



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

## **ESTIMACIÓN ROBUSTA EN MODELOS PARCIALMENTE LINEALES GENERALIZADOS**

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

**Lic. Daniela A. Rodriguez**

Directora de tesis: Dra. Graciela L. Boente Boente.

Lugar de trabajo: Instituto de Cálculo, FCEyN, UBA.

Buenos Aires, diciembre de 2007.

## Estimación Robusta en Modelos Parcialmente Lineales Generalizados

En esta tesis, introducimos una nueva clase de estimadores robustos para las componentes paramétricas y noparamétricas bajo dos modelos parcialmente lineales generalizados. En el primero, las observaciones independientes  $(y_i, \mathbf{x}_i, t_i)$ ,  $1 \leq i \leq n$  satisfacen  $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  con  $\mu_i = H(\eta(t_i) + \mathbf{x}_i^T \boldsymbol{\beta})$ , para una función de distribución  $F$  y una función de vínculo  $H$  conocidas, donde  $t_i \in \mathbb{R}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$ . La función  $\eta : \mathbb{R} \rightarrow \mathbb{R}$  y el parámetro  $\boldsymbol{\beta}$  son las cantidades a estimar. Los estimadores robustos se basan en un procedimiento en dos pasos en el que valores grandes de la *deviance* o de los residuos de Pearson se controlan a través de una función de escores acotada. Los estimadores robustos de  $\boldsymbol{\beta}$  resultan ser  $n^{1/2}$ -consistentes y asintóticamente normales. El comportamiento de estos estimadores se compara con el de los estimadores clásicamente usados, a través de un estudio de Monte Carlo. Por otra parte, la función de influencia empírica permite estudiar la sensibilidad de los estimadores.

El modelo generalizado parcialmente lineal de índice simple, generaliza el anterior pues las observaciones independientes son tales que  $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  con  $\mu_i = H(\eta(\boldsymbol{\alpha}^T \mathbf{t}_i) + \mathbf{x}_i^T \boldsymbol{\beta})$ , donde ahora  $\mathbf{t}_i \in \mathbb{R}^q$ ,  $\mathbf{x}_i \in \mathbb{R}^p$  y la función  $\eta : \mathbb{R} \rightarrow \mathbb{R}$  y los parámetros  $\boldsymbol{\beta}$  y  $\boldsymbol{\alpha}$  ( $\|\boldsymbol{\alpha}\| = 1$ ) son desconocidos y se desean estimar. Introducimos dos familias de estimadores robustos que resultan ser consistentes y asintóticamente normales. Calculamos también su función de influencia empírica.

Todas las propuestas dadas mejoran el comportamiento de los estimadores clásicos en presencia de observaciones atípicas.

*Palabras Claves:* Estimadores de Núcleos; Estimadores Robustos; Modelos Parcialmente Lineales; Suavizadores; Tasa de convergencia.

## Robust Estimation in Generalized Partially Linear Models

In this thesis, we introduce a new class of robust estimates for the parametric and nonparametric components under two generalized partially linear model. In the first one, the data  $(y_i, \mathbf{x}_i, t_i)$ ,  $1 \leq i \leq n$ , are modeled by  $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  with  $\mu_i = H(\eta(t_i) + \mathbf{x}_i^T \boldsymbol{\beta})$ , for some known distribution function  $F$  and link function  $H$ , where  $t_i \in \mathbb{R}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$ . The function  $\eta : \mathbb{R} \rightarrow \mathbb{R}$  and the parameter  $\boldsymbol{\beta}$  are unknown and to be estimated. The robust estimators are based on a two step procedure, where large values of the deviance or Pearson residuals are bounded through a score function. It is shown that the estimates of  $\boldsymbol{\beta}$  are root- $n$  consistent and asymptotically normal. Through a Monte Carlo study, we compare the performance of these estimators with that of the classical ones. Besides, through their empirical influence function we study the sensitivity of the estimators.

The generalized partially linear single index model generalizes the previous one since the independent observations are such that  $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  with  $\mu_i = H(\eta(\boldsymbol{\alpha}^T \mathbf{t}_i) + \mathbf{x}_i^T \boldsymbol{\beta})$ , where now  $\mathbf{t}_i \in \mathbb{R}^q$ ,  $\mathbf{x}_i \in \mathbb{R}^p$  and  $\eta : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\alpha}$  ( $\|\boldsymbol{\alpha}\| = 1$ ) are the unknown parameters to be estimated. Two families of robust estimators are introduced which turn out to be consistent and asymptotically normally distributed. Their empirical influence function is also computed.

The robust proposals improve the behavior of the classical ones when outliers are present.

*Keywords and phrases:* Kernel Weights; Partly linear models; Rate of convergence; Robust estimation; Smoothing.

## Agradecimientos

Mucha gente me acompañó y formó parte de esta etapa y a todos ellos quiero agradecerles.

A Graciela, por su generosidad, porque sin su ayuda y su confianza este trabajo no hubiera sido posible.

A mis viejos Diana y Jorge, por apoyarme y estar siempre.

A mis hermanos, Mariano (por hacerme engordar) y Sebastián por ser “mis cables a tierra”.

A la gente del Instituto de Cálculo . En particular, Mariela, Maru, Andrea, Andres y Patu por los almuerzos y charlas compartidas porque junto a ellos todo se hizo más divertido.

A mis amigas de siempre Carolina y Silvana.

A Mariela por tomar siempre el primer mate, por escucharme, por sus consejos y sugerencias, por su amistad.

A los compañeros y amigos de la facu, Agustin, Vivi, Pablo M., Julieta, los Rela, los Perrucci, Santiago, Fernando, Caro...representando a todos con los que compartimos salidas, deportes, cine, asados....

A la familia, mis viejos, mis hermanos, mi sobrino, coco, mis abuelos, la familia de patagones, tíos y primos, los que están cerca, los que están lejos y los que ya no están porque están SIEMPRE.

Al Colo por darme la felicidad de despertarme cada mañana.

# Indice

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Modelo Generalizados Parcialmente Lineales</b>	<b>4</b>
2.1	Introducción . . . . .	4
2.2	Estimación . . . . .	5
2.3	Consistencia . . . . .	7
2.4	Función de influencia empírica . . . . .	9
2.5	Distribución Asintótica . . . . .	11
2.6	Resultados numéricos . . . . .	17
2.6.1	Estudio de Monte Carlo . . . . .	17
2.6.2	Función de Influencia Empírica . . . . .	21
<b>3</b>	<b>Modelo Generalizados Parcialmente Lineales con Índice Simple</b>	<b>24</b>
3.1	Introducción . . . . .	24
3.2	Estimación . . . . .	25
3.2.1	Propuesta 1 . . . . .	25
3.2.2	Propuesta 2 . . . . .	26
3.3	Consistencia . . . . .	27
3.3.1	Consistencia de los estimadores de la primera propuesta . . . . .	29
3.3.2	Consistencia de los estimadores de la segunda propuesta . . . . .	30
3.4	Función de influencia empírica . . . . .	31
3.5	Distribución Asintótica . . . . .	39
3.5.1	Distribución asintótica de los estimadores de la primer propuesta . . . . .	41
3.5.2	Distribución asintótica de los estimadores de la segunda propuesta . . . . .	49

<b>A</b>	<b>Apéndice</b>	<b>55</b>
A.1	Demostración del Lema 2.5.1. . . . .	55
A.2	Demostración del Lema 3.4.3. . . . .	56
A.3	Demostración del Lema 3.5.1. . . . .	59
A.4	Demostración del Lema 3.5.2. . . . .	59
A.5	Demostración del Lema 3.5.5. . . . .	60

# Capítulo 1

## Introducción

La mayoría de los procedimientos estadísticos clásicos están basados en modelos con hipótesis rígidas, tales como errores normales y observaciones equidistribuidas, entre otros. Bajo estas hipótesis se deducen procedimientos óptimos. Por ejemplo, para el caso de regresión con errores normales, el procedimiento óptimo es el de mínimos cuadrados; para modelos paramétricos en general, los procedimientos óptimos clásicos son los estimadores de máxima verosimilitud. Sin embargo, estos métodos son muy sensibles al incumplimiento de las hipótesis que los generaron, tales como la presencia en la muestra de observaciones atípicas. En efecto, pequeñas desviaciones del modelo central, como puede ser el modelo de errores normales o el modelo logístico, manifestadas por unas pocas observaciones atípicas pueden invalidar las conclusiones basadas en estos procedimientos. Los procedimientos estadísticos robustos tienen como objetivo permitir inferencias válidas cuando el modelo no se cumple exactamente y al mismo tiempo ser altamente eficientes bajo el modelo central.

Por otra parte, los modelos clásicamente usados son paramétricos y la suposición es que la muestra de observaciones proviene de una familia paramétrica conocida. En estos casos, el problema es estimar los parámetros desconocidos o hallar tests de hipótesis o intervalos de confianza para los parámetros. Esta suposición puede ser relativamente fuerte porque el modelo paramétrico supuesto puede no ser el correcto si existe alguno (los datos pueden ser tales que no exista una familia paramétrica adecuada que dé un buen ajuste), además los métodos estadísticos desarrollados para un modelo paramétrico particular pueden llevar a conclusiones erróneas cuando se aplican a un modelo ligeramente perturbado (falta de robustez respecto del modelo). Estos problemas llevaron a la tendencia de desarrollar además de procedimientos estadísticos robustos, métodos semiparamétricos para analizar los datos.

El modelo lineal generalizado supone que se tienen observaciones  $(y_i, \mathbf{x}_i, t_i)$ ,  $1 \leq i \leq n$ , independientes y con la misma distribución que  $(y, \mathbf{x}, t) \in \mathbb{R}^{p+2}$ . Se supone además que la distribución condicional de  $y|(\mathbf{x}, t)$  pertenece a una familia exponencial de la forma

$$\exp[y\theta(\mathbf{x}, t) - B(\theta(\mathbf{x}, t)) + C(y)]$$

para funciones  $B$  y  $C$  conocidas donde la media se modela linealmente a través de una función inversa de vínculo,  $g$ , es decir,  $g(\mu(\mathbf{x}, t)) = \alpha_0 + \mathbf{x}^T \boldsymbol{\beta} + \gamma t$  con  $\mu(\mathbf{x}, t) = E(y|(\mathbf{x}, t)) = B'(\theta(\mathbf{x}, t))$ , y  $B'$  la derivada de  $B$  (ver McCullagh y Nelder, 1989).

Para estos modelos, la función de influencia de los estimadores clásicos basados en cuasi-verosimilitud es no acotada. Grandes desviaciones de las variables respuesta respecto de su media, medidas a través de los residuos de Pearson, o datos atípicos en el espacio de las covariables pueden tener una gran influencia en los estimadores de máxima verosimilitud. Esos outliers o potenciales outliers para los modelos lineales generalizados deben ser detectados y controlados por procedimientos robustos como los considerados por Stefanski, Carroll y Ruppert (1986), Künsch, Stefanski y Carroll (1989), Bianco y Yohai (1995), Cantoni y Ronchetti (2001) y Bianco, García Ben y Yohai (2005).

En algunas aplicaciones, el modelo lineal es insuficiente para explicar la relación entre la variable respuesta y las covariables asociadas. Una extensión natural, que sufre de la así llamada *maldición de la dimensión*, consiste en modelar la media noparamétricamente en las covariables. Una estrategia alternativa consiste en modelar linealmente la mayoría de los predictores mientras que uno o un pequeño número se modelan en forma noparamétrica. Esta es la aproximación que se abordará en esta tesis, de modo que la relación vendrá dada por el *modelo semiparamétrico parcialmente lineal generalizado* siguiente:  $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  con  $\mu_i = H(\eta(t_i) + \mathbf{x}_i^\top \boldsymbol{\beta})$  y  $Var(y_i | (\mathbf{x}_i, t_i)) = V(\mu_i)$ , donde  $H = g^{-1}$  es la función de vínculo conocida,  $V$  es conocida,  $\eta(t)$  es la componente no paramétrica desconocida que se supone una función continua y  $\boldsymbol{\beta} \in \mathbb{R}^p$  es un parámetro desconocido que se desea estimar.

Un caso particular de estos modelos lo constituyen los *modelos de regresión parcialmente lineales* en los que la función vínculo es la identidad. Más precisamente, en estos modelos las observaciones verifican  $y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \eta(t_i) + \varepsilon_i$ , donde los errores son independientes e independientes de los efectos  $(\mathbf{x}_i, t_i)$  y donde en muchos casos se supone que  $x_{ij} = \phi_j(t_i) + z_{ij}$ , para  $1 \leq j \leq p$ . Estimadores de tipo splines y propuestas basadas en polinomios locales, en núcleos o en vecinos más cercanos para modelos de regresión parcialmente lineales fueron considerados por varios autores para el caso de diseños fijos y aleatorios (ver, por ejemplo, Härdle, Liang y Gao (2000)).

En el caso de *modelos parcialmente lineal generalizados*, propuestas basadas en núcleos fueron dadas por Severini y Staniswalis (1994) que aplicaron el concepto de perfiles de verosimilitud generalizada introducido por Severini y Wong (1992). En muchos casos, la componente noparamétrica depende de un vector de covariables  $\mathbf{t} \in \mathbb{R}^q$ , sólo a través de una proyección  $\boldsymbol{\alpha}^\top \mathbf{t}$ , donde el parámetro  $\boldsymbol{\alpha}$  es desconocido. Este modelo se conoce como *modelo parcialmente lineal generalizado de índice simple* y fue introducido por Carroll, Fan, Gijbels y Wand (1997) quienes dieron una propuesta basada en cuasi-verosimilitud. En el mismo se supone, entonces, que las observaciones  $(y_i, \mathbf{x}_i, \mathbf{t}_i) \in \mathbb{R}^{p+q+1}$  son tales que  $y_i | (\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$  con  $\mu_i = H(\eta(\boldsymbol{\alpha}^\top \mathbf{t}_i) + \mathbf{x}_i^\top \boldsymbol{\beta})$ ,  $Var(y_i | (\mathbf{x}_i, \mathbf{t}_i)) = V(\mu_i)$ ,  $H = g^{-1}$  es la función de vínculo conocida,  $V$  es conocida,  $\eta(t)$  es la componente no paramétrica desconocida que se supone una función continua y  $\boldsymbol{\beta} \in \mathbb{R}^p$  y  $\boldsymbol{\alpha} \in \mathbb{R}^q$  son parámetros desconocidos que se desea estimar. Observemos que el *modelos parcialmente lineal generalizado* corresponde al caso  $\boldsymbol{\alpha} = \mathbf{1}$  conocido.

Como en el caso lineal, todos estos estimadores son muy sensibles a observaciones atípicas. En el caso de modelos de regresión parcialmente lineales, estimadores robustos basados en M-estimadores locales fueron considerados por Bianco y Boente (2004) mientras que una propuesta robusta para modelos parcialmente lineales generalizados, basada en el concepto de perfiles, fue dada por Boente, He y Zhou (2006).

En el Capítulo 2, pondremos estimadores de dos pasos basados en procedimientos de cuasi-



verosimilitud locales relacionados con los de Carroll, Fan, Gijbels y Wand (1997), con funciones de escores que ponderarán las observaciones atípicas para el *modelo semiparamétrico parcialmente lineal generalizado*. Por otro lado, obtendremos resultados sobre el comportamiento asintótico de las propuestas dadas, más precisamente probaremos que los estimadores propuestos son consistentes y asintóticamente normales. Para estudiar, la robustez de nuestra propuesta, obtendremos la función de influencia empírica y estudiaremos sus propiedades.

En el Capítulo 3, adaptaremos los estimadores propuestos en el capítulo anterior al *modelo parcialmente lineal generalizado de índice simple*. Daremos dos propuestas de estimación robusta para estos modelos más generales. Al igual que en el capítulo anterior estudiaremos propiedades asintóticas de los estimadores resultantes de ambas propuestas y obtendremos la función de influencia empírica. El Apéndice contiene las demostraciones de algunos de los lemas enunciados en las distintas Secciones.

## Capítulo 2

# Modelo Generalizados Parcialmente Lineales

### 2.1 Introducción

Consideremos  $y$ , una variable aleatoria Bernoulli y supongamos que deseamos modelar la probabilidad de ocurrencia del evento  $\{y = 1\}$  a través de una relación lineal de una variable explicativa  $\mathbf{x}$  de dimensión  $p$ . El modelo de regresión lineal habitual, supone que  $\mu(x) = E(y|\mathbf{x} = x)$  puede ser modelada de la siguiente manera  $\mu(\mathbf{x}) = \beta^T \mathbf{x}$ . Claramente, en el caso de variables Bernoulli este modelo no resulta apropiado ya que  $\mu(x) = P(y = 1|\mathbf{x} = x)$  y por lo tanto toma valores entre  $(0, 1)$  mientras que  $\beta^T \mathbf{x}$  podría tomar valores en toda la recta real. Una solución a este problema fue estudiada por Nelder y Wedderburn (1972). La idea consiste en transformar  $E(y|\mathbf{x} = x)$  con una función cuya imagen sea la recta real y luego modelar esta transformación linealmente en las variables explicativas. Por ejemplo en el caso de variables Bernoulli, suele utilizarse la función logit, definida como  $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$ . De esta forma, el modelo resulta

$$\text{logit}(\mu(\mathbf{x})) = \beta^T \mathbf{x}$$

que es comúnmente conocido como modelo de regresión logística.

De un modo más general, podemos definir los *Modelos Lineales Generalizados*. En estos modelos la variable de respuesta  $y$  tiene distribución condicional a  $\mathbf{x} = x$  perteneciente a una familia exponencial con media  $\mu(x)$ . Más precisamente, la función de densidad de  $y|\mathbf{x} = x$  puede escribirse de la siguiente manera

$$f_{y|\mathbf{x}=x}(y) = \exp\{y\theta(x) - B(\theta(x)) + C(y)\}$$

donde  $B$  y  $C$  son funciones conocidas y por lo tanto  $\mu(x) = B'(\theta(x))$  con  $B'$  la derivada de  $B$ . En estos modelos la media es modelada linealmente a través de una función  $g$  cuya inversa  $H = g^{-1}$  es usualmente llamada vínculo. Es decir,

$$g(\mu(\mathbf{x})) = \beta^T \mathbf{x} . \tag{2.1}$$

Como mencionamos en la introducción, en muchas oportunidades suponer esta dependencia lineal resulta insuficiente para explicar la relación subyacente entre  $y$  e  $\mathbf{x}$ . Por tal motivo y para evitar

el problema de “la maldición de la dimensión” que poseen las técnicas puramente no paramétricas es que surgen los *Modelos Parcialmente Lineales Generalizados*. En este enfoque se supone que contamos con un par  $(\mathbf{x}, t)$  de variables explicativas y con una variable de respuesta  $y$  donde la distribución  $y$  condicional a  $(\mathbf{x}, t)$  pertenece como antes a una familia exponencial. Es decir,  $y|(\mathbf{x}, t) \sim F(\cdot, \mu(\mathbf{x}, t))$  con  $\mu(\mathbf{x}, t) = H(\boldsymbol{\beta}^\top \mathbf{x} + \eta(t))$  y  $\text{VAR}(y|(\mathbf{x}, t)) = V(\mu(\mathbf{x}, t))$  donde  $H = g^{-1}$  es la función vínculo conocida. Este modelo combina la fácil interpretación de los modelos lineales con la flexibilidad de los modelos noparamétricos.

Estimadores para  $\boldsymbol{\beta}$  y  $\eta$  en *Modelos Parcialmente Lineales Generalizados* (GPLM) fueron estudiados por diversos autores como Härdle, Liang y Gao (2000), Härdle, Müller, Sperlich y Werwatz (2004). Las procedimientos clásicos de estimación si bien son eficientes bajo el modelo central pueden verse seriamente influenciados por la presencia de un pequeño grupo de observaciones atípicas. Debido a este hecho, Boente, He y Zhou (2006) propusieron una familia de estimadores para el modelo GPLM resistente a datos anómalos en la muestra.

En este Capítulo, proponemos un nuevo método de estimación en modelos GPLM que puede ser descrito en dos pasos. Estudiaremos distintas propiedades como la consistencia y la distribución asintótica y a fin de estudiar la robustez de nuestra propuesta derivaremos la función de influencia empírica.

## 2.2 Estimación

Sean  $(y_i, \mathbf{x}_i, t_i) \in \mathbb{R}^{p+2}$  observaciones independientes tales que  $y_i|(\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  con  $\mu_i = H(\eta(t_i) + \mathbf{x}_i^\top \boldsymbol{\beta})$  y  $\text{VAR}(y_i|(\mathbf{x}_i, t_i)) = V(\mu_i)$ . Sean  $\eta_0(t)$  y  $\boldsymbol{\beta}_0$  los verdaderos parámetros y indiquemos por  $E_0$  la esperanza bajo el verdadero modelo, entonces  $E_0(y|(\mathbf{x}, t)) = H(\eta_0(t) + \mathbf{x}^\top \boldsymbol{\beta}_0)$ . Sea  $w_1 : \mathbb{R}^p \rightarrow \mathbb{R}$  una función de pesos que controla las covariables  $\mathbf{x}$ 's de alta palanca y  $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}$  una función de pérdida. Definamos,

$$S_n(a, \boldsymbol{\beta}, t) = \sum_{i=1}^n W_i(t) \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i) \quad (2.2)$$

$$S(a, \boldsymbol{\beta}, \tau) = E_0 [\rho(y, \mathbf{x}^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}) | t = \tau], \quad (2.3)$$

donde  $W_i(t)$  son pesos que dependen de la cercanía de  $t_i$  a  $t$  y que tomaremos basados en núcleos, o sea,

$$W_i(t) = \frac{K\left(\frac{t-t_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h}\right)}.$$

Supongamos que  $w_1(\cdot)$  y  $\rho(\cdot)$  son tales que  $S(\eta_0(t), \boldsymbol{\beta}_0, t) = \min_{a, \boldsymbol{\beta}} S(a, \boldsymbol{\beta}, t)$ , entonces una forma de estimar  $\eta_0(t)$  y  $\boldsymbol{\beta}_0$  sería estimar  $S(a, \boldsymbol{\beta}, t)$  por  $S_n(a, \boldsymbol{\beta}, t)$  y luego minimizar  $S_n(a, \boldsymbol{\beta}, t)$ . Este procedimiento no da lugar a un estimador de  $\boldsymbol{\beta}_0$  ya que el mínimo en  $\boldsymbol{\beta}$  de  $S_n(a, \boldsymbol{\beta}, t)$  depende de  $t$ , y por otro lado su velocidad de convergencia asintótica no es  $\sqrt{n}$ .

Con el objetivo de obtener estimadores de  $\boldsymbol{\beta}_0$  cuya velocidad de convergencia sea  $\sqrt{n}$ , definimos para cada  $\boldsymbol{\beta} \in \mathbb{R}^p$  y cada función continua  $v : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$F_n(\boldsymbol{\beta}, v) = \frac{1}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + v(t_i)) w_2(\mathbf{x}_i) \quad (2.4)$$

$$F(\boldsymbol{\beta}, v) = E_0 [\rho(y, \mathbf{x}^T \boldsymbol{\beta} + v(t)) w_2(\mathbf{x})] , \quad (2.5)$$

donde  $w_2(\cdot)$  es una función de peso que controla los puntos de alta palanca en las  $\mathbf{x}_i$ . Supondremos que se satisface la consistencia de Fisher, es decir  $F(\boldsymbol{\beta}_0, \eta_0) = \min_{\boldsymbol{\beta}} F(\boldsymbol{\beta}, \eta_0)$  y además que  $\boldsymbol{\beta}_0$  es el único mínimo de  $F(\boldsymbol{\beta}, \eta_0)$ .

Bianco y Yohai (1995) estudiaron condiciones para garantizar la Fisher consistencia para el modelo de regresión logístico, proponiendo funciones de pérdida de la forma,

$$\rho(y, u) = \phi[-\ln f(y, H(u)) + A(y)] + G(H(u)) , \quad (2.6)$$

donde  $\phi$  es una función acotada, continua y no decreciente con derivada continua  $\varphi$ , y  $f(\cdot, s)$  es la densidad de la función de distribución  $F(\cdot, s)$  donde  $y | (\mathbf{x}, t) \sim F(\cdot, H(\eta_0(t) + \mathbf{x}^T \boldsymbol{\beta}_0))$ . La función  $A(y)$  es introducida con el objetivo que en el caso clásico se tenga la deviance y puede ser tomada de la siguiente manera  $A(y) = \ln(f(y, y))$ . Finalmente,  $G$  se usa para garantizar la consistencia de Fisher y satisface,

$$\begin{aligned} G'(s) &= \int \varphi[-\ln f(y, s) + A(y)] f'(y, s) d\mu(y) \\ &= E_s (\varphi[-\ln f(y, s) + A(y)] f'(y, s) / f(y, s)) , \end{aligned}$$

donde  $E_s$  indica que la esperanza es respecto de la distribución  $F(\cdot, s)$ . Croux y Haesbroeck (2002) también estudiaron la consistencia de Fisher para el modelo de regresión logístico proponiendo una nueva familia de funciones de  $\phi$ . Bianco, García Ben y Yohai (2005) estudiaron el caso de familias de distribuciones continuas, con densidad fuertemente unimodal como por ejemplo la distribución log-gama, probando que el término de corrección dado a través de  $G$ , puede ser eliminado.

Las distintas opciones de función de pérdidas, consideradas anteriormente se basan en controlar altos valores de deviance. Cantoni y Ronchetti (2001) introdujeron otra clase de funciones de pérdidas, que controlan valores atípicos de residuos de Pearson. Este enfoque se basa en robustificar la cuasi-verosimilitud. Sea  $r(y, \mu) = (y - \mu) V^{-1/2}(\mu)$  los residuos de Pearson con  $\text{VAR}(y_i | (\mathbf{x}_i, t_i)) = V(\mu_i)$ . Llamemos  $\nu(y, \mu) = V^{-1/2}(\mu) \psi_c(r(y, \mu))$ , donde  $\psi_c$  es una función impar no decreciente y  $c$  una constante de ajuste, como por ejemplo la función de Huber. Entonces la función de pérdida propuesta sería,

$$\rho(y, u) = - \left[ \int_{s_0}^{H(u)} \nu(y, s) ds + G(H(u)) \right] , \quad (2.7)$$

donde  $s_0$  es tal que  $\nu(y, s_0) = 0$  y  $G(s)$ , el término de corrección que garantiza la consistencia de Fisher satisface  $G'(s) = -E_s(\nu(y, s))$ .

La familia de estimadores robustos para modelos parcialmente lineales generalizados que estudiaremos en este capítulo puede definirse como un procedimiento en dos pasos de la siguiente manera.

- **Paso 1:** Para cada valor de  $t$  fijo, sea

$$\left( \tilde{\boldsymbol{\beta}}(t), \hat{\eta}(t) \right) = \underset{a, \boldsymbol{\beta}}{\operatorname{argmin}} S_n(a, \boldsymbol{\beta}, t) \quad (2.8)$$

• **Paso 2:** Definimos el estimador de  $\beta_0$  como

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} F_n(\beta, \hat{\eta}). \quad (2.9)$$

Luego, si la función  $\rho$  es continuamente diferenciable y si llamamos  $\Psi(y, u) = \frac{\partial}{\partial u} \rho(y, u)$ ,  $(\beta_0, \eta_0(t))$  y  $(\tilde{\beta}(t), \hat{\eta}(t))$  resuelven respectivamente las siguientes ecuaciones de estimación  $S^1(a, \beta, t) = 0$  y  $S_n^1(a, \beta, t) = 0$  donde

$$S^1(a, \beta, \tau) = E(\Psi(y, \mathbf{x}^\top \beta + a) w_1(\mathbf{x}) \mathbf{z} | t = \tau) \quad (2.10)$$

$$S_n^1(a, \beta, t) = \sum_{i=1}^n W_i(t) \Psi(y_i, \mathbf{x}_i^\top \beta + a) \mathbf{z}_i w_1(\mathbf{x}_i), \quad (2.11)$$

y  $\mathbf{z} = (1, \mathbf{x}^\top)^\top$ . Por otra parte, el estimador propuesto del parámetro de regresión  $\hat{\beta}$ , es solución de  $F_n^1(\beta, \hat{\eta}) = 0$  y  $\beta_0$  es solución de  $F^1(\beta_0, \eta_0) = 0$  donde

$$F^1(\beta, v) = E(\Psi(y, \mathbf{x}^\top \beta + v(t_i)) w_2(\mathbf{x}) \mathbf{x}) \quad (2.12)$$

$$F_n^1(\beta, v) = \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \beta + v(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i. \quad (2.13)$$

Este enfoque mejora computacionalmente la propuesta de Boente, He y Zhou (2006). En efecto, dichos autores consideran una propuesta basada en perfiles y definen para cada  $\beta \in \mathbb{R}^p$ ,  $\hat{\eta}_\beta(t) = \operatorname{argmin}_a S_n(a, \beta, t)$ . El estimador de  $\beta$  se define entonces como  $\hat{\beta} = \operatorname{argmin}_\beta F_n(\beta, \hat{\eta}_\beta)$ , mientras que el de  $\eta$  resulta ser  $\hat{\eta} = \hat{\eta}_{\hat{\beta}}$ . En nuestra propuesta  $\hat{\eta}$  se calcula una sola vez en el primer paso y por lo tanto, no es necesaria la minimización de  $S_n$  para una grilla de valores de  $\beta$  de modo a obtener  $\hat{\eta}_\beta$  para cada uno de ellos, lo cual facilita el computo del estimador.

## 2.3 Consistencia

En esta Sección mostraremos que los estimadores definidos en (2.8) y (2.9) bajo algunas condiciones de regularidad son consistentes. Supondremos que  $t \in \mathcal{T}$  y que  $\mathcal{T}_0 \subset \mathcal{T}$  es un conjunto compacto. Entonces si  $v$  es una función continua  $v : \mathcal{T}_0 \rightarrow \mathbb{R}$  indiquemos por  $\|v\|_{0, \infty} = \sup_{t \in \mathcal{T}_0} |v(t)|$  y  $\|v\|_\infty = \sup_{t \in \mathcal{T}} |v(t)|$ . Esta última cantidad puede ser no finita.

**C1.**  $\rho(y, a)$  es una función continua y acotada y  $w_1$  y  $w_2$  están acotadas.

**C2.** El núcleo  $K : \mathbb{R} \rightarrow \mathbb{R}$  es una función par, no negativa, acotada y continua con derivada acotada, que satisface  $\int K(u) du = 1$ ,  $\int u^2 K(u) du < \infty$   $|u|K(u) \rightarrow 0$  si  $|u| \rightarrow \infty$ .

**C3.** La sucesión  $h_n$  satisface  $h_n \rightarrow 0$ ,  $nh_n / \log(n) \rightarrow \infty$ .

**C4.**  $f_T$ , la densidad marginal de  $t_1$  es una función acotada en  $\mathcal{T}$  y dado un conjunto compacto  $\mathcal{T}_0 \subset \mathcal{T}$  existe una constante positiva  $A_1(\mathcal{T}_0)$  tal que  $A_1(\mathcal{T}_0) < f_T(t)$  para todo  $t \in \mathcal{T}_0$ .

**C5.**  $S(a, \boldsymbol{\beta}, t)$  satisface la siguiente condición de equicontinuidad: dado  $\epsilon > 0$  existe  $\delta > 0$  tal que

$$|t_1 - t_2| < \delta \text{ y } \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| < \delta, t_i \in \mathcal{T}_0 \text{ y } \boldsymbol{\beta}_i \in \mathcal{K} \text{ para } i = 1, 2 \Rightarrow \sup_{a \in \mathbb{R}} |S(a, \boldsymbol{\beta}_1, t_1) - S(a, \boldsymbol{\beta}_2, t_2)| < \epsilon.$$

**C6.**  $S(a, \boldsymbol{\beta}, t)$  es una función continua y  $\eta_0(t)$  es continua y diferenciable.

**Lema 2.3.1.** Sea  $\mathcal{K} \subset \mathbb{R}^p$  y  $\mathcal{T}_0 \subset \mathcal{T}$  compactos tales que  $\mathcal{T}_\delta \subset \mathcal{T}$  donde  $\mathcal{T}_\delta$  es la clausura de una  $\delta$  vecindad de  $\mathcal{T}_0$ . Supongamos que se satisfacen **C1** - **C6** y la siguiente condición,

i)  $K$  es de variación acotada

ii) La familia de funciones  $\mathcal{F} = \{f(y, \mathbf{x}) = \rho(y, \mathbf{x}^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, a \in \mathbb{R}\}$  tiene número de cubrimiento  $N(\epsilon, \mathcal{F}, L^1(\mathcal{Q})) \leq A\epsilon^{-W}$ , para cualquier  $\mathcal{Q}$  y  $0 < \epsilon < 1$

Luego se tiene,  $\|\hat{\eta} - \eta_0\|_{0,\infty} \xrightarrow{a.s.} 0$ .

DEMOSTRACIÓN DEL LEMA 2.3.1.

En forma análoga al Teorema 3.1 de Boente, He y Zhou (2006) se tiene que

$$\sup_{\substack{\boldsymbol{\beta} \in \mathbb{R}^p \\ a \in \mathbb{R}}} \|S_n(a, \boldsymbol{\beta}, \cdot) - S(a, \boldsymbol{\beta}, \cdot)\|_{0,\infty} \xrightarrow{a.s.} 0.$$

Luego el Lema A.1 de Carrol, Fan, Gijbels y Wand (1997) implica que  $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\hat{\eta} - \eta_0\|_{0,\infty} \xrightarrow{a.s.} 0$   
□

**Teorema 2.3.2.** Sea  $\hat{\boldsymbol{\beta}}$  el argumento que minimiza  $F_n(\boldsymbol{\beta}, \hat{\eta})$  donde  $F_n(\boldsymbol{\beta}, \eta)$  está definida en (2.4).

Supongamos que  $\Psi(y, u) = \frac{\partial}{\partial u} \rho(y, u)$  es acotada y que  $\hat{\eta}$  satisface

$$\|\hat{\eta} - \eta_0\|_{0,\infty} \xrightarrow{a.s.} 0. \quad (2.14)$$

Supongamos que se satisface **C1** y la siguiente condición; la familia de funciones  $\mathcal{H} = \{f_{\boldsymbol{\beta}}(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^\top \boldsymbol{\beta} + \eta_0(t)) w_2(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p\}$  tiene número de cubrimiento finito,  $N(\epsilon, \mathcal{H}, L^1(\mathcal{Q})) \leq \infty$ , para cualquier  $\mathcal{Q}$  y  $0 < \epsilon < 1$  entonces,

a)  $\sup_{\boldsymbol{\beta} \in \mathbb{R}^p} |F_n(\boldsymbol{\beta}, \hat{\eta}) - F(\boldsymbol{\beta}, \eta_0)| \xrightarrow{a.s.} 0$  para cualquier conjunto compacto  $\mathcal{K}$

b) Si,  $F(\boldsymbol{\beta}, \eta_0)$  tiene un único mínimo en  $(\boldsymbol{\beta}_0, \eta_0)$  y  $\lim_{\|\boldsymbol{\beta}\| \rightarrow \infty} F(\boldsymbol{\beta}, \eta_0) > F(\boldsymbol{\beta}_0, \eta_0)$ , entonces  $\hat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$ .

DEMOSTRACIÓN DEL TEOREMA 2.3.2.

a) Dado  $\epsilon > 0$ , sea  $\mathcal{T}_0$  un conjunto compacto tal que  $P(t_1 \notin \mathcal{T}_0) < \epsilon$  entonces,

$$\sup_{\boldsymbol{\beta} \in \mathbb{R}^p} |F_n(\boldsymbol{\beta}, \hat{\eta}) - F_n(\boldsymbol{\beta}, \eta_0)| \leq \|\hat{\eta} - \eta_0\|_{0,\infty} \|w_2\|_\infty \|\Psi\|_\infty + 2\|w_2\|_\infty \|\rho\|_\infty \frac{1}{n} \sum_{i=1}^n I_{(t_i \notin \mathcal{T}_0)}$$

y usando la condición (2.14) y la ley fuerte de los grandes números se tiene que,

$$\sup_{\beta \in \mathbb{R}^p} |F_n(\beta, \hat{\eta}) - F_n(\beta, \eta_0)| \xrightarrow{a.s.} 0.$$

Luego, por la hipótesis sobre el número de cubrimiento para la clase  $\mathcal{H}$  del Teorema 2.4.1 de Van der Vaart y Wellner, obtenemos que

$$\sup_{\beta \in \mathbb{R}^p} |F_n(\beta, \eta_0) - F(\beta, \eta_0)| \xrightarrow{a.s.} 0.$$

b) Sea  $\hat{\beta}_k$  una subsucesión de  $\hat{\beta}$  tal que  $\hat{\beta}_k \rightarrow \beta^*$ , supongamos sin perder generalidad que  $\hat{\beta} \rightarrow \beta^*$ . Si  $\|\beta^*\| < \infty$ , de a) y de la continuidad de  $F(\beta, \eta_0)$  deducimos que  $F_n(\hat{\beta}, \hat{\eta}) - F(\beta^*, \eta_0) \xrightarrow{a.s.} 0$  y  $F_n(\beta_0, \hat{\eta}) - F(\beta_0, \eta_0) \xrightarrow{a.s.} 0$ . Como  $F_n(\beta_0, \hat{\eta}) \geq F_n(\hat{\beta}, \hat{\eta})$  y  $F$  tiene un único mínimo en  $(\beta_0, \eta_0)$ , fácilmente se obtiene  $\beta^* = \beta_0$ .

Supongamos que  $\|\beta^*\| = \infty$  luego igual que antes tenemos que  $F_n(\hat{\beta}, \hat{\eta}) - F(\hat{\beta}, \eta_0) \xrightarrow{a.s.} 0$ ,  $F_n(\beta_0, \hat{\eta}) - F(\beta_0, \eta_0) \xrightarrow{a.s.} 0$  y  $F_n(\beta_0, \hat{\eta}) \geq F_n(\hat{\beta}, \hat{\eta})$  entonces  $\lim_{n \rightarrow \infty} F(\hat{\beta}, \eta_0) - F(\beta_0, \eta_0) \leq 0$ , absurdo luego  $\|\beta^*\| < \infty \square$

## 2.4 Función de influencia empírica

Uno de los objetivos de la estadística robusta es dar procedimientos de estimación que sean menos sensibles a datos atípicos. La función de influencia empírica, ver Tukey (1970), es una medida de robustez, que refleja el comportamiento del estimador cuando cambiamos un elemento de la muestra por una nueva observación que no proviene del modelo original. Mallows (1974) consideró una función de influencia para muestras finitas que corresponde a la función de influencia definida por Hampel (1974) (ver Hampel, Ronchetti, Rousseeuw y Stahel, 1982) evaluada en la distribución empírica de los datos. Para el caso de regresión no paramétrica, Tamine (2002) propuso una función de influencia suavizada para estimadores no paramétricos de regresión basados en núcleos, pero bajo el supuesto que el parámetro de suavizado  $h$  es fijo. Por otra parte, Manchester (1996) introdujó un método gráfico para reflejar la sensibilidad de un estimador bajo contaminación.

Para medir la influencia de una observación atípica, consideraremos una versión empírica de la función de influencia definida de la siguiente manera. Sea  $\{(y_i, \mathbf{x}_i, t_i)\}_{1 \leq i \leq n}$  una muestra aleatoria y consideremos  $\hat{\beta}$  el estimador del parámetro de regresión basado en esa muestra. Observemos que si llamamos  $P_n$  a la medida empírica que le asigna masa puntual  $1/n$  a cada elemento de la muestra resulta que  $\hat{\beta} = \hat{\beta}(P_n)$ . Por otro lado, consideremos  $P_{n,j}$  a la medida empírica que le asigna masa  $\frac{1-\varepsilon}{n}$  a cada observación  $(y_i, \mathbf{x}_i, t_i)$ ,  $1 \leq i \leq n$   $i \neq j$ , y a la observación  $(y_j, \mathbf{x}_j, t_j)$  masa  $\frac{(n-1)\varepsilon+1}{n}$ . Indiquemos  $\hat{\beta}_{j,\varepsilon}$  al estimador del parámetro de regresión para esta nueva muestra. Entonces, la función de influencia empírica (EIF) para  $\hat{\beta}$  es:

$$\text{EIF}_j(\hat{\beta}) = \lim_{\varepsilon \rightarrow 0} \frac{\hat{\beta}_{j,\varepsilon} - \hat{\beta}}{\varepsilon}$$

En adelante sólo notaremos a  $\hat{\beta}_{j,\varepsilon}$  por  $\hat{\beta}_\varepsilon$  omitiendo por simplicidad el subíndice  $j$ . Observemos que  $\text{EIF}_j(\hat{\beta}) = \left. \frac{\partial}{\partial \varepsilon} \hat{\beta}_\varepsilon \right|_{\varepsilon=0}$ . Luego, con el objetivo de obtener  $\text{EIF}_j(\hat{\beta})$  para el estimador propuesto,

calculemos  $\left. \frac{\partial}{\partial \varepsilon} \tilde{\boldsymbol{\beta}}_\varepsilon \right|_{\varepsilon=0}$ . En primer lugar, notemos que  $\tilde{\boldsymbol{\beta}}_\varepsilon$  satisface

$$\frac{(1-\varepsilon)}{n} \sum_{i=1}^n \Psi \left( y_i, \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}_\varepsilon + \hat{\eta}_{\varepsilon,j}(t_i) \right) w_2(\mathbf{x}_i) \mathbf{x}_i + \varepsilon \Psi \left( y_j, \mathbf{x}_j^\top \tilde{\boldsymbol{\beta}}_\varepsilon + \hat{\eta}_{\varepsilon,j}(t_j) \right) w_2(\mathbf{x}_j) \mathbf{x}_j = 0 \quad (2.15)$$

donde  $(\tilde{\boldsymbol{\beta}}_\varepsilon(t), \hat{\eta}_{\varepsilon,j}(t))$  satisfacen

$$\begin{aligned} 0 &= \frac{1-\varepsilon}{n} \sum_{i=1}^n K \left( \frac{t_i - t}{h} \right) \Psi \left( y_i, \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}_\varepsilon(t) + \hat{\eta}_{\varepsilon,j}(t) \right) w_1(\mathbf{x}_i) \mathbf{z}_i \\ &+ \varepsilon K \left( \frac{t_j - t}{h} \right) \Psi \left( y_j, \mathbf{x}_j^\top \tilde{\boldsymbol{\beta}}_\varepsilon(t) + \hat{\eta}_{\varepsilon,j}(t) \right) w_1(\mathbf{x}_j) \mathbf{z}_j \end{aligned} \quad (2.16)$$

es decir  $\tilde{\boldsymbol{\beta}}_\varepsilon(t)$  y  $\hat{\eta}_{\varepsilon,j}(t)$  corresponden a los estimadores obtenidos en el primer paso con la nueva muestra. Entonces derivando respecto de  $\varepsilon$  la ecuación (2.15) y evaluando en  $\varepsilon = 0$  tenemos,

$$\begin{aligned} 0 &= -\frac{1}{n} \sum_{i=1}^n \Psi \left( y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i) \right) w_2(\mathbf{x}_i) \mathbf{x}_i \\ &+ \frac{1}{n} \sum_{i=1}^n \chi \left( y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i) \right) w_2(\mathbf{x}_i) \mathbf{x}_i [\mathbf{x}_i^\top \text{EIF}_j(\hat{\boldsymbol{\beta}}) + \text{EIF}_j(\hat{\eta})(t_i)] \\ &+ \Psi \left( y_j, \mathbf{x}_j^\top \hat{\boldsymbol{\beta}} + \hat{\eta}(t_j) \right) w_2(\mathbf{x}_j) \mathbf{x}_j \\ &= \frac{1}{n} \sum_{i=1}^n \chi \left( y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i) \right) w_2(\mathbf{x}_i) \mathbf{x}_i [\mathbf{x}_i^\top \text{EIF}_j(\hat{\boldsymbol{\beta}}) + \text{EIF}_j(\hat{\eta})(t_i)] \\ &+ \Psi \left( y_j, \mathbf{x}_j^\top \hat{\boldsymbol{\beta}} + \text{EIF}_j(\hat{\eta})(t_j) \right) w_2(\mathbf{x}_j) \mathbf{x}_j \end{aligned} \quad (2.17)$$

donde  $\text{EIF}_j(\hat{\eta})(t) = \left. \frac{\partial}{\partial \varepsilon} \hat{\eta}_{\varepsilon,j}(t) \right|_{\varepsilon=0}$  y  $\chi(y, u) = \frac{\partial}{\partial u} \Psi(y, u)$ . Luego para calcular la influencia empírica de  $\hat{\boldsymbol{\beta}}$  es necesario obtener  $\text{EIF}_j(\hat{\eta})(t_i)$  para  $1 \leq i \leq n$ . Entonces derivamos respecto de  $\varepsilon$  y evaluamos en  $\varepsilon = 0$  la ecuación (2.16) y obtenemos,

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n K \left( \frac{t_i - t}{h} \right) \chi \left( y_i, \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}(t) + \hat{\eta}(t) \right) w_1(\mathbf{x}_i) \mathbf{z}_i \mathbf{z}_i^\top \left( \left. \frac{\partial}{\partial \varepsilon} \tilde{\boldsymbol{\beta}}_\varepsilon(t) \right|_{\varepsilon=0} \right) \\ &+ K \left( \frac{t_j - t}{h} \right) \Psi \left( y_j, \mathbf{x}_j^\top \tilde{\boldsymbol{\beta}}(t) + \hat{\eta}(t) \right) w_1(\mathbf{x}_j) \mathbf{z}_j \end{aligned}$$

luego  $\text{EIF}_j(\hat{\eta})(t)$  es el primer elemento del vector

$$v_j = - \left\{ \frac{1}{n} \sum_{i=1}^n K \left( \frac{t_i - t}{h} \right) \chi \left( y_i, \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}(t) + \hat{\eta}(t) \right) w_1(\mathbf{x}_i) \mathbf{z}_i \mathbf{z}_i^\top \right\}^{-1} K \left( \frac{t_j - t}{h} \right) \Psi \left( y_j, \mathbf{x}_j^\top \tilde{\boldsymbol{\beta}}(t) + \hat{\eta}(t) \right) w_1(\mathbf{x}_j) \mathbf{z}_j \quad (2.18)$$

Finalmente de (2.17) despejamos  $\text{EIF}_j(\hat{\boldsymbol{\beta}})$  obteniendo



$$\begin{aligned} \text{EIF}_j(\widehat{\boldsymbol{\beta}}) &= - \left\{ \frac{1}{n} \sum_{i=1}^n \chi \left( y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \widehat{\eta}(t_i) \right) w_2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T \right\}^{-1} \left\{ \Psi \left( y_j, \mathbf{x}_j^T \widehat{\boldsymbol{\beta}} + \widehat{\eta}(t_j) \right) w_2(\mathbf{x}_j) \mathbf{x}_j \right. \\ &\quad \left. + \frac{1}{n} \sum_{i=1}^n \chi \left( y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \widehat{\eta}(t_i) \right) w_2(\mathbf{x}_i) \mathbf{x}_i \text{EIF}_j(\widehat{\eta})(t_i) \right\} \end{aligned} \quad (2.19)$$

donde  $\text{EIF}_j(\widehat{\eta})(t_i)$  es el primer elemento de  $v_j$  definido en (2.18).

En la Sección 2.6.1 describimos dos ejemplos de modelos particulares y para una muestra de cada uno de estos modelos calculamos la función de influencia empírica. En cada caso consideramos una grilla de puntos con diferentes valores de  $(y_0, x_0, t_0)$  y calculamos la función de influencia empírica para cada uno de los puntos de la grilla.

## 2.5 Distribución Asintótica

En esta Sección mostraremos que bajo ciertas hipótesis, el estimador propuesto para el parámetro de regresión es asintóticamente normal. Para simplificar la notación llamaremos:

$$\chi(y, a) = \frac{\partial}{\partial u} \Psi(y, u) \quad \text{y} \quad \chi_1(y, a) = \frac{\partial^2}{\partial u^2} \Psi(y, u)$$

Sea  $\boldsymbol{\Sigma}$  la matriz  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_2^T + \boldsymbol{\Sigma}_3$  con

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= E_0 \left\{ \Psi^2(y_1, \mathbf{x}_1^T \boldsymbol{\beta}_0 + \eta_0(t_1)) w_2^2(\mathbf{x}_1) \mathbf{x}_1 \mathbf{x}_1^T \right\} \\ \boldsymbol{\Sigma}_2 &= E_0 \left\{ G_1(t_1) f_T(t_1) \Psi(y_1, \mathbf{x}_1^T \boldsymbol{\beta}_0 + \eta_0(t_1)) w_2(\mathbf{x}_1) \gamma(t_1) \mathbf{x}_1^T \right\} \\ \boldsymbol{\Sigma}_3 &= E_0 \left\{ G_1^2(t_1) \gamma(t_1) \gamma(t_1)^T f_T^2(t_1) \right\} \end{aligned}$$

donde  $\gamma(\tau) = E(\chi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t)) w_2(\mathbf{x}) \mathbf{x} f_T^{-1}(t) | t = \tau)$  y  $G_i(\tau)$  el primer elemento de  $A(t)^{-1} \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(\tau)) w_1(\mathbf{x}_i) \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix}$  con  $A(t) = E \left( \chi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t)) w_1(\mathbf{x}) \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{x}^T \end{pmatrix} | t = \tau \right)$ . Supondremos además que  $\mathcal{T}$  es compacto.

**N1.**  $\eta_0$  y  $\widehat{\eta}(t)$  son continuas y derivables con derivadas continuas, tales que  $n^{\frac{1}{4}} \|\widehat{\eta} - \eta_0\|_\infty \xrightarrow{p} 0$  y  $\|\widehat{\eta}' - \eta_0'\|_\infty \xrightarrow{p} 0$

**N2.** Las funciones  $\Psi$ ,  $\chi$ ,  $\chi_1$ ,  $w_2$ ,  $\psi_1(\mathbf{x}) = \mathbf{x} w_1(\mathbf{x})$  y  $\psi_2(\mathbf{x}) = \mathbf{x} w_2(\mathbf{x})$  son continuas y acotadas.

**N3.** La matriz  $\mathbf{A} = E_0[\chi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t)) \mathbf{x} \mathbf{x}^T w_2(\mathbf{x})]$  y  $A(t)$  definida anteriormente son no singulares.

**N4.** La matriz  $\boldsymbol{\Sigma}$  es definida positiva.

**N5.**  $E_0 \{ \Psi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t)) | (\mathbf{x}, t) \} = 0$ .

**N6.**  $E_0(w_2(\mathbf{x}) \|\mathbf{x}\|^2) < \infty$ .

**N7.**  $f_T$  la densidad marginal de  $t_1$  y su derivada son funciones continuas y acotadas. Además,  $\inf_{t \in \mathcal{T}} f_T(t) > 0$ .

**N8.** a)  $m_1(t, \tau) = E(G_1(t_2)|t_1 = t, t_2 = \tau)$ ,  $m_1'(t, \tau) = \frac{\partial}{\partial \tau} m_1(t, \tau)$  y  $m_1''(t, \tau) = \frac{\partial^2}{\partial \tau^2} m_1(t, \tau)$  son funciones continuas y acotadas.

b)  $m_2(\tau_1, \tau_2, \tau_3, \tau_4) = E([G_i(t_j) - G_i(t_i)][G_s(t_l) - G_s(t_s)]|t_i = \tau_1, t_j = \tau_2, t_s = \tau_3, t_l = \tau_4)$  y  $m_2'(\tau_1, \tau_2, \tau_3, \tau_4) = \frac{\partial}{\partial \tau_1} m_2(\tau_1, \tau_2, \tau_3, \tau_4)$  son funciones continuas y acotadas.

c)  $\gamma(t)$  tiene derivada continua.

**N9.** El núcleo  $K : \mathbb{R} \rightarrow \mathbb{R}$  es una función par, no negativa, acotada y Lipschitz de orden uno, que satisface  $\int K(u)du = 1$ ,  $\int uK(u)du = 0$  y  $\int u^2K(u)du < \infty$ .

Para demostrar que el estimador propuesto para el parámetro de regresión es asintóticamente normal necesitaremos el siguiente lema que demostraremos en el Apéndice.

**Lema 2.5.1.** Supongamos que se satisface **N1**, **N2**, **N3**, **N7**,  $K$  es un núcleo simétrico y  $E(w_1(\mathbf{x})\|\mathbf{x}\|^3) < \infty$ . Si  $nh^4 \rightarrow 0$  y  $\lim_{n \rightarrow \infty} \frac{nh^2}{\log(1/h)} = +\infty$  y  $n^{1/4}\|\tilde{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0\| \xrightarrow{p} 0$  entonces se tiene

$$\sup_{t \in \mathcal{T}} |\hat{\eta}(t) - \eta_0(t) - \frac{1}{nf_T(t)} \sum_{j=1}^n K_h(t - t_j) G_j(t)| = o_p(n^{-1/2})$$

donde  $G_j(t)$  está definido anteriormente.

**Teorema 2.5.2.** Supongamos que  $t_i$  son variables aleatorias y su distribución tiene soporte en un conjunto compacto  $\mathcal{T}$ . Supongamos que se satisfacen **N1** a **N9**, si  $nh^4 \rightarrow 0$  entonces si  $\hat{\boldsymbol{\beta}}$  es una solución de (2.13) consistente y se satisface la conclusión del Lema 2.5.1, se tiene que

$$n^{\frac{1}{2}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{A}^{-1} \boldsymbol{\Sigma} (\mathbf{A}^{-1})^{\top}),$$

donde  $\mathbf{A}$  y  $\boldsymbol{\Sigma}$  están definidas anteriormente.

**DEMOSTRACIÓN DEL TEOREMA 2.5.2.** Sea  $\hat{\boldsymbol{\beta}}$  solución de  $F_n^1(\boldsymbol{\beta}, \hat{\eta}) = 0$  definida en (2.13). Haciendo un desarrollo de Taylor de orden uno se tiene

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^{\top} \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i \\ &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^{\top} \boldsymbol{\beta}_0 + \hat{\eta}(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i + \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^{\top} \boldsymbol{\beta}^* + \hat{\eta}(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^{\top} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \end{aligned}$$

con  $\boldsymbol{\beta}^*$  un punto intermedio entre  $\boldsymbol{\beta}_0$  y  $\hat{\boldsymbol{\beta}}$ . Entonces, es suficiente mostrar que,

$$\begin{aligned} \mathbf{A}_n &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^{\top} \boldsymbol{\beta}^* + \hat{\eta}(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^{\top} \xrightarrow{p} \mathbf{A} \\ \mathbf{B}_n &= \frac{\sqrt{n}}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^{\top} \boldsymbol{\beta}_0 + \hat{\eta}(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}$$

donde  $\mathbf{A}$  y  $\mathbf{\Sigma}$  están anteriormente.  $\mathbf{A}_n$  puede escribirse como

$$\begin{aligned}\mathbf{A}_n &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}^* + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top + \frac{1}{n} \sum_{i=1}^n \chi_1(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}^* + \xi_{ni}) w_2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top (\hat{\eta}(t_i) - \eta_0(t_i)) \\ &= \mathbf{A}_n^{(1)} + \mathbf{A}_n^{(2)}\end{aligned}$$

donde  $\xi_{ni}$  es un punto intermedio entre  $\hat{\eta}(t_i)$  y  $\eta_0(t_i)$ . Por **N1**, **N2** y **N6** se tiene que  $\mathbf{A}_n^{(2)} \xrightarrow{a.s.} 0$ . **N2** implica que  $\lambda(\boldsymbol{\beta}) = E(\chi(y, \mathbf{x}'\boldsymbol{\beta} + \eta_0(t)) w_2(\mathbf{x}) \mathbf{x} \mathbf{x}^\top)$  es una función continua de  $\boldsymbol{\beta}$ , del hecho que  $\boldsymbol{\beta}^* \xrightarrow{p} \boldsymbol{\beta}_0$  se deduce que

$$E[\chi(y, \mathbf{x}^\top \boldsymbol{\beta}^* + \eta_0(t)) w_2(\mathbf{x}) \mathbf{x} \mathbf{x}^\top] - E[\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t)) w_2(\mathbf{x}) \mathbf{x} \mathbf{x}^\top] \xrightarrow{p} 0.$$

Luego, es suficiente mostrar que,

$$\frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}^* + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top - E[\chi(y, \mathbf{x}^\top \boldsymbol{\beta}^* + \eta_0(t)) w_2(\mathbf{x}) \mathbf{x} \mathbf{x}^\top] \xrightarrow{p} 0.$$

Consideremos la siguiente clase de funciones  $\mathcal{H} = \{\chi(y, \mathbf{x}^\top \boldsymbol{\beta} + \eta_0(t)) w_2(\mathbf{x}) \mathbf{x} \mathbf{x}^\top \mid \boldsymbol{\beta} \in \mathcal{K}\}$  con  $\mathcal{K}$  un entorno compacto de  $\boldsymbol{\beta}_0$ . Usando los mismos argumentos que en el Lema 1 de Bianco y Boente (2002) se tiene que  $\mathbf{A}_n^{(1)} \xrightarrow{p} \mathbf{A}$ .

Para obtener la distribución asintótica de  $\hat{\boldsymbol{\beta}}$  es suficiente analizar el comportamiento de  $\mathbf{B}_n$ . Sea,

$$\mathbf{C}_n = \frac{\sqrt{n}}{n} \sum_{i=1}^n [\Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i + \gamma(t_i) G_i(t_i) f_T(t_i)]$$

donde  $G_i(\tau)$  el primer elemento de  $\left[ E \left( \chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t)) w_1(\mathbf{x}) \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{x}^\top \end{pmatrix} \mid t = \tau \right) \right]^{-1} \Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(\tau)) w_1(\mathbf{x}_i) \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix}$  y  $\gamma(\tau) = E(\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t)) w_2(\mathbf{x}) \mathbf{x} f_T^{-1}(t) \mid t = \tau)$ .

De, **N5** se sigue que  $\mathbf{C}_n$  tiene distribución asintótica normal con matriz de covarianza  $\mathbf{\Sigma}$ . Luego basta demostrar que  $\mathbf{B}_n - \mathbf{C}_n \xrightarrow{p} 0$ . Es fácil ver que  $\mathbf{B}_n - \mathbf{C}_n = \mathbf{B}_n^{(1)} + \mathbf{B}_n^{(2)}$  donde

$$\begin{aligned}\mathbf{B}_n^{(1)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i [\hat{\eta}(t_i) - \eta_0(t_i)] - \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma(t_i) G_i(t_i) f_T(t_i) \\ &= \mathbf{B}_n^{(3)} - \mathbf{B}_n^{(4)} \\ \mathbf{B}_n^{(2)} &= \frac{1}{2n} \sum_{i=1}^n \chi_1(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \xi_{in}) w_2(\mathbf{x}_i) \mathbf{x}_i \left( n^{\frac{1}{4}} [\hat{\eta}(t_i) - \eta_0(t_i)] \right)^2\end{aligned}$$

donde  $\xi_{in}$  es un punto intermedio entre  $\hat{\eta}(t_i)$  y  $\eta_0(t_i)$ . De **N1** y **N2** se sigue que  $\mathbf{B}_n^{(2)} \xrightarrow{p} 0$ . Luego, para finalizar bastará con probar que  $\mathbf{B}_n^{(1)} \xrightarrow{p} 0$ . Por el Lema 2.5.1. se tiene

$$\begin{aligned}\mathbf{B}_n^{(3)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i [\hat{\eta}(t_i) - \eta_0(t_i)] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i\end{aligned}$$

$$\begin{aligned}
& \times \left\{ \frac{1}{nhf_T(t_i)} \sum_{j=1}^n G_j(t_i) K\left(\frac{t_i - t_j}{h}\right) + o_p(n^{-1/2}) \right\} \\
& = h^{-1} n^{-\frac{3}{2}} \sum_{i=1}^n \sum_{j=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i f_T^{-1}(t_i) G_j(t_i) K\left(\frac{t_i - t_j}{h}\right) + o_p(1) \\
& = \mathbf{B}_n^{(5)} + o_p(1)
\end{aligned}$$

Entonces basta ver que  $\mathbf{B}_n^{(5)} - \mathbf{B}_n^{(4)} \xrightarrow{p} 0$ .

Sea  $R(y_i, \mathbf{x}_i, t_i) = \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t_i)) w_2(\mathbf{x}_i) \mathbf{x}_i f_T^{-1}(t_i)$  entonces,

$$\begin{aligned}
\mathbf{B}_n^{(5)} - \mathbf{B}_n^{(4)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right) R(y_j, x_j, t_j) G_i(t_j) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma(t_i) G_i(t_i) f_T(t_i) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right) [R(y_j, x_j, t_j) G_i(t_j) - \gamma(t_i) G_i(t_i)] \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right) - f_T(t_i) \right] \gamma(t_i) G_i(t_i) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right) [R(y_j, x_j, t_j) - \gamma(t_j)] G_i(t_j) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right) [\gamma(t_j) - \gamma(t_i)] G_i(t_j) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right) \gamma(t_i) [G_i(t_j) - G_i(t_i)] \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t_i - t_j}{h}\right) - f_T(t_i) \right] \gamma(t_i) G_i(t_i) \\
&= \mathbf{B}_n^{(6)} + \mathbf{B}_n^{(7)} + \mathbf{B}_n^{(8)} + \mathbf{B}_n^{(9)}
\end{aligned}$$

Por **N5**,  $E(G_1(t_1)|t_1) = 0$  luego, usando **N7**,  $\sup_{t \in \mathcal{T}} \left| \frac{1}{nh} \sum_{j=1}^n K\left(\frac{t - t_j}{h}\right) - f_T(t) \right| \xrightarrow{a.s.} 0$  y el Lema 6.6.7 de Härdle, Liang y Gao (2000) obtenemos  $\mathbf{B}_n^{(9)} \xrightarrow{p} 0$ .

Para obtener la convergencia de  $\mathbf{B}_n^{(8)}$  calculamos su esperanza y su varianza. Consideremos  $m_1(t, \tau)$  definida en **N8**, observemos que  $m_1(t, t) = 0$ , entonces

$$\begin{aligned}
E(\mathbf{B}_n^{(8)}) &= \frac{n(n-1)}{n\sqrt{n}} E\left( f_T(t) \gamma(t) m_1(t_1, t_2) \frac{1}{h} K\left(\frac{t_1 - t_2}{h}\right) \right) \\
&= \frac{n-1}{\sqrt{n}} \int_{\mathcal{T}_0} \gamma(t) f_T(t) \int_{\mathcal{T}_0} m_1(t, \tau) \frac{1}{h} K\left(\frac{t - \tau}{h}\right) f_T(\tau) d\tau dt.
\end{aligned}$$

Usando **N7**, **N8** y **N9** se tiene,

$$\begin{aligned}
& \int m_1(t, \tau) \frac{1}{h} K\left(\frac{t-\tau}{h}\right) f_T(\tau) d\tau = \int m_1(t, t-uh) K(u) f_T(t-uh) du \\
& = \int [-uhm_1'(t, t) + u^2 h^2 m_1''(t, \xi_1)] K(u) [f_T(t) - uhf_T'(\xi_2)] du \\
& = h^2 [m_1'(t, t) \int u^2 K(u) f_T'(\xi_2) du + f_T(t) \int u^2 K(u) m_1''(t, \xi_1) du - h \int u^3 K(u) m_1''(t, \xi_1) f_T'(\xi_2) du] \\
& = O(h^2)
\end{aligned}$$

entonces por **N2** como  $\mathcal{T}$  es compacto  $E(\mathbf{B}_n^{(8)}) = \frac{n-1}{\sqrt{n}} O(h^2) = O(n^{1/2} h^2) \rightarrow 0$ . Para el cálculo de la varianza, el desarrollo es similar. Observemos primero que

$$\mathbf{B}_n^{(8)} = \frac{1}{n^{3/2}} \sum_{i \neq j} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \gamma(t_i) [G_i(t_j) - G_i(t_i)] = \frac{1}{n^{3/2}} \sum_{i \neq j} Z_{ij}$$

entonces

$$\begin{aligned}
\text{Cov}(\mathbf{B}_n^{(8)}, \mathbf{B}_n^{(8)}) & = \frac{1}{n^3} \sum_{i \neq j} \text{Cov}(Z_{ij}, Z_{ij}) + \frac{1}{n^3} \sum_{i \neq j} \sum_{l \neq s} \text{Cov}(Z_{ij}, Z_{ls}) \\
& = \frac{n(n-1)}{n^3} \text{Cov}(Z_{12}, Z_{12}) + \frac{n^2(n-1)}{n^3} [\text{Cov}(Z_{12}, Z_{13}) + \text{Cov}(Z_{12}, Z_{32})]
\end{aligned}$$

ahora

$$\begin{aligned}
\text{tr}(\text{Cov}(Z_{12}, Z_{12})) & \leq \frac{1}{h^2} E \left( K^2 \left( \frac{t_1 - t_2}{h} \right) \text{tr}(\gamma(t_1) \gamma^T(t_1)) [G_1(t_1) - G_1(t_2)]^2 \right) \\
& = \frac{1}{h^2} E \left( K^2 \left( \frac{t_1 - t_2}{h} \right) \text{tr}(\gamma(t_1) \gamma^T(t_1)) [G_1(t_1) - G_1(t_2)]^2 \right) \\
& = \frac{1}{h^2} \int K^2 \left( \frac{u-v}{h} \right) \text{tr}(\gamma(u) \gamma^T(u)) [G_1(u) - G_1(v)]^2 f_T(u) f_T(v) du dv \\
& = \frac{-1}{h} \int K^2(z) \text{tr}(\gamma(u) \gamma^T(u)) [G_1(u) - G_1(u-zh)]^2 f_T(u) f_T(u-zh) du dz \\
& = \frac{-1}{h} \int K^2(z) \text{tr}(\gamma(u) \gamma^T(u)) [G_1'(\xi) zh]^2 f_T(u) f_T(u-zh) du dz
\end{aligned}$$

obteniendo así que  $\text{tr}(\text{Cov}(Z_{12}, Z_{12})) = O(h)$  y por lo tanto  $n(n-1) \text{tr}(\text{Cov}(Z_{12}, Z_{12}))/n^3 \rightarrow 0$  cuando  $n \rightarrow \infty$ . Del mismo modo calculemos  $\text{Cov}(Z_{12}, Z_{13})$ .

$$\begin{aligned}
E(Z_{12}) & = \int \frac{1}{h} K\left(\frac{u-v}{h}\right) \gamma(u) [G_1(u) - G_1(v)] f_T(u) f_T(v) du dv \\
& = - \int K(z) \gamma(u) [G_1(u) - G_1(u-zh)] f_T(u) f_T(u-zh) du dz \\
& = O(h)
\end{aligned}$$

$$\begin{aligned}
\text{tr}(E(Z_{12} Z_{13}^T)) & = \int \frac{1}{h^2} K\left(\frac{u-v}{h}\right) K\left(\frac{u-z}{h}\right) \text{tr}(\gamma(u) \gamma^T(u)) [G_1(u) - G_1(v)] [G_1(u) - G_1(z)] \\
& \quad f_T(u) f_T(v) f_T(u) f_T(z) du dv dz
\end{aligned}$$

$$\begin{aligned}
&= \int K(z)K(v)tr(\gamma(u)\gamma^T(u))[G_1(u) - G_1(u - hv)][G_1(u) - G_1(u - hz)] \\
&\quad f_T(u)f_T(v)f_T(u - hv)f_T(u - hz) du dv dz \\
&= \int K(z)K(v)tr(\gamma(u)\gamma^T(u))[G_1'(\xi_1)hv][G_1'(\xi_2)hz]f_T(u)f_T(v)f_T(u - hv)f_T(u - hz)dudvdz \\
&= O(h^2)
\end{aligned}$$

obteniendo entonces  $tr(\text{COV}(Z_{12}, Z_{13})) = O(h^2)$  y de forma análoga  $tr(\text{COV}(Z_{12}, Z_{32})) = O(h^2)$  y por lo tanto tenemos,

$$tr(\text{COV}(\mathbf{B}_n^{(8)}, \mathbf{B}_n^{(8)})) = \frac{n-1}{n^2}tr(\text{VAR}(Z_{12})) + \frac{n-1}{n}tr[\text{COV}(Z_{12}, Z_{13}) + \text{COV}(Z_{12,r}, Z_{32})] = O(hn^{-1} + h^2)$$

La convergencia de  $\mathbf{B}_n^{(6)}$  y  $\mathbf{B}_n^{(7)}$  la obtenemos de la misma forma que para  $\mathbf{B}_n^{(8)}$ . Calculemos en primer lugar la esperanza,

$$\begin{aligned}
E(\mathbf{B}_n^{(7)}) &= \frac{n(n-1)}{n\sqrt{n}}E\left(\frac{1}{h}K\left(\frac{t_1-t_2}{h}\right)[\gamma(t_2) - \gamma(t_1)]G_1(t_2)\right) \\
&= \frac{(n-1)}{\sqrt{n}}E\left(\frac{1}{h}K\left(\frac{t_1-t_2}{h}\right)[\gamma(t_1) - \gamma(t_2)]E(G_1(t_2)|t_1, t_2)\right) \\
&= \frac{(n-1)}{\sqrt{n}}E\left(\frac{1}{h}K\left(\frac{t_1-t_2}{h}\right)[\gamma(t_1) - \gamma(t_2)]m_1(t_1, t_2)\right) \\
&= \frac{n-1}{\sqrt{n}}\int\frac{1}{h}K\left(\frac{u-v}{h}\right)[\gamma(u) - \gamma(v)]m_1(u, v)f_T(u)f_T(v)dudv
\end{aligned}$$

haciendo un cambio de variables, realizando un desarrollo de Taylor y usando el hecho que  $m_1(u, u) = 0$  obtenemos,

$$\begin{aligned}
E(\mathbf{B}_n^{(7)}) &= \frac{n-1}{\sqrt{n}}\int\frac{1}{h}K\left(\frac{u-v}{h}\right)[\gamma(u) - \gamma(v)]m_1(u, v)f_T(u)f_T(v)dudv \\
&= \frac{n-1}{\sqrt{n}}\int f_T(u)\int K(z)\gamma'(\xi_2)hz\left[m_1(u, u) - hzm_1'(u, z) + \frac{h^2z^2}{2}m_1''(u, \xi_2)\right]f_T(u - hz)dzdu
\end{aligned}$$

entonces por **N8** y como  $\gamma$  es derivable y su derivada continua, obtenemos que  $E(\mathbf{B}_n^{(7)}) = O(n^{1/2}h^2)$ . Con los mismo argumentos obtenemos que la traza de la matriz de varianza converge a 0 por lo tanto  $\mathbf{B}_n^{(7)} \xrightarrow{p} 0$ .

Por último calculemos la esperanza de  $\mathbf{B}_n^{(6)}$ ,

$$\begin{aligned}
E(\mathbf{B}_n^{(6)}) &= \frac{n(n-1)}{n\sqrt{n}}E\left(\frac{1}{h}K\left(\frac{t_1-t_2}{h}\right)[R(y_2, \mathbf{x}_2, t_2) - \gamma(t_2)]G_1(t_2)\right) \\
&= \frac{n-1}{\sqrt{n}}E\left(\frac{1}{h}K\left(\frac{t_1-t_2}{h}\right)G_1(t_2)E(R(y_2, \mathbf{x}_2, t_2) - \gamma(t_2)|\mathbf{x}_1, y_1, t_1, t_2)\right) \\
&= \frac{n-1}{\sqrt{n}}E\left(\frac{1}{h}K\left(\frac{t_1-t_2}{h}\right)G_1(t_2)[E(R(y_2, \mathbf{x}_2, t_2)|t_2) - \gamma(t_2)]\right) = 0
\end{aligned}$$

Para la varianza los términos de covarianzas son cero por el mismo argumento que la esperanza. Entonces tenemos que  $tr(\text{VAR}(\mathbf{B}_n^{(6)})) = O((nh)^{-1})$ .  $\square$

## 2.6 Resultados numéricos

### 2.6.1 Estudio de Monte Carlo

Esta Sección contiene los resultados de un estudio de simulación preliminar diseñado con el objetivo de evaluar el procedimiento de estimación propuesto en este Capítulo. En este estudio se consideraron dos modelos logísticos parcialmente lineales generalizados. Con el objetivo de comparar su performance bajo contaminación, para cada uno de ellos se calcularon los estimadores robustos propuestos, indicados por MO en tablas y gráficos y los estimadores de cuasi-verosimilitud, indicados por QL, propuestos por Carroll, Fan, Gijbels y Wand (1997), que son una alternativa a los de perfiles propuestos por Severini y Staniswalis (1994).

Los estimadores MO elegidos fueron aquellos que controlan la deviance y fueron calculados usando la función de escores propuesta por Croux y Haesbroeck (2002) con constante  $c = 0.5$ . Las funciones de pesos  $w_1$  y  $w_2$  utilizadas para controlar las variables de alta palanca fueron elegidas iguales y del siguiente modo,

$$w_1^2(x_i) = w_2^2(x_i) = \frac{1}{1 + (x_i - M_n)^2}$$

donde  $M_n$  corresponde a la mediana de las  $x_i$ 's. Las covariables  $x_i$  se tomaron univariadas.

Para el **Estudio 1** se generaron 1000 muestras de tamaño 100 y se tomó una ventana de  $h = 0.3$ . En el **Estudio 2**, como en Boente, He y Zhou (2006) y para comparar los resultados obtenidos en este estudio con los obtenidos en el **Estudio 3** de dichos autores, se generaron 100 muestras de tamaño 200 con parámetro de suavizado  $h = 0.1$ . En ambos casos se utilizó el función núcleo de Epanechnikov,  $K(t) = \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}t^2\right) I_{(-\sqrt{5}, \sqrt{5})}(t)$ . Los modelos considerados fueron los siguientes,

**Estudio 1:**  $x \sim \mathcal{U}(-1, 1)$ ,  $t \sim \mathcal{U}(0, 1)$  e  $y|(x, t) \sim Bi(1, p(x, t))$  donde  $\log(p(x, t)/(1 - p(x, t))) = 2x + \sin(2\pi t)$ . Es decir,  $\beta_0 = 2$  y  $\eta_0(t) = \sin(2\pi t)$ .

**Estudio 2:** Primero se generaron pares de una distribución normal bivariada,  $(x_i, t_i) \sim N((0, 1/2), \Sigma)$  donde se truncó la variable  $t \in [1/4, 3/4]$  con

$$\Sigma = \begin{pmatrix} 1 & 1/(6\sqrt{3}) \\ 1/(6\sqrt{3}) & 1/36 \end{pmatrix}.$$

La variable de respuesta fue generada de la siguiente manera

$$y_i = \begin{cases} 1, & \beta_0 x_i + \eta_0(t_i) + \epsilon_i \geq 0 \\ 0, & \beta_0 x_i + \eta_0(t_i) + \epsilon_i < 0 \end{cases}$$

donde  $\beta_0 = 2$ ,  $\eta_0(t) = 2 \sin(4\pi t)$  y  $\epsilon_i$  son independientes con distribución logística.

Para cada conjunto de datos generado siguiendo los modelos de los estudios 1 y 2 se consideró la siguiente contaminación. En primer lugar, se generó una muestra de variables uniformes independientes,  $u_i \sim \mathcal{U}(0, 1)$  para  $1 \leq i \leq n$ . A continuación, se consideró el siguiente esquema de contaminación

$$x_i = \begin{cases} x_i & \text{si } u_i \leq 0.90 \\ \text{una nueva observación de una } N(10, 1) & \text{si } u_i > 0.90 \end{cases}$$

$$y_i = \begin{cases} y_i & \text{si } u_i \leq 0.90 \\ \text{una nueva observación de una } Bi(1, 0.05) & \text{si } u_i > 0.90 \end{cases}$$

Las Tablas 1 y 2 contienen los resultados obtenidos para cada estudio de simulación. Para los estimadores de  $\beta_0$ , se consideraron las siguientes medidas resumen: media, desvío estándar y mediana calculadas sobre las replicaciones. Para estudiar el comportamiento de los estimadores de la componente no paramétrica  $\eta_0$ , que indicamos por  $\hat{\eta}_{QL}$  y  $\hat{\eta}_{MO}$ , consideramos el error cuadrático medio (ECM) y el error cuadrático mediano (ECMed) definidos por

$$ECM(\hat{\eta}) = \frac{1}{n} \sum_{i=1}^n [\hat{\eta}(t_i) - \eta(t_i)]^2 \quad ECMed(\hat{\eta}) = \text{mediana}(|\hat{\eta}(t_i) - \eta(t_i)|^2).$$

En las Figuras 1 y 2 se presentan los boxplots de los estimadores del parámetro de regresión  $\beta_0$  para comparar su comportamiento. Cada figura contiene 4 boxplots que corresponden a los estimadores clásicos y robustos obtenidos bajo el modelo sin contaminación y bajo contaminación que indicamos por  $C_0$  y  $C_1$ , respectivamente. En todos los casos la línea roja corresponde al verdadero valor del parámetro regresión que para ambos estudios es  $\beta_0 = 2$ .

Tabla 1: Estimadores de  $\beta_0$

	Estudio 1		Estudio 2		
	$\hat{\beta}_{QL}$	$\hat{\beta}_{MO}$	$\hat{\beta}_{QL}$	$\hat{\beta}_{MO}$	
media	2.0813	2.0945	1.9322	1.8849	$C_0$
desvío	0.5300	0.5863	0.3725	0.4059	
mediana	2.045	2.0366	1.8915	1.8330	
ecm	0.2875	0.3526	0.1433	0.1780	
media	-0.1595	1.9238	-0.1700	1.7737	$C_1$
desvío	0.0680	0.7656	0.0643	0.4573	
mediana	-0.1653	1.9429	-0.1819	1.6540	
ecm	4.6690	0.5919	4.7130	0.2603	

Tabla 2: Estimadores de  $\eta_0$ .

	Estudio 1				
	$ECM(\hat{\eta}_{QL})$	$ECMed(\hat{\eta}_{QL})$	$ECM(\hat{\eta}_{MO})$	$ECMed(\hat{\eta}_{MO})$	
media	0.2678	0.1291	0.3074	0.1364	$C_0$
media	0.2347	0.1494	0.4267	0.1497	$C_1$
	Estudio 2				
	$ECM(\hat{\eta}_{QL})$	$ECMed(\hat{\eta}_{QL})$	$ECM(\hat{\eta}_{MO})$	$ECMed(\hat{\eta}_{MO})$	
media	0.5504	0.1356	0.8009	0.1413	$C_0$
media	18.0007	0.2063	1.2699	0.2247	$C_1$



Figura 1: Boxplot del parámetro de regresión del **Estudio 1**. (a) corresponde a  $C_0$  y (b) a  $C_1$

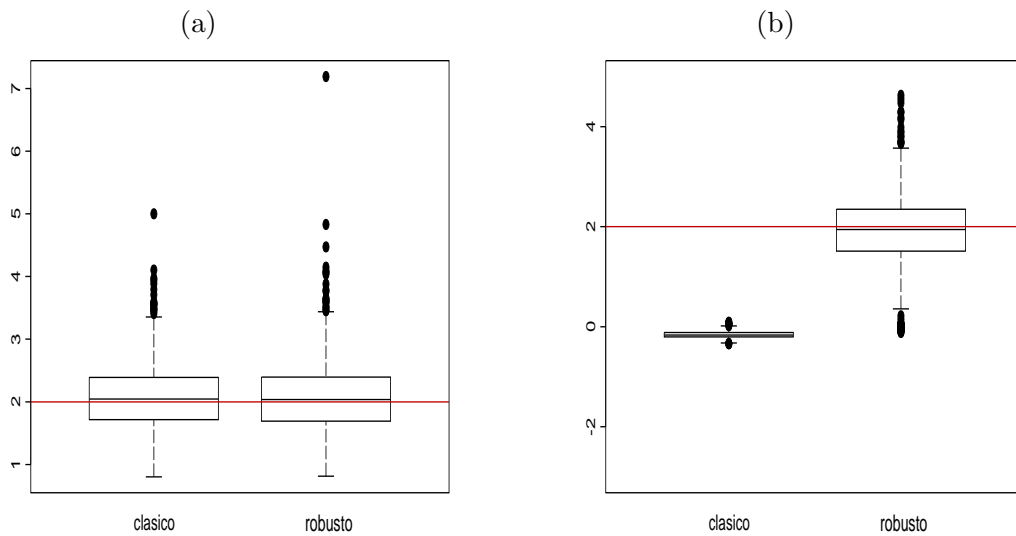
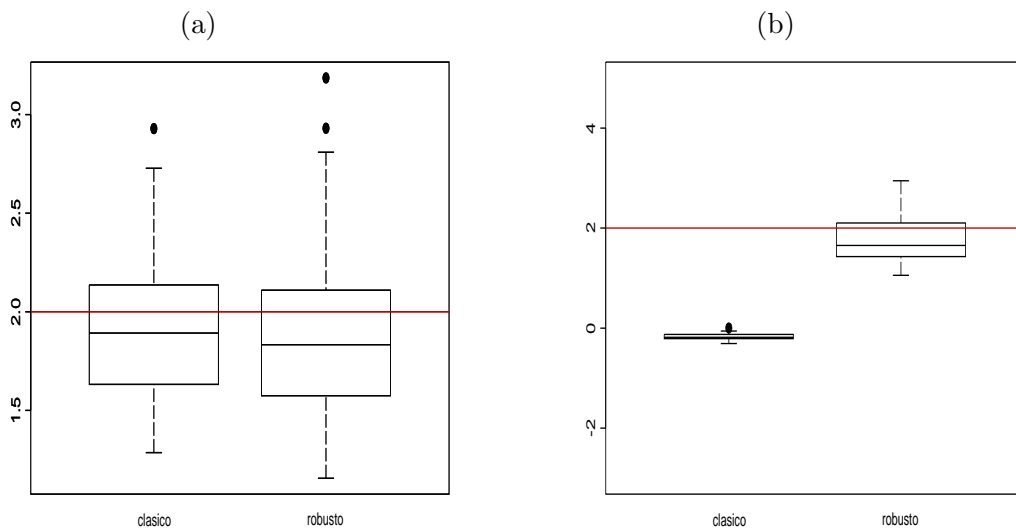


Figura 2: Boxplot del parámetro de regresión del **Estudio 2**. (a) corresponde a  $C_0$  y (b) a  $C_1$



Tanto la Tabla 1 como la Figura 1 muestran el pésimo comportamiento de  $\hat{\beta}_{QL}$  ante la presencia de datos atípicos con alta palanca. El estimador robusto de  $\beta_0$  es más estable ante esta perturbación y su comportamiento bajo  $C_0$  es similar al del estimador de cuasi-verosimilitud siendo su error cuadrático medio un 20% mayor. Por otra parte, bajo contaminación el error cuadrático del estimador  $\hat{\beta}_{QL}$  es casi 8 veces mayor que el del robusto ya que es igual a 4.6680 mientras que el estimador robusto  $\hat{\beta}_{MO}$  tiene un error cuadrático de 0.5919.

Respecto del **Estudio 2**, el estimador de  $\beta_0$  presenta más sesgo que en el **Estudio 1**. Este hecho ya se observaba en el trabajo de Boente, He y Zhou (2006) y puede explicarse en parte,

porque el diseño para las variables  $t$  en el **Estudio 2** no es equiespaciado. Por lo tanto, una ventana adaptativa podría ser beneficiosa en este modelo. Sin embargo, como en el **Estudio 1** el error cuadrático medio de  $\hat{\beta}_{\text{MO}}$  es sólo un 24% mayor que el del estimador clásico bajo  $C_0$  mientras que bajo  $C_1$ , la situación se invierte siendo el error cuadrático de  $\hat{\beta}_{\text{QL}}$  casi seis veces mayor que el de  $\hat{\beta}_{\text{MO}}$ . Vale la pena mencionar que en ambos estudios, el desvío estándar del estimador de cuasi-verosimilitud se reduce notablemente, concentrando los valores del estimador cerca del valor estimado de modo que un test para el parámetro de regresión rechazaría que el mismo fuese igual a 2. Finalmente, respecto de los estimadores introducidos en Boente, He y Zhou (2006), los estimadores del parámetro  $\beta$  introducidos en esta tesis muestran una ligera mejoría tanto en sesgo como en varianza respecto de los introducidos por esos autores tanto bajo  $C_0$  como bajo  $C_1$ . Este beneficio va en detrimento de la estimación de  $\eta$  ya que su error cuadrático es mucho mayor (más del doble en el caso robusto) que el de la propuesta dada por Boente, He y Zhou (2006). Debemos mencionar que si bien la propuesta de esta tesis considera a  $\eta$  como un parámetro de ruido y el objetivo se centra en la estimación del parámetro de regresión  $\beta$ , un método de estimación que combinaría las ventajas de ambos procedimientos podría ser el siguiente:

- Calcule el estimador de  $\beta$  como se definió en (2.8) y (2.9). Sea  $\hat{\beta}$  el estimador resultante.
- Calcule  $\eta_{\hat{\beta}}(t) = \operatorname{argmin}_a S_n(a, \hat{\beta}, t)$

Este procedimiento tiene las mismas propiedades asintóticas que las vistas anteriormente pero permite mejorar el error cuadrático del estimador de  $\eta$ . Por otra parte, no es tan costoso computacionalmente como la propuesta de Boente, He y Zhou (2006). La tabla siguiente contiene los resultados para cada estudio de simulación de el procedimiento descrito anteriormente que permite para mejorar la estimación de  $\eta_0$

Tabla 3: Estimadores de  $\eta_0$  mejorados.

	Estudio 1				
	ECM( $\hat{\eta}_{\text{QL}}$ )	ECMed( $\hat{\eta}_{\text{QL}}$ )	ECM( $\hat{\eta}_{\text{MO}}$ )	ECMed( $\hat{\eta}_{\text{MO}}$ )	
media	0.2106	0.1244	0.2214	0.1321	$C_0$
media	0.2343	0.1535	0.2478	0.1458	$C_1$
	Estudio 2				
	ECM( $\hat{\eta}_{\text{QL}}$ )	ECMed( $\hat{\eta}_{\text{QL}}$ )	ECM( $\hat{\eta}_{\text{MO}}$ )	ECMed( $\hat{\eta}_{\text{MO}}$ )	
media	0.2603	0.136	0.2768	0.1462	$C_0$
media	1.1046	0.2245	0.593	0.2048	$C_1$

Para concluir el estudio queremos mencionar que, en el Estudio 1, con respecto a la estimación de  $\eta$ , todos los procedimientos parecen ser estables, debido a que la magnitud de anomalía  $y$  es muy limitada en este caso. Sin embargo, en el Estudio 2, el estimador clásico de  $\eta$  muestra un pésimo comportamiento bajo  $C_1$ . Este fenómeno puede explicarse por el hecho de que  $\eta(t)$  representa localmente la ordenada al origen en el modelo logístico y por lo tanto, es muy sensible a datos de alta palanca, como ha sido ampliamente estudiado en el modelo de regresión logístico (ver por ejemplo, Maronna, Martin y Yohai (2006))

### 2.6.2 Función de Influencia Empírica

En esta Sección, se grafica la función de influencia empírica en los dos modelos descritos en la Sección 2.6.1. Es decir, para cada uno de esos modelos se generó una muestra  $(y_i, x_i, t_i)$ ,  $1 \leq i \leq n$ , de tamaño  $n = 100$ . Se tomó una grilla de valores  $(y_0, x_0, t_0)$  definida por  $y_0 = 0$  e  $y_0 = 1$ , ya que  $y_i|(x_i, t_i)$  es binomial,  $t_0 \in \{0, 0.1, 0.2, \dots, 1\}$  y  $x_0$  toma valores en una grilla de tamaño 20 equiespaciada en  $[-10, 10]$ .

En consecuencia, para  $y_0 = 1$  e  $y_0 = 0$  tenemos una grilla de 220 puntos  $(x_0, t_0)$  y para cada uno de estos puntos calculamos la función de influencia empírica dada por (2.19) y obtenida en la Sección 2.4. Este influencia se calculó tanto para los estimadores clásicos correspondientes a la cuasi-verosimilitud (con  $\Psi \equiv \text{id}$ ) como para el caso robusto con la función de escores propuesta por Croux y Haesbroeck (2002) como en la Sección 2.6.1. Se tomó una ventana igual a 0.3 en ambos casos y las funciones de pesos y el núcleo de la misma forma que en el estudio de simulación. Notese que el mayor tamaño de la ventana se debe a que en este caso se tomaron muestras de tamaño 100 en ambos estudios. Las Figuras 3, 4, 5 y 6 muestran los resultados obtenidos.

Figura 3: Función de influencia empírica para el **Estudio 1** correspondientes a  $y_0 = 1$  (a) Estimador clásico (b) Estimador robusto.

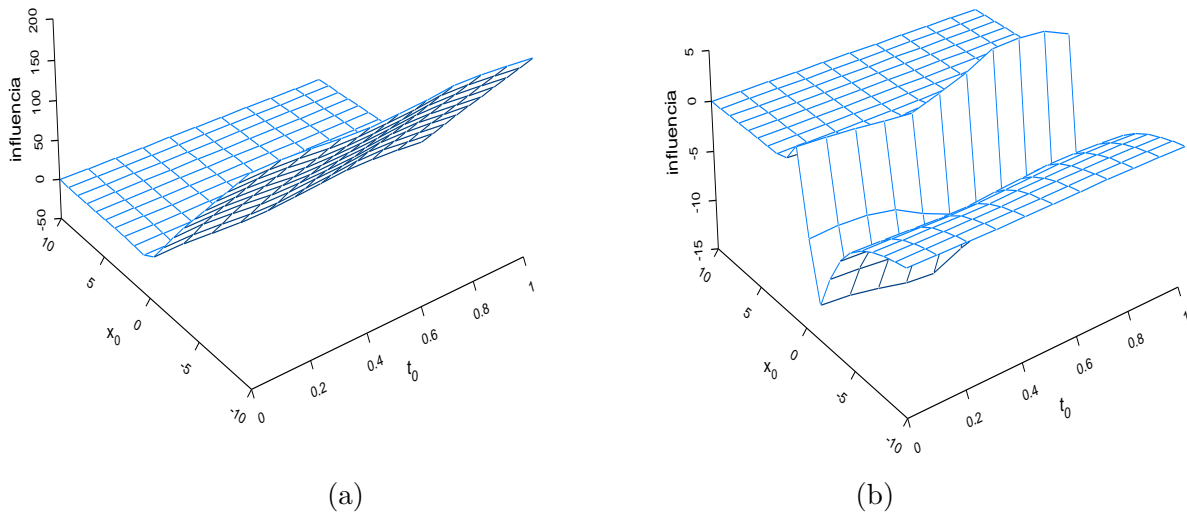
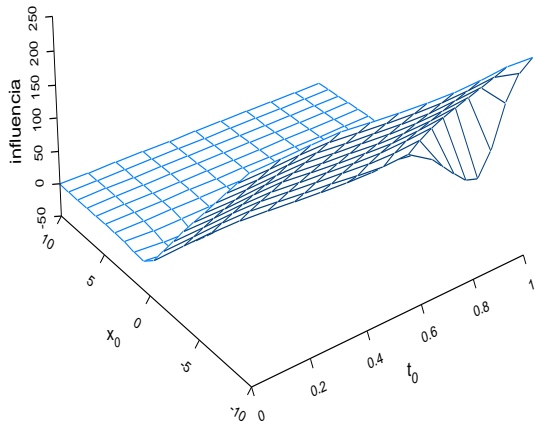
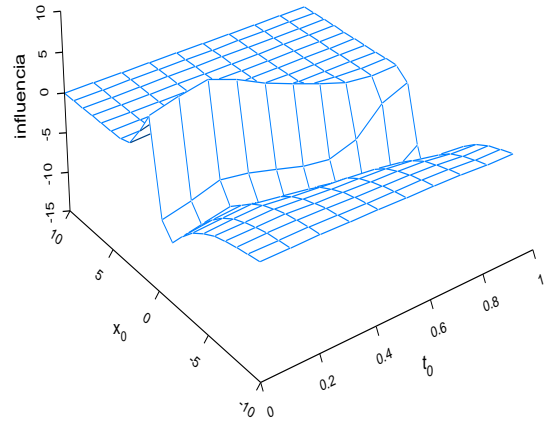


Figura 4: Función de influencia empírica para el **Estudio 2** correspondientes a  $y_0 = 1$  (a) Estimador clásico (b) Estimador robusto.

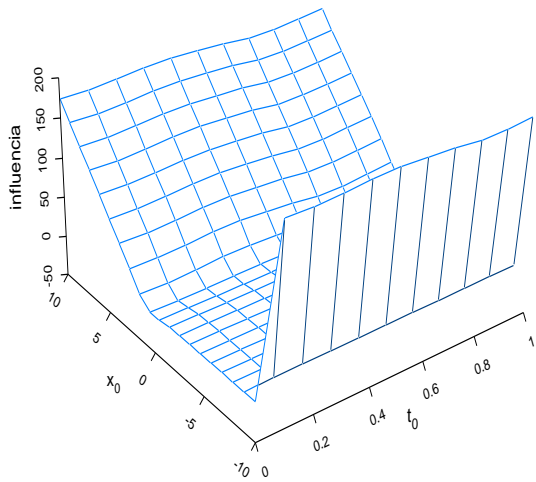


(a)

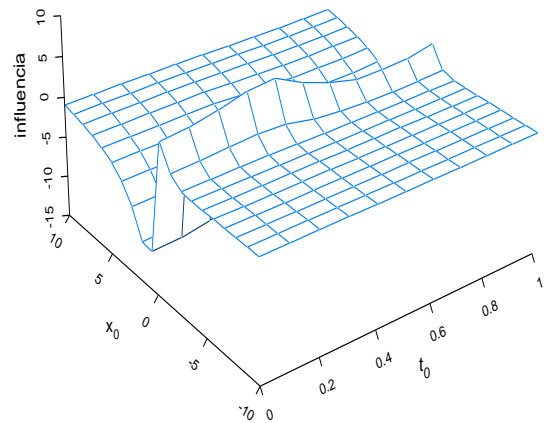


(b)

Figura 5: Función de influencia empírica para el **Estudio 1** correspondientes a  $y_0 = 0$  (a) Estimador clásico (b) Estimador robusto.

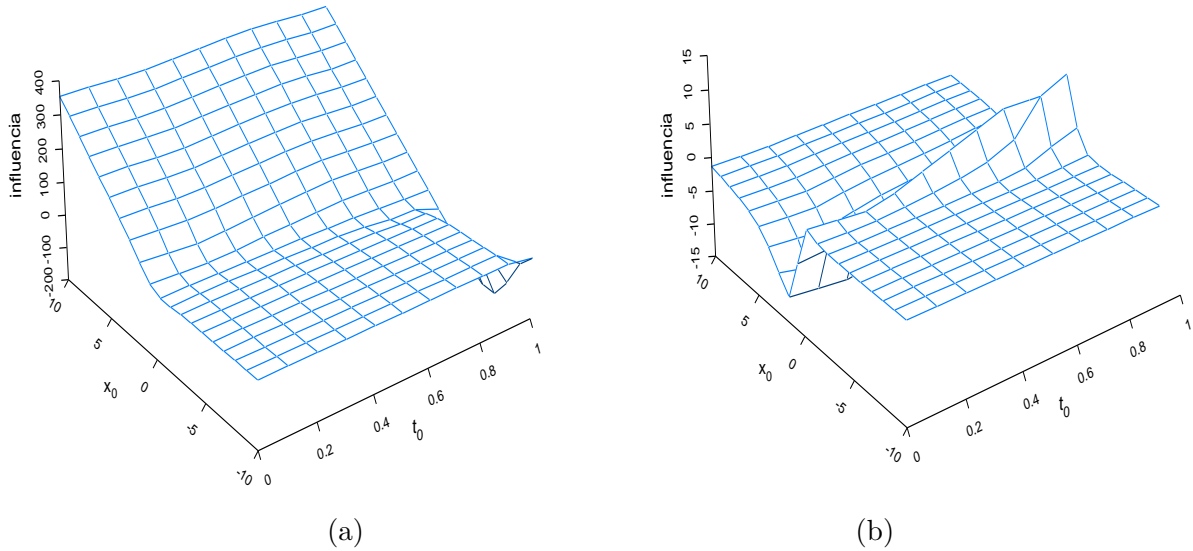


(a)



(b)

Figura6 : Función de influencia empírica para el **Estudio 2** correspondientes a  $y_0 = 0$  (a) Estimador clásico (b) Estimador robusto.



Las Figuras 3, 4, 5 y 6 muestran que la función de influencia del estimador clásico no es acotada, como es bien sabido. Valores negativos de valor absoluto grande son muy influyentes para la estimación de  $\beta$  cuando  $y_0 = 1$ . Por otra parte, el estimador robusto tiene función de influencia empírica acotada. En ambos casos, el máximo valor absoluto de la influencia se alcanza para valores de  $x$  cercanos a  $-2.5$ , siendo esta influencia negativa. Por otra parte, para ambos casos, los puntos cercanos a  $0.5$  son los valores en los que la función de influencia empírica es máxima. Por lo tanto, los valores cercanos a  $-2.5$  son los que se espera producirán el mayor sesgo, siendo éste negativo pero acotado para contaminaciones infinitesimales. El mismo análisis se puede hacer cuando  $y_0 = 0$ .

## Capítulo 3

# Modelo Generalizados Parcialmente Lineales con Índice Simple

### 3.1 Introducción

En el Capítulo anterior se estudiaron *los modelos parcialmente lineales generalizados*, que buscan explicar una variable de respuesta  $y$  con variables explicativas  $(\mathbf{x}, t)$  donde  $\mathbf{x} \in \mathbb{R}^p$  y  $t \in \mathbb{R}$  a través de una relación lineal en  $\mathbf{x}$  y de forma no paramétrica en  $t$ . Una extensión natural que es estudiar este modelo cuando la variable  $t$  toma valores solo en la recta real sino que posee dimensión mayor que 1,  $\mathbf{t} \in \mathbb{R}^q$ . Claramente, uno podría imitar el método propuesto anteriormente, considerando un estimador no paramétrico pero ahora basado en una muestra  $\mathbf{t}_1, \dots, \mathbf{t}_n$  de dimensión  $q$ . Sin embargo, es bien sabido que esta propuesta sufriría del problema conocido como “maldición de la dimensión”, ya que necesitaríamos de un muestra de gran tamaño para garantizar que el suavizado local que se realiza en los métodos de estimación no paramétricos como vecinos más cercanos o núcleos puedan ser llevados a cabo con éxito.

Una propuesta alternativa para solucionar este problema fue estudiada por Carroll, Fan, Gijbels y Wand (1997) que consideraron un modelo donde la variable de respuesta  $y$  depende en forma no paramétrica de  $\mathbf{t}$  a través de una proyección; este modelo es conocido como *Modelo parcialmente lineal generalizado con índice simple* (GPLMI). Más precisamente, este nuevo modelo supone que contamos con  $(y, \mathbf{x}, \mathbf{t})$  un vector aleatorio tal que  $y \in \mathbb{R}$ ,  $\mathbf{x} \in \mathbb{R}^p$  y  $\mathbf{t} \in \mathbb{R}^q$  donde la distribución de  $y$  condicional  $(\mathbf{x}, \mathbf{t})$  pertenece a una familia exponencial con la media modelada a través de una proyección del vector  $\mathbf{t}$ . Es decir,  $y | (\mathbf{x}, \mathbf{t}) \sim F(\cdot, \mu(\mathbf{x}, \mathbf{t}))$  con  $\mu(\mathbf{x}, \mathbf{t}) = H(\boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\alpha}^T \mathbf{t})$  y  $\text{VAR}(y | (\mathbf{x}, \mathbf{t})) = V(\mu(\mathbf{x}, \mathbf{t}))$  donde  $H = g^{-1}$  es la función vínculo conocida. Claramente, para poder identificar a  $\boldsymbol{\alpha}$  y  $\eta$  es necesario contar con la restricción que  $\|\boldsymbol{\alpha}\| = 1$ .

Estimadores para  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$  y  $\eta$  en *Modelos parcialmente lineal generalizado con índice simple* fueron considerados por Carroll, Fan, Gijbels y Wand (1997). En este Capítulo, introduciremos dos propuestas de estimación robusta para este modelo, además estudiaremos sus propiedades asintóticas y derivaremos la función de influencia empírica para una de ellas.

## 3.2 Estimación

Sean  $(y_i, \mathbf{x}_i, \mathbf{t}_i)$  observaciones independientes tales que  $y_i \in \mathbb{R}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$  y  $\mathbf{t}_i \in \mathbb{R}^q$  y supondremos que  $y_i | (\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$  donde  $\mu_i = H(\mathbf{x}_i^\top \boldsymbol{\beta} + \eta(\boldsymbol{\alpha}^\top \mathbf{t}_i))$  y  $\text{VAR}(y_i | (\mathbf{x}_i, \mathbf{t}_i)) = V(\mu_i)$ . Sean  $\eta_0, \boldsymbol{\beta}_0$  y  $\boldsymbol{\alpha}_0$  los verdaderos parámetros y notemos por  $E_0$  la esperanza bajo el verdadero modelo, entonces  $E_0(y_i | (\mathbf{x}_i, \mathbf{t}_i)) = H(\mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_i))$ . Definamos,

$$R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \sum_{i=1}^n W_{\alpha i}(u) \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i) \quad (3.1)$$

$$R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = E_0[\rho(y, \mathbf{x}^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}) | \boldsymbol{\alpha}^\top \mathbf{t} = u], \quad (3.2)$$

donde  $w_1 : \mathbb{R}^p \rightarrow \mathbb{R}$  es una función de pesos que controla las covariables  $\mathbf{x}'$ s de alta palanca y  $\rho : \mathbb{R} \rightarrow \mathbb{R}$  una función de pérdida. Los pesos  $W_{\alpha i}(u)$  dependen de la cercanía de  $u$  y  $\mathbf{t}_i$  proyectado en la dirección  $\boldsymbol{\alpha}$ , i.e.  $\boldsymbol{\alpha}^\top \mathbf{t}_i$  y que tomaremos basados en núcleos

$$W_{\alpha i}(u) = \frac{K\left(\frac{\boldsymbol{\alpha}^\top \mathbf{t}_i - u}{h}\right)}{\sum_{j=1}^n K\left(\frac{\boldsymbol{\alpha}^\top \mathbf{t}_j - u}{h}\right)}$$

con  $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$ . Supongamos que  $\eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  es tal que  $\eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) = \underset{a \in \mathbb{R}}{\text{argmin}} R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)$ . Por otro lado, definamos las siguientes funciones,

$$G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = \frac{1}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + v(\boldsymbol{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \quad (3.3)$$

$$G(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = E_0[\rho(y, \mathbf{x}^\top \boldsymbol{\beta} + v(\boldsymbol{\alpha}^\top \mathbf{t})) w_2(\mathbf{x})] \quad (3.4)$$

donde  $w_2(\cdot)$  es una función de pesos que controla los puntos de alta palanca en las  $\mathbf{x}'$ s. Supongamos que  $\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0$  satisfacen  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = \underset{\boldsymbol{\alpha}, \boldsymbol{\beta}}{\text{argmin}} G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_0)$  es decir se satisface la Fisher consistencia.

Diversos autores, establecieron diferentes condiciones que garantizan la consistencia de Fisher, en el Capítulo anterior se resumieron algunas de las propuestas dadas para modelos parcialmente lineales generalizados, que también son válidas para los modelos parcialmente lineales generalizados con índice simple estudiados en este capítulo.

A continuación introduciremos dos familias de estimadores robustos en modelos parcialmente lineales generalizados con índice simple.

### 3.2.1 Propuesta 1

Esta primer propuesta esta basada en el método de perfiles que fue introducido por Severini y Wong (1992), estudiado por van der Vaart (1988) y aplicado más recientemente por Severini y Staniswalis (1994). El procedimiento de estimación que proponemos en primer lugar puede ser definido en dos pasos de la siguiente manera:

- **Paso 1:** Para cada  $\boldsymbol{\beta}$  y  $\boldsymbol{\alpha}$  fijos, con  $\|\boldsymbol{\alpha}\| = 1$  sea

$$\begin{aligned} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) &= \underset{a \in \mathbb{R}}{\text{argmin}} \sum_{i=1}^n W_{\alpha i}(u) \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i) \\ &= \underset{a \in \mathbb{R}}{\text{argmin}} R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) \end{aligned}$$

- **Paso 2:** Definamos el estimador de  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$  como

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underset{\|\boldsymbol{\alpha}\|=1, \boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(\boldsymbol{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i)$$

### 3.2.2 Propuesta 2

En la siguiente propuesta definimos un estimador de tres pasos. El cálculo del estimador propuesto depende de estimadores iniciales, que pueden ser obtenidos a partir de la propuesta anterior.

- **Paso 1:** Calculamos un estimador inicial robusto y consistente de  $\boldsymbol{\beta}_0$  y un estimador inicial robusto, consistente y equivariante por transformaciones ortogonales de  $\boldsymbol{\alpha}_0$ , que llamaremos  $\hat{\boldsymbol{\beta}}_R$  y  $\hat{\boldsymbol{\alpha}}_{R1}$ . Sea  $\hat{\boldsymbol{\alpha}}_R = \hat{\boldsymbol{\alpha}}_{R1} / \|\hat{\boldsymbol{\alpha}}_{R1}\|$ .

- **Paso 2:** Definamos un estimador de  $\eta_0$  como

$$\begin{aligned} \hat{\eta}_{\hat{\boldsymbol{\alpha}}_R, \hat{\boldsymbol{\beta}}_R}(u) &= \underset{a \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n W_{\hat{\boldsymbol{\alpha}}_R^i}(u) \rho(y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_R + a) w_1(\mathbf{x}_i) \\ &= \underset{a \in \mathbb{R}}{\operatorname{argmin}} R_n(\hat{\boldsymbol{\alpha}}_R, \hat{\boldsymbol{\beta}}_R, a, u). \end{aligned}$$

- **Paso 3:** Definimos estimadores para  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$  como

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underset{\boldsymbol{\alpha}, \boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}_R, \hat{\boldsymbol{\beta}}_R}(\boldsymbol{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i).$$

En ambas propuestas, el problema de minimización da origen a puntos críticos de las ecuaciones derivadas. Si  $\rho(y, u)$  es diferenciable y notamos por  $\Psi(y, u) = \partial \rho(y, u) / \partial u$ . Sean,

$$R^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = E(\Psi(y, \mathbf{x}^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}) | \boldsymbol{\alpha}^\top \mathbf{t} = u) \quad (3.5)$$

$$R_n^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \sum_{i=1}^n \frac{K_h(\boldsymbol{\alpha}^\top \mathbf{t}_i - u)}{\sum_{j=1}^n K_h(\boldsymbol{\alpha}^\top \mathbf{t}_j - u)} \Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i) \quad (3.6)$$

$$F^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}) = E \left( \Psi(y, \mathbf{x}^\top \boldsymbol{\beta} + \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(\boldsymbol{\alpha}^\top \mathbf{t})) w_2(\mathbf{x}) \left( \begin{array}{c} \mathbf{x} + \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) |_{u=\boldsymbol{\alpha}^\top \mathbf{t}} \\ \frac{\partial}{\partial \boldsymbol{\alpha}} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) |_{u=\boldsymbol{\alpha}^\top \mathbf{t}} + \frac{\partial}{\partial u} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) |_{u=\boldsymbol{\alpha}^\top \mathbf{t}} \end{array} \right) \right) + \theta \begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} \quad (3.7)$$

$$F_n^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(\boldsymbol{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \left( \begin{array}{c} \mathbf{x}_i + \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) |_{u=\boldsymbol{\alpha}^\top \mathbf{t}_i} \\ \frac{\partial}{\partial \boldsymbol{\alpha}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) |_{u=\boldsymbol{\alpha}^\top \mathbf{t}_i} + \frac{\partial}{\partial u} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) |_{u=\boldsymbol{\alpha}^\top \mathbf{t}_i} \end{array} \right) + \theta \begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix}. \quad (3.8)$$

donde  $\theta$  es el multiplicador de Lagrange debido a la restricción  $\|\boldsymbol{\alpha}\|^2 = 1$ . Entonces, los estimadores pueden ser obtenidos como solución de las ecuaciones derivadas. Más precisamente,  $\eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  y  $\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  satisfacen  $R^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u), u) = 0$  y  $R_n^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u), u) = 0$ , respectivamente. Y además,  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$  y  $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$  satisfacen  $F^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}) = \mathbf{0}$  y  $F_n^1(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}) = \mathbf{0}$ , respectivamente.

Análogamente, para la propuesta dos  $\eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  y  $\hat{\eta}_{\hat{\boldsymbol{\alpha}}_R, \hat{\boldsymbol{\beta}}_R}(u)$  satisfacen  $R^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u), u) = 0$  y  $R_n^1(\hat{\boldsymbol{\alpha}}_R, \hat{\boldsymbol{\beta}}_R, \hat{\eta}_{\hat{\boldsymbol{\alpha}}_R, \hat{\boldsymbol{\beta}}_R}(u), u) = 0$  respectivamente y  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$  y  $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$  satisfacen  $G^1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}) = \mathbf{0}$



y  $G_n^1(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \eta_{\widehat{\boldsymbol{\alpha}_R}, \widehat{\boldsymbol{\beta}_R}}) = \mathbf{0}$  respectivamente con

$$G^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = E \left( \Psi(y, \mathbf{x}^\top \boldsymbol{\beta} + v(\boldsymbol{\alpha}^\top \mathbf{t})) w_2(\mathbf{x}) \begin{pmatrix} \mathbf{x} \\ v'(\boldsymbol{\alpha}^\top \mathbf{t}) \mathbf{t} \end{pmatrix} \right) \quad (3.9)$$

$$G_n^1(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + v(\boldsymbol{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \begin{pmatrix} \mathbf{x}_i \\ v'(\boldsymbol{\alpha}^\top \mathbf{t}_i) \mathbf{t}_i \end{pmatrix}. \quad (3.10)$$

### 3.3 Consistencia

Al igual que en la Sección 2.3, el objetivo de esta parte es mostrar que bajo algunas condiciones de regularidad ambas propuestas de estimadores de  $\boldsymbol{\alpha}_0$  y  $\boldsymbol{\beta}_0$  presentados en la Sección anterior son consistentes. En esta parte supondremos que  $\mathbf{t} \in \mathcal{T} \subset \mathbb{R}^q$ . Sea  $\mathcal{T}_0 \subset \mathcal{T}$  un conjunto compacto, consideremos el conjunto  $\mathcal{U}(\mathcal{T}_0) = \{\boldsymbol{\alpha}^\top \mathbf{t} : \mathbf{t} \in \mathcal{T}_0, \|\boldsymbol{\alpha}\| = 1\}$  entonces si  $v$  es una función continua  $v : \mathcal{U}(\mathcal{T}_0) \rightarrow \mathbb{R}$  notaremos por  $\|v\|_{0,\infty} = \sup_{u \in \mathcal{U}(\mathcal{T}_0)} |v(u)|$ .

- D1.** La función  $\rho(y, a)$  es una función continua y acotada y  $w_1$  y  $w_2$  están acotadas.
- D2.** El núcleo  $K : \mathbb{R} \rightarrow \mathbb{R}$  es una función par, no negativa, acotada y continua con derivada acotada, que satisface  $\int K(u) du = 1$ ,  $\int u^2 K(u) du < \infty$   $|u|K(u) \rightarrow 0$  si  $|u| \rightarrow \infty$ .
- D3.** La sucesión  $h_n$  satisface  $h_n \rightarrow 0$ ,  $nh_n / \log(n) \rightarrow \infty$ .
- D4.**  $f_{\mathbf{T}}(\mathbf{t})$  la densidad marginal de  $\mathbf{t}_1$  es acotada. Dado un conjunto compacto  $\mathcal{T}_0 \subset \mathcal{T}$  existe una constante positiva  $A_1(\mathcal{U}(\mathcal{T}_0))$  tal que  $A_1(\mathcal{U}(\mathcal{T}_0)) < f_{\boldsymbol{\alpha}}(u)$  para todo  $u \in \mathcal{U}(\mathcal{T}_0)$  y  $\|\boldsymbol{\alpha}\| = 1$  donde  $f_{\boldsymbol{\alpha}}$  es la densidad marginal de  $\boldsymbol{\alpha}^\top \mathbf{t}_1$ .
- D5.**  $R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)$  satisface la siguiente condición de equicontinuidad: dado  $\epsilon > 0$  existe  $\delta > 0$  tal que

$$|u_1 - u_2| < \delta, \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| < \delta \text{ and } \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| < \delta, u_i \in \mathcal{U}(\mathcal{T}_0), \boldsymbol{\beta}_i \in \mathcal{K}, \boldsymbol{\alpha}_i \in \mathcal{S}^1 \text{ con } i = 1, 2$$

$$\text{entonces } \sup_{a \in \mathbb{R}} |R(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, a, u_1) - R(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2, a, u_2)| < \epsilon.$$

**D6.**  $R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)$  es una función continua y  $\eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  es función continua en  $(\boldsymbol{\alpha}, \boldsymbol{\beta}, u)$ .

**D7.** Los estimadores iniciales  $\widehat{\boldsymbol{\alpha}}_R$  y  $\widehat{\boldsymbol{\beta}}_R$  de  $\boldsymbol{\alpha}_0$  y  $\boldsymbol{\beta}_0$  son consistentes.

**Observaciones:** **D4** se cumple si  $f_{\mathbf{T}}(\mathbf{t}) > B_1(\mathcal{T}_0) \forall \mathbf{t} \in \mathcal{T}_0$ . La hipótesis **D7** es necesaria para probar la consistencia de los estimadores de la segunda propuesta, en primer lugar mostraremos la consistencia de los estimadores de la primera propuesta, de esta forma estos estimadores podrían ser considerados los estimadores iniciales de la segunda propuesta.

**Lema 3.3.1.** Sea  $\mathcal{K} \subset \mathbb{R}^p$  un conjunto compacto,  $\mathcal{S}^1 = \{\boldsymbol{\alpha} \in \mathbb{R}^q \mid \|\boldsymbol{\alpha}\| = 1\}$  y  $\mathcal{T}_0 \subset \mathcal{T}$  un conjunto compacto tal que  $\exists \delta_0 > 0$   $\mathcal{T}_{\delta_0} \subset \mathcal{T}$ , donde  $\mathcal{T}_{\delta_0}$  es la clausura de una  $\delta_0$  vecindad de  $\mathcal{T}_0$ . Supongamos que se satisfacen **D1** a **D6** y que vale la siguiente condición

i)  $K$  es de variación acotada

ii) la familia de funciones  $\mathcal{F} = \{f(y, \mathbf{x}) = \rho(y, \mathbf{x}^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}), \boldsymbol{\beta} \in \mathcal{K}, a \in \mathbb{R}\}$  tiene número de cubrimiento  $N(\epsilon, \mathcal{F}, L^1(\mathcal{Q})) \leq A\epsilon^{-W}$ , para cualquier medida de probabilidad  $\mathcal{Q}$  y  $0 < \epsilon < 1$ .

Entonces se tiene:

$$a) \sup_{a \in \mathbb{R}, \boldsymbol{\alpha} \in \mathcal{S}^1, \boldsymbol{\beta} \in \mathcal{K}} \|R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) - R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot)\|_{0, \infty} \xrightarrow{a.s.} 0$$

$$b) \text{ Si } \inf_{\substack{\boldsymbol{\alpha} \in \mathcal{S}^1, \boldsymbol{\beta} \in \mathcal{K} \\ u \in \mathcal{U}(\mathcal{T}_0)}} \left[ \lim_{|a| \rightarrow \infty} R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) - R(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u), u) \right] > 0, \text{ entonces}$$

$$\sup_{\boldsymbol{\alpha} \in \mathcal{S}^1, \boldsymbol{\beta} \in \mathcal{K}} \|\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}} - \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\|_{0, \infty} \xrightarrow{a.s.} 0.$$

DEMOSTRACIÓN DEL LEMA 3.3.1.

$$a) \text{ Sea } R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^\top \mathbf{t}_i - u) \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i), \quad R_{0n}(\boldsymbol{\alpha}, u) = \frac{1}{n} \sum_{i=1}^n K_h(u - \boldsymbol{\alpha}^\top \mathbf{t}_i)$$

$$\text{con } K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right). \text{ Entonces } R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \frac{R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)}{R_{0n}(\boldsymbol{\alpha}, u)}. \text{ Ahora,}$$

$$\begin{aligned} \sup_{\substack{\boldsymbol{\alpha} \in \mathcal{S}^1, \boldsymbol{\beta} \in \mathcal{K} \\ a \in \mathbb{R}}} \|R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) - R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot)\|_{0, \infty} &\leq \left[ \sup_{\substack{\boldsymbol{\alpha} \in \mathcal{S}^1, \boldsymbol{\beta} \in \mathcal{K} \\ a \in \mathbb{R}}} \|R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) - E(R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot))\|_{0, \infty} + \right. \\ &\quad + \sup_{\substack{\boldsymbol{\alpha} \in \mathcal{S}^1, \boldsymbol{\beta} \in \mathcal{K} \\ a \in \mathbb{R}}} \|E(R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot)) - R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) E(R_{0n}(\boldsymbol{\alpha}, \cdot))\|_{0, \infty} + \\ &\quad \left. + \|\rho\|_{\infty} \sup_{\boldsymbol{\alpha} \in \mathcal{S}^1} \|R_{0n}(\boldsymbol{\alpha}, \cdot) - E(R_{0n}(\boldsymbol{\alpha}, \cdot))\|_{0, \infty} \right] / \inf_{\boldsymbol{\alpha} \in \mathcal{S}^1, u \in \mathcal{U}(\mathcal{T}_0)} R_{0n}(\boldsymbol{\alpha}, u) \end{aligned}$$

donde  $\|\rho\|_{\infty} = \sup_{y, a} |\rho(y, a)|$ . Para  $n$  suficientemente grande,

$$\begin{aligned} \inf_{\boldsymbol{\alpha} \in \mathcal{S}^1, u \in \mathcal{U}(\mathcal{T}_0)} R_{0n}(\boldsymbol{\alpha}, u) &\geq \inf_{\boldsymbol{\alpha} \in \mathcal{S}^1, u \in \mathcal{U}(\mathcal{T}_0)} E(R_{0n}(\boldsymbol{\alpha}, u)) - \sup_{\boldsymbol{\alpha} \in \mathcal{S}^1} \|R_{0n}(\boldsymbol{\alpha}, \cdot) - E(R_{0n}(\boldsymbol{\alpha}, \cdot))\|_{0, \infty} \\ E(R_{0n}(\boldsymbol{\alpha}, u)) &= \frac{1}{h} \int K\left(\frac{v-u}{h}\right) f_{\boldsymbol{\alpha}}(v) dv = \int K(v) f_{\boldsymbol{\alpha}}(hv + u) dv \end{aligned}$$

Consideremos  $\delta < \delta_0$  y sea  $\mathcal{U}_{\delta} = \{u + s : u \in \mathcal{U}(\mathcal{T}_0) \text{ } \|s\| \leq \delta\}$  y sea  $R$  tal que  $\int_{|v| \leq R} K(v) dv > 1/2$  entonces si  $h \leq \delta/R$  se tiene que  $hv + u \in \mathcal{U}_{\delta}$  y como  $\mathcal{U}_{\delta}$  es un conjunto compacto y por **D4** tenemos

$$E(R_{0n}(\boldsymbol{\alpha}, u)) > \frac{1}{2} A_1(\mathcal{U}_{\delta}).$$

Entonces es suficiente mostrar,

$$\sup_{\substack{\alpha \in \mathcal{S}^1, \beta \in \mathcal{K} \\ a \in \mathcal{R}}} \|R_{1n}(\alpha, \beta, a, \cdot) - E(R_{1n}(\alpha, \beta, a, \cdot))\|_{0, \infty} \xrightarrow{a.s.} 0 \quad (3.11)$$

$$\sup_{\alpha \in \mathcal{S}^1} \|R_{0n}(\alpha, \cdot) - E(R_{0n}(\alpha, \cdot))\|_{0, \infty} \xrightarrow{a.s.} 0 \quad (3.12)$$

$$\sup_{\substack{\alpha \in \mathcal{S}^1, \beta \in \mathcal{K} \\ a \in \mathcal{R}}} \|E(R_{1n}(\alpha, \beta, a, \cdot)) - R(\alpha, \beta, a, \cdot)E(R_{0n}(\alpha, \cdot))\|_{0, \infty} \rightarrow 0 \quad (3.13)$$

Usando el Teorema 37 de Pollard (1984) y **D1** se tiene (3.12).

Por otro lado, si  $u_1 = \alpha^t \mathbf{t}_1$

$$\begin{aligned} |E(R_{1n}(\alpha, \beta, a, u)) - R(\alpha, \beta, a, u)E(R_{0n}(\alpha, u))| &= |E(K_h(u_1 - u)[R(\alpha, \beta, a, u_1) - R(\alpha, \beta, a, u)])| \\ &= \left| \int \frac{1}{h} K_h(\tau - u)[R(\alpha, \beta, a, \tau) - R(\alpha, \beta, a, u)] f_\alpha(\tau) d\tau \right| \\ &\leq \|f_\alpha\|_\infty \int K(v) |R(\alpha, \beta, a, u - vh) - R(\alpha, \beta, a, u)| dv \end{aligned}$$

Como  $f_{\mathbf{T}}$  es acotada y  $\|\alpha\| = 1$ , haciendo un cambio de variables resulta  $\sup_{\alpha \in \mathcal{S}^1} \|f_\alpha\|_\infty < \infty$ , luego por **D5** se tiene (3.13).

Por último para probar (3.11), consideremos la clase de funciones

$$\mathcal{F}_n = \left\{ g_{\mathbf{t}, a, \alpha, \beta, h}(y, \mathbf{x}, v) = B^{-1} \rho(y, \mathbf{x}^t \beta + a) w_1(\mathbf{x}) K\left(\frac{\alpha^t \mathbf{t} - v}{h}\right) = B^{-1} \rho(y, \mathbf{x}^t \beta + a) w_1(\mathbf{x}) \widetilde{K}_{\alpha, h, \mathbf{t}}(v) \right\}$$

con  $B = \|\rho\|_\infty \|w_1\|_\infty \|K\|_\infty$ . La demostración de (3.11) se obtiene entonces como en el Teorema de 3.1 de Boente, He y Zhou (2006).

Parte b) Es análoga al Teorema 3.1 de Boente, He y Zhou (2006).  $\square$

### 3.3.1 Consistencia de los estimadores de la primera propuesta

**Lema 3.3.2.** Sea  $\widehat{\alpha}, \widehat{\beta}$  el argumento que minimiza  $G_n(\alpha, \beta, \widehat{\eta}_{\alpha, \beta})$  donde  $G_n(\alpha, \beta, v)$  esta definido en (3.3) y  $\widehat{\eta}_{\alpha, \beta}$  verifica

$$\sup_{\alpha \in \mathcal{S}^1, \beta \in \mathcal{K}} \|\widehat{\eta}_{\alpha, \beta} - \eta_{\alpha, \beta}\|_{0, \infty} \xrightarrow{a.s.} 0. \quad (3.14)$$

Si se satisface **D1** y **D6** entonces

$$a) \sup_{a, \alpha \in \mathcal{S}^1, \beta \in \mathcal{K}} |G_n(\alpha, \beta, \widehat{\eta}_{a, b}) - G(\alpha, \beta, \eta_{a, b})| \xrightarrow{a.s.} 0 \text{ para cualquier conjunto compacto } \mathcal{K}.$$

b) Si existe un conjunto compacto  $\mathcal{K}_1$  tal que  $\lim_{m \rightarrow \infty} P\left(\bigcap_{n \geq m} \widehat{\beta} \in \mathcal{K}_1\right)$  y  $G(\alpha, \beta, \eta_{\alpha, \beta})$  tiene un único mínimo en  $(\alpha_0, \beta_0)$ , entonces  $\widehat{\alpha} \xrightarrow{a.s.} \alpha_0$  y  $\widehat{\beta} \xrightarrow{a.s.} \beta_0$  donde  $\widehat{\alpha}$  y  $\widehat{\beta}$  son los estimadores de la primer propuesta.

## DEMOSTRACIÓN DEL LEMA 3.3.2.

a) Dado  $\varepsilon > 0$ , sea  $\mathcal{T}_0$  un conjunto compacto tal que  $P(t \notin \mathcal{T}_0) < \varepsilon$  entonces se tiene que,

$$\begin{aligned} \sup_{a, \alpha \in \mathcal{S}^1, b, \beta \in \mathcal{K}} |G_n(\alpha, \beta, \hat{\eta}_{a,b}) - G_n(\alpha, \beta, \eta_{a,b})| &\leq \sup_{a, \alpha \in \mathcal{S}^1, b, \beta \in \mathcal{K}} \|\hat{\eta}_{a,b} - \eta_{a,b}\|_{0,\infty} \|w_2\|_\infty \|\Psi\|_\infty \\ &+ 2\|w_2\|_\infty \|\rho\|_\infty \frac{1}{n} \sum_{i=1}^n I_{(\mathbf{t}_i \notin \mathcal{T}_0)} \end{aligned}$$

y usando (3.14) y La Ley Fuerte de los Grandes Números tenemos que,

$$\sup_{a, \alpha \in \mathcal{S}^1, b, \beta \in \mathcal{K}} |G_n(\alpha, \beta, \hat{\eta}_{a,b}) - G_n(\alpha, \beta, \eta_{a,b})| \xrightarrow{a.s.} 0.$$

Entonces es suficiente probar que  $\sup_{a, \alpha \in \mathcal{S}^1, b, \beta \in \mathcal{K}} |G_n(\alpha, \beta, \eta_{a,b}) - G(\alpha, \beta, \eta_{a,b})| \xrightarrow{a.s.} 0$ . Consideremos la familia de funciones

$$\mathcal{H} = \{f_{\alpha, \beta}(y, \mathbf{x}, \mathbf{t}) = \rho(y, \mathbf{x}^\top \beta + \eta_{a,b}(\alpha^\top \mathbf{t})) w_2(\mathbf{x}) \quad b, \beta \in \mathcal{K} \quad a, \alpha \in \mathcal{S}^1\}$$

luego usando el Teorema 3 del Capítulo 2 de Pollard, la compacidad de  $\mathcal{K}$ , **D1** y **D6** con los mismos argumentos que en el Lema 1 de Bianco y Boente (2002) se tiene que

$$\sup_{a, \alpha \in \mathcal{S}^1; b, \beta \in \mathcal{K}} |G_n(\alpha, \beta, \eta_{a,b}) - G(\alpha, \beta, \eta_{a,b})| \xrightarrow{a.s.} 0.$$

b) Como  $(\hat{\alpha}, \hat{\beta})$  esta últimamente en un compacto, supongamos por simplicidad que  $(\hat{\alpha}, \hat{\beta})$  es la subsucesión convergente,  $(\hat{\alpha}, \hat{\beta}) \xrightarrow{a.s.} (\alpha^*, \beta^*)$ , por la parte a) y la continuidad de  $G(\alpha, \beta, v)$  tenemos que  $G_n(\hat{\alpha}, \hat{\beta}, \hat{\eta}_{\hat{\alpha}, \hat{\beta}}) - G(\alpha^*, \beta^*, \eta_{\alpha^*, \beta^*}) \xrightarrow{a.s.} 0$  y  $G_n(\alpha_0, \beta_0, \hat{\eta}_{\alpha_0, \beta_0}) - G(\alpha_0, \beta_0, \eta_0) \xrightarrow{a.s.} 0$ . Como  $G_n(\alpha_0, \beta_0, \hat{\eta}_{\alpha_0, \beta_0}) \geq G_n(\hat{\alpha}, \hat{\beta}, \hat{\eta}_{\hat{\alpha}, \hat{\beta}})$  y  $G(\alpha, \beta, \eta_{\alpha, \beta})$  tiene un único mínimo en  $(\alpha_0, \beta_0)$ , fácilmente se obtiene b).  $\square$

### 3.3.2 Consistencia de los estimadores de la segunda propuesta

**Lema 3.3.3.** Sea  $(\hat{\alpha}, \hat{\beta})$  el argumento que minimiza  $G_n(\alpha, \beta, \hat{\eta}_{\hat{\alpha}, \hat{\beta}})$  donde  $G_n(\alpha, \beta, v)$  esta definido en (3.3) y  $\hat{\eta}_{\alpha, \beta}$  verifica

$$\sup_{\alpha \in \mathcal{S}^1, \beta \in \mathcal{K}} \|\hat{\eta}_{\alpha, \beta} - \eta_{\alpha, \beta}\|_{0,\infty} \xrightarrow{a.s.} 0. \quad (3.15)$$

Supongamos que:

- a) se satisface **D1** y **D7**.
- b) existe un conjunto compacto  $\mathcal{K}_1$  tal que  $\lim_{m \rightarrow \infty} P\left(\bigcap_{n \geq m} \hat{\beta} \in \mathcal{K}_1\right) = 1$ .
- c)  $G(\alpha, \beta, \eta_0)$  tiene un único mínimo en  $(\alpha_0, \beta_0)$ .

Entonces  $\widehat{\boldsymbol{\alpha}} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$  y  $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$  donde  $\widehat{\boldsymbol{\alpha}}$  y  $\widehat{\boldsymbol{\beta}}$  son los estimadores correspondientes a la segunda propuesta.

DEMOSTRACIÓN DEL LEMA 3.3.3.

Como  $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$  esta últimamente en un compacto, entonces podemos sin pérdida de generalidad suponer que  $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) \xrightarrow{a.s.} (\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ , luego por **D7**, la continuidad de  $\eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$  y el Lema 3.3.2. parte a) tenemos que  $G_n(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\eta}_{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}}) - G(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \eta_0) \xrightarrow{a.s.} 0$  y  $G_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}) - G(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \eta_0) \xrightarrow{a.s.} 0$ .

Finalmente como  $G_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}) \geq G_n(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\eta}_{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}})$  y  $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_0)$  tiene un único mínimo en  $(\boldsymbol{\beta}_0, \eta_0)$ , se tiene lo que queríamos probar.  $\square$

### 3.4 Función de influencia empírica

Como en la Sección (2.4), en esta parte obtendremos la función de influencia empírica de  $\widehat{\boldsymbol{\alpha}}$  y  $\widehat{\boldsymbol{\beta}}$  los estimadores resultantes de la primer propuesta. Con el objetivo de simplificar el cálculo de la función de influencia empírica, en primer lugar analizaremos como afecta la aplicación de transformaciones ortogonales de las observaciones a los estimadores  $\widehat{\boldsymbol{\alpha}}$  y  $\widehat{\boldsymbol{\beta}}$ . Así mismo, analizaremos la invarianza de transformación ortogonales en el problema teórico. Más precisamente, si  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$  minimizan  $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}})$  estudiaremos como se ven afectados  $\boldsymbol{\alpha}_0$  y  $\boldsymbol{\beta}_0$  cuando en lugar de considerar las variables  $(y, \mathbf{x}, \mathbf{t})$  tomamos  $(y, \mathbf{x}, \Gamma \mathbf{t})$  siendo  $\Gamma$  una matriz ortogonal en  $\mathbb{R}^{q \times q}$ . Así mismo si  $\widehat{\boldsymbol{\alpha}}$  y  $\widehat{\boldsymbol{\beta}}$  corresponden a los estimadores obtenidos a partir de una muestra aleatoria  $(y_i, \mathbf{x}_i, \mathbf{t}_i)$  estudiaremos los estimadores resultantes a partir de una muestra  $(y_i, \mathbf{x}_i, \Gamma \mathbf{t}_i)$ .

**Lema 3.4.1.** *Supongamos que  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$  minimizan  $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}})$ , sea  $\widetilde{\mathbf{t}} = \Gamma \mathbf{t}$  donde  $\Gamma^T \Gamma = I$  consideremos*

$$\widetilde{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \widetilde{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}) = E_0 \left[ \rho \left( y, \mathbf{x}^T \boldsymbol{\beta} + \widetilde{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(\boldsymbol{\alpha}^T \widetilde{\mathbf{t}}) \right) w_2(\mathbf{x}) \right]$$

donde  $\widetilde{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) = \operatorname{argmin}_a E_0 \left[ \rho(y, \mathbf{x}^T \boldsymbol{\beta} + a) w_1(\mathbf{x}) | \boldsymbol{\alpha}^T \widetilde{\mathbf{t}} = u \right]$ .

Si  $(\widetilde{\boldsymbol{\alpha}}_0, \widetilde{\boldsymbol{\beta}}_0) = \operatorname{argmin}_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^q \\ \boldsymbol{\beta} \in \mathbb{R}^p \\ \|\boldsymbol{\alpha}\|=1}} \widetilde{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \widetilde{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}})$  y  $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}})$  tiene un único mínimo en  $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$  entonces

tenemos que  $\widetilde{\boldsymbol{\alpha}}_0 = \Gamma \boldsymbol{\alpha}_0$  y  $\widetilde{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0$ .

DEMOSTRACIÓN DEL LEMA 3.4.1.

En primer lugar observemos que  $\widetilde{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) = \eta_{\Gamma^T \boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  pues

$$\begin{aligned} \widetilde{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) &= \operatorname{argmin}_a E_0 \left[ \rho(y, \mathbf{x}^T \boldsymbol{\beta} + a) w_1(\mathbf{x}) | \boldsymbol{\alpha}^T \widetilde{\mathbf{t}} = u \right] \\ &= \operatorname{argmin}_a E_0 \left[ \rho(y, \mathbf{x}^T \boldsymbol{\beta} + a) w_1(\mathbf{x}) | (\Gamma^T \boldsymbol{\alpha})^T \mathbf{t} = u \right] \\ &= \eta_{\Gamma^T \boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \end{aligned}$$

Entonces, tenemos que,

$$E_0 \left[ \rho \left( y, \mathbf{x}^T \widetilde{\boldsymbol{\beta}}_0 + \widetilde{\eta}_{\widetilde{\boldsymbol{\alpha}}_0, \widetilde{\boldsymbol{\beta}}_0}(\widetilde{\boldsymbol{\alpha}}_0^T \widetilde{\mathbf{t}}) \right) w_2(\mathbf{x}) \right] = E_0 \left[ \rho \left( y, \mathbf{x}^T \widetilde{\boldsymbol{\beta}}_0 + \eta_{\Gamma^T \widetilde{\boldsymbol{\alpha}}_0, \widetilde{\boldsymbol{\beta}}_0}((\Gamma^T \widetilde{\boldsymbol{\alpha}}_0)^T \mathbf{t}) \right) w_2(\mathbf{x}) \right] \quad (3.16)$$

Por otro lado tenemos que

$$\begin{aligned} \mathbb{E}_0 \left[ \rho \left( y, \mathbf{x}^\top \tilde{\boldsymbol{\beta}}_0 + \tilde{\eta}_{\alpha_0, \tilde{\boldsymbol{\beta}}_0}(\tilde{\boldsymbol{\alpha}}_0^\top \tilde{\mathbf{t}}) \right) w_2(\mathbf{x}) \right] &= \min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^q \|\boldsymbol{\alpha}\|=1 \\ \boldsymbol{\beta} \in \mathbb{R}^p}} \mathbb{E}_0 \left[ \rho \left( y, \mathbf{x}^\top \boldsymbol{\beta} + \tilde{\eta}_{\alpha, \boldsymbol{\beta}}(\boldsymbol{\alpha}^\top \tilde{\mathbf{t}}) \right) w_2(\mathbf{x}) \right] \\ \min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^q \|\boldsymbol{\alpha}\|=1 \\ \boldsymbol{\beta} \in \mathbb{R}^p}} \mathbb{E}_0 \left[ \rho \left( y, \mathbf{x}^\top \boldsymbol{\beta} + \eta_{\Gamma^\top \boldsymbol{\alpha}, \boldsymbol{\beta}}((\Gamma^\top \boldsymbol{\alpha})^\top \mathbf{t}) \right) w_2(\mathbf{x}) \right] &= \min_{\substack{\Gamma^\top \boldsymbol{\alpha} \in \mathbb{R}^q \|\Gamma^\top \boldsymbol{\alpha}\|=1 \\ \boldsymbol{\beta} \in \mathbb{R}^p}} \mathbb{E}_0 \left[ \rho \left( y, \mathbf{x}^\top \boldsymbol{\beta} + \eta_{\Gamma^\top \boldsymbol{\alpha}, \boldsymbol{\beta}}((\Gamma^\top \boldsymbol{\alpha})^\top \mathbf{t}) \right) w_2(\mathbf{x}) \right] \\ \min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^q \|\boldsymbol{\alpha}\|=1 \\ \boldsymbol{\beta} \in \mathbb{R}^p}} \mathbb{E}_0 \left[ \rho \left( y, \mathbf{x}^\top \boldsymbol{\beta} + \eta_{\alpha, \boldsymbol{\beta}}(\boldsymbol{\alpha}^\top \mathbf{t}) \right) w_2(\mathbf{x}) \right] &= \mathbb{E}_0 \left[ \rho \left( y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}) \right) w_2(\mathbf{x}) \right] \end{aligned}$$

luego de (3.16) se tiene que  $\Gamma^\top \tilde{\boldsymbol{\alpha}}_0 = \boldsymbol{\alpha}_0$  y  $\tilde{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0$  como queríamos probar.  $\square$

**Lema 3.4.2.** *Supongamos que  $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$  minimizan  $G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\alpha, \boldsymbol{\beta}})$ , sea  $\tilde{\mathbf{t}}_i = \Gamma \mathbf{t}_i$  donde  $\Gamma^\top \Gamma = I$  consideremos*

$$\tilde{G}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\alpha, \boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^n \rho \left( y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\eta}_{\alpha, \boldsymbol{\beta}}(\boldsymbol{\alpha}^\top \tilde{\mathbf{t}}_i) \right) w_2(\mathbf{x}_i)$$

$$\text{donde } \hat{\eta}_{\alpha, \boldsymbol{\beta}}(u) = \arg \min_{a \in \mathbb{R}} \sum_{i=1}^n \frac{K \left( \frac{\boldsymbol{\alpha}^\top \tilde{\mathbf{t}}_i - u}{h} \right)}{\sum_{j=1}^n K \left( \frac{\boldsymbol{\alpha}^\top \tilde{\mathbf{t}}_j - u}{h} \right)} \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i).$$

Si  $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underset{\substack{\boldsymbol{\alpha} \in \mathbb{R}^q \|\boldsymbol{\alpha}\|=1 \\ \boldsymbol{\beta} \in \mathbb{R}^p}}{\operatorname{argmin}} \tilde{G}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\alpha, \boldsymbol{\beta}})$  y  $G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\alpha, \boldsymbol{\beta}})$  tiene un único mínimo en  $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$  entonces

$$\hat{\boldsymbol{\alpha}} = \Gamma \hat{\boldsymbol{\alpha}} \text{ y } \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}.$$

DEMOSTRACIÓN DEL LEMA 3.4.2.

La demostración es análoga a la anterior. Observamos primero que  $\hat{\eta}_{\alpha, \boldsymbol{\beta}}(u) = \hat{\eta}_{\Gamma^\top \boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  pues

$$\begin{aligned} \hat{\eta}_{\alpha, \boldsymbol{\beta}}(u) &= \underset{a \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n \frac{K_h \left( \boldsymbol{\alpha}^\top \tilde{\mathbf{t}}_i - u \right)}{\sum_{j=1}^n K_h \left( \boldsymbol{\alpha}^\top \tilde{\mathbf{t}}_j - u \right)} \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i) \\ &= \underset{a \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n \frac{K_h \left( (\Gamma^\top \boldsymbol{\alpha})^\top \mathbf{t}_i - u \right)}{\sum_{j=1}^n K_h \left( (\Gamma^\top \boldsymbol{\alpha})^\top \mathbf{t}_j - u \right)} \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + a) w_1(\mathbf{x}_i) \\ &= \hat{\eta}_{\Gamma^\top \boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \end{aligned}$$

Entonces tenemos que,

$$\sum_{i=1}^n \rho \left( y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(u) \right) w_2(\mathbf{x}_i) = \sum_{i=1}^n \rho \left( y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}_{\Gamma^\top \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(u) \right) w_2(\mathbf{x}_i) \quad (3.17)$$

Por otro lado es fácil ver que

$$\sum_{i=1}^n \rho \left( y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(u) \right) w_2(\mathbf{x}_i) = \sum_{i=1}^n \rho \left( y_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(u) \right) w_2(\mathbf{x}_i)$$

entonces de (3.17) se tiene que  $(\Gamma^\top \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \square$

OBSERVACIÓN: Como consecuencia de lo estudiado anteriormente observemos que bastará con calcular la función de influencia empírica en el caso en que  $\hat{\boldsymbol{\alpha}}$  es el  $q$ -ésimo vector de la base canónica de dimensión  $q$  que llamaremos  $\mathbf{e}_q \in \mathbb{R}^q$ .

Pues si  $\hat{\boldsymbol{\alpha}} \neq \mathbf{e}_q$ , como  $\|\hat{\boldsymbol{\alpha}}\| = 1$  tomemos  $\hat{\gamma}_1, \dots, \hat{\gamma}_{q-1}$  tales que junto con  $\hat{\boldsymbol{\alpha}}$  formen una base ortonormal y llamemos  $\hat{\Gamma}$  a la matriz cuyas columnas corresponden a los elementos de la base, es decir  $\hat{\Gamma} = (\hat{\gamma}_1 \dots \hat{\gamma}_{q-1} \hat{\boldsymbol{\alpha}})$ . Entonces consideremos  $\{(y_i, \mathbf{x}_i, \hat{\Gamma}^T \mathbf{t}_i)\}_{1 \leq i \leq n}$  la muestra transformada y  $\hat{\boldsymbol{\alpha}}^*$  y  $\hat{\boldsymbol{\beta}}^*$  los estimadores asociados a dicha muestra. Luego bastará observar que por lo demostrado anteriormente  $\hat{\boldsymbol{\alpha}}_\varepsilon = \hat{\Gamma}^T \hat{\boldsymbol{\alpha}}_\varepsilon^*$  y  $\hat{\boldsymbol{\beta}}_\varepsilon = \hat{\boldsymbol{\beta}}_\varepsilon^*$  por lo tanto  $\text{EIF}(\hat{\boldsymbol{\alpha}})_j = \Gamma^T \text{EIF}(\hat{\boldsymbol{\alpha}}^*)_j$  y  $\text{EIF}(\hat{\boldsymbol{\beta}}) = \text{EIF}(\hat{\boldsymbol{\beta}}^*)$ . Entonces nos concentremos solamente en obtener la función de influencia empírica en el caso  $\hat{\boldsymbol{\alpha}} = \mathbf{e}_q$ .

Al igual que en la Sección 2.4 notaremos por  $\hat{\boldsymbol{\alpha}}_\varepsilon$ ,  $\hat{\boldsymbol{\beta}}_\varepsilon$  y  $\hat{\eta}_{\alpha, \beta}^{(\varepsilon)}(u)$  a los estimadores de la muestra transformada omitiendo nuevamente el subíndice  $j$ . También introduciremos la siguiente notación que usaremos tanto en esta Sección como en la siguiente.

$$\begin{aligned} \chi(y, a) &= \frac{\partial}{\partial a} \Psi(y, a) \\ \chi_1(y, a) &= \frac{\partial^2}{\partial a^2} \Psi(y, a) \\ \tau &= (\boldsymbol{\alpha}, \boldsymbol{\beta}, u) \\ \tau_0 &= (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0^T \mathbf{t}) \\ \tau_{0i} &= (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0^T \mathbf{t}_i) \\ \hat{\tau}_i &= (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \\ \nu_i(\tau) &= \begin{pmatrix} \mathbf{x}_i + \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\alpha, \beta}(s)|_{(\alpha, \beta, s) = \tau} \\ \frac{\partial}{\partial \boldsymbol{\alpha}} \eta_{\alpha, \beta}(s)|_{(\alpha, \beta, s) = \tau} + \frac{\partial}{\partial u} \eta_{\alpha, \beta}(s)|_{(\alpha, \beta, s) = \tau} \mathbf{t}_i \end{pmatrix} \\ \hat{\nu}_i(\tau) &= \begin{pmatrix} \mathbf{x}_i + \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\alpha, \beta}(s)|_{(\alpha, \beta, s) = \tau} \\ \frac{\partial}{\partial \boldsymbol{\alpha}} \hat{\eta}_{\alpha, \beta}(s)|_{(\alpha, \beta, s) = \tau} + \frac{\partial}{\partial u} \hat{\eta}_{\alpha, \beta}(s)|_{(\alpha, \beta, s) = \tau} \mathbf{t}_i \end{pmatrix} \end{aligned}$$

### Desarrollo para la obtención de la función de influencia empírica:

Antes de comenzar con el desarrollo formal para obtener la curva de influencia notemos que bajo el supuesto  $\hat{\boldsymbol{\alpha}} = \mathbf{e}_q$ , como  $\|\hat{\boldsymbol{\alpha}}_\varepsilon\| = 1$  para todo  $\varepsilon$ , se deduce fácilmente que  $\hat{\boldsymbol{\alpha}}^T \text{EIF}_j(\hat{\boldsymbol{\alpha}}) = 0$ , entonces si  $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_q)$ , tenemos que  $\text{EIF}_j(\hat{\alpha}_q) = 0$ . Luego si notamos por  $\hat{\boldsymbol{\alpha}}^{(q-1)} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{q-1})$  bastará con obtener  $\text{EIF}_j(\hat{\boldsymbol{\beta}})$  y  $\text{EIF}_j(\hat{\boldsymbol{\alpha}}^{(q-1)})$ .

Este desarrollo lo realizaremos en dos etapas, en primer lugar mostraremos que existen dos matrices aleatorias  $\hat{\mathbf{A}}$  y  $\hat{\mathbf{B}}$  en  $\mathbb{R}^{(p+q) \times (p+q)}$  que definen implícitamente  $\text{EIF}_j(\hat{\boldsymbol{\beta}})$  y  $\text{EIF}_j(\hat{\boldsymbol{\alpha}})$  de la siguiente forma;

$$P_\alpha \hat{\mathbf{A}} = - \left\{ P_\alpha \hat{\mathbf{B}} + \theta \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \right\} \begin{pmatrix} \text{EIF}_j(\hat{\boldsymbol{\beta}}) \\ \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \end{pmatrix}$$

$$\text{con } P_\alpha = \begin{pmatrix} I & 0 \\ 0 & I - \hat{\boldsymbol{\alpha}} \hat{\boldsymbol{\alpha}}^T \end{pmatrix}.$$

En la segunda etapa estudiaremos el sistema obtenido en el paso anterior para obtener expresiones explícitas de  $\text{EIF}_j(\hat{\beta})$  y  $\text{EIF}_j(\hat{\alpha})$ .

**Etapla 1:** Observemos que  $(\hat{\beta}_\varepsilon, \hat{\alpha}_\varepsilon)$  resuelven

$$\begin{aligned} 0 &= \theta_\varepsilon \begin{pmatrix} 0 \\ \hat{\alpha}_\varepsilon \end{pmatrix} \\ &+ \frac{1-\varepsilon}{n} \sum_{i=1}^n \Psi \left( y_i, \mathbf{x}_i^\top \hat{\beta}_\varepsilon + \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(\hat{\alpha}_\varepsilon^\top \mathbf{t}_i) \right) w_2(\mathbf{x}_i) \begin{pmatrix} \mathbf{x}_i + \frac{\partial}{\partial \beta} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_i} \\ \frac{\partial}{\partial \alpha} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_i} + \mathbf{t}_i \frac{\partial}{\partial u} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_i} \end{pmatrix} \\ &+ \varepsilon \Psi \left( y_j, \mathbf{x}_j^\top \hat{\beta}_\varepsilon + \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(\hat{\alpha}'_\varepsilon \mathbf{t}_j) \right) w_2(\mathbf{x}_j) \begin{pmatrix} \mathbf{x}_j + \frac{\partial}{\partial \beta} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_j} \\ \frac{\partial}{\partial \alpha} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_j} + \mathbf{t}_j \frac{\partial}{\partial u} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_j} \end{pmatrix}, \end{aligned} \quad (3.18)$$

donde  $\hat{\eta}_{\alpha, \beta}^{(\varepsilon)}(u)$  satisface

$$\begin{aligned} 0 &= \frac{1-\varepsilon}{n} \sum_{i=1}^n K_h(\alpha^\top \mathbf{t}_i - u) \Psi \left( y_i, \mathbf{x}_i^\top \beta + \hat{\eta}_{\alpha, \beta}^{(\varepsilon)}(u) \right) w_1(\mathbf{x}_i) \\ &+ \varepsilon K_h(\alpha^\top \mathbf{t}_j - u) \Psi \left( y_j, \mathbf{x}_j^\top \beta + \hat{\eta}_{\alpha, \beta}^{(\varepsilon)}(u) \right) w_1(\mathbf{x}_j). \end{aligned} \quad (3.19)$$

Llamemos  $\phi(\varepsilon)$  a

$$\begin{aligned} \phi(\varepsilon) &= \frac{1-\varepsilon}{n} \sum_{i=1}^n \Psi \left( y_i, \mathbf{x}_i^\top \hat{\beta}_\varepsilon + \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(\hat{\alpha}_\varepsilon^\top \mathbf{t}_i) \right) w_2(\mathbf{x}_i) \begin{pmatrix} \mathbf{x}_i + \frac{\partial}{\partial \beta} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_i} \\ \frac{\partial}{\partial \alpha} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_i} + \mathbf{t}_i \frac{\partial}{\partial u} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_i} \end{pmatrix} \\ &+ \varepsilon \Psi \left( y_j, \mathbf{x}_j^\top \hat{\beta}_\varepsilon + \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(\hat{\alpha}'_\varepsilon \mathbf{t}_j) \right) w_2(\mathbf{x}_j) \begin{pmatrix} \mathbf{x}_j + \frac{\partial}{\partial \beta} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_j} \\ \frac{\partial}{\partial \alpha} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_j} + \mathbf{t}_j \frac{\partial}{\partial u} \hat{\eta}_{\hat{\alpha}_\varepsilon, \hat{\beta}_\varepsilon}^{(\varepsilon)}(u) \Big|_{u=\hat{\alpha}_\varepsilon^\top \mathbf{t}_j} \end{pmatrix} \end{aligned} \quad (3.20)$$

luego (3.18) es equivalente a  $\theta_\varepsilon \begin{pmatrix} 0 \\ \hat{\alpha}_\varepsilon \end{pmatrix} + \phi(\varepsilon) = 0$ . Derivando (3.18) en  $\varepsilon$  y evaluando en 0 se tiene,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \varepsilon} \left[ \theta_\varepsilon \begin{pmatrix} 0 \\ \hat{\alpha}_\varepsilon \end{pmatrix} + \phi(\varepsilon) \right] \Big|_{\varepsilon=0} \\ &= \frac{\partial}{\partial \varepsilon} (\theta_\varepsilon) \Big|_{\varepsilon=0} \begin{pmatrix} 0 \\ \hat{\alpha} \end{pmatrix} + \theta \begin{pmatrix} 0 \\ \text{EIF}_j(\hat{\alpha}) \end{pmatrix} + \frac{\partial}{\partial \varepsilon} \phi(\varepsilon) \Big|_{\varepsilon=0} \end{aligned}$$

Entonces como  $\|\hat{\alpha}\| = 1$ ,  $P_{\hat{\alpha}} \begin{pmatrix} 0 \\ \hat{\alpha} \end{pmatrix} = 0$  y  $\hat{\alpha}^\top \text{EIF}_j(\hat{\alpha}) = 0$  de donde,

$$P_{\hat{\alpha}} \frac{\partial}{\partial \varepsilon} \phi(\varepsilon) \Big|_{\varepsilon=0} = -\theta \begin{pmatrix} 0 \\ \text{EIF}_j(\hat{\alpha}) \end{pmatrix} \quad (3.21)$$



Para simplificar la notación llamemos

$$\begin{aligned}
g(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) &= \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}^{(\varepsilon)}(u) & g_{\boldsymbol{\alpha}}(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) &= \frac{\partial}{\partial \boldsymbol{\alpha}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}^{(\varepsilon)}(u) \\
g_{\boldsymbol{\beta}}(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) &= \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}^{(\varepsilon)}(u) & g_u(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) &= \frac{\partial}{\partial u} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}^{(\varepsilon)}(u) \\
G_i &= \frac{\partial}{\partial \varepsilon} g(\varepsilon, \hat{\boldsymbol{\alpha}}_{\varepsilon}, \hat{\boldsymbol{\beta}}_{\varepsilon}, \hat{\boldsymbol{\alpha}}_{\varepsilon}^T \mathbf{t}_i) \Big|_{\varepsilon=0} & G_{\boldsymbol{\alpha}, i} &= \frac{\partial}{\partial \varepsilon} g_{\boldsymbol{\alpha}}(\varepsilon, \hat{\boldsymbol{\alpha}}_{\varepsilon}, \hat{\boldsymbol{\beta}}_{\varepsilon}, \hat{\boldsymbol{\alpha}}_{\varepsilon}^T \mathbf{t}_i) \Big|_{\varepsilon=0} \\
G_{\boldsymbol{\beta}, i} &= \frac{\partial}{\partial \varepsilon} g_{\boldsymbol{\beta}}(\varepsilon, \hat{\boldsymbol{\alpha}}_{\varepsilon}, \hat{\boldsymbol{\beta}}_{\varepsilon}, \hat{\boldsymbol{\alpha}}_{\varepsilon}^T \mathbf{t}_i) \Big|_{\varepsilon=0} & G_{u, i} &= \frac{\partial}{\partial \varepsilon} g_u(\varepsilon, \hat{\boldsymbol{\alpha}}_{\varepsilon}, \hat{\boldsymbol{\beta}}_{\varepsilon}, \hat{\boldsymbol{\alpha}}_{\varepsilon}^T \mathbf{t}_i) \Big|_{\varepsilon=0}.
\end{aligned}$$

Calculemos  $\frac{\partial}{\partial \varepsilon} \phi(\varepsilon) \Big|_{\varepsilon=0}$

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon} \phi(\varepsilon) \Big|_{\varepsilon=0} &= -\frac{1}{n} \sum_{i=1}^n \Psi \left( y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \right) w_2(\mathbf{x}_i) \hat{\nu}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \\
&+ \frac{1}{n} \sum_{i=1}^n \chi \left( y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \right) (\mathbf{x}_i^T \text{EIF}_j(\hat{\boldsymbol{\beta}}) + G_i) w_2(\mathbf{x}_i) \hat{\nu}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \\
&+ \frac{1}{n} \sum_{i=1}^n \Psi \left( y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \right) w_2(\mathbf{x}_i) \begin{pmatrix} G_{\boldsymbol{\beta}, i} \\ G_{\boldsymbol{\alpha}, i} + \mathbf{t}_i G_{u, i} \end{pmatrix} \\
&+ \Psi \left( y_j, \mathbf{x}_j^T \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\alpha}}^T \mathbf{t}_j) \right) w_2(\mathbf{x}_j) \hat{\nu}_j(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}^T \mathbf{t}_j)
\end{aligned}$$

Usando que  $\theta \binom{0}{\hat{\boldsymbol{\alpha}}} + \phi(0) = 0$

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon} \phi(\varepsilon) \Big|_{\varepsilon=0} &= \theta \binom{0}{\hat{\boldsymbol{\alpha}}} + \frac{1}{n} \sum_{i=1}^n \chi \left( y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \right) (\mathbf{x}_i^T \text{EIF}_j(\hat{\boldsymbol{\beta}}) + G_i) w_2(\mathbf{x}_i) \hat{\nu}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \\
&+ \frac{1}{n} \sum_{i=1}^n \Psi \left( y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\alpha}}^T \mathbf{t}_i) \right) w_2(\mathbf{x}_i) \begin{pmatrix} G_{\boldsymbol{\beta}, i} \\ G_{\boldsymbol{\alpha}, i} + \mathbf{t}_i G_{u, i} \end{pmatrix} \\
&+ \Psi \left( y_j, \mathbf{x}_j^T \hat{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}}(\hat{\boldsymbol{\alpha}}^T \mathbf{t}_j) \right) w_2(\mathbf{x}_j) \hat{\nu}_j(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}^T \mathbf{t}_j) \tag{3.22}
\end{aligned}$$

Entonces es necesario calcular  $G_i, G_{\boldsymbol{\beta}, i}, G_{\boldsymbol{\alpha}, i}$  y  $G_{u, i}$ .

**Lema 3.4.3.** Si  $\frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) w_1(\mathbf{x}_i)$  es no singular y  $\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)$  es continua y tiene derivadas parciales continuas respecto de  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  y  $u$  entonces,

$$\begin{aligned}
G_i &= \text{EIF}_j(\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} + \frac{\partial}{\partial \boldsymbol{\alpha}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) + \frac{\partial}{\partial \boldsymbol{\beta}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\beta}}) + \frac{\partial}{\partial u} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \mathbf{t}_i \\
G_{\boldsymbol{\alpha}, i} &= \text{EIF}_j \left( \frac{\partial}{\partial \boldsymbol{\alpha}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \right) \Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial \boldsymbol{\alpha}^T \partial \boldsymbol{\alpha}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \\
&+ \frac{\partial^2}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\alpha}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\beta}}) + \frac{\partial^2}{\partial u \partial \boldsymbol{\alpha}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \mathbf{t}_i \\
G_{\boldsymbol{\beta}, i} &= \text{EIF}_j \left( \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \right) \Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \\
&+ \frac{\partial^2}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\beta}}) + \frac{\partial^2}{\partial u \partial \boldsymbol{\beta}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \mathbf{t}_i
\end{aligned}$$

$$\begin{aligned}
G_{u,i} &= \text{EIF}_j\left(\frac{\partial}{\partial u}\hat{\eta}_{\alpha,\beta}(u)\right)\Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial\alpha\partial u}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha}) \\
&+ \frac{\partial^2}{\partial\beta\partial u}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\beta}) + \frac{\partial^2}{\partial u^2}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha})\mathbf{t}_i
\end{aligned}$$

La demostración del Lema anterior se encuentra en el Apéndice donde también se dan las expresiones de  $\text{EIF}_j(\hat{\eta}_{\alpha,\beta}(u))$ ,  $\text{EIF}_j(\frac{\partial}{\partial\beta}\hat{\eta}_{\alpha,\beta}(u))$ ,  $\text{EIF}_j(\frac{\partial}{\partial\alpha}\hat{\eta}_{\alpha,\beta}(u))$  y  $\text{EIF}_j(\frac{\partial}{\partial u}\hat{\eta}_{\alpha,\beta}(u))$ .

Ahora debemos reemplazar las expresiones de  $G_i$ ,  $G_{\beta,i}$ ,  $G_{\alpha,i}$  y  $G_{u,i}$  obtenidas en el Lema anterior en (3.22), para esto primero calculamos;

$$\begin{aligned}
\mathbf{x}_i^\top \text{EIF}_j(\hat{\beta}) + G_i &= \text{EIF}_j(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} + [\mathbf{x}_i^\top + \frac{\partial}{\partial\beta}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i}] \text{EIF}_j(\hat{\beta}) \\
&+ \text{EIF}_j(\hat{\alpha})\left[\frac{\partial}{\partial\alpha}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} + \frac{\partial}{\partial u}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i}\mathbf{t}_i\right] \\
&= \text{EIF}_j\left(\frac{\partial}{\partial u}\hat{\eta}_{\alpha,\beta}(u)\right)\Big|_{\tau=\hat{\tau}_i} + \hat{\nu}_i(\hat{\tau}_i)^\top \begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}) \end{pmatrix}
\end{aligned}$$

Si llamamos  $\mathbf{A}(\hat{\tau}_i)$  a la matriz de dimensión  $(p+q) \times (p+q)$  tal que

$$\begin{aligned}
\mathbf{A}(\hat{\tau}_i)_{11} &= \frac{\partial^2}{\partial\beta\partial\beta^\top}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \\
\mathbf{A}(\hat{\tau}_i)_{12} &= \frac{\partial^2}{\partial\alpha\partial\beta^\top}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial u\partial\beta}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \mathbf{t}_i^\top \\
\mathbf{A}(\hat{\tau}_i)_{21} &= \frac{\partial^2}{\partial\beta\partial\alpha^\top}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial\beta\partial u}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \mathbf{t}_i^\top \\
\mathbf{A}(\hat{\tau}_i)_{22} &= \frac{\partial^2}{\partial\alpha\partial\alpha^\top}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial u\partial\alpha}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \mathbf{t}_i^\top + \frac{\partial^2}{\partial\alpha\partial u}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \mathbf{t}_i^\top \\
&+ \frac{\partial^2}{\partial u^2}(\hat{\eta}_{\alpha,\beta}(u))\Big|_{\tau=\hat{\tau}_i} \mathbf{t}_i\mathbf{t}_i^\top
\end{aligned}$$

entonces

$$\begin{pmatrix} G_{\beta,i} \\ G_{\alpha,i} + G_{u,i}\mathbf{t}_i \end{pmatrix} = \begin{pmatrix} \text{EIF}_j\left(\frac{\partial}{\partial\beta}\hat{\eta}_{\alpha,\beta}(u)\right)\Big|_{\tau=\hat{\tau}_i} \\ \text{EIF}_j\left(\frac{\partial}{\partial\alpha}\hat{\eta}_{\alpha,\beta}(u)\right)\Big|_{\tau=\hat{\tau}_i} + \text{EIF}_j\left(\frac{\partial}{\partial u}\hat{\eta}_{\alpha,\beta}(u)\right)\Big|_{\tau=\hat{\tau}_i} \mathbf{t}_i \end{pmatrix} + \mathbf{A}(\hat{\tau}_i) \begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}) \end{pmatrix}$$

ahora si reemplazamos en (3.22),

$$\frac{\partial}{\partial\varepsilon}\phi(\varepsilon)\Big|_{\varepsilon=0} = \theta\left(\begin{pmatrix} 0 \\ \hat{\alpha} \end{pmatrix}\right)$$



entonces

$$P_{\hat{\alpha}} \hat{\mathbf{A}} = - \left\{ P_{\hat{\alpha}} \hat{\mathbf{B}} + \theta \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \right\} \begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}) \end{pmatrix} \quad (3.23)$$

**Etapla 2:** A continuación analizaremos si  $P_{\hat{\alpha}} \hat{\mathbf{B}} + \theta \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}$  es asintóticamente inversible. Para esto supondremos que  $w_1 \equiv w_2$ , por lo tanto  $\text{E}(\Psi(y, \mathbf{x}^T \beta + \eta_{\alpha, \beta}(u)) w_2(\mathbf{x}) | \alpha^T \mathbf{t} = u) = 0$  y deducimos entonces que

$$\frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \hat{\beta} + \hat{\eta}_{\hat{\alpha}, \hat{\beta}}(\hat{\alpha}^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \mathbf{A}(\hat{\tau}_i) \xrightarrow{p} 0$$

y  $\text{E}(\chi(y, \mathbf{x}^T \beta + \eta_{\alpha, \beta}(u)) (\mathbf{x}^T + \frac{\partial}{\partial \beta} \eta_{\alpha, \beta}(u)) w_2(\mathbf{x}) | \alpha^T \mathbf{t} = u) = 0$ . Luego se sigue fácilmente que,

$$\frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \hat{\beta} + \hat{\eta}_{\hat{\alpha}, \hat{\beta}}(\hat{\alpha}^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\nu}_i(\hat{\tau}_i) \hat{\nu}_i(\hat{\tau}_i)^T \xrightarrow{a.s.} \mathbf{B}$$

donde

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & 0 \\ 0 & \mathbf{B}_2 \end{pmatrix} \quad (3.24)$$

$$\begin{aligned} \mathbf{B}_{11} &= \text{E} \left\{ \chi(y, \mathbf{x}^T \beta_0 + \eta_0(\alpha_0^T \mathbf{t})) \left( \mathbf{x}^T + \frac{\partial}{\partial \beta} \eta_0(u) |_{u=\alpha_0^T \mathbf{t}} \right) \left( \mathbf{x}^T + \frac{\partial}{\partial \beta} \eta_0(u) |_{u=\alpha_0^T \mathbf{t}} \right)^T w_2(\mathbf{x}) \right\} \\ \mathbf{B}_2 &= \text{E} \left\{ \chi(y, \mathbf{x}^T \beta_0 + \eta_0(\alpha_0^T \mathbf{t})) \left( \frac{\partial}{\partial \alpha} \eta_0(u) |_{u=\alpha_0^T \mathbf{t}} + \mathbf{t} \frac{\partial}{\partial u} \eta_0(u) |_{u=\alpha_0^T \mathbf{t}} \right) \right. \\ &\quad \left. \left( \frac{\partial}{\partial \alpha} \eta_0(u) |_{u=\alpha_0^T \mathbf{t}} + \mathbf{t} \frac{\partial}{\partial u} \eta_0(u) |_{u=\alpha_0^T \mathbf{t}} \right)^T w_2(\mathbf{x}) \right\} \end{aligned}$$

por otro lado  $\theta = -\hat{\alpha}^T \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \hat{\beta} + \hat{\eta}_{\hat{\alpha}, \hat{\beta}}(\hat{\alpha}^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\nu}_i(\hat{\tau}_i)$  por lo tanto tenemos que  $\theta \xrightarrow{p} 0$  y  $P_{\hat{\alpha}} \hat{\mathbf{B}} + \theta \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \xrightarrow{a.s.} P_{\hat{\alpha}} \mathbf{B}$

Bastará entonces analizar la inversibilidad de  $P_{\hat{\alpha}} \mathbf{B}$ . Pero como  $\text{EIF}_j(\hat{\alpha}_q) = 0$  y solo necesitamos calcular  $\text{EIF}_j(\hat{\alpha}_j)$  para  $1 \leq j \leq q-1$ , es suficiente mostrar que la sub-matriz superior izquierda de dimensión  $(p+q-1) \times (p+q-1)$  de  $P_{\hat{\alpha}} \mathbf{B}$  es inversible, es decir si en (3.24) llamamos  $\mathbf{B}_2 = \begin{pmatrix} \mathbf{B}_{21} & \mathbf{B}_{22} \\ \mathbf{B}_{23} & \mathbf{B}_{24} \end{pmatrix}$  donde  $\mathbf{B}_{21}$  es una matriz de  $(q-1) \times (q-1)$  entonces bastará ver que  $\begin{pmatrix} \mathbf{B}_{11} & 0 \\ 0 & \mathbf{B}_{21} \end{pmatrix}$  es inversible pero esto se deduce del hecho que  $\mathbf{B}_{11}$  y  $\mathbf{B}_{21}$  son no singulares.

Finalmente de (3.23) podemos despejar  $\text{EIF}_j(\hat{\beta})$  y  $\text{EIF}_j(\hat{\alpha}^{(q-1)})$

$$P_{\hat{\alpha}} \hat{\mathbf{A}} = - \left\{ P_{\hat{\alpha}} \hat{\mathbf{B}} + \theta \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \right\} \begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}) \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{B}}_{11} & \hat{\mathbf{C}}_1 & \hat{\mathbf{C}}_2 \\ \hat{\mathbf{C}}_3 & \hat{\mathbf{B}}_{21} & \hat{\mathbf{B}}_{22} \\ \hat{\mathbf{C}}_4 & \hat{\mathbf{B}}_{23} & \hat{\mathbf{B}}_{24} \end{pmatrix} \begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}^{(q-1)}) \\ \text{EIF}_j(\hat{\alpha}_q) \end{pmatrix}$$

Observar que la última fila de  $P_{\hat{\alpha}} \hat{A}$  es cero pues la última fila de  $P_{\hat{\alpha}}$  es cero, es decir,  $P_{\hat{\alpha}} \hat{A} = \begin{pmatrix} (P_{\hat{\alpha}} \hat{A})_{p+q-1} \\ 0 \dots 0 \end{pmatrix}$  y  $\text{EIF}_j(\hat{\alpha}_q)$  también es cero. Luego

$$\begin{pmatrix} (P_{\hat{\alpha}} \hat{A})_{p+q-1} \\ 0 \dots 0 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{B}}_{11} & \hat{\mathbf{C}}_1 & \hat{\mathbf{C}}_2 \\ \hat{\mathbf{C}}_3 & \hat{\mathbf{B}}_{21} & \hat{\mathbf{B}}_{22} \\ \hat{\mathbf{C}}_4 & \hat{\mathbf{B}}_{23} & \hat{\mathbf{B}}_{24} \end{pmatrix} \begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}^{q-1}) \\ 0 \end{pmatrix}$$

entonces

$$(P_{\hat{\alpha}} \hat{A})_{p+q-1} = \begin{pmatrix} \hat{\mathbf{B}}_{11} & \hat{\mathbf{C}}_1 \\ \hat{\mathbf{C}}_3 & \hat{\mathbf{B}}_{21} \end{pmatrix} \begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}^{q-1}) \end{pmatrix}$$

y por lo visto antes  $\begin{pmatrix} \hat{\mathbf{B}}_{11} & \hat{\mathbf{C}}_1 \\ \hat{\mathbf{C}}_3 & \hat{\mathbf{B}}_{21} \end{pmatrix}$  es asintóticamente inversible tenemos que

$$\begin{pmatrix} \text{EIF}_j(\hat{\beta}) \\ \text{EIF}_j(\hat{\alpha}^{q-1}) \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{B}}_{11} & \hat{\mathbf{C}}_1 \\ \hat{\mathbf{C}}_3 & \hat{\mathbf{B}}_{21} \end{pmatrix}^{-1} (P_{\hat{\alpha}} \hat{A})_{p+q-1}$$

### 3.5 Distribución Asintótica

En esta Sección mostraremos que bajo ciertos supuestos los estimadores propuestos en las Secciones 3.2.1 y 3.2.2 son asintóticamente normales. Como en la Sección 2.5 supondremos que  $\mathbf{t} \in \mathcal{T}$  y que  $\mathcal{T}$  es compacto. Llamemos  $\Sigma = \Sigma^{(1)} + \Sigma^{(2)} + [\Sigma^{(2)}]^\top + \Sigma^{(3)}$  donde

$$\begin{aligned} \Sigma^{(1)} &= E_0 \left\{ \Psi^2(y_1, \mathbf{x}_1^\top \beta_0 + \eta_0(\alpha_0^\top \mathbf{t}_1)) w_2^2(\mathbf{x}_1) \nu(\tau_0) \nu(\tau_0)^\top \right\} \\ \Sigma^{(2)} &= E_0 \left\{ G_1(\alpha_0^\top \mathbf{t}_1) f_{\alpha_0}(\alpha_0^\top \mathbf{t}_1) \Psi(y_1, \mathbf{x}_1^\top \beta_0 + \eta_0(\alpha_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \gamma(\mathbf{t}_1) \nu(\tau_0)^\top \right\} \\ \Sigma^{(3)} &= E_0 \left\{ G_1^2(\alpha_0^\top \mathbf{t}_1) \gamma(\mathbf{t}_1) \gamma(\mathbf{t}_1)^\top f_{\alpha_0}(\alpha_0^\top \mathbf{t}_1) \right\} \end{aligned}$$

donde  $G_i(\mathbf{u})$  corresponde al primer elemento de

$$[E(\chi(y, \mathbf{x}^\top \beta_0 + \eta_0(\alpha_0^\top \mathbf{t})) w_1(\mathbf{x}) \nu(\tau_0) \nu(\tau_0)^\top | \mathbf{t} = \mathbf{u})]^{-1} \Psi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_0(\mathbf{u})) w_1(\mathbf{x}_i) \nu_i(\tau_{0i})$$

$$\text{y } \gamma(\mathbf{u}) = E(\chi(y, \mathbf{x}^\top \beta_0 + \eta_0(\alpha_0^\top \mathbf{t})) w_2(\mathbf{x}) \mathbf{x} f_{\alpha_0}^{-1}(\alpha_0^\top \mathbf{t}) | \mathbf{t} = \mathbf{u}).$$

El siguiente conjunto de hipótesis es necesario para el desarrollo de la distribución asintótica.

**M1.** a) Las funciones  $\hat{\eta}_{\alpha, \beta}(u)$  y  $\eta_{\alpha, \beta}(u)$  son continuas y con derivadas terceras continuas con respecto a  $\alpha, \beta$  y  $u$ .

b)  $\left\| \hat{\eta}_{\hat{\alpha}, \hat{\beta}} - \eta_0 \right\|_\infty \xrightarrow{p} 0$ , para  $\hat{\alpha}$  y  $\hat{\beta}$  estimadores consistentes de  $\alpha_0$  y  $\beta_0$ .

c) Para cada  $\mathbf{t} \in \mathcal{T}$ ,  $\alpha$  y  $\beta$ ,  $\hat{\eta}_{\alpha, \beta}(\alpha^\top \mathbf{t}) \xrightarrow{p} \eta_{\alpha, \beta}(\alpha^\top \mathbf{t})$ . Además,  $n^{\frac{1}{4}} \left\| \hat{\eta}_{\alpha_0, \beta_0} - \eta_0 \right\|_\infty \xrightarrow{p} 0$ ,

$$\left\| \frac{\partial}{\partial u} \hat{\eta}_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} - \frac{\partial}{\partial u} \eta_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} \right\|_\infty \xrightarrow{p} 0$$

$$\left\| \frac{\partial^2}{\partial u^2} \hat{\eta}_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} - \frac{\partial^2}{\partial u^2} \eta_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} \right\|_{\infty} \xrightarrow{p} 0$$

y

$$\left\| \frac{\partial^3}{\partial u^3} \hat{\eta}_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} - \frac{\partial^3}{\partial u^3} \eta_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} \right\|_{\infty} \xrightarrow{p} 0.$$

Para  $1 \leq j \leq p$ 

$$n^{\frac{1}{4}} \left\| \frac{\partial}{\partial \beta_j} \hat{\eta}_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} - \frac{\partial}{\partial \beta_j} \eta_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} \right\|_{\infty} \xrightarrow{p} 0$$

y

$$n^{\frac{1}{4}} \left\| \frac{\partial}{\partial \alpha_j} \hat{\eta}_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} - \frac{\partial}{\partial \alpha_j} \eta_{\alpha, \beta} \Big|_{\alpha=\alpha_0, \beta=\beta_0} \right\|_{\infty} \xrightarrow{p} 0$$

**M2.** Las funciones  $\Psi$ ,  $\chi$ ,  $\chi_1$ ,  $w_2$  y  $\psi_2(\mathbf{x}) = \mathbf{x} w_2(\mathbf{x})$  son continuas y acotadas.

**M3.** Sea  $\mathbf{B} = E_0 [\chi(y, \mathbf{x}^\top \beta_0 + \eta_0(\alpha_0^\top \mathbf{t})) w_2(\mathbf{x}) \nu(\tau_0) \nu(\tau_0)^\top]$  donde

$$\nu(\tau_0) = \begin{pmatrix} \mathbf{x} + \frac{\partial}{\partial \beta} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = \tau_0} \\ \frac{\partial}{\partial \alpha} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = \tau_0} + \frac{\partial}{\partial u} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = \tau_0} \mathbf{t} \end{pmatrix}$$

la submatriz superior izquierda de  $\mathbf{B}$  de dimensión  $(p+q-1) \times (p+q-1)$  que llamaremos  $\mathbf{B}_1$  es no singular.

**M4.** La submatriz superior izquierda de  $\Sigma$  de dimensión  $(p+q-1) \times (p+q-1)$  que llamaremos  $\Sigma_1$  es definida positiva.

**M5.**  $E_0 \{ \Psi(y, \mathbf{x}^\top \beta_0 + \eta_0(\mathbf{t})) | (\mathbf{x}, \mathbf{t}) \} = 0$ .

**M6.**  $E_0 (w_2(\mathbf{x}) \|\mathbf{x}\|^2) < \infty$ .

**M7.**  $f_{\alpha_0}$  la densidad marginal de  $\alpha_0^\top \mathbf{t}_1$  y su derivada son funciones continuas y acotadas. Más aún,  $\inf_{\mathbf{t} \in \mathcal{T}} f_{\alpha_0}(\alpha_0^\top \mathbf{t}) > 0$ .

**M8.** a)  $m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = E(G_1(\alpha_0^\top \mathbf{t}_2)) | \mathbf{t}_1 = \boldsymbol{\nu}_1, \mathbf{t}_2 = \boldsymbol{\nu}_2)$ ,  $m'_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \frac{\partial}{\partial \boldsymbol{\nu}_2} m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$  y  $m''_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \frac{\partial^2}{\partial \boldsymbol{\nu}_2 \partial \boldsymbol{\nu}_2} m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$  son funciones continuas y acotadas.

b)  $m_2(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4) = E([G_i(\alpha_0^\top \mathbf{t}_j) - G_i(\alpha_0^\top \mathbf{t}_i)][G_s(\alpha_0^\top \mathbf{t}_l) - G_s(\alpha_0^\top \mathbf{t}_s)] | \mathbf{t}_i = \boldsymbol{\nu}_1, \mathbf{t}_j = \boldsymbol{\nu}_2, \mathbf{t}_s = \boldsymbol{\nu}_3, \mathbf{t}_l = \boldsymbol{\nu}_4)$  y  $m'_2(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4) = \frac{\partial}{\partial \boldsymbol{\nu}_1} m_2(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)$  son funciones continuas y acotadas.

**M9.** El núcleo  $K : \mathbb{R} \rightarrow \mathbb{R}$  es una función par, no negativa, acotada, Lipschitz y de variación acotada, que satisface  $\int K(u) du = 1$  y  $\int u^2 K(u) du < \infty$ .

### 3.5.1 Distribución asintótica de los estimadores de la primer propuesta

Para obtener la distribución asintótica de los estimadores de la primer propuesta necesitamos los siguientes lemas cuyas demostraciones se encuentran en el Apéndice.

**Lema 3.5.1.** Sea  $(y_i, \mathbf{x}_i, \mathbf{t}_i)$  observaciones independientes tales que  $y_i | (\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$  con  $\mu_i = H(\mathbf{x}_i^\top \boldsymbol{\beta} + \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_i))$  y  $\text{VAR}(y_i | (\mathbf{x}_i, \mathbf{t}_i)) = V(\mu_i)$ . Supongamos que  $\mathbf{t}_i$  son vectores aleatorios con distribución con soporte compacto  $\mathcal{T}$  y que se satisfacen **M1**, **M2**, **M5** y **M6**. Sean  $\tilde{\boldsymbol{\alpha}}$  y  $\tilde{\boldsymbol{\beta}}$  tales que  $\tilde{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}_0$  y  $\tilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ . Entonces  $\mathbf{B}_n \xrightarrow{p} \mathbf{B}$  donde  $\mathbf{B}$  es definida en **M3** y

$$\begin{aligned} \mathbf{B}_n &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}} + \hat{\eta}_{\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\alpha}}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\nu}_i(\tilde{\tau}_i) \hat{\nu}_i(\tilde{\tau}_i)^\top \\ &+ \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}} + \hat{\eta}_{\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}}(\tilde{\boldsymbol{\alpha}}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\vartheta}_i(\tilde{\tau}_i) \end{aligned}$$

donde

$$\hat{\vartheta}_i(\tau) = \left( \begin{array}{cc} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) & \frac{\partial^2}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}^\top} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) + \frac{\partial^2}{\partial u \partial \boldsymbol{\beta}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \mathbf{t}_i^\top \\ \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) + \frac{\partial^2}{\partial \boldsymbol{\beta} \partial u} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \mathbf{t}_i & \frac{\partial^2}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) + 2 \frac{\partial^2}{\partial \boldsymbol{\alpha} \partial u} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \mathbf{t}_i + \frac{\partial^2}{\partial u^2} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \mathbf{t}_i \mathbf{t}_i^\top \end{array} \right) \Big|_{(\boldsymbol{\alpha}, \boldsymbol{\beta}, u) = \tau}$$

**Lema 3.5.2.** Supongamos que se satisface **M1**, **M2**,  $E(w_1(\mathbf{x}) \|\mathbf{x}\|^3) < \infty$  y  $K$  es un núcleo simétrico y acotado. Si  $nh^4 \rightarrow 0$  y  $\frac{nh^2}{\log(1/h)} \rightarrow \infty$  cuando  $n \rightarrow \infty$  se tiene

$$\sup_{\mathbf{t} \in \mathcal{T}} |\hat{\eta}_{\boldsymbol{\alpha}_0 \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}) - \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}) - \frac{1}{nf_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t})} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t})| = o_p(n^{-1/2})$$

donde  $G_j(u)$  esta definido anteriormente.

**Teorema 3.5.3.** Supongamos que  $\{\mathbf{t}_i\}_{1 \leq i \leq n}$  son variables aleatorias con distribución en un conjunto compacto  $\mathcal{T}$ , el parámetro de suavizado satisface  $nh^4 \rightarrow 0$  y  $\frac{nh^2}{\log(1/h)} \rightarrow \infty$ . Si;

- i) se satisfacen **M1** a **M9**.
- ii)  $\hat{\boldsymbol{\alpha}}$  y  $\hat{\boldsymbol{\beta}}$  los estimadores definidos en (3.2.1) son consistentes.
- iii) se satisface la conclusión del Lema 3.5.2.

entonces se tiene,

a)

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\alpha}}^{(q-1)} - \boldsymbol{\alpha}_0^{(q-1)} \end{pmatrix} \xrightarrow{D} N(0, \mathbf{B}_1^{-1} \boldsymbol{\Sigma}_1 (\mathbf{B}_1^{-1})^\top)$$

donde  $\hat{\boldsymbol{\alpha}}^{(q-1)} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{q-1})$ ,  $\boldsymbol{\alpha}_0^{(q-1)} = (\alpha_{01}, \dots, \alpha_{0q-1})$  y  $\mathbf{B}_1$  y  $\boldsymbol{\Sigma}_1$  están definidas en **M3** y **M4** respectivamente

$$b) \sqrt{n}(\hat{\alpha}_q - \alpha_{0q}) \xrightarrow{p} 0$$

DEMOSTRACIÓN DEL TEOREMA 3.5.3. En primer lugar observemos que por el Lema 3.4.1. es suficiente realizar la demostración en el caso en que  $\alpha_0 = \mathbf{e}_q$ .

Sean  $(\hat{\alpha}, \hat{\beta})$  solución de  $F_n^1(\hat{\alpha}, \hat{\beta}, \hat{\eta}_{\hat{\alpha}, \hat{\beta}}) = 0$ , es decir  $\hat{\alpha}$  y  $\hat{\beta}$  satisfacen,

$$0 = \theta \begin{pmatrix} 0 \\ \hat{\alpha} \end{pmatrix} + \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \hat{\beta} + \hat{\eta}_{\hat{\alpha}, \hat{\beta}}(\hat{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\nu}_i(\hat{\tau}_i) \quad (3.25)$$

llamemos  $\mathbf{D}_n = \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \hat{\beta} + \hat{\eta}_{\hat{\alpha}, \hat{\beta}}(\hat{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\nu}_i(\hat{\tau}_i)$  mediante un desarrollo de Taylor alrededor de  $(\hat{\alpha}, \hat{\beta}, \hat{\alpha}^\top \mathbf{t}_i)$  es fácil ver que  $\mathbf{D}_n = \hat{\mathbf{U}}_n + \mathbf{B}_n \begin{pmatrix} \sqrt{n}(\hat{\beta} - \beta_0) \\ \sqrt{n}(\hat{\alpha} - \alpha_0) \end{pmatrix}$  donde

$$\begin{aligned} \hat{\mathbf{U}}_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \beta_0 + \hat{\eta}_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\nu}_i(\tau_{0i}) \\ \mathbf{B}_n &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \tilde{\beta} + \hat{\eta}_{\tilde{\alpha}, \tilde{\beta}}(\tilde{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\nu}_i(\tilde{\tau}_i) \hat{\nu}_i(\tilde{\tau}_i)^\top \\ &\quad + \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \tilde{\beta} + \hat{\eta}_{\tilde{\alpha}, \tilde{\beta}}(\tilde{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\vartheta}_i(\tilde{\tau}_i) \end{aligned}$$

donde  $\hat{\vartheta}_i(\tau)$  esta definida en Lema 3.5.1,  $\tilde{\alpha}$  es un punto intermedio entre  $\alpha_0$  y  $\hat{\alpha}$ ,  $\tilde{\beta}$  es un punto intermedio entre  $\beta_0$  y  $\hat{\beta}$  y  $\tilde{\tau}_i = (\tilde{\alpha}, \tilde{\beta}, \tilde{\alpha}^\top \mathbf{t}_i)$ . Por el Lema 3.5.1. tenemos que  $\mathbf{B}_n \xrightarrow{p} \mathbf{B}$  donde  $\mathbf{B}$  esta definida en **M3**. Luego debemos hallar la distribución asintótica de  $\hat{\mathbf{U}}_n$ . Sea

$$\mathbf{U}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i})$$

$\mathbf{U}_n$  es asintóticamente normal. Luego es suficiente analizar el comportamiento de  $\hat{\mathbf{U}}_n - \mathbf{U}_n$ , para esto consideremos la siguiente expansión  $\mathbf{U}_n - \hat{\mathbf{U}}_n = \mathbf{U}_{1n} + \mathbf{U}_{2n} + \mathbf{U}_{3n} + \mathbf{U}_{4n}$  donde,

$$\begin{aligned} \mathbf{U}_{1n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) [\nu_i(\tau_{0i}) - \hat{\nu}_i(\tau_{0i})] \\ \mathbf{U}_{2n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) [\hat{\eta}_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i) - \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)] [\nu_i(\tau_{0i}) - \hat{\nu}_i(\tau_{0i})] \\ \mathbf{U}_{3n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) [\hat{\eta}_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i) - \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)] \\ \mathbf{U}_{4n} &= \frac{1}{2n} \sum_{i=1}^n \chi_1(y_i, \mathbf{x}_i^\top \beta_0 + \xi_{in}) w_2(\mathbf{x}_i) \hat{\nu}_i(\tau_{0i}) [n^{1/4} (\hat{\eta}_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i) - \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i))]^2 \end{aligned}$$

donde  $\xi_{in}$  son puntos intermedios entre  $\eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)$  y  $\hat{\eta}_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)$ . Bajo **M1** y **M2** se tiene que  $\mathbf{U}_{4n} \xrightarrow{p} 0$ .



Para probar que  $\mathbf{U}_{2n} \xrightarrow{p} 0$ , en virtud de **M1** y **M2** bastará mostrar que  $\mathbf{U}_{2n}^* \xrightarrow{p} 0$  con

$$\mathbf{U}_{2n}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) [\hat{\eta}_{\alpha, \beta}(u) - \eta_{\alpha, \beta}(u)] \Big|_{\tau=\tau_{0i}} \mathbf{t}_i$$

Notemos que por el Lema 3.5.2. tenemos que  $\mathbf{U}_{2n}^* = \mathbf{U}_{21n} + o_p(1)$  donde

$$\mathbf{U}_{21n} = n^{-\frac{3}{2}} \sum_{1 \leq i, j \leq n} \frac{K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i)}{f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j)} \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \left[ \frac{\partial}{\partial u} \hat{\eta}_{\alpha, \beta}(u) - \frac{\partial}{\partial u} \eta_{\alpha, \beta}(u) \right] \Big|_{\tau=\tau_{0i}} \mathbf{t}_i$$

En el Lema 3.5.4., que se encuentra al finalizar la demostración de este teorema probaremos que  $\mathbf{U}_{21n} \xrightarrow{p} 0$ . Por lo tanto, sólo debemos analizar el comportamiento asintótico de  $\mathbf{U}_{1n}$  y  $\mathbf{U}_{3n}$ .

En primer lugar probaremos que  $\mathbf{U}_{1n} \xrightarrow{p} 0$ , definamos

$$J_n(v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) v(\mathbf{t}_i)$$

Sea  $\mathcal{V} = \{v \in \mathcal{C}^1(\mathcal{T}) : \|v\|_\infty \leq 1 \text{ } \|v'\|_\infty \leq 1\}$ . Observemos que, para cualquier medida de probabilidad  $\mathcal{Q}$ , el “bracketing number”  $N_{[]}(\epsilon, \mathcal{V}, L^2(\mathcal{Q}))$  y por lo tanto, el número de cubrimiento  $N(\epsilon, \mathcal{V}, L^2(\mathcal{Q}))$ , satisfacen

$$\log N(\epsilon/2, \mathcal{V}, L^2(\mathcal{Q})) \leq \log N_{[]}(\epsilon, \mathcal{V}, L^2(\mathcal{Q})) \leq C \epsilon^{-1},$$

para todo  $0 < \epsilon < 2$ , donde la constante  $C$  es independiente de la medida de probabilidad  $\mathcal{Q}$ . (ver corolario 2.7.2 en Van der Vaart y Wellner (1996)).

Consideremos la clase de funciones

$$\mathcal{F} = \{g_v(y, \mathbf{x}, \mathbf{t}) = \Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t})) w_2(\mathbf{x}) v(\mathbf{t}) \quad v \in \mathcal{V}\}.$$

Si  $g_v \in \mathcal{F}$  entonces  $g_v$  está acotada por  $M = \|\Psi\|_\infty \|w_2\|_\infty$  y además por **M5**  $E(g_v(y_i, \mathbf{x}_i, \mathbf{t}_i)) = 0$ .

Llamemos  $\|g\|_{\mathcal{Q}, 2} = (E_{\mathcal{Q}}(g^2))^{1/2}$ . Es fácil ver que dado  $v \in \mathcal{V}$  y  $\|v_s - v\|_{\mathcal{Q}, 2} \leq \epsilon$  se tiene que  $\|g_{v_s} - g_v\|_{\mathcal{Q}, 2} \leq M\epsilon$  entonces  $N(\epsilon M, \mathcal{F}, L^2(\mathcal{Q})) \leq N(\epsilon, \mathcal{V}, L^2(\mathcal{Q}))$ . Luego la clase de funciones tiene número de cubrimiento finito.

Consideremos  $\mathcal{G}$ , una clase de funciones y llamemos  $\mathcal{J}(\delta, \mathcal{G})$  a la integral,

$$\mathcal{J}(\delta, \mathcal{G}) = \sup_{\mathcal{Q}} \int_0^\delta \sqrt{1 + \log N(\epsilon \|G\|_{\mathcal{Q}, 2}, \mathcal{G}, L^2(\mathcal{Q}))} d\epsilon$$

donde el supremo se toma sobre todas las medidas de probabilidad discreta  $\mathcal{Q}$  y  $G$  satisface que  $\forall g \in \mathcal{G} \quad \|g\|_{\mathcal{Q}, 2} \leq G$  y  $\|G\|_{\mathcal{Q}, 2} > 0$ .

La función  $\mathcal{J}$  es decreciente,  $\mathcal{J}(0, \mathcal{G}) = 0$ ,  $\mathcal{J}(1, \mathcal{G}) < \infty$  y  $\mathcal{J}(\delta, \mathcal{G}) \rightarrow 0$  cuando  $\delta \rightarrow 0$  para la clase de funciones que satisfacen la condición uniforme de entropía. Además si  $\mathcal{G}_0 \subset \mathcal{G}$  se tiene que  $\mathcal{J}(\delta, \mathcal{G}_0) \leq \mathcal{J}(\delta, \mathcal{G})$ .

Sea  $\epsilon > 0$  y  $0 < \delta < 1$  consideremos la subclase de  $\mathcal{F}$ ,  $\mathcal{F}_\delta = \{g_v(y, \mathbf{x}, \mathbf{t}) \in \mathcal{F} \text{ con } \|g_v\|_\infty \leq \delta\}$  y

$$\hat{v}_0(t) = \left( \begin{array}{c} \frac{\partial}{\partial \beta} \eta_{\alpha, \beta}(s) - \frac{\partial}{\partial \beta} \hat{\eta}_{\alpha, \beta}(s) \\ \frac{\partial}{\partial \alpha} \eta_{\alpha, \beta}(s) - \frac{\partial}{\partial \alpha} \hat{\eta}_{\alpha, \beta}(s) + \left( \frac{\partial}{\partial u} \eta_{\alpha, \beta}(s) - \frac{\partial}{\partial u} \hat{\eta}_{\alpha, \beta}(s) \right) \mathbf{t} \end{array} \right) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t})}$$

entonces como  $\widehat{v}_0(\mathbf{t}_i) = \nu_i(\tau_{0i}) - \widehat{\nu}_i(\tau_{0i})$  por **M1** tenemos que  $\sup_{\mathbf{t} \in \mathcal{T}} |\widehat{v}_0(\mathbf{t})| \xrightarrow{p} 0$ . Luego para  $n$  suficientemente grande  $P(\widehat{v}_0 \in \mathcal{V} \text{ y } \|\widehat{v}_0\|_\infty < \delta) > 1 - \delta/2$ .

Por otro lado,

$$\sup_{g \in \mathcal{F}_\delta} \frac{1}{n} \sum_{i=1}^n g^2(y_i, \mathbf{x}_i, \mathbf{t}_i) \leq M\delta^2$$

entonces la desigualdad maximal para el número de cubrimiento implica que

$$\begin{aligned} P(|J_n(\widehat{v}_0)| > \varepsilon) &\leq P(|J_n(\widehat{v}_0)| > \varepsilon, \widehat{v}_0 \in \mathcal{V} \text{ y } \|\widehat{v}_0\|_\infty < \delta) + \delta \\ &\leq P\left(\sup_{g \in \mathcal{F}_\delta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n g(y_i, \mathbf{x}_i, \mathbf{t}_i) \right| > \varepsilon\right) + \delta \\ &\leq \varepsilon^{-1} E\left(\sup_{g \in \mathcal{F}_\delta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n g(y_i, \mathbf{x}_i, \mathbf{t}_i) \right|\right) + \delta \\ &\leq \varepsilon^{-1} D M \mathcal{J}(\delta, \mathcal{F}) + \delta \end{aligned}$$

donde  $D$  es una constante independiente de  $n$ .

Luego, usando que la clase de funciones  $\mathcal{F}$  satisface la condición de entropía uniforme se tiene que  $\lim_{\delta \rightarrow 0} \mathcal{J}(\delta, \mathcal{F}) = 0$  por lo tanto  $\mathbf{U}_{1n} = J_n(\widehat{v}_0) \xrightarrow{p} 0$  como queríamos probar.

Finalmente para concluir el análisis del comportamiento asintótico de  $\widehat{\mathbf{U}}_n - \mathbf{U}_n$ , debemos analizar el comportamiento asintótico de  $\mathbf{U}_{3n}$ , para ello utilizaremos el Lema 3.5.2.

$$\begin{aligned} \mathbf{U}_{3n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) [\widehat{\eta}_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) - \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) \left[ \frac{1}{n f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \right] \\ &+ o_p(1) \\ &= n^{-3/2} \sum_{i=1}^n \sum_{j=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)^{-1} G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \\ &+ o_p(1) \\ &= \mathbf{U}_{3n}^{(1)} + o_p(1) \end{aligned}$$

Sea  $R(y_i, \mathbf{x}_i, \mathbf{t}_i) = \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)^{-1}$  y  $\gamma(\mathbf{t}) = E(R(y_1, \mathbf{x}_1, \mathbf{t}_1) | \mathbf{t}_1 = \mathbf{t})$  entonces, se cumple  $\mathbf{U}_{3n}^{(1)} - \frac{1}{\sqrt{n}} \sum_{i=1}^n G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \gamma(\mathbf{t}_i) f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \xrightarrow{p} 0$  pues,

$$\begin{aligned} &\mathbf{U}_{3n}^{(1)} - \frac{1}{\sqrt{n}} \sum_{i=1}^n G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \gamma(\mathbf{t}_i) f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) = \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) [R(y_j, \mathbf{x}_j, \mathbf{t}_j) G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_j) - \gamma(\mathbf{t}_i) G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)] \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) - f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \right] \gamma(\mathbf{t}_i) G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) [R(y_j, x_j, \mathbf{t}_j) - \gamma(\mathbf{t}_j)] G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_j) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) [\gamma(\mathbf{t}_j) - \gamma(\mathbf{t}_i)] G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_j) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) [G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_j) - G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)] \gamma(\mathbf{t}_i) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_j - \boldsymbol{\alpha}_0^\top \mathbf{t}_i) - f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \right] G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \gamma(\mathbf{t}_i).
\end{aligned}$$

Con los mismos argumentos que en el Teorema 2.5.2, es fácil ver que cada sumando de la expresión anterior converge a 0 en probabilidad. Por lo tanto concluimos que

$$\widehat{\mathbf{U}}_n - \mathbf{U}_n + \frac{1}{\sqrt{n}} \sum_{i=1}^n G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \gamma(\mathbf{t}_i) f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \xrightarrow{p} 0,$$

como deseábamos, de donde  $\widehat{\mathbf{U}}_n$  es asintóticamente normal con media 0 y matriz de covarianzas  $\boldsymbol{\Sigma}$  definida al comenzar la Sección.

Para concluir la demostración, de (3.25) tenemos que

$$0 = P_{\widehat{\boldsymbol{\alpha}}} \widehat{\mathbf{U}}_n + P_{\widehat{\boldsymbol{\alpha}}} \mathbf{B}_n \sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \end{pmatrix} \quad (3.26)$$

como  $\boldsymbol{\alpha}_0 = \mathbf{e}_q$  tenemos que  $P_{\widehat{\boldsymbol{\alpha}}} \xrightarrow{p} P_{\boldsymbol{\alpha}}$  donde

$$P_{\boldsymbol{\alpha}} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

entonces si  $\mathbf{B}_n = \begin{pmatrix} \mathbf{B}_{n1} & \mathbf{B}_{n2} \\ \mathbf{B}_{n3} & \mathbf{B}_{n4} \end{pmatrix}$  donde  $\mathbf{B}_{n1} \in \mathbb{R}^{(p+q-1) \times (p+q-1)}$ ,  $\mathbf{B}_{n2}, \mathbf{B}_{n3}^\top \in \mathbb{R}^{(p+q-1) \times 1}$  y  $\mathbf{B}_{n4} \in$

$\mathbb{R}$  y supongamos además que  $\mathbf{B}_{n1} = \begin{pmatrix} \mathbf{B}_{n1}^1 & \mathbf{B}_{n1}^2 \\ \mathbf{B}_{n1}^3 & \mathbf{B}_{n1}^4 \end{pmatrix}$  con  $\mathbf{B}_{n1}^1 \in \mathbb{R}^{p \times p}$ . Llamemos

$P_{\widehat{\boldsymbol{\alpha}}}^{(p+q-1)} = \begin{pmatrix} p_{\widehat{\boldsymbol{\alpha}}_1} \\ \vdots \\ p_{\widehat{\boldsymbol{\alpha}}_{p+q-1}} \end{pmatrix}$  donde  $p_{\widehat{\boldsymbol{\alpha}}_i}$  corresponden a las filas de  $P_{\widehat{\boldsymbol{\alpha}}}$ . Entonces por el Lema 3.5.1.

$P_{\widehat{\boldsymbol{\alpha}}}^{(p+q-1)} \mathbf{B}_n \xrightarrow{p} \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix}$  donde  $\mathbf{B}_1$  está definida en **M3**.

Por otro lado, como  $\|\widehat{\boldsymbol{\alpha}}\| = 1$  entonces,

$$\begin{aligned}
1 &= [n^{-1/2} \sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \boldsymbol{\alpha}_0]^\top [n^{-1/2} \sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \boldsymbol{\alpha}_0] \\
1 &= \sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top [n^{-1/2} \sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + 2\boldsymbol{\alpha}_0] + 1
\end{aligned}$$

luego,  $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top(\hat{\boldsymbol{\alpha}} + \boldsymbol{\alpha}_0) = 0$ . Ahora juntando esta ecuación con (3.26) tenemos;

$$\begin{pmatrix} -P_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)}\hat{\mathbf{U}}_n \\ 0 \end{pmatrix} = \begin{pmatrix} P_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)}\mathbf{B}_n \\ 0 \end{pmatrix} \begin{pmatrix} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ \sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \end{pmatrix}$$

como  $\hat{\boldsymbol{\alpha}}$  es consistente y  $\mathbf{B}_1$  es no singular entonces  $P_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)}\mathbf{B}_n$  resulta asintóticamente inversible. Luego  $\sqrt{n}(\hat{\boldsymbol{\alpha}}_q - \boldsymbol{\alpha}_{0q}) \xrightarrow{p} 0$  y

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\alpha}}^{(q-1)} - \boldsymbol{\alpha}_0^{(q-1)} \end{pmatrix} \xrightarrow{D} N(0, \mathbf{B}_1^{-1}\boldsymbol{\Sigma}_1(\mathbf{B}_1^{-1})^\top)$$

donde  $\mathbf{B}_1$  y  $\boldsymbol{\Sigma}_1$  están definidas en M3 y M4.□

**Lema 3.5.4.** Supongamos que se satisfacen las hipótesis del Teorema 3.5.3 luego  $\mathbf{U}_{12n} \xrightarrow{p} 0$  con

$$\mathbf{U}_{12n} = \frac{1}{n^{3/2}} \sum_{1 \leq i, j \leq n} \frac{K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_i - \boldsymbol{\alpha}_0^\top \mathbf{t}_j)}{f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j)} \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \left[ \frac{\partial}{\partial u} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) - \frac{\partial}{\partial u} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \right] \Big|_{\tau=\tau_{0i}} \mathbf{t}_i$$

DEMOSTRACIÓN DEL LEMA 3.5.4.

Notemos que  $\mathbf{U}_{12n}$  puede escribirse como  $\mathbf{U}_{12n} = \frac{n-1}{n} \mathbf{U}_{12n}^{(1)} + \mathbf{U}_{12n}^{(2)}$  con

$$\begin{aligned} \mathbf{U}_{12n}^{(1)} &= \frac{\sqrt{n}}{n(n-1)} \sum_{i \neq j} \frac{K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_i - \boldsymbol{\alpha}_0^\top \mathbf{t}_j)}{f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j)} \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \hat{v}(\mathbf{t}_i) \\ \mathbf{U}_{12n}^{(2)} &= \frac{K(0)}{hn\sqrt{n}} \sum_{i=1}^n \frac{1}{f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j)} \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \hat{v}(\mathbf{t}_i) \end{aligned}$$

donde  $\hat{v}(\mathbf{t}_i) = [\frac{\partial}{\partial u} \hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) - \frac{\partial}{\partial u} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)]|_{\tau=\tau_{0i}} \mathbf{t}_i$ . Del hecho que  $nh^2 \rightarrow \infty$  y  $\|\hat{v}\|_\infty \xrightarrow{p} 0$  es fácil ver que  $\mathbf{U}_{12n}^{(2)} \xrightarrow{p} 0$ .

Para ver que  $\mathbf{U}_{12n}^{(1)} \xrightarrow{p} 0$  consideremos los procesos  $U_n(\cdot)$  y  $V_n(\cdot)$  definidos por

$$\begin{aligned} U_n(g_v) &= \frac{\sqrt{n}}{n(n-1)} \sum_{i \neq j} \frac{K\left(\frac{\boldsymbol{\alpha}_0^\top \mathbf{t}_i - \boldsymbol{\alpha}_0^\top \mathbf{t}_j}{h}\right)}{f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j)} \left[ g(y_i, \mathbf{x}_i, \mathbf{t}_i, y_j, \mathbf{x}_j, \mathbf{t}_j) - E(g(y, \mathbf{x}, \mathbf{t}, y', \mathbf{x}', \mathbf{t}') | \mathbf{t} = \mathbf{t}_i, \mathbf{t}' = \mathbf{t}_j) \right] v(\mathbf{t}_i) \\ V_n(g_v) &= \frac{\sqrt{n}}{n(n-1)} \sum_{i \neq j} \frac{K\left(\frac{\boldsymbol{\alpha}_0^\top \mathbf{t}_i - \boldsymbol{\alpha}_0^\top \mathbf{t}_j}{h}\right)}{f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j)} E(g(y, \mathbf{x}, \mathbf{t}, y', \mathbf{x}', \mathbf{t}') | \mathbf{t} = \mathbf{t}_i, \mathbf{t}' = \mathbf{t}_j) v(\mathbf{t}_i) \end{aligned}$$

con  $g(y_i, \mathbf{x}_i, \mathbf{t}_i, y_j, \mathbf{x}_j, \mathbf{t}_j) = \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)$  y donde  $(y, \mathbf{x}, \mathbf{t}) \sim (y', \mathbf{x}', \mathbf{t}')$  e independientes entre sí. Luego, bastará probar que

i)  $\frac{1}{h} U_n(g_{\hat{v}}) \xrightarrow{p} 0$

ii)  $\frac{1}{h} V_n(g_{\hat{v}}) \xrightarrow{p} 0$

En primer lugar veamos i). Consideremos la clase de funciones  $\mathcal{G}$  definida por

$$\mathcal{G} = \left\{ g_v(y_i, \mathbf{x}_i, \mathbf{t}_i, y_j, \mathbf{x}_j, \mathbf{t}_j) = K \left( \frac{\boldsymbol{\alpha}_0^\top \mathbf{t}_i - \boldsymbol{\alpha}_0^\top \mathbf{t}_j}{h} \right) \left[ \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \right. \right. \\ \left. \left. - E(\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t})) w_2(\mathbf{x}) G_j(\boldsymbol{\alpha}_0^\top \mathbf{t}) | \mathbf{t} = \mathbf{t}_i \mathbf{t}_j = \mathbf{t}_j) \right] v(\mathbf{t}_i) / f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j) \text{ con } v \in C^2[0, 1], \|v\|_{(2)} \leq 1 \right\}$$

donde  $\|v\|_{(2)}^2 = \|v\|_\infty^2 + \|v'\|_\infty^2 + \|v''\|_\infty^2$ . Observemos que podemos suponer que si  $g_v \in \mathcal{G}$ ,  $\|g_v\|_\infty \leq 1$  y  $g_v$  es simétrica, de lo contrario bastará con considerar

$$g^*(y_i, \mathbf{x}_i, \mathbf{t}_i, y_j, \mathbf{x}_j, \mathbf{t}_j) = \frac{g_v(y_i, \mathbf{x}_i, \mathbf{t}_i, y_j, \mathbf{x}_j, \mathbf{t}_j) + g_v(y_j, \mathbf{x}_j, \mathbf{t}_j, y_i, \mathbf{x}_i, \mathbf{t}_i)}{2}$$

y dividir a  $g^*$  por  $\|K\|_\infty \|\chi\|_\infty \|G_2\|_\infty / \inf_{\mathbf{t} \in \mathcal{T}} f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t})$ . Por otro lado consideremos las clases de funciones

$$\mathcal{K} = \left\{ k_h(\mathbf{t}, \boldsymbol{\tau}) = K \left( \frac{\boldsymbol{\alpha}_0^\top \mathbf{t} - \boldsymbol{\alpha}_0^\top \boldsymbol{\tau}}{h} \right), \quad \|\boldsymbol{\alpha}_0\| = 1 \right\} \\ \mathcal{V} = \{v \in C^2[0, 1], \quad \|v\|_{(2)} \leq 1\}$$

y sean  $P_n$  la medida empírica asociada a  $\{(y_i, \mathbf{x}_i, \mathbf{t}_i)\}_{1 \leq i \leq n}$  y  $T_n$  la medida empírica asociada a  $\{(y_i, \mathbf{x}_i, \mathbf{t}_i, y_j, \mathbf{x}_j, \mathbf{t}_j)\}_{1 \leq i \neq j \leq n}$ .

Por ser  $K$  de variación acotada, se tiene que  $\log N(\epsilon, T_n, \mathcal{K}, 1) \leq \epsilon^{-1/2}$  (Pollard (1984)). Por otra parte, del Teorema 2.7.1 de Van der Vaart y Wellner (1996) se deduce que  $\log N(\epsilon, P_n, \mathcal{V}, 1) \leq \epsilon^{-1/2}$ , de donde

$$\log N(\epsilon, T_n, \mathcal{G}, 1) \leq \log N(\epsilon, T_n, \mathcal{K}, 1) \log N(\epsilon, P_n, \mathcal{V}, 1)$$

con lo cual  $J(s) = \int_0^s \log N(\epsilon, T_n, \mathcal{G}, 1) d\epsilon \leq 4\sqrt{s}$ .

Usando los mismo argumentos que en el Teorema 5 de Nolan y Pollard (1988), consideremos

$$\tilde{g}_v(y, \mathbf{x}, \mathbf{t}, y', \mathbf{x}', \mathbf{t}') = g_v(y, \mathbf{x}, \mathbf{t}, y', \mathbf{x}', \mathbf{t}') - P(g_v(y, \mathbf{x}, \mathbf{t}, \cdot, \cdot, \cdot)) - P(g_v(\cdot, \cdot, \cdot, y', \mathbf{x}', \mathbf{t}')) - P \otimes P(g_v)$$

con  $g_v \in \mathcal{G}$  entonces  $U_n(g_v) = U_n(\tilde{g}_v) + 2n^{1/2}(P_n - P) \otimes P g_v$ .

Veamos primero que  $U_n(\tilde{g}_v)/h \xrightarrow{p} 0$ . Sea  $\epsilon > 0$  fijo, como  $\|\hat{v}\| \xrightarrow{p} 0$  tenemos que dado  $\delta > 0$  existe  $n_0$  tal que si  $n \geq n_0$  entonces

$$P(|U_n(\tilde{g}_v)/h| > \epsilon) \leq P(|U_n(\tilde{g}_v)/h| > \epsilon, \|\hat{v}\| \leq 1) + \delta \\ \leq P\left(\sup_{g_v \in \tilde{\mathcal{G}}_1} |U_n(\tilde{g}_v)| > h\epsilon\right) + \delta \\ \leq \frac{1}{h\epsilon} E\left(\sup_{g_v \in \tilde{\mathcal{G}}_1} |U_n(\tilde{g}_v)|\right) + \delta$$

donde  $\tilde{\mathcal{G}}_\delta = \{\tilde{g}_v : g_v \in \mathcal{G} \text{ y } \|v\|_2 \leq \delta\}$ . Por el Lema 4 de Nolan y Pollard (1988) existe una constante  $C$  tal que

$$E\left(\sup_{g_v \in \tilde{\mathcal{G}}_1} |U_n(\tilde{g}_v)|\right) \leq 2Cn^{-1/2} E\left(\theta_n + \tau_n J\left(\frac{\theta_n}{\tau_n}\right)\right)$$

con  $\theta_n = 1$  y  $\tau_n \leq \frac{1}{4}$  entonces  $E\left(\theta_n + \tau_n J\left(\frac{\theta_n}{\tau_n}\right)\right) \leq 3$ . Por lo tanto,

$$\frac{1}{h\varepsilon} E\left(\sup_{g_v \in \tilde{\mathcal{G}}_1} |U_n(\tilde{g}_v)|\right) \leq 6Cn^{-1/2}(h\varepsilon)^{-1}$$

con lo cual como  $nh^2 \rightarrow \infty$  tenemos que  $E\left(\sup_{g_v \in \tilde{\mathcal{G}}_1} |U_n(\tilde{g}_v)|\right)/h \rightarrow 0$  de donde

$$P\left(\frac{1}{h} |U_n(\tilde{g}_v)| > \varepsilon\right) \rightarrow 0.$$

Veamos ahora que  $h^{-1}n^{1/2}(P_n - P) \otimes Pg_{\tilde{v}} \xrightarrow{p} 0$ , para esto consideremos la clase de funciones  $\mathcal{G}_{h\delta} = \{g_{hv}(y, \mathbf{x}, \mathbf{t}) = \frac{1}{h}Pg_v(\cdot, \cdot, \cdot, y, \mathbf{x}, \mathbf{t}) \text{ con } g_v \in \mathcal{G} \text{ y } \|v\|_\infty \leq \delta\}$  es fácil ver que  $F \equiv 1$  es una mayorante para  $\mathcal{G}_{h\delta}$  y que

$$\begin{aligned} \sup_{g_{hv} \in \mathcal{G}_{h\delta}} n^{-1} \sum_{i=1}^n g_{hv}^2(y_i, \mathbf{x}_i, \mathbf{t}_i) &\leq \delta^2 \\ \log N(\varepsilon, P_n, \mathcal{G}_{h\delta}, 1) &\leq \log N(\varepsilon, P_n, \mathcal{V}_\delta, \delta) \leq \varepsilon^{-1/2} \end{aligned}$$

donde  $\mathcal{V}_\delta$  es la clase de funciones  $C^2[0, 1]$  con  $\|v\|_{(2)} \leq 1$  y  $\|v\|_\infty \leq \delta$ . Entonces tenemos que  $\mathcal{J}(\delta, \mathcal{G}_{h\delta}) \leq 2\sqrt{\delta}$  y por la desigualdad maximal para el número de cubrimiento concluimos que,

$$\begin{aligned} P(h^{-1}n^{1/2} |(P_n - P) \otimes Pg_{\tilde{v}}| > \varepsilon) &\leq P(h^{-1}n^{1/2} |(P_n - P) \otimes Pg_{\tilde{v}}| > \varepsilon, \|\hat{v}\| \leq \delta) + \delta \\ &\leq P\left(\sup_{g_{hv} \in \mathcal{G}_{h\delta}} |n^{1/2}(P_n - P)g_{hv}| > \varepsilon\right) + \delta \\ &\leq \frac{1}{\varepsilon} E\left(\sup_{g_{hv} \in \mathcal{G}_{h\delta}} |n^{1/2}(P_n - P)g_{hv}|\right) + \delta \\ &\leq \frac{1}{\varepsilon} D\mathcal{J}(\delta, \mathcal{G}_{h\delta}) + \delta \\ &\leq 2D\varepsilon^{-1}\sqrt{\delta} + \delta \end{aligned}$$

con  $D$  una constante independiente de  $\delta$ . Entonces  $P(h^{-1}n^{1/2} |(P_n - P) \otimes Pg_{\tilde{v}}| > \varepsilon) \xrightarrow{p} 0$ . Y por lo tanto hemos probado i).

Resta entonces ver ii). Primero observemos que utilizando los mismos argumentos que en Lema 3.5.2. podemos probar que

$$\frac{1}{h}V_n(g_v) - \frac{\sqrt{n}}{n} \sum_i E(g(y, \mathbf{x}, \mathbf{t}, y, \mathbf{x}, \mathbf{t}) | \mathbf{t} = \mathbf{t}_i) v(\mathbf{t}_i) \xrightarrow{p} 0$$

Finalmente, como la clase de funciones  $\mathcal{V}$  definida anteriormente, satisface la condición uniforme de entropía (ver Van der Vaart y Wellner (1996)) se tiene que el proceso  $Z_n$  definido

$$Z_n(v) = \frac{\sqrt{n}}{n} \sum_i E(g(y, \mathbf{x}, \mathbf{t}, y, \mathbf{x}, \mathbf{t}) | \mathbf{t} = \mathbf{t}_i) v(\mathbf{t}_i)$$

es Donsker. En particular, es ajustado luego como  $\|\hat{v}\| \xrightarrow{p} 0$  tenemos que  $P(|Z_n(\hat{v})| > \varepsilon) \rightarrow 0$ .  $\square$

### 3.5.2 Distribución asintótica de los estimadores de la segunda propuesta

En esta Sección necesitaremos la siguiente notación

$$\begin{aligned}\hat{\tau}_{R0i} &= (\hat{\alpha}_R, \hat{\beta}_R, \alpha_0^\top t_i) \\ \hat{\tau}_{Ri} &= (\hat{\alpha}_R, \hat{\beta}_R, \hat{\alpha}^\top t_i) \\ \lambda_i(\tau) &= \begin{pmatrix} \mathbf{x}_i \\ \frac{\partial}{\partial u} \eta_{a,b}(s)|_{(a,b,s)=\tau} t_i \end{pmatrix} \\ \hat{\lambda}_i(\tau) &= \begin{pmatrix} \mathbf{x}_i \\ \frac{\partial}{\partial u} \hat{\eta}_{a,b}(s)|_{(a,b,s)=\tau} t_i \end{pmatrix}\end{aligned}$$

**Lema 3.5.5.** Sea  $(y_i, \mathbf{x}_i, \mathbf{t}_i)$  observaciones independientes tales que  $y_i | (\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$  con  $\mu_i = H(\mathbf{x}_i^\top \beta + \eta_0(\alpha_0^\top \mathbf{t}_i))$  y  $\text{VAR}(y_i | (\mathbf{x}_i, \mathbf{t}_i)) = V(\mu_i)$ . Supongamos que  $\mathbf{t}_i$  son variables aleatorias con distribución con soporte compacto  $\mathcal{T}$  y que se satisfacen **M1**, **M2**, **M5** y **M6**. Sean  $\tilde{\alpha}$ ,  $\hat{\alpha}_R$ ,  $\tilde{\beta}$  y  $\hat{\beta}_R$  estimadores débilmente consistentes de  $\alpha_0$  y  $\beta_0$  respectivamente. Entonces  $\mathbf{C}_n \xrightarrow{p} \mathbf{C}$  donde  $\mathbf{C} = E_0 [\chi(y_1, \mathbf{x}_1^\top \beta_0 + \eta_0(\alpha_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_0) \lambda_1(\tau_0)^\top]$  y

$$\begin{aligned}\mathbf{C}_n &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \tilde{\beta} + \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\tilde{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\lambda}_i(\tilde{\tau}_{Ri}) \hat{\lambda}_i(\tilde{\tau}_{Ri})^\top \\ &+ \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \tilde{\beta} + \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\tilde{\alpha}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \hat{\lambda}_i^1(\tilde{\tau}_{Ri})\end{aligned}$$

con  $\tilde{\tau}_{Ri} = (\hat{\alpha}_R, \hat{\beta}_R, \tilde{\alpha}^\top \mathbf{t}_i)$  y

$$\hat{\lambda}_i^1(\tau) = \left( \begin{array}{c} 0 \\ \frac{\partial^2}{\partial u^2} \hat{\eta}_{\alpha, \beta}(u) \mathbf{t}_i \mathbf{t}_i^\top \end{array} \right) \Big|_{\tau=\tau}$$

La demostración del Lema 3.5.5. utiliza los mismos argumentos que la demostración del Lema 3.5.1., los detalles de la misma se encuentran en el Apéndice.

**Observación:** En la Sección anterior probamos que los estimadores de la primer propuesta son asintóticamente normales más precisamente hemos demostrado que

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\alpha} - \alpha_0 \end{pmatrix} = \begin{pmatrix} B_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n [\Psi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_0(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) - G_i(\alpha_0^\top \mathbf{t}_i) \gamma(\mathbf{t}_i) f_{\alpha_0}(\alpha_0^\top \mathbf{t}_i)] \\ 0 \end{pmatrix} + o_p(1)$$

es decir probamos que podemos linealizar los estimador en el sentido que existe  $\varphi$  tal que  $E(\varphi(y_1, \mathbf{x}_1, \mathbf{t}_1)) = 0$  y

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\alpha} - \alpha_0 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(y_i, \mathbf{x}_i, \mathbf{t}_i) + o_p(1). \quad (3.27)$$

Con el objetivo de probar que los estimadores de la segunda propuesta son asintóticamente normales supondremos que los estimadores iniciales necesarios para la obtención de estos pueden linealizarse,

es decir existe una función  $\varphi$  tal que  $E(\varphi(y_1, \mathbf{x}_1, \mathbf{t}_1)) = 0$  y  $\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}_0 \end{pmatrix}$  puede escribirse como (3.27).

Llamemos  $W = \Psi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}) + \tilde{\varphi}(y_1, \mathbf{x}_1, \mathbf{t}_1)$  donde

$$\begin{aligned} \tilde{\varphi}(y_1, \mathbf{x}_1, \mathbf{t}_1) &= E \left( \chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}) \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\alpha}} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(s) |_{(\boldsymbol{\alpha}, \boldsymbol{\beta}, s) = (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0^\top \mathbf{t}_1)} \\ \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(s) |_{(\boldsymbol{\alpha}, \boldsymbol{\beta}, s) = (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0^\top \mathbf{t}_1)} \end{pmatrix} \right) \\ &\times \varphi(y_1, \mathbf{x}_1, \mathbf{t}_1) + G_1(\boldsymbol{\alpha}_0^\top \mathbf{t}_1) f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_1) \tilde{\gamma}(\mathbf{t}_1), \end{aligned}$$

y  $\tilde{\gamma}(\mathbf{t}) = E(\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}) | \mathbf{t} = \mathbf{t}_1)$ . Notemos que por **M5** y (3.27),  $E(W) = 0$  y definamos,  $\Gamma = E[WW^\top]$ .

Para la demostración de la distribución asintótica de los estimadores de la segunda propuesta además de algunas de las hipótesis consideradas en la Sección necesitaremos los siguientes supuestos.

**M10.** Sea  $\mathbf{C} = E_0 [\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_0) \lambda_1(\tau_0)^\top]$ , la submatriz superior izquierda de  $\mathbf{C}$  de dimensión  $(p+q-1) \times (p+q-1)$  que llamaremos  $\mathbf{C}_1$  es no singular

**M11.** La submatriz superior izquierda de  $\Gamma$  de dimensión  $(p+q-1) \times (p+q-1)$  que llamaremos  $\Gamma_1$  es no singular.

**Teorema 3.5.6.** Supongamos que  $\{\mathbf{t}_i\}_{1 \leq i \leq n}$  son variables aleatorias con distribución en un conjunto compacto  $\mathcal{T}$ , el parámetro de suavizado satisface  $nh^4 \rightarrow 0$  y  $nh^2/\log(1/h) \rightarrow \infty$ . Si

- i) se satisfacen **M1**, **M2**, **M5** a **M11**,
- ii) los estimadores iniciales,  $\widehat{\boldsymbol{\alpha}}_R$  y  $\widehat{\boldsymbol{\beta}}_R$ , satisfacen (3.27),
- iii) los estimadores  $\widehat{\boldsymbol{\alpha}}$  y  $\widehat{\boldsymbol{\beta}}$  definidos en (3.2.2) son consistentes,
- iv) y se satisface la conclusión del Lema 3.5.2

entonces se tiene,

a)

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\alpha}}^{(q-1)} - \boldsymbol{\alpha}_0^{(q-1)} \end{pmatrix} \xrightarrow{D} N(0, \mathbf{C}_1^{-1} \Gamma_1 (\mathbf{C}_1^{-1})^\top)$$

donde  $\widehat{\boldsymbol{\alpha}}^{(q-1)} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_{q-1})$ ,  $\boldsymbol{\alpha}_0^{(q-1)} = (\alpha_{01}, \dots, \alpha_{0q-1})$  y  $\mathbf{C}_1$  y  $\Gamma_1$  están definidas en **M10** y **M11** respectivamente

b)  $\sqrt{n}(\widehat{\alpha}_q - \alpha_{0q}) \xrightarrow{P} 0$

**Observación:** Al igual que en la Sección 3.4, donde analizamos el comportamiento de los estimadores de la Propuesta 1 bajo transformaciones ortogonales, es fácil ver que los estimadores de la segunda propuesta preservan la misma propiedad, tomando el estimador inicial de  $\boldsymbol{\alpha}$  invariante por



transformaciones ortogonales. Debido a este hecho bastará realizar la demostración del Teorema 3.5.6 en el caso en que  $\alpha_0 = \mathbf{e}_q$ .

#### DEMOSTRACIÓN DEL TEOREMA 3.5.6

Sean  $(\hat{\alpha}, \hat{\beta})$  solución de  $G_n^1(\hat{\alpha}, \hat{\beta}, \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}) = 0$ , es decir  $\hat{\alpha}$  y  $\hat{\beta}$  satisfacen,

$$0 = \theta_n \begin{pmatrix} 0 \\ \hat{\alpha} \end{pmatrix} + \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \hat{\beta} + \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\hat{\alpha}^T t_i)) w_2(\mathbf{x}_i) \hat{\lambda}_i(\hat{\tau}_{Ri}) \quad (3.28)$$

llamemos  $D_n = \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \hat{\beta} + \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\hat{\alpha}^T t_i)) w_2(\mathbf{x}_i) \hat{\lambda}_i(\hat{\tau}_{Ri})$  mediante un desarrollo de Taylor alrededor de  $(\hat{\alpha}, \hat{\beta})$  es fácil ver que  $D_n = \hat{\mathbf{V}}_n + \mathbf{C}_n \begin{pmatrix} \sqrt{n}(\hat{\beta} - \beta_0) \\ \sqrt{n}(\hat{\alpha} - \alpha_0) \end{pmatrix}$  donde

$$\begin{aligned} \hat{\mathbf{V}}_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \beta_0 + \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\alpha_0^T t_i)) w_2(\mathbf{x}_i) \hat{\lambda}_i(\hat{\tau}_{R0i}) \\ \mathbf{C}_n &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \tilde{\beta} + \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\tilde{\alpha}^T t_i)) w_2(\mathbf{x}_i) \hat{\lambda}_i(\tilde{\tau}_{Ri}) \hat{\lambda}_i(\tilde{\tau}_{Ri})^T \\ &\quad + \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \tilde{\beta} + \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\tilde{\alpha}^T t_i)) w_2(\mathbf{x}_i) \hat{\lambda}_i^1(\tilde{\tau}_{Ri}) \end{aligned}$$

donde  $\lambda_i^1(\tau)$  esta definida en Lema 3.5.4,  $\tilde{\alpha}$  es un punto intermedio entre  $\alpha_0$  y  $\hat{\alpha}$ ,  $\tilde{\beta}$  es un punto intermedio entre  $\beta_0$  y  $\hat{\beta}$  y  $\tilde{\tau}_{Ri} = (\hat{\alpha}_R, \hat{\beta}_R, \tilde{\alpha}^T t_i)$ . Por el Lema 3.5.5. tenemos que  $\mathbf{C}_n \xrightarrow{p} \mathbf{C}$  donde  $\mathbf{C}$  esta definida en **M10**. Luego debemos hallar la distribución asintótica de  $\hat{\mathbf{V}}_n$ . Sea

$$\mathbf{V}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i})$$

$\mathbf{V}_n$  es asintóticamente normal. Luego es suficiente analizar el comportamiento de  $\hat{\mathbf{V}}_n - \mathbf{V}_n$ , para esto consideremos la siguiente expansión  $\hat{\mathbf{V}}_n - \mathbf{V}_n = \mathbf{V}_{1n} + \mathbf{V}_{2n} + \mathbf{V}_{3n} + \mathbf{V}_{4n}$  donde,

$$\begin{aligned} \mathbf{V}_{1n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i)) w_2(\mathbf{x}_i) [\hat{\lambda}_i(\hat{\tau}_{0i}) - \lambda_i(\tau_{0i})] \\ \mathbf{V}_{2n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i)) w_2(\mathbf{x}_i) [\hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\alpha_0^T t_i) - \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i)] [\hat{\lambda}_i(\hat{\tau}_{R0i}) - \lambda_i(\tau_{0i})] \\ \mathbf{V}_{3n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \beta_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) [\hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\alpha_0^T t_i) - \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i)] \\ \mathbf{V}_{4n} &= \frac{1}{2n} \sum_{i=1}^n \chi_1(y_i, \mathbf{x}_i^T \beta_0 + \xi_{in}) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) [n^{1/4} (\hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\alpha_0^T t_i) - \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i))]^2 \\ \mathbf{V}_{5n} &= \frac{1}{2n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \beta_0 + \zeta_{in}) w_2(\mathbf{x}_i) [n^{1/4} (\hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\alpha_0^T t_i) - \eta_{\alpha_0, \beta_0}(\alpha_0^T t_i))]^2 [\hat{\lambda}_i(\hat{\tau}_{R0i}) - \lambda_i(\tau_{0i})] \end{aligned}$$

donde  $\xi_{in}$  y  $\zeta_{in}$  son puntos intermedios entre  $\eta_{\alpha_0, \beta_0}(\alpha_0^\top t_i)$  y  $\hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(\alpha_0^\top t_i)$ .

Usando argumentos análogos a los utilizados en el Teorema 3.5.3, tenemos que  $\mathbf{V}_{1n} \xrightarrow{p} 0$  por otro lado, de **M1**, **M2** y el hecho que  $n^{1/4} \|\hat{\alpha}_R - \alpha_0\| \xrightarrow{a.s.} 0$  y  $n^{1/4} \|\hat{\beta}_R - \beta_0\| \xrightarrow{a.s.} 0$  concluimos que  $\mathbf{V}_{in} \xrightarrow{p} 0$  para  $i = 4, 5$ .

Luego para obtener el comportamiento asintótico de  $\widehat{\mathbf{V}}_n - \mathbf{V}_n$  probaremos que

i)  $\mathbf{V}_{2n} \xrightarrow{p} 0$

ii) Existe una función  $\tilde{\varphi}$  tal que  $\mathbf{V}_{3n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\varphi}(y_i, \mathbf{x}_i, \mathbf{t}_i) + o_p(1)$ .

Por lo tanto probando i) y ii) podemos concluir que,

$$\widehat{\mathbf{V}}_n - \mathbf{V}_n - \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\varphi}(y_i, \mathbf{x}_i, \mathbf{t}_i) \xrightarrow{p} 0$$

entonces  $\widehat{\mathbf{V}}_n$  es asintóticamente normal con media 0 y matriz de covarianzas  $\Gamma$  definida al comenzar la Sección. Finalmente, la demostración se completa del mismo modo que el Teorema 3.5.3, multiplicando por  $P_{\hat{\alpha}}$  la igualdad (3.28) obtenemos

$$0 = P_{\hat{\alpha}} \widehat{\mathbf{V}}_n + P_{\hat{\alpha}} \mathbf{C}_n \sqrt{n} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\alpha} - \alpha_0 \end{pmatrix}$$

Luego de **M10** tenemos que  $\sqrt{n}(\hat{\alpha}_q - \alpha_{0q}) \xrightarrow{p} 0$  y

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\alpha}^1 - \alpha_0^1 \end{pmatrix} \xrightarrow{D} N(0, \mathbf{C}_1^{-1} \Gamma_1 (\mathbf{C}_1^{-1})^\top)$$

donde  $\mathbf{C}_1$  y  $\Gamma_1$  están definidas en **M10** y **M11**, respectivamente.

Entonces solo falta demostrar i) y ii), en primer lugar notemos que por el Lema 3.5.2. tenemos que

$$\begin{aligned} \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(u) - \eta_{\alpha_0, \beta_0}(u) &= \hat{\eta}_{\hat{\alpha}_R, \hat{\beta}_R}(u) - \hat{\eta}_{\alpha_0, \beta_0}(u) + \hat{\eta}_{\alpha_0, \beta_0}(u) - \eta_{\alpha_0, \beta_0}(u) \\ &= \frac{\partial}{\partial \alpha} \hat{\eta}_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, u)} (\hat{\alpha}_R - \alpha_0) \\ &\quad + \frac{\partial}{\partial \beta} \hat{\eta}_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, u)} (\hat{\beta}_R - \beta_0) \\ &\quad + \frac{1}{n f_{\alpha_0}(u)} \sum_{j=1}^n K_h \left( \frac{\alpha_0^\top \mathbf{t}_j - u}{h} \right) G_j(u) + o_P(n^{-1/2}) \end{aligned} \tag{3.29}$$

Luego con el objetivo de probar ii), de (3.29) obtenemos la siguiente expansión,

$$\mathbf{V}_{3n} = \mathbf{V}_{31n} + \mathbf{V}_{32n} + \mathbf{V}_{33n} + \mathbf{V}_{34n} + \mathbf{V}_{35n} + o_P(n^{-1/2})$$

donde

$$\begin{aligned}
\mathbf{V}_{31n} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) \frac{\partial}{\partial \boldsymbol{\alpha}} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_i)} \sqrt{n}(\widehat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}_0) \\
\mathbf{V}_{32n} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_i)} \sqrt{n}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) \\
\mathbf{V}_{33n} &= \frac{1}{n^{3/2}} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_j)} K_h(\boldsymbol{\alpha}_0^\top \mathbf{t}_i - \boldsymbol{\alpha}_0^\top \mathbf{t}_j) \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) G_j(\boldsymbol{\alpha}^\top \mathbf{t}_i) \\
\mathbf{V}_{34n} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) \frac{\partial}{\partial \boldsymbol{\alpha}} [\widehat{\eta}_{\alpha, \beta}(s) - \eta_{\alpha, \beta}(s)] \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_i)} \sqrt{n}(\widehat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}_0) \\
\mathbf{V}_{35n} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) \frac{\partial}{\partial \boldsymbol{\beta}} [\widehat{\eta}_{\alpha, \beta}(s) - \eta_{\alpha, \beta}(s)] \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_i)} \sqrt{n}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0)
\end{aligned}$$

Por **M1**, **M2** y el hecho que  $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}_0) = O_P(1)$  y  $\sqrt{n}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) = O_P(1)$  fácilmente obtenemos que  $\mathbf{V}_{3in} \xrightarrow{p} 0$  para  $i = 4, 5$ , con los mismos argumentos usados en el Teorema 3.5.3 tenemos que

$$\mathbf{V}_{33n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n G_i(\boldsymbol{\alpha}^\top \mathbf{t}_i) f_{\alpha_0}(\boldsymbol{\alpha}^\top \mathbf{t}_i) \widetilde{\gamma}(\mathbf{t}_i) + o_p(1)$$

con  $\widetilde{\gamma}(\mathbf{t}) = E(\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}) | \mathbf{t} = \mathbf{t}_1)$ . Luego, como los estimadores iniciales  $\widehat{\boldsymbol{\beta}}_R$  y  $\widehat{\boldsymbol{\alpha}}_R$  satisfacen la condición de linealización, (3.27) y por la Ley de los Grandes Números tenemos

$$\begin{aligned}
\mathbf{V}_{31n} + \mathbf{V}_{32n} &= E \left( \chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}) \left( \begin{array}{c} \frac{\partial}{\partial \boldsymbol{\alpha}} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_1)} \\ \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_1)} \end{array} \right) \right) \\
&\times \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(y_i, \mathbf{x}_i, \mathbf{t}_i) + o_p(1)
\end{aligned}$$

y por lo tanto

$$\mathbf{V}_{3n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{\varphi}(y_i, \mathbf{x}_i, \mathbf{t}_i) + o_p(1)$$

con

$$\begin{aligned}
\widetilde{\varphi}(y_i, \mathbf{x}_i, \mathbf{t}_i) &= E \left( \chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}) \left( \begin{array}{c} \frac{\partial}{\partial \boldsymbol{\alpha}} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_1)} \\ \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_1)} \end{array} \right) \right) \varphi(y_i, \mathbf{x}_i, \mathbf{t}_i) \\
&+ G_i(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) f_{\alpha_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i) \widetilde{\gamma}(\mathbf{t}_i).
\end{aligned}$$

Para concluir la demostración veamos i). De (3.29) podemos escribir a  $\mathbf{V}_{2n} = \mathbf{V}_{21n} + \mathbf{V}_{22n} + \mathbf{V}_{23n} + o_P(n^{-1/2})$  donde

$$\mathbf{V}_{21n} = \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \frac{\partial}{\partial \boldsymbol{\alpha}} \widehat{\eta}_{\alpha, \beta}(s) \Big|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_i)} [\widehat{\lambda}_i(\widehat{\tau}_{R0i}) - \lambda_i(\tau_{0i})] \sqrt{n}(\widehat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}_0)$$

$$\mathbf{V}_{22n} = \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\alpha, \beta}(s)(s)|_{(\alpha, \beta, s) = (\alpha_0, \beta_0, \alpha_0^\top \mathbf{t}_i)} [\hat{\lambda}_i(\hat{\tau}_{R0i}) - \lambda_i(\tau_{0i})] \sqrt{n}(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0)$$

$$\mathbf{V}_{23n} = \frac{1}{n^{3/2}} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{f_{\alpha_0}(\alpha_0^\top \mathbf{t}_j)} K_h(\alpha_0^\top \mathbf{t}_i - \alpha_0^\top \mathbf{t}_j) \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_j(\alpha_0^\top \mathbf{t}_i) [\hat{\lambda}_i(\hat{\tau}_{R0i}) - \lambda_i(\tau_{0i})]$$

Por **M1**, tenemos que  $|\hat{\lambda}_1(\hat{\tau}_{R01}) - \lambda_1(\tau_{01})| \xrightarrow{p} 0$  luego como  $\sqrt{n}(\hat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}_0) = O_P(1)$  y  $\sqrt{n}(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}_0) = O_P(1)$  y por **M2**, fácilmente obtenemos  $\mathbf{V}_{2in} \xrightarrow{p} 0$  para  $i = 1, 2$ . Finalmente la demostración de que  $\mathbf{V}_{23n} \xrightarrow{p} 0$  se encuentra en el Lema 3.5.7.  $\square$

**Lema 3.5.7.** Supongamos que se satisfacen las hipótesis del Teorema 3.5.6 luego

$$\frac{1}{n^{3/2}} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{f_{\alpha_0}(\alpha_0^\top \mathbf{t}_j)} K_h(\alpha_0^\top \mathbf{t}_i - \alpha_0^\top \mathbf{t}_j) \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_{\alpha_0, \beta_0}(\alpha_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) G_j(\alpha_0^\top \mathbf{t}_i) [\hat{\lambda}_i(\hat{\tau}_{R0i}) - \lambda_i(\tau_{0i})] \xrightarrow{p} 0$$

La demostración del Lema 3.5.7. es análoga a la demostración del Lema 3.5.4.

# Apéndice A

## Apéndice

### A.1 Demostración del Lema 2.5.1.

Sea  $\Delta_n(t) = \begin{pmatrix} \tilde{\beta}(t) - \beta_0 \\ \hat{\eta}(t) - \eta_0(t) \end{pmatrix}^\top$ , entonces como  $S_n^1(\tilde{\beta}(t), \hat{\eta}(t)) = 0$  y haciendo un desarrollo de Taylor de orden dos alrededor de  $(\beta_0, \eta_0(t))$  tenemos que,

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_0(t)) K_h(t - t_i) w_1(\mathbf{x}_i) \mathbf{z}_i^\top \\ &+ \Delta_n(t) \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \beta_0 + \eta_0(t)) K_h(t - t_i) w_1(\mathbf{x}_i) \mathbf{z}_i \mathbf{z}_i^\top \\ &+ \frac{1}{2nh} \sum_{i=1}^n \chi_1(y_i, \xi_i(t)) (\Delta_n(t) \mathbf{z}_i)^2 \mathbf{z}_i^\top \\ &= W_n(t) + A_n(t) \Delta_n(t) + C_n(t) \end{aligned}$$

con  $\xi_i(t)$  un punto intermedio entre  $\begin{pmatrix} \tilde{\beta}(t) \\ \hat{\eta}(t) \end{pmatrix}^\top \mathbf{z}_i$  y  $\begin{pmatrix} \beta_0 \\ \eta_0(t) \end{pmatrix}^\top \mathbf{z}_i$ . Luego vemos que,

- i)  $C_n(t) = O_P(\|\Delta_n(t)\|^2)$
- ii)  $A_n(t) = -A(t) + O_P(h^2 + (nh)^{-1/2} \log^{1/2}(\frac{1}{h}))$

donde  $A(t) = f_T(t) E(w_1(\mathbf{x}_1) \chi(y_1, \mathbf{x}_1^\top \beta_0 + \eta_0(t)) \mathbf{z}_1 \mathbf{z}_1^\top | t_1 = t)$ .

En primer lugar, veamos i). Por **N2**, el hecho que  $E(\|\mathbf{x}_1\|^3 w_1(\mathbf{x}_1)) < \infty$  y  $K$  es acotada tenemos que,

$$|C_n(t)| \leq \frac{1}{2} \|\Delta_n(t)\|^2 \|\chi_1\|_\infty \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t_i - t}{h}\right) \|\mathbf{z}_i\|^3 w_1(\mathbf{x}_i).$$

Luego, como  $\mathcal{T}$  es compacto y  $E(w_1(\mathbf{x}) \|\mathbf{z}\|^3) < \infty$  es fácil ver que  $\sup_{t \in \mathcal{T}} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t_i - t}{h}\right) \|\mathbf{z}_i\|^3 w_1(\mathbf{x}_i) = O_p(1)$  y por lo tanto queda probado i).

Para probar ii), usando el Lema **A.2** de Carrol, Fan, Gijbels y Wand (1997b) obtenemos que  $A_n(t) = E(A_n(t)) + O_P(c_n \log^{1/2}(\frac{1}{h}))$ . Calculemos entonces  $E(A_n(t))$ ,

$$\begin{aligned} E(A_n(t)) &= \frac{1}{nh} \sum_{i=1}^n E \left( K \left( \frac{t_i - t}{h} \right) \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t)) w_1(\mathbf{x}_i) \mathbf{z}_i \mathbf{z}_i^\top \right) \\ &= \frac{1}{nh} \sum_{i=1}^n E \left( K \left( \frac{t_i - t}{h} \right) E(\chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t)) w_1(\mathbf{x}_i) \mathbf{z}_i \mathbf{z}_i^\top | t_i) \right) \\ &= \frac{1}{nh} \sum_{i=1}^n E \left( K \left( \frac{t_i - t}{h} \right) g(t_i) \right) \\ &= E \left( \frac{1}{h} K \left( \frac{t_1 - t}{h} \right) g(t_1) \right) \end{aligned}$$

donde  $g(\tau) = E(w_1(\mathbf{x}_i) \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta}_0 + \eta_0(t)) \mathbf{z}_i \mathbf{z}_i^\top | t_i = \tau)$ . Haciendo cambio de variables, un desarrollo de Taylor y usando el hecho de que el núcleo es simétrico se tiene que  $E \left( \frac{1}{h} K \left( \frac{t_1 - t}{h} \right) g(t_1) \right) = g(t) f_T(t) + O(h^2)$  obteniendo así que

$$A_n(t) = f_T(t) E(w_1(\mathbf{x}_1) \chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(t)) \mathbf{z}_1 \mathbf{z}_1^\top | t_1 = t) + O_P(h^2 + (nh)^{-1/2} \log^{1/2}(1/h))$$

Como conclusión de i) y ii), tenemos que

$$0 = W_n(t) - A(t) \Delta_n(t) [1 + O_P(h^2 + (nh)^{-1/2} \log^{1/2}(\frac{1}{h}))] + O_P(\|\Delta_n(t)\|^2)$$

entonces

$$\Delta_n(t) + A(t)^{-1} W_n(t) = \Delta_n(t) O_P(h^2 + c_n \log^{1/2}(\frac{1}{h})) + O_P(\|\Delta_n(t)\|^2).$$

Por lo tanto, por **N1**, el hecho que  $n^{1/4} \|\tilde{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0\|_\infty \xrightarrow{p} 0$ ,  $nh^4 \rightarrow 0$  y  $\frac{nh^2}{\log(1/h)} \rightarrow \infty$  si  $n \rightarrow \infty$  concluye la demostración del Lema.  $\square$

## A.2 Demostración del Lema 3.4.3.

Para calcular  $G_i, G_{\beta,i}, G_{\alpha,i}$  y  $G_{u,i}$  basta con utilizar la Regla de la Cadena. Y así por ejemplo, tenemos

$$\begin{aligned} G_i &= \frac{\partial}{\partial \varepsilon} g(\varepsilon, \hat{\boldsymbol{\alpha}}_\varepsilon, \hat{\boldsymbol{\beta}}_\varepsilon, \hat{\boldsymbol{\alpha}}'_\varepsilon t_i) \Big|_{\varepsilon=0} = \frac{\partial}{\partial \varepsilon} g(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) \Big|_{(\varepsilon, \tau)=(0, \hat{\tau}_i)} + \frac{\partial}{\partial \boldsymbol{\alpha}} g(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) \Big|_{(\varepsilon, \tau)=(0, \hat{\tau}_i)} \frac{\partial}{\partial \varepsilon} \hat{\boldsymbol{\alpha}}_\varepsilon \Big|_{\varepsilon=0} \\ &+ \frac{\partial}{\partial \boldsymbol{\beta}} g(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) \Big|_{\varepsilon=0, \hat{\tau}_i = \hat{\tau}_i} \frac{\partial}{\partial \varepsilon} \hat{\boldsymbol{\beta}}_\varepsilon \Big|_{\varepsilon=0} + \frac{\partial}{\partial u} g(\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\beta}, u) \Big|_{(\varepsilon, \tau)=(0, \hat{\tau}_i)} \frac{\partial}{\partial \varepsilon} \hat{\boldsymbol{\alpha}}_\varepsilon \Big|_{\varepsilon=0} \mathbf{t}_i \end{aligned}$$

entonces

$$\begin{aligned} G_i &= \text{EIF}_j(\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau = \hat{\tau}_i} + \frac{\partial}{\partial \boldsymbol{\alpha}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau = \hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \\ &+ \frac{\partial}{\partial \boldsymbol{\beta}} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau = \hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\beta}}) + \frac{\partial}{\partial u} (\hat{\eta}_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u)) \Big|_{\tau = \hat{\tau}_i} \text{EIF}_j(\hat{\boldsymbol{\alpha}}) \mathbf{t}_i \end{aligned}$$

y del mismo modo para  $G_{\beta,i}$ ,  $G_{\alpha,i}$  y  $G_{u,i}$

$$\begin{aligned}
G_{\alpha,i} &= \text{EIF}_j \left( \frac{\partial}{\partial \alpha} \hat{\eta}_{\alpha,\beta}(u) \right) \Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial \alpha \partial \alpha} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha}) \\
&+ \frac{\partial^2}{\partial \beta \partial \alpha} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\beta}) + \frac{\partial^2}{\partial u \partial \alpha} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha}) \mathbf{t}_i \\
G_{\beta,i} &= \text{EIF}_j \left( \frac{\partial}{\partial \beta} \hat{\eta}_{\alpha,\beta}(u) \right) \Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial \alpha \partial \beta} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha}) \\
&+ \frac{\partial^2}{\partial \beta \partial \beta} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\beta}) + \frac{\partial^2}{\partial u \partial \beta} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha}) \mathbf{t}_i \\
G_{u,i} &= \text{EIF}_j \left( \frac{\partial}{\partial u} \hat{\eta}_{\alpha,\beta}(u) \right) \Big|_{\tau=\hat{\tau}_i} + \frac{\partial^2}{\partial \alpha \partial u} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha}) \\
&+ \frac{\partial^2}{\partial \beta \partial u} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\beta}) + \frac{\partial^2}{\partial u^2} (\hat{\eta}_{\alpha,\beta}(u)) \Big|_{\tau=\hat{\tau}_i} \text{EIF}_j(\hat{\alpha}) \mathbf{t}_i
\end{aligned}$$

Luego para obtener explícitamente  $G_i$ ,  $G_{\beta,i}$ ,  $G_{\alpha,i}$  y  $G_{u,i}$  es necesario calcular cada una de las funciones de influencias empíricas involucradas.

De (3.19) derivando respecto de  $\varepsilon$  y evaluando en  $\varepsilon = 0$  se tiene que,

$$\text{EIF}_j(\hat{\eta}_{\alpha,\beta}(u)) = \frac{\partial}{\partial \varepsilon} g(\varepsilon, \alpha, \beta, u) \Big|_{\varepsilon=0} = - \frac{K_h(\alpha^\top \mathbf{t}_j - u) \Psi(y_j, \mathbf{x}_j^\top \beta + \hat{\eta}_{\alpha,\beta}(u)) w_1(\mathbf{x}_j)}{\frac{1}{n} \sum_{i=1}^n K_h(\alpha^\top \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^\top \beta + \hat{\eta}_{\alpha,\beta}(u)) w_1(\mathbf{x}_i)}$$

Ahora nuevamente, en (3.19) primero derivo respecto de  $\alpha$ , entonces tengo que;

$$\begin{aligned}
0 &= \frac{1-\varepsilon}{n} \sum_{i=1}^n \frac{1}{h} K'_h(\alpha^\top \mathbf{t}_i - u) \mathbf{t}_i \Psi(y_i, \mathbf{x}_i^\top \beta + \hat{\eta}_{\alpha,\beta}^{(\varepsilon)}(u)) w_1(\mathbf{x}_i) \\
&+ \frac{1-\varepsilon}{n} \sum_{i=1}^n K_h(\alpha^\top \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^\top \beta + \hat{\eta}_{\alpha,\beta}^{(\varepsilon)}(u)) g_\alpha(\varepsilon, \alpha, \beta, u) w_1(\mathbf{x}_i) \\
&+ \varepsilon w_1(\mathbf{x}_j) \left\{ \frac{1}{h} K'_h(\alpha^\top \mathbf{t}_j - u) \mathbf{t}_j \Psi(y_j, \mathbf{x}_j^\top \beta + \hat{\eta}_{\alpha,\beta}^{(\varepsilon)}(u)) \right. \\
&+ \left. K_h(\alpha^\top \mathbf{t}_j - u) \chi(y_j, \mathbf{x}_j^\top \beta + \hat{\eta}_{\alpha,\beta}^{(\varepsilon)}(u)) g_\alpha(\varepsilon, \alpha, \beta, u) \right\} \tag{A.1}
\end{aligned}$$

Entonces derivando respecto de  $\varepsilon$ , evaluando en  $\varepsilon = 0$  y usando (A.1) con  $\varepsilon = 0$ ; podemos despejar  $\text{EIF}_j(\frac{\partial}{\partial \alpha}(\hat{\eta}_{\alpha,\beta}(u)))$  como,

$$\text{EIF}_j \left( \frac{\partial}{\partial \alpha} (\hat{\eta}_{\alpha,\beta}(u)) \right) = - \left\{ \frac{1}{n} \sum_{i=1}^n K_h(\alpha^\top \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^\top \beta + \hat{\eta}_{\alpha,\beta}(u)) w_1(\mathbf{x}_i) \right\}^{-1} \times$$

$$\begin{aligned}
& \times \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K'_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \mathbf{t}_i \chi(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_i) \text{EIF}_j(\hat{\eta}_{\alpha, \beta}(u)) \right. \\
& + \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi_1(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_i) \text{EIF}_j(\hat{\eta}_{\alpha, \beta}(u)) \frac{\partial}{\partial \boldsymbol{\alpha}}(\hat{\eta}_{\alpha, \beta}(u)) \\
& + \frac{1}{h} K'_h(\boldsymbol{\alpha}^T \mathbf{t}_j - u) \mathbf{t}_j \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_j) \\
& \left. + K_h(\boldsymbol{\alpha}^T \mathbf{t}_j - u) \chi(y_j, \mathbf{x}_j^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_j) \frac{\partial}{\partial \boldsymbol{\alpha}}(\hat{\eta}_{\alpha, \beta}(u)) \right\}
\end{aligned}$$

Del mismo modo para calcular  $\text{EIF}_j(\frac{\partial}{\partial \boldsymbol{\beta}}(\hat{\eta}_{\alpha, \beta}(u)))$  derivamos respecto de  $\boldsymbol{\beta}$  y luego respecto de  $\varepsilon$

$$\begin{aligned}
0 &= \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi_1(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) (\mathbf{x}_i^T + \frac{\partial}{\partial \boldsymbol{\beta}}(\hat{\eta}_{\alpha, \beta}(u))) w_1(\mathbf{x}_i) \text{EIF}_j(\hat{\eta}_{\alpha, \beta}(u)) \\
& + \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_i) \text{EIF}_j(\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\alpha, \beta}(u)) \\
& + K_h(\boldsymbol{\alpha}^T \mathbf{t}_j - u) \chi(y_j, \mathbf{x}_j^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) (\mathbf{x}_j^T + \frac{\partial}{\partial \boldsymbol{\beta}}(\hat{\eta}_{\alpha, \beta}(u))) w_1(\mathbf{x}_j)
\end{aligned}$$

entonces

$$\begin{aligned}
\text{EIF}_j(\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\alpha, \beta}(u)) &= - \left\{ \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_i) \right\}^{-1} \times \\
& \times \left\{ \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi_1(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) (\mathbf{x}_i^T + \frac{\partial}{\partial \boldsymbol{\beta}}(\hat{\eta}_{\alpha, \beta}(u))) w_1(\mathbf{x}_i) \text{EIF}_j(\hat{\eta}_{\alpha, \beta}(u)) \right. \\
& \left. + K_h(\boldsymbol{\alpha}^T \mathbf{t}_j - u) \chi(y_j, \mathbf{x}_j^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) (\mathbf{x}_j^T + \frac{\partial}{\partial \boldsymbol{\beta}}(\hat{\eta}_{\alpha, \beta}(u))) w_1(\mathbf{x}_j) \right\}
\end{aligned}$$

por último y del mismo modo que en los casos anteriores obtenemos,

$$\begin{aligned}
\text{EIF}_j(\frac{\partial}{\partial u} \hat{\eta}_{\alpha, \beta}(u)) &= - \left\{ \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_i) \right\}^{-1} \times \\
& \times \left\{ \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi_1(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_i) \frac{\partial}{\partial u}(\hat{\eta}_{\alpha, \beta}(u)) \text{EIF}_j(\hat{\eta}_{\alpha, \beta}(u)) \right. \\
& - \frac{1}{nh} \sum_{i=1}^n K'_h(\boldsymbol{\alpha}^T \mathbf{t}_i - u) \chi(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_i) \text{EIF}_j(\hat{\eta}_{\alpha, \beta}(u)) \\
& - \frac{1}{h} K'_h(\boldsymbol{\alpha}^T \mathbf{t}_j - u) \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_j) \\
& \left. + K_h(\boldsymbol{\alpha}^T \mathbf{t}_j - u) \chi(y_j, \mathbf{x}_j^T \boldsymbol{\beta} + \hat{\eta}_{\alpha, \beta}(u)) w_1(\mathbf{x}_j) \frac{\partial}{\partial u}(\hat{\eta}_{\alpha, \beta}(u)) \right\} \square
\end{aligned}$$



### A.3 Demostración del Lema 3.5.1.

En primer lugar notemos por

$$\vartheta_i(\tau) = \left( \begin{array}{cc} \frac{\partial^2}{\partial \beta \partial \beta^T} \eta_{\alpha, \beta}(u) & \frac{\partial^2}{\partial \alpha \partial \beta^T} \eta_{\alpha, \beta}(u) + \frac{\partial^2}{\partial u \partial \beta} \eta_{\alpha, \beta}(u) \mathbf{t}_i^T \\ \frac{\partial^2}{\partial \beta \partial \alpha} \eta_{\alpha, \beta}(u) + \frac{\partial^2}{\partial \beta \partial u} \eta_{\alpha, \beta}(u) \mathbf{t}_i & \frac{\partial^2}{\partial \alpha \partial \alpha^T} \eta_{\alpha, \beta}(u) + 2 \frac{\partial^2}{\partial \alpha \partial u} \eta_{\alpha, \beta}(u) \mathbf{t}_i + \frac{\partial^2}{\partial u^2} \eta_{\alpha, \beta}(u) \mathbf{t}_i \mathbf{t}_i^T \end{array} \right) \Big|_{\tau=\tau}$$

Luego, notar que  $\mathbf{B}_n$  puede escribirse como  $\mathbf{B}_n = \sum_{i=1}^6 \mathbf{B}_n^{(i)}$  donde

$$\begin{aligned} \mathbf{B}_n^{(1)} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \tilde{\beta} + \eta_0(\alpha_0^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) \nu_i(\tau_{0i})^T \\ \mathbf{B}_n^{(2)} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \tilde{\beta} + \hat{\eta}_{\tilde{\alpha}, \tilde{\beta}}(\tilde{\alpha}^T \mathbf{t}_i)) w_2(\mathbf{x}_i) [\hat{\nu}_i(\tilde{\tau}_i) \hat{\nu}_i(\tilde{\tau}_i)^T - \nu_i(\tau_{0i}) \nu_i(\tau_{0i})^T] \\ \mathbf{B}_n^{(3)} &= \frac{1}{n} \sum_{i=1}^n \chi_1(y_i, \mathbf{x}_i^T \tilde{\beta} + \xi_{1in}) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) \nu_i(\tau_{0i})^T [\hat{\eta}_{\tilde{\alpha}, \tilde{\beta}}(\tilde{\alpha}^T \mathbf{t}_i) - \eta_0(\alpha_0^T \mathbf{t}_i)] \\ \mathbf{B}_n^{(4)} &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \tilde{\beta} + \eta_0(\alpha_0^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \vartheta_i(\tau_{0i}) \\ \mathbf{B}_n^{(5)} &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \tilde{\beta} + \hat{\eta}_{\tilde{\alpha}, \tilde{\beta}}(\tilde{\alpha}^T \mathbf{t}_i)) w_2(\mathbf{x}_i) [\hat{\vartheta}_i(\tilde{\tau}_i) - \vartheta_i(\tau_{0i})] \\ \mathbf{B}_n^{(6)} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \tilde{\beta} + \xi_{2in}) w_2(\mathbf{x}_i) \vartheta_i(\tau_{0i}) [\hat{\eta}_{\tilde{\alpha}, \tilde{\beta}}(\tilde{\alpha}^T \mathbf{t}_i) - \eta_0(\alpha_0^T \mathbf{t}_i)] \end{aligned}$$

donde  $\xi_{1in}$  y  $\xi_{2in}$  son puntos intermedios entre  $\hat{\eta}_{\tilde{\alpha}, \tilde{\beta}}(\tilde{\alpha}^T \mathbf{t}_i)$  y  $\eta_0(\alpha_0^T \mathbf{t}_i)$ . Usando **M1**, **M2**, **M6** y el hecho que  $\tilde{\alpha}$  y  $\tilde{\beta}$  son consistentes, se tiene que  $\mathbf{B}_n^{(2)}$ ,  $\mathbf{B}_n^{(3)}$ ,  $\mathbf{B}_n^{(5)}$  y  $\mathbf{B}_n^{(6)}$  convergen a 0 en probabilidad.

Por el Teorema de convergencia mayorada tenemos que

$$E(n^{-1} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \tilde{\beta} + \eta_0(\alpha_0^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) \nu_i(\tau_{0i})^T) \rightarrow E(\chi(y, \mathbf{x}^T \beta_0 + \eta_0(\alpha_0^T \mathbf{t})) w_2(\mathbf{x}) \nu(\tau_0) \nu(\tau_0)^T)$$

luego por el Teorema 3 en el Capítulo 2 de Pollard (1984) obtenemos

$$\sup_{\beta} |n^{-1} \sum_{i=1}^n [\chi(y_i, \mathbf{x}_i^T \tilde{\beta} + \eta_0(\alpha_0^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \nu_i(\tau_{0i}) \nu_i(\tau_{0i})^T - E(\chi(y, \mathbf{x}^T \beta_0 + \eta_0(\alpha_0^T \mathbf{t})) w_2(\mathbf{x}) \nu(\tau_0) \nu(\tau_0)^T)]| \xrightarrow{a.s.} 0$$

entonces tenemos que  $\mathbf{B}_n^{(2)} \xrightarrow{p} \mathbf{B}$  con  $\mathbf{B}$  definida en **M3**. Con los mismos argumentos utilizados para probar la convergencia de  $\mathbf{B}_n^{(2)}$  y por **M5** obtenemos que  $\mathbf{B}_n^{(4)} \xrightarrow{p} 0$ .  $\square$

### A.4 Demostración del Lema 3.5.2.

Del hecho que,  $R_n^1(\alpha_0, \beta_0, \hat{\eta}_{\alpha_0, \beta_0}(u), u) = 0$  haciendo un desarrollo de Taylor alrededor de  $\eta_0(u)$  tenemos que,

$$\begin{aligned}
0 &= \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}_0^\top t_i - u) \Psi(y_i, \mathbf{x}_i \boldsymbol{\beta}_0 + \widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(u)) w_1(\mathbf{x}_i) \\
&= \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}_0^\top t_i - u) \Psi(y_i, \mathbf{x}_i \boldsymbol{\beta}_0 + \eta_0(u)) w_1(\mathbf{x}_i) \\
&\quad + \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}_0^\top t_i - u) \chi(y_i, \mathbf{x}_i \boldsymbol{\beta}_0 + \eta_0(u)) w_1(\mathbf{x}_i) (\widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(u) - \eta_0(u)) \\
&\quad + \frac{1}{2n} \sum_{i=1}^n K_h(\boldsymbol{\alpha}_0^\top t_i - u) \chi_1(y_i, \mathbf{x}_i \boldsymbol{\beta}_0 + \xi_n) w_1(\mathbf{x}_i) (\widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(u) - \eta_0(u))^2 \\
&= W_n(u) + A_n(u) (\widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(u) - \eta_0(u)) + C_n(u)
\end{aligned}$$

donde  $\xi_n$  es un punto intermedio entre  $\widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(u)$  y  $\eta_0(u)$ . Luego igual que en Lema 2.5.1 es fácil ver que,

$$\text{i) } C_n(u) = O_P(\|\widehat{\eta}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0} - \eta_0\|^2)$$

$$\text{ii) } A_n(u) = -A(u) + O_P(h^2 + (nh)^{-1/2} \log^{1/2}(\frac{1}{h}))$$

donde  $A(u) = f_{\boldsymbol{\alpha}_0}(u) E(\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(u)) w_1(\mathbf{x}_1) | \boldsymbol{\alpha}_0^\top \mathbf{t}_1 = u)$ . Finalmente, de **M1** y como  $nh^4 \rightarrow 0$ ,  $\frac{nh^2}{\log(1/h)} \rightarrow \infty$  si  $n \rightarrow \infty$  concluimos la demostración.  $\square$

## A.5 Demostración del Lema 3.5.5.

En primer lugar notemos por

$$\lambda_i^1(\tau) = \left( \begin{array}{c} 0 \\ \frac{\partial^2}{\partial u^2} \eta_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(u) \mathbf{t}_i \mathbf{t}_i^\top \end{array} \right) \Big|_{\tau=\tau}$$

Luego, notar que  $\mathbf{C}_n$  puede escribirse como  $\mathbf{C}_n = \sum_{i=1}^6 \mathbf{C}_n^{(i)}$  donde

$$\begin{aligned}
\mathbf{C}_n^{(1)} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}} + \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) \\
\mathbf{C}_n^{(2)} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}} + \widehat{\eta}_{\widehat{\boldsymbol{\alpha}}_R, \widehat{\boldsymbol{\beta}}_R}(\widetilde{\boldsymbol{\alpha}}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) [\widehat{\lambda}_i(\widetilde{\tau}_i) - \lambda_i(\tau_{0i})] \\
\mathbf{C}_n^{(3)} &= \frac{1}{n} \sum_{i=1}^n \chi_1(y_i, \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}} + \xi_{1in}) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i}) [\widehat{\eta}_{\widehat{\boldsymbol{\alpha}}_R, \widehat{\boldsymbol{\beta}}_R}(\widetilde{\boldsymbol{\alpha}}^\top \mathbf{t}_i) - \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)] \\
\mathbf{C}_n^{(4)} &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}} + \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i^1(\tau_{0i}) \\
\mathbf{C}_n^{(5)} &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}} + \widehat{\eta}_{\widehat{\boldsymbol{\alpha}}_R, \widehat{\boldsymbol{\beta}}_R}(\widetilde{\boldsymbol{\alpha}}^\top \mathbf{t}_i)) w_2(\mathbf{x}_i) [\lambda_i^1(\widetilde{\tau}_i) - \lambda_i^1(\tau_{0i})] \\
\mathbf{B}_n^{(6)} &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}} + \xi_{2in}) w_2(\mathbf{x}_i) \lambda_i^1(\tau_{0i}) [\widehat{\eta}_{\widehat{\boldsymbol{\alpha}}_R, \widehat{\boldsymbol{\beta}}_R}(\widetilde{\boldsymbol{\alpha}}^\top \mathbf{t}_i) - \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{t}_i)]
\end{aligned}$$

donde  $\xi_{1in}$  y  $\xi_{2in}$  son puntos intermedios entre  $\widehat{\eta}_{\widehat{\alpha}_R, \widehat{\beta}_R}(\widetilde{\alpha}^T \mathbf{t}_i)$  y  $\eta_0(\alpha_0^T \mathbf{t}_i)$ . Usando **M1**, **M2**, **M6** y el hecho que  $\widetilde{\alpha}$ ,  $\widehat{\alpha}_R$ ,  $\widehat{\beta}_R$  y  $\widetilde{\beta}$  son consistentes, se tiene que  $\mathbf{C}_n^{(2)}$ ,  $\mathbf{C}_n^{(3)}$ ,  $\mathbf{C}_n^{(5)}$  y  $\mathbf{C}_n^{(6)}$  convergen a 0 en probabilidad.

Por el Teorema de convergencia mayorada tenemos que

$$E(n^{-1} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \widetilde{\beta} + \eta_0(\alpha_0^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i})) \rightarrow E(\chi(y_1, \mathbf{x}_1^T \beta_0 + \eta_0(\alpha_0^T \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}))$$

luego por el Teorema 3 en el Capítulo 2 de Pollard (1984) obtenemos

$$\sup_{\beta} |n^{-1} \sum_{i=1}^n [\chi(y_i, \mathbf{x}_i^T \widetilde{\beta} + \eta_0(\alpha_0^T \mathbf{t}_i)) w_2(\mathbf{x}_i) \lambda_i(\tau_{0i})] - E(\chi(y_1, \mathbf{x}_1^T \beta_0 + \eta_0(\alpha_0^T \mathbf{t}_1)) w_2(\mathbf{x}_1) \lambda_1(\tau_{01}))]| \xrightarrow{a.s.} 0$$

entonces tenemos que  $\mathbf{C}_n^{(2)} \xrightarrow{p} \mathbf{C}$  con  $\mathbf{C}$  definida en **M10**. Con los mismos argumentos utilizados para probar la convergencia de  $\mathbf{C}_n^{(2)}$  y por **M5** obtenemos que  $\mathbf{C}_n^{(4)} \xrightarrow{p} 0$ .  $\square$

# Bibliografía

- [1] Bianco, A. y Boente, G. (2002). On the asymptotic behavior of one-step estimation. *Statistics and Probability Letters*, **60**, 33-47.
- [2] Bianco, A. y Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Planning and Inference*, **122**, 229-252.
- [3] Bianco, A., García Ben, M. y Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *The Canadian Journal of Statistics*, **33**, 511-528.
- [4] Bianco, A. y Yohai, V. (1995). Robust estimation in the logistic regression model. *Lecture Notes in Statistics*, **109**, 17-34. Springer-Verlag, New York.
- [5] Boente, G., He, X. y Zhou, J. (2006). Robust Estimates in Generalized Partially Linear Models. *The Annals of Statistics*, **34**, 2856-2878.
- [6] Cantoni, E. y Ronchetti, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.* **96** , 1022-1030.
- [7] Carroll, R., Fan, J., Gijbels, I. y Wand, M. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.*, **92** , 477-489.
- [8] Carroll, R., Fan, J., Gijbels, I. y Wand, M. (1997b). Generalized partially linear single-index models. *Reporte Técnico* .
- [9] Croux, C. y Haesbroeck, G. (2002). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics & Data Analysis*, **44**, 273-295.
- [10] Künsch, H., Stefanski, L. y Carroll, R. (1989). Conditionally unbiased bounded influence estimation in general regression models with applications to generalized linear models. *J. Amer. Assoc.* **84**, 460-466.
- [11] Härdle, W., Liang, H. y Gao, J. (2000). *Partially Linear Models*. Phisica-Verlag.
- [12] Härdle, W., Müller, M., Sperlich, S. y Werwatz, A. (2004). *Nonparametric ans Semiparametric Models*. Springer-Verlag.
- [13] Mallows, C. (1974). On Some Topics in Robustness. *Memorandum, Bell Laboratories, Murray Hill*, N.J.

- [14] Manchester, L. (1996). Empirical Influence for robust smoothing. *Austral. J. Statist.*, **38**, 275-296.
- [15] Maronna, R. A., Martin, R. D. y Yohai, V. (2006), “Robust statistics: Theory and methods“. John Wiley, New York.
- [16] McCullagh, P. y Nelder, J. (1989). Generalized Linear Models. (2nd ed.) *London: Chapman and Hall*.
- [17] Nelder, J. y Wedderburn, R. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society Series A*. **135** , 370-384.
- [18] Noland, D. y Pollard, D. (1988). Functional Limit Theorem for U-processes. *Annals of Statistics*, **16**, 3, 1291-1298.
- [19] Pollard, D. (1984). *Convergence of Stochastic Processes*. (Springer-Verlag, New York).
- [20] Severini, T. y Staniswalis, J. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.*, **89** , 501-511.
- [21] Severini, T. y Wong, W. (1992). Profile Likelihood and Conditionally parametric models. *Annals of Statistics*, **20**, 4, 1768-1802.
- [22] Stefanski, L., Carroll, R. y Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika*, **73**, 413-424.
- [23] Tamine, J. (2002). Smoothed influence function: another view at robust nonparametric regression. Discussion paper 62, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.
- [24] Tukey, J. (1970). *Exploratory Data Analysis*, mimeographed preliminar edition.
- [25] Van der Vaart, A. y Wellner, J. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. New York: Springer.
- [26] Van der Vaart, A. (1988). Estimating a real parameter in a class of semiparametric models. *Annals of Statistics*, **14**, 88-123.