



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

## **Estimadores robustos en modelos parcialmente no lineales.**

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

**Andrés Leandro Muñoz**

Director de tesis: Dra. Daniela Rodríguez.  
Consejero de estudios: Dra. Graciela Boente Boente.

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

Fecha de defensa: 3 de agosto de 2018.

## Estimadores robustos en modelos parcialmente no lineales

La inferencia estadística es el conjunto de métodos y técnicas que permiten inferir, a partir de la información empírica proporcionada por una muestra, cual es la relación de dependencia existente entre una variable respuesta y una o más variables independientes.

Algunos de los modelos que surgen con suma frecuencia en el análisis estadístico son, por un lado, los modelos de regresión no lineales y, por otro, los llamados modelos no paramétricos. Con el deseo de sacar provecho de lo mejor de cada uno de ellos surgen los modelos semiparamétricos parcialmente no lineales los cuales extienden y fusionan dichos modelos. Es bien sabido que los métodos de estimación clásica para los modelos recién mencionados son altamente sensibles a la presencia de datos atípicos y es entonces que entra en escena la estadística robusta.

En esta tesis, el objetivo principal es estudiar procedimientos de estimación robustos para el modelo de regresión parcialmente no lineal. A tal fin proponemos una familia de estimadores robustos obtenidos con un procedimiento de tres pasos. Para la propuesta de estimación estudiamos propiedades teóricas y su comportamiento computacional. Por un lado demostramos, bajo supuestos muy generales, la consistencia y la distribución asintótica. Por otro lado, estos resultados teóricos se completan con un estudio de simulación con el objetivo de evaluar el comportamiento de la propuesta robusta frente a distintas perturbaciones del modelo y tamaños de muestra finitos.

Finalmente, con el objetivo de ilustrar el uso de los estimadores propuestos, presentamos el ajuste de un modelo parcialmente no lineal a un conjunto reales.

*Palabras Claves* : Regresión no lineal. Modelos no paramétricos. Modelos semiparamétricos parcialmente no lineales. Estimación robusta.

## Robust Estimators for partly non linear models

The statistical inference is the set of technics and methods that allows to deduce, from the empirical information given by a sample, which is the relationship between a dependent variable and one or more independent variables.

Some of the models that arise frequently in the statistical analysis are, both, nonlinear regression models as well as the so called nonparametrical ones. In order to take advantage of the best of each one of them, there emerges the semiparametric partially nonlinear models. These last ones extend and merge such models. It is well known that the classic estimation methods for the already mentioned models, are extremely sensitive to the presence of atypical observations and therefore, comes the concept of robust statistics into scene.

The main aim of this thesis, was to study robust procedures of estimation for the partially nonlinear regression model. For that, we propose a family of robust estimators obtained by a three-step procedure. For the estimation proposal, we study theoretical properties and their computational behavior. On the one hand, we show under very general assumptions, the consistency and asymptotic distribution. On the other hand, those theoretical results are completed by a simulation study to evaluate the behavior of the robust proposal under different scenarios of contaminations and finite sample size.

Eventually, in order to illustrate the use of our proposal, we present a fit of a partially non-linear model to a set of real data.

Key words: Nonlinear regression. Nonparametric models. Semiparametric partially non-linear models. Robust estimation.

## Agradecimientos

No hubiese podido llegar a esta instancia sin la ayuda y cariño de mucha gente que me ha acompañado a lo largo de todo este camino. Este agradecimiento es para todos y cada uno de ellos.

En primer lugar, agradezco infinitamente a mi directora Daniela por haberme aceptado como su alumno, por su paciencia y por su siempre tan buena predisposición. A Graciela por su fundamental participación en este proyecto. Gracias!!

Por otra parte, agradezco a los jurados Ana Bianco, Marcela Svarc y María Dolores Jiménez Gamero por tomarse el tiempo de leer con detalle mi trabajo. Agradezco sus observaciones y correcciones las cuales han sido sumamente útiles para mí.

A mis compañeros y amigos de la facultad y, en particular, a toda la gente linda del Instituto del Cálculo por todos estos años juntos.

A mis compañeros y amigos del ITBA y de la UTN-FRA por el apoyo.

A mis amigos de toda la vida, a los de siempre, esos que sin preguntar nada están incondicionalmente sin titubear.

A mis viejos, José y Susana, por lo mucho que hicieron por mí. A mis abuelos, mis hermanos y a todos mis sobrinos.

Por último y muy especialmente a Fabi, el gran amor de mi vida. Nada de esto ni de aquello hubiese sido posible sin ella.

Gracias a todos de corazón !!

# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Nociones Previas</b>	<b>4</b>
2.1. Modelo de Regresión Noparamétrica . . . . .	4
2.1.1. Estimación Clásica . . . . .	4
2.1.2. Estimación Robusta . . . . .	6
2.2. Modelo de Regresión Paramétrica No Lineal . . . . .	7
2.2.1. Estimación Clásica . . . . .	7
2.2.2. Estimación Robusta . . . . .	8
<b>3. Modelos Semiparamétricos Parcialmente No Lineales</b>	<b>11</b>
3.1. Introducción . . . . .	11
3.2. Estimación Clásica . . . . .	12
3.3. Propuesta de Estimación Robusta . . . . .	13
<b>4. Propiedades Asintóticas</b>	<b>15</b>
4.1. Procesos Empíricos . . . . .	15
4.2. Consistencia . . . . .	18
4.3. Distribución Asintótica . . . . .	24
<b>5. Estudio de Simulación</b>	<b>35</b>
<b>6. Ejemplo con datos reales</b>	<b>44</b>
<b>7. Apéndice</b>	<b>51</b>
7.1. Demostraciones de la Sección 4.2 . . . . .	51
7.2. Demostraciones de la Sección 4.3 . . . . .	57



# Capítulo 1

## Introducción

El propósito del estudio de los métodos de regresión es construir modelos matemáticos que permitan explicar la relación de dependencia existente entre una variable respuesta y una o más variables independientes.

Una situación frecuente en el análisis estadístico es la de los modelos de regresión no lineales en los que los datos consisten en respuestas  $y_i$  que se sabe están relacionadas con un vector  $\mathbf{x}_i \in \mathbb{R}^p$  de covariables a través de la siguiente ecuación

$$y_i = g(\mathbf{x}_i, \beta_0) + \epsilon_i \quad 1 \leq i \leq n, \quad (1.1)$$

donde  $\beta_0 \subseteq \mathbb{R}^q$  es el vector de parámetros desconocidos a estimar,  $\epsilon_i$  son variables aleatorias independientes e independientes de las covariables  $\mathbf{x}_i$  para todo  $1 \leq i \leq n$  y, finalmente, se supone a la función de regresión  $g$  una función conocida.

Un estimador clásico de  $\beta_0$  para este modelo está basado en el método de mínimos cuadrados, es decir

$$\hat{\beta}_n = \arg \min_{\beta} \frac{1}{n} \sum_{i=1}^n (y_i - g(\mathbf{x}_i, \beta))^2.$$

Es conocido en la literatura que este estimador tienen propiedades óptimas bajo ciertos supuestos de regularidad, pero es muy sensible a observaciones atípicas. Si la distribución no es exactamente normal, el ajuste puede variar mucho. Por esta razón, se estudian estimadores robustos que sean más estables frente a distintas perturbaciones del modelo y que a su vez sean eficientes bajo normalidad.

Existen diferentes propuestas de estimadores robustos en modelos de regresión no lineales. Fraiman (1983) estudió estimadores con influencia acotada, Sakata y White (2001) introdujeron  $S$ -estimadores para modelos de regresión no lineal con observaciones dependientes, Stromberg (1993) propuso algoritmos para calcular  $MM$ -estimadores. Mas recientemente, Fasano (2009) extendió los  $M$  y  $MM$ -estimadores al caso de regresión no lineal y estudió sus propiedades asintóticas. Por otro lado, Bianco y Spano (2018) estudiaron este modelo bajo respuestas faltantes obteniendo resultados asintóticos relajando algunos supuestos.

Varios ejemplos y aplicaciones de estos modelos pueden verse en Bates y Watts (1988) y Seber y Wild (1989).

Otro modelo muy frecuente estudiado en la literatura y aplicado en diferentes situaciones es el modelo de regresión no paramétrica el cual supone que los datos  $(y_i, \mathbf{x}_i) \in \mathbb{R}^{1+p}$  están relacionados a través de una función de regresión  $m_0$ , de la cual sólo se suponen hipótesis de suavidad, es decir, no asume para  $m_0$  ninguna forma predefinida, mediante la siguiente ecuación

$$y_i = m_0(\mathbf{x}_i) + \epsilon_i \quad 1 \leq i \leq n, \quad (1.2)$$

donde  $\epsilon_i$  son variables aleatorias independientes e independientes de las covariables  $\mathbf{x}_i$  para todo  $1 \leq i \leq n$  y  $m_0$  la función a ser estimada. Es evidente que este modelo engloba prácticamente todas las posibles relaciones entre  $y$  con  $\mathbf{x}$ , incluso las propuestas lineal y no lineal. Estimadores noparamétricos clásicos para el modelo de regresión no paramétrica fueron introducidos por Naradaya (1964) y Watson (1964), en forma independiente, y Collomb (1981). Mientras que estimadores robustos para la función de regresión fueron estudiados por Boente y Fraiman (1989). Si bien estos modelos son más flexibles que los modelos lineales o no lineales, la regresión noparamétrica en varias variables posee dificultades teóricas y computacionales, conocidas como la “maldición de la dimensión”, que está asociada al hecho de que los entornos de un punto  $\mathbf{x} \in \mathbb{R}^p$  se hacen cada vez más malos para valores de  $p$  grandes. Es decir, se necesita un número exponencialmente mayor de datos para que dichos entornos contengan observaciones de la muestra.

Un primer paso en el intento de vincular las bondades de ambos modelos descriptos anteriormente es considerar los modelos conocidos como “modelos parcialmente lineales”, donde una parte de las covariables se modela linealmente y otra parte se modela noparamétricamente. En estos modelos, se supone que las observaciones siguen una relación del tipo

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i) + \epsilon_i \quad 1 \leq i \leq n \quad (1.3)$$

donde los errores son independientes e independientes de  $(\mathbf{x}_i^T, t_i) \in \mathbb{R}^{p+1}$ . Para los modelos parcialmente lineales fueron estudiadas propuestas basadas en polinomios locales, en núcleos o en vecinos más cercanos (ver por ejemplo, Engle et al., (1986); Heckman, (1986); Chen, (1988); Speckman, (1988); Ma et al., (2006), Fan y Li, (2004); Hu et al., (2004); Wang et al., (2005) y Härdle et al., (2000), para un review). Por otro lado, estimadores robustos basados en núcleos y en vecinos más cercanos con núcleos, fueron propuestos por Bianco y Boente (2004).

Li y Nie (2008) consideraron una extensión de los modelos parcialmente lineales al caso de modelos parcialmente no lineales en los cuales las respuestas satisfacen la siguiente relación

$$y_i = g(\mathbf{x}_i, \boldsymbol{\beta}_0) + \eta_0(t_i) + \epsilon_i \quad (1.4)$$

donde la función  $g$  es una función conocida salvo por el parámetro  $\boldsymbol{\beta}_0 \in \mathbb{R}^q$ ,  $\eta_0$  es una función desconocida con propiedades de regularidad, los errores son independientes e independientes de  $(\mathbf{x}_i^T, t_i) \in \mathbb{R}^{p+1}$ .

Los estimadores propuestos en Li y Nie (2008) se basan en perfiles no lineales y utilizan estimadores de mínimos cuadrados. La idea de mínimos cuadrados con perfiles no lineales se



relaciona con el principio de perfiles de verosimilitud generalizados introducidos por Severini y Wong (1992). Por estar basados en mínimos cuadrados o en promedios locales, todos los procedimientos mencionados anteriormente son muy sensibles a observaciones atípicas.

Tanto los modelos parcialmente lineales como los parcialmente no lineales, pueden incluirse en una familia más general de modelos que son los llamados modelos de regresión semiparamétricos. Esta clase de modelos son útiles para el análisis de datos porque mantienen la fácil interpretación de los modelos paramétricos y la flexibilidad de los modelos no paramétricos, pero evitando la maldición de la dimensión. Ruppert et al., (2003) estudian procedimientos de estimación y muchas aplicaciones de modelos de regresión semiparamétricos. Por otra parte, en modelos parcialmente lineales generalizados, o sea, cuando  $\mathbb{E}(y_i | \mathbf{x}_i, t_i) = H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))$  donde  $H$  es una función conocida,  $\eta_0$  es una función desconocida y  $\boldsymbol{\beta}_0$  es un parámetro desconocido, estimadores robustos basados en perfiles fueron dados por Boente et al., (2006) y por Boente y Rodriguez (2010).

El objetivo de este trabajo es aportar procedimientos de estimación robusta en modelos parcialmente no lineales (1.4). Para ello introduciremos estimadores basados en núcleos que incluirán pesos en las covariables los cuales permitirán controlar observaciones de alto leverage. Para mostrar las bondades de la propuesta se estudiarán sus propiedades asintóticas tales como consistencia y distribución asintótica. Asimismo, analizaremos el buen funcionamiento de la misma y su comparación con los métodos clásicos llevando a cabo un estudio de simulación bajo distintos tipos de contaminaciones. Por último, con el objetivo de ilustrar el uso de los estimadores propuestos, estudiaremos el ajuste de un modelo parcialmente no lineal al conjunto de datos meteorológicos “*airquality*” obtenidos del Departamento de Conservación del Estado de Nueva York los cuales se encuentran disponibles en la librería del software R “*robustbase*”.

Esta tesis se encuentra organizada de la siguiente manera. A lo largo del Capítulo 2 recordamos los principales métodos de estimación tanto clásica como robusta para modelos de regresión paramétricos (1.1) y no paramétricos (1.2). En el Capítulo 3, introducimos los modelos parcialmente no lineales (1.4), describimos procedimientos de estimación clásica ya estudiados y presentamos nuestra propuesta de estimación robusta. Por otra parte, en el Capítulo 4 probamos resultados de consistencia y presentamos resultados sobre su distribución asintótica. En el Capítulo 5 se presentan los resultados de un estudio de simulación cuyo principal objetivo consistió en analizar el comportamiento de la propuesta robusta y poder hacer una comparación respecto de sus versiones clásicas. Para ello consideramos tanto datos sin contaminar como distintos tipos de contaminaciones. Finalmente, en el Capítulo 6 presentamos una aplicación de nuestra propuesta a un conjunto de datos reales.

# Capítulo 2

## Nociones Previas

En este capítulo se describen diversos procedimientos existentes en la literatura, tanto clásicos como robustos de estimación para la función de regresión en los modelos no-paramétricos puros y del parámetro de regresión en los modelos paramétricos. Estos métodos de estimación que describiremos a continuación ayudarán a la comprensión de la propuesta de estimación para los modelos parcialmente no lineales que se desarrollará en el siguiente capítulo.

### 2.1. Modelo de Regresión Noparamétrica

La teoría y los métodos de suavizado han cobrado un gran auge en las últimas décadas unido al avance en materia computacional. Una revisión de los mismos se puede encontrar por ejemplo en los libros de Wand y Jones (1995), Fan y Gijbels (1996) y Loader (1999). En esta sección describiremos algunos de los métodos más relevantes de estimación, tanto clásicos como robustos, para la función de regresión bajo un modelo de regresión noparamétrico. Como se mencionó en la introducción este modelo está definido como

$$y = m_0(\mathbf{x}) + \epsilon, \tag{2.1}$$

donde  $(y, \mathbf{x}^T)$  son variables aleatorias en  $\mathbb{R}^{1+p}$ ,  $m_0 : \mathbb{R}^p \rightarrow \mathbb{R}$  una función suave y  $\epsilon$  una variable aleatoria en  $\mathbb{R}$  independiente de  $\mathbf{x}$ .

Resulta de gran interés la estimación de la función  $m_0$  a partir de una muestra  $(y_i, \mathbf{x}_i^T), 1 \leq i \leq n$ . En las siguientes subsecciones mostraremos los métodos clásicos de estimación de  $m_0$  y luego seguiremos por los métodos de estimación robustos.

#### 2.1.1. Estimación Clásica

Sea una muestra  $(y_i, \mathbf{x}_i^T), 1 \leq i \leq n$  bajo el modelo (2.1). Los primeros estimadores no paramétricos de regresión son los sencillos estimadores de tipo núcleo propuestos por Nadaraya (1964) y Watson (1964) de manera independiente. Para la descripción de los

mismos necesitaremos  $\mathcal{K}$ , una función núcleo en  $\mathbb{R}^p$ , es decir, una función no negativa e integrable, y  $h_n$  una sucesión de números positivos convergente a cero (llamada ventana). Luego, para cada  $\mathbf{x} \in \mathbb{R}^p$  el estimador de Nadaraya-Watson está dado por

$$\widehat{m}(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}) \left( \sum_{j=1}^n w_j(\mathbf{x}) \right)^{-1} y_i, \quad (2.2)$$

donde  $w_i(\mathbf{x})$  son los pesos basados en núcleos dados por

$$w_i(\mathbf{x}) = \mathcal{K} \left( \frac{\mathbf{x} - \mathbf{x}_i}{h_n} \right). \quad (2.3)$$

Una elección habitual para el núcleo  $\mathcal{K}$  es tomarlo como  $\mathcal{K}(\cdot) = K(\|\cdot\|)$ , con  $K$  una función núcleo univariada de soporte  $[0, 1]$ , y  $\|\cdot\|$  la norma euclídea en  $\mathbb{R}^p$ . Notar que en este caso el estimador dado en (2.2) queda definida como

$$\widehat{m}(\mathbf{x}) = \frac{\sum_{i=1}^n K \left( \frac{\|\mathbf{x} - \mathbf{x}_i\|}{h_n} \right) y_i}{\sum_{i=1}^n K \left( \frac{\|\mathbf{x} - \mathbf{x}_i\|}{h_n} \right)}. \quad (2.4)$$

Como se puede apreciar en (2.4), para cada  $\mathbf{x} \in \mathbb{R}^p$ ,  $\widehat{m}(\mathbf{x})$  resulta un promedio pesado de las variables  $y_i$  cuyos pesos dependen de las distancias euclídas entre  $\mathbf{x}$  y cada una de las observaciones  $\mathbf{x}_i$ . Por lo tanto sólo aquellas observaciones que disten de  $\mathbf{x}$  menos que el valor de  $h_n$  serán tenidas en cuenta para la estimación.

Dichos estimadores se han ido refinando y perfeccionando dentro de los denominados métodos de regresión basados en polinomios locales, convirtiéndose en uno de los métodos más empleados por diversos analistas en la actualidad. Por simplicidad en la escritura describiremos sólo el caso univariado. Si suponemos que la función de regresión  $m_0$  tiene  $l$  derivadas en un punto  $x_0 \in \mathbb{R}$ , con el fin de estimar  $m_0$  localmente mediante polinomios de grado  $l$  consideraremos el problema de mínimos cuadrados ponderados

$$\arg \min_{\alpha} \sum_{i=1}^n \sum_{j=0}^l (y_i - \alpha_j (x - x_0)^j)^2 w_i(x) \left( \sum_{k=1}^n w_k(x) \right)^{-1}. \quad (2.5)$$

Así, una vez obtenidas estimaciones de los coeficientes  $\alpha_j$  con  $j = 0, \dots, l$  vemos que la estimación del término independiente  $\widehat{\alpha}_0$  será un estimador de  $m_0(x_0)$  y el resto de los coeficientes  $\widehat{\alpha}_j$ ! proporcionarán estimaciones de sus derivadas  $m_0^{(j)}(x_0)$  con  $j = 1, \dots, l$ . Por otro lado, se puede ver fácilmente que si realizamos un ajuste con polinomios constantes obtenemos el estimador de Nadaraya-Watson dado en (2.2). También cuando el ajuste polinomial es de grado uno, se obtiene el denominado estimador lineal local.

### 2.1.2. Estimación Robusta

Como puede observarse en (2.5) los métodos de estimación clásicos están basados en el método de mínimos cuadrados y por lo tanto son altamente sensibles a cualquier variación de los datos. Pequeñas desviaciones podrían producir grandes fluctuaciones en el estimador volviéndolo completamente inestable.

Un primer acercamiento a los estimadores robustos en el caso de variables independientes fue dado por Tsybakov (1982) y Härdle (1984), quienes estudiaron propiedades asintóticas puntuales de una versión robusta del método de Nadaraya-Watson cuando la escala es conocida. Más adelante, Härdle y Tsybakov (1988) extendieron sus resultados previos a M-estimadores equivariantes de escala mediante estimación simultánea de la función de regresión y de escala. Boente y Fraiman (1989) consideraron estimadores noparamétricos robustos equivariantes por escala usando pesos de vecinos más cercanos y pesos basados en núcleos. Para ello definen  $m_M(\mathbf{x})$  la solución de

$$\mathbb{E} \left[ \psi \left( \frac{y - m_M(\mathbf{x})}{s(\mathbf{x})} \right) \mid \mathbf{x} = \mathbf{x} \right] = 0, \quad (2.6)$$

siendo  $s(\mathbf{x})$  una medida de escala robusta y  $\psi$  una función impar, acotada y continua,  $s(\mathbf{x})$  puede tomarse como la mediana local de los desvíos absolutos respecto de la mediana local de la distribución de la variable  $y$  condicional a  $\mathbf{x} = \mathbf{x}$ , que denotamos  $F(y|\mathbf{x} = \mathbf{x})$ . Notemos que aquí estamos realizando un abuso de notación utilizando la variable también como argumento de la función de distribución, entendemos que queda claro según el contexto. Mas precisamente,

$$s(\mathbf{x}) = \text{mediana} (|Y - \tilde{m}(\mathbf{x})| \mid \mathbf{x} = \mathbf{x}) = \text{MAD}(F(y|\mathbf{x} = \mathbf{x})), \quad (2.7)$$

donde  $\tilde{m}(\mathbf{x}) = \text{mediana} (F(y|\mathbf{x} = \mathbf{x}))$ . Por otro lado, consideran el estimador de la función de distribución empírica condicional de las variables respuestas definida como

$$\hat{F}(y|\mathbf{x} = \mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}) I_{(-\infty, y]}(y_i) \left( \sum_{j=1}^n w_j(\mathbf{x}) \right)^{-1}, \quad (2.8)$$

donde  $w_i(\mathbf{x})$  son los pesos basados en núcleos definidos en (2.3) tomando la misma ventana para todas las componentes.  $\hat{F}(y|\mathbf{x} = \mathbf{x})$  provee un estimador de  $F(y|\mathbf{x} = \mathbf{x})$  y fue estudiado por Cheng y Chu (1996).

Obervemos que un posible estimador de  $m_M(\mathbf{x})$  estudiado en Boente y Fraiman (1995) es considerar la mediana local calculada como la mediana de  $\hat{F}(y|\mathbf{x} = \mathbf{x})$ . Un hecho interesante de la mediana local es que, a diferencia de otros estimadores robustos, no necesita un estimador consistente de escala cuando ésta es desconocida.

Notemos que (2.6) es un funcional aplicado a  $F(y|\mathbf{x} = \mathbf{x})$ , luego resulta intuitivo imputar  $\hat{F}(y|\mathbf{x} = \mathbf{x})$  en (2.6) a fin de obtener un estimador de  $m_M(\mathbf{x})$ . De esta forma, se define un estimador robusto local para la función de regresión,  $\hat{m}_M(\mathbf{x})$ , como la solución en  $a$  de

$$\sum_{i=1}^n w_i(\mathbf{x}) \psi \left( \frac{y_i - a}{\widehat{s}(\mathbf{x})} \right) = 0, \quad (2.9)$$

donde  $\widehat{s}(\mathbf{x})$  es un estimador de escala robusto. Elecciones posibles de la función de escores  $\psi$  son la función de Huber o la función bicuadrada de Tukey, mientras que la escala  $\widehat{s}(\mathbf{x})$  puede tomarse como la mediana local de los desvíos absolutos respecto de la mediana local  $\text{MAD}(\widehat{F}(y|\mathbf{x} = \mathbf{x}))$ , es decir, la MAD local definida en (2.7) respecto de la distribución  $\widehat{F}(y|\mathbf{x} = \mathbf{x})$  definida en (2.8). Las propiedades de consistencia y distribución asintótica fueron estudiadas en Boente y Fraiman (1989) para el caso de observaciones independientes, mientras que órdenes de convergencia fuerte pueden verse en Boente y Fraiman (1990,1991).

## 2.2. Modelo de Regresión Paramétrica No Lineal

Como mencionamos en la introducción, dada una muestra aleatoria  $(y_i, \mathbf{x}_i^T) \in \mathbb{R}^{1+p}$ , con  $1 \leq i \leq n$  decimos que sigue un modelo de regresión no lineal si

$$y_i = g(\mathbf{x}_i, \boldsymbol{\beta}_0) + \varepsilon_i \quad (2.10)$$

donde  $g$  es una función conocida que no es lineal en el vector de parámetros  $\boldsymbol{\beta}_0 \in \mathcal{B} \subseteq \mathbb{R}^q$ , el cual es desconocido y es el objetivo de interés,  $\mathcal{B}$  es el espacio de parámetros y  $\varepsilon_i$  son los errores independientes de  $\mathbf{x}_i$ .

### 2.2.1. Estimación Clásica

En la estimación clásica se acostumbra suponer que los errores son independientes e idénticamente distribuidos con media cero y varianza  $\sigma_0^2$  desconocida. Como mencionamos en la introducción, el estimador clásico está basado en el método de mínimos cuadrados, es decir

$$\widehat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^n (y_i - g(\mathbf{x}_i, \boldsymbol{\beta}))^2. \quad (2.11)$$

Notemos que si derivamos la ecuación (2.11) obtenemos que  $\widehat{\boldsymbol{\beta}}_n$  debe cumplir con el siguiente sistema de ecuaciones

$$\sum_{i=1}^n r_i(\boldsymbol{\beta}) \nabla_{\boldsymbol{\beta}} g(\mathbf{x}_i, \boldsymbol{\beta}) = 0, \quad (2.12)$$

donde  $r_i(\boldsymbol{\beta}) = y_i - g(\mathbf{x}_i, \boldsymbol{\beta})$  son los residuos y  $\nabla_{\boldsymbol{\beta}} g(\mathbf{x}, \boldsymbol{\beta})$  es el gradiente de la función  $g$  con respecto a  $\boldsymbol{\beta}$ .

El estimador  $\widehat{\boldsymbol{\beta}}_n$  es óptimo siempre y cuando los errores tengan distribución normal. Si la distribución no es exactamente normal resulta sensible a la presencia de observaciones atípicas, lo que puede provocar que el ajuste no resulte adecuado. Por esta razón, se utilizan estimadores robustos que son más estables ante perturbaciones del modelo y qué, a su vez, son altamente eficientes bajo normalidad.

### 2.2.2. Estimación Robusta

Una buena parte de la teoría de robustez se ocupa del problema de estimación resistente a datos atípicos en modelos de regresión. Existen muchas propuestas de estimadores robustos de regresión y en esta sección nos ocuparemos sólo de algunas de ellas.

Una forma de controlar las observaciones atípicas en (2.11) es reemplazar la función cuadrática por una  $\rho$ -función en el sentido de Maronna et al. (2006) cuya definición damos a continuación.

**Definición 2.2.1.** Sea  $\rho$  una función definida en  $\mathbb{R}$ , decimos que  $\rho$  es una  $\rho$ -función en el sentido de Maronna et al. (2006) si satisface las siguientes condiciones

1.  $\rho(u)$  es creciente como función del  $|u|$ .
2.  $\rho(0) = 0$ .
3.  $\rho(u)$  es estrictamente creciente para  $u > 0$  tal que  $\rho(u) < \sup_u \rho(u)$ .
4. si  $\rho$  es acotada, se supone además que  $\sup_u \rho(u) = 1$ .

**Definición 2.2.2.** Llamaremos  $\Psi$ -función a cualquier función que sea la derivada de una  $\rho$ -función. Esto implica en particular que  $\Psi$  es impar y  $\Psi(u) \geq 0$  si  $u \geq 0$ .

Una clase importante de estimadores robustos para regresión lineal son los  $M$ -estimadores introducidos por Huber (1973). Dentro de esta clase podemos distinguir los  $S$ -estimadores introducidos por Rousseeuw y Yohai (1984) y los  $MM$ -estimadores propuestos por Yohai (1987). Estos estimadores tienen la propiedad de minimizar una función objetivo que depende de los datos únicamente a través de los residuos.

En lo que sigue resumiremos las definiciones de los  $M$ ,  $S$  y  $MM$ -estimadores de regresión para el modelo (2.10).

**Definición 2.2.3.** Dada una muestra aleatoria  $(y_i, \mathbf{x}_i^T) \in \mathbb{R}^{1+p}$  de tamaño  $n$  bajo el modelo (2.10) se define el  $M$ -estimador de regresión por

$$\hat{\beta}_M = \arg \min_{\beta \in \mathbb{R}^q} \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{y_i - g(\mathbf{x}_i, \beta)}{s_n} \right), \quad (2.13)$$

donde  $s_n$  es un estimador robusto preliminar de escala de los residuos y  $\rho$  es una  $\rho$ -función.

Para definir los  $S$ -estimadores necesitaremos, previamente, definir los llamados  $M$ -estimadores de escala.

**Definición 2.2.4.** Sea  $\rho$  una  $\rho$ -función en el sentido de Maronna et al. (2006), luego dado  $\mathbf{r}^T = (r_1, \dots, r_n)$  se define el correspondiente  $M$ -estimador de escala  $s_n(\mathbf{r})$  como la solución en  $s$  de la siguiente ecuación

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{r_i}{s} \right) = b, \quad (2.14)$$

donde  $b$  es una constante entre 0 y 1.

**Observación 2.2.5.** Si se desea que el estimador de escala  $s_n(\mathbf{r})$  definido en (2.14) estime la desviación típica bajo normalidad entonces la constante  $b$  se debe elegir como  $b = \mathbb{E}_\phi(\rho(u))$  con  $\phi$  la función de distribución de una normal estándar.

Por otro lado, Yohai y Zamar (1988) prueban que el punto de ruptura de  $s_n(\mathbf{r})$  es el mínimo entre  $b$  y  $1 - b$ .

**Definición 2.2.6.** Dada una muestra aleatoria  $(y_i, \mathbf{x}_i) \in \mathbb{R}^{1+p}$  de tamaño  $n$  bajo el modelo (2.10) se define el  $S$ -estimador de regresión por

$$\hat{\boldsymbol{\beta}}_S = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^q} s_n(\mathbf{r}(\boldsymbol{\beta})), \quad (2.15)$$

donde  $\mathbf{r}(\boldsymbol{\beta}) = (r_1(\boldsymbol{\beta}), \dots, r_n(\boldsymbol{\beta}))$  es el vector de los  $n$  residuos observados con componentes  $r_i(\boldsymbol{\beta}) = y_i - g(\mathbf{x}_i, \boldsymbol{\beta})$  y  $s_n(\mathbf{r})$  es la M-escala definida por (2.14).

$S$ -estimadores muy populares son los llamados *LMS* y *LTS* que se obtienen, los primeros, al minimizar la mediana de los residuos al cuadrado y, los segundos, al minimizar un subconjunto apropiado de la suma de los residuos al cuadrado. El punto de ruptura de estos estimadores para modelos no lineales fue estudiado por Stromberg y Ruppert (1992) y en Stromberg (1993) se presenta un algoritmo para su cómputo. Los  $S$ -estimadores no pueden ser estables y eficientes al mismo tiempo, pero tienen la importante ventaja de poder calcularse directamente a partir de los datos sin necesidad de estimadores iniciales de regresión ni de escala. Por esta razón, los  $S$ -estimadores son a menudo utilizados como estimadores iniciales en los algoritmos de cálculo de estimadores robustos.

Con el objetivo de obtener un estimador con alto punto de ruptura y eficiente bajo distribución normal, Yohai (1987) introdujo los MM-estimadores para modelos de regresión lineal. Luego, Fasano (2009) extendió este concepto para el caso de un modelo de regresión no lineal de la siguiente manera

**Paso 1** Calcular un estimador inicial  $\hat{\boldsymbol{\beta}}_1$  de  $\boldsymbol{\beta}_0$  con alto punto de ruptura.

**Paso 2** Calcular los residuos  $r_i(\hat{\boldsymbol{\beta}}_1) = y_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_1)$  y el M-estimador de escala basado en la muestra de los residuos observados  $s_n = s_n(\mathbf{r}(\hat{\boldsymbol{\beta}}_1))$  definido por (2.14) usando  $\rho_0$  una  $\rho$ -función acotada y  $b = 0.5$ .

**Paso 3** Sea  $\rho_1$  otra  $\rho$ -función acotada tal que cumpla  $\rho_1 \leq \rho_0$  luego el MM-estimador se define como

$$\hat{\boldsymbol{\beta}}_{MM} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^q} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{r_i(\boldsymbol{\beta})}{s_n} \right), \quad (2.16)$$

que satisface  $L(\hat{\boldsymbol{\beta}}_{MM}) \leq L(\hat{\boldsymbol{\beta}}_1)$  siendo  $L(\boldsymbol{\beta}) = \sum_{i=1}^n \rho_1 \left( \frac{r_i(\boldsymbol{\beta})}{s_n} \right)$ .

**Observación 2.2.7.** Si derivamos con respecto a  $\boldsymbol{\beta}$  la función objetivo definida en (2.16) entonces  $\hat{\boldsymbol{\beta}}_{MM}$  es la solución de

$$\sum_{i=1}^n \Psi_1 \left( \frac{r_i(\boldsymbol{\beta})}{s_n} \right) \nabla_{\boldsymbol{\beta}} g(\mathbf{x}_i, \boldsymbol{\beta}) = 0, \quad (2.17)$$

donde  $\nabla_{\beta}g(\mathbf{x},\beta)$  es el gradiente de la función  $g$  con respecto a  $\beta$  y  $\Psi_1(u) = \rho_1'(u)$ .

**Observación 2.2.8.** *Respecto al estimador inicial en el **Paso 1** distintas alternativas son posibles, la realidad es que puede utilizarse cualquier estimador consistente con punto de ruptura alto que no requiera de una escala previa para ser equivariante. En la práctica se utilizan en general  $S$ -estimadores definidos en (2.15) calibrados de tal manera de obtener un alto punto de ruptura.*



## Capítulo 3

# Modelos Semiparamétricos Parcialmente No Lineales

### 3.1. Introducción

La inferencia estadística generalmente se focaliza sobre modelos de regresión que son puramente paramétricos (2.10) o puramente no paramétricas (2.1).

Un modelo paramétrico razonable produce inferencias precisas, pero un modelo paramétrico erróneo posiblemente conducirá a conclusiones equivocadas. Paralelamente, los modelos noparamétricos, si bien están asociados con alta estabilidad, tienen menor precisión. Además el uso de estos encuentra en las aplicaciones a datos reales la dificultad conocida como la “maldición de la dimensión”, que está asociada al hecho de que los entornos de un punto  $\mathbf{x} \in \mathbb{R}^p$  se hacen cada vez más malos para valores de  $p$  grandes. Es decir, como ya fue mencionado anteriormente, se necesita un número exponencialmente mayor de datos para que dichos entornos contengan observaciones de la muestra. Por esta razón, se introdujeron los modelos semiparamétricos parcialmente lineales (1.3) que combinan una componente de regresión lineal y una componente no paramétrica. Luego se generalizan estos modelos a través de los modelos parcialmente no lineales, los cuales son una extensión natural de los parcialmente lineales permitiendo flexibilizar la componente paramétrica lineal en una función no lineal del parámetro.

En este sentido sea  $(y, \mathbf{x}^T, t)$ , con  $y \in \mathbb{R}$ ,  $\mathbf{x} \in \mathbb{R}^p$  y  $t \in \mathbb{R}$ , un vector aleatorio con distribución  $H_0$  y sea  $(y_i, \mathbf{x}_i^T, t_i)$  con  $1 \leq i \leq n$  una muestra aleatoria proveniente de dicha distribución, diremos que la muestra sigue un modelo parcialmente no lineal si

$$y_i = g(\mathbf{x}_i, \boldsymbol{\beta}_0) + \eta_0(t_i) + \epsilon_i \quad 1 \leq i \leq n, \quad (3.1)$$

donde por un lado,  $g$  es una función conocida salvo por el parámetro de regresión  $\boldsymbol{\beta}_0 \in \mathbb{R}^q$  y no lineal en dicho parámetro, por otro lado  $\eta_0$  es una función desconocida con ciertas propiedades de regularidad.

A lo largo de este trabajo asumiremos que los errores  $\epsilon_i$  son independientes, con distri-

bución simétrica  $F_0(\cdot) = \tilde{F}_0\left(\frac{\cdot}{\sigma_0}\right)$  e independientes de  $(\mathbf{x}_i, t_i) \in \mathbb{R}^{p+1}$ , es decir  $H_0(y, \mathbf{x}, t) = G_0(\mathbf{x}, t)F_0(y - g(\mathbf{x}, \beta_0) - \eta_0(t))$ .

Como hemos mencionado en la introducción, el objetivo principal de este trabajo será estudiar procedimientos de estimación robustos basados en núcleos para el parámetro de interés  $\beta_0$  e inferir un procedimiento de estimación para la función  $\eta_0$ . Antes de pasar a la propuesta en sí introduciremos algunas definiciones y un poco de notación necesarias para la comprensión de la misma.

Supongamos que tenemos una muestra aleatoria  $(y_i, \mathbf{x}_i^T, t_i) \in \mathbb{R}^{1+p+1}$ , con  $1 \leq i \leq n$ , que cumple el modelo parcialmente no lineal. Dado un valor fijo de  $\beta \in \mathbb{R}^q$  llamamos  $v_\beta = y - g(\mathbf{x}, \beta)$  y  $F_\beta(v|t = \tau)$  a la función de distribución de la variable  $v_\beta$  condicional a  $t = \tau$ . Consideremos  $v_{i,\beta} = y_i - g(\mathbf{x}_i, \beta)$  una muestra aleatoria con la misma distribución que  $v_\beta$  y definamos, análogamente a (2.8), el estimador de la función de distribución de la variable  $v_\beta$  condicional a  $t = \tau$  a partir de su distribución empírica como

$$\hat{F}_\beta(v|t = \tau) = \sum_{i=1}^n w_i(\tau) I_{(-\infty, v]}(v_{i,\beta}) \left( \sum_{j=1}^n w_j(\tau) \right)^{-1}, \quad (3.2)$$

donde  $w_i(\tau)$  son los pesos basados en núcleos calculados para las variables  $t_i$   $1 \leq i \leq n$  definidos en (2.3).

## 3.2. Estimación Clásica

Supongamos que tenemos una muestra aleatoria  $(y_i, \mathbf{x}_i, t_i)$ , con  $1 \leq i \leq n$ , bajo el modelo (3.1). En primer lugar notemos que en el enfoque clásico los supuestos habituales para los errores se reducen a asumir que  $\epsilon_i$  es una variable aleatoria con media cero y varianza  $\sigma_0^2$ . Bajo estos supuestos, estimadores clásicos fueron introducidos por Lie y Nie (2008). Esta propuesta puede ser descripta en dos pasos, del siguiente modo. Supongamos que fijamos un valor de  $\beta$ , luego podemos pensar que tenemos para cada  $\beta$  un modelo no paramétrico

$$v_{i,\beta} = \eta_\beta(t_i) + \epsilon_i \quad \forall 1 \leq i \leq n,$$

siendo  $v_{i,\beta} = y_i - g(\mathbf{x}_i, \beta)$ . Por lo tanto, podemos aplicar un método no paramétrico de estimación para obtener un estimador de la función  $\eta_\beta$ . Lie y Nie (2008) consideran el método de estimación basado en polinomios locales de orden uno descrito en (2.5) obteniendo, de esta manera, estimadores  $\hat{\eta}_\beta$  y  $\hat{\eta}'_\beta$  para  $\eta_\beta$  y  $\eta'_\beta$ , respectivamente. Es importante no perder de vista que hasta aquí la propuesta resulta un estimador para  $\eta_\beta$  que depende del valor fijado para  $\beta$ .

El segundo paso en el proceso de estimación consiste en proveer un estimador para el parámetro de regresión. Para ello, consideran la siguiente función de pérdida

$$Q(\beta) = \sum_{i=1}^n \left( y_i - g(\mathbf{x}_i, \beta) - \hat{\eta}_\beta(t_i) \right)^2. \quad (3.3)$$

De este modo, definen un estimador para el parámetro de regresión  $\beta_0$ , como aquel  $\hat{\beta}$  que minimiza (3.3) y finalmente definen un estimador para la función no paramétrica de regresión  $\eta_0$  como  $\hat{\eta} = \hat{\eta}_{\hat{\beta}}$ .

Li y Nie (2008) estudian propiedades como la consistencia y la distribución asintótica del estimador de regresión propuesto.

### 3.3. Propuesta de Estimación Robusta

Como podemos observar el procedimiento de estimación descrito en la sección anterior propuesto por Li y Nie (2008) está basado en el método de mínimos cuadrados o regresión local lineal y por lo tanto, como mencionamos anteriormente, es altamente sensible a datos atípicos. En atención a este hecho, nuestra propuesta de estimación está basada en una combinación del método descrito por Lie y Nie (2008), pero teniendo en cuenta procedimientos de estimación robustos en cada paso. Siguiendo este camino, nuestra propuesta puede ser descrita como un procedimiento de tres pasos que detallamos a continuación.

Sean  $\rho$ ,  $\rho_0$  y  $\rho_1$   $\rho$ -funciones como las definidas en 2.2.1 y, además, sea  $W$  una función real no negativa, la cual jugará el rol de controlar a las covariables  $\mathbf{x}_i \in \mathbb{R}^p$  de alto leverage. Esta última idea surge de observar que en la ecuación (2.12) una observación atípica puede tener influencia en la determinación del estimador de mínimos cuadrados tanto a través del residuo como del leverage. Además desde (2.17) notamos que  $\hat{\beta}$  puede ser aún afectado por observaciones con alto leverage.

Luego, dada  $(y_i, \mathbf{x}_i^T, t_i) \in \mathbb{R}^{1+p+1}$ , con  $1 \leq i \leq n$ , una muestra aleatoria bajo el modelo (3.1) nuestra propuesta de estimación consiste en llevar a cabo los siguientes tres pasos:

**Paso 1.** Para cada  $\tau$  y  $\beta$ , definamos  $\hat{\eta}_{\beta}(\tau)$  como la solución de:

$$\arg \min_{a \in \mathbb{R}} \sum_{i=1}^n \rho \left( \frac{y_i - g(\mathbf{x}_i, \beta) - a}{\hat{s}_{\beta}(\tau)} \right) w_i(\tau),$$

donde  $\hat{s}_{\beta}(\tau)$  es un estimador de escala local obtenido de la distribución empírica condicional definida en (3.2) y  $w_i(\tau) = \mathcal{K} \left( \frac{\tau - t_i}{h_n} \right)$  los pesos definidos en (2.3).

**Paso 2.** Para cada  $\beta$ , consideremos  $\hat{r}_i(\beta) = y_i - g(\mathbf{x}_i, \beta) - \hat{\eta}_{\beta}(t_i)$  y  $\hat{s}_0(\beta)$  la solución en  $s_0$  de  $\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{\hat{r}_i(\beta)}{s_0} \right) = b$ . Sea

$$s_n = \min_{\beta} \hat{s}_0(\beta).$$

Luego, si  $\rho_1 \leq \rho_0$ , definimos el estimador de  $\beta_0$  como

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \rho_1 \left( \frac{\hat{r}_i(\beta)}{s_n} \right) W(\mathbf{x}_i). \quad (3.4)$$

**Paso 3.** El estimador de la componente no paramétrica  $\eta_0$  se define como

$$\hat{\eta}(\tau) = \hat{\eta}_{\hat{\beta}}(\tau). \quad (3.5)$$

**Observación 3.3.1.** a) En cada uno de los pasos anteriores, las funciones  $\rho$ ,  $\rho_0$  y  $\rho_1$  pueden ser elegidas dentro de la familia de funciones bicuadradas de Tukey.

b) Para el **Paso 1** el estimador de escala local  $\hat{s}_{\beta}(\tau)$  puede tomarse como la mediana local de los desvíos absolutos respecto de la mediana local definida en (2.7) de la distribución condicional  $\hat{F}_{\beta}(v|t = \tau)$  definida en (3.2), es decir,

$$\hat{s}_{\beta}(\tau) = \text{mediana} (|y - g(\mathbf{x}, \beta) - \hat{m}_{\beta}(\tau)|t = \tau) = \text{MAD}(\hat{F}_{\beta}(v|t = \tau)), \quad (3.6)$$

donde  $\hat{m}_{\beta}(\tau) = \text{mediana} (\hat{F}_{\beta}(v|t = \tau))$ .

c) El **Paso 2** es un procedimiento de MM-estimación como el definido en (2.16) donde,  $s_n$  es la M-escala previa la cual se obtiene a partir de los pseudo-residuos  $\hat{r}_i(\beta) = y_i - g(\mathbf{x}_i, \beta) - \hat{\eta}_{\beta}(t_i)$ .

Por otro lado, como ya se mencionó anteriormente, con el objetivo de controlar observaciones con alto leverage introducimos en la ecuación (2.16) una función peso  $W$ , que será una función no negativa en las covariables  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . Una posibilidad para obtener el control buscado del gradiente de la función de regresión sería penalizar directamente las covariables de alta planca. En este caso, la función  $W$  podría elegirse de tal manera de penalizar la distancia de Mahalanobis robusta de las covariables. Para esto, será necesario estimar los parámetros de posición y escala de  $\mathbf{x}$ , digamos  $\mu$  y matriz de escala  $\Sigma$ . Así, si  $\hat{\mu}$  y  $\hat{\Sigma}$  son estimadores robustos de  $\mu$  y  $\Sigma$  respectivamente entonces podrían considerarse pesos de la forma

$$W(\mathbf{x}) = \kappa \left( (\mathbf{x} - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbf{x} - \hat{\mu}) \right),$$

donde  $\kappa$  es una función indicadora o características de algún intervalo.

d) En este sentido, notemos que si derivamos con respecto a  $\beta$  la función definida en (3.4), entonces  $\hat{\beta}$  es solución del siguiente sistema de ecuaciones

$$\sum_{i=1}^n \Psi_1 \left( \frac{\hat{r}_i(\beta)}{s_n} \right) W(\mathbf{x}_i) \left[ \nabla_{\beta} \left( g(\mathbf{x}_i, \beta) + \hat{\eta}_{\beta}(t_i) \right) \right]^T = 0, \quad (3.7)$$

donde  $\Psi_1 = \rho_1'$  y  $\nabla_{\beta}(\varphi) = \left( \frac{\partial \varphi}{\partial \beta_1}, \dots, \frac{\partial \varphi}{\partial \beta_q} \right)^T$ .

# Capítulo 4

## Propiedades Asintóticas

Sean  $(y_i, \mathbf{x}_i^T, t_i) \in \mathbb{R}^{1+p+1}$ , con  $1 \leq i \leq n$ , vectores aleatorios independientes con distribución  $H_0$ , que siguen un modelo de regresión parcialmente no lineal (3.1), tal como fue descrito en la sección anterior. A lo largo de este capítulo, estudiaremos las propiedades de los estimadores definidos en (3.4) y (3.5).

### 4.1. Procesos Empíricos

Para las demostraciones de los resultados asintóticos, tanto para la consistencia fuerte como para la convergencia en distribución, se han usado algunas herramientas de Procesos Empíricos. A continuación presentamos un breve resumen de las definiciones, lemas y teoremas sobre el tema que fueron necesarias en esta tesis. Para más detalles se puede ver Pollard (1984) y van der Vaart y Weller (1996).

**Definición 4.1.1.** Sea  $\mathcal{F}$  una clase de funciones, decimos que  $\mathcal{F}$  es permisible si puede ser indexada por un conjunto  $\Gamma$ , es decir,  $\mathcal{F} = \{f(\cdot, \gamma) : \gamma \in \Gamma\}$ .

**Definición 4.1.2.** Un función envolvente de una clase  $\mathcal{F}$  es una función medible  $F$  tal que  $|f| \leq F$  para toda  $f \in \mathcal{F}$ .

Sea  $L^r(\mathbb{P})$  el conjunto de funciones  $f : \Omega \rightarrow \mathbb{R}$  tales que  $\|f\|_{r, \mathbb{P}} = \mathbb{E}_{\mathbb{P}}(|f|^r)^{1/r} < \infty$ .

**Definición 4.1.3.** Dadas dos funciones  $l$  y  $u$ , el bracket  $[l, u]$  se define como el conjunto de todas las funciones  $f$  tal que  $l \leq f \leq u$ . Un  $\epsilon$ -bracket en  $L^r(\mathbb{P})$  es un bracket  $[l, u]$  tal que  $\|u - l\|_{r, \mathbb{P}} < \epsilon$ .

**Definición 4.1.4.** El número de bracketing  $N_{[\cdot]}(\epsilon, \mathcal{F}, L^r(\mathbb{P}))$  es el mínimo número de  $\epsilon$ -brackets en  $L^r(\mathbb{P})$  necesarios para cubrir a  $\mathcal{F}$ . La entropía con bracketing es el logaritmo del número de bracketing.

**Definición 4.1.5.** El número de cubrimiento  $N(\epsilon, \mathcal{F}, L^r(\mathbb{P}))$  es el mínimo número de bolas en  $L^r(\mathbb{P})$  de radio  $\epsilon$  necesarias para cubrir a  $\mathcal{F}$ . La entropía es el logaritmo del número de cubrimiento.

**Teorema 4.1.6.** *Supongamos que para todo  $\epsilon > 0$  existe una clase finita de funciones  $\mathcal{F}_\epsilon$  que contiene aproximaciones superiores e inferiores para cada  $f$  en  $\mathcal{F}$  para las cuales:*

$$f_{\epsilon,L} \leq f \leq f_{\epsilon,U} \quad \text{y} \quad \mathbb{E}_{\mathbb{P}}(f_{\epsilon,U} - f_{\epsilon,L}) < \epsilon.$$

*Luego,  $\sup_{f \in \mathcal{F}} |\mathbb{E}_{\mathbb{P}_n}(f) - \mathbb{E}_{\mathbb{P}}(f)| \xrightarrow{c.s.} 0$ , donde  $\mathbb{P}_n$  es la distribución empírica.*

**Teorema 4.1.7.** *Sea  $\mathcal{F}$  una clase de funciones permisibles con función envolvente  $F$  tal que  $\mathbb{E}_{\mathbb{P}}(F) < \infty$ . Si  $\mathbb{P}_n$  es la distribución empírica y  $N_{[\cdot]}(\epsilon, \mathcal{F}, L^1(\mathbb{P})) < \infty$  o  $\log N(\epsilon, \mathcal{F}, L^1(\mathbb{P})) = o_p(n)$  para cada  $\epsilon > 0$  fijo, entonces  $\sup_{f \in \mathcal{F}} |\mathbb{E}_{\mathbb{P}_n}(f) - \mathbb{E}_{\mathbb{P}}(f)| \xrightarrow{c.s.} 0$ .*

La siguiente observación es un caso particular del Corolario 2.7.2 de Van der Vaart and Wellner (1996).

**Observación 4.1.8.** *Sea  $\Omega$  un conjunto convexo y acotado en  $\mathbb{R}$  y sea  $\mathcal{H} = \{h \in C^1(\Omega) : \|h\|_\infty \leq 1 \quad \|h'\|_\infty \leq 1\}$  entonces para cualquier medida de probabilidad  $\mathbb{P}$  el número de bracketing  $N_{[\cdot]}(\mathcal{H}, \epsilon, L^2(\mathbb{P}))$  y por ende el número de cubrimiento  $N(\mathcal{H}, \epsilon, L^2(\mathbb{P}))$  satisfacen*

$$\log N(\epsilon/2, \mathcal{H}, L^2(\mathbb{P})) \leq \log N_{[\cdot]}(\epsilon, \mathcal{H}, L^2(\mathbb{P})) \leq K\epsilon^{-1}$$

*para  $0 < \epsilon < 2$  donde la constante  $K$  es independiente de la medida  $\mathbb{P}$ .*

**Teorema 4.1.9.** *Clase de funciones Lipschitz en un parámetro.*

*Sea  $\mathcal{F} = \{f_t : t \in T\}$  una clase de funciones tal que existe  $F$  fija que verifica  $|f_t(\mathbf{x}) - f_s(\mathbf{x})| \leq d(t, s)F(\mathbf{x})$  para alguna métrica  $d$  en  $T$ . Entoces para cualquier norma  $\|\cdot\|$  sobre  $\mathcal{F}$ ,*

$$N_{[\cdot]}(2\epsilon\|F\|, \mathcal{F}, \|\cdot\|) \leq N(\epsilon, T, d).$$

**Definición 4.1.10.** Diremos que una clase de funciones  $\mathcal{F}$  es Euclídea para la función envolvente  $F$  en  $L^1(\mathbb{P})$  si existen constantes  $A$  y  $V$  tales que

$$N(\epsilon\|F\|_{1,\mathbb{P}}, \mathcal{F}, L^1(\mathbb{P})) \leq A\epsilon^{-V}, \quad 0 < \epsilon \leq 1.$$

**Proposición 4.1.11.** *Si  $\mathcal{F}$  y  $\mathcal{G}$  son clases Euclídeas para las funciones envolventes  $F$  y  $G$  respectivamente, entonces*

*a) la clase  $\{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$  es Euclídea para la envolvente  $F + G$ .*

*b) la clase  $\{f \cdot g : f \in \mathcal{F}, g \in \mathcal{G}\}$  es Euclídea para la envolvente  $F \cdot G$ .*

**Definición 4.1.12.** Sea  $\mathcal{F}$  una familia de funciones  $f$ . Llamamos gráfico de  $f \in \mathcal{F}$  al conjunto  $G_f = \{(x, t) : 0 \leq t \leq f(x) \text{ ó } f(x) \leq t \leq 0\}$ .

**Definición 4.1.13.** Se dice que una clase de conjuntos  $\mathcal{D}$  sobre algún espacio  $S$  tiene discriminante polinomial si dado  $S_0$  un conjunto de  $N$  puntos de  $S$  entonces existen a lo sumo  $p(N)$  conjuntos de la forma  $S_0 \cap \mathcal{D}$ . Al polinomio  $p$  se lo llama discriminante polinomial y a  $\mathcal{D}$  clase polinomial de conjuntos.

**Proposición 4.1.14.** *Sea  $\mathcal{F}$  una clase de funciones con función envolvente  $F$  en  $L^1(\mathbb{P})$ . Si los gráficos de las funciones en  $\mathcal{F}$  forman una clase polinomial de conjuntos entonces la familia  $\mathcal{F}$  es Euclídea para la función envolvente  $F$ .*

**Definición 4.1.15.** Sea  $\mathcal{F}$  una familia de funciones con envolvente  $F$ . Se define a la integral de bracketing por  $J_{[\cdot]}(\delta, \mathcal{F}) = \int_0^\delta \sqrt{1 + \log(N_{[\cdot]}(2\epsilon\|F\|_{2,\mathbb{P}}, \mathcal{F}, L^2(\mathbb{P})))} d\epsilon$ .

**Definición 4.1.16.** Se dice que una familia  $\mathcal{F}$  de funciones con envolvente  $F$  satisface la condición de entropía de bracketing si  $\int_0^\infty \sqrt{\log(N_{[\cdot]}(2\epsilon\|F\|_{2,\mathbb{P}}, \mathcal{F}, L^2(\mathbb{P})))} d\epsilon < \infty$ .

**Proposición 4.1.17.** *La función  $J_{[\cdot]}(\delta, \mathcal{F})$  es creciente,  $J_{[\cdot]}(0, \mathcal{F}) = 0$ ,  $J_{[\cdot]}(1, \mathcal{F}) < \infty$  y  $J_{[\cdot]}(\delta, \mathcal{F}) \rightarrow 0$  cuando  $\delta \rightarrow 0$  si la familia satisface la condición de entropía de bracketing pues*

$$\begin{aligned} J_{[\cdot]}(\delta, \mathcal{F}) &\leq \int_0^\delta \sqrt{2} \left(1 + \sqrt{\log(N_{[\cdot]}(2\epsilon\|F\|_{2,\mathbb{P}}, \mathcal{F}, L^2(\mathbb{P})))}\right) d\epsilon \\ &\leq \sqrt{2} \left(\delta + \int_0^\delta \sqrt{\log(N_{[\cdot]}(2\epsilon\|F\|_{2,\mathbb{P}}, \mathcal{F}, L^2(\mathbb{P})))} d\epsilon\right). \end{aligned}$$

**Proposición 4.1.18.** *Desigualdad Maximal para Números de Bracketing.*

*Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$  vectores aleatorios i.i.d con distribución  $\mathbb{P}$ . Sea  $\mathcal{F}$  una clase de funciones  $\mathbb{P}$ -medibles con envolvente  $F$  tal que  $\|F\|_{2,\mathbb{P}} < \infty$ ,  $\mathbb{E}_{\mathbb{P}}(f) = 0$  para todo  $f \in \mathcal{F}$  y dado  $\delta > 0$ , sea*

$$a(\delta) = \frac{\delta\|F\|_{2,\mathbb{P}}}{\sqrt{1 + \log(N_{[\cdot]}(2\delta\|F\|_{2,\mathbb{P}}, \mathcal{F}, L^2(\mathbb{P})))}}$$

*Si  $\|f\|_{2,\mathbb{P}} < \delta\|F\|_{2,\mathbb{P}}$  para toda  $f \in \mathcal{F}$  entonces existe una constante  $D$ , independiente de  $n$ , tal que*

$$\begin{aligned} \left\| \sup_{f \in \mathcal{F}} |\sqrt{n}\mathbb{E}_{\mathbb{P}_n}(f)| \right\|_{1,\mathbb{P}} &\leq DJ_{[\cdot]}(\delta, \mathcal{F})\|F\|_{2,\mathbb{P}} + \sqrt{n}\mathbb{E}_{\mathbb{P}} \left( FI_{\{F > \sqrt{na}(\delta)\}} \right) \\ &\leq DJ_{[\cdot]}(1, \mathcal{F})\|F\|_{2,\mathbb{P}}. \end{aligned}$$

**Lema 4.1.19.** *Para cada  $n$ , sea  $\mathcal{F}_n$  una clase permisible de funciones con número de cubrimiento que satisface:*

$\sup_Q N(\epsilon, \mathcal{F}_n, L^1(Q)) \leq A\epsilon^{-W}$  para todo  $0 < \epsilon < 1$  con  $A$  y  $W$  constantes independientes de  $n$ .

*Sea  $\{\alpha_n\}$  una sucesión no creciente y positiva para la cual  $\frac{n\delta_n^2\alpha_n^2}{\log n} \rightarrow \infty$ . Si  $|f| \leq 1$  y  $(\mathbb{E}_{\mathbb{P}} f^2)^{\frac{1}{2}} \leq \delta_n$  para cada  $f \in \mathcal{F}_n$  luego:*

$$\frac{1}{\delta_n^2\alpha_n} \sup_{\mathcal{F}_n} |\mathbb{E}_{\mathbb{P}_n}(f) - \mathbb{E}_{\mathbb{P}}(f)| \xrightarrow{c.s.} 0$$

## 4.2. Consistencia

En primer lugar comenzaremos por estudiar la consistencia de Fisher de la propuesta presentada en el capítulo anterior. Sea  $(y, \mathbf{x}, t)$  un vector aleatorio que sigue el modelo (3.1), consideremos el siguiente funcional,

$$S(\boldsymbol{\beta}, a, \tau) = \mathbb{E} \left( \rho \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}) - a}{s_{\boldsymbol{\beta}}(t)} \right) \middle| t = \tau \right) \quad (4.1)$$

donde  $s_{\boldsymbol{\beta}}(\tau)$  es una escala robusta respecto a la distribución condicional a  $t = \tau$  de  $v_{\boldsymbol{\beta}} = y - g(\mathbf{x}, \boldsymbol{\beta})$  la cual denotaremos por  $F_{\boldsymbol{\beta}}(\cdot | t = \tau)$  y definamos

$$\eta_{\boldsymbol{\beta}}(\tau) = \arg \min_a S(\boldsymbol{\beta}, a, \tau). \quad (4.2)$$

En adelante asumiremos que la función de regresión  $g$  es una función continua en la variable  $\boldsymbol{\beta}$  y la función de peso  $W$  es una función acotada y no negativa. Además aceptaremos la siguiente hipótesis que resulta esencial a la hora de identificar a los parámetros del modelo,

FS.1  $P\left(\mathbf{x}, t : \mathbf{x} \in S_W; \quad g(\mathbf{x}, \boldsymbol{\beta}) + \eta_{\boldsymbol{\beta}}(t) \neq g(\mathbf{x}, \boldsymbol{\beta}^*) + \eta_{\boldsymbol{\beta}^*}(t)\right) > 0$  para todo  $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$ , donde  $S_W$  es el soporte de la función de pesos  $W$ .

Notemos que  $F_{\boldsymbol{\beta}_0}(u | t = \tau) = F_0(u - \eta_0(\tau))$ , por lo tanto si  $F_0$  es simétrica respecto del 0 con función de densidad  $f_0$  unimodal y además la función  $\rho$  es una  $\rho$ -función como las descriptas en (2.2.1) entonces se tiene que  $\eta_0(\tau) = \eta_{\boldsymbol{\beta}_0}(\tau)$ .

Por otro lado, definamos también el funcional

$$T(\boldsymbol{\beta}, \sigma) = \mathbb{E} \left( \rho_1 \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}) - \eta_{\boldsymbol{\beta}}(t)}{\sigma} \right) W(\mathbf{x}) \right). \quad (4.3)$$

Por lo tanto, la consistencia de Fisher se reduce a probar que  $\boldsymbol{\beta}_0 = \arg \min T(\boldsymbol{\beta}, \sigma_0)$  hecho que se desprende directamente de la Proposición 3.1.1 de Spano (2016), bajo los mismos supuestos para  $F_0$  y la función  $\rho_1$  descriptos en el párrafo anterior, el hecho que  $\eta_0 = \eta_{\boldsymbol{\beta}_0}$  y de FS.1.

**Observación 4.2.1.** *Para más detalles con respecto a estos resultados se recomienda ver Teorema 10.2 de Maronna et.al. (2006) y Fassano et.al (2012).*

En lo que resta de esta sección probaremos la consistencia del estimador  $\widehat{\boldsymbol{\beta}}$  propuesto para el parámetro de regresión  $\boldsymbol{\beta}_0$  en nuestro modelo dado en (3.1). Para ello introduciremos algo más de notación que utilizaremos a lo largo de las demostraciones.

Dados  $\boldsymbol{\beta} \in \mathbb{R}^q$ ,  $\tau \in \mathbb{R}$  y  $\sigma \in \mathbb{R}_+$ , denotemos con

$$\lambda(\boldsymbol{\beta}, \tau, a, \sigma) = \mathbb{E} \left( \Psi \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}) - a}{\sigma} \right) \middle| t = \tau \right) \quad (4.4)$$



donde  $\Psi = \rho'$ , notemos que  $\eta_{\beta}(\tau)$  es la única solución en  $a$  de  $\lambda(\beta, \tau, a, s_{\beta}(\tau)) = 0$  donde  $s_{\beta}(\tau)$  es una escala robusta, como la que definimos anteriormente. A lo largo de este trabajo consideraremos como medida de escala robusta condicional  $s_{\beta}(\tau)$  a la mediana de los desvíos absolutos a la mediana respecto de la distribución condicional  $F_{\beta}(\cdot|t = \tau)$ , es decir,

$$s_{\beta}(\tau) = \text{MAD}\left(F_{\beta}(\cdot|t = \tau)\right) = \text{mediana}\left(|y - g(\mathbf{x}, \beta) - m_{\beta}(\tau)| \mid t = \tau\right), \quad (4.5)$$

donde  $m_{\beta}(\tau)$  es la mediana respecto de  $F_{\beta}(\cdot|t = \tau)$ . Por otra parte, si consideramos la distribución condicional empírica  $\widehat{F}_{\beta}(\cdot|t = \tau)$  definida en (3.2) basada en la muestra  $v_{i,\beta} = y_i - g(\mathbf{x}_i, \beta)$  para  $1 \leq i \leq n$  y en los pesos  $w_i(\tau)$  definidos en (2.3) para las variables  $t_i$   $1 \leq i \leq n$ , luego podemos definir el funcional empírico

$$\widehat{\lambda}(\beta, \tau, a, \sigma) = \int \Psi\left(\frac{v-a}{\sigma}\right) d\widehat{F}_{\beta}(v|t = \tau) = \sum_{i=1}^n \Psi\left(\frac{v_{i,\beta} - a}{\sigma}\right) \frac{w_i(\tau)}{\sum_{j=1}^n w_j(\tau)} \quad (4.6)$$

Denotemos también por  $\widehat{s}_{\beta}(\tau) = \text{MAD}\left(\widehat{F}_{\beta}(\cdot|t = \tau)\right)$ , es decir, un estimador de la escala local  $s_{\beta}(\tau)$ . Por lo tanto, con esta notación, el estimador  $\widehat{\eta}_{\beta}(\tau)$  definido en el paso 1 del procedimiento de estimación descrito en la Sección 3.3 puede pensarse como la solución en  $a$  de  $\widehat{\lambda}(\beta, \tau, a, \widehat{s}_{\beta}(\tau)) = 0$ , y claramente constituye un estimador de su versión poblacional  $\eta_{\beta}(\tau)$  definida anteriormente.

Asumiremos el siguiente conjunto de hipótesis:

- C1. La función de densidad de  $t$ ,  $f_t$  es acotada. Además, dado  $T_0$  compacto en  $\mathbb{R}$  existe una constante  $C(T_0) > 0$  tal que  $f_t(\tau) > C(T_0) \quad \forall \tau \in T_0$ .
- C2. El núcleo  $K : \mathbb{R} \rightarrow \mathbb{R}$  es una función par, no negativa, Lipschitz de orden uno, acotada con soporte  $[-1, 1]$  tal que  $\int K(u)du = 1$  y  $\int u^2 K(u)du < \infty$ .
- C3.  $F_{\beta}(v|t = \tau)$  es simétrica alrededor de  $\eta_{\beta}(\tau)$  con función de densidad unimodal para todo  $\beta$ .
- C4. Sean  $T_0$  un conjunto compacto en  $\mathbb{R}$  y  $\mathcal{B}$  un conjunto compacto en  $\mathbb{R}^q$ . Entonces,

1. para cada  $v$  fijo se verifica que en un entorno de  $T_0$

$$\forall \epsilon > 0 \exists \delta > 0 : |\tau' - \tau| < \delta \Rightarrow \sup_{\beta \in \mathcal{B}} |F_{\beta}(v|t = \tau) - F_{\beta}(v|t = \tau')| < \epsilon$$

2. se verifica la siguiente condición de equicontinuidad

$$\forall \epsilon > 0 \exists \delta > 0 : |u - v| < \delta \Rightarrow \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |F_{\beta}(u|t = \tau) - F_{\beta}(v|t = \tau)| < \epsilon.$$

3. para cada  $u$  fijo se verifica la siguiente condición de continuidad uniforme

$$\forall \epsilon > 0 \exists \delta > 0 : |\tau - \tau'| < \delta \text{ y } |\beta_1 - \beta_2| < \delta \Rightarrow |F_{\beta_1}(u|t = \tau) - F_{\beta_2}(u|t = \tau')| < \epsilon.$$

**C5.**  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  es una función impar, estrictamente creciente, acotada y continuamente diferenciable.

**C6.**  $\rho_1 : \mathbb{R} \rightarrow \mathbb{R}$  es una  $\rho$ -función acotada con derivada acotada.

**C7.** La sucesión  $h_n$  es tal que  $h_n \rightarrow 0$  y  $\frac{nh_n}{\log(n)} \rightarrow \infty$  de manera creciente.

Para obtener la consistencia del estimador del parámetro de regresión estudiaremos la convergencia uniforme de  $\hat{\eta}_\beta$  a  $\eta_\beta$ . Para ello será necesario obtener resultados de convergencia de  $\hat{F}_\beta(\cdot|t = \tau)$  a  $F_\beta(\cdot|t = \tau)$  y estudiar como a partir de este hecho podemos obtener propiedades sobre la estimación de la escala local  $\hat{s}_\beta(\tau)$  definida a partir del MAD de  $\hat{F}_\beta(\cdot|t = \tau)$ . Estos resultados serán obtenidos utilizando una serie de lemas y teoremas que enunciaremos a continuación y cuyas demostraciones serán relegadas al Apéndice.

**Lema 4.2.2.** Sea  $T_0 \subset \mathbb{R}$  un conjunto compacto y  $R_{0n}(\tau) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{\tau-t_i}{h_n}\right)$  entonces dado  $\epsilon > 0$  se verifica que

$$\sum_{n \in \mathbb{N}} \mathbb{P} \left( \sup_{\tau \in T_0} |R_{0n}(\tau) - \mathbb{E}R_{0n}(\tau)| > \epsilon \right) < \infty$$

**Lema 4.2.3.** Si  $T_0 \subset \mathbb{R}$  y  $\mathcal{B} \subset \mathbb{R}^q$  son conjuntos compactos, entonces la hipótesis **C4.3** implica que dado  $\epsilon > 0$  existen números reales  $a, b$  tales que, para todo  $\tau \in T_0$  y para todo  $\beta \in \mathcal{B}$  se cumple que  $F_\beta(b|t = \tau) > 1 - \epsilon$  y  $F_\beta(a|t = \tau) < \epsilon$ .

**Lema 4.2.4.** Las hipótesis **C4.3** y **C5** implican que  $\lambda(\beta, \tau, a, \sigma)$  y  $\eta_\beta(\tau)$  son funciones continuas de variables  $(\beta, \tau, a, \sigma)$  y  $(\beta, \tau)$  respectivamente.

**Lema 4.2.5.** Sean  $T_0 \subset \mathbb{R}$  y  $\mathcal{B} \subset \mathbb{R}^q$  conjuntos compactos y sea  $F_{n,\beta}(v|t = \tau)$  una sucesión de funciones de distribución condicional que verifica

$$\sup_{v \in \mathbb{R}} \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |F_{n,\beta}(v|t = \tau) - F_\beta(v|t = \tau)| \rightarrow 0 \quad (4.7)$$

luego, si  $F_\beta(v|t = \tau)$  verifica las hipótesis **C4.2** y **C4.3** entonces existen constantes  $0 < A \leq B$  tales que  $s_{n,\beta}(\tau) = \text{MAD}(F_{n,\beta}(\cdot|t = \tau))$  verifica  $A \leq s_{n,\beta}(\tau) \leq B$  para todo  $\tau \in T_0$ , para todo  $\beta \in \mathcal{B}$  y para todo  $n > n_0$ .

**Teorema 4.2.6.** Sean  $\mathcal{B} \subset \mathbb{R}^q$  y  $T_0 \subset \mathbb{R}$  conjuntos compactos. Supongamos que valen las hipótesis **C1** – **C6** y además se cumple la siguiente condición

**F.E.** La familia de funciones  $\mathcal{F} = \{f_{v,\beta}(y, \mathbf{x}) = I_{(-\infty, v]}(y - g(\mathbf{x}, \beta)), v \in \mathbb{R}, \beta \in \mathcal{B}\}$  tiene número de cubrimiento  $N(\epsilon, \mathcal{F}, L^1(Q)) \leq A\epsilon^{-W}$  para cualquier probabilidad  $Q$  y para todo  $0 < \epsilon < 1$ , entonces

$$\sup_{v \in \mathbb{R}} \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |\hat{F}_\beta(v|t = \tau) - F_\beta(v|t = \tau)| \rightarrow 0 \quad \text{ctp.}$$

**Teorema 4.2.7.** Sean  $T_0 \subset \mathbb{R}$  y  $\mathcal{B} \subset \mathbb{R}^q$  conjuntos compactos. Supongamos que la función  $K : \mathbb{R} \rightarrow \mathbb{R}$  es de variación acotada. Entonces, bajo las hipótesis **C1-C5** y **C7**, se tiene que

$$\sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |\widehat{\eta}_\beta - \eta_\beta| \rightarrow 0 \quad \text{ctp}$$

**Demostración:** Como  $F_\beta(v|t = \tau)$  es simétrica alrededor de  $\eta_\beta(\tau)$  y  $\Psi$  es una función impar y acotada tenemos que  $\lambda(\beta, \tau, \eta_\beta(\tau), \sigma) = 0$  para todo  $\sigma > 0$ . Dado  $a \in \mathbb{R}$  definamos por  $\Psi_a(u) = \Psi\left(\frac{u - \eta_\beta(\tau) - a}{\widehat{s}_\beta(\tau)}\right)$ , luego por **C5** se tiene que  $\Psi_a$  es una función de variación acotada en  $\mathbb{R}$  para cada  $\beta \in \mathcal{B}$  y  $\tau \in T_0$ . Además se puede ver que,  $\|\Psi_a\|_V = \|\Psi\|_V / \widehat{s}_\beta(\tau)$  donde  $\|\Psi\|_V$  indica la norma de variación de la función  $\Psi$ . El Lema 4.2.5 y el Teorema 4.2.6 implican que existen constantes reales positivas  $A$  y  $B$  tales que  $A < \widehat{s}_\beta(\tau) < B$  para todo  $\beta \in \mathcal{B}$  y  $\tau \in T_0$ . Luego, usando el hecho de que  $|\int \Psi dG| \leq \|\Psi\|_V \|G\|_\infty$  donde  $G$  es una función continua y acotada, obtenemos que

$$\begin{aligned} & |\lambda(\beta, \tau, \eta_\beta(\tau) + a, \widehat{s}_\beta(\tau)) - \widehat{\lambda}(\beta, \tau, \eta_\beta(\tau) + a, \widehat{s}_\beta(\tau))| \\ & \leq \|\Psi_a\|_V \sup_{v \in \mathbb{R}} |F_\beta(v|t = \tau) - \widehat{F}_\beta(v|t = \tau)|, \end{aligned}$$

y por lo tanto

$$\begin{aligned} & \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |\lambda(\beta, \tau, \eta_\beta(\tau) + a, \widehat{s}_\beta(\tau)) - \widehat{\lambda}(\beta, \tau, \eta_\beta(\tau) + a, \widehat{s}_\beta(\tau))| \\ & \leq \frac{1}{A} \|\Psi\|_V \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} \sup_{v \in \mathbb{R}} |F_\beta(v|t = \tau) - \widehat{F}_\beta(v|t = \tau)|. \end{aligned}$$

Luego, el Teorema 4.2.6 implica que

$$\sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |\lambda(\beta, \tau, \eta_\beta(\tau) + a, \widehat{s}_\beta(\tau)) - \widehat{\lambda}(\beta, \tau, \eta_\beta(\tau) + a, \widehat{s}_\beta(\tau))| \xrightarrow{c.s} 0 \quad (4.8)$$

Por otro lado, dado  $\epsilon > 0$  tenemos que **C3** y **C5** implican

$$\lambda(\beta, \tau, \eta_\beta(\tau) + \epsilon, \sigma) < 0 < \lambda(\beta, \tau, \eta_\beta(\tau) - \epsilon, \sigma).$$

Además, de la compacidad de  $\mathcal{B}$  y  $T_0$ , la continuidad de  $\lambda(\beta, \tau, \eta_\beta(\tau) \pm \epsilon, \sigma)$  en  $(\beta, \tau, \sigma)$  y de  $\eta_\beta(\tau)$  en  $(\beta, \tau)$  dadas por el lema 4.2.4 se deduce que

$$\lambda_1 = \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} \sup_{A \leq \sigma \leq B} \lambda(\beta, \tau, \eta_\beta(\tau) + \epsilon, \sigma) < 0 < \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} \sup_{A \leq \sigma \leq B} \lambda(\beta, \tau, \eta_\beta(\tau) - \epsilon, \sigma) = \lambda_2. \quad (4.9)$$

Usando (4.8) y el hecho de que  $\mathbb{P}(A < \widehat{s}_\beta(\tau) < B) = 1$  a partir de  $n_0 \in \mathbb{N}$  suficientemente grande, entonces de (4.9) obtenemos que para todo  $\beta \in \mathcal{B}$ ,  $\tau \in T_0$  y para todo  $n > n_0$

$$\widehat{\lambda}(\beta, \tau, \eta_\beta(\tau) + \epsilon, \widehat{s}_\beta(\tau)) < \frac{\lambda_1}{2} < 0 < \frac{\lambda_2}{2} < \widehat{\lambda}(\beta, \tau, \eta_\beta(\tau) - \epsilon, \widehat{s}_\beta(\tau)).$$

Por último basta observar que como  $\widehat{\lambda}(\boldsymbol{\beta}, \tau, a, \widehat{s}_\beta(\tau))$  es estrictamente decreciente en  $a$  entonces  $\eta_\beta(\tau) - \epsilon < \widehat{\eta}_\beta(\tau) < \eta_\beta(\tau) + \epsilon$  casi seguramente para todo  $\boldsymbol{\beta} \in \mathcal{B}$  y  $\tau \in T_0$  luego  $\sup_{\boldsymbol{\beta} \in \mathcal{B}} \sup_{\tau \in T_0} |\widehat{\eta}_\beta - \eta_\beta| \xrightarrow{c.s.} 0$  como se quería demostrar.

El siguiente teorema muestra que el estimador de regresión es fuertemente consistente. Para ello será necesaria la siguiente condición de entropía:

**CE** La clase de funciones  $\mathcal{F} = \{g_{\boldsymbol{\beta}, \sigma}(y, \mathbf{x}, t) = \rho_1 \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}) - \eta_\beta(t)}{\sigma} \right) W(\mathbf{x}), \boldsymbol{\beta} \in \mathcal{B}, \sigma \in \mathcal{S}\}$  donde  $\mathcal{S}$  es un entorno de  $\sigma_0$ , tiene número de bracketing finito, es decir,  $N_{[\cdot]}(\epsilon, \mathcal{F}, L^1(H_0)) < \infty$  o  $\log N(\epsilon, \mathcal{F}, L^1(H_n)) = o_p(n)$  donde  $H_0$  es la distribución de  $(y, \mathbf{x}, t)$  y  $H_n$  es la distribución empírica correspondiente.

**Observación 4.2.8.** La condición de entropía **CE** se puede verificar en variadas situaciones que, por supuesto, dependerán, por un lado, de la características de la función de regresión  $g$ , de la estructura de los espacios  $\mathcal{B}$  y  $\mathcal{S}$  y, por otro lado, de las propiedades heredadas por la función  $\eta_\beta(\tau)$ . Por ejemplo, notar que, al ser la función de peso  $W$  acotada e independiente de los parámetros y el espacio  $\mathcal{S}$  un intervalo real acotado, luego si la función de pérdida  $\rho_1$  es Lipschitz con constante  $L$  y llamando  $\mathcal{F}^* = \{g(\mathbf{x}, \boldsymbol{\beta}) + \eta_\beta(t), \boldsymbol{\beta} \in \mathcal{B}\}$  entonces, siguiendo a Anthony y Bartlett (1999), sección 14.3, obtenemos que  $N(\epsilon, \mathcal{F}, L^1(H_0)) \leq N(\epsilon/L, \mathcal{F}^*, L^1(H_0))$ .

**Teorema 4.2.9.** Sean  $(y_i, \mathbf{x}_i, t_i)$  vectores aleatorios independientes e idénticamente distribuidos que satisfacen el modelo (3.1). Entonces si

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}} \sup_{\tau \in T_0} |\widehat{\eta}_\beta(\tau) - \eta_\beta(\tau)| \xrightarrow{ctp} 0 \quad (4.10)$$

donde  $\eta_\beta$  y  $\widehat{\eta}_\beta$  fueron definidas en (4.2) y el Paso 1 del proceso de estimación respectivamente,  $s_n$  es un estimador consistente para  $\sigma_0$ ,  $E(|W(\mathbf{x})|) < \infty$  y además se satisface la condición de entropía **CE**, entonces el estimador  $\widehat{\boldsymbol{\beta}}$  propuesto en (3.4) resulta un estimador consistente de  $\boldsymbol{\beta}_0$ .

**Demostración:** Consideremos

$$T_n(\boldsymbol{\beta}, \sigma) = \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}) - \widehat{\eta}_\beta(t_i)}{\sigma} \right) W(\mathbf{x}_i),$$

$$\widetilde{T}_n(\boldsymbol{\beta}, \sigma) = \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}) - \eta_\beta(t_i)}{\sigma} \right) W(\mathbf{x}_i).$$

En primer lugar mostremos que si

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}} |T_n(\boldsymbol{\beta}, s_n) - T(\boldsymbol{\beta}, \sigma_0)| \xrightarrow{ctp} 0 \quad (4.11)$$

donde el funcional  $T$  fue definido en (4.3), entonces  $\widehat{\boldsymbol{\beta}} \xrightarrow{ctp} \boldsymbol{\beta}_0$ .

Para ello observemos que  $C6$ , la continuidad de las funciones  $g$  y  $\eta$ , el hecho que  $W$  es una función acotada y el Teorema de Convergencia Dominada garantizan la continuidad del funcional  $T(\beta, \sigma_0)$ . Luego, sea  $\widehat{\beta}_k$  una subsucesión de  $\widehat{\beta}$  tal que  $\widehat{\beta}_k \rightarrow \beta^*$ , supongamos sin perder generalidad que  $\widehat{\beta} \rightarrow \beta^*$ . Si  $\|\beta^*\| < \infty$ , de (4.11) y la continuidad de  $T(\beta, \sigma_0)$  deducimos que  $T_n(\widehat{\beta}, s_n) - T(\beta^*, \sigma_0) \xrightarrow{ctp} 0$  y  $T_n(\beta_0, s_n) - T(\beta_0, \sigma_0) \xrightarrow{ctp} 0$ . Como  $T_n(\beta_0, s_n) \geq T_n(\widehat{\beta}, s_n)$  y  $T$  tiene un único mínimo en  $\beta_0$ , fácilmente se obtiene  $\beta^* = \beta_0$ .

Supongamos que  $\|\beta^*\| = \infty$  luego igual que antes tenemos que  $T_n(\widehat{\beta}, s_n) - T(\widehat{\beta}, \sigma_0) \xrightarrow{ctp} 0$ ,  $T_n(\beta_0, s_n) - T(\beta_0, \sigma_0) \xrightarrow{ctp} 0$  y  $T_n(\beta_0, s_n) \geq T_n(\widehat{\beta}, s_n)$  entonces  $\lim_{n \rightarrow \infty} T(\widehat{\beta}, \sigma_0) - T(\beta_0, \sigma_0) \leq 0$ , llegando a una contradicción, por lo tanto  $\|\beta^*\| < \infty$ .

Por lo tanto la demostración del Teorema se reduce a probar (4.11). Para ello veremos

$$\sup_{\beta \in \mathcal{B}} |T_n(\beta, s_n) - \widetilde{T}_n(\beta, s_n)| \xrightarrow{ctp} 0, \quad (4.12)$$

$$\sup_{\beta \in \mathcal{B}} |\widetilde{T}_n(\beta, s_n) - T(\beta, s_n)| \xrightarrow{ctp} 0, \quad (4.13)$$

$$\sup_{\beta \in \mathcal{B}} |T(\beta, s_n) - T(\beta, \sigma_0)| \xrightarrow{ctp} 0. \quad (4.14)$$

Comencemos por ver (4.12). Dado  $\varepsilon > 0$  existe  $\Gamma_0$  un conjunto compacto tal que  $P(t \in \Gamma_0) > 1 - \varepsilon$ . Sea  $V_n = n^{-1} \sum_{i=1}^n I(t_i \notin \Gamma_0)$ ,  $\rho'_1 = \Psi_1$  y  $M > 0$  tal que  $1/M < s_n$  luego

$$\sup_{\beta \in \mathcal{B}} |T_n(\beta, s_n) - \widetilde{T}_n(\beta, s_n)| \leq \|W\|_\infty \{M \|\Psi_1\|_\infty \sup_{\beta \in \mathcal{B}} \sup_{\tau \in \Gamma_0} |\widehat{\eta}_\beta - \eta_\beta| + 2\|\rho_1\|_\infty V_n\}.$$

Por lo tanto, usando (4.10) y la Ley de los Grandes Números tenemos (4.12)

Veamos ahora (4.13), notemos que

$$\sup_{\beta \in \mathcal{B}} |\widetilde{T}_n(\beta, s_n) - T(\beta, s_n)| \leq \sup_{\beta \in \mathcal{B}, \sigma \in \mathcal{S}} |\widetilde{T}_n(\beta, \sigma) - T(\beta, \sigma)|.$$

Luego, usando la condición de entropía **CE**, el hecho que  $E(|W(\mathbf{x})|) < \infty$ ,  $C6$  y el Teorema 4.1.7 se obtiene lo deseado. Finalmente (4.14) se deduce de  $C6$  junto con del Teorema de Convergencia Dominada.

**Teorema 4.2.10.** *Sea  $\widehat{\beta}$  un estimador consistente de  $\beta_0$  entonces el estimador de la componente no paramétrica definido en (3.5) verifica que*

$$\sup_{\tau \in T_0} \|\widehat{\eta}_{\widehat{\beta}} - \eta_0\| \xrightarrow{ctp} 0.$$

**Demostración:** La demostración es una consecuencia inmediata del teorema (4.2.7), la continuidad de  $\eta_\beta$  y el hecho que  $\eta_{\beta_0}(\tau) = \eta_0(\tau)$ .

### 4.3. Distribución Asintótica

En esta sección deduciremos la distribución asintótica del estimador propuesto del parámetro de regresión  $\widehat{\boldsymbol{\beta}}$  para el modelo estudiado.

Para simplificar la notación notaremos por  $|\mathbf{A}|$  a la norma euclídea de cualquier matriz o vector. Además, dados  $k$  y  $\epsilon$  constantes positivas, definimos  $\mathcal{T}_k = \{\tau \in \mathbb{R} : |\tau| \leq k\}$ ,  $\mathcal{X}_k = \{\mathbf{x} \in \mathbb{R}^p : |\mathbf{x}| \leq k\}$  y  $\mathcal{B}$  y  $\mathcal{S}$  entornos de  $\boldsymbol{\beta}_0$  y  $\sigma_0$  respectivamente, es decir,  $\mathcal{B} = \{\boldsymbol{\beta} \in \mathbb{R}^q : |\boldsymbol{\beta} - \boldsymbol{\beta}_0| \leq \epsilon\}$  y  $\mathcal{S} = \{s \in \mathbb{R} > 0 : |s - \sigma_0| \leq \epsilon\}$ .

Sea  $\mathbf{h} : \mathbb{R}^q \rightarrow \mathbb{R}$  una función, notaremos por  $\nabla_{\mathbf{z}} h(\mathbf{z})$  y  $\mathbf{H}_{\mathbf{z}} h(\mathbf{z})$  al vector gradiente y la matriz hessiana, respectivamente con respecto a  $\mathbf{z}$  y a modo de simplificar la notación en lo que sigue llamaremos

$$\begin{aligned}\widehat{v}(\tau, \boldsymbol{\beta}) &= \eta_{\boldsymbol{\beta}}(\tau) - \widehat{\eta}_{\boldsymbol{\beta}}(\tau) \\ \widehat{\boldsymbol{v}}(\tau, \boldsymbol{\beta}) &= \nabla_{\boldsymbol{\beta}}(\eta_{\boldsymbol{\beta}}(\tau) - \widehat{\eta}_{\boldsymbol{\beta}}(\tau)) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}} = (\widehat{v}_1(\tau, \boldsymbol{\beta}), \dots, \widehat{v}_q(\tau, \boldsymbol{\beta})), \\ \mathbf{z}(\mathbf{s}, \tau, \boldsymbol{\beta}) &= \nabla_{\boldsymbol{\beta}}(g(\mathbf{s}, \boldsymbol{\beta}) + \eta_{\boldsymbol{\beta}}(\tau)) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}} = (z_1(\mathbf{s}, \tau, \boldsymbol{\beta}), \dots, z_q(\mathbf{s}, \tau, \boldsymbol{\beta})), \\ \widehat{\mathbf{z}}(\mathbf{s}, \tau, \boldsymbol{\beta}) &= \nabla_{\boldsymbol{\beta}}(g(\mathbf{s}, \boldsymbol{\beta}) + \widehat{\eta}_{\boldsymbol{\beta}}(\tau)) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}} = (\widehat{z}_1(\mathbf{s}, \tau, \boldsymbol{\beta}), \dots, \widehat{z}_q(\mathbf{s}, \tau, \boldsymbol{\beta})).\end{aligned}$$

Asumiremos, a partir de ahora, que la variable  $t$  tiene distribución con soporte compacto y convexo  $T$  y, también, consideraremos el siguiente conjunto de hipótesis:

**N1.** La función  $\rho'_1 = \Psi_1$  es impar, continua y acotada.

**N2.** Las funciones  $\Psi'_1$ ,  $\Psi''_1$ ,  $\Psi_2 = t\Psi'_1$  y  $\Psi_3 = t\Psi''_1$  son continuas y acotadas.

**N3.** La función de peso  $W$  es acotada.

**N4.** 1. La función de regresión  $g(\mathbf{x}, \boldsymbol{\beta})$  es de clase  $C^2$  como función de  $\boldsymbol{\beta}$  y verifica la siguiente condición de equicontinuidad, dados  $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{B}$ ,

$$|g(\mathbf{x}, \boldsymbol{\beta}_1) - g(\mathbf{x}, \boldsymbol{\beta}_2)| \leq \mathbf{F}(\mathbf{x})|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|$$

con  $\mathbf{F}$  una función real que verifica  $\sup_{\mathbf{x} \in \mathcal{X}_k} |\mathbf{F}(\mathbf{x})| < \infty$  para todo  $k > 0$  y, además, pediremos que  $\sup_{\mathbf{x} \in \mathcal{X}_k} g(\mathbf{x}, \boldsymbol{\beta}_0) < \infty$  para todo  $k > 0$ .

2. Para todo  $1 \leq j, l \leq q$  se satisface la siguiente condición de equicontinuidad

$$\left| \frac{\partial^2}{\partial \beta_j \partial \beta_l} g(\mathbf{x}, \boldsymbol{\beta}_1) - \frac{\partial^2}{\partial \beta_j \partial \beta_l} g(\mathbf{x}, \boldsymbol{\beta}_2) \right| \leq \mathbf{R}(\mathbf{x})|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2| \text{ para } \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{B}.$$

con  $\mathbf{R}$  una función real que verifica  $\mathbb{E}[\mathbf{R}^2(\mathbf{x})] < \infty$  y  $\sup_{\mathbf{x} \in \mathcal{X}_k} |W(\mathbf{x})\mathbf{R}(\mathbf{x})| < \infty$  para todo  $k > 0$ .

3. Para todo  $1 \leq j, l \leq q$  y se satisface la siguiente condición

$$\left| \frac{\partial}{\partial \beta_j} g(\mathbf{x}, \boldsymbol{\beta}_0) \right| \leq \mathbf{H}_1(\mathbf{x}) \quad \text{y} \quad \left| \frac{\partial^2}{\partial \beta_j \partial \beta_l} g(\mathbf{x}, \boldsymbol{\beta}_0) \right| \leq \mathbf{H}_2(\mathbf{x}),$$

con  $\mathbf{H}_i$  funciones reales que verifican  $\mathbb{E} [\mathbf{H}_i^2(\mathbf{x})] < \infty$  para  $i = 1, 2$  y, además,

$$\sup_{\mathbf{x} \in \mathcal{X}_k} |\sqrt{W(\mathbf{x})} \mathbf{H}_1(\mathbf{x})| < \infty \quad \text{y} \quad \sup_{\mathbf{x} \in \mathcal{X}_k} |W(\mathbf{x}) \mathbf{H}_2(\mathbf{x})| < \infty$$

para todo  $k > 0$ .

**N5.** Las funciones  $\hat{\eta}_\beta(\tau)$  y  $\eta_\beta(\tau)$  son de clase  $C^1$  como funciones de  $(\boldsymbol{\beta}, \tau)$  y de clase  $C^2$  como funciones de  $\boldsymbol{\beta}$  tal que  $\frac{\partial^2}{\partial \beta_i \partial \beta_l} \eta_{\boldsymbol{\beta}_0}(\tau)$  es acotada. Además para todo  $1 \leq j, l \leq q$  se satisface la siguiente condición de equicontinuidad

$$\forall \epsilon > 0, \exists \delta > 0 : |\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2| < \delta \Rightarrow \left\| \frac{\partial^2}{\partial \beta_j \partial \beta_l} \eta_{\boldsymbol{\beta}_1} - \frac{\partial^2}{\partial \beta_j \partial \beta_l} \eta_{\boldsymbol{\beta}_2} \right\|_\infty < \epsilon.$$

**N6.** 1.  $\|\hat{\eta}_{\hat{\boldsymbol{\beta}}} - \eta_0\|_\infty \xrightarrow{P} 0$  para cualquier estimador consistente  $\hat{\boldsymbol{\beta}}$  de  $\boldsymbol{\beta}_0$ .

2. Para cada  $\tau \in T$  y  $\boldsymbol{\beta} \in \mathcal{B}$ ,  $\hat{v}(\tau, \boldsymbol{\beta}) \xrightarrow{P} 0$ . Además,  $n^{\frac{1}{4}} \|\hat{v}(\tau, \boldsymbol{\beta}_0)\|_\infty \xrightarrow{P} 0$  y  $n^{\frac{1}{4}} \|\hat{v}_j(\tau, \boldsymbol{\beta}_0)\|_\infty \xrightarrow{P} 0$  para todo  $1 \leq j \leq q$ .

3.  $\left\| \frac{\partial \hat{v}(\tau, \boldsymbol{\beta}_0)}{\partial \tau} \right\|_\infty \xrightarrow{P} 0$  y  $\left\| \frac{\partial \hat{v}_j(\tau, \boldsymbol{\beta}_0)}{\partial \tau} \right\|_\infty \xrightarrow{P} 0$  para todo  $1 \leq j \leq q$ .

4.  $\sup_{\boldsymbol{\beta} \in \mathcal{B}} \|\hat{\boldsymbol{v}}(\cdot, \boldsymbol{\beta})\|_\infty \xrightarrow{P} 0$  y  $\sup_{\boldsymbol{\beta} \in \mathcal{B}} \left\| \frac{\partial}{\partial \beta_l} \hat{\boldsymbol{v}}(\cdot, \boldsymbol{\beta}) \right\|_\infty \xrightarrow{P} 0$  para todo  $1 \leq l \leq q$ .

**N7.**

$$\mathbb{E} [W(\mathbf{x}) |\mathbf{z}(\mathbf{x}, t, \boldsymbol{\beta}_0)|^T \mathbf{z}(\mathbf{x}, t, \boldsymbol{\beta}_0)] < \infty$$

$$\mathbb{E} \left[ W(\mathbf{x}) | \mathbf{H}_\beta(g(\mathbf{x}, \boldsymbol{\beta}) + \eta_\beta(t)) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right] < \infty$$

**N8.** La matriz  $\Sigma$  es definida positiva donde

$$\Sigma = \mathbb{E} \left[ \Psi_1^2 \left( \frac{\epsilon}{\sigma_0} \right) \right] \mathbb{E} \left[ W^2(\mathbf{x}) \mathbf{z}(\mathbf{x}, t, \boldsymbol{\beta}_0) \mathbf{z}(\mathbf{x}, t, \boldsymbol{\beta}_0)^T \right].$$

**N9.** La matriz  $A$  es no singular donde

$$A = \frac{1}{\sigma_0} \mathbb{E} \left[ \Psi_1' \left( \frac{\epsilon}{\sigma_0} \right) \right] \mathbb{E} \left[ W(\mathbf{x}) \mathbf{z}(\mathbf{x}, t, \boldsymbol{\beta}_0) \mathbf{z}(\mathbf{x}, t, \boldsymbol{\beta}_0)^T \right].$$

**N10.**  $\mathbb{E} [W(\mathbf{x}) \mathbf{z}(\mathbf{x}, t, \boldsymbol{\beta}_0) | t] = 0$ .

**Observación 4.3.1.** Las condiciones **N5** y **N6.4** implican que para cualquier estimador consistente  $\tilde{\beta}$  de  $\beta_0$  tenemos que  $\Delta_n \xrightarrow{P} 0$  y  $\Lambda_n \xrightarrow{P} 0$  con

$$\begin{aligned} a) \quad \Delta_n &= \max_{1 \leq j \leq q} \left\| \frac{\partial}{\partial \beta_j} \hat{\eta}_{\tilde{\beta}} - \frac{\partial}{\partial \beta_j} \eta_{\beta_0} \right\|_{\infty} \\ b) \quad \Lambda_n &= \max_{1 \leq j, l \leq q} \left\| \frac{\partial^2}{\partial \beta_j \partial \beta_l} \hat{\eta}_{\tilde{\beta}} - \frac{\partial^2}{\partial \beta_j \partial \beta_l} \eta_{\beta_0} \right\|_{\infty} \end{aligned}$$

**Observación 4.3.2.** Los supuestos **N1-N3** son condiciones estándar sobre la función de pérdida y la función de peso en modelos de regresión. Lo supuesto en **N5** y **N6** es similar a lo pedido en Boente-He et al. (2006).

La condición **N4.1** se cumple, por ejemplo si  $g(\mathbf{x}, \beta)$  tiene derivadas parciales continuas con respecto a  $\beta$  como funciones de las covariables, **N4.2** se verifica si  $\frac{\partial^3}{\partial \beta_j \partial \beta_l \partial \beta_k} g(\mathbf{x}, \beta)$  es continua como función de las covariables para todo  $1 \leq j, l, k \leq q$  y con una elección adecuada de la función peso  $W$ .

La condición **N7** es usada para asegurar la consistencia de estimadores de la matriz  $A$  basados en estimadores consistentes previos del parámetro  $\beta_0$  y de la función  $\eta_{\beta}$  (ver Lema 4.3.4).

Además notar que **N10** se satisface si, por ejemplo,  $W \equiv 1$ . Efectivamente, como  $\eta_{\beta}(\tau)$  minimiza a  $S(\beta, a, \tau)$  para cada  $\tau$ , y cada  $\beta$  entonces satisface

$$\mathbb{E} \left( \Psi_1 \left( \frac{y - g(\mathbf{x}, \beta) - \eta_{\beta}(\tau)}{s_{\beta}(t)} \right) \middle| t = \tau \right) = 0,$$

luego derivando con respecto a  $\beta$  obtenemos

$$\mathbb{E} \left( \Psi_1' \left( \frac{y - g(\mathbf{x}, \beta) - \eta_{\beta}(\tau)}{s_{\beta}(t)} \right) \nabla_{\beta} \left( \frac{y - g(\mathbf{x}, \beta) - \eta_{\beta}(\tau)}{s_{\beta}(t)} \right) \middle| t = \tau \right) = 0,$$

por último basta con evaluar en  $\beta_0$ , considerar **N1** y usar la independencia entre  $(\mathbf{x}, t)$  y  $\epsilon$ .

El siguiente lema y su demostración resultan de una pequeña adaptación del Lema 3.4.1 presentado en Spano (2016). El lema presenta resultados que pueden ser aplicados en un contexto general de convergencia en probabilidad para ciertos procesos empíricos.

**Lema 4.3.3.** Consideremos las funciones  $\rho : \mathbb{R} \rightarrow \mathbb{R}$  y  $M(\mathbf{x}, t, \beta) : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^q \rightarrow \mathbb{R}^{q \times q}$  tales que cumplan las siguientes hipótesis:

**L1.**  $\rho$  es una función continua y acotada.

**L2.** Existe  $\epsilon > 0$  tal que  $\mathbb{E} \left( \sup_{\beta \in \mathcal{B}} |M(\mathbf{x}, t, \beta) W(\mathbf{x})| \right) < \infty$ .

**L3.** La función  $M(\mathbf{x}, t, \beta)$  es equicontinua como función de  $\beta$  en el siguiente sentido, dado  $\epsilon > 0$  existe  $\eta > 0$  tal que si  $|\beta_1 - \beta_2| < \eta$  y  $|\beta_i - \beta_0| < \eta$  con  $i = 1, 2$  entonces  $|M(\mathbf{x}, t, \beta_1) - M(\mathbf{x}, t, \beta_2)| < \epsilon$  si  $\mathbf{x} \in \mathcal{X}_k$  y  $t \in \mathcal{T}_k$ . Además, pediremos que  $\sup_{\mathbf{x} \in \mathcal{X}_k, t \in \mathcal{T}_k} |M(\mathbf{x}, t, \beta_0) W(\mathbf{x})| < \infty$  para todo  $k > 0$ .



Si  $(y_i, \mathbf{x}_i^T, t_i) \in \mathbb{R}^{1+p+1}$  son vectores aleatorios i.i.d. bajo el modelo no lineal (3.1),  $s_n$  y  $\tilde{\boldsymbol{\beta}}$  son estimadores consistentes de  $\sigma_0$  y  $\boldsymbol{\beta}_0$  respectivamente y se satisface **N4.1** entonces,  $D_n \xrightarrow{P} D$  siendo

$$D_n = \frac{1}{n} \sum_{i=1}^n \frac{1}{s_n} \rho\left(\frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \eta_0(t_i)}{s_n}\right) M(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) W(\mathbf{x}_i),$$

$$D = \frac{1}{\sigma_0} E \rho\left(\frac{\epsilon}{\sigma_0}\right) E[M(\mathbf{x}, t, \boldsymbol{\beta}_0) W(\mathbf{x})].$$

**Demostración:** Llamemos a  $I(y, \mathbf{x}, t, \boldsymbol{\beta}, s) = \rho\left(\frac{y - g(\mathbf{x}, \boldsymbol{\beta}) - \eta_0(t)}{s}\right) M(\mathbf{x}, t, \boldsymbol{\beta}) W(\mathbf{x})$ . Luego bastará ver que

- a)  $\lim_{n \rightarrow \infty} \mathbb{E} \left[ \frac{1}{s_n} I(y, \mathbf{x}, t, \tilde{\boldsymbol{\beta}}, s_n) \right] = D$
- b)  $\sup_{\boldsymbol{\beta} \in \mathcal{B}, s \in \mathcal{S}} \left| \frac{1}{n} \sum_{i=1}^n I(y_i, \mathbf{x}_i, t_i, \boldsymbol{\beta}, s) - \mathbb{E}[I(y, \mathbf{x}, t, \boldsymbol{\beta}, s)] \right| \xrightarrow{P} 0$ .

El item **a)** se desprende de **L1, L2, L3, N4.1**, de la consistencia de  $\tilde{\boldsymbol{\beta}}$  y de  $s_n$ , de la independencia entre  $(\mathbf{x}, t)$  y  $\epsilon$  y del Teorema de Convergencia Dominada pues  $|I(y, \mathbf{x}, t, \boldsymbol{\beta}, s)| \leq \|\rho\| \sup_{\boldsymbol{\beta} \in \mathcal{B}} |M(\mathbf{x}, t, \boldsymbol{\beta}) W(\mathbf{x})|$  que es integrable.

Para probar el item **b)** estudiaremos la convergencia de cada elemento  $I^{ij}$  de la matriz  $I$ . Por el Teorema 4.1.6 bastará probar que, dado  $\epsilon > 0$  existe una clase finita de funciones  $\mathcal{F}_\epsilon$  tal que para todo  $\boldsymbol{\beta} \in \mathcal{B}$  y  $s \in \mathcal{S}$  existen  $I_{\epsilon, l}^{ij} \in \mathcal{F}_\epsilon$  y  $I_{\epsilon, u}^{ij} \in \mathcal{F}_\epsilon$  tales que

$$\mathbb{E} \left( I_{\epsilon, u}^{ij}(y, \mathbf{x}, t) - I_{\epsilon, l}^{ij}(y, \mathbf{x}, t) \right) < \epsilon,$$

$$I_{\epsilon, u}^{ij}(y, \mathbf{x}, t) \leq I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}, s) \leq I_{\epsilon, l}^{ij}(y, \mathbf{x}, t).$$

Sea  $\epsilon > 0$  podemos tomar  $k \in \mathbb{N}$  tal que si definimos  $A_k = \left\{ (y, \mathbf{x}, t) : \left| \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma_0} \right| \leq k, |\mathbf{x}| \leq k, |t| \leq k \right\}$  entonces  $\mathbb{P}(A_k) > 1 - \epsilon_1$  y además, como vale **L2**,  $\mathbb{E} \left[ \sup_{\boldsymbol{\beta} \in \mathcal{B}} |M(\mathbf{x}, t, \boldsymbol{\beta}) W(\mathbf{x})| I_{A_k^c} \right] < \epsilon_1$  siendo  $\epsilon_1 = \epsilon/5 \|\rho\|$ . Ahora, si definimos  $C = \sup_{\boldsymbol{\beta} \in \mathcal{B}} |g(\mathbf{x}, \boldsymbol{\beta})| + \sup_{\mathcal{T}_k} |\eta_0(t)|$ , y consideramos  $k_1 = \sigma_0 k + C$  entonces, llamando  $B_k = \{|y| \leq k_1, |\mathbf{x}| \leq k, |t| \leq k\}$ , tenemos que  $B_k \supseteq A_k$ . Por otro lado, sea  $M > 0$  tal que, para todo  $\boldsymbol{\beta} \in \mathcal{B}$  y  $s \in \mathcal{S}$  se verifica que  $\|\boldsymbol{\beta}\| \leq M$  y  $M^{-1} < s < M$ . Luego, si  $(y, \mathbf{x}, t) \in A_k$ ,  $\boldsymbol{\beta} \in \mathcal{B}$  y  $s \in \mathcal{S}$  entonces  $\left| \frac{y - g(\mathbf{x}, \boldsymbol{\beta}) - \eta_0(t)}{s} \right| \leq M(k_1 + C) = b_1$ .

Por **L3** existe  $b_2 \geq 0$  tal que  $\sup_{B_k, \boldsymbol{\beta} \in \mathcal{B}} |M(\mathbf{x}, t, \boldsymbol{\beta}) W(\mathbf{x})| \leq b_2$ . De **L1** tenemos que la función  $g(a, \mathbf{z}) = \rho(a) \cdot \mathbf{z}$  es uniformemente continua en  $C_k = \{|a| \leq b_1, |\mathbf{z}| \leq b_2\}$  y, por lo tanto, existe  $\tau$  tal que  $|\rho(a_1) \cdot \mathbf{z}_1 - \rho(a_2) \cdot \mathbf{z}_2| < \epsilon/10$  si  $|a_1 - a_2| < \tau$ ,  $|\mathbf{z}_1 - \mathbf{z}_2| < \tau$ ,  $(a_1, \mathbf{z}_1) \in C_k$  y  $(a_2, \mathbf{z}_2) \in C_k$ .

Como valen **L3** y **N4.1**, si  $(y, \mathbf{x}, t) \in B_k$  entonces existe  $\gamma > 0$  tal que si  $|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2| < \gamma$  con  $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{B}$  y  $|s_1 - s_2| < \gamma$  con  $s_1, s_2 \in \mathcal{S}$ , entonces

$$\left| \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_1) - \eta_0(t)}{s_1} - \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_2) - \eta_0(t)}{s_2} \right| < \tau,$$

$$\left| M(\mathbf{x}, t, \boldsymbol{\beta}_1) - M(\mathbf{x}, t, \boldsymbol{\beta}_2) \right| < \tau.$$

Sean  $(\mathcal{B}_q)_{q=1}^N$  una colección finita de bolas cuyos centros  $\boldsymbol{\beta}_q$  pertezcan a  $\mathcal{B}$  de radio menor a  $\gamma$  y  $(\mathcal{S}_m)_{m=1}^N$  otra colección finita de bolas cuyos centros  $s_m$  pertezcan a  $\mathcal{S}$  de radio menor a  $\gamma$  tales que  $\mathcal{B} = \bigcup_{q=1}^N \mathcal{B}_q$  y  $\mathcal{S} = \bigcup_{m=1}^N \mathcal{S}_m$ . Entonces dados  $\boldsymbol{\beta} \in \mathcal{B}$  y  $s \in \mathcal{S}$  existen índices  $(q, m)$  tales que  $\boldsymbol{\beta} \in \mathcal{B}_q$  y  $s \in \mathcal{S}_m$  entonces

$$\begin{aligned} |I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}, s) - I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}_q, s_m)| &\leq |I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}, \sigma) - I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}_q, s_m)| \mathbb{I}_{A_k} \\ &\quad + 2\|\rho\| \sup_{\boldsymbol{\beta} \in \mathcal{B}} |M(\mathbf{x}, t, \boldsymbol{\beta})W(\mathbf{x})| \mathbb{I}_{A_k^c} \\ &\leq \frac{\epsilon}{10} + 2\|\rho\| \sup_{\boldsymbol{\beta} \in \mathcal{B}} |M(\mathbf{x}, t, \boldsymbol{\beta})W(\mathbf{x})| \mathbb{I}_{A_k^c}. \end{aligned}$$

Entonces, definiendo  $\phi(y, \mathbf{x}, t) = \frac{\epsilon}{10} + 2\|\rho\| \sup_{\boldsymbol{\beta} \in \mathcal{B}} |M(\mathbf{x}, t, \boldsymbol{\beta})W(\mathbf{x})| \mathbb{I}_{A_k^c}$  y luego considerando a  $I_{\epsilon, l}^{ij}(y, \mathbf{x}, t) = I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}_q, s_m) - \phi(y, \mathbf{x}, t)$  y  $I_{\epsilon, u}^{ij}(y, \mathbf{x}, t) = I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}_q, s_m) + \phi(y, \mathbf{x}, t)$  tenemos que  $\mathcal{F}_\epsilon = \{I_{\epsilon, l}^{ij}, I_{\epsilon, u}^{ij}\}$  es una familia con cardinal  $N^2$ ,  $I_{\epsilon, u}^{ij}(y, \mathbf{x}, t) \leq I^{ij}(y, \mathbf{x}, t, \boldsymbol{\beta}, s) \leq I_{\epsilon, l}^{ij}(y, \mathbf{x}, t)$  y  $\mathbb{E} \left( I_{\epsilon, u}^{ij}(y, \mathbf{x}, t) - I_{\epsilon, l}^{ij}(y, \mathbf{x}, t) \right) \leq \frac{\epsilon}{5} + 4\|\rho\| \frac{\epsilon}{5\|\rho\|} \leq \epsilon$ . Por lo tanto  $D_n \xrightarrow{P} 0$  como queríamos ver.

**Lema 4.3.4.** Sean  $(y_i, \mathbf{x}_i, t_i)$  observaciones independientes que cumplen el modelo (3.1) y sean  $\tilde{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}$  y  $s_n \xrightarrow{P} \sigma_0$ . Supongamos que se satisfacen las hipótesis **N1** hasta **N5**, **N6.1**, **N6.4** y **N7** entonces  $\mathbf{A}_n \xrightarrow{P} \mathbf{A}$ , donde  $\mathbf{A}$  está definida en **N9** y

$$\begin{aligned} \mathbf{A}_n = &\left\{ -\frac{1}{n} \sum_{i=1}^n \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \hat{\eta}_{\tilde{\boldsymbol{\beta}}}(t_i)}{s_n} \right) \frac{1}{s_n} W(\mathbf{x}_i) \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}})^T \right. \\ &\left. + \frac{1}{n} \sum_{i=1}^n \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \hat{\eta}_{\tilde{\boldsymbol{\beta}}}(t_i)}{s_n} \right) W(\mathbf{x}_i) \left[ \mathbf{H}_{\boldsymbol{\beta}}(g(\mathbf{x}_i, \boldsymbol{\beta}) + \hat{\eta}_{\boldsymbol{\beta}}(t_i)) \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} \right]^T \right\}. \end{aligned}$$

**Demostración:** Consideremos a las funciones  $\varphi(a) = \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - a}{s_n} \right)$  y  $\tilde{\varphi}(a) = \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - a}{s_n} \right)$ . Luego, mediante desarrollos de Taylor de primer orden para  $\varphi$  y  $\tilde{\varphi}$  centrados en  $\eta_0(t_i)$  y

algunos manejos algebraicos se puede ver que  $\mathbf{A}_n = \sum_{i=1}^n \mathbf{A}_n^{(j)}$  donde

$$\begin{aligned} \mathbf{A}_n^{(1)} &= -\frac{1}{n} \sum_{i=1}^n \frac{1}{s_n} \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \eta_0(t_i)}{s_n} \right) W(\mathbf{x}_i) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0)^T, \\ \mathbf{A}_n^{(2)} &= \frac{1}{n} \sum_{i=1}^n \frac{1}{s_n} \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \eta_0(t_i)}{s_n} \right) W(\mathbf{x}_i) \left[ \mathbf{H}_\beta \left( g(\mathbf{x}_i, \boldsymbol{\beta}) + \eta_\beta(t_i) \right) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right]^T, \\ \mathbf{A}_n^{(3)} &= \frac{1}{n} \sum_{i=1}^n \frac{1}{s_n^2} \Psi_1'' \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \xi_{1,i}}{s_n} \right) W(\mathbf{x}_i) [\eta_0(t_i) - \hat{\eta}_{\tilde{\boldsymbol{\beta}}}(t_i)] \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0)^T, \\ \mathbf{A}_n^{(4)} &= \frac{1}{n} \sum_{i=1}^n \frac{1}{s_n} \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \xi_{2,i}}{s_n} \right) W(\mathbf{x}_i) [\eta_0(t_i) - \hat{\eta}_{\tilde{\boldsymbol{\beta}}}(t_i)] \left[ \mathbf{H}_\beta \left( g(\mathbf{x}_i, \boldsymbol{\beta}) + \eta_\beta(t_i) \right) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right]^T, \\ \mathbf{A}_n^{(5)} &= \frac{1}{n} \sum_{i=1}^n \frac{1}{s_n} \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \hat{\eta}_{\tilde{\boldsymbol{\beta}}}(t_i)}{s_n} \right) W(\mathbf{x}_i) \left[ \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}})^T - \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0)^T \right], \\ \mathbf{A}_n^{(6)} &= -\frac{1}{n} \sum_{i=1}^n \frac{1}{s_n} \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \tilde{\boldsymbol{\beta}}) - \hat{\eta}_{\tilde{\boldsymbol{\beta}}}(t_i)}{s_n} \right) W(\mathbf{x}_i) \left[ \mathbf{H}_\beta \left( g(\mathbf{x}_i, \boldsymbol{\beta}) + \hat{\eta}_\beta(t_i) \right) \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} - \mathbf{H}_\beta \left( g(\mathbf{x}_i, \boldsymbol{\beta}) + \eta_\beta(t_i) \right) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right]^T, \end{aligned}$$

donde  $\xi_{1,i}$  y  $\xi_{2,i}$  son puntos intermedios entre  $\hat{\eta}_\beta(t_i)$  y  $\eta_0(t_i)$ . Usando **N2**, **N3**, **N4.3**, **N5**, **N7** y el Lema 4.3.3 es sencillo ver que  $\mathbf{A}_n^{(1)} \xrightarrow{P} \mathbf{A}$ . Veamos entonces que  $\mathbf{A}_n^{(j)} \xrightarrow{P} 0$  para todo  $2 \leq j \leq 6$ .

Para obtener convergencia de  $\mathbf{A}_n^{(2)}$  basta observar la imparidad de  $\Psi_1$  dada en **N1**, **N3**, **N4.3**, **N5**, **N7** y el Lema 4.3.3. Usando **N2**, **N6.1** y **N7** junto con la LGN obtenemos la convergencia de  $\mathbf{A}_n^{(3)}$  y  $\mathbf{A}_n^{(4)}$ . Para ver la convergencia de  $\mathbf{A}_n^{(5)}$  notar que

$$\begin{aligned} \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}})^T - \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0)^T &= \left( \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) - \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \right) \left( \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) - \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \right)^T \\ &\quad + \left( \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) - \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \right) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0)^T \\ &\quad + \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \left( \hat{\mathbf{z}}(\mathbf{x}_i, t_i, \tilde{\boldsymbol{\beta}}) - \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) \right)^T, \end{aligned}$$

luego usando **N2**, **N3**, **N4**, **N7** y la Observación 4.3.1 a), obtenemos lo deseado. Por último usando **N1**, **N3**, **N4.2** y la Observación 4.3.1 b), concluimos con la convergencia de  $\mathbf{A}_n^{(6)}$ .

**Teorema 4.3.5.** *Supongamos que se satisfacen las hipótesis **N1** hasta **N10** y que  $s_n \xrightarrow{P} \sigma$ . Luego para cualquier estimador  $\hat{\boldsymbol{\beta}}$  solución de (3.7) consistente de  $\boldsymbol{\beta}_0$  se tiene que*

$$\sqrt{n} \left( \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \xrightarrow{D} \mathbf{N} \left( 0, A^{-1} \Sigma A^{-1T} \right)$$

donde las matrices  $\Sigma$  y  $A$  están definidas en **N16** y **N17** respectivamente.

### Demostración:

Definamos

$$L_n(\sigma, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}) - \eta_{\boldsymbol{\beta}}(t_i)}{\sigma} \right) W(\mathbf{x}_i) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta})^T,$$

$$\widehat{L}_n(\sigma, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}) - \widehat{\eta}_{\boldsymbol{\beta}}(t_i)}{\sigma} \right) W(\mathbf{x}_i) \widehat{\mathbf{z}}(\mathbf{x}_i, t_i, \boldsymbol{\beta})^T.$$

Notemos que  $\widehat{\boldsymbol{\beta}}$  satisface  $\widehat{L}_n(s_n, \widehat{\boldsymbol{\beta}}) = 0$  o equivalentemente si llamamos  $\widehat{L}_{n,j}$  a la  $j$ -ésima coordenada de  $\widehat{L}_n$ , entonces  $\widehat{\boldsymbol{\beta}}$  verifica  $\widehat{L}_{n,j}(s_n, \widehat{\boldsymbol{\beta}}) = 0$  para cada  $1 \leq j \leq q$ . Haciendo una expansión de Taylor de primer orden centrada en  $\boldsymbol{\beta} = \boldsymbol{\beta}_0$  de  $\widehat{L}_{n,j}(s_n, \boldsymbol{\beta})$  obtenemos:

$$\widehat{L}_{n,j}(s_n, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \widehat{\eta}_{\boldsymbol{\beta}_0}(t_i)}{s_n} \right) W(\mathbf{x}_i) \widehat{z}_j(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0) + A_n^{(j)}(\widetilde{\boldsymbol{\beta}})(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

donde

$$A_n^{(j)}(\widetilde{\boldsymbol{\beta}}) = \left\{ \begin{aligned} & \frac{-1}{n} \frac{1}{s_n} \sum_{i=1}^n \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \widetilde{\boldsymbol{\beta}}) - \widehat{\eta}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{s_n} \right) W(\mathbf{x}_i) \widehat{z}_j(\mathbf{x}_i, t_i, \widetilde{\boldsymbol{\beta}}) \widehat{\mathbf{z}}(\mathbf{x}_i, t_i, \widetilde{\boldsymbol{\beta}})^T \\ & + \frac{1}{n} \sum_{i=1}^n \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \widetilde{\boldsymbol{\beta}}) - \widehat{\eta}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{s_n} \right) W(\mathbf{x}_i) \nabla_{\boldsymbol{\beta}} \widehat{z}_j(\mathbf{x}_i, t_i, \boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}^T \end{aligned} \right\},$$

con  $\widetilde{\boldsymbol{\beta}}$  un punto intermedio entre  $\boldsymbol{\beta}$  y  $\boldsymbol{\beta}_0$ . Por lo tanto, si  $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$  y  $A_n \in \mathbb{R}^{q \times q}$  es la matriz cuya  $j$ -ésima fila es  $A_n^{(j)}(\widetilde{\boldsymbol{\beta}})$  obtenemos que

$$-\sqrt{n} A_n^{-1} \widehat{L}_n(s_n, \boldsymbol{\beta}_0) = \sqrt{n} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).$$

Del Lema 4.3.4 sabemos que  $A_n \xrightarrow{P} A$  donde la matriz  $A$  está definida en N9. Luego en orden de obtener la distribución asintótica de  $\widehat{\boldsymbol{\beta}}$  será suficiente analizar el comportamiento asintótico de  $\widehat{L}_n(s_n, \boldsymbol{\beta}_0)$ . Para ello veremos que

1.  $\sqrt{n} L_n(\sigma_0, \boldsymbol{\beta}_0) \xrightarrow{D} N(0, \Sigma)$  donde  $\Sigma$  es la matriz definida en N8.
2.  $\sqrt{n} [\widehat{L}_n(s_n, \boldsymbol{\beta}_0) - L_n(s_n, \boldsymbol{\beta}_0)] \xrightarrow{P} 0$ .
3.  $\sqrt{n} [L_n(s_n, \boldsymbol{\beta}_0) - L_n(\sigma_0, \boldsymbol{\beta}_0)] \xrightarrow{P} 0$ .

El primer punto se desprende fácilmente de N1, la simetría y la independencia de  $\epsilon$  y del Teorema Central del Límite.

Para demostrar el punto 2 usaremos argumentos análogos a los presentados en Boente, He et.all (2006). Notar que mediante una expansión de Taylor podemos escribir

$$\sqrt{n} [\widehat{L}_n(s_n, \boldsymbol{\beta}_0) - L_n(s_n, \boldsymbol{\beta}_0)] = L_n^1 + L_n^2 + L_n^3 + L_n^4,$$

donde,

$$\begin{aligned} L_n^1 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{s_n} \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \eta_0(t_i)}{s_n} \right) W(\mathbf{x}_i) \widehat{v}(t_i, \boldsymbol{\beta}_0) \mathbf{z}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0), \\ L_n^2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \eta_0(t_i)}{s_n} \right) W(\mathbf{x}_i) \widehat{v}(t_i, \boldsymbol{\beta}_0), \\ L_n^3 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{s_n} \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \eta_0(t_i)}{s_n} \right) W(\mathbf{x}_i) \widehat{v}(t_i, \boldsymbol{\beta}_0) \widehat{v}(t_i, \boldsymbol{\beta}_0), \\ L_n^4 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{s_n^2} \Psi_1'' \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \xi_i}{s_n} \right) W(\mathbf{x}_i) \widehat{v}(t_i, \boldsymbol{\beta}_0)^2 \widehat{\mathbf{z}}(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0). \end{aligned}$$

De las condiciones N2, N3, N6.2, N6.4, N7 y del hecho que  $s_n \rightarrow \sigma$  se desprende que  $L_n^3 \xrightarrow{P} 0$  y  $L_n^4 \xrightarrow{P} 0$ .

Veamos en lo que sigue que  $L_n^j \xrightarrow{P} 0$  para  $j = 1, 2$ .

Para demostrar la convergencia de  $L_n^1$  llamemos  $z_j(\mathbf{s}, \tau, \boldsymbol{\beta})$  a la  $j$ -coordenada de  $\mathbf{z}(\mathbf{s}, \tau, \boldsymbol{\beta})$  para cada  $1 \leq j \leq q$  y definamos

$$J_{n,1}^j(\sigma, v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{\sigma} \Psi_1' \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \eta_0(t_i)}{\sigma} \right) W(\mathbf{x}_i) v(t_i) z_j(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0),$$

luego bastará probar que para cualquier  $\epsilon > 0$ ,  $\mathbb{P}(|J_{n,1}^j(s_n, \widehat{v}(\tau, \boldsymbol{\beta}_0))| > \epsilon) \rightarrow 0$  para cada  $1 \leq j \leq q$ .

Sean los conjuntos  $\mathcal{V} = \{v \in C^1(T) : \|v\|_\infty \leq 1 \text{ y } \|v'\|_\infty \leq 1\}$  y  $\mathcal{P} = (\frac{\sigma_0}{2}, 2\sigma_0)$ . Notemos por  $|(v, \sigma)| = \|v\| + |\sigma|$  a la norma sobre  $\mathcal{V} \times \mathcal{P}$  y definamos la siguiente familia de funciones

$$\mathcal{F}_1^j = \left\{ f_{1,v,\sigma}^j(y, \mathbf{x}, t) = \frac{1}{\sigma} \Psi_1' \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma} \right) W(\mathbf{x}) v(t) z_j(\mathbf{x}, t, \boldsymbol{\beta}_0), v \in \mathcal{V}, \sigma \in \mathcal{P} \right\}$$

Mediante un desarrollo de Taylor de primer orden en la variable  $\sigma$ , usando N2, N3 y algunos manejos algebraicos obtenemos que

$$\begin{aligned} |f_{1,v_l,\sigma_s}^j(y, \mathbf{x}, t) - f_{1,v,\sigma}^j(y, \mathbf{x}, t)| &\leq \sqrt{W(\mathbf{x})} |z_j(\mathbf{x}, t, \boldsymbol{\beta}_0)| \left\{ \left| \frac{1}{\sigma} \Psi_1' \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma} \right) \sqrt{W(\mathbf{x})} |v(t) - v_l(t)| \right. \right. \\ &\quad + \frac{1}{(\sigma^*)^2} \Psi_1'' \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma^*} \right) \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma^*} \right) \sqrt{W(\mathbf{x})} v_l(t) |\sigma_s - \sigma| \\ &\quad \left. \left. + \frac{1}{(\sigma^*)^2} \Psi_1' \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma^*} \right) \sqrt{W(\mathbf{x})} v_l(t) |\sigma_s - \sigma| \right\} \\ &\leq \sqrt{W(\mathbf{x})} |z_j(\mathbf{x}, t, \boldsymbol{\beta}_0)| \left( \frac{4 + 2\sigma_0}{\sigma_0^2} \right) \|\sqrt{W}\|_\infty \left[ 2\|\Psi_1'\|_\infty + \|\Psi_3\|_\infty \right] |(v_l, \sigma_s) - (v, \sigma)| \end{aligned}$$

donde  $\sigma^*$  es un punto intermedio entre  $\sigma$  y  $\sigma_s$ .

Luego llamando  $A_1 = \frac{4+2\sigma_0}{\sigma_0^2} \|\sqrt{W}\|_\infty \left[ 2\|\Psi'_1\|_\infty + \|\Psi_3\|_\infty \right]$  y  $S(\mathbf{x}, t) = A_1 \sqrt{W(\mathbf{x})} |z_j(\mathbf{x}, t, \boldsymbol{\beta}_0)|$  tenemos que, por un lado, **N7** implica que la función  $S$  es una envolvente en  $L^2$  para la familia  $\mathcal{F}_1^j$  y además se verifica que

$$|f_{1,v_l,\sigma_s}^j - f_{1,v,\sigma}^j| \leq S(\mathbf{x}, t) |(v_l, \sigma_s) - (v, \sigma)| \quad (4.15)$$

Por otro lado, como vale **N10** tenemos que  $\mathbb{E}(f(y, \mathbf{x}, t)) = 0$  para todo  $f \in \mathcal{F}_1^j$ . Luego, debido a que vale (4.15), por el teorema (4.1.9), tenemos que la familia  $\mathcal{F}_1^j$  verifica la siguiente la condición de bracketing

$$N_{[\cdot]}(2\epsilon \|S\|_{\mathbb{P},2}, \mathcal{F}_1^j, L^2(\mathbb{P})) \leq N(\epsilon, \mathcal{V}, L^2(\mathbb{P})) N(\epsilon, \mathcal{P}, |\cdot|)$$

y, por lo tanto, por la proposición (4.1.17) junto con la observación (4.1.8) se tiene que la familia  $\mathcal{F}_1^j$  satisface que

$$J_{[\cdot]}(\delta, \mathcal{F}_1^j) \xrightarrow{\delta \rightarrow 0} 0. \quad (4.16)$$

Además, si  $\mathcal{F}_0 \subseteq \mathcal{F}_1^j$  y la envolvente  $S$  es usada para  $\mathcal{F}_0$  se tiene que  $J_{[\cdot]}(\delta, \mathcal{F}_0) \leq J_{[\cdot]}(\delta, \mathcal{F}_1^j)$ .

Por consiguiente, usando **N6.2**, **N6.3** y del hecho que  $s_n \xrightarrow{P} \sigma_0$  tenemos que dado  $\delta > 0$  existe  $n_0 \in \mathbb{N}$  tal que para todo  $n > n_0$  se cumple

$$\mathbb{P}(|\sigma_0 - s_n| > \delta) < \frac{\delta}{2} \quad \text{y} \quad \mathbb{P}(\hat{v}(\tau, \boldsymbol{\beta}_0) \in \mathcal{V} \text{ y } \|\hat{v}(\cdot, \boldsymbol{\beta}_0)\|_\infty > \delta) < \frac{\delta}{2},$$

por lo tanto, dado  $\epsilon > 0$  tenemos que

$$\begin{aligned} \mathbb{P}\left(|J_{n,1}^j(s_n, \hat{v}(\tau, \boldsymbol{\beta}_0))| > \epsilon\right) &\leq \mathbb{P}\left(|J_{n,1}^j(s_n, \hat{v}(\tau, \boldsymbol{\beta}_0))| > \epsilon, |\sigma_0 - s_n| < \delta, \|\hat{v}(\tau, \boldsymbol{\beta}_0)\|_\infty < \delta\right) \\ &\quad + \mathbb{P}\left(|\sigma_0 - s_n| > \delta \text{ ó } \|\hat{v}(\tau, \boldsymbol{\beta}_0)\|_\infty > \delta\right) \\ &\leq \mathbb{P}\left(\sup_{|\sigma_0 - \sigma| < \delta, \|v\|_\infty < \delta} |J_{n,1}^j(\sigma, v)| > \epsilon\right) + \delta \\ &\leq \frac{1}{\epsilon} \mathbb{E}\left(\sup_{|\sigma_0 - \sigma| < \delta, \|v\|_\infty < \delta} |J_{n,1}^j(\sigma, v)|\right) + \delta. \end{aligned}$$

Luego es suficiente ver que

$$\lim_{n \rightarrow \infty} \lim_{\delta \rightarrow 0} \mathbb{E}\left(\sup_{|\sigma_0 - \sigma| < \delta, \|v\|_\infty < \delta} |J_{n,1}^j(\sigma, v)|\right) = 0.$$

Dado  $\delta > 0$ , consideremos la subclase

$$\mathcal{F}_{\delta,1}^j = \left\{ f_{1,v,\sigma}^j(y, \mathbf{x}, t) \in \mathcal{F}_1^j : \|v\|_\infty < \delta \right\}$$

Es sencillo ver que la subclase  $\mathcal{F}_{\delta,1}^j$  cumple que  $\|f_{1,v,\sigma}^j\|_{2,\mathbb{P}} \leq \delta \|F\|_{2,\mathbb{P}}$  para toda  $f_{1,v,\sigma}^j \in \mathcal{F}_{\delta,1}^j$  donde la función envolvente  $S(\mathbf{x}, t)$  fue definida anteriormente. Luego, de la desigualdad maximal enunciada en la Proposición (4.1.18), tenemos que

$$\begin{aligned} \mathbb{E} \left( \sup_{|\sigma_0 - \sigma| < \delta, \|v\| < \delta} |J_{n,1}^j(\sigma, v)| \right) &\leq D_1 J_{[\cdot]}(\delta, \mathcal{F}_{\delta,1}^j) \|S\|_{2,\mathbb{P}} + \sqrt{n} \mathbb{E} \left( S(\mathbf{x}, t) I_{\{F(\mathbf{x}, t) > \sqrt{na}(\delta)\}} \right) \\ &\leq D_1 J_{[\cdot]}(\delta, \mathcal{F}_1^j) \|S\|_{2,\mathbb{P}} + \frac{1}{a(\delta)} \mathbb{E} \left( S^2(\mathbf{x}, t) I_{\{S(\mathbf{x}, t) > \sqrt{na}(\delta)\}} \right). \end{aligned}$$

Dado que la función  $S$  tiene segundo momento finito se desprende que para  $\delta$  fijo,

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}} \left( S^2 I_{\{F > \sqrt{na}(\delta)\}} \right) = 0,$$

entonces

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \sup_{|\sigma_0 - \sigma| < \delta, \|v\| < \delta} |J_{n,1}^j(\sigma, v)| \right) \leq D_1 J_{[\cdot]}(\delta, \mathcal{F}_1^j) \|S\|_{2,\mathbb{P}}.$$

Luego, por lo dicho en (4.16) sabemos que  $J_{[\cdot]}(\delta, \mathcal{F}_1^j) \xrightarrow{\delta \rightarrow 0} 0$  concluyendo así lo que se quería probar.

Similarmente, para demostrar la convergencia de  $L_n^2$ , definamos por  $\widehat{\nu}_{j,0}(\tau)$  a la  $j$ -coordenada de  $\widehat{\nu}(\tau, \beta_0)$  y

$$J_{n,2}(\sigma, v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \beta_0) - \eta_0(t_i)}{\sigma} \right) W(\mathbf{x}_i) v(t_i),$$

luego bastará probar que para cualquier  $\epsilon > 0$ ,  $\mathbb{P}(|J_{n,2}(s_n, \widehat{\nu}_{j,0}(\tau))| > \epsilon) \rightarrow 0$ . Para ello llamaremos

$$\begin{aligned} \mathcal{F}_2 &= \left\{ f_{2,v,\sigma}(y, \mathbf{x}, t) = \Psi_1 \left( \frac{y - g(\mathbf{x}, \beta_0) - \eta_0(t)}{\sigma} \right) W(\mathbf{x}) v(t), v \in \mathcal{V}, \sigma \in \mathcal{P} \right\} \quad \text{y} \\ \mathcal{F}_{\delta,2} &= \{ f_{2,v,\sigma}(y, \mathbf{x}, t) \in \mathcal{F}_2 : \|v\|_\infty < \delta \text{ y } |\sigma - \sigma_0| < \delta \}. \end{aligned}$$

Notar que por **N1**, la simetría de  $\epsilon$  y su independencia con  $(\mathbf{x}, t)$  obtenemos que  $\mathbb{E}(f_{2,v,\sigma}) = 0$ , y junto con **N3** tenemos que existe una constante positiva  $C_2$  tal que  $|f_{2,v,\sigma}(y, \mathbf{x}, t)| \leq C_2$  para todo  $v \in \mathcal{V}$  y  $\sigma \in \mathcal{P}$  y por último observemos que bajo **N2**,

$$|f_{2,v_l,\sigma_s} - f_{2,v,\sigma}| \leq A_2 |(v_l, \sigma_s) - (v, \sigma)|,$$

con  $A_2 = \|W\|_\infty \left[ \|\Psi_1\|_\infty + \frac{2}{\sigma_0} \|\Psi_2\|_\infty \right]$ .

Luego, siguiendo la línea de razonamiento anterior, tenemos que

$$J_{[\cdot]}(\delta, \mathcal{F}_2) \xrightarrow{\delta \rightarrow 0} 0.$$

Por consiguiente, para cada  $1 \leq j \leq q$ , por **N6.2**, **N6.3** y el hecho de que  $s_n \xrightarrow{P} \sigma_0$ , entonces dado  $\delta > 0$  existe  $n_0 \in \mathbb{N}$  tal que para todo  $n > n_0$  se cumple

$$\mathbb{P}(|\sigma_0 - s_n| > \delta) < \frac{\delta}{2} \quad \text{y} \quad \mathbb{P}(\hat{\nu}_{j,0} \in \mathcal{V} \text{ y } \|\hat{\nu}_{j,0}\|_\infty > \delta) < \frac{\delta}{2}.$$

Trabajando de la misma manera que la descrita anteriormente tenemos para cada  $1 \leq j \leq q$  dado  $\epsilon > 0$ , existe  $D_2$  y  $n_0 \in \mathbb{N}$  tal que para todo  $n > n_0$

$$\mathbb{P}(|J_{n,2}(s_n, \hat{\nu}_{j,0})| > \epsilon) \leq \frac{1}{\epsilon} D_2 J_{[\cdot]}(\delta, \mathcal{F}_2) + \delta.$$

Por lo tanto,  $L_n^2 = (J_{n,2}(s_n, v_{0,1}), \dots, J_{n,2}(s_n, v_{0,q})) \xrightarrow{P} 0$  como se quería demostrar.

Para finalizar faltaría probar el punto 3, es decir que  $\sqrt{n}[L_n(s_n, \beta_0) - L_n(\sigma_0, \beta_0)] \xrightarrow{P} 0$ . Este resultado se obtiene usando argumentos similares al anterior por lo que su demostración será delegada al apéndice.



## Capítulo 5

# Estudio de Simulación

En este Capítulo se describirán los resultados de un estudio de simulación cuyo objetivo es medir la performance del estimador robusto para un modelo parcialmente no lineal propuesto en (3.3), al que notaremos por ROB, y comparar su comportamiento versus una de sus versiones de estimación clásica descrita en (3.2) la que notaremos por CLA.

Para llevar a cabo los distintos pasos de nuestro estimador definido en (3.3) se consideró por un lado, en el **Paso 1**, la función núcleo de Epanechnikov para el cálculo de los pesos  $w_i$ , es decir,

$$K(t) = 3/4(1 - t^2)I_{|t| \leq 1} \quad (5.1)$$

con diferentes valores para la ventana  $h$ . Además para la  $\rho$ -función se eligió la  $\rho_k$ -función de Huber, es decir

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

con constante de calibración  $k = 1,345$ . Por último para  $\widehat{s}_\beta(t)$  usamos la MAD local muestral definida en (3.6).

Por otro lado, en el **Paso 2**, se usaron para  $\rho_0$  y  $\rho_1$  una  $\rho$ -función dentro de la familia bicuadrada de Tukey definida por

$$\rho_c = \min\{1 - (1 - (x/c)^2)^3, 1\}$$

con constante de calibración  $c = 4$  y se consideró  $b = 0.5$ . Además, para el cálculo de los pesos  $W(\mathbf{x})$  se penalizó la distancia de Mahalanobis robusta de la covariable  $\mathbf{x}$ , para lo cual se computaron estimadores robustos,  $\widehat{\mu}$  y  $\widehat{\Sigma}$ , de los parámetros de posición y escala, respectivamente, basados en  $\mathbf{x}_1, \dots, \mathbf{x}_n$  y se consideró como función penalizadora con constante de calibración  $d = \chi_{0.975,3}^2$  a

$$\kappa(x) = I_{(|x| \leq d)} (1 - (x/d)^2)^2,$$

finalmente los pesos fueron calculados de la forma

$$W(\mathbf{x}) = \kappa \left( (\mathbf{x} - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}) \right).$$

Para medir la performance del estimador propuesto  $\hat{\boldsymbol{\beta}}$  para  $\boldsymbol{\beta}_0$  en base a los resultados obtenidos calculamos como medidas resumen el desvío estándar (SD), el sesgo (*BIAS*) y el error cuadrático medio (*MSE*) para cada una de sus componentes. Por otro lado, para medir la performance de la estimación no paramétrica  $\hat{\eta}$  para  $\eta_0$  calculamos la media, la mediana y el desvío estándar del error  $MSE(\hat{\eta})$ , que denotamos por *M.MSE*, *Med.MSE* y *SD.MSE* respectivamente, definido de la siguiente manera

$$MSE(\hat{\eta}) = \frac{1}{n} \sum_{i=1}^n (\hat{\eta}(t_i) - \eta_0(t_i))^2.$$

Para este estudio de simulación se realizaron  $N = 1000$  replicaciones en las que se generaron muestras aleatorias independientes  $(y_i, \mathbf{x}_i^T, t_i)$  de tamaño  $n = 100$  para el modelo parcialmente no lineal (3.1) donde consideramos como  $g(\mathbf{x}, \boldsymbol{\beta}) = (\mathbf{x}^T \boldsymbol{\beta}_0)^2$  con  $\boldsymbol{\beta}_0 = (1, 2, 1.5)^T$  y  $\eta_0(t) = \sin(2\pi t)$  con  $t \sim U[0, 1]$ . Para los procedimientos de suavizado se utilizaron distintos valores para la ventana  $h$ , sólo presentaremos los resultados obtenidos para el valor  $h = 0.3$  dado que los resultados obtenidos para cada una de ellas son muy similares entre si. Los estimadores propuestos se evaluaron tanto bajo muestras sin contaminar, a la que llamaremos  $C_0$ , como también bajo muestras contaminadas  $C_1, C_2, C_3$  y  $C_4$  definidas a contaminación. Los 1000 conjuntos de datos fueron generados para cada uno de los siguientes cinco escenarios considerados en Jiang et al. (2017):

$C_0$ :  $\mathbf{x} = (x_1, x_2, x_3) \sim N_3(\mathbf{0}, \mathbf{I}_3)$  y la variable  $\epsilon \sim N(0, 1)$ .

$C_1$ :  $\mathbf{x} = (x_1, x_2, x_3) \sim N_3(\mathbf{0}, \mathbf{I}_3)$  y la variable  $\epsilon \sim t_1$ , la distribución  $\mathcal{T}$  de Student con un grado de libertad, es decir la distribución estándar Cauchy.

$C_2$ :  $\mathbf{x} = (x_1, x_2, x_3) \sim N_3(\mathbf{0}, \mathbf{I}_3)$  y la variable  $\epsilon \sim 0.9N(0, 1) + 0.1N(0, 10^2)$ .

$C_3$ :  $\mathbf{x} = (x_1, x_2, x_3) \sim N_3(\mathbf{0}, \mathbf{I}_3)$  reemplazando un 10% de los valores originales por valores atípicos  $(5, 5, 5)$  y la variable  $\epsilon \sim t_1$ .

$C_4$ :  $\mathbf{x} = (x_1, x_2, x_3) \sim N_3(\mathbf{0}, \mathbf{I}_3)$  reemplazando un 10% de los valores originales por valores atípicos  $(5, 5, 5)$  y la variable  $\epsilon \sim N(0, 1)$ .

El primer escenario es para evaluar la performance de nuestra propuesta y poder compararlo con el método clásico cuando los errores provienen de una distribución normal y no hay valores atípicos en la muestra. El segundo y tercer escenario consideran errores con colas pesadas. El cuarto escenario estudia la robustez del estimador propuesto bajo la presencia de outliers tanto en la variable  $\mathbf{x}$  como en los errores. Por último, el quinto escenario considera outliers de alto leverage en la variable  $\mathbf{x}$ .

Los siguientes gráficos y tablas resumen los resultados de la simulación. Los boxplot y las estimaciones de la densidad corresponden a las propuestas de estimación de los parámetros  $\beta_{01}$ ,  $\beta_{02}$  y  $\beta_{03}$  bajo los distintos escenarios considerados, mientras que las tablas muestran las medidas resumen  $SD$ ,  $BIAS$  y el  $MSE$  para cada uno de ellos y  $M.MSE$ ,  $Med.MSE$  y  $SD.MSE$  para la función  $\eta_0$ . Por otro lado, en la Figura 5.8 se muestra el gráfico para las estimaciones clásica y robusta de la función  $\eta_0$  en una de las mil replicaciones versus la gráfica de la verdadera función en cada uno de los escenarios propuestos.

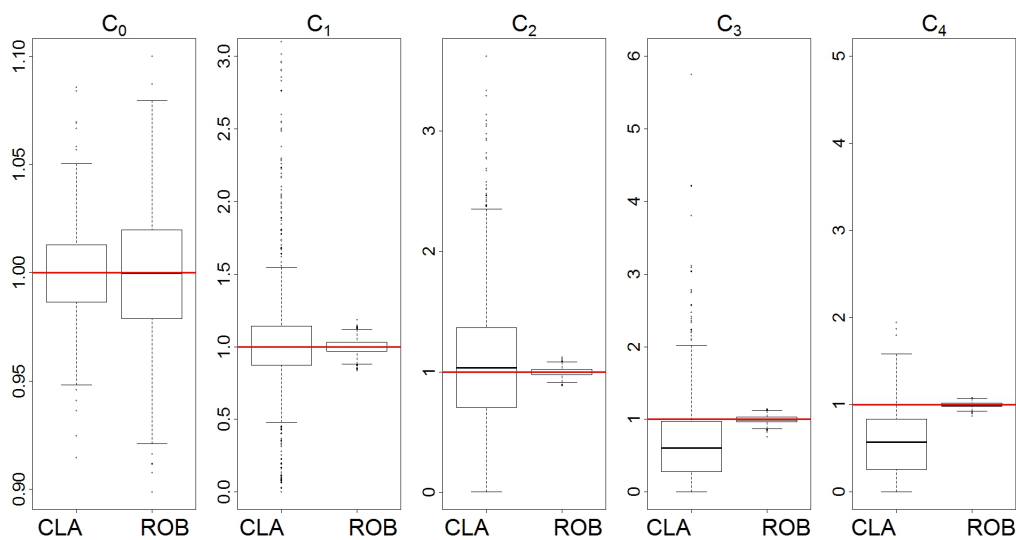


Figura 5.1: Boxplots de  $\hat{\beta}_1$  por el método CLA vs. el método ROB bajo  $C_0$  y las contaminaciones  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_4$  de izquierda a derecha. La línea roja se encuentra sobre el verdadero valor de  $\beta_{01}$ .

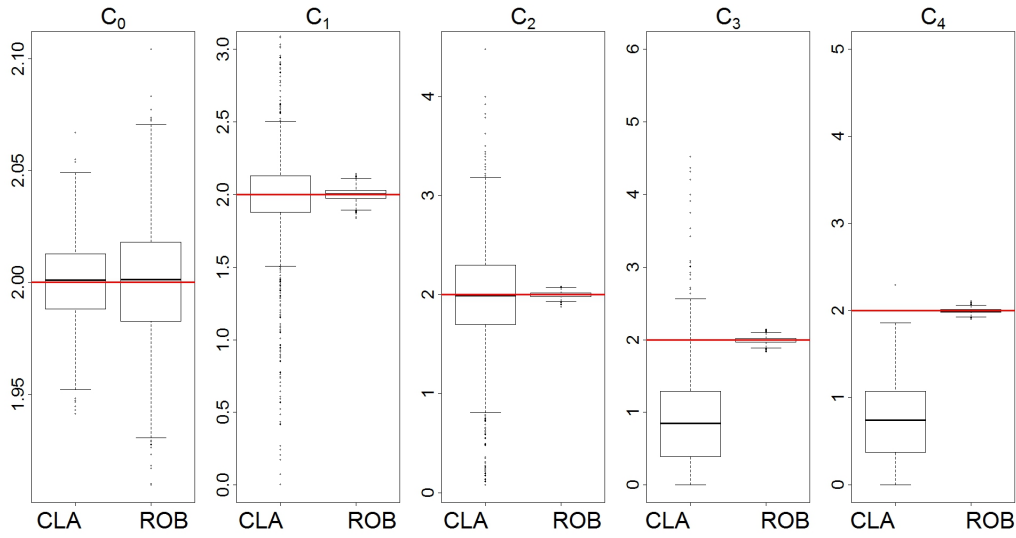


Figura 5.2: Boxplots de  $\hat{\beta}_2$  por el método CLA vs. el método ROB bajo  $C_0$  y las contaminaciones  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_4$  de izquierda a derecha. La línea roja se encuentra sobre el verdadero valor de  $\beta_{02}$ .

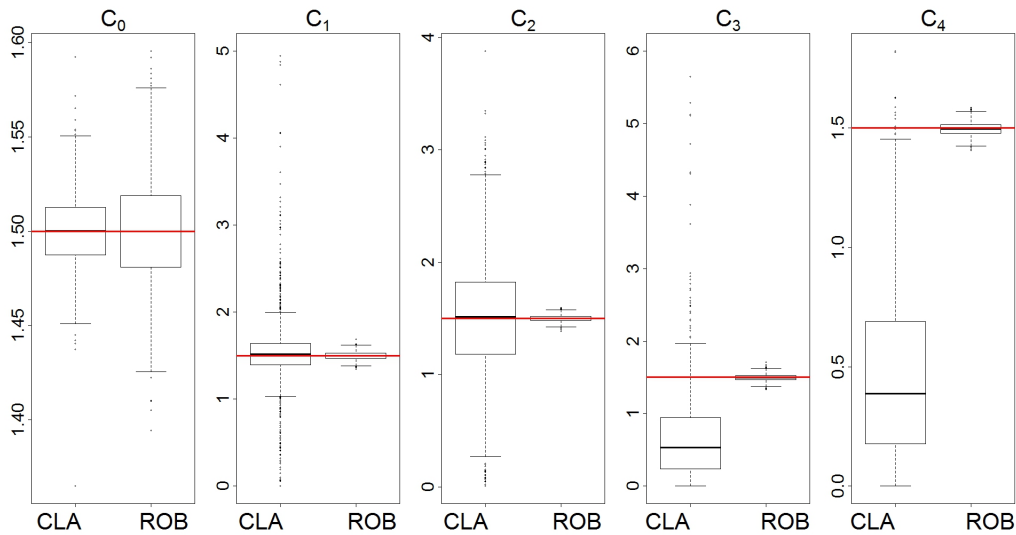


Figura 5.3: Boxplots de  $\hat{\beta}_3$  por el método CLA vs. el método ROB bajo  $C_0$  y las contaminaciones  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_4$  de izquierda a derecha. La línea roja se encuentra sobre el verdadero valor de  $\beta_{03}$ .

	Método	$\beta_{01} = 1$			$\beta_{02} = 2$			$\beta_{03} = 1.5$		
		<i>BIAS</i>	<i>SD</i>	<i>MSE</i>	<i>BIAS</i>	<i>SD</i>	<i>MSE</i>	<i>BIAS</i>	<i>SD</i>	<i>MSE</i>
$C_0$	CLA	0.0005	0.0206	0.0004	-0.0001	0.0187	0.0003	-0.0003	0.002	0.0004
	ROB	0.0004	0.0312	0.0010	-0.0090	0.0274	0.0007	0.0002	0.0297	0.0009
$C_1$	CLA	-0.0820	0.6081	0.3762	-0.0270	0.6081	0.3701	-0.0710	0.6712	0.4552
	ROB	-0.0004	0.0500	0.0025	-0.0009	0.0445	0.0020	-0.0005	0.0474	0.0022
$C_2$	CLA	-0.0600	0.5653	0.3229	0.036	0.5839	0.342	-0.001	0.593	0.3513
	ROB	0.0008	0.0314	0.0010	-0.0002	0.0274	0.0008	0.0015	0.0297	0.0009
$C_3$	CLA	0.2785	0.7084	0.5789	1.073	0.7754	1.7520	0.8057	0.7620	1.2292
	ROB	0.0010	0.0513	0.0026	0.0077	0.0458	0.0022	0.0044	0.0507	0.0026
$C_4$	CLA	0.4255	0.3621	0.3120	1.2535	0.4350	1.7612	1.0360	0.3570	1.2010
	ROB	0.0015	0.0281	0.0008	0.0043	0.0271	0.0008	0.0040	0.0284	0.0008

Cuadro 5.1: Medidas resumen para las estimaciones de los parámetros  $\beta_{01}$ ,  $\beta_{02}$ ,  $\beta_{03}$  bajo los escenarios  $C_0$ ,  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_4$  descritos anteriormente.

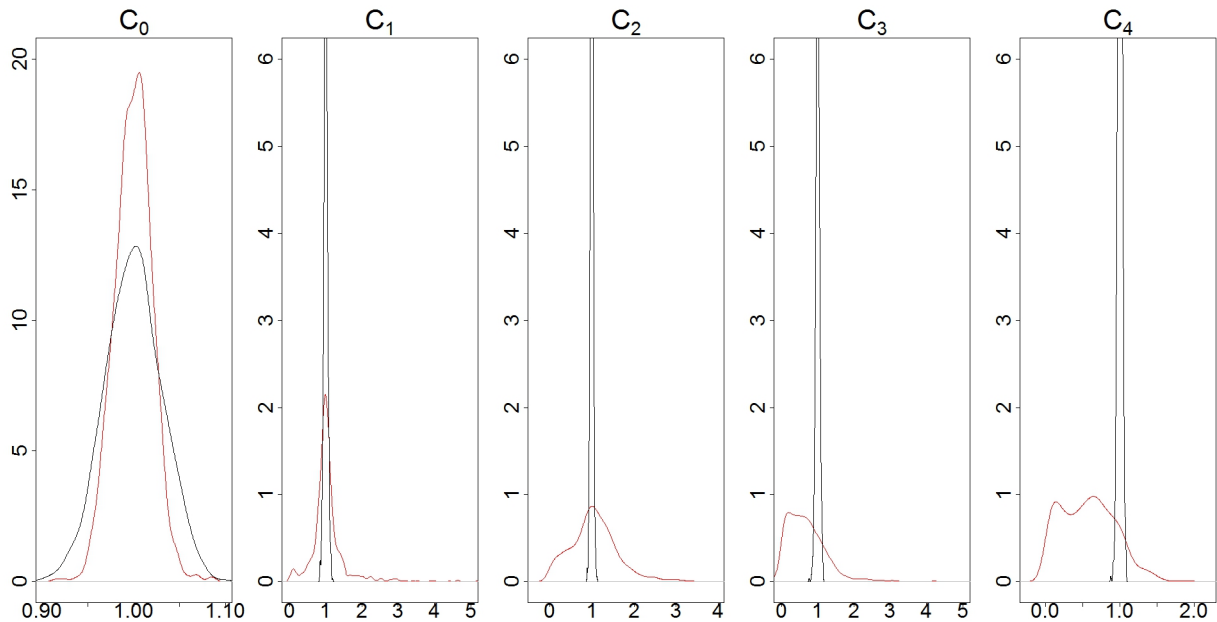


Figura 5.4: Estimación de la densidad dados los 1000 valores estimados de  $\beta_{01}$  para el método clásico en rojo y robusto en negro bajo  $C_0$ ,  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_4$ .

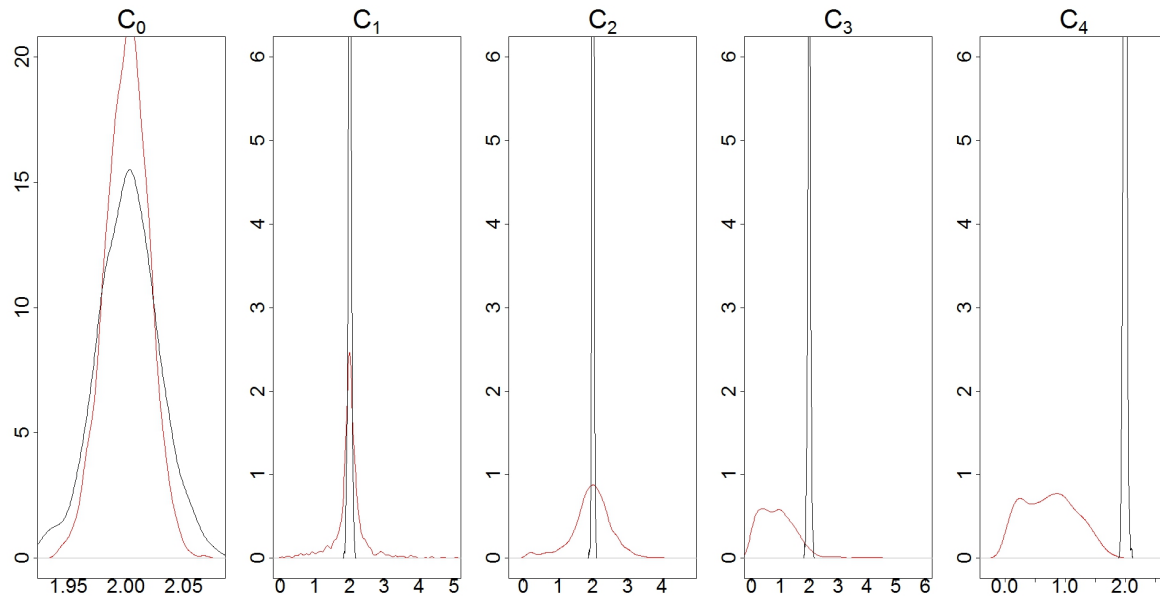


Figura 5.5: Estimación de la densidad dados los 1000 valores estimados de  $\beta_{02}$  para el método clásico en rojo y robusto en negro bajo  $C_0$ ,  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_4$ .

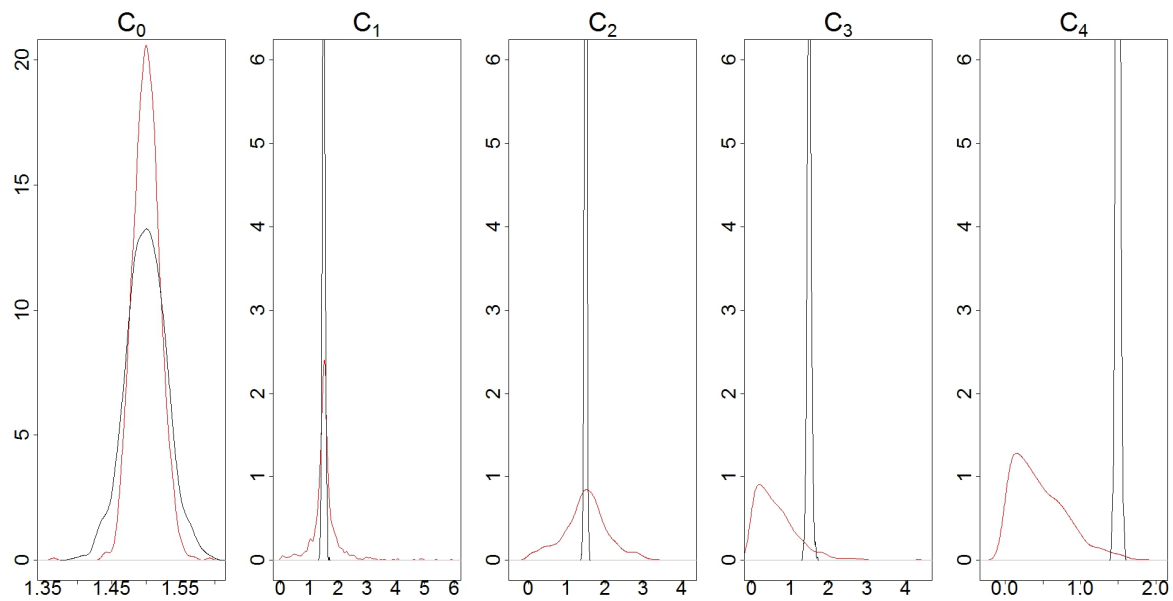


Figura 5.6: Estimación de la densidad dados los 1000 valores estimados de  $\beta_{03}$  para el método clásico en rojo y robusto en negro bajo  $C_0$ ,  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_4$ .

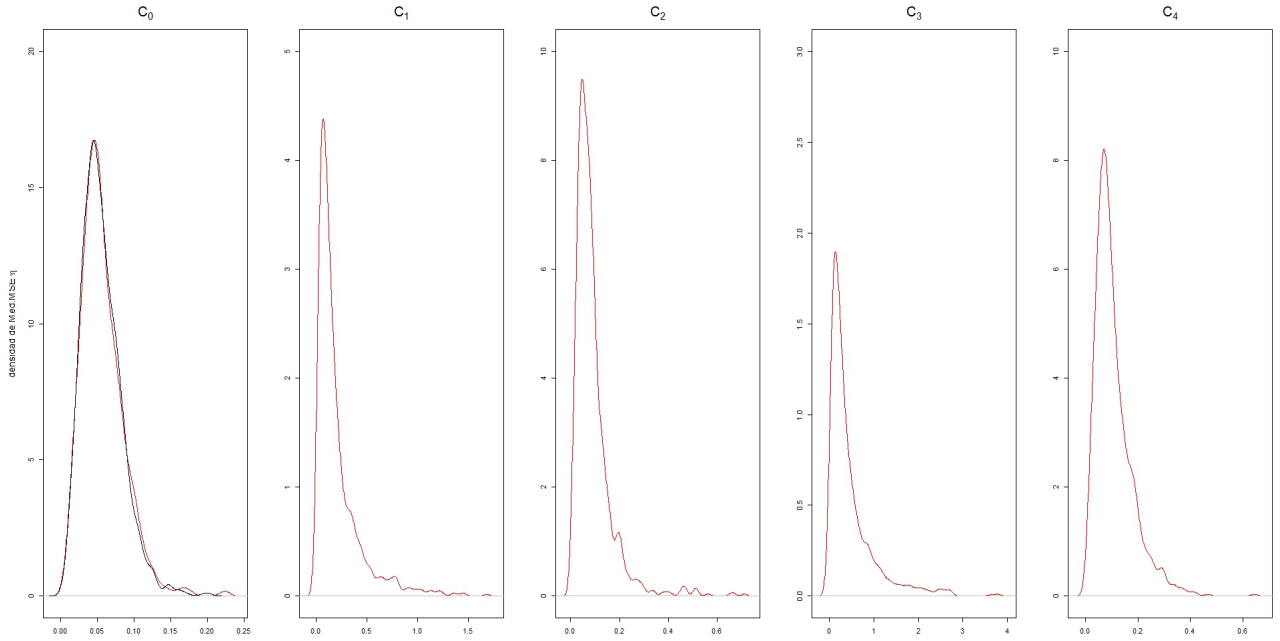


Figura 5.7: Estimación de la densidad de  $Med.MSE(\hat{\eta})$  obtenidas bajo los cinco escenarios descritos anteriormente. La curva en negro corresponde al método CLA y las dadas en rojo corresponden al método ROB.

Estimación para $\eta_0$				
	Método	$M.MSE$	$SD.MSE$	$Med.MSE$
$C_0$	CLA	0.0926	0.0331	0.0869
	ROB	0.0997	0.0437	0.0923
$C_1$	CLA	5.629	8328	3.668
	ROB	0.1783	0.1275	0.1438
$C_2$	CLA	30.9236	35.0628	20.3381
	ROB	0.1084	0.0511	0.0981
$C_3$	CLA	573.5674	7893.1058	31.9639
	ROB	0.3012	0.2765	0.2257
$C_4$	CLA	31.5823	13.9275	29.4050
	ROB	0.1286	0.0651	0.1151

Cuadro 5.2: Medidas resumen para la función no paramétrica  $\eta_0$  bajo los cinco escenarios descritos anteriormente.

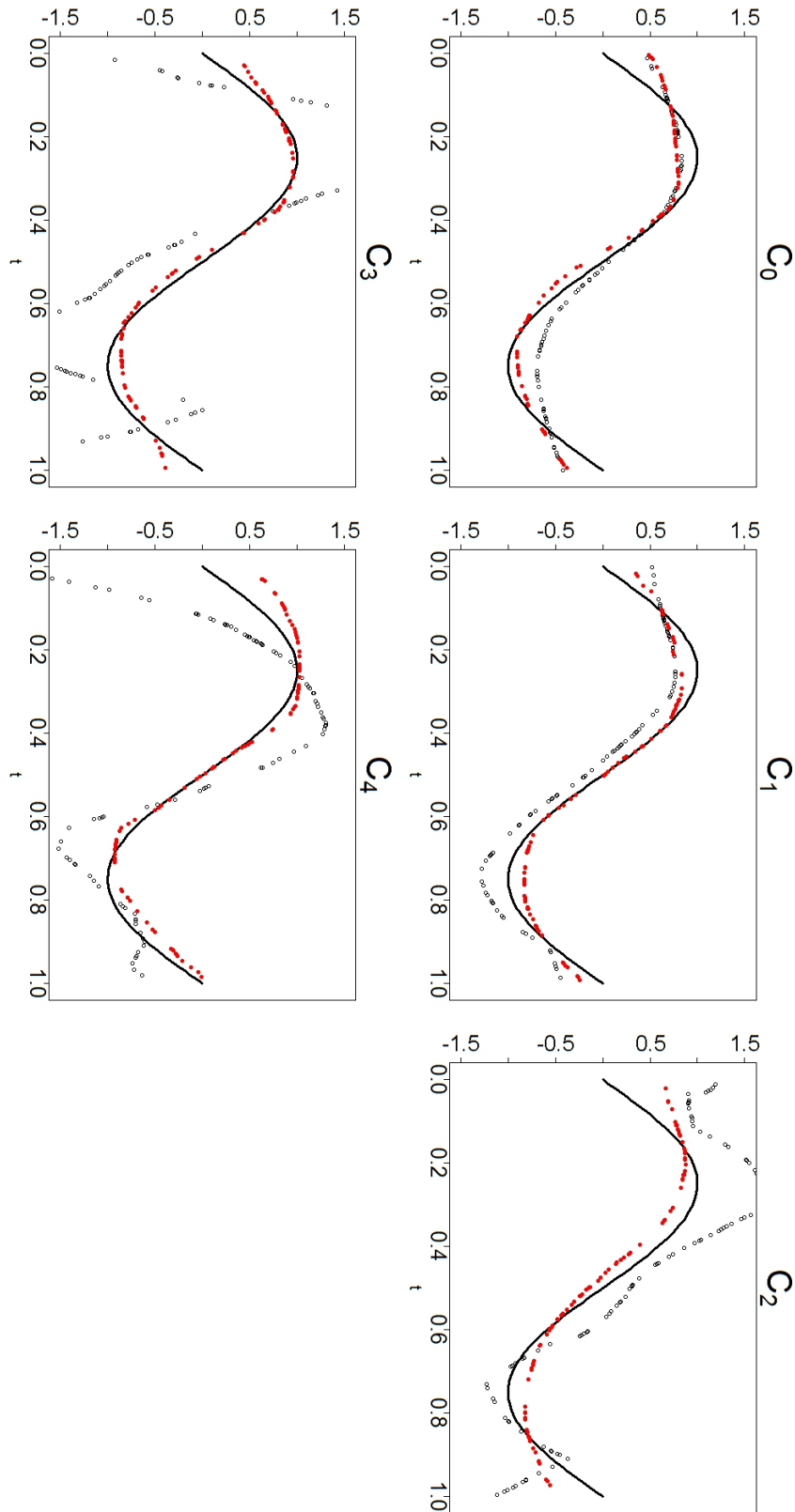


Figura 5.8: Con trazo negro se muestra la curva real  $\eta_0$ , en puntos rojos y negros se muestran sus valores estimados con el método ROB y CLA respectivamente.



En los Cuadros 5.1 y 5.2 podemos observar que bajo el escenario  $C_0$ , si bien, como es de esperar, existe una diferencia entre los estimadores, ésta es mínima. Es decir, el comportamiento del estimador robusto propuesto, tanto para el parámetro  $\beta_0$  como para la función  $\eta_0$ , es similar al del estimador de mínimos cuadrados cuando la muestra está sin contaminar.

Al introducir distintas alteraciones en la muestra podemos ver como el estimador de mínimos cuadrados es severamente afectado en comparación del estimador robusto.

Por un lado, en la estimación para el parámetro de regresión  $\beta_0$  usando el método clásico bajo los escenarios  $C_1$  y  $C_2$  podemos ver que existe un aumento considerable en los valores de los *ECM*'s pero no así en los valores de los sesgos (*BIAS*). Es decir, este aumento pareciera ser por el incremento en los valores de los desvíos (*SD*). Esta situación se ve reflejada, no sólo en las filas correspondientes a  $C_1$  y  $C_2$  del Cuadro 5.1 sino también puede ser observada gráficamente en los boxplot correspondientes dados en las Figuras 5.1, 5.2, 5.3 y en la Figuras 5.4, 5.5 y 5.6 en las que se muestran las estimaciones de las densidad de los estimadores de los parámetros de regresión.

Bajo los escenarios propuestos en  $C_3$  y  $C_4$  es donde se encuentran las mayores diferencias entre ambos estimadores. Podemos observar que, usando el método clásico, los *ECM*'s del parámetro  $\beta$  aumentan hasta en diez mil veces su valor con respecto al escenario  $C_0$  dado, a diferencia de los escenarios  $C_1$  y  $C_2$ , por un aumento en los valores del sesgo (*BIAS*), mientras que el estimador robusto se conserva dentro de rangos razonables.

Por otro lado, podemos observar que la estimación robusta (ROB) para la función  $\eta_0$  bajo una muestra sin contaminar tiene un comportamiento similar a la estimación clásica (CLA). Bajo los distintos escenarios de contaminación propuestos el estimador clásico para  $\eta_0$  se ve afectado a niveles exagerados perdiendo casi el sentido de la "buena" estimación, mientras que el estimador robusto conserva la línea como puede observarse en los valores resúmenes del Cuadro 5.2 y en las Figuras 5.7 y 5.8.

## Capítulo 6

# Ejemplo con datos reales

En esta sección, con el objetivo de ilustrar el uso de los estimadores propuestos, estudiaremos el ajuste de un modelo parcialmente no lineal (3.1) al conjunto de datos meteorológicos “*airquality*” obtenidos del Departamento de Conservación del Estado de Nueva York los cuales se encuentran disponibles en el Software R . Con la intención de estudiar la calidad del aire en el área metropolitana de Nueva York fueron registrados diariamente entre el 1° de mayo y el 30 de septiembre de 1973 un total de 153 observaciones, cada registro constaba de mediciones de cuatro variables: ozono (O), radiación solar, promedio de la velocidad del viento (V) y la temperatura máxima medida en grados Fahrenheit (T). Es importante resaltar que este conjunto de datos posee observaciones faltantes, entre las 153 observaciones se encuentran 37 con datos faltantes, para nuestro análisis y a modo de ilustrar nuestro procedimiento trabajaremos sólo con las 116 que fueron registradas de modo completo.

Este conjunto de datos fue introducido en Chambers et al. (1983). Cleveland (1985) estudió la relación entre el ozono y la velocidad del viento detectando una relación no lineal en la cual a medida que la velocidad aumenta, el ozono decrece. Esto se debe al incremento de ventilación que se produce cuando la velocidad del viento es mayor. En Spano (2016) retoman el análisis propuesto por Cleveland (1985) pero, a diferencia de éste, no descartan los datos incompletos si no que trabajan en un contexto donde puede haber datos faltantes. Luego de probar diferentes modelos obtienen que utilizar una regresión no lineal con una función de decrecimiento exponencial dado por la siguiente función de regresión

$$g(V, \beta_1, \beta_2, \beta_3) = \beta_1 \exp(\beta_2 V) + \beta_3 \quad (6.1)$$

resulta un modelo adecuado para explicar la relación entre el ozono y la velocidad del viento. Asimismo, realizan un estudio para detectar valores atípicos en la muestra donde se identifican y explican la presencia de 5 outliers, que corresponden a las observaciones número 86, 100, 101, 121 y 126 de la muestra original.

En lo que sigue discutiremos si resulta razonable la inclusión de la variable temperatura a fin de obtener un mejor ajuste de la variable ozono. Para ello consideraremos un modelo de regresión no paramétrico definido en (2.1),

$$O = m(V, T) + \epsilon$$

donde  $m : \mathbb{R}^2 \rightarrow \mathbb{R}$  es una función suave y  $\epsilon$  una variable aleatoria en  $\mathbb{R}$ .

Para un primer análisis a modo descriptivo estimaremos la función  $m$  con el procedimiento de estimación robusto descrito en (2.1.2) y tomando como función núcleo el de Epanechnikov (5.1). Realizamos el ajuste para distintos valores de ventanas y como se puede apreciar en la Figura 6.1, debido a que el rango de las observaciones temperatura y velocidad del viento es considerablemente diferente usamos dos ventanas diferentes.

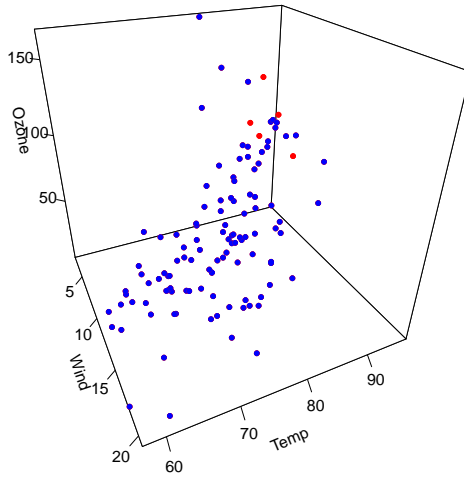


Figura 6.1: Conjunto de datos correspondientes al ozono, la temperatura y la velocidad del viento, en rojo se pueden apreciar los puntos considerados atípicos en Spano (2016).

La Figura 6.2 ilustra algunos de los ajustes obtenidos para diferentes valores de ventanas, en ella se observa por un lado que a pesar de considerar diferentes valores de ancho de banda el comportamiento de la función de regresión estimada no presenta un cambio considerable de forma. Por otro, lado observamos claramente que la variable ozono tiene una dependencia importante para diferentes condiciones de temperatura.

Luego, a partir del análisis anterior consideraremos la inclusión de la variable temperatura al modelo de regresión no lineal considerado en Spano (2016), proponiendo una dependencia no paramétrica con la variable ozono, es decir ajustaremos el siguiente modelo parcialmente no lineal

$$O = \beta_1 \exp(\beta_2 V) + \eta(T) + \epsilon. \quad (6.2)$$

Notemos que a diferencia del modelo puramente no lineal, en este modelo no consideramos el parámetro  $\beta_3$  ya que será modelado como parte de la función  $\eta$  evitando un problema de identificabilidad.

En la Tabla 6.1 resumiremos el error cuadrático medio ( $ECM$ ) y error cuadrático mediano ( $ECMed$ ) calculados con el fin de evaluar los diferentes ajustes de los modelos con-

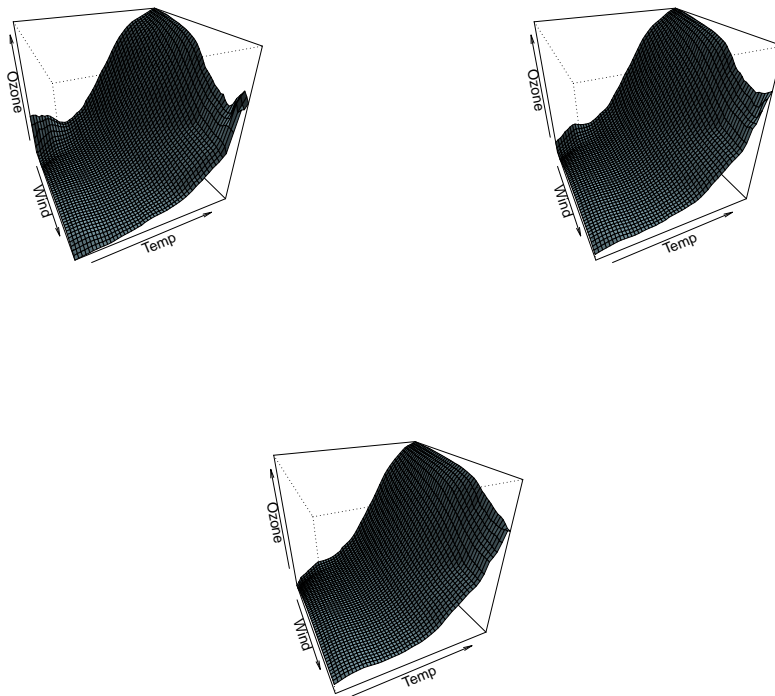


Figura 6.2: Estimador noparamétrico de la regresión usando diferentes ventanas.

siderados. Mas específicamente, calculamos

$$ECM(\hat{m}) = \frac{1}{n} \sum_{i=1}^n (O_i - \hat{m}(V_i, T_i))^2,$$

$$ECMed(\hat{m}) = \text{median}((O_i - \hat{m}(V_i, T_i))^2),$$

donde  $\hat{m}(V, T)$  corresponderá al estimador de la función de regresión en el punto  $(V, T)$  bajo el modelo calculado. En particular, comparamos nuestro procedimiento de estimación utilizando:

- un modelo puramente no paramétrico que notamos como (*NP*), para las variables regresoras velocidad del viento y temperatura que estimamos utilizando estimadores robustos.
- un modelo no lineal para la variable velocidad del viento estimado a partir del procedimiento robusto presentado en Spano (2016) que denotamos (*NLR*).
- un modelo parcialmente no lineal estimado (*PNLR*) a partir del procedimiento robusto el cual presentamos en (3.3).

	$\hat{m}_{NP}$	$\hat{m}_{NLR}$	$\hat{m}_{PNLR}$
<i>ECM</i>	376.6914	903.6621	275.6815
<i>ECMed</i>	70.6501	126.3261	53.41841

Cuadro 6.1: *ECM* y *ECMed* para los diferentes modelos considerados.

Tanto para el cálculo del estimador *NP* como de nuestro procedimiento *PNLR* que requieren de la selección de un parámetro de suavizado se han computado los estimadores para diferentes valores de ventanas y se ha elegido aquella con menor *ECM*. En la Figura 6.3 observamos el resultado del *ECM* y el *ECMed* de nuestra propuesta de estimación para diferentes valores de ventana. Bajo el modelo *PNLR*, los valores estimados para  $\beta_1$  y  $\beta_2$  corresponden a 186.8735 y  $-0.3524$ , respectivamente.

Como se observa en la Tabla 6.1, modelar la variable ozono a partir de la variable velocidad del viento y temperatura conduce a un mejor ajuste. Esto se ve plasmado al comparar el *ECMed* obtenido bajo un modelo puramente no lineal frente al *ECMed* de los estimadores obtenidos en el modelo parcialmente no lineal, cabe destacar que corresponde aquí comparar el *ECMed* en lugar del *ECM* ya que debido a la presencia de outliers esta última no representa una medida adecuada. Por otro lado en la Figura 6.3 observamos que aún los valores más altos de *ECMed* obtenidos para diferentes ventanas se conservan por debajo del 126.3261 que corresponde al *ECMed* del estimador bajo el modelo no lineal. Finalmente si comparamos la estimación bajo el modelo parcialmente no lineal frente al estimador bajo un modelo puramente no paramétrico, notamos una reducción del *ECMed*. Esto pone de manifiesto que el modelo propuesto parece adecuarse a la variable a ser explicada.

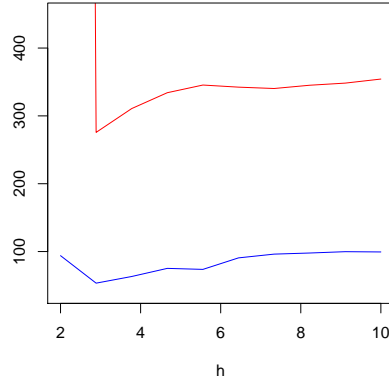


Figura 6.3:  $ECM$  en rojo y  $ECMed$  en azul del estimador  $\hat{m}_{PNLR}$  calculados para diferentes anchos de banda.

Para finalizar el análisis, utilizaremos los estimadores robustos calculados para detectar datos atípicos en la muestra. En la Figura 6.5 se resume con dos gráficos los residuos estandarizados basados obtenidos. En base a los mismos podemos detectar 5 outliers, que corresponden a las observaciones número 30, 62, 86, 117 y 126 de la muestra original que se muestran en la Figura 6.4

Las observaciones 86 y 126 también fueron detectadas como atípicas en Spano (2016); una de ellas corresponde a un nivel muy bajo de velocidad del viento asociada a un valor relativamente bajo de ozono. Dos de las observaciones detectadas se corresponden a altos valores de ozono pero bajas velocidades de viento. Así mismo en todos los casos las temperaturas correspondientes a los outliers encontrados conservan valores medios altos.

Por último eliminamos las 5 observaciones detectadas como atípicas y calculamos el estimador clásico obteniendo un  $ECM = 125.4424$  y  $ECMed = 49.204$  y los parámetros de  $\beta$  fueron estimados con  $\hat{\beta}_1 = 112.2661$  y  $\hat{\beta}_2 = -0.3175$ . Intentando una comparación más clara de ambos ajustes, es decir el ajuste clásico sin outliers frente al ajuste robusto con la muestra completa, calculamos para cada ajuste residuos parciales de la componente no paramétrica y de la paramétrica. Es decir, obtuvimos  $res_i = O_i - \underline{g}(\hat{\beta}, V_i)$  y  $\tilde{res}_i = O_i - \hat{\eta}(T_i)$  para cada observación  $i$  de la muestra, donde los estimadores  $\hat{\beta}$  y  $\hat{\eta}$  fueron calculados de acuerdo al procedimiento clásico y robusto según correspondía. La Figura 6.6, resume el ajuste de la componente no paramétrica  $\eta$  y la componente paramétrica  $g(\cdot, \beta)$ . Como se puede observar en la figura, el ajuste robusto frente a la presencia de outliers tiene un comportamiento similar al ajuste clásico obtenido en la muestra sin datos atípicos.

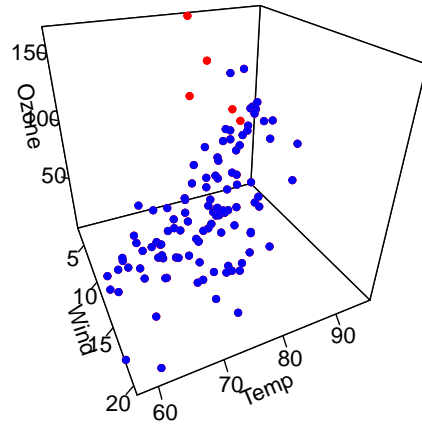


Figura 6.4: Conjunto de datos correspondientes al ozono, la temperatura y la velocidad del viento, en rojo se pueden apreciar los puntos considerados atípicos.

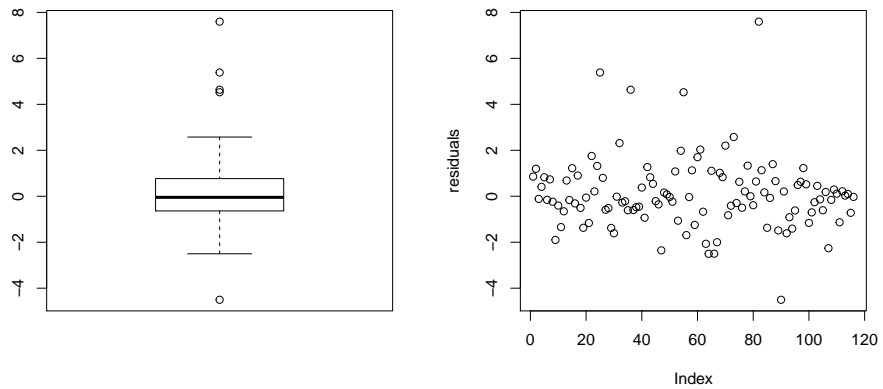


Figura 6.5: Gráficos de residuos estandarizados versus el índice y boxplot de los residuos estandarizados.

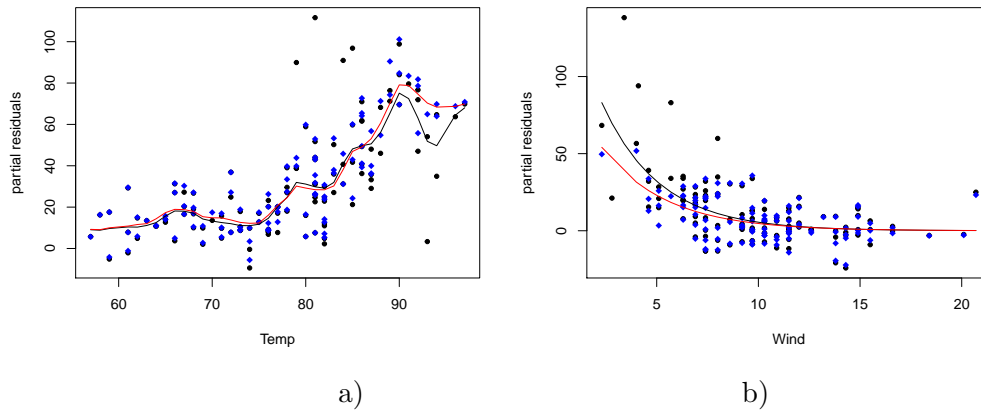


Figura 6.6: a) temperatura versus  $res_i$  (puntos negros robusto y diamantes azules clásico), la línea negra corresponde al ajuste robusto de  $\eta$  y en rojo el ajuste clásico de  $\eta$ . b) velocidad del viento versus  $\tilde{res}_i$  (puntos negros robusto y diamantes azules clásico), la línea negra corresponde al ajuste robusto de  $g(\cdot, \beta)$  y en rojo el ajuste clásico.



# Capítulo 7

## Apéndice

### 7.1. Demostraciones de la Sección 4.2

Comenzaremos esta sección demostrando el Lema 4.2.2 para ello utilizaremos la siguiente desigualdad.

**Lema 7.1.1.** *Desigualdad de Bernstein*

Sean  $X_1, \dots, X_n$  variables aleatorias independientes con esperanza 0 y varianza finita, si para todo  $1 \leq i \leq n$  se verifica que  $|X_i| \leq M$  casi seguramente luego

$$\mathbb{P} \left( \left| \sum_{i=1}^n X_i \right| > \epsilon \right) \leq 2 \exp \left( - \frac{3\epsilon^2}{2(3V + M\epsilon)} \right)$$

donde  $V > \sum_{i=1}^n \text{Var}(X_i)$ .

**Demostración del Lema 4.2.2:** Sea  $\tau \in T_0$ , llamemos  $w_i(\tau) = K \left( \frac{\tau - t_i}{h_n} \right) - \mathbb{E} \left( K \left( \frac{\tau - t_i}{h_n} \right) \right)$  y  $S_n(\tau) = \sum_{i=1}^n w_i(\tau)$ . Como  $K$  es una función acotada entonces existe  $\mathcal{A}_1 \in \mathbb{R}$  tal que para todo  $1 \leq i \leq n$  se cumple que  $|w_i| \leq \mathcal{A}_1$ , además por **C1** y **C2** tenemos que  $\mathbb{E}(w_i) = 0$  y  $\text{Var}(w_i) \leq \|K\|_\infty \|f_t\|_\infty h_n$  para todo  $1 \leq i \leq n$ . Luego usando **C7** y la desigualdad de Bernstein (7.1.1) tenemos que dado  $\epsilon > 0$  existen  $n_1 \in \mathbb{N}$  y alguna constante positiva  $\alpha \in \mathbb{R}$  tal que para todo  $n > n_1$  se verifica que  $\mathbb{P}(\frac{1}{nh_n} |S_n(\tau)| > \epsilon) \leq \exp(-\alpha nh_n)$  y como la cota no depende de  $\tau$  llegamos a

$$\sup_{\tau \in T_0} \mathbb{P} \left( \frac{1}{nh_n} |S_n(\tau)| > \epsilon \right) \leq 2 \exp(-\alpha nh_n). \quad (7.1)$$

Por otro lado como  $T_0$  es un conjunto compacto, podemos considerar una colección finita de bolas  $B_j = B_{h_n^\gamma}(\tau_j)$  con centros  $\tau_j \in T_0$  y radio  $h_n^\gamma$  con  $\gamma = 3$  tal que  $T_0 \subseteq \bigcup_{j=1}^l B_j$ . Luego

$$\sup_{\tau \in T_0} |S_n(\tau)| \leq \max_{1 \leq j \leq l} \sup_{\tau \in B_j} |S_n(\tau) - S_n(\tau_j)| + \max_{1 \leq j \leq l} |S_n(\tau_j)|.$$

Por **C2** la función  $K$  es Lipschitz de orden uno, llamemos  $\|K\|_L$  a su constante Lipschitz. Entonces, para cada  $\tau \in B_j$  tenemos que

$$\frac{1}{nh_n} |S_n(\tau) - S_n(\tau_j)| \leq \frac{1}{nh_n} \|K\|_L nh_n^{\gamma-1} = Ch_n^{\gamma-2}. \quad (7.2)$$

Luego, como  $h_n \rightarrow 0$  entonces existe  $n_2 \in \mathbb{N}$  tal que para todo  $n > n_2$  tenemos que

$$\max_{1 \leq j \leq l} \sup_{\tau \in B_j} \frac{1}{nh_n} |S_n(\tau) - S_n(\tau_j)| < \epsilon.$$

Finalmente por **C7**, (7.1) y (7.2) existe  $n_0 = \max\{n_1, n_2\}$  tal que para todo  $n > n_0$  se verifica

$$\begin{aligned} \mathbb{P} \left( \frac{1}{nh_n} \sup_{\tau \in T_0} |S_n(\tau)| > 2\epsilon \right) &\leq \mathbb{P} \left( \max_{1 \leq j \leq l} \frac{1}{nh_n} |S_n(\tau_j)| > \epsilon \right) \\ &\leq \sum_{j=1}^l \mathbb{P} \left( \frac{1}{nh_n} |S_n(\tau_j)| > \epsilon \right) \leq 2l \exp(-\alpha nh_n) \leq Cn^{-2} \end{aligned}$$

lo cual concluye la demostración.

**Demostración del Lema 4.2.3:** Veamos que, dado  $\epsilon > 0$ , existe  $a$  real tal que, para todo  $\tau \in T_0$  y para todo  $\beta \in \mathcal{B}$  se verifica que  $F_\beta(a|t = \tau) < \epsilon$ . Efectivamente, sean  $\tilde{\beta} \in \mathcal{B}$  y  $\tilde{\tau} \in T_0$  fijos entonces existe  $a(\tilde{\beta}, \tilde{\tau}) \in \mathbb{R}$  tal que  $F_{\tilde{\beta}}(a(\tilde{\beta}, \tilde{\tau})|t = \tilde{\tau}) < \epsilon$ . Como por **C4.3**,  $F_\beta(v|t = \tau)$  es una función continua de  $(\beta, \tau)$ , también existe  $\delta = \delta(\tilde{\beta}, \tilde{\tau}) > 0$  tal que si  $|\beta - \tilde{\beta}| < \delta$  y  $|\tau - \tilde{\tau}| < \delta$  entonces  $F_\beta(a(\tilde{\beta}, \tilde{\tau})|t = \tau) < \epsilon$ . Trivialmente  $T_0 \times \mathcal{B} \subseteq \bigcup B_{\delta(\beta, \tau)}(\beta, \tau)$ , luego, como  $T_0 \times \mathcal{B}$  es un conjunto compacto existen  $(\beta_1, \tau_1), \dots, (\beta_m, \tau_m) \in T_0 \times \mathcal{B}$  tales que  $T_0 \times \mathcal{B} \subseteq \bigcup_{i=1}^m B_{\delta(\beta_i, \tau_i)}(\beta_i, \tau_i)$ . Sean  $a(\beta_1, \tau_1), \dots, a(\beta_m, \tau_m) \in \mathbb{R}$  tales que  $F_{\beta_i}(a(\beta_i, \tau_i)|t = \tau_i) < \epsilon$  para  $1 \leq i \leq m$ . Definamos  $a = \min_{1 \leq i \leq m} (a(\beta_i, \tau_i))$ . Por último notar que dado  $(\beta, \tau)$  existe  $1 \leq i \leq m$  tal que  $(\beta, \tau) \subseteq B_{\delta(\beta_i, \tau_i)}(\beta_i, \tau_i)$  de donde se deduce que  $F_\beta(a|t = \tau) \leq F_{\beta_i}(a(\beta_i, \tau_i)|t = \tau) < \epsilon$ .

La otra desigualdad se demuestra análogamente.

**Demostración del Lema 4.2.4:** Para la continuidad de  $\lambda$  veamos que dada  $(\beta_n, \tau_n, a_n, \sigma_n) \rightarrow (\beta, \tau, a, \sigma)$  se cumple que  $\lambda(\beta_n, \tau_n, a_n, \sigma_n) \rightarrow \lambda(\beta, \tau, a, \sigma)$  cuando  $n \rightarrow \infty$ . Para ello bastará ver que

- a)  $\int \Psi \left( \frac{v-a}{\sigma} \right) dF_{\beta_n}(v|t = \tau_n) \rightarrow \int \Psi \left( \frac{v-a}{\sigma} \right) dF_\beta(v|t = \tau)$ .
- b)  $\int \left[ \Psi \left( \frac{v-a_n}{\sigma_n} \right) - \Psi \left( \frac{v-a}{\sigma} \right) \right] dF_{\beta_n}(v|t = \tau_n) \rightarrow 0$ .

Para el punto **a)** basta con notar que como valen **C4.3** y **C5** estamos bajo las condiciones del Lema de Portmanteau.

Veamos el punto **b**). Para ello tomemos  $\epsilon > 0$ , el Lema 4.2.3 implica que existe  $n_0 \in \mathbb{N}$  y constantes  $a$  y  $b$  tales que  $F_{\beta_n}(b|t = \tau_n) > 1 - \epsilon/4\|\Psi\|_\infty$  y  $F_{\beta_n}(a|t = \tau_n) < \epsilon/4\|\Psi\|_\infty$  para todo  $n > n_0$ . Por otro lado, por **C5** sabemos que existe  $n_1 \in \mathbb{N}$  tal que  $|\Psi\left(\frac{v-a_n}{\sigma_n}\right) - \Psi\left(\frac{v-a}{\sigma}\right)| < \epsilon/2$  para todo  $n > n_1$  y para todo  $v$  en  $[a, b]$ . Luego, tomando  $n = \max\{n_0, n_1\}$  llegamos a

$$\begin{aligned} \int \left[ \Psi\left(\frac{v-a_n}{\sigma_n}\right) - \Psi\left(\frac{v-a}{\sigma}\right) \right] dF_{\beta_n}(v|t = \tau_n) &\leq (\epsilon/2) \int_{[a,b]} dF_{\beta_n}(v|t = \tau_n) \\ &\quad + 2\|\Psi\|_\infty \int_{[a,b]^c} dF_{\beta_n}(v|t = \tau_n) \\ &\leq \epsilon. \end{aligned}$$

concluyendo así lo que queríamos probar. Por otro lado, para ver la continuidad de  $\eta_\beta(\tau)$  observemos que como  $\Psi$  es estrictamente creciente tenemos que para todo  $\zeta > 0$

$$\lambda(\beta, \tau, \eta_\beta(\tau) - \zeta, \sigma) > \lambda(\beta, \tau, \eta_\beta(\tau), \sigma) > \lambda(\beta, \tau, \eta_\beta(\tau) + \zeta, \sigma).$$

Debido a la continuidad de  $\lambda$  sabemos que  $\forall \epsilon > 0, \exists \delta > 0$  tal que  $|(\beta_1, \tau_1) - (\beta, \tau)| < \delta$  entonces

$$|\lambda(\beta_1, \tau_1, \eta_\beta(\tau) \pm \zeta, \sigma) - \lambda(\beta, \tau, \eta_\beta(\tau) \pm \zeta, \sigma)| < \epsilon \quad (7.3)$$

Sea  $\epsilon = \min\{\lambda(\beta, \tau, \eta_\beta(\tau), \sigma) - \lambda(\beta, \tau, \eta_\beta(\tau) + \zeta, \sigma), \lambda(\beta, \tau, \eta_\beta(\tau) - \zeta, \sigma) - \lambda(\beta, \tau, \eta_\beta(\tau), \sigma)\}$  entonces (7.3) implica que existe  $\delta_1 > 0$  tal que si  $|(\beta_1, \tau_1) - (\beta, \tau)| < \delta_1$  se tiene que

$$\lambda(\beta_1, \tau_1, \eta_\beta(\tau) + \zeta, \sigma) < \lambda(\beta, \tau, \eta_\beta(\tau) + \zeta, \sigma) + \epsilon < \lambda(\beta_1, \tau_1, \eta_{\beta_1}(\tau_1), \sigma)$$

y, por lo tanto,  $\eta_{\beta_1}(\tau_1) - \eta_\beta(\tau) < \zeta$ . Análogamente se ve que existe  $\delta_2 > 0$  tal que si  $|(\beta_1, \tau_1) - (\beta, \tau)| < \delta_2$  entonces  $-\zeta < \eta_{\beta_1}(\tau_1) - \eta_\beta(\tau)$ . Por lo tanto, dado  $\zeta > 0$  si  $\delta = \min\{\delta_1, \delta_2\}$  se tiene que si

$$|(\beta_1, \tau_1) - (\beta, \tau)| < \delta \Rightarrow |\eta_\beta(\tau) - \eta_{\beta_1}(\tau_1)| < \zeta.$$

**Demostración del Lema 4.2.5:** Por Lema 4.2.3 existen constantes  $a, b$  tales que  $F_\beta(b|t = \tau) > \frac{7}{8}$  y  $F_\beta(a|t = \tau) < \frac{1}{8}$  para todo  $\tau \in T_0$  y para todo  $\beta \in \mathcal{B}$ . Por (4.7) sabemos que, dado  $\epsilon > 0$  existe  $n_0 \in \mathbb{N}$  tal que  $-\epsilon < F_{n,\beta}(v|t = \tau) - F_\beta(v|t = \tau) < \epsilon$  para todo  $n > n_0$ . En particular tomando  $\epsilon = \frac{1}{8}$ ,  $v = a$  y  $v = b$  llegamos a  $F_{n,\beta}(a|t = \tau) < \frac{1}{4}$  y  $F_{n,\beta}(b|t = \tau) > \frac{3}{4}$  para todo  $n > n_0$ , para todo  $\tau \in T_0$  y para todo  $\beta \in \mathcal{B}$ , luego se deduce que  $a < m_{n,\beta}(\tau) < b$ , donde  $m_{n,\beta}(\tau)$  es la mediana respecto a la distribución  $F_{n,\beta}(\cdot|t = \tau)$ . Es sencillo ver que la elección de  $a, b$  y  $n_0$  implican además que  $s_{n,\beta}(\tau) \leq b - a$  para todo  $n > n_0$ , para todo  $\tau \in T_0$  y para todo  $\beta \in \mathcal{B}$ . Como vale **C4.2** tenemos que para cada  $\tau$  y  $\beta$  fijos  $F_\beta(v|t = \tau)$  es una función de distribución continua, luego dados  $\beta_0 \in \mathcal{B}$  y  $\tau_0 \in T_0$  existen constantes  $a(\beta_0, \tau_0)$  y  $b(\beta_0, \tau_0)$  tales que  $F_{\beta_0}(a(\beta_0, \tau_0)|t = \tau_0) = \frac{1}{3}$  y  $F_{\beta_0}(b(\beta_0, \tau_0)|t = \tau_0) = \frac{7}{10}$ . Por **C4.3** sabemos que existe  $\delta_0 = \delta(\beta_0, \tau_0)$  tal que si  $|\beta - \beta_0| < \delta$  y  $|\tau - \tau_0| < \delta$  entonces

$$\frac{1}{3} - \epsilon < F_\beta(a(\beta_0, \tau_0)|t = \tau) < \frac{1}{3} + \epsilon \quad y \quad \frac{7}{10} - \epsilon < F_\beta(b(\beta_0, \tau_0)|t = \tau) < \frac{7}{10} + \epsilon.$$

Sea  $B_h(c)$  la bola de centro  $c$  y radio  $h$ . Trivialmente  $T_0 \times \mathcal{B} \subseteq \bigcup B_{\delta(\beta, \tau)}(\beta, \tau)$ , luego como  $T_0 \times \mathcal{B}$  en un conjunto compacto existen  $(\beta_1, \tau_1), \dots, (\beta_m, \tau_m) \in T_0 \times \mathcal{B}$  tales que  $T_0 \times \mathcal{B} \subseteq \bigcup_{i=1}^m B_{\delta(\beta_i, \tau_i)}(\beta_i, \tau_i)$ . Para cada  $(\beta_i, \tau_i)$ , con  $1 \leq i \leq m$ , existen  $a(\beta_i, \tau_i)$  y  $b(\beta_i, \tau_i)$  tales que  $F_{\beta_i}(a(\beta_i, \tau_i)|t = \tau_i) = \frac{1}{3}$  y  $F_{\beta_i}(b(\beta_i, \tau_i)|t = \tau_i) = \frac{7}{10}$ , por lo tanto, para todo  $(\beta, \tau) \in B_{\delta(\beta_i, \tau_i)}(\beta_i, \tau_i)$  tenemos que

$$\frac{1}{3} - \epsilon < F_{\beta}(a(\beta_i, \tau_i)|t = \tau) < \frac{1}{3} + \epsilon \quad y \quad \frac{7}{10} - \epsilon < F_{\beta}(b(\beta_i, \tau_i)|t = \tau) < \frac{7}{10} + \epsilon.$$

Por C4.2 existe  $\eta > 0$  tal que

$$\frac{1}{3} - 2\epsilon < F_{\beta}(a(\beta_i, \tau_i) - \eta|t = \tau) < \frac{1}{3} + 2\epsilon \quad y \quad \frac{7}{10} - 2\epsilon < F_{\beta}(b(\beta_i, \tau_i) + \eta|t = \tau) < \frac{7}{10} + 2\epsilon,$$

y por (4.7) tenemos que para todo  $(\beta, \tau) \in B_{\delta(\beta_i, \tau_i)}(\beta_i, \tau_i)$ , para todo  $n > n_0$  y  $1 \leq i \leq m$

$$\frac{1}{3} - 3\epsilon < F_{n, \beta}(a(\beta_i, \tau_i) - \eta|t = \tau) < \frac{1}{3} + 3\epsilon \quad y \quad \frac{7}{10} - 3\epsilon < F_{n, \beta}(b(\beta_i, \tau_i) + \eta|t = \tau) < \frac{7}{10} + 3\epsilon.$$

Finalmente, tomando  $\epsilon = \frac{1}{60}$  obtenemos que para todo  $n > n_0$ ,  $F_{n, \beta}(a(\beta_i, t_i)|t = \tau) < \frac{1}{2}$ ,  $F_{n, \beta}(a(\beta_i, \tau_i) - \eta|t = \tau) > \frac{1}{4}$ ,  $F_{n, \beta}(b(\beta_i, \tau_i)|t = \tau) > \frac{1}{2}$  y  $F_{n, \beta}(b(\beta_i, \tau_i) + \eta|t = \tau) < \frac{3}{4}$  para todo  $n > n_0$  y para todo  $1 \leq i \leq m$ . Por lo tanto  $a(\beta_i, \tau_i) < m_{n, \beta}(\tau) < b(\beta_i, \tau_i)$  para todo  $(\beta, \tau) \in B_{\delta_i}(\beta_i, \tau_i)$  y  $s_{n, \beta}(\tau) > \eta$  para todo  $\tau \in T_0$  y para todo  $\beta \in \mathcal{B}$  si  $n > n_0$  concluyendo así la demostración.

**Demostración del Teorema 4.2.6:** Notemos que si consideramos para cada  $v \in \mathbb{R}$  fijo  $z_{i, \beta} = I_{(-\infty, v]}(v_{i, \beta})$  y  $K$  la función núcleo, entonces  $\widehat{F}_{\beta}(\cdot|t = \tau)$  puede escribirse como el cociente entre  $R_{1n}(\beta, \tau)$  y  $R_{0n}(\tau)$  donde

$$R_{1n}(\beta, \tau) = \frac{1}{nh_n} \sum_{i=1}^n z_{i, \beta} K\left(\frac{\tau - t_i}{h_n}\right) \quad y \quad R_{0n}(\tau) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{\tau - t_i}{h_n}\right),$$

mediante algunos sencillos manejos algebraicos, llegamos a la siguiente desigualdad:

$$\begin{aligned} |\widehat{F}_{\beta}(v|t = \tau) - F_{\beta}(v|t = \tau)| &\leq \left[ |R_{1n}(\beta, \tau) - \mathbb{E}(R_{1n}(\beta, \tau))| \right. \\ &\quad + |\mathbb{E}(R_{1n}(\beta, \tau)) - F_{\beta}(v|t = \tau)\mathbb{E}(R_{0n}(\tau))| \\ &\quad \left. + |F_{\beta}(v|t = \tau)| |\mathbb{E}(R_{0n}(\tau)) - R_{0n}(\tau)| \right] \frac{1}{R_{0n}(\tau)} \end{aligned}$$

por lo tanto, bastará probar que:

$$\sup_{v \in \mathbb{R}} \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |R_{1n}(\beta, \tau) - \mathbb{E}(R_{1n}(\beta, \tau))| \rightarrow 0 \quad \text{ctp} \quad (7.4)$$

$$\sup_{\tau \in T_0} |\mathbb{E}(R_{0n}(\tau)) - R_{0n}(\tau)| \rightarrow 0 \quad \text{ctp} \quad (7.5)$$

$$\sup_{v \in \mathbb{R}} \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |\mathbb{E}(R_{1n}(\beta, \tau)) - F_\beta(v|\tau)\mathbb{E}(R_{0n}(\tau))| \rightarrow 0 \quad \text{ctp} \quad (7.6)$$

$$\text{Para algún } a > 0 \text{ se verifica que } \sum_{n \geq 1} \mathbb{P} \left( \inf_{\tau \in T_0} R_{0n}(\tau) \leq a \right) < \infty \quad (7.7)$$

Empecemos probando (7.7). Para ello observemos que **C1** y **C2** implican que

$$\mathbb{E}(R_{0n}(\tau)) > \inf_{\tau \in T_0} f_t(\tau) = I > 0.$$

Por otro lado se puede ver que

$$\inf_{\tau \in T_0} R_{0n}(\tau) > I - \sup_{\tau \in T_0} |\mathbb{E}(R_{0n}(\tau)) - R_{0n}(\tau)|.$$

Entonces, tomando  $a = \frac{I}{2}$  llegamos a que

$$\mathbb{P} \left( \inf_{\tau \in T_0} R_{0n}(\tau) \leq a \right) \leq \mathbb{P} \left( \sup_{\tau \in T_0} |\mathbb{E}(R_{0n}(\tau)) - R_{0n}(\tau)| \geq a \right)$$

por último, usando el Lema 4.2.2 se concluye lo que queríamos ver. La prueba de (7.5) también resulta una consecuencia inmediata del Lema 4.2.2 junto con el lema de Borel-Cantelli.

Para probar (7.4) veamos que estamos bajo las condiciones del Lema 4.1.19 . Para ello, para cada  $n \in \mathbb{N}$ , definamos la siguiente familia de funciones

$$\mathcal{F}_n = \left\{ f_{v,\beta,\tau,h_n}(y, \mathbf{x}, t) = \|K\|_\infty^{-1} I_{(-\infty, v]}(y - g(\mathbf{x}, \beta)) K \left( \frac{t - \tau}{h_n} \right), v \in \mathbb{R}, \beta \in \mathcal{B}, \tau \in T_0 \right\}.$$

Por un lado, como vale **C2**, siguiendo a Pollard (1984), Capítulo 2, Problema [27], junto a la Proposición 4.1.14 tenemos que para cada  $n \in \mathbb{N}$ , la familia  $\left\{ K \left( \frac{\cdot - \tau}{h_n} \right), \tau \in T_0 \right\}$  es Euclídea. Este hecho empalmado a la condición **F.E.** y a la Proposición 4.1.11 , implica que para todo  $0 < \epsilon < 1$  el número de cubrimiento de  $\mathcal{F}_n$  verifica que  $\sup_{\mathbb{Q}} N(\epsilon, \mathcal{F}_n, L^1(\mathbb{Q})) \leq A_1 \epsilon^{-W_1}$  con  $\mathbb{Q}$  medida de probabilidad y  $A_1$  y  $W_1$  constantes independientes de  $n$ .

Por otro lado para cada  $f_{v,\beta,\tau,h_n} \in \mathcal{F}_n$  se verifica que  $|f_{v,\beta,\tau,h_n}| \leq 1$  y, además

$$\mathbb{E}(f_{v,\beta,\tau,h_n}^2) \leq \frac{1}{\|K\|_\infty^2} \int K^2 \left( \frac{u - \tau}{h_n} \right) f_t(u) du \leq \frac{\|f_t\|_\infty}{\|K\|_\infty} h_n \int K(u) du, \quad (7.8)$$

por lo tanto, por (7.8), **C2**, **C6** y llamando  $C = \frac{\|f_t\|_\infty}{\|K\|_\infty}$ ,  $\delta(n) = Ch_n^{1/2}$  y  $\alpha(n) = \left( \frac{nh_n}{\log(n)} \right)^{-1/4}$  estamos bajo las hipótesis del Lema 4.1.19. Luego, para concluir, basta observar que

$$\begin{aligned} \|K\|_\infty \sup_{v \in \mathbb{R}} \sup_{\beta \in \mathcal{B}} \sup_{\tau \in T_0} |R_{1n}(\beta, \tau) - \mathbb{E}(R_{1n}(\beta, \tau))| &= \frac{1}{h_n} \sup_{f \in \mathcal{F}_n} |\mathbb{E}_{\mathbb{P}_n}(f) - \mathbb{E}_{\mathbb{P}}(f)| \\ &\leq \frac{1}{\delta_n^2 \alpha_n} \sup_{f \in \mathcal{F}_n} |\mathbb{E}_{\mathbb{P}_n}(f) - \mathbb{E}_{\mathbb{P}}(f)|. \end{aligned}$$

Finalmente demostremos (7.6) , para ello notemos que

$$\begin{aligned} \mathbb{E}(R_{1n}(\beta, \tau)) &= \frac{1}{nh_n} \sum_{i=1}^n \mathbb{E} \left( K \left( \frac{\tau - t_i}{h_n} \right) F_\beta(v|t = t_i) \right), \\ F_\beta(v|t = \tau) \mathbb{E}(R_{0n}(\tau)) &= \frac{1}{nh_n} \sum_{i=1}^n \mathbb{E} \left( K \left( \frac{\tau - t_i}{h_n} \right) F_\beta(v|t = \tau) \right). \end{aligned}$$

Luego

$$\begin{aligned} \left| \mathbb{E}(R_{1n}(\beta, \tau)) - F_\beta(v|t = \tau) \mathbb{E}(R_{0n}(\tau)) \right| &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \frac{1}{h_n} K \left( \frac{\tau - t_i}{h_n} \right) \left( F_\beta(v|t = t_i) - F_\beta(v|t = \tau) \right) \right] \\ &\leq \int \frac{1}{h_n} K \left( \frac{\tau - u}{h_n} \right) |F_\beta(v|t = u) - F_\beta(v|t = \tau)| f_t(u) du \\ &\leq A_1(\beta, \tau) + A_2(\beta, \tau) \end{aligned}$$

Donde,

$$\begin{aligned} A_1(\beta, \tau) &= \int_{T_0^c} \frac{1}{h_n} K \left( \frac{\tau - u}{h_n} \right) |F_\beta(v|t = u) - F_\beta(v|t = \tau)| f_t(u) du, \\ A_2(\beta, \tau) &= \int_{\overline{T_0^c}} \frac{1}{h_n} K \left( \frac{\tau - u}{h_n} \right) |F_\beta(v|t = u) - F_\beta(v|t = \tau)| f_t(u) du, \end{aligned}$$

siendo  $T_0^c$  el entorno de  $T_0$  establecido en C4.1 y  $\overline{T_0^c}$  su complemento, es decir,  $T_0^c \subset \mathbb{R}$  es un conjunto abierto tal que  $T_0 \subset T_0^c$  y  $d(\overline{T_0^c}, T_0) > \epsilon$  siendo  $d$  la distancia entre conjuntos.

Acotemos primero  $A_2(\beta, \tau)$ . Como  $u \in \overline{T_0^c}$  y  $h_n$  converge a cero entonces existe  $n_0$  natural tal que  $\frac{|\tau - u|}{h_n} \notin [0, 1]$  para todo  $n > n_0$ , luego por C2 tenemos que  $K \left( \frac{\tau - u}{h_n} \right) = 0$  y por lo tanto  $A_2(\beta, \tau) = 0$  para todo  $n > n_0$ , para todo  $\beta \in \mathcal{B}$  y para todo  $\tau \in T_0$ .

Veamos como acotar  $A_1(\beta, \tau)$ . Por C4.1 sabemos que dado  $\eta > 0$  existe  $\delta > 0$  tal que si  $u \in T_0^c$  y  $|\tau - u| < \delta$  entonces  $\sup_{\beta \in \mathcal{B}} |F_\beta(v|t = u) - F_\beta(v|t = \tau)| < \eta$ .

Por otro lado, como  $h_n \rightarrow 0$  sabemos que existe  $n_1 \in \mathbb{N}$  tal que  $h_n < \delta$  para todo  $n > n_1$ , luego, como la función  $K$  tiene soporte compacto en  $[-1, 1]$  tenemos que  $K \left( \frac{\tau - u}{h_n} \right) = 0$  para todo  $u \in \overline{B_{h_n}(\tau)}$  y para todo  $n > n_1$ . Por lo tanto

$$A_1(\beta, \tau) = \int_{T_0^c \cap B_{h_n}(\tau)} \frac{1}{h_n} K \left( \frac{\tau - u}{h_n} \right) |F_\beta(v|t = u) - F_\beta(v|t = \tau)| f_t(u) du$$

Además, podemos tomar  $n_2 \in \mathbb{N}$  tal que  $h_n < \epsilon$  para todo  $n > n_2$  resultando  $T_0^\epsilon \cap B_{h_n}(\tau) = B_{h_n}(\tau)$  si  $n > n_2$ . Por último tomando  $n = \max\{n_1, n_2\}$  obtenemos que

$$A_1(\boldsymbol{\beta}, \tau) \leq \frac{1}{h_n} \|K\|_\infty \int_{B_{h_n}(\tau)} |F_{\boldsymbol{\beta}}(v|t = u) - F_{\boldsymbol{\beta}}(v|t = \tau)| f_T(u) du \leq 2\eta \|f_t\|_\infty \|K\|_\infty.$$

Por lo tanto deducimos que para cada  $v \in \mathbb{R}$  se cumple que

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}} \sup_{\tau \in T_0} |\widehat{F}_{\boldsymbol{\beta}}(v|t = \tau) - F_{\boldsymbol{\beta}}(v|t = \tau)| \rightarrow 0 \quad \text{ctp.} \quad (7.9)$$

Finalmente, para cada  $q \in \mathbb{Q}$  definimos  $\mathcal{N}_q = \{\omega \in \Omega : \sup_{\boldsymbol{\beta} \in \mathcal{K}} \sup_{\tau \in T_0} |\widehat{F}_{\boldsymbol{\beta}}(q|t = \tau) - F_{\boldsymbol{\beta}}(q|t = \tau)| \not\rightarrow 0\}$  y  $\mathcal{N} = \bigcup_{q \in \mathbb{Q}} \mathcal{N}_q$ . Luego (7.9) implica que  $\mathbb{P}(\mathcal{N}) = 0$ . Sea  $\omega \notin \mathcal{N}$ , entonces  $\sup_{\boldsymbol{\beta} \in \mathcal{B}} \sup_{\tau \in T_0} |\widehat{F}_{\boldsymbol{\beta}}(q|t = \tau) - F_{\boldsymbol{\beta}}(q|t = \tau)| \rightarrow 0$  para todo  $q \in \mathbb{Q}$ . Dado  $\epsilon > 0$ , por el Lema 4.2.3 sabemos que existen  $a, b \in \mathbb{Q}$  tales que  $F_{\boldsymbol{\beta}}(a|t = \tau) < \epsilon$  y  $F_{\boldsymbol{\beta}}(b|t = \tau) > 1 - \epsilon$  para todo  $\boldsymbol{\beta} \in \mathcal{B}$  y  $\tau \in T_0$ . Por otro lado, la equicontinuidad pedida en C4.2 implica que existen  $a = v_1 < v_2 < \dots < v_l = b$ , con  $v_i \in \mathbb{Q}$  tal que  $|v_i - v_{i-1}| < \delta$  y  $|F_{\boldsymbol{\beta}}(v|t = \tau) - F_{\boldsymbol{\beta}}(v_i|t = \tau)| < \epsilon$  para todo  $\boldsymbol{\beta} \in \mathcal{B}$  y  $\tau \in T_0$ , y para todo  $v$  tal que  $|v - v_i| < \delta$ . Sea  $n_0 \in \mathbb{N}$  tal que para todo  $n > n_0$ ,  $\max_{1 \leq i \leq l} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \sup_{\tau \in T_0} |\widehat{F}_{\boldsymbol{\beta}}(v_i|t = \tau) - F_{\boldsymbol{\beta}}(v_i|t = \tau)| < \epsilon$ . Entonces es sencillo ver que  $\sup_{v \in \mathbb{R}} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \sup_{\tau \in T_0} |\widehat{F}_{\boldsymbol{\beta}}(v|t = \tau) - F_{\boldsymbol{\beta}}(v|t = \tau)| < 2\epsilon$  para todo  $n > n_0$ , lo cual concluye la demostración.

## 7.2. Demostraciones de la Sección 4.3

**Demostración del Teorema 4.3.5:** Para concluir con la demostración del teorema falta ver que  $\sqrt{n}[L_n(s_n, \boldsymbol{\beta}_0) - L_n(\sigma_0, \boldsymbol{\beta}_0)] \xrightarrow{P} 0$ .

Para ello llamemos  $z_j(\mathbf{s}, \tau, \boldsymbol{\beta})$  a la  $j$ -ésima coordenada de  $\mathbf{z}(\mathbf{s}, \tau, \boldsymbol{\beta})$  para cada  $1 \leq j \leq q$  y definamos

$$J_n(\sigma) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \eta_0(t_i)}{\sigma} \right) - \Psi_1 \left( \frac{y_i - g(\mathbf{x}_i, \boldsymbol{\beta}_0) - \eta_0(t_i)}{\sigma_0} \right) \right] W(\mathbf{x}_i) z_j(\mathbf{x}_i, t_i, \boldsymbol{\beta}_0),$$

luego bastará probar que para cualquier  $\epsilon > 0$ ,  $\mathbb{P}(|J_n(s_n)| > \epsilon) \rightarrow 0$  para cada  $1 \leq j \leq q$ .

Sea  $\mathcal{P} = (\frac{\sigma_0}{2}, 2\sigma_0)$  y definamos la siguiente familia de funciones

$$\mathcal{F}^j = \left\{ f_\sigma^j(y, \mathbf{x}, t) = \left[ \Psi_1 \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma} \right) - \Psi_1 \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma_0} \right) \right] W(\mathbf{x}) z_j(\mathbf{x}, t, \boldsymbol{\beta}_0), \sigma \in \mathcal{P} \right\}$$

Usando el Teorema de Lagrange para  $\Psi_1$  en la variable  $\sigma$  y N2 obtenemos que

$$\begin{aligned} |f_{\sigma_s}^j(y, \mathbf{x}, t) - f_\sigma^j(y, \mathbf{x}, t)| &\leq \sqrt{W(\mathbf{x})} |z_j(\mathbf{x}, t, \boldsymbol{\beta}_0)| \left| \frac{1}{\sigma^*} \Psi_1' \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma^*} \right) \left( \frac{y - g(\mathbf{x}, \boldsymbol{\beta}_0) - \eta_0(t)}{\sigma^*} \right) \right| |\sigma_s - \sigma| \\ &\leq \frac{2}{\sigma_0} \|\sqrt{W}\|_\infty \|\Psi_2\|_\infty \sqrt{W(\mathbf{x})} |z_j(\mathbf{x}, t, \boldsymbol{\beta}_0)| |\sigma_s - \sigma|. \end{aligned}$$

donde  $\sigma^*$  es un punto intermedio entre  $\sigma$  y  $\sigma_s$ .

Luego llamando  $A = \frac{2}{\sigma_0} \|\sqrt{W}\|_\infty \|\Psi_2\|_\infty$  y  $S(\mathbf{x}, t) = A\sqrt{W(\mathbf{x})}z_j(\mathbf{x}, t, \boldsymbol{\beta}_0)$  tenemos que, por un lado, N7 implica que la función  $S$  es una envolvente en  $L^2$  para la familia  $\mathcal{F}^j$  y además se verifica que

$$|f_{\sigma_s}^j - f_\sigma^j| \leq S(\mathbf{x}, t)|\sigma_s - \sigma| \quad (7.10)$$

Por otro lado, como vale N10 tenemos que  $\mathbb{E}(f(y, \mathbf{x}, t)) = 0$  para todo  $f \in \mathcal{F}^j$ .

Asimismo, debido a que vale (7.10), por el Teorema 4.1.9, tenemos que la familia  $\mathcal{F}^j$  verifica la siguiente la condición de bracketing

$$N_{[\cdot]}(2\epsilon\|S\|_{\mathbb{P},2}, \mathcal{F}^j, L^2(\mathbb{P})) \leq N(\epsilon, \mathcal{P}, |\cdot|)$$

y, por lo tanto, por la proposición (4.1.17) se tiene que la familia  $\mathcal{F}^j$  satisface que

$$J_{[\cdot]}(\delta, \mathcal{F}^j) \xrightarrow{\delta \rightarrow 0} 0. \quad (7.11)$$

Por consiguiente, del hecho que  $s_n \xrightarrow{P} \sigma_0$  tenemos que dado  $\delta > 0$  existe  $n_0 \in \mathbb{N}$  tal que para todo  $n > n_0$  se cumple

$$\mathbb{P}(|\sigma_0 - s_n| > \delta) < \frac{\delta}{2}.$$

Por lo tanto, dado  $\epsilon > 0$  tenemos que

$$\begin{aligned} \mathbb{P}(|J_n^j(s_n)| > \epsilon) &\leq \mathbb{P}(|J_n^j(s_n)| > \epsilon, |\sigma_0 - s_n| < \delta) + \mathbb{P}(|\sigma_0 - s_n| > \delta) \\ &\leq \mathbb{P}\left(\sup_{|\sigma_0 - \sigma| < \delta} |J_n^j(\sigma)| > \epsilon\right) + \delta \\ &\leq \frac{1}{\epsilon} \mathbb{E}\left(\sup_{|\sigma_0 - \sigma| < \delta} |J_n^j(\sigma)|\right) + \delta. \end{aligned}$$

Luego es suficiente ver que

$$\lim_{n \rightarrow \infty} \lim_{\delta \rightarrow 0} \mathbb{E}\left(\sup_{|\sigma_0 - \sigma| < \delta} |J_n^j(\sigma)|\right) = 0.$$

Dado  $\delta > 0$ , consideremos la subclase

$$\mathcal{F}_\delta^j = \{f_\sigma^j(y, \mathbf{x}, t) \in \mathcal{F}^j : |\sigma_0 - \sigma| < \delta\}$$

Es sencillo ver que la subclase  $\mathcal{F}_\delta^j$  cumple que  $\|f_\sigma^j\|_{2,\mathbb{P}} \leq \delta\|S\|_{2,\mathbb{P}}$  para toda  $f_\sigma^j \in \mathcal{F}_\delta^j$  donde la función envolvente  $S(\mathbf{x}, t)$  fue definida anteriormente. Luego, de la desigualdad maximal enunciada en la Proposición 4.1.18, tenemos que



$$\begin{aligned} \mathbb{E} \left( \sup_{|\sigma_0 - \sigma| < \delta} |J_n^j(\sigma)| \right) &\leq DJ_{[\cdot]}(\delta, \mathcal{F}_\delta^j) \|S\|_{2, \mathbb{P}} + \sqrt{n} \mathbb{E} \left( S(\mathbf{x}, t) I_{\{S(\mathbf{x}, t) > \sqrt{na}(\delta)\}} \right) \\ &\leq DJ_{[\cdot]}(\delta, \mathcal{F}^j) \|S\|_{2, \mathbb{P}} + \frac{1}{a(\delta)} \mathbb{E} \left( S^2(\mathbf{x}, t) I_{\{S(\mathbf{x}, t) > \sqrt{na}(\delta)\}} \right). \end{aligned}$$

Dado que la función  $S$  tiene segundo momento finito se desprende que para  $\delta$  fijo,

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}} \left( S^2 I_{\{S > \sqrt{na}(\delta)\}} \right) = 0,$$

entonces

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \sup_{|\sigma_0 - \sigma| < \delta} |J_n^j(\sigma)| \right) \leq DJ_{[\cdot]}(\delta, \mathcal{F}^j) \|S\|_{2, \mathbb{P}}.$$

Luego, por lo dicho en (7.11) sabemos que  $J_{[\cdot]}(\delta, \mathcal{F}^j) \xrightarrow{\delta \rightarrow 0} 0$  concluyendo así lo que se quería probar.



# Bibliografía

- [1] Anthony, M. y Bartlett, P. (1999). *Neural Network Learning: Theoretical Foundations*, Cambridge University Press.
- [2] Bates, D. y Watts, D. (1988). *Nonlinear Regression Analysis and Its Applications*.
- [3] Bianco, A. y Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *Journal of Statistical Planning and Inference*, **122**, 229-252.
- [4] Bianco, A. y Spano, P. (2017). Robust inference for nonlinear regression models. *Journal TEST* , <https://doi.org/10.1007/s11749-017-0570-2>.
- [5] Boente, G. y Fraiman, R. (1989). Robust Nonparametric Regression Estimation. *Journal of Multivariate Analysis*, **29**, 180-198.
- [6] Boente, G. y Fraiman, R. (1991). Strong Uniform Convergence Rates for Some Robust Equivariant Nonparametric Regression Estimates for Mixing Processes. *International Statistical Review*, **59**, 355-372.
- [7] Boente, G. y Fraiman, R. (1995). Asymptotic distribution of smoothers based on local means and local medians under dependence. *Journal of Multivariate Analysis*, **54**, 77-90.
- [8] Boente, G., He, X. y Zhou, J. (2006). Robust estimates in generalized partially linear models. *The Annals of Statistics*, **34**, 2856-2878.
- [9] Boente, G. y Rodriguez, D. (2010). Robust inference in generalized partially linear models. *Computational Statistics and Data Analysis*, **54**, 2942-2966.
- [10] Chambers, J. M., Cleveland, W.S., Kleiner, B. y Tukey, P. A. (1983). *Graphical Methods for Data Analysis*. Belmont, CA: Wadsworth.
- [11] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *The Annals of Statistics*, **16**, 136-146.
- [12] Cheng, P.E. y Chu, C.K. (1996). Kernel estimation of distribution functions and quantiles con missing data. *Statistica Sinica* , **6**, 63-78.

- [13] Cleveland, W. (1985). *The elements of graphing data*, Bell Telephone Laboratories Inc., New Jersey.
- [14] Collomb, A. (1981). Estimation nonparametrique de la regression, revue bibliographique. *International Statistical Review*, **49**, 73-93.
- [15] Engle R., Granger C., Rice J. y Weiss A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, **81**, 310-320.
- [16] Fan J. y Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- [17] Fan, J. y Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, **99**, 710-723.
- [18] Fasano, V. (2009). Teoría asintótica de estimadores robustos en regresión no lineal. Tesis Doctoral, Departamento de Matemáticas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata.
- [19] Fasano, V., Marona, R., Sued, M. y Yohai, V. (2012). Continuity and differentiability of regression M functionals. *Bernoulli*, **18**, 1284-1309.
- [20] Fraiman, R. (1983). General M-estimator and applications to bounded influence estimation for non-linear regression. *Communication in Statistics-Theory and Methods*, **22**, 2617-2631.
- [21] Green, P.J., Chapman y Hall. *Nonparametric Regression and Generalized Linear Models, a roughness penalty approach*, London.
- [22] Härdle, W. (1984). Robust regression function estimation. *Journal of Multivariate Analysis*, **14**, 169-180.
- [23] Härdle, W., Liang, H. y Gao, J. (2000). *Partially Linear Models*. Springer-Verlag.
- [24] Härdle, W. y Tsybakov, A. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *The Annals of Statistics*, **16**, 120-135.
- [25] Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73-101.
- [26] Huber, P.J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Journal of Multivariate Analysis*, **1**, 799-821.
- [27] Huet, S., Bouvier, V., Poursat, M. y Jolivet, E. (2004). *Statistical tools for nonlinear regression a practical guide with S-PLUS and R examples*. Springer Series in Statistics, New York.

- [28] Jiang, Y. y Tian, GL. (2017). A robust and efficient estimation method for partially nonlinear models via a new MM algorithm. *Stat Papers*, **108**, 632-643.
- [29] Li, R. y Nie, L. (2008). Efficient Statistical Inference Procedures for Partially Non-linear Models and their Applications. *Biometrics*, **64**, 904-911.
- [30] Loader, C. (1999). Local Regression and Likelihood. Springer, New York.
- [31] Ma, Y., Chiou, J.M. y Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika* , **93**, 75-84.
- [32] Maronna, R., Martin, D. y Yohai, V. (2006). *Robust statistics: Theory and methods*, Wiley, New York.
- [33] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, 141-142.
- [34] Pollard, D. (1984). Convergence of Stochastic Processes, Springer-Verlag, New York Inc.
- [35] Rousseuw, P.J. y Yohai, V.J. (1984). Robust regression by means of S-estimators. *Lecture Notes in Statist*, **26**, 256-272. Springer, New York.
- [36] Ruppert, D. y Sheather, S.J. (1995). Effective Bandwidth Selector for Local Least Squares Regression. *Journal of the American Statistical Association*, **90**, 1257-1270.
- [37] Ruppert, D., Wand M.P. y Carroll, R.J. (2003). Semiparametric Regression. New York, ,Cambridge University Press.
- [38] Sakata, S. y White, H. (2001). S-Estimation of Nonlinear Regression Models With Dependent and Heterogeneous Observations. *Journal of Econometrics* , **103**, 5-72.
- [39] Seber, G. y Wild, C. (1989). Nonlinear regression, John Wiley and Sons, Inc.
- [40] Severini, T.A. y Wong W.H. (1992). Profile likelihood and conditionally parametric models. *The Annals of Statistics*, **20**, 1768-1802.
- [41] Spano, P. (2016). Inferencia robusta en modelos no lineales con respuestas faltantes. Tesis Doctoral, Departamento de Matemáticas, Facultad de Ciencias Exactas, Universidad de Buenos Aires.
- [42] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B* ,**50**, 413-436.
- [43] Stromberg, A.J. (1993). Computation of High Breakdown Nonlinear Regression Parameters. *Journal of the American Statistical Association*, **88**, 237-244.
- [44] Stromberg, A. y Ruppert, D. (1992). Breakdown in nonlinear regression. *Journal of the American Statistical Association*, **87**, 991-997

- [45] Tsybakov, A. (1982). Robust Estimates of a Function. *Problems Inform. Transmission*, **18**, 190-201.
- [46] Vahnovan, V. (2013). Estimadores robustos en modelos de regresión noparamétricos funcionales y en modelos semi-funcionales parcialmente lineales. Tesis Doctoral, Departamento de Matemáticas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata.
- [47] van der Geer, S. (2000). Empirical processes in M-estimation, Cambridge University Press.
- [48] van der Vaart, A. (1998), Asymptotic Statistics, Cambridge University Press
- [49] van der Vaart, A. y Wellner, J. (1996). Weak Convergence and Empirical Processes, Springer-Verlag New York Inc.
- [50] Wand, M.P. y Jones, M.C. (1995). Kernel Smoothing. Chapman and Hall, London.
- [51] Wang, N., Carroll, R.J. y Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal clustered data. *Journal of the American Statistical Association*, **100**, 147-157.
- [52] Watson, G.S. (1964). Smooth Regression Analysis. *Sankhya, The Indian Journal of Statistics. Series A*, **26**, 359-372.
- [53] Yohai, V. (1987). High Breakdown point and high efficiency estimates for regression. *Journal of the American Statistical Association*, **15**, 642-656.