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

Inferencia robusta para modelos parcialmente lineales de índice simple con escala desconocida

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

Lic. María Florencia Statti

Directora de tesis: Dra. Ana M. Bianco
Consejera de estudios: Dra. Graciela L. Boente Boente

Lugar de trabajo: Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales

Buenos Aires, 21 de abril de 2023.

Inferencia robusta para modelos parcialmente lineales de índice simple con escala desconocida

En el problema de regresión, un clásico en Estadística, se desea encontrar una relación funcional entre una variable de respuesta y un conjunto de covariables o variables explicativas. Uno de los modelos paramétricos más difundidos es el modelo lineal y una alternativa muy conocida es la de los modelos no paramétricos. Ambos representan dos ramas muy importantes del análisis de regresión. Lo mejor de los dos mundos puede encontrarse en opciones semiparamétricas, tal es el caso del Modelo Parcialmente Lineal de Índice Simple en el que se combinan ambos enfoques.

Por un lado, este modelo asume que la variable de respuesta y se relaciona con dos vectores de covariables $\mathbf{x} \in \mathbb{R}^p$ y $\mathbf{t} \in \mathbb{R}^q$ satisfaciendo la ecuación

$$y \;=\; \boldsymbol{\beta}_0^t \mathbf{x} + \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}) + \sigma_0 \epsilon \;,$$

donde la función $\eta_0 : \mathcal{I} \to \mathbb{R}$, $\mathcal{I}$ intervalo real acotado, el vector de parámetros $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \in \mathbb{R}^p \times \mathbb{R}^q$ y el parámetro *nuisance* σ_0 son desconocidos. Asumimos que $\|\boldsymbol{\theta}_0\| = 1$ y que su primera componente es positiva, a los fines de identificabilidad. Este tipo de modelos tienen un gran potencial en las aplicaciones cuando se piensa en ajustar una regresión no paramétrica múltiple, ya que a través del índice simple logran reducir a uno la dimensionalidad de la componente no paramétrica y permite, al mismo tiempo, capturar una posible relación no lineal entre y y $\mathbf{t}$ a través de la función η_0.

Por otro lado, es bien sabido que los métodos de estimación clásica suelen ser altamente sensibles a la presencia de datos atípicos y por este motivo son necesarios procedimientos alternativas robustos.

El principal objetivo de esta Tesis es estudiar métodos de inferencia robustos para el Modelo Parcialmente Lineal de Índice Simple. Para ello proponemos una familia de estimadores robustos obtenidos con un procedimiento de dos pasos y un estimador inicial a fin de poder comenzar el procedimiento con estimadores preliminares resistentes a datos atípicos. Para el estimador propuesto estudiamos su comportamiento asintótico, deduciendo la consistencia y la distribución asintótica. Estos resultados teóricos se completan con un estudio de simulación con el objetivo de evaluar el comportamiento de la propuesta robusta frente a distintas contaminaciones y en muestras finitas.

Por otro lado, introducimos una familia de tests robustos para chequear hipótesis que involucren al parámetro de la componente lineal $\boldsymbol{\beta}_0$ utilizando un estadístico de tipo Wald basado en el estimador propuesto. Deducimos su distribución asintótica bajo la hipótesis nula y realizamos un estudio de simulación preliminar para evaluar el comportamiento de los tests propuestos.

Palabras Claves: Modelo Parcialmente Lineal de Índice Simple; Estimación robusta; Comportamiento asintótico; Tests de hipótesis robustos.

Robust inference for partially linear single index models with unknown scale

In regression analysis, a classic topic in Statistics, it is desired to find a functional relationship between a response variable and a set of covariates or explanatory variables. One of the most widespread parametric models is the linear one, while non–parametric models are a very well–known alternative. They both represent two very important branches of regression analysis. The best of both worlds can be found among semiparametric models, as it is the case of Single Index Partially Linear Model, where both approaches are combined. In this model it is assumed that the response variable y is related to two covariate vectors $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$ through the equation

$$y \;=\; \boldsymbol{\beta}_0^t \mathbf{x} + \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}) + \sigma_0 \epsilon \;,$$

where the univariate real function $\eta_0 : \mathcal{I} \to \mathbb{R}$, with $\mathcal{I}$ a bounded real interval, the vector of parameters $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \in \mathbb{R}^p \times \mathbb{R}^q$ and the *nuisance* parameter σ_0 are unknown. We assume that $\|\boldsymbol{\theta}_0\| = 1$ and that its first component is positive, for identifiability purposes. These models have great potential in the case of multiple non–parametric regression, since they manage to reduce the dimensionality of the non–parametric component to 1 through the single index and at the same time, capture a possible non–linear relationship between y and $\mathbf{t}$ through the function η_0.

It is known that classical estimation methods are usually highly sensitive to the presence of outliers and robust procedures are welcome in this scenario. The main objective of this thesis is to study robust inference methods for the Partially Linear Single Index Model. For this, we propose a family of robust estimators obtained through a two–step procedure and an initial estimator in order to be able to start the procedure with preliminary estimators resistant to outliers. For the proposed estimator we study its asymptotic behaviour, deriving consistency and their asymptotic distribution. These theoretical results are completed with a simulation study in order to evaluate the behaviour of our robust proposal under different contamination schemes and in finite samples.

On the other hand, we introduce a family of robust tests to check hypotheses involving the linear component parameter $\boldsymbol{\beta}_0$ using a Wald–type statistic based on the proposed estimator. We derive its asymptotic distribution under the null hypothesis and we carried out a preliminary simulation study to evaluate the behaviour of the proposed test.

Keywords: Partially Linear Single Index Model; Robust Estimation; Asymptotic Behaviour; Robust Hypotheses Testing.

Agradecimientos

A mi Mamá y mi Papá, por darme Vida, por amarme tanto, siempre creyendo en mis capacidades y ayudándome a desenvolverme en el mundo. También por inculcarme el valor del esfuerzo, de la perseverancia, de la empatía, de la generosidad. Gracias Mami por ayudarnos a Marcos y a mí con la crianza de Faustino y ser una abuela genial. Gracias Papi por estar feliz de este logro, tu ausencia física es insoslayable y te extrañamos.

A Faustino, mi hijo, por llegar a mi Vida en la mitad del doctorado y hacerla más feliz y desafiante, llena de amor y muchos aprendizajes, por ayudarme a transformarme y ser una mejor versión de mí.

A Marcos, por creer en mí y seguirme, y haber construído juntos la familia que tenemos, sosteniéndola con amor, escucha, empatía, tolerancia y paciencia, transformándonos todos los días en una mejor versión de nosotros.

A Vane, mi hermanAmiga, por sus incontables consejos y saber que siempre dispongo de su contención y buena energía cuando la necesito, por valorar mi transformación y ayudarme mucho en ese proceso.

A Marian y Lucho, mis hermanos, que tan orgullosos están de mí y siempre me alentaron en mi vida profesional.

A Alma y Agustín, mis maravillosos sobris por alegrarme tanto y por la dulzura que derraman.

A mis geniales amigas: Lucía, Luz, Jime, Ale, Flor, Romi y Pau, por ayudarme a descomprimir en momentos de mucha angustia, escuchándome, aconsejándome, preocupándose por mí y haciéndome sentir muy querida y valorada.

A Millie y Pablo, mis suegros, por ayudarnos a Marcos y a mí con la crianza de Faustino y ser abuelos amorosos.

A Ana, mi directora de doctorado, por haberme dirigido con paciencia y confianza, tranquilizándome cuando lo necesité y con mucha empatía en cuanto a la maternidad.

A Mariela, cuya vocación docente y empuje me convocó a desafiarme y lograr en mí un mejor ejercicio docente: con mayor dedicación y compromiso, y el entusiasmo de siempre. Además porque siendo su alumna, y sin conocerme tanto, me brindó posibilidades de crecimiento profesional. Es una referente para mí.

A Graciela, por ser mi consejera de estudios.

A toda la Comunidad del Instituto de Cálculo, en especial a Marie, Mari, Maru y Dani.

También a Lida, Sole, Jemi, Luciana, Inés, Guille, Fede, Oscar, Diana, Willy, Naza y muchas/os más. ¡Es un gusto trabajar acá!

A Gonza y Manu, compañeros durante el doctorado, y a Pablo T. compañero de oficina que me asistió en temas computacionales en innumerables situaciones.

A la y los integrantes del jurado de esta Tesis, por evaluarla y por sus devoluciones y amables comentarios durante la defensa.

A la Argentina y sus habitantes, el país donde elijo vivir y con mucho entusiasmo y orgullo lo hago, un país con educación superior pública, gratuita y de excelencia y cuya política pública sostenida a lo largo del tiempo le ha permitido a muchas/os habitantes del suelo argentino lograr un ascenso social, además de intelectual.

Para el GRAN pueblo argentino

Índice general

1. Introducción **13**

2. Preliminares **17**
2.1. Estimación Paramétrica: Modelo Lineal 17
2.2. Estimación No Paramétrica de la Regresión 20
2.3. Estimación de la componente de Índice Simple 21

3. Propuesta y Propiedades **23**
3.1. Modelo 23
3.2. Estimación 23
3.2.1. Propuesta 23
3.2.2. Estimación Inicial 25
3.3. Consistencia 26
3.3.1. Consistencia del estimador propuesto 26
3.3.2. Sobre el estimador inicial 37

4. Distribución Asintótica **53**
4.1. Resultados Principales 53
4.2. Apéndice 73

5. Estudio de Simulación **77**
5.1. Condiciones del Experimento Numérico 77
5.2. Resultados 78

6. Aplicación: ejemplo con datos reales **105**
6.1. Una actualización del estimador de σ_0 105
6.2. Estimaciones 106

7. Tests de Hipótesis **109**
7.1. Estadístico 109
7.2. Estimación de la matriz de covarianzas 111
7.2.1. Estimador de $\boldsymbol{\Delta}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ usando una estimación de $\boldsymbol{\beta}_0$ 111

7.2.2. Estimador de $\boldsymbol{\Delta}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ usando el valor de $\boldsymbol{\beta}_0$ bajo H_0 114
7.3. Estudio de Simulación: Nivel Empírico 115

Bibliografía **119**

Que los tesoros que descubra en mí
hagan la vida de otras personas más felices.

(parafraseando a) **Lama Tarthang Tulku**

Extraído del libro "Una Escuela en Ronda",
de Magdalena Fleitas

Capítulo 1

Introducción

El modelo lineal es ampliamente utilizado en los análisis de datos de regresión que así lo permitan, debido a su vasta flexibilidad y excelente interpretación en variadas situaciones. El modelo de regresión no paramétrico ha permitido estudiar una gran cantidad de fenómenos que quedaban relegados, pero sufre el problema de la *maldición de la dimensionalidad*. Para combatirla, Ichimura en su tesis de doctorado en 1987, introduce el *Modelo de Índice Simple* (MIS) que supone que el vector aleatorio explicativo en cuestión, $\mathbf{t}$, puede ser colapsado en un índice simple, $\boldsymbol{\theta}_0^t\mathbf{t}$, a través de una función no paramétrica, η_0, que es suave y no constante.

El *Modelo Parcialmente Lineal de Índice Simple* (MPLIS) combina la facilidad del modelo lineal con la sofisticación de un modelo no paramétrico que evita la *maldición de la dimensionalidad*. Sean $(y_i, \mathbf{x}_i, \mathbf{t}_i)$ observaciones independientes cumpliendo un MPLIS, donde

$$y_i \;=\; \boldsymbol{\beta}_0^t\mathbf{x}_i + \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_i) + \sigma_0\epsilon_i\,, \tag{1.1}$$

siendo el vector aleatorio $(\mathbf{x}_i^t, \mathbf{t}_i^t) \in \mathbb{R}^p \times \mathbb{R}^q$ independiente del error ϵ_i. Tanto la función real univariada continua η_0 como el vector de parámetros $(\boldsymbol{\beta}_0^t, \boldsymbol{\theta}_0^t) \in \mathbb{R}^p \times \mathbb{R}^q$ y el parámetro *nuisance* σ_0, son desconocidos. Con el propósito de que el modelo sea identificable, se supone que $\|\boldsymbol{\theta}_0\| = 1$, ya que al ser η_0 desconocida, sólo la dirección del vector $\boldsymbol{\theta}_0$ puede ser reconocida, y además supondremos, sin pérdida de generalidad, que la primera componente, θ_{01}, es positiva.

El contexto clásico supone que los momentos del error cumplen que $\mathbb{E}(\epsilon|(\mathbf{x}, \mathbf{t})) = 0$ y $\mathbb{E}(\epsilon^2|(\mathbf{x}, \mathbf{t})) < \infty$. En este trabajo se logra, mediante procedimientos robustos, mitigar el efecto que observaciones atípicas producen al estimar los parámetros del modelo (1.1) y hacer inferencia sobre los mismos, en particular contrastar hipótesis sobre ellos, asumiendo que el error tiene distribución simétrica alrededor de 0, un supuesto habitual en el marco de robustez.

Según la bibliografía consultada, Carroll et al. (1997) son los primeros que proponen estimadores de los parámetros de un *Modelo Parcialmente Lineal Generalizado de Índice Simple* (MPLGIS), o sea en un contexto más específico que un MPLIS, ya que con la función de enlace identidad proponen estimadores de los parámetros de un MPLIS con σ_0 conocido. Como resaltan Liang et al. (2010), este trabajo pionero plantea un algoritmo de *backfitting* pero donde los estimadores resultantes pueden ser computacionalmente inestables,

como indican Yu y Ruppert (2002), además de que la función no paramétrica debe ser subsuavizada para reducir el sesgo de los estimadores paramétricos. Yu y Ruppert (2002), segundos en la cronología pionera del MPLIS, estiman a η_0 mediante un procedimiento de spline penalizado (P-spline) que es rápido y computacionalmente estable pero cuyos estimadores no resultan eficientes ya que la covarianza asintótica no alcanza la cota de eficiencia semiparamétrica que se especifica en Carroll et al. (1997), como bien remarcan Liang et al. (2010).

Xia y Härdle (2006) proponen estimadores de todos los parámetros, incluso la función no paramétrica η_0, del MPLIS minimizando una única expresión, gran virtud que se confieren respecto de Carroll et al. (1997). Argumentan que esa es la razón por la cual no tienen que subsuavizar. Sin embargo, esa expresión no brinda fórmula cerrada para obtener los estimadores y es por ello que, en ese trabajo, es introducido el algoritmo PLSIM.

Justamente la búsqueda de Liang et al. (2010) es introducir un procedimiento a partir del método de perfiles de Severini y Wong (1992) que no sea iterativo. Ese trabajo propone el enfoque de perfiles de mínimos cuadrados que permite obtener estimadores eficientes que proveen la cota de eficiencia semiparamétrica antes mencionada. En este artículo estiman a η_0, usando la técnica de regresión lineal local introducida por Fan y Gijbels (1996), dependiendo de una variable, que será el perfil, y luego aplican el método de Newton-Raphson para obtener un minimizante, en el perfil, de una suma de cuadrados de residuos. También basándose en el método de perfiles, Boente y Rodriguez (2012) introducen estimadores robustos para un MPLGIS utilizando un procedimiento de tres pasos que en su primer paso requiere estimadores robustos y consistentes.

A diferencia de los trabajos de Carroll et al. (1997), Yu y Ruppert (2002) y Xia y Härdle (2006) en donde es necesario estimar consistentemente a η_0 para garantizar la consistencia de los estimadores de $\boldsymbol{\beta}_0$ y $\boldsymbol{\theta}_0$, Ma y Zhu (2013) presentan estimadores consistentes de los parámetros del término lineal y del índice simple aún cuando la componente no paramétrica η_0 tiene una estimación que no es consistente o está mal especificada.

Yang y Yang (2014) introducen estimadores robustos de los parámetros de un MPLIS utilizando a la moda como medida de posición robusta. Su propuesta extiende la de Liu et al. (2013) para MIS.

Inspirado en la propuesta realizada por Boente y Rodriguez (2012), el procedimiento de estimación robusto que se propone en esta Tesis utiliza un *método de perfiles robustos* basado en un algoritmo de dos pasos que en su primer paso necesita de un estimador preliminar de σ que sea robusto y consistente. Se prueba consistencia y distribución asintótica de los estimadores propuestos, y con ésto se introduce la posibilidad de hacer tests de hipótesis. Las hipótesis a contrastar de manera robusta que se estudian en esta Tesis son:

$$H_0 : \boldsymbol{\beta}_0 = \boldsymbol{\beta}^* \quad \text{contra} \quad H_1 : \boldsymbol{\beta}_0 \neq \boldsymbol{\beta}^* \,, \tag{1.2}$$

cuyo estadístico es de tipo Wald y comprende una forma cuadrática basada en la matriz de covarianza asintótica del estimador propuesto para la componente lineal. Se estudia la distribución del estadístico bajo H_0 y se proponen dos implementaciones que varían en la estimación de la matriz en la que está basado. Se estudió el comportamiento de los niveles

empíricos de estas pruebas de hipótesis, tanto para el procedimiento clásico como el aquí propuesto y estudiado, ya sea en muestras sin contaminar como en contaminadas.

La estructura de este trabajo se describe a continuación. En el Capítulo 2 se hará un breve y conciso resumen que permitirá repasar los contenidos que resultan necesarios para una fluída lectura de esta Tesis; en el Capítulo 3 se presentará el método de estimación que, entre otros temas, esta Tesis propone y se derivará su consistencia. En el Capítulo 4 se desarrollará la distribución asintótica de los estimadores propuestos; mientras en el Capítulo 5 se evaluará el desempeño de esta nueva propuesta simulando variados escenarios de contaminación, con el objetivo de verificar sus propiedades robustas y compararlos con los estimadores tradicionales para muestras finitas. En el Capítulo 6 se aplicará el procedimiento de estimación robusto a datos reales y se compararán los resultados de este análisis con los obtenidos usando la metodología clásica. En el Capítulo 7 se propondrá un test de tipo Wald para chequear hipótesis que involucran al parámetro de la componente lineal y se darán dos formas de construir el estadístico. Asimismo, se presentarán los resultados de un estudio de simulación que permita evaluar los niveles empíricos alcanzados en distintas situaciones.

Capítulo 2

Preliminares

En este capítulo se describen métodos de estimación para la función de regresión propuestos en la literatura, tanto clásica como robusta, en modelos paramétricos y en no paramétricos, que se extienden a modelos semiparamétricos como es de interés en esta Tesis. Esta descripción facilitará la comprensión de nuestra propuesta de estimación robusta que se presentará en el próximo Capítulo.

2.1. Estimación Paramétrica: Modelo Lineal

Un problema clásico que aparece en distintos escenarios es el de encontrar relaciones funcionales entre variables. Sea y una variable aleatoria que está asociada a otras variables aleatorias $x_1, \ldots, x_p$ y a otros factores no identificados que ocurren aleatoriamente, de tal manera que la relación puede escribirse como

$$y = m(x_1, \ldots, x_p) + \sigma\epsilon \tag{2.1}$$

donde m es una función de regresión, en general desconocida, y el término del error ϵ es una variable aleatoria que representa el efecto de factores desconocidos y aleatorios sobre y. Con frecuencia se denomina respuesta a la variable y, mientras que a $x_1, \ldots, x_p$ se las denomina variables independientes o explicativas, o covariables.

El interés suele estar centrado en la estimación de la función de regresión m. Se suelen hacer supuestos que limitan su elección a una familia de funciones de regresión estableciendo que la función m pertenece a una familia $\mathcal{M}$ de funciones de $\mathbb{R}^p \to \mathbb{R}$.

En el caso del modelo (2.1) con m restringida a una familia $\mathcal{M}$, se llaman modelos de regresión y pueden clasificarse en paramétricos y no paramétricos. Se dice que el modelo de regresión es paramétrico si la familia $\mathcal{M}$ puede escribirse como

$$\mathcal{M} = \{m(x_1, \ldots, x_p, \beta_1, \ldots, \beta_p) : \boldsymbol{\beta} = (\beta_1, \ldots, \beta_p) \in \Gamma \subset \mathbb{R}^p\}$$

es decir, si la función de regresión está indizada por un número finito de parámetros. En particular, se dice que un modelo de regresión es lineal si es de la forma

$$m(x_1, \ldots, x_p) = \beta_1 x_1 + \cdots + \beta_p x_p \,,$$

siendo $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ el vector de parámetros desconocidos a estimar.

Supongamos que disponemos de n observaciones $(y_1, x_{11}, \ldots, x_{1p}), \ldots, (y_n, x_{n1}, \ldots, x_{np})$ que satisfacen el modelo

$$y_i = \beta_{01} x_{i1} + \cdots + \beta_{0p} x_{ip} + \epsilon_i = \boldsymbol{\beta}_0^t \mathbf{x}_i + \sigma_0 \epsilon_i \qquad , \qquad 1 \leq i \leq n \tag{2.2}$$

donde hemos notado $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^t$. En la teoría clásica, los errores cumplen

$$\mathbb{E}(\epsilon_i) = 0, \quad \mathbb{V}(\epsilon_i) = 1, \quad \operatorname{cov}(\epsilon_i, \epsilon_j) = 0 \text{ si } i \neq j \,.$$

La forma clásica de estimar al vector de parámetros $\boldsymbol{\beta}_0$ es mediante el método de mínimos cuadrados que consiste en minimizar, en $\boldsymbol{\beta}$, la función

$$\sum_{i=1}^{n} (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2 \,. \tag{2.3}$$

Es muy conocido el hecho de que los estimadores de mínimos cuadrados no son robustos, en el sentido de que son muy sensibles a la presencia de datos atípicos. Si bien son óptimos bajo normalidad, pueden ser ineficientes y sesgados bajo no normalidad. Muchas alternativas robustas han sido propuestas a fin de aliviar este problema resultando en estimadores estables frente a perturbaciones del modelo y a la vez eficientes bajo el modelo central.

El problema en la función objetivo en (2.3) es que los puntos atípicos están alejados de la mayoría de los datos y pueden dar origen a residuos grandes que podrían dominar la minimización. Una manera natural de controlar a estos puntos es reemplazando la función cuadrática por una función de un crecimiento más lento, tal como es el caso de una $\rho-$función (ver Maronna et al. (2019)), que se define de la siguiente forma.

Definición 2.1.1. Una ρ-función denota a una función $\rho : \mathbb{R} \to \mathbb{R}$ cumpliendo las siguientes condiciones:
R1. $\rho(x)$ es una función no decreciente de $|x|$.
R2. $\rho(0) = 0$.
R3. $\rho(x)$ es creciente para $x > 0$ tal que $\rho(x) < \lim\limits_{x \to +\infty} \rho(x)$.

Los $M-$estimadores, introducidos por Huber (1973), fueron pioneros en el campo de la estimación robusta en el contexto del modelo lineal. Fueron seguidos por su generalización, los $GM-$estimadores que controlan el leverage de las covariables mediante pesos. Estos estimadores tienen dos formas típicas, una debida a Mallows (1975) y otra debida a Schweppe, Wildes

y Rom (1970). En búsqueda de mayor robustez, surgieron los S-estimadores introducidos por Rousseeuw y Yohai (1984) y por una versión superadora presentada por Yohai (1987) conocida como $MM-$estimadores, capaces de lograr eficiencia y robustez simultáneamente. Recordaremos brevemente su definición.

Dada una muestra de tamaño n, $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$, que satisface el modelo (2.2) se define un $M-$estimador de regresión como

$$\widehat{\boldsymbol{\beta}}_M = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t \boldsymbol{\beta}}{\widehat{\sigma}_n}\right). \tag{2.4}$$

donde $\widehat{\sigma}_n$ es un estimador robusto de escala de los residuos y la función ρ es una $\rho-$función tal como hemos definido más arriba.

Dados $r_1, \ldots, r_n$, un $M-$estimador de escala se define como la solución positiva $\widehat{\sigma}_n(\mathbf{r})$ en s de

$$\sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right) = b\,, \tag{2.5}$$

donde ρ es una $\rho-$función y b es una constante positiva menor a 1. Cabe destacar que, si llamamos Φ a la función de distribución normal, debe tomarse $b = \mathbb{E}_\Phi(\rho\,(u))$ para que el $M-$estimador de escala resultante de (2.5) coincida con el desvío estándar bajo normalidad y que su punto de ruptura, es decir la máxima proporción de datos atípicos que el estimador tolera, sea mín$(b, 1-b)$ (Yohai y Zamar (1988)).

A partir de la noción de $M-$estimador de escala, introducida por Huber (1981), podemos definir los $S-$estimadores de regresión para una muestra $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$, que sigue el modelo (2.2), como

$$\widehat{\boldsymbol{\beta}}_S = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \widehat{\sigma}_n\left(\mathbf{r}(\boldsymbol{\beta})\right)\,, \tag{2.6}$$

donde $\mathbf{r}(\boldsymbol{\beta})$ es el vector de residuos con componentes $r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{\beta}^t \mathbf{x}_i$ y $\widehat{\sigma}_n(\mathbf{r})$ es la $M-$escala definida en (2.5). Estos estimadores suelen usarse como valores iniciales para otros métodos de estimación, tales como los $MM-$estimadores, debido a su estabilidad frente a datos anómalos. Dado que no son resistentes y eficientes al mismo tiempo, ése suele ser su mayor uso.

Precisamente, con el objetivo de obtener un estimador con alto punto de ruptura y que sea altamente eficiente bajo la distribución normal, Yohai (1987) introdujo los $MM-$estimadores del parámetro de un modelo lineal. Un $MM-$estimador puede definirse a través de los siguientes tres pasos:

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

Paso 2. Calcular los residuos $r_i(\widehat{\boldsymbol{\beta}}_1) = y_i - \widehat{\boldsymbol{\beta}}_1^t \mathbf{x}_i$ y computar la $M-$escala basada en la muestra de los residuos observados usando una $\rho-$función ρ_0 con $b = 0{,}5||\rho_0||_\infty$, es decir $\widehat{\sigma}$ es tal que

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{r_i(\widehat{\boldsymbol{\beta}}_1)}{\widehat{\sigma}}\right) = b.$$

Paso 3. Sea ρ_1 otra $\rho-$función tal que $\rho_1(t) \leq \rho_0(t)$, $||\rho_1||_\infty = ||\rho_0||_\infty$, luego el $MM-$estimador se define como

$$\widehat{\boldsymbol{\beta}}_{MM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_1 \left(\frac{r_i(\boldsymbol{\beta})}{\widehat{\sigma}} \right). \tag{2.7}$$

2.2. Estimación No Paramétrica de la Regresión

En el enfoque no paramétrico de la regresión no se realizan supuestos de forma de la función que relaciona a la respuesta con las covariables tal como se hace en el modelo lineal. De manera que la estimación de la función m dada en (2.1) estará guiada por lo que sugieran los datos empíricamente y sólo se asumirá que m es suave. En el contexto clásico, se supone que el término del error en (2.1) cumple que $\mathbb{E}(\epsilon|\mathbf{x}) = 0$, por lo que $m(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$.

Los estimadores de núcleos propuestos independientemente por Nadaraya (1964) y Watson (1964) pueden escribirse como un promedio pesado en el que la ponderación depende de la cercanía al punto en el que se desea realizar la estimación de la regresión. De esta forma, si $\mathbb{K} : \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ es una función integrable y $h > 0$ es una ventana, tenemos que para $\mathbf{u} \in \mathbb{R}^p$ el estimador de Naradaya–Watson está dado por

$$\widehat{m}(\mathbf{u}) = \sum_{i=1}^{n} W_i^h(\mathbf{u}) y_i \,,$$

donde los pesos $W_i^h(\mathbf{u})$ están definidos por

$$W_i^h(\mathbf{u}) = \frac{\mathbb{K}\left(\dfrac{\mathbf{x}_i - \mathbf{u}}{h} \right)}{\displaystyle\sum_{j=1}^{n} \mathbb{K}\left(\frac{\mathbf{x}_j - \mathbf{u}}{h} \right)} \,.$$

Una elección frecuente es tomar al núcleo $\mathbb{K}$ como $\mathbb{K}(\cdot) = K(\|\cdot\|)$, siendo K una función de densidad univariada y $\|\cdot\|$ la norma euclídea en $\mathbb{R}^p$, con lo que el estimador resulta

$$\widehat{m}(\mathbf{u}) = \frac{\displaystyle\sum_{i=1}^{n} K\left(\frac{\|\mathbf{x}_i - \mathbf{u}\|}{h} \right) y_i}{\displaystyle\sum_{j=1}^{n} K\left(\frac{\|\mathbf{x}_j - \mathbf{u}\|}{h} \right)} \,.$$

Al tratarse de promedios, estos estimadores pueden verse afectados por respuestas atípicas. Härdle (1984) estudia una versión robusta no equivariante de los estimadores de Nadaraya–Watson, mientras que una versión con propiedades de equivarianza fueron considerados en

Boente y Fraiman (1989). Los $M-$estimadores locales son la solución de

$$\sum_{i=1}^{n} W_i^h(\mathbf{x}_i)\psi\left(\frac{y_i - \widehat{m}(\mathbf{x}_i)}{\widehat{s}(\mathbf{x}_i)}\right) = 0\,,$$

donde ψ es la derivada de una $\rho-$función, es decir, $\psi = \rho'$ y $\widehat{s}(\mathbf{x})$ es un estimador de escala robusto. Con frecuencia $\widehat{s}(\mathbf{u})$ se elige como la mediana local de los desvíos absolutos respecto de la mediana local, es decir como la MAD de la distribución empírica condicional de $y|\mathbf{x} = \mathbf{u}$.

2.3. Estimación de la componente de Índice Simple

En un modelo de índice simple se tiene que

$$\mathbb{E}(y|\mathbf{t}) = \eta_0(\boldsymbol{\theta}_0^t\mathbf{t})\,, \tag{2.8}$$

donde $\boldsymbol{\theta}_0 \in \mathbb{R}^q$ es desconocido y $\eta_0 : \mathbb{R} \to \mathbb{R}$ es una función suave desconocida. Se denomina índice simple a $\boldsymbol{\theta}_0^t\mathbf{t}$. Notemos que la esperanza en (2.8) depende de $\mathbf{t}$ a través del índice simple, es decir que podemos considerar equivalentemente al modelo

$$\mathbb{E}(y|\boldsymbol{\theta}_0^t\mathbf{t}) = \eta_0(\boldsymbol{\theta}_0^t\mathbf{t})\,. \tag{2.9}$$

Con el propósito de que el modelo sea identificable, se supone que $\|\boldsymbol{\theta}_0\| = 1$, ya que al ser η_0 desconocida, sólo la dirección del vector $\boldsymbol{\theta}_0$ puede ser reconocida, y además supondremos, sin pérdida de generalidad, que la primera componente, θ_{01}, es positiva.

Un modelo un poco más general, con $g(\mathbf{t};\boldsymbol{\theta}_0)$ siendo g una función conocida en lugar de $\boldsymbol{\theta}_0^t\mathbf{t}$, fue introducido por Ichimura en su tesis de doctorado y publicado luego (Ichimura, 1993). En aquel trabajo, se propone el siguiente argumento heurístico basado en la varianza condicional: minimizar en $\boldsymbol{\theta}$ la función objetivo

$$J(\boldsymbol{\theta}) = \mathbb{E}((y - \mathbb{E}(y|\boldsymbol{\theta}^t\mathbf{t}))^2)\,.$$

La propuesta pionera de Ichimura tiene la gran desventaja de ser difícil de calcular en la práctica. Justamente, en la búsqueda de un cálculo más eficiente de estimadores del modelo (2.8), es que surgieron nuevos procedimientos, uno de los más relevantes es el MAVE (por sus siglas en inglés: minimum average variance estimation) ideado por Xia y Härdle (2006), en donde estiman los parámetros del modelo a partir de la minimización de una sola función de pérdida, la función objetivo, utilizando una aproximación lineal local. Sin embargo, necesitan de dos ecuaciones con las que iterar para obtener los estimadores, algoritmo PLSIM llaman a ese procedimiento.

Como ya fue mencionado, Carroll et al. (1997) introducen el modelo parcialmente lineal generalizado de índice simple. Esta generalización natural del modelo lineal generalizado,

que incluye el modelo tratado en esta Tesis, representa una reducción de la dimensión en tanto asume que el efecto del vector de covariables $\mathbf{t}$ se condensa mediante el índice simple $\boldsymbol{\theta}_0^t\mathbf{t}$, reduciendo así a 1 la dimensión en la que se estima la componente no paramétrica. En su trabajo, Carroll et al. (1997), combinan las técnicas descriptas más arriba para estimar las distintas componentes del modelo. Sin embargo, los métodos de estimación clásicos no controlan el impacto de datos atípicos presentes en la muestra y, en la mayoría de los casos, en la etapa de estimación les es indiferente contar con un buen estimador del parámetro *nuisance* σ_0. Por el contrario, los métodos de estimación robustos necesitan de una buena estimación de σ_0 para poder determinar cuándo un residuo es grande y asignarle poca influencia a la observación correspondiente, y así lograr disminuir el impacto de estos datos sobre los estimadores resultantes.

Un primer enfoque que se ocupa de minimizar el impacto de datos atípicos en la variable respuesta de un MIS es el presentado por Delecroix et al. (2006). Su limitación radica en que asume conocido el parámetro *nuisance* σ_0. Boente y Rodriguez (2012) proponen estimadores robustos para un modelo parcialmente lineal generalizado de índice simple, pero también asumen que los parámetros auxiliares son conocidos. Recientemente, Agostinelli et al. (2020) proponen un procedimiento de estimación robusto donde consideran al parámetro *nuisance* desconocido. Extienden la esencia de los MM-estimadores a un MIS, con todas las sutilezas que el nuevo modelo planteado requiere.

Capítulo 3

Propuesta y Propiedades

3.1. Modelo

Consideremos el *Modelo Parcialmente Lineal de Índice Simple* (MPLIS) en el que se observa un vector $(y, \mathbf{x}, \mathbf{t})$, donde la variable respuesta y se relaciona con los dos vectores de covariables $\mathbf{x}$ y $\mathbf{t}$ mediante la ecuación

$$y = \boldsymbol{\beta}_0^t \mathbf{x} + \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}) + \sigma_0 \epsilon \,, \tag{3.1}$$

siendo $\mathbf{x} \in \mathbb{R}^p$ y $\mathbf{t} \in \mathbb{R}^q$, y donde $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, $\boldsymbol{\theta}_0 \in \mathbb{R}^q$ y $\sigma_0 \in \mathbb{R}$ son parámetros desconocidos y la función real univariada continua η_0 también lo es. Además asumiremos que el error ϵ es independiente del vector de covariables $(\mathbf{x}, \mathbf{t})$.

Para que el modelo sea identificable, supondremos que $||\boldsymbol{\theta}_0|| = 1$ y que su primera componente es positiva, ya que por el hecho de que η_0 sea desconocida, sólo la dirección del vector $\boldsymbol{\theta}_0$ puede ser reconocida. De ahora en más, llamaremos $\mathcal{S}^1 = \{\boldsymbol{\theta} \in \mathbb{R}^q : ||\boldsymbol{\theta}|| = 1\}$.

3.2. Estimación

3.2.1. Propuesta

Sea $\{(y_i, \mathbf{x}_i, \mathbf{t}_i)\}_{1 \leq i \leq n} \subset \mathbb{R}^{p+q+1}$ una muestra aleatoria de vectores que siguen la misma distribución que $(y, \mathbf{x}, \mathbf{t})$, o sea

$$y_i = \boldsymbol{\beta}_0^t \mathbf{x}_i + \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}_i) + \sigma_0 \epsilon_i \,, \tag{3.2}$$

y los errores ϵ_i son idénticamente distribuidos, independientes e independientes de las covariables $(\mathbf{x}_i, \mathbf{t}_i)$. Respecto de la distribución de los errores, sólo asumiremos que es simétrica respecto de 0. Como es usual en los modelos parcialmente lineales, supondremos que el vector de unos, $\mathbf{1}_n$, no está en el espacio generado por las columnas de la matriz que tiene a las $\mathbf{x}_i$ en sus filas. En otras palabras, para que el modelo sea identificable, $\boldsymbol{\beta}_0$ no incluirá a la ordenada al origen. Denotemos $\mathbb{P}$ y $\mathbb{E}$ a la probabilidad y la esperanza respectivamente, bajo este modelo central.

Para la presentación de los estimadores consideraremos la clase de funciones de pérdida, definida en la Definición 2.1.1, que reemplazará a la clásica pérdida cuadrática a fin de controlar los residuos grandes y potencialmente influyentes, y nos permitirá construir funciones de peso adecuadas.

Sean $w_i : \mathbb{R}^p \to \mathbb{R}$ para $i = 1, 2$ dos funciones de peso que introduciremos a fin de controlar posibles puntos $\mathbf{x}_i$ con alto leverage o palanca, habitualmente se eligen w_i acotadas y no negativas.

Consideremos los pesos $W^h_{\boldsymbol{\theta},i}(u)$ que dependen de la cercanía de u a la proyección de $\mathbf{t}$ sobre la dirección $\boldsymbol{\theta}$. Consideraremos pesos basados en un núcleo no negativo y simétrico K, definidos por

$$W^h_{\boldsymbol{\theta},i}(u) = W^h(u, \boldsymbol{\theta}^t\mathbf{t}_i) = \frac{K\left(\dfrac{\boldsymbol{\theta}^t\mathbf{t}_i - u}{h}\right)}{\displaystyle\sum_{j=1}^{n} K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_j - u}{h}\right)},$$

siendo h el parámetro de suavizado o ventana.

Para estimar los parámetros mencionados, utilizaremos el siguiente procedimiento robusto basado en perfiles que utiliza un estimador preliminar $\widehat{\sigma}$ del parámetro de escala σ_0 e involucra a una función de pérdida $\rho_1 : \mathbb{R} \to \mathbb{R}$, ρ-función como las ya introducidas. La definición de los estimadores propuestos está dada por el siguiente algoritmo de dos pasos:

Paso 1: Computamos $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ como

$$\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} W^h_{\boldsymbol{\theta},i}(u)\rho_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - a}{\widehat{\sigma}}\right) w_1(\mathbf{x}_i). \tag{3.3}$$

Paso 2: Computamos el estimador $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right)$ de $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ como

$$\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right) = \underset{\boldsymbol{\beta}\,\in\,\mathbb{R}^p,\ \boldsymbol{\theta}\,\in\,\mathcal{S}^1}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\,, \tag{3.4}$$

donde $\mathcal{S}^1 = \{\boldsymbol{\theta} \in \mathbb{R}^q : ||\boldsymbol{\theta}|| = 1\}$.

Observación 3.2.1. Cabe mencionar que este procedimiento en pasos sólo involucra suavizadores no paramétricos unidimensionales, evitando así la posible *maldición de la dimensionalidad* asociada a la alta dimensionalidad de las covariables $\mathbf{t}$. Por otro lado, dado que en la mayoría de los núcleos más populares, el máximo de K se alcanza en 0, la contribución de la $i-$ésima observación al suavizador está determinada por la cercanía entre u y $\boldsymbol{\theta}^t\mathbf{t}_i$. En nuestros experimentos numéricos y análisis de datos reales la función ρ_1 fue elegida dentro de la familia bicuadrada de Tukey, es decir $\rho_1(r) = \rho_T(r/c_1)$, siendo $\rho_T(r) = \text{mín}(1, 3r^2 - 3r^4 + r^6)$.

3.2.2. Estimación Inicial

El procedimiento propuesto requiere en el **Paso 1** de un estimador inicial del parámetro *nuisance* σ_0 que interviene en la determinación del tamaño de los residuos y por lo tanto en la influencia final que cada observación tendrá en la estimación. Una estimación del parámetro de escala puede ser obtenida mediante un $S-$estimador preliminar.

Los estimadores iniciales podrían calcularse estimando la función η_0 de distintas maneras a partir de la distribución condicional. En nuestro caso nos focalizaremos en una propuesta en la que se computa un $M-$estimador local basado en la distribución condicional y un $S-$estimador para obtener una estimación de la escala presente en el modelo. Para ello definiremos el siguiente procedimiento de tres pasos basado en dos ρ-funciones: ρ y ρ_0.

Para cada valor de $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathcal{S}^1$ y $u \in \mathbb{R}$, llamemos $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}$ a la función de distribución acumulada de $y - \boldsymbol{\beta}^t\mathbf{x}|\boldsymbol{\theta}^t\mathbf{t} = u$ y consideremos la función de distribución empírica condicional definida por

$$F_{n,\boldsymbol{\beta},\boldsymbol{\theta},u}(r) = \hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r) = \sum_{i=1}^{n} \mathbb{I}_{(-\infty,r]}(y_i - \boldsymbol{\beta}^t\mathbf{x}_i) W^h_{\boldsymbol{\theta},i}(u). \tag{3.5}$$

El siguiente procedimiento de tres pasos permite computar un estimador preliminar de escala:

Paso I.1: Para cada valor de $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathcal{S}^1$ y $u \in \mathbb{R}$, computar $\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ como un $M-$estimador de la función de distribución empírica condicional dada en (3.5), es decir

$$\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - a}{s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}\right) W^h_{\boldsymbol{\theta},i}(u) \tag{3.6}$$

siendo $s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ un estimador robusto de la escala local y h una ventana dada.
Paso I.2: Para cada $\boldsymbol{\beta} \in \mathbb{R}^q$ y $\boldsymbol{\theta} \in \mathcal{S}^1$, sea $\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}) > 0$ solución de

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t}_i)}{\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta})}\right) = b\,. \tag{3.7}$$

Paso I.3: El estimador inicial $\widehat{\sigma}$ de σ_0 está dado por

$$\widehat{\sigma} = \tilde{\sigma}\left(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}\right) \quad \text{donde} \quad \left(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}\right) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p,\,\boldsymbol{\theta}\in\mathcal{S}^1}{\operatorname{argmin}} \tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta})\,. \tag{3.8}$$

Cabe notar que $s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ podría ser cualquier estimador de la escala local robusto, sin embargo por simplicidad en nuestros experimentos numéricos y en la aplicación nos hemos focalizado en el uso de la mediana local de las desviaciones absolutas a la mediana local, $MAD(\widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u})$, donde $\widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}$ es la función de distribución empírica condicional definida en (3.5). Con lo cual, de aquí en adelante $s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = MAD(\widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u})$.

Este procedimiento provee de un estimador de la escala σ_0 dado por $\widehat{\sigma}$ y a la vez estimadores iniciales robustos $\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}\right)$ que eventualmente pueden usarse en el procedimiento de estimación de $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ introducido en la sección anterior como punto de partida en el proceso iterativo. Usaremos como valor de la constante $b = 1/2$.

Es usual que se tome $\rho_0(r) = \rho_T(r/c_0)$, donde c_0 es una constante de calibración necesaria para obtener un estimador de escala Fisher consistente. Así, cuando la elección de la función de pérdida corresponde a ρ_T, la elección $c_0 = 1{,}54764$ y $b = 1/2$ resulta en un estimador de escala Fisher–consistente bajo la distribución normal y de punto de ruptura 50 %. Luego, a fin de mejorar la eficiencia del estimador propuesto cuando se utiliza el estimador inicial resultante de los **Pasos I.1** a **I.3**, se usa como constante de calibración del estimador final $c_1 > c_0$, siendo $\rho_1(r) = \rho_T(r/c_1)$.

Observación 3.2.2. Notemos que el estimador de escala preliminar depende también de una ventana y para no hacer más engorrosa la notación, es que no se visibiliza su dependencia. Sin embargo, es importante señalar que al ser un estimador preliminar podría tener un núcleo y una ventana distintos al estimador definido en (3.6).

3.3. Consistencia

3.3.1. Consistencia del estimador propuesto

Asumamos que $\mathbf{t} \in \mathcal{T} \subset \mathbb{R}^q$, luego dado $\mathcal{T}_0 \subset \mathcal{T}$ un conjunto compacto, llamaremos $\mathcal{U}(\mathcal{T}_0)$ al conjunto $\mathcal{U}(\mathcal{T}_0) = \{\boldsymbol{\theta}^t\mathbf{t} : \mathbf{t} \in \mathcal{T}_0, ||\boldsymbol{\theta}|| = 1\}$. Para $v : \mathcal{U}(\mathcal{T}_0) \to \mathbb{R}$ una función continua, notaremos $||v||_{0,\infty} = \sup_{u \in \mathcal{U}(\mathcal{T}_0)} |v(u)|$.

A fin de simplificar la notación denotaremos cuando sea necesario

$$\phi(y, u, s) = \rho_1\left(\frac{y-u}{s}\right)$$

y también llamaremos

$$K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right) .$$

Consideremos los siguientes funcionales relacionados a los estimadores introducidos en la sección anterior. Dados $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathcal{S}^1$, $s \in \mathbb{R}_{>0}$, $a \in \mathbb{R}$ y $u \in \mathbb{R}$ definimos

$$H(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u) = \mathbb{E}\left(\phi\left(y, \boldsymbol{\beta}^t\mathbf{x} + a, s\right) w_1(\mathbf{x}) \,\middle|\, \boldsymbol{\theta}^t\mathbf{t} = u\right), \tag{3.9}$$

y además, si v es una función continua tal que $v : \mathbb{R} \to \mathbb{R}$, llamamos

$$G(\boldsymbol{\beta}, \boldsymbol{\theta}, s, v) = \mathbb{E}\left(\phi\left(y, \boldsymbol{\beta}^t\mathbf{x} + v(\boldsymbol{\theta}^t\mathbf{t}), s\right) w_2(\mathbf{x})\right) . \tag{3.10}$$

Notemos que los estimadores definidos en el Paso 1 y en el 2 quedan determinados por las versiones muestrales de los funcionales definidos más arriba, es decir:

$$H_n(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u) = \sum_{i=1}^{n} \phi\left(y_i, \boldsymbol{\beta}^t \mathbf{x}_i + a, s\right) W^h_{\boldsymbol{\theta},i}(u) w_1(\mathbf{x}_i)\,, \tag{3.11}$$

y

$$G_n(\boldsymbol{\beta}, \boldsymbol{\theta}, s, v) = \frac{1}{n} \sum_{i=1}^{n} \phi\left(y_i, \boldsymbol{\beta}^t \mathbf{x}_i + v(\boldsymbol{\theta}^t \mathbf{t}_i), s\right) w_2(\mathbf{x}_i)\,. \tag{3.12}$$

Consideraremos $\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}}\, H(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)$ y $\eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}}\, H(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_0, a, u)$ de aquí en más. Asumiremos que las funciones de peso y la función de pérdida ρ_1 son tales que vale la Fisher–consistencia de los estimadores definidos a través de los Pasos 1 y 2. En el contexto de un modelo parcialmente lineal generalizado de índice simple con parámetro auxiliar conocido, Boente y Rodriguez (2012) analizan condiciones para garantizar la Fisher–consistencia de sus estimadores y descartan que el parámetro de la componente lineal, $\boldsymbol{\beta}$, pueda contener una intercept, lo que sigue siendo válido en nuestro caso. Más aún, estas autoras notan que cuando la función de pérdida es de la forma $\phi(y, u) = \rho(y - u)$, siendo ρ una $\rho-$función con derivada impar, la Fisher consistencia es válida bajo simetría de los errores. El Lema 3.3.1 a continuación prueba la Fisher–consistencia bajo condiciones de regularidad.

Consideremos las siguientes condiciones.

F1. El error representado por la variable aleatoria $\epsilon \sim F_0$ tiene una densidad $f_0(v)$ que es par, no creciente en $|v|$ y estrictamente creciente en $|v|$ en un entorno de 0.

F2. Dados $(\boldsymbol{\beta}_1, \boldsymbol{\theta}_1) \neq (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, donde $\boldsymbol{\beta}_j \in \mathbb{R}^p$, $\boldsymbol{\theta}_j \in \mathcal{S}^1$ para $j = 0, 1$ y $\eta_1 \not\equiv \eta_0$, tenemos que el conjunto $\mathcal{N} = \left\{(\mathbf{x}, \mathbf{t}) : (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^t \mathbf{x} + \eta_1(\boldsymbol{\theta}_1^t \mathbf{t}) - \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}) = 0\right\}$ cumple $\mathbb{P}(\mathcal{N} \cup \{w_2(\mathbf{x}) = 0\}) < 1$.

Como ocurre en el contexto del modelo lineal, en el modelo parcialmente lineal de índice simple al contemplar funciones ρ_1 redescendientes, para obtener Fisher–consistencia son necesarios supuestos un poco más restrictivos sobre la distribución de los errores. La condición **F1** es necesaria por el uso de funciones de pérdida de tipo redescendientes que son introducidas a fin de controlar residuos grandes y requieren supuestos más fuertes para obtener unicidad, como la simetría y la unimodalidad de la distribución subyacente. Cabe destacar que si a es una constante positiva y ϵ cumple el supuesto **F1**, entonces $a\,\epsilon$ también lo hace. De hecho, si ρ_1 es una $\rho-$función , bajo **F1**, usando el Lema 3.1 de Yohai (1985) se obtiene que si para $a > 0$ consideramos $\lambda(m, a) = \mathbb{E}\{\rho_1[(\epsilon - m)/a]\}$, entonces $0 = \underset{m}{\operatorname{argmin}}\, \lambda(m, a)$.

Respecto de **F2**, cuando w_2 es positiva, se reduce a pedir $\mathbb{P}\big(\{(\mathbf{x}, \mathbf{t}) : \boldsymbol{\beta}_0^t \mathbf{x} + \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}) \neq \boldsymbol{\beta}_1^t \mathbf{x} + \eta_1(\boldsymbol{\theta}_1^t \mathbf{t})\}\big) > 0$, para $(\boldsymbol{\beta}1, \boldsymbol{\theta}_1) \neq (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, $\eta_1 \not\equiv \eta_0$. Esta condición está relacionada a

la identificabilidad de la función de regresión $\mu(\mathbf{x},\mathbf{t}) = \boldsymbol{\beta}^t\mathbf{x} + \eta(\boldsymbol{\theta}^t\mathbf{t})$, donde $\boldsymbol{\theta} \in \mathcal{S}^1$ y cuya primera componente es positiva, y al hecho de que $(\boldsymbol{\beta},\boldsymbol{\theta})$ y η sean únicos. Xia, Tong y Li (1999) en su Teorema 1 deducen condiciones para garantizar esta unicidad. Cuando w_2 es positiva, las condiciones de estos autores traducidas a nuestro contexto corresponden a que $\mathbf{t}$ tiene una densidad positiva sobre un conjunto convexo abierto $\mathcal{T} \subset \mathbb{R}^q$ y además que $\eta_0''(u) \neq 0$ sobre un subintervalo abierto de $u \in \{\boldsymbol{\theta}^t\mathbf{t} : \mathbf{t} \in \mathcal{T}\}$.

Lema 3.3.1. *Consideremos el vector $(y,\mathbf{x},\mathbf{t})$ que cumple el modelo (3.1). Si se cumplen **F1**,**F2**, ρ_1 es una $\rho-$función y w_2 es no negativa, entonces para todo $s > 0$ tenemos que el único mínimo de $G(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta)$ se alcanza en $G(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,s,\eta_0)$ para $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathcal{S}^1$.*

Demostración: Consideremos $(\boldsymbol{\beta}_1,\boldsymbol{\theta}_1) \neq (\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$, donde $\boldsymbol{\beta}_j \in \mathbb{R}^p$, $\boldsymbol{\theta}_j \in \mathcal{S}^1$ para $j = 0,1$ y $\eta_1 \not\equiv \eta_0$. Tomemos el conjunto $\mathcal{N} = \left\{(\mathbf{x},\mathbf{t}) : (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^t\mathbf{x} + \eta_1(\boldsymbol{\theta}_1^t\mathbf{t}) - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}) = 0\right\}$.

Sumando y restando $\boldsymbol{\beta}_0^t\mathbf{x} + \eta_0(\boldsymbol{\theta}_0^t\mathbf{t})$, obtenemos

$$\begin{aligned} G(\boldsymbol{\beta}_1,\boldsymbol{\theta}_1,s,\eta_1) &= \mathbb{E}\left(\phi\left(y,\boldsymbol{\beta}_1^t\mathbf{x} + \eta_1(\boldsymbol{\theta}_1^t\mathbf{t}),s\right)w_2(\mathbf{x})\right) = \mathbb{E}\left[\rho_1\left(\frac{y - \boldsymbol{\beta}_1^t\mathbf{x} - \eta_1(\boldsymbol{\theta}_1^t\mathbf{t}))}{s}\right)w_2(\mathbf{x})\right] \\ &= \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s} - \frac{(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)^t\mathbf{x} + \eta_1(\boldsymbol{\theta}_1^t\mathbf{t}) - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t})}{s}\right)w_2(\mathbf{x})\right] \\ &= \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s} - \frac{\mu_1(\mathbf{x},\mathbf{t}) - \mu_0(\mathbf{x},\mathbf{t})}{s}\right)w_2(\mathbf{x})\left(\mathbb{I}_{\mathcal{N}}(\mathbf{x},\mathbf{t}) + \mathbb{I}_{\mathcal{N}^c}(\mathbf{x},\mathbf{t})\right)\right], \end{aligned}$$

donde hemos llamado $\mu_j(\mathbf{x},\mathbf{t}) = \boldsymbol{\beta}_j^t\mathbf{x} + \eta_j(\boldsymbol{\theta}_j^t\mathbf{t})$, $j = 0,1$. Luego, si llamamos $A(\mathbf{x},\mathbf{t}) = (\mu_1(\mathbf{x},\mathbf{t}) - \mu_0(\mathbf{x},\mathbf{t}))/s$, tomando probabilidad condicional a $(\mathbf{x},\mathbf{t})$, sigue por la independencia entre ϵ y las covariables que

$$\begin{aligned} G(\boldsymbol{\beta}_1,\boldsymbol{\theta}_1,s,\eta_1) &= \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s}\right)\right]\mathbb{E}\left[w_2(\mathbf{x})\mathbb{I}_{\mathcal{N}}(\mathbf{x},\mathbf{t})\right] \\ &+ \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s} - A(\mathbf{x},\mathbf{t})\right)w_2(\mathbf{x})\mathbb{I}_{\mathcal{N}^c}(\mathbf{x},\mathbf{t})\right]. \end{aligned}$$

Notemos que en el segundo término $A(\mathbf{x},\mathbf{t}) \neq 0$, luego en virtud del mencionado Lema 3.1 de Yohai (1985) y la independencia entre el error y el vector de covariables $(\mathbf{x},\mathbf{t})$, tenemos que

$$\mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s} - A(\mathbf{x},\mathbf{t})\right)\middle|(\mathbf{x},\mathbf{t})\right] > \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s}\right)\right].$$

Finalmente, usando esta cota y el hecho de que por **F2** el conjunto $\mathcal{N}^c \cap \{w_2(\mathbf{x}) > 0\}$ tiene probabilidad positiva, obtenemos que

$$G(\boldsymbol{\beta}_1,\boldsymbol{\theta}_1,s,\eta_1) > \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s}\right)\right]\mathbb{E}\left[w_2(\mathbf{x})\mathbb{I}_{\mathcal{N}}(\mathbf{x},\mathbf{t})\right] + \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s}\right)\right]\mathbb{E}\left[w_2(\mathbf{x})\mathbb{I}_{\mathcal{N}^c}(\mathbf{x},\mathbf{t})\right],$$

es decir

$$G(\boldsymbol{\beta}_1, \boldsymbol{\theta}_1, s, \eta_1) \;\; > \;\; \mathbb{E}\left[\rho_1\left(\epsilon\frac{\sigma_0}{s}\right)\right]\mathbb{E}\left[w_2(\mathbf{x})\right] = G(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, s, \eta_0)\,,$$

como queríamos demostrar. ■

Otra propiedad interesante de estudiar es la de invariancia por transformaciones ortogonales en la variable del índice simple. Nos interesa saber cuál es el efecto de considerar un vector transformado $(y, \mathbf{x}, \boldsymbol{\Gamma}\mathbf{t})$, siendo $\boldsymbol{\Gamma}$ una matriz ortogonal en $\mathbb{R}^{q\times q}$, en lugar de las variables originales $(y, \mathbf{x}, \mathbf{t})$.

Lema 3.3.2. *Supongamos que $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ es el único mínimo de $G(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_0, \eta_{\boldsymbol{\beta},\boldsymbol{\theta}})$. Sean $\boldsymbol{\Gamma} \in \mathbb{R}^{q\times q}$ tal que $\boldsymbol{\Gamma}^t\boldsymbol{\Gamma} = \mathbb{I}_q$, siendo $\mathbb{I}_q$ matriz identidad de $q \times q$, $\tilde{\mathbf{t}} = \boldsymbol{\Gamma}\mathbf{t}$ y*

$$\tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_0, \overline{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) = \mathbb{E}\left(\rho_1\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - \overline{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}\left(\boldsymbol{\theta}^t\tilde{\mathbf{t}}\right)}{\sigma_0}\right)w_2(\mathbf{x})\right)\,,$$

donde $\overline{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \underset{a\in\mathbb{R}}{argmin}\,\mathbb{E}\left(\rho_1\left((y - \boldsymbol{\beta}^t\mathbf{x} - a)/\sigma_0\right)w_1(\mathbf{x}) \,\middle|\, \boldsymbol{\theta}^t\tilde{\mathbf{t}} = u\right)$. Luego, si $\left(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{\theta}}_0\right) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\theta}\in\mathcal{S}^1}{argmin}\,\tilde{G}(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_0, \overline{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}})$, entonces $\tilde{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0$ y $\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\Gamma}\boldsymbol{\theta}_0$.

Demostración: Notemos que por la ortogonalidad de la matriz $\boldsymbol{\Gamma}$ y por la definición de $\tilde{t}$ resulta

$$\mathbb{E}\left(\rho_1\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - a}{\sigma_0}\right)w_1(\mathbf{x}) \,\middle|\, \boldsymbol{\theta}^t\tilde{\mathbf{t}} = u\right) = \mathbb{E}\left(\rho_1\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - a}{\sigma_0}\right)w_1(\mathbf{x}) \,\middle|\, (\boldsymbol{\Gamma}^t\boldsymbol{\theta})^t\mathbf{t} = u\right)\,,$$

por lo que es inmediato que $\overline{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}} = \eta_{\boldsymbol{\beta},\boldsymbol{\Gamma}^t\boldsymbol{\theta}}$. Esto permite deducir que $\tilde{G}\left(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{\theta}}_0, \sigma_0, \overline{\eta}_{\tilde{\boldsymbol{\beta}}_0,\tilde{\boldsymbol{\theta}}_0}\right) = G\left(\tilde{\boldsymbol{\beta}}_0, \boldsymbol{\Gamma}^t\tilde{\boldsymbol{\theta}}_0, \sigma_0, \eta_{\tilde{\boldsymbol{\beta}}_0,\boldsymbol{\Gamma}^t\tilde{\boldsymbol{\theta}}_0}\right)$. Por otro lado, se tiene que

$$\begin{aligned}
\tilde{G}\left(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{\theta}}_0, \sigma_0, \overline{\eta}_{\tilde{\boldsymbol{\beta}}_0,\tilde{\boldsymbol{\theta}}_0}\right) &= \min_{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\theta}\in\mathcal{S}^1}\mathbb{E}\left(\rho_1\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - \overline{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}\left(\boldsymbol{\theta}^t\tilde{\mathbf{t}}\right)}{\sigma_0}\right)w_2(\mathbf{x})\right)\\
&= \min_{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\theta}\in\mathcal{S}^1}\mathbb{E}\left(\rho_1\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - \eta_{\boldsymbol{\beta},\Gamma^t\boldsymbol{\theta}}\left(\left(\Gamma^t\boldsymbol{\theta}\right)^t\mathbf{t}\right)}{\sigma_0}\right)w_2(\mathbf{x})\right)\\
&= \min_{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\theta}\in\mathcal{S}^1}\mathbb{E}\left(\rho_1\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - \eta_{\boldsymbol{\beta},\boldsymbol{\theta}}\left(\boldsymbol{\theta}^t\mathbf{t}\right)}{\sigma_0}\right)w_2(\mathbf{x})\right)\\
&= G(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0, \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0})\,.
\end{aligned}$$

Finalmente, $G\left(\tilde{\boldsymbol{\beta}}_0, \boldsymbol{\Gamma}^t\tilde{\boldsymbol{\theta}}_0, \sigma_0, \eta_{\tilde{\boldsymbol{\beta}}_0,\boldsymbol{\Gamma}^t\tilde{\boldsymbol{\theta}}_0}\right) = G(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0, \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0})$ y sabiendo que $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ es el único vector que realiza el mínimo, concluímos que $\tilde{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0$ y $\boldsymbol{\Gamma}^t\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_0$.■

Notemos que si llamamos $F_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u}$ a la función de distribución de $y - \boldsymbol{\beta}_0^t\mathbf{x}|\boldsymbol{\theta}_0^t\mathbf{t} = u$, entonces $F_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u}(r) = F_0((r - \eta_0(u))/\sigma_0)$, siendo F_0 la distribución del error. Luego, como bajo **F1**, F_0 es simétrica alrededor de 0 y con densidad f_0 unimodal, si ρ_1 es una $\rho-$función , entonces $\eta_0(u) = \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u)$.

Probaremos la consistencia de los estimadores propuestos a través de los Pasos 1 y 2 bajo el siguiente conjunto de condiciones.

C1. El núcleo $K : \mathbb{R} \to \mathbb{R}$ es una función par, no negativa, acotada, continua, de variación acotada y satisface $\int K(u)du = 1$, $\int u^2K(u)du < \infty$ y $\lim_{|u|\to\infty} |u|K(u) = 0$.

C2. Dada una sucesión de ventanas $h = h_n$, se verifica que $\lim_{n\to\infty} h = 0$ y que $\lim_{n\to\infty} nh/\log(n) = \infty$.

C3. i) La densidad marginal de $\mathbf{t}$, $f_{\mathbf{t}}$, es acotada en $\mathcal{T}$.

ii) Para cada $\boldsymbol{\theta} \in \mathcal{S}^1$, $f_{\boldsymbol{\theta}}$, la densidad marginal de $\boldsymbol{\theta}^t\mathbf{t}$, es acotada. Dado un conjunto compacto $\mathcal{T}_0 \subset \mathcal{T}$, existe una constante positiva $A_1(\mathcal{U}(\mathcal{T}_0))$ tal que para todo $u \in \mathcal{U}(\mathcal{T}_0), \boldsymbol{\theta} \in \mathcal{S}^1$, se tiene que $f_{\boldsymbol{\theta}}(u) > A_1(\mathcal{U}(\mathcal{T}_0))$.

C4. ρ_1 es una función continua. Las funciones de peso w_1 y w_2 son no negativas y acotadas.

C5. $H(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)$ satisface la siguiente condición de equicontinuidad: dados los conjuntos compactos $\mathcal{K} \subset \mathbb{R}^p$, $\mathcal{T}_0 \subset \mathcal{T}$, $\mathcal{S} \subset \mathbb{R}_{>0}$, para cualquier $\varepsilon > 0$ existe $\delta > 0$ tal que si $||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|| < \delta$, $||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| < \delta$, $|s_1 - s_2| < \delta$ y $|u_1 - u_2| < \delta$ donde $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{K}$, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{S}^1$, $s_1, s_2 \in \mathcal{S}$ y $u_1, u_2 \in \mathcal{U}(\mathcal{T}_0)$ entonces

$$\sup_{a\in\mathbb{R}} |H(\boldsymbol{\beta}_1, \boldsymbol{\theta}_1, s_1, a, u_1) - H(\boldsymbol{\beta}_2, \boldsymbol{\theta}_2, s_2, a, u_2)| < \varepsilon\,.$$

C6. $H(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)$ es una función continua en $(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)$ y $\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u)$ es una función continua en $(\boldsymbol{\beta}, \boldsymbol{\theta}, s, u)$.

C7. ρ_1 es una función con derivada acotada ψ_1.

El conjunto de hipótesis **C1** a **C7** son habituales en el marco de modelos semiparamétricos y también en el de robustez, ver Severini y Staniswalis (1994), Carroll et al. (1997) y Boente y Rodriguez (2012). Los supuestos **C1** y **C2** sobre el núcleo y la ventana son los clásicos

en el contexto no paramétrico. Elecciones tales como el núcleo de Epanechnikov dado por $K(z) = 3/4\ (1-z^2)\ \mathbb{I}_{[-1,1]}(z)$ y el gaussiano, cumplen **C1**. Por otro lado, **C2** se satisface, por ejemplo, si $h = n^{-c}$ para $c \in (0,1)$. Respecto de **C3**, esta hipótesis es habitual en modelos de índice simple. Como consecuencia de i), tenemos que $\sup_{\boldsymbol{\theta}\in\mathcal{S}^1} ||f_{\boldsymbol{\theta}}||_\infty < \infty$. Por otro lado, si $f_{\mathbf{t}}(\underline{\mathbf{t}}) > B_{\mathcal{T}_0}$ para $\underline{\mathbf{t}} \in \mathcal{T}_0$, entonces vale ii). Las condiciones **C4** y **C7** son satisfechas por las funciones de pérdida que suelen usarse en el contexto de robustez, como por ejemplo la función bicuadrada de Tukey, ρ_T, que ya hemos presentado y que garantiza que $\phi(y,u,s)$ es continua y acotada. Si $(y,\mathbf{x})|\boldsymbol{\theta}^t\mathbf{t} = u$ tiene distribución continua respecto de u, $H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)$ resulta continua. Boente y Rodriguez (2012) observan que cuando la distribución de $\mathbf{x}|\boldsymbol{\theta}^t\mathbf{t} = u$ es continua respecto de $(\boldsymbol{\theta},u)$ y ρ_1 es continuamente diferenciable, el Teorema de la Función Implícita implica que $\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u)$ es continua, cumpliéndose **C6**.

El siguiente lema es una generalización del Teorema 3.1 de Boente, He y Zhou (2006).

Teorema 3.3.3. *Sean $\mathcal{K} \subset \mathbb{R}^p$, $\mathcal{S} \subset \mathbb{R}_{>0}$ conjuntos compactos y $\mathcal{T}_0 \subset \mathcal{T}$ un conjunto compacto tal que existe $\delta_0 > 0$ para el cual $\mathcal{T}_{\delta_0} \subset \mathcal{T}$, donde $\mathcal{T}_{\delta_0}$ es la clausura de una δ_0-vecindad de $\mathcal{T}_0$. Supongamos que se satisfacen las condiciones **C1**-**C6** y se cumple la siguiente condición*

(C) la familia de funciones $\mathcal{F} = \{f(y,\mathbf{x}) = \phi\left(y, \boldsymbol{\beta}^t\mathbf{x} + a, s\right) w_1(\mathbf{x}), \boldsymbol{\beta} \in \mathcal{K}, s \in \mathcal{S}, a \in \mathbb{R}\}$ tiene un número de cubrimiento tal que $\sup_{\mathcal{Q}} N(\varepsilon,\mathcal{F},L^1(\mathcal{Q})) \le A\varepsilon^{-W}$, para $0<\varepsilon<1$, $\mathcal{Q}$ una medida de probabilidad para $(y,\mathbf{x})$ y $A, W > 0$.

Luego, tenemos que

a) $$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},a\in\mathbb{R}} ||H_n(\boldsymbol{\beta},\boldsymbol{\theta},s,a,\cdot) - H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,\cdot)|| \xrightarrow{c.t.p.} 0\,.$$

b) Si $\inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0),s\in\mathcal{S}} \left[\lim_{|a|\to\infty} H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u) - H(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u),u)\right] > 0$ *y $H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)$ tiene un único mínimo en $\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u)$, entonces*

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S}} ||\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta},s} - \eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}||_{0,\infty} \xrightarrow{c.t.p.} 0\,, \tag{3.13}$$

donde $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u) = \underset{a}{argmin}\, H_n(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)$.

Demostración:

a) Usando la notación que introdujimos al principio del capítulo, sean

$$H_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u) = \frac{1}{n}\sum_{i=1}^n K_h(u - \boldsymbol{\theta}^t\mathbf{t}_i)\phi\left(y_i, \boldsymbol{\beta}^t\mathbf{x}_i + a, s\right) w_1(\mathbf{x}_i)$$

$$F_{0n}(\boldsymbol{\theta},u) = \frac{1}{n}\sum_{i=1}^n K_h\left(u - \boldsymbol{\theta}^t\mathbf{t}_i\right)\,.$$

Entonces $H_n(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u) = \dfrac{H_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)}{F_{0n}(\boldsymbol{\theta}, u)}$. Luego,

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},a\in\mathbb{R}} ||H_n(\boldsymbol{\beta}, \boldsymbol{\theta},s, a, \cdot) - H(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot)||_{0,\infty} \leq$$

$$\left[\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},a\in\mathbb{R}} ||H_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot) - \mathbb{E}(H_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot))|_{0,\infty}\right.$$

$$+ \sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},a\in\mathbb{R}} ||\mathbb{E}(H_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot)) - H(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot)\mathbb{E}(F_{0n}(\boldsymbol{\theta}, \cdot))||_{0,\infty}$$

$$\left. + ||\rho_1||_\infty ||w_1||_\infty \sup_{\boldsymbol{\theta}\in\mathcal{S}^1} ||F_{0n}(\boldsymbol{\theta}, \cdot) - \mathbb{E}(F_{0n}(\boldsymbol{\theta}, \cdot))||_{0,\infty}\right] \left[\inf_{\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} F_{0n}(\boldsymbol{\theta}, u)\right]^{-1}.$$

En cuanto al denominador tengamos en cuenta que para n suficientemente grande, se cumple que

$$\inf_{\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} F_{0n}(\boldsymbol{\theta}, u) \geq \inf_{\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} \mathbb{E}(F_{0n}(\boldsymbol{\theta}, u)) - \sup_{\boldsymbol{\theta}\in\mathcal{S}^1} ||F_{0n}(\boldsymbol{\theta}, \cdot) - \mathbb{E}(F_{0n}(\boldsymbol{\theta}, \cdot))||_{0,\infty}\,, \quad (3.14)$$

donde

$$\mathbb{E}(F_{0n}(\boldsymbol{\theta}, u)) = \frac{1}{h}\int K\left(\frac{v-u}{h}\right) f_{\boldsymbol{\theta}}(v)\, dv = \int K(w) f_{\boldsymbol{\theta}}(hw+u)\, dw\,.$$

Dado $\delta < \delta_0$, sea $\mathcal{U}_\delta = \{u+s : u \in \mathcal{U}(\mathcal{T}_0), ||s|| \leq \delta\}$ y $M > 0$ de manera que $\int_{|w|\leq M} K(w)dw > 1/2$, luego considerando $h \leq \delta/M$ se tiene que $hw + u \in \mathcal{U}_\delta$. Como consecuencia de la compacidad de $\mathcal{U}_\delta$ y de la condición **C3**, resulta que para todo $\boldsymbol{\theta} \in \mathcal{S}^1, u \in \mathcal{U}(\mathcal{T}_0)$

$$\mathbb{E}(F_{0n}(\boldsymbol{\theta}, u)) > \frac{1}{2} A_1(\mathcal{U}_\delta)\,. \quad (3.15)$$

Por lo tanto, es suficiente demostrar que

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},a\in\mathbb{R}} ||H_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot) - \mathbb{E}(H_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot))||_{0,\infty} \xrightarrow{c.t.p.} 0\,, \quad (3.16)$$

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},a\in\mathbb{R}} ||\mathbb{E}(H_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot)) - H(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot)\mathbb{E}(F_{0n}(\boldsymbol{\theta}, \cdot))||_{0,\infty} \to 0\,, \quad (3.17)$$

$$\sup_{\boldsymbol{\theta}\in\mathcal{S}^1} ||F_{0n}(\boldsymbol{\theta}, \cdot) - \mathbb{E}(F_{0n}(\boldsymbol{\theta}, \cdot))||_{0,\infty} \xrightarrow{c.t.p.} 0\,. \quad (3.18)$$

La convergencia de (3.18) se obtiene a partir de las condiciones sobre el núcleo, la sucesión de ventanas y la densidad marginal $f_{\boldsymbol{\theta}}$ requeridas de **C1** a **C3** y el Teorema 37 de Pollard (1984).

Respecto de (3.17), tenemos que para $\tilde{u} = \boldsymbol{\theta}^t \mathbf{t}$ vale que

$$\begin{aligned} |\mathbb{E}(H_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)) &- H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)\mathbb{E}(F_{0n}(\boldsymbol{\theta},u))| \\ &= \left|\frac{1}{h}\mathbb{E}\left(K\left(\frac{\tilde{u}-u}{h}\right)[H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,\tilde{u}) - H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)]\right)\right| \\ &= \left|\int \frac{1}{h}K\left(\frac{r-u}{h}\right)[H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,r) - H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)]f_{\boldsymbol{\theta}}(r)\,dr\right| \\ &\leq \|f_{\boldsymbol{\theta}}\|_\infty \int K(w)\,|H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u-wh) - H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)|\,dw\,, \end{aligned}$$

donde la última desigualdad se obtiene después del cambio de variables $w = (r-u)/h$. Por la condición de integrabilidad del núcleo requerida en **C1**, el hecho de que **C3** i) implica que $\sup_{\boldsymbol{\theta}\in\mathcal{S}^1}\|f_{\boldsymbol{\theta}}\|_\infty < \infty$, la condición de equicontinuidad de **C5** y el hecho de que $h \to 0$, obtenemos la convergencia de (3.17).

Probaremos ahora (3.16). Para ello consideremos la clase de funciones

$$\begin{aligned} \mathcal{G}_n = \Bigg\{ & g_{v,\boldsymbol{\beta},\boldsymbol{\theta},s,a,h}(y,\mathbf{x},\mathbf{t}) = \mathcal{C}\,\phi(y,\mathbf{x}^t\boldsymbol{\beta}+a,s)w_1(\mathbf{x})K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}-v}{h}\right), \boldsymbol{\beta}\in\mathcal{K}, s\in\mathcal{S}, \\ & a\in\mathbb{R}, \boldsymbol{\theta}\in\mathcal{S}^1, v\in\mathcal{U}(\mathcal{T}_0)\Big\}\,, \end{aligned}$$

donde $\mathcal{C} = 1/(\|\rho_1\|_\infty\|w_1\|_\infty\|K\|_\infty)$. Dado que K es de variación acotada, que $0 \leq K\left((\boldsymbol{\theta}^t\mathbf{t}-v)/h\right) \leq \|K\|_\infty$ y que por hipótesis la familia $\mathcal{F}$ satisface la condición (C) del enunciado, para $0 < \varepsilon < 1$ resulta que $\sup_{\mathcal{Q}} N(\varepsilon,\mathcal{G}_n,L^1(\mathcal{Q})) \leq A_1\varepsilon^{-W_1}$, con $W_1 > 0$ y A_1 independientes de n para cualquier medida de probabilidad $\mathcal{Q}$. Notemos que $|g_{v,\boldsymbol{\beta},\boldsymbol{\theta},s,a,h}| \leq 1$ y

$$\mathbb{E}(g^2_{v,\boldsymbol{\beta},\boldsymbol{\theta},s,a,h}(y,\mathbf{x},\mathbf{t})) \leq \frac{h}{\|K\|_\infty}\sup_{\boldsymbol{\theta}\in\mathcal{S}^1}\|f_{\boldsymbol{\theta}}\|_\infty = \mathcal{D}h$$

ya que

$$\begin{aligned} \mathbb{E}(g^2_{v,\boldsymbol{\beta},\boldsymbol{\theta},s,a,h}(y,\mathbf{x},\mathbf{t})) &\leq \frac{1}{\|K\|^2_\infty}\int K^2\left(\frac{u-v}{h}\right)f_{\boldsymbol{\theta}}(u)\,du \\ &= \frac{1}{\|K\|^2_\infty}\int K^2(w)f_{\boldsymbol{\theta}}(hw+v)h\,dw \leq \frac{h}{\|K\|_\infty}\,||f_{\boldsymbol{\theta}}||_\infty\int K(w)dw\,. \end{aligned}$$

Luego, por el Teorema 37 de Pollard (1984) tomando $\delta_n^2 = \mathcal{D}h$ y $\alpha_n = 1$, sabiendo que se cumple **C2**, tenemos que

$$\frac{1}{h}\sup_{\mathcal{G}_n}\left|\frac{1}{n}\sum_{i=1}^n g_{v,\boldsymbol{\beta},\boldsymbol{\theta},s,a,h}(y_i,\mathbf{x}_i,\mathbf{t}_i) - \mathbb{E}(g_{v,\boldsymbol{\beta},\boldsymbol{\theta},s,a,h}(y_1,\mathbf{x}_1,\mathbf{t}_1))\right| \xrightarrow{c.t.p.} 0$$

de donde resulta (3.16).

b) La continuidad de $\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u)$ implica que es acotada sobre $\mathcal{K}\times\mathcal{S}^1\times\mathcal{S}\times\mathcal{U}(\mathcal{T}_0)$ y que existe $\mathcal{B}=\mathcal{B}(\mathcal{K},\mathcal{S}^1,\mathcal{S},\mathcal{U}(\mathcal{T}_0))$, conjunto compacto, tal que $\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u)\in\mathcal{B}$ si $\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S}$ y $u\in\mathcal{U}(\mathcal{T}_0)$.

Supongamos que $\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S}}||\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta},s}-\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}||_{0,\infty}$ no converge a 0 en un conjunto Ω_0 tal que $\mathbb{P}(\Omega_0)>0$. Luego para cada $\omega\in\Omega_0$, existe una sucesión $\{(\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k,u_k)\}_{k\in\mathbb{N}}\subset\mathcal{K}\times\mathcal{S}^1\times\mathcal{S}\times\mathcal{U}(\mathcal{T}_0)$ tal que $\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)-\eta_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)\to c\neq 0$. Dado que $\mathcal{K}$, $\mathcal{S}^1$, $\mathcal{S}$ y $\mathcal{U}(\mathcal{T}_0)$ son compactos, sin pérdida de generalidad podemos suponer que $(\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k,u_k)\to(\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L,u_L)\in\mathcal{K}\times\mathcal{S}^1\times\mathcal{S}\times\mathcal{U}(\mathcal{T}_0)$. Así, $\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)-\eta_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)\to c\neq 0$ y por la continuidad resulta que $\eta_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)\to\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L)$. En consecuencia, $\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)-\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L)\to c$.

En primera instancia, supongamos que $c<\infty$. Dado $\varepsilon>0$, para n y k suficientemente grandes, como se cumple la hipótesis **C5** y por la continuidad tenemos que

$$H(\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L,\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L)+c,u_L)<H(\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k,\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k),u_k)+\frac{\varepsilon}{2}\,,$$

y además por el ítem *a*) y la definición de $\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)$, sigue que

$$\begin{aligned}
H(\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k,\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k),u_k) &< H_n(\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k,\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k),u_k)+\frac{\varepsilon}{8}\\
&< H_n(\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k,\widehat{\eta}_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L),u_k)+\frac{\varepsilon}{8}\\
&< H(\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k,\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L),u_k)+\frac{\varepsilon}{4}\\
&< H(\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L,\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L),u_L)+\frac{\varepsilon}{2}
\end{aligned}$$

Por lo tanto, concluímos que

$$H(\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L,\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L)+c,u_L)<H(\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L,\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L),u_L)+\varepsilon$$

contradiciendo la unicidad dada en *b*) a menos que $c=0$.

Supongamos ahora que $c=\infty$ y por lo tanto, $\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k)\to\infty$. Llamemos

$$\begin{aligned}
\mathcal{I} &= \inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},u\in\mathcal{U}(\mathcal{T}_0)}\left[\lim_{|a|\to\infty}H(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)-H(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u),u)\right]\\
R_1 &= H(\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L,\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k),u_L)\\
R_2 &= H(\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L,\eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L),u_L)\,.
\end{aligned}$$

Por hipótesis en b) tenemos que $\mathcal{I} > 0$. Luego,

$$\lim_{|a|\to\infty} H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, a, u_L) - H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_L) \geq \mathcal{I}\,.$$

En consecuencia, para k suficientemente grande se cumple que

$$R_1 > R_2 + \mathcal{I}/2\,. \tag{3.19}$$

Debido a la condición de equicontinuidad **C5**, dado $\varepsilon > 0$ para k suficientemente grande

$$H(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_k, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_k) \leq H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_L) + \frac{\varepsilon}{4}$$

y además

$$H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k), u_L) \leq H(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_k, \widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k), u_k) + \frac{\varepsilon}{4}.$$

Usando a) y la definición de $\widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}$, obtenemos que

$$\begin{aligned} H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k), u_L) &\leq H_n(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_k, \widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k), u_k) + \frac{\varepsilon}{2} \\ &\leq H_n(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_k, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_k) + \frac{\varepsilon}{2}\,. \end{aligned}$$

Nuevamente, usando a) resulta que

$$\begin{aligned} H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k), u_L) &\leq H_n(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_k, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_k) + \frac{\varepsilon}{2} \\ &\leq H(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_k, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_k) + \frac{3\varepsilon}{4} \\ &\leq H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_L) + \varepsilon \end{aligned}$$

Resumiendo, para k suficientemente grande

$$H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \widehat{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k,s_k}(u_k), u_L) \leq H(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, s_L, \eta_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L,s_L}(u_L), u_L) + \varepsilon\,,$$

lo que contradice (3.19). ■

El siguiente teorema prueba la consistencia fuerte de los estimadores propuestos a partir de la implementación de los Pasos 1 y 2.

Cabe destacar que el supuesto de que el estimador de la componente lineal, $\widehat{\boldsymbol{\beta}}$, yace últimamente en un compacto, presente en dicho teorema, es una condición también requerida en la bibliografía de estimación robusta en modelos semiparamétricos, tal como en Boente, He y Zhou (2006) para modelos parcialmente lineales generalizados y en Boente y Rodriguez (2012) para modelos parcialmente lineales generalizados de índice simple.

Teorema 3.3.4. *Sean $\mathcal{K} \subset \mathbb{R}^p$ y $\mathcal{S} \subset \mathbb{R}_{>0}$ conjuntos compactos. Sean $\widehat{\sigma}$ un estimador consistente de σ_0 y $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right)$ el argumento que minimiza $G_n\big(\boldsymbol{\beta}, \boldsymbol{\theta}, \widehat{\sigma}, \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta},\widehat{\sigma}}\big)$ donde $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta},s}(u)$ satisface*

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S}} ||\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta},s} - \eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}||_{0,\infty} \xrightarrow{c.t.p.} 0\,. \tag{3.20}$$

*Si se cumplen **C4**, **C6** y **C7**, entonces*

a) $\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{K},\boldsymbol{\theta},\mathbf{a}\in\mathcal{S}^1,s\in\mathcal{S}} |G_n(\boldsymbol{\beta},\boldsymbol{\theta},s,\widehat{\eta}_{\mathbf{b},\mathbf{a},s}) - G(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\mathbf{b},\mathbf{a},s})| \xrightarrow{c.t.p.} 0.$

b) Si existe un conjunto compacto $\mathcal{K}_1 \subset \mathbb{R}^p$ tal que $\lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n\geq m} \widehat{\boldsymbol{\beta}} \in \mathcal{K}_1\right) = 1$ y $G\big(\boldsymbol{\beta},\boldsymbol{\theta},\sigma_0,\eta_{\boldsymbol{\beta},\boldsymbol{\theta},\sigma_0}\big)$ tiene un único mínimo en $(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$, entonces $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}) \xrightarrow{c.t.p.} (\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$.

Demostración:

a) Notemos que

$$\begin{aligned} |G_n(\boldsymbol{\beta},\boldsymbol{\theta},s,\widehat{\eta}_{\mathbf{b},\mathbf{a},s}) - G(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\mathbf{b},\mathbf{a},s})| &\leq |G_n(\boldsymbol{\beta},\boldsymbol{\theta},s,\widehat{\eta}_{\mathbf{b},\mathbf{a},s}) - G_n(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\mathbf{b},\mathbf{a},s})| \\ &+ |G_n(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\mathbf{b},\mathbf{a},s}) - G(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\mathbf{b},\mathbf{a},s})| \\ &= |G_{1n}| + |G_{2n}| \end{aligned}$$

Comenzamos por acotar el supremo del primer término. Sean $\varepsilon > 0$ y $\mathcal{T}_0$ un compacto tal que $\mathbb{P}(\mathbf{t} \notin \mathcal{T}_0) < \varepsilon$, tenemos que

$$\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{K},\boldsymbol{\theta},\mathbf{a}\in\mathcal{S}^1,s\in\mathcal{S}} |G_{1n}| \leq \mathcal{C}_1 \sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{K},\boldsymbol{\theta},\mathbf{a}\in\mathcal{S}^1,s\in\mathcal{S}} ||\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta},s} - \eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}||_{0,\infty} + \mathcal{C}_2\, \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\mathcal{T}_0^c}(\mathbf{t}_i)\,,$$

donde $\mathcal{C}_1 = ||\psi_1||_\infty ||w_2||_\infty / \min(\mathcal{S})$ y $\mathcal{C}_2 = 2||\rho_1||_\infty ||w_2||_\infty$. Por lo tanto, usando (3.20) y la Ley Fuerte de los Grandes Números, obtenemos que $\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{K},\boldsymbol{\theta},\mathbf{a}\in\mathcal{S}^1,s\in\mathcal{S}} |G_{1n}| \xrightarrow{c.t.p.} 0\,.$

Ahora sólo resta probar que $\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{K},\boldsymbol{\theta},\mathbf{a}\in\mathcal{S}^1,s\in\mathcal{S}} |G_{2n}| \xrightarrow{c.t.p.} 0$. Por la compacidad de $\mathcal{K}$, $\mathcal{S}$ y $\mathcal{S}^1$, **C4** y la continuidad de $\eta_{\mathbf{b},\mathbf{a},s}$ dada en **C6**, usando el Teorema 3 del Capítulo II de Pollard (1984) y argumentos similares a los empleados en el Lema 1 de Bianco y Boente (2002), obtenemos que

$$\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{K},\boldsymbol{\theta},\mathbf{a}\in\mathcal{S}^1,s\in\mathcal{S}} |G_n(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\mathbf{b},\mathbf{a},s}) - G(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta_{\mathbf{b},\mathbf{a},s})| \xrightarrow{c.t.p.} 0\,,$$

con lo que concluye la prueba de *a)*.

b) Como $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right)$ yace últimamente en un compacto, existe una subsucesión convergente $\left(\widehat{\boldsymbol{\beta}}_k, \widehat{\boldsymbol{\theta}}_k\right)$ a $\left(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*\right) \in \mathcal{K}_1 \times \mathcal{S}^1$, es decir, $\left(\widehat{\boldsymbol{\beta}}_k, \widehat{\boldsymbol{\theta}}_k\right) \xrightarrow{c.t.p.} \left(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*\right)$. A fin de aliviar la notación y sin pérdida de generalidad, asumamos que $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right) \xrightarrow{c.t.p.} \left(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*\right)$.

Por la consistencia del estimador de escala, la continuidad de $\eta_{\boldsymbol{\beta},\boldsymbol{\theta},s}$ y de G, usando el ítem *a*) de este Teorema, tenemos que

$$G_n\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}, \widehat{\sigma}, \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{\sigma}}\right) - G(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*, \sigma_0, \eta_{\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,\sigma_0}) \xrightarrow{c.t.p.} 0$$

y también que

$$G_n\left(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \widehat{\sigma}, \widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\widehat{\sigma}}\right) - G(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0, \eta_0) \xrightarrow{c.t.p.} 0\,,$$

teniendo en cuenta que $\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\sigma_0} = \eta_0$.

Dado que $G_n\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}, \widehat{\sigma}, \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}},\widehat{\sigma}}\right) \leq G_n\left(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \widehat{\sigma}, \widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\widehat{\sigma}}\right)$ y que $G\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_0, \eta_{\boldsymbol{\beta},\boldsymbol{\theta},\sigma_0}\right)$ tiene un único mínimo en $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, resulta *b*). ■

3.3.2. Sobre el estimador inicial

Consideremos el siguiente funcional relacionado al **Paso I.1** del estimador inicial

$$M(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u) = \mathbb{E}\left(\rho\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - a}{s}\right) \,\middle|\, \boldsymbol{\theta}^t\mathbf{t} = u\right), \tag{3.21}$$

siendo $s > 0$. Llamaremos $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \operatorname{argmin} M(\boldsymbol{\beta}, \boldsymbol{\theta}, S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), a, u)$ donde $S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ es una medida de escala robusta respecto de la distribución de $y - \boldsymbol{\beta}^t\mathbf{x}$ condicional a $\boldsymbol{\theta}^t\mathbf{t} = u$, cuya función de distribución denotamos $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}$.

Tomaremos como escala robusta condicional a la mediana de los desvíos absolutos respecto de la mediana condicional, es decir respecto de la mediana de la distribución condicional $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}$, más precisamente

$$S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = MAD(F_{\boldsymbol{\beta},\boldsymbol{\theta},u}) = \text{mediana}\left(|y - \boldsymbol{\beta}^t\mathbf{x} - m_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)| \,\middle|\, \boldsymbol{\theta}^t\mathbf{t} = u\right)$$

siendo $m_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ la mediana respecto de $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}$.

Notemos que si consideramos la versión empírica de $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}$, es decir $\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}$ definida en (3.5), el estimador inicial definido en el **Paso I.1** involucra a la versión muestral de estos dos funcionales. Para favorecer la comprensión y aliviar la notación, es que utilizaremos $\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}$ sin hacer explícita su dependencia con n, el tamaño de la muestra aleatoria.

Asociado al funcional M introducimos para $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathcal{S}^1$, $s > 0$ el funcional

$$\Lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u) = \mathbb{E}\left(\psi\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - a}{s}\right) \,\middle|\, \boldsymbol{\theta}^t\mathbf{t} = u\right), \tag{3.22}$$

donde $\psi(t) = \rho'(t)$. Consideramos $\Lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, S_{\boldsymbol{\beta},\boldsymbol{\theta}}, a, u) = 0$ y su única solución $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$.

Trabajaremos con el siguiente conjunto de hipótesis:

I1. $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)$ es una función continua de $(\boldsymbol{\beta},\boldsymbol{\theta},u)$, simétrica alrededor de $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ y con función de densidad unimodal para todo $\boldsymbol{\beta}$ y $\boldsymbol{\theta}$.

I2. Sean $\mathcal{K} \subset \mathbb{R}^p$ y $\mathcal{T}_0 \subset \mathcal{T}$ conjuntos compactos. Entonces,

i) se verifica la siguiente condición: para todo $\varepsilon > 0$, existe $\delta > 0$ tal que si $|r_1 - r_2| < \delta$, se tiene que

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} \sup_{u\in\mathcal{U}(\mathcal{T}_0)} |F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r_1) - F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r_2)| < \varepsilon\,.$$

ii) se verifica la siguiente condición de equicontinuidad: para todo $\varepsilon > 0$, existe $\delta > 0$ tal que si
$|u_1 - u_2| < \delta$, entonces para todo $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$ se tiene que

$$\sup_{r\in\mathbb{R}} |F_{\boldsymbol{\beta},\boldsymbol{\theta},u_1}(r) - F_{\boldsymbol{\beta},\boldsymbol{\theta},u_2}(r)| < \varepsilon\,.$$

iii) para cada r fijo se verifica la siguiente condición de continuidad uniforme:
para todo $\varepsilon > 0$, existe $\delta > 0$ tal que si $|u_1 - u_2| < \delta$, $\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| < \delta$, $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < \delta$, $\boldsymbol{\beta}_i \in \mathcal{K}$, $\boldsymbol{\theta}_i \in \mathcal{S}^1$, $u_i \in \mathcal{U}(\mathcal{T}_0)$, se tiene que

$$|F_{\boldsymbol{\beta}_1,\boldsymbol{\theta}_1,u_1}(r) - F_{\boldsymbol{\beta}_2,\boldsymbol{\theta}_2,u_2}(r)| < \varepsilon\,.$$

I3. $\psi : \mathbb{R} \to \mathbb{R}$ es una función continuamente diferenciable, impar, estrictamente creciente y acotada.

I4. $\Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)$ y $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ son funciones continuas de $(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)$ y de $(\boldsymbol{\beta},\boldsymbol{\theta},u)$, respectivamente.

Cabe destacar que si ρ es una $\rho-$función, entonces si se cumple **I1**, para todo $s > 0$, $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \arg\min_a \mathbb{E}\left(\rho\left((y - \boldsymbol{\beta}^t\mathbf{x} - a)/s\right) \,\middle|\, \boldsymbol{\theta}^t\mathbf{t} = u\right)$. Por otro lado, **I4** es cierta si se cumplen **I2** iii) e **I3**.

Estudiaremos la consistencia del estimador $\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}$, definido en (3.6), a $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}$ para lo que necesitaremos probar algunas propiedades de convergencia del estimador de la distribución empírica condicional y del estimador de escala propuesto. Algunos de los resultados que siguen son generalizaciones de resultados obtenidos en Boente y Fraiman (1989 y 1991).

En particular, el teorema a continuación prueba la convergencia uniforme de la distribución empírica condicional. Su demostración puede hallarse en el Apéndice.

Teorema 3.3.5. *Sea $\mathcal{K} \subset \mathbb{R}^p$ un conjunto compacto y sea $\mathcal{T}_0 \subset \mathcal{T}$ un conjunto compacto que cumple que existe $\delta_0 > 0$ tal que $\mathcal{T}_{\delta_0} \subset \mathcal{T}$, donde $\mathcal{T}_{\delta_0}$ es la clausura de una δ_0 vecindad de $\mathcal{T}_0$. Si se satisfacen las condiciones **C1**-**C3** e **I2** ii), entonces*

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} \sup_{u\in\mathcal{U}(\mathcal{T}_0)} \sup_{r\in\mathbb{R}} |\widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r) - F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)| \xrightarrow{c.t.p.} 0 \text{ para todo } \mathcal{K} \subset \mathbb{R}^p \text{ conjunto compacto.} \quad (3.23)$$

Las demostraciones de los siguientes lemas auxiliares pueden hallarse en el Apéndice.

Lema 3.3.6. *Sean $\mathcal{K} \subset \mathbb{R}^p$ y $\mathcal{T}_0 \subset \mathcal{T}$ conjuntos compactos. Dado $\varepsilon > 0$, si se satisface **I2** iii), entonces existen constantes c y d tales que para todo $\boldsymbol{\beta} \in \mathcal{K}$, $\boldsymbol{\theta} \in \mathcal{S}^1$, $u \in \mathcal{U}(\mathcal{T}_0)$, $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c) < \varepsilon$ y $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d) > 1 - \varepsilon$.*

Lema 3.3.7. *Sean $\mathcal{K} \subset \mathbb{R}^p$ y $\mathcal{T}_0 \subset \mathcal{T}$ conjuntos compactos y $\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)$ una sucesión de funciones de distribución condicional para la cual*

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} \sup_{u\in\mathcal{U}(\mathcal{T}_0)} \sup_{r\in\mathbb{R}} |\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r) - F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)| \xrightarrow{c.t.p.} 0\,. \tag{3.24}$$

*Si $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)$ satisface **I2** i) y iii), entonces existen $n_0 \in \mathbb{N}$ y constantes positivas A y B tales que $s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = MAD(\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u})$ verifica que $A < s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) < B$ para todo $\boldsymbol{\beta} \in \mathcal{K}$, $\boldsymbol{\theta} \in \mathcal{S}^1$, $u \in \mathcal{U}(\mathcal{T}_0)$ y $n \geq n_0$.*

El siguiente teorema da como resultado la consistencia del estimador no paramétrico. El Teorema 3.3.14 dará más adelante una prueba alternativa sin el requerimiento solicitado en **I3** sobre la función de score ψ.

Teorema 3.3.8. *Sean $\mathcal{K} \subset \mathbb{R}^p$ y $\mathcal{T}_0 \subset \mathcal{T}$ conjuntos compactos. Si se cumplen **C1**-**C3** e **I1**-**I4**, entonces*

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} ||\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}||_{0,\infty} \xrightarrow{c.t.p.} 0 \tag{3.25}$$

Demostración:

Notemos que para todo $s > 0$, $\Lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), u) = 0$ ya que por **I3** ψ es una función acotada e impar y por **I1** $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)$ es simétrica alrededor de $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$.

Si dados $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1, u \in \mathcal{U}(\mathcal{T}_0)$, llamamos $\psi_b(v) = \psi\left((v - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) - b)/s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\right)$, de **I3** resulta que $\psi_b(v)$ es de variación acotada y que la norma de variación acotada de ψ_b cumple $\|\psi_b\|_V = \|\psi\|_V / s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$. Luego, si definimos

$$\widehat{\Lambda}(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u) = \int \psi\left(\frac{r-a}{s}\right) d\widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)\,,$$

tenemos que

$$\begin{aligned} |\Lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) + b, u) &- \widehat{\Lambda}(\boldsymbol{\beta}, \boldsymbol{\theta}, s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) + b, u)| \\ &\leq \|\psi_b\|_V \sup_{r\in\mathbb{R}} |F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r) - \widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)|\,, \end{aligned}$$

ya que $|\int \psi_b dH| \leq \|\psi_b\|_V \|H\|_\infty$ si H es continua a derecha y acotada.

En consecuencia, como por el Lema 3.3.7 existen constantes positivas A y B tales que $A < s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) < B$ para todo $\boldsymbol{\beta} \in \mathcal{K}$, $\boldsymbol{\theta} \in \mathcal{S}^1$, $u \in \mathcal{U}(\mathcal{T}_0)$, resulta

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} |\Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)+b,u) &- \widehat{\Lambda}(\boldsymbol{\beta},\boldsymbol{\theta},s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)+b,u)| \\ &\leq \frac{\|\psi\|_V}{A} \sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} \sup_{r\in\mathbb{R}} |F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r) - \widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)|\,.\end{aligned}$$

Esta desigualdad y el Teorema 3.3.5 implican que, fijado $b \in \mathbb{R}$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} |\Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)+b,u) - \widehat{\Lambda}(\boldsymbol{\beta},\boldsymbol{\theta},s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)+b,u)| \xrightarrow{c.t.p.} 0\,. \quad (3.26)$$

Dado $\varepsilon > 0$, por **I1**, **I3** y la continuidad de $\Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s,a,u)$ y de $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}$ dadas en **I4**, tenemos que

$$\Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)+\varepsilon,u) < 0 < \Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)-\varepsilon,u)$$

y por lo tanto, usando la compacidad de $\mathcal{K}, \mathcal{S}^1$ y $\mathcal{U}(\mathcal{T}_0)$ obtenemos que

$$\begin{aligned}L_1 &= \sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} \sup_{A<s<B} \Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)+\varepsilon,u) < 0 < \\ &< \inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} \inf_{A<s<B} \Lambda(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)-\varepsilon,u) = L_2\end{aligned}$$

Del hecho de que $A < s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) < B$ casi seguramente si $n \geq n_0$ y de la convergencia establecida en (3.26), obtenemos que

$$\widehat{\Lambda}(\boldsymbol{\beta},\boldsymbol{\theta},s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)+\varepsilon,u) < \frac{L_1}{2} < 0 < \frac{L_2}{2} < \widehat{\Lambda}(\boldsymbol{\beta},\boldsymbol{\theta},s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)-\varepsilon,u)$$

para todo $\boldsymbol{\beta} \in \mathcal{K}$, $\boldsymbol{\theta} \in \mathcal{S}^1$, $u \in \mathcal{U}(\mathcal{T}_0)$, $n \geq n_0$.

Recordando que $\widehat{\Lambda}(\boldsymbol{\beta},\boldsymbol{\theta},s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u),a,u)$ es estrictamente decreciente en a, obtenemos que $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) - \varepsilon < \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) < \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) + \varepsilon$ casi seguramente sobre $\mathcal{K}$, $\mathcal{S}^1$, $\mathcal{U}(\mathcal{T}_0)$ y por lo tanto,

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} |\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)| \xrightarrow{c.t.p.} 0\,.$$

como queríamos demostrar. ■

Bajo hipótesis adicionales puede estudiarse la convergencia de los estimadores paramétricos iniciales resultantes de los **Pasos I.2** e **I.3**. Para ello introducimos los funcionales

$$L(\boldsymbol{\beta},\boldsymbol{\theta},s,v) = \mathbb{E}\left(\rho_0\left(\frac{y-\boldsymbol{\beta}^t\mathbf{x}-v(\boldsymbol{\theta}^t\mathbf{t})}{s}\right)\right), \quad (3.27)$$

$$\sigma(\boldsymbol{\beta},\boldsymbol{\theta}) = \text{ínf}\left\{s > 0 : L(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) < \frac{1}{2}\right\}, \quad (3.28)$$

cuyas versiones muestrales son

$$L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, s, v) = \frac{1}{n} \sum_{i=1}^{n} \rho_0 \left(\frac{y_i - \boldsymbol{\beta}^t \mathbf{x}_i - v(\boldsymbol{\theta}^t \mathbf{t}_i)}{s} \right), \tag{3.29}$$

$$\tilde{\tau}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \text{ínf} \left\{ s > 0 : L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) < \frac{1}{2} \right\}. \tag{3.30}$$

Observación 3.3.9. Por definición, de acuerdo con (3.8) tenemos que $\tilde{\sigma}(\boldsymbol{\beta}, \boldsymbol{\theta}) \leq \tilde{\tau}(\boldsymbol{\beta}, \boldsymbol{\theta})$. Luego usando que $L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \cdot, \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}})$ es una función decreciente, obtenemos que $1/2 = L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \tilde{\sigma}(\boldsymbol{\beta}, \boldsymbol{\theta}), \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) \geq L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \tilde{\tau}(\boldsymbol{\beta}, \boldsymbol{\theta}), \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}})$. Como consecuencia de la definición de $\tilde{\tau}$ (ínfimo de la pre-imagen por un conjunto abierto de una función continua), concluímos que $L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \tilde{\tau}(\boldsymbol{\beta}, \boldsymbol{\theta}), \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) = 1/2$ y entonces $\tilde{\tau}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \tilde{\sigma}(\boldsymbol{\beta}, \boldsymbol{\theta})$.

Consideremos el siguiente conjunto de supuestos.

I5. ρ_0 es una ρ-función continuamente diferenciable y acotada, su derivada $\psi_0(u) = \rho_0'(u)$ y la función $\gamma_0(u) = \psi_0(u)\, u$ son acotadas. Además, $\mathbb{E}(\rho_0(\epsilon)) = 1/2$.

I6. $||\rho_0||_\infty > 1/2$ y para todo $\mathcal{K} \subset \mathbb{R}^p$ conjunto compacto

$$\sup_{\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1} \mathbb{P}(|y - \boldsymbol{\beta}^t \mathbf{x} + \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t \mathbf{t})| = 0) < 1 - \frac{1}{2\,||\rho_0||_\infty}.$$

I7. $\lim_{\alpha \to \infty} \sup_{\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1} \left(\mathbb{P}(|y| > \alpha) + \mathbb{P}(|\boldsymbol{\beta}^t \mathbf{x} + \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t \mathbf{t})| > \alpha) \right) = 0 \quad \forall\, \mathcal{K} \subset \mathbb{R}^p$ conjunto compacto.

La condición **I5** es habitual en el contexto de robustez, mientras que **I6** e **I7** son adaptaciones de las condiciones establecidas en la hipótesis 5.3 de Sakata y White (2001) a nuestro marco de trabajo y al supuesto de independencia entre las observaciones con el que trabajamos.

El siguiente lema es una adaptación parcial del Lema 5.1 de Sakata y White (2001) y su demostración se halla en el Apéndice.

Lema 3.3.10. *Sea $\mathcal{K} \subset \mathbb{R}^p$ un conjunto compacto. Si se cumplen las condiciones **I5**-**I7**, entonces se tiene que*

a) existen dos constantes $s_g > s_c > 0$ tales que para todo $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$, $s_c \leq \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) \leq s_g$.

b) Para cada $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$, $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \cdot, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) : (0, +\infty) \to [0, +\infty)$ es una función derivable.

c) Existe $0 < \delta_s < s_c$ tal que

$$\text{ínf}_{\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1} \; \text{ínf}_{s \in (\sigma(\boldsymbol{\beta},\boldsymbol{\theta}) - \delta_s, \sigma(\boldsymbol{\beta},\boldsymbol{\theta}) + \delta_s)} \left| \frac{\partial}{\partial s} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \right| > 0, \tag{3.31}$$

donde $\frac{\partial}{\partial s} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}})$ es la derivada parcial de $L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}})$ respecto de s.

Observación 3.3.11. Por el ítem *a*) del lema enunciado y el Lema 3.1 (c) de Sakata y White (2001), para todo compacto $\mathcal{K} \subset \mathbb{R}^p$, $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}), \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) = 1/2$ para cualquier $\boldsymbol{\beta} \in \mathcal{K}$ y $\boldsymbol{\theta} \in \mathcal{S}^1$.

Teorema 3.3.12. *Sea $\mathcal{K} \subset \mathbb{R}^p$ un conjunto compacto y sea $\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}\right)$ el argumento que minimiza $\tilde{\sigma}(\boldsymbol{\beta}, \boldsymbol{\theta})$. Si la función $\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}$ verifica que*

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} ||\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}||_{0,\infty} \xrightarrow{c.t.p.} 0$$

*y se satisfacen **I4**-**I7**,*

a) entonces $\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} |\tilde{\sigma}(\boldsymbol{\beta}, \boldsymbol{\theta}) - \sigma(\boldsymbol{\beta}, \boldsymbol{\theta})| \xrightarrow{c.t.p.} 0.$

b) si además existe $\mathcal{K}_1 \subset \mathbb{R}^p$ conjunto compacto tal que $\lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n\geq m} \tilde{\boldsymbol{\beta}} \in \mathcal{K}_1\right) = 1$ *y $\sigma(\boldsymbol{\beta}, \boldsymbol{\theta})$ es continua en $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ y tiene un único mínimo en $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, entonces $\tilde{\boldsymbol{\beta}} \xrightarrow{c.t.p.} \boldsymbol{\beta}_0$, $\tilde{\boldsymbol{\theta}} \xrightarrow{c.t.p.} \boldsymbol{\theta}_0$ y $\tilde{\sigma}\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}\right) \xrightarrow{c.t.p.} \sigma_0$.*

Demostración:

a) Tomemos las constantes s_c, s_g y δ_s como en los ítems *a*) y *c*) del Lema 3.3.10 y consideremos $0 < \varepsilon < \delta_s$ arbitrario. Definamos

$$A_2 = \inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} \inf_{s\in(\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\delta_s,\sigma(\boldsymbol{\beta},\boldsymbol{\theta})+\delta_s)} \left|\frac{\partial}{\partial s} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}})\right|.$$

Luego por el ítem *c*) del Lema 3.3.10 y usando que $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \cdot, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}})$ es una función decreciente, se tiene que para todo $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) - L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}), \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \leq -A_2\,\varepsilon\,, \tag{3.32}$$

y por lo tanto,

$$\begin{aligned} L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) - \frac{1}{2} &= L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) - L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}), \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \\ &= L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) - L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \\ &\quad + L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) - L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}), \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \\ &\leq L_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) - L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \varepsilon, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) - A_2\,\varepsilon\,, \end{aligned}$$

donde hemos usado la Observación 3.3.11.

Por **I4** sabemos que $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ es continua en $(\boldsymbol{\beta},\boldsymbol{\theta},u)$. De este hecho y del ítem a) del Lema 3.3.10, análogamente al ítem a) del Teorema 3.3.4 obtenemos que

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} |L_n(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta})+\varepsilon,\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) - L(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta})+\varepsilon,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}})| \xrightarrow{c.t.p.} 0\,.$$

Por lo que se llega a que en un conjunto de probabilidad 1, para n suficientemente grande

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} L_n(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta})+\varepsilon,\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) - \frac{1}{2} \leq -\frac{1}{2}\,A_2\,\varepsilon < 0\,,$$

es decir, para todo $\boldsymbol{\beta}\in\mathcal{K}$, $\boldsymbol{\theta}\in\mathcal{S}^1$

$$\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}) \leq \sigma(\boldsymbol{\beta},\boldsymbol{\theta})+\varepsilon\,,$$

o sea, que

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} [\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}) - \sigma(\boldsymbol{\beta},\boldsymbol{\theta})] \leq \varepsilon\,.$$

De manera similar, sabemos que dado $\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1$

$$L(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\varepsilon,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) - L(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta}),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \geq A_2\,\varepsilon \tag{3.33}$$

y entonces

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} L_n(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\varepsilon,\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) > \frac{1}{2}\,.$$

Si $\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}) < \sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\varepsilon$, tendríamos que

$$L_n(\boldsymbol{\beta},\boldsymbol{\theta},\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}),\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) \geq L_n(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\varepsilon,\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}) > \frac{1}{2}$$

contradiciendo la Observación 3.3.9. Luego, $\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}) \geq \sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\varepsilon$ si $\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1$ y por lo tanto

$$\inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} [\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}) - \sigma(\boldsymbol{\beta},\boldsymbol{\theta})] \geq -\varepsilon\,.$$

En consecuencia, $\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} |\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta})-\sigma(\boldsymbol{\beta},\boldsymbol{\theta})| \leq \varepsilon$ y tal como queríamos demostrar resulta que $\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} |\tilde{\sigma}(\boldsymbol{\beta},\boldsymbol{\theta}) - \sigma(\boldsymbol{\beta},\boldsymbol{\theta})| \xrightarrow{c.t.p.} 0$.

b) Como $\big(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}\big)$ está últimamente en un compacto, contiene una subsucesión convergente, por simplicidad supongamos que $\big(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}\big)$ es la subsucesión convergente a $\big(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*\big)$ c.t.p. Por el ítem a) y la continuidad de $\sigma(\boldsymbol{\beta},\boldsymbol{\theta})$, tenemos que

$$\tilde{\sigma}\big(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}\big) - \sigma\big(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*\big) \xrightarrow{c.t.p.} 0 \tag{3.34}$$

y además $\tilde{\sigma}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - \sigma(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \xrightarrow{c.t.p.} 0$. Como $\tilde{\sigma}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \geq \tilde{\sigma}\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}\right)$ y $\sigma(\boldsymbol{\beta}, \boldsymbol{\theta})$ tiene un único mínimo en $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, entonces $\left(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*\right) = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$. Luego, $\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}\right) \xrightarrow{c.t.p.} (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ y, de (3.34), $\tilde{\sigma}\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}\right) \xrightarrow{c.t.p.} \sigma(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \sigma_0$, quedando demostrado el ítem b). ■

Observación 3.3.13. El supuesto de continuidad de $\sigma(\boldsymbol{\beta}, \boldsymbol{\theta})$ puede probarse bajo condiciones de regularidad. Consideremos la función $f(\boldsymbol{\beta}, \boldsymbol{\theta}, s) = L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) - 1/2$ para $(\boldsymbol{\beta}, \boldsymbol{\theta}, s)$ en un abierto de $\mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}_{>0}$ que contenga a $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0)$. Por un lado, la condición **I5** garantiza que $f(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0) = 0$, y por el otro, nos asegura que

$$\frac{\partial f}{\partial s}(\boldsymbol{\beta}, \boldsymbol{\theta}, s) = -\frac{1}{s}\mathbb{E}\left[\psi_0\left(\frac{y - \boldsymbol{\beta}^t\mathbf{x} - \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})}{s}\right)\frac{y - \boldsymbol{\beta}^t\mathbf{x} - \tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})}{s}\right].$$

Resulta que $\frac{\partial f}{\partial s}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0) = -\mathbb{E}(\psi_0(\epsilon)\epsilon)/\sigma_0 \neq 0$ pues $\psi_0(u)u$ es una función par y ϵ tiene distribución simétrica alrededor del 0. Por lo tanto, por el Teorema de la Función Implícita existen $U \subset \mathbb{R}^{p+q}$ y $V \subset \mathbb{R}$ conjuntos abiertos tales que $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \in U$ y $\sigma_0 \in V$ y una única función $g : U \longrightarrow V$ tal que $g(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \sigma_0$ y $f(\boldsymbol{\beta}, \boldsymbol{\theta}, g(\boldsymbol{\beta}, \boldsymbol{\theta})) = 0$ para todo $(\boldsymbol{\beta}, \boldsymbol{\theta}) \in U$. Más aún, si $(\boldsymbol{\beta}, \boldsymbol{\theta}) \in U$ y $\sigma \in V$ tal que $f(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma) = 0$, entonces $\sigma = g(\boldsymbol{\beta}, \boldsymbol{\theta})$, con $g \in \mathcal{C}^1(U)$. Por la Observación 3.3.11, tenemos que $f(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta})) = 0$ para todo $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$; esto nos permite considerar que $g = \sigma$, y entonces resulta que $\sigma(\boldsymbol{\beta}, \boldsymbol{\theta})$ es continua en un entorno de $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$.

El siguiente teorema prueba la consistencia del estimador no paramétrico bajo otros requerimientos de regularidad y en ese sentido es un resultado alternativo al Teorema 3.3.8. A continuación se enuncian las condiciones que se utilizarán.

I8. ρ es una función continua y acotada.

I9. $M(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)$ satisface la siguiente condición de equicontinuidad: dados los conjuntos compactos $\mathcal{K} \subset \mathbb{R}^p$, $\mathcal{T}_0 \subset \mathcal{T}$ y $\mathcal{S} \subset \mathbb{R}_{>0}$, para cualquier $\varepsilon > 0$ existe $\delta > 0$ tal que si $||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|| < \delta$, $||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| < \delta$, $|s_1 - s_2| < \delta$ y $|u_1 - u_2| < \delta$ donde $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{K}$, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{S}^1$, $s_1, s_2 \in \mathcal{S}$ y $u_1, u_2 \in \mathcal{U}(\mathcal{T}_0)$ entonces

$$\sup_{a \in \mathbb{R}} |M(\boldsymbol{\beta}_1, \boldsymbol{\theta}_1, s_1, a, u_1) - M(\boldsymbol{\beta}_2, \boldsymbol{\theta}_2, s_2, a, u_2)| < \varepsilon\,.$$

I10. $M(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)$ es una función continua en $(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, u)$ y $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ es una función continua en $(\boldsymbol{\beta}, \boldsymbol{\theta}, u)$.

El siguiente teorema requerirá que la escala robusta condicional $S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = MAD(F_{\boldsymbol{\beta},\boldsymbol{\theta},u})$ sea una función continua y que $\sup_{\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1} ||s_{\boldsymbol{\beta},\boldsymbol{\theta}} - S_{\boldsymbol{\beta},\boldsymbol{\theta}}||_{0,\infty} \xrightarrow{c.t.p.} 0$. Estos resultados pueden ser probados utilizando técnicas análogas a las empleadas en Boente y Martinez (2017).

Teorema 3.3.14. *Sean $\mathcal{K} \subset \mathbb{R}^p$, $\mathcal{S} \subset \mathbb{R}_{>0}$ conjuntos compactos y $\mathcal{T}_0 \subset \mathcal{T}$ un conjunto compacto que cumple que existe $\delta_0 > 0$ tal que $\mathcal{T}_{\delta_0} \subset \mathcal{T}$, donde $\mathcal{T}_{\delta_0}$ es la clausura de una δ_0-vecindad de $\mathcal{T}_0$. Supongamos que se satisfacen las condiciones **C1-C3**, **I8-I10** y la siguiente*

(C) la familia de funciones $\mathfrak{F} = \{f(y, \mathbf{x}) = \rho\left((y - \boldsymbol{\beta}^t\mathbf{x} - a)/s\right) : \boldsymbol{\beta} \in \mathcal{K}, s \in \mathcal{S}, a \in \mathbb{R}\}$ tiene un número de cubrimiento tal que $\sup_{\mathcal{Q}} N(\varepsilon, \mathfrak{F}, L^1(\mathcal{Q})) \leq A\varepsilon^{-W}$, para $0 < \varepsilon < 1$, $\mathcal{Q}$ una medida de probabilidad para $(y, \mathbf{x})$ y $A, W > 0$.

Luego, tenemos que

a) $$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,s\in\mathcal{S},a\in\mathbb{R}} ||M_n(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot) - M(\boldsymbol{\beta}, \boldsymbol{\theta}, s, a, \cdot)||_{0,\infty} \xrightarrow{c.t.p.} 0\,.$$

b) Si $$\inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1,u\in\mathcal{U}(\mathcal{T}_0)} \left[\lim_{|a|\to\infty} M(\boldsymbol{\beta}, \boldsymbol{\theta}, S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), a, u) - M(\boldsymbol{\beta}, \boldsymbol{\theta}, S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), u)\right] > 0$$ *y la función $M(\boldsymbol{\beta}, \boldsymbol{\theta}, S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u), a, u)$ tiene un único mínimo en $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$, entonces*

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} ||\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}||_{0,\infty} \xrightarrow{c.t.p.} 0\,. \tag{3.35}$$

Demostración:

a) Análoga a la demostración del ítem a) del Teorema 3.3.3.

b) La continuidad de $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ junto con su dominio compacto implican que $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ es acotada en $\mathcal{K} \times \mathcal{S}^1 \times \mathcal{U}(\mathcal{T}_0)$. Entonces, existe $\mathcal{A}(\mathcal{K}, \mathcal{T}_0)$ conjunto compacto tal que $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) \in \mathcal{A}(\mathcal{K}, \mathcal{T}_0)$ para todo $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$ y $u \in \mathcal{U}(\mathcal{T}_0)$.
Sea $\tilde{\Omega}$ tal que $\mathbb{P}(\tilde{\Omega}) = 1$ y $\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} ||s_{\boldsymbol{\beta},\boldsymbol{\theta}} - S_{\boldsymbol{\beta},\boldsymbol{\theta}}||_{0,\infty} \to 0$ para todo $\omega \in \tilde{\Omega}$. Supongamos que
$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} ||\tilde{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}||_{0,\infty}$ no converge a 0 en un conjunto Ω_0 tal que $\mathbb{P}(\Omega_0) > 0$. Luego para cada $\omega \in \Omega_0 \cap \tilde{\Omega}$, existe $\{(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, u_k)\}_{k\in\mathbb{N}} \subset \mathcal{K}\times\mathcal{S}^1\times\mathcal{U}(\mathcal{T}_0)$, que sin pérdida de generalidad podemos suponer tal que $(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, u_k) \to (\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, u_L)$. Así, $\tilde{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) - \eta^*_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) \to c \neq 0$, $\eta^*_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) \to \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L)$,

$$s_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) - S_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) \to 0\,, \tag{3.36}$$

$$s_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L) - S_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L) \to 0 \tag{3.37}$$

y

$$S_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) \to S_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L). \tag{3.38}$$

Por lo tanto $\tilde{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) - \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L) \to c$ y $s_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) \to S_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L)$.
Cuando $c \in \mathbb{R}_{\neq 0}$, para k y n suficientemente grandes y $\varepsilon > 0$, sabiendo que se cumple la

hipótesis **I9** y que $S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ es una función continua, tenemos que
$M(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, S_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L) + c, u_L) < M(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, S_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), \tilde{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), u_k) + \frac{\varepsilon}{8}$.
Además por (3.36), el ítem a) y la definición de $\tilde{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k)$,
$M(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, S_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), \tilde{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), u_k) < M(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), \tilde{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), u_k) + \frac{\varepsilon}{8}$

$< M_n(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), \tilde{\eta}_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), u_k) + \frac{\varepsilon}{4} \leq M_n(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), u_k) + \frac{\varepsilon}{4}$

$< M(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, s_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), u_k) + \frac{3}{8}\varepsilon < M(\boldsymbol{\beta}_k, \boldsymbol{\theta}_k, S_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k), \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), u_k) + \frac{\varepsilon}{2}$

$< M(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, S_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), u_L) + \frac{7}{8}\varepsilon$.
Por lo tanto, llegamos a que

$$M(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, S_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L) + c, u_L) < M(\boldsymbol{\beta}_L, \boldsymbol{\theta}_L, S_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), \eta^*_{\boldsymbol{\beta}_L,\boldsymbol{\theta}_L}(u_L), u_L) + \varepsilon.$$

Debido a la buena definición de $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ por el ítem b), resulta que $c = 0$, lo cual es un absurdo.
Si $c = \infty$, $\tilde{\eta}^*_{\boldsymbol{\beta}_k,\boldsymbol{\theta}_k}(u_k) \to \infty$. Debido a la condición de equicontinuidad **I9**, a (3.38), a (3.36) y a (3.37), de manera análoga al Teorema 3.3.3 se llega a una contradicción.

En conclusión, se prueba (3.35). ■

Observación 3.3.15. Del reciente Teorema enunciado, utilizando la condición **I8**, se deduce que para cada conjunto compacto $\mathcal{K} \subset \mathbb{R}^p$ y para cada conjunto compacto $\mathcal{S} \subset \mathbb{R}_{>0}$, la familia de funciones $\mathfrak{F}$ con ρ una ρ-función, forma una clase euclídea, entonces es inmediato probar que tiene un número de cubrimiento satisfaciendo $\sup_{\mathcal{Q}} N(\varepsilon, \mathfrak{F}, L^1(\mathcal{Q})) \leq A\varepsilon^{-W}$, para $0 < \varepsilon < 1$, con $\mathcal{Q}$ medida de probabilidad para $(y, \mathbf{x})$ y $A, W > 0$.

Observación 3.3.16. El Teorema anterior puede generalizarse para otros estimadores de escala robustos $e(\boldsymbol{\beta}, \boldsymbol{\theta}, u)$ en lugar de $S_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ modificando el ítem i) y las hipótesis del b) de la siguiente manera:
i) la función $M(\boldsymbol{\beta}, \boldsymbol{\theta}, e(\boldsymbol{\beta}, \boldsymbol{\theta}, u), a, u)$ tiene un único mínimo en $\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$, siendo $e(\boldsymbol{\beta}, \boldsymbol{\theta}, u)$ un estimador de escala robusto de $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}$.
b) se pide además que $e(\boldsymbol{\beta}, \boldsymbol{\theta}, u)$ sea una función continua, $e_n(\boldsymbol{\beta}, \boldsymbol{\theta}, u)$ sea un estimador de escala robusto de $\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}$ y

$$\sup_{\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1} ||e_n(\boldsymbol{\beta}, \boldsymbol{\theta}, \cdot) - e(\boldsymbol{\beta}, \boldsymbol{\theta}, \cdot)||_{0,\infty} \xrightarrow{c.t.p.} 0\,.$$

Apéndice

Demostración del Teorema 3.3.5:

Denominemos

$$F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{(-\infty,r]}(y_i-\boldsymbol{\beta}^t\mathbf{x}_i)K_h(\boldsymbol{\theta}^t\mathbf{t}_i-u) \text{ y } F_{0n}(\boldsymbol{\theta},u)=\frac{1}{n}\sum_{i=1}^{n}K_h(\boldsymbol{\theta}^t\mathbf{t}_i-u)$$

siendo como ya definimos $K_h(u)=K\left(\frac{u}{h}\right)/h$. Notemos que $\widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)=F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u)/F_{0n}(\boldsymbol{\theta},u)$ y por lo tanto

$$\begin{aligned}|\widehat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)-F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)|\leq&[|F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u)-\mathbb{E}(F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u))|\\&+|\mathbb{E}(F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u))-F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)\mathbb{E}(F_{0n}(\boldsymbol{\theta},u))|\\&+|F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)||F_{0n}(\boldsymbol{\theta},u)-\mathbb{E}(F_{0n}(\boldsymbol{\theta},u))|]/F_{0n}(\boldsymbol{\theta},u)\end{aligned}$$

Como en la prueba del Teorema 3.3.3, para demostrar (3.23) alcanza con probar que

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1}\sup_{u\in\mathcal{U}(\mathcal{T}_0)}\sup_{r\in\mathbb{R}}|F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u)-\mathbb{E}(F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u))|\xrightarrow{c.t.p.}0\,, \tag{3.39}$$

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1}\sup_{u\in\mathcal{U}(\mathcal{T}_0)}\sup_{r\in\mathbb{R}}|\mathbb{E}(F_{1n}(\boldsymbol{\beta},\boldsymbol{\theta},r,u))-F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)\mathbb{E}(F_{0n}(\boldsymbol{\theta},u))|\to 0\,. \tag{3.40}$$

Las pruebas de (3.39) y la de (3.40) siguen usando argumentos análogos a los empleados en el Teorema 3.3.3 para demostrar (3.16) y (3.17), respectivamente . ■

Demostración del Lema 3.3.6:

Dados $\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*$ existe un $c^*=c(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*)$ tal que $F_{\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*}(c^*)<\varepsilon/2$. Como por **I2** iii) $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c)$ es continua en $(\boldsymbol{\beta},\boldsymbol{\theta},u)$, entonces existe un $\delta=\delta(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*)>0$ tal que si $|u^*-u|<\delta$, $\|\boldsymbol{\beta}^*-\boldsymbol{\beta}\|<\delta$, $\|\boldsymbol{\theta}^*-\boldsymbol{\theta}\|<\delta$, entonces $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c^*)<\varepsilon$.

Sea $B_r(a)$ la bola de centro a y radio r. Como $\mathcal{K}\times\mathcal{S}^1\times\mathcal{U}(\mathcal{T}_0)\subset\cup B_{\delta(\boldsymbol{\beta},\boldsymbol{\theta},u)}(\boldsymbol{\beta},\boldsymbol{\theta},u)$, por la compacidad existen $(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i),1\leq i\leq k$, tales que $\mathcal{K}\times\mathcal{S}^1\times\mathcal{U}(\mathcal{T}_0)\subset\cup_{i=1}^{k}B_{\delta(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)}(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)$.

Sean $c_i=c(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)\in\mathbb{R}$ tales que $F_{\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i}(c_i)<\varepsilon/2$ para todo $i=1,\ldots,k$. Tomando $c=\min_{1\leq i\leq k}c_i$ obtenemos la primera desigualdad, ya que dado $(\boldsymbol{\beta},\boldsymbol{\theta},u)$, para algún i, $(\boldsymbol{\beta},\boldsymbol{\theta},u)\in B_{\delta(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)}(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)$ y en consecuencia $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c)\leq F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c_i)<\varepsilon$. La segunda cota se obtiene de forma análoga. ■

Demostración del Lema 3.3.7:

Como consecuencia del Lema 3.3.6, al cumplirse **I2** iii) tenemos que existen constantes $c<d$ tales que $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c)<1/8$ y $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d)>7/8$ si $\boldsymbol{\beta}\in\mathcal{K}$, $\boldsymbol{\theta}\in\mathcal{S}^1$, $u\in\mathcal{U}(\mathcal{T}_0)$.

Dada la convergencia de sucesión de distribuciones condicionales que resulta de (3.24), tenemos que dado $\varepsilon>0$, existe n_0 tal que si $n\geq n_0$ entonces $-\varepsilon<\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)-F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(r)<\varepsilon$. Por lo tanto, si elegimos $\varepsilon=1/8$, para $n\geq n_0$ se obtiene que $\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c)<1/4$ y $\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d)>3/4$,

por lo que si $\hat{m}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)(u)$ denota a la mediana respecto de la distribución $\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}$, entonces $c < \hat{m}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)(u) < d$, de donde resulta que $s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) < d - c$, con lo cual basta tomar $B = d - c$.

Por otro lado, por **I2** i) resulta que $F_{\boldsymbol{\beta},\boldsymbol{\theta},u}$ es una función de distribución continua para $(\boldsymbol{\beta},\boldsymbol{\theta},u)$ fijo. Por esta razón, fijado $(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*)$, $\boldsymbol{\beta}^* \in \mathcal{K}$, $\boldsymbol{\theta}^* \in \mathcal{S}^1$, $u^* \in \mathcal{U}(\mathcal{T}_0)$, existen constantes $c^* = c(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*)$ y $d^* = d(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*)$ para las que

$$F_{\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*}(c^*) = \frac{1}{3} \quad , \quad F_{\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*}(d^*) = \frac{7}{10}\,.$$

Además, por **I2** iii) sabemos que existe un $\delta^* = \delta(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*,u^*)$ tal que si $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| < \delta^*$, $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| < \delta^*$ y $|u^* - u| < \delta^*$, entonces

$$\frac{1}{3} - \varepsilon < F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c^*) < \frac{1}{3} + \varepsilon \quad \text{y} \quad \frac{7}{10} - \varepsilon < F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d^*) < \frac{7}{10} + \varepsilon\,.$$

Razonando como en el lema anterior, por la compacidad existen $(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i), 1 \le i \le k$, tales que $\mathcal{K} \times \mathcal{S}^1 \times \mathcal{U}(\mathcal{T}_0) \subset \cup_{i=1}^k B_{\delta(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)}(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)$ y para cada uno de ellos existen c_i y d_i tales que $F_{\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i}(c_i) = 1/3$ y $F_{\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i}(d_i) = 7/10$, de manera que si $(\boldsymbol{\beta},\boldsymbol{\theta},u) \in B_{\delta(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)}(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)$ entonces

$$\frac{1}{3} - \varepsilon < F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c_i) < \frac{1}{3} + \varepsilon \quad \text{y} \quad \frac{7}{10} - \varepsilon < F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d_i) < \frac{7}{10} + \varepsilon\,.$$

Asimismo, como por **I2** i), existe un $\nu > 0$ para el cual

$$\frac{1}{3} - 2\varepsilon < F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c_i - \nu) < \frac{1}{3} + 2\varepsilon \quad \text{y} \quad \frac{7}{10} - 2\varepsilon < F_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d_i + \nu) < \frac{7}{10} + 2\varepsilon\,.$$

En consecuencia, la convergencia de la sucesión de distribuciones condicionales dado $(\boldsymbol{\beta},\boldsymbol{\theta},u) \in B_{\delta(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)}(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i), 1 \le i \le k$, para $n \ge n_0$ resulta que

$$\frac{1}{3} - 3\varepsilon < \hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c_i - \nu) < \frac{1}{3} + 3\varepsilon \quad \text{y} \quad \frac{7}{10} - 3\varepsilon < \hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d_i + \nu) < \frac{7}{10} + 3\varepsilon\,.$$

Por último, tomando $\varepsilon < 1/60$, para $n \ge n_0$, $1 \le i \le k$, obtenemos que

$$\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c_i) < \frac{1}{2} \quad , \quad \hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(c_i - \nu) > \frac{1}{4}$$

y además

$$\hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d_i) > \frac{1}{2} \quad , \quad \hat{F}_{\boldsymbol{\beta},\boldsymbol{\theta},u}(d_i + \nu) < \frac{3}{4}\,,$$

de donde se desprende que $c_i \le \hat{m}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)(u) \le d_i$ para todo $(\boldsymbol{\beta},\boldsymbol{\theta},u) \in B_{\delta(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i)}(\boldsymbol{\beta}_i,\boldsymbol{\theta}_i,u_i), 1 \le i \le k$, y además $s_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) > \nu$ para todo $\boldsymbol{\beta} \in \mathcal{K}$, $\boldsymbol{\theta} \in \mathcal{S}^1$, $u \in \mathcal{U}(\mathcal{T}_0)$, $n \ge n_0$ y alcanza con tomar $A = \nu$. ■

Demostración del Lema 3.3.10:

a) Por la condición **I6**, existen $\alpha > 0$ y $\varepsilon > 0$ tales que

$$\inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} \mathbb{P}(|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha) \geq \frac{1}{2\,||\rho_0||_\infty} + \varepsilon\,.$$

Usando esta desigualdad y el hecho de que ρ_0 es una ρ-función que cumple **R1** y **R2**, obtenemos que para cada $(\boldsymbol{\beta}, \boldsymbol{\theta}, s) \in \mathcal{K} \times \mathcal{S}^1 \times \mathbb{R}_{>0}$

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) &\geq \rho_0\left(\frac{\alpha}{s}\right)\mathbb{P}(|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha) + \rho_0(0)\,\mathbb{P}(|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| \leq \alpha) \\ &\geq \rho_0\left(\frac{\alpha}{s}\right)\left(\frac{1}{2\,||\rho_0||_\infty} + \varepsilon\right). \end{aligned}$$

En consecuencia,

$$\lim_{s\searrow 0}\inf_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \geq ||\rho_0||_\infty\left(\frac{1}{2\,||\rho_0||_\infty} + \varepsilon\right) = \frac{1}{2} + ||\rho_0||_\infty\,\varepsilon > \frac{1}{2}\,.$$

Luego, existe $s_c > 0$ en un entorno del 0 tal que para cualquier $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$, $L(\boldsymbol{\beta}, \boldsymbol{\theta}, s_c, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) > 1/2$, lo que implica que $\sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) \geq s_c$.

Para probar la segunda desigualdad, usaremos que para cada $(\boldsymbol{\beta}, \boldsymbol{\theta}) \in \mathcal{K} \times \mathcal{S}^1$ y cada real $\alpha > 0$,

$$\mathbb{P}(|y| + |\boldsymbol{\beta}^t\mathbf{x} + \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha) \leq \mathbb{P}(|y| > \alpha/2) + \mathbb{P}(|\boldsymbol{\beta}^t\mathbf{x} + \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha/2)\,.$$

Al ser y una variable aleatoria y cumplirse la condición **I7**, el miembro derecho de la desigualdad puede ser tan chico como se quiera uniformemente sobre $\mathcal{K}\times\mathcal{S}^1$, para valores de α suficientemente grandes, por lo que para $\tilde{\epsilon} = 1/(4||\rho_0||_\infty)$, existe $\alpha_0 > 0$ tal que

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\theta})\in\mathcal{K}\times\mathcal{S}^1} \mathbb{P}(|y| + |\boldsymbol{\beta}^t\mathbf{x} + \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha_0) < \tilde{\epsilon}\,.$$

Luego, dado $s > 0$,

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) &\leq ||\rho_0||_\infty\,\mathbb{P}(|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha_0) + \rho_0\left(\frac{\alpha_0}{s}\right)\mathbb{P}(|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| \leq \alpha_0) \\ &\leq ||\rho_0||_\infty\,\mathbb{P}(|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha_0) + \rho_0\left(\frac{\alpha_0}{s}\right), \end{aligned}$$

de donde resulta que

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\theta})\in\mathcal{K}\times\mathcal{S}^1} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \leq ||\rho_0||_\infty\,\tilde{\epsilon} + \rho_0\left(\frac{\alpha_0}{s}\right) = \frac{1}{4} + \rho_0\left(\frac{\alpha_0}{s}\right).$$

Como consecuencia, se obtiene que

$$\lim_{s\to+\infty} \sup_{(\boldsymbol{\beta},\boldsymbol{\theta})\in\mathcal{K}\times\mathcal{S}^1} L(\boldsymbol{\beta},\boldsymbol{\theta},s,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) \leq \frac{1}{4} < \frac{1}{2}\,,$$

y por lo tanto, existe s_g suficientemente grande tal que $L(\boldsymbol{\beta},\boldsymbol{\theta},s_g,\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}) < 1/2$. En conclusión, $\sigma(\boldsymbol{\beta},\boldsymbol{\theta}) \leq s_g$ cuando $(\boldsymbol{\beta},\boldsymbol{\theta}) \in \mathcal{K}\times\mathcal{S}^1$.

b) Sean $\boldsymbol{\beta}\in\mathcal{K}$ y $\boldsymbol{\theta}\in\mathcal{S}^1$ fijos. Por la condición **I5**, dado $s>0$ se tiene que

$$\left|\frac{\partial}{\partial s}\left(\rho_0\left(\frac{y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})}{s}\right)\right)\right| = \frac{1}{s^2}\left(\psi_0\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s}\right)\right)|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|$$
$$\leq \frac{1}{s}\,\|\gamma_0\|_\infty\,.$$

Ahora bien, dado $s_0>0$ arbitrario, la desigualdad anterior implica que para cualquier $s\in[s_0/2,+\infty)$

$$\left|\frac{\partial}{\partial s}\left(\rho_0\left(\frac{y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})}{s}\right)\right)\right| \leq 2\,s_0^{-1}\,\|\gamma_0\|_\infty$$

y por lo tanto, en un entorno de s_0 tenemos que $|\frac{\partial}{\partial s}(\rho_0((y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t}))/s))|$ es integrable. Luego, aplicando el Teorema de la Convergencia Dominada resulta que

$$\frac{\partial}{\partial s}\left[\mathbb{E}\left(\rho_0\left(\frac{y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})}{s}\right)\right)\right] = \mathbb{E}\left[\frac{\partial}{\partial s}\left(\rho_0\left(\frac{y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})}{s}\right)\right)\right].$$

Al ser s_0 arbitrario, queda probado *b*).

c) Consideremos

$$\varepsilon = \frac{1}{2}\left(\left(1-\frac{1}{2\,||\rho_0||_\infty}\right) - \lim_{\alpha\searrow 0}\sup_{\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1}\mathbb{P}(|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|\leq\alpha)\right), \tag{3.41}$$

que es un número positivo por la condición **I6**, y tomemos

$$d = \frac{1}{||\rho_0||_\infty^{-1}+\varepsilon}\,, \tag{3.42}$$

que verifica $0<d<||\rho_0||_\infty$. Como ρ_0 es una ρ-función, por **R3** e **I5** existe un único $\tilde{\alpha}>0$ tal que $\rho_0(\tilde{\alpha})=d$. Fijemos $0<\delta_s<s_c$ suficientemente chico, siendo s_c el límite inferior del ítem *a*) de este lema, tal que

$$\rho_0\left(\tilde{\alpha}\,\frac{s_c}{s_c-\delta_s}\right) < ||\rho_0||_\infty\,.$$

Como por el ítem a) de este lema $\sigma(\boldsymbol{\beta},\boldsymbol{\theta})$ es positiva para cualquier $\boldsymbol{\beta}\in\mathcal{K}$ y $\boldsymbol{\theta}\in\mathcal{S}^1$, por el Lema 3.1 (c) de Sakata y White (2001), $L(\boldsymbol{\beta},\boldsymbol{\theta},\sigma(\boldsymbol{\beta},\boldsymbol{\theta}),\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}})=1/2$. Por lo tanto,

$$\begin{aligned}\frac{1}{2} &\geq \rho_0(\tilde{\alpha})\mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}>\tilde{\alpha}\right)+\rho_0(0)\,\mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}\leq\tilde{\alpha}\right)\\ &\geq \rho_0(\tilde{\alpha})\mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}>\tilde{\alpha}\right).\end{aligned}$$

Luego

$$\mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}>\tilde{\alpha}\right)\leq\frac{1}{2\,\rho_0(\tilde{\alpha})}=\frac{1}{2\,||\rho_0||_\infty}+\frac{\varepsilon}{2},\tag{3.43}$$

donde la igualdad se deduce de que $\rho_0(\tilde{\alpha})=d$, con d definido en (3.42). De esta forma, para cualquier $s\in[\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\delta_s,\sigma(\boldsymbol{\beta},\boldsymbol{\theta})+\delta_s]$ resulta que

$$\begin{aligned}\mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s}>\tilde{\alpha}\,\frac{s_c}{s_c-\delta_s}\right) &\leq \mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\delta_s}>\tilde{\alpha}\,\frac{s_c}{s_c-\delta_s}\right)\\ &= \mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}>\tilde{\alpha}\,\frac{s_c}{s_c-\delta_s}\frac{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\delta_s}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}\right)\\ &\leq \mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}>\tilde{\alpha}\,\frac{s_c}{s_c-\delta_s}\frac{s_c-\delta_s}{s_c}\right)\\ &= \mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{\sigma(\boldsymbol{\beta},\boldsymbol{\theta})}>\tilde{\alpha}\right).\end{aligned}$$

De (3.43) y esta última cadena de desigualdades, obtenemos que para $s\in[\sigma(\boldsymbol{\beta},\boldsymbol{\theta})-\delta_s,\sigma(\boldsymbol{\beta},\boldsymbol{\theta})+\delta_s]$ y $\alpha_1=\tilde{\alpha}\,s_c/(s_c-\delta_s)$,

$$\mathbb{P}\left(\frac{|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s}>\alpha_1\right)\leq\frac{1}{2\,||\rho_0||_\infty}+\frac{\varepsilon}{2}\,.\tag{3.44}$$

Por otro lado, la definición de ε en (3.41) implica que existe $0<\alpha_2<\alpha_1$ tal que para todo $\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1$,

$$\mathbb{P}(|y-\boldsymbol{\beta}^t\mathbf{x}-\eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|>\alpha_2\,(s_g+\delta_s))\geq\frac{1}{2\,||\rho_0||_\infty}+\varepsilon,$$

siendo s_g el límite superior dado en el ítem a) de este lema. De manera análoga a la cadena de desigualdades anterior, se tiene que para cualquier $\boldsymbol{\beta}\in\mathcal{K},\boldsymbol{\theta}\in\mathcal{S}^1$ y

$s \in [\sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) - \delta_s, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \delta_s]$

$$\mathbb{P}\left(\frac{|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s} > \alpha_2\right) \geq \mathbb{P}(|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})| > \alpha_2\,(s_g + \delta_s))\,.$$

Por lo tanto, con las dos últimas desigualdades obtenemos que

$$\mathbb{P}\left(\frac{|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s} > \alpha_2\right) \geq \frac{1}{2\,||\rho_0||_\infty} + \varepsilon\,. \tag{3.45}$$

Combinando (3.44) y (3.45), obtenemos que

$$\mathbb{P}\left(\alpha_2 \leq \frac{|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s} \leq \alpha_1\right) \geq \frac{\varepsilon}{2}\,,$$

con lo cual para cualquier $\boldsymbol{\beta} \in \mathcal{K}, \boldsymbol{\theta} \in \mathcal{S}^1$ y $s \in [\sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) - \delta_s, \sigma(\boldsymbol{\beta}, \boldsymbol{\theta}) + \delta_s]$ resulta condicionando

$$\begin{aligned}
&\left|\frac{\partial}{\partial s} L(\boldsymbol{\beta}, \boldsymbol{\theta}, s, \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}})\right| = s^{-2}\mathbb{E}\left[\psi_0\left(\frac{|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s}\right)|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|\right] \\
&\geq s^{-2}\mathbb{E}\left[\psi_0\left(\frac{|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s}\right)|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|\right| \\
&\alpha_2 \leq \frac{|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s} \leq \alpha_1\Bigg]\,\mathbb{P}\left(\alpha_2 \leq \frac{|y - \boldsymbol{\beta}^t\mathbf{x} - \eta^*_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t})|}{s} \leq \alpha_1\right) \\
&\geq s_g^{-2} \inf_{x \in [\alpha_2, \alpha_1]} \psi_0(x)\,\frac{\varepsilon}{2} > 0\,,
\end{aligned}$$

donde la última desigualdad es cierta independientemente de $\boldsymbol{\beta}, \boldsymbol{\theta}$ y s y tenemos que $\inf_{x \in [\alpha_2, \alpha_1]} \psi_0(x) > 0$ debido a la definición de ρ-función (ítem **R3**). Tomando ínfimo se prueba lo buscado. ■

Capítulo 4

Distribución Asintótica

4.1. Resultados Principales

En esta sección probaremos que, bajo condiciones de regularidad, los estimadores que se proponen, se distribuyen asintóticamente normal. Para ello, necesitaremos suponer que $\mathbf{t} \in \mathcal{T}$ siendo $\mathcal{T} \subset \mathbb{R}^q$ un conjunto compacto. Denominemos $\mathcal{U}_0(\mathcal{T}) = \{\boldsymbol{\theta}_0^t \mathbf{t} : \mathbf{t} \in \mathcal{T}\}$, y para $v : A \to \mathbb{R}$ una función continua sobre $A \subset \mathbb{R}$ un conjunto compacto, notaremos $||v||_{A,\infty} = \sup_{u \in A} |v(u)|$.

El estimador $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right)$ de $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ verifica

$$\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\theta} \in \mathcal{S}^1}{\operatorname{argmin}} \; G_n\big(\boldsymbol{\beta}, \boldsymbol{\theta}, \widehat{\sigma}, \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}\big). \tag{4.1}$$

Sabemos que

$$\nabla_{(\boldsymbol{\beta},\boldsymbol{\theta})} G_n\big(\boldsymbol{\beta}, \boldsymbol{\theta}, \widehat{\sigma}, \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}\big)\big|_{(\boldsymbol{\beta},\boldsymbol{\theta})=(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}})} + \iota_n \begin{pmatrix} \mathbf{0} \\ \widehat{\boldsymbol{\theta}} \end{pmatrix} = \mathbf{0}$$

debido a la restricción $||\boldsymbol{\theta}|| = 1$, siendo ι_n el multiplicador de Lagrange asociado, y puesto que todas las funciones asociadas son diferenciables.

Si llamamos

$$\Psi_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{\widehat{\sigma}} \frac{1}{n} \sum_{i=1}^{n} \psi_1 \left(\frac{y_i - \boldsymbol{\beta}^t \mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t \mathbf{t}_i)}{\widehat{\sigma}} \right) \left(\mathbf{x}_i + \frac{\partial}{\partial \boldsymbol{\beta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{u=\boldsymbol{\theta}^t \mathbf{t}_i} \right) w_2(\mathbf{x}_i) \,,$$

y

$$\Psi_{2n}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{\widehat{\sigma}} \frac{1}{n} \sum_{i=1}^{n} \psi_1 \left(\frac{y_i - \boldsymbol{\beta}^t \mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t \mathbf{t}_i)}{\widehat{\sigma}} \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{u=\boldsymbol{\theta}^t \mathbf{t}_i} + \frac{\partial}{\partial u} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{u=\boldsymbol{\theta}^t \mathbf{t}_i} \mathbf{t}_i \right) w_2(\mathbf{x}_i) \,,$$

tenemos

$$\mathbf{0} = \Psi_n(\boldsymbol{\beta}, \boldsymbol{\theta}) + \iota_n \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \Psi_{1n}(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ \Psi_{2n}(\boldsymbol{\beta}, \boldsymbol{\theta}) \end{pmatrix} + \iota_n \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\theta} \end{pmatrix}. \tag{4.2}$$

Haciendo una expansión de Taylor de orden 1

$$\mathbf{0} = \left[\iota_n \begin{pmatrix} \mathbf{0} \\ \widehat{\boldsymbol{\theta}} \end{pmatrix} + \sqrt{n}\Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \mathbf{M}_n \sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \right] \frac{1}{\sqrt{n}} \tag{4.3}$$

con $\mathbf{M}_n = D\Psi_n \left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}} \right)$, siendo $D\Psi_n$ la matriz diferencial de Ψ_n, $\tilde{\boldsymbol{\beta}}$ un punto intermedio entre $\boldsymbol{\beta}_0$ y $\widehat{\boldsymbol{\beta}}$ y $\tilde{\boldsymbol{\theta}}$ intermedio entre $\boldsymbol{\theta}_0$ y $\widehat{\boldsymbol{\theta}}$.

Llamaremos

$$\widehat{\boldsymbol{\lambda}}(\mathbf{x}, \mathbf{t}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, u_0) = \begin{pmatrix} \mathbf{x} + \frac{\partial}{\partial \boldsymbol{\beta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u_0)} \\ \frac{\partial}{\partial \boldsymbol{\theta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u_0)} + \frac{\partial}{\partial u} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u_0)} \mathbf{t} \end{pmatrix}$$

Para simplificar la notación, denotemos $\boldsymbol{\tau} = (\mathbf{b}, \mathbf{a}, r)$, siendo $\mathbf{b} \in \mathbb{R}^p$, $\mathbf{a} \in \mathcal{S}^1$, $r \in \mathbb{R}$ y definamos

$$\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}) := \widehat{\boldsymbol{\lambda}}(\mathbf{x}_i, \mathbf{t}_i, \boldsymbol{\tau}) = \begin{pmatrix} \mathbf{x}_i + \frac{\partial}{\partial \boldsymbol{\beta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \\ \frac{\partial}{\partial \boldsymbol{\theta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} + \frac{\partial}{\partial u} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \mathbf{t}_i \end{pmatrix}$$

y

$$\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}) := \widehat{\boldsymbol{\Gamma}}(\mathbf{t}_i, \boldsymbol{\tau}) = \begin{pmatrix} (\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{11} & (\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{12} \\ (\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{21} & (\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{22} \end{pmatrix}$$

donde

$$(\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{11} = \frac{\partial^2}{\partial \boldsymbol{\beta}^2} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}}$$

$$(\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{12} = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} + \frac{\partial^2}{\partial \boldsymbol{\beta} \partial u} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \mathbf{t}_i^t$$

$$(\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{21} = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\beta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} + \mathbf{t}_i \left(\frac{\partial^2}{\partial u \partial \boldsymbol{\beta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \right)^t$$

y

$$\begin{aligned} (\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\tau}))_{22} &= \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \right)^t + \mathbf{t}_i \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial u} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \right)^t \\ &+ \frac{\partial^2}{\partial u \partial \boldsymbol{\theta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \mathbf{t}_i^t + \left(\frac{\partial^2}{\partial u^2} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}} \right) \mathbf{t}_i \mathbf{t}_i^t \, . \end{aligned}$$

Asimismo, llamaremos $\boldsymbol{\lambda}_i(\boldsymbol{\tau})$ y $\boldsymbol{\Gamma}_i(\boldsymbol{\tau})$ a las respectivas versiones de $\widehat{\boldsymbol{\lambda}}_i$ y $\widehat{\boldsymbol{\Gamma}}_i$ que involucran a la función $\eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ en lugar de $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$.

Para obtener la expresión de la distribución asintótica, definiremos

$$\widehat{\mathbf{V}}_n = \sqrt{n}\Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i})\,, \tag{4.4}$$

siendo para cada $i \in \mathbb{N}$, $\boldsymbol{\tau}_{0i} = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\theta}_0^t\mathbf{t}_i)$ y la matriz diferencial de Ψ_n evaluada en $(\boldsymbol{\beta}, \boldsymbol{\theta})$:

$$\begin{aligned} D\Psi_n(\boldsymbol{\beta}, \boldsymbol{\theta}) =& \frac{1}{\widehat{\sigma}}\frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}\left(\boldsymbol{\theta}^t\mathbf{t}_i\right)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\theta}^t\mathbf{t}_i) \\ &- \frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}\left(\boldsymbol{\theta}^t\mathbf{t}_i\right)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\theta}^t\mathbf{t}_i)\widehat{\boldsymbol{\lambda}}_i^t(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\theta}^t\mathbf{t}_i)\,. \end{aligned} \tag{4.5}$$

Por lo tanto, resulta que

$$\begin{aligned} \mathbf{M}_n =& \frac{1}{\widehat{\sigma}}\frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}}(\tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\widehat{\boldsymbol{\Gamma}}_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}^t\mathbf{t}_i) \\ &- \frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}}(\tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\widehat{\boldsymbol{\lambda}}_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)\widehat{\boldsymbol{\lambda}}_i^t(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)\,. \end{aligned} \tag{4.6}$$

Denotemos

$$\mathbf{M} = -\frac{1}{\sigma_0^2}\mathbb{E}(\chi_1(\epsilon_1)w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})\boldsymbol{\lambda}_1^t(\boldsymbol{\tau}_{01}))\,, \tag{4.7}$$

$$\mathbf{V}_n = \frac{1}{\sigma_0}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_1(\epsilon_i)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) \tag{4.8}$$

y para $j \in \mathbb{N}$ definamos

$$E_j(u) := E(y_j, \mathbf{x}_j, u) = \frac{\sigma_0}{A^1(u)}\psi_1\left(\frac{y_j - \boldsymbol{\beta}_0^t\mathbf{x}_j - \eta_0(u)}{\sigma_0}\right) w_1(\mathbf{x}_j) \tag{4.9}$$

donde $A^1(u) = \mathbb{E}(\chi_1(\epsilon_1)w_1(\mathbf{x}_1)|\boldsymbol{\theta}_0^t\mathbf{t}_1 = u)$. El comportamiento asintótico de los estimadores requerirá también de

$$\boldsymbol{\gamma}(\mathbf{t}) = \mathbb{E}\left(\frac{1}{f_{\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_1\right)}\chi_1(\epsilon_1)w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})\middle|\mathbf{t}_1 = \mathbf{t}\right)\,, \tag{4.10}$$

donde recordemos que $f_{\boldsymbol{\theta}_0}$ es la densidad marginal de $\boldsymbol{\theta}_0^t\mathbf{t}$,

$$\mathbf{w} = \frac{1}{\sigma_0}\psi_1(\epsilon_1)w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) - \frac{1}{\sigma_0^2}E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)\boldsymbol{\gamma}(\mathbf{t}_1)f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_1)\,,$$

$$\boldsymbol{\Sigma} = \mathbb{E}\left(\mathbf{w}\mathbf{w}^t\right) . \tag{4.11}$$

Definimos las función

$$m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \mathbb{E}(E_1(\boldsymbol{\theta}_0^t \mathbf{t}_2) \mid (\mathbf{t}_1, \mathbf{t}_2) = (\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)) . \tag{4.12}$$

N1. $\mathcal{T} \subset \mathbb{R}^q$ es un conjunto compacto, convexo y con interior no vacío.

N2. i) Las funciones $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ y $\eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ son tres veces diferenciables en $(\boldsymbol{\beta}, \boldsymbol{\theta}, u)$ con derivadas parciales terceras continuas.

ii) $||\widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}} - \eta_0||_\infty \xrightarrow{P} 0$ siendo $\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}\right)$ un estimador consistente de $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$.

iii) Para cada $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\theta} \in \mathbb{R}^q$ y $\mathbf{t} \in \mathcal{T}$ fijos, tenemos que $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t}) - \eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(\boldsymbol{\theta}^t\mathbf{t}) \xrightarrow{P} 0$ y $n^{1/4}||\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0} - \eta_0||_\infty \xrightarrow{P} 0$. Asimismo, para $\mathbf{v} = (\boldsymbol{\beta}, \boldsymbol{\theta}, u)$

$$n^{1/4} \left\| \frac{\partial}{\partial v_j} \left(\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \eta_{\boldsymbol{\beta},\boldsymbol{\theta}}\right)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta})=(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)} \right\|_\infty \xrightarrow{P} 0 \quad \text{para } j = 1, ..., p+q+1$$

y

$$\left\| \frac{\partial^k}{\partial u^k} \left(\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \eta_{\boldsymbol{\beta},\boldsymbol{\theta}}\right)\Big|_{(\boldsymbol{\beta},\boldsymbol{\theta})=(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)} \right\|_\infty \xrightarrow{P} 0 \quad k = 2, 3.$$

N3. ρ_1 es una ρ-función tal que $\rho_1 \in \mathcal{C}^3(\mathbb{R})$. Sus sucesivas derivadas son $\psi_1 := \rho_1'$, $\chi_1 := \rho_1''$ y χ_1', todas acotadas. Además las funciones $\chi_1(e)e$, $\chi_1'(e)e$ y $\chi_1'(e)e^2$ son acotadas.

N4. Las funciones w_1 y w_2 son no negativas, continuas y acotadas. Además, la función $w_2(\mathbf{x})\mathbf{x}$ es continua y acotada.

N5. $\mathbb{E}(w_2(\mathbf{x})||\mathbf{x}||^2) < \infty$.

N6. $f_{\boldsymbol{\theta}_0}$ es una función acotada y derivable con derivada acotada y tal que $\inf_{\mathbf{t}\in\mathcal{T}} f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}) > 0$.

N7. Las funciones $E(y, \mathbf{x}, u)$, su derivada parcial de primer y segundo orden respecto de u, $m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$, $m_1'(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \dfrac{\partial m_1}{\partial \boldsymbol{\nu}_2}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$ y $m_1''(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \dfrac{\partial^2 m_1}{\partial \boldsymbol{\nu}_2 \partial \boldsymbol{\nu}_2^t}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$ son continuas y acotadas.

N8. $K : \mathbb{R} \to \mathbb{R}$ es una función no negativa, par, acotada y de variación acotada tal que $\int K(u)\, du = 1$, $\int uK(u)\, du = 0$, $\int u^2K(u)\, du < \infty$ y $\int u^3K(u)\, du < \infty$.

N9. La matriz $\mathbf{M}_1 \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ es no singular, donde $\mathbf{M}_1$ es la submatriz cuadrada superior izquierda de $\mathbf{M}$ definida en (4.7).

N10. La matriz $\mathbf{\Sigma}_1 \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ es no singular, siendo $\mathbf{\Sigma}_1$ la submatriz cuadrada superior izquierda de $\mathbf{\Sigma}$ definida en (4.11).

Observación 4.1.1. Los supuestos anteriores son similares a los realizados en el contexto de robustez en modelos semiparamétricos, tales como Rodriguez (2008), Boente y Rodriguez (2012) y Agostinelli et al. (2020). Más precisamente, la condición **N1** es habitual en el contexto de robustez en modelos de índice simple, ver, por ejemplo, Agostinelli et al. (2020). La condición **N2** es similar al supuesto **M1** de Boente y Rodriguez (2012). En la Observación 4.1 del citado trabajo se nota que el hecho de que el núcleo sea continuamente diferenciable y el Teorema de la Función Implícita implican que $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ es continuamente diferenciable en u y dan argumentos para derivar la convergencia uniforme requerida. Por otro lado, las condiciones **N3**, **N4**, **N5**, **N9** y **N10** son habituales en el ámbito de la robustez en el campo de modelos de regresión y resultan adaptaciones naturales al modelo tratado. En particular, **N9** y **N10** son suposiciones comunes de las regresiones robustas que permiten obtener estimadores de $\boldsymbol{\beta}_0$ y $\boldsymbol{\theta}_0$ con tasa $\sqrt{n}$. Por último, las condiciones **N6**-**N8** son habituales en robustez en el ámbito de modelos de índice simple, ver por ejemplo Boente y Rodriguez (2012) y Agostinelli et al. (2020).

El siguiente lema da la convergencia en probabilidad de la matriz $\mathbf{M}_n$.

Lema 4.1.2. *Sean $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\theta}}$ y $\widehat{\sigma}$ estimadores débilmente consistentes de $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ y σ_0, respectivamente. Supongamos que $\mathcal{T} \subset \mathbb{R}^q$ es un conjunto compacto, que la sucesión de vectores aleatorios $\{\mathbf{t}_i\}_{i\in\mathbb{N}}$ está contenida en $\mathcal{T}$ y que se satisfacen las condiciones **N2**-**N5**. Entonces $\mathbf{M}_n \xrightarrow{P} \mathbf{M}$, cuyas respectivas expresiones se encuentran en (4.6) y (4.7).*

Demostración: Denotemos por $\tilde{\boldsymbol{\tau}}_i = (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)$ con $i = 1, ..., n$. Para demostrar que $\mathbf{M}_n$ converge en probabilidad a $\mathbf{M}$, descompondremos $\mathbf{M}_n = \sum_{i=1}^{6} \mathbf{M}_n^{(i)}$ siendo

$$\mathbf{M}_n^{(1)} = \frac{1}{\widehat{\sigma}}\frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\mathbf{\Gamma}_i(\boldsymbol{\tau}_{0i})\,,$$

$$\mathbf{M}_n^{(2)} = \frac{1}{\widehat{\sigma}}\frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}}(\tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\left[\widehat{\mathbf{\Gamma}}_i(\tilde{\boldsymbol{\tau}}_i) - \mathbf{\Gamma}_i(\boldsymbol{\tau}_{0i})\right],$$

$$\mathbf{M}_n^{(3)} = -\frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \tilde{\zeta}_i}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\mathbf{\Gamma}(\boldsymbol{\tau}_{0i})\left[\widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}}(\tilde{\boldsymbol{\theta}}^t\mathbf{t}_i) - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right],$$

$$\mathbf{M}_n^{(4)} = -\frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\boldsymbol{\lambda}_i^t(\boldsymbol{\tau}_{0i})\,,$$

$$\mathbf{M}_n^{(5)} = -\frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}}(\tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\left[\widehat{\boldsymbol{\lambda}}_i(\tilde{\boldsymbol{\tau}}_i)\widehat{\boldsymbol{\lambda}}_i^t(\tilde{\boldsymbol{\tau}}_i) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\boldsymbol{\lambda}_i^t(\boldsymbol{\tau}_{0i})\right],$$

$$\mathbf{M}_n^{(6)} = \frac{1}{\widehat{\sigma}^3}\frac{1}{n}\sum_{i=1}^{n}\chi_1'\left(\frac{y_i - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_i - \tilde{\xi}_i}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\boldsymbol{\lambda}_i^t(\boldsymbol{\tau}_{0i})\left[\widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}}(\tilde{\boldsymbol{\theta}}^t\mathbf{t}_i) - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right],$$

donde para cada $i \in \mathbb{N}$, tanto $\tilde{\xi}_i$ como $\tilde{\zeta}_i$ son puntos intermedios entre $\widehat{\eta}_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\theta}}}(\tilde{\boldsymbol{\theta}}^t\mathbf{t}_i)$ y $\eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_i)$.

Por las condiciones **N2-N5** y la consistencia débil de los estimadores $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\theta}}$ y $\widehat{\sigma}$, se puede probar en forma directa que $\mathbf{M}_n^{(j)} \xrightarrow{P} 0$ para $j = 2, 3, 5, 6$.

Del Teorema de Convergencia Dominada y la consistencia de $\tilde{\boldsymbol{\beta}}$ y $\widehat{\sigma}$ obtenemos que

$$\mathbb{E}\left[\frac{1}{\widehat{\sigma}^2}\chi_1\left(\frac{y_1 - \tilde{\boldsymbol{\beta}}^t\mathbf{x}_1 - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_1)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})\boldsymbol{\lambda}_1^t(\boldsymbol{\tau}_{01})\right] \to \mathbb{E}\left[\frac{1}{\sigma_0^2}\chi_1\left(\epsilon_1\right) w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})\boldsymbol{\lambda}_1^t(\boldsymbol{\tau}_{01})\right].$$

Usando argumentos similares a los empleados en el Lema 1 de Bianco y Boente (2002) resulta que para cualquier conjunto compacto $\mathcal{K} \subset \mathbb{R}^p$ y para cualquier intervalo compacto $\mathcal{S} \subset \mathbb{R}_{>0}$

$$\sup_{\boldsymbol{\beta}\in\mathcal{K},\sigma\in\mathcal{S}}\left|\frac{1}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\sigma}\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\boldsymbol{\lambda}_i^t(\boldsymbol{\tau}_{0i}) - \mathbb{E}\left[\frac{1}{\sigma^2}\chi_1\left(\frac{y_1 - \boldsymbol{\beta}^t\mathbf{x}_1 - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_1)}{\sigma}\right) w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})\boldsymbol{\lambda}_1^t(\boldsymbol{\tau}_{01})\right]\right| \xrightarrow{P} 0\,.$$

Por lo tanto, se tiene que $\mathbf{M}_n^{(4)} \xrightarrow{P} \mathbf{M}$. De la misma forma, teniendo en cuenta que ψ_1 es una función impar y acotada, que los errores se distribuyen de manera simétrica respecto del 0 y son independientes de las covariables, resulta que

$$\mathbb{E}\left[\psi_1\left(\frac{y_1 - \boldsymbol{\beta}_0^t\mathbf{x}_1 - \eta_0(\boldsymbol{\theta}_0^t\mathbf{t}_1)}{\sigma}\right) w_2(\mathbf{x}_1)\boldsymbol{\Gamma}(\boldsymbol{\tau}_{01})\right] = 0\,,$$

por lo tanto de la misma manera que con $\mathbf{M}_n^{(4)}$, concluímos que $\mathbf{M}_n^{(1)} \xrightarrow{P} 0$, quedando demostrado el lema. ■

El resultado del siguiente lema será necesario para la obtención de la distribución asintótica de nuestro estimador. En el ítem i) se obtiene un resultado similar al resultado parcial que se establece en la prueba del Teorema 2 de Carroll et al. (1997) y en el Lema 4.1 de Boente y Rodriguez (2010), mientras que la parte ii) será útil en la demostración del Teorema 4.1.4.

L1. $K : \mathbb{R} \to \mathbb{R}$ es una función no negativa, par y acotada, satisfaciendo una condición de Lipschitz de orden uno y tal que $\int K(u)du = 1$ y $\int uK(u)du = 0$.

L2. $\inf_{u\in\mathcal{U}_0(\mathcal{T})} \mathbb{E}(\chi_1(\epsilon_1)w_1(\mathbf{x}_1)|\boldsymbol{\theta}_0^t\mathbf{t}_1 = u) > 0$ siendo $\mathcal{U}_0(\mathcal{T}) := \{\boldsymbol{\theta}_0^t\mathbf{t} : \mathbf{t} \in \mathcal{T}\}$.

Lema 4.1.3. *Supongamos que* $n^{1/4}||\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0} - \eta_0||_\infty \xrightarrow{P} 0$ *y que* $\widehat{\sigma}$ *un estimador tal que* $n^{1/4}(\widehat{\sigma} - \sigma_0) \xrightarrow{P} 0$.

i) Si se cumplen las condiciones ***N1****,* ***N3****,* ***N4****,* ***N6****,* ***L1*** *y* ***L2****,* $nh^4 \longrightarrow 0$ *y* $nh^2/\ln^2(h) \longrightarrow +\infty$*, entonces*

$$\sup_{\tilde{\mathbf{t}}\in\mathcal{T}} \left| \widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\tilde{\mathbf{t}}) - \eta_0(\boldsymbol{\theta}_0^t\tilde{\mathbf{t}}) - \frac{1}{n f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\tilde{\mathbf{t}})} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t\mathbf{t}_j - \boldsymbol{\theta}_0^t\tilde{\mathbf{t}}) E_j(\boldsymbol{\theta}_0^t\tilde{\mathbf{t}}) \right| = o_P\left(n^{-1/2}\right) , \tag{4.13}$$

donde $E_j(u)$ *está definida en* (4.9)*.*

ii) Bajo ***N3*** *y* ***N5*** *, si* $\tilde{\sigma}$ *es un punto intermedio entre* $\widehat{\sigma}$ *y* σ_0*, entonces*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \chi_1'\left(\frac{\sigma_0\epsilon_i}{\tilde{\sigma}}\right)\left(\frac{\sigma_0\epsilon_i}{\tilde{\sigma}}\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) \left[\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right) - \widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)\right](\widehat{\sigma} - \sigma_0) \xrightarrow{P} 0\,. \tag{4.14}$$

Demostración: *i*) Tenemos que

$$\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - a}{\widehat{\sigma}}\right) w_1(\mathbf{x}_i) W^h_{\boldsymbol{\theta}_0,i}(u)\,,$$

que corresponde al estimador introducido en el Paso 1 de la propuesta, pero usando los verdaderos valores de los parámetros de regresión e índice simple. Derivando respecto de la variable a y evaluando en este estimador, tenemos que para todo u

$$\frac{1}{n}\sum_{i=1}^{n} \psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u)}{\widehat{\sigma}}\right) w_1(\mathbf{x}_i) W^h_{\boldsymbol{\theta}_0,i}(u) = 0$$

donde $\displaystyle\sum_{j=1}^{n} K\left(\frac{\boldsymbol{\theta}_0^t\mathbf{t}_j - u}{h}\right) \neq 0$. Mediante una expansión de Taylor de orden 2 alrededor de $(\sigma_0, \eta_0(u))$ resulta

$$\begin{aligned} 0 =& \mathcal{W}_n(u) + A_n^1(u)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u)) + A_n^2(u)(\widehat{\sigma} - \sigma_0) \\ &+ \frac{1}{2}[A_n^3(u)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u))^2 + A_n^4(u)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u))(\widehat{\sigma} - \sigma_0) + A_n^5(u)(\widehat{\sigma} - \sigma_0)^2] \end{aligned} \tag{4.15}$$

donde

$$\begin{aligned}
\mathcal{W}_n(u) &= \frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_0(u)}{\sigma_0}\right)w_1(\mathbf{x}_i)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-u)\\
A_n^1(u) &= -\frac{1}{\sigma_0}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_0(u)}{\sigma_0}\right)w_1(\mathbf{x}_i)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-u)\\
A_n^2(u) &= -\frac{1}{\sigma_0}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_0(u)}{\sigma_0}\right)\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_0(u)}{\sigma_0}w_1(\mathbf{x}_i)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-u)\\
A_n^3(u) &= \frac{1}{\tilde{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1'\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\tilde{\xi}}{\tilde{\sigma}}\right)w_1(\mathbf{x}_i)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-u)
\end{aligned}$$

$$\begin{aligned}
A_n^4(u) &= \frac{2}{\tilde{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\left[\chi_1'\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\tilde{\xi}}{\tilde{\sigma}}\right)\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\tilde{\xi}}{\tilde{\sigma}}+\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\tilde{\xi}}{\tilde{\sigma}}\right)\right]\\
&\quad w_1(\mathbf{x}_i)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-u)\\
A_n^5(u) &= \frac{1}{\tilde{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\iota\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\tilde{\xi}}{\tilde{\sigma}}\right)w_1(\mathbf{x}_i)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-u)
\end{aligned}$$

con $\iota(z)=\chi_1'(z)z^2+2\chi_1(z)z$ y $(\tilde{\sigma},\tilde{\xi})$ un punto en el segmento que une $(\widehat{\sigma},\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u))$ con $(\sigma_0,\eta_0(u))$.
Como $\mathcal{T}$ es un conjunto compacto, $nh\longrightarrow+\infty$ y se cumplen las condiciones **N1** y **N6**, usando el Lema A.1 de Boente y Rodriguez (2010), tenemos que

$$\sup_{\tilde{\mathbf{t}}\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-\boldsymbol{\theta}_0^t\tilde{\mathbf{t}})-f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\tilde{\mathbf{t}})\right|\xrightarrow{P}0\,.$$

Por lo tanto,

$$\sup_{\tilde{\mathbf{t}}\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^{n}K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i-\boldsymbol{\theta}_0^t\tilde{\mathbf{t}})=O_P(1)\,. \tag{4.16}$$

Usando (4.16), que $\widehat{\sigma}$ es un estimador consistente de σ_0 y las condiciones **L1** y **N3**, resulta que para $j=3,4,5$ $||A_n^j||_{\mathcal{U}_0(\mathcal{T}),\infty}=O_P(1)$.

Ahora veamos que para $j=1,2$ se tiene que $||A_n^j-f_{\boldsymbol{\theta}_0}A^j||_{\mathcal{U}_0(\mathcal{T}),\infty}=O_P(h)+o_P\left(n^{-1/4}\right)$. Como $nh^4\longrightarrow 0$ y $nh^2/\ln^2(h)\longrightarrow+\infty$, podemos usar nuevamente el Lema A.1 de Boente y Rodriguez (2010), pero en este caso el ítem b), obteniendo que

$A_n^j(u) = \mathbb{E}(A_n^j(u)) + o_P\left(n^{-1/4}\right)$ uniformemente para $u \in \mathcal{U}_0(\mathcal{T})$. Calculemos $\mathbb{E}(A_n^1(u))$:

$$\begin{aligned}\mathbb{E}(A_n^1(u)) &= -\frac{1}{\sigma_0}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\chi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \eta_0(u)}{\sigma_0}\right)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i - u)w_1(\mathbf{x}_i)\right)\\ &= -\frac{1}{\sigma_0}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i - u)\mathbb{E}\left(\chi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \eta_0(u)}{\sigma_0}\right)w_1(\mathbf{x}_i)\middle|\boldsymbol{\theta}_0^t\mathbf{t}_i\right)\right)\\ &= -\frac{1}{\sigma_0}\mathbb{E}(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i - u)\tilde{A}^1(u, \boldsymbol{\theta}_0^t\mathbf{t}_i))\,.\end{aligned}$$

Mediante un cambio de variables en el integrando, el Teorema del Valor Medio y el hecho de que el núcleo K es no negativo, par y cumple $\int uK(u)du = 0$, deducimos que

$$\sup_{u\in\mathcal{U}_0(\mathcal{T})}|\mathbb{E}(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_i - u)\tilde{A}^1(u, \boldsymbol{\theta}_0^t\mathbf{t}_i)) - A^1(u)f_{\boldsymbol{\theta}_0}(u)| = O(h)\,.$$

Luego, $A_n^1(u) = -A^1(u)f_{\boldsymbol{\theta}_0}(u)/\sigma_0 + O(h) + o_P\left(n^{-1/4}\right)$.

Como la función $\chi_1(e)e + \psi_1(e)$ es una función impar y acotada, ϵ tiene distribución simétrica respecto de 0 y los errores son independientes de las covariables, resulta que $A^2(u) = \mathbb{E}[(\chi_1(\epsilon_1)\epsilon_1 + \psi_1(\epsilon_1))w_1(\mathbf{x}_1)|\boldsymbol{\theta}_0^t\mathbf{t}_1 = u] = 0$. Por lo tanto, obtenemos que $A_n^2(u) = O(h) + o_P\left(n^{-1/4}\right)$. Reemplazando lo obtenido en (4.15), conseguimos la ecuación:

$$\begin{aligned}-W_n(u) =& -A^1(u)f_{\boldsymbol{\theta}_0}(u)/\sigma_0\left(1 + O(h) + o_P\left(n^{-1/4}\right)\right)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u))\\ &+ \left(O(h) + o_P\left(n^{-1/4}\right)\right)(\widehat{\sigma} - \sigma_0) + O_P(1)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u))^2\\ &+ O_P(1)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u))(\widehat{\sigma} - \sigma_0) + O_P(1)(\widehat{\sigma} - \sigma_0)^2\,.\end{aligned}$$

Por las condiciones **N6** y **L2**, $\inf_{u\in\mathcal{U}_0(\mathcal{T})} A^1(u)f_{\boldsymbol{\theta}_0}(u) > 0$. Luego para $u \in \mathcal{U}_0(\mathcal{T})$,

$$\begin{aligned}\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u) - \frac{1}{nf_{\boldsymbol{\theta}_0}(u)}\sum_{j=1}^{n}K_h(\boldsymbol{\theta}_0^t\mathbf{t}_j - u)E_j(u) =& \left(O(h) + o_P\left(n^{-1/4}\right)\right)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u) + \widehat{\sigma} - \sigma_0)\\ &+ O_P(1)(\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u) - \eta_0(u) + \widehat{\sigma} - \sigma_0)^2.\end{aligned}$$

En conclusión, probamos i) usando que $nh^4 \longrightarrow 0$, que $n^{1/4}||\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0} - \eta_0||_\infty \xrightarrow{P} 0$ y que $n^{1/4}(\widehat{\sigma} - \sigma_0) \xrightarrow{P} 0$.

ii) El resultado es una consecuencia directa de **N3**, **N5** y de que $n^{1/4}||\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0} - \eta_0||_\infty \xrightarrow{P} 0$ y $n^{1/4}(\widehat{\sigma} - \sigma_0) \xrightarrow{P} 0$. ■

En la demostración del siguiente teorema haremos uso de algunas nociones relacionadas a procesos empíricos. Haremos referencia a resultados que pueden consultarse en van der Vaart y Wellner (1996) y Pollard (1984), pero a los fines de completitud recordamos algunas definiciones y resultados.

La noción de $\mathbb{Q}-$medibilidad de una clase de funciones $\mathcal{F}$ puede hallarse en van der Vaart y Wellner (1996) pág. 110 y es necesaria para garantizar la medibilidad de supremos sobre la clase $\mathcal{F}$.

Sea $\mathcal{F}$ una clase de funciones con envolvente F, es decir $|f| \leq F, \forall f \in \mathcal{F}$, y $\mathbb{Q}$ una medida de probabilidad. Dadas dos funciones f y g, el *bracket* $[f, g]$ es el conjunto de funciones ℓ con $f \leq \ell \leq g$ y un $\varepsilon-$bracket es un bracket $[f, g]$ con $\|g - f\|_{\mathbb{Q},2} < \varepsilon$, donde $\|f\|_{\mathbb{Q},r} = (E_{\mathbb{Q}}(|f|^r))^{\frac{1}{r}}$.

Denotamos $N_{[\,]}(\varepsilon, \mathcal{F}, L^2(\mathbb{Q}))$, al número bracketing, es decir al mínimo número de $\varepsilon-$brackets necesarios para cubrir $\mathcal{F}$. Las cotas inferiores y superiores pueden no pertenecer a la clase $\mathcal{F}$, pero deben tener norma $\|.\|_{\mathbb{Q},2}$ finita. Llamaremos $J_{[\,]}(\delta, \mathcal{F})$ a la integral bracketing definida como

$$J_{[\,]}(\delta, \mathcal{F}) = \int_0^{\delta} \sqrt{1 + \log\left(N_{[\,]}\left(\varepsilon \|F\|_{\mathbb{Q},2}, \mathcal{F}, L^2(\mathbb{Q})\right)\right)}\, d\varepsilon\,.$$

La función $J_{[\,]}$ es creciente, $J_{[\,]}(0, \mathcal{F}) = 0$ y $J_{[\,]}(1, \mathcal{F}) < \infty$. Sean $\mathbf{X}_1, \ldots, \mathbf{X}_n$ vectores aleatorios i.i.d. con distribución común P, $J_{[\,]}(\delta, \mathcal{F}) \to 0$ cuando $\delta \to 0$ para clases de funciones $\mathcal{F}$ que satisfacen la condición de entropía bracketing, esto es,

$$\int_0^{\infty} \sqrt{\log\left(N_{[\,]}\left(\varepsilon \|F\|_{P,2}, \mathcal{F}, L^2(P)\right)\right)}\, d\epsilon \, < \infty\,. \tag{4.17}$$

Un resultado que será de gran utilidad está dado por la siguiente desigualdad.

Desigualdad Maximal para Números Bracketing. *Sean $\mathbf{X}_1, \ldots, \mathbf{X}_n$ vectores aleatorios i.i.d. con distribución común P. Sea $\mathcal{F}$ una clase $\mathbb{P}-$medible de funciones con envolvente F tal que $\|F\|_{\mathbb{P},2} = [\mathbb{E}_{\mathbb{P}}(F^2)]^{\frac{1}{2}} < \infty$. Dado $\delta > 0$, sea*

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

Luego, si $\|f\|_{\mathbb{P},2} < \delta \|F\|_{\mathbb{P},2}$ para toda $f \in \mathcal{F}$, existe una constante A independiente de n, tal que

$$\left\| \sup_{f \in \mathcal{F}} |T_n(f)| \right\|_{\mathbb{P},1} \leq A\, J_{[\,]}(\delta, \mathcal{F})\, \|F\|_{\mathbb{P},2} + \sqrt{n}\, \mathbb{E}_{\mathbb{P}}\left(F\, I_{\{F > \sqrt{n}\, a(\delta)\}}\right) \leq A\, J_{[\,]}(1, \mathcal{F})\, \|F\|_{\mathbb{P},2}\,,$$

donde $T_n(f) = \sqrt{n}\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \left(f(\mathbf{X}_i) - \mathbb{E}\left(f(\mathbf{X}_1)\right)\right)\right)$.

Otro concepto relacionado es el del número de cubrimiento. Denotamos $N(\varepsilon, \mathcal{F}, L^2(\mathbb{Q}))$ al número de cubrimiento, esto es al mínimo número de bolas de radio ε,
$\mathcal{B}(\varepsilon, g) = \{h : \|h - g\|_{\mathbb{Q},2} < \varepsilon\}$, necesarias para cubrir $\mathcal{F}$. Los centros de las bolas no necesitan pertenecer a $\mathcal{F}$, pero ellos deben tener norma $\|.\|_{\mathbb{Q},2}$ finita.

En particular, las clases de funciones monótonas y acotadas y las clases de funciones que son Lipschitz en un parámetro satisfacen (4.17) si, por ejemplo, el espacio paramétrico es

acotado y tiene un número de cubrimiento finito (ver van der Vaart y Wellner (1996) pág. 164).

Definimos la integral

$$J(\delta, \mathcal{F}) = \sup_{\mathbb{Q}} \int_0^{\delta} \sqrt{1 + \log\left(N\left(\varepsilon \left\|F\right\|_{\mathbb{Q},2}, \mathcal{F}, L^2(\mathbb{Q})\right)\right)}\, d\varepsilon\,,$$

donde el supremo es tomado sobre todas las medidas de probabilidad discretas $\mathbb{Q}$ con $\|F\|_{\mathbb{Q},2} > 0$. La función J es creciente, $J(0, \mathcal{F}) = 0$, $J(1, \mathcal{F}) < \infty$ y además $J(\delta, \mathcal{F}) \to 0$ a medida que $\delta \to 0$ para las clases de funciones $\mathcal{F}$ que satisfacen la condición uniforme de entropía, es decir,

$$\int_0^{\infty} \sup_{\mathbb{Q}} \sqrt{\log\left(N\left(\varepsilon \left\|F\right\|_{\mathbb{Q},2}, \mathcal{F}, L^2(\mathbb{Q})\right)\right)}\, d\varepsilon \; < \infty\,. \tag{4.18}$$

En particular, las clases de funciones de Vapnik-Cervonenkis satisfacen (4.18). La siguiente es la versión de la desigualdad maximal para el número de cubrimiento.

Desigualdad Maximal para el número de cubrimiento. *Sean $\mathbf{X}_1, \ldots, \mathbf{X}_n$ vectores aleatorios i.i.d. con distribución común P. Sea $\mathcal{F}$ una clase $\mathbb{P}-$medible de funciones con envolvente F tal que $\|F\|_{\mathbb{P},2} = [\mathbb{E}_{\mathbb{P}}(F^2)]^{\frac{1}{2}} < \infty$. Luego, existe una constante C_1 que no depende de n, tal que*

$$\begin{aligned} \left\| \sup_{f \in \mathcal{F}} |T_n(f)| \, \right\|_{\mathbb{P},2} &\leq C_1 \left\| J\left(\delta_n, \mathcal{F}\right) \left[\frac{1}{n} \sum_{i=1}^{n} F^2\left(X_i\right) \right]^{\frac{1}{2}} \right\|_{\mathbb{P},2} \\ &\leq C_1\, J(1, \mathcal{F})\, \|F\|_{\mathbb{P},\,2}\,. \end{aligned}$$

siendo

$$\delta_n = \frac{\sup_{f \in \mathcal{F}} \sqrt{\dfrac{1}{n} \sum_{i=1}^{n} f^2(\mathbf{X}_i)}}{\sqrt{\dfrac{1}{n} \sum_{i=1}^{n} F^2(\mathbf{X}_i)}}\,.$$

De aquí en más, dado un vector $\boldsymbol{\theta} \in \mathbb{R}^q$, llamaremos $\boldsymbol{\theta}^{(q-1)}$ a sus primeras $q-1$ componentes, así por ejemplo, $\widehat{\boldsymbol{\theta}}^{(q-1)} = (\hat{\theta}_1, ..., \hat{\theta}_{q-1})^t$ y $\boldsymbol{\theta}_0^{(q-1)} = (\theta_{01}, ..., \theta_{0q-1})^t$.

Teorema 4.1.4. *Sea $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right)$ una solución de (3.4) que estime consistentemente a $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$. Si se cumplen las condiciones **N1-N10**, la sucesión de vectores aleatorios $\{\mathbf{t}_i\}_{i \in \mathbb{N}}$ está contenida en $\mathcal{T}$, i) y ii) del Lema 4.1.3 se verifican, $\widehat{\sigma} \xrightarrow{P} \sigma_0$, $\lim\limits_{n\to\infty} nh = +\infty$ y $\lim\limits_{n\to\infty} nh^4 = 0$, entonces*

a) $\sqrt{n}(\hat{\theta}_q - \theta_{0q}) \xrightarrow{P} 0\,.$

b) $\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}^{(q-1)} - \boldsymbol{\theta}_0^{(q-1)} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{M}_1^{-1}\boldsymbol{\Sigma}_1\mathbf{M}_1^{-1})$ *donde* $\mathbf{M}_1$ *y* $\boldsymbol{\Sigma}_1$ *están dadas en* ***N9*** *y* ***N10****, respectivamente.*

Demostración: Como consecuencia del Lema 3.3.2, bastará con demostrar que el resultado vale para $\boldsymbol{\theta}_0 = \mathbf{e}_q$ el q-ésimo vector canónico de $\mathbb{R}^q$.

De (4.3) tenemos que

$$\mathbf{0} = \iota_n \begin{pmatrix} \mathbf{0} \\ \widehat{\boldsymbol{\theta}} \end{pmatrix} + \frac{1}{\widehat{\sigma}}\sqrt{n}\, \widehat{L}_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \mathbf{M}_n\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \tag{4.19}$$

donde

$$\widehat{L}_n(s, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \frac{1}{n}\sum_{i=1}^{n} \psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{s}\right) w_2(\mathbf{x}_i)\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i})\,.$$

y $\mathbf{M}_n$ está definida en (4.6). Notar que $\widehat{\mathbf{V}}_n$ definida en (4.4) verifica que $\widehat{\mathbf{V}}_n = \sqrt{n}\, \widehat{L}_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)/\widehat{\sigma}$.

A partir del Lema 4.1.2 sabemos que $\mathbf{M}_n \xrightarrow{P} \mathbf{M}$, donde $\mathbf{M}$ fue definida en (4.7). Por lo tanto, en virtud de que $\widehat{\sigma} \xrightarrow{P} \sigma_0$, para obtener la distribución asintótica de nuestro estimador será suficiente estudiar el comportamiento de $\widehat{L}_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$.

Sea

$$L_n(s, \boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \psi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \eta_{\boldsymbol{\beta},\boldsymbol{\theta}_0}(\boldsymbol{\theta}^t\mathbf{t}_i)}{s}\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\,,$$

luego notemos que

$$\begin{aligned} \sqrt{n}\widehat{L}_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= \sqrt{n}\, L_n(\sigma_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \sqrt{n}\, [L_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - L_n(\sigma_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)] \\ &+ \sqrt{n}\, \left[\widehat{L}_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - L_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\right]\,. \end{aligned}$$

Comenzaremos por estudiar el comportamiento asintótico de

i) $\sqrt{n}L_n(\sigma_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$

ii) $\sqrt{n}\,[L_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - L_n(\sigma_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)]$

iii) $\sqrt{n}\left[\widehat{L}_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) - L_n(\widehat{\sigma}, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\right]$.

Respecto del ítem i), cabe observar que cada término de $L_n(\sigma_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ tiene esperanza 0 y segundo momento finito dado que ψ_1 es acotada y se cumple **N2** i) y **N5**. Luego, $\sqrt{n}L_n(\sigma_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ es asintóticamente normal y por lo tanto, es un $O_p(1)$.

Respecto del ítem ii), veremos que $\sqrt{n}\,[L_n(\widehat{\sigma},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) - L_n(\sigma_0,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)] \xrightarrow{P} 0$. Observemos que

$$\begin{aligned}
&\sqrt{n}\,[L_n(\widehat{\sigma},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) - L_n(\sigma_0,\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)] \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right) - \psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\sigma_0}\right)\right] w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\,,
\end{aligned}$$

Consideremos la $j-$ésima coordenada de esta diferencia y definamos el proceso

$$J_n^j(\sigma) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi_1\left(\frac{y - \boldsymbol{\beta}_0^t\mathbf{x}_i - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\sigma}\right) - \psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\sigma_0}\right)\right] w_2(\mathbf{x}_i)(\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}))_j\,,$$

con lo cual bastará probar que para todo $j = 1,\ldots,p+q$ se cumple que $J_n^j(\widehat{\sigma}) \xrightarrow{P} 0$, o dicho de otra forma que dado $\varepsilon > 0$,

$$\mathbb{P}(|J_n^j(\sigma)| > \varepsilon) \to 0\,. \tag{4.20}$$

Para ello tomemos $\mathcal{S} = (\sigma_0/2, 2\sigma_0)$ y definamos la siguiente familia de funciones

$$\begin{aligned}
\mathcal{H}_j \;=\; & \left\{h_\sigma^j(y,\mathbf{x},\mathbf{t}) = \left[\psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x} - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t})}{\sigma}\right) - \psi_1\left(\frac{y - \boldsymbol{\beta}_0^t\mathbf{x} - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t})}{\sigma_0}\right)\right]\times \right. \\
& \left. w_2(\mathbf{x})\lambda_j(\mathbf{x},\mathbf{t},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\boldsymbol{\theta}_0^t\mathbf{t}), \sigma \in \mathcal{S}\right\},
\end{aligned}$$

siendo $\lambda_j(\mathbf{x},\mathbf{t},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\boldsymbol{\theta}_0^t\mathbf{t})$ la $j-$ésima coordenda de $\boldsymbol{\lambda}(\mathbf{x},\mathbf{t},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\boldsymbol{\theta}_0^t\mathbf{t})$. Por el Teorema del Valor Medio obtenemos que

$$\begin{aligned}
|h_{\sigma^*}^j(y,\mathbf{x},\mathbf{t}) - h_\sigma^j(y,\mathbf{x},\mathbf{t})| \;\le\; & \frac{1}{\tilde{\sigma}}\left|\chi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x} - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t})}{\tilde{\sigma}}\right)\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x} - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t})}{\tilde{\sigma}}\right| \times \\
& |w_2(\mathbf{x})\lambda_j(\mathbf{x},\mathbf{t},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\boldsymbol{\theta}_0^t\mathbf{t})(\sigma^* - \sigma)|\,,
\end{aligned}$$

siendo $\tilde{\sigma}$ un punto intermedio entre σ^* y σ. Si notamos $\varphi(u) = u\,\chi_1(u)$, por la condición **N3** tenemos que φ es acotada y por lo tanto, para $\sigma,\sigma^* \in \mathcal{S}$ vale que

$$|h_{\sigma^*}^j(y,\mathbf{x},\mathbf{t}) - h_\sigma^j(y,\mathbf{x},\mathbf{t})| \;\le\; \frac{2}{\sigma_0}\|\sqrt{w_2}\|_\infty\|\varphi\|_\infty|\sqrt{w_2(\mathbf{x})}\lambda_j(\mathbf{x},\mathbf{t},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\boldsymbol{\theta}_0^t\mathbf{t})||\sigma^* - \sigma|\,,$$

Entonces, si notamos $B_1 = 2\|\sqrt{w_2}\|_\infty\|\varphi\|_\infty/\sigma_0$ y $D(y,\mathbf{x},\mathbf{t}) = B_1|\sqrt{w_2(\mathbf{x})}\lambda_j(\mathbf{x},\mathbf{t},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\boldsymbol{\theta}_0^t\mathbf{t})|$ resulta que

$$|h_{\sigma^*}^j(y,\mathbf{x},\mathbf{t}) - h_\sigma^j(y,\mathbf{x},\mathbf{t})| \;\le\; D(y,\mathbf{x},\mathbf{t})|\sigma^* - \sigma|\,, \tag{4.21}$$

siendo D además una envolvente de la clase $\mathcal{H}_j$ en L^2 por **N2 i)** y **N5**. Asimismo, cabe observar que

$$h_\sigma^j(y,\mathbf{x},\mathbf{t}) = \left[\psi_1\left(\frac{\sigma_0\epsilon}{\sigma}\right) - \psi_1(\epsilon)\right] w_2(\mathbf{x})\lambda_j(\mathbf{x},\mathbf{t},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,\boldsymbol{\theta}_0^t\mathbf{t})$$

y por lo tanto por la imparidad de ψ_1, la simetría de los errores y la independencia entre estos y las covariables tenemos que $\mathbb{E}(h_\sigma^j(y,\mathbf{x},\mathbf{t}))=0$. A partir de (4.21), por el Teorema 2.7.11 de van der Vaart y Wellner (1996) el número de bracketing de la familia $\mathcal{H}_j$ satisface la siguiente condición

$$N_{[\,]}(2\varepsilon\|D\|_{\mathbb{Q},2},\mathcal{H}_j,L^2(\mathbb{Q}))\leq N(\varepsilon,\mathcal{S},|\cdot|)\,.$$

Notemos que para alguna constante C_2 que sólo depende de σ_0, $N(\varepsilon,\mathcal{S},|\cdot|))<C_2/\varepsilon^2$. Por lo tanto, obtenemos que las clases de funciones $\mathcal{H}_j$ cumplen

$$\mathcal{J}_{[\,]}(\delta,\mathcal{H}_j)\to 0 \text{ si } \delta\to 0\,. \tag{4.22}$$

Como $\widehat{\sigma}$ es un estimador consistente, tenemos que, siendo $\delta>0$, $\mathbb{P}(|\widehat{\sigma}-\sigma_0|<\delta)>1-\delta/2$ para $n>n_0$, por lo que dado $\varepsilon>0$, tenemos que

$$\begin{aligned}
\mathbb{P}\left(\left|J_n^j(\widehat{\sigma})\right|>\varepsilon\right) &\leq \mathbb{P}\left(\left|J_n^j(\widehat{\sigma})\right|>\epsilon,|\sigma_0-\widehat{\sigma}|<\delta\right)+\mathbb{P}\left(|\sigma_0-\widehat{\sigma}|>\delta\right)\\
&\leq \mathbb{P}\left(\sup_{|\sigma_0-\sigma|<\delta}\left|J_n^j(\sigma)\right|>\varepsilon\right)+\frac{\delta}{2}\\
&\leq \frac{1}{\varepsilon}\mathbb{E}\left(\sup_{|\sigma_0-\sigma|<\delta}\left|J_n^j(\sigma)\right|\right)+\frac{\delta}{2}
\end{aligned}$$

En consecuencia, será suficiente comprobar que

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

Para ello consideremos la subfamilia de funciones de $\mathcal{H}_j$ definida por

$$\mathcal{H}_j^\delta=\left\{h_\sigma^j(y,\mathbf{x},\mathbf{t})\in\mathcal{H}_j:|\sigma-\sigma_0|<\delta\right\}.$$

Cabe notar que si $h_\sigma^j\in\mathcal{H}_j^\delta$, entonces $\|h_\sigma^j\|_{\mathbb{Q},2}<\delta\|D\|_{\mathbb{Q},2}$ y por lo tanto, en virtud de la desigualdad maximal existe una constante A que no depende de n tal que

$$\begin{aligned}
\mathbb{E}\left(\sup_{|\sigma_0-\sigma|<\delta}\left|J_n^j(\sigma)\right|\right) &\leq AJ_{[\,]}\left(\delta,\mathcal{H}_j^\delta\right)\|D\|_{\mathbb{Q},2}+\sqrt{n}\mathbb{E}\left(D(y,\mathbf{x},\mathbf{t})I_{\{D(y,\mathbf{x},\mathbf{t})>\sqrt{n}a(\delta)\}}\right)\\
&\leq AJ_{[\,]}\left(\delta,\mathcal{H}_j\right)\|D\|_{\mathbb{Q},2}+\frac{1}{a(\delta)}\mathbb{E}\left(D^2(y,\mathbf{x},\mathbf{t})I_{\{D(y,\mathbf{x},\mathbf{t})>\sqrt{n}a(\delta)\}}\right). \quad (4.23)
\end{aligned}$$

Dado que la función envolvente $D(y,\mathbf{x},\mathbf{t})$ tiene segundo momento finito, el segundo término de (4.23) tiende a 0 para $\delta>0$ fijo y en consecuencia

$$\mathbb{E}\left(\sup_{|\sigma_0-\sigma|<\delta}\left|J_n^j(\sigma)\right|\right)\leq AJ_{[\,]}\left(\delta,\mathcal{H}_j\right)\|D\|_{\mathbb{Q},2}+o(1)$$

y (4.20) sigue de (4.22).

Respecto de iii), notemos que tras realizar un desarrollo de Taylor resulta

$$\begin{aligned}
& \sqrt{n}\left[\widehat{L}_n(\widehat{\sigma},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) - L_n(\widehat{\sigma},\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right] \\
= \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right)\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i}) - \psi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right] w_2(\mathbf{x}_i) \\
= \; & L_{1n}+L_{2n}+L_{3n}+L_{4n}+L_{5n}\,,
\end{aligned}$$

siendo

$$\begin{aligned}
L_{1n} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\left[\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i})-\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right] \\
L_{2n} &= \frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\left[\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right]\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) \\
L_{3n} &= \frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\left[\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right]\left[\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i})-\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right] \\
L_{4n} &= \frac{1}{2}\frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1'\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\xi_i}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\left[n^{1/4}\left(\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right)\right]^2\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) \\
L_{5n} &= \frac{1}{2}\frac{1}{\widehat{\sigma}^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1'\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\xi_i}{\widehat{\sigma}}\right) w_2(\mathbf{x}_i)\left[n^{1/4}\left(\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right)\right]^2\left[\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i})-\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right]\,,
\end{aligned}$$

con ξ_i un punto intermedio entre $\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)$ y $\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)$.

A partir de las hipótesis **N2**, **N3** y **N4** obtenemos que $L_{3n} \xrightarrow{P} 0$, $L_{4n} \xrightarrow{P} 0$ y $L_{5n} \xrightarrow{P} 0$.

Veamos cómo se comporta L_{2n}. Observemos que

$$\begin{aligned}
L_{2n} &= \frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\sigma_0}\right) w_2(\mathbf{x}_i)\left[\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right]\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) \\
&+ \frac{1}{\widehat{\sigma}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\widehat{\sigma}}\right)-\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\sigma_0}\right)\right] \\
&\quad w_2(\mathbf{x}_i)\left[\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)\right]\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) \\
&= \frac{1}{\widehat{\sigma}}\mathbf{V}_{1n}+\frac{1}{\widehat{\sigma}}\mathbf{V}_{2n}\,. \qquad (4.24)
\end{aligned}$$

Comencemos por el segundo término de (4.24). Por el Teorema del Valor Medio resulta que

$$\begin{aligned}\mathbf{V}_{2n} &= -\frac{1}{\tilde{\sigma}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1'\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)}{\tilde{\sigma}}\right)\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)}{\tilde{\sigma}}\right)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\\ &\quad\left[\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)\right]\left[\tilde{\sigma}-\sigma_0\right],\end{aligned}$$

siendo $\tilde{\sigma}$ un punto intermedio entre $\widehat{\sigma}$ y σ_0. Luego, como se cumple (4.14) y $\widehat{\sigma}$ es un estimador consistente, resulta que $\mathbf{V}_{2n}\xrightarrow{P}0$ y lo mismo ocurre con el segundo término de (4.24).

Respecto del primer término de (4.24), notemos que

$$\mathbf{V}_{1n}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1\left(\epsilon_i\right)w_2(\mathbf{x}_i)\left[\eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)-\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)\right]\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\,.$$

En virtud de que se cumple (4.13), para $u\in\{\boldsymbol{\theta}_0^t\mathbf{t}:\mathbf{t}\in\mathcal{T}\}$ y $E_j(u)$ definida en (4.9), se cumple que

$$\widehat{\eta}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(u)-\eta_0(u)=\frac{1}{nf_{\boldsymbol{\theta}_0}(u)}\sum_{j=1}^{n}K_h(\boldsymbol{\theta}_0^t\mathbf{t}_j-u)E_j(u)+o_P(n^{-1/2}),$$

por lo tanto,

$$\begin{aligned}\mathbf{V}_{1n}=&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1(\epsilon_i)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\frac{1}{nf_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}\sum_{j=1}^{n}K_h(\boldsymbol{\theta}_0^t\mathbf{t}_j-\boldsymbol{\theta}_0^t\mathbf{t}_i)E_j(\boldsymbol{\theta}_0^t\mathbf{t}_i)\\ &-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1(\epsilon_i)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})o_P(n^{-1/2})\end{aligned}$$

y bajo **N2-N5** tenemos que

$$\mathbf{V}_{1n}=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi_1(\epsilon_i)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\frac{1}{nf_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}\sum_{j=1}^{n}K_h(\boldsymbol{\theta}_0^t\mathbf{t}_j-\boldsymbol{\theta}_0^t\mathbf{t}_i)E_j(\boldsymbol{\theta}_0^t\mathbf{t}_i)+o_P(1)\,.$$

Definamos

$$\begin{aligned}\tilde{\mathbf{V}}_{1n}=&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\chi_1(\epsilon_i)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\frac{1}{nf_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}\sum_{j=1}^{n}K_h(\boldsymbol{\theta}_0^t\mathbf{t}_j-\boldsymbol{\theta}_0^t\mathbf{t}_i)E_j(\boldsymbol{\theta}_0^t\mathbf{t}_i)\\ &+E_i(\boldsymbol{\theta}_0^t\mathbf{t}_i)\boldsymbol{\gamma}(\mathbf{t}_i)f_{\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t\mathbf{t}_i\right)\Big].\end{aligned}$$

y llamemos $C(y_i, \mathbf{x}_i, \mathbf{t}_i) = \chi_1(\epsilon_i) w_2(\mathbf{x}_i) \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) / f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t \mathbf{t}_i)$. Observemos que $\boldsymbol{\gamma}(\cdot)$ definida en (4.10) cumple $\boldsymbol{\gamma}(\mathbf{t}) = \mathbb{E}(C(y_1, \mathbf{x}_1, \mathbf{t}_1) | \mathbf{t}_1 = \mathbf{t})$ y que al ser K una función par

$$
\begin{aligned}
\tilde{\mathbf{V}}_{1n} &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t \mathbf{t}_i - \boldsymbol{\theta}_0^t \mathbf{t}_j) [C(y_j, \mathbf{x}_j, \mathbf{t}_j) E_i(\boldsymbol{\theta}_0^t \mathbf{t}_j) - \boldsymbol{\gamma}(\mathbf{t}_i) E_i(\boldsymbol{\theta}_0^t \mathbf{t}_i)] \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{1}{n} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t \mathbf{t}_i - \boldsymbol{\theta}_0^t \mathbf{t}_j) - f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t \mathbf{t}_i) \right] \boldsymbol{\gamma}(\mathbf{t}_i) E_i(\boldsymbol{\theta}_0^t \mathbf{t}_i) \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t \mathbf{t}_i - \boldsymbol{\theta}_0^t \mathbf{t}_j) [C(y_j, \mathbf{x}_j, \mathbf{t}_j) - \boldsymbol{\gamma}(\mathbf{t}_j)] E_i(\boldsymbol{\theta}_0^t \mathbf{t}_j) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t \mathbf{t}_i - \boldsymbol{\theta}_0^t \mathbf{t}_j) [\boldsymbol{\gamma}(\mathbf{t}_j) - \boldsymbol{\gamma}(\mathbf{t}_i)] E_i(\boldsymbol{\theta}_0^t \mathbf{t}_j) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t \mathbf{t}_i - \boldsymbol{\theta}_0^t \mathbf{t}_j) [E_i(\boldsymbol{\theta}_0^t \mathbf{t}_j) - E_i(\boldsymbol{\theta}_0^t \mathbf{t}_i)] \boldsymbol{\gamma}(\mathbf{t}_i) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{1}{n} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t \mathbf{t}_i - \boldsymbol{\theta}_0^t \mathbf{t}_j) - f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t \mathbf{t}_i) \right] \boldsymbol{\gamma}(\mathbf{t}_i) E_i(\boldsymbol{\theta}_0^t \mathbf{t}_i) \\
&= \tilde{\mathbf{V}}_{1n}^{(1)} + \tilde{\mathbf{V}}_{1n}^{(2)} + \tilde{\mathbf{V}}_{1n}^{(3)} + \tilde{\mathbf{V}}_{1n}^{(4)} .
\end{aligned}
$$

Observemos que $\mathbb{E}(E_i(\boldsymbol{\theta}_0^t \mathbf{t}_1) | \boldsymbol{\theta}_0^t \mathbf{t}_1) = 0$, entonces usando **N6** y el hecho de que

$$
\sup_{\mathbf{t} \in \mathcal{T}} \left| \frac{1}{n} \sum_{j=1}^{n} K_h(\boldsymbol{\theta}_0^t \mathbf{t} - \boldsymbol{\theta}_0^t \mathbf{t}_j) - f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t \mathbf{t}) \right| \xrightarrow{P} 0 ,
$$

por el Lema 6.6.7 de Härdle et al. (2000), obtenemos que $\tilde{\mathbf{V}}_{1n}^{(4)} \xrightarrow{P} 0$. Para probar que $\tilde{\mathbf{V}}_{1n}^{(j)} \xrightarrow{P} 0$ con $j = 1, 2$ y 3 bastará estudiar la esperanza y la traza de la matriz de covarianzas de cada una de ellas, tal como se realiza en el apéndice de este capítulo.

Por lo que resulta que $\mathbf{V}_{1n} - \tilde{\mathbf{V}}_{1n} \xrightarrow{P} 0$ y como $\sum_{i=1}^{n} E_i(\boldsymbol{\theta}_0^t \mathbf{t}_i) \boldsymbol{\gamma}(\mathbf{t}_i) f_{\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t \mathbf{t}_i\right)) / \sqrt{n}$ es un $O_p(1)$ y $\widehat{\sigma}$ es un estimador consistente, obtenemos que L_{2n} se comporta asintóticamente como $-\sum_{i=1}^{n} E_i(\boldsymbol{\theta}_0^t \mathbf{t}_i) \boldsymbol{\gamma}(\mathbf{t}_i) f_{\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t \mathbf{t}_i\right) / (\sigma_0 \sqrt{n})$.

Veamos que $L_{1n} \xrightarrow{P} 0$. Usaremos argumentos similares a los empleados para la deducción en ii) y para ello definiremos clases de funciones adecuadas.

Para comenzar, notemos que

$$\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i})-\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}) = \begin{pmatrix} (\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) - \frac{\partial}{\partial\boldsymbol{\beta}}\eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u))\big|_{(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u)=\boldsymbol{\tau}} \\ (\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) - \frac{\partial}{\partial\boldsymbol{\theta}}\eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u))\big|_{(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u)=\boldsymbol{\tau}} + (\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) - \frac{\partial}{\partial u}\eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u))\big|_{(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0,u)=\boldsymbol{\tau}}\, \mathbf{t}_i \end{pmatrix}$$

depende de $\mathbf{t}_i$. Definimos $\boldsymbol{\vartheta}(\mathbf{t}_i) = \widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}_{0i}) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})$ y como en ii) trabajaremos con cada coordenada $\boldsymbol{\vartheta}_j$.

Para cada coordenada consideremos el proceso

$$\tilde{J}_n(\sigma, v) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_1\left(\frac{y_i - \boldsymbol{\beta}_0^t\mathbf{x}_i - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t}_i)}{\sigma}\right) w_2(\mathbf{x}_i)v(\mathbf{t}_i)\,.$$

Sean $\mathcal{S} = (\frac{\sigma_0}{2}, 2\sigma_0)$ el entorno considerado más arriba y $\mathcal{V} = \{v \in \mathcal{C}^1(\mathcal{T}) : \|v\|_\infty \leq 1, \|v'\|_\infty \leq 1\}$. Consideremos la norma sobre $\mathcal{S}\times\mathcal{V}$ como $|(\sigma, v)| = |\sigma| + \|v\|_\infty$. Cabe observar que por el Corolario 2.7.2 de van der Vaart y Wellner (1996) para cualquier medida de probabilidad $\mathbb{Q}$ se cumple que

$$\log N\left(\frac{\varepsilon}{2}, \mathcal{V}, L^2(\mathbb{Q})\right) \leq \log N_{[\,]}(\varepsilon, \mathcal{V}, L^2(\mathbb{Q})) \leq C\varepsilon^{-1}\,,$$

para $0 < \varepsilon < 2$, donde C es una constante que no depende de $\mathbb{Q}$.

Consideremos la familia de funciones definida por

$$\mathcal{H} = \left\{h_{\sigma,v}(y, \mathbf{x}, \mathbf{t}) = \psi_1\left(\frac{y - \boldsymbol{\beta}_0^t\mathbf{x} - \eta_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\mathbf{t})}{\sigma}\right) w_2(\mathbf{x})v(\mathbf{t}), \sigma \in \mathcal{S}, v \in \mathcal{V}\right\}.$$

Notemos que dado que ψ_1 es impar, que los errores tienen distribución simétrica alrededor del 0 y son independientes de las covariables, tenemos que $\mathbb{E}(h_{\sigma,v}(y, \mathbf{x}, \mathbf{t})) = 0$. Como ψ_1 y w_2 son acotadas y v y su derivada son continuas sobre un compacto, $h_{\sigma,v}$ resulta acotada, es decir existe una constante B tal que $|h_{\sigma,v}| < B$ si $\sigma \in \mathcal{S}, v \in \mathcal{V}$. Además, tras un desarrollo de Taylor y algunos manejos algebraicos, obtenemos que

$$\begin{aligned} |h_{\sigma,v}(y, \mathbf{x}, \mathbf{t}) - h_{\sigma^*,v^*}(y, \mathbf{x}, \mathbf{t})| &\leq \|w_2\|_\infty\left[\|\psi_1\|_\infty + \frac{2}{\sigma_0}\|\varphi\|_\infty\right]|(\sigma, v) - (\sigma^*, v^*)| \\ &\leq B_1|(\sigma, v) - (\sigma^*, v^*)|\,. \end{aligned}$$

A partir de esta desigualdad la clase resulta de Lipschitz y por Teorema 2.7.11 de van der Vaart y Wellner (1996) tenemos que dado $0 < \varepsilon < 2$

$$N_{[\,]}(\varepsilon B_1, \mathcal{H}, L^2(\mathbb{Q})) \leq N\left(\frac{\varepsilon}{2}, \mathcal{V}, L^2(\mathbb{Q})\right) N\left(\frac{\varepsilon}{2}, \mathcal{S}, |\cdot|\right)\,,$$

lo que implica que

$$N\left(\frac{\varepsilon}{2}B_1, \mathcal{H}, L^2(\mathbb{Q})\right) \leq N(\varepsilon, \mathcal{V}, L^2(\mathbb{Q}))\, N(\varepsilon, \mathcal{S}, |\cdot|))\,.$$

Con lo cual la familia tiene entropía finita ya que el número de cubrimiento de la clase queda acotado por los respectivos del intervalo $\mathcal{S}$ y de $\mathcal{V}$ y por lo tanto,

$$\mathcal{J}(\delta, \mathcal{H}) \to 0 \text{ si } \delta \to 0\,. \tag{4.25}$$

Observemos que por **N2** ii), para $j \in \{1, ..., p+q\}$, $\sup_{\mathbf{t}\in\mathcal{T}} |\boldsymbol{\vartheta}_j(\mathbf{t})| \xrightarrow{P} 0$ y por lo tanto, para n suficientemente grande, entonces

$$\mathbb{P}(\boldsymbol{\vartheta}_j \in \mathcal{V}, \|\boldsymbol{\vartheta}_j\|_\infty < \delta) > 1 - \frac{\delta}{2}\,.$$

Como $\widehat{\sigma}$ es un estimador consistente, tenemos que $\mathbb{P}(|\widehat{\sigma} - \sigma_0| < \delta) > 1 - \delta/2$ para n suficientemente grande, por lo que dado $\varepsilon > 0$, tenemos que

$$\begin{aligned}
\mathbb{P}\left(\left|\tilde{J}_n(\widehat{\sigma}, \boldsymbol{\vartheta}_j)\right| > \varepsilon\right) &\leq \mathbb{P}\left(\left|\tilde{J}_n(\widehat{\sigma}, \boldsymbol{\vartheta}_j)\right| > \epsilon, |\sigma_0 - \widehat{\sigma}| < \delta, \|\boldsymbol{\vartheta}_j\|_\infty < \delta\right) + \mathbb{P}\left(|\sigma_0 - \widehat{\sigma}| > \delta \text{ o } \|\boldsymbol{\vartheta}_j\|_\infty > \delta\right) \\
&\leq \mathbb{P}\left(\sup_{|\sigma_0-\sigma|<\delta, \|v\|_\infty > \delta} \left|\tilde{J}_n(\sigma, v)\right| > \varepsilon\right) + \delta \\
&\leq \frac{1}{\varepsilon}\mathbb{E}\left(\sup_{|\sigma_0-\sigma|<\delta, \|v\|_\infty > \delta} \left|\tilde{J}_n(\sigma, v)\right|\right) + \delta
\end{aligned}$$

Por lo tanto, será suficiente probar que

$$\lim_{n\to\infty} \lim_{\delta\to 0} \mathbb{E}\left(\sup_{|\sigma_0-\sigma|<\delta, \|v\|_\infty > \delta} \left|\tilde{J}_n(\sigma)\right|\right) = 0\,.$$

Para ello basta tomar la subfamilia de funciones de $\mathcal{H}$ definida por

$$\mathcal{H}^\delta = \left\{h_{\sigma,v}(y, \mathbf{x}, \mathbf{t}) \in \mathcal{H} : \|v\|_\infty < \delta\right\},$$

y notar que

$$\sup_{h\in\mathcal{H}^\delta} \frac{1}{n}\sum_{i=1}^n h^2(y, \mathbf{x}, \mathbf{t}) \leq A_1\delta^2\,,$$

para $A_1 = \|\psi_1\|_\infty^2 \|w_2\|_\infty^2$. Entonces,

$$\mathbb{P}\left(\left|\tilde{J}_n(\widehat{\sigma}, \boldsymbol{\vartheta}_j)\right| > \varepsilon\right) \leq \frac{1}{\varepsilon} B_2 A_1 \mathcal{J}(\delta, \mathcal{H}) + \delta$$

donde B_2 es una constante independiente de n, de donde sigue que cada coordenada de L_{1n} converge en probabilidad a 0.

En conclusión, a partir de los resultados anteriores y la consistencia de $\widehat{\sigma}$ a σ_0 tenemos que

$$\widehat{\mathbf{V}}_n - \frac{1}{\sigma_0}\sqrt{n} L_n(\sigma_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n}}\sum_{i=1}^n E_i(\boldsymbol{\theta}_0^t \mathbf{t}_i)\boldsymbol{\gamma}(\mathbf{t}_i) f_{\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t \mathbf{t}_i\right) \xrightarrow{P} 0\,. \tag{4.26}$$

Para finalizar la demostración, consideremos para cada $\boldsymbol{\theta}$, la matriz de proyección $\mathbf{P}_{\boldsymbol{\theta}}$ dada por $\mathbf{P}_{\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0}_{\mathbb{R}^{p\times q}} \\ \mathbf{0}_{\mathbb{R}^{q\times p}} & \mathbf{I}_q - \boldsymbol{\theta}\boldsymbol{\theta}^t \end{pmatrix}$. En la identidad (4.19), multiplicamos a izquierda por $\mathbf{P}_{\widehat{\boldsymbol{\theta}}}$, obteniendo en virtud de la ortogonalidad

$$\mathbf{0}_{\mathbb{R}^{p+q}} = \mathbf{P}_{\widehat{\boldsymbol{\theta}}}\widehat{\mathbf{V}}_n + \mathbf{P}_{\widehat{\boldsymbol{\theta}}}\mathbf{M}_n\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix}. \tag{4.27}$$

Debido a la consistencia de $\widehat{\boldsymbol{\theta}}$, $\mathbf{P}_{\widehat{\boldsymbol{\theta}}} \xrightarrow{P} \mathbf{P}_{\boldsymbol{\theta}_0}$. Por otro lado, como podemos suponer que $\boldsymbol{\theta}_0 = \mathbf{e}_q$ por la invariancia del parámetro $\boldsymbol{\theta}$ por transformaciones ortogonales, Lema 3.3.2, resulta que $\mathbf{P}_{\boldsymbol{\theta}_0} = \begin{pmatrix} \mathbf{I}_{p+q-1} & 0 \\ 0 & 0 \end{pmatrix}$. Por lo tanto, usando (4.26) llegamos a que ¡

$$\mathbf{P}_{\widehat{\boldsymbol{\theta}}}\widehat{\mathbf{V}}_n \xrightarrow{P} \begin{pmatrix} \mathbf{W} \\ 0 \end{pmatrix},$$

donde $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_1)$ definida en la condición **N10**. Además, como consecuencia del Lema 4.1.2, tenemos que $\mathbf{P}_{\widehat{\boldsymbol{\theta}}}\mathbf{M}_n \xrightarrow{P} \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ 0 & 0 \end{pmatrix}$ siendo $(\mathbf{M}_1\mathbf{M}_2)$ las primeras $p+q-1$ filas de la matriz $\mathbf{M}$, $\mathbf{M}_1$ matriz cuadrada y $\mathbf{M}_2$ vector columna; en definitiva, juntando estas dos convergencias con (4.27) resulta que

$$(\mathbf{M}_1\mathbf{M}_2)\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{P} \mathbf{W}. \tag{4.28}$$

Además, podemos reescribir a $\sqrt{n}(\widehat{\boldsymbol{\theta}} + \boldsymbol{\theta}_0)^t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = 0$ de manera de conformar la siguiente identidad

$$2\boldsymbol{\theta}_0^t\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\sqrt{n}||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||^2. \tag{4.29}$$

Para probar el ítem a) del corriente teorema usaremos lo siguiente:

– $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathcal{O}_P(1)$ como consecuencia de (4.28),

– $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ por la consistencia de $\widehat{\boldsymbol{\theta}}$,

– $\boldsymbol{\theta}_0 = \mathbf{e}_q$,

ya que garantiza, utilizando (4.29), que

$$\sqrt{n}(\hat{\theta}_q - \theta_{0q}) = \boldsymbol{\theta}_0^t\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\frac{\sqrt{n}}{2}||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||^2 \xrightarrow{P} 0. \tag{4.30}$$

En conclusión, utilizando (4.28) y (4.30):

$$\mathbf{M}_1\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}^{(q-1)} - \boldsymbol{\theta}_0^{(q-1)} \end{pmatrix} = \mathbf{M}_1\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}^{(q-1)} - \boldsymbol{\theta}_0^{(q-1)} \end{pmatrix} + \mathbf{M}_2\sqrt{n}(\hat{\theta}_q - \theta_{0q}) - \mathbf{M}_2\sqrt{n}(\hat{\theta}_q - \theta_{0q}) \xrightarrow{P} \mathbf{W},$$

junto con **N9**, prueban el ítem b). ■

4.2. Apéndice

Comenzamos por estudiar el comportamiento de $\tilde{\mathbf{V}}_{1n}^{(3)}$. Calcularemos su esperanza y la traza de su matriz de covarianzas y veremos su convergencia a 0.

Consideremos $m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$ dada en (4.12). Por la condición **N3**, ψ_1 es una función impar y acotada, además como los errores se distribuyen de manera simétrica respecto del 0 y son independientes de las covariables, podemos observar que $m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_1) = 0$. Ahora bien, es conveniente también observar que $m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$ es una función que depende de $(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$ a través de $(\boldsymbol{\theta}_0^t\boldsymbol{\nu}_1, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_2)$, con lo cual se puede definir la función $m_0 : \mathbb{R}^2 \to \mathbb{R}$ tal que $m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = m_0(\boldsymbol{\theta}_0^t\boldsymbol{\nu}_1, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_2)$. Por lo tanto,

$$\begin{aligned}
\mathbb{E}\left(-\tilde{\mathbf{V}}_{1n}^{(3)}\right) &= \frac{n(n-1)}{n\sqrt{n}}\mathbb{E}(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1)[E_1(\boldsymbol{\theta}_0^t\mathbf{t}_2) - E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)]\boldsymbol{\gamma}(\mathbf{t}_1)) \\
&= \frac{n-1}{\sqrt{n}}\mathbb{E}(\mathbb{E}(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1)[E_1(\boldsymbol{\theta}_0^t\mathbf{t}_2) - E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)]\boldsymbol{\gamma}(\mathbf{t}_1)|(\mathbf{t}_1, \mathbf{t}_2))) \\
&= \frac{n-1}{\sqrt{n}}\mathbb{E}(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1)m_1(\mathbf{t}_1, \mathbf{t}_2)\boldsymbol{\gamma}(\mathbf{t}_1)) \\
&= \frac{n-1}{\sqrt{n}}\int_{\mathcal{T}} \boldsymbol{\gamma}(\boldsymbol{\zeta}) f_{\mathbf{t}}(\boldsymbol{\zeta}) \int_{\mathcal{U}_0(\mathcal{T})} m_0(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, v) K_h(v - \boldsymbol{\theta}_0^t\boldsymbol{\zeta}) f_{\boldsymbol{\theta}_0}(v) dv d\boldsymbol{\zeta}\,.
\end{aligned}$$

De las condiciones **N6** a **N8**, se desprende que

$$\begin{aligned}
&\int_{\mathcal{U}_0(\mathcal{T})} m_0(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, v) K_h(v - \boldsymbol{\theta}_0^t\boldsymbol{\zeta}) f_{\boldsymbol{\theta}_0}(v) dv \\
= &\int_{\mathcal{V}_0(\mathcal{T},\boldsymbol{\zeta},h)} m_0(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, uh + \boldsymbol{\theta}_0^t\boldsymbol{\zeta}) K(u) f_{\boldsymbol{\theta}_0}(uh + \boldsymbol{\theta}_0^t\boldsymbol{\zeta}) du \\
= &\int_{\mathcal{V}_0(\mathcal{T},\boldsymbol{\zeta},h)} \left[m_0'(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, \boldsymbol{\theta}_0^t\boldsymbol{\zeta})uh + \frac{m_0''(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, \boldsymbol{\theta}_0^t\boldsymbol{\iota}_1^{int})}{2}u^2h^2\right] K(u)[f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}) + uh f'_{\boldsymbol{\theta}_0^t\mathbf{t}}(\boldsymbol{\theta}_0^t\boldsymbol{\iota}_1^{int})] du \\
= &\; h^2\left[m_0'(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, \boldsymbol{\theta}_0^t\boldsymbol{\zeta})\int_{\mathcal{V}_0(\mathcal{T},\boldsymbol{\zeta},h)} u^2 K(u) f'_{\boldsymbol{\theta}_0^t\mathbf{t}}(\boldsymbol{\theta}_0^t\boldsymbol{\iota}_1^{int}) du + \frac{f_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0^t\boldsymbol{\zeta})}{2}\int_{\mathcal{V}_0(\mathcal{T},\boldsymbol{\zeta},h)} u^2 K(u) m_0''(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, \boldsymbol{\theta}_0^t\boldsymbol{\iota}_1^{int}) du \right. \\
&\left. + h\int_{\mathcal{V}_0(\mathcal{T},\boldsymbol{\zeta},h)} u^3 K(u) m_0''(\boldsymbol{\theta}_0^t\boldsymbol{\zeta}, \boldsymbol{\theta}_0^t\boldsymbol{\iota}_1^{int}) f'_{\boldsymbol{\theta}_0^t\mathbf{t}}(\boldsymbol{\theta}_0^t\boldsymbol{\iota}_1^{int}) du\right] = O(h^2)
\end{aligned}$$

definiendo $\mathcal{V}_0(\mathcal{T}, \boldsymbol{\zeta}, h) := \{(\boldsymbol{\theta}_0^t\mathbf{t} - \boldsymbol{\theta}_0^t\boldsymbol{\zeta})/h : \mathbf{t} \in \mathcal{T}\}$ al conjunto donde integramos luego del cambio de variables $u = (v - \boldsymbol{\theta}_0^t\boldsymbol{\zeta})/h$. Como consecuencia de las condiciones **N1** a **N4** y la hipótesis $nh^4 \to 0$, llegamos a que $\mathbb{E}\left(-\tilde{\mathbf{V}}_{1n}^{(3)}\right) = (n-1)O(h^2)/\sqrt{n} = O(\sqrt{n}h^2) \to 0$.

Para el cálculo de la traza de la matriz de covarianzas, observemos que

$$-\tilde{\mathbf{V}}_{1n}^{(3)} = \frac{1}{n^{3/2}}\sum_{i\neq j} K_h(\boldsymbol{\theta}_0^t\mathbf{t}_j - \boldsymbol{\theta}_0^t\mathbf{t}_i)[E_i(\boldsymbol{\theta}_0^t\mathbf{t}_j) - E_i(\boldsymbol{\theta}_0^t\mathbf{t}_i)]\boldsymbol{\gamma}(\mathbf{t}_i) = \frac{1}{n^{3/2}}\sum_{i\neq j} Z_{ij}\,.$$

Por lo tanto,

$$\begin{aligned}
\operatorname{cov}\left(-\tilde{\mathbf{V}}_{1n}^{(3)}, -\tilde{\mathbf{V}}_{1n}^{(3)}\right) &= \frac{1}{n^3}\sum_{i\neq j}\operatorname{cov}(Z_{ij}, Z_{ij}) + \frac{1}{n^3}\sum_{i\neq j}\operatorname{cov}(Z_{ij}, Z_{ji}) \\
+ \quad & \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq k, j\neq i, k\neq i}\operatorname{cov}(Z_{ij}, Z_{ik}) + \frac{1}{n^3}\sum_{j=1}^{n}\sum_{i\neq k, i\neq j, k\neq j}\operatorname{cov}(Z_{ij}, Z_{kj}) \\
+ \quad & \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq k, j\neq i, k\neq i}\operatorname{cov}(Z_{ji}, Z_{ik}) + \frac{1}{n^3}\sum_{j=1}^{n}\sum_{i\neq k, i\neq j, k\neq j}\operatorname{cov}(Z_{ji}, Z_{kj}) \\
= \quad & \frac{n(n-1)}{n^3}[\operatorname{cov}(Z_{12}, Z_{12}) + \operatorname{cov}(Z_{12}, Z_{21})] \\
+ \quad & \frac{n(n-1)(n-2)}{n^3}[\operatorname{cov}(Z_{12}, Z_{13}) + \operatorname{cov}(Z_{12}, Z_{32}) + \operatorname{cov}(Z_{21}, Z_{13}) + \operatorname{cov}(Z_{21}, Z_{32})]
\end{aligned}$$

Necesitamos computar $\operatorname{tr}\left(\operatorname{cov}\left(-\tilde{\mathbf{V}}_{1n}^{(3)}, -\tilde{\mathbf{V}}_{1n}^{(3)}\right)\right)$, luego vamos a acotar cada una de las trazas de las matrices de covarianzas detalladas.

Comencemos por el primer término. Para ello, recordemos que la traza conmuta con la esperanza. Para cada $l \in \{1, .., q\}$ denotemos c_l a la constante no nula tal que $\gamma_l^2(\boldsymbol{\zeta}) \leq c_l$ para todo $\boldsymbol{\zeta} \in \mathcal{T}$, siendo γ_l la $l-$ésima coordenada de $\boldsymbol{\gamma}$, y $c = \sum_{l=1}^{q} c_l$.

Definimos la función

$$m_{is}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4) = \mathbb{E}([E_i(\boldsymbol{\theta}_0^t\mathbf{t}_j) - E_i(\boldsymbol{\theta}_0^t\mathbf{t}_i)][E_s(\boldsymbol{\theta}_0^t\mathbf{t}_l) - E_s(\boldsymbol{\theta}_0^t\mathbf{t}_s)] \mid (\mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_s, \mathbf{t}_l) = (\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)).$$

Es conveniente observar que $m_{11}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)$ es una función que depende de $(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)$ a través de $(\boldsymbol{\theta}_0^t\boldsymbol{\nu}_1, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_2, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_3, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_4)$, con lo cual se puede definir la función $m_{1100} : \mathbb{R}^4 \to \mathbb{R}$ tal que $m_{11}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4) = m_{1100}(\boldsymbol{\theta}_0^t\boldsymbol{\nu}_1, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_2, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_3, \boldsymbol{\theta}_0^t\boldsymbol{\nu}_4)$. De esta forma obtenemos que

$$\begin{aligned}
\operatorname{tr}(\mathbb{E}(Z_{12}Z_{12}^t)) &= \frac{1}{h^2}\mathbb{E}\left(K^2\left(\frac{\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1}{h}\right)[E_1(\boldsymbol{\theta}_0^t\mathbf{t}_2) - E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)]^2\operatorname{tr}(\boldsymbol{\gamma}(\mathbf{t}_1)\boldsymbol{\gamma}^t(\mathbf{t}_1))\right) \\
&= \frac{1}{h^2}\mathbb{E}\left(K^2\left(\frac{\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1}{h}\right)[E_1(\boldsymbol{\theta}_0^t\mathbf{t}_2) - E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)]^2\sum_{l=1}^{q}\gamma_l^2(\mathbf{t}_1)\right) \\
&\leq \frac{c}{h^2}\mathbb{E}\left(K^2\left(\frac{\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1}{h}\right)[E_1(\boldsymbol{\theta}_0^t\mathbf{t}_2) - E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)]^2\right) \\
&= \frac{c}{h^2}\mathbb{E}\left(\mathbb{E}\left(K^2\left(\frac{\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1}{h}\right)[E_1(\boldsymbol{\theta}_0^t\mathbf{t}_2) - E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)]^2\middle|(\mathbf{t}_1, \mathbf{t}_2)\right)\right)
\end{aligned}$$

y por lo tanto

$$\begin{aligned}
\mathrm{tr}(\mathbb{E}(Z_{12}Z_{12}^t)) &\leq \frac{c}{h^2}\mathbb{E}\left(K^2\left(\frac{\boldsymbol{\theta}_0^t\mathbf{t}_2-\boldsymbol{\theta}_0^t\mathbf{t}_1}{h}\right)m_{11}(\mathbf{t}_1,\mathbf{t}_2,\mathbf{t}_1,\mathbf{t}_2)\right)\\
&= \frac{c}{h^2}\mathbb{E}\left(K^2\left(\frac{\boldsymbol{\theta}_0^t\mathbf{t}_2-\boldsymbol{\theta}_0^t\mathbf{t}_1}{h}\right)m_{1100}(\boldsymbol{\theta}_0^t\mathbf{t}_1,\boldsymbol{\theta}_0^t\mathbf{t}_2,\boldsymbol{\theta}_0^t\mathbf{t}_1,\boldsymbol{\theta}_0^t\mathbf{t}_2)\right)\\
&= \frac{c}{h^2}\int_{\mathcal{U}_0(\mathcal{T})}\int_{\mathcal{U}_0(\mathcal{T})}K^2\left(\frac{v-s}{h}\right)m_{1100}(s,v,s,v)f_{\boldsymbol{\theta}_0}(v)f_{\boldsymbol{\theta}_0}(s)dv\,ds\\
&= \frac{c}{h}\int_{\mathcal{U}_0(\mathcal{T})}f_{\boldsymbol{\theta}_0}(s)\int_{\mathcal{W}_0(\mathcal{T},s,h)}K^2(u)m_{1100}(s,uh+s,s,uh+s)f_{\boldsymbol{\theta}_0}(uh+s)du\,ds\ ,
\end{aligned}$$

donde $\mathcal{W}_0(\mathcal{T},s,h)$ es el conjunto que se obtiene luego del cambio de variables. Mediante un desarrollo de Taylor deducimos que

$$\begin{aligned}
&m_{1100}(s,uh+s,s,uh+s)=\mathbb{E}([E_1(\boldsymbol{\theta}_0^t\mathbf{t}_2)-E_1(\boldsymbol{\theta}_0^t\mathbf{t}_1)]^2|(\boldsymbol{\theta}_0^t\mathbf{t}_1,\boldsymbol{\theta}_0^t\mathbf{t}_2)=(s,uh+s))\\
&=u^2h^2\mathbb{E}([E_1'(s)+E_1^{''}(s^{int})uh/2]^2|(\boldsymbol{\theta}_0^t\mathbf{t}_1,\boldsymbol{\theta}_0^t\mathbf{t}_2)=(s,uh+s))\\
&=u^2h^2\tilde{m}_{11}(s,uh+s)\ ,
\end{aligned}$$

siendo s^{int} un punto intermedio entre s y $uh+s$ y $\tilde{m}_{11}$ convenientemente definida. En consecuencia, a partir de **N8** resulta que

$$\begin{aligned}
\mathrm{tr}(\mathbb{E}(Z_{12}Z_{12}^t)) &\leq c\,h\,||K||_\infty\int_{\mathcal{U}_0(\mathcal{T})}f_{\boldsymbol{\theta}_0}(s)\int_{\mathcal{W}_0(\mathcal{T},s,h)}u^2K(u)\tilde{m}_{11}(s,uh+s)f_{\boldsymbol{\theta}_0}(uh+s)du\,ds\\
&= O(h)
\end{aligned}$$

Notemos que $\mathrm{tr}(\mathbb{E}(Z_{12})\mathbb{E}(Z_{12})^t)=\sum_{l=1}^{q}(\mathbb{E}(Z_{12}))_l^2$, con lo cual al ya tener calculada $\mathbb{E}\left(-\tilde{\mathbf{V}}_{1n}^{(3)}\right)$, sabemos que $(\mathbb{E}(Z_{12}))_l=O(h^2)$. Finalmente, $\mathrm{tr}(\mathrm{cov}(Z_{12},Z_{12}))=O(h+h^4)=O(h)$.

Consideremos $\mathrm{cov}(Z_{12},Z_{21})$. Para cada $l\in\{1,..,q\}$ denotemos d_l a la constante no nula tal que $|\gamma_l(\boldsymbol{\zeta})\gamma_l(\boldsymbol{\iota})|\leq d_l\,\forall\,\boldsymbol{\zeta},\boldsymbol{\iota}\in\mathcal{T}$, con lo cual $|\mathrm{tr}(\boldsymbol{\gamma}(\boldsymbol{\zeta})\boldsymbol{\gamma}^t(\boldsymbol{\iota}))|\leq d=\sum_{l=1}^{q}d_l$. De manera similar a lo realizado renglones más arriba y observando que $m_{12}(\mathbf{t}_1,\mathbf{t}_2,\mathbf{t}_2,\mathbf{t}_1)$ puede ser escrita como $\tilde{m}_{12}(\boldsymbol{\theta}_0^t\mathbf{t}_1,\boldsymbol{\theta}_0^t\mathbf{t}_2)$, tenemos entonces que

$$\begin{aligned}
|\mathrm{tr}(\mathbb{E}(Z_{12}Z_{21}^t))| &= |\mathbb{E}(K_h^2(\boldsymbol{\theta}_0^t\mathbf{t}_2-\boldsymbol{\theta}_0^t\mathbf{t}_1)\mathrm{tr}(\boldsymbol{\gamma}(\mathbf{t}_1)\boldsymbol{\gamma}^t(\mathbf{t}_2))m_{12}(\mathbf{t}_1,\mathbf{t}_2,\mathbf{t}_2,\mathbf{t}_1))|\\
&= |\mathbb{E}(K_h^2(\boldsymbol{\theta}_0^t\mathbf{t}_2-\boldsymbol{\theta}_0^t\mathbf{t}_1)\mathrm{tr}(\boldsymbol{\gamma}(\mathbf{t}_1)\boldsymbol{\gamma}^t(\mathbf{t}_2))\tilde{m}_{12}(\boldsymbol{\theta}_0^t\mathbf{t}_1,\boldsymbol{\theta}_0^t\mathbf{t}_2))|\\
&\leq d\,\mathbb{E}(K_h^2(\boldsymbol{\theta}_0^t\mathbf{t}_2-\boldsymbol{\theta}_0^t\mathbf{t}_1)|\tilde{m}_{12}(\boldsymbol{\theta}_0^t\mathbf{t}_1,\boldsymbol{\theta}_0^t\mathbf{t}_2)|)=O(h)\ .
\end{aligned}$$

En definitiva, $\mathrm{tr}(\mathrm{cov}(Z_{12},Z_{21}))=O(h)$.

Análogamente ahora calculemos $\mathrm{tr}(\mathrm{cov}(Z_{12}, Z_{13}))$. Tenemos que

$$
\begin{aligned}
& |\mathrm{tr}(\mathbb{E}(Z_{12}Z_{13}^t))| = |\mathbb{E}(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_3 - \boldsymbol{\theta}_0^t\mathbf{t}_1)\mathrm{tr}(\boldsymbol{\gamma}(\mathbf{t}_1)\boldsymbol{\gamma}^t(\mathbf{t}_1))m_{11}(\mathbf{t}_1,\mathbf{t}_2,\mathbf{t}_1,\mathbf{t}_3))| \\
\leq\ & c\,\mathbb{E}(K_h(\boldsymbol{\theta}_0^t\mathbf{t}_2 - \boldsymbol{\theta}_0^t\mathbf{t}_1)K_h(\boldsymbol{\theta}_0^t\mathbf{t}_3 - \boldsymbol{\theta}_0^t\mathbf{t}_1)|m_{1100}(\boldsymbol{\theta}_0^t\mathbf{t}_1, \boldsymbol{\theta}_0^t\mathbf{t}_2, \boldsymbol{\theta}_0^t\mathbf{t}_1, \boldsymbol{\theta}_0^t\mathbf{t}_3)|) \\
=\ & \frac{c}{h^2}\int_{\mathcal{U}_0(\mathcal{T})}\int_{\mathcal{U}_0(\mathcal{T})}\int_{\mathcal{U}_0(\mathcal{T})} K\left(\frac{v-s}{h}\right)K\left(\frac{r-s}{h}\right)|m_{1100}(s,v,s,r)|f_{\boldsymbol{\theta}_0}(v)f_{\boldsymbol{\theta}_0}(r)f_{\boldsymbol{\theta}_0}(s)dv\,dr\,ds \\
=\ & c\int_{\mathcal{U}_0(\mathcal{T})}\int_{\mathcal{W}_0(\mathcal{T},s,h)}\int_{\mathcal{W}_0(\mathcal{T},s,h)} K(u)K(w)|m_{1100}(s,uh+s,s,wh+s)|f_{\boldsymbol{\theta}_0}(uh+s) \\
& f_{\boldsymbol{\theta}_0}(wh+s)f_{\boldsymbol{\theta}_0}(s)du\,dw\,ds = O(h^2)\,,
\end{aligned}
$$

de donde resulta que $\mathrm{tr}(\mathrm{cov}(Z_{12}, Z_{13})) = O(h^2)$. Con un procedimiento similar, se prueba que $\mathrm{tr}(\mathrm{cov}(Z_{12}, Z_{32})) = O(h^2)$, $\mathrm{tr}(\mathrm{cov}(Z_{21}, Z_{13})) = O(h^2)$ y $\mathrm{tr}(\mathrm{cov}(Z_{21}, Z_{32})) = O(h^2)$.

En conclusión

$$
\begin{aligned}
& \mathrm{tr}\left(\mathrm{cov}\left(-\tilde{\mathbf{V}}_{1n}^{(3)}, -\tilde{\mathbf{V}}_{1n}^{(3)}\right)\right) = \frac{n(n-1)}{n^3}[\mathrm{tr}(\mathbb{V}(Z_{12}) + \mathrm{cov}(Z_{12}, Z_{21}))] \\
& + \frac{n(n-1)(n-2)}{n^3}[\mathrm{tr}(\mathrm{cov}(Z_{12}, Z_{13}) + \mathrm{cov}(Z_{12}, Z_{32}) + \mathrm{cov}(Z_{21}, Z_{13}) + \mathrm{cov}(Z_{21}, Z_{32}))] \\
& = O(hn^{-1}) + O(h^2) \to 0\,.
\end{aligned}
$$

La convergencia a 0 en probabilidad de $\tilde{\mathbf{V}}_{1n}^{(1)}$ y $\tilde{\mathbf{V}}_{1n}^{(2)}$ se obtiene con cálculos y razonamientos análogos a los de la convergencia a 0 de $\tilde{\mathbf{V}}_{1n}^{(3)}$.

Capítulo 5

Estudio de Simulación

5.1. Condiciones del Experimento Numérico

Mediante un estudio de simulación, se comparó el desempeño de los estimadores propuestos en este trabajo con los clásicos respectivos usando al núcleo de Epanechnikov para suavizar. Consideraremos el modelo

$$y_i = \sqrt{5}\, x_i + 5 \sin\left(2\pi \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \mathbf{t}_i\right) + 2\epsilon_i \quad \text{para } i \in \{1, ..., 100\} \tag{5.1}$$

donde $x_i \sim \mathcal{N}(0,1)$, $\mathbf{t}_i \sim \mathcal{U}([0,1] \times [0,1])$ y $\epsilon_i \sim \mathcal{N}(0,1)$. Se generaron 1000 réplicas siguiendo este modelo.

Notemos que dado $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$ con $||\boldsymbol{\theta}|| = 1$, se tiene $\alpha \in [-\pi/2, \pi/2]$ ángulo asociado tal que $\cos(\alpha) = \theta_1$ y $\text{sen}(\alpha) = \theta_2$.

Respecto de los estimadores iniciales, en el Paso I.1, fijando el parámetro de suavizado de la función de pesos en $h = 0,25$ y la función ρ en la familia de las funciones bicuadradas con $k = 3,44$, obtenemos los estimadores iniciales de β_0, de α_0 el ángulo que forma $\boldsymbol{\theta}_0$ con el vector $(1,0)$ desde este último, y de σ_0. A partir de estos estimadores iniciales, tomando $h = 0,18$ como parámetro de suavizado de la función de pesos del Paso 1 y considerando ρ_1 de los Pasos 1 y 2 como la función bicuadrada con $k = 4,685$, calculamos los estimadores finales de β_0 y $\boldsymbol{\theta}_0$.

En el caso clásico, consideramos los parámetros de suavizado $h = 0,20$ y $h = 0,18$ para las estimaciones iniciales y finales, respectivamente.

Es importante destacar que las ventanas mencionadas se eligieron a partir de un análisis previo con varias ventanas fijas en cada uno de los contextos: clásico y robusto. No se efectuaron las selecciones de ventanas mediante el método de convalidación cruzada debido a los tiempos que ello demandaba. En las primeras instancias de simulaciones se combinaron una amplia variedad de modelos y parámetros de los métodos de estimación que permitieron identificar la combinación que mejor se ajustaba a lo que se quería resaltar de cada método. Los parámetros

k de las ρ-funciones que se utilizan en los métodos de estimación robustos se eligieron de manera de tener alta eficiencia bajo el modelo central y robustez suficiente como para penalizar a aquellas observaciones que tienen un comportamiento notablemente diferente al de la mayoría. Cabe tener en cuenta que $k = 3,44$ y $k = 4,685$ brindan, respectivamente, 85 % y 95 % de eficiencia asintótica con respecto a la media bajo normalidad, para más detalles ver Capítulo 2 de Maronna et al. (2019).

Para evaluar la robustez de los estimadores, se generaron muestras siguiendo el modelo dado en (5.1), que denotaremos por C_0, y muestras donde se contamina ya sea a los errores o, a la covariable inserta en la componente lineal y la variable respuesta. Los diferentes escenarios de contaminación se describen a continuación.

- $C_{1;p;\sigma_C^2}$: $\epsilon_1, ..., \epsilon_{100}$ son v.a.i.i.d. $(1-p)\mathcal{N}(0,1) + p\mathcal{N}(0,\sigma_C^2)$ siendo $p \in \{0,1; 0,2\}$ y $\sigma_C^2 \in \{49; 100\}$.
- $C_{2;p}$: $\epsilon_1, ..., \epsilon_{100}$ son v.a.i.i.d. $(1-p)\mathcal{N}(0,1) + p\mathcal{N}(9,1/16)$ siendo $p \in \{0,1; 0,2; 0,3\}$.
- $C_{3;m;\mu_C}$: m observaciones de $\{(y_i, x_i)\}$ se contaminarán de la siguiente manera: a x_i por una variable aleatoria $\mathcal{N}(10, 25/10^4)$ y a y_i por una variable aleatoria $\mathcal{N}(\mu_C, 25/10^4)$ siendo $m \in \{10; 15\}$ y $\mu_C \in \{5; 10; 15\}$.

Notemos que la contaminación de primer tipo lo que hace es generar un porcentaje de errores con la varianza inflada; la de segundo tipo, traslada un porcentaje de los errores alrededor del valor 9; y la de tercer tipo, genera datos anómalos en la covariable x, concentrándolos alrededor del valor 10, y alejándolos así del grueso de los datos, y a la variable respuesta y se la contamina de manera que los pares contaminados no respeten el modelo (5.1).

5.2. Resultados

A continuación se presentarán para cada uno de los escenarios detallados, distintos gráficos de las estimaciones de los parámetros. En el caso de las estimaciones de β_0, gráficos de caja (boxplots). Para las estimaciones del parámetro del índice simple, o sea $\boldsymbol{\theta}_0$, se exponen gráficos de flechas. Estos gráficos de flechas se construyen a partir de las estimaciones de α_0, ángulo asociado a $\boldsymbol{\theta}_0$. Se calcula la mediana muestral de las estimaciones de α_0 para luego obtener la dirección asociada a ese ángulo, que es la flecha de color rojo trazada, asociada a una estimación $\boldsymbol{\theta}_0$; además se calculan los cuartiles inferior y superior y los "bigotes" del gráfico de caja de las estimaciones de α_0 para luego obtener las direcciones asociadas a esos ángulos y representarlas con las flechas de color negro, relativas a las estimaciones de $\boldsymbol{\theta}_0$. Por último, para las estimaciones de η_0 se presentan boxplots funcionales, una herramienta gráfica muy utilizada para representar las estimaciones de una función real univariada, que fue introducida por Sun y Genton (2011). Como $\mathbf{t}$ es un vector aleatorio, de manera de generar estimaciones de η_0 que sean comparables, es que se fijan, en nuestro caso, 100 puntos $\mathbf{v}_1, ..., \mathbf{v}_{100} \in [0,1] \times [0,1]$ y se estiman en cada réplica $\eta_0(\boldsymbol{\theta}_0^t \mathbf{v}_i)$ usando los procedimientos clásico y robusto. En los

boxplots funcionales, la región pintada de fucsia representa la región central, la curva negra representa la curva más profunda y las curvas rojas punteadas corresponden a las curvas atípicas. Para comparar con la verdadera función η_0, es que se la representa superpuesta a los boxplots funcionales en color verde.

La primera observación que cabe señalar de los gráficos de las Figuras 5.1, 5.2 y 5.3 es que en un escenario sin datos atípicos, las estimaciones clásicas, como es de esperar, resultan más precisas que las robustas. En cambio en un escenario con datos atípicos, la sensibilidad de las estimaciones clásicas a ellos, produce estimaciones que resultan inutilizables en la práctica. Justamente la naturaleza de los métodos de estimación robusta yace en generar mecanismos de detección de datos atípicos para poder neutralizar con apropiadas herramientas su impacto en la estimación, y que las estimaciones de los parámetros resulten confiables.

Analizando ahora los distintos escenarios de contaminación, nos encontraremos con las destacables virtudes que los métodos robustos agregan en contextos donde se estiman múltiples párametros como lo es un *Modelo Parcialmente Lineal de Índice Simple*.

Empecemos estudiando lo que acontece en la simulación realizada para las estimaciones del parámetro del término lineal, o sea β_0, observando los gráficos de caja de las Figuras 5.4, 5.7, 5.10, 5.13, 5.16, 5.19 y 5.22. En los escenarios de contaminaciones $C_{1;p;\sigma_C^2}$ y $C_{2;p}$, las estimaciones robustas de β_0 son más precisas que las clásicas ya que tienen menor distancia intercuartil, y tanto estimaciones clásicas como robustas resultan ser estables ya que contienen a β_0 en la caja de los boxplots. Es en el escenario de contaminación $C_{3;m;\mu_C}$, cuyos gráficos corresponden a las Figuras 5.25, 5.28, 5.31, 5.34, 5.37 y 5.40, donde se vislumbra un mayor beneficio de las estimaciones robustas debido a que muestran una excelente estimación frente al pobre desempeño de las clásicas, que se encuentran alejadas de β_0 pero que resultan ser precisas por tener una distancia intercuartil muy chica. Para ilustrar el impacto del procedimiento robusto en este escenario de contaminación, es que en la Figura 5.25 se muestran también los boxplots de las estimaciones iniciales de ambos procedimientos. Es importante destacar en este punto que los estimadores iniciales robustos no cuentan con un mecanismo para detectar contaminaciones en la covariable x, como sí lo tienen los finales, y es por eso que las estimaciones iniciales robustas no resultan tan recomendables. Este hecho permite realzar aún más el muy buen comportamiento de los estimadores finales robustos, ya que a pesar de basar sus estimaciones en estimaciones iniciales desacertadas, pueden detectar datos atípicos en la covariable x y en la respuesta y logrando estimaciones que demuestran el muy buen ejercicio de su característica robusta.

En el caso de las estimaciones del índice simple $\boldsymbol{\theta}_0$, podemos observar en los gráficos de flechas de las Figuras 5.5, 5.8, 5.11, 5.14, 5.17 y 5.20 que tanto en los modelos de contaminación $C_{1;p;\sigma_C^2}$ como en $C_{2;p}$ con $p \in \{0,1;0,2\}$, en las estimaciones clásicas hay direcciones muy alejadas de $\boldsymbol{\theta}_0$ mientras que las estimaciones robustas son estables y precisas. Por otro lado, más allá de que las estimaciones clásicas en el escenario de contaminación $C_{2;0,3}$ siguen siendo muy malas, hay estimaciones robustas muy alejadas de $\boldsymbol{\theta}_0$, observando los gráficos de flechas de la Figura 5.23. No se visualiza en los modelos de contaminación $C_{3;m;\mu_C}$, cuyos gráficos corresponden a las Figuras 5.26, 5.29, 5.32, 5.35, 5.38 y 5.41, una notable ventaja en el

procedimiento robusto respecto del clásico. En términos generales, las estimaciones robustas cumplen con su propósito de identificar datos anómalos para introducir un peso apropiado en el método de estimación y lograr estimar de manera realista al parámetro en cuestión.

Respecto del parámetro infinito–dimensional η_0, los boxplots funcionales representados en la Figura 5.6 del escenario de contaminación $C_{1;0,1;49}$ se realizaron en la misma escala para visualizar la diferencia del conjunto imagen en los procedimientos clásico y robusto. Como se observa, el boxplot funcional clásico tiene una estimación atípica de η_0 y toma valores entre -20 y 25, en cambio el boxplot funcional robusto no tiene estimaciones atípicas de η_0 y toma valores entre -10 y 10. Cabe notar que los restantes boxplots funcionales de la primera forma de contaminación de las Figuras 5.9, 5.12 y 5.15 también tienen las características mencionadas a pesar de no tener misma escala. En los modelos de contaminación $C_{2;p}$ con $p \in \{0,1; 0,2; 0,3\}$, cuyos boxplots funcionales se hallan en las Figuras 5.18, 5.21 y 5.24, más allá de la diferencia de valores que toman las versiones clásica y robusta de los boxplots funcionales, cabe destacar que la verdadera función η_0 ni siquiera está en la región central del boxplot funcional clásico, en contraposición con lo que ocurre para la versión robusta. Por otro lado, el modelo de contaminación $C_{2;0,3}$, en la Figura 5.24, tiene muchas estimaciones robustas que resultan ser atípicas, pero a pesar de ello, la verdadera función η_0 está contenida en la región central del boxplot funcional. Respecto de la tercera contaminación, $C_{3;m;\mu_C}$, en los boxplots funcionales de las Figuras 5.27, 5.30, 5.33, 5.36, 5.39 y 5.42 no se visualizan virtudes de la versión robusta respecto de la clásica, sólo se puede remarcar el defecto de la versión clásica en el caso en que $m = 15$: la verdadera función η_0 tiene partes por fuera de las regiones centrales de los boxplots funcionales.

En resumen, los estimadores robustos propuestos son resistentes ante datos anómalos de naturaleza similar a los indicados en los modelos de contaminación $C_{1;p;\sigma_C^2}$, $C_{2;p}$ y $C_{3;m;\mu_C}$ con $p \in \{0,1; 0,2\}$, $\sigma_C^2 \in \{49, 100\}$, $m \in \{10, 15\}$ y $\mu_C \in \{5, 10, 15\}$. Es decir que en situaciones donde haya hasta un 20 % de datos que fueron medidos con errores anómalos, los estimadores robustos que proponemos para el *Modelo Parcialmente Lineal de Índice Simple* logran su objetivo de estimar correctamente. También lo hacen en aquellas situaciones donde hasta el 15 % de puntos (y, x) toma valores fuera de lo común. Es notable, además, la resistencia a datos anómalos que el estimador de η_0 muestra en esos mismos escenarios de contaminación.

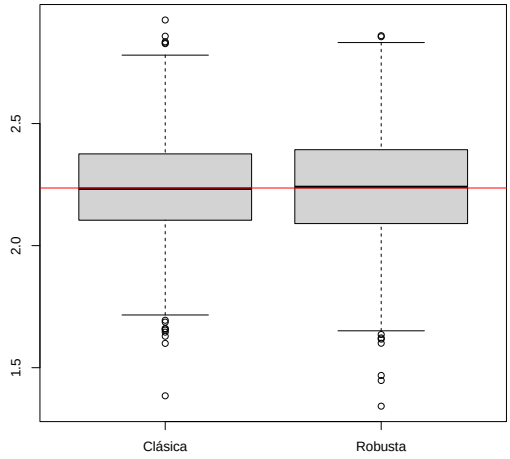

Figura 5.1: Boxplot de las estimaciones de β_0 bajo el modelo (5.1). El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimador robusto.

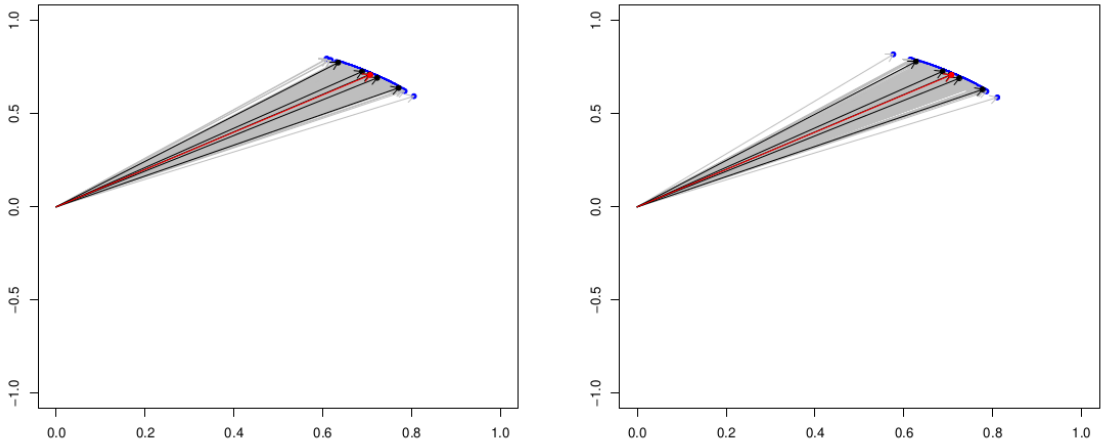

Figura 5.2: Gráfico de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo (5.1). El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

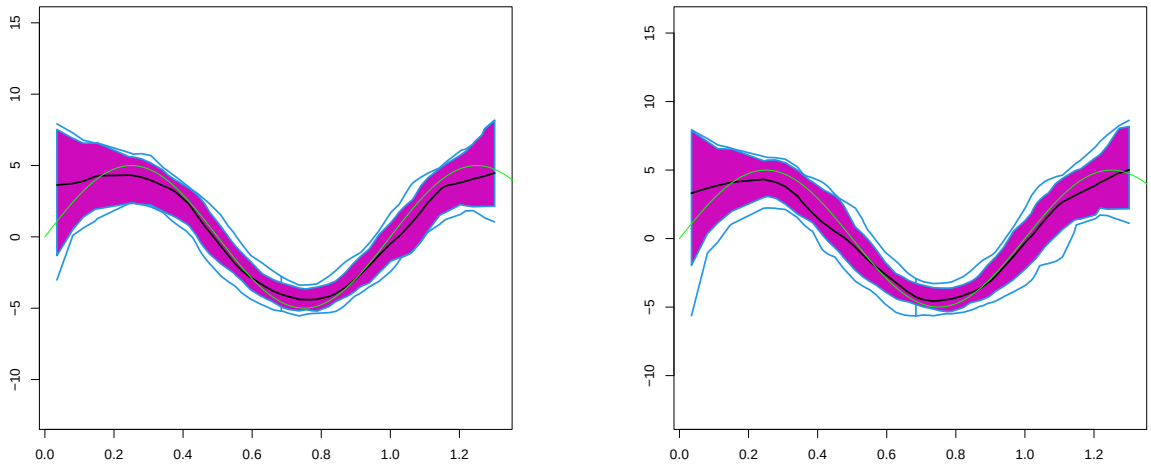

Figura 5.3: Boxplots funcionales de las estimaciones de la función η_0 bajo el modelo (5.1). El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

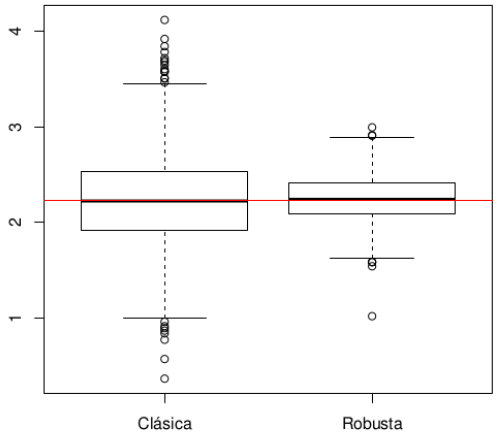

Figura 5.4: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{1;0,1;49}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimador robusto.

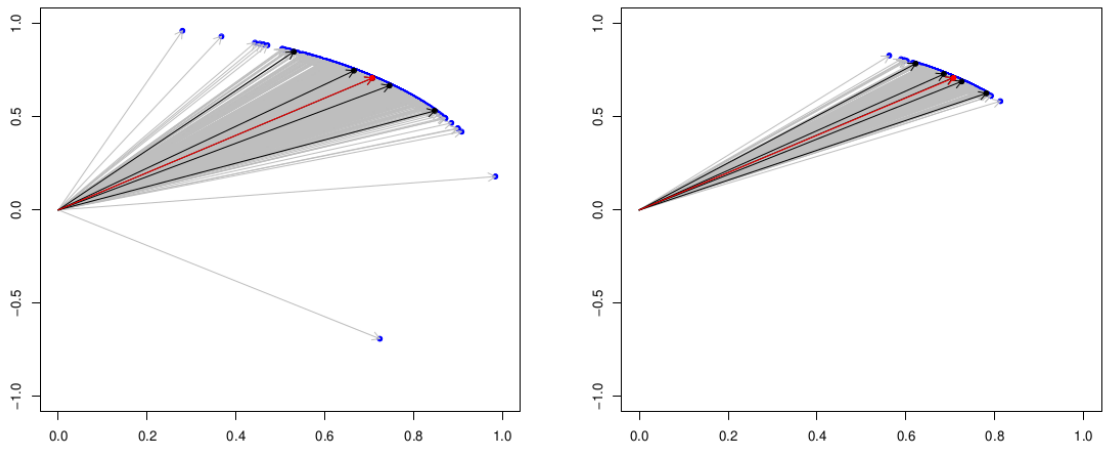

Figura 5.5: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{1;0,1;49}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

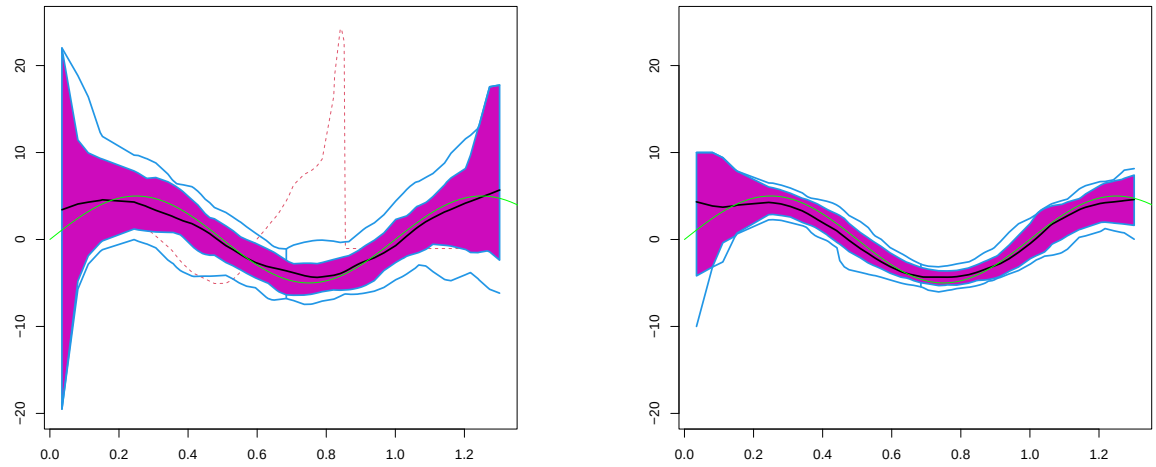

Figura 5.6: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{1;0,1;49}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

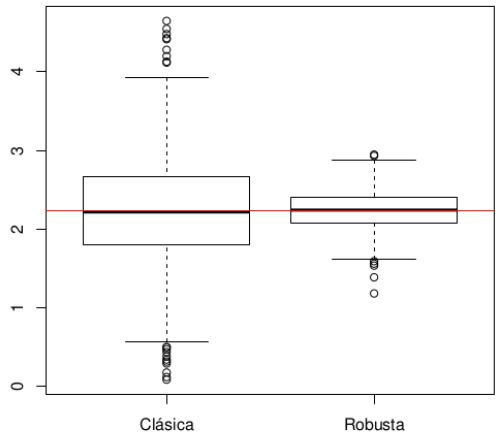

Figura 5.7: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{1;0,1;100}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimador robusto.

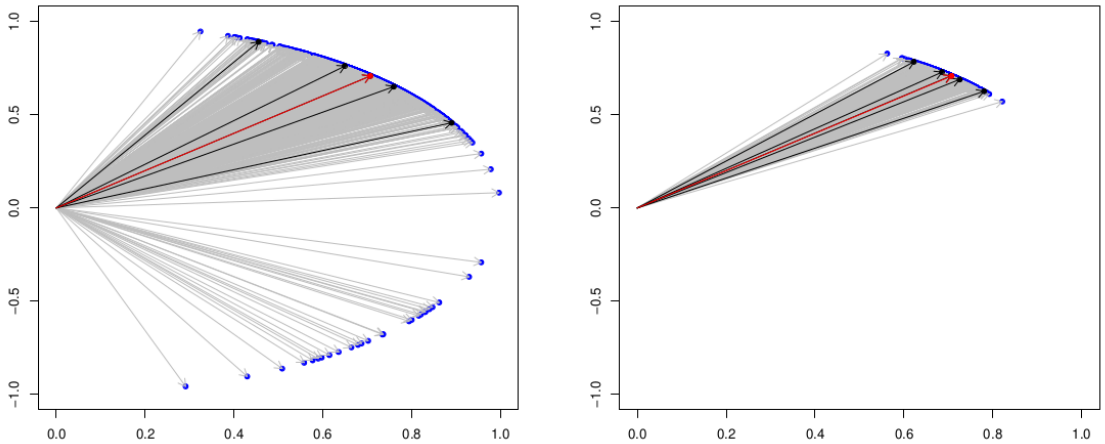

Figura 5.8: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{1;0,1;100}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

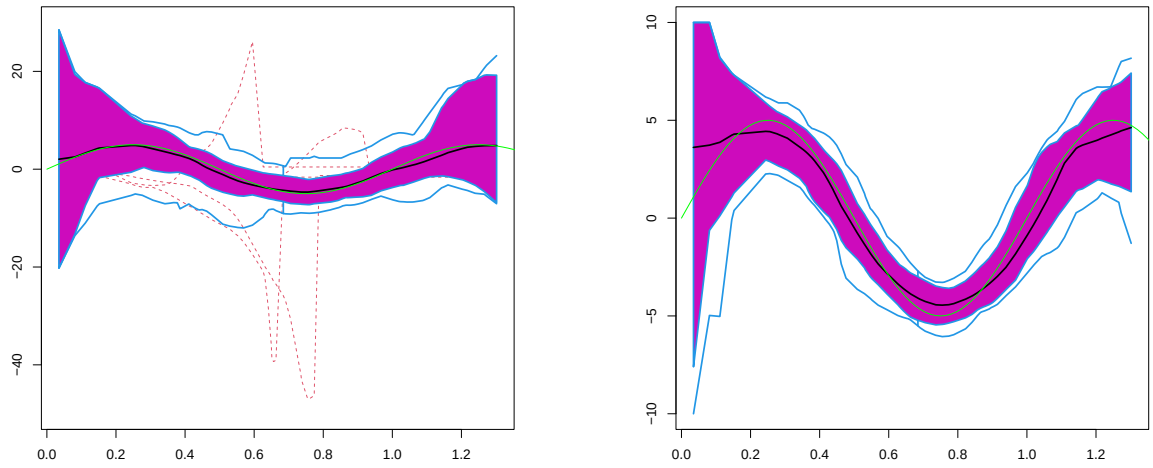

Figura 5.9: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{1;0,1;100}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

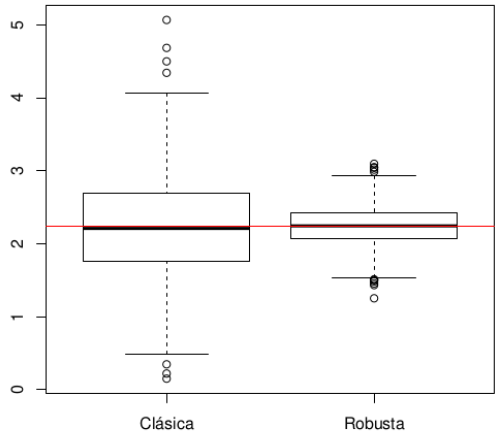

Figura 5.10: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{1;0,2;49}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimador robusto.

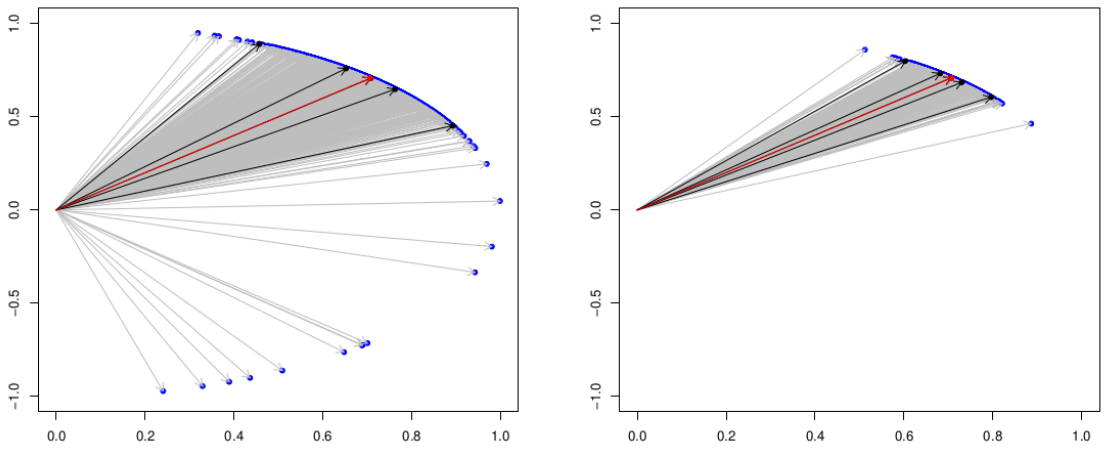

Figura 5.11: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{1;0,2;49}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

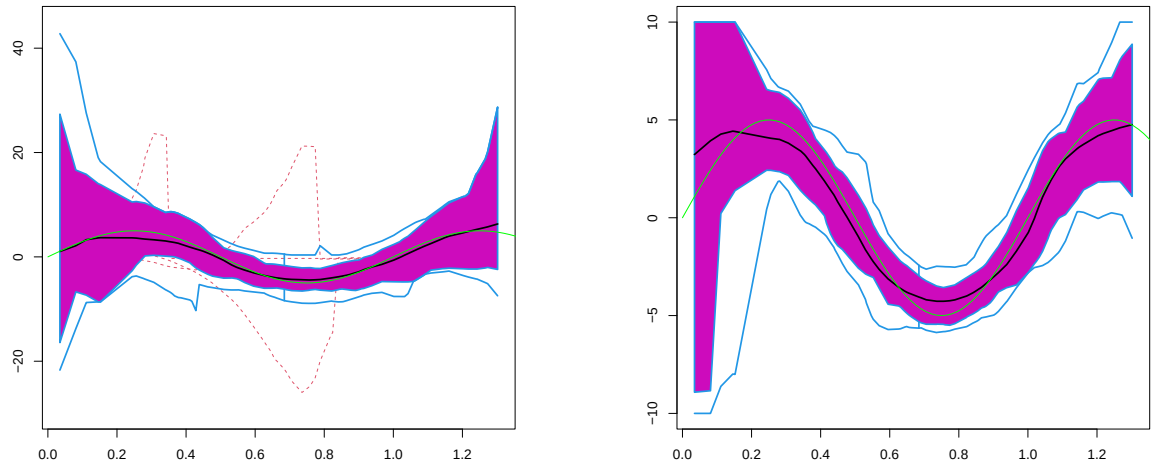

Figura 5.12: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{1;0,2;49}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

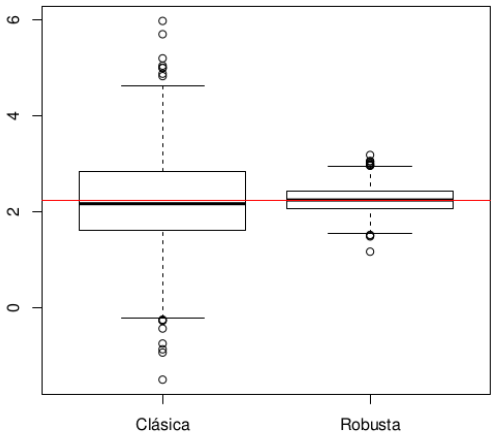

Figura 5.13: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{1;0,2;100}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimador robusto.

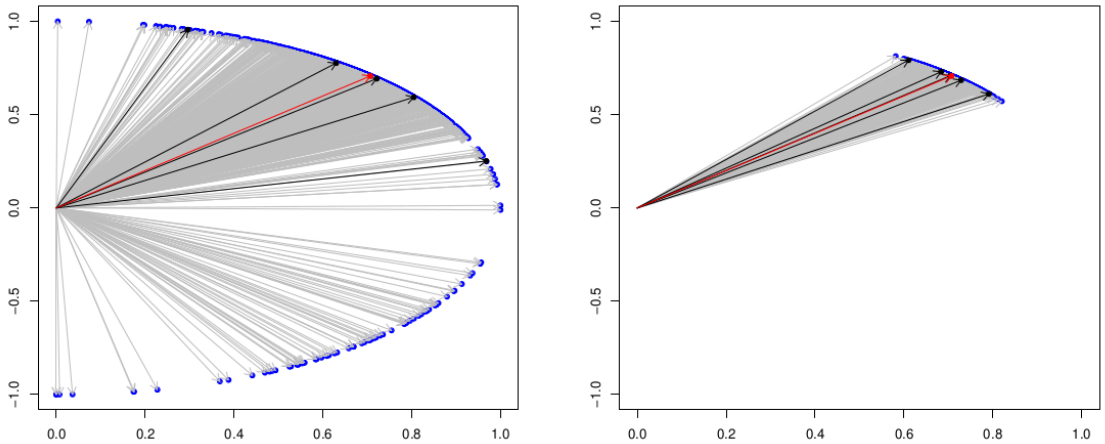

Figura 5.14: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{1;0,2;100}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

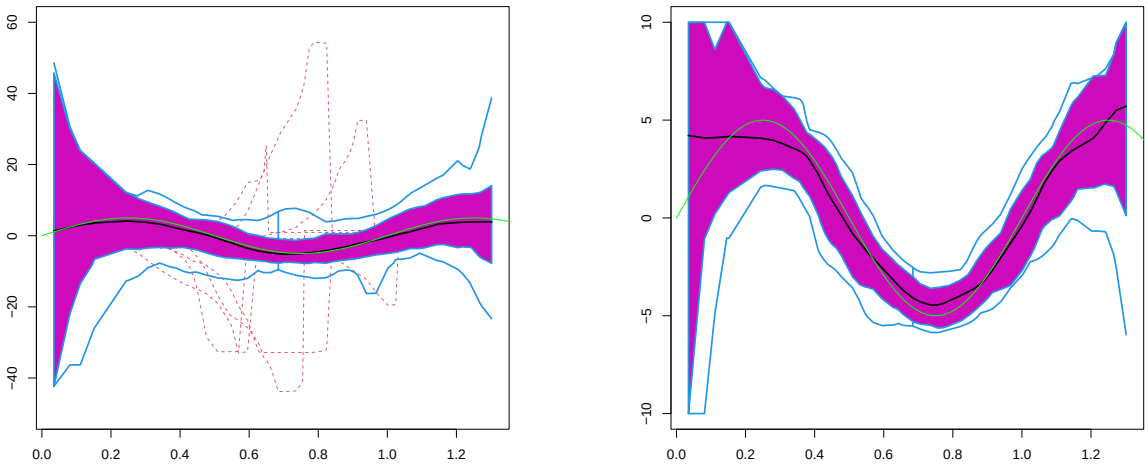

Figura 5.15: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{1;0,2;100}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

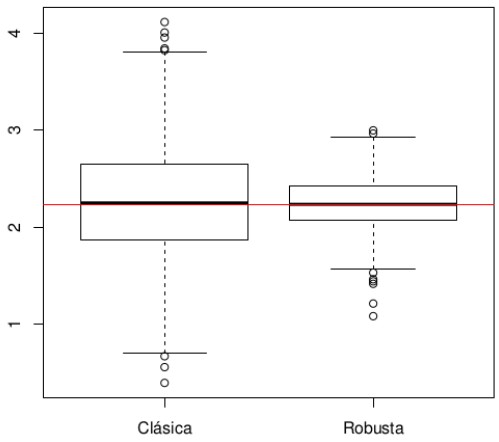

Figura 5.16: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{2;0,1}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimador robusto.

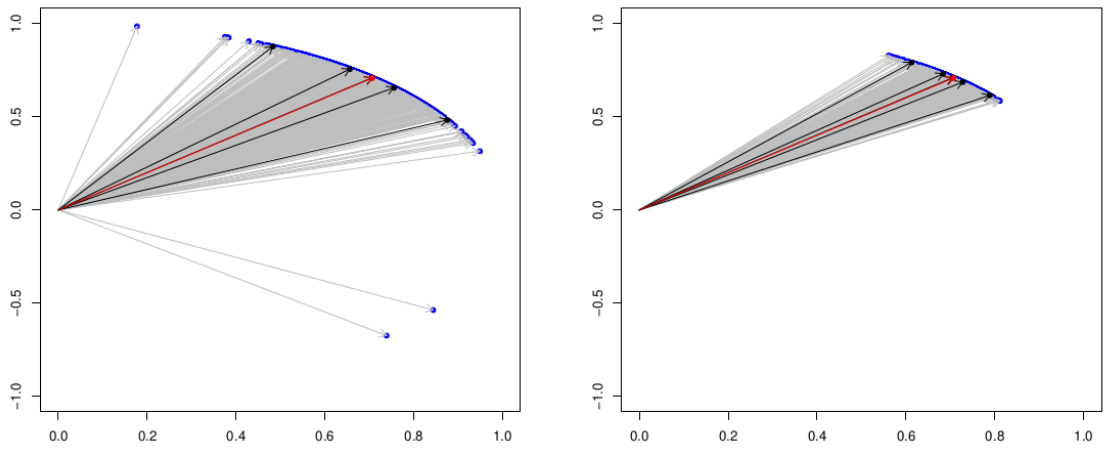

Figura 5.17: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{2;0,1}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

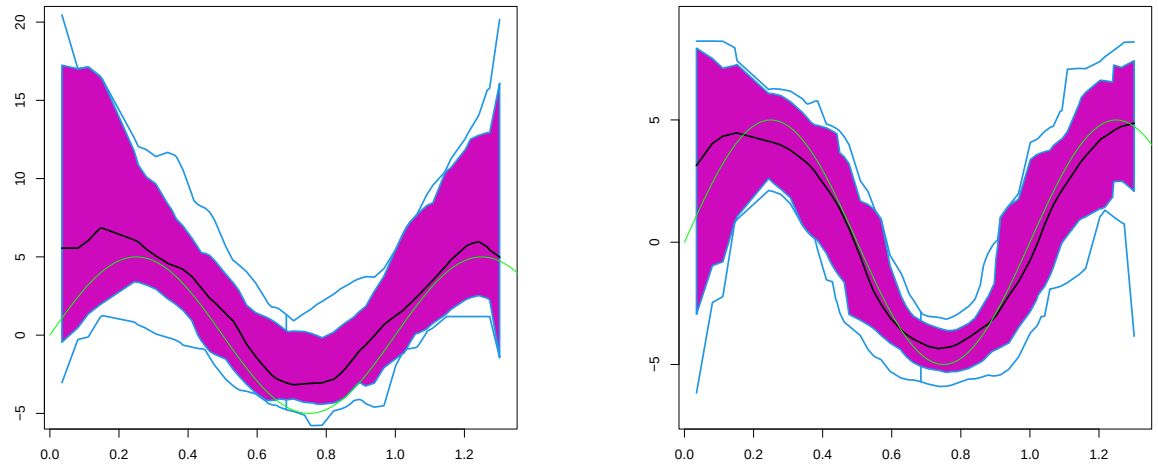

Figura 5.18: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{2;0,1}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

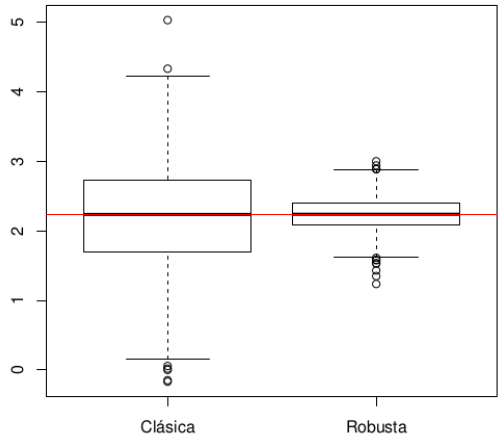

Figura 5.19: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{2;0,2}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

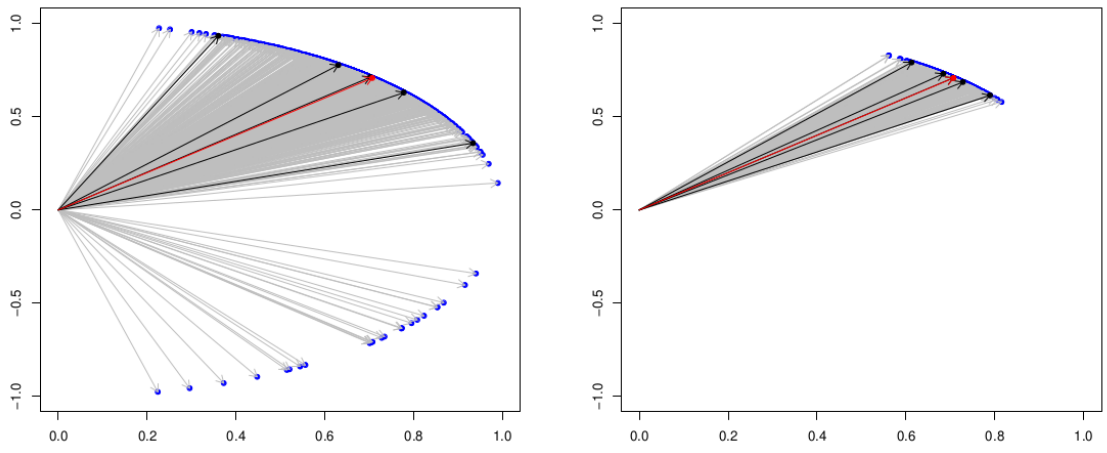

Figura 5.20: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{2;0,2}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

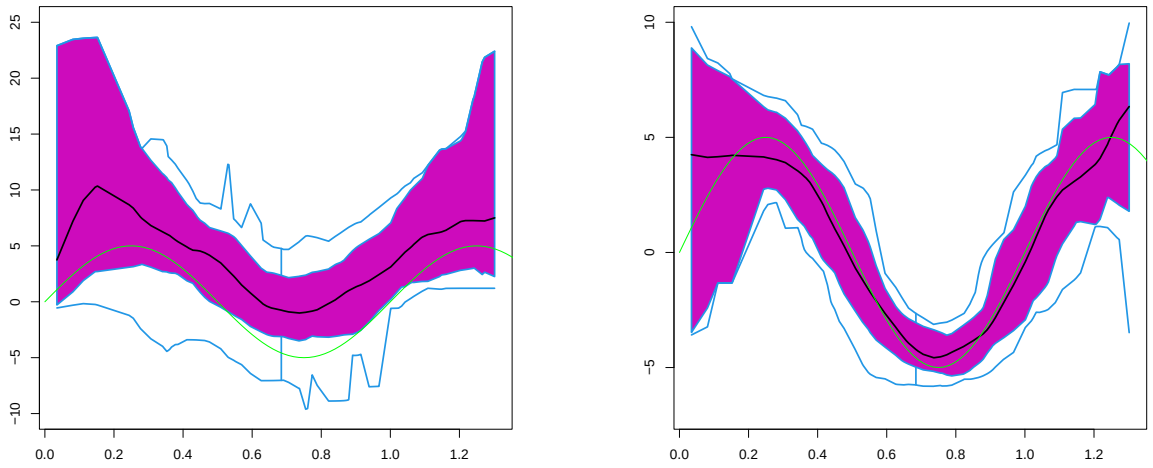

Figura 5.21: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{2;0,2}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

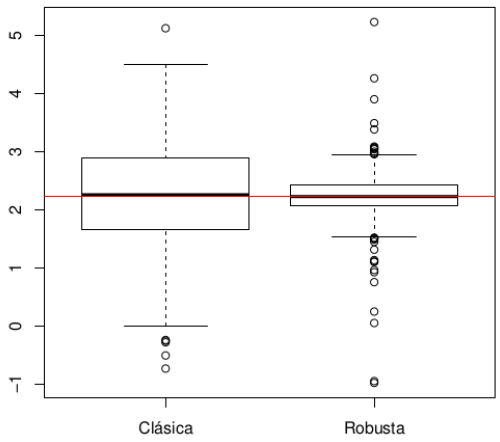

Figura 5.22: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{2;0,3}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

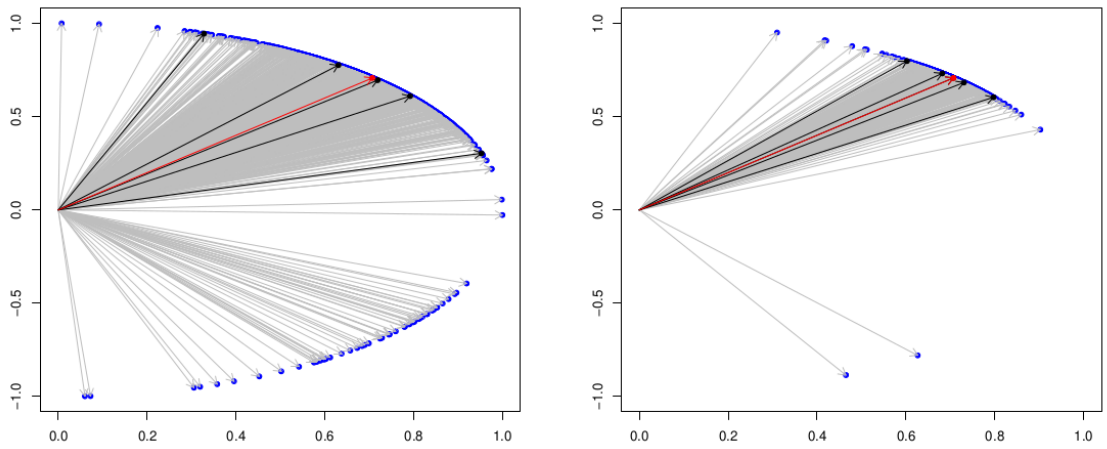

Figura 5.23: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{2;0,3}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

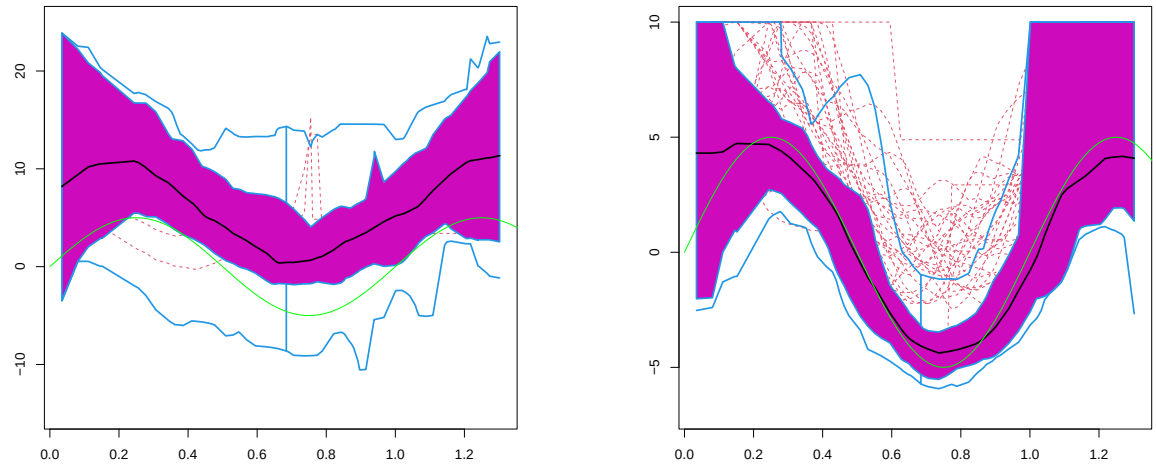

Figura 5.24: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{2;0,3}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

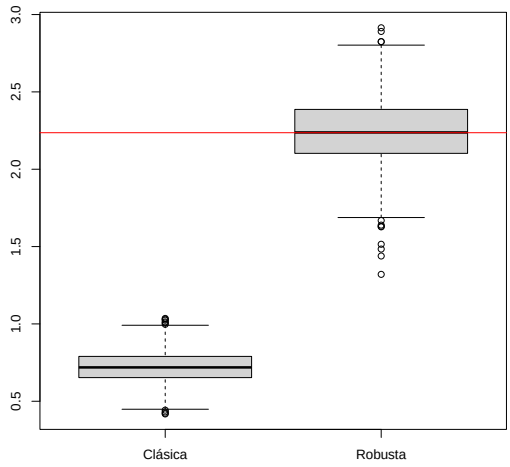

Figura 5.25: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{3;10;5}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

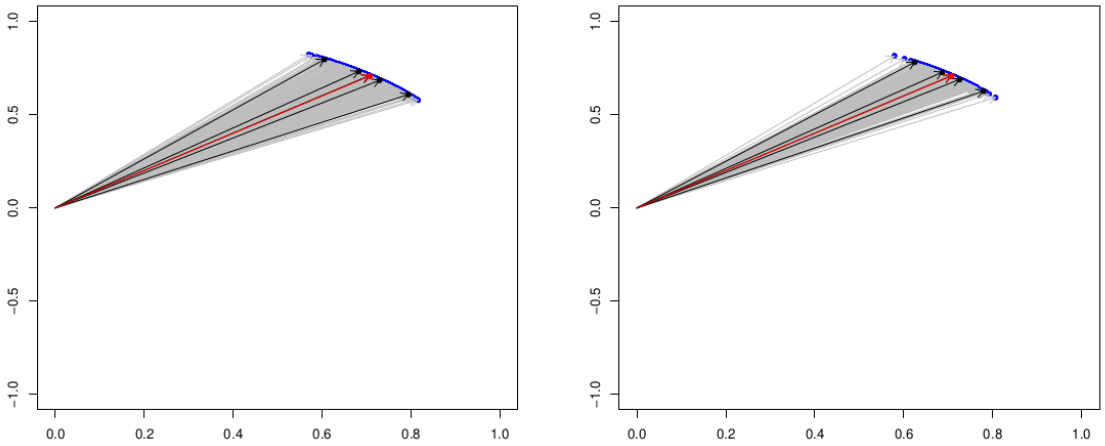

Figura 5.26: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{3;10;5}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

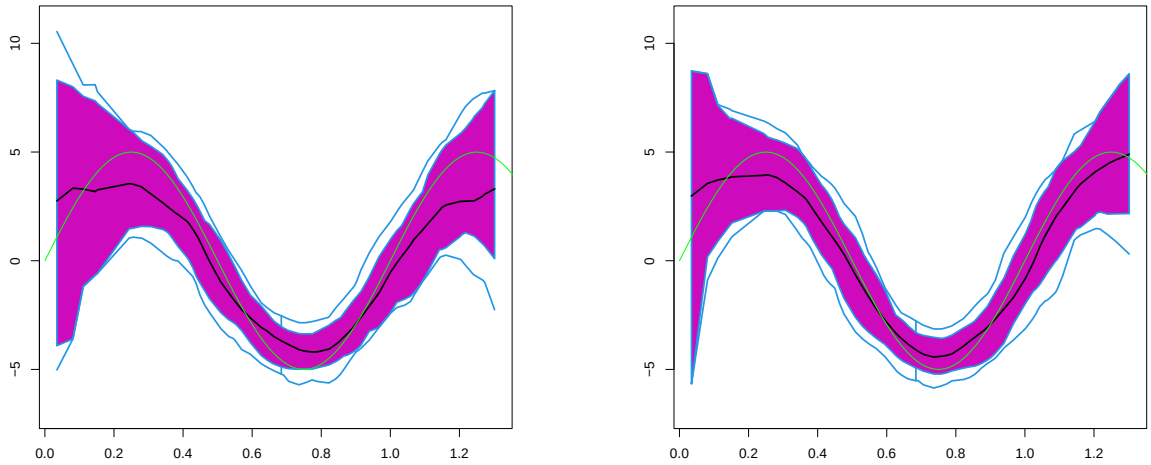

Figura 5.27: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{3;10;5}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

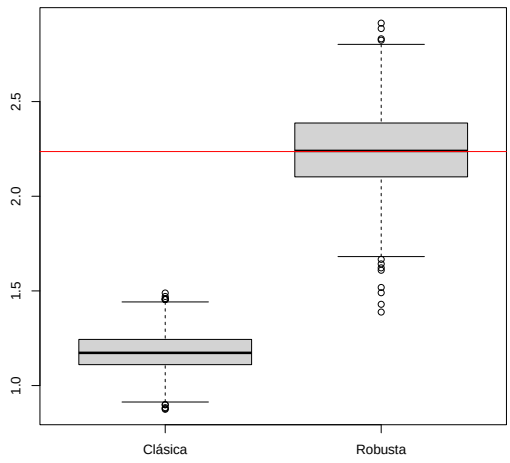

Figura 5.28: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{3;10;10}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

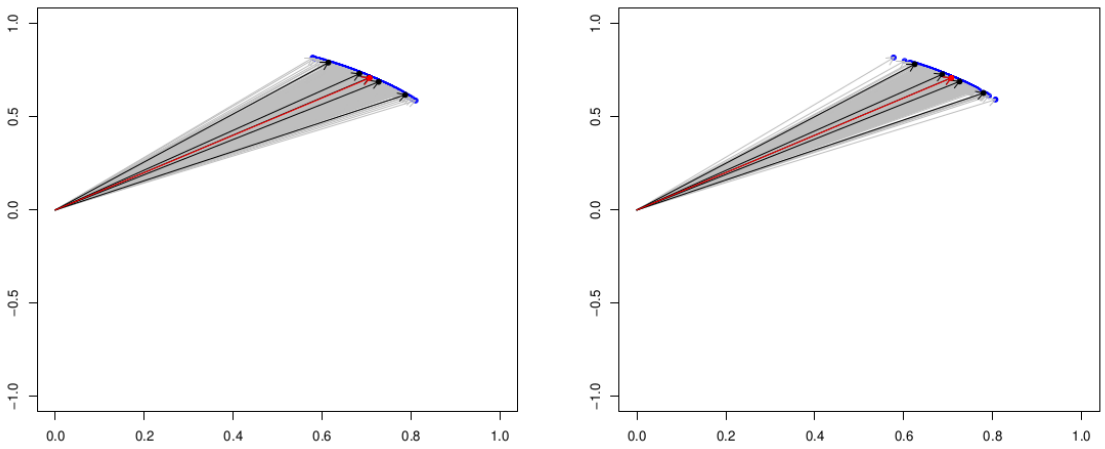

Figura 5.29: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{3;10;10}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

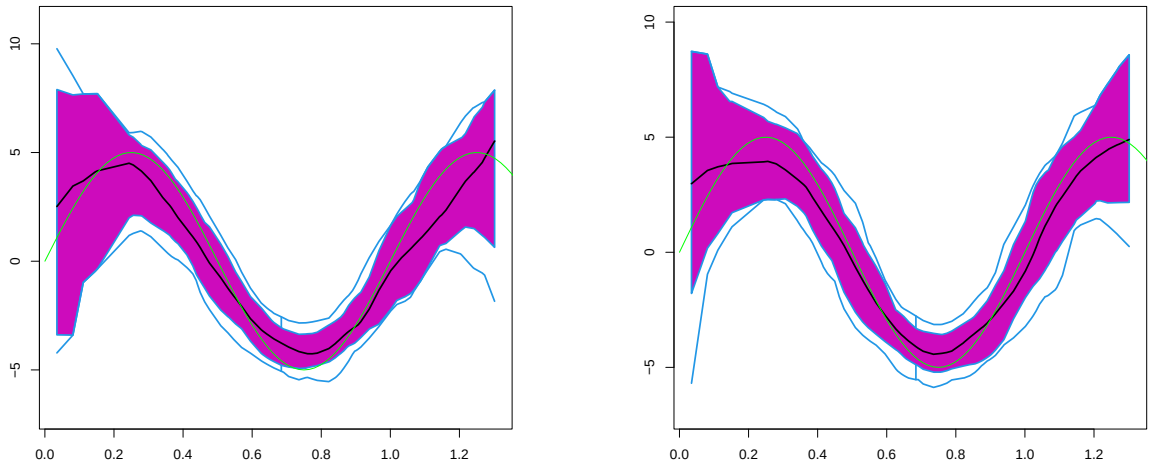

Figura 5.30: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{3;10;10}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

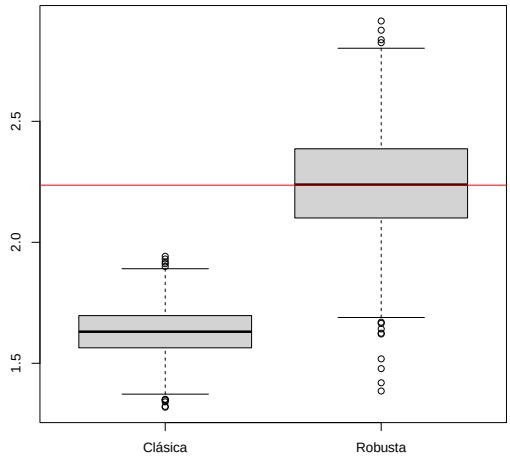

Figura 5.31: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{3;10;15}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

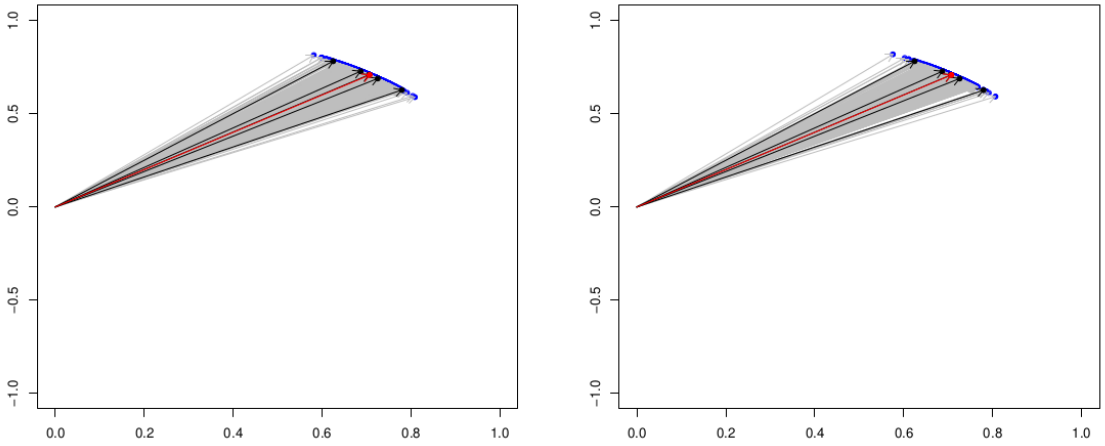

Figura 5.32: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{3;10;15}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

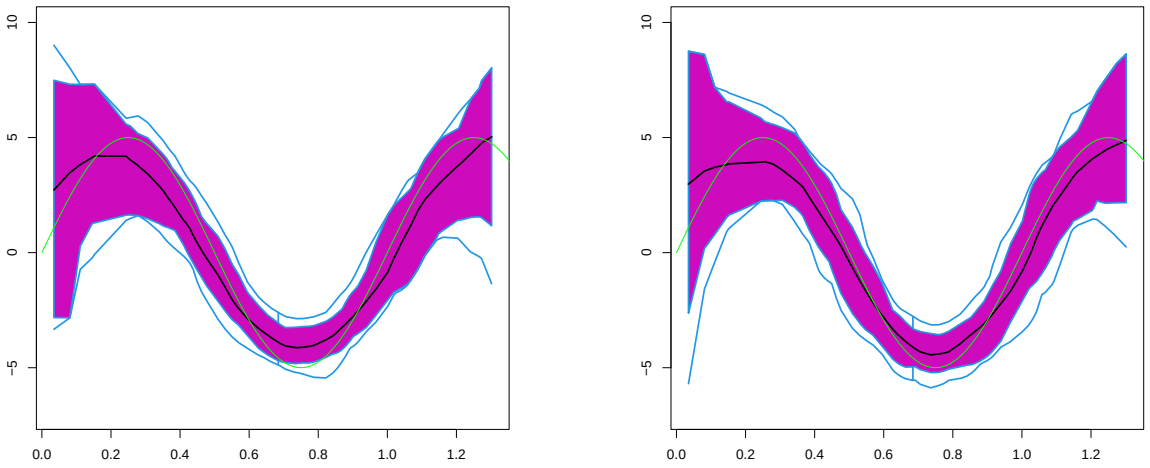

Figura 5.33: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{3;10;15}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

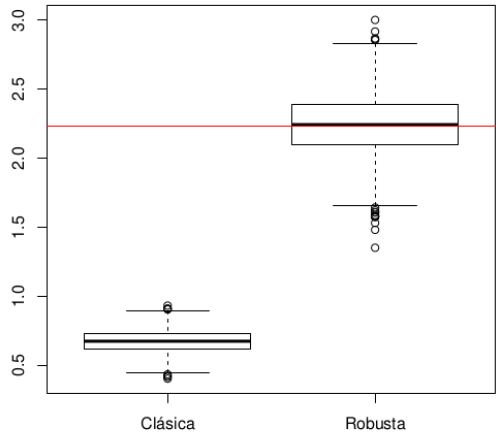

Figura 5.34: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{3;15;5}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

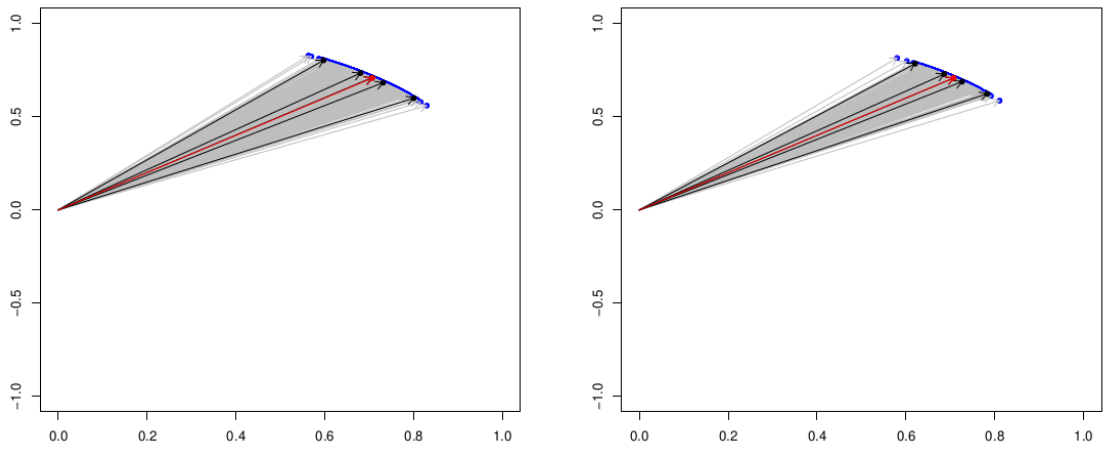

Figura 5.35: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{3;15;5}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

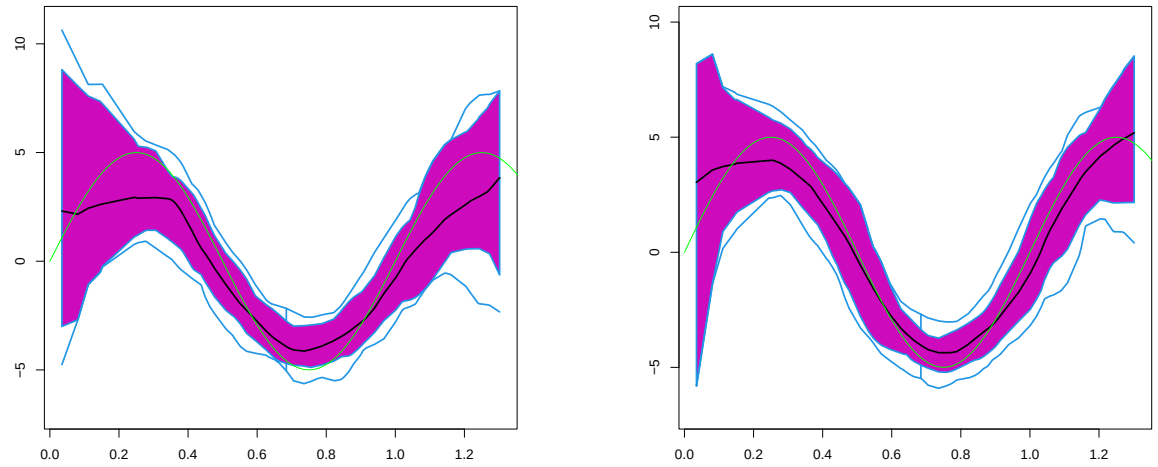

Figura 5.36: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{3;15;5}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

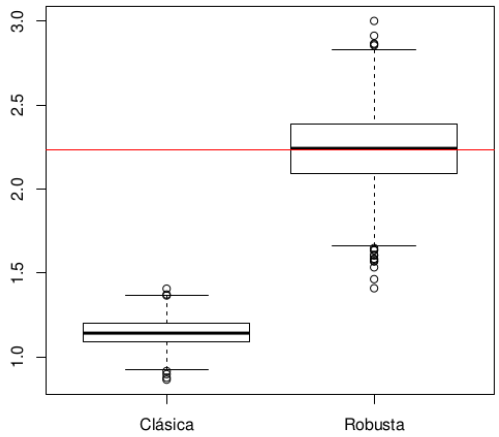

Figura 5.37: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{3;15;10}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

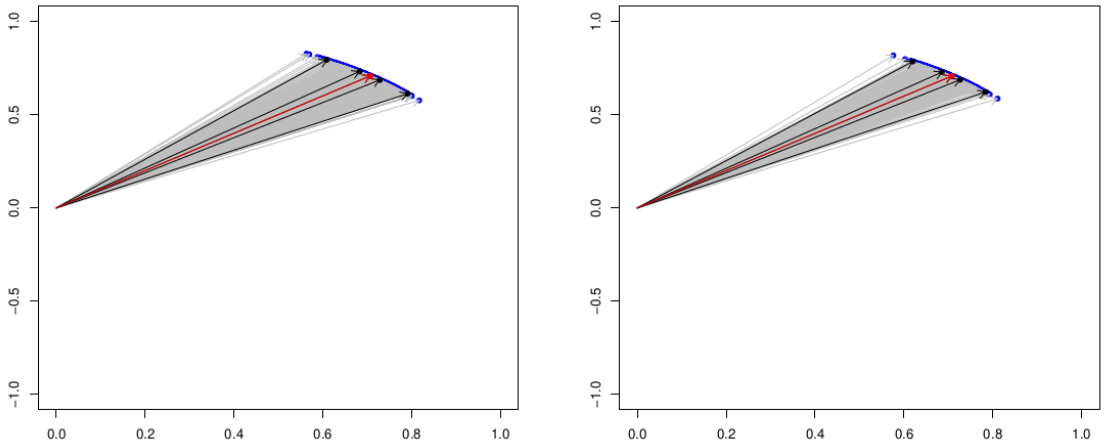

Figura 5.38: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{3;15;10}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

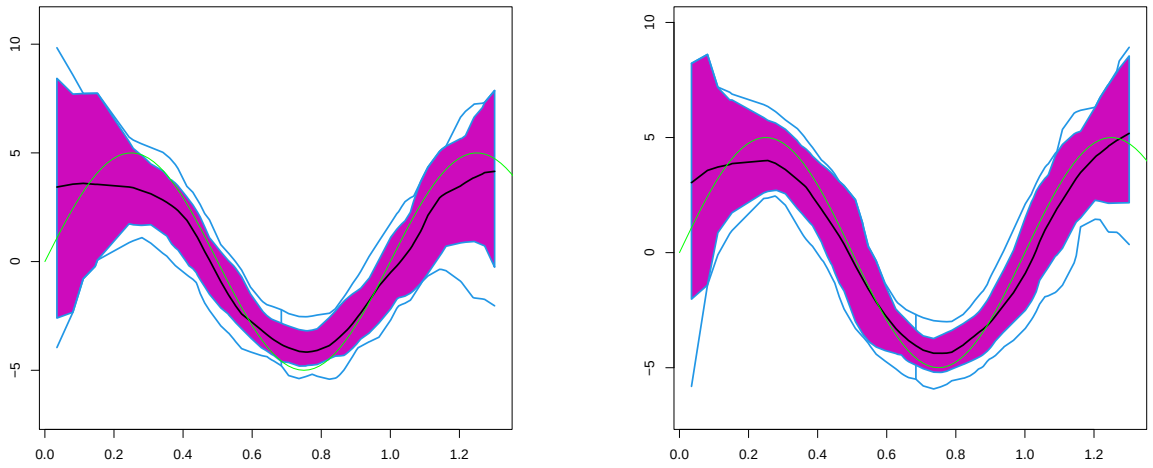

Figura 5.39: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{3;15;10}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

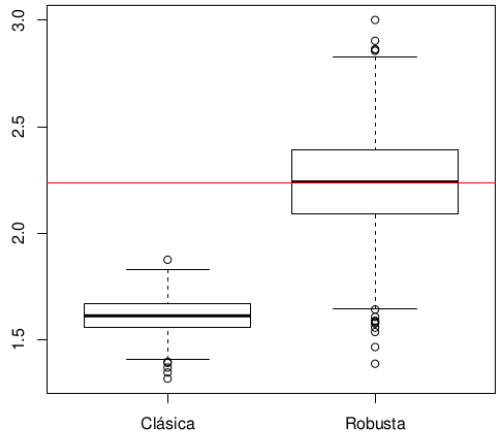

Figura 5.40: Boxplot de las estimaciones de β_0 bajo el modelo de contaminación $C_{3;15;15}$. El boxplot de la izquierda corresponde al estimador clásico y el de la derecha al estimado robusto.

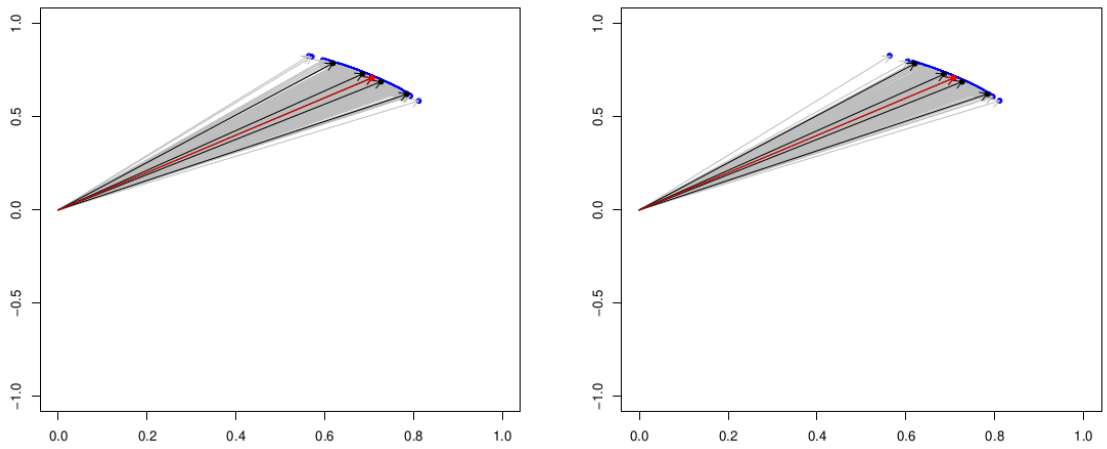

Figura 5.41: Gráficos de flechas de las estimaciones de $\boldsymbol{\theta}_0$ bajo el modelo de contaminación $C_{3;15;15}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las flechas negras representan la mediana, los cuartiles inferior y superior y los bigotes, mientras que la flecha roja indica la verdadera dirección.

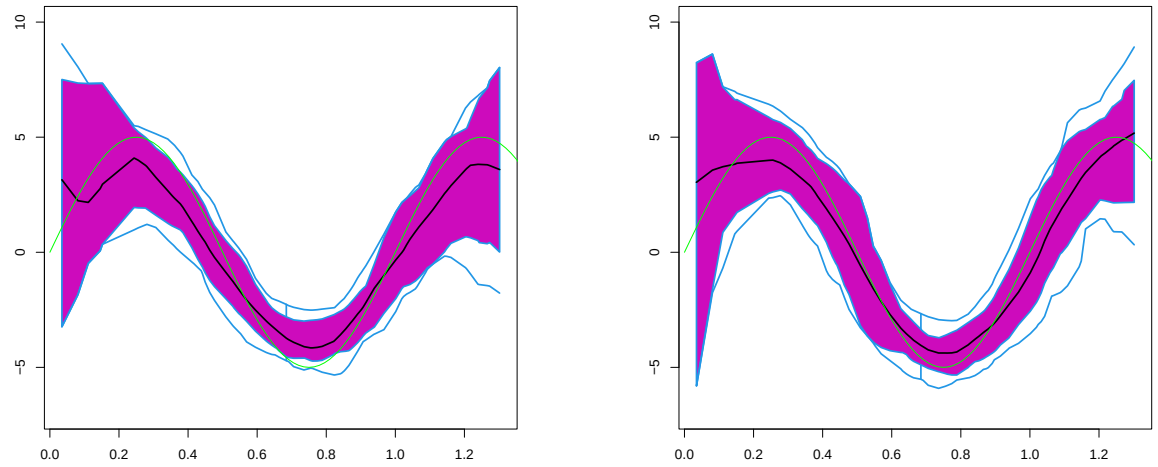

Figura 5.42: Boxplots funcionales de las estimaciones de η_0 bajo el modelo de contaminación $C_{3;15;15}$. El panel de la izquierda corresponde al estimador clásico, mientras que el de la derecha al estimador robusto. Las líneas entrecortadas en rojo indican las curvas identificadas como anómalas por el boxplot funcional, mientras que la curva verde corresponde a la verdadera η_0.

Por último, observemos los Cuadros 5.1 y 5.2, que aparecen en la próxima página y nos brindan medidas resúmenes, clásicas y robustas, de las estimaciones de β_0 y α_0, respectivamente. Recordar que α_0 es el ángulo que forma $\boldsymbol{\theta}_0$ con el vector $(1, 0)$ desde este último. Para empezar a analizar la información que se presenta en los cuadros, contemplemos los primeros dos renglones que se focalizan en el conjunto de datos estudiado sin introducir contaminación alguna. Allí se puede notar que las estimaciones clásicas y robustas se asemejan muchísimo ya que las robustas son superiores en, a lo sumo, un 16 % respecto de las clásicas.

Ahora concentrémonos en el Cuadro 5.1. Para las estimaciones obtenidas en los escenarios de contaminación $C_{1;0,2;100}$ y $C_{2;0,2}$, tanto los desvíos estándares (DE) como las MADs de las estimaciones clásicas son al menos 3 veces las de las robustas; más aún esas estimaciones robustas son, como mucho, sólo el 23 % superior que su correspondiente estimación sin contaminación, lo que permite intuir un excelente desempeño de los estimadores robustos. Sin duda este tipo de situaciones impactan en el error cuadrático medio (ECM), donde las versiones clásicas son entre 9 y 12 veces sus respectivas robustas. Puntualmente, se visualiza la estabilidad del procedimiento robusto comparando los ECMs de los escenarios bajo contaminación con respecto al escenario sin contaminación, donde lo máximo que aumenta es un 51 %, mientras que el procedimiento tradicional tiene ECMs en los escenarios contaminados que son al menos 13 veces el ECM estimado sin contaminar.

Los últimos dos renglones del Cuadro 5.1 muestran las estimaciones para los escenarios de contaminación del tercer tipo, siendo el correlato de los boxplots presentados en la Figura 5.25. Comparando con los primeros dos renglones, se ve que tanto la estimación de la media como del desvío estándar, la mediana y la MAD, en la versión clásica, están subestimadas de tal manera que impacta muy negativamente en la estimación del ECM, que es 52,5 veces la estimación del ECM en el escenario sin contaminación y 44 veces el del correspondiente robusto.

En el caso de las estimaciones de α_0, en el Cuadro 5.2, se observa que la contaminación $C_{1;0,2;100}$ tiene el cociente entre los ECM clásico y robusto igual a 91, más aún el cociente entre los ECM clásico sin contaminación y bajo ésta es igual a 165,67. En realidad, lo mencionado es coherente con las estimaciones de media y desvío estándar obtenidas que son las que influyen en el cálculo de la estimación del ECM. Además la estimación de la mediana que se lee en esa contaminación es el equivalente numérico a lo que se visualiza en el gráfico de flechas, que se encuentra en la Figura 5.14, donde se distingue por un lado la estimación de la mediana y por el otro la verdadera dirección. Bajo el escenario de contaminación $C_{2;0,2}$, las estimaciones de las medidas de dispersión en su versión clásica son por lo menos 3,36 las de la respectiva versión robusta. Para finalizar, como consecuencia de la sobreestimación del desvío estándar, en la contaminación $C_{3;10;5}$, la estimación del ECM clásico es 51 % superior a la de su par robusto.

Contaminación	Versión	Media($\widehat{\beta}$)	DE($\widehat{\beta}$)	ECM($\widehat{\beta}$)	Mediana($\widehat{\beta}$)	MAD($\widehat{\beta}$)
C_0	Clásica	2,237	0,210	0,044	2,233	0,200
	Robusta	2,238	0,225	0,051	2,240	0,224
$C_{1;0,2;100}$	Clásica	2,212	0,962	0,925	2,165	0,939
	Robusta	2,241	0,277	0,077	2,242	0,261
$C_{2;0,2}$	Clásica	2,220	0,764	0,584	2,237	0,770
	Robusta	2,245	0,255	0,065	2,240	0,241
$C_{3;10;5}$	Clásica	0,720	0,103	2,310	0,719	0,102
	Robusta	2,238	0,227	0,052	2,238	0,212

Cuadro 5.1: Medidas resumen para las estimaciones de β_0.

Contaminación	Versión	Media($\widehat{\alpha}$)	DE($\widehat{\alpha}$)	ECM($\widehat{\alpha}$)	Mediana($\widehat{\alpha}$)	MAD($\widehat{\alpha}$)
C_0	Clásica	0,787	0,037	0,00134	0,788	0,0359
	Robusta	0,787	0,0392	0,00154	0,787	0,0393
$C_{1;0,2;100}$	Clásica	0,673	0,457	0,222	0,765	0,189
	Robusta	0,787	0,0494	0,00244	0,788	0,0477
$C_{2;0,2}$	Clásica	0,755	0,295	0,0881	0,792	0,156
	Robusta	0,788	0,0494	0,00244	0,786	0,0464
$C_{3;10;5}$	Clásica	0,787	0,0519	0,00270	0,787	0,0497
	Robusta	0,787	0,0406	0,00165	0,786	0,0404

Cuadro 5.2: Medidas resumen para las estimaciones de α_0 el ángulo que forma $\boldsymbol{\theta}_0$ con el vector $(1, 0)$.

Capítulo 6

Aplicación: ejemplo con datos reales

Utilizaremos un conjunto de datos reales para mostrar una aplicación específica de la propuesta de estimación introducida en esta Tesis. Los 111 datos registrados diariamente corresponden a condiciones meteorológicas en Nueva York entre mayo y septiembre de 1973. Las variables medidas fueron concentración de ozono (O), radiación solar (R), velocidad del viento (V) y temperatura (T). Yu y Ruppert (2002) modelan este conjunto de datos mediante un MPLIS, sin considerar el parámetro de escala del error, para mostrar el buen ajuste que brinda estimar la función no paramétrica η_0 con splines penalizados. Adecuando su planteo a nuestra situación, consideraremos

$$O = \beta_0 R + \eta_0(\theta_{01} V + \theta_{02} T) + \sigma_0 \epsilon \tag{6.1}$$

donde $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02})$ y supondremos que ϵ tiene distribución unimodal y simétrica respecto del 0 .

6.1. Una actualización del estimador de σ_0

Aunque no sea un parámetro de interés para el modelo que se plantea, el comportamiento de las estimaciones de σ_0 nos interesará para hacer las estimaciones de los parámetros usando datos reales y para plantear las pruebas de hipótesis. En ambos objetivos no sólo se utizará el estimador preliminar de escala $\widehat{\sigma}$ si no que también se procederá a utilizar una actualización de la estimación de σ_0, un M-estimador de escala como se introdujo en (2.5), mediante $\widehat{\sigma}_F$ la solución positiva de

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0 \left(\frac{y_i - \widehat{\boldsymbol{\beta}}^t \mathbf{x}_i - \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_i)}{\widehat{\sigma}_F} \right) = \frac{1}{2}\,, \tag{6.2}$$

usando la misma ρ-función ρ_0 que en el Paso I.2 de la propuesta de estimación inicial. Al ser un parámetro de ruido, σ_0 tiene serias dificultades para ser estimado correctamente, en general. En la simulaciones ejecutadas en el Capítulo anterior, aunque no está mencionado, se procedió

a analizar lo que sucedía con las estimaciones obtenidas de este estimador. Se observó que las estimaciones clásicas resultan, en la mayoría de los casos, imprecisas sin siquiera lograr estimar valores alrededor del verdadero valor, mientras que las robustas se mantienen alrededor de él.

6.2. Estimaciones

En primera instancia, se determinaron los parámetros de suavizado, mediante validación cruzada, de las estimaciones tanto clásicas como de las robustas introducida en esta Tesis, para las versiones iniciales y finales. Más precisamente para cada $i \in \{m \in \mathbb{N} : m \leq 111\}$ denotemos por $\left(\widehat{\eta}^{(-i)}_{\widehat{\beta},\widehat{\boldsymbol{\theta}}}, \widehat{\beta}^{(-i)}, \widehat{\boldsymbol{\theta}}^{(-i)}, \widehat{\sigma}^{(-i)}\right)$[1] a la estimación de $(\eta_0, \boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \sigma_0)$ del procedimiento robusto, efectuada extrayendo del conjunto de datos a la i-ésima observación, y calculemos

$$h_{ROB} = \underset{h\in\mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{O_i - \widehat{\beta}^{(-i)} R_i - \widehat{\eta}^{(-i)}_{\widehat{\beta},\widehat{\boldsymbol{\theta}}}\left((V_i, T_i)\widehat{\boldsymbol{\theta}}^{(-i)}\right)}{\widehat{\sigma}^{(-i)}_F}\right) \tag{6.3}$$

siendo $\rho(r) = \rho_T(r/4,685)$ la ρ-función de la familia bicuadrada con $k = 4,685$ y donde $\mathcal{H} \subset \mathbb{R}_{>0}$ es un conjunto finito de posibles ventanas.

En el caso clásico, se calculará

$$h_{CL} = \underset{h\in\mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(O_i - \widehat{\beta}^{(-i)}_{CL} R_i - \widehat{\eta}^{(-i)}_{\widehat{\beta}_{CL},\widehat{\boldsymbol{\theta}}_{CL}}\left((V_i, T_i)\widehat{\boldsymbol{\theta}}^{(-i)}_{CL}\right)\right)^2 \tag{6.4}$$

De esta manera, resultó que $h = 7$ y $h_{CL} = 0,9$.

Una de las habilidades que poseen los procedimientos robustos es identificar datos atípicos. La manera usual de lograrlo es calculando los residuos estandarizados que se derivan de la estimación robusta para luego identificar como observaciones atípicas a aquellas cuyos residuos en módulo son mayores o iguales a 2,5. Se detectan de esta manera 6 datos atípicos en el conjunto de datos analizados: las observaciones 23, 34, 53, 63, 68 y 77. Excluyendo a estas observaciones, se calculó la estimación clásica, que la llamaremos *estimación clásica sin datos atípicos*. Habiendo obtenido la ventana mediante validación cruzada, $h_{CLsda} = 1,9$, se calcula la estimación de esta versión. En el cuadro que figura a continuación se encuentran las estimaciones obtenidas de los parámetros finito-dimensionales de todas las versiones mencionadas.

La Figura 6.1 permite visualizar las estimaciones de la dirección $\boldsymbol{\theta}_0$ en sus versiones clásica (negra), robusta (roja) y clásica sin datos atípicos (verde). Es notable cómo la estimación clásica sin datos atípicos es más cercana a la estimación robusta que a la clásica, lo que sugiere una muy buena efectividad del procedimiento robusto. En la Figura 6.2 se realza la suavidad que los procedimientos robustos introducen en sus estimaciones. Sumamente relevante es que se comprueba en este ejemplo la poca sensibilidad a datos atípicos que tiene la estimación

[1] Obviaremos la dependencia con el parámetro h para que sea más amena la lectura.

	$\widehat{\beta}$	$\widehat{\boldsymbol{\theta}}^t$	$\widehat{\sigma}_F$
ROB	0,0457	(0,8831; -0,4692)	12,257
CL	0,0643	(0,4459; -0,8951)	13,393
$CLsda$	0,0385	(0,8102; -0,5861)	10,446

Cuadro 6.1: Estimaciones de los parámetros finito-dimensionales.

robusta de η_0, representada en color rojo en el gráfico de la derecha de la Figura 6.2. Cabe señalar que la estimación clásica de η_0 sufre la selección de tan pequeña ventana habiendo datos a mayor distancia que ella, y es por eso que estima constante en una parte.

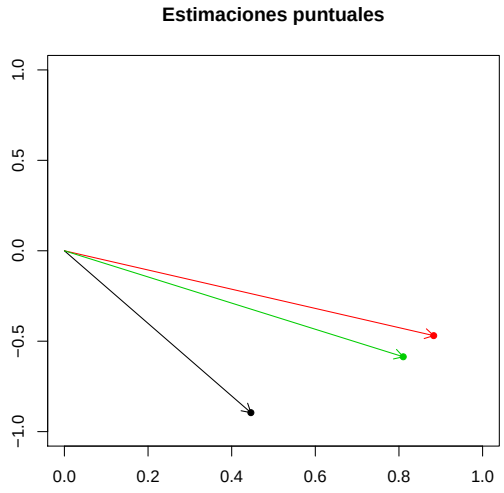

Figura 6.1: Estimaciones de $\boldsymbol{\theta}_0$ para el conjunto de datos reales. En negro la versión clásica, en roja la robusta y en verde la clásica sin datos atípicos.

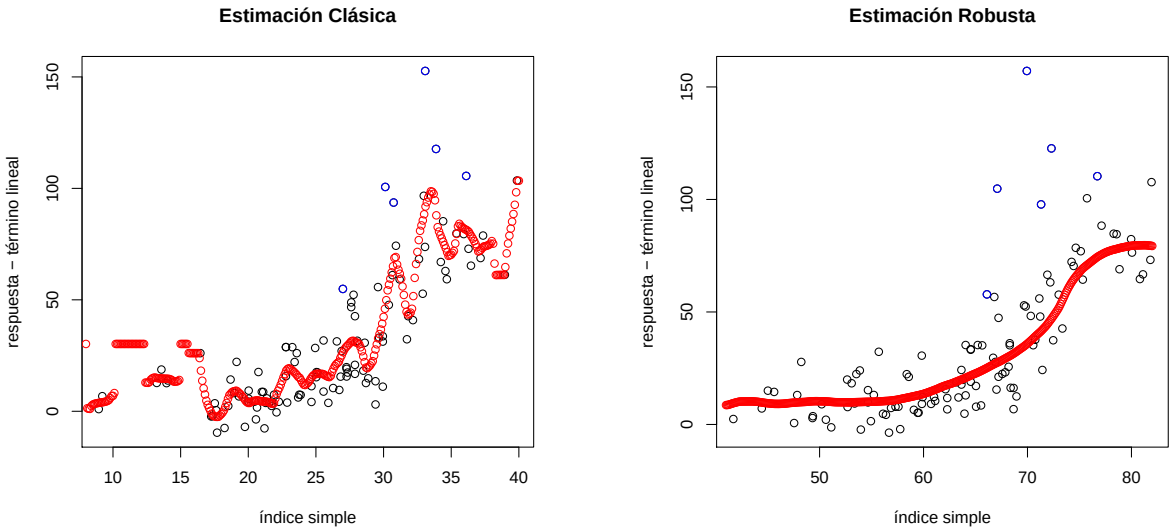

Figura 6.2: En el panel de la izquierda se encuentra el gráfico del índice simple, $(V,T)\widehat{\boldsymbol{\theta}}_{CL}$, contra residuos, $O-\widehat{\beta}_{CL}R$, en la versión clásica. En el panel de la derecha se encuentra el gráfico del índice simple, $(V,T)\widehat{\boldsymbol{\theta}}_{ROB}$, contra residuos, $O-\widehat{\beta}_{ROB}R$, en la versión robusta. En rojo se encuentra graficada la estimación de η_0 en cada escenario. En color azul se resaltan los puntos atípicos detectados.

Capítulo 7

Tests de Hipótesis

7.1. Estadístico

Gran parte de la actividad en robustez concierne al proceso de estimación, pero más allá de desarrollar estimadores robustos, el problema de realizar tests robustos también merece gran atención. De hecho, los tests de hipótesis son parte de la práctica diaria que realiza un usuario. Por ejemplo, cuando se ajusta un modelo lineal, después del proceso de estimación y a fin de completar el análisis, se suelen hacer tests individuales sobre cada parámetro para chequear si es nulo o no, y así facilitar la interpretación del ajuste realizado. Como un primer paso en esta dirección y de forma preliminar, en este capítulo de la Tesis proponemos un estadístico basado en el estimador estudiado que permita chequear hipótesis que involucren a la componente lineal del modelo y explorar su comportamiento.

Como hemos mencionado, en general, los test robustos han recibido un tratamiento menos extendido que la estimación robusta. Sin embargo, es sabido que los procedimientos de tests de hipótesis basados en la metodología clásica suelen heredar su sensibilidad a datos atípicos, en el sentido de que una pequeña cantidad de observaciones puede afectar el nivel o la potencia de los tests.

Es así que desarrollar tests de hipótesis que bajo contaminación retengan un nivel de significación estable es deseable. Los trabajos de Heritier y Ronchetti (1994), Cantoni y Ronchetti (2001) figuran entre los primeros que van en esta dirección en el campo de modelos paramétricos, el primero en un contexto general, mientras que el segundo está más enfocado a un modelo lineal generalizado. Estos autores también investigan la estabilidad del nivel asintótico bajo contaminación. Más recientemente, Bianco, Boente y Martínez (2006) y Bianco y Martínez (2009) estudian tests robustos en el caso del modelo parcialmente lineal y en el modelo logístico, respectivamente. Maronna et al. (2019) tratan el problema de tests robustos y en particular, se ocupan en el modelo lineal de los tests robustos de tipo Wald.

En el caso del modelo estudiado en esta Tesis, la complejidad intrínseca del modelo que presenta una parte paramétrica y otra no paramétrica, hacen que el estudio de tests de hipótesis se vuelva un mayor desafío. Liang et al. (2010) desarrollan pruebas de hipótesis

lineales para los coeficientes lineal e índice simple y proponen un test de bondad de ajuste para la componente no paramétrica. Este trabajo utiliza un método de perfiles que, al basarse en mínimos cuadrados, permite que datos atípicos influyan en la estimación y en consecuencia, en los estadísticos de las pruebas de hipótesis que se consideran allí.

Proponemos pruebas de hipótesis que involucran al parámetro lineal basadas en un estadístico de tipo Wald que utiliza el estimador estudiado en los capítulos anteriores con el objetivo de que sean resistentes a la presencia de un pequeño porcentaje de observaciones anómalas.

Suponemos que tenemos una muestra aleatoria de vectores $(y_i, \mathbf{x}_i, \mathbf{t}_i) \subset \mathbb{R}^{p+q+1}$, $1 \leq i \leq n$, que siguen el modelo (3.2), o sea

$$y_i = \boldsymbol{\beta}_0^t \mathbf{x}_i + \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}_i) + \sigma_0 \epsilon_i \tag{7.1}$$

y el objetivo será decidir entre las hipótesis

$$H_0 : \boldsymbol{\beta}_0 = \boldsymbol{\beta}_* \quad \text{contra} \quad H_1 : \boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_* \,. \tag{7.2}$$

Como antes, dado $\boldsymbol{\theta} \in \mathbb{R}^q$, denotamos por $\boldsymbol{\theta}^{(q-1)} = (\theta_1, ..., \theta_{q-1})^t$. Por el Teorema 4.1.4 bajo condiciones de regularidad, tal como fue establecido en el Capítulo 4, tenemos que

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}^{(q-1)} - \boldsymbol{\theta}_0^{(q-1)} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{M}_1^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \boldsymbol{\Sigma}_1(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \mathbf{M}_1^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)) \,,$$

donde las matrices $\mathbf{M}_1 = \mathbf{M}_1(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ y $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ fueron definidas en **N9** y **N10**, pero hemos hecha explícita su dependencia a los parámetros de la componente lineal y de índice simple. Recordemos que $\mathbf{M}_1(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ es la submatriz cuadrada, no singular, superior izquierda de la matriz definida en (4.7) como

$$\mathbf{M}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \frac{1}{\sigma_0^2} \mathbb{E}(\chi_1(\epsilon_1) w_2(\mathbf{x}_1) \boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) \boldsymbol{\lambda}_1^t(\boldsymbol{\tau}_{01})) \,,$$

siendo $\boldsymbol{\tau}_{01} = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\theta}_0^t \mathbf{t}_1)$ y

$$\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) = \left(\begin{array}{c} \mathbf{x}_1 + \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) \\ \frac{\partial}{\partial \boldsymbol{\theta}} \eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) + \frac{\partial}{\partial u} \eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) \mathbf{t}_1 \end{array} \right) \Bigg|_{(\boldsymbol{\beta},\boldsymbol{\theta},u)=\boldsymbol{\tau}_{01}} .$$

Asimismo, tenemos que $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ es la submatriz cuadrada, no singular, superior izquierda de la matriz definida en (4.11)

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \mathbb{E}\left(\mathbf{w}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \mathbf{w}^t(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\right) \,,$$

siendo

$$\mathbf{w}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \frac{1}{\sigma_0} \psi_1(\epsilon_1) w_2(\mathbf{x}_1) \boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) - \frac{1}{\sigma_0^2} E_1(\boldsymbol{\theta}_0^t \mathbf{t}_1) \boldsymbol{\gamma}(\mathbf{t}_1) f_{\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t \mathbf{t}_1\right)$$

donde $f_{\boldsymbol{\theta}_0}$ es la densidad marginal de $\boldsymbol{\theta}_0^t \mathbf{t}_1$,

$$E_1(u) = \frac{\sigma_0}{A^1(u)} \psi_1 \left(\frac{y_1 - \boldsymbol{\beta}_0^t \mathbf{x}_1 - \eta_0(u)}{\sigma_0} \right) w_1(\mathbf{x}_1) \,,$$

$$A^1(u) = \mathbb{E}(\chi_1(\epsilon_1) w_1(\mathbf{x}) | \boldsymbol{\theta}_0^t \mathbf{t}_1 = u)$$

y

$$\boldsymbol{\gamma}(\tilde{\mathbf{t}}) = \mathbb{E} \left(\frac{1}{f_{\boldsymbol{\theta}_0}\left(\boldsymbol{\theta}_0^t \mathbf{t}_1\right)} \chi_1(\epsilon_1) w_2(\mathbf{x}_1) \boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) \middle| \mathbf{t}_1 = \tilde{\mathbf{t}} \right) .$$

Denotemos por

$$\boldsymbol{\Delta}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \mathbf{M}_1^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \boldsymbol{\Sigma}_1(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) \mathbf{M}_1^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$$

a la matriz de covarianzas asintóticas. Como recuerdan Heritier y Ronchetti (1994) el estadístico de Wald puede definirse a partir de la distribución asintótica como

$$\mathbf{W}_n = n \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \right)^t \boldsymbol{\Delta}_1^{-1} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \right) , \tag{7.3}$$

donde $\boldsymbol{\Delta}_1$ es la submatriz de orden p de $\boldsymbol{\Delta}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ correspondiente a las p coordenadas del parámetro de la componente lineal. A partir del Teorema 4.1.4, es inmediato que bajo H_0, $\mathbf{W}_n = n \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \right)^t \boldsymbol{\Delta}_1^{-1} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \right) \xrightarrow{\mathcal{D}} \chi_p^2$. Como notan estos autores, en la práctica es necesario reemplazar a $\boldsymbol{\Delta}_1^{-1}$ por un estimador consistente, para obtener la siguiente convergencia débil

$$\widehat{\mathbf{W}}_n = n \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \right)^t \widehat{\boldsymbol{\Delta}}_1^{-1} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \right) \xrightarrow{\mathcal{D}} \chi_p^2 . \tag{7.4}$$

A partir de ésto, resulta que la regla que establece rechazar H_0 si $\widehat{\mathbf{W}}_n > \chi_{p,\alpha}^2$, es un test asintótico de nivel α.

A continuación se detallarán dos propuestas de estimación de $\boldsymbol{\Delta}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ que nos permitirán construir estadísticos de tipo Wald del estilo presentado.

7.2. Estimación de la matriz de covarianzas

7.2.1. Estimador de $\boldsymbol{\Delta}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ usando una estimación de $\boldsymbol{\beta}_0$

Estimaremos a $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ por $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}\right)$, a η_0 por $\widehat{\eta}_{\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}}$, a σ_0 por $\widehat{\sigma}_F$ definido en (6.2) y a $\mathbf{M}$ y a $\boldsymbol{\Sigma}$ por sus versiones empíricas. O sea que

$$\widehat{\boldsymbol{\Delta}} = \widehat{\mathbf{M}}_1^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) \widehat{\boldsymbol{\Sigma}}_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) \widehat{\mathbf{M}}_1^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) \,, \tag{7.5}$$

llamando a $\widehat{\mathbf{M}} = \widehat{\mathbf{M}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ y a $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$. Tendremos que

$$\widehat{\mathbf{M}} = \frac{1}{\widehat{\sigma}_F^2} \frac{1}{n} \sum_{i=1}^{n} \chi_1 \left(\frac{y_i - \widehat{\boldsymbol{\beta}}^t \mathbf{x}_i - \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_i)}{\widehat{\sigma}_F} \right) w_2(\mathbf{x}_i) \widehat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_i) \widehat{\boldsymbol{\lambda}}_i^t(\hat{\boldsymbol{\tau}}_i) \tag{7.6}$$

siendo $\hat{\boldsymbol{\tau}}_i = \left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}^t \mathbf{t}_i\right)$ y $\widehat{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \boldsymbol{\theta}, u) = \begin{pmatrix} \mathbf{x}_i + \frac{\partial}{\partial \boldsymbol{\beta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) \\ \frac{\partial}{\partial \boldsymbol{\theta}} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) + \frac{\partial}{\partial u} \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) \mathbf{t}_i \end{pmatrix}$, y

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{w}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) \widehat{\mathbf{w}}_i^t(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) \,, \tag{7.7}$$

con

$$\widehat{\mathbf{w}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = \frac{1}{\widehat{\sigma}_F} \psi_1 \left(\frac{y_i - \widehat{\boldsymbol{\beta}}^t \mathbf{x}_i - \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_i)}{\widehat{\sigma}_F} \right) w_2(\mathbf{x}_i) \widehat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_i) - \frac{1}{\widehat{\sigma}_F^2} \widehat{E}_i(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_i) \widehat{\boldsymbol{\gamma}}(\mathbf{t}_i) \hat{f}_h \left(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_i \right) ,$$

$$\widehat{E}_i(u) = \frac{\widehat{\sigma}_F}{\hat{A}^1(u)} \psi_1 \left(\frac{y_i - \widehat{\boldsymbol{\beta}}^t \mathbf{x}_i - \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}}(u)}{\widehat{\sigma}_F} \right) w_1(\mathbf{x}_i) \,,$$

$$\hat{A}^1(u) = \sum_{j=1}^{n} \chi_1 \left(\frac{y_j - \widehat{\boldsymbol{\beta}}^t \mathbf{x}_j - \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_j)}{\widehat{\sigma}_F} \right) w_1(\mathbf{x}_j) W^h_{\widehat{\boldsymbol{\theta}},j}(u) \,,$$

$$\widehat{\boldsymbol{\gamma}}(\tilde{\mathbf{t}}) = \sum_{j=1}^{n} \frac{1}{\hat{f}_h \left(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_j \right)} \chi_1 \left(\frac{y_j - \widehat{\boldsymbol{\beta}}^t \mathbf{x}_j - \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}^t \mathbf{t}_j)}{\widehat{\sigma}_F} \right) w_2(\mathbf{x}_j) \widehat{\boldsymbol{\lambda}}_j(\hat{\boldsymbol{\tau}}_j) U_j^{\mathbf{h}}(\tilde{\mathbf{t}}) \,,$$

$$\text{donde} \qquad U_j^{\mathbf{h}}(\tilde{\mathbf{t}}) = \frac{\prod_{l=1}^{q} K_l \left(\frac{t_{jl} - \tilde{t}_l}{h_l} \right)}{\sum_{i=1}^{n} \prod_{l=1}^{q} K_l \left(\frac{t_{il} - \tilde{t}_l}{h_l} \right)} \,,$$

siendo para cada l, K_l un núcleo univariado, $\mathbf{h} = (h_1, ..., h_q) \in \mathbb{R}^q_{>0}$ y

$$\hat{f}_h(u) = \frac{1}{nh} \sum_{j=1}^{n} K \left(\frac{\widehat{\boldsymbol{\theta}}^t \mathbf{t}_j - u}{h} \right) .$$

Las derivadas parciales involucradas en estos cálculos las determinamos analíticamente como sigue. Empecemos por las derivadas parciales de $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ que serán los estimadores de las derivadas parciales de $\eta_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$.

En el Paso 1 del procedimiento propuesto, hemos definido a $\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ por

$$\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - a}{\widehat{\sigma}}\right) w_1(\mathbf{x}_i) W^h_{\boldsymbol{\theta},i}(u)$$

con lo cual, verifica que

$$\sum_{i=1}^{n} \psi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right) w_1(\mathbf{x}_i) W^h_{\boldsymbol{\theta},i}(u) = 0\,.$$

Derivando respecto de $\boldsymbol{\beta}$ esta última expresión, obtenemos que

$$\sum_{i=1}^{n} \chi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)\left(\mathbf{x}_i + \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)\right) w_1(\mathbf{x}_i) W^h_{\boldsymbol{\theta},i}(u) = 0\,.$$

lo que nos permite deducir que

$$\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = -\frac{\displaystyle\sum_{i=1}^{n} \chi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right) \mathbf{x}_i w_1(\mathbf{x}_i) W^h_{\boldsymbol{\theta},i}(u)}{\displaystyle\sum_{i=1}^{n} \chi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right) w_1(\mathbf{x}_i) W^h_{\boldsymbol{\theta},i}(u)}\,.$$

Derivando ahora respecto de $\boldsymbol{\theta}$, llegamos a que $\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)$ cumple

$$\sum_{i=1}^{n}\left[-\frac{1}{\widehat{\sigma}}\chi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) W^h_{\boldsymbol{\theta},i}(u) + \psi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)\frac{\partial}{\partial\boldsymbol{\theta}} W^h_{\boldsymbol{\theta},i}(u)\right] w_1(\mathbf{x}_i) = 0$$

y entonces

$$\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u) = \frac{\widehat{\sigma}\displaystyle\sum_{i=1}^{n} \psi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)\frac{\partial}{\partial\boldsymbol{\theta}} W^h_{\boldsymbol{\theta},i}(u) w_1(\mathbf{x}_i)}{\displaystyle\sum_{i=1}^{n} \chi_1\left(\frac{y_i - \boldsymbol{\beta}^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right) W^h_{\boldsymbol{\theta},i}(u) w_1(\mathbf{x}_i)}$$

donde

$$\frac{\partial}{\partial\boldsymbol{\theta}} W^h_{\boldsymbol{\theta},i}(u) = \frac{K'\left(\dfrac{\boldsymbol{\theta}^t\mathbf{t}_i - u}{h}\right)\dfrac{\mathbf{t}_i}{h}\displaystyle\sum_{l=1}^{n} K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_l - u}{h}\right) - K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_i - u}{h}\right)\sum_{l=1}^{n} K'\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_l - u}{h}\right)\frac{\mathbf{t}_l}{h}}{\left(\displaystyle\sum_{l=1}^{n} K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_l - u}{h}\right)\right)^2}$$

siendo K' la derivada de K.

Por último, derivamos respecto de u:

$$\sum_{i=1}^{n}\left[-\frac{1}{\widehat{\sigma}}\chi_1\left(\frac{y_i-\boldsymbol{\beta}^t\mathbf{x}_i-\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)W^h_{\boldsymbol{\theta},i}(u)+\psi_1\left(\frac{y_i-\boldsymbol{\beta}^t\mathbf{x}_i-\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)\frac{\partial}{\partial u}W^h_{\boldsymbol{\theta},i}(u)\right]w_1(\mathbf{x}_i)=0$$

y entonces

$$\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)=\frac{\widehat{\sigma}\sum_{i=1}^{n}\psi_1\left(\frac{y_i-\boldsymbol{\beta}^t\mathbf{x}_i-\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)\frac{\partial}{\partial u}W^h_{\boldsymbol{\theta},i}(u)w_1(\mathbf{x}_i)}{\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\boldsymbol{\beta}^t\mathbf{x}_i-\widehat{\eta}_{\boldsymbol{\beta},\boldsymbol{\theta}}(u)}{\widehat{\sigma}}\right)W^h_{\boldsymbol{\theta},i}(u)w_1(\mathbf{x}_i)}$$

donde

$$\frac{\partial}{\partial u}W^h_{\boldsymbol{\theta},i}(u)=\frac{K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_i-u}{h}\right)\sum_{l=1}^{n}K'\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_l-u}{h}\right)-K'\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_i-u}{h}\right)\sum_{l=1}^{n}K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_l-u}{h}\right)}{h\left(\sum_{l=1}^{n}K\left(\frac{\boldsymbol{\theta}^t\mathbf{t}_l-u}{h}\right)\right)^2}.$$

Siguiendo argumentos similares a los del Lema 4.1.2, se garantiza que fijado $(\boldsymbol{\beta},\boldsymbol{\theta},\sigma)$, $\widehat{\mathbf{M}}\xrightarrow{P}\mathbf{M}(\boldsymbol{\beta},\boldsymbol{\theta})$.

7.2.2. Estimador de $\boldsymbol{\Delta}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)$ usando el valor de $\boldsymbol{\beta}_0$ bajo H_0

Utilizaremos la hipótesis nula, por lo que reemplazaremos $\boldsymbol{\beta}_0$ por $\boldsymbol{\beta}_*$; y estimaremos a $\boldsymbol{\theta}_0$ por $\widehat{\boldsymbol{\theta}}$, a η_0 por $\widehat{\eta}_{\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}}}$, a σ_0 por $\widehat{\sigma}_F$ y a $\mathbf{M}$ y a $\boldsymbol{\Sigma}$ por sus versiones empíricas. Es decir que

$$\widehat{\boldsymbol{\Delta}}=\widehat{\mathbf{M}}_1^{-1}(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}})\widehat{\mathbf{M}}_1^{-1}(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}}),$$

llamando a $\widehat{\mathbf{M}}=\widehat{\mathbf{M}}(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}})$ y a $\widehat{\boldsymbol{\Sigma}}=\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}})$. Tendremos que

$$\widehat{\mathbf{M}}=\frac{1}{\widehat{\sigma}_F^2}\frac{1}{n}\sum_{i=1}^{n}\chi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\widehat{\eta}_{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}}\left(\widehat{\boldsymbol{\theta}}^t\mathbf{t}_i\right)}{\widehat{\sigma}_F}\right)w_2(\mathbf{x}_i)\widehat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_i)\widehat{\boldsymbol{\lambda}}_i^t(\hat{\boldsymbol{\tau}}_i) \tag{7.8}$$

donde $\hat{\boldsymbol{\tau}}_i=\left(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}^t\mathbf{t}_i\right)$, y

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}})=\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{w}}_i(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}})\widehat{\mathbf{w}}_i^t(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}}), \tag{7.9}$$

siendo

$$\widehat{\mathbf{w}}_i(\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}})=\frac{1}{\widehat{\sigma}_F}\psi_1\left(\frac{y_i-\boldsymbol{\beta}_0^t\mathbf{x}_i-\widehat{\eta}_{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}}\left(\widehat{\boldsymbol{\theta}}^t\mathbf{t}_i\right)}{\widehat{\sigma}_F}\right)w_2(\mathbf{x}_i)\widehat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_i)-\frac{1}{\widehat{\sigma}_F^2}\widehat{E}_i\left(\widehat{\boldsymbol{\theta}}^t\mathbf{t}_i\right)\widehat{\boldsymbol{\gamma}}(\mathbf{t}_i)\hat{f}_h\left(\widehat{\boldsymbol{\theta}}^t\mathbf{t}_i\right),$$

$$\widehat{E}_i(u) = \frac{\widehat{\sigma}_F}{\hat{A}^1(u)}\psi_1\left(\frac{y_i - \boldsymbol{\beta}_*^t\mathbf{x}_i - \widehat{\eta}_{\boldsymbol{\beta}_*,\widehat{\boldsymbol{\theta}}}(u)}{\widehat{\sigma}_F}\right) w_1(\mathbf{x}_i)\,,$$

$$\hat{A}^1(\widehat{\boldsymbol{\theta}}, u) = \sum_{j=1}^{n}\chi_1\left(\frac{y_j - \boldsymbol{\beta}_0^t\mathbf{x}_j - \widehat{\eta}_{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}}\left(\widehat{\boldsymbol{\theta}}^t\mathbf{t}_j\right)}{\widehat{\sigma}_F}\right) w_1(\mathbf{x}_j)W^h_{\widehat{\boldsymbol{\theta}},j}(u)\,,$$

$$\widehat{\boldsymbol{\gamma}}(\tilde{\mathbf{t}}) = \sum_{j=1}^{n}\frac{1}{\hat{f}_h\left(\widehat{\boldsymbol{\theta}}^t\mathbf{t}_j\right)}\chi_1\left(\frac{y_j - \boldsymbol{\beta}_0^t\mathbf{x}_j - \widehat{\eta}_{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\theta}}}\left(\widehat{\boldsymbol{\theta}}^t\mathbf{t}_j\right)}{\widehat{\sigma}_F}\right) w_2(\mathbf{x}_j)\widehat{\boldsymbol{\lambda}}_j(\hat{\boldsymbol{\tau}}_j)\frac{\prod_{l=1}^{q}K_l\left(\frac{t_{jl}-\tilde{t}_l}{h_l}\right)}{\sum_{i=1}^{n}\prod_{l=1}^{q}K_l\left(\frac{t_{il}-\tilde{t}_l}{h_l}\right)}\,.$$

Siguiendo argumentos similares a los del Lema 4.1.2, se garantiza que fijado $(\boldsymbol{\beta}_*, \boldsymbol{\theta}, \sigma)$ siendo $\boldsymbol{\beta}_*$ el impuesto en la hipótesis nula, $\widehat{\mathbf{M}} \xrightarrow{P} \mathbf{M}(\boldsymbol{\beta}_*, \boldsymbol{\theta})$.

7.3. Estudio de Simulación: Nivel Empírico

Para evaluar el comportamiento de la propuesta, introducida en este capítulo, de pruebas de hipótesis, realizamos simulaciones para cuantificar niveles de significación y compararlos con los obtenidos en versiones clásicas. Repitiendo el mismo experimento numérico del capítulo anterior, se generaron $N = 1000$ muestras aleatorias de tamaño $n = 100$ que siguen el modelo (5.1). Las hipótesis a contrastar son

$$H_0 : \boldsymbol{\beta} = \sqrt{5} \quad \text{contra} \quad H_1 : \boldsymbol{\beta} \neq \sqrt{5}$$

cuyo estadístico del test es un estadístico de Wald, alguna de las dos versiones presentadas en la sección anterior, con nivel de significación asintótico $\alpha = 0{,}05$. Las primeras simulaciones de las versiones clásicas de las pruebas de hipótesis mostraron que era necesario sobresuavizar para estimar inicialmente a σ_0, como lo hacen González-Manteiga y Aneiros-Pérez (2003) y Aneiros-Pérez et al. (2004). La ventana piloto que se utilizó para hacer ese sobresuavizado fue $h = 0,3$. El resto de las ventanas utilizadas son las mismas que en el estudio de simulación del capítulo anterior. Las siguientes tablas muestran los niveles de significación empíricos que se obtuvieron en los diferentes escenarios.

Se observa entonces el buen desempeño de los tests de hipótesis propuestos en los escenarios de contaminación de tipo 1 y 2 donde los niveles de significación empíricos son cercanos a α. Aunque esta característica no muestra supremacía sobre los resultados de las pruebas de hipótesis en su versión tradicional, sirven para ratificar el uso de la propuesta presentada aquí. Notemos la evidente mejora que visibilizan las pruebas de hipótesis propuestas en el escenario de contaminación 3, frente al paupérrimo desempeño de su correspondiente versión clásica, cuyos niveles de significación empíricos son siempre 1. Sin duda, este desempeño está ligado

Contaminación	Versión clásica	Versión robusta
C_0	0,042	0,063
$C_{1;0,1;49}$	0,049	0,058
$C_{1;0,2;49}$	0,07	0,054
$C_{1;0,1;100}$	0,056	0,062
$C_{1;0,2;100}$	0,069	0,059
$C_{2;0,1}$	0,071	0,093
$C_{2;0,2}$	0,084	0,068
$C_{2;0,3}$	0,092	0,065
$C_{3;10;5}$	1	0,059
$C_{3;10;10}$	1	0,060
$C_{3;10;15}$	1	0,057
$C_{3;20;5}$	1	0,072
$C_{3;20;10}$	1	0,071
$C_{3;20;15}$	1	0,063

Cuadro 7.1: Niveles de significación empíricos, estadístico del test estimando a β.

Contaminación	Versión clásica	Versión robusta
C_0	0,042	0,049
$C_{1;0,1;49}$	0,049	0,049
$C_{1;0,2;49}$	0,069	0,043
$C_{1;0,1;100}$	0,057	0,046
$C_{1;0,2;100}$	0,068	0,048
$C_{2;0,1}$	0,071	0,072
$C_{2;0,2}$	0,083	0,057
$C_{2;0,3}$	0,09	0,053
$C_{3;10;5}$	1	0,047
$C_{3;10;10}$	1	0,046
$C_{3;10;15}$	1	0,046
$C_{3;20;5}$	1	0,072
$C_{3;20;10}$	1	0,056
$C_{3;20;15}$	1	0,050

Cuadro 7.2: Niveles de significación empíricos, estadístico del test bajo H_0.

a la extrema sensibilidad de los tests de hipótesis clásicos a los escenarios de simulación de la contaminación de tipo 3 propuestos, que sistemáticamente rechazan la hipótesis nula en cualquiera de ellos. No sorprende lo que aquí sucede, ya que está íntimamente relacionado a lo que sucedía en los boxplots de las estimaciones de β_0 de la tercera contaminación que se encuentra en el capítulo anterior.

Por último, cabe resaltar que los niveles empíricos calculados en la versión bajo H_0 son siempre menores o iguales a los de la versión estimando β_0. Con lo cual, basadas en este estudio de simulación de los niveles empíricos, es que aconsejamos usar la versión bajo H_0 que brindará una mejor aproximación del nivel de significación asintótico del test de hipótesis contemplado.

Bibliografía

[1] Agostinelli C., Bianco A. y Boente G. (2020) *Robust estimation in single-index models when the errors have a unimodal density with unknown nuisance parameter*. Annals of the Institute of Statistical Mathamatics, 72: 855-893.

[2] Aneiros-Pérez G., González-Manteiga W. y Vieu P. (2004) *Estimation and testing in a partial linear regression model under long-memory dependence*. Bernoulli, 10(1): 49-78.

[3] Bianco A. y Boente G. (2002). *On the asymptotic behavior of one–step estimates in heteroscedastic regression models*. Statistics & Probability Letters, 60: 33-47.

[4] Bianco A., Boente G. y Martinez E. (2006) *Robust tests in semiparametric partly linear models*. Scandinavian Journal of Statistics, 33: 435-450.

[5] Bianco A. y Martínez E. (2009) *Robust testing in the logistic regression model*. Computational Statistics and Data Analysis, 53: 4095 - 4105

[6] Boente G. y Fraiman R. (1989) *Robust nonparametric regression estimation*. Journal of Multivariate Analysis, 29: 180-198.

[7] Boente G. y Fraiman R. (1991) *Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes*. International Statistical Review, 59(3): 355-372.

[8] Boente G., He X. y Zhou J. (2006) *Robust estimates in generalized partially linear models*. The Annals of Statistics, 34(6): 2856-2878.

[9] Boente G. y Martinez A. (2017) *Marginal integration M-estimators for additive models*. Test, 26: 231-260.

[10] Boente G. y Rodriguez D. (2010). *Robust inference in generalized partially linear models*. Computational Statistics and Data Analysis, 54: 2942-2966.

[11] Boente G. y Rodriguez D. (2012). *Robust estimators in generalized partially single index models*. Test, 21: 386-411.

[12] Cantoni E. y Ronchetti E. (2001) *Robust inference for generalized linear models*. Journal of the American Statistical Association, 96: 1022-1030.

[13] Carroll R. J., Fan J., Gijbels I. y Wand M. P. (1997) *Generalized partially linear single-index models*. Journal of the American Statistical Association, 92: 477-489.

[14] Delecroix M., Hristache M. y Patilea V. (2006) *On semiparametric M-estimation in single-index regression*. Journal of Statistical Planning and Inference, 136: 730-769.

[15] Fan J. y Gijbels I. (1996) *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

[16] González-Manteiga W. y Aneiros-Pérez G. (2003). *Testing in partial linear regression models with dependent errors.* Nonparametric Statistics, 15(1): 93-111.

[17] Härdle W. (1984) *Robust regression function estimation*. Journal of Multivariate Analysis, 14: 169-180.

[18] Härdle W., Liang H. y Gao J. (2000) *Partially linear models*.

[19] Heritier S. y Ronchetti E. (1994) *Robust Bounded-Influence Tests in General Parametric Models*. Journal of the American Statistical Association, Vol. 89, No. 427. 897-904.

[20] Huber P. J. (1973) *Robust Regression: Asymptotics, Conjectures and Monte Carlo*. Annals of Statistics, 1: 799-821.

[21] Huber P. (1981) *Robust Statistics*. John Wiley & Sons, Inc.

[22] Ichimura H. (1993) *Semiparametric least squares (SLS) and weighted SLS estimation of single-index models*. Journal of Econometrics, 58: 71-120.

[23] Liang H., Liu X., Li R. y Tsai C. L. (2010) *Estimation and testing for partially linear single-index models*. The Annals of Statistics, 38(6): 3811-3836.

[24] Liu J., Zhang R., Zhao W. y Lv Y. (2013) *A robust and efficient estimation method for single index models*. Journal of Multivariate Analysis, 122: 226-238.

[25] Ma Y. y Zhu L. (2013) *Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates*. Journal of the Royal Statistical Society: Series B, 75(2): 305-322.

[26] Mallows C. L. (1975) *On Some Topics in Robustness*. Murray Hill, NJ: Bell Telephone Laboratories. Memorandum sin publicar.

[27] Maronna R. A., Martin R. D., Salibián-Barrera M. y Yohai V. J. (2019) *Robust statistics: theory and methods (with R)*. Second edition - John Wiley & Sons, Ltd.

[28] Nadaraya E. A. (1964) *On estimating regression*. Theory of Probability and Application, 9: 141-142.

[29] Pollard D. (1984) *Convergence of stochastic processes*. Springer-Verlag New York Inc.

[30] Rodriguez D. (2008) *Estimación robusta en modelos parcialmente lineales generalizados*. Tesis doctoral, Universidad de Buenos Aires. Disponible en http://cms.dm.uba.ar/academico/carreras/doctorado/tesisdanielarodriguez.pdf

[31] Rousseeuw P.J. y Yohai V.J. (1984) *Robust Regression by means of S-estimators*. Robust and Nonlinear Time Series, vol. 26, 256-272. Franke, J., Hardle, W. & Martin, R.D. editors. Lecture Notes in Statistics, Springer, New York.

[32] Sakata S. y White H. (2001) *S-estimation of nonlinear regression models with dependent and heterogeneous observations*. Journal of Econometrics, 103: 5-72.

[33] Schweppe F. C., Wildes J. y Rom D. B. (1970) *Power system static state estimation. Parts I, II, III*. IEEE Trans, on PAS, PAS-89 , 120-135.

[34] Severini T. y Staniswalis J. (1994) *Quasi-likelihood estimation in semiparametric models*. Journal of the American Statistical Association, 89: 501?511.

[35] Severini T. A. y Wong W. H. (1992) *Profile likelihood and conditionally parametric models*. The Annals of Statistics, 20(4): 1768-1802.

[36] Sun Y. y Genton M. G. (2011) *Functional boxplots*. Journal of Computational and Graphical Statistics, 20: 316-334.

[37] Van der Vaart A. W. y Wellner J. A. (1996) *Weak convergence and empirical processes*. Springer-Verlag New York, Inc.

[38] Watson G. S. (1964) *Smooth regression analysis*. Sankhyā Series A, 26: 359-392.

[39] Xia Y. y Härdle W. (2006) *Semi-parametric estimation of partially linear single-index models*. Journal of Multivariate Analysis, 97: 1162-1184.

[40] Xia Y., Tong H. y Li W.K. (1999) *On extended partially linear single-index models*. Biometrika, 86(4): 831?842

[41] Yang H. y Yang J. (2014) *A robust and efficient estimation and variable selection method for partially linear single-index models*. Journal of Multivariate Analysis, 129: 227-242.

[42] Yohai V. J. (1985) *High breakdown point and high efficiency robust estimates for regression*. Technical report 66, Dept. Statistics, Univ. Washington. Disponible en http://www.stat.washington.edu/research/reports/1985/tr066.pdf.

[43] Yohai V.J. (1987) *High Breakdown-Point and High Efficiency Robust Estimates for Regression*. The Annals of Statistics, vol. 15, 642-656.

[44] Yohai V. J. y Zamar R. H. *High breakdown-point estimates of regression by means of the minimization of an efficient scale*. Journal of the American Statistical Association. 83(402): 406-413

[45] Yu Y. y Ruppert D. (2002) *Penalized spline estimation for partially linear single-index models*. Journal of the American Statistical Association, 97: 1042-1054.