup_{T(\delta)}|S_n(\cdot,s,\mathbf{v}) - M_n(s,\mathbf{v})|^2\right\} \leq (18C_2\Lambda(1))^2\sum_{i=1}^{n}\mathbb{E}F_i^2. \qquad \text{(B.21)}$$

Recordemos que en (B.21) lo único que depende de δ son los F_i. Dados ε_1 y ε_2 positivos, podemos elegir un $\delta > 0$ tal que

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}F_i^2 < \frac{\varepsilon_1^2\varepsilon_2}{36C_2\Lambda(1)} \quad \forall n$$

y entonces de (B.21) obtenemos que

$$\frac{1}{n}\mathbb{E}\left\{\sup_{T(\delta)}|S_n(\cdot,s,\mathbf{v}) - M_n(s,\mathbf{v})|^2\right\} \leq \frac{\varepsilon_1^2\varepsilon_2}{2} \quad \forall n. \qquad \text{(B.22)}$$

Entonces

$$\mathbb{P}\left\{\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}(1-w_i)\,u_ix_{ij}-M_n(\nu_n,\boldsymbol{\beta}_{0n})\right|>\varepsilon_1\right\}\leq\mathbb{P}\{(\nu_n,\boldsymbol{\beta}_{0n})\notin T(\delta)\}+\frac{\varepsilon_2}{2}$$

y, eligiendo $n_0(\varepsilon_1,\varepsilon_2)$ tal que

$$\mathbb{P}\{(\nu_n,\boldsymbol{\beta}_{0n})\notin T(\delta)\}<\frac{\varepsilon_2}{2}\ \ \forall n\geq n_0,$$

se tiene que

$$\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}(1-w_i)\,u_ix_{ij}-M_n(\nu_n,\boldsymbol{\beta}_{0n})\right]=o_p(1)\,.$$

Ahora sólo nos queda probar que $M_n(\nu_n,\boldsymbol{\beta}_{0n})=o_p\left(n^{1/2}\right)$, siendo

$$M_n(\nu_n,\boldsymbol{\beta}_{0n})=\sum_{i=1}^{n}\left[\int\mathbb{I}\left(\frac{|u-\mathbf{x}_i'\boldsymbol{\beta}_{0n}|}{s_nt_n}\geq1\right)g\left(\frac{u}{\sigma}\right)du\right]x_{ij}$$

con $g(u)=uf_0(u)$. Haciendo un cambio de variables y usando que g es impar obtenemos que

$$\int\mathbb{I}\left(\frac{|u-\mathbf{x}_i'\boldsymbol{\beta}_{0n}|}{s_nt_n}\geq1\right)g\left(\frac{u}{\sigma}\right)du=-\int_{-s_nt_n}^{s_nt_n}g\left(\frac{u+\mathbf{x}_i'\boldsymbol{\beta}_{0n}}{\sigma}\right)du.$$

Llamemos

$$G(s,t)=\int_{-s}^{s}g\left(\frac{u+t}{\sigma}\right)du.$$

Entonces

$$\begin{aligned}\dot{G}(s,t)&\triangleq\frac{\partial}{\partial t}G(s,t)=g\left(\frac{s+t}{\sigma}\right)-g\left(\frac{-s+t}{\sigma}\right)\\ \ddot{G}(s,t)&\triangleq\frac{\partial^2}{\partial t^2}G(s,t)=\frac{1}{\sigma}g'\left(\frac{s+t}{\sigma}\right)-\frac{1}{\sigma}g'\left(\frac{-s+t}{\sigma}\right)\end{aligned}$$

son ambas continuas en (s,t). Notar además que $G(s,0)=0$ y $\dot{G}(s,0)=2g(s/\sigma)$ para todo s. Así que haciendo un desarrollo de Taylor en la variable t alrededor de $(s_nt_n,0)$ se obtiene que

$$G(s_nt_n,\mathbf{x}_i'\boldsymbol{\beta}_{0n})=2g\left(\frac{s_nt_n}{\sigma}\right)\mathbf{x}_i'\boldsymbol{\beta}_{0n}+\ddot{G}(s_nt_n,\xi_{ni})\,\frac{1}{2}\left(\mathbf{x}_i'\boldsymbol{\beta}_{0n}\right)^2$$

con $|\xi_{ni}| \leq |\mathbf{x}_i'\boldsymbol{\beta}_{0n}|$. Como

$$M_n(\nu_n, \boldsymbol{\beta}_{0n}) = -\sum_{i=1}^{n} G(s_n t_n, \mathbf{x}_i'\boldsymbol{\beta}_{0n})\, x_{ij}$$

obtenemos

$$|M_n(\nu_n, \boldsymbol{\beta}_{0n})| \leq 2g\left(\frac{s_n t_n}{\sigma}\right) \|\boldsymbol{\beta}_{0n}\| \sum_{i=1}^{n} \|\mathbf{x}_i\|^2 + \frac{1}{\sigma} |g'|_\infty \|\boldsymbol{\beta}_{0n}\|^2 \sum_{i=1}^{n} \|\mathbf{x}_i\|^3 .$$

Como $\|\boldsymbol{\beta}_{0n}\| = o_p\left(n^{-1/4}\right)$ por A6, para ver que $M_n(\nu_n, \boldsymbol{\beta}_{0n}) = o_p\left(n^{1/2}\right)$ basta probar que $g(s_n t_n/\sigma) = O_p\left(n^{-1/4}\right)$. Como $t_n \to \infty$ en probabilidad, se tiene que $t_n = \bar{t}_n$ con probabilidad tendiendo a 1 y entonces basta ver que $g(s_n \bar{t}_n/\sigma) = O_p\left(n^{-1/4}\right)$. Recordando que

$$1 - F_n^+(\bar{t}_n) \leq d_n \leq \left|F_n^+ - F_0^+\right|_\infty$$

por definición, finalmente obtenemos que

$$\begin{aligned}
g(s_n\bar{t}_n/\sigma) &\leq C\left(1 - F_0^+(s_n\bar{t}_n/\sigma)\right)^{1/2} \\
&\leq C\left(1 - F_0^+(\bar{t}_n)\right)^{1/2} + C\left|F_0^+(\bar{t}_n) - F_0^+(s_n\bar{t}_n/\sigma)\right|^{1/2} \\
&\leq C\left(d_n + F_n^+(\bar{t}_n) - F_0^+(\bar{t}_n)\right)^{1/2} + C\left|F_0^+(\bar{t}_n) - F_0^+(s_n\bar{t}_n/\sigma)\right|^{1/2} \\
&\leq C\left(2\left|F_n^+ - F_0^+\right|_\infty\right)^{1/2} + \left|u^2 f_0'(u)\right|_\infty^{1/2} \max\{1/s_n, 1/\sigma\}\, |s_n - \sigma|
\end{aligned}$$

y siendo $\left|F_n^+ - F_0^+\right|_\infty = O_p\left(n^{-1/2}\right)$ por Lema 4.2 y $|s_n - \sigma| = o_p\left(n^{-1/4}\right)$ por A6, se obtiene $g(s_n\bar{t}_n/\sigma) = O_p\left(n^{-1/4}\right)$ como queríamos. ∎

Demostración del Teorema 4.6. En realidad este resultado está incluído en la demostración del Teorema 4.4, porque ahí se probó que

$$\sum_{i=1}^{n} w_i \mathbf{x}_i u_i = \sum_{i=1}^{n} w(\nu_n u_i)\, u_i \mathbf{x}_i + n h_2(t^*)\, \Sigma_n \boldsymbol{\beta}_{0n} + o_p\left(n^{1/2}\right)$$

sin usar la suposición A7 hasta ese punto (ver (B.20)). Por lo tanto

$$\Sigma_n^{-1/2} \frac{1}{n} \sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{x}_i' \boldsymbol{\beta}_{1n} = h_2(t^*)\, \Sigma_n^{1/2} \boldsymbol{\beta}_{0n} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w(\nu_n u_i)\, u_i \mathbf{z}_{ni} + o_p\left(n^{-1/2}\right).$$

Invocando el Lema B.5 con $\Gamma = 0_{pp}$ se obtiene que el segundo término en el miembro derecho es $O_p\left(n^{-1/2}\right)$ y con el Lema B.3 se completa la demostración. ∎

C. Cálculo del REWLS a partir del LMS, para contaminaciones puntuales

En esta sección mostramos cómo se puede calcular el REWLS a partir del LMS cuando G_0 es la distribución $N_p(0, \mathrm{I})$, $F_0 = \Phi$ y las contaminaciones son puntuales. Debido a la simetría esférica de G_0 basta con considerar contaminaciones de la forma $\mathbf{z} = (x, 0, \ldots, 0, y)'$ con x e y positivos. Para este tipo de contaminaciones, se puede demostrar que la primera componente del LMS no depende de p y que las demás son cero, de modo que para calcular (3.12) basta tomar $p = 1$. Para esta situación Simpson y Yohai (1998, pág. 1159) dan las expresiones del LMS y del correspondiente estimador de escala en función de (x, y). Sean $d_1 = \Phi^{-1}((3/4 - \varepsilon)/(1 - \varepsilon))$, $d_2 = \Phi^{-1}((3/4 - \varepsilon/2)/(1 - \varepsilon))$ y $d = \Phi^{-1}(3/4)$. Sea β_0 la solución de $d_1^2(1 + \beta^2) = (y_0 - \beta x_0)^2$ más cercana a 0, o sea

$$\beta_0 = \frac{x_0 y_0 - d_1 (x_0^2 + y_0^2 - d_1^2)^{1/2}}{x_0^2 - d_1^2}, \tag{C.1}$$

que está bien definido para $|y_0| \geq d_1$. Si llamamos $T_0(\varepsilon, x_0, y_0)$ al LMS para $H = (1 - \varepsilon) H_0 + \varepsilon \delta_{(x_0, y_0)}$, entonces

$$T_0(\varepsilon, x_0, y_0) = \begin{cases} 0 & \text{si } |y_0| < d_1 \\ \beta_0 & \text{si } d_1 \leq |y_0| \leq d_2 \\ \beta_0 & \text{si } |y_0| > d_2 \text{ y } d_1 (1 + \beta_0^2)^{1/2} \leq d_2 \\ 0 & \text{si } |y_0| > d_2 \text{ y } d_1 (1 + \beta_0^2)^{1/2} > d_2. \end{cases} \tag{C.2}$$

El correspondiente estimador de escala es

$$S(\varepsilon, x_0, y_0) = \begin{cases} d_1/d & \text{si } |y_0| < d_1 \\ d_1 (1 + \beta_0^2)^{1/2}/d & \text{si } d_1 \leq |y_0| \leq d_2 \\ d_1 (1 + \beta_0^2)^{1/2}/d & \text{si } |y_0| > d_2 \text{ y } d_1 (1 + \beta_0^2)^{1/2} \leq d_2 \\ d_2/d & \text{si } |y_0| > d_2 \text{ y } d_1 (1 + \beta_0^2)^{1/2} > d_2. \end{cases}$$

A fin de obtener la fórmula explícita para el REWLS, debemos encontrar d_H y $\bar{t}_H$. También debemos especificar un valor de η. Nosotros usaremos $\eta = 2.5$, aunque para que valgan los resultados siguientes sirve cualquier $\eta > d$.

Sea

$$r_0 = \frac{y_0 - x_0 T_0(\varepsilon, x_0, y_0)}{S(\varepsilon, x_0, y_0)}$$

el residuo estandarizado de (x_0, y_0). Entonces el estimador REWLS, calculado con pesos hard-rejection, es

$$T_1(\varepsilon, x_0, y_0) = \frac{(1 - \varepsilon) E_0\{wxy\} + \varepsilon\, x_0 y_0 \mathbb{I}(|r_0| < t_H)}{(1 - \varepsilon) E_0\{wx^2\} + \varepsilon\, x_0^2 \mathbb{I}(|r_0| < t_H)}. \tag{C.3}$$

El cálculo explícito de (C.3) es tedioso pero directo. Tenemos que considerar por separado los mismos casos que en (C.2). Es importante señalar que las fórmulas explícitas de (C.3) que obtendremos serán válidas también cuando $t_H = 2.5$, es decir, para el estimador a un paso de Rousseeuw.

Caso 1: $|y_0| < d_1$: En este caso $T_0 = 0$ y $S = d_1/d$. Por lo tanto

$$F_H^+(t) = (1-\varepsilon) F_0^+(St) + \varepsilon \{|y_0|/S \leq t\}.$$

Debemos hallar

$$d_H = \sup_{t\geq 0} \max \left\{ F_0^+(t) - F_H^+(t), 0 \right\}.$$

Si definimos

$$F_1(t) = F_0^+(t) - (1-\varepsilon) F_0^+\left(\frac{d_1}{d} t\right) \tag{C.4}$$

entonces

$$d_H = \max \left\{ \sup_{t<|y_0|/S} F_1(t), \sup_{t\geq |y_0|/S} (F_1(t) - \varepsilon), 0 \right\}.$$

Sea $t_1 = \arg\max_{t\geq 0} F_1(t)$. Es fácil ver que

$$t_1 = \left[\frac{-2\ln\{(1-\varepsilon) d_1/d\}}{\{1-(d_1/d)^2\}} \right]^{1/2}$$

Además se puede ver que $F_1'(t)$ se anula únicamente en t_1, por lo tanto F_1 es estrictamente creciente en $(0, t_1)$ y estrictamente decreciente en (t_1, ∞). Como además $|y_0|/S < d$ en el caso que estamos considerando y $d < t_1$, tenemos que

$$d_H = \max \{F_1(|y_0|/S), F_1(t_1) - \varepsilon, 0\}.$$

Como $F_1(d) = \varepsilon$ es $F_1(t_1) > \varepsilon$ y entonces

$$d_H = \max \{F_1(|y_0|/S), F_1(t_1) - \varepsilon\}.$$

Así que debemos considerar dos casos para encontrar $\bar{t}_H$. Recordemos que $\bar{t}_H$ es el menor valor de $t \geq 0$ que satisface

$$F_H^+(t) \geq 1 - d_H.$$

Si $d_H = F_1(|y_0|/S)$, ningún $t < |y_0|/S$ cumple

$$F_H^+(t) = (1-\varepsilon) F_0^+(St) \geq 1 - F_1(|y_0|/S)$$

porque el lado derecho es siempre mayor que $1-\varepsilon$. Así que debemos considerar los $t \geq |y_0|/S$. De la igualdad $F_H^+(\bar{t}_H) = 1 - d_H$ despejamos

$$\bar{t}_H = \frac{1}{S}\left(F_0^+\right)^{-1}\left(\frac{1 - F_1\left(|y_0|/S\right) - \varepsilon}{1-\varepsilon}\right)$$

que se puede ver que es efectivamente cumple $\bar{t}_H \geq |y_0|/S$.

Ahora, si $d_H = F_1(t_1) - \varepsilon$, nuevamente ningún $t < |y_0|/S$ satisface

$$F_H^+(t) \geq 1 - F_1(t_1) + \varepsilon$$

porque, al menos para los valores de ε considerados en la Tabla 1, verificamos numéricamente que $F_1(t_1) < 2\varepsilon$. Entonces debemos considerar los $t \geq |y_0|/S$ y como antes se obtiene

$$\bar{t}_H = \frac{1}{S}\left(F_0^+\right)^{-1}\left(\frac{1 - F_1(t_1)}{1-\varepsilon}\right)$$

que satisface $\bar{t}_H \geq |y_0|/S$.

Para calcular (C.3), notar que $w = \{|y|/S < t_H\}$ es independiente de x y simétrica alrededor de cero en y, así que

$$\begin{aligned} E_0\{wxy\} &= 0, \\ E_0\{wx^2\} &= F_0^+(St_H). \end{aligned}$$

Como $\bar{t}_H \geq |y_0|/S$ el REWLS es

$$T_1(\varepsilon, x_0, y_0) = \frac{\varepsilon x_0 y_0}{(1-\varepsilon)F_0^+(St_H) + \varepsilon x_0^2}.$$

Cuando $t_H = \bar{t}_H$, esto se simplifica a

$$T_1(\varepsilon, x_0, y_0) = \frac{\varepsilon x_0 y_0}{1 - d_H - \varepsilon + \varepsilon x_0^2}.$$

Case 2: $d_1 \leq |y_0| \leq d_2$, o $|y_0| > d_2$ y $d_1\left(1+\beta_0^2\right)^{1/2} \leq d_2$: Ahora $T_0 = \beta_0$ y $S = d_1\left(1+\beta_0^2\right)^{1/2}/d$. Notar que $(y - \beta_0 x)/S \sim N\left(0, d^2/d_1^2\right)$ y $r_0 = d$ en este caso, así que

$$F_H^+(t) = (1-\varepsilon)F_0^+\left(\frac{d_1}{d}t\right) + \varepsilon\{d \leq t\}.$$

Entonces

$$F_0^+(t) - F_H^+(t) = F_1(t) - \varepsilon\{d \leq t\}$$

con F_1 como en (C.4). Entonces

$$d_H = \max\{\varepsilon; F_1(t_1) - \varepsilon\}$$

porque $t_1 > d$ y $F_1(d) = \varepsilon$. Pero ya dijimos que $F_1(t_1) < 2\varepsilon$, así que

$$d_H = \varepsilon$$

en este caso, sin importar el valor que tomen x_0 ni y_0. Es también fácil de ver que para hallar $\bar{t}_H$ solamente tiene sentido considerar los $t \geq d$ y entonces se obtiene

$$\bar{t}_H = \frac{d}{d_1}\left(F_0^+\right)^{-1}\left(\frac{1-2\varepsilon}{1-\varepsilon}\right) = \frac{d}{d_1}\Phi^{-1}\left(\frac{1-1.5\varepsilon}{1-\varepsilon}\right)$$

que es efectivamente mayor que d. Por lo tanto

$$St_H = \sqrt{1+\beta_0^2}\,\Phi^{-1}\left(\frac{1-1.5\varepsilon}{1-\varepsilon}\right)$$

si $\bar{t}_H > \eta$, o

$$St_H = \sqrt{1+\beta_0^2}\,\frac{d_1}{d}\,\eta$$

si $\bar{t}_H \leq \eta$.

Definamos ahora $a(\varepsilon,\beta_0) = E_0\{wxy\}$ y $b(\varepsilon,\beta_0) = E_0\{wx^2\}$. Entonces

$$\begin{aligned} a(\varepsilon,\beta_0) &= 2\textstyle\int_{-\infty}^{\infty} x\varphi(x)\,\varphi(\beta_0 x - St_H)\,dx, \\ b(\varepsilon,\beta_0) &= \textstyle\int_{-\infty}^{\infty} x^2\varphi(x)\left[2\Phi(\beta_0 x + St_H) - 1\right]dx. \end{aligned} \tag{C.5}$$

Estas integrales se evalúan numéricamente. El REWLS es

$$T_1(\varepsilon, x_0, y_0) = \frac{(1-\varepsilon)\,a(\varepsilon,\beta_0) + \varepsilon x_0 y_0}{(1-\varepsilon)\,b(\varepsilon,\beta_0) + \varepsilon x_0^2}.$$

Case 3: $|y_0| > d_2$ y $d_1\left(1+\beta_0^2\right)^{1/2} > d_2$: En este caso $T_0 = 0$ y $S = d_2/d$. Sea

$$F_2(t) = F_0^+(t) - (1-\varepsilon)\,F_0^+\left(\frac{d_2}{d}t\right),$$

de modo que

$$d_H = \max\left\{\sup_{t<|y_0|/S} F_2(t)\,,\ \sup_{t\geq|y_0|/S}\left(F_2(t)-\varepsilon\right),0\right\}.$$

Si $t_2 = \arg\min_{t\geq 0} F_2(t)$, entonces F_2 es decreciente en $(0,t_2)$ y creciente en (t_2,∞) pues $F_2'(t)$ se anula sólo en t_2. Además $F_2(0) = 0$ y $\lim_{t\to\infty} F_2(t) = \varepsilon$, así que $F_2(t) < 0$ en $(0,t_2)$ y $F_2(t) < \varepsilon$ para todo $t \geq 0$. Por lo tanto,

$$d_H = \max\left\{F_2\left(|y_0|/S\right),0\right\}.$$

Si $d_H = F_2(|y_0|/S)$, entonces no puede ser $\bar{t}_H < |y_0|/S$ porque

$$(1-\varepsilon) F_0^+(St) < 1-\varepsilon < 1-d_H, \ \forall t \geq 0.$$

Así que para hallar $\bar{t}_H$ debemos considerar los $t \geq |y_0|/S$. De la ecuación $F_H^+(\bar{t}_H) = 1 - d_H$ se despeja

$$\bar{t}_H = \frac{d}{d_2}\left(F_0^+\right)^{-1}\left(\frac{1-F_2(|y_0|/S)-\varepsilon}{1-\varepsilon}\right). \tag{C.6}$$

Sin embargo, el lado derecho de (C.6) será menor que $|y_0|/S$ para $|y_0|$ suficientemente grande (específicamente para $|y_0|$ tal que $F_0^+(|y_0|/S) > 1-\varepsilon$) así que

$$\bar{t}_H = \max\left\{\frac{d}{d_2}\left(F_0^+\right)^{-1}\left(\frac{1-F_2(|y_0|/S)-\varepsilon}{1-\varepsilon}\right), \frac{|y_0|}{S}\right\}.$$

Por otro lado, si $d_H = 0$ tenemos que $\bar{t}_H = \infty$. El REWLS es

$$T_1(\varepsilon, x_0, y_0) = \frac{\varepsilon x_0 y_0 \mathbb{I}(|y_0|/S < t_H)}{(1-\varepsilon)F_0^+(St_H) + \varepsilon x_0^2 \mathbb{I}(|y_0|/S < t_H)}.$$

Es importante recalcar que para cada ε fijo t_H tomará el valor $|y_0|/S$ para $|y_0|$ suficientemente grande y entonces $T_1(\varepsilon, x_0, y_0) = 0$. ■

El punto (x_0, y_0) donde $T_1(\varepsilon, x_0, y_0)$ alcanza el máximo debe hallarse mediante una búsqueda por grilla. Para la función de sesgo máximo del LMS (que coincide con la de sesgo máximo puntual) se tiene la expresión

$$\mathcal{B}_{\mathrm{T}_0}(\varepsilon) = \left(\frac{d_2^2}{d_1^2} - 1\right)^{1/2}$$

de acuerdo a la fórmula (3.24) de Martin et al. (1989). Claramente, el LMS alcanza su sesgo máximo en el Caso 2, cuando $d_1\left(1+\beta_0^2\right)^{1/2} = d_2$. Dado cualquier x_0, el sesgo máximo del LMS se alcanzará en $y_0 = d_2 + \mathcal{B}_{\mathrm{T}_0}(\varepsilon)\, x_0$. Numéricamente se verificó que el máximo valor de $T_1(\varepsilon, x_0, y_0)$ se alcanza en ese y_0 también. La forma del REWLS en esos puntos es

$$T_1(\varepsilon, x_0) = \frac{(1-\varepsilon)\, a\left(\varepsilon, \mathcal{B}_{\mathrm{T}_0}(\varepsilon)\right) + \varepsilon d_2 x_0 + \varepsilon \mathcal{B}_{\mathrm{T}_0}(\varepsilon)\, x_0^2}{(1-\varepsilon)\, b\left(\varepsilon, \mathcal{B}_{\mathrm{T}_0}(\varepsilon)\right) + \varepsilon x_0^2}.$$

Para un ε fijo el valor de x_0 que maximiza $T_1(\varepsilon, x_0)$ está dado por

$$x_0^*(\varepsilon) = C(\varepsilon) + \left[C(\varepsilon)^2 + \left(\frac{1-\varepsilon}{\varepsilon}\right) b\left(\varepsilon, \mathcal{B}_{\mathrm{T}_0}(\varepsilon)\right)\right]^{1/2}$$

donde

$$C(\varepsilon)=\left(\frac{1-\varepsilon}{\varepsilon}\right)\left(\frac{\mathcal{B}_{\mathrm{T}_0}(\varepsilon)\,b(\varepsilon,\mathcal{B}_{\mathrm{T}_0}(\varepsilon))-a(\varepsilon,\mathcal{B}_{\mathrm{T}_0}(\varepsilon))}{d_2}\right).$$

Entonces la función de sesgo máximo puntual del REWLS sería

$$\mathcal{B}^*_{\mathrm{T}_1}(\varepsilon)=T_1\left(\varepsilon,x^*_0(\varepsilon),d_2+\mathcal{B}_{\mathrm{T}_0}(\varepsilon)\,x^*_0(\varepsilon)\right).$$

Esta expresión es una conjetura, ya que si bien se verificó que es válida para los valores de ε considerados en la Tabla 1, no pudimos demostrar que efectivamente el máximo REWLS para cada x_0 se alcance en $y_0=d_2+\mathcal{B}_{\mathrm{T}_0}(\varepsilon)\,x_0$.

References

[1] **Davies, L.** (1990). The asymptotics of S-estimators in the linear regression model. *Ann. Statist.* **18** 1651-1675.

[2] **Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A.** (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[3] **He, X. and Portnoy, S.** (1992). Reweighted LS estimators converge at the same rate as the initial estimator. *Ann. Statist.* **20** 2161-2167.

[4] **He, X. and Simpson, D.G.** (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *Ann. Statist.* **21** 314-337.

[5] **Martin, R.D., Yohai, V.J. and Zamar, R.H.** (1989). Min-max bias robust regression. *Ann. Statist.* **17** 1608-1630.

[6] **Pollard, D.** (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

[7] **Pollard, D.** (1990). *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 2.

[8] **Rousseeuw, P.J.** (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871-880.

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

[10] **Rousseeuw, P.J. and Yohai, V.** (1984). Robust regression by means of S-estimates. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256-272. Springer-Verlag, New York.

[11] **Simpson, D.G. and Yohai, V.J.** (1998). Functional stability of one-step GM-estimators in approximately linear regression. *Ann. Statist.* **26** 1147-1169.

[12] **Yohai, V.J.** (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15** 642-656.

[13] **Yohai, V.J. and Zamar, R.H.** (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83** 406-413.

[14] **Yohai, V.J. and Zamar, R.H.** (1993). A minimax-bias property of the least α-quantile estimate. *Ann. Statist.* **21** 1824-1842.