

Tesis de Posgrado

Estimadores robustos para regresión no lineal

Fraiman Maus, J. Ricardo

1980

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

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

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Fraiman Maus, J. Ricardo. (1980). Estimadores robustos para regresión no lineal. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_1635_FraimanMaus.pdf

Cita tipo Chicago:

Fraiman Maus, J. Ricardo. "Estimadores robustos para regresión no lineal". Tesis de Doctor. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 1980.
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_1635_FraimanMaus.pdf

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

ESTIMADORES ROBUSTOS PARA REGRESION NO LINEAL

J.Ricardo Fraiman Maus

Trabajo presentado para optar al
título de Doctor en Ciencias Matemáticas

Director: Profesor Víctor J. Yohai

Victor J. Yohai
J. Ricardo Fraiman Maus

935



Agradezco al Profesor Víctor J. Yohai su dedicada orientación, y constantes sugerencias durante la elaboración de este trabajo. También vaya mi gratitud a dos queridos amigos; Elena Ganón y Oscar Bustos, quienes me ayudaron en la parte de simulación. Finalmente, agradezco al IMPA y al CNPO el uso del computador.

A mis padres

INTRODUCCION

1.- Porqué buscar estimadores robustos.-

A) En Mosteller y Tukey (1977) leemos que:

".... la historia de la estadística y del Análisis de Datos es una confusa mezcla de saludable escepticismo e ingenuo optimismo acerca de la forma exacta que tiene la distribución de las observaciones...."

Buena parte de la estimación puntual ha consistido en, dada una función de pérdida, la búsqueda y obtención de estimadores óptimos, en ciertas subclases, (por ejemplo estimadores insesgados, invariantes respecto de ciertas transformaciones) bajo condiciones de normalidad. Posiblemente, ese optimismo se ha visto alentado por las hermosas propiedades de las familias normales, así como de una interpretación abusiva del teorema central del límite.

Pero, Cómo surgió la consideración de tal familia de distribuciones? De Gauss, quien sugirió su uso, Huber (1972) destaca el siguiente párrafo del "Göttingische gelehrte Anzeigen", cuando se refería al estudio del modelo de estimación de una magnitud física, donde casi toda la variabilidad estadística es debida a los errores de medición (modelo de locación) :

"El autor del presente tratado, quien en el año 1797 investigó primero este problema de acuerdo a los principios de la teoría de probabilidades, pronto se dió cuenta que era imposible determinar el valor mas probable de la cantidad desconocida, a menos que la función representante de los errores sea conocida. Pero ya que esto no es así, no hay otro recurso que suponer tal función de una manera hipotética.- Le pareció natural tomar el camino opuesto y buscar aquella función que debe ser tomada como base, a fin de que, para el caso mas simple de todos, se obtenga un procedimiento o regla que generalmente se acepta como bueno: simplemente que el valor mas probable es la media aritmética de muchas observaciones de un mismo fenómeno repetido en iguales condiciones. Esto implicó que la probabilidad de un error X

debe ser tomado como proporcional a una expresión exponencial de la forma $\exp(-hhXX)$ "

El tomar esta distribución normal, lleva, cuando la función de pérdida es la cuadrática, a las propiedades óptimas de \bar{X} (media aritmética) así como a las de los estimadores de mínimos cuadrados en los modelos de regresión.

Si suponemos entonces, que la distribución normal no es necesariamente la distribución "de la Naturaleza", o al menos que no es exactamente esa, dejan de valer los resultados de optimalidad de los estimadores considerados; pero más aún, como veremos en los ejemplos pueden ser muy malos frente a pequeños apartamientos de la normalidad. De igual modo nos interesará estudiar el comportamiento de los estimadores (óptimos) cuando la distribución de las variables esta cerca de una F dada. (F función de distribución).-

Consideremos los tipos de perturbaciones planteados por Huber (1964) y Hampel (1971) y (1977) :

(i) existe un pequeño porcentaje de observaciones con errores groseros.- Un valor puede ser copiado incorrectamente, o mal leído en el correspondiente instrumento, o se esta midiendo en realidad otra magnitud distinta de la que se quiere medir.

Desviaciones grandes que ocurren con probabilidad pequeña.

(ii) errores de redondeo, debidos a la precisión limitada de toda medición. Desviaciones pequeñas que ocurren con probabilidad grande.

(iii) el modelo, generalmente, es concebido solo como una aproximación a la realidad.

(iv) apartamientos de la independencia.-

Queremos obtener procedimientos de estimación que sean estables en entornos de los modelos pre-experimentales que contemplen estos tipos de perturbaciones.-

La situación estadística será entonces, de que tenemos dadas Y_1, \dots, Y_n observaciones cuya distribución conjunta (en R^n) este cerca de la medida producto $(F_\theta)^{\otimes n}$ para $\theta \in \Theta$.- Queremos estimar θ . Esto nos lleva a considerar los siguientes entornos:

ENTORNOS

El modelo inicial será denotado por $(F_\theta)^{\otimes n}_{\theta \in \Theta}$ la medida producto,

y en los casos a), b) y c) la distribución verdadera será $G^{\otimes n}$ con $G \in \mathcal{F}$ y \mathcal{F} descrita por:

a) Modelo de Contaminación

Sea ε real $\in (0, 1/2)$ (que generalmente se tomará cerca de 0)

$$\mathcal{F}_\varepsilon = \{G: G = (1-\varepsilon)F_\theta + \varepsilon H\} \quad \theta \in \Theta \text{ y } H \in \mathcal{M} \text{ donde } \mathcal{M} \text{ es una familia de probabilidades sobre } (R, B_1) \text{ siendo } B_1 \text{ la } \sigma\text{-álgebra de Borel.}$$

Este modelo fue considerado por Huber (1964), para el caso de locación donde obtiene estimadores minimax (minimizan la máxima varianza asintótica en el entorno).-

Se puede interpretar como que el $(1-\varepsilon)$ % de las observaciones tienen distribución F_θ , y un ε % tiene distribución H .-

Se puede interpretar como que el $(1-\varepsilon)$ % de las observaciones tienen distribución F_θ , y un ε % tiene distribución H .-

En el caso de locación, Y_1, \dots, Y_n que verifican

$$Y_i = \theta + U_i \quad \text{donde } \theta \in \Theta$$

$U = (U_1, \dots, U_n)$ es un vector aleatorio de errores es equivalente a que (si las U_i son independientes)

$$G(t) = G^\circ(t-\theta)$$

$$\text{con } G^\circ \quad \mathcal{F}_{G^\circ}(\varepsilon) = \{G^\circ / G^\circ = (1-\varepsilon)F_\theta + \varepsilon H\}$$

donde $H \in \mathcal{M}$ una cierta familia de medidas de probabilidad sobre (R, B_1) ; en términos de la distribución de los errores.

b) Modelo de variación total

$$0 < \varepsilon < 1$$

Va a consistir en la unión para $\theta \in \mathbb{H}$ de las bolas de radio ε y centro F_θ respecto de la distancia en variación total (Kolmogorov)

Si P, Q son dos probabilidades sobre (Ω, \mathcal{A})

$$K(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$$

$$\mathcal{F}_K(\varepsilon) = \{ G / \exists \theta \in \mathbb{H} : K(F_\theta, G) \leq \varepsilon \}$$

El modelo de variación total puede ser considerado como una extensión del modelo contaminado.

En efecto, si $G = (1-\varepsilon)F + \varepsilon H$ entonces $K(G, F_\theta) \leq \varepsilon$

y reciprocamente, si $K(F, G) = \varepsilon$, existe un real positivo ε^*

$0 < \varepsilon^* < 1$ y dos medidas positivas H_1 y H_2 (que no son en general probabilidades) tales que

$$(1 - \varepsilon^*)F + \varepsilon^* H_1 = (1 - \varepsilon^*)G + \varepsilon^* H_2$$

$$\text{Basta tomar } H_1 = 1/\varepsilon (F - \min(F, G)) \quad H_2 = 1/\varepsilon (G - \min(F, G))$$

$$\varepsilon^* = \frac{\varepsilon}{1+\varepsilon} \quad \text{luego } G = F - \varepsilon H_1 + \varepsilon H_2$$

Para el caso de locación

$$\mathcal{F}_{K^\circ}(\varepsilon) = \{ G^\circ / K(G^\circ, F_0) \leq \varepsilon \} \quad \text{con } G(t) = G^\circ(t - \theta)$$

c) Modelo de Prohorov

Sean F, G dos probabilidades sobre $(\mathbb{R}, \mathcal{B}_1)$

La distancia de Prohorov $\pi(F, G) = \inf \{ \varepsilon / \forall B \in \mathcal{B}_1, F(B) \leq G(B^\varepsilon) + \varepsilon \}$

donde $B^\varepsilon = \{ x \in \mathbb{R} / \text{existe } b \in B \text{ con } |x-b| \leq \varepsilon \}$

De los dos ε que intervienen en la distancia de Prohorov, el ε de la métrica (en B^ε) se puede ver como una difusión local de la pro-

babilidad debido a los errores de redondeo, y el que se refiere a la probabilidad sobre (R, B_1) como el que contempla los errores groseros.

$$\mathcal{F}_\pi(\varepsilon) = \{G: \exists \theta \in \mathbb{H}: \pi(F_\theta, G) \leq \varepsilon\}$$

En el caso de locación

$$\mathcal{F}_{\pi_0}(\varepsilon) = \{G^0: \pi(F_0, G^0) \leq \varepsilon\}$$

El modelo de variación permite contemplar el caso de errores groseros, pero no así los de redondeo como vemos en el siguiente ejemplo:

$$x_n \in R \quad x_n \rightarrow x \quad x_n \neq x$$

Entonces $K(\delta_{x_n}, \delta_x) = 1$ para todo n (δ_x = masa puntual en x)

mientras que

$$\pi(\delta_{x_n}, \delta_x) = \min(|x_n - x|, 1) \rightarrow 0 \text{ cuando } n$$

Este modelo fue considerado por Hampel (1968) y (1971) así como el modelo siguiente:

d) Perturbaciones de la independencia

Para ello consideramos la distancia de Prohorov sobre (R^n, B_n)

Sean Q probabilidades sobre (R^n, B_n)

$$\mathcal{F}_\pi(\varepsilon, n) = \{Q / \exists \theta \in \mathbb{H}: \pi(Q, F_\theta^{\otimes n}) \leq \varepsilon\}$$

que nos permite, además, que Q no sea necesariamente una medida producto $G^{\otimes n}$

B) Ejemplos para locación

1) Sea $\varepsilon > 0$ cualquiera.

$$\Phi_0(x) = \text{función de distribución } N(0,1)$$

$H_1(x)$ = función de distribución de Cauchy.

$$G^\varepsilon = (1-\varepsilon)\Phi_0(x) + \varepsilon H_1(x)$$

$M' = \{H: H \text{ continua y simétrica}\}$

G pertenece a cualquiera de los entornos considerados anteriormente con $\Phi_0 = F_0$.- Bajo G

El estimador $\bar{X} = \sum_{i=1}^n X_i / n$, estimador óptimo bajo Φ_0 , no es consistente, peor aun,

$\lim \sup |\bar{X}| = +\infty$ por la ley de los grandes números (Kolmogorov) pues $E(|X_1|) = +\infty$

2) Los pequeños apartamientos de la normalidad, si bien perjudican sensiblemente el comportamiento de la media, son difíciles de detectar como lo muestra el siguiente ejemplo (Huber, C(1977)) :

Supongamos tener 1000 observaciones X_1, \dots, X_{1000} .- Si estas observaciones se distribuyen según una $N(0,1)$

$P(|X| > 2.5) = 0.012$, es decir que cabría esperarse que haya 6 observaciones con valores mayores que 2.5 y 6 con valores menores que -2.5 .-

Ahora bien si $X \sim F$ con $F(x) = 0.01 \Phi(x/3) + 0.99 \Phi(x)$

(contaminación al 1% con $N(0,9)$)

Si $Y \sim N(0,9)$ $P(|Y| > 2.5) = 0.4$

Es plausible entonces que 4 de los puntos provenientes de la contaminación esten fuera de $[-2.5; 2.5]$

Luego, bastará con que 4 de los 990 puntos que siguen $N(0,1)$ y que sería de esperar esten fuera de $[-2.5; 2.5]$ no sean tan grandes, para que en total contemos 12 puntos fuera de $[-2.5; 2.5]$, lo esperado bajo normalidad.

Sin embargo, si la ley de la experiencia es

$(1-\epsilon)N(0,1) + \epsilon N(0,9)$, como la media es un estimador insesgado en este caso, su eficiencia asintótica será

$$= \lim_{n \rightarrow \infty} \frac{1}{I_n(\epsilon, \theta) V_n(\epsilon)} \quad \text{donde} \quad V_n(\epsilon) = \text{var}(\bar{X}) = \frac{1+8\epsilon}{n}$$

y $I_n(\epsilon, \theta)$ = información de Fisher.

Tenemos entonces que para

Contaminación %	0.00	0.02	0.05	0.10
Eficiencia asintótica de \bar{X}	1.00	0.90	0.80	0.70

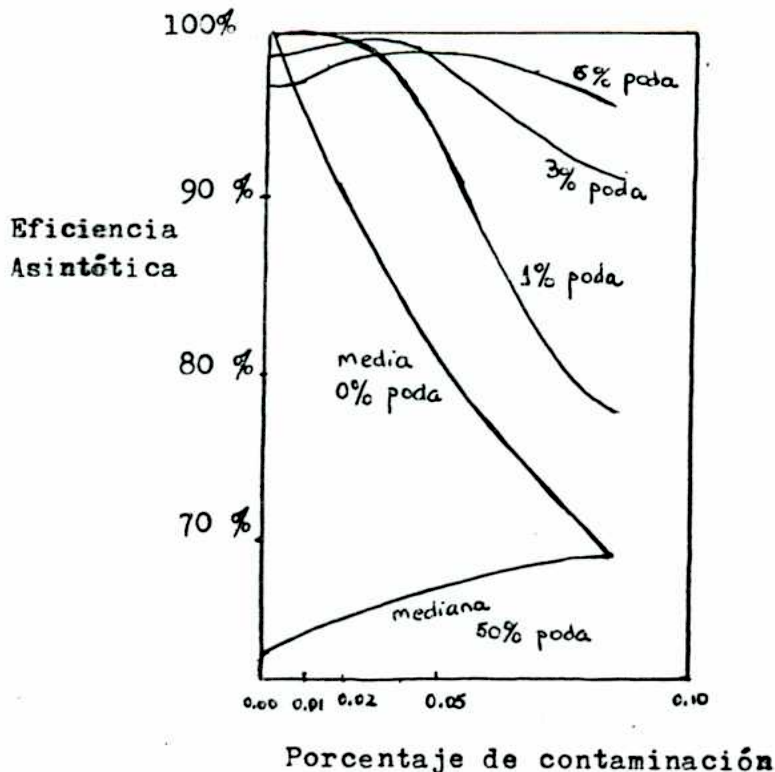
Una modificación muy simple de la media, son las medias α -podadas:

Si $\alpha \in (0, 1/2)$, la media α -podada de X_1, \dots, X_n es

$$\bar{X} = \frac{1}{n-2[\alpha]} (X^{[n\alpha]+1} + X^{[n\alpha]+2} + \dots + X^{n-[n\alpha]})$$

donde $X^{(1)} \leq X^{(2)} \leq \dots \leq X^{(n)}$ son los estadísticos de orden.-

La siguiente gráfica (Tukey) ilustra el comportamiento de la eficiencia asintótica de las medias α -podadas en función del coeficiente de poda α .-



C) Estas consideraciones, nos llevan a buscar, estimadores que sean eficientes bajo los modelos pre-experimentales, pero además estables en un entorno del modelo. O sea, que buscaremos estimadores estables frente a las perturbaciones mencionadas, y que al mismo tiempo sean casi tan eficientes como los óptimos bajo el modelo cuando no las haya.-

Hampel (1971) da una definición de estabilidad:

Definición La sucesión de estimadores $\{T_n\}_{n \geq 1}$ es robusta en F_0 si y solo si para todo $\epsilon > 0$, existe $d > 0$ tal que

si $\pi(F, F_0) < d$ entonces $\pi(L_F(T_n), L_{F_0}(T_n)) < \epsilon$ para todo

n (uniformemente) donde π es la métrica de Prohorov y por $L_F(T_n)$ notamos la distribución de T_n bajo F .

Una sucesión $\{T_n\}_{n \geq 1}$ de estimadores, será una sucesión de funcionales medibles $T_n: \mathcal{F}_n \rightarrow R^k$

donde $\mathcal{F}_n = \{F_n, \dots, P_n, \dots\}$ es el conjunto de medidas de probabilidad discretas cuyos átomos tienen probabilidades igual a $1/n$ o múltiplos de $1/n$ (Funciones de distribución empíricas).-

Hampel describe la robustez en términos de continuidad de T_n con

Teorema 1

Si $\{T_n\}$ verifica

i) T_n es continua como función en Ω^n para todo n

ii) $\{T_n\}$ es continua en F

iii) $\{T_n\}$ es robusta en F

Corolario 1

Sea $T: R^k \rightarrow R^k$ continua en F y $T_n = T|_{\mathcal{F}_n}$

considerada como función en Ω^n continua respecto de la métrica en Ω^n para todo n.

T) $\{T_n\}$ es robusta en F

Además, da algunas medidas de robustez para estimadores, como el punto de ruptura y la curva de influencia = $IC_{T, F_0}(w)$.

Esta segunda, se refiere a la sensibilidad frente a un pequeño porcentaje de contaminación. En la sección IV, hallamos estimadores cuya traza de la matriz de covarianza sea mínima sujeta a la restricción que la sensibilidad $V^* = \sup_{w \in \Omega} |IC_{T, F_0}(w)| \leq a$

Definición

Dado Ω , (espacio métrico completo y separable)

Sea \mathcal{F} = las medidas sobre Ω .

$$IC_{T, F_0}(w) = \lim_{\epsilon \rightarrow 0} \frac{T((1-\epsilon)F_0 + \epsilon\delta_w) - T(F_0)}{\epsilon} \quad (\text{si este límite existe})$$

$$= \lim_{\epsilon \rightarrow 0} \frac{T(F_0 + \epsilon(\delta_w - F_0)) - T(F_0)}{\epsilon} \quad \text{que es la derivada de Gateaux}$$

$$= DT(F_0)(\delta_w - F_0) \text{ .-}$$

Se sabe que cuando T admite derivada de Frechet (sobre las medidas signadas) también existe la derivada de Gateaux y ambas coinciden .-

2.- M-estimadores

El tipo de soluciones consideradas va a ser en la familia de los llamados M-estimadores.

Para el caso de locación, X_1, \dots, X_n variables aleatorias independientes idénticamente distribuidas con distribución $F(x-\theta)$, Huber(1964) propuso la clase de M-estimadores o estimadores de tipo de máxima

verosimilitud, definidos como solución de la ecuación

$$(1.1) \quad \sum_{i=1}^n \psi(x_i - t) = 0$$

donde ψ es una función tal que $E_F(\psi) = 0$

En particular, si F es simétrica y ψ impar y F -integrable se verifica.-

Si $\psi(t) = t$ obtenemos los estimadores de máxima verosimilitud bajo normalidad.-

Huber estudió el comportamiento asintótico y probó que para ψ elegida convenientemente, el estimador resultante tiene propiedades robustas, así como si ψ es monótona no decreciente, bajo hipótesis débiles que es consistente p.p. y asintóticamente normal con media θ , y varianza $V(\psi, F)/n$ donde $V(\psi, F) = E_F(\psi^2)/E_F^2(\psi')$

En particular, si estamos en un entorno de contaminación de la normal Huber(1964) y Hampel (1968) probaron que los estimadores, para

$$\psi_k(t) = \min(|t|, k) \operatorname{sign}(t)$$

son las soluciones del problema minimax en el entorno considerado. Hampel(1968), Andrews (1974), Beaton y Tukey (1974) y Collins(1976) considerando otras propiedades robustas como la curva de influencia, propusieron el uso de funciones ψ no monótonas que se anulen fuera de un compacto, como ser

$$\text{Hampel: } \psi_{a,b,c}(u) = \begin{cases} u & \text{si } |u| \leq a \\ a \operatorname{sg}(u) & \text{si } a \leq |u| \leq b \\ a(c-|u|)\operatorname{sg}(u)/(c-b) & \text{si } b \leq |u| \leq c \\ 0 & \text{si } |u| > c \end{cases}$$

$$\text{Andrews: } \psi_k(u) = \begin{cases} \operatorname{sen}(u/k) & \text{si } |u| \leq k \\ 0 & \text{si } |u| > k \end{cases}$$

$$\text{Tukey: } \psi_k(u) = \begin{cases} u(1-(\frac{u}{k})^2)^2 & \text{si } |u| \leq k \\ 0 & \text{si } |u| > k \end{cases}$$

Estudios numéricos han mostrado que estos estimadores tienen buenas propiedades de eficiencia y robustez, aun para tamaños de muestra pequeña.-

En estos casos la ecuación $E_F(\Psi(x-t)) = 0$ puede tener soluciones diferentes de 0.- Luego, algunas soluciones de la ecuación (1.1) pueden converger a $\theta + t_0$ donde t_0 es una raíz de $E_F(\Psi(x-t)) = 0$ distinta de 0.

Luego, para definir el estimador se debe especificar que solución de la ecuación (1.1) se esta considerando.- Yohai y Klein (1979) estudian este problema, indicando el algoritmo numérico usado para calcularlo y prueban consistencia y normalidad asintótica para M-estimadores iterativos para locación y para el modelo lineal.-

•

3.- Modelos considerados y resultados obtenidos.-

Se considera el modelo de regresión múltiple, no lineal

$$Y = g(\theta, X) + U$$

donde U es el error, Y es la variable dependiente, $Y \in \mathbb{R}$, $X \in \mathbb{R}^q$

$g: \mathbb{R}^k \times \mathbb{R}^q \rightarrow \mathbb{R}$ es una función regular

$\theta \in \mathbb{R}^k$ es el vector de parámetros de la regresión.-

SECCIONES I y II

En las secciones I y II se estudia el modelo cuando se tienen efectos fijos, o sea tenemos observaciones Y_1, \dots, Y_n

$$Y_i = g(\theta_0, X_i) + U_i$$

$X_i \in \mathbb{R}^q$ vector numérico de efectos fijos

U_i son variables aleatorias independientes con distribución común

F desconocida

para M-estimadores iterativos.

Los estimadores usuales para este modelo son los de mínimos cuadrados

solución del sistema

$$\sum_{i=1}^n (Y_i - g(X_i, \hat{\theta}_{mc})) \frac{\partial g(X_i, \hat{\theta}_{mc})}{\partial \theta} = 0$$

que cuando $F = \Phi$ (normal) coinciden con los estimadores de máxima verosimilitud, y tienen propiedades óptimas. Estos estimadores fueron estudiados por Malenvaud (1969) y Jenrich (1970), que obtienen resultados de consistencia y normalidad asintótica.

Estos estimadores, adolecen de los mismos problemas que \bar{X} para el caso de locación - no son robustos.- Si consideramos que la distribución de los errores está en un entorno de Φ , pero no es exactamente Φ , su comportamiento es muy malo; son muy sensibles a distribuciones de los errores con colas mas pesadas. Una observación anómala puede hacer que el ajuste por mínimos cuadrados varíe mucho.

SECCION I

En la sección I cuando $\sigma^2 = \text{var}_F(U)$ es conocida (escala conocida)

la clase de M-estimadores, o estimadores de tipo de máxima verosimilitud considerada por Huber (1964), se extiende al caso de nuestro modelo como fue hecho para el caso del modelo lineal por Relles(1968), Huber (1972), Yohai (1974), Marona y Yohai(1979) y Klein y Yohai (1979) definiendo $\hat{\theta}_n$ como solución de la ecuación:

$$\sum_{j=1}^n \rho(Y_j - g(\hat{\theta}_n, X_j)) = \text{mínimo} \quad \text{donde } \rho: \mathbb{R} \rightarrow \mathbb{R}$$

o si consideramos $\rho = \psi$, como solución del sistema

$$(I) \quad \sum_{j=1}^n \psi(Y_j - g(\hat{\theta}_n, X_j)) \frac{\partial g(\hat{\theta}_n, X_j)}{\partial \theta} = 0$$

En particular, si $\rho = t^2$ obtenemos los estimadores de mínimos cuadrados usuales. Eligiendo convenientemente ψ , obtendremos estimadores robustos, en un entorno de una F dada, y cuya eficiencia en el caso normal (por ejemplo si $F = \Phi$) sea casi tan buena como la del óptimo bajo el modelo.-

Como la ecuación (I) puede tener mas de una solución, y alguna de ellas pueden no converger a θ_0 , para definir los estimadores, y estudiar sus propiedades y su comportamiento asintótico debemos indicar que solución de la ecuación estamos considerando.

Siguiendo el método utilizado por Yohai-Klein (1979) esto se hace indicando el algoritmo numérico que usamos para computarlo. En los casos estudiados, se consideró Newton-Ralphson a partir de un estimador inicial consistente.

Se consideran entonces funciones Ψ no necesariamente monótonas, obteniéndose M-estimadores iterativos.-

En esta sección se prueba que si

$$U^t U = \sum_{j=1}^n \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial g(\theta_0, X_j)}{\partial \theta}^t, \text{ y}$$

$$\hat{\theta}_{n,0} = \begin{cases} \lim_{j \rightarrow \infty} \hat{\theta}_{n,j} & \text{si este límite existe} \\ \hat{\theta}_{n,0} & \text{si no} \end{cases}$$

donde, $\hat{\theta}_{n,j}$ = paso j-ésimo del algoritmo

$\hat{\theta}_{n,0}$ es el estimador inicial

Proposición 1

(i) $\lim_{n \rightarrow \infty} P(\hat{\theta}_n = \lim_{j \rightarrow \infty} \hat{\theta}_{n,j} \text{ y } \hat{\theta}_n \text{ satisface la ecuación I}) = 1$

(ii) $U(\hat{\theta}_n - \theta_0)$ está acotado en probabilidad

Proposición 2

Para todo $a \in \mathbb{R}^k$ con $\|a\| = 1$

$$a^t U(\hat{\theta}_n - \theta_0) \xrightarrow{w} N(0, E_F(\Psi^2) / E_F^2(\Psi))$$

SECCION II

Se considera ahora el problema cuando no se conoce la escala. (σ^2 desconocido).- Como los M-estimadores no son escala equivariantes, luego, no son razonables cuando la escala es desconocida.

Entonces, para obtener estimadores que lo sean, modificamos la ecuación (I) y consideramos

$$(II) \quad \sum_{i=1}^n \psi \left[\frac{Y_i - g(X_i, \hat{\theta}_n)}{s_n} \right] \frac{\partial g(X_i, \hat{\theta}_n)}{\partial \theta} = 0$$

donde s_n es un estimador de escala equivariante y robusto.

$$s_n(ay) = |a| s_n(y)$$

Por ejemplo s_n puede ser el rango intercuartil normalizado de los \bullet residuos

$$\hat{\sigma}_1 = (\hat{U}^{(n-[n/4]+1)} - \hat{U}^{([n/4])}) / 2 \Phi^{-1}(3/4)$$

donde $\hat{U}_t = Y_t - g(X_t, \hat{\theta}_0)$ donde $\hat{\theta}_0$ es un estimador inicial invariante.

o bien s_n puede ser la mediana muestral de los valores absolutos de los residuos normalizados

$$\hat{\sigma}_2 = \text{mediana}(|\hat{U}_j|) / \Phi^{-1}(3/4)$$

donde $U^{(1)} \leq \dots \leq U^{(n)}$ son los estadísticos de orden.-

En esta sección se consideran soluciones iterativas a la ecuación II cuando s_n es un estimador consistente de escala $s_n \rightarrow s_0$

Se define $\hat{\theta}_n = \begin{cases} \lim \hat{\theta}_{n,j} & \text{si existe} \\ \hat{\theta}_{n,0} & \text{si no} \end{cases}$

y usando resultados de tightness en C- espacio de funciones continuas en $[0,1]$ se llevan los resultados de la sección I al caso de escala desconocida.-

Se obtienen así, las proposiciones 3 y 4, análogas a las proposiciones 1 y 2 con $\psi_{s_0} = \psi(\cdot/s_0)$ en lugar de ψ , y s_0 el valor al cual converge s_n .

SECCIONES III y IV

En las secciones III y IV se considera el modelo cuando las X_s son aleatorias, o sea que

$$Y_i = g(X_i, \beta_0) + U_i$$

X_i vectores aleatorios en R^q

U_i variables aleatorias independientes igualmente distribuidas

$Z_i = (Y_i, X_i) \in R^{q+1}$ vectores aleatorios independientes con distribución común F .

$\beta_0 \in R^k$ coeficientes de la regresión.

Para el caso del modelo lineal, Hampel (1979), Krasker (1979) y Krasker y Welch (1979) ya observaron que cuando las X son aleatorias los M -estimadores usuales, no son robustos (curva de influencia no acotada), cuando tenemos perturbaciones en las X_s . - Por ejemplo, si el modelo deja de cumplirse en ciertos períodos anormales, o bien aproximaciones o errores groseros.

SECCION III

En nuestro caso, si $\frac{\partial g(X, \beta)}{\partial \beta}$ no es acotada, los M -estimadores usuales (considerados en las secciones anteriores) no son robustos.

Como en el caso lineal, se considera la familia de GM -estimadores, o estimadores de tipo máxima verosimilitud generalizados, definidos como solución de la ecuación:

$$(III) \quad \sum_{j=1}^n \psi(Y_j, X_j, \hat{\beta}_n) = 0 \quad \text{para } \psi: R \times R^q \times R^k \rightarrow R^k$$

El estimador de mínimos cuadrados $\hat{\beta}_{mc}$ está en la clase con

$$\psi(Y, X, \beta) = (Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta}$$

así como los M-estimadores para

$$\psi(Y, X, \beta) = \Psi(Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta}$$

Si consideramos Ψ acotada, tendremos estimadores robustos.

En realidad, al considerar soluciones de la ecuación III, estamos considerando estimadores para modelos mas generales, que en particular aplicamos al caso de regresión.-

Como en las secciones anteriores, en el caso que la ecuación

$E_{\mathbb{F}}(\psi(Y, X, \beta)) = 0$ no tenga solución única, para definir el estimador $\hat{\beta}_n$ debemos especificar que solución de III estamos considerando.

Esto lo hacemos indicando el algoritmo usado para calcularlo.

Para el caso de Newton-Ralphson

si $\hat{\beta}_{n,j}$ es el paso j-ésimo del algoritmo, en la sección III se prueba que

Proposición 5

(i) $\hat{\beta}_n \xrightarrow{p.p.} \beta_0$

(ii) $\hat{\beta}_n = \lim_{j \rightarrow \infty} \hat{\beta}_{n,j}$ si $n > n_0$ p.p.

Proposición 6

$$n^{1/2} (\hat{\beta}_n - \beta_0) \xrightarrow{w} N(0, \Lambda^{-1} V \Lambda^{-1})$$

con $V = E_{\mathbb{F}}(\psi(Y, X, \beta_0) \psi(Y, X, \beta_0)^t)$

y $\Lambda = E_{\mathbb{F}}(\frac{\partial \psi(Y, X, \beta_0)}{\partial \beta})$ (no singular por hipótesis)

SECCION IV

En la sección IV , se encuentra un resultado análogo al obtenido por Krasker (1979) para el modelo lineal.-

Se encuentra un estimador óptimo en la clase de los GM-estimadores sujeto a una cota en la sensibilidad.

Más precisamente, se encuentra la función ψ de modo que el estimador $\hat{\beta}$ correspondiente a dicha ψ , minimiza la traza de la matriz de covarianza $E(\|\hat{\beta} - \beta_0\|^2) = \text{mínima}$, bajo normalidad, entre todos aquellos tales que su sensibilidad $\psi^+ = \sup_{(Y,X)} |IC(Y,X)| \leq a$

Además se prueba que si existe un estimador óptimo en la subclase de los estimadores de mínimos cuadrados pesados definidos para

$$\psi(Y,X,\beta) = (Y-g(x,\beta))W(Y,X,\beta) \frac{\partial g(x,\psi)}{\partial \beta}$$

donde W es la función de pesos,

en el sentido de definida positiva (para la matriz de covarianza) entre aquellos cuya sensibilidad $\leq a$, este es el correspondiente a la ψ obtenida.-

SECCION V

Finalmente, en la sección V, se hace una comparación por Montecarlo entre el estimador de mínimos cuadrados y dos M-estimadores, los correspondientes a las funciones de Huber y Tukey respectivamente con constantes $k = 1.34$ y $k = 4.68$ respectivamente. Las constantes se eligieron de modo que la eficiencia asintótica bajo normalidad sea del 95 % .-

Se consideraron para los errores las siguientes distribuciones:
distribución normal (0,1)

distribución normal contaminada con normal al 5% y desviación 5

distribución Student con 3 grados de libertad.

Se hicieron 2000 replicaciones para tamaños de muestra 20 y 50.

Se consideraron las funciones de regresión logística y se aplicó

el algoritmo de Newton con hasta 10 iteraciones, a partir del valor verdadero y a partir de estimadores iniciales.

Los resultados obtenidos en las tablas muestran el mejor comportamiento de los estimadores robustos propuestos cuando hay apartamientos de la normalidad, así como cuando el modelo se cumple exactamente su comportamiento es casi equivalente al de mínimos cuadrados; cuando el tamaño de la muestra es pequeño (20 y 50 observaciones).-

Julio de 1980

SECCION I

M-estimadores iterativos para regresión no lineal con efectos fijos y escala conocida.-

Consideremos el modelo de regresión general, con efectos fijos, en que tenemos observaciones Y_1, Y_2, \dots, Y_n que verifican:

$$Y_i = g(\theta_0, X_i) + U_i \quad i = 1, \dots, n \quad \text{donde}$$

$X_i \in R^q$ es un vector numérico

$\theta_0 \in R^k$ es el vector de parámetros que queremos estimar

U_i son variables aleatorias independientes con distribución común F desconocida.

Como ya observamos en la introducción, cuando $F = \Phi$ (normal) los estimadores de mínimos cuadrados, coinciden con los de máxima verosimilitud y tienen propiedades óptimas; pero dicho comportamiento empeora sensiblemente cuando $F \neq \Phi$, aun en los entornos considerados anteriormente.

Ello nos lleva a considerar aquí la familia de M-estimadores definidos como solución de la ecuación:

$$(I)^* \quad \sum_{j=1}^n f(Y_j - g(\hat{\theta}, X_j)) = \text{mínimo} \quad \text{con } f: R \rightarrow R$$

o, si notamos por $\Psi = f'$, la solución del sistema

$$(I) \quad \sum_{j=1}^n \Psi(Y_j - g(\hat{\theta}, X_j)) \frac{\partial g(\hat{\theta}, X_j)}{\partial \theta} = 0$$

En particular, si $f(t) = t^2$ obtenemos los estimadores de mínimos cuadrados. Si consideramos funciones Ψ acotadas, obtendremos estimadores robustos.

Para resolver la ecuación (I), consideraremos M-estimadores iterativos. aplicando el algoritmo de Newton-Raphson a dicha ecuación, a partir de un estimador inicial consistente.

Obtendremos así un estimador que convenientemente normalizado estará acotado en probabilidad (consistencia) y su distribución asintótica será normal.

Notaciones:

$\frac{\partial g(\theta, X_j)}{\partial \theta}$ vector columna

$$P(\theta) = \sum_{j=1}^n \Psi(Y_j - g(\theta, X_j)) \frac{\partial g(\theta, X_j)}{\partial \theta}$$

$$P'(\theta) = \left(\left(\frac{\partial P_j}{\partial \theta_1} \right) \right) \in R^{k \times k}$$

$$B(\theta, X_j) = \frac{\partial g(\theta, X_j)}{\partial \theta} \left(\frac{\partial g(\theta, X_j)}{\partial \theta} \right)^t$$

$$H(\theta, X_j) = \left(\left(\frac{\partial}{\partial \theta_i \partial \theta_1} g(\theta, X_j) \right) \right) \in R^{k \times k}$$

$$h_1(v, z) = \frac{\partial g(\theta_0 + v, X_1)}{\partial \theta} \left(\frac{\partial g(\theta_0 + z, X_1)}{\partial \theta} \right)^t$$

$$P'(\theta) = \sum_{j=1}^n \Psi(Y_j - g(\theta, X_j)) H(\theta, X_j) - \sum_{j=1}^n \Psi'(Y_j - g(\theta, X_j)) \frac{\partial g(\theta, X_j)}{\partial \theta} \frac{\partial^t g(\theta, X_j)}{\partial \theta}$$

Sea U la matriz tal que

$$U^t U = \sum_{j=1}^n \left(\left(\frac{\partial}{\partial \theta_i} g(\theta_0, X_j) \frac{\partial}{\partial \theta_h} g(\theta_0, X_j) \right) \right) = \sum_{j=1}^n B(\theta_0, X_j)$$

que estará definida por hipótesis.

Vamos entonces a normalizar definiendo $\theta^* = \theta_0 + U(\theta - \theta_0)$

por tanto, $\theta = \theta_0 + U^{-1}(\theta^* - \theta_0)$

Queríamos minimizar

$$\sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(\theta^* - \theta_0), X_j))$$

Luego, debe ser

$$P(\theta^*) = \sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(\theta^* - \theta_0), X_j)) U^{-1t} \frac{\partial g(\theta_0 + U^{-1}(\theta^* - \theta_0), X_j)}{\partial \theta} = 0$$

(II)

$$P'(\hat{\theta}^*) = \sum_{j=1}^n \Psi(Y_j - g(\theta_{\bullet} + U^{-1}(\hat{\theta}^* - \theta_{\bullet}), X_j)) U^{-1t} H(\theta_{\bullet} + U^{-1}(\hat{\theta}^* - \theta_{\bullet}), X_j) U^{-1} \\ - \sum_{j=1}^n \Psi'(Y_j - g(\theta_{\bullet} + U^{-1}(\hat{\theta}^* - \theta_{\bullet}), X_j)) U^{-1t} B(\theta_{\bullet} + U^{-1}(\hat{\theta}^* - \theta_{\bullet}), X_j) U^{-1}$$

Luego, si aplicamos el algoritmo de Newton -Ralphson

$$t_{j+1} = t_j - (P'(t_j))^{-1} P(t_j)$$

a nuestro caso tendremos:

$$\hat{\theta}_{n,j+1}^* = \hat{\theta}_{n,j}^* - \left[\sum_{j=1}^n \Psi(Y_j - g(\theta_{\bullet} + U^{-1}(\hat{\theta}_{n,j}^* - \theta_{\bullet}), X_j)) U^{-1t} H(\theta_{\bullet} + U^{-1}(\hat{\theta}_{n,j}^* - \theta_{\bullet}), X_j) U^{-1} \right. \\ \left. - \sum_{j=1}^n \Psi'(Y_j - g(\theta_{\bullet} + U^{-1}(\hat{\theta}_{n,j}^* - \theta_{\bullet}), X_j)) U^{-1t} B(\theta_{\bullet} + U^{-1}(\hat{\theta}_{n,j}^* - \theta_{\bullet}), X_j) U^{-1} \right]^{-1} \\ \cdot \left(\sum_{j=1}^n \Psi'(Y_j - g(\theta_{\bullet} + U^{-1}(\hat{\theta}_{n,j}^* - \theta_{\bullet}), X_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_{\bullet} + U^{-1}(\hat{\theta}_{n,j}^* - \theta_{\bullet}), X_j) \right)$$

Definimos entonces

$$\hat{\theta}_n^* = \begin{cases} \lim_{j \rightarrow \infty} \hat{\theta}_{n,j}^* & \text{si existe} \\ \hat{\theta}_{n,o}^* & \text{si no} \end{cases} \quad \hat{\theta}_n = \begin{cases} \lim_{j \rightarrow \infty} \hat{\theta}_{n,j} & \text{si existe} \\ \hat{\theta}_{n,o} & \text{si no} \end{cases}$$

donde $\hat{\theta}_{n,o}^*$ es el estimador inicial normalizado ($\hat{\theta}_{n,o}^* = \theta_{\bullet} + U(\hat{\theta}_{n,o} - \theta_{\bullet})$)

y $\hat{\theta}_{n,o}$ es el estimador inicial consistente dado, para los algoritmos normalizados y sin normalizar respectivamente.

Definimos además

$$\eta_n(t) = t - (P'(t))^{-1} P(t) \quad \text{para } P \text{ dada en (II)}$$

y para toda $\Gamma \in \mathbb{R}^{k \times k}$ definida positiva

$$\eta_{n\Gamma}(t) = t + \Gamma^{-1} P(t)$$

Llamaremos también $\Gamma_n(t) = P'(t)$

Luego

$$\eta_{n\Gamma}(t) = t + \Gamma^{-1} \left(\sum_{j=1}^n \Psi(Y_j - g(\theta_{\bullet} + U^{-1}(t - \theta_{\bullet}), X_j)) U^{-1} \frac{\partial}{\partial \theta} g(\theta_{\bullet} + U^{-1}(t - \theta_{\bullet}), X_j) \right)$$

y tendremos que

$$\eta_n(t) = \eta_n \Gamma_n(t) \quad (t)$$

Llamaremos además $\Gamma_o = E_F(\Psi') I_k$ donde I_k es la matriz identidad en $R^{k \times k}$.-

HIPOTESIS

A1) Ψ tiene derivada continua salvo en un número finito de puntos y $g \in C^2$

A2) $\int \Psi(x) dF(x) = 0$, y F es continua.

A3) existe $d_o > 0$ tal que

$$\sup_{|a| \leq d_o} |\Psi^2(x-a)| \text{ es } F\text{-integrable.}$$

$$\sup_{|a| \leq d_o} |\Psi'^2(x-a)| \text{ es } F\text{-integrable.}$$

A4) $\hat{\theta}_{n,o}$ es un estimador inicial que verifica que $U(\hat{\theta}_{n,o} - \theta_o)$ está acotado en probabilidad.

A5) Existe $n_o > 0$ tal que si $n \geq n_o$ U es no singular.

A6) $\max_{1 \leq j \leq n} \left\| \frac{\partial}{\partial \theta} g(\theta_o, X_j) \right\|^2 / \sum_{j=1}^n \left\| \frac{\partial}{\partial \theta} g(\theta_o, X_j) \right\|^2 \rightarrow 0$ (condición de tipo Lindenberg)

y las siguientes hipótesis sobre el diseño:

(i) existe $a > 0$ tal que $\frac{\max \text{ autovalor de } U^t U}{\min \text{ autovalor de } U^t U} \leq a$ si $n \geq n_o$

(ii) $\|h(\theta_o, X_j)\| \leq K \left\| \frac{\partial}{\partial \theta} g(\theta_o, X_j) \right\|^2$ para algún K

(iii) $h_j(v, z)$ es equicontinua en $(0,0)$

(iv) existe $D > 0$ tal que $0 < D < \left\| \sum_{j=1}^n h(0,0, X_j) \right\| / n$
 $h_j(0,0) = h(\cdot, 0, X_j)$

(v) $\max \text{ autovalor de } U^t U \rightarrow \infty$ cuando $n \rightarrow \infty$

A7) gradiente $g \neq 0$; y si $V = \{t \in R^k / g(X_1, t) = a\}$ $a \in R$, y L una recta cualquiera, entonces $V \cap L$ esta formado por intervalos o puntos aislados.-

Lema 1

Sean U_1, U_2, \dots una sucesión de variables aleatorias independientes igualmente distribuidas, y para todo r en un entorno de 0, $f_r(U)$ es una función boreliana definida en R . Si además se verifican:

(i) $\lim_{r \rightarrow 0} f_r(U) = f_0(U)$ p.p.

(ii) $E(f_0(U)) = 0$

(iii) Existe $d_0 > 0$ tal que $\sup_{|r| \leq d_0} |f_r(U)| \leq f(U)$ con $E(f^2(U)) < \infty$

(iv) $V_{n,j}$ son números reales $1 \leq j \leq n$, $1 \leq n < \infty$ tales que existe M para el cual $\sum_{j=1}^n |V_{n,j}| \leq M$ para todo n

(v) $\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} |V_{n,j}| = 0$

entonces

T) $\lim_{d \rightarrow 0} \limsup_{n \rightarrow \infty} P\left(\sup_{|r_j| \leq d} \left| \sum_{j=1}^n f_{r_j}(U_j) V_{n,j} \right| > \varepsilon\right) = 0$

La demostración puede verse en Yohai (1974)

Lema 2

Sea (S, d) un espacio métrico. Para cada $\gamma \in \Gamma$

$\eta_\gamma : S \rightarrow S$ es una contracción uniforme en γ , o sea que existe k_0 $0 \leq k_0 < 1$ tal que $d(\eta_\gamma(x), \eta_\gamma(x')) \leq k_0 d(x, x')$ para todo $\gamma \in \Gamma$

Supongamos que además todas tengan el mismo punto fijo x^* .-

Entonces

T) Dado $x_0 \in S$, y una sucesión cualquiera $\gamma_1, \gamma_2, \dots$ en Γ

Si definimos inductivamente

$$x_1 = \eta_{\gamma_1}(x_0) \quad x_{n+1} = \eta_{\gamma_{n+1}}(x_n)$$

se verifica que $x_n \rightarrow x^*$

Demostración

Vamos a probar por inducción que $d(x_n, x^*) \leq k_0^n d(x_0, x^*)$

Para $n=0$ se verifica.

$$d(x_{n+1}, x^*) = d(\gamma_{n+1}(x_n), \gamma_{n+1}(x^*)) \leq k_0 d(x_n, x^*) \leq k_0^{n+1} d(x_0, x^*)$$

esta última desigualdad por la hipótesis inductiva.-

Lema 3

Sea G = subconjunto de matrices definidas positivas cuyo mínimo autovalor es $> d > 0$.-

entonces

T) $\sup_{\Gamma \in G} \|\eta_{n\Gamma}(\theta_0) - \theta_0\|$ está acotado en probabilidad.

Demostración

$$\begin{aligned} \sup_{\Gamma \in G} \|\eta_{n\Gamma}(\theta_0) - \theta_0\| &\leq \max \text{ aut. de } \Gamma^{-1} \left\| \sum_{j=1}^n \psi(U_j) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\| \\ &\leq \frac{1}{d} \left\| \sum_{j=1}^n \psi(U_j) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\| \end{aligned}$$

Luego, por Tchebichev basta ver que

$$E \left(\left\| \sum_{j=1}^n \psi(U_j) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2 \right) \text{ está acotada.}$$

$$\begin{aligned} E \left(\left\| \sum_{j=1}^n \psi(U_j) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2 \right) &= \\ &= E \left(\sum_{j,k} \psi(U_j) \psi(U_k) \left\langle U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta}, U^{-1t} \frac{\partial g(\theta_0, X_k)}{\partial \theta} \right\rangle \right) = \\ &= E \left(\sum_{j=1}^n \psi^2(U_j) \left\| U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2 \right) = E_F(\psi^2) \sum_{j=1}^n \left\| U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2 \\ &= E_F(\psi^2) \sum_{j=1}^n \left(U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right)^t \left(U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right) = \\ &= E_F(\psi^2) \text{ traza } \sum_{j=1}^n U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \left(\frac{\partial g(\theta_0, X_j)}{\partial \theta} \right)^t U^{-1} = \\ &= E_F(\psi^2) \text{ traza } U^{-1t} \sum_{j=1}^n \frac{\partial g(\theta_0, X_j)}{\partial \theta} \left(\frac{\partial g(\theta_0, X_j)}{\partial \theta} \right)^t U^{-1} = \\ &= E_F(\psi^2) \text{ traza } I_k = k E_F(\psi^2) < \infty \end{aligned}$$

Lema 4

Para todo $M > 0$ vale que

$$\sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \Gamma_n(t) - E_F(\Psi') I_k \right\| \right\} \xrightarrow[n \rightarrow \infty]{P} 0$$

(tiende en probabilidad a 0)

Demostración

$$\Gamma_n(t) = \sum_{j=1}^n \psi(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} H(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1} - \sum_{j=1}^n \psi'(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} B(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1}$$

$$I_k = \sum_{j=1}^n U^{-1t} B(\theta_0, X_j) U^{-1}$$

Luego,

$$\begin{aligned} & \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \Gamma_n(t) - E_F(\Psi') I_k \right\| \right\} = \\ & \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n (\psi'(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) - E_F(\Psi')) U^{-1t} B(\theta_0, X_j) U^{-1} \right. \right. \\ & \quad \left. \left. + \sum_{j=1}^n \psi'(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} (B(\theta_0 + U^{-1}(t - \theta_0), X_j) - B(\theta_0, X_j)) U^{-1} \right. \right. \\ & \quad \left. \left. + \sum_{j=1}^n \psi(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} (H(\theta_0 + U^{-1}(t - \theta_0), X_j) - H(\theta_0, X_j)) U^{-1} \right. \right. \\ & \quad \left. \left. + \sum_{j=1}^n \psi(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} H(\theta_0, X_j) U^{-1} \right\| \right\} \leq \\ & \leq \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n (\psi'(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) - E_F(\Psi')) U^{-1t} B(\theta_0, X_j) U^{-1} \right\| \right\} \\ & + \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \psi'(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} H(\theta_0, X_j) U^{-1} \right\| \right\} \\ & + \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \psi'(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} (B(\theta_0 + U^{-1}(t - \theta_0), X_j) - B(\theta_0, X_j)) U^{-1} \right\| \right\} \end{aligned}$$

$$+ \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(t-\theta_0), X_j)) U^{-1t} (H(\theta_0 + U^{-1}(t-\theta_0), X_j) - H(\theta_0, X_j)) U^{-1} \right\| \right\}$$

Llamemosle respectivamente a las sumas (I), (II), (III), (IV).-

Veremos que cada una de ellas tiende en probabilidad a 0.-

$$(a) \lim_{n \rightarrow \infty} P(\sup_{\|t-\theta_0\| \leq M} (I) > \epsilon) = 0$$

Basta verlo para el elemento i, j

Sea $v_{n,h}$ = elemento i, j de la matriz $U^{-1t} B(\theta_0, X_h) U^{-1}$

$$\begin{aligned} & \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{h=1}^n (\Psi'(Y_h - g(\theta_0 + U^{-1}(t-\theta_0), X_h)) - E_F(\Psi')) \right\| v_{n,h} \right\} \\ &= \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{h=1}^n \Psi'(U_h - g(\theta_0 + U^{-1}(t-\theta_0), X_h) + g(\theta_0, X_h)) - E_F(\Psi') \right\| v_{n,h} \right\} \end{aligned}$$

Ahora bien,

$$\left| g(\theta_0, X_h) - g(\theta_0 + U^{-1}(t-\theta_0), X_h) \right| = \left| \frac{\partial}{\partial \theta} g(\bar{\theta}, X_h) U^{-1t} (t-\theta_0) \right| \quad (1)$$

Como $h_1(v, z)$ es equicontinua en $(0,0)$

entonces $h_1(v) = \frac{\partial}{\partial \theta} g(\theta_0 + v, X_1) \left(\frac{\partial}{\partial \theta} g(\theta_0 + v, X_1) \right)^t$ lo es en 0

y por tanto $\|h_1(v)\|$ es equicontinua en 0 pues

$$\begin{aligned} & \left\| h_1(v) - h_1(0) \right\| \geq \left| \|h_1(v)\| - \|h_1(0)\| \right| \quad \text{y como} \\ & \|h_1(v)\| = \left\| \frac{\partial}{\partial \theta} g(\theta_0 + v, X_1) \right\|^2 \end{aligned}$$

tenemos que $\frac{\partial}{\partial \theta} g(\theta_0 + v, X_1) = f_1(v)$ es equicontinua en 0

pues si f_1^2 es equicontinua entonces $|f_1|$ lo es.

Volviendo a la acotación (1), como estamos en $\|t-\theta_0\| \leq M$,

entonces, por la hipótesis A6) (v) $\|U^{-1t}(t-\theta_0)\| \leq d$ si $n \geq n_0$

y $d > 0$ arbitrario.

Entonces, el punto intermedio $\bar{\theta}$, se puede poner como $\bar{\theta} = \theta_0 + v$ con $|v| < d$

$$\begin{aligned} \text{luego, } \max_{1 \leq h \leq n} \left| \frac{\partial g(\bar{\theta}, X_h) U^{-1t}}{\partial \theta} (t - \theta_0) \right| &\leq \max_{1 \leq h \leq n} \left\| \frac{\partial g(\bar{\theta}, X_h)}{\partial \theta} \right\| \|U^{-1t}\| M \leq \\ &\leq \max_{1 \leq h \leq n} \left\{ \left\| \frac{\partial g(\bar{\theta}, X_h)}{\partial \theta} \right\| - \left\| \frac{\partial g(\theta_0, X_h)}{\partial \theta} \right\| \right\} \|U^{-1t}\| M + \left\| \frac{\partial g(\theta_0, X_h)}{\partial \theta} \right\| \|U^{-1t}\| M \end{aligned}$$

y ahora usando equicontinuidad para la primera parte,

$$\leq \varepsilon \|U^{-1t}\| M + \max_{1 \leq j \leq n} \left\| \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\| \|U^{-1t}\| M$$

$$\text{Adem\u00e1s } \max_{1 \leq j \leq n} \left\| \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\| \|U^{-1t}\| M \rightarrow 0 \text{ cuando } n \rightarrow \infty$$

$$\text{pues } \max_{1 \leq j \leq n} \left\| \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2 \|U^{-1t}\|^2 \leq \max_{1 \leq h \leq n} a^2 \left\| \frac{\partial g(\theta_0, X_h)}{\partial \theta} \right\|^2 / \|U^t U\| =$$

$$= \max_{1 \leq j \leq n} \frac{a^2 \left\| \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2}{\sum_{j=1}^n \left\| \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2} \rightarrow 0 \text{ para } n \rightarrow \infty \text{ por la hip\u00f3tesis A6.}$$

Luego, para el elemento i, j

$$\lim_{n \rightarrow \infty} P(\sup_{\|t - \theta_0\| \leq M} (1) > \varepsilon) \leq \lim_{b \rightarrow 0} \limsup_{n \rightarrow \infty} P\left(\sup_{|s_h| \leq b} \left| \sum_{h=1}^n (\psi'(U_h - s_h) - E_F(\psi')) v_{nh} \right| > \varepsilon\right)$$

y ahora aplicaremos el lema 1, con

$$f_{s_h}(u) = \psi'(U_h - s_h) - E_F(\psi') \text{ que verifica las hip\u00f3tesis del}$$

lema, ya que

$$(i) \lim_{r \rightarrow 0} f_r(u) = f_0(u) \text{ p.p. (continuidad de } \psi' \text{ y } F \text{ continua)}$$

$$(ii) E(f_0(u)) = 0$$

$$(iii) \exists d_0 > 0 \text{ tal que } \sup_{|r| \leq d_0} |f_r(U)| \leq f(U) \text{ con } E(f^2(U)) < \infty$$

$$\sup_{|r| \leq d_0} |\psi'(U-r) - E_F(\psi')| \leq \sup_{|r| \leq d_0} |\psi'(U-r)| - E_F(\psi') \text{ que por}$$

la hip\u00f3tesis A3 es de cuadrado integrable.

$$(iv) \sum_{h=1}^n |v_{n,h}| \leq M \quad \text{para todo } n.$$

$$\sum_{h=1}^n |v_{n,h}| = \text{elemento } i,j \text{ de la matriz } \sum_{h=1}^n |U^{-1t} B(\theta_0, X_h) U^{-1}|$$

$$\text{pero } \sum_{h=1}^n \|U^{-1t} B(\theta_0, X_h) U^{-1}\| = \sum_{h=1}^n \left\| U^{-1t} \frac{\partial g(\theta_0, X_h)}{\partial \theta} \right\|^2 = k \quad \text{como ya}$$

vimos en el lema 3.

$$(v) \lim_{n \rightarrow \infty} \max_{1 \leq h \leq n} |v_{n,h}| = 0$$

$$\max_{1 \leq h \leq n} |v_{n,h}| = \max \text{ elemento } i,j \text{ de } |U^{-1t} B(\theta_0, X_h) U^{-1}|$$

$$\text{Sea } Z_h = U^{-1t} \frac{\partial g(\theta_0, X_h)}{\partial \theta}$$

$$\max_{1 \leq h \leq n} |v_{n,h}| = \max_{1 \leq h \leq n} |Z_{ih} Z_{jh}| \leq \max_{1 \leq h \leq n} \|Z_h\|^2 = \max_{1 \leq h \leq n} \left\| U^{-1t} \frac{\partial g(\theta_0, X_h)}{\partial \theta} \right\| \rightarrow 0$$

por la hipótesis A6.-

Luego, por el lema 1, vale (a). —

$$(b) \lim_{n \rightarrow \infty} P(\sup_{\|t - \theta_0\| \leq M} (2) > \varepsilon) = 0$$

igual que en (a), basta verlo para el elemento i, j , y para ello basta mostrar que

$$\lim_{b \rightarrow 0} \limsup_{n \rightarrow \infty} P(\sup_{|s_h| \leq b} \left| \sum_{h=1}^n \Psi(U_h - s_h) W_{n,h} \right| > \varepsilon) = 0$$

donde $W_{n,h}$ = elemento i, j de la matriz $U^{-1t} H(\theta_0, X_h) U^{-1}$

Tomemos $f_s(u) = \Psi(u-s)$

que verifica las hipótesis (i), (ii), (iii) del lema 1, así como

$$(iv) \sum_{h=1}^n |W_{n,h}| \leq M$$

$$\sum_{h=1}^n |W_{n,h}| \leq \sum_{h=1}^n \left\| U^{-1t} H(\theta_0, X_h) U^{-1} \right\|$$

$$\leq a^2 \sum_{h=1}^n \|H(\theta_0, X_h)\| / \|U^t U\| = a^2 \sum_{h=1}^n \|H(\theta_0, X_h)\| / \sum_{h=1}^n \left\| \frac{\partial}{\partial \theta} g(\theta_0, X_h) \right\|^2$$

$$\leq a^2 \cdot C \quad \text{pues por la hipótesis A6} \quad \|H(\theta_0, X_h)\| \leq C \left\| \frac{\partial}{\partial \theta} g(\theta_0, X_h) \right\|^2$$

$$(v) \lim_{n \rightarrow \infty} \max_{1 \leq h \leq n} |W_{n,h}| = 0$$

$$\max_{1 \leq h \leq n} |W_{n,h}| \leq \max_{1 \leq h \leq n} \|U^{-1t} H(\theta_0, X_h) U^{-1}\| \leq \max_{1 \leq h \leq n} \|H(\theta_0, X_h)\| a^2 / \|U^t U\|$$

$$\leq \max_{1 \leq h \leq n} C \left\| \frac{\partial}{\partial \theta} g(\theta_0, X_h) \right\|^2 a^2 / \sum_{h=1}^n \left\| \frac{\partial}{\partial \theta} g(\theta_0, X_h) \right\|^2 \rightarrow 0 \quad \text{por la hipótesis A6.}$$

pótesis A6.--

Luego, por el lema 1, vale (b).--

$$(c) \lim_{n \rightarrow \infty} P(\sup_{\|t - \theta_0\| \leq M} (3) > \epsilon) = 0$$

Como $\sup_{\|s\| \leq d} \Psi'(u-s)$ está acotado en probabilidad, basta probar que

$$\sup_{\|t - \theta_0\| \leq M} \left\{ \sum_{h=1}^n \|U^{-1t} (B(\theta_0 + U^{-1}(t - \theta_0), X_h) - B(\theta_0, X_h)) U^{-1}\| \right\} \rightarrow 0$$

luego, si $n \gg n_0$, basta ver que

$$\sup_{\|v\| \leq d} \left\{ \sum_{h=1}^n \|U^{-1t} (B(\theta_0 + v, X_h) - B(\theta_0, X_h)) U^{-1}\| \right\} \rightarrow 0 \quad \text{cuando } d \rightarrow 0$$

Probaremos algo ligeramente mas general que nos será útil en la demostración de normalidad asintótica.

Veremos que

$$\sup_{\|v\| \leq d} \left\{ \sum_{h=1}^n \|U^{-1t} (h(v, z, X_h) - h(0, 0, X_h)) U^{-1}\| \right\} \rightarrow 0 \quad \text{cuando } d \rightarrow 0$$

$\|z\| \leq d$

que cuando $v = z$ nos queda la expresión anterior,.

$$\sup_{\substack{\|v\| \leq d \\ \|z\| \leq d}} \left\{ \sum_{h=1}^n \| U^{-1t} (h(v, z, X_h) - h(0, 0, X_h)) U^{-1} \| \right\} \leq$$

$$\sup_{\substack{\|v\| \leq d \\ \|z\| \leq d}} \frac{a^2}{\|U^t U\|} \sum_{h=1}^n \| h(v, z, X_h) - h(0, 0, X_h) \|$$

sea d tal que $\|h(v, z, X_h) - h(0, 0, X_h)\| < \varepsilon_1 \quad \forall h, \quad \forall \|v\| \leq d, \|z\| \leq d$

Sea D de la hipótesis A6 (iv)

$$\text{luego } 0 < D < \left\| \sum_{h=1}^n h(0, 0, X_h) \right\| / n$$

$$\text{Luego, si tomamos } \varepsilon_1 < \varepsilon \cdot D / a^2 \quad \text{como } \|U^t U\| = \left\| \sum_{h=1}^n h(0, 0, X_h) \right\|$$

tenemos que se verifica que

$$\sup_{\substack{\|v\| \leq d \\ \|z\| \leq d}} \frac{a^2}{\|U^t U\|} \sum_{h=1}^n \| h(v, z, X_h) - h(0, 0, X_h) \| < \varepsilon$$

Finalmente veremos que

$$(d) \quad \lim_{n \rightarrow \infty} P(\sup_{\|t - \theta_0\| \leq M} (4) > \varepsilon) = 0$$

Como en (c), como $\sup_{|s| \leq d} \psi(u-s)$ esta acotado en probabilidad, basta

probar que

$$\sup_{\|t - \theta_0\| \leq M} \left\{ \sum_{h=1}^n \| U^{-1t} (H(\theta_0 + U^{-1}(t - \theta_0), X_h) - H(\theta_0, X_h)) U^{-1} \| \right\} \rightarrow 0$$

Si $n \geq n_0$, basta con que

$$\sup_{\|v\| \leq d} \left\{ \sum_{h=1}^n \| U^{-1t} (H(\theta_0 + v, X_h) - H(\theta_0, X_h)) U^{-1} \| \right\} \rightarrow 0$$

luego, como en (c) basta con tomar d tal que

$$\|H(\theta_0 + v, X_h) - H(\theta_0, X_h)\| < \varepsilon_1 \quad \text{para todo } h, \text{ y para todo } v, \|v\| \leq d$$

Luego, tomando igual que antes $\varepsilon_1 < \varepsilon D / a^2$ listo.--

Lema 5

Para cada muestra Y_1, Y_2, \dots, Y_n existe una sucesión de

abiertos D_1, \dots, D_r, \dots tales que

$$\bigcup_{j \geq 1} \bar{D}_j = R^k \text{ y tales que}$$

$\eta_{n\Gamma}(t)$ es continuamente diferenciable en D_j $1 \leq j$ para toda Γ

definida positiva.

Más aun, en cada D_j

(a₁)

$$\begin{aligned} \frac{\partial}{\partial t} \eta_{n\Gamma}(t) = & -I_k - \Gamma^{-1} \left(\sum_{j=1}^n \Psi'(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} R(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1} - \right. \\ & \left. - \sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)) U^{-1t} H(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1} \right) \end{aligned}$$

(a₂) Existe $\varepsilon_1(k_0)$ tal que

$$\lim_{n \rightarrow \infty} P \left(\sup_{\substack{\|t - \theta_0\| \leq M \\ t \in \bigcup_{j \geq 1} D_j \\ \|\Gamma - \Gamma_0\| \leq \varepsilon_1}} \left\{ \left\| \frac{\partial}{\partial t} \eta_{n\Gamma}(t) \right\| \right\} > k_0 \right) = 0$$

Demostración.

Sean a_1, \dots, a_p los puntos donde Ψ no es diferenciable.

Para cada muestra Y_1, \dots, Y_n $\frac{\partial}{\partial t} \eta_{n\Gamma}(t)$ está definida y es - (a₁)

para toda Γ definida positiva, salvo en los conjuntos

$$R_{hi} = \left\{ t / g(\theta_0 + U^{-1}(t - \theta_0), X_i) = Y_i - a_h \right\} \quad \begin{matrix} 1 \leq i \leq n \\ 1 \leq h \leq p \end{matrix}$$

que parten a R^k en abiertos densos por la hipótesis A₇.

$$(a_2) \quad \Gamma_0 = E_F(\Psi') I_k$$

Notemos por $Q = \{ (t, \Gamma) / \|t - \theta_0\| \leq M; t \in \bigcup_{j \geq 1} D_j; \|\Gamma - \Gamma_0\| \leq \epsilon_1 \}$

$$\lim_{n \rightarrow \infty} P(\sup_{(t, \Gamma) \in Q} \{ \left\| \frac{\partial}{\partial t} \gamma_{n\Gamma}(t) \right\| \} < k_0) =$$

$$\lim_{n \rightarrow \infty} P(\sup_{(t, \Gamma) \in Q} \{ \left\| \frac{\partial}{\partial t} \gamma_{n\Gamma}(t) - I_k + E_F(\Psi') I_k \Gamma_0^{-1} \right\| \} < k_0) =$$

$$= \lim_{n \rightarrow \infty} P(\sup_{(t, \Gamma) \in Q} \{ \left\| I_k - \Gamma_n(t) \Gamma^{-1} - I_k + E_F(\Psi') I_k \Gamma_0^{-1} \right\| \} < k_0)$$

$$\geq \lim_{n \rightarrow \infty} P(\sup_Q (\|(\Gamma_n(t) - E_F(\Psi')) \Gamma^{-1}\| + \|\Gamma^{-1} - \Gamma_0^{-1}\| |E_F(\Psi')|) < k_0)$$

$$\geq \lim_{n \rightarrow \infty} P(\sup_Q (\|\Gamma^{-1}\| \|\Gamma_n(t) - E_F(\Psi') I_k\| + \|\Gamma^{-1} - \Gamma_0^{-1}\| |E_F(\Psi')|) < k_0) =$$

$$= 1 \quad \text{ya que } \sup_{\|t - \theta_0\| \leq M} \|\Gamma_n(t) - E_F(\Psi') I_k\| < \epsilon \text{ en probabilidad por}$$

el lema 4

$$\text{y } \|\Gamma^{-1} - \Gamma_0^{-1}\| < \epsilon_2 \text{ pues } \|\Gamma - \Gamma_0\| < \epsilon_1 \text{ por hipótesis.---}$$

Lema 6

Sea $0 < k_0 < 1$

(i) Para todo $M > 0$, sea

$$A_{n, M} = \{ \|\gamma_{n\Gamma}(t) - \gamma_{n\Gamma}(t')\| \leq k_0 \|t - t'\|; \|t - \theta_0\| \leq M, \|t' - \theta_0\| \leq M, \|\Gamma - \Gamma_0\| \leq \epsilon_0 \}$$

con ϵ_0 del lema 3.- Entonces

$$P) \lim_{n \rightarrow \infty} P(A_{n, M}) = 1$$

(ii) Para todo $M > 0$ sea $0 < \nu < |E_F(\Psi')|$

$$B_{n, M} = \{ \gamma_{n\Gamma} \text{ lleva } \|t - \theta_0\| \leq M \text{ en si mismo } \forall \Gamma \text{ tal que } \|\Gamma - \Gamma_0\| < \nu \}$$

Dado $d > 0$, existe $M_0(d)$ y ν_0 tal que

$\lim_{n \rightarrow \infty} P(B_{n, M, \rho}) > 1-d$ para todo $M \geq M_0(d)$

(iii) Para todo $M > 0$ y $0 < \rho < |E_F(\Psi')|$

sea $C_{n, M, \rho} = \{ \Gamma_n(t) \text{ lleva } \|t - \theta_0\| \leq M \text{ en } \|\Gamma - \Gamma_0\| < \rho \}$

entonces

T) $\lim_{n \rightarrow \infty} P(C_{n, M, \rho}) = 1$

Demostración

(i) Sabemos por el lema 5 que

$$\lim_{n \rightarrow \infty} P(\sup_{(t, \Gamma) \in Q} \left\| \frac{\partial}{\partial t} \zeta_{n\Gamma}(t) \right\| \leq k_0) = 1$$

Por otro lado, por la hipótesis A7, existe dados t y t' una sucesión finita de puntos t_0, \dots, t_q tales que $t_0 = t$; $t_q = t'$ y

$$[t_{i+1}, t_i] \subset \bar{D}_i \text{ con } (t_{i+1}, t_i) \subset D_i \text{ o bien } (t_{i+1}, t_i) \subset \bar{D}_i - D_i$$

Luego, basta probar que si $(t_{i+1}, t_i) \overset{\circ}{\subset} D_i$ entonces

$$\|\zeta_{n\Gamma}(t_{i+1}) - \zeta_{n\Gamma}(t_i)\| < k_0 \|t_{i+1} - t_i\| \text{ con la misma constan-}$$

te k_0 , pues como $\zeta_{n\Gamma}(t)$ es continua y $\bigcup_i D_i$ es densa valdrá también

cuando $(t_{i+1}, t_i) \subset \bar{D}_i - D_i$

y luego dados t, t' tendremos que como la secuencia es finita y están sobre una recta

$$\begin{aligned} \|\zeta_{n\Gamma}(t) - \zeta_{n\Gamma}(t')\| &\leq \sum_{j=1}^q \|\zeta_{n\Gamma}(t_j) - \zeta_{n\Gamma}(t_{j-1})\| \leq \\ &\leq k_0 \sum_{j=1}^q \|t_j - t_{j-1}\| = k_0 \|t - t'\| \end{aligned}$$

Luego, nos resta probar (i) en cada $D_j \cap \{ \|t - \theta_0\| \leq M \}$ si $\|\Gamma - \Gamma_0\| \leq \epsilon_0$ y $(t, t') \subset D_j$

Notaremos con Q_j el conjunto correspondiente cuando $t \in D_j$ $Q = \bigcup_{j \geq 1} Q_j$

Sean t y t' en dichas condiciones

$$\|\zeta_{n\Gamma}(t) - \zeta_{n\Gamma}(t')\| \leq \sup_{Q_j} \left\{ \left\| \frac{\partial}{\partial t} \zeta_{n\Gamma}(t) \right\|_0 \|t - t'\| \right\} \leq$$

$$\leq \sup_{Q_j} \left\| \frac{\partial}{\partial t} \gamma_{n\Gamma}(t) \right\| \|t-t'\| < k_0 \|t-t'\| \quad \text{como queríamos.}$$

(ii) Por el lema 3 tenemos que

$$P\left(\sup_{\|\Gamma-\Gamma_0\| \leq \mathbb{E}_F(\Psi')/2} \|\gamma_{n\Gamma}(\theta_0) - \theta_0\| > k_1\right) \leq d \quad \text{para todo } n$$

$$\begin{aligned} \|\gamma_{n\Gamma}(t) - \theta_0\| &\leq \|\gamma_{n\Gamma}(t) - \gamma_{n\Gamma}(\theta_0)\| + \|\gamma_{n\Gamma}(\theta_0) - \theta_0\| \leq \\ &\leq k_0 M + k_1 \leq M(1-k_0) + k_0 M = M \quad \text{si } M > k_1/(1-k_0) \end{aligned}$$

Luego si t es tal que $\|t-\theta_0\| \leq M$

$$\sup_{\|\Gamma-\Gamma_0\| \leq \nu_0} \|\gamma_{n\Gamma}(t) - \theta_0\| \leq \sup\{\|\gamma_{n\Gamma}(\theta_0) - \theta_0\|\} + \sup\{\|\gamma_{n\Gamma}(t) - \gamma_{n\Gamma}(\theta_0)\|\} < M$$

Luego, tenemos que

$$\lim_{n \rightarrow \infty} P(B_{n,M,\nu_0}) \geq \lim_{n \rightarrow \infty} P(A_{n,M}) - \lim_{n \rightarrow \infty} P\left(\sup_{\|\Gamma-\Gamma_0\| \leq \mathbb{E}_F(\Psi')/2} \|\gamma_{n\Gamma}(\theta_0) - \theta_0\| > k_1\right) \geq$$

$$\geq 1 - d \quad \text{siendo } \nu_0 = \min(\mathbb{E}_F(\Psi')/2, \epsilon_0) \quad \text{y } \epsilon_0 \text{ del lema 3.-}$$

(iii) es directamente el lema 4.-

Lema 7

Sea $D_{n,M} = \left\{ \gamma_n(t) = t \text{ tiene un \u00fanico punto fijo } t^0 \text{ en } \|t-\theta_0\| \leq M, \text{ y} \right.$
 para cualquier t_0 tal que $\|t_0 - \theta_0\| \leq M$, $\gamma_n^{(k)}(t_0) \rightarrow t^0 \}$
 donde $\gamma_n^{(k)}(t_0)$ se define iterativamente = $\gamma_n(\gamma_n^{(k-1)}(t_0))$

Tenemos entonces que $\forall d > 0$, existe $M_0(d)$ tal que $\forall M \geq M_0(d)$

$$\lim_{n \rightarrow \infty} P(D_{n,M}) \geq 1-d$$

Demostración

Sea $0 < k_0 < 1$

Para ver que $D_{n,M} \supset (A_{n,M} \cap B_{n,M,\mu_0} \cap C_{n,M,\mu_0})$

Si ocurre $A_{n,M}$ como $k_0 < 1$, las $\eta_{n\Gamma}(t)$ son contracciones para cada Γ con $\|\Gamma - \Gamma_0\| < \epsilon_0$.

Como ocurre B_{n,M,μ_0} transforman $\|t - \theta_0\| \leq M$ en si mismo.- Luego, por

el teorema del punto fijo de Banach, tienen un único punto fijo para cada Γ , pero además es el mismo para cada Γ , pues

t_0 es un punto fijo sii $\sum_{j=1}^n \psi(\cdot) U^{-1} \frac{\partial}{\partial \theta} g = 0$ que no depende de Γ .

Finalmente, $\eta_n(t) = t - (P'(t))^{-1} P(t) = t$ sii $P(t) = 0$

Luego, tiene el mismo punto fijo que las $\eta_{n\Gamma}(t)$.

Nos resta ver que $\eta_n^{(k)}(t_0) \rightarrow t_0$

$$\eta_n(t) = \eta_{n\Gamma_n}(t)(t)$$

$$t_1 = \eta_n(t_0) = \eta_{n\Gamma_n}(t_0)(t_0)$$

$$t_k = \eta_n(t_{k-1}) = \eta_{n\Gamma_n}(t_{k-1})(t_{k-1})$$

Como $\eta_{n\Gamma_n}(t)$ son contracciones si $\|\Gamma_n - \Gamma_0\| < \epsilon$, con el mismo pun-

to fijo, por el lema 2, convergen al punto fijo común t_0 ; y

$\|\Gamma_n(t_k) - \Gamma_0\| < \epsilon_2$ si ocurre C_{n,M,μ_0} puesto que como $\|t - \theta_0\| \leq M$

y vale $C_{n,M,\mu_0} \|\Gamma_n(t_0) - \Gamma_0\| < \epsilon$, iterando, listo.---

Proposición 1

(i) $\lim_{n \rightarrow \infty} P(\hat{\theta}_n = \lim_{j \rightarrow \infty} \hat{\theta}_{n,j} \text{ y } \hat{\theta}_n \text{ satisface la ec. original}) = 1$

(ii) $U(\hat{\theta}_n - \theta_0)$ está acotado en probabilidad.-

Demostración

(i) Sea $\hat{\theta}_{n,j}^* = \theta_0 + U(\hat{\theta}_{n,j} - \theta_0)$

$$\hat{\theta}_n^* = \theta_0 + U(\hat{\theta}_n - \theta_0)$$

$\hat{\theta}_n$ satisface la ecuación (I) $\sum_{j=1}^n \Psi(Y_j - g(\theta, X_j)) \frac{\partial}{\partial \theta} g(\theta, X_j) = 0$

si y solo si $\hat{\theta}_n^*$ satisface:

$$\sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(\theta - \theta_0), X_j)) U^{-1} \frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\theta - \theta_0), X_j) = 0 \quad \text{(II)}$$

pues $\theta_0 + U^{-1}(\theta_0 + U(\hat{\theta}_n - \theta_0) - \theta_0) = \hat{\theta}_n$

Luego, para (i) basta probar que

$\lim_{n \rightarrow \infty} P(\lim_{j \rightarrow \infty} \hat{\theta}_{n,j}^* = \hat{\theta}_n^* \text{ y } \hat{\theta}_n^* \text{ satisface la ecuación (II)}) = 1$

pero, $\left\{ \lim_{j \rightarrow \infty} \hat{\theta}_{n,j}^* = \hat{\theta}_n^* \text{ y } \hat{\theta}_n^* \text{ satisface (II)} \right\} \supset D_{n,M} \cap \{ \|\hat{\theta}_{n,0}^* - \theta_0\| \leq M \}$

donde $\hat{\theta}_{n,0}^*$ es el estimador inicial normalizado que definimos antes.

Sea $d > 0$, luego por la hipótesis A4 $\|\hat{\theta}_{n,0}^* - \theta_0\|$ está acotado en probabilidad.

Luego existe M_2 tal que $P(\|\hat{\theta}_{n,0}^* - \theta_0\| \leq M) \geq 1 - d/2$ si $M \geq M_2$

Si tomamos entonces $M_3 = \max(M_0(d/2), M_2)$ con $M_0(d/2)$ del lema 7

tenemos que

$$P(\lim_{j \rightarrow \infty} \hat{\theta}_{n,j}^* = \hat{\theta}_n^* \text{ y } \hat{\theta}_n^* \text{ verifica (II)}) \geq 1 - d$$

(ii) $U(\hat{\theta}_n - \theta_0)$ está acotado en probabilidad.

$U(\hat{\theta}_n - \theta_0)$ está acotado en probabilidad si y solo si

$\|\hat{\theta}_n^* - \theta_0\|$ está acotado en probabilidad.

sea $d > 0$

$$\{\|\hat{\theta}_n^* - \theta_0\| \leq M\} \supset D_{n,M} \cap \{\|\hat{\theta}_{n,0}^* - \theta_0\| \leq M\}$$

luego, tomando M_3 como en (i)

tenemos que

$$\lim_{n \rightarrow \infty} P(\|\hat{\theta}_n^* - \theta_0\| \leq M_3) \geq 1 - d$$

por tanto $\|\hat{\theta}_n^* - \theta_0\|$ está acotado en probabilidad.---

SECCION I B

Normalidad asintótica

Sea $\hat{\theta}^*$ la solución que consideramos de la ecuación

$$(II) \sum_{j=1}^n \Psi(U_j - (g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j) - g(\theta_0, x_j))) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j) = 0$$

Proposición 2

Para todo $a \in \mathbb{R}^k$ con $\|a\| = 1$ vale que si $|\psi''| < K$

$$a^t (\hat{\theta}^* - \theta_0) \xrightarrow{w} N(0, E_F(\psi^2) / E_F^2(\psi'))$$

Demostración

$$\begin{aligned} & \sum_{j=1}^n \Psi(U_j - (g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j) - g(\theta_0, x_j))) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j) = \\ & = \sum_{j=1}^n \Psi(U_j - (g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j) - g(\theta_0, x_j))) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, x_j) + v_n \quad (II)_I \end{aligned}$$

donde v_n es

$$v_n = \sum_{j=1}^n \Psi(U_j - (g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j) - g(\theta_0, x_j))) U^{-1t} \left(\frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j) - \frac{\partial}{\partial \theta} g(\theta_0, x_j) \right)$$

y $v_n \xrightarrow{p} 0$ ya que

$$v_n = \sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(\hat{\theta}^* - \theta_0), x_j)) U^{-1t} H(\bar{\theta}_j, x_j) U^{-1} (\hat{\theta}^* - \theta_0)$$

y $\|\hat{\theta}^* - \theta_0\|$ está acotado en probabilidad por la proposición 1

y del lema 4 de consistencia tenemos que

$$\sup_{\|t - \theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), x_j)) U^{-1t} H(\theta_0, x_j) U^{-1} \right\| \right\} \xrightarrow{p} 0 \quad y$$

$$\sup_{\|t - \theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(t - \theta_0), x_j)) U^{-1t} (H(\theta_0 + U^{-1}(t - \theta_0), x_j) - H(\theta_0, x_j)) U^{-1} \right\| \right\}$$

$\|t - \theta_0\| \leq M$ tiende en probabilidad a 0

Luego $\sum_{j=1}^n \Psi(Y_j - g(\theta_0 + U^{-1}(\hat{\theta}^v - \theta_0), X_j)) U^{-1} H(\bar{\theta}_j, X_j) U^{-1} \xrightarrow{r} 0$

y por lo tanto V_n lo hace.

Sea $c_\Psi(u, x) = \frac{\Psi(u+x) - \Psi(u)}{x}$

$\Psi(u+x) = c_\Psi(u, x)x + \Psi(u)$

Definimos $\Psi'(x) = 0$ donde no está definida.

notaremos

$$|g(\theta_0 + U^{-1}(\hat{\theta}^v - \theta_0), X_j) - g(\theta_0, X_j)| = \left| \frac{\partial}{\partial \theta} g(\bar{\theta}_j, X_j) U^{-1t} (\hat{\theta}^v - \theta_0) \right| = |I_g(X_j)|$$

Sustituyendo en (II)₁ .a tenemos que:

$$\sum_{j=1}^n \Psi(U_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a - \sum_{j=1}^n c_\Psi(U_j, I_g) I_g(X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a + V_n = 0$$

$$\text{Llamaremos } W_1 = \sum_{j=1}^n \Psi(U_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a$$

$$W_2 = \sum_{j=1}^n c_\Psi(U_j, I_g) (\hat{\theta}^v - \theta_0)^t U^{-1t} \frac{\partial}{\partial \theta} g(\bar{\theta}_j, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a =$$

$$= (\hat{\theta}^v - \theta_0)^t \left[\sum_{j=1}^n c_\Psi(U_j, I_g) U^{-1t} \frac{\partial}{\partial \theta} g(\bar{\theta}_j, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right] a =$$

$$= (\hat{\theta}^v - \theta_0)^t W_2^* a$$

Llamaremos además

$$G_2 = \sum_{j=1}^n \Psi'(U_j) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1}$$

Tenemos entonces que $\boxed{W_1 - W_2 + V_n = 0}$

$$(a_1) \quad W_1 \xrightarrow{w} N(0, E_F(\Psi^2))$$

pues $E(W_1) = 0$

$$\text{var}(W_1) = \sum_{j=1}^n E_F(\Psi^2) \left(\frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \right)^2 =$$

$$\begin{aligned}
 &= E_F(\Psi^2) \sum_{j=1}^n a^t U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a = \\
 &= E_F(\Psi^2) a^t \left[\sum_{j=1}^n U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right] a = E_F(\Psi^2) a^t I_k a = \\
 &= E_F(\Psi^2)
 \end{aligned}$$

y los coeficientes

$$\begin{aligned}
 \max_{1 \leq j \leq n} \left| \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \right|^2 &\leq \max_{1 \leq j \leq n} \|U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j)\|^2 \|a\|^2 = \\
 &= \max_{1 \leq j \leq n} \|U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j)\|^2 \rightarrow 0 \quad \text{por hipótesis.}
 \end{aligned}$$

Luego se verifica la condición de Lindenberg y vale (a_1)

Por tanto, tenemos que

$$W_1 / E_F(\Psi^2) \rightarrow N(0, E_F(\Psi^2) / E_F(\Psi^2))$$

$$(a_2) \text{ Tenemos que } (\hat{\theta}^* - \theta_0)^t W_2^* a = W_1 + V_n$$

Consideremos

$$\begin{aligned}
 (\hat{\theta}^* - \theta_0)^t G_2 a &= (\hat{\theta}^* - \theta_0)^t \left[\sum_{j=1}^n (\Psi'(U_j) - E_F(\Psi')) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right] a \\
 &+ (\hat{\theta}^* - \theta_0)^t E_F(\Psi') \left[\sum_{j=1}^n U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right] a
 \end{aligned}$$

El segundo sumando = $(\hat{\theta}^* - \theta_0)^t E_F(\Psi') a$

y el primero $\rightarrow 0$ en probabilidad, pues

si llamamos

$$W_3 = \sum_{j=1}^n (\Psi'(U_j) - E_F(\Psi')) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a$$

tenemos que $E(W_3) = 0$, y su matriz de covarianza es

$$\text{cov}(W_3) = \text{var}(\Psi') \sum_{j=1}^n \left(\frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \right)^2 U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1}$$

$$\text{Luego, como traza } \sum_{i=1}^n z_i z_i^t = \sum_{i=1}^n z_i^t z_i = \sum_{i=1}^n \|z_i\|^2$$

$$\text{Tenemos que } E(\|W_3\|^2) = \text{var}(\Psi') \sum_{j=1}^n \left(\frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \right)^2 \|U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j)\|^2$$

$$\leq \text{var}(\Psi') \max_{1 \leq j \leq n} \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right\|^2 \sum_{j=1}^n \left(\frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \right)^2 =$$

$$= \text{var}(\Psi') \max_{1 \leq j \leq n} \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right\|^2 \rightarrow 0$$

y como por la proposición 1 $(\hat{\theta}^* - \theta_0)$ está acotada en probabilidad

$$(\hat{\theta}^* - \theta_0)^t W_3 \xrightarrow{P} 0$$

Luego, si probamos que $(\hat{\theta}^* - \theta_0)^t (W_2^* - G_2) a \rightarrow 0$

Tendremos probada la proposición 2

$$\text{ya que } (\hat{\theta}^* - \theta_0)^t W_2^* a = W_1 + V_n \xrightarrow{W} N(0, E_F(\Psi^2))$$

$$\text{y entonces valdría que } (\hat{\theta}^* - \theta_0)^t W_2 a \xrightarrow{P} E_F(\Psi') (\hat{\theta}^* - \theta_0)^t a$$

Por tanto nos queda probar

$$(a_3) \quad (\hat{\theta}^* - \theta_0)^t (W_2^* - G_2) a \xrightarrow{P} 0$$

$$W_2^* - G_2 = \sum_{j=1}^n (c_{\Psi}(U_j, I_g) - \Psi'(U_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1}$$

$$+ \sum_{j=1}^n c_{\Psi}(U_j, I_g) U^{-1t} \left(\frac{\partial}{\partial \theta} g(\bar{\theta}_j, X_j) - \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1}$$

Llamemosle sumas I y II respectivamente.

Consideremos la suma I

$|I_g| \leq d$ si $n \geq n_0$ como ya probamos en la parte de consistencia.

Sea entonces t un punto de discontinuidad de $\Psi'(\cdot)$.

(Como son finitos veremos luego que basta considerar uno y repetir el argumento)

La suma I la descomponemos en:

$$\sum_{j=1}^n (c_{\Psi}(U_j, I_g) - \Psi'(U_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} I_{[t-d, t+d]}(U_j) +$$

$$+ \sum_{j=1}^n (\psi(U_j, I_g) - \psi'(U_j)) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} I_{[t-d, t+d]}^c (U_j)$$

Donde por $I_{[,]}$ notamos el indicador del conjunto.

Tomemos el término correspondiente a esta segunda suma, que notamos

$$\begin{aligned} & (\hat{\theta} - \theta_0)^t (W_2 - G_2)_2 a = \\ & = \frac{1}{2} (\hat{\theta} - \theta_0)^t \sum_{j=1}^n \psi''(U_j - \{j I_g) I_g U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} a I_{[t-d, t+d]}^c (U_j) \end{aligned}$$

que sustituyendo $I_g(X_j) = (\hat{\theta} - \theta_0)^t U^{-1t} \frac{\partial g(\bar{\theta}_j, X_j)}{\partial \theta}$ nos queda

$$\begin{aligned} & = \frac{1}{2} (\hat{\theta} - \theta_0)^t \sum_{j=1}^n \psi''(U_j - \{j I_g) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} a \frac{\partial^t g(\bar{\theta}_j, X_j)}{\partial \theta} U^{-1} (\hat{\theta} - \theta_0) I_{[t-d, t+d]}^c (U_j) \\ & = \frac{1}{2} (\hat{\theta} - \theta_0)^t \left[\sum_{j=1}^n \psi''(U_j - \{j I_g) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} \left(\frac{\partial^t g(\bar{\theta}_j, X_j)}{\partial \theta} U^{-1} a \right) I_{[t-d, t+d]}^c \right] (\hat{\theta} - \theta_0) \\ & = \frac{1}{2} (\hat{\theta} - \theta_0)^t W_4 (\hat{\theta} - \theta_0) \end{aligned}$$

Vamos a ver entonces, que $\|W_4\|^2 \xrightarrow{P} 0$, usando el hecho de que

$$|\psi''| < K$$

$$\text{Llamaremos } \bar{U}_j = U_j - \{j I_g$$

$$b_i = \frac{\partial g(\bar{\theta}_i, X_i)}{\partial \theta} U^{-1} a$$

$$z_i = U^{-1t} \frac{\partial g(\theta_0, X_i)}{\partial \theta} \quad \text{luego,}$$

$$W_4 = \sum_{j=1}^n \psi''(\bar{U}_j) b_j z_j z_j^t I_{[t-d, t+d]}^c (U_j)$$

$$\|W_4\|^2 = \sum_{k=1}^n \sum_{l=1}^n \left(\sum_{j=1}^n \psi''(\bar{U}_j) I_{[t-d, t+d]}^c (U_j) b_j z_j^k z_j^l \right)^2 =$$

$$= \sum_{k=1}^n \sum_{l=1}^n \sum_{i=1}^n \sum_{j=1}^n \psi''(\bar{U}_i) I(U_i) \psi''(\bar{U}_j) I(U_j) b_i b_j z_i^k z_i^l z_j^k z_j^l =$$

$$= \sum_{i=1}^n \sum_{j=1}^n b_i b_j \psi''(\bar{U}_i) \psi''(\bar{U}_j) I(U_i) I(U_j) \sum_{k=1}^n z_i^k z_j^k \sum_{l=1}^n z_i^l z_j^l$$

$$= \sum_{i=1}^n \sum_{j=1}^n b_i b_j \psi''(\bar{U}_i) \psi''(\bar{U}_j) I(U_i) I(U_j) \langle z_i, z_j \rangle^2$$

$$\begin{aligned} & \sum_{i=1}^n b_i^2 \psi''(\bar{u}_i) I_{t,c}(u_i) \|z_i\|^4 + \sum_{j \neq i} |b_i| |b_j| |\psi''(\bar{u}_i)| |\psi''(\bar{u}_j)| \langle z_i, z_j \rangle^2 \\ & \leq K^2 \sum_{i=1}^n b_i^2 \|z_i\|^4 + K^2 \sum_{j \neq i} |b_i| |b_j| \langle z_i, z_j \rangle^2 \\ & \leq K^2 \max_{1 \leq j \leq n} \left\| U^{-1} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right\|^2 \sum_{i=1}^n b_i^2 + K^2 \sum_{i \neq j} |b_i| |b_j| \langle z_i, z_j \rangle^2 \end{aligned}$$

Luego, el primer sumando $\rightarrow 0$; y lo mismo ocurre con el segundo ya que

$$\begin{aligned} \sum_{i \neq j} |b_i| |b_j| \langle z_i, z_j \rangle^2 & \leq \sum_{i \neq j} |b_i| |b_j| \|z_i\|^2 \|z_j\|^2 < \left(\sum_{i=1}^n |b_i| \|z_i\|^2 \right)^2 \\ & \leq \left(\sum_{i=1}^n |b_i|^2 \right) \left(\sum_{i=1}^n \|z_i\|^4 \right) \leq \sum_{i=1}^n |b_i|^2 \sum_{i=1}^n \|z_i\|^2 \max_{1 \leq j \leq n} \left\| U^{-1} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right\|^2 \end{aligned}$$

que tiende a 0.

(usamos dos veces la desigualdad de Cauchy-Swartz, y el hecho que

$$\sum_{i=1}^n |b_i|^2 \rightarrow 1; \quad \sum_{i=1}^n \|z_i\|^2 = k)$$

Luego $\|W_4\|^2 \xrightarrow{P} 0$

Consideremos ahora el término correspondiente a la primera suma.

$$(W_2 - G_2)_1 = \sum_{j=1}^n (c_\psi(u_j, I) - \psi'(u_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) I_{[t-d, t+d]}(u_j)$$

Luego, $(\hat{\theta}^* - \theta_0)(W_2^* - G_2)_1 \leq$ en probabilidad que

$$\leq CM \sum_{j=1}^n \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right\| I_{[t-d, t+d]}(u_j) \quad \text{si } n \geq n_0$$

ya que $|c_\psi(t, I) - \psi'(t)|$ está acotado pues ψ' tiene derivadas laterales finitas y está acotada y $\|\hat{\theta}^* - \theta_0\| \leq M$ en probabilidad.

Ahora bien,

$$\begin{aligned} & E \left(\sum_{i=1}^n \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right\| I_{[t-d, t+d]}(u_i) \right) = \\ & = F([t-d, t+d]) \sum_{i=1}^n \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_i) \frac{\partial}{\partial \theta} g(\theta_0, X_i) U^{-1} \right\| \end{aligned}$$

$$\begin{aligned}
 &= F([t-d, t+d]) \cdot k \\
 \text{y } \text{var} \left(\sum_{j=1}^n \left\| U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} \right\| I_{[t-d, t+d]}(U_j) \right) &= \\
 \sum_{i=1}^n \left\| U^{-1t} \frac{\partial g(\theta_0, X_i)}{\partial \theta} \frac{\partial^t g(\theta_0, X_i)}{\partial \theta} U^{-1} \right\|^2 F([t-d, t+d]) (1 - F([t-d, t+d])) &= \\
 \max_{1 \leq j \leq n} \left\| U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2 \sum_{j=1}^n \left\| U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} \right\| &= \\
 = k \max_{1 \leq j \leq n} \left\| U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right\|^2 \rightarrow 0 &
 \end{aligned}$$

Luego, como $F([t-d, t+d]) \cdot k < \epsilon$, eligiendo d , listo.—

Consideremos ahora el caso en que tenemos r puntos de discontinuidad de ψ' .— Sean t_1, t_2, \dots, t_r tales puntos.

Elegimos d tal que los intervalos

$$[t_1-d, t_1+d]; \dots; [t_r-d, t_r+d] \text{ sean disjuntos, y formamos las}$$

sumas

$$\sum_{j=1}^n (c_{\psi}(U_j, I_g) - \psi'(U_j)) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} I_{[t_j-d, t_j+d]}(U_j)$$

para $j=1, \dots, r$

y la suma

$$\sum_{j=1}^n (c(U_j, I_g) - \psi'(U_j)) U^{-1t} \frac{\partial g(\theta_0, X_j)}{\partial \theta} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1} I_{\left(\bigcup_{j=1}^r [t_j-d, t_j+d]\right)}(U_j)$$

Para esta última suma es lo mismo que con

$$(\hat{\theta}^* - \theta_0)^t (W_2 - G_2)_2 \mathbf{a}$$

y para cada j $j=1, \dots, r$

$$\text{hacemos lo mismo que con } (\hat{\theta}^* - \theta_0)^t (W_2 - G_2)_1 \mathbf{a}$$

que como son un número finito, la suma también $\xrightarrow{P} 0$

Luego hemos probado que la suma I $\rightarrow 0$ en probabilidad.

Nos resta ver que la suma II también lo hace.—

La suma II es

$$\sum_{j=1}^n c_{\psi}(U_j, I_g) U^{-1t} \left(\frac{\partial g(\bar{\theta}_j, X_j)}{\partial \theta} - \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right) \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1}$$

Su norma es

$$\begin{aligned} &\leq c \sum_{j=1}^n \|U^{-1t} \left(\frac{\partial g(\bar{\theta}_j, X_j)}{\partial \theta} - \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right) \frac{\partial^t g(\theta_0, X_j)}{\partial \theta} U^{-1}\| = \\ &= c \sum_{j=1}^n \|U^{-1t} \left(\frac{\partial g(\bar{\theta}_j, X_j)}{\partial \theta} - \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right)\| \|U^{-1t} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta}\| \\ &\leq c \max_{1 \leq j \leq n} \|U^{-1t} \frac{\partial^t g(\theta_0, X_j)}{\partial \theta}\| \sum_{j=1}^n \|U^{-1t} \left(\frac{\partial g(\bar{\theta}_j, X_j)}{\partial \theta} - \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right)\| \end{aligned}$$

$$\bar{\theta}_j - \theta_0 = U^{-1}(t_j - \theta_0) \quad \text{con } \|t_j - \theta_0\| \leq M \quad \text{en probabilidad}$$

$$\sum_{j=1}^n \|U^{-1t} \left(\frac{\partial g(\bar{\theta}_j, X_j)}{\partial \theta} - \frac{\partial g(\theta_0, X_j)}{\partial \theta} \right)\| = \sum_{j=1}^n \|U^{-1t} H(\bar{\theta}_j, X_j) U^{-1}(t_j - \theta_0)\|$$

que en probabilidad es

$$\leq M c \sum_{j=1}^n \|U^{-1t} H(\bar{\theta}_j, X_j) U^{-1}\| \xrightarrow{P} 0$$

como ya vimos con V_n .-

Luego hemos probado la proposición 2.-

SECCION II

M-estimadores iterativos para regresión no lineal con efectos fijos y escala desconocida.-

Consideremos el modelo de regresión general, con efectos fijos, en que tenemos observaciones Y_1, Y_2, \dots, Y_n que verifican:

$$Y_i = g(\theta_0, X_i) + U_i \quad i=1, \dots, n \text{ donde}$$

$X_i \in R^q$ es un vector numérico

$\theta_0 \in R^k$ es el vector de parámetros que queremos estimar

U_i son variables aleatorias independientes con distribución común F desconocida, de las cuales ahora supondremos además que no se conoce su varianza.

Como ya observamos en la introducción, los M-estimadores considerados no son invariantes por cambios de escala, luego para que lo sean, consideraremos en este caso, la solución del sistema:

$$(II) \quad \sum_{j=1}^n \psi \left[\frac{Y_j - g(\theta, X_j)}{s_n} \right] \frac{\partial g(\theta, X_j)}{\partial \theta} = 0$$

donde s_n es un estimador robusto de escala, escala equivariante, o sea que $s_n(aY) = |a|s_n(Y)$ donde $Y=(Y_1, \dots, Y_n)$

como los considerados en la introducción.-

Para resolver la ecuación (II), como en la sección I, consideraremos M-estimadores iterativos, aplicando el algoritmo de Newton-Ralphson a la ecuación (II), a partir de un estimador inicial consistente y un estimador de escala consistente y robusto, siguiendo un camino análogo al del caso de escala conocida.-

En este caso, el algoritmo para calcular la solución de (II) será

$$\tilde{\theta}_{n,j+1} = \tilde{\theta}_{n,j} - (P'(\tilde{\theta}_{n,j}))^{-1} P(\tilde{\theta}_{n,j})$$

$$\text{donde } P(\theta) = s_n \sum_{j=1}^n \psi \left[\frac{Y_j - g(\theta, X_j)}{s_n} \right] \frac{\partial g(\theta, X_j)}{\partial \theta}$$

$$P'(\theta) = \sum_{j=1}^n s_n \psi \left[\frac{Y_j - g(\theta, X_j)}{s_n} \right] H(\theta, X_j) - \sum_{j=1}^n \psi' \left[\frac{Y_j - g(\theta, X_j)}{s_n} \right] B(\theta, X_j)$$

Definimos también

$$\tilde{\theta}_n = \begin{cases} \lim_{j \rightarrow \infty} \tilde{\theta}_{n,j} & \text{si existe} \\ \tilde{\theta}_{n,0} & \text{si no} \end{cases}$$

Donde $\tilde{\theta}_{n,0}$ es un estimador inicial dado.

HIPOTESIS

B0 - s_n es un estimador invariante de escala, $s_n \xrightarrow{P} s_0 > 0$

B1 = A1 de la sección I

B2 - F es continua y $\int \psi(x/s_0) dF(x) = 0$.

B3 - Existen d_0 y d_1 tales que

$$\begin{aligned} \sup_{|a| \leq d_0} \left| \frac{\psi^2(x-a)}{b} \right| & \text{ es } F\text{-integrable} \\ |b-s_0| & \leq d_1 \end{aligned}$$

$$\begin{aligned} \sup_{|a| \leq d_0} \left| \frac{\psi'^2(x-a)}{b} \right| & \text{ es } F\text{-integrable} \\ |b-s_0| & \leq d_1 \end{aligned}$$

B4 - $\tilde{\theta}_{n,0}$ es un estimador inicial tal que $U(\tilde{\theta}_{n,0} - \theta_0)$ está acotado en probabilidad.

B5 = A5

B6 = A6 y A7

B7 - Ψ es constante fuera de un compacto, y $\lim \Psi(t) \leq 0$ aunque no necesariamente la misma cte. a izq. $t \rightarrow -\infty$ que a derecha.

Definimos, análogamente al caso de escala conocida, el algoritmo de Newton-Ralphson para la ecuación (II) normalizada, así como la familia de contracciones asintóticas.

$$P^{\circ}(t) = s_n \sum_{j=1}^n \Psi \left(\frac{Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)}{s_n} \right) U^{-1t} \frac{\partial g(\theta_0 + U^{-1}(t - \theta_0), X_j)}{\partial \theta}$$

$$P^{\circ\circ}(t) = \Gamma_n^{\circ}(t) = \sum_{j=1}^n s_n \Psi \left[\frac{Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)}{s_n} \right] U^{-1t} H(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1} \\ - \sum_{j=1}^n \Psi' \left[\frac{Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)}{s_n} \right] U^{-1t} B(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1}$$

$$\eta_n^{\circ}(t) = t - (P^{\circ\circ}(t))^{-1} P^{\circ}(t)$$

$$\eta_{n\Gamma}^{\circ}(t) = t - \Gamma^{-1} P^{\circ}(t)$$

$$\Gamma_0^{\circ} = E_{F(s_0)}(\Psi'_{s_0}) I_k$$

donde $\Psi_a = a \Psi(. / a)$

Lema A

Supongamos que Ψ verifica la hipótesis B7 y que $s_n \xrightarrow{P} s_0$.
Entonces, para todo $a \in \mathbb{R}^k$, con $\|a\| = 1$ vale que

$$T) \sum_{j=1}^n (\Psi_{s_0}(U_j) - \Psi_{s_n}(U_j)) \frac{\partial g(\theta_0, X_j)}{\partial \theta} U^{-1} a \xrightarrow{P} 0$$

Demostración

Sea $0 \leq t \leq I$

Definimos

$$H_t(u) = \Psi_{s_0}(u) - \Psi_{s_0+t}(u) = s_0 \Psi(u/s_0) - (s_0+t) \Psi(u/s_0+t)$$

$$I_t(u) = \Psi_{s_0}(u) - \Psi_{s_0-t}(u)$$

Como $s_n \xrightarrow{P} s_0$, basta probar que

$$\lim_{n \rightarrow \infty} \limsup_{d \rightarrow 0} P\left(\sup_{0 \leq t \leq d} \left| \sum_{j=1}^n H_t(U_j) \frac{\partial g(\theta_0, X_j)}{\partial \theta} U^{-1} a \right| \geq \varepsilon\right) = 0$$

$$\lim_{n \rightarrow \infty} \limsup_{d \rightarrow 0} P\left(\sup_{0 \leq t \leq d} \left| \sum_{j=1}^n I_t(U_j) \frac{\partial g(\theta_0, X_j)}{\partial \theta} U^{-1} a \right| \geq \varepsilon\right) = 0$$

Sea entonces

$$J_n(t) = \sum_{j=1}^n H_t(U_j) \frac{\partial g(\theta_0, X_j)}{\partial \theta} U^{-1} a \quad \text{una sucesión de variables}$$

aleatorias en C , espacio de funciones continuas $[0,1]$.

De acuerdo con el teorema 12.3 (Billingsley (1)), para probar que

$J_n(t)$ es tight, es suficiente con probar que se verifican las

siguientes dos condiciones:

1) $J_n(0)$ lo es

2) Existe F continua y no decreciente, y $\gamma > 0$, $\alpha > 1$, tales que

$$E\left(|J_n(t_2) - J_n(t_1)|^\gamma\right) \leq (F(t_2) - F(t_1))^\alpha$$

$J_n(0) = 0$ luego es tight.

$$E((J_n(t_2) - J_n(t_1))^2) = E((H_{t_2}(U_1) - H_{t_1}(U_1))^2) \sum_{i=1}^n \left(\frac{\partial g(\theta_0, X_i)}{\partial \theta} U^{-1} a\right)^2$$

$$= E((H_{t_2}(U_1) - H_{t_1}(U_1))^2)$$

$$|H_{t_2}(u) - H_{t_1}(u)| = \left| (s_0 + t_1) \Psi(u/s_0 + t_1) - (s_0 + t_2) \Psi(u/s_0 + t_2) \right|$$

Sea $a > 0$, tal que $-a$ está fuera del compacto y $\Psi(-a) \leq 0$

$$\Psi(u) = \int_{-a}^u \Psi'(t) dt + \Psi(-a)$$

Luego, sea $t_2 > t_1$

$$\begin{aligned} |H_{t_2}(u) - H_{t_1}(u)| &= \left| (s_0 + t_1) \left[\int_{-a}^{u/s_0 + t_1} \psi'(t) dt + \psi(-a) \right] - \right. \\ &\quad \left. - (s_0 + t_2) \left[\int_{-a}^{u/s_0 + t_2} \psi'(t) dt + \psi(-a) \right] \right| = \\ &= \left| (t_1 - t_2) \psi(-a) + (s_0 + t_1) \int_{-a}^{u/s_0 + t_1} \psi'(t) dt - (s_0 + t_2) \int_{-a}^{u/s_0 + t_2} \psi'(t) dt \right| \end{aligned}$$

si hacemos el cambio $(s_0 + t_1)t = s$ y $(s_0 + t_2)t = s$ y acotamos

$$\begin{aligned} &\leq \left| \psi(-a)(t_2 - t_1) + \int_{-(s_0 + t_1)a}^u \psi'(s/s_0 + t_1) ds - \int_{-(s_0 + t_2)a}^u \psi'(s/s_0 + t_2) ds \right| \\ &\leq |\psi(-a)|(t_2 - t_1) + \int_{-(s_0 + t_2)a}^{-(s_0 + t_1)a} |\psi'(s/s_0 + t_2)| ds + \int_{-(s_0 + t_1)a}^u \left| \psi'(s/s_0 + t_1) - \psi'(s/s_0 + t_2) \right| ds \end{aligned}$$

$$\leq |\psi(-a)|(t_2 - t_1) + aK(-s_0 - t_1 + s_0 + t_2) + \int_{-(s_0 + t_1)a}^b |\psi'(s/s_0 + t_1) - \psi'(s/s_0 + t_2)| ds$$

donde b es tal que si $s \geq b$ $\psi'(s/s_0 + 1) = 0$ y además $b > a(s_0 + 1)$

(por tanto la diferencia en $|| = 0$)

luego

$$\leq (aK + |\psi(-a)|)(t_2 - t_1) + \int_{-(s_0 + t_1)a}^b \sup |\psi''(s)| \left| \frac{s}{s_0 + t_1} - \frac{s}{s_0 + t_2} \right| ds$$

$$= (aK + |\psi(-a)|)(t_2 - t_1) + K_1 \left(\frac{1}{s_0 + t_1} - \frac{1}{s_0 + t_2} \right) \int_{-(s_0 + t_1)a}^b s^2/2 ds$$

$$= K^*(t_2 - t_1) + K_1 \left(\frac{-1}{s_0 + t_2} - \left(\frac{-1}{s_0 + t_1} \right) \right) (b^2/2 - a^2(s_0 + 1)^2/2)$$

$$\leq K^*(t_2 - t_1) + K_1 \left(\frac{-1}{s_0 + t_2} - \left(\frac{-1}{s_0 + t_1} \right) \right) b^2/2 \quad (\text{pues } b > a(s_0 + t_1))$$

$$= K^* t_2 - b^2 K_1 / (s_0 + t_2)^2 - (K^* t_1 - b^2 K_1 / (s_0 + t_1)^2) =$$

$$= F(t_2) - F(t_1) \quad \text{con} \quad F(t) = K^* t - b^2 K_1 / 2(s_0 + t) \quad \text{continua y monótona.}$$

Luego,

$$E((H_{t_2}(U_1) - H_{t_1}(U_1))^2) \leq (F(t_2) - F(t_1))^2$$

y por tanto $J_n(t)$ es tight

Pero como $J_n(t)$ es tight y $J_n(0) = 0$

$$\lim_{n \rightarrow \infty} \limsup_{d \rightarrow 0} P(\sup_{0 \leq t \leq d} |J_n(t)| > \epsilon) = 0$$

De igual forma para la suma con $I_t(U_j)$.-

Lema 3 (b)

Bajo B0, B1, B2 y B3 (i) y B7

Sea $G =$ subconjunto de matrices definidas positivas cuyo mínimo autovalor es $> d > 0$.

Entonces

T) $\sup_{\Gamma \in G} \|\eta_{n\Gamma}^0(\theta_0) - \theta_0\|$ está acotado en probabilidad.-

Demostración

$$\|\eta_{n\Gamma}^0(\theta_0) - \theta_0\| = \Gamma^{-1} \sum_{i=1}^n \psi_{s_n}(U_i) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_i)$$

$$\leq \max \text{autovalor de } \Gamma^{-1} \left\| \sum_{i=1}^n \psi_{s_n}(U_i) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_i) \right\|$$

$$\leq 1/d \left(\left\| \sum_{i=1}^n \psi_{s_0}(U_i) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_i) \right\| + \left\| \sum_{i=1}^n (\psi_{s_n} - \psi_{s_0}) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_i) \right\| \right)$$

y el primer sumando está acotado en probabilidad (por el lema 3 de la sección I) y el segundo por el lema A $\xrightarrow{P} 0$.

Lema 4 (b)

Bajo B3, B5 y B6

vale que $\sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \Gamma_n^\circ(t) - E_F(\psi'_{s_0}) I_k \right\| \right\} \xrightarrow{P} 0$

Demostración

Haciendo la misma acotación que en el lema 4 (sección I) , basta probar que si llamamos

$$(I) = \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \left(\frac{\psi' \left(\frac{Y_j - g(\theta_0 + U^{-1}(t-\theta_0), X_j)}{s_n} \right)}{s_n} - E_F(\psi'_{s_0}) \right) U^{-1t} B(\theta_0, X_j) U^{-1} \right\| \right\}$$

$$(II) = \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \psi' \left(\frac{Y_j - g(\theta_0 + U^{-1}(t-\theta_0), X_j)}{s_n} \right) U^{-1t} H(\theta_0, X_j) U^{-1} \right\| \right\}$$

$$(III) = \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \psi' \left(\frac{Y_j - g(\theta_0 + U^{-1}(t-\theta_0), X_j)}{s_n} \right) U^{-1t} (B(\theta_0 + U^{-1}(t-\theta_0), X_j) - B(\theta_0, X_j)) U^{-1} \right\| \right\}$$

$$(IV) = \sup_{\|t-\theta_0\| \leq M} \left\{ \left\| \sum_{j=1}^n \frac{\psi' \left(\frac{Y_j - g(\theta_0 + U^{-1}(t-\theta_0), X_j)}{s_n} \right)}{s_n} U^{-1t} (H(\theta_0 + U^{-1}(t-\theta_0)) - H(\theta_0, X_j)) U^{-1} \right\| \right\}$$

$$(I) \xrightarrow{P} 0 ; (II) \xrightarrow{P} 0 ; (III) \xrightarrow{P} 0 ; (IV) \xrightarrow{P} 0$$

Para probar que (I) $\xrightarrow{P} 0$, basta con ver que

$$\lim_{a \rightarrow 0} \lim_{b \rightarrow 0} \limsup P \left(\sup_{\substack{|w_{i,j}| \leq b \\ |s - \frac{1}{2}c| \leq a}} \left| \sum_{h=1}^n (\psi'(U_h - w_h/s) - E_F(\psi'_{s_0})) V_{n,h} \right| > \epsilon \right) = 0$$

siendo $V_{n,h}$ el elemento i, j de la matriz $U^{-1t} B(\theta_0, X_h) U^{-1}$ definido en el lema 4 (sección I).

(usamos el hecho que $s_n \xrightarrow{P} s_0$ y que $|g(\theta_0, X_j) - g(\theta_0 + U^{-1}(t - \theta_0), X_j)| \rightarrow 0$ uniformemente en j para $\|t - \theta_0\| \leq M$)

Luego, si aplicamos el lema 1, con

$$f_{w,s}(u) = \psi' \left(\frac{u-w}{s} \right) - E_F(\psi' \Big|_{s_0})$$

que verifica las hipótesis (i), (ii), y (iii) del lema 1 (usamos B3) así como las hipótesis (iv) y (v) se verifican pues son los mismos coeficientes que en el lema correspondiente de la sección I.

Luego (I) $\xrightarrow{P} 0$.-

Para ver que (II) $\xrightarrow{P} 0$, de igual forma basta ver que

$$\lim_{a \rightarrow 0} \lim_{b \rightarrow 0} \limsup_{n \rightarrow \infty} P \left(\sup_{\substack{|w_i| \leq b \\ |s - s_0| \leq a}} \left| \sum_{h=1}^n s \psi \left(\frac{u - w_i}{s} \right) W_{n,h} \right| > \varepsilon \right) = 0$$

donde $W_{n,h}$ es el elemento i,j de la matriz $U^{-1t} H(\theta_0, X_h) U^{-1}$

Luego, si tomamos $f_{w,s}(u) = s \psi((u-w)/s)$ y aplicamos el lema 1 , obtenemos el resultado querido.

Finalmente para (III) $\xrightarrow{P} 0$ y (IV) $\xrightarrow{P} 0$, como en este caso también ψ' y $s_n \psi$ están acotadas en probabilidad, se reduce al lema

4 de la sección I.--

Lema 5 (b)

Bajo B0 a B3, vale el lema 5 con $\eta_{n \cap}^0$ en lugar de $\eta_{n \cap}$, o sea para cada muestra Y_1, \dots, Y_n , existen abiertos D_j para $j \geq 1$

$$\text{tal que } \bigcup_{j \geq 1} \bar{D}_j = R^k \text{ y tal que}$$

$\eta_{n \cap}^0(t)$ es continuamente diferenciable en cada D_j $1 \leq j$ $\forall n$ def > 0

Más aun, en cada D_j

$$(b_1) \frac{\partial}{\partial t} \gamma_{n\Gamma}^{\circ}(t) = I_k^{-1} \Gamma^{-1} \left[\sum_{j=1}^n \psi' \left(\frac{Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)}{s_n} \right) U^{-1t} B(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1} \right. \\ \left. - \sum_{j=1}^n s_n \psi \left(\frac{Y_j - g(\theta_0 + U^{-1}(t - \theta_0), X_j)}{s_n} \right) U^{-1t} H(\theta_0 + U^{-1}(t - \theta_0), X_j) U^{-1} \right]$$

(b₂) Existe $\epsilon_1(k_0)$ tal que

$$\lim_{n \rightarrow \infty} P(\sup_{(t, \Gamma) \in Q} \left\| \frac{\partial}{\partial t} \gamma_{n\Gamma}^{\circ}(t) \right\| > k_0) = 0$$

Demostración

Para cada muestra Y_1, \dots, Y_n $\frac{\partial}{\partial t} \gamma_{n\Gamma}^{\circ}(t)$ está definida y es = (b₁) para toda Γ definida positiva, salvo en los conjuntos

$$R_{ji} = \left\{ t / \frac{Y_j - g(t, X_j)}{s_n} = a_i \right\} = \left\{ t / g(t, X_j) = s_n a_i + Y_j \right\}$$

= $\left\{ t / g(t, X_j) = r_{ij} \right\}$.- Donde a_i son los puntos de discontinuidad

de ψ° . - Y ahora estamos en las condiciones del lema 5 de la sección

I.-

(b₂) es igual a la demostración del lema 5 (a₂) usando en este

caso el lema 4 (b).-

Lema 6

Bajo B0 a B3 para todo $M > 0$

(i) Sea $A_{n,M}^{\circ} = \left\{ \left\| \gamma_{n\Gamma}^{\circ}(t) - \gamma_{n\Gamma}^{\circ}(t') \right\| \leq k_0 \|t - t'\|, \|t - \theta_0\| \leq M, \|t' - \theta_0\| \leq M, \|\Gamma - \Gamma_0\| \leq \epsilon_0 \right\}$

ϵ_0 del lema 4 (b) Entonces $\lim_{n \rightarrow \infty} P(A_{n,M}^{\circ}) = 1$

(ii) Para todo $M > 0$, sea $0 < \mu < |E_{\Gamma}(\psi'_{s_0})|$

$B_{n,M}^{\circ} = \left\{ \gamma_{n\Gamma}^{\circ}(t) \text{ lleva } \|t - \theta_0\| \leq M \text{ en si mismo } \forall \Gamma \text{ tq. } \|\Gamma - \Gamma_0\| \leq \mu \right\}$

Dado $d > 0$, existe $M_0(d)$ y μ_0 tal que

$$\lim_{n \rightarrow \infty} P(B_{n, M, \mu}^{\circ}) > 1-d \text{ para todo } M \geq M_0(d)$$

(iii) Para todo $M > 0$, $0 < \mu < |E_F(\psi'_{s_0})|$

Sea $C_{n, M}^{\circ} = \{ \Gamma_n^{\circ}(t) \text{ lleva } \|t - \theta_0\| \leq M \text{ en } \|\Gamma - \Gamma_0^{\circ}\| \leq \mu \}$

Entonces $\lim_{n \rightarrow \infty} P(C_{n, M, \mu}^{\circ}) = 1$

Demostración

Igual al lema 6 de la sección I, pero usando los lemas 5, 3 y 4 (b).

Lema 7 (b)

Sea $D_{n, M}^{\circ} = \left\{ \eta_n^{\circ}(t) = t \text{ tiene un único punto fijo } t^* \text{ en } \|t - \theta_0\| \leq M \text{ y} \right.$
 $\left. \text{para cualquier } t_0 \text{ tal que } \|t_0 - \theta_0\| \leq M \quad \eta_n^{\circ}(t_0) \rightarrow t^* \right\}$

Para todo $d > 0$, existe $M_0(d)$ tal que $M \geq M_0(d)$

$$\lim_{n \rightarrow \infty} P(D_{n, M}^{\circ}) > 1-d$$

Demostración

Es la misma que la del lema 7 de la sección I, pero con los conjuntos
 •.—

Proposición 3

Bajo B0 a B7

(i) $\lim_n P(\tilde{\theta}_n = \lim_{j \rightarrow \infty} \tilde{\theta}_{n, j} \text{ y } \tilde{\theta}_n \text{ satisface la ec. II}) = 1$

(ii) $U(\tilde{\theta}_n - \theta_0)$ está acotado en probabilidad.—

Demostración

Igual a la proposición 1, pero con los lemas (b).---

SECCION II B

Normalidad Asintótica.-

Sea $\hat{\theta}_n^v$ la solución que consideramos de la ecuación

$$\sum_{j=1}^n \Psi_{s_0} (Y_j - g(\theta_0 + U^{-1}(\hat{\theta}_n^v - \theta_0)), X_j) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\hat{\theta}_n^v - \theta_0), X_j) = 0$$

donde $\Psi_{s_0} = \Psi(\cdot/s_0)$

Luego, por la proposición 2, (sección I), tenemos que

$$\forall a, \|a\|=1 \quad a^t (\hat{\theta}_n^v - \theta_0) \xrightarrow{w} N(0, E_F(\Psi_{s_0}^2) / E_F^2(\Psi_{s_0}')) \quad .-$$

$\tilde{\theta}_n^*$, será entonces, la solución correspondiente a la ecuación

$$\sum_{j=1}^n \Psi_{s_n} (Y_j - g(\theta_0 + U^{-1}(\tilde{\theta}_n^* - \theta_0)), X_j) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\tilde{\theta}_n^* - \theta_0), X_j) = 0$$

De la proposición 2 tenemos que si:

$$W_1(\Psi_{s_0}) = \sum_{j=1}^n \Psi_{s_0}(U_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a$$

$$W_2^*(\Psi_{s_0}) = \sum_{j=1}^n c_{\Psi_{s_0}}(U_j, I_g) U^{-1t} \frac{\partial}{\partial \theta} g(\bar{\theta}_j, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1}$$

$$V_n = \sum_{j=1}^n \Psi_{s_0}(U_j - (g(\theta_0 + U^{-1}(\tilde{\theta}_n^* - \theta_0), X_j) - g(\theta_0, X_j))) U^{-1t} \left[\frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\tilde{\theta}_n^* - \theta_0), X_j) - \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right]$$

Se verificaba que

$$W_1(\Psi_{s_0}) = (\hat{\theta}_n^* - \theta_0)^t W_2(\Psi_{s_0}) a + V_n$$

y entonces tendremos que vale también que

$$W_1(\Psi_{s_n}) = (\tilde{\theta}_n^* - \theta_0)^t W_2(\Psi_{s_n}) a + W_n$$

donde W_n se obtiene de V_n sustituyendo Ψ_{s_0} por Ψ_{s_n} y lo mismo con

$$W_1(\Psi_{s_n}) \text{ y } W_2(\Psi_{s_n})$$

Proposición 4

Para todo a con $\|a\| = 1$, si $s_n \xrightarrow{P} s_0$, tenemos que

$$T) a^t (\tilde{\theta}_n^* - \theta_0) \rightarrow N(0, E_F(\psi_{s_0}^2) / E_F(\psi_{s_0}'))$$

Demostración

$$a) W_1(\psi_{s_n}) \rightarrow N(0, E_F(\psi_{s_0}^2))$$

Para ello basta ver que

$$W_1(\psi_{s_n}) - W_1(\psi_{s_0}) \xrightarrow{P} 0$$

$$\begin{aligned} W_1(\psi_{s_n}) - W_1(\psi_{s_0}) &= \sum_{i=1}^n (\psi_{s_0}(U_i) - \psi_{s_n}(U_i)) \frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\tilde{\theta}_n^* - \theta_0), X_j) U^{-1} a \\ &= \sum_{i=1}^n (\psi_{s_0}(U_i) - \psi_{s_n}(U_i)) \frac{\partial}{\partial \theta} g(\theta_0, X_i) U^{-1} a + \\ &+ \sum_{i=1}^n (\psi_{s_0}(U_i) - \psi_{s_n}(U_i)) \left(\frac{\partial}{\partial \theta} g(\theta_0 + U^{-1}(\tilde{\theta}_n^* - \theta_0), X_i) - \frac{\partial}{\partial \theta} g(\theta_0, X_i) \right) U^{-1} a \end{aligned}$$

La segunda suma tiende en probabilidad a 0, pues ya vimos cuando estudiamos normalidad asintótica con escala conocida que

$$\sum_{j=1}^n \|U^{-1} \left(\frac{\partial}{\partial \theta} g(\theta_j, X_j) - \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right)\| \xrightarrow{P} 0$$

y $[\psi_{s_0} - \psi_{s_n}]$ está acotado.

P

Por otro lado, que la primera suma $\xrightarrow{P} 0$ no es otra cosa que el lema A.-

b) $W_n \xrightarrow{P} 0$ pues cuando estudiamos el caso de V_n , con ψ_{s_0} para ver que $V_n \rightarrow 0$ usamos solamente que ψ_{s_0} estaba acotada, lo que se verifica también en este caso.-

$$c) \quad (\tilde{\theta}_n^* - \theta_0)^t W_2^* (\psi_{s_n}) a \xrightarrow{P} E_F(\psi'_{s_0}) (\tilde{\theta}_n^* - \theta_0)^t a$$

con lo que obtendremos la proposición 4 usando el teorema de Slutsky

$$\text{pues} \quad (\tilde{\theta}_n^* - \theta_0)^t W_2^* (\psi_{s_n}) a = W_1 + W_n \rightarrow N(0, E_F(\psi_{s_0}^2))$$

$$\begin{aligned} (\tilde{\theta}_n^* - \theta_0)^t W_2^* (\psi_{s_n}) a &= (\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n c_{\psi_{s_n}} (U_j, I_g) U^{-1t} \frac{\partial}{\partial \theta} g(\bar{\theta}_j, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \\ &= (\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n (c_{\psi_{s_n}} (U_j, I_g) - \psi'_{s_n}(U_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \\ &+ (\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n c_{\psi_{s_n}} (U_j, I_g) U^{-1t} \left(\frac{\partial}{\partial \theta} g(\bar{\theta}_j, X_j) - \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \\ &+ (\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n \psi'_{s_n}(U_j) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \end{aligned}$$

La segunda suma tiende en probabilidad a 0, pues nuevamente en la demostración correspondiente a escala conocida, usamos solo que c_{s_0}

estaba acotada. - (Que llamábamos suma II)

La tercera suma es $(\tilde{\theta}_n^* - \theta_0)^t G_2(\psi_{s_n}) a$ que veremos que tiende en probabilidad a $E_F(\psi'_{s_0}) (\tilde{\theta}_n^* - \theta_0)^t a$

Y veremos que la primera suma tiende en probabilidad a 0. -

$$c_1) \quad (\tilde{\theta}_n^* - \theta_0)^t G_2(\psi_{s_n}) a \xrightarrow{P} E_F(\psi'_{s_0}) (\tilde{\theta}_n^* - \theta_0)^t a$$

$$\begin{aligned} (\tilde{\theta}_n^* - \theta_0)^t G_2(\psi_{s_n}) a &= (\tilde{\theta}_n^* - \theta_0)^t E_F(\psi'_{s_0}) \left(\sum_{j=1}^n U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right) a \\ &+ (\tilde{\theta}_n^* - \theta_0)^t \left(\sum_{j=1}^n (\psi'(U_j/s_n) - E_F(\psi'_{s_0})) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right) a \end{aligned}$$

El primer sumando = $(\tilde{\theta}_n^* - \theta_0)^t E_F(\psi'_{s_0}) a$

Al segundo lo descomponemos en las siguientes dos sumas.

$$(\tilde{\theta}_n^* - \theta_0)^t \left[\sum_{j=1}^n (\psi'(U_j/s_n) - \psi'(U_j/s_0)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right] a$$

$$+ (\tilde{\theta}_n^* - \theta_0)^t \left[\sum_{j=1}^n (\psi'(U_j/s_0) - E_F(\psi'_{s_0})) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right] a$$

La segunda suma $\xrightarrow{P} 0$, pues $(\tilde{\theta}_n^* - \theta_0)$ está acotado en probabilidad,

$$y \sum_{j=1}^n (\psi'(U_j/s_0) - E_F(\psi'_{s_0})) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \xrightarrow{P} 0$$

por ser la suma correspondiente para escala conocida con $\psi = \psi_{s_0}$

Como $(\tilde{\theta}_n^* - \theta_0)$ está acotado en probabilidad, para la primer suma, basta-

rá ver que para todo $a \in \mathbb{R}^k$, con $\|a\| = 1$ se verifica que:

$$a^t \sum_{j=1}^n (\psi'(U_j/s_n) - \psi'(U_j/s_0)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \xrightarrow{P} 0$$

Para ello usaremos nuevamente un argumento de tightness en C.-

Sea $0 \leq t \leq 1$ Definimos

$$H_t(u) = \psi'(u/s_0) - \psi'(u/(s_0+t))$$

$$I_t(u) = \psi'(u/s_0) - \psi'(u/(s_0-t))$$

Como $s_n \xrightarrow{P} s_0$ basta probar que

$$\lim_{n \rightarrow \infty} \limsup_{d \rightarrow 0} P(\sup_{0 \leq t \leq d} \left| a^t \sum_{j=1}^n H_t(U_j) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \right| \geq \varepsilon) = 0$$

$$\lim_{n \rightarrow \infty} \limsup_{d \rightarrow 0} P(\sup_{0 \leq t \leq d} \left| a^t \sum_{j=1}^n I_t(U_j) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \right| \geq \varepsilon) = 0$$

Si definimos entonces

$$J_n(t) = \sum_{j=1}^n H_t(U_j) a^t U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a$$

como $J_n(0) = 0$, bastará con probar como en el lema A que la sucesión

$J_n(t)$ de variables aleatorias en C es tight.

Luego, por el teorema 12.3 (Billingsley) nos bastará probar que

$$E((J_n(t_2) - J_n(t_1))^2) \leq (F(t_2) - F(t_1))^2 \quad \text{con } F \text{ continua y monótona.}$$

(pues como $J_n(0) = 0$ es tight)

$$\begin{aligned} E((J_n(t_2) - J_n(t_1))^2) &= E((H_{t_2}(U_1) - H_{t_1}(U_1))^2) \sum_{j=1}^n (a^t U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1a})^2 \\ &= k E((H_{t_2}(U_1) - H_{t_1}(U_1))^2) \end{aligned}$$

Sea b tal que si $u \geq b$ $\psi'(u/s_0 + 1) = 0$ y si $u \leq -b$ ocurra lo mismo.

Luego $\psi'(u/s_0 + t) = 0$ para todo $t \in [0, 1]$ si $|u| \geq b$. - (Existe por la hipótesis sobre ψ . - (B7). - Sean $0 \leq t_1 \leq t_2 \leq 1$

Luego,

$$\begin{aligned} |H_{t_2}(u) - H_{t_1}(u)| &= |\psi'(u/s_0 + t_1) - \psi'(u/s_0 + t_2)| = \\ &= \left| \int_{u/(s_0 + t_2)}^{u/(s_0 + t_1)} \psi''(t) dt \right| \leq \int_{u/(s_0 + t_2)}^{u/(s_0 + t_1)} |\psi''(t)| dt \leq \\ &\leq \int_{u/(s_0 + t_2)}^{u/(s_0 + t_1)} \sup |\psi''| dt \leq K \left[\frac{u}{s_0 + t_1} - \frac{u}{s_0 + t_2} \right] \end{aligned}$$

Tenemos entonces que si $|u| \geq b$ $|H_{t_2}(u) - H_{t_1}(u)| = 0$

y si $|u| < b$ es $\leq K|b| \left(\frac{1}{s_0 + t_1} - \frac{1}{s_0 + t_2} \right) = F(t_2) - F(t_1)$

con $F(t) = \frac{-K|b|}{s_0 + t}$ continua y monótona; y la acotación vale $\forall u$.

Por tanto, $E((H_{t_2}(U_1) - H_{t_1}(U_1))^2) \leq (F(t_2) - F(t_1))^2$ como queríamos,

y $J_n(t)$ es tight con $J_n(0) = 0$.

De igual modo para la suma con $I_t(U_j)$.-

Finalmente

(c₂)

$$(\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n (c_{\psi_{s_n}}(U_j, I_g) - \psi'_{s_n}(U_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a \xrightarrow{P} 0$$

$|I_g| \leq d$ si $n \geq n_0$ en probabilidad.

Sea t un punto de discontinuidad de ψ' .

(Como antes consideraremos un solo punto y luego para finitos)

Como $\lim_{n \rightarrow \infty} P(|s_n - s_0| > d_2) = 0 \quad \forall d_2 > 0$

nuedo tomar d_1 tal que $|I_g/s_n| \leq d_1$ si $n \geq n_1$ en probabilidad.

Y separamos en las siguientes dos sumas:

$$\begin{aligned} & (\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n (c_{\psi_{s_n}}(U_j, I_g) - \psi'_{s_n}(U_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a I_{[s_n(t-d_1), s_n(t+d_1)]}(U_j) \\ & + \\ & (\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n (c_{\psi_{s_n}}(U_j, I_g) - \psi'_{s_n}(U_j)) U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} a I_{[s_n(t-d_1), s_n(t+d_1)]}^c(U_j) \end{aligned}$$

Como $|c_{\psi_{s_n}} - \psi'_{s_n}|$ está acotado en probabilidad

la primera suma es

$$\leq M C \sum_{j=1}^n \| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \| I_{[s_n(t-d_1), s_n(t+d_1)]}(U_j)$$

en probabilidad es

$$\leq M C \sum_{j=1}^n \| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \| I_{[(s_0-1)(t-d_1), (s_0+1)(t+d_1)]}(U_j)$$

que tiende en probabilidad a 0, pues

$$E\left(\sum_{j=1}^n \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right\| I_{[(s_0-1)(t-d_1), (s_0+1)(t+d_1)]}^{(U_j)} \right) =$$

$$= k F([(s_0-1)(t-d_1), (s_0+1)(t+d_1)]) \leq \epsilon \quad \text{eligiendo } d_1; \quad \gamma$$

$$\text{var}\left(\sum_{j=1}^n \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right\| I_{[(s_0-1)(t-d_1), (s_0+1)(t+d_1)]}^{(U_j)} \right) =$$

$$= \sum_{j=1}^n \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right\|^2 F(A) (1 - F(A))$$

$$\max_{1 \leq j \leq n} \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \right\|^2 \sum_{j=1}^n \left\| U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} \right\| \xrightarrow{P} 0$$

$$\text{Donde } A = [(s_0-1)(t-d_1), (s_0+1)(t+d_1)]$$

La segunda suma =

$$= \frac{1}{2} (\tilde{\theta}_n^* - \theta_0)^t \sum_{j=1}^n \psi''(U_j - I_g) I_g U^{-1t} \frac{\partial}{\partial \theta} g(\theta_0, X_j) \frac{\partial}{\partial \theta} g(\theta_0, X_j) U^{-1} I_{A_n^c}^{(U_j)}$$

donde A_n es el intervalo $[s_n(t-d_1), s_n(t+d_1)]$

Y esta suma $\xrightarrow{P} 0$, pues en este caso vale la misma demostración hecha para el caso de escala conocida, pues usabamos solo el hecho que ψ'' estuviera acotada asi como el indicador.

Por tanto, nos resta ver el caso en que tenemos r puntos de discontinuidad de ψ' t_1, \dots, t_r que se resuelve de igual modo que con escala conocida tomando entornos disjuntos (en probabilidad) de los puntos de discontinuidad.-

SECCION III

GM-estimadores iterativos para regresión no lineal con efectos aleatorios.

Consideremos ahora el modelo de regresión general, con efectos aleatorios, en que tenemos observaciones Y_1, \dots, Y_n que verifican:

$$Y_i = g(\beta_0, X_i) + U_i \quad i=1, \dots, n \quad \text{donde}$$

X_i son vectores aleatorios en R^q $\beta_0 \in R^k$ vector de parámetros

U_i son variables aleatorias independientes e igualmente distribuidas

$(Y_i, X_i) \in R^{q+1}$ son vectores aleatorios independientes, con distribución común F.-

Como ya observamos en la introducción los M-estimadores usuales resultan no robustos (curva de influencia no acotada) cuando tenemos contaminación en las X_i , si $\frac{\partial g(X, \beta)}{\partial \beta}$ no es acotada.-

Por ello consideraremos una clase mas amplia, la de los GM-estimadores definida como solución de la ecuación:

$$(III) \quad \sum_{i=1}^n \psi(Y_i, X_i, \hat{\beta}_n) = 0 \quad \text{para} \quad \psi: R \times R^q \times R^k \rightarrow R^k$$

Si ψ es acotada, resultan estimadores robustos.- El estimador de mínimos cuadrados $\hat{\beta}_n^0$ pertenece a la familia considerada pues está definido para $\psi(Y, X, \beta) = (Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta}$.- Asimismo, los M-estimadores usuales también están en la familia para $\psi(Y, X, \beta) = - \psi(Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta}$.-

En el caso que la ecuación $E_F(\psi(Y, X, \beta)) = 0$ no tenga una única solución, para definir el estimador $\hat{\beta}_n$ debemos especificar que solución de (III) estamos considerando. Esto puede ser hecho indicando el algoritmo usado para calcularlo.- Vamos a probar consistencia y normalidad asintótica para GM-estimadores iterativos.-

Definimos entonces

$$P(\beta) = \sum_{i=1}^n \varphi(Y_i, X_i, \beta)$$

y el algoritmo de Newton-Raphson para este caso será:

$$\hat{\beta}_{n,j+1} = \hat{\beta}_{n,j} - \left(\sum_{i=1}^n \frac{\partial \varphi(Y_i, X_i, \hat{\beta}_{n,j})}{\partial \beta} \right)^{-1} \sum_{i=1}^n \varphi(Y_i, X_i, \hat{\beta}_{n,j})$$

y definimos

$$\hat{\beta}_n = \begin{cases} \lim_{j \rightarrow \infty} \hat{\beta}_{n,j} & \text{si el límite existe} \\ \hat{\beta}_{n,0} & \text{sino} \end{cases}$$

donde $\hat{\beta}_{n,0}$ es un estimador inicial consistente del cual se parte en el procedimiento iterativo.-

HIPOTESIS

A1) Para cada β fijo, φ es continua y $\frac{\partial \varphi}{\partial \beta}(Y, X, \beta)$ existe y es continua salvo en un número finito de puntos.-

A2) F continua, $E_F(\varphi(Y, X, \beta)) = 0$ y $E\left(\frac{\partial \varphi}{\partial \beta}(Y, X, \beta)\right)$ no singular.

A3) Existe $d_0 > 0$ tal que

$$\sup_{\|a - \beta_0\| \leq d_0} \|\varphi(Y, X, a)\| \text{ es } F\text{-integrable}$$

$$\sup_{\|a - \beta_0\| \leq d_0} \left\| \frac{\partial \varphi}{\partial \beta}(Y, X, a) \right\| \text{ es } F\text{-integrable (y de cuadrado}$$

integrable para normalidad asintótica)

A4) $\hat{\beta}_{n,0}$ es un estimador inicial, consistente p.p. F

Definimos

$$\eta_n(t) = t - \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta} \varphi(Y_i, X_i, t) \right)^{-1} \sum_{i=1}^n \varphi(Y_i, X_i, t) / n$$

y para $\Gamma \in R^k \times R^k$ definida positiva

$$\eta_{n\Gamma}(t) = t - \Gamma^{-1} \left(\sum_{i=1}^n \varphi(Y_i, X_i, t) / n \right)$$

$$\Gamma_n(t) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial t} \varphi(Y_i, X_i, t)$$

Usaremos los lemas 1 y 2 de la sección I.-

Lema 3

Sea G = subconjunto de matrices definidas positivas cuyo mínimo autovalor es $> d > 0$

Bajo A2 vale que

$$T) \sup_{\Gamma \in G} \|\eta_{n\Gamma}(\beta_0) - \beta_0\| \rightarrow 0 \quad \text{p.p. F}$$

Demostración

$$\begin{aligned} \sup_{\Gamma \in G} \|\eta_{n\Gamma}(\beta_0) - \beta_0\| &\leq \max. \text{autovalor de } \Gamma^{-1} \left\| \sum_{i=1}^n \varphi(Y_i, X_i, \beta_0) / n \right\| \\ &\leq \frac{1}{d} \left\| \sum_{i=1}^n \varphi(Y_i, X_i, \beta_0) / n \right\| \xrightarrow{\text{p.p.}} 0 \quad \text{por la ley de los grandes números y la hipótesis A2.} \end{aligned}$$

Lema 4

$$\text{Sea } C_n = \left\{ \sup_{\|t - \beta_0\| \leq d(\epsilon)} \left\| \Gamma_n(t) - E\left(\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0)\right) \right\| > \epsilon \right\}$$

entonces

$$T) P(\limsup C_n) = 0$$

Demostración

Basta verlo para cada elemento i, j .- $(\varphi^1, \varphi^2, \dots, \varphi^k) = \varphi$

$$(\Gamma_n(t))_{ij} = \frac{1}{n} \sum_{h=1}^n \frac{\partial}{\partial \beta_j} \varphi^i(Y_h, X_h, t)$$

$$(E(\frac{\partial}{\partial \beta} (Y, X, \beta_0)))_{ij} = E(\frac{\partial}{\partial \beta_j} \varphi^i(Y, X, \beta_0))$$

Usaremos el lema 2 (en R^{q+1}) (Yohai 1974)

Sean $Z_1 = (Y_1, X_1), \dots, Z_n = (Y_n, X_n)$ vectores aleatorios independientes igualmente distribuidos en R^{q+1}

C compacto, y $\{f_k\}_{k \in C}$ $f_k : R^{q+1} \rightarrow R$ medibles $C \subset R^p$

que verifican:

(i) $|f_k| \leq f$ con $E(f(Z)) < \infty$

(ii) $\lim_{t \rightarrow k} f_t(Z) = f_k(z)$ p.p. para todo $k \in C$

(iii) $|E(f_k(Z))| \leq A$ para todo $k \in C$

Entonces

$$T) \limsup_{n \rightarrow \infty} \sup_{k \in C} \left| \frac{\sum_{i=1}^n f_k(Z_i)}{n} \right| \leq A \text{ p.p.}$$

Sea entonces $f_t(Z) = \frac{\partial}{\partial \beta_j} \varphi^i_t(Y, X) - E_F(\frac{\partial}{\partial \beta_j} \varphi^i(Y, X, \beta_0))$

$$C = \{ \|t - \beta_0\| \leq d(\varepsilon) \}$$

Verificaremos que estamos en las condiciones del lema 2.

$$1) |f_t(Z)| \leq \sup_{\|t - \beta_0\| \leq d_0} \left| \frac{\partial}{\partial \beta_j} \varphi^i(Y, X, t) \right| - E_F(\frac{\partial}{\partial \beta_j} \varphi^i(Y, X, \beta_0))$$

que es F -integrable por la hipótesis A3.

$$2) \text{ basta ver que } \lim_{t \rightarrow k} \frac{\partial}{\partial \beta_j} \varphi^i(Y, X, t) = \frac{\partial}{\partial \beta_j} \varphi^i(Y, X, k) \text{ p.p. } F$$

que ocurre si y solo si

$$P(\{ \omega \in \Omega / \frac{\partial}{\partial \beta_j} \varphi^i(Y(\omega), X(\omega), k) \text{ no es continua} \}) = 0 \text{ si } k \in C, \text{ que}$$

se verifica, pues F es continua y $\frac{\partial}{\partial \beta} \varphi(Y, X, k)$ tiene un número finito de discontinuidades para cada k fijo por la hipótesis A1.-

$$3) \left| E\left(\frac{\partial}{\partial \beta_j} \varphi^i(Y, X, k)\right) - E\left(\frac{\partial}{\partial \beta_j} \varphi^i(Y, X, \beta_0)\right) \right| < \varepsilon \quad \text{si } \|k - \beta_0\| \leq d(\varepsilon) = d_0$$

$$\text{pues } \frac{\partial}{\partial \beta_j} \varphi^i(Y, X, k) \xrightarrow[k \rightarrow \beta_0]{} \frac{\partial}{\partial \beta_j} \varphi^i(Y, X, \beta_0) \quad \text{p.p. } F$$

y convergencia dominada ($d_0 < d$ de la hipótesis A3)

Luego, por el lema 2

$$\lim_{n \rightarrow \infty} \sup_{\|t - \beta_0\| \leq d(\varepsilon)} \left| \frac{1}{n} \sum_{h=1}^n \frac{\partial}{\partial \beta_j} \varphi^i(Y_h, X_h, t) - E_F\left(\frac{\partial}{\partial \beta_j} \varphi^i(Y, X, \beta_0)\right) \right| < \varepsilon \quad \text{p.p. } F$$

por tanto

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{m > n} \left\{ \sup_{\|t - \beta_0\| \leq d(\varepsilon)} \left| \frac{1}{m} \sum_{h=1}^m \frac{\partial}{\partial \beta_j} \varphi^i(Y_h, X_h, t) - E_F\left(\frac{\partial}{\partial \beta_j} \varphi^i(Y, X, \beta_0)\right) \right| \geq \varepsilon \right\}\right) = 0$$

Lema 5

Para cada muestra (Y_i, X_i) $\eta_{n\Gamma}(t)$ es continuamente diferenciable

salvo en un número finito (aleatorio) de puntos D_1, \dots, D_r y

(a) en $R^k - \bigcup_j D_j$ vale que

$$\frac{\partial}{\partial t} \eta_{n\Gamma}(t) = I - \Gamma^{-1} \sum_{i=1}^n \frac{\partial}{\partial t} \varphi(Y_i, X_i, t) / n$$

(b) Para cada $k_0 > 0$, existe $\varepsilon_1(k_0)$ y d_1 tal que

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{(t, \Gamma) \in Q} \sup_{(t, \Gamma) \in Q} \left\| \frac{\partial}{\partial t} \eta_{n\Gamma}(t) \right\| > k_0\right) = 0$$

donde $Q = \left\{ (t, \Gamma) / \|t - \beta_0\| \leq d; t \in (R^k - \bigcup_{j=1}^r D_j); \|\Gamma - \Gamma_0\| < \varepsilon_1 \right\}$

$$y \quad \Gamma_0 = E_F \left(\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0) \right)$$

(Luego, si $k_0 < 1$, $\gamma_n \Gamma$ será una contracción p.p. uniforme entoda la familia)

Demcstración

Dados $(Y_1, X_1), \dots, (Y_n, X_n)$, para esta muestra

Sean t_1, \dots, t_p los puntos de R^k donde φ no es diferenciable. Para

cada muestra $(Y_1, X_1), \dots, (Y_n, X_n)$ $\frac{\partial}{\partial t} \gamma_n \Gamma(t)$ esta definida y es = (a)

para toda Γ definida positiva salvo en los puntos t_1, \dots, t_p .

(b)

$$\begin{aligned} \text{En } \left(\bigcup_{j=1}^p D_j \right)^c \quad \frac{\partial}{\partial t} \gamma_n \Gamma(t) &= I - \Gamma^{-1} \sum_{i=1}^n \frac{\partial}{\partial t} \varphi(Y_i, X_i, t) / n = \\ &= I - \Gamma^{-1} \sum_{i=1}^n \frac{\partial}{\partial t} \varphi(Y_i, X_i, t) / n - I_k + \Gamma_0^{-1} E_F \left(\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0) \right) \end{aligned}$$

Luego,

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left(\bigcap_{m \geq n} \sup_{(t, \Gamma) \in Q} \left\{ \left\| \frac{\partial}{\partial t} \gamma_n \Gamma(t) \right\| < k_0 \right\} \right) &= \\ = \lim_{n \rightarrow \infty} P \left(\bigcap_{m \geq n} \sup_{(t, \Gamma) \in Q} \left\{ \left\| \Gamma^{-1} \Gamma_m(t) - \Gamma_0^{-1} E_F \left(\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0) \right) \right\| < k_0 \right\} \right) &= \\ \lim_{n \rightarrow \infty} P \left(\bigcap_{m \geq n} \sup_Q \left\{ \left\| (\Gamma_n(t) - \Gamma_0) \Gamma^{-1} \right\| + \left\| \Gamma^{-1} - \Gamma_0^{-1} \right\| \left\| E_F \left(\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0) \right) \right\| \right\} < k_0 \right) &= \\ \lim_{n \rightarrow \infty} P \left(\bigcap_{m \geq n} \sup_Q \left\{ \left\| \Gamma^{-1} \right\| \left\| \Gamma_n(t) - \Gamma_0 \right\| + \left\| \Gamma^{-1} - \Gamma_0^{-1} \right\| \left\| E_F \left(\frac{\partial}{\partial \beta} \varphi \right) \right\| \right\} < k_0 \right) &= \\ = 1 \quad \text{pues } \left\| \Gamma^{-1} - \Gamma_0^{-1} \right\| < \varepsilon_2 \quad \text{pues } \left\| \Gamma - \Gamma_0 \right\| < \varepsilon_1, \quad \left\| \Gamma^{-1} \right\| \text{ está acotada} & \\ \text{y, por el lema 4 } \lim_{n \rightarrow \infty} P \left(\bigcap_{m \geq n} \sup_{\|t - \beta_0\| \leq d(\varepsilon)} \left\| \Gamma_n(t) - \Gamma_0 \right\| < \varepsilon_1 \right) = 1 & \end{aligned}$$

Lema 6

Sea $0 < k_0 < 1$

(i) Sea $A_{n,d,\epsilon} = \{ \|\gamma_{n\Gamma}(t) - \gamma_{n\Gamma}(t')\| < k_0 \|t - t'\| ; \|t - \beta_0\| \leq d, \|t' - \beta_0\| \leq d; \|\Gamma - \Gamma_0\| < \epsilon \}$

Entonces, existe $\epsilon_2, d_2 > 0$ tales que

$$\lim_{n \rightarrow \infty} P\left(\bigcap_{m \geq n} A_{m,d,\epsilon_2}\right) = 1 \text{ para todo } d \leq d_2$$

(ii) Sea $B_{n,d,\epsilon} = \{ \gamma_{n\Gamma}(t) \text{ va de } \|t - \beta_0\| < d \text{ en si mismo } \forall \Gamma \text{ tq } \|\Gamma - \Gamma_0\| < \epsilon \}$

$$\text{Entonces } \lim_{n \rightarrow \infty} P\left(\bigcap_{m \geq n} B_{m,d,\epsilon_2}\right) = 1 \text{ para todo } d \leq d_2$$

(iii) Sea $C_{n,d,\epsilon} = \{ \gamma_n(t) \text{ va de } \{t / \|t - \beta_0\| \leq d\} \text{ en } \{\Gamma / \|\Gamma - \Gamma_0\| \leq \epsilon\} \}$

$$\text{Entonces, } \lim_{n \rightarrow \infty} P\left(\bigcap_{m \geq n} C_{m,d,\epsilon_2}\right) = 1$$

Demostración

Por el lema anterior, sabemos que

$$\lim_{n \rightarrow \infty} P\left(\bigcap_{m \geq n} \sup_{(t,r) \in Q} \left\| \frac{\partial}{\partial t} \gamma_{n\Gamma}(t) \right\| \leq k_0\right) = 1 ; \text{ luego en } Q$$

$$\|\gamma_{n\Gamma}(t) - \gamma_{n\Gamma}(t')\| \leq \sup_Q \left\{ \left\| \frac{\partial}{\partial t} \gamma_{n\Gamma}(t) \right\|_0 \|t - t'\| \right\}$$

$$\leq \sup_Q \left\| \frac{\partial}{\partial t} \gamma_{n\Gamma}(t) \right\| \|t - t'\| \leq k_0 \|t - t'\|$$

y como $\gamma_{n\Gamma}(t)$ es continua, vale también en D_1, \dots, D_r y por tanto es Lipchitziana en todo R^k como queríamos.

(ii)

$$\|\gamma_{n\Gamma}(t) - \beta_0\| = \|\gamma_{n\Gamma}(t) - \gamma_{n\Gamma}(\beta_0) + \gamma_{n\Gamma}(\beta_0) - \beta_0\| \leq$$

$$\leq \|\gamma_{n\Gamma}(t) - \gamma_{n\Gamma}(\beta_0)\| + \|\gamma_{n\Gamma}(\beta_0) - \beta_0\| \leq k_0 \|t - \beta_0\| + \epsilon < d$$

por parte i) y lema 3

si $\|\gamma_{n\Gamma}(\beta_0) - \beta_0\| < (1-k_0)d$ ($k_0d + (1-k_0)d = d$)

Luego, $A_{n,d,\epsilon} \cap \{\|\gamma_{n\Gamma}(\beta_0) - \beta_0\| \leq (1-k_0)d\} \subset B_{n,d,\epsilon}$

Por tanto

$$\bigcap_{m > n} A_{m,d,\epsilon} \cap \{\|\gamma_{n\Gamma}(\beta_0) - \beta_0\| < (1-k_0)d\} \subset \bigcap_{m > n} B_{m,d,\epsilon}$$

$$y \lim_{n \rightarrow \infty} P\left(\bigcap_{m > n} A_{m,d,\epsilon} \cap \{\|\gamma_{n\Gamma}(\beta_0) - \beta_0\| < (1-k_0)d\}\right) = 1$$

(iii) Es directamente el lema 4.-

Lema 7

Sea $D_{n,d} = \{\gamma_n(t)$ tiene un único punto fijo t^* en $\|t - \beta_0\| < d$ y para
 cualquier t_0 con $\|t_0 - \beta_0\| < d$, $\gamma_n^{(k)}(t_0) \rightarrow t^*\}$

$$\text{donde } \gamma_n^{(k)}(t_0) = \gamma_n^{(k-1)}(\gamma_n(t_0))$$

Entonces, existe d_2 tal que para todo $d \leq d_2$

$$T) \lim_{n \rightarrow \infty} P\left(\bigcap_{m > n} D_{m,d}\right) = 1$$

Demostración

Sea $0 < k_0 < 1$

Basta ver que $D_{n,d} \subset (A_{n,d,\epsilon_2} \cap B_{n,d,\epsilon_2} \cap C_{n,d,\epsilon_2})$

Si ocurre A_{n,d,ϵ_2} como $k_0 < 1$, las $\gamma_{n\Gamma}$ son contracciones para cada Γ con $\|\Gamma - \Gamma_0\| < \epsilon_2$

Como ocurre B_{n,d,ϵ_2} transforman $\|t - \beta_0\| < d$ en si mismo.

Luego, por el teorema del punto fijo de Banach, tienen un único punto fijo para cada Γ , pero además, es el mismo $\forall \Gamma$, pues si t es un punto fijo vale que

$$\eta_{n\Gamma}(t) = t - \Gamma^{-1} \sum_{i=1}^n \varphi(Y_i, X_i, t) / n = t \quad \text{que ocurre si}$$

$$\sum_{i=1}^n \varphi(Y_i, X_i, t) = 0 \quad \text{que no depende de } \Gamma \text{ .-}$$

Finalmente, como $\eta_n(t) = t - \left(\frac{\partial P(t)}{\partial t}\right)^{-1} P(t) = t$ si

$P(t) = \sum_{i=1}^n \varphi(Y_i, X_i, t) = 0$ entonces $\eta_n(t)$ tiene un único punto fijo t^* en $\|t - \beta_0\| < d$

Nos resta ver que $\eta_n^{(k)}(t_0) \rightarrow t^*$

$$\text{aplicar } \eta_n(t_0) = \eta_{n\Gamma_n}(t_0)$$

$$t_1 = \eta_n(t_0) = \eta_{n\Gamma_n}(t_0)$$

.....

$$t_k = \eta_n(t_{k-1}) = \eta_{n\Gamma_n}(t_{k-1})$$

y como $\eta_{n\Gamma_n}(t)$ son contracciones, con el mismo punto fijo común t^*

si $\|\Gamma_n - \Gamma_0\| < \epsilon$, y $\|\Gamma_n(t_k) - \Gamma_0\| < \epsilon_2$ si ocurre $C_{n,d,\epsilon}$ pues como

$\|t - \beta_0\| < d$ y vale $C_{n,d,\epsilon}$ entonces, $\|\Gamma_n(t_0) - \Gamma_0\| < \epsilon$, iterando listo.-

Proposición 5

$$(i) \quad \hat{\beta}_n \xrightarrow{\text{p.p.}} \beta_0$$

$$(ii) \quad \hat{\beta}_n = \lim_{j \rightarrow \infty} \hat{\beta}_{n,j} \quad \text{si } n \geq n_0 \quad \text{p.p.}$$

Demostración

Sea $\hat{\beta}_{n,0}$ el estimador inicial consistente.

$$D_{n,d} \cap \{\|\hat{\beta}_{n,0} - \beta_0\| < d\} \subset \{\|\hat{\beta}_n - \beta_0\| < d\}$$

Luego,

$$\bigcap_{n \geq n_0} D_{n,d} \cap \{ \|\hat{\beta}_{n,0} - \beta_0\| < d \} \subset \bigcap_{n \geq n_0} \{ \|\hat{\beta}_n - \beta_0\| < d \}$$

por tanto, $\hat{\beta}_n \xrightarrow{\text{p.p.}} \beta_0$

y además

$$\bigcap_{n \geq n_0} D_{n,d} \cap \{ \|\hat{\beta}_{n,0} - \beta_0\| < d \} \subset \bigcap_{n \geq n_0} \{ \lim_{j \rightarrow \infty} \hat{\beta}_{n,j} = \hat{\beta}_n \text{ y } \hat{\beta}_n \text{ verifica } P(\hat{\beta}_n) = 0 \}$$

luego, $\hat{\beta}_n = \lim_{j \rightarrow \infty} \hat{\beta}_{n,j}$ si $n \geq n_0$ p.p.

Normalidad Asintótica

Modelo con efectos aleatorios.-

Sea $\hat{\beta}_n$ la solución que consideramos de la ecuación

$$\sum_{i=1}^n \varphi(Y_i, X_i, \hat{\beta}_n) = 0 \quad Y_i = g(\beta_0, X_i) + U_i$$

Por la proposición 5 $\hat{\beta}_n$ verifica que

$$n^{-1/2} \sum_{i=1}^n \varphi(Y_i, X_i, \hat{\beta}_n) \rightarrow 0 \text{ p.p.}$$

Sea $V = E_{\mathbb{F}}(\varphi(Y, X, \beta_0) \varphi(Y, X, \beta_0)^t)$ (está definida por hipótesis)

$$\Lambda = E_{\mathbb{F}}\left(\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0)\right) \text{ no singular (A2)}$$

Sea $\lambda(\beta) = E_{\mathbb{F}}(\varphi(Y, X, \beta))$

$$U(Y, X, \beta, d) = \sup_{|\tau - \beta| \leq d} |\varphi(Y, X, \tau) - \varphi(Y, X, \beta)|$$

De acuerdo con Huber (1967) (pag. 226-227) y corolario sobre normalidad asintótica se tiene que si se cumplen:

N1) $\varphi(Y, X, \beta)$ es \mathcal{Q} -medible y $\varphi(y, x, \beta)$ es separable en el sentido de Doob: Existe un conjunto N de \mathbb{F} -medida 0, y un subconjunto numerable $\mathbb{H}' \subset \mathbb{H}$ tal que para cada abierto $U \subset \mathbb{H}$, y cada intervalo cerrado A , los conjuntos

$$\{z / \varphi(Y, X, \beta) \in A \quad \forall \beta \in U\} \quad , \quad \{z / \varphi(Y, X, \beta) \in A \quad \forall \beta \in U \cap \mathbb{H}'\}$$

difieren en a lo sumo un subconjunto de N .-

N2) $\exists \beta_0$ tal que $\lambda(\beta_0) = 0$

N3) Existen a, b, c, d_0 estrictamente positivos tales que:

(i) $|\lambda(\beta)| = E_{\mathbb{F}}(\varphi(Y, X, \beta)) \geq a|\beta - \beta_0|$ para $|\beta - \beta_0| \leq d_0$

(ii) $E(U(Y, X, \beta, d)) \leq bd$ para $|\beta - \beta_0| + d \leq d_0$ $d \geq 0$

(iii) $E(U^2(Y, X, \beta, d)) \leq cd$ para $|\beta - \beta_0| + d \leq d_0$ $d \geq 0$

$$N4) \quad E(\|\varphi(Y, X, \beta_0)\|^2)$$

Si además $\hat{\beta}_n \xrightarrow{P} \beta_0$ y Λ es no singular

Tenemos que

$$n^{-1/2} (\hat{\beta}_n - \beta_0) \xrightarrow{w} N(0, \Lambda^{-1} V \Lambda^{-1})$$

Luego, para nuestro caso, basta ver que se verifican las hipótesis de Huber.-

Nos falta solo verificar N3.

(i) Como $\lambda(\beta_0) = 0$

$$\lambda(\beta) = \|\lambda(\beta) - \lambda(\beta_0)\| = \|E(\varphi(Y, X, \beta) - \varphi(Y, X, \beta_0))\| \geq a \|\beta - \beta_0\|$$

pues φ es diferenciable en β_0 (Por ejemplo $a = E(\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0)) / 2$)

(ii)

$$E(\sup_{\|\tau - \beta\| \leq d} \|\varphi(Y, X, \tau) - \varphi(Y, X, \beta)\|) \leq bd$$

Sabemos por la hipótesis A3 que $\sup_{\|a - \beta_0\| < \delta_0} \|\frac{\partial}{\partial \beta} \varphi(Y, X, a)\|$ es F inte-

grable.

Si $d_0 < \delta_0$, como $\|\tau - \beta_0\| < d_0$ y $\|\beta - \beta_0\| < d_0$ pues $d \leq d_0 - \|\beta - \beta_0\|$

$$\leq E(\sup_{\|a - \beta_0\| < \delta_0} \|\frac{\partial}{\partial \beta} \varphi(Y, X, a)\| \|\tau - \beta\|) \leq bd \quad \text{pues } \|\tau - \beta\| \leq d$$

(iii) Idem con $\sup_{\|a - \beta_0\| < \delta_0} \|\frac{\partial}{\partial \beta} \varphi(Y, X, a)\|^2$

Luego, tenemos probado el resultado sobre normalidad asintótica.-

SECCION IV

GM-estimadores óptimos sujetos a una cota en la sensibilidad.-

En esta sección hallaremos en la clase de los GM-estimadores, que verifiquen que su sensibilidad sea \leq que una constante prefijada, aquellos estimadores que minimizan la traza de la matriz de covarianza asintótica bajo normalidad.-

Sea entonces el modelo

$$Y_i = g(X_i, \beta_0) + U_i \quad i=1, \dots, n \quad Y_i \in \mathbb{R}, \quad \beta_0 \in \mathbb{R}^k$$

$X_i \in \mathbb{R}^q$ vectores aleatorios independientes igualmente distribuidos,

con distribución H

U_i variables aleatorias independientes con distribución $N(0, \sigma^2)$

e independientes de las X.

$g: \mathbb{R}^q \times \mathbb{R}^k \rightarrow \mathbb{R}$ con derivadas respecto de β continuas.

$$E\left(\frac{\partial g(X, \beta)}{\partial \beta} \frac{\partial^t g(X, \beta)}{\partial \beta} \right) < \infty$$

Como ya vimos en la sección anterior, consideramos la clase de los GM-estimadores para β , definidos por la ecuación

$$0 = \sum_{i=1}^n \psi(Y_i, X_i, \hat{\beta}) \quad \text{para } \psi: \mathbb{R} \times \mathbb{R}^q \times \mathbb{R}^k \rightarrow \mathbb{R}^k$$

La curva de influencia, para estos estimadores, esta dada por

$$IC \psi(Y_0, X_0) = \left(E\left(-\frac{\partial \psi(Y, X, \beta)}{\partial \beta} \right) \right)^{-1} \psi(Y_0, X_0, \beta)$$

Bajo las hipótesis consideradas en la sección III, un GM-estimador será consistente hacia el β_0 del modelo si

$E(\psi(Y, X, \beta)) = 0 \quad \forall \beta$, donde la esperanza se toma respecto de la distribución de (Y, X) en $\mathbb{R} \times \mathbb{R}^q$.

Consideraremos los GM-estimadores que lo verifican.-

La sensibilidad de un estimador dado por ψ

$$\gamma_{\psi}^* = \sup_{(Y, X)} \|IC_{\psi}(Y, X)\|$$

y la matriz de covarianza asintótica

$$V_{\psi}^* = E(IC_{\psi}(Y, X) IC_{\psi}(Y, X)^t)$$

Observaciones:

1) Tomando $\psi^*(Y, X, \beta) = (E(-\frac{\partial}{\partial \beta} \psi(Y, X, \beta)))^{-1} \psi(Y, X, \beta)$

se obtiene el mismo estimador $\hat{\beta}$ de β_0 , (pues las ecuaciones difieren en una matriz constante), pero además se verifica que $E(-\frac{\partial}{\partial \beta} \psi^*(Y, X, \beta)) = I_k$; por tanto, podemos suponer, sin pérdida

de generalidad que $E(-\frac{\partial}{\partial \beta} \psi(Y, X, \beta)) = I_k$

2) Para el caso de mínimos cuadrados

$$IC(Y, X) = E(-\frac{\partial}{\partial \beta} (Y - g(X, \beta)) \frac{\partial}{\partial \beta} g(X, \beta))^{-1} \psi(Y, X, \beta) =$$

$$= (E(\frac{\partial}{\partial \beta} g(X, \beta_0) \frac{\partial^t}{\partial \beta} g(X, \beta_0)) - E((Y - g(X, \beta_0)) H(X, \beta_0))^{-1} (Y - g(X, \beta_0)) \frac{\partial}{\partial \beta} g(X, \beta_0))$$

pero como $E((Y - g(X, \beta_0)) H(X, \beta_0)) = 0$ pues U_1 es independiente de X

y $E(U_1) = 0$ nos queda

$$= E(\frac{\partial}{\partial \beta} g(X, \beta_0) \frac{\partial^t}{\partial \beta} g(X, \beta_0))^{-1} (Y - g(X, \beta_0)) \frac{\partial}{\partial \beta} g(X, \beta_0) =$$

$$= (Y - g(X, \beta_0)) Q^{-1} \frac{\partial}{\partial \beta} g(X, \beta_0)$$

donde notamos por $Q = E(\frac{\partial}{\partial \beta} g(X, \beta_0) \frac{\partial^t}{\partial \beta} g(X, \beta_0))$

Luego, para mínimos cuadrados

$$V^* = E((Y - g(X, \beta_0))^2 Q^{-1} \frac{\partial}{\partial \beta} g(X, \beta_0) \frac{\partial^t}{\partial \beta} g(X, \beta_0) Q^{-1}) = Q^{-1} \sigma^2 E(\frac{\partial}{\partial \beta} g \frac{\partial^t}{\partial \beta} g) Q^{-1} =$$

$$= \sigma^2 Q^{-1}$$

Queremos encontrar ψ de modo que:

traza V^* = mínima, sujeta a una cota en la sensibilidad

$$\|IC_{\psi}(Y, X)\| \leq a \quad \text{y} \quad E(\psi(Y, X, \beta)) = 0$$

Lema I

Bajo las hipótesis consideradas, si

$E(\psi(Y, X, \beta_0)) = 0 \quad \forall \beta_0$ y ψ es acotada, entonces

$$T) \quad E\left(\frac{\partial}{\partial \beta} \psi(Y, X, \beta_0)\right) = E\left(\frac{1}{\sigma^2} (Y - g(X, \beta_0)) \psi(Y, X, \beta_0) \frac{\partial}{\partial \beta} g(X, \beta_0)\right)$$

Demostración

Sea H la distribución de X en R^q y m la medida de Lebesgue en R .

Entonces, con respecto a la medida producto $m \times H$ en $R \times R^q$, la distribución de (Y, X) tiene densidad $\frac{1}{\sigma} \Phi'((Y - g(X, \beta))/\sigma)$ donde $\Phi(x)$

es la función de distribución $N(0, 1)$.

Por tanto, tenemos que

$$\frac{\partial}{\partial \beta} \int \psi(Y, X, \beta) \frac{1}{\sigma} \Phi'((Y - g(X, \beta))/\sigma) d(m \times H)(Y, X) \Big|_{\beta_0} = 0$$

Como ψ es acotada y $X^\alpha \Phi'(t)$ es integrable, podemos derivar dentro de la integral, y por tanto tenemos que

$$\begin{aligned} 0 &= \int \frac{\partial}{\partial \beta} \psi(Y, X, \beta) \Big|_{\beta_0} \frac{1}{\sigma} \Phi'((Y - g(X, \beta_0))/\sigma) d(m \times H)(Y, X) + \\ &+ \psi(Y, X, \beta_0) \frac{1}{\sigma} \frac{\partial}{\partial \beta} \Phi'((Y - g(X, \beta_0))/\sigma) \Big|_{\beta_0} d(m \times H)(Y, X) = \\ &= E\left(\frac{\partial}{\partial \beta} \psi(Y, X, \beta_0)\right) + \int \psi(Y, X, \beta_0) \frac{1}{\sigma^3} (Y - g(X, \beta_0)) \Phi'((Y - g(X, \beta_0))/\sigma) \frac{\partial}{\partial \beta} g(X, \beta_0) d(m \times H) \\ &= E\left(\frac{\partial}{\partial \beta} \psi(Y, X, \beta_0)\right) + E\left(\frac{1}{\sigma^2} (Y - g(X, \beta_0)) \psi(Y, X, \beta_0) \frac{\partial}{\partial \beta} g(X, \beta_0)\right) \end{aligned}$$

como queríamos.-

Proposición 7

La solución de traza V^* = mínimo,

sujeto a

$$\begin{cases} E\left(-\frac{\partial}{\partial \beta} \varphi(Y, X, \beta_0)\right) = I_k & (i) \\ E(\varphi(Y, X, \beta_0)) = 0 & (ii) \\ \|\varphi(Y, X, \beta_0)\| \leq a \quad \forall (Y, X) & (iii) \end{cases}$$

se obtiene para $\varphi = \Psi_a((Y-g(X, \beta))A^{-1}\frac{\partial}{\partial \beta}g(X, \beta))$

donde $\Psi_a: R^k \rightarrow R^k$

$$\Psi_a(Z) = \begin{cases} Z & \text{si } \|Z\| \leq a \\ \frac{aZ}{\|Z\|} & \text{si } \|Z\| > a \end{cases}$$

y la matriz A es solución de

$$A = E\left(2\Phi\left(\frac{a}{\sigma}\|A^{-1}\frac{\partial}{\partial \beta}g(X, \beta)\|\right) - 1\right) \frac{\partial}{\partial \beta}g(X, \beta) \frac{\partial^t}{\partial \beta}g(X, \beta)$$

(que luego veremos que existe)

Observaciones:

A se ha tomado de modo que $E\left(-\frac{\partial}{\partial \beta} \Psi_a\left((Y-g(X, \beta))A^{-1}\frac{\partial}{\partial \beta}g(X, \beta)\right)\right) = I_k$

2) Bajo la restricción (i), (iii) no es otra cosa que

$$\|IC_\varphi(Y, X)\| \leq a \quad \forall (Y, X)$$

Demostración

Sea A^{-1} una matriz simétrica cualquiera. Bajo las restricciones (i)

$$V_\varphi^* = E(\varphi \varphi^t)$$

$$\begin{aligned} \boxed{I} &= E\left((Y-g(X, \beta))A^{-1}\frac{\partial}{\partial \beta}g(X, \beta) - \varphi(Y, X, \beta_0)\right)\left(\left((Y-g(X, \beta))A^{-1}\frac{\partial}{\partial \beta}g(X, \beta) - \varphi(Y, X, \beta_0)\right)^t\right) \\ &= E\left((Y-g(X, \beta))^2 A^{-1}\frac{\partial}{\partial \beta}g(X, \beta) \frac{\partial^t}{\partial \beta}g(X, \beta) A^{-1}\right) + E(\varphi \varphi^t) - \\ &\quad - E\left((Y-g(X, \beta))A^{-1}\frac{\partial}{\partial \beta}g(X, \beta) \varphi^t\right) - E\left((Y-g(X, \beta))\varphi \frac{\partial^t}{\partial \beta}g(X, \beta) A^{-1}\right) \end{aligned}$$

$$= \sigma^2 E(A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial}{\partial \beta} g(X, \beta) A^{-1}) + E(\psi \psi^t) - 2 A^{-1} \sigma^2$$

por el lema I y ψ verificando (i), además de que $E((Y-g(X, \beta))^2) = \sigma^2$
 $(Y-g(X, \beta))$ es independiente de X

Luego, es lo mismo minimizar bajo las restricciones traza V_{ψ}^*
 que traza de \boxed{I} .- Pero

$$\text{traza } \boxed{I} = E(\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) - \psi(Y, X, \beta) \|^2) = \text{mínimo, suje-}$$

$$\text{to a } \|\psi(Y, X, \beta)\| \leq a$$

$$\text{tiene por solución a } \psi_a \left((Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right)$$

donde A es tal que

$$E\left(-\frac{\partial}{\partial \beta} \psi_a \left((Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right) \right) = I_k$$

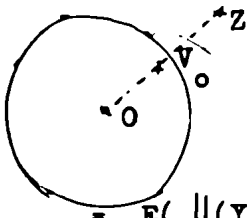
$$\text{pues si llamamos } D_a = \left\{ (Y, X) / \left\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right\| > a \right\}$$

$$E(\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) - \psi(Y, X, \beta) \|^2) \geq$$

$$\geq E(I_{D_a} \| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) - \psi(Y, X, \beta) \|^2) \geq$$

$$\geq E(I_{D_a} \left\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) - \frac{a(Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta)}{\left\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right\|} \right\|^2) =$$

$$\text{pues si } \|Z\| > a, \quad \inf_{V \in B(0, a)} \|Z - V\| = \|Z - V_0\| \quad \text{donde } V_0 = \frac{a Z}{\|Z\|}$$



pero esta última no es otra cosa que

$$= E(\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) - \psi_a \left((Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right) \|^2)$$

y como ψ_a verifica la condición (ii) por ser los residuos normales
 (simétricos) y ψ_a impar, obtenemos lo que queríamos.-

Resta ver entonces que existe A tal que

$$A = E \left(\left(2 \Phi \left(a/\sigma \left\| A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right\| \right) - 1 \right) \frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial^t}{\partial \beta} g(X, \beta) \right)$$

y que para dicha A ,

$$E \left(\frac{-\partial}{\partial \beta} \psi_a \left((Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right) \right) = I_k$$

Por el lema I, como $E(\psi_a) = 0 \forall \beta_0$ y es acotada

$$E \left(\frac{-\partial}{\partial \beta} \psi_a (Y, X, \beta_0) \right) = E \left(\frac{1}{\sigma^2} (Y-g(X, \beta_0)) \psi_a \left((Y-g(X, \beta_0)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta_0) \right) \frac{\partial^t}{\partial \beta} g(X, \beta_0) \right)$$

Sea $B = \left\{ (Y, X) / \left\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right\| < a \right\}$ $I_B(Y, X)$ = indicador del conjunto B.

$$= E \left(I_B(Y, X) \frac{1}{\sigma^2} (Y-g(X, \beta)) \psi_a \left((Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right) \frac{\partial^t}{\partial \beta} g(X, \beta) \right) +$$

$$+ E \left((1-I_B) \frac{1}{\sigma^2} (Y-g(X, \beta)) a \frac{(Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta)}{\left\| (Y-g(X, \beta)) A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right\|} \frac{\partial^t}{\partial \beta} g(X, \beta) \right)$$

$$= E \left(A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial^t}{\partial \beta} g(X, \beta) E_{Y/X} \left(I_B (Y-g(X, \beta))^2 / \sigma^2 \right) \right) +$$

$$+ E \left(\frac{1}{\sigma^2} a A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial^t}{\partial \beta} g(X, \beta) / \left\| A^{-1} \frac{\partial}{\partial \beta} g(X, \beta) \right\| \cdot E_{Y/X} \left((Y-g(X, \beta)) (1-I_B) \right) \right)$$

$$E_{Y/X} \left(I_B (Y-g(X, \beta))^2 / \sigma^2 \right) = \int_{-a/\|A^{-1} \frac{\partial}{\partial \beta} g(X, \beta)\|}^{a/\|A^{-1} \frac{\partial}{\partial \beta} g(X, \beta)\|} \frac{u^2}{\sigma^3} \Phi'(u/\sigma) du =$$

$$= 2 \int_0^{a/\|A^{-1} \frac{\partial}{\partial \beta} g(X, \beta)\|} \frac{u^2}{\sigma^3} \Phi'(u/\sigma) du = 2 \int_0^{a/\sigma \|A^{-1} \frac{\partial}{\partial \beta} g\|} \frac{z^2 \Phi'(z)}{\sigma^3} dz$$

$$= -2 \int_0^{a/\sigma \|A^{-1} \frac{\partial}{\partial \beta} g\|} z \Phi''(z) dz \quad (\text{donde hicimos } z=u/\sigma, \text{ y usamos que}$$

$\Phi''(z) = -z \Phi'(z)$) y ahora integrando por partes, tenemos que =

$$= -2 \left(\frac{a}{\sigma \|A^{-1} \frac{\partial g}{\partial \beta}\|} \Phi' \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) + \frac{1}{2} - \Phi \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) \right)$$

$$E_{Y/X} \left((1 - I_B) |Y - g(X, \beta)| \right) = 2 \int_{a/\|A^{-1} \frac{\partial g}{\partial \beta}\|}^{\infty} \frac{u}{\sigma} \Phi'(u/\sigma) du = 2\sigma \int_{a/\|A^{-1} \frac{\partial g}{\partial \beta}\| \sigma}^{\infty} z \Phi'(z) dz =$$

$$= 2\sigma \Phi' \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right)$$

Luego,

$$E \left(2A^{-1} \frac{\partial g}{\partial \beta} \frac{\partial}{\partial \beta} \left[\left(\Phi \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) - \frac{a}{\sigma \|A^{-1} \frac{\partial g}{\partial \beta}\|} \Phi' \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) - \frac{1}{2} \right] + \right.$$

$$\left. + A^{-1} \frac{\partial g}{\partial \beta} \frac{\partial}{\partial \beta} \left[\frac{a}{\sigma \|A^{-1} \frac{\partial g}{\partial \beta}\|} 2\sigma \Phi' \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) \right] \right) =$$

$$= E \left(A^{-1} \frac{\partial g}{\partial \beta} \frac{\partial}{\partial \beta} g(X, \beta) \left(2\Phi \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) - 1 \right) \right)$$

que para que sea $= I_k$ debe ser

$$A = E \left(\left(2\Phi \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) - 1 \right) \frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial}{\partial \beta} g(X, \beta) \right)$$

Finalmente, veamos que tal A existe.-

Proposición 8

La solución de $A = E \left(\left(2\Phi \left(\frac{a}{\sigma} \|A^{-1} \frac{\partial g}{\partial \beta}\| \right) - 1 \right) \frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial}{\partial \beta} g(X, \beta) \right)$
(i)

existe si a es suficientemente grande.-

(ii) $A \rightarrow E \left(\frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial}{\partial \beta} g(X, \beta) \right)$ para $a \rightarrow \infty$

Demostración

Sea $B(Q, d)$ la bola de centro Q y radio d, o sea

$$\{V / \|Q - V\| < d\} \quad \text{siendo } |V| = \sup_{\|z\|=1} \|Vz\|$$

donde d es suficientemente pequeño de modo que para toda $V \in B(Q, d)$

sea inversible.-

$$\text{Sea } J(V, a) = E((2\Phi(a/\sigma \|V^{-1} \frac{\partial g}{\partial \beta}(X, \beta)\|) - 1) \frac{\partial g}{\partial \beta}(X, \beta) \frac{\partial^t g}{\partial \beta}(X, \beta))$$

Veremos que existe A , a tal que $J(A, a) = A$

Sea $\epsilon > 0$, luego, existe BCR^q compacto, tal que

$$\begin{aligned} E(|(1-I_B(x)) \frac{\partial g}{\partial \beta}(X, \beta) \frac{\partial^t g}{\partial \beta}(X, \beta)|) < \epsilon \quad \text{pues} \quad E(\|\frac{\partial g}{\partial \beta}(X, \beta) \frac{\partial^t g}{\partial \beta}(X, \beta)\|) < \infty \\ \|J(V, a) - J(Q, a)\| &= \| E((2\Phi(a/\sigma \|V^{-1} \frac{\partial g}{\partial \beta}\|) \frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta}) - E(\frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta}) + E(\frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta}) \\ &\quad - E(2\Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|) \frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta}) \| = \\ &= 2 \| E((\Phi(a/\sigma \|V^{-1} \frac{\partial g}{\partial \beta}\|) - \Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|)) \frac{\partial g}{\partial \beta}(X, \beta) \frac{\partial^t g}{\partial \beta}(X, \beta)) \| \leq \\ &\leq 2 E(I_B |\Phi(a/\sigma \|V^{-1} \frac{\partial g}{\partial \beta}\|) - \Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|) | \| \frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta} \|) + \\ &+ 2 E((1-I_B) |\Phi(a/\sigma \|V^{-1} \frac{\partial g}{\partial \beta}\|) - \Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|) | \| \frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta} \|) \\ &\leq \epsilon + 2 E(I_B |\Phi(a/\sigma \|V^{-1} \frac{\partial g}{\partial \beta}\|) - \Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|) | \| \frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta} \|) \end{aligned}$$

Pero como $\|V^{-1} \frac{\partial g}{\partial \beta}(X, \beta)\|$ está acotado cuando $X \in B$ y $V \in B(Q, d)$ pues $\frac{\partial g}{\partial \beta}(X, \beta)$ es continua, y la transformación que manda $V \rightarrow V^{-1}$ es continua, vale que

$$|\Phi(a/\sigma \|V^{-1} \frac{\partial g}{\partial \beta}\|) - \Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|) | < \epsilon' \quad \text{si } a \geq a_0 \text{ uniformemente en } V \text{ y } X$$

Por tanto,

$$\sup_{V \in B(Q, d)} \|J(V, a) - J(Q, a)\| < d/2 \quad \text{si } a \geq a_0$$

$$\text{Luego, como } \|J(Q, a) - Q\| = E((2\Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|) - 2) \frac{\partial g}{\partial \beta} \frac{\partial^t g}{\partial \beta}) ;$$

$$\text{y } 2\Phi(a/\sigma \|Q^{-1} \frac{\partial g}{\partial \beta}\|) - 2 \rightarrow 0 \quad \text{p.p. cuando } a \rightarrow \infty$$

Por tanto, $J(Q,a) - Q \rightarrow 0$ para $a \rightarrow \infty$

Luego, existe a_1 tal que $\|J(Q,a) - Q\| < d/2$ si $a \geq a_1$

Por tanto,

$$\sup_{V \in B(Q,d)} \|J(V,a) - Q\| < d \quad \text{si } a \geq \max(a_0, a_1)$$

y por tanto, la transformación $V \rightarrow J(V,a)$ es continua y manda $B(Q,d)$ en sí misma para $a \geq \max(a_0, a_1)$.-

Por tanto, por el teorema del punto fijo de Broewer, tiene un punto fijo A.-

Además, como $A_a \in B(Q,d)$ si $a \geq \max(a_0, a_1)$

$A_a \rightarrow Q$ para $a \rightarrow \infty$.-

Observación 1:

En la mayoría de los casos tanto σ , como la distribución de los X son desconocidas. Luego, A puede elegirse como la solución de

$$A = \frac{1}{n} \sum_{i=1}^n \left(2 \Phi(a/\hat{\sigma}_0 \|A^{-1} \frac{\partial g(X_i, \hat{\beta}_0)}{\partial \beta}\|) - 1 \right) \frac{\partial g(X_i, \hat{\beta}_0)}{\partial \beta} \frac{\partial^t g(X_i, \hat{\beta}_0)}{\partial \beta}$$

donde $\hat{\sigma}_0$ es un estimador robusto de σ , así como $\hat{\beta}_0$ lo es de β_0 .-

Observación 2:

La matriz A existe sii $a \geq \sqrt{\pi}/2 \frac{\sigma}{E\left(\frac{\partial g(X, \beta_0)}{\partial \beta}\right)}$

Dem.

$$A = E\left(\left(2 \Phi\left(a/\sigma \|A^{-1} \frac{\partial g}{\partial \beta}\right) - 1\right) \frac{\partial g(X, \beta_0)}{\partial \beta} \frac{\partial^t g(X, \beta_0)}{\partial \beta}\right)$$

multiplicando por A^{-1} y tomando norma

$$1 \leq E\left(2\left(\Phi\left(a/\sigma \|A^{-1} \frac{\partial g}{\partial \beta}\right) - \frac{1}{2}\right) A^{-1} \frac{\partial g(X, \beta_0)}{\partial \beta} \frac{\partial^t g(X, \beta_0)}{\partial \beta}\right)$$

y como Φ es cóncava, y $\Phi(0) = 1/2$ es

$$\leq E(2\Phi'(0) \frac{a}{\sqrt{\|A^{-1} \frac{\partial g}{\partial \beta}\|}} \|A^{-1} \frac{\partial g}{\partial \beta} \frac{t}{\|A^{-1} \frac{\partial g}{\partial \beta}\|} \|) \leq 2\Phi'(0) \frac{a}{\sqrt{\|A^{-1} \frac{\partial g}{\partial \beta}\|}} E(\|A^{-1} \frac{\partial g}{\partial \beta}\| \frac{t}{\|A^{-1} \frac{\partial g}{\partial \beta}\|} / \|A^{-1} \frac{\partial g}{\partial \beta}\|)$$

$$= 2\Phi'(0) \frac{a}{\sqrt{\|A^{-1} \frac{\partial g}{\partial \beta}\|}} E(\|\frac{\partial g}{\partial \beta}(X, \beta)\|) = \frac{2}{\sqrt{2\pi}} \frac{a}{\sqrt{\|A^{-1} \frac{\partial g}{\partial \beta}\|}} E(\|\frac{\partial g}{\partial \beta}(X, \beta)\|)$$

por tanto $a \geq \sqrt{\pi/2} \frac{\sqrt{\|A^{-1} \frac{\partial g}{\partial \beta}\|}}{E(\|\frac{\partial g}{\partial \beta}(X, \beta)\|)}$

Observación 3

La función Ψ_a obtenida, se puede poner (a menos de una matriz constante, como la correspondiente a mínimos cuadrados pesados, por tanto el estimador resultante es el mismo.

La familia de mínimos cuadrados pesados viene dada para este caso por considerar funciones Ψ de la forma:

$$\Psi(Y, X, \beta) = w(Y, X, \beta) (Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta}$$

con $w: R^q \times R^k \rightarrow R$ función de peso.

Ψ_a se puede poner como:

$$\Psi_a((Y - g(X, \beta)) A^{-1} \frac{\partial g(X, \beta)}{\partial \beta}) = (Y - g(X, \beta)) A^{-1} \frac{\partial g(X, \beta)}{\partial \beta} w_1(\|(Y - g(X, \beta)) A^{-1} \frac{\partial g(X, \beta)}{\partial \beta}\|)$$

siendo

$$w_1(t) = \begin{cases} 1 & \text{si } t \leq a \\ a/t & \text{si } t > a \end{cases}$$

Luego, si hay un óptimo en el sentido de definidas positivas (para la matriz de covarianza), en la clase de mínimos cuadrados pesados, en particular debe minimizar la traza de la matriz de covarianza asintótica, y por tanto coincidir con el estimador dado por Ψ_a .

SECCION V

Resultados de Simulación.-

Se consideró para el modelo con efectos fijos, la familia de funciones de regresión logística (a un parámetro) dada por:

$$g(x, \theta) = \frac{1}{1 + e^{-\theta x}}$$

para estudiar el comportamiento de los estimadores robustos propuestos, y su relación con el de mínimos cuadrados cuando el tamaño de la muestra es pequeño.

Para ello, se tomó como valor verdadero del parámetro $\theta = 1$, y a partir de él se generaron muestras Y_1, \dots, Y_{n_1} generando los residuos U_i

$$Y_i = \frac{1}{1 + e^{-X_i}} + U_i/10$$

cuando los U_i eran:

Normales (0,1)

normal contaminada con normal al 5% y desviación 5
($\Phi(x) = 0.95\Phi(x) + 0.05\Phi(x/5)$)

Student con tres grados de libertad

Para los efectos fijos, se tomó un diseño simétrico

$$X_i = 0.4 Z + 1.25 \quad i=1, \dots, n_1/2$$

$$X_{n_1/2+i} = -X_i$$

donde Z es una variable con distribución $N(0,1)$; o sea que X_1 tiene distribución $N(1.25, 0.16)$

Se consideraron tamaños de muestra $n_1 = 20$ y 50 , y para cada tipo de residuo se hicieron $2000 = NR$ replicaciones con tamaño de muestra 20 y 2000 con tamaño 50 .

Fijado el tipo de residuo, para cada una de las 2000 muestras Y_1, \dots, Y_{n_1} se calcularon aplicando el algoritmo de Newton con

hasta 10 iteraciones los estimadores de:

mínimos cuadrados

M-estimador con $\psi = \psi_k$ (Huber) $K=0.1346$

M-estimador con $\psi = \psi_A$ (Tukey) $A=0.4685$ donde

$$\psi_k(t) = \begin{cases} t & \text{si } |t| \leq k \\ k & \text{si } t > k \\ -k & \text{si } t < -k \end{cases}$$

$$\psi_A(t) = \begin{cases} t(1 - (t/A)^2)^2 & \text{si } |t| \leq A \\ 0 & \text{si } |t| > A \end{cases}$$

obteniéndose muestras de tamaño 2000 de dichos estimadores, que notaremos T_i^{MC} ; T_i^{HU} ; T_i^{TU} $i=1, \dots, 2000$ respectivamente a partir de las cuales se hace la comparación.

Llamemos $B_i^{(j)} = |T_i^{(j)} - 1|$ $i=1, \dots, 2000$ $j=MC, HU, TU$ a los desvíos absolutos.

Se calcularon para cada caso (MC, HU, TU) :

la media

el error cuadrático medio

y las siguientes medidas robustas de dispersión:

mediana $B_i^{(j)}$

percentil 0.75 de $B_i^{(j)}$ = $B'_{(0.75 \times 2000)}^{(j)}$ donde $B'_{(h)}^{(j)}$ es el

vector $B_i^{(j)}$ ordenado (en orden creciente)

percentil 0.90 de $B_i^{(j)}$

percentil 0.98 de $B_i^{(j)}$

encontrándose los resultados obtenidos en las tablas que se encuentran a continuación.

La elección de los valores de las constantes K y A para las funciones Ψ_K y Ψ_A , se hizo de modo que la eficiencia asintótica bajo normalidad sea del 95% (de acuerdo con las fórmulas de varianza asintótica obtenidas en la proposición 2).

De las tablas obtenidas, para las primeras 6 se partió en el método de Newton del valor verdadero (sin estimador inicial) para muestras de tamaño 20 (las. tres tablas) y tamaño 50.

En el caso de las tablas 6 a 9, se partió de estimadores iniciales, tomando uno robusto para T_i^{HU} , y para T_i^{TU} se tomó como valor inicial el valor final de T_i^{HU} ; para muestras de tamaño 20.-

Para el caso de mínimos cuadrados, el estimador inicial es del tipo del propuesto por Hartley (1965) (promediar y tomar imagen inversa) mientras que como estimador inicial robusto se tomó la mediana de la distribución que toma valores $-\ln(1/Y_i - 1) / X_i$ con pesos $|X_i| / \sum_{i=1}^n |X_i|$ (tomamos como mediana el 1er. valor que su probabilidad acumulada (a izquierda) supere 0.5).-

Los programas principales que se usaron son los que se denominaron F.JUN 26 para el caso sin estimador inicial

F.JUN 27 para el caso con estimador inicial.

La biblioteca de subrutinas F.S.JUN 26 (BIB.JUN 26) que contiene a las subrutinas: CALCUL, RTNI (Newton), FCT (Calcula la función y su derivada), SORT, RATVEC (genera Student), GENERA (genera normal contaminada), FUNCTION RANORM(T), FUNCTION ALEA(T), RANDU, COMPAR (calcula las medidas que se encuentran en las tablas), CALCUI,SESTIN(calcula estimadores iniciales).-

En las tablas se observá que también para muestras pequeñas, el comportamiento bajo normalidad es casi equivalente; mientras que en los otros casos la eficiencia llega a ser mas del doble que la del estimador de mínimos cuadrados. Se resume en el siguiente cuadro la eficiencia relativa:

N O R M A L I D A D

		ECM(MC)/ECM(HU)	Percentiles 0.75	Percentiles 0.90
T ^{HU} HUBER	n=20	0.948	0.945	0.987
	n=50	0.953	0.932	0.918
	n=20 con est. inicial	0.968	0.969	0.970
		ECM(MC)/ECM(TU)	Percentiles 0.75	Percentiles 0.90
T ^{TU} TUKEY	n=20	0.945	0.962	0.976
	n=50	0.933	0.936	0.918
	n=20 con est. inicial	0.964	0.964	0.976

N O R M A L C O N T A M I N A D A

		ECM(MC)/ECM(HU)	Percentiles 0.75	Percentiles 0.90
T ^{HU} HUBER	n=20	1.881	1.583	1.682
	n=50	1.778	1.642	1.635
	n=20 con E. I.	1.969	1.579	2.057
		ECM(MC)/ECM(TU)	Percentiles 0.75	Percentiles 0.90
T ^{TU} TUKEY	n=20	1.923	1.585	1.626
	n=50	1.836	1.716	1.745
	n=20 CON EI	2.048	1.669	2.057

S T U D E N T T R E S G R A D O S

		ECM(MC)/ECM(HU)	Percentiles 0.75	Percentiles 0.90
T ^{HU} HUBER	n=20	2.241	1.623	1.708
	n=50	1.949	1.680	1.827
	n=20 con EI	1.7514	1.643	1.662
		ECM(MC)/ECM(TU)	Percentiles 0.75	Percentiles 0.90
T ^{TU} TUKEY	n=20	2.25	1.657	1.718
	n=50	1.968	1.756	1.847
	n=20 con EI	1.718	1.594	1.623

A continuación se encuentran las tablas obtenidas para los distintos casos.-

* * * * *
 * COMPARACION POR MUESTREO DEL COMPORTAMIENTO DE *
 * ESTIMADORES ITERATIVOS PARA REGRESION LOGIT LINEAL *
 * * * * *

TABLA 7.1

FUNCION DE REGRESION = LOGISTICA

RESIDUOS CON DISTRIBUCION = NORMAL 0

	MEGIA	ERROR	CUADRATICO	MEDIA	PERCENTILES	VALORES ABSOLUTOS	VALORES ABSOLUTOS
	MEGIA	MEGIA	MEGIA	MEGIA	MEGIA	MEGIA	MEGIA
ESTIMADOR	1.0033	0.01324460	0.0798	0.2747	0.1959	0.1390	
WILCOX							
CUADRADOS							
WILCOX	1.0035	0.01326467	0.0790	0.2804	0.1381	0.1337	
TUKEY	1.0034	0.01401337	0.0787	0.2805	0.1392	0.1325	

* * * * *
 * * * * *
 * * * * *
 * * * * *
 * * * * *
 * * * * *
 * * * * *
 * * * * *
 * * * * *
 * * * * *

* COMPARACION POR MUESTREO DEL COMPORTAMIENTO DE *
 * ESTIMADORES ITERATIVOS PARA REGRESION NO LINEAL *
 * * * * *

TABLE 2.3

FUNCIÓN DE REGRESION = LOGISTICA
 RESIDUOS CON DISTRIBUCION = NORMAL CONTAMINADA

	MEDIA	ERROR CUADRATICO MEDIO	MEDIANA VALORES ABSOLUTOS	PERCENTILES CENTRADOS	VALORES ABSOLUTOS CENTRADOS
ESTIMADOR	1.0121	0.016909	0.1051	0.4533	0.2727
MINIMOS CUADRADOS					0.1809
M-ESTIMADOR					0.75
HUBER	1.0090	0.01625216	0.0936	0.3125	0.2112
M-ESTIMADOR					0.1437
TUCKY	1.0099	0.01648395	0.0826	0.3007	0.2139
					0.1436

$n_1 = 50$

* * * * *
 * COMPARACION POR MUESTREO DEL COMPORTAMIENTO DE *
 * ESTIMADORES ITERATIVOS PARA REGRESION NO LINEAL *
 * * * * *

TABLA 2.1
 FUNCION DE REGRESION = LOGISTICA
 RESIDUOS CON DISTRIBUCION = NORMAL

	MEDIA	ERRORES CUADRATICOS MEDIO	MEDIANA VALORES ABSOLUTOS CENTRADOS	PERCENTILES VALORES ABSOLUTOS CENTRADOS	
ESTIMADOR	0.9998	0.00504711	0.0458	0.1701	0.1171
MINIMOS CUADRADOS					0.0806
M-ESTIMADOR	1.0001	0.00541230	0.0490	0.1744	0.1220
HURER					0.0935
M-ESTIMADOR	1.0001	0.00540775	0.0475	0.1744	0.1222
TUKEY					0.0833

COMPARACION POR MONTECARLO DEL COMPORTAMIENTO DE ESTIMADORES ITERATIVOS PARA REGRESION NO LINEAL

TABLA 2.3

FUNCION DE REGRESION = LOGISTICA
RESIDUOS CON DISTRIBUCION = NORMAL CONTAMINADA

	MEDIA	ERRORES	MEDIANA	PERCENTILES	VALORES ABSOLUTOS	VALORES ABSOLUTOS
		CUADRATICOS	VALORES	CENTRADOS	CENTRADOS	CENTRADOS
		MEG	ABSOLUTOS	0.90	0.90	0.75
ESTIMADOR	1.0011	0.01150404	0.0707	0.2583	0.1710	0.1192
MINIMOS CUADRADOS						
M-ESTIMADOR						
HUPFER	1.0005	0.00666340	0.0524	0.1783	0.1242	0.0930
M-ESTIMADOR						
TUKY	1.0017	0.00625025	0.0534	0.1791	0.1299	0.0910

* * * * *
 * COMPARACION POR MULTICARLO DEL COMPORTAMIENTO DE *
 * ESTIMADORES ITERATIVOS PARA REGRESION NO LINEAL *
 * * * * *
 * * * * *
 * * * * *

$n_j = 50$

TARLA 2.4

FUNCION DE REGRESION = LOGISTICA
 RESIDUOS CON DISTRIBUCION = STUDENT TRESGRADOS

	MEDIA	VARIAZION CUADRATICA MEDIO	MEDIANA	PERCENTILES	VALORES ABSOLUTOS CENTRADOS	VALORES ABSOLUTOS CENTRADOS
ESTIMADOR	1.0037	0.01595380	0.0722	0.3083	0.1972	0.1351
M-ESTIMADOR	1.0021	0.00912573	0.0520	0.2063	0.1459	0.1050
HUBER	1.0022	0.00810343	0.0607	0.2118	0.1451	0.1028
TUKEY	1.0022	0.00810343	0.0607	0.2118	0.1451	0.1028

CON ESTIMADOR INICIAL

TABLA 2.4

FUNCION DE REGRESION = LOGISTICA

RESIDUOS CON DISTRIBUCION = STUDENT TRESGRADOS

	MEIA	ERROR CUADRATICO MEDIO	MEJANA VALORES ABSOLUTOS CENTRADOS	PERCENTILES	VALORES ABSOLUTOS CENTRADOS
ESTIMADOR	1.0147	0.0359335	0.1197	0.3041	0.2093
MINIMOS CUADRADOS					
M-ESTIMADOR					
HUBER	1.0107	0.0211227	0.0954	0.3439	0.1633
M-FSTIMADOR					
TUKEY	1.0105	0.02153234	0.0949	0.3400	0.1658

REFERENCIAS

- Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972) . Robust estimates of location: Survey and Advances. Princeton Univ. Press.
- Beaton, A.E. and Tukey, J.W. (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 16, 147-185.
- Billingsley, P. (1968) Convergence of probability measures. J.Wiley.
- Collins, J.R. (1976) Robust estimation of a location parameter in the presence of asymmetry. The Annals of Statistics 4, 68-85.
- Hampel, F.R. (1968) Contribution to the theory of robust estimation. Ph.D. Dissertation, Univ. of California, Berkeley.
- Hampel, F.R. (1974) The influence curve and its role in robust estimation. Journal of the American Statistical Association 69 383-393.
- Hartley, H.O. and Booker, A. (1965) Non-linear least squares estimation. Annals of Mathematical Statistics 36, 638-650.
- Huber, P.J. (1964) Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73-101.
- Huber, P.J. (1967) The behavior of maximum likelihood estimates under non-standart conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. 221-233.
- Huber, P.J. (1973) Robust regression: asymptotics, conjectures and Monte Carlo. Annals of Statistics 1, 799-821.
- Huber, P.J. (1977) Robust Statistical Procedures. Society for Industrial and Applied Mathematics, Philadelphia.
- Holland, P.W. and Welsch, R.E. (1977) Robust regression using iteratively reweighted least squares. Communications in Statistics, Theory and Methods, 9, 813-827.
- Jenrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics 2, 633-643.
- Krasker, W.S. (1978) Estimation in linear regression models with disparate data points. Manuscripto no publicado.
- Klein, R. and Yohai, V.J. (1979) Asymptotic behavior of iterative M-estimators for the linear model. A aparecer.

- Malenvaud, E. (1970) The consistency of nonlinear regressions
The Annals of Mathematical Statistics. 3, 956-969.
- Relles, D.A. (1968) Robust regression by modified least squares.
Ph.D. Dissertation Yale University.
- Souza, G. (1979) Statistical Inference in nonlinear models: a pseudo
likelihood approach. Ph.D. Dissertation, North Carolina State
University.
- Yohai, V.J. (1974) Robust estimation in the linear model. The Annals
of Statistics 2, 562-567.
- Yohai, V.J. and Maronna, R. (1979) Asymptotic behavior of M-estimators
for the linear model. Annals of Statistics, 7.
- Welsch R.E., Dennis J.E. Jr, Gay D.M. (1977) An adaptive nonlinear least
squares algorithm. National Bureau of Economic Research, Inc.
Working Paper No. 196
- Welsch R.E. and Becker R.A. (1975) Robust nonlinear regression using
the Dogleg algorithm. National Bureau of Economic Research, Inc.
Working Paper No. 76.

Yohai
1/1/81
it is correct
Maronna