



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

**Estimación robusta en modelos semiparamétricos bajo restricciones de
monotonía**

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

Pablo Vena

Directores de tesis: Dra. Graciela Boente, Dr. Matías Salibián-Barrera.

Buenos Aires, 30 de Septiembre de 2019.

Estimación robusta en modelos semiparamétricos bajo restricciones de monotonía

Resumen

Los modelos semiparamétricos se introducen cuando el modelo lineal resulta insuficiente para explicar la relación entre la variable de respuesta y sus covariables asociadas. Por un lado, este enfoque ha sido usado para extender los modelos lineales generalizados. Por otro lado, frente al creciente interés y desarrollo de técnicas basadas en datos funcionales, los modelos semiparamétricos han extendido el modelo parcialmente lineal al contexto funcional combinando componentes paramétricas y no paramétricas. En ambos casos, más allá de la suavidad, no son usuales los supuestos sobre la componente no paramétrica como podría ser la monotonía. Asimismo, las propuestas estudiadas previamente son sensibles a datos atípicos. Esta tesis se divide en dos partes.

En la primera, consideramos observaciones que siguen un modelo parcialmente lineal generalizado isotónico donde la media de la variable de respuesta se modela, a través de una función de vínculo, linealmente en algunas covariables y de forma no paramétrica en una función, que suponemos monótona, de un regresor univariado. Definimos una familia de estimadores robustos para la componente no paramétrica y para el parámetro de regresión relacionado con el término lineal. Los estimadores robustos se basan en una aproximación por splines combinada con una función de pérdida que controla valores grandes de la *deviance*. Bajo condiciones de regularidad, obtenemos resultados de consistencia para el estimador de la componente no paramétrica así como resultados de consistencia y distribución asintótica para los estimadores del parámetro de regresión. A través de un estudio de Monte Carlo investigamos el desempeño de los estimadores propuestos bajo un modelo de regresión parcialmente lineal log-Gamma con una componente no paramétrica creciente. La propuesta se ilustra en un conjunto de datos reales.

En la segunda parte, consideramos estimadores equivariantes para los parámetros de un modelo de regresión lineal semifuncional. Bajo este modelo, la respuesta real es modelada linealmente con una variable explicativa funcional y no paraméricamente en una covariable univariada. Se definen MM-estimadores aproximando tanto el parámetro de regresión como la componente no paramétrica por medio de B-splines. Consideramos también la situación en que la componente no paramétrica es monótona a través de una modificación de los estimadores obtenidos sin restricciones de monotonía. Un estudio numérico muestra las ventajas de la metodología propuesta para muestras finitas bajo diferentes esquemas de contaminación. Asimismo, el análisis de un conjunto de datos reales permite mostrar la estabilidad del método propuesto y su capacidad para permitir detectar de datos atípicos.

Palabras Clave: B-splines, Deviance, Regresión isotónica, Modelos parcialmente lineales, Modelos Lineales Funcionales, Estimación Robusta

Robust estimators in semiparametric models under monotony constraints

Abstract

Semiparametric models are introduced when the linear model is insufficient to explain the relationship between the response variable and its associated covariates. On one hand, this approach has been used to extend generalized linear models. On the other hand, facing the growing interest and development of techniques based on functional data, semiparametric models have been extended to the functional setting combining parametric and non parametric components through different partly linear models. In both cases, beyond smoothness, conditions on the non parametric component such as monotonicity are not usual. Also, recent studied proposals are sensitive to atypical observations.

This thesis is divided in two parts. In the first one, we consider observations following an isotonic generalized partly linear model where the mean of the responses is modelled, through a link function, linearly on some covariates and nonparametrically on an univariate regressor in such a way that the nonparametric component is assumed to be a monotone function. A class of robust estimates for the monotone nonparametric component and for the regression parameter, related to the linear one, is defined. The robust estimators are based on a spline approach combined with a loss function which controls large values of the deviance. Under regularity conditions, we derive consistency results for the nonparametric function estimators as well as consistency and asymptotic distribution results for the regression parameter estimators. Through a Monte Carlo study, we investigate the performance of the proposed estimators under a partly linear log-Gamma regression model with increasing nonparametric component. The proposal is illustrated on a real data set.

In the second part, we consider robust equivariant estimators for the parameters of a semi-functional linear regression model. Under this model, a real-valued response is modelled linearly on a functional explanatory variable and nonparametrically on an univariate covariate which is assumed to be increasing, for instance. The MM-estimators are obtained by approximating both the regression parameter and the non-parametric component using monotone B-splines. The situation in which the nonparametric component is monotone is also considered. Numerical experiments show the advantage of the proposed methodology for finite samples under different contamination schemes. We also illustrate the stability of proposed method and its ability to detect outliers on a real data set.

Keywords: B-splines, Deviance, Isotonic Regression, Partly Linear Models, Functional Linear Models, Robust Estimation

Índice general

I	Modelo de regresión parcialmente lineal generalizado	1
1.	Introducción	3
2.	Estimadores robustos para modelos parcialmente lineales generalizados isotónicos	7
2.1.	Introducción	7
2.2.	Estimadores basados en B-splines monótonos	8
2.3.	Estimadores monotonizados basados en B-splines	9
2.4.	La función de pérdida	11
2.5.	El caso particular del modelo de regresión log-Gamma	13
2.5.1.	Algoritmo para el caso log-Gamma	14
2.6.	Sobre la elección de la cantidad de splines	17
2.7.	Apéndice: Sobre la Fisher-consistencia	17
2.7.1.	Modelo parcialmente lineal con errores simétricos	17
2.7.2.	Modelo parcialmente lineal con errores asimétricos	19
2.7.3.	Modelo parcialmente lineal logístico	20
3.	Consistencia de los estimadores	23
3.1.	Nociones previas: Procesos empíricos	23
3.2.	Resultados de consistencia	27
3.3.	Tasas de convergencia	28
3.4.	Apéndice A: Demostración de los Teoremas 3.9 y 3.11	31
3.4.1.	Demostración del Teorema 3.9	31

3.4.2.	Demostración del Teorema 3.11	32
3.5.	Apéndice B: Algunos comentarios sobre los supuestos	35
3.5.1.	Sobre C0 a C3	35
3.5.2.	Sobre las condiciones de entropía	36
3.5.3.	Condiciones que garantizan C9	45
4.	Normalidad asintótica del estimador del parámetro β_0	47
4.1.	Distribución asintótica de $\hat{\beta}$ y $\hat{\beta}_{INI}$	47
4.2.	Estimación de la matriz de covarianza asintótica	49
4.3.	Apéndice: Demostración del Teorema 4.1	51
5.	Función de influencia empírica	63
5.1.	Función de influencia de los estimadores definidos en la Sección 2.2	63
5.2.	La función de influencia empírica bajo un modelo de regresión parcialmente lineal log-Gamma	65
5.3.	Algunas observaciones sobre el estimador monotonizado definido en la Sección 2.3	73
5.4.	Apéndice: Demostración de (5.6) y (5.7)	74
6.	Estudio numérico y análisis de un conjunto de datos reales	77
6.1.	Estudio de Monte Carlo	77
6.2.	Aplicación a un conjunto de datos reales: Costos Hospitalarios	92
7.	Consideraciones finales	95
II	Modelo de regresión parcialmente lineal semi-funcional	97
8.	Introducción	99
9.	Estimadores robustos para modelos parcialmente lineales semifuncionales	105
9.1.	Introducción	105
9.2.	Estimadores basados en B-splines	106

9.2.1. Estimadores basados en B-splines monótonos	108
9.2.2. Estimadores monotonizados basados en B-splines	108
9.3. Estimadores basados en componentes principales	110
9.4. Selección del modelo	113
9.5. Algunos comentarios adicionales	114
10. Consistencia de los estimadores basados en B-splines	115
10.1. Consistencia	116
10.2. Tasas de convergencia	119
10.3. Apéndice A: Demostraciones	120
10.3.1. Demostración del Lema 10.1	120
10.3.2. Demostración del Lema 10.2	121
10.3.3. Demostración de la Proposición 10.3	122
10.3.4. Demostración del Teorema 10.4	123
10.3.5. Demostración del Teorema 10.6	126
10.3.6. Demostración del Teorema 10.7	128
10.3.7. Demostración del Lema 10.8	133
11. Estudio numérico y análisis de un conjunto de datos reales	135
11.1. Estudio de Monte Carlo	135
11.1.1. Esquemas de simulación	136
11.1.2. Sobre los criterios de selección de las dimensiones de las bases	139
11.1.3. Resultados para muestras contaminadas	144
11.2. Aplicación a un conjunto de datos reales: TECATOR	146
11.3. Apéndice: Tablas sin poda.	150
12. Consideraciones finales	153

Parte I

Modelo de regresión parcialmente lineal generalizado

Capítulo 1

Introducción

Los modelos semiparamétricos suelen utilizarse cuando el modelo lineal resulta insuficiente para explicar la relación entre la variable de respuesta y sus covariables asociadas. Este enfoque ha sido considerado también para extender los modelos lineales generalizados a contextos más generales para permitir que la mayoría de los predictores sea modelada linealmente mientras un número pequeño de ellos sea modelado de manera no paramétrica.

A lo largo de la primera parte de la tesis, trabajaremos con observaciones $(y_i, \mathbf{x}_i^T, t_i)^T$ que satisfacen un modelo semiparamétrico parcialmente lineal generalizado que abreviaremos GPLM. Bajo este modelo, suponemos que $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i, \kappa_0)$ con $\text{VAR}(y_i | (\mathbf{x}_i, t_i)) = A^2(\kappa_0) V^2(\mu_i)$, donde A y V son funciones conocidas y $\mu_i = \mathbb{E}(y_i | (\mathbf{x}_i, t_i)) = \mu(\mathbf{x}_i, t_i)$ es tal que

$$\mu(\mathbf{x}, t) = H(\mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t)) , \quad (1.1)$$

siendo H^{-1} una función de vínculo conocida, $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ un parámetro desconocido y η_0 una función continua desconocida con soporte en un intervalo compacto \mathcal{I} , que supondremos, sin pérdida de generalidad, igual a $[0, 1]$. El parámetro κ_0 indica usualmente un parámetro de ruido, o *nuisance* por su nombre en inglés, y pertenece a un subconjunto de \mathbb{R} . Por esa razón, supondremos que $\kappa_0 \in \mathcal{K}$, donde $\mathcal{K} \subset \mathbb{R}$ es un conjunto abierto.

Cuando $H(u) = u$, el GPLM se reduce al modelo de regresión parcialmente lineal que ha sido ampliamente estudiado. En este caso, κ_0 suele corresponder al parámetro de escala (ver Härdle *et al.*, 2000). Estimadores robustos para el GPLM han sido propuestos, por ejemplo, por Boente *et al.* (2006) y por Boente y Rodríguez (2010), suponiendo que el parámetro κ_0 es conocido. Sin embargo, los trabajos anteriores no imponen ninguna restricción sobre la función η_0 . En la primera parte de esta tesis, consideraremos la situación en la cual hay restricciones en la componente no paramétrica η_0 . Más precisamente, supondremos que en el modelo (1.1) la función η_0 es monótona y, sin pérdida de generalidad, no decreciente. La mayoría de los trabajos sobre modelos parcialmente lineales generalizados suponen que η_0 es una función suave desconocida, aunque en muchas aplicaciones la monotonía es una propiedad natural en la función a estimar. Cuando $\boldsymbol{\beta}_0 = \mathbf{0}$, algunos ejemplos en los cuales la restricción de monotonía aparece naturalmente, pueden encontrarse en Ramsay (1988) quien estudió, entre otros ejemplos, la relación entre la incidencia del síndrome de Down y la edad de la madre, que se encuentra representada en la Figura 1.1(a). En dicha Figura, se observa que cuanto mayor es la madre más alta es la incidencia de esta condición. Otros ejemplos pueden verse

en He y Shi (1998). En la Sección 6.2 presentaremos un conjunto de datos analizado previamente en Marazzi y Yohai (2004) mediante un modelo lineal generalizado. El objetivo de dichos autores es estudiar la relación entre el costo hospitalario y ciertas variables explicativas entre las que se encuentra la duración de la estadía del paciente en días. En la Sección 6.2, modelaremos la relación entre el costo y la duración de la estadía de forma no paramétrica mientras que las demás variables entrarán al modelo en forma lineal. El supuesto de monotonía sobre η_0 es natural para este conjunto de datos puesto que el costo hospitalario aumenta para estadías más largas, como se observa en la Figura 1.1(b).

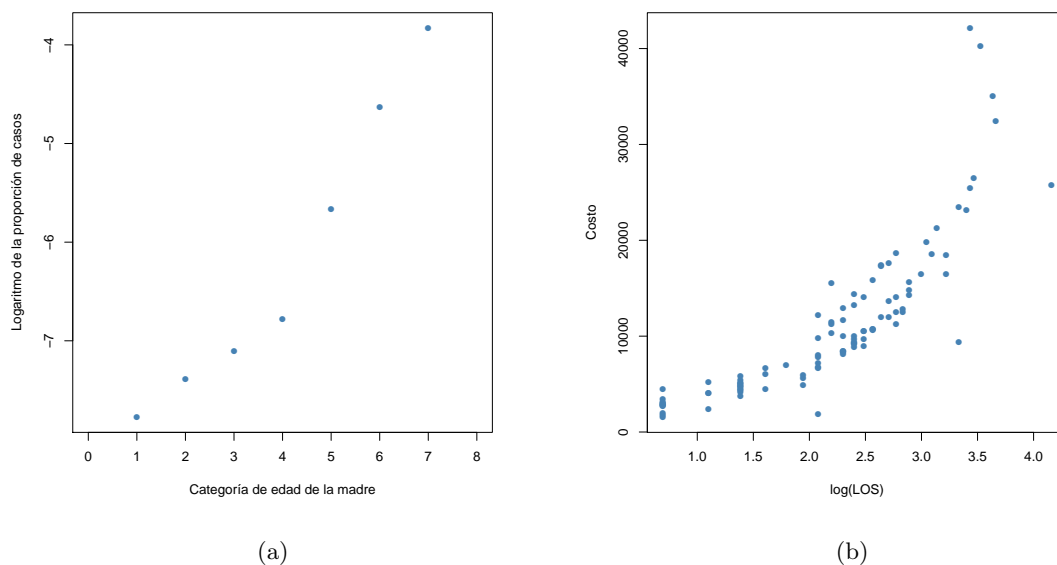


Figura 1.1: (a) Relación entre el logaritmo de la proporción de nacimientos con Síndrome de Down y la edad de la madre, Ramsay (1988). (b) Costo hospitalario en función del logaritmo de la longitud de la estadía, Marazzi y Yohai (2004).

La mayor parte de los desarrollos en estimación bajo restricciones de monotonía fueron hechos para modelos de regresión no paramétrica o para modelos de regresión parcialmente lineales. En este contexto, dos enfoques son posibles. Uno de ellos solo supone que la función η_0 es monótona, dando por estimador de η_0 una función no decreciente constante por intervalos, mientras que el otro impone condiciones de suavidad y provee estimaciones basadas en núcleos o en splines. Entre los trabajos que consideran el primer enfoque podemos mencionar, entre otros, a Huang (2002) y Sun *et al.* (2012) quienes consideraron la estimación bajo restricciones en un modelo parcialmente lineal isotónico. Siguiendo el segundo punto de vista que impone condiciones de suavidad a la función η_0 , Lu (2010) propuso un estimador de mínimos cuadrados basado en B-splines. Recientemente, Lu (2015) consideró un enfoque basado en splines para generalizar la propuesta anterior a modelos parcialmente lineales generalizados con restricciones de monotonía. Todos estos métodos son sensibles a datos atípicos. Cuando $\beta_0 = \mathbf{0}$, η_0 es monótona y $H(u) = u$, es decir, bajo el modelo de regresión no paramétrico isotónico, se han dado algunas propuestas de estimadores resistentes a datos atípicos. Para el caso del modelo de regresión isotónico sin condiciones de suavidad, podemos mencionar entre otros a Wang y Huang (2002) quienes han propuesto un procedimiento de estimación robusto basado en la regresión de la mediana minimizando la distancia L_1 entre el vector de valores observados y el

de valores ajustados. Para mejorar el estimador L_1 manteniendo buenas propiedades de robustez, Álvarez y Yohai (2012) consideraron M-estimadores. Por otro lado, cuando se imponen condiciones de suavidad en la función η_0 del modelo de regresión parcialmente lineal isotónico, He y Shi (1998) proponen un método basado en B-splines monótonos y en la distancia L_1 , mientras que Du *et al.* (2013) consideraron M-estimadores basados en B-splines monótonos cuando el parámetro de escala es conocido y los errores tienen una distribución simétrica. Sin embargo, para el conjunto de datos de costos hospitalarios que consideraremos en la Sección 6.2, es razonable modelar los errores a través de una distribución asimétrica como la distribución log-Gamma. Por esta razón, la propuesta considerada en Du *et al.* (2013) no es apropiada. Más aun, para estos datos, el parámetro de forma es desconocido y debe ser estimado para poder calibrar los estimadores robustos y atenuar el peso de residuos grandes.

En esta primera parte, daremos un marco general para proponer una familia de estimadores para el parámetro de regresión β_0 y la función de regresión monótona η_0 bajo el modelo GPLM (1.1) cuando el parámetro de ruido es desconocido y la función η_0 es una función monótona suave. Este modelo incluye al modelo de regresión parcialmente lineal con escala desconocida y errores simétricos y además al modelo de regresión parcialmente lineal isotónico log-Gamma con parámetro de forma desconocido. En este sentido, los estimadores a considerar generalizan la propuesta dada en Du *et al.* (2013) ya que consideran un estimador preliminar de escala para calibrar el tamaño de los residuos y extienden la propuesta de esos autores al caso de modelos con errores asimétricos.

La primera parte se organiza como sigue. El Capítulo 2 describe los estimadores robustos propuestos y que son nuestro objeto de estudio. En particular, como nuestro enfoque está basado en B-splines, describimos un método robusto para elegir la cantidad de nodos en la Sección 2.6. Por otra parte, el caso particular del modelo log-Gamma se describe en la Sección 2.5. Resultados de consistencia y de tasa de convergencia para los estimadores propuestos se detallan en el Capítulo 3, mientras que la distribución asintótica del parámetro de regresión se obtiene en el Capítulo 4. En la Sección 3.5, se estudian condiciones que garantizan la validez de los supuestos necesarios para obtener resultados de consistencia y de tasas para algunos modelos particulares. En el Capítulo 5, calculamos la función de influencia empírica que permite estudiar la sensibilidad local frente a datos atípicos. En el Capítulo 6, se presentan los resultados de un estudio de simulación y el análisis de un conjunto de datos reales. Más precisamente, en la Sección 6.1, reportamos un estudio numérico para examinar las propiedades en muestras finitas de los procedimientos propuestos tanto para muestras generadas bajo el modelo log-Gamma como para muestras contaminadas. El análisis de los datos de costo hospitalario se presenta en la Sección 6.2. Finalmente, en el Capítulo 7 damos algunos comentarios finales y conclusiones sobre esta primera parte. Para facilitar la lectura, las demostraciones fueron relegadas a los apéndices de cada capítulo.

Capítulo 2

Estimadores robustos para modelos parcialmente lineales generalizados isotónicos

2.1. Introducción

Como mencionamos en el Capítulo 1, el objetivo de esta tesis es proponer estimadores robustos para el modelo parcialmente lineal generalizado isotónico (1.1) cuando el parámetro de ruido es desconocido y la función η_0 es una función monótona suave. Para ello es necesario elegir qué método de suavizado utilizaremos y qué técnica para obtener estimadores monótonos aplicaremos. Respecto de lo primero, las opciones usuales incluyen estimadores de núcleos como los considerados, por ejemplo, en Härdle y Tsybakov (1988) y Boente y Fraiman (1988) para modelos de regresión no paramétricos, por Bianco y Boente (2004) en modelos parcialmente lineales y por Boente *et al.* (2006) y Boente y Rodríguez (2010) en modelos parcialmente lineales generalizados. Todos estos estimadores no imponen restricciones de monotonía en la función η_0 . Otra familia de estimadores robustos en modelos de regresión no paramétrica es la basada en splines. El primer enfoque para obtener estimadores resistentes a datos atípicos consistió en reemplazar la pérdida cuadrática por la pérdida L_1 . El ajuste resultante aproxima la mediana condicional, en lugar de la media condicional. Como es bien sabido, si la distribución de los errores es simétrica y existe primer momento, la esperanza condicional coincide con la mediana de la distribución condicional. Esta idea fue considerada por He y Shi (1994) quienes, más generalmente, definieron regresión por cuantiles basada en splines.

Como mencionamos, estos estimadores no imponen restricciones y una manera de abordar el problema de estimar una función monótona es la de imponer restricciones en el proceso de estimación. Bajo un modelo parcialmente lineal, este enfoque fue considerado por He y Shi (1998) que combinaron B-splines monótonos con distancia L_1 para estimar la función η_0 y por Du *et al.* (2013) que utilizaron M-estimadores basados en B-splines monótonos cuando la escala es conocida. En este capítulo, presentaremos una generalización de estos estimadores para incluir estimadores en modelos de regresión parcialmente lineales isotónicos con errores asimétricos y, más generalmente,

los modelos de regresión parcialmente lineales generalizados (1.1).

Otro enfoque posible al problema de estimar una función monótona consiste en estimar primero el parámetro β_0 y la función η_0 sin restricciones, como fue hecho en Boente *et al.* (2006) y Boente y Rodríguez (2010), para luego aplicar el proceso de monotonización considerado en Dette *et al.* (2006) y Neumeyer (2007). En lugar de utilizar estimadores basados en núcleos como en las propuestas de Boente *et al.* (2006) y Boente y Rodríguez (2010), consideraremos estimadores robustos basados en B-splines que no han sido estudiados en el contexto de modelos parcialmente lineales generalizados sin restricciones y transformaremos los estimadores de η_0 para obtener un estimador monótono.

2.2. Estimadores basados en B-splines monótonos

Sean $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$ una función de pérdida y $w : \mathbb{R}^p \rightarrow \mathbb{R}$ una función de pesos para controlar las covariables \mathbf{x} de alta palanca. Definamos las funciones

$$L_n(\beta, g, a) = \frac{1}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^T \beta + g(t_i), a) w(\mathbf{x}_i) \quad \text{y} \quad L(\beta, g, a) = \mathbb{E} \rho(y, \mathbf{x}^T \beta + g(t), a) w(\mathbf{x}), \quad (2.1)$$

donde $(y, \mathbf{x}^T, t)^T$ es un vector aleatorio con la misma distribución que $(y_i, \mathbf{x}_i^T, t_i)^T$. Claramente, una opción es tomar $w \equiv 1$ que corresponde a no controlar las covariables de alta palanca y atenuar solamente el peso de residuos grandes. Esta elección es la que se ha tomado, por ejemplo, en el modelo lineal al considerar los MM-estimadores. Sin embargo, es de esperar que al introducir pesos en las covariables se obtenga un beneficio al realizar inferencia sobre el parámetro β_0 , ya sea al computar regiones de confianza o tests de hipótesis de tipo Wald. Las ventajas de los estimadores pesados en el contexto de modelos lineales generalizados han sido descritas por Bianco *et al.* (2013b) donde se observa que los pesos también intervienen en el control del crecimiento de la matriz de covarianza asintótica estimada, necesaria para realizar inferencia.

Como en Lu (2010, 2015) y Du *et al.* (2013), consideremos $\mathcal{T}_n = \{\tau_i\}_{i=1}^{m_n+2\ell}$, donde

$$0 = \tau_1 = \dots = \tau_\ell < \tau_{\ell+1} < \dots < \tau_{m_n+\ell+1} = \dots = \tau_{m_n+2\ell} = 1$$

es una sucesión de nodos que particionan el intervalo cerrado $[0, 1]$ en $m_n + 1$ subintervalos $\mathcal{I}_i = [\tau_{\ell+i}, \tau_{\ell+i+1})$, para $i = 0, \dots, m_n - 1$ y $\mathcal{I}_{m_n} = [\tau_{m_n+\ell}, \tau_{m_n+\ell+1}]$. Más aun, llamemos $\mathcal{S}_n(\mathcal{T}_n, \ell)$ a la clase de splines de orden $\ell > 1$ con nodos en \mathcal{T}_n . Si η_0 es r veces derivable, tomaremos $\ell \geq r + 2$. De acuerdo al Corolario 4.10 de Schumaker (1981), para cualquier $g \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ existe una clase de B-splines $\{B_j : 1 \leq j \leq k_n\}$, con $k_n = m_n + \ell$, tal que $g = \sum_{j=1}^{k_n} \lambda_j B_j$. Por otra parte, según el Teorema 5.9 de Schumaker (1981), el spline g es monótono no decreciente en $[0, 1]$ si los coeficientes $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{k_n})^T$ son no decrecientes, es decir, cuando $\lambda_1 \leq \dots \leq \lambda_{k_n}$. Sin embargo, como se menciona en Lu (2015), la restricción de desigualdad sobre los coeficientes no es una condición necesaria aunque sí lo es para B-splines cuadráticos ($\ell = 3$).

Por lo tanto, podemos definir la subclase de splines monótonos no decrecientes en $[0, 1]$, $\mathcal{M}_n(\mathcal{T}_n, \ell)$, que es una subclase de $\mathcal{S}_n(\mathcal{T}_n, \ell)$, como

$$\mathcal{M}_n(\mathcal{T}_n, \ell) = \left\{ \sum_{i=j}^{k_n} \lambda_j B_j : \lambda_1 \leq \dots \leq \lambda_{k_n} \right\},$$

donde las restricciones de no decrecimiento se imponen en los coeficientes para garantizar la monotonía. De esta forma, la función η_0 puede ser aproximada como $\eta_0(t) \approx \boldsymbol{\lambda}^\top \mathbf{B}(t)$ con $\mathbf{B}(t) = (B_1(t), \dots, B_{k_n}(t))^\top$ el vector de la base de funciones B-splines y $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{k_n})^\top$ el vector de coeficientes tal que $\boldsymbol{\lambda}^\top \mathbf{B} \in \mathcal{M}_n(\mathcal{T}_n, \ell)$.

Esto sugiere que los estimadores de $(\boldsymbol{\beta}_0, \eta_0)$ pueden ser obtenidos minimizando $L_n(\boldsymbol{\beta}, g, \hat{\kappa})$ sobre $\boldsymbol{\beta} \in \mathbb{R}^p$ y $g \in \mathcal{M}_n(\mathcal{T}_n, \ell)$, donde $\hat{\kappa}$ es un estimador robusto y consistente de κ_0 . Por ejemplo, $\hat{\kappa}$ puede calcularse previamente estimando los parámetros sin la restricción de monotonía. Más precisamente, definimos los estimadores $(\hat{\boldsymbol{\beta}}, \hat{\eta}) = (\hat{\boldsymbol{\beta}}, \sum_{j=1}^{k_n} \hat{\lambda}_j B_j) = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}^\top \mathbf{B})$ a través de los valores $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ tales que

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}}{\operatorname{argmín}} L_n \left(\boldsymbol{\beta}, \sum_{j=1}^{k_n} \lambda_j B_j, \hat{\kappa} \right), \quad (2.2)$$

donde $\mathcal{L}_{k_n} = \{\boldsymbol{\lambda} \in \mathbb{R}^{k_n} : \lambda_1 \leq \dots \leq \lambda_{k_n}\}$. Si llamamos $\mathbf{B}_i = (B_1(t_i), \dots, B_{k_n}(t_i))$, tenemos que

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}}{\operatorname{argmín}} \frac{1}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{B}_i^\top \boldsymbol{\lambda}, \hat{\kappa}) w(\mathbf{x}_i).$$

Sea

$$\mathcal{G} = \{g : g \text{ es una función monótona no decreciente en } [0,1]\}. \quad (2.3)$$

De ahora en más, al considerar los estimadores definidos en (2.2), supondremos que el funcional asociado a estos estimadores es consistente en el sentido de Fisher, o Fisher-consistente por brevedad, es decir que $(\boldsymbol{\beta}_0, \eta_0)$ es el único mínimo de $L(\boldsymbol{\beta}, g, \kappa_0)$, o sea,

$$L(\boldsymbol{\beta}_0, \eta_0, \kappa_0) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, g \in \mathcal{G}}{\operatorname{mín}} L(\boldsymbol{\beta}, g, \kappa_0), \quad (2.4)$$

con $L(\boldsymbol{\beta}_0, \eta_0, \kappa_0) < L(\boldsymbol{\beta}, g, \kappa_0)$ para todo $(\boldsymbol{\beta}, g) \neq (\boldsymbol{\beta}_0, \eta_0)$ y $g \in \mathcal{G}$. Esta condición es usual en robustez y establece que los parámetros que se desean estimar son efectivamente los parámetros del modelo. En Bianco *et al.* (2013a) se pide una condición similar para modelos lineales generalizados, mientras que para modelos parcialmente lineales generalizados Boente *et al.* (2006) y Boente y Rodríguez (2010) describen condiciones generales que aseguran $L(\boldsymbol{\beta}_0, \eta_0, \kappa_0) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, g \in \mathcal{G}}{\operatorname{mín}} L(\boldsymbol{\beta}, g, \kappa_0)$. En la Sección 2.7, presentamos resultados sobre la consistencia de Fisher (2.4) para algunas familias de distribuciones.

2.3. Estimadores monotonizados basados en B-splines

En esta sección, propondremos estimadores robustos basados en el procedimiento de monotonización de Dette *et al.* (2006) y estudiado en Neumeyer (2007). Para ello recordaremos la definición de este método. Dada una función medible $f : [a, b] \rightarrow \mathbb{R}$, definimos la función $\Upsilon(f) : \mathbb{R} \rightarrow \mathbb{R}$ como

$$\Upsilon(f)(u) = \int_a^b \mathbb{I}_{\{f(z) \leq u\}} dz + a \quad u \in \mathbb{R}. \quad (2.5)$$

Si f es estrictamente creciente, la función $\Upsilon(f)\mathbb{I}_{[f(a), f(b)]}$ es la inversa f^{-1} de la función f . Por otra parte, si f es no decreciente, entonces $\Upsilon(f)\mathbb{I}_{[f(a), f(b)]}$ corresponde a la inversa generalizada $f^{-1}(u) = \inf\{z : f(z) > u\}$ que puede tener puntos de salto cuando f es constante en intervalos. Como menciona Neumeyer (2007), para cualquier función f , $\Upsilon(f)$ es siempre medible y creciente.

Para una función $\eta : [0, 1] \rightarrow \mathbb{R}$, Dette *et al.* (2006) y Neumeyer (2007) consideran la modificación creciente $\eta_h : [0, 1] \rightarrow \mathbb{R}$ de η como

$$\eta_h = \Upsilon \left(\Upsilon(\eta) \mathbb{I}_{[\eta(0), \eta(1)]} \right) \mathbb{I}_{[0,1]},$$

que es tal que $\eta_h = \eta$ para cualquier función η no decreciente. Vale la pena mencionar que como $v = \Upsilon(\eta) \mathbb{I}_{[\eta(0), \eta(1)]}$ es una función no decreciente, entonces $\Upsilon(v)(\cdot) = \inf\{z : v(z) > \cdot\}$.

Por lo tanto, podemos definir estimadores monotonizados basados en B-splines como sigue. En primer lugar, definimos estimadores iniciales de β_0 y η_0 sin imponer restricciones, es decir, definimos $(\hat{\beta}_{\text{INI}}, \hat{\eta}_{\text{INI}}) = (\hat{\beta}_{\text{INI}}, \sum_{j=1}^{k_n} \hat{\lambda}_{\text{INI},j} B_j) = (\hat{\beta}, \hat{\lambda}_{\text{INI}}^T \mathbf{B})$ donde $(\hat{\beta}_{\text{INI}}, \hat{\lambda}_{\text{INI}})$ son tales que

$$(\hat{\beta}_{\text{INI}}, \hat{\lambda}_{\text{INI}}) = \underset{(\beta, \lambda) \in \mathbb{R}^{p+k_n}}{\text{argmín}} L_n \left(\beta, \sum_{j=1}^{k_n} \lambda_j B_j, \hat{\kappa} \right). \quad (2.6)$$

Un estimador monótono para η_0 puede construirse teniendo en cuenta que $\eta_{\text{INI}} : [0, 1] \rightarrow \mathbb{R}$ y de define como

$$\hat{\eta}_h = \Upsilon \left(\Upsilon(\hat{\eta}_{\text{INI}}) \mathbb{I}_{[\hat{\eta}_{\text{INI}}(0), \hat{\eta}_{\text{INI}}(1)]} \right) \mathbb{I}_{[0,1]}. \quad (2.7)$$

Como se menciona en Neumeyer (2007), la expresión $\Upsilon(\eta)$ puede aproximarse por

$$\Upsilon(\eta)(u) \approx \tilde{\Upsilon}(\eta)(u) = \frac{1}{N} \sum_{j=1}^N \mathbb{I}_{\{\eta(j/N) \leq u\}}, \quad (2.8)$$

lo que lleva a la aproximación discreta del estimador

$$\hat{\eta}_{h,D} = \Upsilon \left(\tilde{\Upsilon}(\hat{\eta}_{\text{INI}}) \mathbb{I}_{[\hat{\eta}_{\text{INI}}(0), \hat{\eta}_{\text{INI}}(1)]} \right) \mathbb{I}_{[0,1]}.$$

El Teorema 3.1 (a) de Neumeyer (2007) establece que si η_0 es una función no decreciente, entonces existe una constante $C > 0$ tal que $\|\hat{\eta}_h - \eta_0\|_\infty \leq C \|\hat{\eta}_{\text{INI}} - \eta_0\|_\infty$. Por otra parte, si además η_0 es derivable y $\inf_{0 \leq t \leq 1} \eta_0'(t) > 0$, entonces existe una constante $C > 0$ tal que

$$\|\hat{\eta}_{h,D} - \eta_0\|_\infty \leq C \left(\|\hat{\eta}_{\text{INI}} - \eta_0\|_\infty + \frac{1}{N} \right).$$

Estos dos resultados implican que el estimador monotonizado $\hat{\eta}_h$ será uniformemente consistente si el estimador inicial lo es y su versión discreta $\hat{\eta}_{h,D}$ lo será si además $N \rightarrow \infty$. Por esta razón, bastará estudiar el comportamiento de $\hat{\eta}_{\text{INI}}$. Para ello, necesitaremos que el funcional asociado a los estimadores definidos en (2.6) sea Fisher-consistente, es decir que al considerar estos estimadores supondremos que

$$L(\beta_0, \eta_0, \kappa_0) = \underset{\beta \in \mathbb{R}^p, g \in \mathcal{G}_0}{\text{mín}} L(\beta, g, \kappa_0), \quad (2.9)$$

con $L(\beta_0, \eta_0, \kappa_0) < L(\beta, g, \kappa_0)$ para todo $(\beta, g) \neq (\beta_0, \eta_0)$ y $g \in \mathcal{G}_0$ con

$$\mathcal{G}_0 = \{g : [0, 1] \rightarrow \mathbb{R} \text{ medible}\}. \quad (2.10)$$

Esta condición es más fuerte que (2.4), pero como veremos en el Apéndice 2.7 se cumple para varias familias de distribuciones. Al estimador final obtenido por este procedimiento lo llamaremos $\hat{\eta}_{\text{MOD}}$.

2.4. La función de pérdida

Los estimadores definidos en (2.2) y (2.6) dependen de la función de pérdida ρ que determinará cuán resistentes resultan los estimadores ante la presencia de datos atípicos en la muestra.

Como fue descrito por varios autores, bajo un modelo lineal generalizado completamente paramétrico, la función de pérdida ρ se elige para acotar los valores grandes de la *deviance* o bien de los residuos de Pearson. Bianco y Yohai (1996), Croux y Haesbroeck (2003), Bianco *et al.* (2005) y Cantoni y Ronchetti (2001) presentan diferentes elecciones para la función de pérdida, siguiendo estos dos enfoques. Por otro lado, funciones de *score* acotadas óptimas han sido estudiadas en Stefanski *et al.* (1986). A continuación, recordaremos brevemente la definición de la familia que acota la *deviance* que es la función de pérdida usada en nuestro estudio de simulación. Más detalles pueden verse, por ejemplo, en Boente *et al.* (2006) que considera esta familia de funciones de pérdida para estimar parámetros de un modelo parcialmente lineal generalizado usando un enfoque de perfiles basados en núcleos.

Sea φ_a una función acotada, no decreciente, no constante, con derivadas continuas φ'_a , siendo a la constante de calibración. Típicamente, φ_a es una función que se comporta como la función identidad en un entorno del 0 pero acota valores grandes de la *deviance*. Denotamos como $f(\cdot, s)$ a la densidad de la función de distribución $F(\cdot, s)$ con $y | (\mathbf{x}, t) \sim F(\cdot, H(\eta(t) + \mathbf{x}^T \boldsymbol{\beta}))$, donde por simplicidad hemos omitido al parámetro κ_0 . En este contexto, los *estimadores robustos basados en deviance* corresponden a la siguiente elección de la función $\rho(y, u, a)$

$$\rho(y, u, a) = \varphi_a[-\log f(y, H(u)) + \log f(y, y)] + G_a(H(u)). \quad (2.11)$$

El término de corrección G_a está dado por

$$G'_a(s) = \mathbb{E}_s \left(\varphi'_a[-\log f(y, s) + \log f(y, y)] \frac{f'(y, s)}{f(y, s)} \right),$$

donde \mathbb{E}_s indica que la esperanza se calcula cuando $y \sim F(\cdot, s)$ y $f'(y, s)$ es una notación compacta para $\partial f(y, s)/\partial s$. Por ejemplo, en un modelo lineal generalizado, el estimador de máxima verosimilitud corresponde a la elección de $\varphi_a(s) = s$, $G_a(u) = 0$ y $w \equiv 1$. Para una función general φ_a , se incluye el factor de corrección para garantizar la consistencia de Fisher bajo el modelo verdadero como ocurre en el modelo lineal generalizado. Si el factor de corrección fuera elegido igual a 0, los resultados del Capítulo 3 solo asegurarían que los estimadores resultan consistentes a los minimizadores de $L(\boldsymbol{\beta}, g, \kappa_0)$. Sin embargo, como se discute en Bianco *et al.* (2005), el término de corrección G_a puede omitirse cuando consideramos una familia continua de distribuciones con función de densidad fuertemente unimodal. En este caso, κ_0 puede jugar el rol de constante de calibración. Por ejemplo, para la distribución Gamma la constante de calibración depende del parámetro de forma, entonces si la forma es desconocida debemos considerar un estimador inicial. Sobre este punto volveremos en la Sección 2.5.

Para los modelos de regresión Poisson y regresión logística κ_0 no necesita ser estimado puesto que no existe un parámetro adicional y podemos suponer $\kappa_0 = 1$ y $\varphi_a(s) = \varphi(s)$. Más aun, como señalan Croux y Haesbroeck (2003), para garantizar la existencia de una solución en el modelo logístico, además de la condición de solapamiento requerida para el estimador de máxima verosimilitud, la derivada φ' de la función $\varphi(s)$ debe satisfacer restricciones adicionales: φ' debe ser creciente en

$(-\infty, A_0]$ y decreciente en $[A_0, +\infty)$ para algún $A_0 > 0$ o creciente en \mathbb{R} y también cumplir que $\lim_{s \rightarrow +\infty} \varphi'(st)/\varphi'(-s) = \infty$ para cualquier $t > 0$. Un ejemplo de función φ que satisface estas condiciones está dado en ese trabajo.

Por otro lado, cuando la función de vínculo es igual a la función identidad, la función de pérdida cuadrática usual es reemplazada por una ρ -función luego de escalar los residuos para controlar el efecto de residuos grandes. En este caso, el modelo puede escribirse como

$$y = \beta_0^T \mathbf{x} + \eta_0(t) + u, \quad (2.12)$$

donde la componente no paramétrica es monótona y el error u es independiente de las covariables (\mathbf{x}, t) . Dos situaciones deben distinguirse según la distribución de los errores.

Para errores simétricos, κ_0 adopta el rol de parámetro de escala de los errores y la función de pérdida puede definirse como

$$\rho(y, s, a) = \phi\left(\frac{y-s}{a}\right), \quad (2.13)$$

donde $\phi : \mathbb{R} \rightarrow [0, \infty)$ es una ρ -función definida como en Maronna *et al.* (2019). A continuación, damos su definición porque la mencionaremos recurrentemente.

Definición 2.1. *Decimos que $\phi : \mathbb{R} \rightarrow [0, \infty)$ es una ρ -función definida como en Maronna *et al.* (2019) si es una función no nula, continua, par, no decreciente con $\phi(0) = 0$ y tal que $\phi(u) < \phi(v)$ cuando $0 \leq u < v$ con $\phi(v) < \sup_s \phi(s)$. Si ϕ es acotada supondremos que $\|\phi\|_\infty = 1$.*

Por ejemplo, una tal ρ -función es la función bicuadrada de Tukey

$$\phi(y) = \phi_T(y) = \min(3y^2 - 3y^4 + y^6, 1).$$

Al igual que en el conjunto de datos de costos hospitalarios analizados en la Sección 6.2, podría resultar de interés estimar los parámetros del modelo parcialmente lineal isotónico cuando los errores tienen distribución asimétrica. Por esa razón, consideraremos estimadores bajo el modelo (2.12) cuando la densidad de los errores es de la forma

$$f_0(u, \alpha_0) = Q(\alpha_0) \exp^{\alpha_0 v(u)}, \quad (2.14)$$

con $\alpha_0 > 0$ un parámetro desconocido y v una función continua, tal que la densidad tiene una única moda en u_0 , es decir, v tiene un único máximo en u_0 . En este caso, los estimadores robustos a considerar quedarán definidos usando la función de pérdida

$$\rho(y, s, a) = \phi\left(\frac{\sqrt{d}(y-s)}{a}\right), \quad (2.15)$$

donde $d(s) = v(u_0) - v(s)$ y ϕ una ρ -función. Una característica atractiva de esta familia de distribuciones es que permite modelar errores asimétricos como los del modelo de regresión log-Gamma. Bajo un modelo de regresión lineal, la densidad exponencial (2.14) ha sido considerada por Bianco *et al.* (2005) para extender los MM-estimadores del modelo usual de regresión lineal al caso de un modelo con errores asimétricos. La constante a es la constante de calibración relacionada con el parámetro α_0 y debe ser estimada para calibrar los estimadores robustos y reducir el peso de los residuos *deviance* grandes. Notemos que para errores simétricos no es necesario suponer conocida la densidad de los errores como sí debe conocerse en el caso de errores asimétricos o más generalmente para modelos parcialmente lineales generalizados. Discutiremos en detalle el modelo log-Gamma en la sección siguiente.

2.5. El caso particular del modelo de regresión log–Gamma

La distribución Gamma, con el logaritmo como función de vínculo, juega un rol importante entre los modelos lineales generalizados y es usualmente llamado regresión log–Gamma. Una descripción del mismo puede verse, por ejemplo, en el Capítulo 8 de McCullagh y Nelder (1989).

Para cualquier $\alpha > 0$ y $\mu > 0$, indicaremos por $\Gamma(\alpha, \mu)$ a la parametrización de la distribución Gamma dada por la densidad

$$f(u, \alpha, \mu) = \frac{1}{\mu^\alpha \Gamma(\alpha)} \alpha^\alpha u^{\alpha-1} \exp[-(\alpha/\mu)u] \mathbb{I}_{u \geq 0} .$$

Denotemos como $(z_i, \mathbf{x}_i^\top, t_i)^\top$ a observaciones que siguen un modelo parcialmente lineal log–Gamma, es decir, tales que $z_i | (\mathbf{x}_i, t_i) \sim \Gamma(\alpha_0, \mu_i)$, con $\mu_i = \mathbb{E}(z_i | (\mathbf{x}_i, t_i))$ y función de vínculo $\log(\mu_i) = \boldsymbol{\beta}_0^\top \mathbf{x}_i + \eta_0(t_i)$. En este caso, las respuestas pueden transformarse para modelarlas mediante un modelo de regresión parcialmente lineal con errores asimétricos (ver, por ejemplo, Cantoni y Ronchetti, 2006 para el caso $\eta_0 \equiv 0$). Sean $y_i = \log(z_i)$ las respuestas transformadas, entonces y_i satisface el modelo (2.12) con u_i independientes de (\mathbf{x}_i, t_i) y $u_i \sim \log(\Gamma(\alpha_0, 1))$ con densidad

$$f_0(u, \alpha_0) = \frac{\alpha_0^{\alpha_0}}{\Gamma(\alpha_0)} \exp[\alpha_0(u - \exp(u))] .$$

Esta densidad es asimétrica y unimodal con máximo en $u_0 = 0$ y es un caso particular de la familia de densidades (2.14). Para modelos lineales totalmente paramétricos, o sea, cuando $\eta_0(t) = \gamma_0 t$, una descripción de los estimadores robustos basados en residuos *deviance* fue dada en Bianco *et al.* (2005), mientras que Heritier *et al.* (2009) consideran M–estimadores basados en los residuos de Pearson. A continuación, describiremos cómo adaptar los estimadores basados en acotar los residuos *deviance* a la presente situación.

Para incluir al modelo log–Gamma, presentaremos el marco general de estimadores bajo el modelo de regresión parcialmente lineal isotónico (2.12) cuando los errores tienen densidad dada por (2.14). Llamemos $d_i(\boldsymbol{\beta}, \eta, \alpha)$ a la *deviance* de la i -ésima observación,

$$d_i(\boldsymbol{\beta}, \eta, \alpha) = 2\alpha d(y_i - (\mathbf{x}_i^\top \boldsymbol{\beta} + \eta(t_i))) ,$$

donde $d(u) = v(u_0) - v(u)$. El caso log–Gamma corresponde a $d(u) = \exp(u) - u - 1$.

En este contexto, los estimadores clásicos que consideraremos no están basados en la cuasiverosimilitud sino en la *deviance* y corresponden a la elección $\varphi_a(u) = \varphi(u) = u$ en (2.11) ya que no requieren una constante de calibración. Por lo tanto, la función de pérdida es igual a $\rho(y, s, a) = \rho(y, s) = d(y-s)$ y sus derivadas son $\Psi(y, s) = \partial \rho(y, s) / \partial s = v'(y-s)$, $\chi(y, s) = \partial \Psi(y, s) / \partial s = -v''(y-s)$. En particular, para el modelo log–Gamma tenemos que $\Psi(y, s) = 1 - \exp(y-s)$ y $\chi(y, s) = \exp(y-s)$. Por lo tanto, si $\mathbf{B}(t)$ y \mathbf{B}_i se definen como en la Sección 2.2, los estimadores clásicos sin restricciones de $(\boldsymbol{\beta}_0, \eta_0)$ pueden obtenerse como $(\hat{\boldsymbol{\beta}}_{\text{INI,CL}}, \hat{\eta}_{\text{INI,CL}})$ con $\hat{\eta}_{\text{INI,CL}}(t) = \hat{\boldsymbol{\lambda}}_{\text{INI,CL}}^\top \mathbf{B}(t)$ y

$$(\hat{\boldsymbol{\beta}}_{\text{INI,CL}}, \hat{\boldsymbol{\lambda}}_{\text{INI,CL}}) = \underset{(\boldsymbol{\beta}, \boldsymbol{\lambda}) \in \mathbb{R}^{p+k_n}}{\text{argmín}} \sum_{i=1}^n d(y_i - (\mathbf{x}_i^\top \boldsymbol{\beta} + \boldsymbol{\lambda}^\top \mathbf{B}_i)) ,$$

mientras que los basados en B–splines monótonos, se definen como $(\hat{\boldsymbol{\beta}}_{\text{CL}}, \hat{\eta}_{\text{CL}})$ con $\hat{\eta}_{\text{CL}}(t) = \hat{\boldsymbol{\lambda}}_{\text{CL}}^\top \mathbf{B}(t)$

$$(\hat{\boldsymbol{\beta}}_{\text{CL}}, \hat{\boldsymbol{\lambda}}_{\text{CL}}) = \underset{(\boldsymbol{\beta}, \boldsymbol{\lambda}) \in \mathbb{R}^p \times \mathcal{L}_{k_n}}{\text{argmín}} \sum_{i=1}^n d(y_i - (\mathbf{x}_i^\top \boldsymbol{\beta} + \boldsymbol{\lambda}^\top \mathbf{B}_i)) .$$

Por otro lado, como se menciona en la Sección 2.4, los estimadores robustos puede ser definidos tomando la función de pérdida dada en (2.15) que controla los valores grandes de la *deviance* a través de la ρ -función ϕ . Por lo tanto, la constante de calibración a debe ser elegida. Con esta notación, el estimador clásico se recupera tomando $\phi(u) = u^2$ y $a = 1$.

2.5.1. Algoritmo para el caso log-Gamma

Para dar un algoritmo que calcule los estimadores con una constante adaptativa consideramos la situación en la que k_n está fijo. La Sección 2.6 describe como elegir la cantidad de nodos usando un criterio de información robusto. Buscamos $\boldsymbol{\lambda}$ tal que $\sum_{i=j}^{k_n} \lambda_j B_j(t)$ sea una buena aproximación para $\eta_0(t)$. Como en Bianco *et al.* (2005), puede considerarse un procedimiento de tres pasos para calcular el estimador inicial de los parámetros. Primero notemos que, como la constante de calibración de la función de pérdida depende del parámetro desconocido α_0 , Bianco *et al.* (2005) introducen una sucesión adaptativa de constantes de calibración $\hat{c}_{M,n}$ para definir una sucesión de M-estimadores $\hat{\boldsymbol{\theta}}_{M,n} = (\hat{\boldsymbol{\beta}}_{M,n}, \hat{\boldsymbol{\lambda}}_{M,n})$. La constante de calibración permite por un lado, reducir los residuos *deviance* grandes asegurando que los estimadores resultantes sean robustos, y por el otro, calibrar los estimadores para que alcancen una eficiencia determinada. Si la sucesión de constantes de calibración es tal que $\hat{c}_{M,n} \xrightarrow{p} c_0$, como veremos en el Capítulo 4, los estimadores

$$\hat{\boldsymbol{\theta}}_{M,n} = \underset{(\boldsymbol{\beta}, \boldsymbol{\lambda}) \in \mathbb{R}^{p+k_n}}{\operatorname{argmín}} \sum_{i=1}^n \phi \left(\frac{\sqrt{d(y_i - [\mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}_i])}}{\hat{c}_{M,n}} \right),$$

son tales que $\hat{\boldsymbol{\beta}}_{M,n}$ tiene matriz de covarianza asintótica dada por $(B(\phi, \alpha_0, c_0)/A^2(\phi, \alpha_0, c_0)) \boldsymbol{\Sigma}_0$, donde $D(\alpha_0) \boldsymbol{\Sigma}_0$ es la matriz de covarianza asintótica de los estimadores clásicos obtenidos cuando $\phi(u) = u^2$ y $D(\alpha_0)$ es una constante dependiendo solo de α_0 . Para el modelo log-Gamma $D(\alpha_0) = 1/\alpha_0$. Las constantes $B(\phi, \alpha_0, c_0)$ y $A(\phi, \alpha_0, c_0)$ dependen solo de la derivada de la función de *score* ϕ y del parámetro de forma α_0 , pero no de las covariables. De esta forma, los estimadores de $\boldsymbol{\beta}_0$ pueden ser calibrados para alcanzar una determinada eficiencia respecto del estimador clásico. En adelante, llamemos $C_e(\alpha_0)$ al valor de la constante de calibración c_0 tal que el M-estimador tiene eficiencia e con respecto al clásico.

En nuestra modificación, consideramos el siguiente algoritmo de cuatro pasos para calcular un MM-estimador generalizado. Cabe destacar que el método descrito en lo que sigue es la propuesta considerada en Bianco *et al.* (2005) aplicada a la aproximación de dimensión finita de η_0 pero tomando en cuenta las restricciones de orden.

- **Paso 1.** Primero calculamos un S -estimador inicial $(\tilde{\boldsymbol{\beta}}_n, \tilde{\boldsymbol{\lambda}}_n)$ y el correspondiente estimador de escala $\hat{\sigma}_n$ tomando $b = \sup \phi/2$. Para ser más precisos, para cada valor de $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ sea $\sigma_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$ el M-estimador de escala de $\sqrt{d(y_i - [\mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}_i])}$ dado por

$$\frac{1}{n} \sum_{i=1}^n \phi \left(\frac{\sqrt{d(y_i - [\mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}_i])}}{\sigma_n(\boldsymbol{\beta}, \boldsymbol{\lambda})} \right) = b,$$

donde ϕ es la función bicuadrada de Tukey ϕ_T .

El S -estimador de $(\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0)$ para el modelo considerado está definido a través de un S -estimador calculado sin restricciones

$$(\tilde{\boldsymbol{\beta}}_n, \tilde{\boldsymbol{\lambda}}_n) = \underset{(\boldsymbol{\beta}, \boldsymbol{\lambda}) \in \mathbb{R}^{p+k_n}}{\operatorname{argmín}} \sigma_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$$

y el correspondiente estimador de escala se define como $\hat{\sigma}_n = \min_{(\boldsymbol{\beta}, \boldsymbol{\lambda}) \in \mathbb{R}^{p+k_n}} \sigma_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$. Más aun, definamos un estimador preliminar de η_0 como $\tilde{\eta}(t) = \sum_{j=1}^{k_n} \tilde{\lambda}_{n,j} B_j(t)$.

Sea u una variable aleatoria con densidad (2.14) y sea $\sigma^*(\alpha)$ la solución de

$$\mathbb{E}_G \left[\phi \left(\frac{\sqrt{d(u_1)}}{\sigma^*(\alpha)} \right) \right] = b.$$

Argumentos similares a los considerados por el Teorema 5 de Bianco *et al.* (2005) combinados con los resultados del Teorema 3.9 permiten mostrar que $\tilde{\boldsymbol{\beta}}_n \xrightarrow{a.s.} \boldsymbol{\beta}_0$, $\|\tilde{\eta} - \eta_0\|_{\mathcal{F}}^2 \xrightarrow{a.s.} 0$ y $\hat{\sigma}_n \xrightarrow{a.s.} \sigma^*(\alpha)$, bajo condiciones de regularidad. Más aun, como en Bianco *et al.* (2005), si v es una función diferenciable, $\sigma^*(\alpha)$ es una función continua y estrictamente decreciente, con lo cual podemos definir un estimador de α_0 como $\hat{\alpha}_n = \sigma^{*-1}(\hat{\sigma}_n)$ dando origen a un estimador fuertemente consistente para α_0 .

- **Paso 2.** En el segundo paso, calculamos $\hat{\alpha}_n = \sigma^{*-1}(\hat{\sigma}_n)$ y

$$\hat{c}_n = \max(\hat{\sigma}_n, C_e(\hat{\alpha}_n)) = \max(\hat{\sigma}_n, C_e(\sigma^{*-1}(\hat{\sigma}_n))).$$

Entonces, $\hat{c}_n \xrightarrow{p} c_0 = \max\{\sigma^*(\alpha_0), C_e(\alpha_0)\}$.

- **Paso 3.** Sea $(\hat{\boldsymbol{\beta}}_{\text{INI}}, \hat{\boldsymbol{\lambda}}_{\text{INI}})$ un MM-estimador adaptativo sin restricciones definido por

$$(\hat{\boldsymbol{\beta}}_{\text{INI}}, \hat{\boldsymbol{\lambda}}_{\text{INI}}) = \underset{(\boldsymbol{\beta}, \boldsymbol{\lambda}) \in \mathbb{R}^{p+k_n}}{\operatorname{argmín}} \sum_{i=1}^n \phi \left(\frac{\sqrt{d(y_i - [\mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}_i])}}{\hat{c}_n} \right) w(\mathbf{x}_i),$$

donde la función $w(\mathbf{x})$ controla puntos de alta palanca en el espacio de las covariables \mathbf{x} .

- **Paso 4.** Si $\hat{\lambda}_{\text{INI},1} \leq \hat{\lambda}_{\text{INI},2} \leq \dots \leq \hat{\lambda}_{\text{INI},k_n}$, los estimadores finales son $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\text{INI}}$ y $\hat{\eta}(t) = \sum_{j=1}^{k_n} \hat{\lambda}_{\text{INI},j} B_j(t)$. De otra manera, los estimadores finales pueden obtenerse ya sea, considerando el procedimiento de monotonización de Dette *et al.* (2006) descrito en la Sección 2.3 o mediante una minimización con restricciones como se presentó en la Sección 2.2. En este último caso, utilizamos un algoritmo estándar de minimización no lineal con restricciones tomando como valor inicial $(\hat{\boldsymbol{\beta}}^{(0)}, \boldsymbol{\lambda}^{(0)}) = (\hat{\boldsymbol{\beta}}_{\text{INI}}, \boldsymbol{\lambda}^{(0)})$, donde $\boldsymbol{\lambda}^{(0)} \in \mathcal{L}_{k_n}$. Una posible elección posible para $\boldsymbol{\lambda}^{(0)}$ es $\lambda_1^{(0)} = \lambda_2^{(0)} = 0$ y $\lambda_i^{(0)} = i - 2$ para $i = 3, \dots, k_n$, en cuyo caso la matriz \mathbf{A} abajo descripta sería igual a $\mathbf{A} = (1, -1, 0, \dots, 0)$.

Una vez que el valor inicial $(\hat{\boldsymbol{\beta}}^{(0)}, \boldsymbol{\lambda}^{(0)})$ con $\boldsymbol{\lambda}^{(0)} \in \mathcal{L}_{k_n}$ ha sido elegido a través de **Paso 1** a **Paso 3**, la optimización con restricciones descripta en el **Paso 4** puede implementarse, por ejemplo, a través del paquete `alabama`¹ de R.

Describimos brevemente un algoritmo, similar al considerado en Lu (2015), para aproximar el minimizador con restricciones de $L_n(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{c}_n)$.

¹Disponible en <http://cran.r-project.org/web/packages/alabama/index.html>.

- Sea $\widehat{\nabla}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\widehat{\nabla}_1(\boldsymbol{\beta}, \boldsymbol{\lambda})^\top, \widehat{\nabla}_2(\boldsymbol{\beta}, \boldsymbol{\lambda})^\top)^\top$ y $\widehat{\mathbf{H}}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\widehat{\mathbf{H}}_{ij}(\boldsymbol{\beta}, \boldsymbol{\lambda}))_{1 \leq i, j \leq 2}$ el vector gradiente y la matriz Hessiana de la función objetivo, esto es,

$$\begin{aligned}\widehat{\nabla}_1(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{B}_i^\top \boldsymbol{\lambda}, \widehat{c}_n) w(\mathbf{x}_i) \mathbf{x}_i, \\ \widehat{\nabla}_2(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{B}_i^\top \boldsymbol{\lambda}, \widehat{c}_n) w(\mathbf{x}_i) \mathbf{B}_i, \\ \widehat{\mathbf{H}}_{11}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{B}_i^\top \boldsymbol{\lambda}, \widehat{c}_n) w(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top, \\ \widehat{\mathbf{H}}_{12}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{B}_i^\top \boldsymbol{\lambda}, \widehat{c}_n) w(\mathbf{x}_i) \mathbf{B}_i \mathbf{x}_i^\top, \\ \widehat{\mathbf{H}}_{21}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{B}_i^\top \boldsymbol{\lambda}, \widehat{c}_n) w(\mathbf{x}_i) \mathbf{x}_i \mathbf{B}_i^\top, \\ \widehat{\mathbf{H}}_{22}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{B}_i^\top \boldsymbol{\lambda}, \widehat{c}_n) w(\mathbf{x}_i) \mathbf{B}_i \mathbf{B}_i^\top,\end{aligned}$$

donde

$$\Psi(y, s, a) = \partial \rho(y, s, a) / \partial s = \frac{1}{2a \sqrt{d(y-s)}} \phi' \left(\frac{\sqrt{d(y-s)}}{a} \right) v'(y-s)$$

con ϕ' la primera derivada de ϕ y $\chi(y, u, a) = \partial \Psi(y, u, a) / \partial u$. Sea $\mathcal{A} = \{i_1, \dots, i_m\}$ el conjunto de índices tales que $\lambda_{i_j}^{(0)} = \lambda_{i_j+1}^{(0)}$. Si $m > 0$ definimos la matriz activa $\mathbf{A} \in \mathbb{R}^{m \times (k_n+p)}$ como la matriz cuya j -ésima fila es el vector cuyo i_j -ésimo elemento es igual a 1, el $(i_j + 1)$ -ésimo elemento igual a -1 y el resto es igual a 0.

- Fijemos un valor inicial $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})$ y llamemos $\widehat{\mathbf{H}} = \widehat{\mathbf{H}}(\boldsymbol{\theta})$, $\widehat{\nabla} = \widehat{\nabla}(\boldsymbol{\theta})$. En el primer paso, $\boldsymbol{\theta} = (\widehat{\boldsymbol{\beta}}_{\text{INI}}, \boldsymbol{\lambda}^{(0)})$.
- **Paso 4.1.** Obtenga la dirección factible como

$$\boldsymbol{\eta} = \left(\mathbf{I} - \widehat{\mathbf{H}}^{-1} \mathbf{A}^\top \left(\mathbf{A} \widehat{\mathbf{H}}^{-1} \mathbf{A}^\top \right)^{-1} \mathbf{A} \right) \widehat{\mathbf{H}}^{-1} \widehat{\nabla}$$

- **Paso 4.2.** Si $\|\boldsymbol{\eta}\| < \epsilon$ para algún $\epsilon > 0$ suficientemente chico, calcule los multiplicadores de Lagrange $\boldsymbol{\nu} = - \left(\mathbf{A} \widehat{\mathbf{H}}^{-1} \mathbf{A}^\top \right)^{-1} \mathbf{A} \widehat{\mathbf{H}}^{-1} \widehat{\nabla}$. Sea ν_i la i -ésima componente de $\boldsymbol{\nu}$.
 - Si $\nu_i \geq 0$, para todo $i \in \mathcal{A}$, entonces $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$.
 - Si existe al menos un $i \in \mathcal{A}$ tal que $\nu_i < 0$, determine el índice correspondiente al valor más grande de ν_i , elimínelo de \mathcal{A} y vuelva al **Paso 4.1**.

- **Paso 4.3** Calcule

$$\xi_1 = \min_{\eta_i > \eta_{i+1}, i \notin \mathcal{A}, 1 \leq i \leq k_n-1} \frac{-(\lambda_{i+1} - \lambda_i)}{\eta_{i+1} - \eta_i}$$

y obtenga el menor r tal que $L_n(\boldsymbol{\theta} + 2^{-r} \boldsymbol{\eta}, \widehat{\kappa}) < L_n(\boldsymbol{\theta}, \widehat{\kappa})$. Reemplace $\boldsymbol{\theta}$ por $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \min(2^{-r}, \xi_1) \boldsymbol{\eta}$, actualice \mathcal{A} y \mathbf{A} y vuelva al **Paso 4.1**.

2.6. Sobre la elección de la cantidad de splines

Un problema importante a considerar es la elección del número de nodos y su ubicación para definir el espacio de B-splines. La selección de los nodos es más importante para la estimación de η_0 que para la de β_0 . Un enfoque es usar nodos uniformes que es el planteo empleado en nuestro estudio de simulación. Los nodos uniformes son usualmente suficientes cuando la función η_0 no exhibe cambios dramáticos en sus derivadas. Por otro lado, nodos no uniformes son deseables cuando la función tiene comportamientos locales disímiles en diferentes regiones. Otra perspectiva usual es ubicar los nodos según los cuantiles de las observaciones t_i tomados en percentiles uniformes.

El número de nodos m_n o equivalentemente la dimensión de la base k_n (recordemos que $k_n = m_n + \ell$) puede determinarse por un criterio de selección de modelos tal como una versión robusta del criterio de Akaike usado en Lu (2015) o la versión robusta del criterio *BIC* de Schwartz (1978) ya considerado en He y Shi (1996) y He *et al.* (2002).

Describiremos a continuación, la adaptación del criterio *BIC* a nuestro problema. Supongamos que $(\hat{\beta}^{(k)}, \hat{\lambda}^{(k)})$ es el estimador que resuelve (2.2) (o respectivamente (2.6)) usando un espacio de splines k -dimensional. Definimos un criterio *BIC* robusto como

$$BIC(k) = \frac{1}{n} \sum_{i=1}^n \rho \left(y_i, \mathbf{x}_i^T \hat{\beta}^{(k)} + \sum_{j=1}^k \hat{\lambda}_j^{(k)} B_j(t_i), \hat{\kappa} \right) w(\mathbf{x}_i) + \frac{\log n}{2n} (k + p) \quad (2.16)$$

de forma que valores grandes de este criterio indican un mal ajuste.

Como es usual en los procedimientos basados en splines, el número de nodos debe aumentar lentamente con el tamaño de muestra n para alcanzar la tasa de convergencia óptima. Cuando se supone que η_0 es dos veces continuamente diferenciable y se consideran splines cúbicos ($\ell = 4$), como en nuestro estudio de simulación, de acuerdo a las tasas de convergencia dadas en el Teorema 3.11, un criterio posible es buscar el primer mínimo local de $BIC(k)$ (es decir, el menor k) en el rango de $\max(n^{1/5}/2, 4) \leq k \leq 8 + 2n^{1/5}$. Dentro de este rango, existe usualmente un solo mínimo local. Cabe notar también que el mínimo global de $BIC(k)$ ocurre para el modelo saturado cuando $k = n - p$, por lo que $BIC(k)$ es un criterio válido solo para un rango limitado de valores de k .

2.7. Apéndice: Sobre la Fisher-consistencia

En este apéndice, daremos condiciones que aseguran la consistencia de Fisher de los estimadores propuestos para algunas distribuciones. Más precisamente, mostraremos que $L(\beta_0, \eta_0, \kappa_0) = \min_{\beta \in \mathbb{R}^p, g \in \mathcal{G}_0} L(\beta, g, \kappa_0)$ donde $L(\beta, g, a)$ y \mathcal{G}_0 está definidas en (2.1) y (2.10), respectivamente.

2.7.1. Modelo parcialmente lineal con errores simétricos

El modelo parcialmente lineal corresponde a la función de vínculo identidad, $H(s) = s$ y puede escribirse como en (2.12). Como se mencionó en la Sección 2.4, en el caso de errores simétricos, la función de pérdida puede tomarse como $\rho(y, s, a) = \phi((y - s)/a)$ para una función ϕ apropiada. En este caso, el parámetro de ruido κ_0 juega el rol del parámetro de escala. Por esta razón, en

este apartado consideramos errores con distribución simétrica, una función ϕ par y reescribiremos el modelo como

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i) + \kappa_0 \varepsilon_i,$$

donde ε_i son independientes de (\mathbf{x}_i, t_i) y κ_0 es el parámetro de escala, es decir que $u_i = \kappa_0 \varepsilon_i$.

Para obtener resultados de consistencia de Fisher requerimos los siguientes supuestos.

F1 Para casi todo t_0 , $\mathbb{P}(\mathbf{x}^T \boldsymbol{\beta} = c \cup w(\mathbf{x}) = 0 | t = t_0) < 1$, para todo $\boldsymbol{\beta} \in \mathbb{R}^p$, $c \in \mathbb{R}$, $(\boldsymbol{\beta}, c) \neq \mathbf{0}$.

F2 La variable aleatoria ε tiene una función de densidad $g_0(u)$ par, no creciente en $|u|$, y estrictamente decreciente en $|u|$ en un entorno de 0.

F3 La función $\phi : \mathbb{R} \rightarrow [0, \infty)$ es una función continua, no decreciente y par tal que $\phi(0) = 0$. Más aun, si $0 \leq s < v$ con $\phi(v) < \sup_s \phi(s)$ entonces $\phi(s) < \phi(v)$. Cuando ϕ es acotada suponemos que $\sup_s \phi(s) = 1$.

La condición (**F1**) no permite que $\boldsymbol{\beta}_0$ incluya un ordenada al origen para que el modelo resulte identificable. El supuesto **F3** dice que la función ϕ es una ρ -función. El siguiente Lema establece la consistencia de Fisher de los estimadores propuestos.

Lema 2.1. *Consideremos el modelo de regresión parcialmente lineal $y_i = \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i) + \kappa_0 \varepsilon_i$ donde los errores tienen distribución simétrica. Si se cumplen **F1** a **F3**, entonces, para cualquier $a > 0$, $(\boldsymbol{\beta}_0, \eta_0)$ es el único minimizador sobre $\mathbb{R}^p \times \mathcal{G}_0$ de*

$$L(\boldsymbol{\beta}, g, a) = \mathbb{E} \phi \left(\frac{y - \mathbf{x}^T \boldsymbol{\beta} - g(t)}{a} \right) w(\mathbf{x}).$$

DEMOSTRACIÓN. Sea $\Upsilon(\mathbf{x}, t) = \mathbf{x}^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + g(t) - \eta_0(t)$, entonces

$$L(\boldsymbol{\beta}, \eta, a) = \mathbb{E} \phi \left(\frac{\kappa_0}{a} \varepsilon - \frac{\Upsilon(\mathbf{x}, t)}{a} \right) w(\mathbf{x})$$

Indiquemos por $\mathcal{A}_0 = \{(\mathbf{x}, t) : \Upsilon(\mathbf{x}, t) = 0\}$ y $b(\mathbf{x}, t) = \Upsilon(\mathbf{x}, t)/a$. Como los errores son independientes de las covariables

$$L(\boldsymbol{\beta}, \eta, a) = \mathbb{E} \phi \left(\varepsilon \frac{\kappa_0}{a} \right) \mathbb{E} (w(\mathbf{x}) \mathbb{I}_{\mathcal{A}_0}(\mathbf{x}, t)) + \mathbb{E} \left\{ \mathbb{E} \left[\phi \left(\varepsilon \frac{\kappa_0}{a} - b(\mathbf{x}, t) \right) \middle| (\mathbf{x}, t) \right] w(\mathbf{x}) \mathbb{I}_{\mathcal{A}_0^c}(\mathbf{x}, t) \right\}.$$

Observemos que $\tilde{\varepsilon} = \varepsilon \kappa_0/a$ también satisface **F2**, por lo tanto, el Lema 3.1 de Yohai (1987) junto con **F3** implican que, para todo $b \neq 0$, vale la siguiente desigualdad estricta

$$\mathbb{E} \left[\phi \left(\varepsilon \frac{\kappa_0}{a} - b \right) \right] > \mathbb{E} \left[\phi \left(\varepsilon \frac{\kappa_0}{a} \right) \right]. \quad (2.17)$$

Luego, para cualquier $(\mathbf{x}, t) \in \mathcal{A}_0^c$ tenemos que

$$\mathbb{E} \left[\phi \left(\varepsilon \frac{\kappa_0}{a} - b(\mathbf{x}, t) \right) \middle| (\mathbf{x}, t) = (\mathbf{x}_0, t_0) \right] = \mathbb{E} \left[\phi \left(\varepsilon \frac{\kappa_0}{a} - b(\mathbf{x}_0, t_0) \right) \right] > \mathbb{E} \left[\phi \left(\varepsilon \frac{\kappa_0}{a} \right) \right]$$

donde la primer igualdad se debe a la independencia entre los errores y las covariables.

Notemos que **F1** implica que $\mathbb{P}(\mathcal{A}_0^c \cap \{w(\mathbf{x}) \neq 0\}) > 0$. En conclusión, resulta que

$$\begin{aligned} L(\boldsymbol{\beta}, \eta, a) &> \mathbb{E} \phi \left(\varepsilon \frac{\kappa_0}{a} \right) \mathbb{E} (w(\mathbf{x}) \mathbb{I}_{\mathcal{A}_0}(\mathbf{x}, t)) + \mathbb{E} \left\{ \mathbb{E} \left[\phi \left(\varepsilon \frac{\kappa_0}{a} \right) \right] w(\mathbf{x}) \mathbb{I}_{\mathcal{A}_0^c}(\mathbf{x}, t) \right\} = \mathbb{E} \left(\phi \left(\varepsilon \frac{\kappa_0}{a} \right) w(\mathbf{x}) \right) \\ &> L(\boldsymbol{\beta}_0, \eta_0, a), \end{aligned}$$

lo que concluye la demostración. ■

2.7.2. Modelo parcialmente lineal con errores asimétricos

Como se mencionó en la Sección 2.5, bajo un modelo parcialmente lineal generalizado con respuestas con distribución gamma, es decir cuando $z_i | \mathbf{x}_i \sim \Gamma(\alpha, \mu_i)$ con $\mu_i = \mathbb{E}(z_i | (\mathbf{x}_i, t_i))$ y $\log(\mu_i) = \boldsymbol{\beta}_0^\top \mathbf{x}_i + \eta_0(t_i)$, las respuestas pueden transformarse mediante $y_i = \log(z_i)$ para trabajar con el modelo de regresión con errores asimétricos dado por (2.12), donde u_i y (\mathbf{x}_i, t_i) son independientes.

Recordemos que bajo un modelo log-Gamma los errores son tales que $u_i \sim \log(\Gamma(\alpha, 1))$ y su densidad es una función fuertemente unimodal. En este contexto, la función de pérdida es igual a $\rho(y, s, a) = \phi \left(\sqrt{d(y-s)/a} \right)$, donde $d(u) = \exp(u) - u - 1$.

Derivaremos un resultado de consistencia de Fisher más amplio que incluye otras distribuciones asimétricas con densidades fuertemente unimodales para los errores. Por esta razón, consideramos el siguiente supuesto

F4 La variable aleatoria u tiene función de densidad $g_0(u) > 0$ para todo u , continua y estrictamente unimodal.

Al igual que para errores simétricos, el siguiente lema da un resultado de consistencia de Fisher fuerte porque muestra que para cualquier parámetro de ruido los parámetros verdaderos $(\boldsymbol{\beta}_0, \eta_0)$ minimizan la función objetivo. Este resultado corresponde a la condición requerida en la Sección 3.3 para evitar pedir orden de consistencia al estimador del parámetro de ruido.

Lema 2.2. *Consideremos el modelo de regresión parcialmente lineal (2.12) donde la densidad del error satisface **F4**. Si se cumplen **F1** y **F3**, entonces, para todo $a > 0$, $(\boldsymbol{\beta}_0, \eta_0)$ es el único minimizador sobre $\mathbb{R}^p \times \mathcal{G}_0$ de*

$$L(\boldsymbol{\beta}, g, a) = \mathbb{E} \left[\phi \left(\frac{\sqrt{d(y - \mathbf{x}^\top \boldsymbol{\beta} - g(t))}}{a} \right) w(\mathbf{x}) \right].$$

DEMOSTRACIÓN. Como antes, sea $\Upsilon(\mathbf{x}, t) = \mathbf{x}^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + g(t) - \eta_0(t)$ y $\mathcal{A}_0 = \{(\mathbf{x}, t) : \Upsilon(\mathbf{x}, t) = 0\}$. Entonces, observando que

$$L(\boldsymbol{\beta}, g, a) = \mathbb{E} \left(\phi \left(\frac{\sqrt{d(u + \Upsilon(\mathbf{x}, t))}}{a} \right) w(\mathbf{x}) \right)$$

y usando que los errores son independientes de las covariables, concluimos que

$$L(\boldsymbol{\beta}, g, a) = \mathbb{E} \left(\phi \left(\frac{\sqrt{d(u)}}{a} \right) \right) \mathbb{E} (w(\mathbf{x}) \mathbb{I}_{\mathcal{A}_0}(\mathbf{x}, t)) + \mathbb{E} \left\{ \mathbb{E} \left[\phi \left(\frac{\sqrt{d(u + \Upsilon(\mathbf{x}, t))}}{a} \right) \middle| (\mathbf{x}, t) \right] w(\mathbf{x}) \mathbb{I}_{\mathcal{A}_0^c}(\mathbf{x}, t) \right\}. \quad (2.18)$$

Usando que los errores verifican **F4**, el Lema 1 en Bianco *et al.* (2005) permite acotar el segundo término en (2.18). En efecto, para cualquier $(\mathbf{x}, t) \in \mathcal{A}_0^c$ y para cualquier $a > 0$ fijo, tenemos que

$$\mathbb{E} \left(\phi \left(\frac{\sqrt{d(u + \Phi(\mathbf{x}, t))}}{a} \right) \middle| (\mathbf{x}, t) \right) > \mathbb{E} \left(\phi \left(\frac{\sqrt{d(u)}}{a} \right) \middle| (\mathbf{x}, t) \right) = \mathbb{E} \left(\phi \left(\frac{\sqrt{d(u)}}{a} \right) \right),$$

donde la última desigualdad se sigue de la independencia de los errores y las covariables. Usando **F1**, resulta que la desigualdad estricta ocurre en un conjunto con probabilidad positiva y se obtiene el resultado como en el Lema 2.1. ■

2.7.3. Modelo parcialmente lineal logístico

En el modelo parcialmente lineal logístico isotónico, suponemos que $y_i | (\mathbf{x}_i, t_i) = (\mathbf{x}_0, t_0) \sim \text{Bi}(1, \pi(\mathbf{x}_0, t_0))$ donde

$$\pi(\mathbf{x}, t) = H(\boldsymbol{\beta}_0^T \mathbf{x} + \eta_0(t))$$

con $H(s) = 1/(1 + \exp(s))$ y η_0 monótona. Para este modelo, la función de pérdida ρ dada en (2.11) puede escribirse como

$$\rho(y, u) = y \varphi(-\log(H(u))) + (1 - y) \varphi(-\log(1 - H(u))) + G(H(u)), \quad (2.19)$$

con $G(t) = G_1(t) + G_1(1 - t)$, $G_1(t) = \int_0^t \varphi'(-\log u) du$ y $H(u) = (1 + \exp(-u))^{-1}$.

Como en Boente *et al.* (2006), la consistencia de Fisher puede derivarse fácilmente para la función de pérdida dada en (2.11), cuando φ verifica las condiciones de regularidad enunciadas en Bianco y Yohai (1996), $w(\mathbf{x}) > 0$ para todo \mathbf{x} y $\mathbb{P}(\mathbf{x}^T \boldsymbol{\beta} = c | t = t_0) = 0$, para todo $(\boldsymbol{\beta}, c) \neq \mathbf{0}$. Más generalmente, como en los dos casos anteriores, supondremos que se cumple la condición **F1**.

Lema 2.3. *Sea $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}$ definida como en (2.19) donde la función $\varphi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ es tal que $\varphi(0) = 0$ y*

a) $\varphi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ es acotada con derivada continua y acotada φ' .

b) $\varphi'(t) \geq 0$ y existe para algún $c \geq \log 2$ tal que $\varphi'(t) > 0$ para todo $0 < t < c$.

*Supongamos además que se cumple **F1**, entonces $(\boldsymbol{\beta}_0, \eta_0)$ es el único mínimo de $L(\boldsymbol{\beta}, g)$ sobre $\mathbb{R}^p \times \mathcal{G}_0$.*

DEMOSTRACIÓN. La demostración es una consecuencia directa del Lema 2.1 de Bianco y Yohai (1996) y **F1**. Como en dicho Lema, sea z una variable aleatoria Bernoulli tal que $\mathbb{P}(z = 1) = \pi_0$ y definamos

$$M(\pi_0, \pi) = \mathbb{E}z\varphi(-\log \pi) + (1 - z)\varphi(-\log[1 - \pi]) + G(\pi).$$

Entonces, se cumple que $M(\pi_0, \pi_0) < M(\pi_0, \pi)$ para cualquier $\pi \neq \pi_0$. Tomando esperanza condicional y observando que $\mathbb{P}(y = 1 | (\mathbf{x}, t)) = H(\mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t))$ resulta que

$$\mathbb{E}\rho(y, \mathbf{x}^T \boldsymbol{\beta} + g(t))w(\mathbf{x}) = \mathbb{E}w(\mathbf{x})M(H(\mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t)), H(\mathbf{x}^T \boldsymbol{\beta} + g(t))).$$

Para un valor fijo de (\mathbf{x}, t) , sea $\pi = H(\mathbf{x}^T \boldsymbol{\beta} + g(t))$ y $\pi_0 = H(\mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t))$, la función $M(\pi_0, \pi)$ alcanza su único mínimo cuando $\pi = \pi_0$ y la demostración sigue inmediatamente a partir del supuesto **F1** como en el Lema 2.1. ■

Capítulo 3

Consistencia de los estimadores

En este capítulo, obtendremos, bajo ciertas condiciones de regularidad, resultados de consistencia y tasas de convergencia para los estimadores definidos en el Capítulo 2. Comenzaremos fijando la notación a utilizar.

Sea $\|\cdot\|$ la norma euclídea de \mathbb{R}^p y $\|f\|_2^2 = (\mathbb{E}f^2(t))^{1/2}$. Para cualquier función continua $v : \mathbb{R} \rightarrow \mathbb{R}$ denotamos por $\|v\|_\infty = \sup_t |v(t)|$ y por \mathcal{G} la clase de funciones monótonas no decrecientes en $[0, 1]$ definida en (2.3).

De ahora en adelante, cuando κ_0 sea desconocido, \mathcal{V} refiere a un entorno de κ_0 con clausura $\overline{\mathcal{V}}$ estrictamente incluida en \mathcal{K} . En caso contrario, cuando el parámetro de ruido es conocido, $\widehat{\kappa} = \kappa_0$ y $\mathcal{V} = \{\kappa_0\}$. Definimos la familia de funciones

$$\mathcal{F}_n = \{f(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}(t), a) w(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, a \in \mathcal{V}\}. \quad (3.1)$$

y para cualquier medida \mathbb{Q} consideramos $N(\epsilon, \mathcal{F}_n, L_r(\mathbb{Q}))$ y $N_{[\cdot]}(\epsilon, \mathcal{F}_n, L_r(\mathbb{Q}))$ los números de cubrimiento y cubrimiento corchete de la clase \mathcal{F}_n con respecto a la distancia en $L_r(\mathbb{Q})$, definidos como, por ejemplo, en van der Vaart y Wellner (1996) y cuyas nociones recordaremos en la Sección 3.1.

3.1. Nociones previas: Procesos empíricos

Con el objetivo de obtener resultados de consistencia y tasas de convergencia para los estimadores definidos en el Capítulo 2 y de distribución asintótica para el estimador de $\boldsymbol{\beta}_0$, necesitamos algunas definiciones y resultados sobre procesos empíricos. A los fines de completitud de esta tesis, presentamos en esta sección un resumen de estas nociones, en su mayoría, extraídos de van der Vaart y Wellner (1996), van der Geer (2000) y Kosorok (2008).

Sea (\mathcal{H}, d) un espacio semimétrico, es decir, \mathcal{H} es un espacio (posiblemente de dimensión infinita) y d define una semimétrica, o sea, $d : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ es tal que $d(x, y) \geq 0$, $d(x, y) = d(y, x)$ y $d(x, z) \leq d(x, y) + d(y, z)$. Esta situación incluye el caso de espacios normados. Indicaremos $B(x, \epsilon) = \{y \in \mathcal{H} : d(x, y) \leq \epsilon\}$.

Definición 3.1. Una ϵ -red para un subconjunto \mathcal{S} de \mathcal{H} es un conjunto de puntos $\{x_k\}_{k \in J}$ con la

propiedad de que para cada $x \in \mathcal{S}$ existe $k \in J$ tal que $x \in B(x_k, \epsilon)$, es decir, $\mathcal{S} \subset \cup_{j \in J} B(x_k, \epsilon)$.
 Obsérvese que los puntos x_k no necesariamente son elementos de \mathcal{S} .

Definición 3.2. Dado un subconjunto $\mathcal{S} \subset \mathcal{H}$ y $\epsilon > 0$, sea $N(\epsilon, \mathcal{S}, d)$ el mínimo número de bolas abiertas de radio ϵ necesarias para cubrir a \mathcal{S} . La ϵ -entropía de Kolmogorov del conjunto \mathcal{S} está dada por $\log(N(\epsilon, \mathcal{S}, d))$.

La noción de entropía fue introducida por Kolmogorov (ver, Kolmogorov y Tikhomirov, 1959) y representa una medida de la complejidad de un conjunto, en el sentido que, una alta entropía significa que se necesita mucha información para describir el conjunto con una exactitud ϵ . Por lo tanto, la elección de la estructura topológica juega un papel crucial en la búsqueda de resultados asintóticos. En la teoría de procesos empíricos, se utilizan mayoritariamente, las distancias generadas por las normas L_r respecto de una medida de probabilidad. Vamos a repasar esas nociones.

Sea X_1, \dots, X_n una muestra aleatoria de una distribución de probabilidad \mathbb{P} sobre un espacio medible $(\mathcal{X}, \mathcal{A})$. La distribución empírica es la medida uniforme discreta de las observaciones y la denotaremos por $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ donde δ_x es la masa puntual en x . Dada una función medible $f : \mathcal{X} \rightarrow \mathbb{R}$, indicamos por $P_n f$ a la esperanza de f bajo la medida empírica y por $\mathbb{P}f$ a la esperanza de f bajo \mathbb{P} , es decir,

$$P_n f = \frac{1}{n} \sum_{i=1}^n f(X_i), \quad \mathbb{P}f = \int f d\mathbb{P}.$$

El proceso empírico evaluado en f está definido como $\mathbb{G}_n f = \sqrt{n}(P_n f - \mathbb{P}f)$. El supremo de esos procesos estocásticos sobre una clase dada \mathcal{F} de funciones medibles se denotará por $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{G}_n f|$.

Dada una medida de probabilidad \mathbb{P} sobre $(\mathcal{X}, \mathcal{A})$, indicaremos por $L_r(\mathbb{P})$ el conjunto de funciones $g : \mathcal{X} \rightarrow \mathbb{R}$ tales que $\|g\|_{r, \mathbb{P}} = (\int |g|^r d\mathbb{P})^{1/r} = (\mathbb{E}_{\mathbb{P}}|g(X)|^r)^{1/r} < \infty$.

Definición 3.3. Una función envolvente de una clase \mathcal{F} es una función medible $F : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ tal que $|f(x)| \leq F(x)$ para toda $f \in \mathcal{F}$ y $x \in \mathcal{X}$.

Como sabemos, una forma relativamente simple de medir el tamaño de una clase \mathcal{F} es en términos de la entropía. Consideraremos la entropía corchete, o entropía con bracketing relativa a la norma $L_r(\mathbb{P})$.

Definición 3.4. Dadas dos funciones ℓ y u , el bracket o corchete $[\ell, u]$ es el conjunto de todas las funciones f tales que $\ell \leq f \leq u$. Más generalmente, dada $f \in \mathcal{F}$, diremos que pertenece al corchete $[\ell, u]$ si $\mathbb{P}(\ell(X) \leq f(X) \leq u(X)) = 1$. Un ϵ -corchete en $L_r(\mathbb{P})$ es un corchete $[\ell, u]$ tal que $\mathbb{P}(\ell(X) \leq u(X)) = 1$ y $\|u - \ell\|_{r, \mathbb{P}} < \epsilon$.

Definición 3.5. El número de cubrimiento corchete $N_{[\cdot]}(\epsilon, \mathcal{F}, L_r(\mathbb{P}))$ es el mínimo número de ϵ -corchetes en $L_r(\mathbb{P})$ necesarios para cubrir \mathcal{F} . La entropía corchete es el logaritmo del número de cubrimiento corchete.

Definición 3.6. La integral corchete está definida como

$$J_{[\]}(\delta, \mathcal{F}, L_2(\mathbb{P})) = \int_0^\delta \sqrt{1 + \log N_{[\]}(\epsilon, \mathcal{F}, L_2(\mathbb{P}))} d\epsilon.$$

En $L_r(\mathbb{P})$ definimos la ϵ -bola alrededor de una función $f \in L_r(\mathbb{P})$ como el conjunto $\{g \in L_r(\mathbb{P}) : \|g - f\|_{r, \mathbb{P}} < \epsilon\}$.

Definición 3.7. El número de cubrimiento indicado $N(\epsilon, \mathcal{F}, L_r(\mathbb{P}))$ es el menor número de ϵ -bolas en $L_r(\mathbb{P})$ necesarias para cubrir \mathcal{F} . La entropía es el logaritmo del número de cubrimiento.

Cabe notar que en la definición anterior los centros de las bolas no necesitan pertenecer a \mathcal{F} aunque duplicando el radio se puede suponer que de hecho son elementos de \mathcal{F} . Como se mencionó anteriormente, el número de cubrimiento se puede interpretar como la riqueza que tiene la clase \mathcal{F} a escala ϵ . Por otra parte, es fácil ver que $N(\epsilon, \mathcal{F}, L_r(\mathbb{P})) \leq N_{[\]}(2\epsilon, \mathcal{F}, L_r(\mathbb{P}))$. Para clases de funciones indexadas por un parámetro en un espacio semimétrico y que son Lipschitz en el mismo, el número de cubrimiento corchete puede mayorarse por el cubrimiento del espacio de parámetros. El siguiente resultado resume lo dicho.

Lema 3.1. Sea Θ un espacio semimétrico con semimétrica d y $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ una clase de funciones. Supongamos que existe una función $F : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$, $F \in L_r(\mathbb{P})$, tal que, para todo $\theta_1, \theta_2 \in \Theta$, $|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq d(\theta_1, \theta_2)F(x)$, entonces $N_{[\]}(2\epsilon\|F\|_{r, \mathbb{P}}, \mathcal{F}, L_r(\mathbb{P})) \leq N(\epsilon, \Theta, d)$.

A diferencia de lo que suele ocurrir en el contexto paramétrico, las clases con las que trabajaremos, como la clase \mathcal{F}_n definida en (3.1) no son clases fijas sino que varían con el tamaño de la muestra, por esta razón los resultados sobre procesos empíricos a utilizar deben contemplar esta particularidad.

Lema 3.2. Sea \mathcal{G}_n una colección de funciones, que depende de n , con envolvente $G_n = \sup_{g \in \mathcal{G}_n} |g|$. Supongamos que para alguna sucesión $b_n \geq 1$ tal que $b_n = o(n^{1/2})$ tenemos que para todo $\epsilon > 0$

$$\frac{b_n^2}{n} \log(N(\epsilon, \mathcal{G}_n, L_1(P_n))) \xrightarrow{P} 0$$

y

$$\limsup_{n \rightarrow \infty} \int_{G_n > b_n} G_n d\mathbb{P} = 0.$$

Entonces, $\sup_{g \in \mathcal{G}_n} |P_n g - \mathbb{P} f| \xrightarrow{P} 0$.

Definición 3.8. Sea \mathcal{D} una colección de subconjuntos de \mathcal{X} . Para $\xi_1, \dots, \xi_n \in \mathcal{X}$ definimos

$$\Delta^{\mathcal{D}}(\xi_1, \dots, \xi_n) = \text{card}\{D \cap \{\xi_1, \dots, \xi_n\} : D \in \mathcal{D}\},$$

es decir, $\Delta^{\mathcal{D}}(\xi_1, \dots, \xi_n)$ es el número de conjuntos diferentes de la forma $D \cap \{\xi_1, \dots, \xi_n\}$ con $D \in \mathcal{D}$. Más aun, definamos $m^{\mathcal{D}}(n) = \sup \{\Delta^{\mathcal{D}}(\xi_1, \dots, \xi_n) : \xi_1, \dots, \xi_n \in \mathcal{X}\}$ y

$$V(\mathcal{D}) = \inf \{n \geq 1 : m^{\mathcal{D}}(n) < 2^n\}.$$

$V(\mathcal{D})$ se llama el índice de la clase \mathcal{D} . Se dice que la colección \mathcal{D} es una clase de Vapnik-Chervonenkis (VC) si $V(\mathcal{D}) < \infty$.

Un semiespacio es un conjunto de la forma $\{x \in \mathbb{R}^d : \langle x, u \rangle \leq c\}$ para $u \in \mathbb{R}^d$ fijo y $c \in \mathbb{R}$.

Lema 3.3. a) La colección de semiespacios en \mathbb{R}^d es una clase VC de índice $d + 2$.

b) La colección de todas las bolas cerradas de \mathbb{R}^d es una clase VC de índice $d + 2$.

Definición 3.9. El subgrafo de una función $f : \mathcal{X} \rightarrow \mathbb{R}$ es el subconjunto de $\mathcal{X} \times \mathbb{R}$ dado por $\{(x, t) : t < f(x)\}$.

Sea \mathcal{F} una colección de funciones medibles sobre \mathcal{X} . La clase \mathcal{F} se dice una clase VC de funciones si la colección de todos los subgrafos de las funciones de \mathcal{F} forman una clase VC de conjuntos en $\mathcal{X} \times \mathbb{R}$.

Lema 3.4. Todo espacio vectorial \mathcal{F} de funciones medibles $f : \mathcal{X} \rightarrow \mathbb{R}$ de dimensión finita es una clase VC de índice menor o igual que $\dim(\mathcal{F}) + 2$.

Lema 3.5. Sean \mathcal{F} y \mathcal{G} clases VC de funciones en un conjunto \mathcal{X} con índices $V_{\mathcal{F}}$ y $V_{\mathcal{G}}$. Sean $g : \mathcal{X} \rightarrow \mathbb{R}$, $\phi : \mathbb{R} \rightarrow \mathbb{R}$ y $\psi : \mathcal{Z} \rightarrow \mathcal{X}$ funciones fijas. Entonces

a) $\mathcal{F} \wedge \mathcal{G} \equiv \{f \wedge g : f \in \mathcal{F}, g \in \mathcal{G}\}$ es una clase VC con índice menor o igual a $V_{\mathcal{F}} + V_{\mathcal{G}} - 1$;

b) $\mathcal{F} \vee \mathcal{G}$ es una clase VC con índice menor o igual a $V_{\mathcal{F}} + V_{\mathcal{G}} - 1$;

c) $\{\mathcal{F} > 0\} \equiv \{\{f > 0\} : f \in \mathcal{F}\}$ es una clase VC de conjuntos con índice $V_{\mathcal{F}}$;

d) $-\mathcal{F}$ es una clase VC con índice $V_{\mathcal{F}}$;

e) $\mathcal{F} + \mathcal{G} \equiv \{f + g : f \in \mathcal{F}\}$ es VC con índice $V_{\mathcal{F}}$;

f) $\mathcal{F} \cdot \mathcal{G} \equiv \{fg : f \in \mathcal{F}\}$ es VC con índice $2V_{\mathcal{F}} - 1$;

g) $\mathcal{F} \circ \psi \equiv \{f(\psi) : f \in \mathcal{F}\}$ es VC con índice menor o igual a $V_{\mathcal{F}}$;

h) $\phi \circ \mathcal{F}$ es VC con índice menor o igual a $V_{\mathcal{F}}$ para ϕ monótona;

Lema 3.6. Sea \mathcal{F} una clase VC de funciones con envolvente medible F y $r \geq 1$. Para cualquier medida de probabilidad \mathbb{Q} tal que $\|F\|_{r, \mathbb{Q}} > 0$, tenemos que

$$N(\epsilon \|F\|_{r, \mathbb{Q}}, \mathcal{F}, L_r(\mathbb{Q})) \leq K V(\mathcal{F}) (16e)^{V(\mathcal{F})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)},$$

para una constante universal K y $0 < \epsilon < 1$.

Lema 3.7. Para cada n , sean \mathbb{M}_n y M_n procesos estocásticos indexados por un conjunto Θ . Sea $\theta_n \in \Theta$ (posiblemente aleatorio) y $0 \leq \delta_n < \eta$ arbitrarios, y sea $\theta \mapsto d_n(\theta, \theta_n)$ una función arbitraria

(posiblemente aleatoria) de $\Theta_n \subset \Theta$ a $[0, \infty)$. Supongamos que, para cada n y $\delta_n < \delta \leq \eta$,

$$\sup_{\delta/2 < d_n(\theta, \theta_n) \leq \delta, \theta \in \Theta_n} M_n(\theta) - M_n(\theta_n) \leq -\delta^2$$

$$\mathbb{E}^* \sup_{\delta/2 < d_n(\theta, \theta_n) \leq \delta, \theta \in \Theta_n} \sqrt{n} [(\mathbb{M}_n - M_n)(\theta) - (\mathbb{M}_n - M_n)(\theta_n)]^+ \lesssim \phi_n(\delta),$$

para funciones ϕ_n tales que $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ es decreciente en (δ_n, η) , para algún $\alpha < 2$. Sea $r_n \lesssim \delta_n^{-1}$ tal que $r_n^2 \phi_n r_n^{-1} \leq \sqrt{n}$, para todo n . Si la sucesión $\hat{\theta}_n$ toma valores en Θ_n , satisface $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_n) - O_{\mathbb{P}}(r_n^{-2})$ y $d_n(\hat{\theta}_n, \theta_n)$ converge a cero en probabilidad exterior, entonces $r_n d_n(\hat{\theta}_n, \theta_n) = O_{\mathbb{P}}^*(1)$. Si las condiciones anteriores son válidas para $\eta = \infty$, entonces la condición de consistencia de $\hat{\theta}_n$ no es necesaria.

Lema 3.8. Sea \mathcal{F} una clase de funciones medibles tales que $\mathbb{P}f^2 < \delta^2$ y $\|f\|_\infty \leq M$ para toda $f \in \mathcal{F}$. Entonces

$$\mathbb{E}_{\mathbb{P}}^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[]}(\delta, \mathcal{F}, L_2(\mathbb{P})) \left(1 + \frac{J_{[]}(\delta, \mathcal{F}, L_2(\mathbb{P}))}{\delta^2 \sqrt{n}} M \right).$$

3.2. Resultados de consistencia

Para derivar resultados de consistencia de nuestra propuesta en el contexto considerado, necesitaremos los siguientes supuestos. Teniendo en cuenta el marco general que abordamos, su validez debe ser analizada para cada modelo en particular. En la Sección 3.5, presentamos algunos ejemplos de funciones de pérdida ρ y de modelos para los cuales estas hipótesis se cumplen.

C0 Los estimadores $\hat{\kappa}$ de κ_0 son fuertemente consistentes.

C1 $\rho(y, u, a)$ y $w(\cdot)$ son funciones acotadas no negativas y $\rho(y, u, a)$ es una función continua. Más aun, $L^*(\beta, \lambda, a) = L(\beta, \sum_{j=1}^{k_n} \lambda_j B_j, a)$ satisface la siguiente condición de equicontinuidad: para cualquier $\epsilon > 0$ existe $\delta > 0$ tal que para cualquier $a_1, a_2 \in \bar{\mathcal{V}}$,

$$|a_1 - a_2| < \delta \Rightarrow \sup_{\beta \in \mathbb{R}^k, \lambda \in \mathcal{L}_{k_n}} |L^*(\beta, \lambda, a_1) - L^*(\beta, \lambda, a_2)| < \epsilon.$$

C2 a) La función η_0 es no decreciente

b) La derivada r -ésima de η_0 satisface una condición Lipschitz en $[0, 1]$, con $r \geq 1$, es decir,

$$\eta_0 \in \mathcal{H}_r = \{g \in C^r[0, 1] : \|g^{(j)}\|_\infty \leq C_1, 0 \leq j \leq r \text{ y } |g^{(r)}(z_1) - g^{(r)}(z_2)| \leq C_2 |z_1 - z_2|\}.$$

C3 El máximo espaciado entre nodos es de orden $O(n^{-\nu})$, donde $0 < \nu < 1/2$. Más aun, el cociente entre el espaciado máximo y mínimo está acotado uniformemente.

C4 Para algún $\epsilon_1 > 0$, la clase de funciones \mathcal{F}_n definida en (3.1) es tal que, para $0 < \epsilon < \epsilon_1$, $\log N(\epsilon, \mathcal{F}_n, L_1(P_n)) = O_{\mathbb{P}}((k_n + p) \log(1/\epsilon))$.

Por simplicidad, denotemos como $L(\boldsymbol{\theta}_0, \kappa_0) = L(\boldsymbol{\beta}_0, \eta_0, \kappa_0)$, donde $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \eta_0)$ y como $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ a los estimadores definidos a través de (2.2) con $\widehat{\eta}(t) = \sum_{j=1}^{k_n} \widehat{\lambda}_j B_j(t)$. En forma similar definimos como $\widehat{\boldsymbol{\theta}}_{\text{INI}} = (\widehat{\boldsymbol{\beta}}_{\text{INI}}, \widehat{\eta}_{\text{INI}})$ a los estimadores sin restricciones definidos en (2.6).

Para medir la cercanía entre los estimadores y los parámetros poblacionales consideramos la métrica $\pi^2(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}) = \|\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}\|^2 + \|\eta_0 - \widehat{\eta}\|_{\mathcal{F}}^2$ donde $\|\cdot\|_{\mathcal{F}}$ representa una norma en el espacio de funciones $\mathcal{F} = \{g : [0, 1] \rightarrow \mathbb{R}, \text{ tal que } g \text{ es una función continua}\}$ que se elige usualmente como $\|f\|_2 = (\mathbb{E}f^2(t))^{1/2}$ o $\|f\|_{\infty} = \sup_{t \in [0, 1]} |f(t)|$. Sea $\mathcal{A}_{\epsilon} = \{\boldsymbol{\theta} = (\boldsymbol{\beta}, g) : \boldsymbol{\beta} \in \mathbb{R}^p, g \in \mathcal{G} \cap \mathcal{F}, \pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) > \epsilon\}$. El siguiente teorema establece la consistencia de los estimadores $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ basados en B-splines monótonos de orden $\ell \geq r + 2$.

Teorema 3.9 (Consistencia). *Sean $(y_i, \mathbf{x}_i, t_i)^T$ observaciones i.i.d. que satisfacen (1.1). Supongamos que se cumplen **C0** a **C4**, que para todo $\epsilon > 0$, $\inf_{\boldsymbol{\theta} \in \mathcal{A}_{\epsilon}} L(\boldsymbol{\theta}, \kappa_0) > L(\boldsymbol{\theta}_0, \kappa_0)$ y que $k_n = O(n^{\nu})$ con $1/(2r + 2) < \nu < 1/(2r)$. Entonces, se cumple que $\pi(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}) \xrightarrow{a.s.} 0$.*

Un resultado análogo puede obtenerse para los estimadores sin restricciones $\widehat{\boldsymbol{\theta}}_{\text{INI}} = (\widehat{\boldsymbol{\beta}}_{\text{INI}}, \widehat{\eta}_{\text{INI}})$. Para ello, bastará eliminar las restricciones de orden en los requerimientos anteriores, en particular, no será necesario pedir que η_0 sea monótona. Más precisamente, consideraremos los supuestos.

D1 $\rho(y, u, a)$ y $w(\cdot)$ son funciones acotadas no negativas y $\rho(y, u, a)$ es una función continua. Más aun, $L^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, a) = L(\boldsymbol{\beta}, \sum_{j=1}^{k_n} \lambda_j B_j, a)$ satisface la siguiente condición de equicontinuidad: para cualquier $\epsilon > 0$ existe $\delta > 0$ tal que para cualquier $a_1, a_2 \in \bar{\mathcal{V}}$,

$$|a_1 - a_2| < \delta \Rightarrow \sup_{\boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\lambda} \in \mathbb{R}^{k_n}} |L^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, a_1) - L^*(\boldsymbol{\beta}, \boldsymbol{\lambda}, a_2)| < \epsilon.$$

D4 Para algún $\epsilon_1 > 0$, la clase de funciones $\widetilde{\mathcal{F}}_n$ definida como

$$\widetilde{\mathcal{F}}_n = \{f(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}(t), a) w(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, a \in \mathcal{V}\}. \quad (3.2)$$

es tal que, para $0 < \epsilon < \epsilon_1$, $\log N(\epsilon, \widetilde{\mathcal{F}}_n, L_1(P_n)) = O_{\mathbb{P}}((k_n + p) \log(1/\epsilon))$.

Teorema 3.10. *Sean $(y_i, \mathbf{x}_i, t_i)^T$ observaciones i.i.d. que satisfacen (1.1). Supongamos que se cumplen **C0**, **D1**, **C2b**), **C3** y **D4**. Si para todo $\epsilon > 0$, $\inf_{\boldsymbol{\theta} \in \widetilde{\mathcal{A}}_{\epsilon}} L(\boldsymbol{\theta}, \kappa_0) > L(\boldsymbol{\theta}_0, \kappa_0)$ y $k_n = O(n^{\nu})$ con $1/(2r + 2) < \nu < 1/(2r)$ donde $\widetilde{\mathcal{A}}_{\epsilon} = \{\boldsymbol{\theta} = (\boldsymbol{\beta}, g) : \boldsymbol{\beta} \in \mathbb{R}^p, g \in \mathcal{F}, \pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) > \epsilon\}$, entonces se tiene que $\pi(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}_{\text{INI}}) \xrightarrow{a.s.} 0$.*

La demostración del Teorema 3.10 se omite pues se obtiene utilizando los mismos argumentos que en la demostración del Teorema 3.9, observando que, como $\ell \geq r + 2$, existe $g_n \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ tal que $\|g_n - \eta_0\|_{\infty} = O(n^{-r\nu})$, para $1/(2r + 2) < \nu < 1/(2r)$ (ver Schumaker, 1981), en lugar de usar la aproximación monótona de η_0 dada por el Lema A1 de Lu *et al.* (2007).

3.3. Tasas de convergencia

Para derivar las tasas de convergencia, tomamos como norma $\|\cdot\|_{\mathcal{F}}$ en el espacio de funciones \mathcal{F} , la norma $L_{\varphi}(\mathbb{Q})$, con $2 \leq \varphi \leq \infty$, donde $t \sim \mathbb{Q}$. Quedan incluidas como posibles normas

$\|f\|_{\mathcal{F}}^2 = \|f\|_2^2 = \mathbb{E}f^2(t)$ o $\|f\|_{\mathcal{F}} = \|f\|_{\infty}$, en cuyo caso $\pi^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^2 + \|\eta_1 - \eta_2\|_{\wp}^2$ con $\wp = 2$ o $\wp = \infty$, respectivamente. En este contexto, definimos la distancia

$$\pi_{\mathbb{P}}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbb{E} \left(w(\mathbf{x}) [\mathbf{x}^T(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) + \eta_1(t) - \eta_2(t)]^2 \right),$$

donde para $j = 1, 2$, $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j, \eta_j) \in \Theta = \mathbb{R}^p \times \mathcal{G}$.

Para obtener la tasa de convergencia de nuestros estimadores, necesitaremos algunas hipótesis adicionales. Entre ellas, pediremos dos condiciones distintas sobre la entropía corchete y dependiendo de ellas, las tasas de convergencia obtenidas serán más débiles o más fuertes. En la Sección 3.5, se discuten condiciones sobre la función ρ que garantizan la validez de estos supuestos para algunos modelos.

Para evitar pedir tasa de convergencia al estimador $\hat{\kappa}$ de κ_0 supondremos que $L(\boldsymbol{\beta}_0, \eta_0, a) < L(\boldsymbol{\beta}, g, a)$ para todo $a \in \mathcal{V}$ y para $(\boldsymbol{\beta}, g) \neq (\boldsymbol{\beta}_0, \eta_0)$, $g \in \mathcal{G}$. Esta condición implica la consistencia de Fisher y vale, por ejemplo, para el modelo parcialmente lineal tal como se mostró en la Sección 2.7. Una alternativa a esta hipótesis consistiría en requerir una condición de consistencia de Fisher condicional. Ambas opciones se aunan en el supuesto **C10**.

De ahora en adelante, para $\boldsymbol{\lambda} \in \mathbb{R}^{k_n}$, $g_{\boldsymbol{\lambda}}(t)$ referirá al spline $g_{\boldsymbol{\lambda}}(t) = \boldsymbol{\lambda}^T \mathbf{B}(t)$.

C5 Sea $\mathcal{G}_{n,c,\boldsymbol{\lambda}_0} = \{f(y, \mathbf{x}, t) = [\rho(y, \mathbf{x}^T \boldsymbol{\beta} + g_{\boldsymbol{\lambda}}(t), a) - \rho(y, \mathbf{x}^T \boldsymbol{\beta}_0 + g_{\boldsymbol{\lambda}_0}(t), a)] w(\mathbf{x}), \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon_0, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, a \in \mathcal{V}, \pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,\boldsymbol{\lambda}_0}, \boldsymbol{\theta}_{\boldsymbol{\lambda}}) \leq c\}$, con $\boldsymbol{\theta}_{0,\boldsymbol{\lambda}_0} = (\boldsymbol{\beta}_0, g_{\boldsymbol{\lambda}_0})$ y $\boldsymbol{\theta}_{\boldsymbol{\lambda}} = (\boldsymbol{\beta}, g_{\boldsymbol{\lambda}})$. Para alguna constante $C_2 > 0$ independiente de n , $\boldsymbol{\lambda}_0 \in \mathcal{L}_{k_n}$ y $\epsilon > 0$, tenemos que $N_{[\cdot]}(\epsilon, \mathcal{G}_{n,c,\boldsymbol{\lambda}_0}, L_2(P)) \leq C_2 (c/\epsilon)^{k_n+p+1}$.

C6 Para $n \geq n_0$, la familia de funciones $\mathcal{F}_{n,c}^* = \{f(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \boldsymbol{\beta} + g_{\boldsymbol{\lambda}}(t), a) w(\mathbf{x}), \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, a \in \mathcal{V}, \pi(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{\boldsymbol{\lambda}}) \leq c \text{ donde } \boldsymbol{\theta}_{\boldsymbol{\lambda}} = (\boldsymbol{\beta}, g_{\boldsymbol{\lambda}})\}$ es tal que, para cualquier $0 < \epsilon < 1$, $N_{[\cdot]}(\epsilon, \mathcal{F}_{n,c}^*, L_2(P)) \leq C_2/\epsilon^{k_n+p+1}$, para alguna constante $C_2 > 0$ independiente de n y ϵ .

C7 La función ρ es dos veces continuamente diferenciable con respecto a su segundo argumento con derivadas $\Psi(y, u, a) = \partial \rho(y, u, a)/\partial u$ y $\chi(y, u, a) = \partial \Psi(y, u, a)/\partial u$ tales que

$$\|\Psi\|_{\infty, \mathcal{V}} = \sup_{y \in \mathbb{R}, u \in \mathbb{R}, a \in \mathcal{V}} |\Psi(y, u, a)| < \infty \text{ y } \|\chi\|_{\infty, \mathcal{V}} = \sup_{y \in \mathbb{R}, u \in \mathbb{R}, a \in \mathcal{V}} |\chi(y, u, a)| < \infty.$$

C8 $\mathbb{E}w(\mathbf{x}) \|\mathbf{x}\|^2 < \infty$.

C9 Existe $\epsilon_0 > 0$, un entorno \mathcal{V} de κ_0 y una constante positiva C_0 , tal que para cualquier $\boldsymbol{\theta} \in \mathbb{R}^p \times \mathcal{M}_n(\mathcal{T}_n, \ell)$ con $\pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \epsilon_0$ y cualquier $a \in \mathcal{V}$, $L(\boldsymbol{\theta}, a) - L(\boldsymbol{\theta}_0, a) \geq C_0 \pi_{\mathbb{P}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$.

C10 Una de las siguientes afirmaciones se cumple

- a) para cualquier $a \in \mathcal{V}$, $L(\boldsymbol{\beta}_0, \eta_0, a) < L(\boldsymbol{\beta}, g, a)$ para $(\boldsymbol{\beta}, g) \neq (\boldsymbol{\beta}_0, \eta_0)$, $g \in \mathcal{G}$.
- b) $\mathbb{E} \{\Psi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t), a) | (\mathbf{x}, t)\} = 0$, casi seguramente, para cualquier $a \in \mathcal{V}$.

Teorema 3.11 (Tasas de convergencia). *Sean $(y_i, \mathbf{x}_i, t_i)^T$ observaciones i.i.d. que satisfacen (1.1) y $k_n = O(n^\nu)$ para $1/(2r+2) < \nu < 1/(2r)$. Supongamos que **C1** a **C3** y **C7** a **C10** se cumplen y que $\pi(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$. Entonces, tenemos que*

- a) si se cumple **C5**, $\gamma_n \pi_{\mathbb{P}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$, donde $\gamma_n = n^{\min(r\nu, (1-\nu)/2)}$. Por lo tanto, si $\nu = 1/(1+2r)$, los estimadores convergen con tasa óptima $n^{r/(1+2r)}$.
- b) si se cumple **C6**, $\gamma_n \pi_{\mathbb{P}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$, para cualquier γ_n , tal que $\gamma_n \leq O(n^{r\nu})$ y $\gamma_n \log(\gamma_n) \leq O(n^{(1-\nu)/2})$.

Observación 3.1. Cabe mencionar que bajo **C5**, el hecho que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(n^{-\min(r\nu, (1-\nu)/2)})$ junto con el Lema 7 de Stone (1987) implican que $\|\widehat{\eta} - \eta_0\|_{\infty} \xrightarrow{p} 0$, dando un resultado similar al obtenido en el Teorema 1 de Lu (2015).

Observemos que la condición **C10b**) es análoga a la consistencia de Fisher condicional de Künsch *et al.* (1989), mientras que la condición **C5** es similar al supuesto C3' de Shen y Wong (1994). El Lema 3.20 de la Sección 3.5 da condiciones bajo las cuales se cumple **C9**. ♣

Como en la Sección 3.2, el Teorema 3.12 permite obtener tasas de convergencia para los estimadores $\widehat{\boldsymbol{\theta}}_{\text{INI}} = (\widehat{\boldsymbol{\beta}}_{\text{INI}}, \widehat{\eta}_{\text{INI}})$ basados en B-splines sin imponer restricciones de monotonía. Como antes, se omite su demostración pues sigue los mismos pasos que los considerados en la demostración del Teorema 3.11 utilizando nuevamente que existe $g_n \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ tal que $\|g_n - \eta_0\|_{\infty} = O(n^{-r\nu})$, para $1/(2r+2) < \nu < 1/(2r)$. Para ello, debemos eliminar las restricciones de orden en los requerimientos anteriores. Más precisamente, consideraremos los supuestos.

- D5** Sea $\widetilde{\mathcal{G}}_{n,c,\lambda_0} = \{f(y, \mathbf{x}, t) = [\rho(y, \mathbf{x}^T \boldsymbol{\beta} + g_{\lambda}(t), a) - \rho(y, \mathbf{x}^T \boldsymbol{\beta}_0 + g_{\lambda_0}(t), a)] w(\mathbf{x}), \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon_0, \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, a \in \mathcal{V}, \pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,\lambda_0}, \boldsymbol{\theta}_{\lambda}) \leq c\}$, con $\boldsymbol{\theta}_{0,\lambda_0} = (\boldsymbol{\beta}_0, g_{\lambda_0})$ y $\boldsymbol{\theta}_{\lambda} = (\boldsymbol{\beta}, g_{\lambda})$. Para alguna constante $C_2 > 0$ independiente de n , $\boldsymbol{\lambda}_0 \in \mathbb{R}^{k_n}$ y $\epsilon > 0$, tenemos que $N_{[\cdot]}(\epsilon, \widetilde{\mathcal{G}}_{n,c,\lambda_0}, L_2(P)) \leq C_2 (c/\epsilon)^{k_n+p+1}$.
- D6** Para $n \geq n_0$, la familia de funciones $\widetilde{\mathcal{F}}_{n,c}^* = \{f(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \boldsymbol{\beta} + g_{\lambda}(t), a) w(\mathbf{x}), \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, a \in \mathcal{V}, \pi(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{\lambda}) \leq c\}$ donde $\boldsymbol{\theta}_{\lambda} = (\boldsymbol{\beta}, g_{\lambda})$ es tal que, para cualquier $0 < \epsilon < 1$, $N_{[\cdot]}(\epsilon, \widetilde{\mathcal{F}}_{n,c}^*, L_2(P)) \leq C_2/\epsilon^{k_n+p+1}$, para alguna constante $C_2 > 0$ independiente de n y ϵ .
- D9** Existe $\epsilon_0 > 0$, un entorno \mathcal{V} de κ_0 y una constante positiva C_0 , tal que para cualquier $\boldsymbol{\theta} \in \mathbb{R}^p \times \mathcal{S}_n(\mathcal{T}_n, \ell)$ con $\pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \epsilon_0$ y cualquier $a \in \mathcal{V}$, $L(\boldsymbol{\theta}, a) - L(\boldsymbol{\theta}_0, a) \geq C_0 \pi_{\mathbb{P}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$.

D10 Una de las siguientes afirmaciones se cumple

- a) para cualquier $a \in \mathcal{V}$, $L(\boldsymbol{\beta}_0, \eta_0, a) < L(\boldsymbol{\beta}, g, a)$ para $(\boldsymbol{\beta}, g) \neq (\boldsymbol{\beta}_0, \eta_0)$, $g \in \mathcal{G}_0$.
- b) $\mathbb{E} \{\Psi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t), a) | (\mathbf{x}, t)\} = 0$, casi seguramente, para cualquier $a \in \mathcal{V}$.

Observemos que **D5** y **D6** corresponden a las hipótesis **C5** y **C6** pero sin exigir que los coeficientes tengan una restricción de orden. Lo mismo ocurre con **D9** y **D10a**) que son las versiones adaptadas al caso sin restricciones de **C9** y **C10a**). Más aun, **D10a**) implica que $\mathbb{E} \{\Psi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t), a) | t\} = 0$, casi seguramente, para cualquier $a \in \mathcal{V}$, por lo tanto, la consistencia de Fisher condicional dada en **D10b**) es una condición algo más fuerte que **D10a**).

Teorema 3.12. Sean $(y_i, \mathbf{x}_i, t_i)^T$ observaciones i.i.d. que satisfacen (1.1) y $k_n = O(n^{\nu})$ para $1/(2r+2) < \nu < 1/(2r)$. Supongamos que valen **C1**, **C2b**), **C3**, **C7**, **C8**, **D9** y **D10** y que $\pi(\widehat{\boldsymbol{\theta}}_{\text{INI}}, \boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$. Entonces, tenemos que

- a) si se cumple **D5**, $\gamma_n \pi_{\mathbb{P}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}_{\text{INI}}) = O_{\mathbb{P}}(1)$, donde $\gamma_n = n^{\min(r\nu, (1-\nu)/2)}$. Luego, si $\nu = 1/(1+2r)$, los estimadores convergen con tasa óptima $n^{r/(1+2r)}$.
- b) si se cumple **D6**, $\gamma_n \pi_{\mathbb{P}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}_{\text{INI}}) = O_{\mathbb{P}}(1)$, para cualquier γ_n , tal que $\gamma_n \leq O(n^{r\nu})$ y $\gamma_n \log(\gamma_n) \leq O(n^{(1-\nu)/2})$.

Como se mencionó en la Observación 3.1, bajo **D5**, se obtiene que $\|\widehat{\eta}_{\text{INI}} - \eta_0\|_{\infty} \xrightarrow{p} 0$. Por lo tanto, si η_0 es una función no decreciente y $\widehat{\eta}_{\text{inc}}$ es el estimador definido en (2.7), del Teorema 3.1 (a) de Neumeyer (2007) se deduce que $\|\widehat{\eta}_t - \eta_0\|_{\infty} \xrightarrow{p} 0$. Por otra parte, si además η_0 es derivable, $\inf_{0 \leq t \leq 1} \eta_0'(t) > 0$ y la cantidad de puntos donde se calcula la discretización $N \rightarrow \infty$, entonces $\|\widehat{\eta}_{\text{I,D}} - \eta_0\|_{\infty} \xrightarrow{p} 0$.

3.4. Apéndice A: Demostración de los Teoremas 3.9 y 3.11

En este apéndice se presentan las demostraciones de los teoremas que establecen la consistencia de los estimadores y su tasa de convergencia. Llamaremos $\|\rho\|_{\infty} = \sup_{y \in \mathbb{R}, u \in \mathbb{R}, a \in \mathcal{V}} \rho(y, u, a)$ y $\|w\|_{\infty} = \sup_{\mathbf{x} \in \mathbb{R}^p} w(\mathbf{x})$.

3.4.1. Demostración del Teorema 3.9

Sea $V_{\boldsymbol{\beta}, g, a} = \rho(y, \mathbf{x}^T \boldsymbol{\beta} + g(t), a) w(\mathbf{x})$ y llamemos P a la medida de probabilidad de (y, \mathbf{x}, t) y P_n a su correspondiente medida empírica. Luego, $L_n(\boldsymbol{\beta}, g, a) = P_n V_{\boldsymbol{\beta}, g, a}$ y $L(\boldsymbol{\beta}, g, a) = P V_{\boldsymbol{\beta}, g, a}$.

La consistencia de $\widehat{\kappa}$ implica que dado cualquier entorno \mathcal{V} de κ_0 , existe un conjunto de medida nula $\mathcal{N}_{\mathcal{V}}$ tal que, para todo $\omega \notin \mathcal{N}_{\mathcal{V}}$, $\widehat{\kappa} \in \mathcal{V}$ para todo $n \geq n_0$ y algún $n_0 \in \mathbb{N}$.

La demostración sigue pasos similares a los utilizados para el Teorema 5.7 of van der Vaart (1998). Comenzaremos mostrando que

$$A_n = \sup_{\boldsymbol{\beta} \in \mathbb{R}^p, g \in \mathcal{M}_n(\mathcal{T}_n, \ell), a \in \mathcal{V}} |L_n(\boldsymbol{\beta}, g, a) - L(\boldsymbol{\beta}, g, a)| \xrightarrow{a.s.} 0. \quad (3.3)$$

Notemos que $A_n = \sup_{f \in \mathcal{F}_n} (P_n - P)f$, donde \mathcal{F}_n está definida en (3.1). Más aun, **C1** implica que $\sup_{f \in \mathcal{F}_n} |f| = \|\rho\|_{\infty} \|w\|_{\infty}$, mientras que **C4** y el hecho de que $k_n = O(n^{\nu})$ con $\nu < 1/(2r) < 1$ implica que

$$\frac{1}{n} \log N(\epsilon, \mathcal{F}_n, L_1(P_n)) = O_{\mathbb{P}}(1) \frac{k_n + p}{n} \log \left(\frac{1}{\epsilon} \right) \xrightarrow{p} 0.$$

Por lo tanto, utilizando el Lema 3.2 con $b_n = \max(1, 2 \|\rho\|_{\infty} \|w\|_{\infty})$, deducimos que (3.3) vale.

Como $L(\boldsymbol{\theta}_0, \kappa_0) = \inf_{\boldsymbol{\beta} \in \mathbb{R}^p, g \in \mathcal{G}} L(\boldsymbol{\beta}, g, \kappa_0)$, donde $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \eta_0)$, tenemos que

$$0 \leq L(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}_0) - L(\boldsymbol{\theta}_0, \kappa_0) = \sum_{j=1}^3 A_{n,j}, \quad (3.4)$$

con $A_{n,1} = L(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) - L_n(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$, $A_{n,2} = L_n(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) - L(\boldsymbol{\theta}_0, \kappa_0)$ y $A_{n,3} = L(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}_0) - L(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$. Notando que $|A_{n,1}| \leq A_n$, obtenemos que $A_{n,1} = o_{\text{a.s.}}(1)$. Por otro lado, como $L(\widehat{\boldsymbol{\theta}}, a) = L^*(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda}}, a)$ la equicontinuidad de L^* pedida en **C1** y la consistencia de $\widehat{\kappa}$ implican que $A_{n,3} = o_{\text{a.s.}}(1)$.

Acotaremos ahora $A_{n,2}$. Usando el Lema A1 de Lu *et al.* (2007), tenemos que existe $g_n \in \mathcal{M}_n(\mathcal{T}_n, \ell)$, tal que $\|g_n - \eta_0\|_\infty = O(n^{-r\nu})$, para $1/(2r+2) < \nu < 1/(2r)$. Sea $\boldsymbol{\theta}_{0,n} = (\boldsymbol{\beta}_0, g_n)$, $S_{n,1} = (P_n - P)V_{\boldsymbol{\beta}_0, g_n, \hat{\kappa}}$ y $S_{n,2} = L(\boldsymbol{\theta}_{0,n}, \hat{\kappa}) - L(\boldsymbol{\theta}_0, \kappa_0)$. Notemos que $S_{n,1} \leq A_n$, de manera que a partir de (3.3), concluimos que $S_{n,1} \xrightarrow{a.s.} 0$. Por otra parte, si escribimos $S_{n,2} = \sum_{j=1}^2 S_{n,2}^{(j)}$ donde $S_{n,2}^{(1)} = L(\boldsymbol{\theta}_{0,n}, \hat{\kappa}) - L(\boldsymbol{\theta}_{0,n}, \kappa_0)$ y $S_{n,2}^{(2)} = L(\boldsymbol{\theta}_{0,n}, \kappa_0) - L(\boldsymbol{\theta}_0, \kappa_0)$, la continuidad de ρ junto con el hecho de que $\|g_n - \eta_0\|_\infty \rightarrow 0$ y el Teorema de Convergencia Dominada implican que $S_{n,2}^{(2)} \rightarrow 0$, mientras la continuidad de ρ y el hecho que es acotada junto con la consistencia $\hat{\kappa}$ llevan a que $S_{n,2}^{(1)} = o_{a.s.}(1)$. Resumiendo, hemos mostrado que $S_{n,j} = o_{a.s.}(1)$ para $j = 1, 2$.

Usando que $\hat{\boldsymbol{\theta}}$ minimiza L_n en $\mathbb{R}^p \times \mathcal{M}_n(\mathcal{T}_n, \ell)$ obtenemos que

$$A_{n,2} = L_n(\hat{\boldsymbol{\theta}}, \hat{\kappa}) - L(\boldsymbol{\theta}_0, \kappa_0) \leq L_n(\boldsymbol{\theta}_n, \hat{\kappa}) - L(\boldsymbol{\theta}_0, \kappa_0) = S_{n,1} + S_{n,2}. \quad (3.5)$$

Por lo tanto, usando que $A_{n,j} = o_{a.s.}(1)$, para $j = 1, 3$ y $S_{n,j} = o_{a.s.}(1)$, para $j = 1, 2$, (3.4) y (3.5) nos permiten concluir que

$$0 \leq L(\hat{\boldsymbol{\theta}}, \kappa_0) - L(\boldsymbol{\theta}_0, \kappa_0) = \sum_{j=1}^3 A_{n,j} \leq o_{a.s.}(1),$$

de donde obtenemos que $L(\hat{\boldsymbol{\theta}}, \kappa_0) \xrightarrow{a.s.} L(\boldsymbol{\theta}_0, \kappa_0)$ que junto con $\inf_{\boldsymbol{\theta} \in \mathcal{A}_\epsilon} L(\boldsymbol{\theta}, \kappa_0) > L(\boldsymbol{\theta}_0, \kappa_0)$ implican que $\pi(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$, concluyendo la demostración. \blacksquare

3.4.2. Demostración del Teorema 3.11

Definamos las funciones $M_1(s) = L(\boldsymbol{\beta}_0 + s\boldsymbol{\beta}, \eta_0, a)$ y $M_2(s) = L(\boldsymbol{\beta}_0, \eta_0 + sg, a)$ y notemos que $M_1'(0) = \mathbb{E}[w(\mathbf{x})\Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), a)\mathbf{x}^\top \boldsymbol{\beta}]$ y $M_2'(0) = \mathbb{E}[w(\mathbf{x})\Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), a)g(t)]$. Cuando se cumple **C10a**, $M_1(s)$ y $M_2(s)$ tienen un mínimo en $s = 0$, para cualquier $\boldsymbol{\beta} \in \mathbb{R}^p$ y $g \in \mathcal{G}$. Luego, $M_1'(0) = 0$ y $M_2'(0) = 0$, es decir, para todo $a \in \mathcal{V}$,

$$\mathbb{E}[w(\mathbf{x})\Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), a)\mathbf{x}] = \mathbf{0} \quad (3.6)$$

$$\mathbb{E}[w(\mathbf{x})\Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), a)g(t)] = 0 \quad \text{para todo } g \in \mathcal{G}. \quad (3.7)$$

Claramente, (3.6) y (3.7) también valen bajo **C10b**.

Para probar el Teorema 3.11 bajo ambos conjuntos de supuestos enunciaremos los pasos comunes al comienzo y luego continuaremos la demostración según se verifique **C5** o **C6**.

Sea $\Theta_n = \mathbb{R}^p \times \mathcal{M}_n(\mathcal{T}_n, \ell) \cap \{\boldsymbol{\theta} = (\boldsymbol{\beta}, g) \in \Theta : \pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \epsilon_0\}$, donde $\Theta = \mathbb{R}^p \times \mathcal{G}$. Salvo por un conjunto de probabilidad nula, $\hat{\boldsymbol{\theta}} \in \Theta_n$ para n suficientemente grande. Como en la demostración del Teorema 3.9, sea $g_n \in \mathcal{M}_n(\mathcal{T}_n, \ell)$, $g_n(t) = \boldsymbol{\lambda}_n^\top \mathbf{B}(t)$, tal que $\|g_n - \eta_0\|_\infty = O(n^{-r\nu})$ para $1/(2r+2) < \nu < 1/(2r)$, y llamemos $\boldsymbol{\theta}_{0,n} = (\boldsymbol{\beta}_0, g_n)$.

Para obtener la tasa de convergencia del estimador $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\eta})$ aplicaremos el Lema 3.7 que corresponde al Teorema 3.4.1 de van der Vaart y Wellner (1996). Siguiendo la notación de dicho Lema, llamemos $M(\boldsymbol{\theta}) = -L(\boldsymbol{\theta}, \hat{\kappa})$ y $\mathbb{M}_n(\boldsymbol{\theta}) = -L_n(\boldsymbol{\theta}, \hat{\kappa})$ y, para $\boldsymbol{\theta} \in \Theta_n$, sea $d_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \pi_{\mathbb{P}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. La función M es aleatoria debido al estimador $\hat{\kappa}$ del parámetro de ruido. Sea $\delta_n = A\|\eta_0 - g_n\|_{\mathcal{F}}$, donde $A = 4\sqrt{(C_0/\|w\|_\infty + A_0)/C_0}$ con $A_0 = \|w\|_\infty\|\chi\|_\infty/2$ y C_0 dados en **C9**.

Usando que $|(L_n(\boldsymbol{\theta}, \widehat{\kappa}) - L(\boldsymbol{\theta}, \widehat{\kappa})) - (L_n(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}) - L(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}))| = |(\mathbb{M}_n - M)(\boldsymbol{\theta}) - (\mathbb{M}_n - M)(\boldsymbol{\theta}_{0,n})|$, para utilizar el Lema 3.7, deberemos mostrar que existe una función ϕ_n tal que $\phi_n(\delta)/\delta^\nu$ es decreciente en (δ_n, ∞) para algún $\nu < 2$ y que para cualquier $\delta > \delta_n$,

$$\sup_{\boldsymbol{\theta} \in \Theta_{n,\delta}} L(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}) - L(\boldsymbol{\theta}, \widehat{\kappa}) = \sup_{\boldsymbol{\theta} \in \Theta_{n,\delta}} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_{0,n}) \lesssim -\delta^2 \quad (3.8)$$

$$\mathbb{E}^* \sup_{\boldsymbol{\theta} \in \Theta_{n,\delta}} \sqrt{n} |(L_n(\boldsymbol{\theta}, \widehat{\kappa}) - L(\boldsymbol{\theta}, \widehat{\kappa})) - (L_n(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}) - L(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}))| \lesssim \phi_n(\delta) \quad (3.9)$$

$$d_n(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_{0,n}) \xrightarrow{p} 0 \quad (3.10)$$

donde el símbolo \lesssim significa *menor o igual salvo un factor constante*, \mathbb{E}^* es la esperanza exterior y $\Theta_{n,\delta} = \{\boldsymbol{\theta} \in \Theta_n : \delta/2 < d_n(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) \leq \delta\}$.

El supuesto **C9** y el hecho de que $\widehat{\kappa} \xrightarrow{a.s.} \kappa_0$ implican que, salvo un conjunto de probabilidad nula, para cualquier $\boldsymbol{\theta} \in \Theta_n$, $L(\boldsymbol{\theta}, \widehat{\kappa}) - L(\boldsymbol{\theta}_0, \widehat{\kappa}) \geq C_0 \pi_{\mathbb{P}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. Por otro lado, usando (3.7) obtenemos que $\mathbb{E}[w(\mathbf{x})\Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), a)(g_n(t) - \eta_0(t))] = 0$, de donde

$$\begin{aligned} L(\boldsymbol{\theta}_{0,n}, a) - L(\boldsymbol{\theta}_0, a) &= \mathbb{E}[w(\mathbf{x})\Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), a)(g_n(t) - \eta_0(t))] \\ &\quad + \frac{1}{2} \mathbb{E}\left[w(\mathbf{x})\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \tilde{\eta}(t), a)(g_n(t) - \eta_0(t))^2\right] \\ &= \frac{1}{2} \mathbb{E}\left[w(\mathbf{x})\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \tilde{\eta}(t), a)(g_n(t) - \eta_0(t))^2\right] \\ &\leq \frac{1}{2} \|w\|_\infty \|\chi\|_\infty \mathbb{E}(g_n(t) - \eta_0(t))^2 = A_0 \|g_n - \eta_0\|_2^2 \leq A_0 \|g_n - \eta_0\|_{\mathcal{F}}^2 = O(n^{-2r\nu}), \end{aligned}$$

con $A_0 = \|w\|_\infty \|\chi\|_\infty / 2$ y $\tilde{\eta}(t)$ un valor intermedio entre $\eta_0(t)$ y $g_n(t)$. Luego, usando que $d_n^2(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) \leq 2d_n^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + 2d_n^2(\boldsymbol{\theta}_{0,n}, \boldsymbol{\theta}_0) \leq 2d_n^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + 2\|w\|_\infty \|g_n - \eta_0\|_2^2 \leq 2d_n^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + 2\|w\|_\infty \|g_n - \eta_0\|_{\mathcal{F}}^2$ y que $\delta/2 < d_n(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n})$ obtenemos que

$$\begin{aligned} L(\boldsymbol{\theta}, \widehat{\kappa}) - L(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}) &\geq C_0 d_n^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) - A_0 \|g_n - \eta_0\|_{\mathcal{F}}^2 \geq \frac{C_0}{2} d_n^2(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) - \left(\frac{C_0}{\|w\|_\infty} + A_0\right) \|g_n - \eta_0\|_{\mathcal{F}}^2 \\ &\geq \frac{C_0}{8} \delta^2 - \frac{1}{A^2} \left(\frac{C_0}{\|w\|_\infty} + A_0\right) \delta_n^2 = \frac{C_0}{8} \delta^2 - \frac{C_0}{16} \delta_n^2 \geq \frac{C_0}{16} \delta^2, \end{aligned}$$

concluyendo la demostración de (3.8).

En el siguiente paso, debemos hallar $\phi_n(\delta)$ tal que $\phi_n(\delta)/\delta$ sea decreciente en δ y se cumpla (3.9). Por la consistencia de $\widehat{\kappa}$, con probabilidad 1 para n suficientemente grande

$$\begin{aligned} \sqrt{n} |(L_n(\boldsymbol{\theta}, \widehat{\kappa}) - L(\boldsymbol{\theta}, \widehat{\kappa})) - (L_n(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}) - L(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}))| &\leq \\ &\sup_{a \in \mathcal{V}} \sqrt{n} |(L_n(\boldsymbol{\theta}, a) - L(\boldsymbol{\theta}, a)) - (L_n(\boldsymbol{\theta}_{0,n}, a) - L(\boldsymbol{\theta}_{0,n}, a))|. \end{aligned}$$

Definamos la clase de funciones

$$\mathcal{F}_{n,\delta} = \{V_{\boldsymbol{\theta},a} - V_{\boldsymbol{\theta}_{0,n},a} : \frac{\delta}{2} \leq d_n(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) \leq \delta, \boldsymbol{\theta} \in \Theta_n, a \in \mathcal{V}\} = \{V_{\boldsymbol{\theta},a} - V_{\boldsymbol{\theta}_{0,n},a} : \boldsymbol{\theta} \in \Theta_{n,\delta}, a \in \mathcal{V}\},$$

con $V_{\boldsymbol{\theta},a} = \rho(y, \mathbf{x}^\top \boldsymbol{\beta} + g(t), a) w(\mathbf{x})$, para $\boldsymbol{\theta} = (\boldsymbol{\beta}, g)$. La desigualdad (3.9) involucra un proceso empírico indexado por $\mathcal{F}_{n,\delta}$, ya que

$$\mathbb{E}^* \sup_{\boldsymbol{\theta} \in \Theta_{n,\delta}} \sqrt{n} |(L_n(\boldsymbol{\theta}, \widehat{\kappa}) - L(\boldsymbol{\theta}, \widehat{\kappa})) - (L_n(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}) - L(\boldsymbol{\theta}_{0,n}, \widehat{\kappa}))| \leq \mathbb{E}^* \sup_{f \in \mathcal{F}_{n,\delta}} \sqrt{n} |(P_n - P)f|.$$

Para cualquier $f \in \mathcal{F}_{n,\delta}$ tenemos que $\|f\|_\infty \leq A_1 = 2\|\rho\|_\infty\|w\|_\infty$. Más aun, si $A_2 = \|\psi\|_\infty\|w\|_\infty$ usando que

$$|V_{\theta,a} - V_{\theta_{0,n},a}| \leq \|\psi\|_\infty w(\mathbf{x}) |\mathbf{x}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + g(t) - g_n(t)|,$$

y que $\pi_{\mathbb{P}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) = d_n(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) \leq \delta$, obtenemos que

$$Pf^2 \leq \|\psi\|_\infty \mathbb{E} \left(w^2(\mathbf{x}) [\mathbf{x}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + g(t) - g_n(t)]^2 \right) \leq A_2 \pi_{\mathbb{P}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) \leq A_2 \delta^2.$$

El Lema 3.8 permite obtener la cota

$$\mathbb{E}^* \sup_{f \in \mathcal{F}_{n,\delta}} \sqrt{n} |(P_n - P)f| \leq J_{[\cdot]} \left(A_2^{1/2} \delta, \mathcal{F}_{n,\delta}, L_2(P) \right) \left(1 + A_1 \frac{J_{[\cdot]}(A_2^{1/2} \delta, \mathcal{F}_{n,\delta}, L_2(P))}{A_2 \delta^2 \sqrt{n}} \right),$$

donde $J_{[\cdot]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[\cdot]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon$ es la integral corchete.

a) Supongamos ahora que **C5** se cumple. Para cualquier $\boldsymbol{\theta} = (\boldsymbol{\beta}, g) \in \Theta_{n,\delta}$, g puede escribirse como $g = \boldsymbol{\lambda}^\top \mathbf{B}$ para algún $\boldsymbol{\lambda} \in \mathcal{L}_{k_n}$, entonces

$$d_n^2(\boldsymbol{\theta}, \boldsymbol{\theta}_{0,n}) = \mathbb{E} \left(w(\mathbf{x}) [\mathbf{x}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + (\boldsymbol{\lambda} - \boldsymbol{\lambda}_n)^\top \mathbf{B}(t)]^2 \right).$$

Por lo tanto, $\mathcal{F}_{n,\delta} \subset \mathcal{G}_{n,c,\boldsymbol{\lambda}_n}$ con $c = \delta$ y la cota dada en **C5** conlleva a

$$N_{[\cdot]}(\epsilon, \mathcal{F}_{n,\delta}, L_2(P)) \leq C_2 \left(\frac{\delta}{\epsilon} \right)^{k_n + p + 1},$$

lo que implica que

$$J_{[\cdot]}(A_2^{1/2} \delta, \mathcal{F}_{n,\delta}, L_2(P)) \lesssim \delta \sqrt{k_n + p + 1}.$$

Si llamamos $q_n = k_n + p + 1$, obtenemos que

$$\mathbb{E}^* \sup_{\boldsymbol{\theta} \in \Theta_{n,\delta}} |\mathbb{G}_n V_{\boldsymbol{\theta}_{0,n}, \kappa_0} - \mathbb{G}_n V_{\boldsymbol{\theta}, \kappa_0}| \leq A_3 \left[\delta q_n^{1/2} + \frac{q_n}{\sqrt{n}} \right],$$

para alguna constante A_3 independiente de n y δ . Eligiendo

$$\phi_n(\delta) = \delta q_n^{1/2} + \frac{q_n}{\sqrt{n}},$$

tenemos que $\phi_n(\delta)/\delta$ es decreciente en δ , concluyendo la demostración de (3.9). El hecho de que $\pi(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$ implica que $\pi_{\mathbb{P}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$ de donde, utilizando que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,n}, \boldsymbol{\theta}_0) \rightarrow 0$, se deduce que (3.10) vale.

Sea $\gamma_n = O(n^{\min(r\nu, (1-\nu)/2)})$, entonces $\gamma_n \lesssim \delta_n^{-1}$, donde $\delta_n = A\|\eta_0 - g_n\|_{\mathcal{F}} = O(n^{-r\nu})$. Tenemos que mostrar que $\gamma_n^2 \phi_n(1/\gamma_n) \lesssim \sqrt{n}$. Notemos que

$$\gamma_n^2 \phi_n \left(\frac{1}{\gamma_n} \right) = \gamma_n q_n^{1/2} + \gamma_n^2 \frac{q_n}{\sqrt{n}} = \sqrt{n} a_n (1 + a_n),$$

donde $a_n = \gamma_n q_n^{1/2} / \sqrt{n}$. De esta forma, para derivar que $\gamma_n^2 \phi_n(1/\gamma_n) \lesssim \sqrt{n}$, es suficiente mostrar que $a_n = O(1)$, lo cual se deduce fácilmente del hecho que $k_n = O(n^\nu)$ y $\gamma_n = O(n^\varsigma)$ con $\varsigma = \min(r\nu, (1-\nu)/2)$.

La condición $\mathbb{M}_n(\widehat{\boldsymbol{\theta}}) \geq \mathbb{M}_n(\boldsymbol{\theta}_{0,n}) - O_{\mathbb{P}}(\gamma_n^{-2})$ requerida por el Lema 3.7 se verifica trivialmente porque $\widehat{\boldsymbol{\theta}}_n$ minimiza $L_n(\boldsymbol{\theta}, \widehat{\kappa})$. Por lo tanto, obtenemos que $\gamma_n^2 d_n^2(\boldsymbol{\theta}_{0,n}, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$.

Finalmente, $d_n(\boldsymbol{\theta}_{0,n}, \boldsymbol{\theta}_0) \leq \|w\|_\infty^{1/2} \|g_n - \eta_0\|_\infty = O(n^{-r\nu}) \leq \gamma_n$, junto a $\gamma_n^2 d_n^2(\boldsymbol{\theta}_{0,n}, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$ y la desigualdad triangular conducen a $\gamma_n^2 d_n^2(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$, concluyendo la demostración.

b) Supongamos ahora que se cumple **C6**. Usando que cualquier $f \in \mathcal{F}_{n,\delta}$ puede escribirse como $f = f_1 - f_2$ con $f_j \in \mathcal{F}_{n,\epsilon_0}^*$ y la cota dada en **C6**, tenemos que

$$N_{[]}(\epsilon, \mathcal{F}_{n,\delta}, L_2(P)) \leq C_2^2 \frac{1}{\epsilon^{2(k_n+p+1)}}.$$

Esto implica que

$$J_{[]} (A_2^{1/2} \delta, \mathcal{F}_{n,\delta}, L_2(P)) \lesssim \delta \log \left(\frac{1}{\delta} \right) \sqrt{k_n + p + 1}.$$

Si llamamos $q_n = k_n + p + 1$, obtenemos

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta_{n,\delta}} |\mathbb{G}_n V_{\boldsymbol{\theta}_{0,n}, \kappa_0} - \mathbb{G}_n V_{\boldsymbol{\theta}, \kappa_0}| \leq A \left(q_n^{1/2} \delta \log \left(\frac{1}{\delta} \right) + n^{-1/2} q_n \left[\log \left(\frac{1}{\delta} \right) \right]^2 \right).$$

Eligiendo

$$\phi_n(\delta) = q_n^{1/2} \delta \log \left(\frac{1}{\delta} \right) + n^{-1/2} q_n \left[\log \left(\frac{1}{\delta} \right) \right]^2,$$

obtenemos que $\phi_n(\delta)/\delta$ es decreciente en δ .

El Lema 3.7 permite concluir que $\gamma_n^2 d_n^2(\boldsymbol{\theta}_{0,n}, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$, donde γ_n es cualquier sucesión que satisface $\gamma_n \lesssim \delta_n^{-1}$ con $\delta_n = \pi(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0,n}) = O(n^{-r\nu})$ y $\gamma_n^2 \phi_n(1/\gamma_n) \leq \sqrt{n}$. La primera condición implica que $\gamma_n \leq O(n^{r\nu})$, mientras que de la segunda se deduce

$$\gamma_n^2 \left(q_n^{1/2} \gamma_n^{-1} \log(\gamma_n) + q_n n^{-1/2} [\log(\gamma_n)]^2 \right) \leq n^{1/2}.$$

Luego, usando que $k_n = O(n^\nu)$ concluimos que $\gamma_n \log(\gamma_n) \leq O(n^{(1-\nu)/2})$. Finalmente, como antes, la condición $\mathbb{M}_n(\widehat{\boldsymbol{\theta}}) \geq \mathbb{M}_n(\boldsymbol{\theta}_0) - O_{\mathbb{P}}(r_n^{-2})$ requerida por el Lema 3.7 se verifica trivialmente porque $\widehat{\boldsymbol{\theta}}_n$ minimiza $L_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\kappa}})$.

Como en a), $d_n(\boldsymbol{\theta}_{0,n}, \boldsymbol{\theta}_0) \leq \|w\|_\infty^{1/2} \|g_n - \eta_0\|_\infty = O(n^{-r\nu}) \leq \gamma_n$ junto con $\gamma_n^2 d_n^2(\boldsymbol{\theta}_{0,n}, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$ y la desigualdad triangular conducen a $\gamma_n^2 d_n^2(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(1)$. \blacksquare

3.5. Apéndice B: Algunos comentarios sobre los supuestos

En esta sección, discutiremos la validez de los supuestos necesarios para obtener los resultados de consistencia y de tasa de convergencia de los estimadores propuestos.

3.5.1. Sobre C0 a C3

Como se mencionó anteriormente, para el modelo logístico y el modelo Poisson, no existe parámetro de ruido κ_0 , pudiendo considerarse $\kappa_0 = 1$ conocido y no necesitando ser estimado. Por esta razón, **C0** puede omitirse. Por otro lado, bajo el modelo parcialmente lineal (2.12) con errores simétricos, el parámetro de escala κ_0 puede ser estimado usando cualquier estimador robusto de escala calculado sin utilizar la restricción de monotonía. Para ser más precisos, sean, por ejemplo, $(\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ los estimadores robustos de $(\boldsymbol{\beta}_0, \eta_0)$ definidos en Bianco y Boente (2004) y definamos los residuos como $r_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} - \widehat{\eta}(t_i)$. El estimador de escala $\widehat{\kappa}$ puede tomarse como $\widehat{\kappa} = \text{mediana}_{1 \leq i \leq n} |r_i|$,

o como un estimador de escala basado en una ρ -función como en Maronna *et al.* (2019). Argumentos similares a los utilizados en el Corolario S.3.1 de Boente y Martínez (2017), permiten probar que $\widehat{\kappa} \xrightarrow{a.s.} \kappa_0$.

Análogamente, para el modelo parcialmente lineal con errores log-Gamma o más generalmente para errores con densidad (2.14), el parámetro κ_0 corresponde a la constante de calibración y se relaciona con el parámetro de forma, como se menciona en la Sección 2.4. Puede ser estimado usando un S-estimador preliminar calculado sin imponer la restricción de monotonía, como se describe en la Sección 2.5. Cálculos directos permiten verificar que en este caso también se cumple **C0**.

El supuesto **C1** es un requisito estándar porque establece que la función de peso controla valores grandes de las covariables y que la función de pérdida acota residuos grandes. La equicontinuidad pedida en **C1** permite trabajar con el parámetro de ruido en un contexto general. Una condición similar aparece en Bianco *et al.* (2013a). Para el caso particular de un modelo de regresión parcialmente lineal (2.12), cuando la función $\rho(y, u, a)$ está dada por (2.13) o (2.15), la condición de equicontinuidad se cumple, por ejemplo, si ϕ es continuamente diferenciable con derivada primera ϕ' tal que $s\phi'(s)$ está acotada. Si κ_0 es conocido, la equicontinuidad no es necesaria.

La condiciones **C2** y **C3** refieren a la suavidad de la componente no paramétrica y el espaciado entre nodos. Son análogas a las consideradas, por ejemplo, en Lu (2010, 2015). Por otro lado, las condiciones $\inf_{\boldsymbol{\theta} \in \mathcal{A}_\epsilon} L(\boldsymbol{\theta}, \kappa_0) > L(\boldsymbol{\theta}_0, \kappa_0)$ o $\inf_{\boldsymbol{\theta} \in \widetilde{\mathcal{A}}_\epsilon} L(\boldsymbol{\theta}, \kappa_0) > L(\boldsymbol{\theta}_0, \kappa_0)$ solicitadas para obtener la consistencia de los estimadores $\widehat{\boldsymbol{\theta}}$ y $\widehat{\boldsymbol{\theta}}_{\text{INI}}$, respectivamente, aseguran que $L(\boldsymbol{\theta}_0, \kappa_0)$ no alcanza el valor mínimo en infinito. Cabe mencionar que la hipótesis $\inf_{\boldsymbol{\theta} \in \widetilde{\mathcal{A}}_\epsilon} L(\boldsymbol{\theta}, \kappa_0) > L(\boldsymbol{\theta}_0, \kappa_0)$ es también un requisito en Boente *et al.* (2006) y Boente y Rodríguez (2010) para garantizar la consistencia fuerte de los estimadores allí considerados.

3.5.2. Sobre las condiciones de entropía

En esta sección damos algunos ejemplos de funciones de pérdida que satisfacen los supuestos **C4**, **C5** y **C6**. Como se menciona en la Observación 4.3, argumentos similares sirven para mostrar que **N6**, **N7(a)** y **N8** se verifican cuando la función de escores $\Psi(y, s, a)$ es Lipschitz en (s, a) uniformemente en y .

Sobre el supuesto **C4**

Consideremos el caso del modelo parcialmente lineal isotónico (2.12) así como el caso del modelo logístico parcialmente lineal.

Lema 3.13. *Supongamos que $(y, \mathbf{x}^T, t)^T$ satisface el modelo parcialmente lineal isotónico (2.12). Sea $\rho(y, s, a)$ la función de pérdida dada en (2.13) o (2.15) con ϕ una ρ -función acotada y consideremos las clases de funciones definidas en (3.1) y (3.2), esto es,*

$$\begin{aligned} \mathcal{F}_n &= \{f(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}(t), a) w(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, a \in \mathcal{V}\}, \\ \widetilde{\mathcal{F}}_n &= \{f(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \boldsymbol{\beta} + \boldsymbol{\lambda}^T \mathbf{B}(t), a) w(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, a \in \mathcal{V}\}. \end{aligned}$$

a) Si $\rho(y, s, a)$ está dada por (2.13), entonces se cumplen **C4** y **D4**.

b) Supongamos que $\rho(y, s, a)$ está dada por (2.15) y que los errores tienen densidad dada por (2.14). Si ϕ es una función continuamente diferenciable tal que $\zeta(s) = s\phi'(s)$ es acotada, entonces \mathcal{F}_n satisface **C4** y **D4**.

DEMOSTRACIÓN. Bastará probar que se cumple **D4**. Considerando que ϕ y w son funciones acotadas y que dividir por una constante no modifica la definición de los estimadores, sin pérdida de generalidad suponemos que $\|\phi\|_\infty = 1$ y $\|w\|_\infty = 1$.

La prueba de a) usa los mismos argumentos considerados en la demostración del Lema 4.2.2 en Smucler (2016). Los incluimos por claridad. La familia de residuos escalados

$$\mathcal{R} = \left\{ \frac{y - \mathbf{x}^\top \boldsymbol{\beta} - \boldsymbol{\lambda}^\top \mathbf{B}(t)}{a} : \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, a \in \mathcal{V} \right\}$$

es un subconjunto del espacio vectorial de todas las funciones en $p + k_n + 1$ variables. Por el Lema 3.4, \mathcal{R} tiene índice VC a lo sumo $p + k_n + 3$.

Recordemos que $\rho(y, s, a) = \phi((y - s)/a)$ con ϕ una ρ -función. Luego, ϕ es una función par, no decreciente en $[0, +\infty)$ y puede ser escrita como $\phi = \phi^{(1)} + \phi^{(2)}$, donde $\phi^{(1)}(x) = \phi(x)\mathbb{I}_{[0, +\infty)}(x)$ es no decreciente y $\phi^{(2)}(x) = \phi(x)\mathbb{I}_{(-\infty, 0)}(x)$ es no creciente. Usando las propiedades de permanencia para clases VC dadas en el Lema 3.5, obtenemos que las clases de funciones $\mathcal{R}_\phi^{(1)} = \phi^{(1)} \circ \mathcal{R}$ y $\mathcal{R}_\phi^{(2)} = \phi^{(2)} \circ \mathcal{R}$ son clases VC con índice a lo sumo $p + k_n + 3$. Para $j = 1, 2$, definimos la familia de funciones

$$\mathcal{F}_n^{(j)} = \left\{ f(y, \mathbf{x}, t) = \phi^{(j)} \left(\frac{y - \mathbf{x}^\top \boldsymbol{\beta} - g_\lambda(t)}{a} \right) w(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, g_\lambda(t) = \boldsymbol{\lambda}^\top \mathbf{B}(t) \text{ con } \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, a \in \mathcal{V} \right\}.$$

Entonces, $\mathcal{F}_n^{(j)} = \mathcal{R}_\phi^{(j)} \cdot w$ y por lo tanto, el Lema 3.5 implica que $\mathcal{F}_n^{(j)}$ es una clase VC con índice a lo sumo $q_n = 2(p + k_n + 3) - 1$.

Más aun, las clases $\mathcal{F}_n^{(j)}$ tienen envolvente 1 porque $\|\phi\|_\infty = \|w\|_\infty = 1$. Luego, el Lema 3.6 implica que existe una constante universal K tal que, para cualquier medida de probabilidad \mathbb{Q} sobre \mathbb{R}^{p+k_n+1} y cualquier $0 < \epsilon < 1$, tenemos que

$$N(\epsilon, \mathcal{F}_n^{(j)}, L_1(\mathbb{Q})) \leq K q_n (16e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n - 1}.$$

Notemos que $\mathcal{F}_n^{(1)} + \mathcal{F}_n^{(2)}$ tiene envolvente constante e igual a 2, por lo tanto,

$$\begin{aligned} N(2\epsilon, \mathcal{F}_n^{(1)} + \mathcal{F}_n^{(2)}, L_1(\mathbb{Q})) &\leq N(\epsilon, \mathcal{F}_n^{(1)}, L_1(\mathbb{Q})) \times N(\epsilon, \mathcal{F}_n^{(2)}, L_1(\mathbb{Q})) \\ &\leq \left[K q_n (16e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n - 1} \right]^2. \end{aligned}$$

Finalmente, notando que $\tilde{\mathcal{F}}_n$ tiene envolvente constante igual a 1 y que $\tilde{\mathcal{F}}_n \subset \mathcal{F}_n^{(1)} + \mathcal{F}_n^{(2)}$ y suponiendo sin pérdida de generalidad que $K > 1$, obtenemos que

$$\begin{aligned} \log \left(N(\epsilon, \tilde{\mathcal{F}}_n, L_1(\mathbb{Q})) \right) &\leq \log \left[K q_n (16e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n - 1} \right]^2 \\ &\leq 2 \left\{ \log(K) + \log q_n + q_n \log(16e) + (q_n - 1) \log \left(\frac{1}{\epsilon} \right) \right\} \\ &\leq 2 \left\{ q_n \left[\log(K) + 1 + \log(16e) + \log \left(\frac{1}{\epsilon} \right) \right] \right\} \leq C(p + k_n) \log \left(\frac{1}{\epsilon} \right) \end{aligned}$$

para $\epsilon < \epsilon_1$ y alguna constante C , concluyendo la prueba de a).

b) Como antes, la clase $\tilde{\mathcal{F}}_n$ tiene envolvente constante igual a 1. Recordemos que, para errores con distribución (2.14), la función de pérdida está dada por

$$\rho(y, s, a) = \phi \left(\frac{\sqrt{d(y - s)}}{a} \right),$$

con $d(s) = v(u_0) - v(s)$, donde v es una función continua con única moda en u_0 , esto es, $d(s) = v(u_0) - v(s)$ es decreciente en $(-\infty, u_0)$ y creciente en $[u_0, +\infty)$.

Por simplicidad, llamamos $\varphi(s) = \phi(\sqrt{s})$. Para cada $a \in \mathbb{R}$ fijo definamos

$$\mathcal{F}_{n,a} = \left\{ f(y, \mathbf{x}, t) = \varphi \left(\frac{d(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))}{a^2} \right) w(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, g_\lambda(t) = \boldsymbol{\lambda}^T \mathbf{B}(t) \text{ con } \boldsymbol{\lambda} \in \mathbb{R}^{k_n} \right\}.$$

Notemos que $\mathcal{F}_{n,a} \subset \mathcal{F}_{n,a}^{(1)} + \mathcal{F}_{n,a}^{(2)}$, donde $\mathcal{F}_{n,a}^{(j)} = \mathcal{G}_a^{(j)} \cdot w$ con

$$\begin{aligned} \mathcal{G}_a^{(1)} &= \left\{ f(y, \mathbf{x}, t) = \varphi \left(\frac{d(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))}{a^2} \right) \mathbb{I}_{[u_0, \infty)}(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t)), \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^{k_n} \right\} \\ \mathcal{G}_a^{(2)} &= \left\{ f(y, \mathbf{x}, t) = \varphi \left(\frac{d(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))}{a^2} \right) \mathbb{I}_{(-\infty, u_0]}(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t)), \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^{k_n} \right\}. \end{aligned}$$

Queremos ver que existe una constante universal $K > 1$ tal que, para cualquier medida de probabilidad \mathbb{Q} en \mathbb{R}^{p+k_n+1} , $a \in \mathbb{R}$ y cualquier $0 < \epsilon < 1$, tenemos que

$$N(2\epsilon, \mathcal{F}_{n,a}, L_1(\mathbb{Q})) \leq \left[K q_n (16e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n - 1} \right]^2 \quad (3.11)$$

con $q_n = 2v_n - 1$ y $v_n = 2(p + k_n) + 7$.

Argumentando como en a), bastará con mostrar que $\mathcal{G}_a^{(1)}$ y $\mathcal{G}_a^{(2)}$ son clases VC con índice menor o igual a $v_n = 2(p + k_n) + 7$. Obtendremos solamente el resultado para $\mathcal{G}_a^{(1)}$ por ser el otro muy similar. Como, a está fijo, llamemos por simplicidad

$$f_{\boldsymbol{\beta}, \boldsymbol{\lambda}}(y, \mathbf{x}, t) = \varphi \left(\frac{d(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))}{a^2} \right) \mathbb{I}_{[u_0, \infty)}(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))$$

y notemos que $\phi(0) = 0$ implica que

$$f_{\boldsymbol{\beta}, \boldsymbol{\lambda}}(y, \mathbf{x}, t) = \varphi \left(\frac{d(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))}{a^2} \mathbb{I}_{[u_0, \infty)}(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t)) \right).$$

Por lo tanto, usando que la función φ es no decreciente en $[0, +\infty)$, bastará ver que la clase

$$\mathcal{R}_a^{(1)} = \{g_{\beta, \lambda}(y, \mathbf{x}, t) = \frac{d(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))}{a^2} \mathbb{I}_{[u_0, \infty)}(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t)), \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathbb{R}^{k_n}\}$$

es una clase VC con índice menor o igual a $v_n = 2(p + k_n) + 7$.

Supongamos que $\mathcal{R}_a^{(1)}$ tiene índice VC, posiblemente infinito, mayor a v_n . Entonces, existen v_n puntos $(y_i, \mathbf{x}_i, t_i, r_i)$, $1 \leq i \leq v_n$, con $r_i \in \mathbb{R}$, que pueden ser desmenuzados por los subgrafos de las funciones de $\mathcal{R}_a^{(1)}$. Claramente, al menos $p + k_n + 4$ puntos r_i deben tener el mismo signo.

Supongamos primero que $r_i < 0$, para $1 \leq i \leq s$ con $s \geq p + k_n + 4$. Entre todos los subconjuntos de $\mathcal{D}_s = \{(y_i, \mathbf{x}_i, t_i, r_i), 1 \leq i \leq s\}$, los subgrafos de las funciones $f_{\beta, \lambda}(y, \mathbf{x}, t)$ solo eligen el conjunto \mathcal{D}_s porque $r_i < 0$ y $f_{\beta, \lambda}(y, \mathbf{x}, t) \geq 0$. Esto lleva a una contradicción, de manera que los v_n son tales que al menos $p + k_n + 4$ puntos r_i son mayores o iguales a 0.

Sin pérdida de generalidad, supongamos que $r_i \geq 0$, para $1 \leq i \leq s$ y algún $s \geq p + k_n + 4$. Dado cualquier subconjunto $I \subset \{1, \dots, s\}$, existe un subgrafo que lo elige, esto es, existe $(\boldsymbol{\beta}_I, \boldsymbol{\lambda}_I)$ tal que $f_{\beta_I, \lambda_I}(y_i, \mathbf{x}_i, t_i) > r_i$ si y solo si $i \in I$. Considerando que $\mathbb{I}_{[u_0, \infty)}(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t)) = 0$ cuando $y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t) < u_0$, tenemos que $u_{i, I} = y_i - \mathbf{x}_i^T \boldsymbol{\beta}_I - \boldsymbol{\lambda}_I^T \mathbf{B}(t_i) > u_0$. Observemos que $f_{\beta_I, \lambda_I}(y_i, \mathbf{x}_i, t_i) > r_i$ es equivalente a $d(u_{i, I}) > a^2 r_i$ y considerando que $d(s)$ es creciente en $[u_0, +\infty)$ resulta $u_{i, I} > d^{-1}(a^2 r_i)$. Resumiendo, tenemos que para cualquier subconjunto $I \subset \{1, \dots, s\}$, existe $(\boldsymbol{\beta}_I, \boldsymbol{\lambda}_I)$ tal que

$$y_i - \mathbf{x}_i^T \boldsymbol{\beta}_I - \boldsymbol{\lambda}_I^T \mathbf{B}(t_i) > d^{-1}(a^2 r_i) \iff i \in I.$$

Llamemos $\tilde{\mathbf{u}}_i = (y_i, \mathbf{x}_i^T, \mathbf{B}(t_i)^T, z_i)^T \in \mathbb{R}^{p+k_n+2}$ con $z_i = d^{-1}(a^2 r_i)$. Definiendo $\boldsymbol{\gamma}_I = (1, -\boldsymbol{\beta}_I, -\boldsymbol{\lambda}_I, -1)$, obtenemos que

$$\boldsymbol{\gamma}_I^T \tilde{\mathbf{u}}_i > 0 \iff i \in I.$$

Por lo tanto, los semiespacios de dimension $p + k_n + 2$ pueden desmenuzar un conjunto de $p + k_n + 4$ elementos, lo que contradice el hecho de que los semiespacios de dimensión d tienen índice VC igual a $d + 2$, ver Lema 3.3. De esta forma, obtenemos que $\mathcal{R}_a^{(1)}$ tiene índice VC menor o igual a v_n concluyendo la demostración de (3.11).

Para terminar la demostración de b), notemos que $\mathcal{F}_n = \cup_{a \in \mathcal{V}} \mathcal{F}_{n, a}$. Sea

$$f_{\beta, \lambda, a}(y, \mathbf{x}, t) = \phi \left(\frac{\sqrt{d(y - \mathbf{x}^T \boldsymbol{\beta} - g_\lambda(t))}}{a} \right) w(\mathbf{x}).$$

Recordemos que \mathcal{V} es un entorno de $\kappa_0 > 0$, sin pérdida de generalidad podemos suponer que existe $0 < \kappa_1 < \kappa_0 < \kappa_2$ tal que $\mathcal{V} \subset [\kappa_1, \kappa_2]$. Por lo tanto, dado $\epsilon > 0$, existen $M = M(\epsilon)$ y puntos a_1, \dots, a_M tal que $\mathcal{V} \subset \cup_{j=1}^M \mathcal{B}(a_j, \kappa_1 \epsilon / \|\zeta\|_\infty)$, donde $M(\epsilon) \leq K_{\mathcal{V}} / \epsilon$, para alguna constante $K_{\mathcal{V}}$, y $\mathcal{B}(a, \epsilon) = \{u : |u - a| < \epsilon\}$.

Sea $\epsilon > 0$ y $a \in \mathcal{V}$. Por lo anterior, existe $1 \leq j_a \leq M(\epsilon)$ tal que $|a - a_{j_a}| < \epsilon / \|\zeta\|_\infty$. El Teorema del Valor Medio junto con el hecho que $\zeta(s) = s\phi'(s)$ es acotada y $\|w\|_\infty = 1$ implican que, para cualquier $\beta \in \mathbb{R}^p$, $\lambda \in \mathbb{R}^{k_n}$,

$$|f_{\beta, \lambda, a}(y, \mathbf{x}, t) - f_{\beta, \lambda, a_{j_a}}(y, \mathbf{x}, t)| \leq \|w\|_\infty \|\zeta\|_\infty \frac{|a - a_{j_a}|}{\kappa_1} < \epsilon,$$

lo que implica que $\|f_{\beta, \lambda, a} - f_{\beta, \lambda, a_{j_a}}\|_{L_1(\mathbb{Q})} < \epsilon$.

Consideremos las clases de funciones \mathcal{F}_{n, a_j} para $1 \leq j \leq M(\epsilon)$ y sea $N_j(\epsilon) = N(\epsilon, \mathcal{F}_{n, a_j}, L_1(\mathbb{Q}))$. Para cualquier β, λ existe $1 \leq m \leq N_{j_a}(\epsilon)$ y $f_m = f_{\beta_m, \lambda_m, a_{j_a}} \in \mathcal{F}_{n, a_{j_a}}$ tal que $\|f_{\beta, \lambda, a_{j_a}} - f_m\|_{L_1(\mathbb{Q})} < \epsilon$, de donde se deduce que $\|f_{\beta, \lambda, a} - f_{\beta_m, \lambda_m, a_{j_a}}\|_{L_1(\mathbb{Q})} < 2\epsilon$, lo que implica que $N(2\epsilon, \mathcal{F}_n, L_1(\mathbb{Q})) \leq \sum_{j=1}^{M(\epsilon)} N_j(\epsilon)$. Usando (3.11), obtenemos la cota

$$N(2\epsilon, \mathcal{F}_n, L_1(\mathbb{Q})) \leq M(\epsilon) \left[K q_n (32e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n - 1} \right]^2 \leq K_{\mathcal{V}} \left[K q_n (32e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n - \frac{1}{2}} \right]^2$$

concluyendo la demostración. ■

Lema 3.14. *Supongamos que $(y, \mathbf{x}^\top, t)^\top$ es tal que $y | (\mathbf{x}, t) \sim \text{Bi}(1, \pi(\mathbf{x}, t))$ donde*

$$\pi(\mathbf{x}, t) = H(\mathbf{x}^\top \beta + \eta_0(t))$$

con $H(u) = 1/(1 + \exp(-u))$. Consideremos la función de pérdida ρ dada por (2.19), es decir,

$$\rho(y, u) = y\varphi(-\log[H(u)]) + (1-y)\varphi(-\log[1-H(u)]) + G(H(u)) + G(1-H(u)),$$

con $G(t) = \int_0^t \varphi'(-\log u) du$. Supongamos que $\varphi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ es una función acotada con derivada continua y acotada φ' tal que $\varphi'(t) \geq 0$ y $\varphi(0) = 0$. Entonces, las clases de funciones definida en (3.1) y (3.2), satisfacen **C4** y **D4**, respectivamente.

DEMOSTRACIÓN. La demostración sigue los mismos argumentos usados en la demostración del Lema 3.13, considerando que $\rho_0(u) = \rho(0, u)$ es creciente, $\rho_1(u) = \rho(1, u)$ es una función decreciente y que la variable y toma solo los valores 0 or 1. ■

Sobre los supuestos C5 y C6

En esta sección, estudiaremos condiciones sobre la función de pérdida que aseguran que **C5** y **C6** se verifican para algunos modelos.

En primer lugar, observemos que los argumentos considerados en Shen y Wong (1994) al analizar el Caso 3 en la página 596 implican que la entropía corchete de la clase

$$\mathcal{H}_{c, \lambda_0} = \{g_\lambda(t) = \lambda^\top \mathbf{B}(t), \lambda \in \mathcal{L}_{k_n}, \|g_\lambda - g_{\lambda_0}\|_2 \leq c\} \quad (3.12)$$

puede acotarse por

$$\log N_{[]}(\epsilon, \mathcal{H}_{c, \lambda_0}, L_2(P)) \leq C_1 k_n \log(c/\epsilon), \quad (3.13)$$

para $0 < \epsilon < c$ y la constante C_1 es independiente de c , λ_0 y n . Un resultado análogo vale para la clase

$$\tilde{\mathcal{H}}_{c, \lambda_0} = \{g_\lambda(t) = \boldsymbol{\lambda}^\top \mathbf{B}(t), \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, \|g_\lambda - g_{\lambda_0}\|_2 \leq c\}. \quad (3.14)$$

El siguiente lema será útil. Para cualquier matriz simétrica $\mathbf{A} \in \mathbb{R}^{p \times p}$, llamemos $\zeta_1(\mathbf{A}) \geq \dots \geq \zeta_p(\mathbf{A})$ a los autovalores ordenados de \mathbf{A} .

Lema 3.15. *Sea $\mathbf{f}(t) = \mathbb{E}(\mathbf{x}|t)$. Supongamos que $\mathbb{P}(\mathbf{x} = \mathbf{f}(t)) < 1$ y que $\mathbf{S}_x = \mathbb{E}\mathbf{x}\mathbf{x}^\top$ es no singular.*

a) Sea $c = \sup_{\|\mathbf{a}\|=1} F(\mathbf{a})$ con

$$F(\mathbf{a}) = \frac{\mathbf{a}^\top \mathbb{E}\mathbf{f}(t)\mathbf{f}(t)^\top \mathbf{a}}{\mathbf{a}^\top \mathbb{E}\mathbf{x}\mathbf{x}^\top \mathbf{a}}.$$

Entonces, $0 \leq c < 1$.

b) Si $w \equiv 1$, existe una constante $C > 0$ tal que, para cualquier $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \eta_1)$ y $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, \eta_2)$, tenemos que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq C\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Más aun, $C = (1 - \sqrt{c}) \min(1, \zeta_p(\mathbf{S}_x))$.

Entonces, si $\pi_{\mathbb{P}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$ también $\pi(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$.

DEMOSTRACIÓN. a) Recordemos que para cualquier par de variables aleatorias U y Z , tenemos que

$$\mathbb{E}U^2 = \mathbb{E}(\mathbb{E}^2(U|Z)) + \mathbb{E}\{\mathbb{E}(U|Z) - U\}^2.$$

Tomando $U = \mathbf{a}^\top \mathbf{x}$ y $Z = t$, obtenemos inmediatamente que

$$\mathbf{a}^\top \mathbb{E}\mathbf{x}\mathbf{x}^\top \mathbf{a} = \mathbf{a}^\top \mathbb{E}\mathbf{f}(t)\mathbf{f}(t)^\top \mathbf{a} + \mathbf{a}^\top \mathbb{E}[(\mathbf{x} - \mathbf{f}(t))(\mathbf{x} - \mathbf{f}(t))^\top] \mathbf{a},$$

para cualquier $\mathbf{a} \in \mathbb{R}^p$. Luego, usando que $\mathbb{P}(\mathbf{x} = \mathbf{f}(t)) < 1$, deducimos que, para cualquier $\|\mathbf{a}\| = 1$, $0 \leq F(\mathbf{a}) < 1$. Además, el hecho de que \mathbf{S}_x sea no singular implica que la función

$$F(\mathbf{a}) = \frac{\mathbf{a}^\top \mathbb{E}\mathbf{f}(t)\mathbf{f}(t)^\top \mathbf{a}}{\mathbf{a}^\top \mathbb{E}\mathbf{x}\mathbf{x}^\top \mathbf{a}}$$

es continua en la bola unitaria, alcanzando su máximo en un punto \mathbf{a}_{\max} , lo que implica que $c = F(\mathbf{a}_{\max}) < 1$.

b) Usaremos el Lema 25.86 de van der Vaart (1998). Sean $g_1 = \eta_1(t) - \eta_2(t)$ y $g_2 = \mathbf{x}^\top(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$. La desigualdad de Cauchy–Schwartz implica que

$$|\mathbb{E}g_1g_2|^2 = |\mathbb{E}g_1\mathbb{E}(g_2|t)|^2 \leq \mathbb{E}g_1^2 \mathbb{E}[\mathbb{E}^2(g_2|t)].$$

Elegiendo $\mathbf{a} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$, obtenemos que $\mathbb{E}[\mathbb{E}^2(g_2|t)] = F(\mathbf{a})\mathbb{E}g_2^2$. Entonces, usando que $F(\mathbf{a}) = F(\mathbf{a}/\|\mathbf{a}\|) \leq c$, concluimos que $|\mathbb{E}g_1g_2|^2 \leq c\mathbb{E}g_1^2\mathbb{E}g_2^2$. El Lema 25.86 de van der Vaart (1998) implica que, para $C = 1 - \sqrt{c}$, vale la siguiente desigualdad

$$\mathbb{E}(g_1 + g_2)^2 \geq C \{\mathbb{E}g_1^2 + \mathbb{E}g_2^2\},$$

es decir,

$$\pi_{\mathbb{P}}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq C \left\{ \mathbb{E} (\eta_1(t) - \eta_2(t))^2 + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)^{\text{T}} \mathbb{E} \mathbf{x} \mathbf{x}^{\text{T}} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \right\}$$

y la demostración se concluye usando que el menor autovalor de $\mathbb{E} \mathbf{x} \mathbf{x}^{\text{T}}$ es positivo. \blacksquare

Observación 3.2. Cuando $w \equiv 1$, **N5** implica que $\mathbb{E} \mathbf{x} \mathbf{x}^{\text{T}}$ es no singular. Por otro lado, el supuesto $\mathbb{P}(\mathbf{x} = \mathbf{f}(t)) < 1$, en el Lema 3.15, significa que \mathbf{x} no se predice completamente por t asegurando que el modelo sea identificable. Más aun, bajo el modelo (2.12), **N5** implica que $\mathbb{P}(\mathbf{x} = \mathbf{f}(t)) < 1$ se cumple. Entonces, el Teorema 3.11a) junto con el Lema 3.15 implican que $\pi(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$. \clubsuit

Lema 3.16. *Supongamos que \mathbf{x} es independiente de t , $\mathbf{S}_{\mathbf{x}} = \mathbb{E} w(\mathbf{x}) \mathbf{x} \mathbf{x}^{\text{T}}$ es no singular y $\mathbb{E} w(\mathbf{x}) \mathbf{x} = \mathbf{0}$. Entonces, para cualquier $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \eta_1)$ y $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, \eta_2)$, tenemos que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq C \pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, donde $C = \min(\mathbb{E} w(\mathbf{x}), \zeta_p(\mathbf{S}_{\mathbf{x}}))$. Por lo tanto, si $\pi_{\mathbb{P}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$ también tenemos que $\pi(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$.*

DEMOSTRACIÓN. Por la independencia de \mathbf{x} y t

$$\begin{aligned} \pi_{\mathbb{P}}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \mathbb{E} w(\mathbf{x}) [\mathbf{x}^{\text{T}} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)]^2 + \mathbb{E} w(\mathbf{x}) \mathbb{E} [\eta_1(t) - \eta_2(t)]^2 + 2 \mathbb{E} \mathbf{x}^{\text{T}} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) [\eta_1(t) - \eta_2(t)] \\ &= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)^{\text{T}} \mathbf{S}_{\mathbf{x}} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) + \mathbb{E} w(\mathbf{x}) \mathbb{E} [\eta_1(t) - \eta_2(t)]^2, \end{aligned}$$

porque $\mathbb{E} w(\mathbf{x}) \mathbf{x} = \mathbf{0}$. El resultado sigue inmediatamente. \blacksquare

Cuando la función de peso es tal que $w(\mathbf{x}) = w(-\mathbf{x})$, la condición $\mathbb{E} w(\mathbf{x}) \mathbf{x} = \mathbf{0}$ se verifica para vectores aleatorios que tengan distribución simétrica alrededor de $\mathbf{0}$. Para familias de distribuciones elípticas, se toma usualmente como función de peso $w(\mathbf{x}) = W((\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^{\text{T}} \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}))$ donde $\boldsymbol{\mu}_{\mathbf{x}}$ y $\boldsymbol{\Sigma}_{\mathbf{x}}$ indican los parámetros de posición y dispersión de la distribución elíptica y la función $W : \mathbb{R} \rightarrow \mathbb{R}$ es tal que $W(s) \geq 0$. Por lo tanto, si $\boldsymbol{\mu} = \mathbf{0}$, las condiciones $w(\mathbf{x}) = w(-\mathbf{x})$ y $\mathbb{E} w(\mathbf{x}) \mathbf{x} = \mathbf{0}$ se cumplen automáticamente. En el estudio de simulación reportado en el Capítulo 6.1, x y t fueron elegidas independientes entre sí y $x \sim N(0, 1)$, de forma tal que la conclusión del Lema 3.16 se cumple para los datos generado sin contaminación en nuestro estudio.

Veamos ahora cuándo se cumple **C5**. Consideraremos primero el modelo parcialmente lineal isotónico (2.12). Sin pérdida de generalidad supondremos que $\mathcal{V} \subset [\kappa_1, \kappa_2]$ con $0 < \kappa_1 \leq \kappa_0 \leq \kappa_2$. Para hacer explícita la dependencia en $\boldsymbol{\lambda}$, llamaremos $\boldsymbol{\theta}_{0, \lambda_0} = (\boldsymbol{\beta}_0, g_{\lambda_0})$ y $\boldsymbol{\theta}_{\lambda} = (\boldsymbol{\beta}, g_{\lambda})$.

Lema 3.17. *Supongamos que $(y, \mathbf{x}^{\text{T}}, t)^{\text{T}}$ verifica el modelo parcialmente lineal isotónico (2.12). Sea $\rho(y, s, a)$ la función dada en (2.13) o (2.15) con ϕ una ρ -función diferenciable y acotada tal que $\phi'(s)$ y $\zeta(s) = s\phi'(s)$ están acotadas. Supongamos que w está acotada y sin pérdida de generalidad, supongamos que $\|w\|_{\infty} = 1$. Para un $\boldsymbol{\lambda}_0 \in \mathcal{L}_{k_n}$ fijo, consideremos las clases de funciones definidas*

en **C5** y **D5**, es decir,

$$\begin{aligned} \mathcal{G}_{n,c,\lambda_0} &= \{f_{\lambda,\beta,a}(y, \mathbf{x}, t) = [\rho(y, \mathbf{x}^\top \boldsymbol{\beta} + g_\lambda(t), a) - \rho(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + g_{\lambda_0}(t), a)] w(\mathbf{x}), \\ &\quad \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon_0, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, a \in \mathcal{V}, \pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,\lambda_0}, \boldsymbol{\theta}_\lambda) \leq c\}. \end{aligned} \quad (3.15)$$

$$\begin{aligned} \tilde{\mathcal{G}}_{n,c,\lambda_0} &= \{f_{\lambda,\beta,a}(y, \mathbf{x}, t) = [\rho(y, \mathbf{x}^\top \boldsymbol{\beta} + g_\lambda(t), a) - \rho(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + g_{\lambda_0}(t), a)] w(\mathbf{x}), \\ &\quad \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon_0, \boldsymbol{\lambda} \in \mathbb{R}^{k_n}, a \in \mathcal{V}, \pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,\lambda_0}, \boldsymbol{\theta}_\lambda) \leq c\}. \end{aligned} \quad (3.16)$$

Supongamos que $\mathbb{E}w^2(\mathbf{x})\|\mathbf{x}\|^2 < \infty$ y que, para cualquier $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, g_1)$ y $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, g_2)$, $\pi_{\mathbb{P}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq C\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ para alguna constante $C > 0$. Entonces, tenemos que se cumplen **C5** y **D5**, si se verifica alguna de las siguiente condiciones

a $\rho(y, s, a)$ está dada por (2.13),

b $\rho(y, s, a)$ está dada por (2.15), los errores tienen densidad dada por (2.14), v es derivable y

$$\psi^*(s, a) = \frac{1}{2a} \phi' \left(\frac{\sqrt{d(s)}}{a} \right) \frac{v'(s)}{\sqrt{d(s)}} \quad (3.17)$$

está acotada.

Observemos que los Lemas 3.15 y 3.16 dan condiciones que aseguran que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq C\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, como se requiere en el Lema 3.17.

DEMOSTRACIÓN. Basta probar el resultado para $\tilde{\mathcal{G}}_{n,c,\lambda_0}$. Usando que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq C\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ tenemos que, para cualquier $f_{\lambda,\beta,a} \in \tilde{\mathcal{G}}_{n,c,\lambda_0}$, $g_\lambda = \boldsymbol{\lambda}^\top \mathbf{B}(t) \in \mathcal{H}_{c_1,\lambda_0}$, con $c_1 = c/C$ y $\tilde{\mathcal{H}}_{c_1,\lambda_0}$ definida en (3.14). Por lo tanto, usando (3.13), para cualquier $0 < \epsilon < c_1$, $\tilde{\mathcal{H}}_{c_1,\lambda_0}$ puede cubrirse por una cantidad finita $M_1(\epsilon) \leq (c_1/\epsilon)^{C_1 k_n}$ de ϵ -corchetes $\{[g_{j,L}, g_{j,U}], 1 \leq j \leq M_1(\epsilon)\}$.

Por otro lado, el conjunto $\{\mathbf{b} : \|\mathbf{b} - \boldsymbol{\beta}_0\| < \epsilon_0\}$ puede cubrirse por $M_2(\epsilon) \leq C_2(\epsilon_0/\epsilon)^p$ bolas de radio ϵ y centro $\boldsymbol{\beta}^{(m)}$, $1 \leq m \leq M_2(\epsilon)$. Similarmente, el conjunto $[\kappa_1, \kappa_2]$ puede ser cubierto por $M_3(\epsilon) \leq C_3(1/\epsilon)$ bolas de radio ϵ y centro $a^{(s)}$, $1 \leq s \leq M_3(\epsilon)$.

Es fácil ver que bajo el modelo (2.12), la función $\Psi(y, s, a) = \partial \rho(y, s, a) / \partial s$ puede escribirse como $\Psi(y, s, a) = \psi^*(y - s, a)$, donde $\psi^*(s, a) = -\phi'(s/a)$, cuando ρ está dada por (2.13), mientras que ψ^* está dada por (3.17) si $\rho(y, s, a)$ está definida en (2.15).

Más aun, cuando $\rho(y, s, a) = \phi((y - s)/a)$, tenemos que

$$\frac{\partial}{\partial a} \rho(y, s, a) = -\frac{1}{a} \zeta \left(\frac{y - s}{a} \right).$$

Por otro lado, cuando $\rho(y, s, a)$ está definida en (2.15), obtenemos que

$$\frac{\partial}{\partial a} \rho(y, s, a) = -\frac{1}{a} \zeta \left(\frac{\sqrt{d(y - s)}}{a} \right).$$

Luego, en ambas situaciones $|(\partial/\partial a)\rho(y, s, a)| \leq \|\zeta\|_\infty/\kappa_1$.

Dada $f_{\lambda, \beta, a} \in \tilde{\mathcal{G}}_{n, c, \lambda_0}$, sea j, m y s tales que g_λ pertenece al ϵ -corchete $[g_{j, L}, g_{j, U}]$, $\|\beta - \beta^{(m)}\| < \epsilon$ y $|a - a^{(s)}| < \epsilon$. Llamemos

$$f_{j, m, s}(y, \mathbf{x}, t) = \left[\rho(y, \mathbf{x}^T \beta^{(m)} + g_{j, U}(t), a^{(s)}) - \rho(y, \mathbf{x}^T \beta_0 + g_{\lambda_0}(t), a^{(s)}) \right] w(\mathbf{x})$$

donde $g_{\lambda_0}(t) = \lambda_0^T \mathbf{B}(t)$ y $f_{j, m}(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \beta^{(m)} + g_{j, U}(t), a) - \rho(y, \mathbf{x}^T \beta_0 + g_{\lambda_0}(t), a)$. Usando un desarrollo de Taylor de orden 1 y que $\|w\|_\infty = 1$, tenemos que

$$\begin{aligned} |f_{\lambda, \beta, a} - f_{j, m, s}| &\leq |f_{\lambda, \beta, a} - f_{j, m}| + |f_{j, m} - f_{j, m, s}| \\ &\leq \|\psi^*\|_\infty \left\{ w(\mathbf{x}) \|\mathbf{x}\| \left\| \beta - \beta^{(m)} \right\| + |g_\lambda(t) - g_{j, U}(t)| \right\} + 2 \frac{\|\zeta\|_\infty}{\kappa_1} |a - a^{(s)}| \\ &\leq \epsilon \left(\|\psi^*\|_\infty w(\mathbf{x}) \|\mathbf{x}\| + 2 \frac{\|\zeta\|_\infty}{\kappa_1} \right) + \|\psi^*\|_\infty (g_{j, L}(t) - g_{j, U}(t)), \end{aligned}$$

donde la última desigualdad sigue del hecho de que $g_{j, L} \leq g_\lambda \leq g_{j, U}$, $\|\beta - \beta^{(m)}\| < \epsilon$ y $|a - a^{(s)}| < \epsilon$.

Definamos las funciones

$$\begin{aligned} \varphi_{j, m, s}^{(U)}(y, \mathbf{x}, t) &= f_{j, m, s}(y, \mathbf{x}, t) + \epsilon \left(\|\psi^*\|_\infty w(\mathbf{x}) \|\mathbf{x}\| + 2 \frac{\|\zeta\|_\infty}{\kappa_1} \right) + \|\psi^*\|_\infty (g_{j, L}(t) - g_{j, U}(t)), \\ \varphi_{j, m, s}^{(L)}(y, \mathbf{x}, t) &= f_{j, m, s}(y, \mathbf{x}, t) - \epsilon \left(\|\psi^*\|_\infty w(\mathbf{x}) \|\mathbf{x}\| + 2 \frac{\|\zeta\|_\infty}{\kappa_1} \right) - \|\psi^*\|_\infty (g_{j, L}(t) - g_{j, U}(t)). \end{aligned}$$

Entonces, $\varphi_{j, m, s}^{(L)} \leq f_{\lambda, \beta, a} \leq \varphi_{j, m, s}^{(U)}$. Considerando que $\|g_{j, L} - g_{j, U}\|_2^2 = \mathbb{E}(g_{j, U} - g_{j, L})^2 < \epsilon^2$, obtenemos que

$$\begin{aligned} \left\| \varphi_{j, m, s}^{(U)} - \varphi_{j, m, s}^{(L)} \right\|_2 &\leq 2 \left\{ \epsilon \left(\|\psi^*\|_\infty (\mathbb{E} w^2(\mathbf{x}) \|\mathbf{x}\|^2)^{1/2} + 2 \frac{\|\zeta\|_\infty}{\kappa_1} \right) + \|\psi^*\|_\infty \|g_{j, L} - g_{j, U}\|_2 \right\} \\ &\leq 2\epsilon \left(\|\psi^*\|_\infty (\mathbb{E} w^2(\mathbf{x}) \|\mathbf{x}\|^2)^{1/2} + 2 \frac{\|\zeta\|_\infty}{\kappa_1} + \|\psi^*\|_\infty \right) = A_1 \epsilon, \end{aligned}$$

lo que significa que el número total de corchetes de tamaño $A_1 \epsilon$ requeridos para cubrir $\mathcal{G}_{n, c, \lambda_0}$ está acotado por $\prod_{i=1}^3 M_i(\epsilon) \leq A(c_1/\epsilon)^{C_1 k_n + p + 1}$ concluyendo la demostración. \blacksquare

Ahora estudiamos el cubrimiento corchete de las clases de funciones

$$\begin{aligned} \mathcal{F}_{n, c}^* &= \{f_{\beta, \lambda, a}(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \beta + g_\lambda(t), a) w(\mathbf{x}), \lambda \in \mathcal{L}_{k_n}, a \in \mathcal{V}, \pi(\theta_0, \theta) \leq c\}, \\ \tilde{\mathcal{F}}_{n, c}^* &= \{f_{\beta, \lambda, a}(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^T \beta + g_\lambda(t), a) w(\mathbf{x}), \lambda \in \mathbb{R}^{k_n}, a \in \mathcal{V}, \pi(\theta_0, \theta) \leq c\}, \end{aligned}$$

definidas en **C6** y **D6**, respectivamente, donde $\pi^2(\theta_1, \theta_2) = \|\beta_1 - \beta_2\|^2 + \|\eta_1 - \eta_2\|_{\mathcal{F}}^2$ y $\|\cdot\|_{\mathcal{F}}$ es la norma $L_\varphi(\mathbb{Q})$, con $2 \leq \varphi \leq \infty$. Usando que $\|\eta_1 - \eta_2\|_2^2 \leq \|\eta_1 - \eta_2\|_{\mathcal{F}}^2$, es suficiente mostrar que **C6** o **D6** se cumplen cuando $\pi^2(\theta_1, \theta_2) = \|\beta_1 - \beta_2\|^2 + \|\eta_1 - \eta_2\|_2^2$.

Es claro que para cualquier $f_{\beta, \lambda, a} \in \mathcal{F}_{n, c}^*$, se cumple $\|\beta - \beta_0\| < c$. Más aun, sea $g_n = g_{\lambda_0, n} \in \mathcal{M}_n(\mathcal{T}_n, \ell)$ tal que $\|g_n - \eta_0\|_\infty = O(n^{-r/(1+2r)})$. Entonces, para $n \geq n_0$, $\|g_n - \eta_0\|_\infty \leq c$. Por lo tanto, cuando $\|g\|_{\mathcal{F}} = \|g\|_2$, tenemos que si $f_{\beta, \lambda, a} \in \mathcal{F}_{n, c}^*$, la función $g_\lambda \in \mathcal{H}_{2c, \lambda_0, n}$ donde $\mathcal{H}_{2c, \lambda_0, n}$ está definida en (3.14). De forma similar se obtiene que si $f_{\beta, \lambda, a} \in \tilde{\mathcal{F}}_{n, c}^*$, entonces $g_\lambda \in \tilde{\mathcal{H}}_{2c, \lambda_0, n}$. Por lo tanto, usando argumentos similares a los considerados en el Lema 3.17, obtenemos el siguiente resultado.

Lema 3.18. *Supongamos que $(y, \mathbf{x}^T, t)^T$ satisface el modelo parcialmente lineal isotónico (2.12). Sea $\rho(y, s, a)$ dada en (2.13) o (2.15) con ϕ una ρ -función acotada y diferenciable tal que $\phi'(s)$ y $\zeta(s) = s\phi'(s)$ son acotadas. Para $\boldsymbol{\lambda}_0 \in \mathcal{L}_{k_n}$ fijo, consideramos la clase de funciones definidas en **C6** y **D6**. Supongamos, sin pérdida de generalidad, que $\mathcal{V} \subset [\kappa_1, \kappa_2]$ con $0 < \kappa_1 < \kappa_0 < \kappa_2$. Entonces, si $\mathbb{E}w^2(\mathbf{x}) \|\mathbf{x}\|^2 < \infty$, las hipótesis **C6** y **D6** se cumplen, para $n \geq n_0$, en cualquier de los siguientes casos*

a) $\rho(y, s, a)$ está dada por (2.13)

b) $\rho(y, s, a)$ está dada por (2.15), los errores tienen densidad dada por (2.14), v es derivable y la función $\psi^*(s, a)$ definida en (3.17) es acotada.

Al considerar un modelo logístico parcialmente lineal, observando que $\Psi(y, t) = \partial\rho(y, t)/\partial t$ puede escribirse como $\Psi(y, t) = -[y - H(t)]\nu(t)$ con

$$\nu(t) = \varphi'(-\log H(t)) [1 - H(t)] + \varphi'(-\log [1 - H(t)]) H(t),$$

argumentos similares a los considerados en el Lema 3.17 permiten mostrar el siguiente resultado.

Lema 3.19. *Supongamos que $(y, \mathbf{x}^T, t)^T$ es tal que $y | (\mathbf{x}, t) \sim \text{Bi}(1, \pi(\mathbf{x}, t))$ donde*

$$\pi(\mathbf{x}, t) = H(\mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t))$$

con $H(u) = 1/(1 + \exp(-u))$. Consideremos la función de pérdida ρ dada en (2.19). Supongamos que $\varphi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ es una función acotada con derivada continua y acotada φ' tal que $\varphi'(t) \geq 0$ y $\varphi(0) = 0$.

a) Si $\mathbb{E}w^2(\mathbf{x}) \|\mathbf{x}\|^2 < \infty$ y existe una constante $C > 0$ tal que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \geq C\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, para cualquier $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, g_1)$ y $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, g_2)$, entonces se cumplen **C5** y **D5**.

b) Si $\mathbb{E}w^2(\mathbf{x}) \|\mathbf{x}\|^2 < \infty$, se cumple **C6** y **D6**.

3.5.3. Condiciones que garantizan **C9**

El siguiente Lema establece condiciones que aseguran la validez de **C9**. Un resultado análogo vale para **D9**.

Lema 3.20. *Supongamos que se cumple **C10** y que $\rho(y, u, a)$ es dos veces continuamente diferenciable con respecto a u .*

a) Si la función $\chi(y, u, a) = \partial^2 \rho(y, u, a) / \partial u^2$ es tal que existe $\epsilon_0 > 0$ y un entorno \mathcal{V} de κ_0 tal que

$$C_0 = \inf_{a \in \mathcal{V}} \inf_{\substack{\pi^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \epsilon_0 \\ \boldsymbol{\theta} \in \mathbb{R}^p \times \mathcal{G}}} \inf_{(\mathbf{x}_0, t_0) \in \mathcal{S}_w \times [0, 1]} \mathbb{E}(\chi(y, \mathbf{x}^T \boldsymbol{\beta} + g(t), a) | (\mathbf{x}, t) = (\mathbf{x}_0, t_0)) > 0, \quad (3.18)$$

donde \mathcal{S}_w es el soporte de la función w , entonces se cumple **C9**.

b) Más aun, supongamos que $\pi^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^2 + \|\eta_1 - \eta_2\|_\infty^2$, que w tiene soporte acotado $\mathcal{S}_w \subset \{\|\mathbf{x}\| \leq A_1\}$ o bien que $\mathbb{P}(\|\mathbf{x}\| \leq A_1) = 1$ y que para alguna constante positiva A_2

$$C_0 = \inf_{a \in \mathcal{V}} \inf_{(\mathbf{x}_0, t_0) \in \mathcal{S}_w \times [0,1]} \inf_{|s-s_0| < A_2} \mathbb{E}(\chi(y, s, a) | (\mathbf{x}, t) = (\mathbf{x}_0, t_0)) > 0, \quad (3.19)$$

donde $s_0 = \mathbf{x}_0^\top \boldsymbol{\beta}_0 + \eta_0(t_0)$, entonces se cumple **C9**.

La condición (3.18) es la contraparte robusta del supuesto de que la varianza condicional de $y | (\mathbf{x}, t)$ está acotada inferiormente lejos de 0 usada en el Teorema 1 en Lu (2015). El supuesto (3.19) se verifica si, por ejemplo, $\mathbb{E}(\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), a) | (\mathbf{x}, t) = (\mathbf{x}_0, t_0)) > 0$ y la función $\chi(y, s, a)$ es continua en todos sus argumentos. Estas dos condiciones se verifican, por ejemplo, bajo el modelo parcialmente lineal (2.12) tanto para errores simétricos como para errores con densidad (2.14), cuando las funciones ϕ y v satisfacen los supuestos **N3** y **N5** requeridos para obtener la normalidad asintótica del estimador $\widehat{\boldsymbol{\beta}}$ del parámetro de regresión.

DEMOSTRACIÓN DEL LEMA 3.20. Para cualquier $\boldsymbol{\theta} \in \mathbb{R}^p \times \mathcal{M}_n(\mathcal{T}_n, \ell)$, sea $M_{\boldsymbol{\theta}}(s) = L(\boldsymbol{\theta}_0 + s(\boldsymbol{\theta} - \boldsymbol{\theta}_0), a)$, entonces $M_{\boldsymbol{\theta}}(1) = L(\boldsymbol{\theta}, a)$ y $M_{\boldsymbol{\theta}}(0) = L(\boldsymbol{\theta}_0, a)$. Más aun, llamando $b(\mathbf{x}, t) = \mathbf{x}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + g(t) - \eta_0(t)$, tenemos que

$$\begin{aligned} M'_{\boldsymbol{\theta}}(s) &= \mathbb{E} [w(\mathbf{x}) \Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t) + s b(\mathbf{x}, t), a) b(\mathbf{x}, t)] \\ M''_{\boldsymbol{\theta}}(s) &= \mathbb{E} [w(\mathbf{x}) \chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t) + s b(\mathbf{x}, t), a) b^2(\mathbf{x}, t)]. \end{aligned}$$

El supuesto **C10** implica que $M'_{\boldsymbol{\theta}}(0) = 0$. Por lo tanto, usando un desarrollo de Taylor de orden dos obtenemos que para algún $0 < \xi < 1$, $M_{\boldsymbol{\theta}}(1) - M_{\boldsymbol{\theta}}(0) = M''_{\boldsymbol{\theta}}(\xi)/2$.

a) Sea $\boldsymbol{\beta}_\xi = \boldsymbol{\beta}_0 + \xi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ y $g_\xi = \eta_0 + \xi(g - \eta_0) = (1 - \xi)\eta_0 + \xi g$, luego $\boldsymbol{\theta}_\xi = (\boldsymbol{\beta}_\xi, g_\xi) \in \Theta$ para $g \in \mathcal{G}$ y $\pi(\boldsymbol{\theta}_\xi, \boldsymbol{\theta}_0) = \xi \pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. Entonces, para $a \in \mathcal{V}$, y $\boldsymbol{\theta} \in \mathbb{R}^p \times \mathcal{M}_n(\mathcal{T}_n, \ell)$, tales que $\pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \epsilon_0$, tenemos que

$$\begin{aligned} L(\boldsymbol{\theta}, a) - L(\boldsymbol{\theta}_0, a) &= M_{\boldsymbol{\theta}}(1) - M_{\boldsymbol{\theta}}(0) = \frac{1}{2} \mathbb{E} [w(\mathbf{x}) \chi(y, \mathbf{x}^\top \boldsymbol{\beta}_\xi + g_\xi(t), a) b^2(\mathbf{x}, t)] \\ &= \frac{1}{2} \mathbb{E} \left[w(\mathbf{x}) \mathbb{E} \left\{ \chi(y, \mathbf{x}^\top \boldsymbol{\beta}_\xi + g_\xi(t), a) \middle| (\mathbf{x}, t) \right\} b^2(\mathbf{x}, t) \mathbb{I}_{\mathcal{S}_w \times [0,1]}(\mathbf{x}, t) \right] \\ &\geq C_0 \mathbb{E} w(\mathbf{x}) b^2(\mathbf{x}, t) = C_0 \pi_{\mathbb{P}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \end{aligned}$$

donde usamos que $\pi(\boldsymbol{\theta}_\xi, \boldsymbol{\theta}_0) < \epsilon_0$ y (3.18), concluyendo la demostración de a).

b) Supongamos que $\pi^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^2 + \|\eta_1 - \eta_2\|_\infty^2$ y que se cumple (3.19). Sea $s_0 = \mathbf{x}_0^\top \boldsymbol{\beta}_0 + \eta_0(t_0)$ con $\mathbf{x}_0 \in \mathcal{S}_w$. Usando que $|\mathbf{x}_0^\top \boldsymbol{\beta}_\xi + g_\xi(t) - s_0| \leq A_1 \|\boldsymbol{\beta}_\xi - \boldsymbol{\beta}_0\| + |g_\xi(t_0) - \eta_0(t_0)|$, obtenemos que $|\mathbf{x}_0^\top \boldsymbol{\beta}_\xi + g_\xi(t) - s_0| \leq A_2$, siempre que $\pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \epsilon$, con $\epsilon_0 < A_2/(1 + A_1)$. La demostración continúa como en a) usando (3.19). \blacksquare

Capítulo 4

Normalidad asintótica del estimador del parámetro β_0

En este Capítulo, estudiaremos la distribución asintótica del estimador del parámetro de regresión. Consideraremos dos situaciones: (i) el parámetro de ruido es conocido, como en el modelo logístico y Poisson, o (ii) cuando κ_0 es desconocido, supondremos que $H(t) = t$, es decir que las observaciones siguen el modelo parcialmente lineal (2.12). Más aun, consideraremos las situación de errores con distribución simétrica o con densidad (2.14). En el primer caso, la escala κ_0 es desconocida lo que provee un contexto más general al considerado en Du *et al.* (2013). En el segundo caso, como se describe en la Sección 2.5, el parámetro de ruido κ_0 depende de α_0 y corresponde a la constante de calibración. Como en el Capítulo 3, cuando κ_0 es conocido, tomamos $\hat{\kappa} = \kappa_0$.

4.1. Distribución asintótica de $\hat{\beta}$ y $\hat{\beta}_{\text{INI}}$

Recordemos que al considerar el modelo de regresión parcialmente lineal con errores simétricos, la función ρ está dada por (2.13). Por otro lado, para la situación en que el error u tiene una densidad de la forma (2.14), la función de pérdida está definida a través de (2.15). En particular, para el modelo log-Gamma $u_0 = 0$ y $d(s) = \exp(s) - s - 1$.

Es importante mencionar que los Lemas 2.1 y 2.2 muestran que, bajo el modelo (2.12), la condición de Fisher-consistencia fuerte requerida en el supuesto **N1** se cumple, bajo condiciones leves.

Supongamos que $\mathbb{E} \left(\chi(y, \mathbf{x}^T \beta_0 + \eta_0(t), \kappa_0) w(\mathbf{x}) \mid t \right) \neq 0$ y sea \mathbf{h}^* la función

$$\mathbf{h}^*(t) = \frac{\mathbb{E} \left(\chi(y, \mathbf{x}^T \beta_0 + \eta_0(t), \kappa_0) w(\mathbf{x}) \mathbf{x} \mid t \right)}{\mathbb{E} \left(\chi(y, \mathbf{x}^T \beta_0 + \eta_0(t), \kappa_0) w(\mathbf{x}) \mid t \right)}. \quad (4.1)$$

Denotemos como $\mathbf{A}(\boldsymbol{\theta}_0, \kappa_0)$ y $\mathbf{D}(\boldsymbol{\theta}_0, \kappa_0)$ a las matrices simétricas

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta}_0, \kappa_0) &= \mathbb{E} \chi(y, \mathbf{x}^T \beta_0 + \eta_0(t), \kappa_0) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [\mathbf{x} - \mathbf{h}^*(t)]^T, \\ \mathbf{D}(\boldsymbol{\theta}_0, \kappa_0) &= \mathbb{E} \Psi^2(y, \mathbf{x}^T \beta_0 + \eta_0(t), \kappa_0) w^2(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [\mathbf{x} - \mathbf{h}^*(t)]^T. \end{aligned}$$

En **N5** supondremos que la matriz $\mathbf{A}(\boldsymbol{\theta}_0, \kappa_0)$ es no singular lo que significa, en particular, que \mathbf{x} no puede ser perfectamente predicha por t .

Teorema 4.1 (Distribución asintótica). *Supongamos que **C2**, **C3** y **N0** a **N8** (en el apéndice) se verifican y que $k_n = O(n^\nu)$ con $\nu = 1/(1+2r)$. Sea $\hat{\kappa}$ un estimador preliminar consistente de κ_0 y sean $(\hat{\boldsymbol{\beta}}, \hat{\eta})$ los estimadores de $(\boldsymbol{\beta}_0, \eta_0)$ definidos en (2.2). Supongamos que $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ y que $\pi_{\mathbb{P}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$. Más aun, supongamos que $\mathbb{P}(\hat{\lambda}_j > \hat{\lambda}_{j-1}, \text{ para } 2 \leq j \leq k_n) \rightarrow 1$, donde $\hat{\eta}(t) = \sum_{j=1}^{k_n} \hat{\lambda}_j B_j(t)$. Entonces,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0)),$$

con $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0) = \mathbf{A}^{-1}(\boldsymbol{\theta}_0, \kappa_0) \mathbf{D}(\boldsymbol{\theta}_0, \kappa_0) \mathbf{A}^{-1}(\boldsymbol{\theta}_0, \kappa_0)$.

El Lema 4.2 del Apéndice da condiciones bajo las cuales los coeficientes de $\hat{\eta}$ satisfacen una desigualdad estricta. Notemos que $\pi_{\mathbb{P}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$ implica que $\|\hat{\eta} - \eta_0\|_{\infty} \xrightarrow{p} 0$.

Observación 4.1. Como en el Capítulo 3, es posible obtener la distribución asintótica del estimador $\hat{\boldsymbol{\beta}}_{\text{INI}}$ definido en la Sección 2.3. Más precisamente, supongamos que $k_n = O(n^\nu)$ con $\nu = 1/(1+2r)$, que $\hat{\kappa}$ un estimador preliminar consistente de κ_0 , que $\hat{\boldsymbol{\beta}}_{\text{INI}} \xrightarrow{p} \boldsymbol{\beta}_0$ y que $\pi_{\mathbb{P}}(\hat{\boldsymbol{\theta}}_{\text{INI}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$. Luego, bajo **C2**, **C3**, **N0**, **N1** y **N3** a **N8** donde, en la definición de las clases de funciones $\mathcal{E}_{n,j,h,\delta,\lambda_0}$, $\mathcal{F}_{n,j,\delta,\lambda_0}$ y $\mathcal{G}_{n,j,\delta,\lambda_0}$ dadas en (4.7), (4.8) y (4.9) se reemplaza $\boldsymbol{\lambda} \in \mathcal{L}_{k_n}$ por $\boldsymbol{\lambda} \in \mathbb{R}^{k_n}$, se cumple que $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{INI}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0))$, es decir, que $\hat{\boldsymbol{\beta}}_{\text{INI}}$ tiene la misma distribución asintótica que $\hat{\boldsymbol{\beta}}$. ♣

Observación 4.2. Cabe notar que, bajo el modelo (2.12), las funciones $\Psi(y, s, a) = \partial \rho(y, s, a) / \partial s$ y $\chi(y, s, a) = \partial \Psi(y, s, a) / \partial s$ pueden escribirse como $\Psi(y, s, a) = \psi^*(y-s, a)$ y $\chi(y, s, a) = \chi^*(y-s, a)$ para funciones adecuadas ψ^* y χ^* . En particular, cuando la función ρ está dada por (2.13) y ϕ es dos veces continuamente diferenciable, tenemos que $\psi^*(s, a) = -\phi'(s/a)$ y $\chi^*(s, a) = \phi''(s/a)$. Para el caso de errores con densidad (2.14), usando que ρ está definida a través de (2.15) y suponiendo que ϕ y v son funciones dos veces continuamente diferenciables, obtenemos que

$$\begin{aligned} \psi^*(s, a) &= \frac{1}{2a} \phi' \left(\frac{\sqrt{d(s)}}{a} \right) \frac{v'(s)}{\sqrt{d(s)}}, \\ \chi^*(s, a) &= \frac{1}{4a^2} \phi'' \left(\frac{\sqrt{d(s)}}{a} \right) \frac{[v'(s)]^2}{d(s)} - \frac{1}{4a} \phi' \left(\frac{\sqrt{d(s)}}{a} \right) \left\{ \frac{2v''(s)}{\sqrt{d(s)}} + \frac{[v'(s)]^2}{d(s)^{\frac{3}{2}}} \right\}. \end{aligned}$$

Tanto para errores simétricos como para errores con densidad (2.14), los estimadores clásicos se obtienen tomando $\phi(s) = \phi_{\text{CL}}(s) = s^2/2$. Por lo tanto, los estimadores no dependen de la elección de una constante de calibración preliminar, es decir que para los estimadores clásicos podemos tomar $a = \kappa_0 = 1$ al definir la función de pérdida. De esta forma, para errores simétricos, $\psi_{\text{CL}}^*(s) = -s$ y $\chi_{\text{CL}}^*(s) = 1$, mientras que para errores asimétricos $\psi_{\text{CL}}^*(s) = v'(s)$ y $\chi_{\text{CL}}^*(s) = -v''(s)$.

Más aun, bajo el modelo (2.12) el vector $\mathbf{h}^*(t)$, definido en (4.1), y las matrices $\mathbf{A}(\boldsymbol{\theta}_0, \kappa_0)$ y $\mathbf{D}(\boldsymbol{\theta}_0, \kappa_0)$ tienen una expresión más simple debida a la independencia entre los errores y las covariables. Más precisamente, a partir de

$$\mathbb{E} \left(\chi(y, \mathbf{x}^T \boldsymbol{\beta}_o + \eta_0(t), \kappa_0) w(\mathbf{x}) \mathbf{x} \middle| t \right) = \mathbb{E} \left(\chi^*(u, \kappa_0) w(\mathbf{x}) \mathbf{x} \middle| t \right) = \mathbb{E}(\chi^*(u, \kappa_0)) \mathbb{E} \left(w(\mathbf{x}) \mathbf{x} \middle| t \right),$$

obtenemos que

$$\mathbf{h}^*(t) = \frac{\mathbb{E} \left(w(\mathbf{x}) \mathbf{x} \middle| t \right)}{\mathbb{E} \left(w(\mathbf{x}) \middle| t \right)}, \quad (4.2)$$

lo que implica que $\mathbf{A}(\boldsymbol{\theta}_0, \kappa_0) = \mathbb{E} \chi^*(u, \kappa_0) \mathbf{A}_1$ y $\mathbf{D}(\boldsymbol{\theta}_0, \kappa_0) = \mathbb{E} \psi^{*2}(u, \kappa_0) \mathbf{D}_1$, donde las matrices \mathbf{A}_1 y \mathbf{D}_1 se definen como $\mathbf{A}_1 = \mathbb{E} w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [\mathbf{x} - \mathbf{h}^*(t)]^T$ y $\mathbf{D}_1 = \mathbb{E} w^2(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [\mathbf{x} - \mathbf{h}^*(t)]^T$. Por lo tanto, la matriz de covarianza asintótica está dada por

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0) = \frac{\mathbb{E} \psi^{*2}(u, \kappa_0)}{\{\mathbb{E} \chi^*(u, \kappa_0)\}^2} \mathbf{A}_1^{-1} \mathbf{D}_1 \mathbf{A}_1^{-1}.$$

Notemos que, como en regresión lineal, la varianza asintótica depende de la función de escores solo a través del término $e(\psi^*) = \mathbb{E} \psi^{*2}(u, \kappa_0) \{\mathbb{E} \chi^*(u, \kappa_0)\}^{-2}$. Por otro lado, cuando $w \equiv 1$ o cuando $w^2 = w$ como en el caso de funciones de peso de tipo *hard rejection*, $\mathbf{A}_1 = \mathbf{D}_1$ y la matriz de covarianza asintótica es igual a $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0) = e(\psi^*) \mathbf{A}_1^{-1}$. Luego, si $w \equiv 1$, la eficiencia asintótica de los estimadores de regresión está dada por

$$\text{eff} = \frac{e(\psi_{\text{CL}}^*)}{e(\psi^*)} = \frac{\mathbb{E} \psi_{\text{CL}}^{*2}(u)}{\{\mathbb{E} \chi_{\text{CL}}^*(u)\}^2} \left\{ \frac{\mathbb{E} \psi^{*2}(u, \kappa_0)}{\{\mathbb{E} \chi^*(u, \kappa_0)\}^2} \right\}^{-1},$$

que es la eficiencia de los estimadores robustos definidos para los modelos de regresión lineal. En particular, cuando los errores tienen distribución log-Gamma, $u \sim \log(\Gamma(\alpha_0, 1))$, la eficiencia puede obtenerse como

$$\text{eff} = \frac{1}{\alpha_0} \left\{ \frac{\mathbb{E} \psi^{*2}(u, \kappa_0)}{\{\mathbb{E} \chi^*(u, \kappa_0)\}^2} \right\}^{-1}.$$

♣

4.2. Estimación de la matriz de covarianza asintótica

Como en Lu (2014), es natural estimar $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0)$ por su contraparte empírica, sumergiendo los estimadores $\hat{\boldsymbol{\theta}}$, $\hat{\kappa}$ y $\hat{\mathbf{h}}^*(t)$ de $\boldsymbol{\theta}_0$, κ_0 y de la función $\mathbf{h}^*(t)$ definida en (4.1), respectivamente. Más precisamente, el estimador de la matriz de covarianza asintótica se obtiene como $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}, \hat{\kappa}) = \hat{\mathbf{A}}^{-1}(\hat{\boldsymbol{\theta}}, \hat{\kappa}) \hat{\mathbf{D}}(\hat{\boldsymbol{\theta}}, \hat{\kappa}) \hat{\mathbf{A}}^{-1}(\hat{\boldsymbol{\theta}}, \hat{\kappa})$, donde

$$\hat{\mathbf{A}}(\hat{\boldsymbol{\theta}}, \hat{\kappa}) = \frac{1}{n} \sum_{i=1}^n \chi(y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i), \hat{\kappa}) w(\mathbf{x}_i) \left[\mathbf{x}_i - \hat{\mathbf{h}}^*(t_i) \right] \left[\mathbf{x}_i - \hat{\mathbf{h}}^*(t_i) \right]^T, \quad (4.3)$$

$$\hat{\mathbf{D}}(\hat{\boldsymbol{\theta}}, \hat{\kappa}) = \frac{1}{n} \sum_{i=1}^n \Psi^2(y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i), \hat{\kappa}) w^2(\mathbf{x}_i) \left[\mathbf{x}_i - \hat{\mathbf{h}}^*(t_i) \right] \left[\mathbf{x}_i - \hat{\mathbf{h}}^*(t_i) \right]^T. \quad (4.4)$$

Para dar un estimador de \mathbf{h}^* en el caso particular del modelo parcialmente lineal isotónico (2.12), usaremos que vale (4.2). Notemos que $\mathbf{h}^*(t)$ minimiza sobre el espacio de funciones medibles p -dimensionales la cantidad $\Upsilon(\mathbf{h}) = \mathbb{E}w(\mathbf{x})\|\mathbf{x} - \mathbf{h}(t)\|^2$. Tomando en cuenta que h_j^* es continuamente diferenciable y que su r -ésima derivada satisface una condición de Lipschitz en $[0, 1]$, podemos aproximar h_j^* usando un elemento en $\mathcal{S}_n(\mathcal{T}_n, \ell)$. De hecho, dado $\boldsymbol{\xi}_j = (\xi_{j,1}, \dots, \xi_{j,k_n})^\top$, denotemos como $h_{j,\boldsymbol{\xi}_j}(t) = \sum_{m=1}^{k_n} \xi_{j,m} B_m(t)$ y $\mathbf{h}_{\boldsymbol{\xi}} = (h_{1,\boldsymbol{\xi}_1}, \dots, h_{p,\boldsymbol{\xi}_p})^\top$. Podemos definir $\widehat{h}_j^*(t)$ como $\widehat{h}_j^*(t) = \sum_{m=1}^{k_n} \widehat{\xi}_{j,m} B_m(t)$, donde $\widehat{\boldsymbol{\xi}}_j = (\widehat{\xi}_{j,1}, \dots, \widehat{\xi}_{j,k_n})^\top$ minimiza sobre $\boldsymbol{\xi}$,

$$\widehat{\Upsilon}(\mathbf{h}_{\boldsymbol{\xi}}) = \sum_{i=1}^n w(\mathbf{x}_i) \|\mathbf{x}_i - \mathbf{h}_{\boldsymbol{\xi}}(t_i)\|^2.$$

Debe mencionarse que, si las covariables \mathbf{x} y t son independientes, entonces $\mathbf{h}^*(t) = \mathbb{E}(w(\mathbf{x}) \mathbf{x}) \{\mathbb{E}w(\mathbf{x})\}^{-1}$ no depende de t y puede estimarse como $\widehat{\mathbf{h}}^*(t) = \sum_{i=1}^n w(\mathbf{x}_i) \mathbf{x}_i \{\sum_{i=1}^n w(\mathbf{x}_i)\}^{-1}$.

Cabe notar que bajo el modelo parcialmente lineal isotónico (2.12), dos estimadores de $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0)$ pueden darse. El usuario podría estimar $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0)$ como $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) = \widehat{\mathbf{A}}^{-1}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) \widehat{\mathbf{D}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) \widehat{\mathbf{A}}^{-1}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ donde $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ y $\widehat{\mathbf{D}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ están dados en (4.3) y (4.4) o utilizando las matrices $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ y $\widehat{\mathbf{D}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ definidas por

$$\widehat{\mathbf{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) = \frac{1}{n} \sum_{i=1}^n \chi^*(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} - \widehat{\eta}(t_i), \widehat{\kappa}) \frac{1}{n} \sum_{i=1}^n w(\mathbf{x}_i) \begin{bmatrix} \mathbf{x}_i - \widehat{\mathbf{h}}^*(t_i) \\ \mathbf{x}_i - \widehat{\mathbf{h}}^*(t_i) \end{bmatrix}^\top, \quad (4.5)$$

$$\widehat{\mathbf{D}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) = \frac{1}{n} \sum_{i=1}^n \left[\Psi^*(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} - \widehat{\eta}(t_i), \widehat{\kappa}) \right]^2 \frac{1}{n} \sum_{i=1}^n w^2(\mathbf{x}_i) \begin{bmatrix} \mathbf{x}_i - \widehat{\mathbf{h}}^*(t_i) \\ \mathbf{x}_i - \widehat{\mathbf{h}}^*(t_i) \end{bmatrix}^\top, \quad (4.6)$$

donde hemos usado que, bajo el modelo (2.12), la independencia entre los errores y las covariables implica que $\mathbf{A}(\boldsymbol{\theta}_0, \kappa_0) = \mathbb{E}\chi^*(u, \kappa_0) \mathbf{A}_1$ y $\mathbf{D}(\boldsymbol{\theta}_0, \kappa_0) = \mathbb{E}\psi^{*2}(u, \kappa_0) \mathbf{D}_1$. En nuestro estudio de simulación, al calcular los errores estándar asintóticos de los estimadores de regresión, usamos los estimadores de $\mathbf{A}(\boldsymbol{\theta}_0, \kappa_0)$ y $\mathbf{D}(\boldsymbol{\theta}_0, \kappa_0)$ dados en (4.5) y (4.6). Argumentos estándar permiten mostrar que $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ provee un estimador consistente de $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0)$ para ambas elecciones de los estimadores $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ y $\widehat{\mathbf{D}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$.

Asimismo, para definir un estimador de $\mathbf{h}^*(t)$ en un contexto general supongamos que, para todo t , $\mathbb{E}\left(\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_o + \eta_0(t), \kappa_0) w(\mathbf{x}) \mid t\right) > 0$. Notando que $\mathbf{h}^*(t)$ minimiza sobre el espacio de funciones medibles p -dimensionales la cantidad $\Upsilon(\mathbf{h}) = \mathbb{E}\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_o + \eta_0(t), \kappa_0) w(\mathbf{x}) \|\mathbf{x} - \mathbf{h}^*(t)\|^2$ y usando **N8**, podemos aproximar h_j^* usando un elemento de $\mathcal{S}_n(\mathcal{T}_n, \ell)$. Como antes, denotemos $h_{j,\boldsymbol{\xi}_j}(t) = \sum_{m=1}^{k_n} \xi_{j,m} B_m(t)$ y $\mathbf{h}_{\boldsymbol{\xi}} = (h_{1,\boldsymbol{\xi}_1}, \dots, h_{p,\boldsymbol{\xi}_p})^\top$ y definamos $\widehat{h}_j^*(t) = \sum_{m=1}^{k_n} \widehat{\xi}_{j,m} B_m(t)$, donde $\widehat{\boldsymbol{\xi}}_j = (\widehat{\xi}_{j,1}, \dots, \widehat{\xi}_{j,k_n})^\top$ minimiza sobre $\boldsymbol{\xi}$

$$\sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}} + \widehat{\eta}(t_i), \widehat{\kappa}) w(\mathbf{x}_i) \|\mathbf{x}_i - \mathbf{h}_{\boldsymbol{\xi}}(t_i)\|^2.$$

El estimador de la matriz de covarianza asintótica se obtiene como

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) = \widehat{\mathbf{A}}^{-1}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) \widehat{\mathbf{D}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa}) \widehat{\mathbf{A}}^{-1}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$$

donde las matrices $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ y $\widehat{\mathbf{D}}(\widehat{\boldsymbol{\theta}}, \widehat{\kappa})$ están definidas en (4.3) y (4.4).

Cuando la condición $\mathbb{E}\left(\chi(y, \mathbf{x}^\top \boldsymbol{\beta}_o + \eta_0(t), \kappa_0) w(\mathbf{x}) \mid t\right) > 0$, para todo t , no se cumple, el numerador y el denominador en $\mathbf{h}^*(t)$ deben estimarse por separado usando B-splines o cualquier otro suavizador.

4.3. Apéndice: Demostración del Teorema 4.1

Recordemos que hemos definido $Pf = \mathbb{E}f(y, \mathbf{x}, t)$ y $P_n f = (1/n) \sum_{i=1}^n f(y_i, \mathbf{x}_i, t_i)$. Cuando la función f depende de un parámetro desconocido γ estimado a través de $\hat{\gamma}$, $Pf_{\hat{\gamma}}$ indica que la esperanza se toma con respecto a $(y, \mathbf{x}^\top, t)^\top$ mientras que $(y_1, \mathbf{x}_1^\top, t_1)^\top, \dots, (y_n, \mathbf{x}_n^\top, t_n)^\top$ están fijos. Es decir, si llamamos $M(\gamma) = Pf_\gamma$ entonces $Pf_{\hat{\gamma}} = M(\hat{\gamma})$.

Por otra parte, para una clase de funciones \mathcal{F} ,

$$J_{[\cdot]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[\cdot]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon$$

indicaba la integral corchete y $\|\mathbb{G}_n\|_{\mathcal{F}} = \sqrt{n} \sup_{f \in \mathcal{F}} |(P_n - P)f|$. Recordemos que la norma Bernstein se define como $\|f\|_{P, \text{BERN}}^2 = 2 \mathbb{E} \{ \exp(|f|) - 1 - |f| \}$. La desigualdad $2(\exp(|u|) - 1 - |u|) \leq u^2 \exp(|u|)$ será útil cuando la función de peso no tenga soporte compacto.

Para $\boldsymbol{\lambda} \in \mathbb{R}^{k_n}$, $g_\lambda(t)$ denota el spline $g_\lambda(t) = \boldsymbol{\lambda}^\top \mathbf{B}(t)$ mientras que indicamos por $\boldsymbol{\theta}_{0, \lambda_0} = (\boldsymbol{\beta}_0, g_{\lambda_0})$ y $\boldsymbol{\theta}_\lambda = (\boldsymbol{\beta}, g_\lambda)$.

Como en la demostración del Teorema 3.9, sea $V_{\boldsymbol{\beta}, g, a} = \rho(y, \mathbf{x}^\top \boldsymbol{\beta} + g(t), a) w(\mathbf{x})$. Llamaremos $\mathbf{V}_{\mathbf{b}, g, a}^\beta = (V_{1, \mathbf{b}, g, a}^\beta, \dots, V_{p, \mathbf{b}, g, a}^\beta)^\top$ a su derivada parcial con respecto a $\boldsymbol{\beta}$ evaluada en \mathbf{b} , es decir,

$$\mathbf{V}_{\mathbf{b}, g, a}^\beta = \frac{\partial V_{\boldsymbol{\beta}, g, a}}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\mathbf{b}} = \Psi(y, \mathbf{x}^\top \mathbf{b} + g(t), a) w(\mathbf{x}) \mathbf{x},$$

mientras que $V_{\boldsymbol{\beta}, g, a}^\eta$ indica al operador definido como

$$V_{\boldsymbol{\beta}, g, a}^\eta[h] = \frac{\partial V_{\boldsymbol{\beta}, g, a}^{\eta+s h, a}}{\partial s} \Big|_{s=0} = \Psi(y, \mathbf{x}^\top \boldsymbol{\beta} + g(t), a) w(\mathbf{x}) h(t),$$

donde $h \in \mathcal{G}_0$ y \mathcal{G}_0 es la clase de funciones medibles sobre $\mathcal{I} = [0, 1]$ definida en (2.10). Más aun, para $\mathbf{h} = (h_1, \dots, h_p) \in \mathcal{H}^p$, denotamos por

$$V_{\boldsymbol{\beta}, g, a}^\eta[\mathbf{h}] = \left(V_{\boldsymbol{\beta}, g, a}^\eta[h_1], \dots, V_{\boldsymbol{\beta}, g, a}^\eta[h_p] \right)^\top$$

y por $\eta_s(t) = \eta_0(t) + sh(t)$ para $s > 0$. Si η_0 es estrictamente creciente, para un s suficientemente chico, cualquier η_s será también creciente si h es derivable con derivada continua.

Definamos para cualquier $\boldsymbol{\beta}, g$ y $a \in \mathcal{K}$ la función

$$\mathbf{W}_{\boldsymbol{\beta}, g, a}^\beta = \Psi(y, \mathbf{x}^\top \boldsymbol{\beta} + g(t), a) w(\mathbf{x}) (\mathbf{x} - \mathbf{h}^*(t)),$$

donde \mathbf{h}^* está definido en (4.1). Para aliviar la notación, llamemos

$$\mathbf{W}^\beta = \mathbf{W}_{\boldsymbol{\beta}_0, \eta_0, \kappa_0}^\beta = \mathbf{V}_{\boldsymbol{\beta}_0, \eta_0, \kappa_0}^\beta - V_{\boldsymbol{\beta}_0, \eta_0, \kappa_0}^\eta[\mathbf{h}^*].$$

Para cada $1 \leq j \leq p$, una función dada $h \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ tal que $\|h_j^* - h\|_\infty < \delta$ y un spline monótono $g_{\lambda_0} \in \mathcal{M}_n(\mathcal{T}_n, \ell)$ llamemos

$$\mathcal{E}_{n, j, h, \delta, \lambda_0} = \{ f(y, \mathbf{x}, t) = V_{\boldsymbol{\beta}, g_\lambda, a}^\eta[h_j^* - h], \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon_0, a \in \mathcal{V}, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, \|g_\lambda - g_{\lambda_0}\|_\infty < \epsilon_0 \}. \quad (4.7)$$

Definimos también

$$\mathcal{F}_{n, j, \delta, \lambda_0} = \{ f(y, \mathbf{x}, t) = V_{\boldsymbol{\beta}, g_\lambda, a}^\eta[h_j^*] - V_{\boldsymbol{\beta}_0, \eta_0, a}^\eta[h_j^*], \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon_0, a \in \mathcal{V}, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, \pi_{\mathbb{P}}(\boldsymbol{\theta}_{0, \lambda_0}, \boldsymbol{\theta}_\lambda) < \delta \} \quad (4.8)$$

$$\mathcal{G}_{n, j, \delta, \lambda_0} = \{ f(y, \mathbf{x}, t) = V_{j, \boldsymbol{\beta}, g_\lambda, a}^\beta - V_{j, \boldsymbol{\beta}_0, \eta_0, a}^\beta, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon_0, a \in \mathcal{V}, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}, \pi_{\mathbb{P}}(\boldsymbol{\theta}_{0, \lambda_0}, \boldsymbol{\theta}_\lambda) < \delta \}. \quad (4.9)$$

Como en el Capítulo 3, entendemos en lo que sigue que si κ_0 es conocido entonces $\widehat{\kappa} = \kappa_0$ y el entorno \mathcal{V} de κ_0 se reduce a κ_0 . Además, en **N3**, la función $\rho(y, s, a)$ corresponde a $\rho(y, s, \kappa_0)$.

Para obtener resultados sobre la distribución asintótica de $\widehat{\boldsymbol{\beta}}$, necesitamos los siguientes supuestos.

N0 Se cumple una de las siguientes condiciones

- a) κ_0 es conocido.
- b) Las observaciones siguen un modelo de regresión parcialmente lineal (2.12) donde el error u es simétrico o tiene función de densidad $f_0(u, \alpha_0)$ dada por (2.14) con un parámetro $\alpha_0 > 0$ desconocido y una función continua v con única moda en u_0 .

N1 Para todo $a \in \mathcal{K}$, $L(\boldsymbol{\beta}_0, \eta_0, a) = \min_{\boldsymbol{\beta} \in \mathbb{R}^p, g \in \mathcal{G}_0} L(\boldsymbol{\beta}, g, a)$ donde \mathcal{G}_0 está definida en (2.10).

N2 La función η_0 es estrictamente creciente y derivable.

N3 La función $\rho(y, s, a)$ es tres veces continuamente diferenciable con respecto a s con derivadas

$$\Psi(y, s, a) = \frac{\partial \rho(y, s, a)}{\partial s}, \quad \chi(y, s, a) = \frac{\partial \Psi(y, s, a)}{\partial s}, \quad \chi_1(y, s, a) = \frac{\partial \chi(y, s, a)}{\partial s},$$

tales que

$$\begin{aligned} \|\Psi\|_{\infty, \mathcal{V}} &= \sup_{y \in \mathbb{R}, s \in \mathbb{R}, a \in \mathcal{V}} |\Psi(y, s, a)| < \infty \\ \|\chi\|_{\infty, \mathcal{V}} &= \sup_{y \in \mathbb{R}, s \in \mathbb{R}, a \in \mathcal{V}} |\chi(y, s, a)| < \infty, \\ \|\chi_1\|_{\infty, \mathcal{V}} &= \sup_{y \in \mathbb{R}, s \in \mathbb{R}, a \in \mathcal{V}} |\chi_1(y, s, a)| < \infty. \end{aligned}$$

Más aun, cuando κ_0 es desconocido, $\chi(y, s, a)$ es una función continua de a .

N4 La función de peso $w(\mathbf{x})$ es acotada y se cumple alguna de las siguientes condiciones

- a) Cuando $r > 1$, $\mathbb{E}w(\mathbf{x})\|\mathbf{x}\|^2 x_j^4 < \infty$, para $1 \leq j \leq p$, mientras que para $r = 1$, $\mathbb{E}w(\mathbf{x})\|\mathbf{x}\|^2 x_j^8 < \infty$ para $1 \leq j \leq p$.
- b) Cuando $r > 1$, $\mathbb{E}w(\mathbf{x})x_j^4 < \infty$, para $1 \leq j \leq p$, mientras que si $r = 1$, $\mathbb{E}w(\mathbf{x})x_j^8 < \infty$, para $1 \leq j \leq p$. Más aun, $\pi(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$, cuando $\pi^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^2 + \|\eta_1 - \eta_2\|_2^2$.

Sin pérdida de generalidad, suponemos que $\|w\|_{\infty} = 1$.

N5 La matriz $\mathbf{A}(\boldsymbol{\theta}_0, \kappa_0) = \mathbb{E}\chi(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t), \kappa_0)w(\mathbf{x})[\mathbf{x} - \mathbf{h}^*(t)][\mathbf{x} - \mathbf{h}^*(t)]^T$ es no singular.

N6 La integral de entropía de las clases $\mathcal{E}_{n,j,h,\delta,\lambda_0}$ y $\mathcal{F}_{n,j,\delta,\lambda_0}$ definidas, respectivamente, en (4.7) y (4.8) pueden acotarse por

$$J_{[\cdot]}(A_1 \delta, \mathcal{E}_{n,j,h,\delta,\lambda_0}, L_2(P)) \leq C \delta \sqrt{k_n} \quad \text{y} \quad J_{[\cdot]}(C_j \delta, \mathcal{F}_{n,j,\delta,\lambda_0}, L_2(P)) \leq C \delta \sqrt{k_n},$$

donde $A_1 = \|\Psi\|_{\infty}$, $C_j = 2\|\chi\|_{\infty}\|h_j^*\|_{\infty}$ y C es una constante genérica independiente de n y δ .

N7 a) Cuando $w(\mathbf{x})\mathbf{x}$ es acotada, la integral de entropía de la clase $\mathcal{G}_{n,j,h,\delta,\lambda_0}$ dada en (4.9) puede acotarse por

$$J_{[\cdot]}(D_j\delta, \mathcal{G}_{n,j,\delta,\lambda_0}, L_2(P)) \leq C\delta\sqrt{k_n},$$

con $D_j = 2\|\chi\|_\infty\|w(\mathbf{x})x_j\|_\infty$ y C es una constante genérica independiente de n y δ .

b) Si $w(\mathbf{x})\mathbf{x}$ no es acotada, $\mathbb{E}w^2(\mathbf{x})\|\mathbf{x}\|^2 \exp(2\|\Psi\|_\infty w(\mathbf{x})\|\mathbf{x}\|) < \infty$ y

$$J_{[\cdot]}(\delta, \mathcal{G}_{n,j,\delta,\lambda_0}, \|\cdot\|_{P,\text{BERN}}) \leq C\delta\sqrt{k_n},$$

donde C es una constante genérica independiente de n y δ .

N8 Para cada $1 \leq j \leq p$, la familia de funciones

$$\mathcal{F}_j = \{f(y, \mathbf{x}, t) = V_{\beta_0, \eta_0, a}^\eta[h_j^*] = \Psi(y, \mathbf{x}^\top \beta_0 + \eta_0(t), a) w(\mathbf{x})h_j^*(t); a \in \mathcal{V}\}$$

es Donsker.

N9 La función h_j^* es continuamente diferenciable y su r -ésima derivada satisface una condición Lipschitz en $[0, 1]$, con $r \geq 1$.

Observación 4.3. En la Sección 2.7 discutimos condiciones que aseguran la consistencia de Fisher fuerte solicitada en **N1**. Esta condición puede reemplazarse por la consistencia Fisher condicional pedida en la hipótesis **C10b**).

El supuesto **N2** nos permite asegurar que $\hat{\eta} + sh$ será no decreciente para cualquier $h \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ (ver Lema 4.2). Esta condición también es un requerimiento en Lu *et al.* (2007, 2009).

La hipótesis **N3** es un requisito estándar para estimadores robustos. Bajo el modelo (2.12), si los errores tienen distribución simétrica, se verifica **N3** cuando la función $\phi : \mathbb{R} \rightarrow [0, \infty)$ es tres veces continuamente diferenciable con derivadas acotadas. Por otra parte, si los errores tienen una densidad dada por (2.14), se debe cumplir además que la función v sea tres veces continuamente diferenciable.

Para las funciones de peso usualmente consideradas en robustez, se cumple **N4(a)** porque $w(\mathbf{x})$ tiene soporte compacto. Un ejemplo de tal función de peso está dada por $w(\mathbf{x}) = w_T(d(\mathbf{x}, \boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_\mathbf{x}))$ donde w_T es la función de pesos bicuadrada y $d(\mathbf{x}, \boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_\mathbf{x})$ es la distancia de Mahalanobis definida como $d^2(\mathbf{x}, \boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})^\top \boldsymbol{\Sigma}_\mathbf{x}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})$ y calculada con los parámetros de posición $\boldsymbol{\mu}_\mathbf{x}$ y dispersión $\boldsymbol{\Sigma}_\mathbf{x}$ de las covariables. Por otro lado, para garantizar que nuestros resultados incluyan el caso $w \equiv 1$ con condiciones sobre los momentos más débiles que las dadas en **N4(a)**, el supuesto **N4(b)** pide también que $\pi(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$. Los Lemas 3.15 y 3.16 del Capítulo 3.2 dan condiciones que aseguran que $\pi(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$ vale.

La condición **N5** evita que cualquier elemento \mathbf{x} sea, casi seguramente, perfectamente predicho por t porque el modelo resultaría completamente no paramétrico. Más aun, es un requerimiento estándar en regresión robusta para obtener estimadores de β_0 con tasa \sqrt{n} .

Argumentos similares a los considerados en la Sección 3.5.2 permiten mostrar que **N6**, **N7(a)** y **N8** se cumplen para algunos modelos, como el parcialmente lineal y el modelo logístico. La cota dada en **N7(b)** es similar a la obtenida en el Lema A.1 en Lu (2015). La validez de este supuesto para los modelos parcialmente lineales puede derivarse combinando los argumentos considerados en la Sección 3.5.2 y en la prueba del Lema A.1 en Lu (2015). Por otro lado, al considerar el modelo parcialmente lineal con parámetro de escala conocido, Du *et al.* (2013) requieren que la covariable \mathbf{x} tenga soporte acotado, $\mathbb{P}(\|\mathbf{x}\| < A) = 1$ para algún $A > 0$, para evitar considerar cubrimientos corchete con respecto a la norma de Bernstein.

Cuando κ_0 es conocido, \mathcal{F}_j se reduce a la función $\Psi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + \eta_0(t), \kappa_0) w(\mathbf{x}) h_j^*(t)$ y **N8** se cumple automáticamente. ♣

Lema 4.2. Sean $(\hat{\boldsymbol{\beta}}, \hat{\eta})$ los estimadores de $(\boldsymbol{\beta}_0, \eta_0)$ definidos en (2.2). Supongamos que **C2**, **C3** y **N2** se cumplen, $r > 2$, $k_n = O(n^{1/(1+2r)})$ y $\|\hat{\eta} - \eta_0\|_2 = O_{\mathbb{P}}(n^{-r/(1+2r)})$. Entonces, si llamamos $\hat{\eta} = \hat{\boldsymbol{\lambda}}^\top \mathbf{B}$, $\mathbb{P}(\hat{\lambda}_i > \hat{\lambda}_{i-1}, 2 \leq i \leq k_n) \rightarrow 1$. Por lo tanto, con probabilidad que tiende a 1, para s suficientemente pequeño, $\hat{\eta} + sh \in \mathcal{M}_n(\mathcal{T}_n, \ell)$ para todo $h \in \mathcal{S}_n(\mathcal{T}_n, \ell)$.

DEMOSTRACIÓN. Para todo spline g_λ , existe una constante $A > 1$ que depende solo de su orden ℓ , tal que $(1/(A k_n^2)) \|\boldsymbol{\lambda}\|_2^2 \leq \|g_\lambda\|_2^2 \leq (A/k_n^2) \|\boldsymbol{\lambda}\|_2^2$. Sea $g_{0,n} = \boldsymbol{\lambda}_0^\top \mathbf{B} \in \mathcal{M}_n(\mathcal{T}_n, \ell)$ tal que $\|g_{0,n} - \eta_0\|_\infty = O(n^{-r/(1+2r)})$. Entonces, $\|\hat{\eta} - g_{0,n}\|_2 = O_{\mathbb{P}}(n^{-r/(1+2r)})$, lo que implica que

$$S^2 = \sum_{i=2}^{k_n} D_i^2 = \sum_{i=2}^{k_n} \left(\hat{\lambda}_i - \hat{\lambda}_{i-1} - (\lambda_{0,i} - \lambda_{0,i-1}) \right)^2 \leq 2A \|\hat{\eta} - g_{0,n}\|_2^2 k_n^2.$$

Usando que $k_n = O(n^{1/(1+2r)})$ obtenemos que $S^2 = O_{\mathbb{P}}(n^{(2-2r)/(1+2r)})$. Por otro lado, notemos que los coeficientes de $\lambda_{0,i}$ son los valores de la función η_0 en una grilla creciente de puntos obtenida a partir de los nodos cuando ℓ es par o promediando dos valores consecutivos de los nodos cuando ℓ es impar. Luego, usando que $C_0 = \inf_s \eta_0'(s) > 0$ y **C3** obtenemos que

$$\begin{aligned} \hat{\lambda}_i - \hat{\lambda}_{i-1} &\geq (\lambda_{0,i} - \lambda_{0,i-1}) - |D_i| \geq (\lambda_{0,i} - \lambda_{0,i-1}) - S \\ &\geq C_0 C_1 n^{-1/(1+2r)} - O_{\mathbb{P}}(n^{(1-r)/(1+2r)}) \\ &\geq n^{-1/(1+2r)} \left\{ C_0 C_1 - O_{\mathbb{P}}(1) n^{(2-r)/(1+2r)} \right\}. \end{aligned}$$

Por lo tanto, como $r > 2$, deducimos que

$$\hat{\lambda}_i - \hat{\lambda}_{i-1} \geq n^{-1/(1+2r)} \{C_0 C_1 - o_{\mathbb{P}}(1)\},$$

con lo cual, $\hat{\lambda}_i - \hat{\lambda}_{i-1} > 0$, $2 \leq i \leq k_n$, con probabilidad que converge a 1. ■

Para probar el Teorema 4.1, verificaremos las condiciones del siguiente lema que es una ligera modificación del Teorema 3 en Zhang *et al.* (2010).

Lema 4.3. Sea $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \eta_0)$ y $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ un estimador consistente de $\boldsymbol{\theta}_0$. Supongamos que se cumplen

$$\mathbf{H1} \quad P_n \mathbf{V}_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\beta = o_{\mathbb{P}}(n^{-1/2}) \text{ y } P_n V_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\eta[\mathbf{h}^*] = o_{\mathbb{P}}(n^{-1/2}),$$

$$\mathbf{H2} \quad \begin{aligned} \text{a) } & (P_n - P)[\mathbf{V}_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\beta - \mathbf{V}_{\boldsymbol{\beta}_0, \eta_0, \kappa_0}^\beta] = o_{\mathbb{P}}(n^{-1/2}) \text{ y} \\ \text{b) } & (P_n - P) \left\{ V_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\eta[\mathbf{h}^*] - V_{\boldsymbol{\beta}_0, \eta_0, \kappa_0}^\eta[\mathbf{h}^*] \right\} = o_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

$$\mathbf{H3} \quad P[\mathbf{W}_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\beta - \mathbf{W}_{\boldsymbol{\beta}_0, \eta_0, \kappa_0}^\beta] = -\mathbf{A}_{\boldsymbol{\theta}_0, \widehat{\kappa}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_{\mathbb{P}}(n^{-1/2}).$$

Si $\mathbf{A}_{\boldsymbol{\theta}_0, \widehat{\kappa}} \xrightarrow{p} \mathbf{A}_{\boldsymbol{\theta}_0, \kappa_0}$ y $\mathbf{A}_{\boldsymbol{\theta}_0, \kappa_0}$ es no singular, resulta que

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = n^{1/2} \mathbf{A}_{\boldsymbol{\theta}_0, \kappa_0}^{-1} P_n \mathbf{W}_{\boldsymbol{\theta}_0, \kappa_0}^\beta + o_{\mathbb{P}}(1).$$

Por lo tanto, si $\mathbf{D}_{\boldsymbol{\theta}_0, \kappa_0} = \mathbb{E} \mathbf{W}_{\boldsymbol{\theta}_0, \kappa_0}^\beta \mathbf{W}_{\boldsymbol{\theta}_0, \kappa_0}^{\beta \top}$, tenemos que

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N\left(\mathbf{0}, \mathbf{A}_{\boldsymbol{\theta}_0, \kappa_0}^{-1} \mathbf{D}_{\boldsymbol{\theta}_0, \kappa_0} \mathbf{A}_{\boldsymbol{\theta}_0, \kappa_0}^{-1 \top}\right).$$

DEMOSTRACIÓN DEL TEOREMA 4.1. En orden, probaremos las hipótesis del Lema 4.3 del cual seguirá el resultado.

i) Comenzaremos probando **H1**. Recordemos que $L_n(\boldsymbol{\beta}, g, a) = P_n V_{\boldsymbol{\beta}, g, a}$, $L(\boldsymbol{\beta}, g, a) = P V_{\boldsymbol{\beta}, g, a}$ y que $(\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ minimiza $L_n(\boldsymbol{\beta}, g, \widehat{\kappa})$ sobre $\boldsymbol{\beta} \in \mathbb{R}^p$ y $g \in \mathcal{M}_n(\mathcal{T}_n, \ell)$. Por lo tanto, $\widehat{\boldsymbol{\beta}}$ minimiza $L_n(\boldsymbol{\beta}, \widehat{\eta}, \widehat{\kappa})$ lo que implica que $P_n \mathbf{V}_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\beta = 0$ de donde se deduce que solo debemos verificar que

$$P_n V_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\eta[\mathbf{h}^*] = o_{\mathbb{P}}(n^{-1/2}). \quad (4.10)$$

Sea $\mathcal{B}_n = \{\omega \in \Omega : \widehat{\lambda}_i > \widehat{\lambda}_{i-1} \text{ for } 2 \leq i \leq k_n\}$. Tenemos que $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{B}_n) = 1$. Entonces, para cualquier $\omega \in \mathcal{B}_n$, se cumple que, para cualquier $h \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ fijo, la función $\widehat{\eta} + sh$ también pertenece $\mathcal{M}_n(\mathcal{T}_n, \ell)$, para s suficientemente chico, digamos $s < s_h$.

Especifiquemos el submodelo paramétrico suave como $\{\widehat{\eta} + sh\}$ para $s > 0$, donde $h \in \mathcal{S}_n(\mathcal{T}_n, \ell)$. Considerando que $(\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ minimiza $L_n(\boldsymbol{\beta}, g, \widehat{\kappa})$ sobre $\boldsymbol{\beta} \in \mathbb{R}^p$ y $g \in \mathcal{M}_n(\mathcal{T}_n, \ell)$, tenemos que $\widehat{\eta}$ minimiza $P_n V_{\widehat{\boldsymbol{\beta}}, g, \widehat{\kappa}}$ sobre $\{g = \widehat{\eta} + sh; s < s_h\}$, con lo cual

$$\frac{\partial}{\partial s} P_n V_{\widehat{\boldsymbol{\beta}}, \widehat{\eta} + sh, \widehat{\kappa}} \Big|_{s=0} = P_n V_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\eta[h] = 0, \quad (4.11)$$

para cualquier $h \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ fijo.

Usando **N9** y **C3**, obtenemos, a partir de Schumaker (1981), que, para $j = 1, \dots, p$, existe $h_{n,j} \in \mathcal{S}_n(\mathcal{T}_n, \ell)$ tal que $\|h_j^* - h_{n,j}\|_\infty = O(n^{-r/(1+2r)})$. Luego, usando (4.11) y que $\mathbb{P}(\mathcal{B}_n) \rightarrow 1$, concluimos que para probar (4.10) basta ver que

$$P_n V_{\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}}^\eta[h_j^* - h_{n,j}] = o_{\mathbb{P}}(n^{-1/2}). \quad (4.12)$$

El término $P_n V_{\widehat{\beta}, \widehat{\eta}, \widehat{\kappa}}^\eta [h_j^* - h_{n,j}]$ puede escribirse como $T_1 + T_2$ donde $T_1 = (P_n - P) V_{\widehat{\beta}, \widehat{\eta}, \widehat{\kappa}}^\eta [h_j^* - h_{n,j}]$ y $T_2 = P V_{\widehat{\beta}, \widehat{\eta}, \widehat{\kappa}}^\eta [h_j^* - h_{n,j}]$.

Como en el Teorema 3.11, sea $g_{0,n} \in \mathcal{M}_n(\mathcal{T}_n, \ell)$ tal que $\|g_{0,n} - \eta_0\|_\infty = O(n^{-r/(1+2r)})$. Llamemos $g_{0,n} = \boldsymbol{\lambda}_{0,n}^\top \mathbf{B}$ y $\boldsymbol{\theta}_{0,n} = (\boldsymbol{\beta}_0, g_{0,n})$. Entonces, usando que $\|\widehat{\eta} - \eta_0\|_\infty \xrightarrow{p} 0$, obtenemos $\|\widehat{\eta} - g_{0,n}\|_\infty \xrightarrow{p} 0$.

Consideremos la familia de funciones $\mathcal{E}_{n,j,h,\delta,\lambda_0}$ definida en (4.7) con $h = h_{n,j} \in \mathcal{S}_n(\mathcal{T}_n, \ell)$, $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_{0,n}$ y $\delta = \delta_n = 2 \|h_j^* - h_{n,j}\|_\infty$. Para no recargar la notación, sea $\mathcal{E}_n = \mathcal{E}_{n,j,h_{n,j},\delta_n,\lambda_{0,n}}$. Usando que $\|w\|_\infty = 1$, tenemos, para cualquier $f \in \mathcal{E}_n$,

$$\|V_{\boldsymbol{\beta}, g_{\boldsymbol{\lambda}}, a}^\eta [h_j^* - h]\|_\infty \leq \|\Psi\|_\infty \|h_j^* - h\|_\infty \leq M(\delta),$$

donde $M(\delta) = \|\Psi\|_\infty \delta = A_1 \delta$. Más aun,

$$P f^2 = \mathbb{E} [\Psi(y, \mathbf{x}^\top \boldsymbol{\beta} + \boldsymbol{\lambda}^\top \mathbf{B}(t), a)(h_j^*(t) - h(t))w(\mathbf{x})]^2 \leq \|\Psi\|_\infty^2 \|h_j^* - h\|_\infty^2 \leq M^2(\delta).$$

El Lema 3.8 implica que

$$\begin{aligned} \mathbb{E}^* \|G_n\|_{\mathcal{E}_n} &\leq J_{[\cdot]}(M(\delta), \mathcal{E}_n, L_2(P)) \left(1 + \frac{J_{[\cdot]}(M(\delta), \mathcal{E}_n, L_2(P))}{M^2(\delta)\sqrt{n}} M(\delta) \right) \\ &= J_{[\cdot]}(M(\delta), \mathcal{E}_{n,j,h,\delta}, L_2(P)) \left(1 + \frac{J_{[\cdot]}(M(\delta), \mathcal{E}_{n,j,h,\delta}, L_2(P))}{M(\delta)\sqrt{n}} \right), \end{aligned}$$

que junto con **N6** conducen a

$$\mathbb{E}^* \|G_n\|_{\mathcal{E}_n} \leq C \delta k_n^{1/2} \left(1 + \frac{C k_n^{1/2}}{\|\Psi\|_\infty \sqrt{n}} \right).$$

Recordemos que $k_n = O(n^{1/(1+2r)})$, $\|h_j^* - h_{n,j}\|_\infty = O(n^{-r/(1+2r)})$. Entonces, usando que $\delta = 2 \|h_j^* - h_{n,j}\|_\infty$, deducimos que

$$\begin{aligned} \mathbb{P} \left(\sqrt{n} |T_1| > \epsilon \cap \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\widehat{\eta} - g_{0,n}\|_\infty < \epsilon_0 \right) &\leq \frac{1}{\epsilon} \mathbb{E}^* \|G_n\|_{\mathcal{E}_{n,j,h,\delta}} \leq \frac{1}{\epsilon} C \delta k_n^{1/2} \left(1 + \frac{C k_n^{1/2}}{\|\Psi\|_\infty \sqrt{n}} \right) \\ &\leq \frac{1}{\epsilon} C_1 n^{-\frac{r}{1+2r}} n^{\frac{1}{2(1+2r)}} \left(1 + \frac{C}{\|\Psi\|_\infty} n^{\frac{1}{2(1+2r)}} n^{-\frac{1}{2}} \right) \\ &\leq \frac{1}{\epsilon} C_1 n^{\frac{1-2r}{1+2r}} \left(1 + \frac{C}{\|\Psi\|_\infty} n^{-\frac{2r}{2(1+2r)}} \right), \end{aligned}$$

que converge a 0 porque $r \geq 1$. Notando que $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\widehat{\eta} - g_{0,n}\|_\infty \xrightarrow{p} 0$, obtenemos que $T_1 = o_{\mathbb{P}}(n^{-1/2})$.

Para concluir la demostración de (4.12) resta mostrar que $T_2 = o_{\mathbb{P}}(n^{-1/2})$. Recordemos que la condición de Fisher-consistencia dada en **N1** implica que $P V_{\boldsymbol{\beta}_0, \eta_0, \widehat{\kappa}}^\eta [h_j^* - h_{n,j}] = 0$, luego,

$$T_2 = P \left(V_{\widehat{\beta}, \widehat{\eta}, \widehat{\kappa}}^\eta - V_{\boldsymbol{\beta}_0, \eta_0, \widehat{\kappa}}^\eta \right) [h_j^* - h_{n,j}].$$

Llamemos ζ_{β_0} y $\zeta_{\eta_0}(t)$ valores intermedios entre β_0 y $\hat{\beta}$ y $\eta_0(t)$ y $\hat{\eta}(t)$, respectivamente. Entonces, usando nuevamente que $\|w\|_\infty = 1$, por el Teorema de Valor Medio deducimos que

$$\begin{aligned}
|T_2| &= |\mathbb{E}\chi(y, \mathbf{x}^\top \zeta_{\beta_0} + \zeta_{\eta_0}(t), \hat{\kappa}) \left[\mathbf{x}^\top (\beta_0 - \hat{\beta}) + (\eta_0 - \hat{\eta}) \right] w(\mathbf{x})(h_j^* - h_{n,j})| \\
&\leq \|\chi\|_\infty \|h_j^* - h_{n,j}\|_\infty \mathbb{E} \left| \mathbf{x}^\top (\beta_0 - \hat{\beta}) + (\eta_0 - \hat{\eta}) \right| w(\mathbf{x}) \\
&\leq \|\chi\|_\infty \|h_j^* - h_{n,j}\|_\infty \left\{ \mathbb{E} \left[\mathbf{x}^\top (\beta_0 - \hat{\beta}) + (\eta_0 - \hat{\eta}) \right]^2 w^2(\mathbf{x}) \right\}^{1/2} \\
&\leq \|\chi\|_\infty \|w\|_\infty^{1/2} \|h_j^* - h_{n,j}\|_\infty \left\{ \mathbb{E} \left[\mathbf{x}^\top (\beta_0 - \hat{\beta}) + (\eta_0 - \hat{\eta}) \right]^2 w(\mathbf{x}) \right\}^{1/2} \\
&\leq \|\chi\|_\infty \|h_j^* - h_{n,j}\|_\infty \pi_{\mathbb{P}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0).
\end{aligned}$$

Usando que $\|h_j^* - h_{n,j}\|_\infty = O(n^{-r/(1+2r)})$ y $\pi_{\mathbb{P}}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$ concluimos que $|T_2| \leq O_{\mathbb{P}}(n^{-2r/(1+2r)}) = o_{\mathbb{P}}(n^{-1/2})$ como se quería probar.

ii) Debemos mostrar que se cumple **H2**. solo mostraremos **H2(b)**, porque la demostración de **H2(a)** es muy similar usando los Lemas 3.8 o 3.8 según sea o no acotada $w(\mathbf{x})$.

Notemos que $(P_n - P) \left\{ V_{\hat{\beta}, \hat{\eta}, \hat{\kappa}}^\eta[\mathbf{h}^*] - V_{\beta_0, \eta_0, \kappa_0}^\eta[\mathbf{h}^*] \right\} = \mathbf{S}_{1,n} + \mathbf{S}_{2,n}$, donde

$$\begin{aligned}
\mathbf{S}_{1,n} &= (P_n - P) \left\{ V_{\hat{\beta}, \hat{\eta}, \hat{\kappa}}^\eta[\mathbf{h}^*] - V_{\beta_0, \eta_0, \hat{\kappa}}^\eta[\mathbf{h}^*] \right\}, \\
\mathbf{S}_{2,n} &= (P_n - P) \left\{ V_{\beta_0, \eta_0, \hat{\kappa}}^\eta[\mathbf{h}^*] - V_{\beta_0, \eta_0, \kappa_0}^\eta[\mathbf{h}^*] \right\}.
\end{aligned}$$

Usando que la familia \mathcal{F}_j definida en **N8** es Donsker y que $\hat{\kappa} \xrightarrow{p} \kappa_0$, resulta $\sqrt{n}S_{2,j,n} = o_{\mathbb{P}}(n^{-1/2})$, donde $\mathbf{S}_{2,n} = (S_{2,1,n}, \dots, S_{2,p,n})^\top$. Para concluir la prueba de **H2(b)**, falta mostrar que $\sqrt{n}S_{1,j,n} = o_{\mathbb{P}}(n^{-1/2})$, para $1 \leq j \leq p$, con $\mathbf{S}_{1,n} = (S_{1,1,n}, \dots, S_{1,p,n})^\top$.

Fijemos $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_{0,n}$ donde $g_{0,n} = \boldsymbol{\lambda}_{0,n}^\top \mathbf{B} \in \mathcal{M}_n(\mathcal{T}_n, \ell)$ es tal que $\|g_{0,n} - \eta_0\|_\infty = O(n^{-r/(1+2r)})$ y tomemos $\delta = \delta_n = n^{-\gamma r/(1+2r)}$ con $\gamma = 3/4$. Entonces, $\pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,n}, \boldsymbol{\theta}_0) < \delta$ para $n \geq n_0$ con $\boldsymbol{\theta}_{0,n} = (\beta_0, g_{0,n})$. Fijemos $n \geq n_0$.

Considerando que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,n}, \hat{\boldsymbol{\theta}}) = O_{\mathbb{P}}(n^{-r/(1+2r)})$, obtenemos $n^{\gamma r/(1+2r)} \pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,n}, \hat{\boldsymbol{\theta}}) \xrightarrow{p} 0$. Además, $\hat{\beta} \xrightarrow{p} \beta_0$, luego, con probabilidad que converge a 1, $\|\hat{\beta} - \beta_0\| < \epsilon_0$ y $\pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,n}, \hat{\boldsymbol{\theta}}) < \delta$, de donde se deduce que $V_{\hat{\beta}, \hat{\eta}, \hat{\kappa}}^\eta[h_j^*] - V_{\beta_0, \eta_0, \hat{\kappa}}^\eta[h_j^*] \in \mathcal{F}_{n,j,\delta_n, \lambda_{0,n}}$.

Por simplicidad, llamemos $\mathcal{F}_n^* = \mathcal{F}_{n,j,\delta_n, \lambda_{0,n}}$ donde $\mathcal{F}_{n,j,\delta, \lambda_0}$ está definida en (4.8). Sea $f(y, \mathbf{x}, t)$ una función en \mathcal{F}_n^* , esto es,

$$f(y, \mathbf{x}, t) = \left[\Psi(y, \mathbf{x}^\top \beta + g_\lambda(t), a) - \Psi(y, \mathbf{x}^\top \beta_0 + \eta_0(t), a) \right] w(\mathbf{x}) h_j^*(t),$$

para algún $\|\beta - \beta_0\| < \epsilon_0$, $a \in \mathcal{V}$ y $g_\lambda \in \mathcal{M}_n(\mathcal{T}_n, \ell)$ tal que $\pi_{\mathbb{P}}(\boldsymbol{\theta}_{0,n}, \boldsymbol{\theta}_\lambda) < \delta$. Entonces, $\|f\|_\infty \leq B_j$ con $B_j = 2\|\Psi\|_\infty \|h_j^*\|_\infty$. Usando un desarrollo de Taylor de orden 2, obtenemos que

$$f(y, \mathbf{x}, t) = \chi(y, \mathbf{x}^\top \zeta_\beta + \zeta_\eta(t), a) \left[\mathbf{x}^\top (\beta - \beta_0) + g_\lambda(t) - \eta_0(t) \right],$$

donde $\zeta_{\beta} = \xi_{\beta_0}\beta_0 + (1 - \xi_{\beta_0})\beta$ y $\zeta_{\eta}(t) = \xi_{\eta_0}\eta_0 + (1 - \xi_{\eta_0})g_{\lambda}$ con $0 < \xi_{\eta_0}, \xi_{\beta_0} < 1$. De la cota $|f(y, \mathbf{x}, t)| \leq \|\chi\|_{\infty} \|h_j^*\|_{\infty} w(\mathbf{x}) |\mathbf{x}^T (\beta - \beta_0) + g_{\lambda}(t) - \eta_0(t)|$ y del hecho que la función de peso w es acotada y que $\pi_{\mathbb{P}}(\theta_0, \theta_{\lambda}) \leq \pi_{\mathbb{P}}(\theta_{0,n}, \theta_{\lambda}) + \pi_{\mathbb{P}}(\theta_0, \theta_{0,n}) \leq 2\delta$, concluimos que

$$Pf^2 \leq \|\chi\|_{\infty}^2 \|h_j^*\|_{\infty}^2 \mathbb{E} \left(w(\mathbf{x}) [\mathbf{x}^T (\beta - \beta_0) + g_{\lambda}(t) - \eta_0(t)]^2 \right) = \|\chi\|_{\infty}^2 \|h_j^*\|_{\infty}^2 \pi_{\mathbb{P}}^2(\theta_0, \theta_{\lambda}) \leq C_j^2 \delta^2,$$

con $C_j^2 = 4\|\chi\|_{\infty}^2 \|h_j^*\|_{\infty}^2$ como se definió en **N6**. Usando nuevamente el Lema 3.8 obtenemos que

$$\mathbb{E}^* \|G_n\|_{\mathcal{F}_n^*} \leq J_{[]} (C_j \delta, \mathcal{F}_n^*, L_2(P)) \left(1 + \frac{J_{[]} (C_j \delta, \mathcal{F}_n^*, L_2(P))}{C_j^2 \delta^2 \sqrt{n}} B_j \right),$$

que junto con **N6** conduce a

$$\mathbb{E}^* \|G_n\|_{\mathcal{F}_n^*} \leq C \delta k_n^{1/2} \left(1 + \frac{B_j C k_n^{1/2}}{C_j^2 \delta \sqrt{n}} \right) = C \delta k_n^{1/2} + \frac{C B_j}{C_j^2} k_n n^{-1/2}.$$

Llamemos $\mathcal{B}_n = \{\|\hat{\beta} - \beta_0\| < \epsilon_0 \cap \pi_{\mathbb{P}}(\theta_{0,n}, \hat{\theta}) < \delta\}$. Entonces, $\mathbb{P}(\mathcal{B}_n) \rightarrow 1$. Usando que $\delta = n^{-\nu r/(1+2r)}$, $k_n = O(n^{1/(1+2r)})$ y la desigualdad de Markov, resulta que

$$\begin{aligned} \mathbb{P}(\sqrt{n} \mathbb{E} |S_{1,j,n}| > \epsilon \cap \mathcal{B}_n) &\leq \mathbb{E}^* \|G_n\|_{\mathcal{F}_n^*} \leq C \delta k_n^{1/2} + \frac{C B_j}{C_j} k_n n^{-1/2} \\ &\leq C^* \left(n^{-(2-3r)/[4(1+2r)]} + n^{-(1-2r)/[2(1+2r)]} \right), \end{aligned}$$

que converge a 0 porque $r \geq 1$. Usando que $\mathbb{P}(\mathcal{B}_n) \rightarrow 1$, obtenemos que $S_{1,j,n} = o_{\mathbb{P}}(n^{-1/2})$, lo que concluye la demostración de **H2(b)**.

iii) Por último, para terminar la demostración, mostraremos que **H3** se verifica. Mediante un desarrollo de Taylor de orden 1 alrededor de (β_0, η_0) tenemos

$$\begin{aligned} \mathbf{W}_{\beta, g, a}^{\beta} &= \mathbf{W}_{\beta_0, \eta_0, a}^{\beta} + \chi(y, \mathbf{x}^T \beta_0 + \eta_0(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [\mathbf{x} - \mathbf{h}^*(t)]^T (\beta - \beta_0) \\ &\quad + \chi(y, \mathbf{x}^T \beta_0 + \eta_0(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] \mathbf{h}^*(t)^T (\beta - \beta_0) \\ &\quad + \chi(y, \mathbf{x}^T \beta_0 + \eta_0(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [g(t) - \eta_0(t)] \\ &\quad + \frac{1}{2} \chi_1(y, \mathbf{x}^T \zeta_{\beta} + \zeta_{\eta}(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] \{ \mathbf{x}^T (\beta - \beta_0) + [g(t) - \eta_0(t)] \}^2 \end{aligned}$$

donde $\zeta_{\beta} = \xi_{\beta_0}\beta_0 + (1 - \xi_{\beta_0})\beta$ y $\zeta_{\eta}(t) = \xi_{\eta_0}\eta_0 + (1 - \xi_{\eta_0})g$ con $0 < \xi_{\eta_0}, \xi_{\beta_0} < 1$. Para cualquier $\theta = (\beta, g)$ y $a \in \mathcal{V}$, denotemos como

$$\begin{aligned} \mathbf{A}_{\theta, a} &= \mathbb{E} \chi(y, \mathbf{x}^T \beta + g(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [\mathbf{x} - \mathbf{h}^*(t)]^T \\ \mathbf{F}_{\theta, a} &= \mathbb{E} \chi(y, \mathbf{x}^T \beta + g(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] \mathbf{h}^*(t)^T \\ \mathbf{E}_{\theta, a}(\eta) &= \mathbb{E} \chi(y, \mathbf{x}^T \beta + g(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [\eta(t) - \eta_0(t)] \end{aligned}$$

Entonces, para cualquier $a \in \mathcal{V}$, se obtiene que

$$P\mathbf{W}_{\hat{\beta}, \hat{\eta}, a}^{\beta} = P\mathbf{W}_{\beta_0, \eta_0, a}^{\beta} + q\mathbf{A}_{\theta_0, a}(\hat{\beta} - \beta_0) + \mathbf{F}_{\theta_0, a}(\beta - \beta_0) + \mathbf{E}_{\theta_0, a}(\hat{\eta}) + \frac{1}{2} \mathbf{R}_n(\hat{\beta}, \hat{\eta}, a),$$

donde $\mathbf{R}_n(\boldsymbol{\beta}, g, a) = (R_{n,1}(\boldsymbol{\beta}, g, a), \dots, R_{n,p}(\boldsymbol{\beta}, g, a))^T$ es igual a

$$\mathbf{R}_n(\boldsymbol{\beta}, g, a) = \mathbb{E} \chi_1(y, \mathbf{x}^T \boldsymbol{\zeta}_\beta + \zeta_\eta(t), a) w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] \{ \mathbf{x}^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + [g(t) - \eta_0(t)] \}^2, \quad (4.13)$$

con $\boldsymbol{\zeta}_\beta = \xi_{\beta_0} \boldsymbol{\beta}_0 + (1 - \xi_{\beta_0}) \boldsymbol{\beta}$ y $\zeta_\eta = \xi_{\eta_0} \eta_0(t) + (1 - \xi_{\eta_0}) g$.

Bajo **N0(a)**, la definición de \mathbf{h}^* implica inmediatamente que $\mathbf{F}_{\boldsymbol{\theta}_0, \kappa_0} = \mathbf{0}$ y $\mathbf{E}_{\boldsymbol{\theta}_0, \kappa_0}(g) = \mathbf{0}$ para cualquier función g . Por otro lado, cuando se cumple **N0(b)**, la función $\chi(y, s, a)$ puede escribirse como $\chi(y, s, a) = \chi^*(y - s, a)$ (ver la Observación 4.2). La independencia entre los errores y las covariables bajo el modelo parcialmente lineal (2.12), implica que, para cualquier $a \in \mathcal{V}$,

$$\begin{aligned} \mathbf{F}_{\boldsymbol{\theta}_0, a} &= \mathbb{E} \chi^*(u, a) \mathbb{E} \{ w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] \mathbf{h}^*(t)^T \} = \mathbf{0} \\ \mathbf{E}_{\boldsymbol{\theta}_0, a}(g) &= \mathbb{E} \chi^*(u, a) \mathbb{E} \{ w(\mathbf{x}) [\mathbf{x} - \mathbf{h}^*(t)] [g(t) - \eta_0(t)] \} = \mathbf{0}. \end{aligned}$$

Por otro lado, **N1** implica que $P\mathbf{W}_{\boldsymbol{\beta}_0, \eta_0, a}^\beta = \mathbf{0}$, luego, obtenemos que

$$P\mathbf{W}_{\hat{\boldsymbol{\beta}}, \hat{\eta}, \hat{\kappa}}^\beta = A_{\boldsymbol{\theta}_0, \hat{\kappa}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \mathbf{R}_n(\hat{\boldsymbol{\beta}}, \hat{\eta}, \hat{\kappa}).$$

De la consistencia de $\hat{\kappa}$ y **N3**, es fácil ver que $A_{\boldsymbol{\theta}_0, \hat{\kappa}} \xrightarrow{p} A_{\boldsymbol{\theta}_0, \kappa_0}$, es decir que para probar la validez de **H3**, solo resta mostrar que $\mathbf{R}_n(\hat{\boldsymbol{\beta}}, \hat{\eta}, \hat{\kappa}) = o_{\mathbb{P}}(n^{-1/2})$.

a) Supongamos que se cumple **N4(a)**. A partir de (4.13), usando que χ_1 está acotada y llamando $b(\mathbf{x}, t) = \mathbf{x}^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + [g(t) - \eta_0(t)]$, tenemos que

$$\begin{aligned} |R_{n,j}(\boldsymbol{\beta}, g, \kappa)| &\leq \|\chi_1\|_\infty \{ \mathbb{E} w(\mathbf{x}) |x_j| b^2(\mathbf{x}, t) + \mathbb{E} w(\mathbf{x}) |h_j^*(t)| b^2(\mathbf{x}, t) \} \\ &\leq \|\chi_1\|_\infty \{ \mathbb{E} w(\mathbf{x}) |x_j| b^2(\mathbf{x}, t) + \|h_j^*\|_\infty \pi_{\mathbb{P}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \} = \|\chi_1\|_\infty \{ A_{1,n}(\boldsymbol{\beta}, g) + A_{2,n}(\boldsymbol{\beta}, g) \}, \end{aligned}$$

donde la última desigualdad sigue de los hechos de que $\pi_{\mathbb{P}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \mathbb{E} w(\mathbf{x}) b^2(\mathbf{x}, t)$, $h_j^*(t)$ es una función acotada y $\|w\|_\infty = 1$.

Usando que $\pi_{\mathbb{P}}^2(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-2r/(1+2r)})$ y $r \geq 1$, obtenemos inmediatamente que $A_{2,n}(\hat{\boldsymbol{\beta}}, \hat{\eta}) = o_{\mathbb{P}}(n^{-1/2})$. solo falta mostrar que $A_{1,n}(\hat{\boldsymbol{\beta}}, \hat{\eta}) = o_{\mathbb{P}}(n^{-1/2})$. La desigualdad de Cauchy–Schwartz implica que

$$\begin{aligned} A_{1,n}(\boldsymbol{\beta}, g) &= \mathbb{E} w^{\frac{1}{2}}(\mathbf{x}) |x_j| |b(\mathbf{x}, t)| w^{\frac{1}{2}}(\mathbf{x}) |b(\mathbf{x}, t)| \\ &\leq \{ \mathbb{E} w(\mathbf{x}) x_j^2 b^2(\mathbf{x}, t) \}^{\frac{1}{2}} \{ \mathbb{E} w(\mathbf{x}) b^2(\mathbf{x}, t) \}^{\frac{1}{2}} = \{ \mathbb{E} w(\mathbf{x}) x_j^2 b^2(\mathbf{x}, t) \}^{\frac{1}{2}} \pi_{\mathbb{P}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0). \end{aligned}$$

Usando nuevamente la desigualdad de Cauchy–Schwartz, obtenemos

$$\mathbb{E} w(\mathbf{x}) x_j^2 b^2(\mathbf{x}, t) = \mathbb{E} w^{\frac{1}{2}}(\mathbf{x}) x_j^2 |b(\mathbf{x}, t)| w^{\frac{1}{2}}(\mathbf{x}) |b(\mathbf{x}, t)| \leq \{ \mathbb{E} w(\mathbf{x}) x_j^4 b^2(\mathbf{x}, t) \}^{\frac{1}{2}} \pi_{\mathbb{P}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

que junto con la desigualdad $(a + b)^2 \leq 2(a^2 + b^2)$ conducen a

$$\begin{aligned} A_{1,n}(\boldsymbol{\beta}, g) &\leq \left\{ \mathbb{E}w(\mathbf{x})x_j^4 b^2(\mathbf{x}, t) \right\}^{\frac{1}{2}} \pi_{\mathbb{P}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \left\{ \mathbb{E}w(\mathbf{x})x_j^4 b^2(\mathbf{x}, t) \right\}^{\frac{1}{4}} \pi_{\mathbb{P}}^{\frac{3}{2}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \\ &\leq 2^{\frac{1}{4}} \pi_{\mathbb{P}}^{\frac{3}{2}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \left\{ \mathbb{E}w(\mathbf{x})x_j^4 (\mathbf{x}^{\top}(\boldsymbol{\beta} - \boldsymbol{\beta}_0))^2 + \mathbb{E}w(\mathbf{x})x_j^4 [g(t) - \eta_0(t)]^2 \right\}^{\frac{1}{4}} \\ &\leq 2^{\frac{1}{4}} \pi_{\mathbb{P}}^{\frac{3}{2}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \left\{ \mathbb{E}w(\mathbf{x})x_j^4 \|\mathbf{x}\|^2 \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \|g - \eta_0\|_{\infty}^2 \mathbb{E}w(\mathbf{x})x_j^4 \right\}^{\frac{1}{4}}. \end{aligned}$$

Usando que $\pi_{\mathbb{P}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$, tenemos que, para $r > 1$,

$$n^{\frac{1}{2}} \pi_{\mathbb{P}}^{\frac{3}{2}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1) n^{\frac{1}{2} - \frac{3r}{2(1+2r)}} = O_{\mathbb{P}}(1) n^{\frac{1-r}{2(1+2r)}} = o_{\mathbb{P}}(1).$$

Por otro lado, el segundo término en la cota para $A_{1,n}(\widehat{\boldsymbol{\beta}}, \widehat{\eta})$, esto es,

$$\left\{ \mathbb{E}w(\mathbf{x})x_j^4 \|\mathbf{x}\|^2 \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + \|\widehat{\eta} - \eta_0\|_{\infty}^2 \mathbb{E}w(\mathbf{x})x_j^4 \right\}^{\frac{1}{4}},$$

también converge a 0 porque, como se notó antes, $\pi_{\mathbb{P}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$ implica que $\|\widehat{\eta} - \eta_0\|_{\infty} \xrightarrow{p} 0$. Entonces, $R_{n,j}(\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}) = o_{\mathbb{P}}(n^{-1/2})$.

Cuando $r = 1$, usamos de nuevo la desigualdad de Cauchy–Schwartz para obtener la cota

$$\mathbb{E}w(\mathbf{x})x_j^4 b^2(\mathbf{x}, t) = \mathbb{E}w^{\frac{1}{2}}(\mathbf{x})x_j^4 |b(\mathbf{x}, t)| w^{\frac{1}{2}}(\mathbf{x}) |b(\mathbf{x}, t)| \leq \left\{ \mathbb{E}w(\mathbf{x})x_j^8 b^2(\mathbf{x}, t) \right\}^{\frac{1}{2}} \pi_{\mathbb{P}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0),$$

que lleva a

$$\begin{aligned} A_{1,n}(\boldsymbol{\beta}, g) &\leq \left\{ \mathbb{E}w(\mathbf{x})x_j^4 b^2(\mathbf{x}, t) \right\}^{\frac{1}{4}} \pi_{\mathbb{P}}^{\frac{3}{2}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \\ &\leq \left\{ \mathbb{E}w(\mathbf{x})x_j^8 b^2(\mathbf{x}, t) \right\}^{\frac{1}{8}} \pi_{\mathbb{P}}^{\frac{7}{4}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \\ &\leq 2^{\frac{1}{8}} \pi_{\mathbb{P}}^{\frac{7}{4}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \left\{ \mathbb{E}w(\mathbf{x})\|\mathbf{x}\|^2 x_j^8 \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \|g - \eta_0\|_{\infty}^2 \mathbb{E}w(\mathbf{x})x_j^8 \right\}^{\frac{1}{8}}. \end{aligned}$$

Usando que $\pi_{\mathbb{P}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-1/3})$, deducimos inmediatamente que $R_{n,j}(\widehat{\boldsymbol{\beta}}, \widehat{\eta}, \widehat{\kappa}) = o_{\mathbb{P}}(n^{-1/2})$, concluyendo la demostración cuando **N4(a)** se cumple.

b) Supongamos que vale **N4(b)**. Usando nuevamente que w y $h_j^*(t)$ son funciones acotadas y que $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \mathbb{E}[g(t) - \eta_0(t)]^2 = \pi^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$, obtenemos las desigualdades

$$\begin{aligned} |R_{n,j}(\boldsymbol{\beta}, g, \kappa)| &\leq \|\chi_1\|_{\infty} \mathbb{E}w(\mathbf{x}) |x_j - h_j^*(t)| \left\{ \mathbf{x}^{\top}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + [g(t) - \eta_0(t)] \right\}^2 \\ &\leq 2\|\chi_1\|_{\infty} \mathbb{E}w(\mathbf{x}) |x_j - h_j^*(t)| \left(\left\{ \mathbf{x}^{\top}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\}^2 + \left\{ [g(t) - \eta_0(t)] \right\}^2 \right) \\ &\leq 2\|\chi_1\|_{\infty} \left\{ \mathbb{E}w(\mathbf{x}) |x_j| \|\mathbf{x}\|^2 \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \mathbb{E}w(\mathbf{x}) |x_j| [g(t) - \eta_0(t)]^2 \right. \\ &\quad \left. + \mathbb{E}w(\mathbf{x}) |h_j^*(t)| [g(t) - \eta_0(t)]^2 + \mathbb{E}w(\mathbf{x}) \|\mathbf{x}\|^2 |h_j^*(t)| \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 \right\} \\ &\leq 2\|\chi_1\|_{\infty} \left\{ \mathbb{E}w(\mathbf{x}) |x_j| \|\mathbf{x}\|^2 \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \mathbb{E}w(\mathbf{x}) |x_j| [g(t) - \eta_0(t)]^2 \right. \\ &\quad \left. + \|h_j^*\|_{\infty} (1 + \mathbb{E}w(\mathbf{x}) \|\mathbf{x}\|^2) \pi^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \right\}. \end{aligned}$$

Recordemos que $\mathbb{E}w(\mathbf{x}) \|\mathbf{x}\|^3 < \infty$. Luego, tenemos que

$$|R_{n,j}(\boldsymbol{\beta}, g, \kappa)| \leq 2\|\chi_1\|_{\infty} \left\{ A \pi^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + \mathbb{E}w(\mathbf{x}) |x_j| [g(t) - \eta_0(t)]^2 \right\},$$

con $A = \mathbb{E}w(\mathbf{x}) \|\mathbf{x}\|^3 + \|h_j^*\|_\infty (1 + \mathbb{E}w(\mathbf{x})\|\mathbf{x}\|^2)$. Argumentando como antes acotamos el segundo término del lado derecho de la desigualdad como

$$\mathbb{E}w(\mathbf{x}) |x_j| [g(t) - \eta_0(t)]^2 \leq \pi^{\frac{3}{2}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \{\mathbb{E}w(\mathbf{x})x_j^4\}^{\frac{1}{4}} \|g - \eta_0\|_\infty^{\frac{1}{2}},$$

cuando $r > 1$, mientras que si $r = 1$, obtenemos que

$$\mathbb{E}w(\mathbf{x}) |x_j| [g(t) - \eta_0(t)]^2 \leq \pi^{\frac{7}{8}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \{\mathbb{E}w(\mathbf{x})x_j^8\}^{\frac{1}{8}} \|g - \eta_0\|_\infty^{\frac{1}{4}}.$$

La conclusión se sigue del hecho de que $\pi(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-r/(1+2r)})$ y $\|\widehat{\eta} - \eta_0\|_\infty \xrightarrow{P} 0$. ■

Capítulo 5

Función de influencia empírica

Uno de los objetivos de un procedimiento robusto es producir estimadores menos sensibles a datos atípicos que los procedimientos clásicos. La función de influencia es una medida de robustez con respecto a datos atípicos. Medidas de diagnóstico y técnicas gráficas para detectar datos atípicos pueden construirse en base a la función de influencia empírica. Para modelos paramétricos este tópico está ampliamente desarrollado. Sin embargo, en la bibliografía no paramétrica ha recibido menos atención. Un enfoque basado en un funcional suavizado para estimadores no paramétricos por núcleos fue introducido por Ait Sahalia (1995) y usado por Tamine (2002) para definir una función de influencia suavizada en regresión no paramétrica. Por otro lado, Manchester (1996) introdujo un método gráfico para mostrar la sensibilidad de un suavizador local. Para medir la influencia de una observación atípica en ambas componentes, la paramétrica y la no paramétrica, seguiremos un enfoque similar al de Boente y Rodríguez (2010) y daremos un enfoque relacionado con la función de influencia empírica definida por Mallows (1974), que es la función de influencia del funcional bajo estudio calculado en la distribución empírica.

Por simplicidad, a través de este capítulo, supondremos que κ_0 es conocido y que se verifica **C7**. Denotaremos $\rho(y, u) = \rho(y, u, \kappa_0)$, $\Psi(y, u) = \Psi(y, u, \kappa_0)$ y $\chi(y, u) = \chi(y, u, \kappa_0)$.

5.1. Función de influencia de los estimadores definidos en la Sección 2.2

Dado un conjunto de datos $\{(y_i, \mathbf{x}_i^T, t_i)^T\}_{1 \leq i \leq n}$ que satisface el modelo semiparamétrico lineal isotónico generalizado, sean $(\hat{\beta}, \hat{\eta})$, con $\hat{\eta}(t) = \sum_{j=1}^{k_n} \hat{\lambda}_j B_j(t)$, los estimadores de (β_0, η_0) basados en este conjunto de datos, esto es, $(\hat{\beta}, \hat{\lambda})$ es la solución de (2.2). Como la función $\hat{\eta}$ está completamente determinada por los coeficientes $\hat{\lambda}$, una vez que la base de splines es fijada, la identificaremos con su parámetro de dimensión finita. Sea $(y_0, \mathbf{x}_0^T, t_0)^T$ un punto de contaminación y llamemos P_n a la medida empírica que da peso $1/n$ a cada punto de la muestra, de modo que $(\hat{\beta}, \hat{\lambda}) = (\hat{\beta}(P_n), \hat{\lambda}(P_n))$. Por otro lado, sea $P_{n,0}$ la medida empírica que da peso $(1 - \epsilon)/n$ a cada (y_i, \mathbf{x}_i, t_i) , $1 \leq i \leq n$, y peso ϵ a la observación $(y_0, \mathbf{x}_0^T, t_0)^T$. Denotemos $(\hat{\beta}_{0,\epsilon}, \hat{\lambda}_{0,\epsilon})$ a los estimadores de los parámetros para

esta nueva muestra, esto es, los valores que resuelven

$$(\widehat{\boldsymbol{\beta}}_{0,\epsilon}, \widehat{\boldsymbol{\lambda}}_{0,\epsilon}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\lambda} \in \mathcal{L}_{k_n}}{\operatorname{argmín}} \frac{1-\epsilon}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{B}_i^T \boldsymbol{\lambda}) w(\mathbf{x}_i) + \epsilon \rho(y_0, \mathbf{x}_0^T \boldsymbol{\beta} + \mathbf{B}_0^T \boldsymbol{\lambda}) w(\mathbf{x}_0) \quad (5.1)$$

donde $\mathbf{B}_0 = (B_1(t_0), \dots, B_{k_n}(t_0))$ y $\mathbf{B}_i = (B_1(t_i), \dots, B_{k_n}(t_i))$.

Entonces, podemos definir la función de influencia empírica (EIF) de $\widehat{\boldsymbol{\beta}}$ en (y_0, \mathbf{x}_0, t_0) como

$$\begin{aligned} \operatorname{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0)) &= \lim_{\epsilon \rightarrow 0} \frac{\widehat{\boldsymbol{\beta}}_{0,\epsilon} - \widehat{\boldsymbol{\beta}}}{\epsilon}, \\ \operatorname{EIF}(\widehat{\boldsymbol{\lambda}}; (y_0, \mathbf{x}_0, t_0)) &= \lim_{\epsilon \rightarrow 0} \frac{\widehat{\boldsymbol{\lambda}}_{0,\epsilon} - \widehat{\boldsymbol{\lambda}}}{\epsilon}. \end{aligned}$$

Podemos construir un gráfico de superficie variando los valores $(y_0, \mathbf{x}_0^T, t_0)^T$ para ver cómo los valores atípicos afectan los estimadores del parámetro de regresión y de la componente no paramétrica $\eta(t)$. En adelante, supondremos que las funciones de influencia $\operatorname{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0))$ y $\operatorname{EIF}(\widehat{\boldsymbol{\lambda}}; (y_0, \mathbf{x}_0, t_0))$ existen.

Para simplificar las próximas cuentas, reparametrizamos el problema con $\boldsymbol{\Delta} = (\Delta_1, \dots, \Delta_{k_n})$ donde $\Delta_{k_n} = \lambda_{k_n}$, $\Delta_j = \lambda_j - \lambda_{j+1} \leq 0$ para $1 \leq j \leq k_n - 1$, de modo que $\lambda_s = \sum_{j=s}^{k_n} \Delta_j$. Notemos que $\sum_{s=1}^{k_n} B_s(t_i) \lambda_s = \sum_{j=1}^{k_n} \Delta_j \sum_{s=1}^j B_s(t_i) = \sum_{j=1}^{k_n} b_{i,j} \Delta_j$, con $b_{i,j} = \sum_{s=1}^j B_s(t_i)$. Entonces, si definimos $\mathbf{b}_i = (b_{i,1}, \dots, b_{i,k_n})^T$, el problema puede ser reescrito como

$$(\widehat{\boldsymbol{\beta}}_{0,\epsilon}, \widehat{\boldsymbol{\Delta}}_{0,\epsilon}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{\Delta} \in \mathcal{D}_{k_n}}{\operatorname{argmín}} \frac{1-\epsilon}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{b}_i^T \boldsymbol{\Delta}) + \epsilon \rho(y_0, \mathbf{x}_0^T \boldsymbol{\beta} + \mathbf{b}_0^T \boldsymbol{\Delta}),$$

donde $\mathcal{D}_{k_n} = \{\boldsymbol{\Delta} \in \mathbb{R}^{k_n} : \Delta_j \leq 0 \text{ for } 1 \leq j \leq k_n - 1\}$. Usando las condiciones de Karush–Kuhn–Tucker (KKT) con las funciones de restricción $g_i(\boldsymbol{\Delta}) = \Delta_i$, $1 \leq i \leq k_n - 1$ y derivando con respecto a \mathbf{b} y $\boldsymbol{\Delta}$ obtenemos que los estimadores $\widehat{\boldsymbol{\beta}}_{0,\epsilon}$ y $\widehat{\boldsymbol{\Delta}}_{0,\epsilon}$ son soluciones del siguiente sistema de ecuaciones

$$\begin{aligned} \frac{1-\epsilon}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_{0,\epsilon} + \mathbf{b}_i^T \widehat{\boldsymbol{\Delta}}_{0,\epsilon}) w(\mathbf{x}_i) \mathbf{x}_i + \epsilon \Psi(y_0, \mathbf{x}_0^T \widehat{\boldsymbol{\beta}}_{0,\epsilon} + \mathbf{b}_0^T \widehat{\boldsymbol{\Delta}}_{0,\epsilon}) w(\mathbf{x}_0) \mathbf{x}_0 &= 0 \\ \frac{1-\epsilon}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_{0,\epsilon} + \mathbf{b}_i^T \widehat{\boldsymbol{\Delta}}_{0,\epsilon}) w(\mathbf{x}_i) \mathbf{b}_i + \epsilon \Psi(y_0, \mathbf{x}_0^T \widehat{\boldsymbol{\beta}}_{0,\epsilon} + \mathbf{b}_0^T \widehat{\boldsymbol{\Delta}}_{0,\epsilon}) w(\mathbf{x}_0) \mathbf{b}_0 + \boldsymbol{\nu}_{0,\epsilon} &= 0 \\ g_s(\widehat{\boldsymbol{\Delta}}_{0,\epsilon}) = \widehat{\Delta}_{0,\epsilon,s} \leq 0 \quad \nu_{0,\epsilon,s} \geq 0 \quad \nu_{0,\epsilon,s} \widehat{\Delta}_{0,\epsilon,s} = 0 \quad 1 \leq s \leq k_n - 1 \end{aligned}$$

donde $\boldsymbol{\nu}_{0,\epsilon} = (\nu_{0,\epsilon,1}, \nu_{0,\epsilon,2}, \dots, \nu_{0,\epsilon,k_n-1}, 0)$ son los multiplicadores KKT. En particular, cuando $\epsilon = 0$, si $\widehat{\boldsymbol{\nu}} = \boldsymbol{\nu}_{0,0}$, tenemos que $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{0,0}$ y $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}_{0,0}$ satisfacen

$$\frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_i^T \widehat{\boldsymbol{\Delta}}) w(\mathbf{x}_i) \mathbf{x}_i = 0, \quad (5.2)$$

$$\frac{1}{n} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_i^T \widehat{\boldsymbol{\Delta}}) w(\mathbf{x}_i) \mathbf{b}_i + \widehat{\boldsymbol{\nu}} = 0, \quad (5.3)$$

$$g_s(\widehat{\boldsymbol{\Delta}}) = \widehat{\Delta}_s \leq 0 \quad \widehat{\nu}_s \geq 0 \quad \widehat{\nu}_s \widehat{\Delta}_s = 0 \quad 1 \leq s \leq k_n - 1. \quad (5.4)$$

Consideremos el conjunto de índices $\mathcal{J}_1 = \{k_n\} \cup \{1 \leq j \leq k_n - 1 : \widehat{\Delta}_j < 0\} = \{k_n\} \cup \{1 \leq j \leq k_n - 1 : \widehat{\lambda}_j < \widehat{\lambda}_{j+1}\}$, $\mathcal{J}_2 = \{1 \leq j \leq k_n - 1 : \widehat{\Delta}_j = 0 \text{ y } \nu_j > 0\}$ y $\mathcal{J}_3 = \{1 \leq j \leq k_n - 1 : \widehat{\Delta}_j = 0 \text{ y } \nu_j = 0\}$.

Usando que para $s \in \mathcal{J}_1$, $s \neq k_n$, $\widehat{\Delta}_s = \widehat{\Delta}_{0,0,s} \neq 0$ y el hecho de que estamos suponiendo que $\widehat{\Delta}_{0,\epsilon}$ es derivable con respecto a ϵ y por lo tanto continua, obtenemos que existe ϵ_1 lo suficientemente chico para que $\widehat{\Delta}_{0,\epsilon,s} < 0$ para $\epsilon < \epsilon_1$. Luego, usando que $\nu_{0,\epsilon,s} \widehat{\Delta}_{0,\epsilon,s} = 0$, obtenemos que $\nu_{0,\epsilon,s} = 0$.

Por otro lado, si $s \in \mathcal{J}_2$, tenemos que $\nu_s = \nu_{0,0,s} \neq 0$ luego para $\epsilon < \epsilon_2$, $\nu_{0,\epsilon,s} > 0$ de manera que $\widehat{\Delta}_{0,\epsilon,s} = 0$, lo que implica que

$$\text{EIF}(\widehat{\Delta}_s; (y_0, \mathbf{x}_0, t_0)) = 0 \quad s \in \mathcal{J}_2. \quad (5.5)$$

Sea $m_\ell = \#\mathcal{J}_\ell$ y llamemos $\mathbf{b}_{\mathcal{J}_\ell}$ al vector de dimensión m_ℓ -ésima con componentes correspondientes a los índices $s \in \mathcal{J}_\ell$. Definamos las matrices

$$\begin{aligned} \mathbf{A}_n &= \frac{1}{n} \sum_{i=1}^n \chi \left(y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_i^T \widehat{\boldsymbol{\Delta}} \right) w(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T \in \mathbb{R}^{p \times p} \\ \mathbf{B}_{n,\mathcal{J}_1} &= \frac{1}{n} \sum_{i=1}^n \chi \left(y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_i^T \widehat{\boldsymbol{\Delta}} \right) w(\mathbf{x}_i) \mathbf{x}_i \mathbf{b}_{i,\mathcal{J}_1}^T \in \mathbb{R}^{p \times m_1} \\ \mathbf{C}_{n,\mathcal{J}_1} &= \frac{1}{n} \sum_{i=1}^n \chi \left(y_i, \mathbf{x}_i^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_i^T \widehat{\boldsymbol{\Delta}} \right) w(\mathbf{x}_i) \mathbf{b}_{i,\mathcal{J}_1} \mathbf{b}_{i,\mathcal{J}_1}^T \in \mathbb{R}^{m_1 \times m_1}. \end{aligned}$$

Llamemos $\mathbf{E}_{n,\mathcal{J}_1} = \mathbf{C}_{n,\mathcal{J}_1} - \mathbf{B}_{n,\mathcal{J}_1}^T \mathbf{A}_n^{-1} \mathbf{B}_{n,\mathcal{J}_1}$ y $\mathbf{F}_{n,\mathcal{J}_1} = \mathbf{A}_n^{-1} \mathbf{B}_{n,\mathcal{J}_1}$.

En el Apéndice 5.4 se muestra que, si $\mathcal{J}_3 = \emptyset$, entonces

$$\begin{aligned} \text{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0)) &= - \left(\mathbf{A}_n^{-1} + \mathbf{F}_{n,\mathcal{J}_1} \mathbf{E}_{n,\mathcal{J}_1}^{-1} \mathbf{F}_{n,\mathcal{J}_1}^T \right) \Psi \left(y_0, \mathbf{x}_0^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_0^T \widehat{\boldsymbol{\Delta}} \right) w(\mathbf{x}_0) \mathbf{x}_0 \\ &\quad + \mathbf{F}_{n,\mathcal{J}_1} \mathbf{E}_{n,\mathcal{J}_1}^{-1} \Psi \left(y_0, \mathbf{x}_0^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_0^T \widehat{\boldsymbol{\Delta}} \right) w(\mathbf{x}_0) \mathbf{b}_{0,\mathcal{J}_1}, \end{aligned} \quad (5.6)$$

$$\begin{aligned} \text{EIF}(\widehat{\boldsymbol{\Delta}}_{\mathcal{J}_1}; (y_0, \mathbf{x}_0, t_0)) &= \mathbf{E}_{n,\mathcal{J}_1}^{-1} \mathbf{F}_{n,\mathcal{J}_1}^T \Psi \left(y_0, \mathbf{x}_0^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_0^T \widehat{\boldsymbol{\Delta}} \right) w(\mathbf{x}_0) \mathbf{x}_0 \\ &\quad - \mathbf{E}_{n,\mathcal{J}_1}^{-1} \Psi \left(y_0, \mathbf{x}_0^T \widehat{\boldsymbol{\beta}} + \mathbf{b}_0^T \widehat{\boldsymbol{\Delta}} \right) w(\mathbf{x}_0) \mathbf{b}_{0,\mathcal{J}_1} \end{aligned} \quad (5.7)$$

Recordemos que $\lambda_s = \sum_{j=s}^{k_n} \Delta_j$, lo que implica que $\text{EIF}(\widehat{\lambda}_s; (y_0, \mathbf{x}_0, t_0)) = \sum_{j=s}^{k_n} \text{EIF}(\widehat{\Delta}_j; (y_0, \mathbf{x}_0, t_0))$. Luego, una expresión para $\text{EIF}(\widehat{\lambda}_s; (y_0, \mathbf{x}_0, t_0))$ puede obtenerse de (5.5) y (5.7), lo que permite calcular la función de influencia empírica de $\widehat{\eta}(t)$, para cada t fijo, como

$$\text{EIF}(\widehat{\eta}(t); (y_0, \mathbf{x}_0, t_0)) = \sum_{s=1}^{k_n} B_s(t) \text{EIF}(\widehat{\lambda}_s; (y_0, \mathbf{x}_0, t_0)) = \sum_{s=1}^{k_n} \text{EIF}(\widehat{\Delta}_s; (y_0, \mathbf{x}_0, t_0)) b_s(t), \quad (5.8)$$

donde $b_s(t) = \sum_{m=1}^s B_m(t)$.

5.2. La función de influencia empírica bajo un modelo de regresión parcialmente lineal log–Gamma

Para ilustrar el comportamiento de estos estimadores consideramos el modelo de regresión parcialmente lineal isotónico log–Gamma descrito en la Sección 2.5. Para $1 \leq i \leq n$, generamos covariables (x_i, t_i) independientes entre sí tales que $x_i \sim \mathcal{N}(0, 1)$, $t_i \sim \mathcal{U}(0, 1)$. La variable de respuesta fue generada como $z_i | (x_i, t_i) \sim \Gamma(3, \lambda_i)$, donde $\mathbb{E}(z_i | (x_i, t_i)) = 3/\lambda_i = \exp\{\beta_0 x_i + \eta_0(t_i)\}$, es decir, $H(a) = \exp(a)$, con $\beta_0 = 2$. El modelo log–Gamma transformado es

$$y_i = \beta_0 x_i + \eta_0(t_i) + u_i,$$

donde $u_i \sim \log(\Gamma(3, 1))$. Consideramos dos elecciones para la componente no paramétrica, $\eta_{0,1}(t) = \sin(\pi t/2)$ y $\eta_{0,2}(t) = \pi t + 0.25 \sin(4\pi t)$ que llamaremos Modelos 1 y 2 respectivamente. El tamaño de muestra se tomó igual a $n = 100$ y las funciones de influencia empírica, dadas por (5.6) y (5.8), fueron calculadas considerando el modelo transformado sobre una grilla definida por $t_0 \in \{0.2, 0.4, 0.6, 0.8\}$ mientras que (y_0, x_0) toma valores en una grilla equidistante en cada eje de tamaño 50×50 sobre $[0, 2.5] \times [-10, 10]$. Consideramos una base de B-splines de orden 4, es decir, polinomios de grado 3, y fijamos el número de elementos de la base k_n igual a 5 para $\eta_{0,1}$ y 8 para $\eta_{0,2}$. Además, cuando calculamos EIF($\hat{\eta}(t); (y_0, \mathbf{x}_0, t_0)$) el punto t fue tomado igual a t_0 . Los estimadores robustos fueron calculados como se describió en la Sección 2.5, esto es, corresponden a aquellos que controlan los valores grandes de la *deviance* usando como función de escores la función bicuadrada de Tukey. Tomando en cuenta que las funciones de influencia empírica dadas en (5.6) y (5.8) fueron obtenidas suponiendo que κ_0 es fijo y que en este contexto, κ_0 corresponde a la constante de calibración, la constante de calibración c en los Pasos 3 y 4, descritos en la Sección 2.5.1, fue fijada en 2. La función de peso w usada para controlar los puntos de alta palanca fue tomada como la función bicuadrada de Tukey con constante de calibración c_w

$$w(x) = \begin{cases} \left(1 - \left[\frac{x - \hat{\mu}_n}{c_w s_n}\right]^2\right)^2 & |x - \hat{\mu}_n| \leq c_w s_n \\ 0 & |x - \hat{\mu}_n| \geq c_w s_n, \end{cases} \quad (5.9)$$

con $\hat{\mu}_n$ la mediana de x_i y $s_n = \text{MAD}(x_i)$, pues consideramos $x_i \in \mathbb{R}$. Dos valores de c_w fueron elegidos $c_w = 4.685$ y $c_w = \sqrt{\chi_{0.975,1}^2} = 2.241$.

solo mostraremos las superficies obtenidas para $t_0 = 6$ porque para los otros valores de t_0 los gráficos resultan similares. Para el modelo log-Gamma elegido, los estimadores clásicos considerados son aquellos basados en la *deviance* como se describió en la Sección 2.5, es decir, que $\Psi(y_0, x_0\beta + v)x_0 = (1 - \exp(y_0 - x_0\beta - v))x_0$. Por lo tanto, la función de influencia empírica no será acotada para cada t_0 fijo, cuando $y_0 \rightarrow \infty$ para x_0 fijo y cuando $x_0 \rightarrow -\infty$ para y_0 fijo. Más aun, seguirá siendo no acotada cuando $x_0 \rightarrow +\infty$, pero a una tasa menor que cuando $x_0 \rightarrow -\infty$.

Las Figuras 5.1 y 5.2 muestran que, para valores negativos grandes de x y valores grandes de y , el valor absoluto de la función de influencia empírica de los estimadores clásicos toma valores grandes. El peor efecto se observa en puntos de alta palanca. Más precisamente, al estimar el parámetro de regresión β la EIF toma valores negativos muy grandes en ambos modelos, por lo tanto, valores negativos de x con valor absoluto grande combinados con los valores grandes de y conducirán a estimaciones con grandes sesgos negativos. Con respecto a la componente no paramétrica, la EIF toma valores positivos grandes, de manera que la función resulta sobreestimada. Por otro lado, el procedimiento robusto lleva a estimadores más estables, pues la función de influencia empírica es igual a 0 para valores grandes de x debido a la función de peso y la influencia permanece acotada incluso para valores grandes de y . Las Figuras 5.3 y 5.4 dan los valores de la función de influencia empírica en el intervalo $(-7, 0)$, para analizar el efecto de las observaciones correspondientes a valores absolutos grandes de la respuesta y , esto es, valores de z cercanos a 0. Las conclusiones descriptas siguen valiendo en este caso.

Para estudiar el comportamiento en el centro del rango de x , la Figura 5.5 reporta la EIF bajo el Modelo 1, se obtienen gráficos similares bajo el Modelo 2. Los gráficos enfatizan el efecto redescendiente de las funciones de escores y de peso usadas en el procedimiento de estimación

robusta.

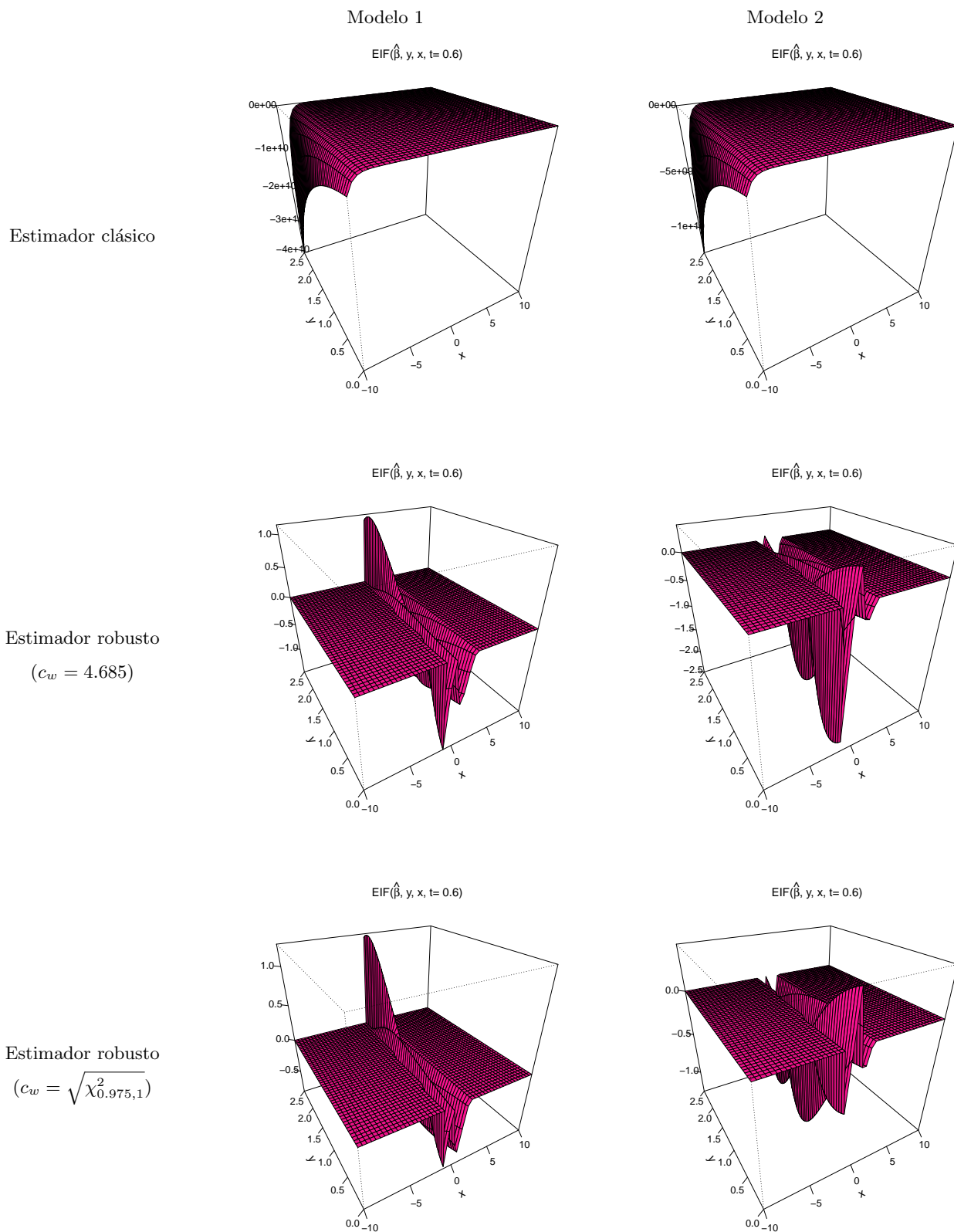


Figura 5.1: Función de influencia empírica de $\hat{\beta}$, bajo un modelo log-Gamma para $t_0 = 0.6$.

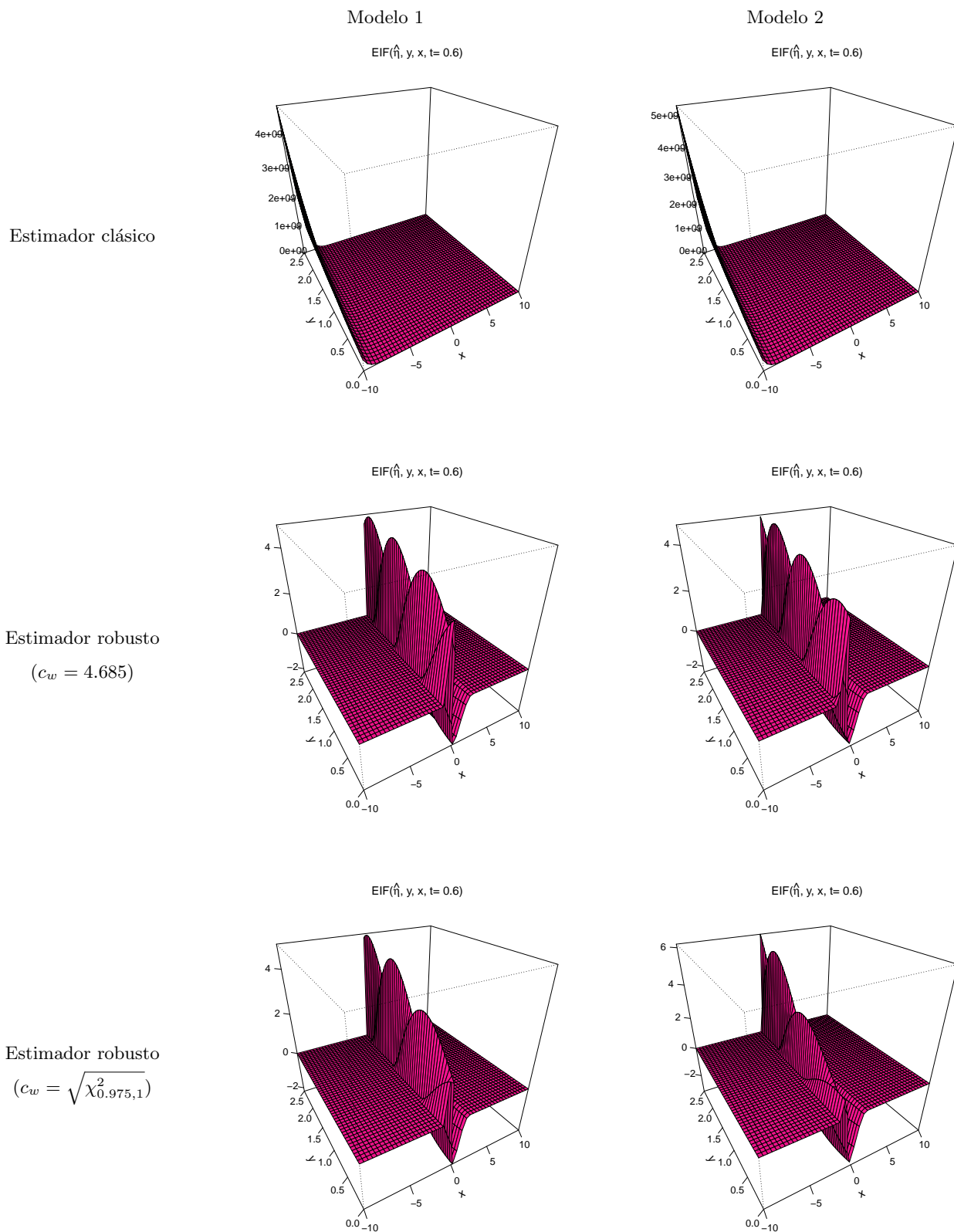


Figura 5.2: Función de influencia empírica de $\hat{\eta}(t_0)$, bajo un modelo log-Gamma para $t_0 = 0.6$.

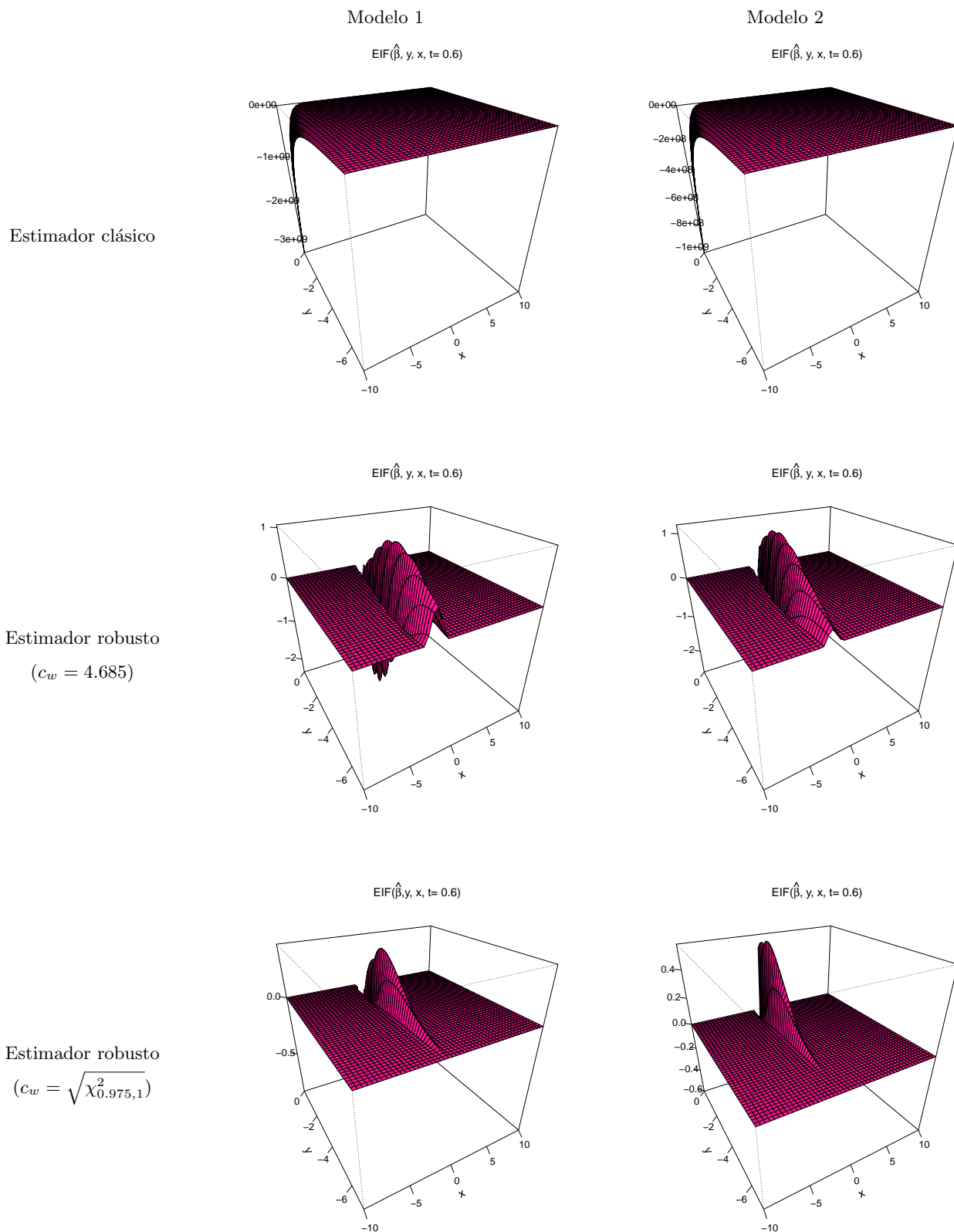


Figura 5.3: Función de influencia empírica de $\hat{\beta}$, bajo un modelo log-Gamma para $t_0 = 0.6$ y $y \in (-7, 0)$.

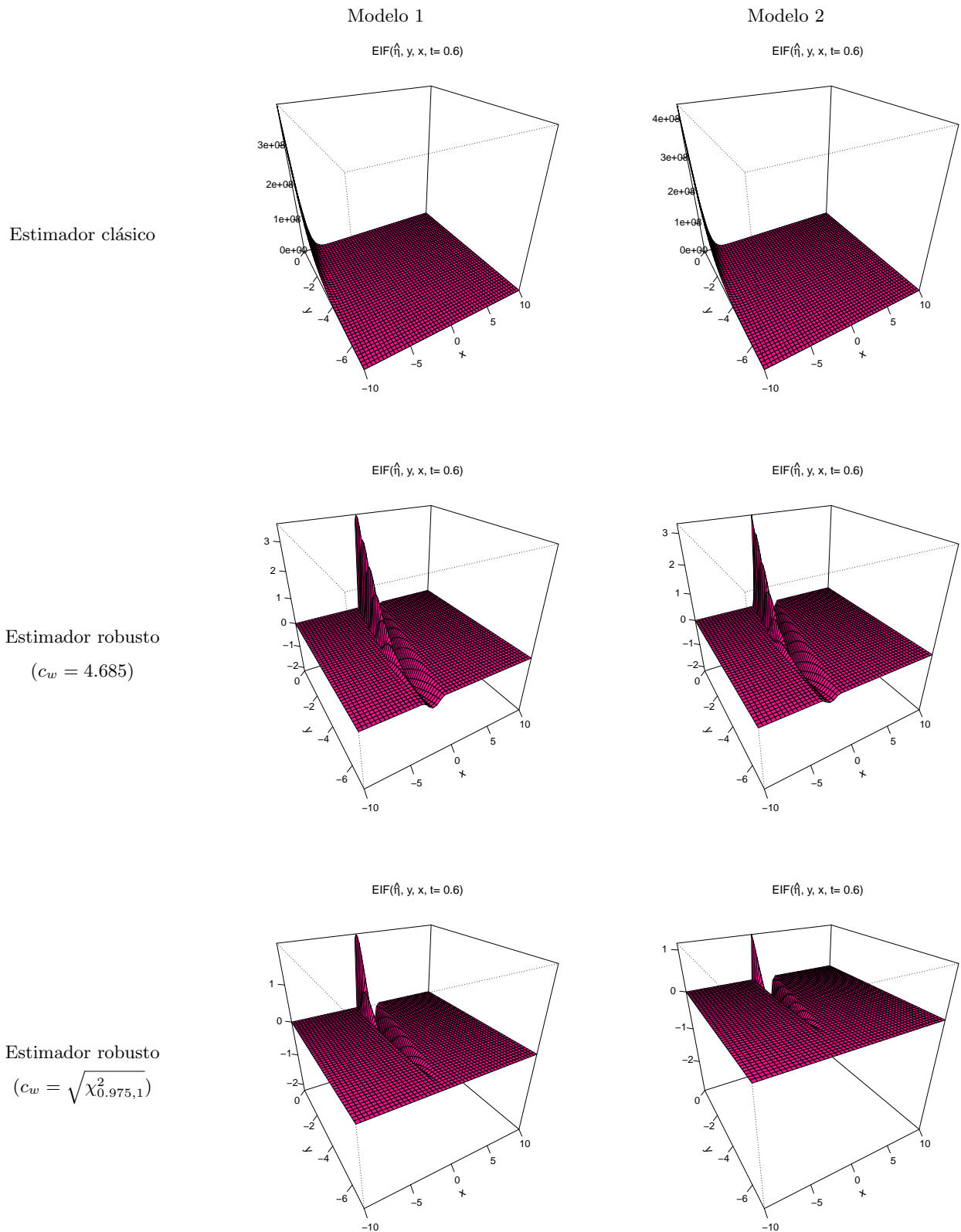


Figura 5.4: Función de influencia empírica de $\hat{\eta}(t_0)$, bajo un modelo log-Gamma cuando $t_0 = 0.6$ y $y \in (-7, 0)$.

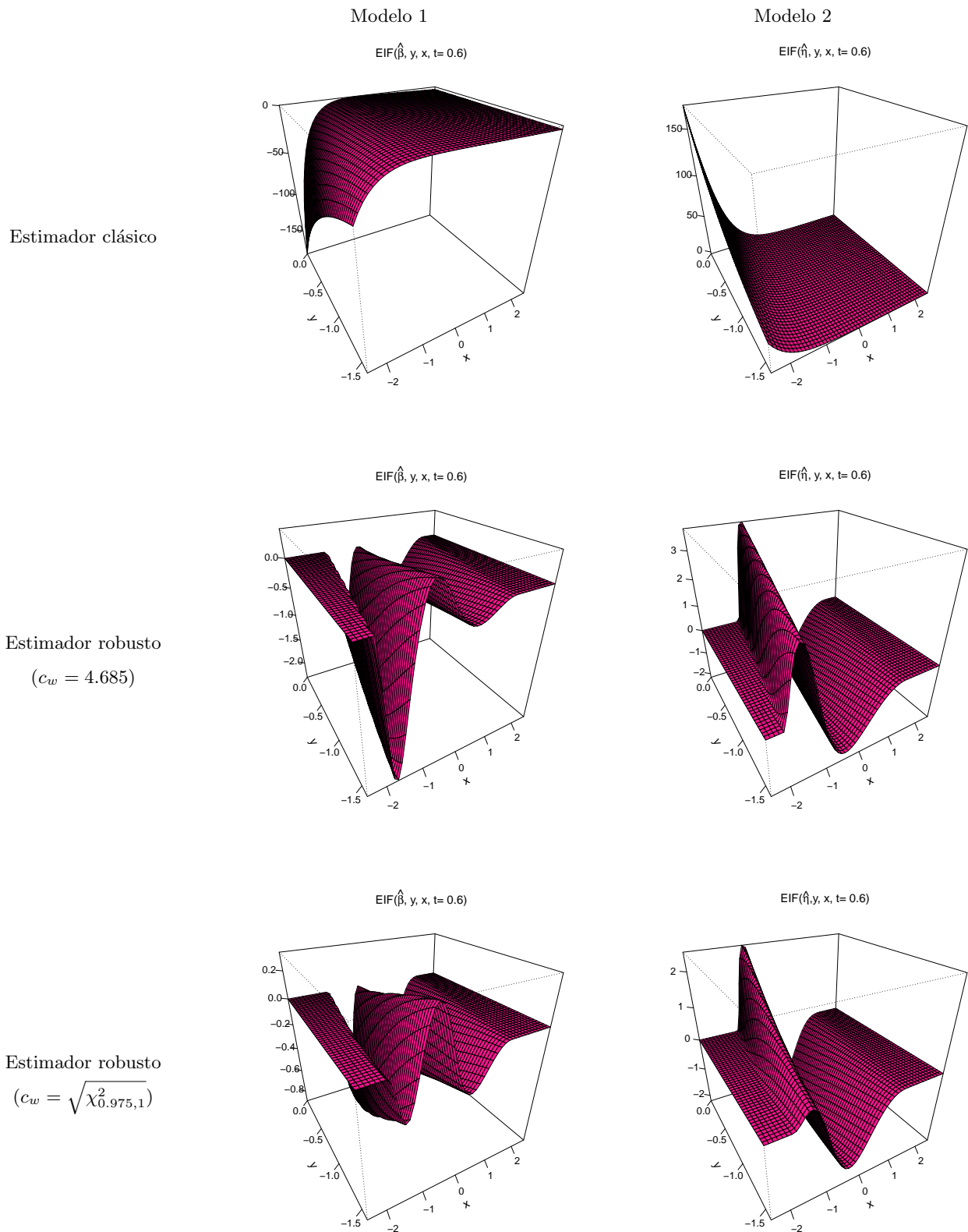


Figura 5.5: Función de influencia empírica de $\hat{\beta}$ (arriba) y $\hat{\eta}(t_0)$ (abajo), bajo un modelo log-Gamma cuando $\eta_{0,1}(t) = \sin(\pi t/2)$, $t_0 = 0.6$, $y \in (-1.6, 0)$ y $x \in (-2.5, 2.5)$.

5.3. Algunas observaciones sobre el estimador monotonizado definido en la Sección 2.3

En este capítulo hemos considerado solamente el caso de los estimadores basados en B-splines monótonos. Argumentos similares a los utilizados en modelos lineales generalizados permitirían obtener la función de influencia empírica de los estimadores $(\beta_{\text{INI}}, \lambda_{\text{INI}})$ calculados sin imponer restricciones de orden y definidos en la Sección 2.3. A partir de la influencia empírica de λ_{INI} se obtiene en forma inmediata la de $\hat{\eta}_{\text{INI}}(t)$, para cada t fijo, como en la expresión (5.8).

Respecto del estimador monotonizado $\hat{\eta}_i$ definido en 2.7, como la función $\Upsilon(\eta)$ está definida a través de una función indicadora, su función de influencia empírica no puede obtenerse como en la Sección 5.1. Una alternativa sería considerar una aproximación suave a la función indicadora, como la considerada en Neumeyer (2007). Más precisamente, indiquemos por K una densidad y por $\mathcal{K}(y) = \int_{-\infty}^y K(u)du$ la primitiva de K . Sea h_n una sucesión de ventanas positivas tales que $\lim_{n \rightarrow \infty} h_n = 0$. Definamos

$$\Phi(\hat{\eta}_{\text{INI}})(u) = \int_0^1 \mathcal{K}\left(\frac{u - \hat{\eta}_{\text{INI}}(z)}{h_n}\right) dz$$

y la modificación creciente suave de $\hat{\eta}_{\text{INI}}$ como

$$\hat{\eta}_{\text{SI}} = \Upsilon\left(\Phi(\hat{\eta}_{\text{INI}})\mathbb{I}_{[\hat{\eta}_{\text{INI}}(0), \hat{\eta}_{\text{INI}}(1)]}\right)\mathbb{I}_{[0,1]} \quad (5.10)$$

que será estrictamente creciente si \mathcal{K} lo es. La expresión $\Phi(\hat{\eta}_{\text{INI}})$ puede aproximarse por

$$\Phi(\hat{\eta}_{\text{INI}})(u) \approx \tilde{\Phi}(\hat{\eta}_{\text{INI}})(u) = \frac{1}{N} \sum_{j=1}^N \mathcal{K}\left(\frac{u - \hat{\eta}_{\text{INI}}(j/N)}{h_n}\right)$$

que es una expresión que facilitaría el cálculo de la función de influencia empírica de $\hat{\eta}_{\text{SI}}$ combinando la derivabilidad de la función \mathcal{K} con los argumentos utilizados para el cálculo de la función de influencia de los cuantiles de una distribución. Efectivamente si \mathcal{K} es estrictamente creciente, la función $\hat{f}(u) = \tilde{\Phi}(\hat{\eta}_{\text{INI}})(u)$ resulta estrictamente monótona por lo que $\Upsilon(\hat{f})$ es simplemente la inversa de \hat{f} . En nuestro estudio de simulación no hemos considerado el estimador definido en (5.10) ya que su cómputo implica la elección de la sucesión de ventanas. Como se menciona en Neumeyer (2007), resultados de consistencia uniformes para $\hat{\eta}_{\text{SI}}$ solo pueden obtenerse si se cuenta con tasas de convergencia uniforme para $\hat{\eta}_{\text{INI}}$ y si se elige la sucesión de ventanas de acuerdo a dicha tasa. Los resultados de consistencia del Capítulo 3 dan tasas de convergencia en distancia $L_2(P)$ para $\hat{\eta}_{\text{INI}}$, los que implican la convergencia uniforme de $\hat{\eta}_{\text{INI}}$ a η_0 en probabilidad pero no permiten deducir tasas de convergencia uniforme. Por esta razón, se omitió el estimador $\hat{\eta}_{\text{SI}}$ de nuestro estudio numérico y tampoco se calcula su influencia empírica.

5.4. Apéndice: Demostración de (5.6) y (5.7)

Como en la Sección 5.1, llamamos $\widehat{\Delta}_{0,\epsilon,\mathcal{J}_\ell} \in \mathbb{R}^{m_\ell}$ al m_ℓ -ésimo vector con componentes correspondientes a los índices $j \in \mathcal{J}_\ell$ y sea

$$\Lambda_\epsilon(y, \mathbf{x}, \mathbf{b}) = \Psi \left(y, \mathbf{x}^\top \widehat{\beta}_{0,\epsilon} + \mathbf{b}_{\mathcal{J}_1}^\top \widehat{\Delta}_{0,\epsilon,\mathcal{J}_1} + \mathbf{b}_{\mathcal{J}_3}^\top \widehat{\Delta}_{0,\epsilon,\mathcal{J}_3} \right) w(\mathbf{x}) = \Psi \left(y, \mathbf{x}^\top \widehat{\beta}_{0,\epsilon} + \mathbf{b}^\top \widehat{\Delta}_{0,\epsilon} \right) w(\mathbf{x}),$$

donde la última igualdad vale porque $\widehat{\Delta}_{0,\epsilon,s} = 0$ para $s \in \mathcal{J}_2$. Entonces

$$\begin{aligned} \frac{1-\epsilon}{n} \sum_{i=1}^n \Lambda_\epsilon(y_i, \mathbf{x}_i, \mathbf{b}) \mathbf{x}_i + \frac{\epsilon}{n} \Lambda_\epsilon(y_0, \mathbf{x}_0, \mathbf{b}) \mathbf{x}_0 &= 0 \\ \frac{1-\epsilon}{n} \sum_{i=1}^n \Lambda_\epsilon(y_i, \mathbf{x}_i, \mathbf{b}) b_{i,s} + \frac{\epsilon}{n} \Lambda_\epsilon(y_0, \mathbf{x}_0, \mathbf{b}) b_{0,s} &= 0, s \in \mathcal{J}_1 \\ \frac{1-\epsilon}{n} \sum_{i=1}^n \Lambda_\epsilon(y_i, \mathbf{x}_i, \mathbf{b}) b_{i,s} + \frac{\epsilon}{n} \Lambda_\epsilon(y_0, \mathbf{x}_0, \mathbf{b}) b_{0,s} + \nu_{0,\epsilon,s} &= 0, s \notin \mathcal{J}_1 \end{aligned} \quad (5.11)$$

$$\begin{aligned} \widehat{\Delta}_{0,\epsilon,s} < 0 & \quad \nu_{0,\epsilon,s} = 0 \quad s \in \mathcal{J}_1 \\ \widehat{\Delta}_{0,\epsilon,s} = 0 & \quad \nu_{0,\epsilon,s} > 0 \quad s \in \mathcal{J}_2 \\ \widehat{\Delta}_{0,\epsilon,s} \leq 0 & \quad \nu_{0,\epsilon,s} \geq 0 \quad \nu_{0,\epsilon,s} \widehat{\Delta}_{0,\epsilon,s} = 0 \quad s \in \mathcal{J}_3. \end{aligned}$$

Multiplicando (5.11) por $\widehat{\Delta}_{0,\epsilon,s}$ y usando que $\nu_{0,\epsilon,s} \widehat{\Delta}_{0,\epsilon,s} = 0$ para $s \notin \mathcal{J}_1$, tenemos que

$$\begin{aligned} \frac{1-\epsilon}{n} \sum_{i=1}^n \Lambda_\epsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) \mathbf{x}_i + \epsilon \Lambda_\epsilon(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{x}_0 &= 0 \\ \frac{1-\epsilon}{n} \sum_{i=1}^n \Lambda_\epsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} + \epsilon \Lambda_\epsilon(y_0, \mathbf{x}_0, \mathbf{b}_0) b_{0,s} &= 0, s \in \mathcal{J}_1 \\ \frac{1-\epsilon}{n} \sum_{i=1}^n \Lambda_\epsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} \widehat{\Delta}_{0,\epsilon,s} + \epsilon \Lambda_\epsilon(y_0, \mathbf{x}_0, \mathbf{b}_0) b_{0,s} \widehat{\Delta}_{0,\epsilon,s} &= 0, s \notin \mathcal{J}_1. \end{aligned}$$

Derivando con respecto a ϵ , evaluando en $\epsilon = 0$ y usando que $\widehat{\Delta}_{0,0} = \widehat{\Delta}$, $\widehat{\beta}_{0,0} = \widehat{\beta}$ junto con (5.2) a (5.4) tenemos que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \Upsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) \mathbf{x}_i \left[\mathbf{x}_i^\top \text{EIF}(\widehat{\beta}) + \mathbf{b}_i^\top \text{EIF}(\widehat{\Delta}) \right] + \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{x}_0 &= 0 \\ \frac{1}{n} \sum_{i=1}^n \Upsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} \left[\mathbf{x}_i^\top \text{EIF}(\widehat{\beta}) + \mathbf{b}_i^\top \text{EIF}(\widehat{\Delta}) \right] + \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) b_{0,s} &= 0, s \in \mathcal{J}_1 \\ \frac{1}{n} \sum_{i=1}^n \left(\Upsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} \left[\mathbf{x}_i^\top \text{EIF}(\widehat{\beta}) + \mathbf{b}_i^\top \text{EIF}(\widehat{\Delta}) \right] \widehat{\Delta}_s + \Lambda(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} \text{EIF}(\widehat{\Delta}_s) \right) \\ &+ \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) b_{0,s} \widehat{\Delta}_s = 0, s \notin \mathcal{J}_1, \end{aligned}$$

donde, por simplicidad, llamamos $\text{EIF}(\widehat{\beta}) = \text{EIF}(\widehat{\beta}; (y_0, \mathbf{x}_0, t_0))$, $\text{EIF}(\widehat{\Delta}) = \text{EIF}(\widehat{\Delta}; (y_0, \mathbf{x}_0, t_0))$ y $\Upsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) = \chi \left(y_i, \mathbf{x}_i^\top \widehat{\beta} + \mathbf{b}_i^\top \widehat{\Delta} \right) w(\mathbf{x}_i)$. Recordemos que $\widehat{\Delta}_s = 0$ para $s \notin \mathcal{J}_1$, de manera que la última ecuación implica

$$\frac{1}{n} \sum_{i=1}^n \Lambda(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} \text{EIF}(\widehat{\Delta}_s) = 0, s \notin \mathcal{J}_1, \quad (5.12)$$

que no es informativa. De hecho, cuando $s \in \mathcal{J}_2$, $\text{EIF}(\widehat{\Delta}_s) = 0$ entonces (5.12) se verifica. Por otro lado, para $s \in \mathcal{J}_3$, por (5.3), tenemos que

$$\frac{1}{n} \sum_{i=1}^n \Lambda(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} + \nu_s = 0,$$

junto con el hecho de que $\nu_j = 0$ para $j \in \mathcal{J}_3$ conduce a $\sum_{i=1}^n \Lambda(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} = 0$ para $s \in \mathcal{J}_3$, de forma que (5.12) se verifica. Notemos que de (5.11) tenemos que

$$\frac{1}{n} \sum_{i=1}^n \Upsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} \left[\mathbf{x}_i^T \text{EIF}(\widehat{\beta}) + \mathbf{b}_i^T \text{EIF}(\widehat{\Delta}) \right] + \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) b_{0,s} = -\frac{\partial}{\partial \epsilon} \nu_{0,\epsilon,s} \Big|_{\epsilon=0}, s \in \mathcal{J}_3.$$

Entonces, el sistema se reduce a

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \Upsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) \mathbf{x}_i \left[\mathbf{x}_i^T \text{EIF}(\widehat{\beta}) + \mathbf{b}_i^T \text{EIF}(\widehat{\Delta}) \right] + \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{x}_0 &= 0, \\ \frac{1}{n} \sum_{i=1}^n \Upsilon(y_i, \mathbf{x}_i, \mathbf{b}_i) b_{i,s} \left[\mathbf{x}_i^T \text{EIF}(\widehat{\beta}) + \mathbf{b}_i^T \text{EIF}(\widehat{\Delta}) \right] + \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) b_{0,s} &= 0, s \in \mathcal{J}_1 \\ \text{EIF}(\widehat{\Delta}_s) &= 0, s \in \mathcal{J}_2. \end{aligned}$$

El hecho de que $\mathcal{J}_3 = \emptyset$, junto con $\text{EIF}(\widehat{\Delta}_s) = 0$ para $s \in \mathcal{J}_2$ implica que $\mathbf{b}_i^T \text{EIF}(\widehat{\Delta}) = \mathbf{b}_{i,\mathcal{J}_1}^T \text{EIF}(\widehat{\Delta}_{\mathcal{J}_1})$, entonces

$$\begin{aligned} \mathbf{A}_n \text{EIF}(\widehat{\beta}) + \mathbf{B}_{n,\mathcal{J}_1} \text{EIF}(\widehat{\Delta}_{\mathcal{J}_1}) + \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{x}_0 &= 0, \\ \mathbf{B}_{n,\mathcal{J}_1}^T \text{EIF}(\widehat{\beta}) + \mathbf{C}_{n,\mathcal{J}_1} \text{EIF}(\widehat{\Delta}_{\mathcal{J}_1}) + \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{b}_{0,\mathcal{J}_1} &= 0, \\ \text{EIF}(\widehat{\Delta}_s) &= 0, s \in \mathcal{J}_2, \end{aligned}$$

es decir, si la matriz

$$\Sigma_{n,\mathcal{J}_1} = \begin{pmatrix} \mathbf{A}_n & \mathbf{B}_{n,\mathcal{J}_1} \\ \mathbf{B}_{n,\mathcal{J}_1}^T & \mathbf{C}_{n,\mathcal{J}_1} \end{pmatrix}$$

es no singular tenemos que

$$\begin{pmatrix} \text{EIF}(\widehat{\beta}) \\ \text{EIF}(\widehat{\Delta}_{\mathcal{J}_1}) \end{pmatrix} = -\Sigma_{n,\mathcal{J}_1}^{-1} \begin{pmatrix} \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{x}_0 \\ \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{b}_{0,\mathcal{J}_1} \end{pmatrix}$$

y $\text{EIF}(\widehat{\Delta}_{\mathcal{J}_2}) = \mathbf{0}$. Usando la expresión para la inversa de una matriz por bloques, la función de influencia empírica de $\widehat{\beta}$ se obtiene como

$$\text{EIF}(\widehat{\beta}) = -\left(\mathbf{A}_n^{-1} + \mathbf{F}_{n,\mathcal{J}_1} \mathbf{E}_{n,\mathcal{J}_1}^{-1} \mathbf{F}_{n,\mathcal{J}_1}^T \right) \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{x}_0 + \mathbf{F}_{n,\mathcal{J}_1} \mathbf{E}_{n,\mathcal{J}_1}^{-1} \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{b}_{0,\mathcal{J}_1},$$

con $\mathbf{E}_{n,\mathcal{J}_1} = \mathbf{C}_{n,\mathcal{J}_1} - \mathbf{B}_{n,\mathcal{J}_1}^T \mathbf{A}_n^{-1} \mathbf{B}_{n,\mathcal{J}_1}$ y $\mathbf{F}_{n,\mathcal{J}_1} = \mathbf{A}_n^{-1} \mathbf{B}_{n,\mathcal{J}_1}$, mientras que $\text{EIF}(\widehat{\Delta}_{\mathcal{J}_1})$ está dado por

$$\text{EIF}(\widehat{\Delta}_{\mathcal{J}_1}) = \mathbf{E}_{n,\mathcal{J}_1}^{-1} \mathbf{F}_{n,\mathcal{J}_1}^T \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{x}_0 - \mathbf{E}_{n,\mathcal{J}_1}^{-1} \Lambda(y_0, \mathbf{x}_0, \mathbf{b}_0) \mathbf{b}_{0,\mathcal{J}_1}.$$

■

Capítulo 6

Estudio numérico y análisis de un conjunto de datos reales

6.1. Estudio de Monte Carlo

En esta sección describimos los resultados de un estudio de simulación diseñado para comparar el desempeño de los estimadores propuestos con los estimadores clásicos bajo un modelo de regresión parcialmente lineal isotónico log–Gamma.

En todas las tablas y figuras, los estimadores robustos son indicados como ROB y sus contrapartes clásicas como CL, en tanto corresponden a estimadores basados en la *deviance*. Los estimadores robustos corresponden a aquellos que controlan los valores grandes de la *deviance* como se describió en la Sección 2.5 y fueron calculados usando la función bicuadrada de Tukey y la función de peso w usada en la Sección 5.2. Por otro lado, para los estimadores clásicos $w \equiv 1$ y $\rho(y, s) = d(y - s)$ con $d(u) = \exp(u) - u - 1$.

Realizamos $NR = 1000$ replicaciones con muestras de tamaño $n = 100$. El modelo central, denotado C_0 , corresponde a elegir (x_i, t_i) independientes entre sí tales que $x_i \sim N(0, 1)$, $t_i \sim \mathcal{U}(0, 1)$. La variable respuesta fue generada como en la Sección 5.2, esto es, el modelo log–Gamma transformado corresponde a $y_i = \beta_0 x_i + \eta_0(t_i) + u_i$, con $u_i \sim \log(\Gamma(3, 1))$, $\beta_0 = 2$ y $\eta_0 = \eta_{0,1}$ o $\eta_0 = \eta_{0,2}$ para los Modelos 1 y 2, respectivamente, donde, $\eta_{0,1}(t) = \sin(\pi t/2)$ o $\eta_{0,2}(t) = \pi t + 0.25 \sin(4\pi t)$.

Para cada muestra generada consideramos tres contaminaciones C_1 , C_2 y C_3 que producen muestras contaminadas $(y_{i,c}, x_{i,c}, t_i)$. Primero generamos una muestra $v_i \sim \mathcal{U}(0, 1)$ para $1 \leq i \leq n$ y luego consideramos los siguientes esquemas de contaminación:

- C_1 introduce puntos dañinos de alta palanca en las covariables x , sin cambiar las respuestas

ya generadas, o sea, $y_{i,c} = y_i$, $1 \leq i \leq n$, mientras

$$x_{i,c} = \begin{cases} x_i & \text{if } v_i \leq 0.90 \\ x_i^* & \text{if } v_i > 0.90, \end{cases}$$

donde $x_i^* \sim N(5, 1/16)$.

- C_2 introduce observaciones atípicas en las respuestas generadas según el modelo pero con una covariable incorrecta x .

$$y_{i,c} = \begin{cases} y_i & \text{if } v_i \leq 0.90 \\ y_i^* & \text{if } v_i > 0.90, \end{cases}$$

donde $y_i^* = \beta_0 x_i^* + \eta_0(t_i) + u_i^*$ con $u_i^* \sim \log(\Gamma(3, 1))$ y x_i^* una nueva de una $N(5, 1/16)$. Notar que las covariables no son contaminadas en esta situación, es decir, $x_{i,c} = x_i$.

- C_3 corresponde a incrementar la varianza de las covariables x y también introduce valores grandes en las respuestas

$$x_{i,c} = \begin{cases} x_i & \text{if } v_i \leq 0.90 \\ \text{una nueva observación de una } N(0, 25) & \text{si } v_i > 0.90, \end{cases} \quad y_{i,c} = \begin{cases} y_i & \text{si } v_i \leq 0.90 \\ y_i^* & \text{si } v_i > 0.90, \end{cases}$$

con $y_i^* = 3 \log(10) + u_i^*$ donde $u_i^* \sim \log(\Gamma(3, 1))$.

La Tabla 6.1 resume los resultados obtenidos con el estimador basado en splines monótonos descrito en la Sección 2.2. El valor de k_n fue elegido como se describe en la Sección 2.6. Dicha tabla reporta la media sobre las replicaciones de $\hat{\beta} - \beta_0$, denotada Sesgo, su desvío estándar, SD, el error cuadrático medio, MSE, esto es, la media sobre las replicaciones de $(\hat{\beta} - \beta_0)^2$. También se reporta la media sobre las replicaciones de los errores estándar estimados, denotados AS.SE y la probabilidad de cubrimiento Cov.Prob para un intervalo de confianza de nivel 95 %, esto es, la proporción de veces que el intervalo de confianza asintótico contiene al valor verdadero β_0 . Para muestras sin contaminar, el error estándar asintótico del estimador clásico es igual a $\sqrt{1/\alpha_0}/\sqrt{n} \approx 0.0577$, cercano a los valores reportados bajo C_0 en la Tabla 6.1 para ambos modelos.

Para estudiar el desempeño de un estimador $\hat{\eta}$ de la función de regresión η_0 consideramos el error cuadrático medio integrado $MISE(\hat{\eta})$, o sea, la media sobre las replicaciones de una aproximación del error cuadrático integrado dado por

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

Por otra parte, los resultados obtenidos para los estimadores monotonizados basados en B-splines, $\hat{\eta}_{MOD}$, descritos en la Sección 2.3, se presentan en las Tablas 6.2 y 6.3, dependiendo de la

manera en que se elige el valor k_n que es la cantidad de splines del estimador inicial sin restricciones $\hat{\eta}_{\text{INI}} = \sum_{j=1}^{k_n} \lambda_j B_j$. Más precisamente, se exploraron dos variantes. En la primera, a partir de los estimadores iniciales sin restricciones $(\hat{\beta}_{\text{INI}}, \hat{\lambda}_{\text{INI}})$, como por ejemplo los calculados en el **Paso 3** del algoritmo presentado en la Sección 2.5.1, se eligió la cantidad de splines k_n como el valor que minimiza $BIC(k_n)$. Se le aplicó luego a este estimador inicial la monotización descrita en la Sección 2.3 para obtener los estimadores finales $(\hat{\beta}, \hat{\eta}_{\text{i,D}})$. En la segunda forma de elegir la cantidad de splines, se invirtieron los dos procesos, es decir, primero se obtiene, para cada valor de k_n , la versión monotizada de los estimadores iniciales calculados sin restricción y luego se elige la cantidad de splines basada en $BIC(k)$ utilizando el estimador monotizado. Las Tablas 6.2 y 6.3 reportan respectivamente las medidas resumen correspondientes a dichas variantes, respectivamente.

El estimador clásico muestra su sensibilidad bajo todas las contaminaciones, siendo peor el efecto en la estimación de la función de regresión η_0 al contaminar las respuestas como en C_2 o C_3 . Para estas dos contaminaciones, el error cuadrático medio de los estimadores clásicos de η_0 es más de mil veces que los obtenidos por el procedimiento robusto que están bastante cerca de los correspondientes bajo C_0 . Por otra parte, contaminar solo las covariables duplica el error cuadrático medio de las estimaciones clásicas $\hat{\eta}_{\text{CL}}$. Por lo tanto, como esperábamos, respuestas grandes afectan a los estimadores de la componente no paramétrica más que los puntos de alta palanca.

Cabe notar que, para el modelo log-Gamma estudiado, tanto el sesgo como la dispersión de los estimadores clásicos de β_0 aumentan bajo C_2 agrandando el error cuadrático medio. Estas observaciones atípicas también afectan el error estándar asintótico que es 7 veces el obtenido para muestras sin contaminar. Aunque el estimador de la varianza asintótica se vea aumentado, el efecto en el sesgo de la contaminación considerada rompe la probabilidad de cubrimiento. Por otro lado, el error cuadrático medio aumentado bajo C_3 se debe principalmente al sesgo y este efecto lleva nuevamente a un muy bajo cubrimiento.

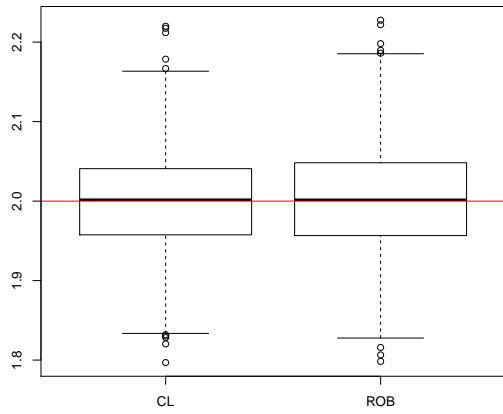
El efecto de las diferentes contaminaciones también es llamativo en las Figuras 6.1 y 6.2 las cuales dan los boxplots de $\hat{\beta}$ bajo los Modelos 1 y 2, respectivamente. Por ejemplo, bajo C_1 y C_3 , todo el boxplot de las estimaciones clásicas está por debajo de la línea horizontal que corresponde al verdadero valor $\beta_0 = 2$. Más aun, las diferencias existentes entre el desvío estándar de $\hat{\beta}$ sobre las replicaciones y la media del error estándar estimado puede ser explicada por los boxplots de $\hat{\beta}$. El efecto de los esquemas de contaminación considerados en los estimadores clásicos aumenta considerablemente el valor de sus desvíos estándar debido a las observaciones atípicas que aparecen en los boxplots, mientras que los errores estándar estimados, relacionados a los estimadores del parámetro de ruido, muestran un aumento menor. Por otro lado, los estimadores robustos son bastante estables a través de los distintos escenarios de contaminación. En particular, la probabilidad

Modelo 1							
Medidas resumen para $\hat{\beta}$							MISE($\hat{\eta}$)
Estimador	Sesgo	SD	MSE	AS.SE	Cov.Prob		
C_0	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096
C_1	CL	-0.5497	0.2170	0.3492	0.0535	0.0050	0.0265
	ROB	-0.0016	0.0706	0.0050	0.0591	0.8850	0.0100
C_2	CL	-1.8359	0.9343	4.2426	0.3781	0.0690	54.3390
	ROB	0.0002	0.0711	0.0051	0.0639	0.9170	0.0103
C_3	CL	-1.9400	0.2721	3.8376	0.1848	0.0100	15.0401
	ROB	0.0043	0.0727	0.0053	0.0598	0.8900	0.0146

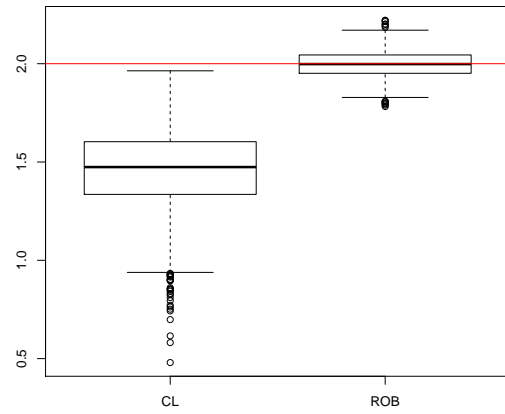
Modelo 2							
Medidas resumen para $\hat{\beta}$							MISE($\hat{\eta}$)
Estimador	Sesgo	SD	MSE	AS.SE	Cov.Prob		
C_0	CL	0.0000	0.0636	0.0040	0.0588	0.9340	0.0324
	ROB	0.0019	0.0700	0.0049	0.0639	0.9190	0.0340
C_1	CL	-0.5549	0.2215	0.3570	0.0542	0.0040	0.0556
	ROB	-0.0020	0.0728	0.0053	0.0609	0.8830	0.0344
C_2	CL	-1.8168	0.9665	4.2340	0.3748	0.0850	52.8369
	ROB	-0.0001	0.0736	0.0054	0.0660	0.9250	0.0348
C_3	CL	-1.9116	0.2581	3.7207	0.1654	0.0090	10.1817
	ROB	0.0020	0.0749	0.0056	0.0618	0.8940	0.0350

Tabla 6.1: Medidas resumen para los estimadores de β_0 y η_0 basados en B-splines monótonos, bajo un modelo log-Gamma. Los estimadores fueron obtenidos cuando k_n es el número de nodos, basado en la muestra, que minimiza $BIC(k)$.

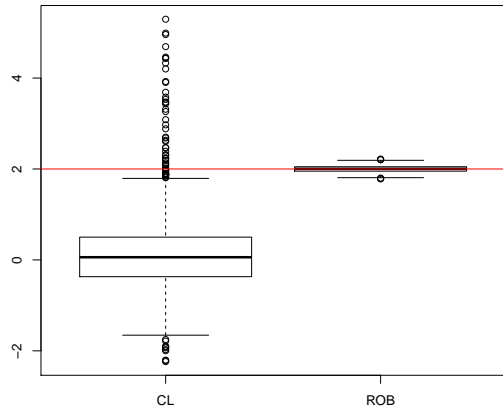
de cubrimiento nunca baja de 0.88 en todos los esquemas de contaminación. Para muestras sin contaminar, la probabilidad de cubrimiento está ligeramente por debajo del procedimiento clásico y este efecto está relacionado a la pérdida de eficiencia. Más aun, la estabilidad del procedimiento robusto es claramente ilustrada en la Figura 6.3 que presenta los estimadores de la densidad de las estimaciones clásicas y robustas, $\hat{\beta}_{CL}$ y $\hat{\beta}_{ROB}$, bajo los diferentes esquemas de contaminación con las ventanas calculadas automáticamente a partir de los datos. La línea sólida negra corresponde a las muestras sin contaminar, mientras que la roja discontinua, la punteada azul y la discontinua punteada granate a las contaminaciones C_1 a C_3 respectivamente. Además, la línea discontinua



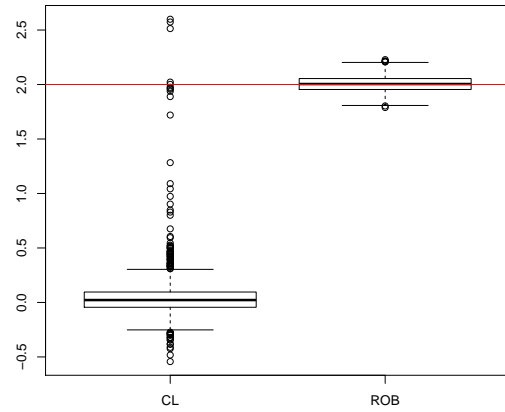
(a) C_0



(b) C_1



(c) C_2



(d) C_3

Figura 6.1: Boxplots de los estimadores $\hat{\beta}$ de β_0 basados en B-splines monótonos, bajo un modelo log-Gamma con $\eta_0 = \eta_{0,1}$.

verde corresponde a la densidad normal con media 2 y desvío estándar igual a 0.0608 y 0.0672 para el estimador clásico y robusto, respectivamente. Notemos que estos valores corresponden a los desvíos estándar de los estimadores $\hat{\beta}_{CL}$ y $\hat{\beta}_{ROB}$ reportados en la Tabla 6.1, para muestras sin contaminar. Al considerar la propuesta robusta, todos los estimadores de la densidad están superpuestos mostrando que las contaminaciones tienen un efecto leve en las estimaciones. Por otro lado, al utilizar el procedimiento clásico basado en la *deviance*, las densidades de las estimaciones calculadas con muestras contaminadas se alejan de las obtenidas con muestras sin contaminar, conduciendo a estimaciones poco fiables. Notemos también que para muestras sin contaminar, tanto para el procedimiento robusto como el clásico, la media sobre las replicaciones de los errores estándar estimados está cerca de los desvíos estándar de las estimaciones obtenidas $\hat{\beta}$. Este resultado empírico

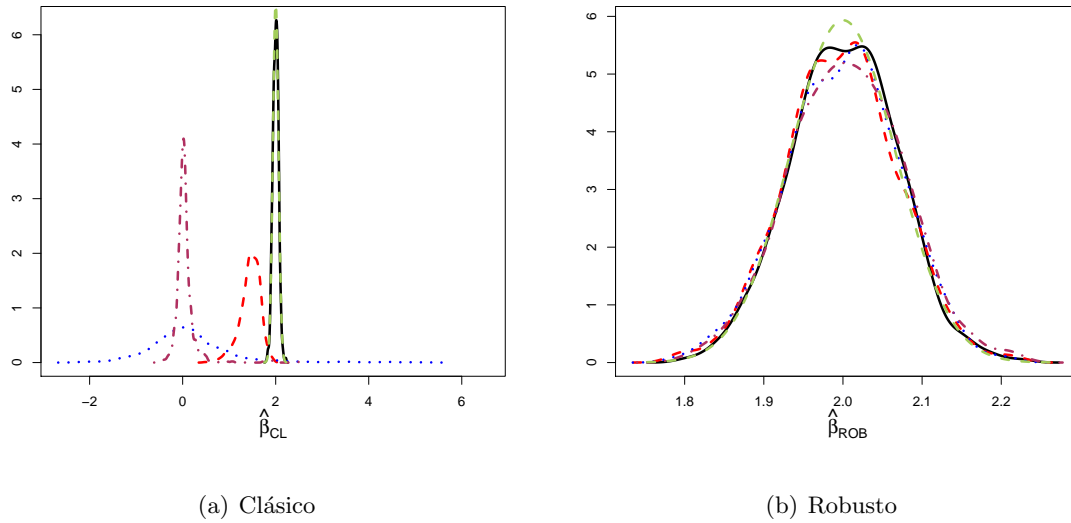


Figura 6.3: Estimador de la densidad de los estimadores clásico y robusto, $\hat{\beta}_{CL}$ y $\hat{\beta}_{ROB}$, de β_0 , bajo un modelo log-Gamma con $\eta_0 = \eta_{0,1}$. La línea sólida negra corresponde a las muestras sin contaminar, mientras que la discontinua roja, la punteada azul y la discontinua punteada granate a las contaminaciones C_1 a C_3 respectivamente. La línea verde es una distribución normal con media igual al parámetro verdadero β_0 y desvío estándar igual al desvío estándar del estimador bajo C_0 .

corresponde a las muestras sin contaminar, mientras que la roja discontinua, la punteada azul y la discontinua punteada granate a las contaminaciones C_1 a C_3 respectivamente. Además, la línea discontinua verde corresponde a la densidad normal con media 2 y desvío estándar igual a los valores de desvíos estándar reportados en las Tablas 6.2 y 6.3. Por otra parte, los boxplots de los estimadores de β_0 dados en las Figuras 6.4, 6.5, 6.6 y 6.7 ilustran claramente que las contaminaciones correspondientes a respuestas grandes C_2 y C_3 , impactan peor en los estimadores del parámetro de regresión que la contaminación C_1 , correspondiente solamente a los puntos de alta palanca. Para los estimadores monotonizados, el efecto es mucho más notorio para el Modelo 2 de lo que se observaba en la Figura 6.2 con los estimadores basados en splines monótonos. Dentro de las variantes de los estimadores basados en B-splines sin restricciones, los resultados prácticamente no ofrecen diferencias tanto para la componente paramétrica como para el término no paramétrico. Por último, vale la pena observar que el método de estimación por B-splines monótonos da estimadores menos variables que el procedimiento de monotonizar estimadores sin restricciones. Más precisamente, el MISE de $\hat{\eta}$ así como los valores de SD de $\hat{\beta}$ (y por lo tanto el MSE) son menores para todos los escenarios al utilizar splines monótonos. Asimismo, la probabilidad de cubrimiento es ligeramente menor cuando se aplica el procedimiento de monotonización en lugar de estimar directamente con restricciones. Por lo antedicho, en base a los resultados observados, el método de estimación por B-splines monótonos resulta más atractivo. Por esta razón, en el análisis de datos reales presentado

en la próxima sección se utiliza este procedimiento.

Modelo 1							
Medidas resumen para $\hat{\beta}$							MISE($\hat{\eta}_{\text{MOD}}$)
Estimador	Sesgo	SD	MSE	AS.SE	Cov.Prob		
C_0	CL	0.0009	0.0613	0.0038	0.0565	0.9280	0.0118
	ROB	-0.0000	0.0921	0.0085	0.0620	0.8060	0.0157
C_1	CL	-0.5656	0.2324	0.3739	0.0547	0.0050	0.0458
	ROB	0.0002	0.0871	0.0076	0.0589	0.8010	0.0155
C_2	CL	-0.8331	1.1950	2.1207	0.7523	0.6710	55.3241
	ROB	0.0003	0.0957	0.0092	0.0641	0.8100	0.0159
C_3	CL	-1.7090	0.4786	3.1494	0.3697	0.1130	17.6784
	ROB	0.0012	0.0902	0.0081	0.0600	0.7950	0.0155

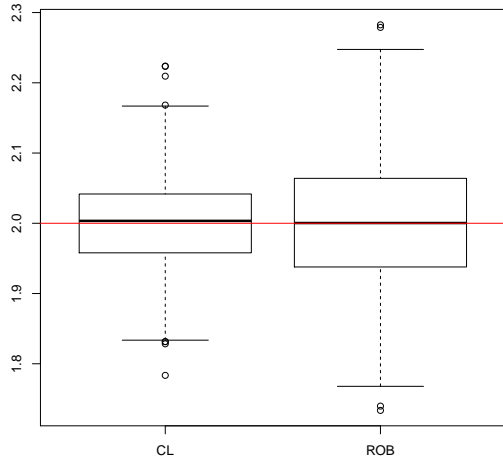
Modelo 2							
Medidas resumen para $\hat{\beta}$							MISE($\hat{\eta}_{\text{MOD}}$)
Estimador	Sesgo	SD	MSE	AS.SE	Cov.Prob		
C_0	CL	0.0007	0.0639	0.0041	0.0585	0.9330	0.0336
	ROB	-0.0015	0.0945	0.0089	0.0639	0.8110	0.0382
C_1	CL	-0.5725	0.2373	0.3839	0.0548	0.0040	0.0665
	ROB	-0.0008	0.0892	0.0080	0.0607	0.8120	0.0378
C_2	CL	-0.8003	1.1671	2.0011	0.7370	0.6800	54.8082
	ROB	-0.0007	0.0980	0.0096	0.0660	0.8030	0.0385
C_3	CL	-1.7018	0.4401	3.0898	0.3415	0.1070	9.7977
	ROB	0.0035	0.0932	0.0087	0.0618	0.8080	0.0382

Tabla 6.2: Medidas resumen para los estimadores de β_0 y η_0 descritos en la Sección 2.3, bajo un modelo log-Gamma. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que minimiza $BIC(k)$ primero y luego aplica el procedimiento de monotización.

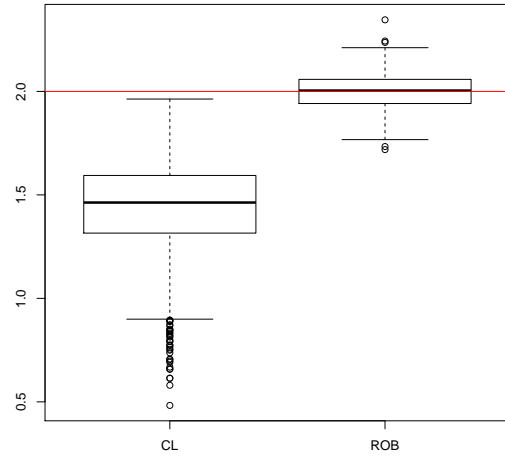
Modelo 1							
Medidas resumen para $\hat{\beta}$							MISE($\hat{\eta}_{\text{MOD}}$)
Estimador	Sesgo	SD	MSE	AS.SE	Cov.Prob		
C_0	CL	0.0009	0.0612	0.0037	0.0565	0.9280	0.0118
	ROB	-0.0000	0.0921	0.0085	0.0620	0.8060	0.0157
C_1	CL	-0.5601	0.2238	0.3638	0.0537	0.0050	0.0353
	ROB	0.0002	0.0871	0.0076	0.0589	0.8010	0.0155
C_2	CL	-1.8721	1.3945	5.4473	0.6797	0.1320	82.4910
	ROB	0.0003	0.0957	0.0092	0.0641	0.8100	0.0159
C_3	CL	-1.9122	0.2953	3.7435	0.2364	0.0170	14.2858
	ROB	0.0012	0.0902	0.0081	0.0600	0.7950	0.0155

Modelo 2							
Medidas resumen para $\hat{\beta}$							MISE($\hat{\eta}_{\text{MOD}}$)
Estimador	Sesgo	SD	MSE	AS.SE	Cov.Prob		
C_0	CL	0.0005	0.0639	0.0041	0.0585	0.9330	0.0337
	ROB	-0.0015	0.0945	0.0089	0.0639	0.8110	0.0382
C_1	CL	-0.5637	0.2274	0.3694	0.0542	0.0040	0.0631
	ROB	-0.0008	0.0892	0.0080	0.0607	0.8120	0.0378
C_2	CL	-1.8905	1.2763	5.2014	0.5290	0.1430	79.6413
	ROB	-0.0007	0.0980	0.0096	0.0660	0.8030	0.0385
C_3	CL	-1.8880	0.2779	3.6416	0.2012	0.0120	9.5721
	ROB	0.0035	0.0932	0.0087	0.0618	0.8080	0.0382

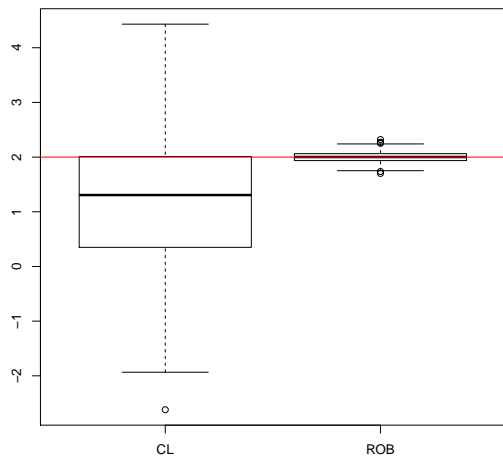
Tabla 6.3: Medidas resumen para los estimadores de β_0 y η_0 descritos en la Sección 2.3, bajo un modelo log-Gamma. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que aplica el procedimiento de monotonización a los estimadores iniciales calculados sin restricción y luego elige aquellos que minimizan $BIC(k)$.



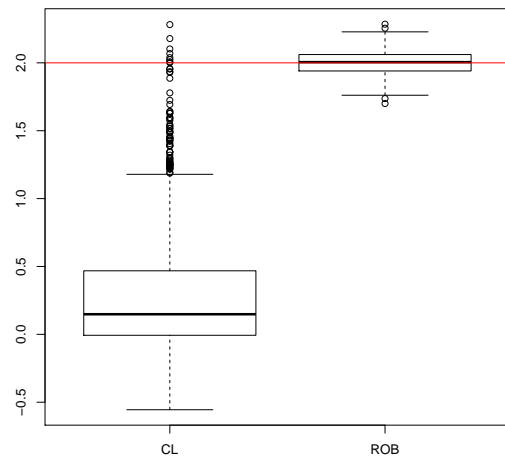
(a) C_0



(b) C_1

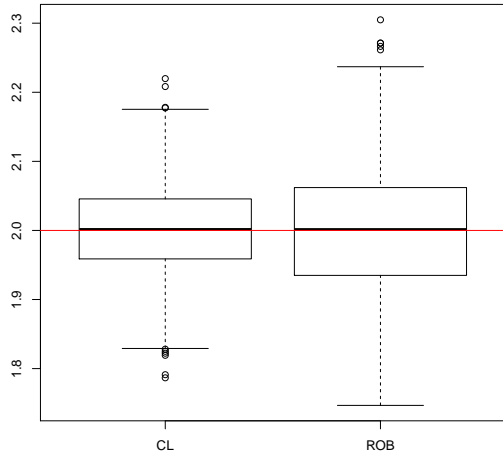


(c) C_2

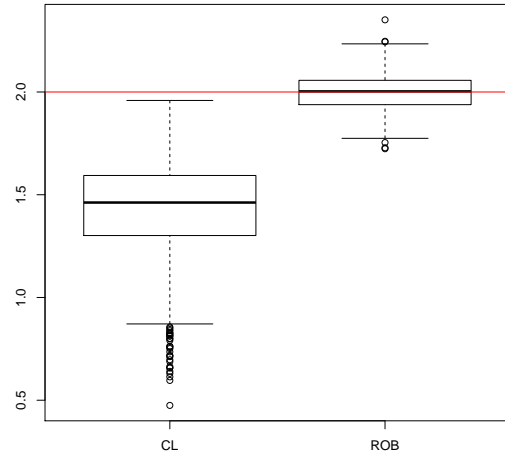


(d) C_3

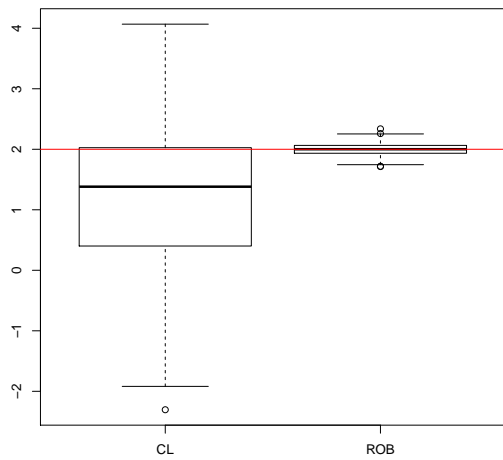
Figura 6.4: Boxplots de los estimadores $\hat{\beta}$ de β_0 basados en B-splines sin restricciones, bajo un modelo log-Gamma con $\eta_0 = \eta_{0,1}$. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que minimiza $BIC(k)$ primero y luego aplica el procedimiento de monotonización.



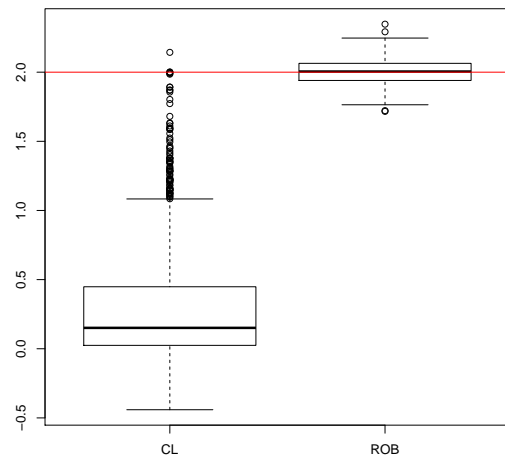
(a) C_0



(b) C_1

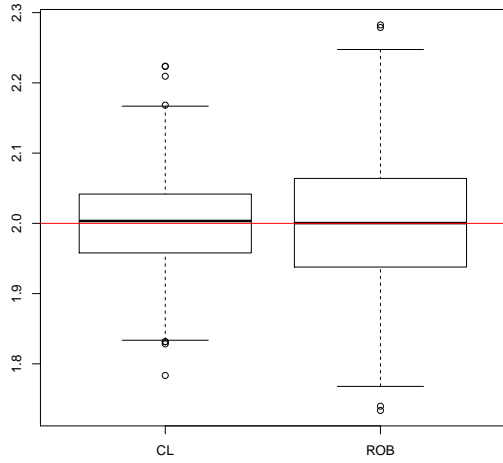


(c) C_2

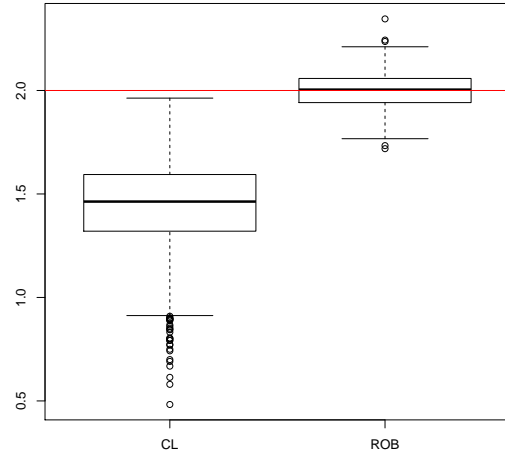


(d) C_3

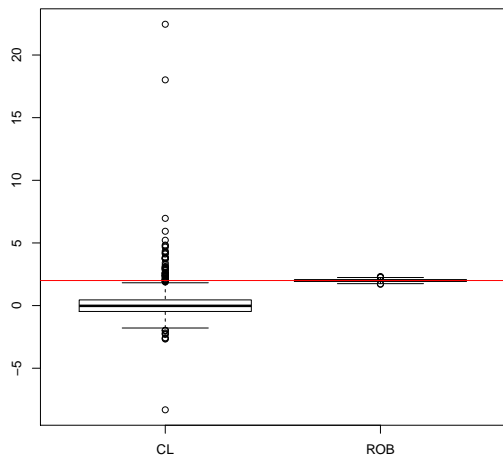
Figura 6.5: Boxplots de los estimadores de $\hat{\beta}$ de β_0 basados en B-splines sin restricciones, bajo un modelo log-Gamma con $\eta_0 = \eta_{0,2}$. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que minimiza $BIC(k)$ primero y luego aplica el procedimiento de monotonización.



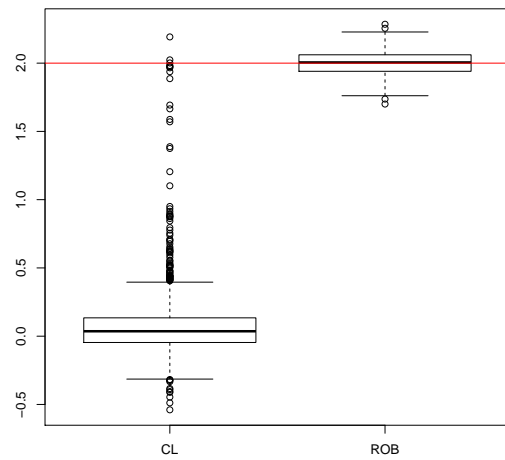
(a) C_0



(b) C_1

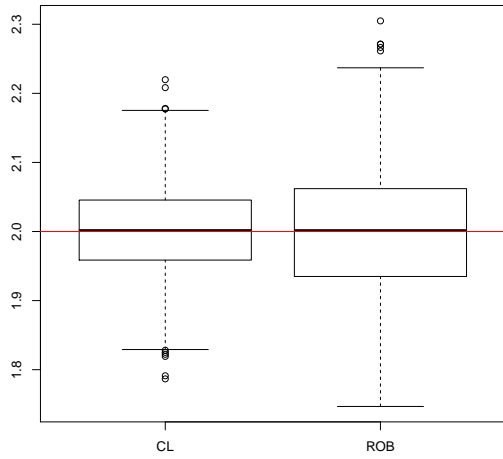


(c) C_2

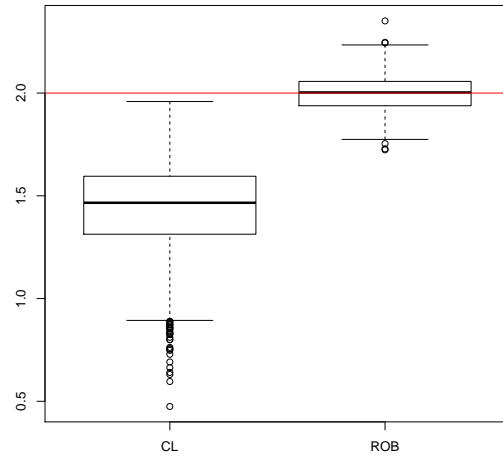


(d) C_3

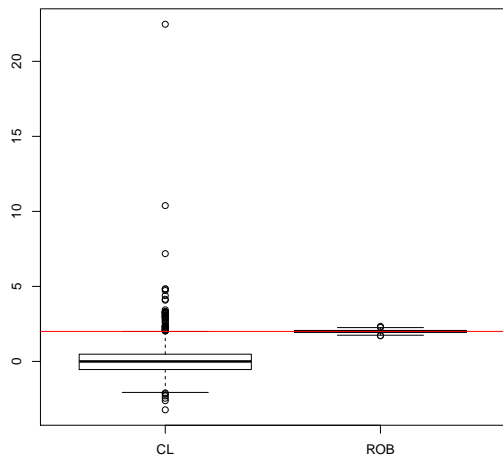
Figura 6.6: Boxplots de los estimadores $\hat{\beta}$ de β_0 basados en B-splines sin restricciones, bajo un modelo log-Gamma con $\eta_0 = \eta_{0,1}$. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que aplica el procedimiento de monotonización a los estimadores iniciales calculados sin restricción y luego elige aquellos que minimizan $BIC(k)$.



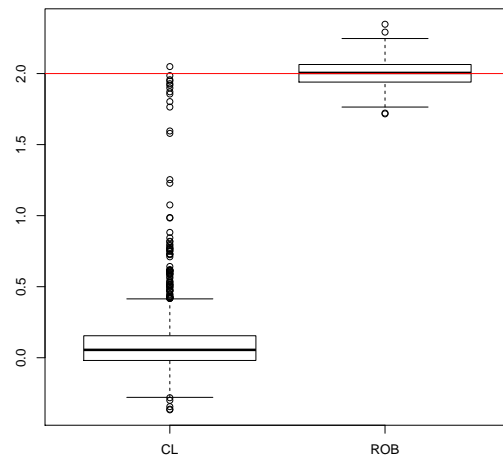
(a) C_0



(b) C_1

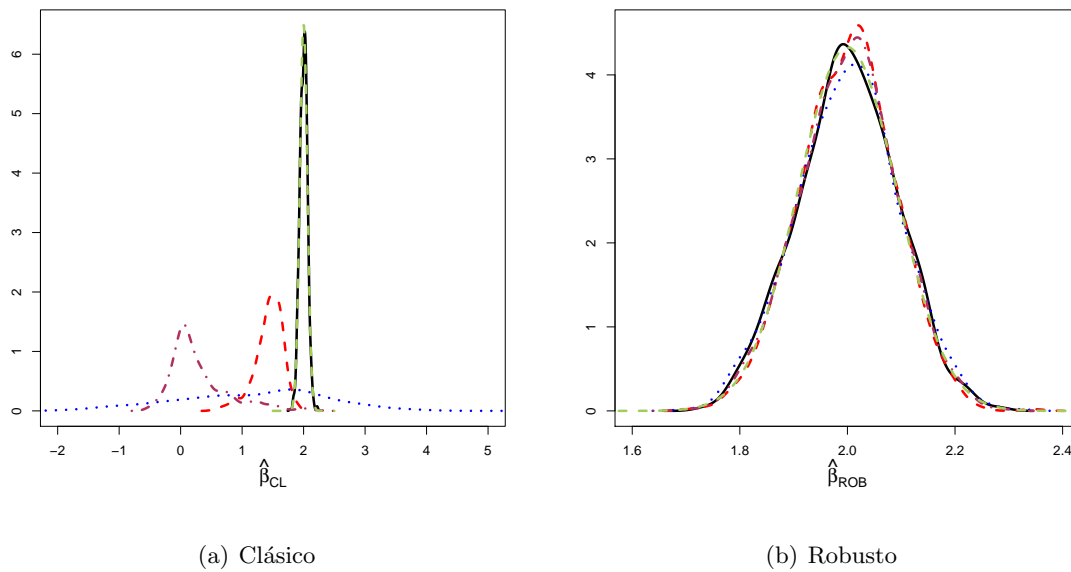


(c) C_2



(d) C_3

Figura 6.7: Boxplots de los estimadores de $\hat{\beta}$ de β_0 basados en B-splines sin restricciones, bajo un modelo log-Gamma con $\eta_0 = \eta_{0,2}$. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que aplica el procedimiento de monotonización a los estimadores iniciales calculados sin restricción y luego elige aquellos que minimizan $BIC(k)$.



(a) Clásico

(b) Robusto

Figura 6.8: Estimador de la densidad de los estimadores clásico y robusto, $\hat{\beta}_{CL}$ y $\hat{\beta}_{ROB}$, de β_0 , bajo un modelo log-Gamma con $\eta_0 = \eta_{0,1}$. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que minimiza $BIC(k)$ primero y luego aplica el procedimiento de monotización. La línea sólida negra corresponde a las muestras sin contaminar, mientras que la discontinua roja, la punteada azul y la discontinua punteada granate a las contaminaciones C_1 a C_3 respectivamente. La línea verde es una distribución normal con media igual al parámetro verdadero β_0 y desvío estándar igual al desvío estándar del estimador bajo C_0 .

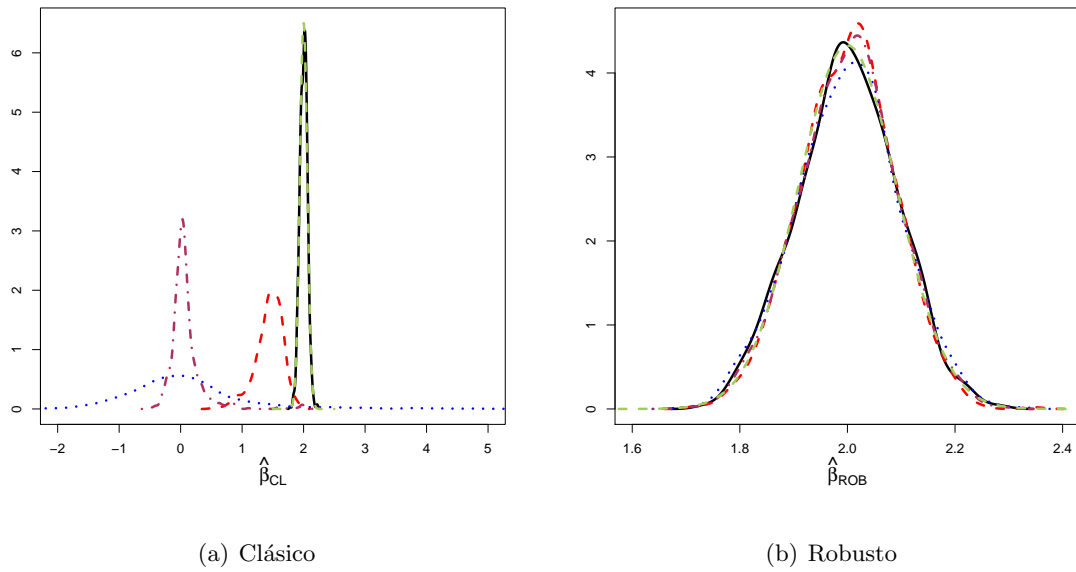


Figura 6.9: Estimador de la densidad de los estimadores clásico y robusto, $\hat{\beta}_{\text{CL}}$ y $\hat{\beta}_{\text{ROB}}$, de β_0 , bajo un modelo log-Gamma con $\eta_0 = \eta_{0,1}$, descritos en la Sección 2.3. Los estimadores corresponden a los obtenidos cuando el valor de k_n se elige de acuerdo a la variante que aplica el procedimiento de monotonización a los estimadores iniciales calculados sin restricción y luego elige aquellos que minimizan $BIC(k)$. La línea sólida negra corresponde a las muestras sin contaminar, mientras que la discontinua roja, la punteada azul y la discontinua punteada granate a las contaminaciones C_1 a C_3 respectivamente. La línea verde es una distribución normal con media igual al parámetro verdadero β_0 y desvío estándar igual al desvío estándar del estimador bajo C_0 .

6.2. Aplicación a un conjunto de datos reales: Costos Hospitalarios

Marazzi y Yohai (2004) introdujeron un conjunto de datos que corresponde a los costos hospitalarios, por problemas de espalda, de 100 pacientes en un hospital Suizo en 1999. Ellos se interesaron en la relación entre el costo hospitalario de la estadía, z , (Costo, en francos suizos) y las siguientes variables administrativas:

- *LOS*: duración de la estadía en días
- *ADM*: tipo de admisión (0 = planificada; 1 = emergencia)
- *INS*: tipo de seguro (0 = regular; 1 = privado)
- *AGE*: edad
- *SEX*: (0 = mujer; 1 = hombre)
- *DEST*: destino del alta (1 = hogar; 0 = otro)

Cantoni y Ronchetti (2006) ajustaron al conjunto de datos completo el modelo $\log(\mathbb{E}(z_i|\mathbf{x}_i)) = \boldsymbol{\gamma}_0^T \mathbf{x}_i$ que para respuestas Gamma es equivalente a $y_i = \log(z_i) = \boldsymbol{\gamma}_0^T \mathbf{x}_i + u_i$, donde $u_i \sim \log \Gamma(\alpha, 1)$ y $\mathbf{x} = (ADM, INS, AGE, SEX, DEST, \log(LOS), 1)$. Usando su propuesta robusta, identificaron 5 datos atípicos correspondientes a las observaciones etiquetadas como 14, 21, 28, 44 y 63, cuyos pesos son menores o iguales a 0.5. Observaron que los puntos atípicos afectaban a las estimaciones clásicas del coeficiente de la variables *INS* y al parámetro de forma. Bianco *et al.* (2013b) también analizaron este conjunto de datos para realizar tests para las covariables *SEX* y *DEST*.

En este ejemplo, nuestro análisis no impone una relación lineal entre y_i y $\log LOS$ pero consideramos el modelo, más general, parcialmente lineal isotónico $y_i = \boldsymbol{\beta}_0^T \mathbf{x}_i + \eta_0(t_i) + u_i$, con $u_i \sim \log(\Gamma(\alpha_0, 1))$, $\mathbf{x} = (ADM, INS, AGE, SEX, DEST)$ y $t = \log(LOS)$. La función η_0 se supone no decreciente lo que resulta natural en este ejemplo, en tanto los costos hospitalarios aumentan para estadías más largas. Los resultados obtenidos para los estimadores de $\boldsymbol{\beta}_0$ basados en B-splines monótonos se reportan en la Tabla 6.4. Para las estimaciones clásicas, denotadas $\widehat{\boldsymbol{\beta}}_{CL}$, el criterio *BIC* eligió $k_n = 4$, mientras que para los robustos, $\widehat{\boldsymbol{\beta}}_{ROB}$, la mejor elección fue $k_n = 5$ y la constante de calibración para la ρ -función que acota las *deviance* resultó igual a $c_\rho = 0.3515$. Como en el ajuste lineal, el estimador clásico de $\boldsymbol{\beta}_0$ es muy sensible a los 5 datos atípicos, que también fueron detectados en nuestro análisis. En particular, los estimadores del parámetro de forma y del coeficiente relacionado con el tipo de seguro se ven altamente afectados. Luego de eliminar estos 5 datos, el estimador clásico $\widehat{\boldsymbol{\beta}}_{CL}^{-\{5\}}$ es muy similar a $\widehat{\boldsymbol{\beta}}_{ROB}$, mostrando el buen desempeño de la propuesta robusta en presencia de datos atípicos. Calculamos los errores estándar asintóticos de las

estimaciones de β que se reportan entre paréntesis. Notemos que los errores estándar asintóticos de los estimadores robustos y los clásicos calculados sin los 5 datos atípicos, $\widehat{\beta}_{\text{CL}}^{-\{5\}}$ son bastante similares mientras que los de los estimadores clásicos $\widehat{\beta}_{\text{CL}}$ se ven aumentados.

La Figura 6.10 muestra el gráfico de los estimadores de η_0 obtenidos usando los estimadores clásicos (línea discontinua roja) y los robustos (línea discontinua azul). También graficamos el ajuste dado por el modelo de regresión lineal log–Gamma cuando usamos un MM–estimador pesado para γ_0 , es decir, que la línea sólida negra corresponde al ajuste $\eta(t) = 0.8892t + 7.1268$. El ajuste lineal parece ser una buena elección para este conjunto de datos, sin embargo, aparecen algunas discrepancias cerca de los bordes que pueden ser causadas por una forma diferente de la función de regresión para valores grandes de $\log(\text{LOS})$. Cabe notar que en este caso, la forma del estimador clásico es bastante cercana a la del robusto y esto puede explicarse principalmente por la estructura isotónica impuesta y el tipo de datos atípicos que afectan principalmente al parámetro de regresión.

	$\widehat{\beta}_{\text{CL}}$		$\widehat{\beta}_{\text{CL}}^{-\{5\}}$		$\widehat{\beta}_{\text{ROB}}$	
<i>ADM</i>	0.2148	(0.0497)	0.2172	(0.0345)	0.1979	(0.0339)
<i>INS</i>	0.0984	(0.0792)	-0.0324	(0.0575)	-0.0207	(0.0537)
<i>AGE</i>	-0.0009	(0.0013)	-0.0016	(0.0009)	-0.0019	(0.0009)
<i>SEX</i>	0.1088	(0.0529)	0.0820	(0.0354)	0.0615	(0.0358)
<i>DEST</i>	-0.1358	(0.0723)	-0.1608	(0.0489)	-0.1673	(0.0493)
$\widehat{\alpha}$	21.0809	-	45.7560	-	46.0088	-

Tabla 6.4: Análisis de los costos hospitalarios bajo un modelo de regresión parcialmente lineal isotónico log–Gamma.

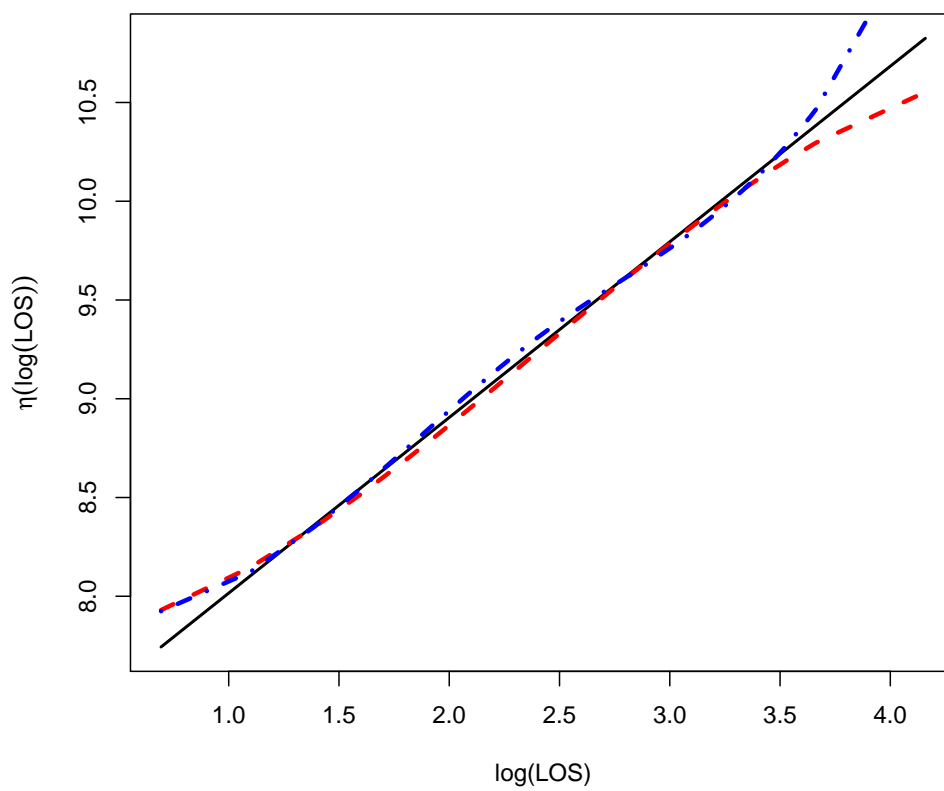


Figura 6.10: Estimadores clásicos (rojo) y robustos (azul) de la función de regresión $\eta(t)$. El ajuste lineal corresponde a la línea sólida negra.

Capítulo 7

Consideraciones finales

El problema de estimar la componente no paramétrica η_0 y el parámetro de regresión β_0 bajo un modelo parcialmente lineal generalizado ha sido extensamente estudiado. Entre otros métodos, los B-splines han sido considerados para aproximar la función desconocida η_0 . Una ventaja de los B-splines es que proveen un procedimiento de estimación que puede extenderse a la situación en que hay restricciones de monotonía en la componente no paramétrica imponiendo restricciones de no decrecimiento en los coeficientes. Para superar la sensibilidad del procedimiento clásico basado en la *deviance* hemos introducido dos familias de estimadores robustos para las componentes de un modelo parcialmente lineal generalizado. La primera está basada en B-splines monótonos y la segunda usa una monotonización del estimador de B-splines sin restricciones, para estimar la componente no paramétrica. En ambos casos, usamos una función de pérdida acotada para controlar los residuos *deviance* grandes. Una de las ventajas de nuestro procedimiento es que también permite incluir modelos con un parámetro de ruido desconocido, como el parámetro de escala en el modelo de regresión parcialmente lineal o el parámetro de forma en el modelo de regresión parcialmente lineal con errores log-Gamma. La estimación del parámetro de ruido es un problema importante porque permite calibrar los estimadores robustos y atenuar el peso de residuos grandes. De hecho, como en regresión lineal, para decidir si una observación es atípica es necesario determinar el tamaño de los residuos que dependen fuertemente del estimador del parámetro de ruido.

Los estimadores obtenidos son consistentes y obtuvimos tasas de convergencia para los estimadores sin restricciones y para los basados en B-splines monótonos. Hemos estudiado también la función de influencia empírica de los estimadores propuestos que permite cuantificar la sensibilidad frente a datos anómalos de los estimadores del parámetro de regresión y de la función de regresión. Como era de esperar, el enfoque clásico basado en la *deviance* no es robusto porque conduce a una función de influencia empírica no acotada, mientras que para nuestra propuesta robusta la influencia es acotada. El comportamiento inadecuado del método clásico cuando hay datos atípicos

en la muestra se confirma también en nuestro estudio de simulación. El procedimiento robusto da origen estimadores más confiables conduciendo a casi los mismos resultados tanto bajo el modelo log–Gamma central o bajo las contaminaciones estudiadas.

En esta primera parte, hemos considerado estimadores bajo un modelo parcialmente lineal generalizado isotónico y, en este sentido, la pregunta sobre cómo verificar la monotonía de la parte no paramétrica se vuelve un tema importante. En algunas situaciones, como las consideradas en la Sección 6.2, este supuesto está motivado por la estructura del problema. Sin embargo, en otros casos, es importante justificarlo a través de un procedimiento de test. La mayor parte de la bibliografía sobre tests de monotonía se enfoca en el modelo usual de regresión no paramétrica. Entre otros, Gijbels *et al.* (2000) consideran un procedimiento basado en signos de las diferencias entre la variable de respuesta que resulta relativamente robusto frente a distribuciones de los errores con colas pesadas. Por otro lado, Birke y Dette (2007) definen un test basado en la distancia L_2 y en la composición de un estimador de la inversa de una función de regresión monótona con una estimación sin restricción de la función de regresión. En el contexto general que hemos estudiado, un enfoque al problema de test podría ser considerar un procedimiento relacionado al definido en Birke y Dette (2007). Más precisamente, definamos $\Upsilon(g)(s) = \int_0^1 \mathbb{I}_{\{g(t) \leq s\}} dt$, para $s \in \mathbb{R}$. Hemos visto que si g es monótona entonces $\Upsilon(g)$ es igual a g^{-1} la inversa generalizada de g . Sea $\hat{\eta}_{\text{INI}}$ el estimador robusto por B–splines de η_0 calculado sin restricciones y definido en la Sección 2.3. Denotemos por $\hat{\Upsilon}^{(0)} = \Upsilon(\hat{\eta}_{\text{INI}})$. En particular, para el modelo log–Gamma $\hat{\eta}_{\text{INI}}$ es definida a través de los Pasos 1 a 3 descritos en la Sección 2.5. Entonces, un test de bondad de ajuste para testear monotonía puede basarse en el estadístico $\int_0^1 [\hat{\Upsilon}^{(0)}(\hat{\eta}_{\text{INI}}(t)) - t]^2 dt$. Un enfoque diferente puede aplicarse bajo un modelo parcialmente lineal con errores simétricos aplicando el procedimiento del signo definido en Gijbels *et al.* (2000) a los residuos parciales $y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ ordenados según los valores crecientes de las covariables t_i . Un análisis cuidadoso de las propiedades asintóticas y del comportamiento en muestras finitas de estos procedimientos va más allá del alcance de esta tesis. Dejamos estos importantes y desafiantes problemas teóricos como trabajo futuro.

Parte II

Modelo de regresión parcialmente lineal semi-funcional

Capítulo 8

Introducción

Los modelos parcialmente lineales pertenecen a la clase de modelos semiparamétricos. Contienen tanto componentes paramétricas como no paramétricas. Por un lado, el modelo parcialmente lineal se ocupa de la *maldición de la dimensión* asociada a los modelos completamente no paramétricos y facilita la interpretación del efecto de las covariables asociadas con la parte lineal. Por otro lado, son más flexibles que los modelos de regresión lineal habituales cuando se cree que algunas covariables se relacionan de manera no lineal con la variable independiente.

En el caso de covariables euclídeas, el modelo parcialmente lineal fue extensamente estudiado. Entre otros, podemos mencionar los trabajos de Ansley y Wecker (1983), Green *et al.* (1985), Denby (1986), Heckman (1986), Engle *et al.* (1986), Rice (1986), Chen (1988), Robinson (1988), Speckman (1988), Chen y Chen (1991), Chen y Shiau (1991 y 1994), Gao y Zhao (1993), Gao y Liang (1995), He y Shi (1996) y Yee y Wild (1996) quienes investigaron algunos resultados asintóticos usando técnicas de splines suavizantes, núcleos, o vecinos más cercanos. Una descripción extensa de los diferentes resultados obtenidos en modelos de regresión parcialmente lineales puede encontrarse en Härdle *et al.* (2000). Propuestas robustas para estos modelos fueron dadas, por ejemplo, en He *et al.* (2002) quienes consideraron M-estimadores basados en splines para observaciones repetidas y por Bianco y Boente (2004) quienes introdujeron un procedimiento de tres pasos basado en núcleos. Una estrategia diferente fue sugerida por Bhattacharya y Zhao (1997) quienes definieron un estimador consistente, con tasa \sqrt{n} , del coeficiente de regresión cuando las covariables relacionadas con la componente lineal pertenecen a un conjunto compacto.

El análisis de datos funcionales aumentó su desarrollo recientemente debido a su aplicabilidad en problemas donde es difícil una formulación en un contexto de observaciones escalares o vectoriales. Algunas áreas de aplicación son campos como biología, econometría, geofísica, ciencias médicas, meteorología y reconocimiento de patrones. En este contexto, los datos son de naturaleza continua,

corresponden a observaciones de curvas aleatorias, superficies o imágenes en vez de escalares o vectores. Aun cuando, en muchos casos, los datos se registran en una grilla densa de puntos, la alta dimensionalidad y la estructura de correlación implican que el uso directo de métodos de dimensión finita pueda fallar si tratamos a las observaciones como vectores aleatorios. Además, el marco dado por el análisis de datos funcionales permite sacar provecho de supuestos cualitativos como la suavidad de las curvas. Por estas razones, en las últimas décadas, se propusieron diferentes métodos para manejar este tipo de datos llamados *datos funcionales*. Una característica de estos datos es que pueden verse como realizaciones de elementos aleatorios que toman valores en un espacio de funciones como $L^2(\mathcal{T})$ o $C^r(\mathcal{T})$, con $\mathcal{T} \subset \mathbb{R}$ un intervalo finito. Un contexto usualmente considerado trata las observaciones como elementos de un espacio de Hilbert separable \mathcal{H} que no necesariamente tiene dimensión finita. Diversos métodos de inferencia para analizar datos funcionales pueden encontrarse en Ramsay y Silverman (2005), Ferraty y Vieu (2006) y Ferraty y Romain (2010). Podemos mencionar también los libros de Horváth y Kokoszka (2012) y Hsing y Eubank (2015), donde se describen diferentes procedimientos para datos funcionales y sus propiedades y los trabajos de Cuevas (2014) y Goia y Vieu (2016) quienes presentan un resumen de los avances recientes en el área de estimación en el caso de dimensión infinita.

Por ejemplo, consideremos el conjunto de datos TECATOR¹ que corresponde a un estudio de control de calidad en la industria alimenticia donde se estudiaron 215 muestras de carne. Cada muestra contiene finas láminas de carne con diferentes contenidos de grasa, proteínas y contenido de humedad. Para cada muestra, se observa una curva espectrométrica \mathcal{X} correspondiente a la absorbancia medida en una grilla equiespaciada de 100 longitudes de onda que varía de 850nm a 1050nm. Los contenidos de grasa, proteínas y de humedad se determinan por métodos analíticos. Este conjunto de datos fue estudiado, entre otros, por Ferraty y Vieu (2002) y Aneiros-Pérez y Vieu (2006), donde pueden verse más detalles. La Tabla 8.1 muestra las primeras observaciones de este conjunto.

Espectro				Proteína	Grasa	Agua
Col 1	Col 2	...	Col 100	Col 101	Col 102	Col 103
2.6177	2.6181	...	0.8996	60.5	22.5	16.7
2.8558	2.8606	...	0.6203	46	40.1	13.5
2.5828	2.5845	...	0.6905	71	8.4	20.5

Tabla 8.1: Primeras tres observaciones del conjunto de datos Tecator.

La Figura 8.1 muestra todas las observaciones correspondientes a la absorbancia. Dicha figura ilustra claramente que, dada la precisión de la grilla, cada observación parece una curva discreti-

¹Disponible en la librería de R `fd` o en el sitio <http://lib.stat.cmu.edu/datasets/tecator>.

zada que puede pensarse como la realización de una curva continua que es función de la longitud de onda. Como mencionan Leurgans, Moyeed y Silverman (1993): *the spectra observed are to all intents and purposes functional observations*². Otra característica de estos datos es que las curvas espectrométricas parecen similares excepto por un corrimiento vertical. Por esta razón, en el análisis de estos datos suelen tomarse las derivadas del espectro que se presentan en la Figura 8.2.

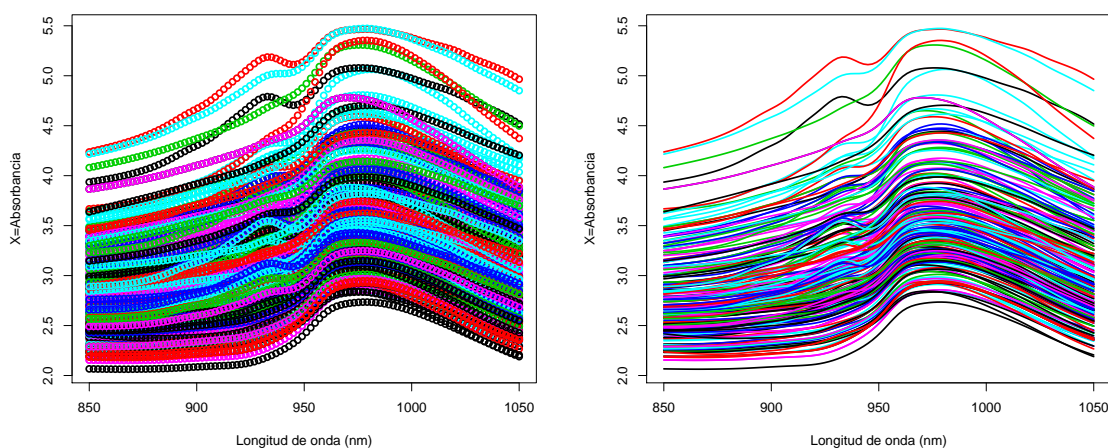
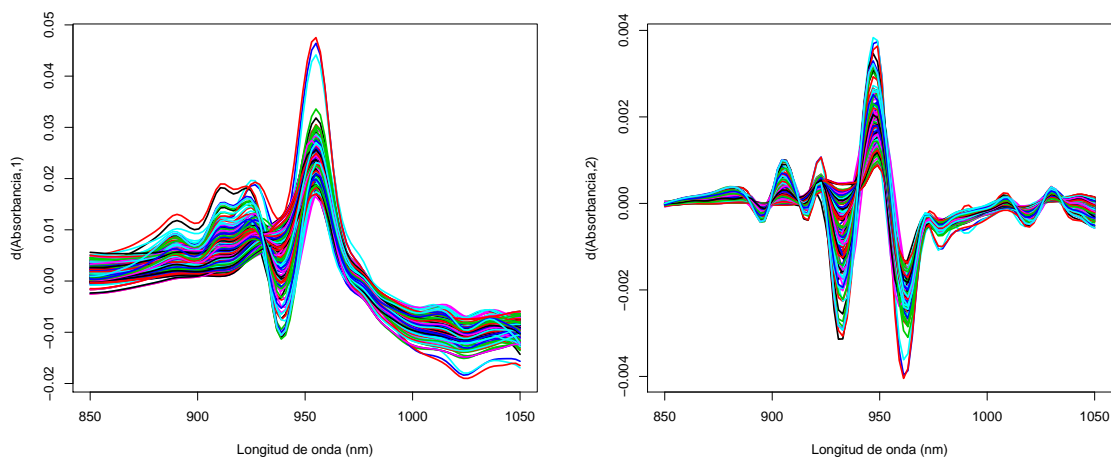


Figura 8.1: Datos TECATOR. (a) Puntos observados. (b) Curvas



(a) Primera derivada.

(b) Segunda derivada.

Figura 8.2: Conjunto de datos reales TECATOR: derivadas de la absorbancia \mathcal{X} .

Cabe mencionar que obtener la curva espectrométrica es más económico en tiempo y costo que el procedimiento analítico para determinar el contenido graso. Por esta razón, resulta interesante obtener un modelo que permita predecir el contenido graso de una muestra a partir de su contenido

²Los espectros observados son, a todos los efectos y fines, observaciones funcionales

de proteína y agua así como del espectro de su absorbancia. Entre los modelos considerados está el modelo parcialmente lineal funcional que describiremos a continuación.

En esta parte de la tesis, el objetivo es modelar la relación entre una respuesta univariada y y un conjunto de covariables (X, \mathbf{z}) , donde $X \in L^2(\mathcal{T})$ es una covariable funcional y $\mathbf{z} \in \mathbb{R}^q$. Por claridad de notación, de ahora en más, las variables aleatorias escalares se indicarán con letras minúsculas, mientras que los elementos aleatorios funcionales con mayúsculas. Dos enfoques fueron considerados en la literatura para extender el modelo parcialmente lineal del caso euclídeo al caso en que existen covariables funcionales. Para capturar las ventajas del modelado parcialmente lineal y el análisis no paramétrico de datos funcionales Aneiros–Pérez y Vieu (2006) introdujeron el modelo de regresión parcialmente lineal funcional que modela la relación entre la respuesta y el predictor funcional X de manera no paramétrica mientras que la variable aleatoria escalar \mathbf{z} se introduce en el modelo a través de una componente lineal. Estimadores robustos para los parámetros de este modelo fueron considerados por Boente y Vahnovan (2017). Un modelo diferente fue estudiado por Lian (2011) quien propuso una extensión del modelo parcialmente lineal llamada modelo lineal semifuncional SFLM en la que se combinan enfoques paramétricos y no paramétricos usando un modelo de regresión funcional para las covariables de dimensión infinita X y un modelo de regresión no paramétrica para las otras covariables que pueden pertenecer a un espacio semimétrico. Su enfoque combina una aproximación de las variables funcionales X usando componentes principales y estimadores de Nadaraya–Watson basados en núcleos. En el caso particular de $q = 1$, es decir, cuando z es una variable escalar, Zhou y Chen (2012) consideraron estimadores basados en splines, mientras que Qingguo (2015) y Huang *et al.* (2015) consideraron M–estimadores para dar dos enfoques más resistente a la presencia de respuestas atípicas.

Siguiendo la idea de Lian (2011), Qingguo (2015) proyecta la covariable funcional en la base de sus primeras autofunciones mientras que aproxima la componente no paramétrica usando B–splines. Por otra parte, Huang *et al.* (2015) usan B–splines para aproximar tanto el parámetro de regresión como la componente no paramétrica. Ambos autores, consideran M–estimadores basados en funciones de scores ψ monótonas una vez realizada la aproximación de dimensión finita. Cabe notar que los estimadores dados en Qingguo (2015) y Huang *et al.* (2015) no son equivariantes por cambios de escala ya que no incorporan ningún estimador preliminar de escala. Como ha sido ampliamente discutido en modelos de regresión en dimensión finita, el tamaño de los residuos que se considerarán atípicos depende de un buen estimador preliminar de escala, en este sentido la estimación de la escala es necesaria para calibrar los estimadores robustos y atenuar el peso de residuos grandes (ver Maronna *et al.*, 2019). Otra desventaja de las propuestas dadas en Qingguo (2015) y Huang *et al.* (2015) es el uso de una ρ -función convexa, es sabido que conduce a estimadores no robustos si existen datos atípicos en las covariables. Por otra parte, la presencia de datos atípicos

en las covariables funcionales podría afectar el procedimiento de estimación propuesto en Qingguo (2015) ya que este autor utiliza las componentes principales clásicas para aproximar a X y estimar el parámetro de regresión. Para superar estos problemas nuestra propuesta usa MM-estimadores tal como han sido definidos en Yohai (1987), esto es, la función de pérdida convexa es reemplazada por una función de pérdida acotada. Por otra parte, consideramos dos familias de estimadores: en la primera se utilizan B-splines para aproximar tanto el parámetro de regresión lineal funcional como la componente no paramétrica, mientras que en la segunda se usan componentes principales robustas para aproximar a X y B-splines para la componente no paramétrica.

La segunda parte de la tesis se organiza como sigue. El Capítulo 9 describe los estimadores robustos propuestos. En particular, como nuestro enfoque está basado en el desarrollo en dos bases, componentes esféricas o B-splines para la componente funcional y B-splines para la componente no paramétrica, describimos un método robusto para elegir la cantidad de nodos en la Sección 9.4. Los resultados de consistencia y de tasas de convergencia para los estimadores que utilizan en B-splines para ambas componentes, se detallan en el Capítulo 10. En el Capítulo 11 reportamos un estudio de simulación realizado con el objetivo de examinar el comportamiento de los procedimientos propuestos bajo distintos esquemas de contaminación, en muestras finitas. En el mismo capítulo, en la Sección 11.2 se presenta el análisis de los datos TECATOR. Finalmente, en el Capítulo 12 damos algunas conclusiones sobre esta segunda parte.

Capítulo 9

Estimadores robustos para modelos parcialmente lineales semifuncionales

9.1. Introducción

Sean y y z variables aleatorias a valores reales y $X \in L^2(\mathcal{T})$ un proceso estocástico, todos ellos definidos en el espacio de probabilidad $(\Omega, \mathcal{B}, \mathbb{P})$. Sea $\langle \cdot, \cdot \rangle$ y $\| \cdot \|$ la norma y producto interno de $L^2(\mathcal{T})$ respectivamente.

El modelo de regresión lineal semifuncional, estudiado en Lian (2011), Zhou y Chen (2012) y Qingguo (2015), entre otros, supone que observamos ternas (y_i, X_i, z_i) , $1 \leq i \leq n$, independientes e idénticamente distribuidas con la misma distribución que (y, X, z) , donde la respuesta y se relaciona con las covariables (X, z) a través del modelo

$$y = \langle X, \beta_0 \rangle + \eta_0(z) + \sigma_0 \epsilon, \quad (9.1)$$

donde $\beta_0 \in L^2(\mathcal{T})$, $\eta_0 : \mathcal{Z} \rightarrow \mathbb{R}$ es una función suave desconocida, \mathcal{T} y \mathcal{Z} son intervalos compactos y ϵ es un error aleatorio independiente de (X, z) . Notemos que para identificar η_0 no incluimos en el modelo una ordenada al origen. Sin pérdida de generalidad, suponemos que $\mathcal{T} = \mathcal{Z} = [0, 1]$.

Cabe mencionar que si, por información previa, sabemos que la función η_0 es monótona, por ejemplo, no decreciente, esta restricción debe tomarse en cuenta al estimar η_0 .

En el contexto clásico, se supone que $\mathbb{E}\epsilon = 0$ y $\mathbb{E}\epsilon^2 = 1$. Como es usual en robustez, evitamos suponer la existencia de momentos para los errores y supondremos que los errores ϵ tienen una distribución $G(\cdot)$ simétrica alrededor de 0, donde G tiene parámetro de escala 1 para identificar la escala de los errores. El parámetro σ_0 denota un parámetro de escala que no necesariamente coincide con la raíz cuadrada de la varianza ya que no suponemos la existencia del segundo momento de los

errores.

Por otra parte, como en el contexto clásico, se supone que el proceso X tiene segundo momento, es decir, que $\mathbb{E}\|X\|^2 < \infty$ y se indica por $\mu = \mathbb{E}X$. Más aun, se supone que el operador de covarianza Γ de X tiene rango infinito, es decir, que todos sus autovalores $\lambda_1 \geq \lambda_2 \geq \dots$ son positivos, para poder identificar la pendiente funcional β_0 . Efectivamente, supongamos por simplicidad que $\mathbb{E}\|X\|^2 < \infty$ y $\mathbb{E}(X) = 0$ y sea Γ el operador de covarianza de X . Llamemos $\lambda_1 \geq \lambda_2 \geq \dots$ a los autovalores de Γ y ϕ_j a la autofunción asociada a λ_j . Si Γ tuviera rango finito k , entonces para $j > k$ tendríamos que $\mathbb{P}(\langle X, \phi_j \rangle = 0) = 1$. En esta situación, β_0 no sería identificable porque $\beta_0 + \phi_j$ también verificaría el modelo (9.1).

9.2. Estimadores basados en B-splines

En esta sección, introduciremos estimadores robustos basados en B-splines para estimar η_0 y β_0 . Para ello, necesitaremos suponer que ambas funciones son suaves, en particular, supondremos que $\eta_0 \in C^r([0, 1])$ y $\beta_0 \in C^r([0, 1])$ donde $C^r([a, b])$ es el espacio de funciones definidas en el intervalo $[a, b]$ que son r veces continuamente diferenciables.

Para obtener una aproximación por splines para cada función, procedemos como en la Sección 2.2 de la primera parte. Tomamos una sucesión de nodos $\mathcal{Z}_n = \{\zeta_{i,n}\}_{i=1}^{m_n+2\ell}$ donde $0 = \zeta_{1,n} = \dots = \zeta_{\ell,n} < \zeta_{\ell+1,n} < \dots < \zeta_{m_n+\ell+1,n} = \dots = \zeta_{m_n+2\ell,n} = 1$ que particionan al intervalo cerrado $[0, 1]$ en $m_n + 1$ subintervalos $\mathcal{I}_i = [\zeta_{\ell+i,n}, \zeta_{\ell+i+1,n})$, para $i = 0, \dots, m_n - 1$ y $\mathcal{I}_{m_n} = [\zeta_{m_n+\ell,n}, \zeta_{m_n+\ell+1,n}]$. Llamamos $\mathcal{S}_n(\mathcal{Z}_n, \ell)$ a la clase de splines de orden $\ell \geq r + 2$ con nodos \mathcal{Z}_n . Según el Corolario 4.10 of Schumaker (1981), para cualquier $g \in \mathcal{S}_n(\mathcal{Z}_n, \ell)$, existe una base de funciones B-splines $\{B_j : 1 \leq j \leq k_n\}$, con $k_n = m_n + \ell$, tal que $g = \sum_{j=1}^{k_n} a_j B_j$.

Si fijamos dos valores $m_n^{(1)}$ y $m_n^{(2)}$ para el número de nodos, automáticamente quedan determinados los valores $k_{n,\beta} = m_n^{(1)} + \ell$ y $k_{n,\eta} = m_n^{(2)} + \ell$ para las dimensiones de las bases $\{B_j^{(1)} : 1 \leq j \leq k_{n,\beta}\}$ y $\{B_j^{(2)} : 1 \leq j \leq k_{n,\eta}\}$. Entonces, para funciones β_0 y η_0 suficientemente suaves, podemos usar las aproximaciones por B-splines $\beta_0(t) \approx \sum_{j=1}^{k_{n,\beta}} b_j B_j^{(1)}(t)$ y $\eta_0(z) \approx \sum_{j=1}^{k_{n,\eta}} a_j B_j^{(2)}(z)$. Por simplicidad en la notación escribiremos $p_1 = k_{n,\beta}$ y $p_2 = k_{n,\eta}$.

Cuando no se imponen restricciones de monotonía sobre la función verdadera η_0 , este enfoque permite definir MM-estimadores basados en B-splines como sigue.

Para cada $\mathbf{b} \in \mathbb{R}^{p_1}$ y $\mathbf{a} \in \mathbb{R}^{p_2}$, llamemos sean $\beta_{\mathbf{b}}(t) = \sum_{j=1}^{p_1} b_j B_j^{(1)}(t)$ y $\eta_{\mathbf{a}}(z) = \sum_{j=1}^{p_2} a_j B_j^{(2)}(z)$ y definamos $r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}) = y_i - \sum_{j=1}^{p_1} b_j x_{ij} - \sum_{j=1}^{p_2} a_j B_j^{(2)}(z_i)$ a los correspondientes residuos, donde $x_{ij} = \langle X_i, B_j^{(1)} \rangle$. Entonces, si $\mathbf{x}_i = (x_{i1}, \dots, x_{ip_1})^T$ y $\mathbf{B}_i = (B_1^{(2)}(z_i), \dots, B_{p_2}^{(2)}(z_i))^T$ escribimos de manera más compacta $r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}) = y_i - \mathbf{b}^T \mathbf{x}_i - \mathbf{a}^T \mathbf{B}_i$.

Sea $\rho : \mathbb{R} \rightarrow \mathbb{R}_+$ una ρ -función definida como en 2.1, es decir, como en Maronna *et al.* (2019): una función par, no decreciente en $|x|$, creciente para $x > 0$ cuando $\rho(x) < \lim_{t \rightarrow +\infty} \rho(t)$ y tal que $\rho(0) = 0$. Llamemos $\rho_c(u) = \rho(u/c)$ con $c > 0$ una constante de calibración. Una elección común es la función de Tukey definida como $\rho(x) = \min(3x^2 - 3x^4 + x^6, 1)$.

Los estimadores se definen a través del procedimiento usual en regresión robusta, Yohai (1987), dado en dos pasos: en el primero obtenemos un estimador inicial posiblemente ineficiente que permite calcular una escala robusta de los residuos, mientras que el segundo calculamos M-estimadores de regresión equivariantes con la escala obtenida anteriormente. A continuación detallamos ambos pasos.

- **Paso S.** Para todo $\mathbf{b} \in \mathbb{R}^{p_1}$ y $\mathbf{a} \in \mathbb{R}^{p_2}$, sea $s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})$ el M-estimador de escala de los residuos $r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})$, $1 \leq i \leq n$, es decir, $s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})$ verifica

$$\frac{1}{n - (p_1 + p_2)} \sum_{i=1}^n \rho_{c_1} \left(\frac{r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})}{s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})} \right) = b.$$

El término de corrección $n - (p_1 + p_2)$ se incluye para ajustar por el posible gran número de parámetros y dar estimadores con menor sesgo. Definamos

$$(\hat{\mathbf{b}}_{\text{INI}}, \hat{\mathbf{a}}_{\text{INI}}) = \underset{\mathbf{b}, \mathbf{a}}{\text{argmín}} s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}),$$

$\hat{\beta}_{\text{INI}} = \beta_{\hat{\mathbf{b}}_{\text{INI}}}(t) = \sum_{j=1}^{p_1} \hat{b}_{j,\text{INI}} B_j^{(1)}(t)$, $\hat{\eta}_{\text{INI}}(z) = \eta_{\hat{\mathbf{a}}_{\text{INI}}}(z) = \sum_{j=1}^{p_2} \hat{a}_{j,\text{INI}} B_j^{(2)}(z)$ y sea

$$\hat{\sigma} = s_n(\hat{\beta}_{\text{INI}}, \hat{\eta}_{\text{INI}}) = \underset{\mathbf{b}, \mathbf{a}}{\text{mín}} s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}). \quad (9.2)$$

Cuando ρ es la función de Tukey, la elección $c_1 = 1.54764$ y $b = 1/2$ da un estimador de escala Fisher-consistente para la distribución normal con punto de ruptura 50%.

- **Paso M.** Usando la función ρ_{c_2} con $c_2 > c_1$ y el estimador de escala $\hat{\sigma}$ del paso anterior calculamos un M-estimador, es decir, definimos

$$(\hat{b}_1, \dots, \hat{b}_{p_1}, \hat{a}_1, \dots, \hat{a}_{p_2}) = \underset{\mathbf{b}, \mathbf{a}}{\text{argmín}} \sum_{i=1}^n \rho_{c_2} \left(\frac{r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})}{\hat{\sigma}} \right). \quad (9.3)$$

Los estimadores finales de la función de regresión β_0 y de la componente no paramétrica η_0 se definen, respectivamente, como

$$\hat{\beta}(t) = \sum_{j=1}^{p_1} \hat{b}_j B_j^{(1)}(t) \quad (9.4)$$

$$\hat{\eta}(z) = \sum_{j=1}^{p_2} \hat{a}_j B_j^{(2)}(z). \quad (9.5)$$

Cuando hay una restricción de monotonía sobre la función η_0 resulta interesante definir estimadores isotónicos de la componente no paramétrica. Exploramos dos enfoques que detallamos en las siguientes secciones.

9.2.1. Estimadores basados en B-splines monótonos

Como en la Sección 2.2, para obtener estimadores que tengan en cuenta la restricción de que η_0 es no decreciente podemos utilizar B-splines monótonos. Sin embargo, los utilizaremos solamente en el **Paso M** del MM-estimador, una vez obtenido, en el **Paso S**, el estimador de escala sin restricciones.

Dado un estimador de escala $\hat{\sigma}$, los estimadores de β_0 y η_0 se obtienen como en (9.4) y (9.5) pero donde los coeficientes $(\hat{b}_1, \dots, \hat{b}_{p_1}, \hat{a}_1, \dots, \hat{a}_{p_2})$ se obtienen como la solución de

$$(\hat{b}_1, \dots, \hat{b}_{p_1}, \hat{a}_1, \dots, \hat{a}_{p_2}) = \underset{\mathbf{b}, a_1 \leq a_2 \leq \dots \leq a_{p_2}}{\operatorname{argmín}} \sum_{i=1}^n \rho_{c_2} \left(\frac{r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})}{\hat{\sigma}} \right).$$

El problema de minimización involucra ahora la restricción $a_1 \leq \dots \leq a_{p_2}$. Para resolverlo utilizamos un procedimiento iterativo de cuadrados mínimos pesados sujeto a $a_1 \leq \dots \leq a_{p_2}$, es decir, iterativamente, aproximamos el problema de minimización por un subproblema cuadrático como se describe en el Algoritmo 1. Vale la pena mencionar que el método iterativo de cuadrados mínimos pesados, indicado en la literatura como IRWLS es un algoritmo habitual para calcular M-estimadores de regresión y condiciones para su convergencia pueden verse en Maronna *et al.* (2019). Nuestro algoritmo solamente incorpora la condición de orden existente sobre los coeficientes a_j al procedimiento usual. Cabe mencionar que en cada paso del algoritmo se usa la misma S-escala $\hat{\sigma}$ computada sin restricciones. A los estimadores obtenidos con este procedimiento los denotaremos

$$\left(\hat{\beta}_{\text{MON}}(t), \hat{\eta}_{\text{MON}}(z) \right) = \left(\sum_{j=1}^{p_1} \hat{b}_j B_j^{(1)}(t), \sum_{j=1}^{p_2} \hat{a}_j B_j^{(2)}(z) \right). \quad (9.6)$$

9.2.2. Estimadores monotonizados basados en B-splines

Para dar una modificación monótona de $\hat{\eta}$ sin estimar nuevamente β , podemos utilizar el procedimiento de monotonización estudiado en Neumeyer (2007) que fue descrito en la Sección 2.3. De esta forma, aplicamos dicho procedimiento sobre los estimadores sin restricciones definidos a través de (9.4) y (9.5), sin necesidad de computar un nuevo estimador para β .

Sean $(\hat{\beta}, \hat{\eta})$ los estimadores de β_0 y η_0 sin restricciones, dados en (9.4) y (9.5). El estimador monótono para η_0 se define como

$$\hat{\eta}_i = \Upsilon \left(\Upsilon(\hat{\eta}) \mathbb{I}_{[\hat{\eta}(0), \hat{\eta}(1)]} \right) \mathbb{I}_{[0,1]}, \quad (9.7)$$

donde Υ está dada en (2.5). Como en la Sección 2.3, utilizando la función $\tilde{\Upsilon}$ dada en (2.5), podemos definir una versión discreta de dichos estimadores como $\hat{\eta}_{i,D} = \Upsilon \left(\tilde{\Upsilon}(\hat{\eta}) \mathbb{I}_{[\hat{\eta}(0), \hat{\eta}(1)]} \right) \mathbb{I}_{[0,1]}$. De esta manera, los estimadores resultantes serán

$$\left(\hat{\beta}(t), \hat{\eta}_{\text{MOD}}(z) \right) = \left(\hat{\beta}(t), \hat{\eta}_{i,D}(z) \right). \quad (9.8)$$

Algoritmo 1 Cálculo del M-estimador con restricciones de monotonía

1: Sea $m = 0$ y $(\mathbf{b}^{(0)}, \mathbf{a}^{(0)})$ la solución de (9.3), esto es, los MM-estimadores sin restricciones y $\hat{\sigma}$ la escala definida en el **Paso S**.

2: **repetir**

3: Dado m calculemos los pesos

$$w_{i,m} = \psi \left(\frac{r_i(\beta_{\mathbf{b}^{(m)}}, \eta_{\mathbf{a}^{(m)}})}{\hat{\sigma}} \right) \frac{\hat{\sigma}}{r_i(\beta_{\mathbf{b}^{(m)}}, \eta_{\mathbf{a}^{(m)}})}$$

con $\psi = \rho'_{c_2}$ y definamos

$$y_{w,i} = w_{i,m}^{1/2} y_i \quad , \quad B_{w,il}^{(1)} = w_{i,m}^{1/2} B_l^{(1)}(t_i) \quad , \quad B_{w,il}^{(2)} = w_{i,m}^{1/2} B_l^{(2)}(z_i)$$

4: Llamemos $\mathbf{v}_i = \left(B_{w,i1}^{(1)}, \dots, B_{w,ip_1}^{(1)}, B_{w,i1}^{(2)}, \dots, B_{w,ip_2}^{(2)} \right)^T$, $\mathbf{y}_w = (y_{w,1}, \dots, y_{w,n})^T$ y $\mathbf{d} = (\mathbf{b}^T, \mathbf{a}^T)^T$. Resolvemos el subproblema cuadrático aproximante con restricciones de monotonía

$$\hat{\mathbf{d}} = \min_{\mathbf{b}, a_1 \leq \dots \leq a_{p_2}} \|\mathbf{y}_w - \mathbf{V}^T \mathbf{d}\|^2 = \min_{\mathbf{b}, a_1 \leq \dots \leq a_{p_2}} \sum_{i=1}^n w_{i,m} \left(y_i - \sum_{j=1}^{p_1} b_j B_j^{(1)}(t_i) - \sum_{j=1}^{p_2} a_j B_j^{(2)}(z_i) \right)^2$$

5: Definimos $\mathbf{b}^{(m+1)}$ como las primeras p_1 componentes de $\hat{\mathbf{d}}$ y $\mathbf{a}^{(m+1)}$ como las últimas.

6: **hasta** convergencia.

9.3. Estimadores basados en componentes principales

Un enfoque diferente fue dado por Qingguo (2015) quien en lugar de usar una base de splines para aproximar el parámetro de regresión lineal funcional, considera la base de componentes principales del proceso X , es decir, las autofunciones de la función de covarianza $K(s, t) = \text{Cov}(X(s), X(t))$. Más precisamente, por el Teorema de Mercer existe un sistema ortonormal completo $\{\phi_k\}_{k \geq 1} \subset L^2(\mathcal{T})$ y una sucesión de números reales $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ tales que

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t) \text{ para } s, t \in \mathcal{T},$$

donde la serie converge absolutamente y uniformemente en $\mathcal{T} \times \mathcal{T}$. Más aun, el desarrollo de Karhunen–Loève del proceso X , permite obtener la siguiente representación

$$X = \mu + \sum_{k \geq 1} \xi_k \phi_k, \quad (9.9)$$

con $\xi_k = \langle X - \mu, \phi_k \rangle$ tales que $\mathbb{E}(\xi_k) = 0$, $\text{COV}(\xi_k, \xi_\ell) = \delta_{k\ell} \lambda_k$. Cabe notar que la convergencia en (9.9) es en media cuadrática y uniforme en \mathcal{T} .

Utilizando (9.9), el modelo (9.1) puede escribirse como

$$y_i = \langle X_i, \beta_0 \rangle + \eta_0(z_i) + \epsilon_i = \sum_{j \geq 1} \beta_{j,0} \xi_{ij} + \eta_1(z_i) + \epsilon_i, \quad (9.10)$$

donde $\beta_{j,0} = \langle \beta, \phi_j \rangle$, $\xi_{ij} = \langle X_i - \mu, \phi_j \rangle$ y $\eta_1(z) = \eta_0(z) + \langle \mu, \beta_0 \rangle$. Observemos que si η_0 es monótona, η_1 también lo es.

Por lo tanto, si estimáramos en forma robusta tanto la posición μ como las autofunciones $\phi_j(t)$ podríamos predecir X_i usando una cantidad finita de componentes, como $\widehat{X}_i = \widehat{\mu} + \sum_{j=1}^{p_1} \widehat{\xi}_{ij} \widehat{\phi}_j$. Cabe mencionar que este procedimiento de aproximación permite reconstruir las curvas usando sus *scores* $\widehat{\xi}_{ij}$ y una proyección en una base de dimensión p_1 pequeña, ya que usualmente un pequeño número de componentes principales da una buena aproximación del proceso. La función de regresión centrada en (9.10), $\langle X_i - \mu, \beta_0 \rangle$, queda aproximada por

$$\langle X_i - \mu, \beta_0 \rangle \approx \langle \widehat{X}_i - \widehat{\mu}, \beta \rangle = \sum_{j=1}^{p_1} b_{j,0} \widehat{\xi}_{ij}, \quad (9.11)$$

donde $b_{j,0} = \langle \beta_0, \widehat{\phi}_j \rangle$ y $\widehat{\xi}_{ij} = \langle X_i - \widehat{\mu}, \widehat{\phi}_j \rangle$.

Una vez obtenido el desarrollo del término funcional en las primeras p_1 componentes, utilizamos B-splines para aproximar a η_1 y definimos los residuos como

$$r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}) = y_i - \sum_{j=1}^{p_1} b_j \widehat{\xi}_{ij} - \sum_{j=1}^{p_2} a_j B_j(z_i),$$

donde, como en la Sección 9.2, $p_2 = k_{n,\eta}$ indica la dimensión de la base de splines $\{B_j : 1 \leq j \leq k_{n,\eta}\}$ utilizada para aproximar η_1 , es decir, tomamos $\mathbf{x}_i = (\hat{\xi}_{i1}, \dots, \hat{\xi}_{ip_1})^T$ en los **Pasos S** y **M** antes descritos. De esta forma, podemos calcular cualquiera de los estimadores propuestos en 9.2. En particular, una vez obtenido el estimador, $\hat{\eta}_1$, de η_1 el estimador de η_0 se define como $\hat{\eta}(z) = \hat{\eta}_1(z) - \langle \hat{\mu}, \hat{\beta} \rangle$, donde $\hat{\beta}$ es el estimador de β .

Un punto importante del método de estimación antes descrito es el cálculo de los estimadores robustos de μ y ϕ_j . En el ámbito robusto, varios funcionales de posición han sido considerados para el proceso X , entre otros la mediana geométrica o espacial definida como

$$\mu_{\text{GM}} = \underset{\theta \in L^2(\mathcal{T})}{\text{argmín}} \mathbb{E} (\|X - \theta\| - \|X\|), \quad (9.12)$$

es la elección usual para centrar los datos cuando se usa el operador de covarianza espacial. El estimador natural de la posición es en este caso, la versión empírica de (9.12), es decir, el valor $\hat{\mu}$ definido como

$$\hat{\mu} = \underset{\theta \in L^2(\mathcal{T})}{\text{argmín}} \frac{1}{n} \sum_{i=1}^n (\|X_i - \theta\| - \|X_i\|) \quad (9.13)$$

que se denomina la mediana espacial muestral. Gervini (2008) mostró que $\hat{\mu}$ puede obtenerse resolviendo un problema de minimización convexa n -dimensional. Más aun, $\hat{\mu}$ es fuertemente consistente respecto de la topología débil en $L^2(\mathcal{T})$, es decir, para todo $u \in L^2(\mathcal{T})$, $\langle \hat{\mu}, u \rangle \xrightarrow{a.s.} \langle \mu_{\text{GM}}, u \rangle$. Cardot *et al.* (2013) propusieron estimar la mediana espacial a través de un algoritmo que puede verse como un algoritmo de gradiente estocástico. El Teorema 3.1 de Cardot *et al.* (2013) muestra que, bajo ciertas condiciones de regularidad, este estimador converge en norma a la mediana geométrica en casi todo punto.

Las autofunciones ϕ_j son llamadas *direcciones principales* y se estiman usualmente como la j -ésima autofunción de la función de covarianza muestral. Sin embargo, estos estimadores son sensibles a la presencia de datos funcionales atípicos. Existen varias propuestas para estimar las direcciones principales en forma robusta, siendo la primera de ellas las componentes principales esféricas definidas por Locantore *et al.* (1999) y estudiadas en Gervini (2008). Otros métodos incluyen el procedimiento de projection-pursuit definido en Bali *et al.* (2011), el enfoque robusto dado en Sawant *et al.* (2012), los M-estimadores basados en splines propuestos por Lee *et al.* (2013) y los S -estimadores introducidos en Boente y Salibián Barrera (2015). En particular, las componentes principales esféricas son las autofunciones del operador de signo espacial muestral, que es simplemente el operador de covarianza muestral de las curvas centradas proyectadas sobre la esfera unitaria. Nuestra recomendación es utilizar las componentes principales esféricas, que definiremos a continuación, por su simplicidad computacional y por sus propiedades en el caso elíptico.

Definamos la función $\widehat{K}^S(s, t)$ y el operador de covarianza muestral esférico asociado, $\widehat{\Gamma}^S$, como

$$\widehat{K}^S(s, t) = \frac{1}{n} \sum_{i=1}^n \frac{(X_i(s) - \widehat{\mu}(s))(X_i(t) - \widehat{\mu}(t))}{\|X_i - \widehat{\mu}\|^2} \quad \widehat{\Gamma}^S = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \widehat{\mu}) \otimes (X_i - \widehat{\mu})}{\|X_i - \widehat{\mu}\|^2}.$$

Gervini (2008) prueba que $\widehat{\Gamma}^S$ es un estimador consistente del operador del signo espacial definido por

$$\Gamma^S = \mathbb{E} \left\{ \frac{(X - \mu_{\text{GM}}) \otimes (X - \mu_{\text{GM}})}{\|X - \mu_{\text{GM}}\|^2} \right\}$$

Sean $\widehat{\phi}_k$ y ϕ_k^S las autofunciones $\widehat{\Gamma}^S$ y Γ^S , respectivamente, ordenadas de manera decreciente según sus autovalores. Las autofunciones $\widehat{\phi}_k$ se denominan componentes principales esféricas muestrales y ϕ_k^S son las direcciones principales esféricas. Mediante $\widehat{\phi}_j$ podemos aproximar las observaciones X_i como $\widehat{X}_i = \widehat{\mu} + \sum_{j=1}^{p_1} \widehat{\xi}_{ij} \widehat{\phi}_j$ donde, como antes, $\widehat{\xi}_{ij} = \langle X_i - \widehat{\mu}, \widehat{\phi}_j \rangle$ son los *scores* predichos centrados.

Cuando se consideran las direcciones principales esféricas pueden darse dos situaciones: la distribución del proceso está concentrada en un espacio de dimensión finita o el rango de Γ^S es infinito. El Teorema 3 de Gervini (2008) establece que si el proceso es de dimensión finita, es decir, si $X = \mu + \sum_{k=1}^q \xi_k \phi_k$ y los *scores* estandarizados $(\xi_1/\sqrt{\lambda_1}, \dots, \xi_q/\sqrt{\lambda_q})$, con $\lambda_1 \geq \dots \geq \lambda_q > 0$, tienen distribución simétrica con marginales intercambiables, entonces las autofunciones de Γ^S son ϕ_j , $1 \leq j \leq q$ y en el mismo orden. Por otra parte, Boente *et al.* (2014) mostraron que las componentes principales esféricas son Fisher-consistentes para cualquier distribución elíptica. Precisamente, si X es un proceso elíptico, $\mathcal{E}(\mu, \Gamma)$, entonces, Γ^S tiene las mismas autofunciones que Γ y en el mismo orden. Observemos que en el contexto que nos interesa, es decir, bajo el modelo (9.1), el proceso X no puede tener una expansión finita ya que no permitiría identificar el parámetro β . Finalmente, Boente *et al.* (2019) obtuvieron la distribución asintótica de las direcciones principales esféricas muestrales. Estas dos propiedades no requieren la existencia de segundos momentos, lo que vuelve adecuado al procedimiento bajo la sospecha de curvas atípicas.

Observación 9.1. Supongamos que $\mathbb{E}\|X\|^2 < \infty$ y llamemos λ_k a los autovalores del operador de covarianza de X . Si indicamos por λ_k^S a los autovalores de Γ^S tenemos que λ_k^S se contrae respecto de λ_k como sigue

$$\lambda_k^S = \lambda_k \mathbb{E} \left(\frac{\xi_k^2}{\sum_{j \leq 1} \lambda_j \xi_j^2} \right). \quad (9.14)$$

Para evitar situaciones en las que los autovalores del operador del signo están muy cerca y no permitan identificar el orden de las autofunciones es mejor ordenar las autofunciones $\widehat{\phi}_k$ según los valores de una escala robusta de los datos proyectados, $\langle X_i - \widehat{\mu}, \widehat{\phi}_k \rangle$, $1 \leq i \leq n$. ♣

Un estudio detallado de las propiedades asintóticas y del comportamiento para muestras finitas va más allá de los lineamientos de esta tesis y serán objeto de estudio futuro.

9.4. Selección del modelo

Al igual que en la primera parte de la tesis, los estimadores propuestos dependen de parámetros de suavizado que son, o bien, el tamaño de las bases de B-splines utilizadas para aproximar el parámetro de regresión y la componente no paramétrica, o bien, la cantidad de componentes esféricas empleadas para estimar β_0 . Por esta razón, resulta importante determinar un criterio que seleccione, en algún sentido, el mejor modelo.

De ahora en más, indicaremos por p_1 la dimensión utilizada para aproximar a β_0 , o sea, la cantidad de parámetros correspondiente al término funcional, y por p_2 a la dimensión usada para estimar η_0 . Sean $\hat{\sigma}$, $\hat{\beta}$ y $\hat{\eta}$ los estimadores obtenidos cuando se utilizan espacios de dimensión p_1 y p_2 respectivamente y sean r_{i,p_1,p_2} los correspondientes residuos, es decir,

$$r_{i,p_1,p_2} = y_i - \langle X_i, \hat{\beta} \rangle - \hat{\eta}(z_i).$$

Qingguo (2015) define un criterio análogo al criterio de información de Schwarz (1978) como

$$BIC(p_1, p_2) = \log \left(\sum_{i=1}^n \rho(r_{i,p_1,p_2}) \right) + \frac{\log n}{2n} (p_1 + p_2).$$

Sin embargo, esta definición adolece del mismo problema que los estimadores propuestos en Qingguo (2015). La escala de los residuos no es tenida en cuenta. Para remediar el problema definimos una versión del mismo criterio que sí la incluye.

$$BIC_\alpha(p_1, p_2) = \log \left(\hat{\sigma}^2 \sum_{i=1}^n \rho \left(\frac{r_{i,p_1,p_2}}{\hat{\sigma}} \right) \right) + \alpha \frac{\log n}{n} (p_1 + p_2). \quad (9.15)$$

Observemos que si ρ fuera la función cuadrática se recupera el BIC clásico. Un criterio alternativo es el AIC propuesto por Akaike (1974)

$$AIC(p_1, p_2) = \log \left(\hat{\sigma}^2 \sum_{i=1}^n \rho \left(\frac{r_{i,p_1,p_2}}{\hat{\sigma}} \right) \right) + \frac{2}{n} (p_1 + p_2). \quad (9.16)$$

Ambos criterios pesan por igual la cantidad de parámetros correspondientes al término funcional y a la componente no paramétrica. Para separar ambos efectos también proponemos un criterio híbrido, que llamamos HIC, como un intermedio entre (9.15) y (9.16) y que puede ser más adecuado cuando se consideran los estimadores basados en componentes principales.

$$HIC(p_1, p_2) = \log \left(\hat{\sigma}^2 \sum_{i=1}^n \rho \left(\frac{r_{i,p_1,p_2}}{\hat{\sigma}} \right) \right) + \frac{2}{n} p_1 + \frac{\log(n)}{2n} p_2. \quad (9.17)$$

En forma similar a los criterios considerados en Ronchetti (1985) y Tharmaratnam y Claeskens (2013) para selección de modelos de regresión, podemos considerar también dos criterios similares

al *BIC* y *AIC* pero que tienen en cuenta en forma separada el efecto de la escala y de la función de pérdida ρ . Indicaremos por BIC^* y AIC^* a estos criterios que se definen como sigue

$$BIC_{\alpha}^*(p_1, p_2) = \log(\widehat{\sigma}^2) + \frac{1}{n} \sum_{i=1}^n \rho\left(\frac{r_{i,p_1,p_2}}{\widehat{\sigma}}\right) + \alpha \frac{\log n}{n} (p_1 + p_2) \quad (9.18)$$

$$AIC^*(p_1, p_2) = \log(\widehat{\sigma}^2) + \frac{1}{n} \sum_{i=1}^n \rho\left(\frac{r_{i,p_1,p_2}}{\widehat{\sigma}}\right) + \frac{2}{n} (p_1 + p_2). \quad (9.19)$$

En todos los casos, valores grandes indican un mal ajuste.

9.5. Algunos comentarios adicionales

Cabe mencionar que los MM-estimadores definidos en la Sección 9.2 y los resultados del Capítulo 10, pueden extenderse fácilmente al caso, en que se tienen observaciones (y_i, X_i, z_i, v_i) independientes e idénticamente distribuídas tales que

$$y_i = \gamma_0 + \langle X_i, \beta_0 \rangle + v_i \eta_0(z_i) + \sigma_0 \epsilon,$$

donde γ_0 juega el papel de la ordenada al origen. En particular, para obtener los MM-estimadores basta definir en el **Paso S** y el **Paso M**, los residuos $r_i(\gamma, \beta_{\mathbf{b}}, \eta_{\mathbf{a}})$ como $r_i(\gamma, \beta_{\mathbf{b}}, \eta_{\mathbf{a}}) = y_i - \gamma - \sum_{j=1}^{p_1} b_j x_{ij} - \sum_{j=1}^{p_2} a_j v_i B_j^{(2)}(z_i) = y_i - \gamma - \mathbf{b}^T \mathbf{x}_i - \mathbf{a}^T (v_i \mathbf{B}_i)$. La consistencia de estos estimadores puede obtenerse en forma análoga a la consistencia de los MM-estimadores para el modelo (9.1) con la hipótesis adicional $\mathbb{E}v^2 < \infty$.

Las propuestas y resultados también pueden extenderse de modo a incluir el modelo

$$y_i = \langle X_i, \beta_0 \rangle + \gamma_0 v_i + \eta_0(z_i) + \sigma_0 \epsilon,$$

que en este caso por razones de identificabilidad de η_0 no incluye ordenada al origen. Para ello bastará considerar en la definición de los estimadores $\mathbf{x}_i = (\langle X_i, B_1^{(1)} \rangle, \dots, \langle X_i, B_{p_1}^{(1)} \rangle, v_i)^T$ o $\mathbf{x}_i = (\widehat{\xi}_{i1}, \dots, \widehat{\xi}_{ip_1}, v_i)^T$ si se utilizan B-splines o direcciones principales, para aproximar la componente lineal. De esta forma se define $r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}) = y_i - \mathbf{b}^T \mathbf{x}_i - \mathbf{a}^T \mathbf{B}_i$, en los **Pasos S** y **M**.

Claramente, nuestras propuestas proveen procedimientos de estimación para el caso del modelo lineal funcional o, más generalmente, para el modelo $y_i = \langle X_i, \beta_0 \rangle + \gamma_0^T \mathbf{z}_i + \sigma_0 \epsilon$, tomando $\mathbf{x}_i = (\langle X_i, B_1^{(1)} \rangle, \dots, \langle X_i, B_{p_1}^{(1)} \rangle, \mathbf{z}_i)^T$. Estimadores robustos basados en splines para el caso en que $\gamma_0 = 0$ fueron dados por Maronna y Yohai (2013) y en este sentido, esta tesis complementa dicho trabajo obteniendo resultados de consistencia y tasas para dichos estimadores.

Capítulo 10

Consistencia de los estimadores basados en B-splines

En este capítulo, obtenemos resultados de consistencia y tasas de convergencia de los estimadores definidos en (9.4) y (9.5). Llamamos $\|\cdot\|_p$ a la norma euclídea de \mathbb{R}^p y por $\langle \cdot, \cdot \rangle$ y $\|\cdot\|$ al producto interno y a la norma de $L^2(0, 1)$, respectivamente. Para cualquier función continua $v : \mathbb{R} \rightarrow \mathbb{R}$ sea $\|v\|_\infty = \sup_t |v(t)|$. A lo largo del capítulo, \mathcal{V} es un entorno de σ_0 cuya clausura $\overline{\mathcal{V}}$ está estrictamente incluida en $(0, \infty)$. Como en la primera parte, para cualquier medida \mathbb{Q} , $N(\epsilon, \mathcal{F}, L_s(\mathbb{Q}))$ y $N_{[\cdot]}(\epsilon, \mathcal{F}, L_s(\mathbb{Q}))$ indican los números de cubrimiento y cubrimiento corchete de una clase de funciones \mathcal{F} con respecto a la distancia de $L_s(\mathbb{Q})$. Sus definiciones pueden consultarse en la Sección 3.1.

Consideraremos el espacio de funciones cuyas derivadas r -ésimas satisfacen una condición de Lipschitz en $[0, 1]$, con $r \geq 1$,

$$\mathcal{L}_r = \mathcal{L}_r([0, 1]) = \left\{ g \in C^r([0, 1]) : \|g^{(j)}\|_\infty < \infty, 0 \leq j \leq r, \text{ y } \sup_{z_1 \neq z_2} \frac{|g^{(r)}(z_1) - g^{(r)}(z_2)|}{|z_1 - z_2|} < \infty \right\},$$

con la norma

$$\|f\|_{\mathcal{L}_r} = \max_{1 \leq j \leq r} \|f^{(j)}\|_\infty + \sup_{x \neq y, x, y \in (0, 1)} \frac{|f^{(r)}(x) - f^{(r)}(y)|}{|x - y|},$$

donde $f^{(j)}$ es la j -ésima derivada de f . La bola unitaria en dicho espacio se indica por $\mathcal{V}_1^{(r)} = \{f \in \mathcal{L}_r([0, 1]) : \|f\|_{\mathcal{L}_r} \leq 1\}$.

Como mencionamos en la Sección 9.1, el parámetro β_0 no está identificado unívocamente si, por ejemplo, el operador de covarianza Γ de X tiene rango finito. Precisamente, como se menciona en Cardot *et al.* (2013), si $\mathcal{N}(\Gamma) = \{x \in L^2(\mathcal{T}), \Gamma x = 0\} \neq \{0\}$, entonces algunos autovalores de Γ son nulos. En la Sección 9.3, hemos mencionado que el desarrollo de Karhunen–Loève del proceso X , permite obtener la representación $X = \mu + \sum_{k \geq 1} \xi_k \phi_k$, donde $\xi_k = \langle X - \mu, \phi_k \rangle$

y $\{\phi_k\}_{k \geq 1} \subset L^2(\mathcal{T})$ son las autofunciones ortonormales de Γ . Por lo tanto, con probabilidad 1 $\langle X - \mu, \beta_0 \rangle = \langle X - \mu, \beta_0 + \alpha_0 \rangle$ para cualquier $\alpha_0 \in \mathcal{N}(\Gamma)$, es decir, no podemos garantizar unicidad para la representación dada en (9.1) y mucho menos para los funcionales asociados a los estimadores definidos en la Sección 9.2.1, salvo que nos restrinjamos a considerar como espacio paramétrico para β_0 el ortogonal del núcleo de Γ . Por esta razón, para obtener resultados de consistencia supondremos que $\mathcal{N}(\Gamma) = \{0\}$.

Las demostraciones de este capítulo pueden encontrarse en el apéndice del mismo.

10.1. Consistencia

Los supuestos que presentamos a continuación son necesarios para obtener los resultados de consistencia. Refieren a la distribución de los errores para asegurar la consistencia de Fisher, a la suavidad del parámetro de regresión y de la componente no paramétrica, así como las tasas a las que crecen las dimensiones de las bases. Estos últimos supuestos son usuales al tratar con aproximaciones por splines. La hipótesis **FC2** permiten asegurar que (β_0, η_0) es el único mínimo de $M(\beta, \eta, \sigma)$, requisito usual en modelos de regresión para obtener estimadores consistentes. Asimismo, **FC8**, que comentamos en la Observación 10.1, es la versión funcional del supuesto (A.3) en Yohai (1987) adaptado a la estructura parcialmente lineal. Para aliviar la notación escribimos ρ en lugar de ρ_{c_1} o ρ_{c_2} .

- FC1** La variable aleatoria ϵ tiene una función de densidad $g_0(t)$ que es par, no creciente en $|t|$, y estrictamente decreciente para $|t|$ en un entorno de 0.
- FC2** Para casi todo z_0 , $\mathbb{P}(\langle X, \beta \rangle = c | z = z_0) < 1$, para toda $\beta \in L^2(0, 1)$, $c \in \mathbb{R}$ y $(\beta, c) \neq 0$.
- FC3** La función $\rho : \mathbb{R} \rightarrow [0, \infty)$ es una ρ -función como en 2.1, es decir, es una función continua, par, no decreciente tal que $\rho(0) = 0$, $\lim_{u \rightarrow +\infty} \rho(u) \neq 0$ y, además, si $0 \leq u < v$ con $\rho(v) < \sup_u \rho(u)$ entonces $\rho(u) < \rho(v)$. Cuando ρ es acotada suponemos que $\sup_u \rho(u) = 1$.
- FC4** La función ρ es diferenciable con derivada acotada ψ tal que $\zeta(u) = u\psi(u)$ también es acotada.
- FC5** Las funciones verdaderas β_0 y η_0 son tales que $\beta_0 \in \mathcal{C}^r([0, 1])$ y $\eta_0 \in \mathcal{C}^r([0, 1])$. Además, sus derivadas r -ésimas satisfacen una condición de Lipschitz en $[0, 1]$, con $r \geq 1$, es decir, $\eta_0, \beta_0 \in \mathcal{L}_r([0, 1])$.
- FC6** Los estimadores $\hat{\sigma}$ de σ_0 son fuertemente consistentes.
- FC7** Los parámetros de suavizado $k_{n,\beta}$ y $k_{n,\eta}$ tienen orden $O(n^\nu)$ con $1/(2r+2) < \nu < 1/(2r)$. Más aun, el cociente entre el espaciado máximo y mínimo está acotado uniformemente.

FC8 Existe $0 < c < 1$ tal que $\mathbb{P}(\langle X, \beta \rangle + \eta(z) = 0) < c$, para cualquier $\beta \in L_1([0, 1])$, $\eta \in \mathcal{L}_1([0, 1])$, $(\beta, \eta) \neq 0$.

Observación 10.1. La condición **FC8** se cumple, por ejemplo, cuando $\mathbb{P}(\langle X, \beta \rangle + \eta(z) = 0) = 0$, para cualquier $\beta \in L^2(0, 1)$, $\eta \in \mathcal{L}_1([0, 1])$, $(\beta, \eta) \neq 0$. La hipótesis **FC8** implica, como condición necesaria, que el núcleo del operador de covarianza de X se reduzca a $\{0\}$. En particular, X no puede tener un desarrollo de Karhunen–Loève finito. Si así fuera, como mencionamos anteriormente, $\mathbb{P}(\langle X, \phi_j \rangle = 0) = 1$, para $j > k$, donde ϕ_j son las autofunciones de Γ asociadas a su j -ésimo autovalor λ_j , con $\lambda_1 \geq \lambda_2 \geq \dots$. En esta situación, β_0 no es identificable porque $\beta_0 + \phi_j$ también verifica el modelo (2.12). Denotamos Γ_{z_0} al operador de covarianza condicional $X|z = z_0$, esto es,

$$\Gamma_{z_0} = \mathbb{E} \{ [X - \mathbb{E}(X|z = z_0)] \otimes [X - \mathbb{E}(X|z = z_0)] | z = z_0 \} .$$

Luego, los supuestos **FC2** y **FC8** se cumplen cuando, para casi todo z_0 , el núcleo de Γ_{z_0} se reduce a $\{0\}$, $\mathcal{N}(\Gamma_{z_0}) = \{x \in L^2(0, 1) : \Gamma_{z_0}x = 0\} = \{0\}$, que es análogo al requisito en Huang et al. (2015). En efecto, supongamos que $\mathcal{N}(\Gamma_{z_0}) = \{0\}$ y llamemos $\mu_{z_0} = \mathbb{E}(X|z = z_0)$, entonces $\Gamma_{z_0} = \mathbb{E} \{ [X - \mu_{z_0}] \otimes [X - \mu_{z_0}] | z = z_0 \}$. Probemos que **FC2** se verifica. Notemos que $\langle \beta, \Gamma_{z_0}\beta \rangle = \mathbb{E}[\langle \beta, X - \mu_{z_0} \rangle]^2$ de manera que $\langle \beta, \Gamma_{z_0}\beta \rangle = 0$ si y solo si $\mathbb{P}(\langle \beta, X - \mu_{z_0} \rangle = 0) = 1$. Supongamos que **FC2** no se cumple, entonces existe $\beta \in L^2(0, 1)$ y $a \in \mathbb{R}$, $(\beta, a) \neq 0$ tales que $\mathbb{P}(\langle X, \beta \rangle = a|z = z_0) = 1$. Luego, en particular, tenemos que $a = \langle \mu_{z_0}, \beta \rangle$, de forma que $\mathbb{P}(\langle X - \mu_{z_0}, \beta \rangle = 0|z = z_0) = 1$ lo que implica que $\langle \beta, \Gamma_{z_0}\beta \rangle = 0$. Luego, usando que Γ_{z_0} es un operador lineal, compacto y autoadjunto con traza finita resulta que $\Gamma_{z_0}^{1/2}\beta = 0$ entonces $\beta \in \mathcal{N}(\Gamma_{z_0})$ llevando a una contradicción. Argumentos similares permiten mostrar que **FC8** se verifica. Más aun, si $\mathbb{P}(\langle X, \beta \rangle = a|z = z_0) = 0$, para cualquier $\beta \in L^2(0, 1)$, $a \in \mathbb{R}$ tal que $(\beta, a) \neq 0$, entonces $\mathcal{N}(\Gamma_{z_0}) = \{0\}$. Luego, nuestros supuestos **FC2** y **FC8** son más débiles que pedir $\mathcal{N}(\Gamma_{z_0}) = \{0\}$. ♣

El Lema 10.1 que enunciamos seguidamente asegura que el funcional asociado a los estimadores considerados cumple la consistencia de Fisher, condición necesaria para garantizar que estimamos las cantidades buscadas.

Lema 10.1 (Consistencia de Fisher). *Supongamos que se cumple **FC1** y que ρ es una función que verifica **FC3**. Sea*

$$M(\beta, \eta, \sigma) = \mathbb{E} \rho \left(\frac{y - \langle X, \beta \rangle - \eta(z)}{\sigma} \right) .$$

Entonces, para cualquier $\sigma > 0$,

a) $M(\beta, \eta, \sigma) \geq M(\beta_0, \eta_0, \sigma)$.

b) Si además se cumple **FC2**, (β_0, η_0) es el único minimizador de $M(\beta, \eta, \sigma)$.

Los siguientes resultados serán útiles para derivar la consistencia de los estimadores propuestos.

En adelante, consideramos la versión empírica $M_n(\beta, \eta, \sigma)$ del funcional $M(\beta, \eta, \sigma)$ definida por

$$M_n(\beta, \eta, \sigma) = \frac{1}{n} \sum_{i=1}^n \rho \left(\frac{y_i - \langle X_i, \beta \rangle - \eta(z_i)}{\sigma} \right) \quad (10.1)$$

y los espacios de splines de dimensión finita para $s = 1, 2$

$$\mathcal{M}_{p_s}^{(s)} = \left\{ \sum_{j=1}^{p_s} b_j B_j^{(s)}(t), \mathbf{b} \in \mathbb{R}^{p_s} \right\}$$

donde, como en la Sección 9.2, hemos indicado por simplicidad, $p_1 = k_{n,\beta}$ y $p_2 = k_{n,\eta}$.

Lema 10.2. *Sea ρ una función acotada que verifica **FC3** y **FC4** y supongamos que se cumple **FC7**.*

Entonces, si $p_1 = k_{n,\beta}$ y $p_2 = k_{n,\eta}$

a) $\sup_{\sigma > 0, \beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} |M_n(\beta, \eta, \sigma) - M(\beta, \eta, \sigma)| \xrightarrow{a.s.} 0$, cuando $n \rightarrow \infty$.

b) Más aun, $\sup_{\sigma > 0, \beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} \left| \frac{1}{n-p_1-p_2} \sum_{i=1}^n \left[\rho \left(\frac{y_i - \langle X_i, \beta \rangle - \eta(z_i)}{\sigma} \right) - M(\beta, \eta, \sigma) \right] \right| \xrightarrow{a.s.} 0$.

Para obtener la consistencia de los MM-estimadores definidos en (9.4) y (9.5), dada una función ρ y un estimador de escala consistente $\hat{\sigma}$, definimos los estimadores

$$\hat{\beta}(t) = \sum_{j=1}^{p_1} \hat{b}_j B_j^{(1)}(t) \quad (10.2)$$

$$\hat{\eta}(z) = \sum_{j=1}^{p_2} \hat{a}_j B_j^{(2)}(z). \quad (10.3)$$

donde $\hat{\mathbf{b}} = (\hat{b}_1, \dots, \hat{b}_{p_1})^T$ y $\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_{p_2})^T$ cumplen

$$(\hat{b}_1, \dots, \hat{b}_{p_1}, \hat{a}_1, \dots, \hat{a}_{p_2}) = \underset{\mathbf{b}, \mathbf{a}}{\operatorname{argmín}} \sum_{i=1}^n \rho \left(\frac{r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})}{\hat{\sigma}} \right).$$

Proposición 10.3. *Sea ρ una función acotada y supongamos que se cumplen los supuestos **FC1** a **FC7**. Si $(\hat{\beta}, \hat{\eta})$ son los estimadores definidos en (10.2) y (10.3) entonces $M(\hat{\beta}, \hat{\eta}, \sigma_0) \xrightarrow{a.s.} M(\beta_0, \eta_0, \sigma_0)$.*

Teorema 10.4. *Sea ρ una función acotada y supongamos que se cumplen **FC1** a **FC7**. Supongamos que $M(\beta_0, \eta_0, \sigma_0) = b_\rho < 1 = \|\rho\|_\infty$, $\mathbb{E}\|X\|^2 < \infty$ y que se cumple **FC8** con $c < 1 - b_\rho$. Entonces, si $(\hat{\beta}, \hat{\eta})$ son los estimadores definidos en (10.2) y (10.3) se cumple que $\|\hat{\beta} - \beta_0\|_\infty + \|\hat{\eta} - \eta_0\|_\infty \xrightarrow{a.s.} 0$.*

Observemos que para cualquier función ρ tal que $\rho \leq \rho_{c_1}$, se tiene $M(\beta_0, \eta_0, \sigma_0) \leq b < 1$. Por lo tanto, si $c < 1 - b$ también se tiene que $c < 1 - b_\rho$. y la condición $c < 1 - b$ fue también un requisito en Yohai (1987) para obtener consistencia de los MM-estimadores.

Si tuviéramos razones para suponer que la función η_0 en (9.1) es monótona, por ejemplo no decreciente, la modificación monótona de $\hat{\eta}$ descrita en la Sección 9.2.2 permite obtener estimadores

monótonos. El siguiente resultado es una consecuencia directa del Teorema 10.4 y del Teorema 3.2 en Neumeyer (2007). Por otra parte, argumentos análogos a los considerados en las demostraciones de la Proposición 10.3 y del Teorema 10.4 permiten mostrar que los estimadores $(\widehat{\beta}_{\text{MON}}, \widehat{\eta}_{\text{MON}})$ definidos en (9.6) también son consistentes.

Corolario 10.5. *Sea ρ una función acotada y supongamos que se cumplen de **FC1** a **FC7**. Además, supongamos que $M(\beta_0, \eta_0, \sigma_0) = b_\rho < 1 = \|\rho\|_\infty$, $\mathbb{E}\|X\|^2 < \infty$ y **FC8** se verifica con $c < 1 - b_\rho$. Sean $(\widehat{\beta}, \widehat{\eta})$ los estimadores definidos en (10.2) y (10.3) y $\widehat{\eta}_1$ la modificación monótona de $\widehat{\eta}$ dada en (9.7). Entonces se tiene que $\|\widehat{\eta}_1 - \eta_0\|_\infty \xrightarrow{a.s.} 0$.*

Resta dar una familia de estimadores de escala que verifique **FC6**. Los MM-estimadores de la Sección 9.2 usan estimadores de escala basados en una escala robusta. Precisamente, hemos definido

$$\widehat{\sigma} = s_n(\widehat{\beta}_{\text{INI}}, \widehat{\eta}_{\text{INI}}) = s_n(\beta_{\widehat{\mathbf{b}}_{\text{INI}}}, \eta_{\widehat{\mathbf{a}}_{\text{INI}}}) = \min_{\mathbf{b}, \mathbf{a}} s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}),$$

donde $s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})$ satisface

$$\frac{1}{n - (p_1 + p_2)} \sum_{i=1}^n \rho_{c_1} \left(\frac{r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})}{s_n(\beta_{\mathbf{b}}, \eta_{\mathbf{a}})} \right) = b.$$

Definamos el M-funcional $S(\beta, \eta)$ asociado. Para ello, denotemos $r(\beta, \eta) = y - \langle X, \beta \rangle - \eta(z)$ a los residuos. El funcional $S(\beta, \eta)$ satisface

$$\mathbb{E} \rho_{c_1} \left(\frac{r(\beta, \eta)}{S(\beta, \eta)} \right) = b.$$

Por simplicidad, de ahora en más, supondremos que los estimadores de escala están calibrados de modo tal que $\sigma_0 = S(\beta_0, \eta_0) = \text{argmín } S(\beta, \eta)$, o sea, $\mathbb{E} \rho_{c_1}(\epsilon) = b$.

El siguiente resultado muestra que bajo ciertas condiciones los estimadores de escala efectivamente satisfacen **FC6**.

Teorema 10.6 (Consistencia de la escala). *Supongamos que la función ρ_{c_1} es acotada y satisface **FC3** y **FC4**. Si $\mathbb{E}(\|X\|) < \infty$ y se cumplen **FC1**, **FC5** y **FC7**, entonces $\widehat{\sigma} \xrightarrow{a.s.} \sigma_0 = S(\beta_0, \eta_0)$.*

10.2. Tasas de convergencia

En esta sección obtendremos tasas de convergencia para los estimadores definidos en (10.2) y (10.3) respecto de la distancia $\pi^2(\theta_1, \theta_2) = \mathbb{E}[\langle X, \beta_1 - \beta_2 \rangle + \eta_1(z) - \eta_2(z)]^2$ donde, de ahora en más, $\theta = (\beta, \eta)$.

Llamemos $\Theta = \mathcal{L}_r([0, 1]) \times \mathcal{L}_r([0, 1])$ y $\Theta_n = \mathcal{M}_{p_1}^{(1)} \times \mathcal{M}_{p_2}^{(2)} \cap \{\theta = (\beta, \eta) \in \Theta : \|\beta - \beta_0\|_\infty + \|\eta - \eta_0\|_\infty \leq \epsilon_0\}$ para $\epsilon_0 > 0$, un valor que fijaremos más adelante. En base a los resultados

dados en el Teorema 10.4, salvo un conjunto de probabilidad nula, para n suficientemente grande, $\hat{\theta} = (\hat{\beta}, \hat{\eta}) \in \Theta_n$.

Para obtener tasas de convergencia necesitaremos los siguientes dos supuestos adicionales.

FC9 Existe $\epsilon_0 > 0$, un entorno \mathcal{V} de σ_0 y una constante positiva C_0 tal que para cualquier $\theta \in \Theta_n$ y cualquier $\sigma \in \mathcal{V}$

$$M(\theta, \sigma) - M(\theta_0, \sigma) \geq C_0 \pi^2(\theta, \theta_0).$$

FC10 La función ψ es continuamente diferenciable con derivada acotada ψ' y $\mathbb{E}\psi'(\epsilon) > 0$.

Teorema 10.7. *Sea ρ una función acotada y supongamos que se cumplen **FC1** a **FC7** y **FC9**. Supongamos que $M(\beta_0, \eta_0, \sigma_0) = b_\rho < 1 = \|\rho\|_\infty$, $\mathbb{E}\|X\|^2 < \infty$ y que **FC8** vale con $c < 1 - b_\rho$. Sean $(\hat{\beta}, \hat{\eta})$ son los estimadores definidos en (10.2) y (10.3). Entonces, tenemos que $\gamma_n \pi(\hat{\theta}, \hat{\theta}_0) = O_{\mathbb{P}}(1)$, donde $\hat{\theta} = (\hat{\beta}, \hat{\eta})$ y $\gamma_n = n^{\min(r\nu, (1-\nu)/2)}$. Por lo tanto, si $\nu = 1/(1+2r)$ los estimadores convergen a la tasa óptima $n^{r/(1+2r)}$.*

El siguiente lema da condiciones bajo las cuales se cumple el supuesto **FC9**.

Lema 10.8. *Sea ρ una función acotada tal que se cumplen **FC3**, **FC4** y **FC10**. Si para alguna constante $C > 0$ vale $\mathbb{P}(\|X\| \leq C) = 1$ entonces se cumple **FC9**.*

Observación 10.2. Observemos que si X y z son independientes y $\mathbb{E}(X) = 0$ se tiene que

$$\pi^2(\theta_1, \theta_2) = \mathbb{E}[\langle X, \beta_1 - \beta_2 \rangle + \eta_1(z) - \eta_2(z)]^2 = \mathbb{E}[\langle X, \beta_1 - \beta_2 \rangle]^2 + \mathbb{E}[\eta_1(z) - \eta_2(z)]^2.$$

Por lo tanto, del Teorema 10.7 deducimos que $\gamma_n^2 \mathbb{E}[\hat{\eta}(z) - \eta_0(z)]^2 = O_{\mathbb{P}}(1)$. Más aun, de la demostración de dicho resultado se deduce que existe $\tilde{\eta} \in \mathcal{M}_{p_2}^{(2)}$ tal que $\|\tilde{\eta} - \eta_0\|_\infty = O(n^{-r\nu})$ y $\gamma_n^2 \mathbb{E}[\hat{\eta}(z) - \tilde{\eta}(z)]^2 = O_{\mathbb{P}}(1)$. Usando que $\hat{\eta} - \tilde{\eta} \in \mathcal{M}_{p_2}^{(2)}$ y el Lema 7 de Stone (1986) podemos concluir que existe una constante $C > 0$ tal que $\|\hat{\eta} - \tilde{\eta}\|_\infty^2 \leq C p_2 \mathbb{E}[\hat{\eta}(z) - \tilde{\eta}(z)]^2$, lo que implica que $p_2^{-1/2} \gamma_n \|\hat{\eta} - \tilde{\eta}\|_\infty = O_{\mathbb{P}}(1)$. Supongamos que $\nu = 1/(1+2r)$ y $r \geq 1$. Teniendo en cuenta que $p_2 = O(n^\nu)$, podemos concluir que $n^{(r-1/2)/(1+2r)} \|\hat{\eta} - \tilde{\eta}\|_\infty = O_{\mathbb{P}}(1)$, lo que conduce a $n^{(r-1/2)/(1+2r)} \|\hat{\eta} - \eta_0\|_\infty = O_{\mathbb{P}}(1)$. Esta tasa es heredada por la modificación monótona de $\hat{\eta}$ cuando η es monótona. ♣

10.3. Apéndice A: Demostraciones

10.3.1. Demostración del Lema 10.1

La demostración es similar a la del Lema 2.1.

DEMOSTRACIÓN. El Lema 3.1 de Yohai (1987) junto con **FC3** y el hecho de que $\tilde{\epsilon} = \epsilon\sigma_0/\sigma$ satisfice el supuesto **FC1** implican que para todo $a \neq 0$,

$$\mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} - a \right) \right] > \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \right]. \quad (10.4)$$

lo que demuestra la parte a).

Para ver la parte b) llamemos $a(X, z) = \Phi(X, z)/\sigma$, con $\Phi(X, z) = \langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z)$ y $\mathcal{A}_0 = \{(X, z) : \Phi(X, z) = 0\}$. Condicionando a (X, z) y usando la independencia entre los errores y las covariables se obtiene que

$$\begin{aligned} M(\beta, \eta, \sigma) &= \mathbb{E} \rho \left(\frac{\epsilon\sigma_0 - \Phi(X, z)}{\sigma} \right) = \mathbb{E} \rho \left(\epsilon \frac{\sigma_0}{\sigma} - \frac{\Phi(X, z)}{\sigma} \right) \\ &= \mathbb{E} \left\{ \rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \mathbb{I}_{\mathcal{A}_0}(X, z) \right\} + \mathbb{E} \left\{ \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} - a(X, z) \right) \mid (X, z) \right] \mathbb{I}_{\mathcal{A}_0^c}(X, z) \right\} \\ &= \mathbb{E} \left(\rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \right) \mathbb{E} \{ \mathbb{I}_{\mathcal{A}_0}(X, z) \} + \mathbb{E} \left\{ \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} - a(X, z) \right) \mid (X, z) \right] \mathbb{I}_{\mathcal{A}_0^c}(X, z) \right\}. \end{aligned}$$

Para todo $(X, z) \notin \mathcal{A}_0$, usando 10.4,

$$\begin{aligned} \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} - a(X, z) \right) \mid (X, z) = (X_0, z_0) \right] &= \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} - a(X_0, z_0) \right) \mid (X, z) = (X_0, z_0) \right] \\ &= \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} - a(X_0, z_0) \right) \right] > \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \right] \end{aligned}$$

donde la última igualdad sigue de la independencia entre los errores y las covariables. Como **FC2** implica que $\mathbb{P}(\mathcal{A}_0^c) > 0$ obtenemos que

$$\begin{aligned} M(\beta, \eta, \sigma) &= \mathbb{E} \rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \mathbb{P}(\mathcal{A}_0) + \mathbb{E} \left\{ \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} - a(X, z) \right) \mid (X, z) \right] \mathbb{I}_{\mathcal{A}_0^c}(X, z) \right\} \\ &> \mathbb{E} \rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \mathbb{P}(\mathcal{A}_0) + \mathbb{E} \left\{ \mathbb{E} \left[\rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \right] \mathbb{I}_{\mathcal{A}_0^c}(X, z) \right\} = \mathbb{E} \rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \mathbb{P}(\mathcal{A}_0) + \mathbb{E} \rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \mathbb{P}(\mathcal{A}_0^c) \\ &> \mathbb{E} \left(\rho \left(\epsilon \frac{\sigma_0}{\sigma} \right) \right) = M(\beta_0, \eta_0, \sigma), \end{aligned}$$

lo que concluye la demostración. ■

10.3.2. Demostración del Lema 10.2

La parte b) se deduce de a) notando que $n/(n - p_1 - p_2) \rightarrow 1$. Para demostrar la parte a) recordemos que llamamos $\mathbf{B}(z) = \left(B_1^{(2)}(z), \dots, B_{p_2}^{(2)}(z) \right)^\top$ y $\mathbf{x} = \left(\langle X, B_1^{(1)} \rangle, \dots, \langle X, B_{p_1}^{(1)} \rangle \right)^\top$.

Definamos la clase de funciones

$$\mathcal{F}_n = \left\{ f(y, \mathbf{x}, z) = \rho \left(\frac{y - \mathbf{b}^\top \mathbf{x} - \mathbf{a}^\top \mathbf{B}(z)}{\sigma} \right), \mathbf{b} \in \mathbb{R}^{p_1}, \mathbf{a} \in \mathbb{R}^{p_2}, \sigma > 0 \right\}.$$

Dado que ρ es acotada y vale **FC3**, la clase de funciones \mathcal{F}_n tiene envolvente 1. Los argumentos utilizados en la demostración de la primera parte de Lema 3.13 permiten acotar, para cualquier

medida de probabilidad \mathbb{Q} , el número de cubrimiento $N(2\epsilon, \mathcal{F}_n, L_1(\mathbb{Q}))$ como

$$N(2\epsilon, \mathcal{F}_n, L_1(\mathbb{Q})) \leq \left[K q_n (16e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n-1} \right]^2, \quad (10.5)$$

donde $q_n = 2(p_1 + p_2 + 3) - 1$. Usando que $\log(q_n)/(p_1 + p_2 + 3) < 1$ y (10.5) y suponiendo, sin pérdida de generalidad, que $K > 1$, tenemos que

$$\begin{aligned} \log(N(2\epsilon, \mathcal{F}_n, L_1(\mathbb{Q}))) &\leq \log \left[K q_n (16e)^{q_n} \left(\frac{1}{\epsilon} \right)^{q_n-1} \right]^2 \\ &\leq 2 \left\{ \log(K) + \log q_n + q_n \log(16e) + (q_n - 1) \log \left(\frac{1}{\epsilon} \right) \right\} \\ &\leq 2 \left\{ q_n \left[\log(K) + 1 + \log(16e) + \log \left(\frac{1}{\epsilon} \right) \right] \right\} \\ &\leq C(p_1 + p_2) \log \left(\frac{1}{\epsilon} \right) \end{aligned}$$

para $\epsilon < \min((16e)^{-1}, e^{-K})$ y alguna constante $C > 0$. Entonces, como $p_1 = O(n^\nu)$ y $p_2 = O(n^\nu)$ con $\nu < 1$, deducimos que

$$\frac{1}{n} \log N(2\epsilon, \mathcal{F}_n, L_1(P_n)) \leq C \frac{p_1 + p_2}{n} \log \left(\frac{1}{\epsilon} \right) \rightarrow 0,$$

que conjuntamente con el Lema 3.2 implica

$$\sup_{\sigma > 0, \beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} |M_n(\beta, \eta, \sigma) - M(\beta, \eta, \sigma)| \xrightarrow{a.s.} 0,$$

concluyendo la demostración.

10.3.3. Demostración de la Proposición 10.3

Por **FC7**, $p_1 = k_{n,\beta}$ y $p_2 = k_{n,\eta}$ son de orden $O(n^\nu)$. Recordemos que denotamos por $r_i(\beta_{\mathbf{b}}, \eta_{\mathbf{a}}) = y_i - \mathbf{b}^T \mathbf{x}_i - \mathbf{a}^T \mathbf{B}_i$, donde $\mathbf{x}_i = (x_{i1}, \dots, x_{ip_1})^T$, $x_{ij} = \langle X_i, B_j^{(1)} \rangle$ y $\mathbf{B}_i = (B_1^{(2)}(z_i), \dots, B_{p_2}^{(2)}(z_i))^T$. El Lema 10.2 implica que

$$A_n = \sup_{\sigma > 0, \beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} |M_n(\beta, \eta, \sigma) - M(\beta, \eta, \sigma)| \xrightarrow{a.s.} 0, \quad (10.6)$$

donde M_n está definida en (10.1). Por otro lado, del Lema 10.1 se deduce que $M(\beta_0, \eta_0, \sigma) = \inf_{\beta, \eta} M(\beta, \eta, \sigma)$, para cualquier $\sigma > 0$. Luego,

$$0 \leq M(\hat{\beta}, \hat{\eta}, \sigma_0) - M(\beta_0, \eta_0, \sigma_0) = \sum_{i=1}^3 A_{n,i}$$

con $A_{n,1} = M(\hat{\beta}, \hat{\eta}, \hat{\sigma}) - M_n(\hat{\beta}, \hat{\eta}, \hat{\sigma})$, $A_{n,2} = M_n(\hat{\beta}, \hat{\eta}, \hat{\sigma}) - M(\beta_0, \eta_0, \sigma_0)$ y $A_{n,3} = M(\hat{\beta}, \hat{\eta}, \sigma_0) - M(\hat{\beta}, \hat{\eta}, \hat{\sigma})$. Notemos que $|A_{n,1}| \leq A_n$ que, junto con (10.6), implica que $A_{n,1} \xrightarrow{a.s.} 0$. Usando un desarrollo de Taylor de orden uno y el supuesto **FC4**, tenemos que

$$|A_{n,3}| \leq \|\zeta\|_\infty \frac{|\sigma_0 - \hat{\sigma}|}{\hat{\xi}},$$

donde $\widehat{\xi} = \theta\sigma_0 + (1 - \theta)\widehat{\sigma}$ es un punto intermedio. Por lo tanto, usando **FC6**, obtenemos que $A_{n,3} \xrightarrow{a.s.} 0$.

Acotemos $A_{n,2}$. Usando **FC5** y **FC7**, obtenemos de Schumaker (1981) que existen $\widetilde{\beta} \in \mathcal{M}_{p_1}^{(1)}$ y $\widetilde{\eta} \in \mathcal{M}_{p_2}^{(2)}$ tales que $\|\widetilde{\beta} - \beta_0\|_\infty = O(n^{-r\nu})$ y $\|\widetilde{\eta} - \eta_0\|_\infty = O(n^{-r\nu})$. Entonces, usando que $(\widehat{\beta}, \widehat{\eta})$ minimizan $M_n(\cdot, \cdot, \widehat{\sigma})$, podemos acotar $A_{n,2}$ por

$$A_{n,2} \leq M_n(\widetilde{\beta}, \widetilde{\eta}, \widehat{\sigma}) - M(\beta_0, \eta_0, \sigma_0) = \sum_{j=1}^3 C_{n,j},$$

donde $C_{n,1} = M_n(\widetilde{\beta}, \widetilde{\eta}, \widehat{\sigma}) - M(\widetilde{\beta}, \widetilde{\eta}, \widehat{\sigma})$, $C_{n,2} = M(\widetilde{\beta}, \widetilde{\eta}, \sigma_0) - M(\beta_0, \eta_0, \sigma_0)$ y $C_{n,3} = M(\widetilde{\beta}, \widetilde{\eta}, \widehat{\sigma}) - M(\widetilde{\beta}, \widetilde{\eta}, \sigma_0)$. La consistencia fuerte de $\widehat{\sigma}$ y el hecho que $\widetilde{\beta} \in \mathcal{M}_{p_1}^{(1)}$ y $\widetilde{\eta} \in \mathcal{M}_{p_2}^{(2)}$ implican que $|C_{n,1}|$ puede acotarse por A_n , por lo tanto, $|C_{n,1}| \xrightarrow{a.s.} 0$. Con un argumento similar al utilizado para acotar $A_{n,3}$, tenemos que $C_{n,3} \xrightarrow{a.s.} 0$. Finalmente, observemos que $\|\widetilde{\beta} - \beta_0\|_\infty + \|\widetilde{\eta} - \eta_0\|_\infty \rightarrow 0$ implica que, para todo (y, X, z) ,

$$y - \langle X, \widetilde{\beta} \rangle + \widetilde{\eta}(z) \rightarrow y - \langle X, \beta_0 \rangle + \eta_0(z).$$

Por lo tanto, el hecho que ρ sea continua y acotada junto con el Teorema de Convergencia Dominada permiten concluir que $C_{n,2} \rightarrow 0$. Luego,

$$0 \leq M(\widehat{\beta}, \widehat{\eta}, \sigma_0) - M(\beta_0, \eta_0, \sigma_0) \xrightarrow{a.s.} 0,$$

o sea, $M(\widehat{\beta}, \widehat{\eta}, \sigma_0) \xrightarrow{a.s.} M(\beta_0, \eta_0, \sigma_0)$, lo que concluye la demostración.

10.3.4. Demostración del Teorema 10.4

Para probar el Teorema 10.4 necesitaremos el siguiente resultado preliminar.

Lema 10.9. *Sea ρ una función acotada tal que $M(\beta_0, \eta_0, \sigma_0) = b_\rho < 1 = \|\rho\|_\infty$. Sea $(\widehat{\beta}_n, \widehat{\eta}_n) \in \mathcal{M}_{p_1}^{(1)} \times \mathcal{M}_{p_2}^{(2)}$ una sucesión aleatoria tal que $M(\widehat{\beta}_n, \widehat{\eta}_n, \sigma_0) \xrightarrow{a.s.} M(\beta_0, \eta_0, \sigma_0)$, donde hemos hecho explícita la dependencia en n por claridad. Supongamos que $\mathbb{E}\|X\|^2 < \infty$ y que se cumplen **FC3** a **FC5** y **FC8** con $c < 1 - b_\rho$. Entonces, existe L tal que*

$$\mathbb{P}\left(\bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \|\widehat{\beta}_n - \beta_0\|_{\mathcal{L}_1} + \|\widehat{\eta}_n - \eta_0\|_{\mathcal{L}_1} \leq L\right) = 1.$$

Demostración. Dado $\delta > 0$, definamos K_δ tal que para todo $K \geq K_\delta$,

$$\mathbb{P}(\|X\| \geq K) < \delta. \tag{10.7}$$

Recordemos que $\mathcal{V}_1^{(1)}$ es un conjunto compacto para la topología de la norma $\|\cdot\|_\infty$, esto es, visto en $\mathcal{C}([0, 1])$. Fijemos $\beta \in \mathcal{V}_1^{(1)}$, $\eta \in \mathcal{V}_1^{(1)}$. Sea $\theta = (\beta, \eta)$ y consideremos $\phi_\theta > 0$ un punto de continuidad de la distribución de $|\langle X, \beta \rangle + \eta(z)|$ tal que

$$\mathbb{P}(|\langle X, \beta \rangle + \eta(z)| < \phi_\theta) < c.$$

Si $\beta^*, \eta^* \in \mathcal{C}([0, 1])$ es tal que $\max(\|\beta^* - \beta\|_\infty, \|\eta^* - \eta\|_\infty) < \vartheta_\theta$, donde $\vartheta_\theta = \phi_\theta/(2(K+1))$, tenemos que

$$\mathbb{P}\left(|\langle X, \beta^* \rangle + \eta^*(z)| \geq \frac{\phi_\theta}{2}\right) \geq \mathbb{P}(|\langle X, \beta \rangle + \eta(z)| \geq \phi_\theta) - \mathbb{P}\left(\vartheta_\theta(\|X\| + 1) \geq \frac{\phi_\theta}{2}\right) = A(\theta).$$

Observando que $A(\theta) > 1 - c - \delta$ obtenemos

$$\inf_{\max(\|\beta^* - \beta\|_\infty, \|\eta^* - \eta\|_\infty) < \vartheta_\theta} \mathbb{P}\left(|\langle X, \beta^* \rangle + \eta^*(z)| \geq \frac{\phi_{\beta, \eta}}{2}\right) \geq A(\theta) > 1 - c - \delta. \quad (10.8)$$

Sea $\mathcal{B}(\theta, \rho)$ representa la bola abierta con centro en θ y radio ρ , $\mathcal{B}(\theta, \rho) = \{(f, g) \in \mathcal{C}([0, 1]) \times \mathcal{C}([0, 1]) : \max(\|f - \beta\|_\infty, \|g - \eta\|_\infty) < \rho\}$. Consideremos el cubrimiento $\cup_{\theta \in \mathcal{V}_1^{(1)} \times \mathcal{V}_1^{(1)}} \mathcal{B}(\theta, \vartheta_\theta)$ de $\mathcal{V}_1^{(1)} \times \mathcal{V}_1^{(1)}$. Como $\mathcal{V}_1^{(1)} \times \mathcal{V}_1^{(1)}$ es un conjunto compacto en $\mathcal{C}([0, 1]) \times \mathcal{C}([0, 1])$, existen $\theta_j = (\beta_j, \eta_j) \in \mathcal{V}_1^{(1)} \times \mathcal{V}_1^{(1)}$, $1 \leq j \leq s$, tales que $\mathcal{V}_1^{(1)} \times \mathcal{V}_1^{(1)} \subset \cup_{j=1}^s \mathcal{B}(\theta_j, \vartheta_j)$ con $\vartheta_j = \vartheta_{\theta_j}$. A partir de (10.8), obtenemos que

$$\min_{1 \leq j \leq s} \inf_{\max(\|\beta - \beta_j\|_\infty, \|\eta - \eta_j\|_\infty) < \vartheta_j} \mathbb{P}\left(|\langle X, \beta \rangle + \eta(z)| > \frac{\phi_j}{2}\right) > 1 - c - \delta.$$

donde $\phi_j = \phi_{\theta_j}$, lo que significa que para cualquier $(\beta, \eta) \in \mathcal{V}_1^{(1)} \times \mathcal{V}_1^{(1)}$ existe $1 \leq j \leq s$ tal que

$$\mathbb{P}\left(|\langle X, \beta \rangle + \eta(z)| > \frac{\phi_j}{2}\right) > 1 - c - \delta. \quad (10.9)$$

Sea \mathcal{N} tal que $\mathbb{P}(\mathcal{N}) = 0$ y para cada $\omega \notin \mathcal{N}$, $M(\widehat{\beta}_n, \widehat{\eta}_n, \sigma_0) \rightarrow M(\beta_0, \eta_0, \sigma_0) = b_\rho$. Fijemos $\omega \notin \mathcal{N}$ y sea $\xi > 0$ tal que $b + \xi < 1 - c$. Luego, existe $n_0 \in \mathbb{N}$ tal que para cada $n \geq n_0$, $M(\widehat{\beta}_n, \widehat{\eta}_n, \sigma_0) \leq b_\rho + \xi/2$.

Para ver que existe $L > 0$ tal que, para $\omega \notin \mathcal{N}$, $\limsup_{n \rightarrow \infty} \|\widehat{\beta}_n - \beta_0\|_{\mathcal{L}_1} + \|\widehat{\eta}_n - \eta_0\|_{\mathcal{L}_1} \leq L$, basta mostrar que existe $L > 0$ tal que

$$\inf_{\|\beta - \beta_0\|_{\mathcal{L}_1} + \|\eta - \eta_0\|_{\mathcal{L}_1} > L} M(\beta, \eta, \sigma_0) \geq b_\rho + \xi.$$

Llamemos $R(u) = \mathbb{E}\rho(\epsilon - u/\sigma_0)$. La independencia entre los errores y las covariables implica que

$$M(\beta, \eta, \sigma_0) = \mathbb{E}\rho\left(\epsilon - \frac{\langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z)}{\sigma_0}\right) = \mathbb{E}R(\langle X, \beta - \beta_0 \rangle + (\eta - \eta_0)(z)).$$

Usando que $\lim_{|u| \rightarrow +\infty} R(u) = 1$, obtenemos que para cualquier $\delta > 0$ existe u_0 tal que para todo u con $|u| \geq u_0$,

$$R(u) > 1 - \delta. \quad (10.10)$$

Elijamos $L > 2u_0/\min_{1 \leq j \leq s}(\phi_j)$ donde ϕ_j está dada en (10.9) y sean $(\beta_k, \eta_k) \in \mathcal{L}_1([0, 1]) \times \mathcal{L}_1([0, 1])$, $k \in \mathbb{N}$, tales que $\nu_k = \|\beta_k - \beta_0\|_{\mathcal{L}_1} + \|\eta_k - \eta_0\|_{\mathcal{L}_1} > L$ y $M(\beta_k, \eta_k, \sigma_0) \rightarrow \inf_{\|\beta - \beta_0\|_{\mathcal{L}_1} + \|\eta - \eta_0\|_{\mathcal{L}_1} > L} M(\beta, \eta, \sigma_0)$. Llamemos $\widetilde{\beta}_k = (\beta_k - \beta_0)/\nu_k$ y $\widetilde{\eta}_k = (\eta_k - \eta_0)/\nu_k$, entonces $(\widetilde{\beta}_k, \widetilde{\eta}_k) \in \mathcal{V}_1^{(1)} \times \mathcal{V}_1^{(1)}$, usando (10.9), obtenemos que existe $1 \leq j = j(k) \leq s$ tal que

$$\mathbb{P}\left(|\langle X, \widetilde{\beta}_k \rangle + \widetilde{\eta}_k(z)| > \frac{\phi_j}{2}\right) > 1 - c - \delta. \quad (10.11)$$

Usando que $\nu_k > L > 2u_0/\phi_j$ y denotando $u_k(X, z) = \nu_k(\langle X, \tilde{\beta}_k \rangle + \tilde{\eta}_k(z))$ obtenemos que $|u_k(X, z)| > u_0$ siempre que $|\langle X, \tilde{\beta}_k \rangle + \tilde{\eta}_k(z)| > \phi_j/2$, que junto con (10.10) permite deducir que

$$\begin{aligned} M(\beta_k, \eta_k, \sigma_0) &= \mathbb{E}R(\langle X, \beta_k \rangle + \eta_k(z)) = \mathbb{E}R(u_k(X, z)) \geq \mathbb{E}\left\{R(u_k(X, z)) \mathbb{I}_{|\langle X, \tilde{\beta}_k \rangle + \tilde{\eta}_k(z)| > \phi_j/2}\right\} \\ &> (1 - \delta) \mathbb{P}\left(\left|\langle X, \tilde{\beta}_k \rangle + \tilde{\eta}_k(z)\right| > \frac{\phi_j}{2}\right) \\ &> (1 - c - \delta)(1 - \delta), \end{aligned}$$

donde la última desigualdad sigue de (10.11). Luego,

$$\inf_{\|\beta - \beta_0\|_{\mathcal{L}_1} + \|\eta - \eta_0\|_{\mathcal{L}_1} > L} M(\beta, \eta, \sigma_0) \geq (1 - c - \delta)(1 - \delta).$$

Como $\lim_{\delta \rightarrow 0} (1 - c - \delta)(1 - \delta) = 1 - c > b + \xi$, podemos elegir δ y en consecuencia L tal que

$$\inf_{\|\beta - \beta_0\|_{\mathcal{L}_1} + \|\eta - \eta_0\|_{\mathcal{L}_1} > L} M(\beta, \eta, \sigma_0) > b_\rho + \xi > M(\hat{\beta}_n, \hat{\eta}_n, \sigma_0),$$

de donde se deduce que $\|\hat{\beta}_n - \beta_0\|_{\mathcal{L}_1} + \|\hat{\eta}_n - \eta_0\|_{\mathcal{L}_1} \leq L$, lo que concluye la demostración. \blacksquare

DEMOSTRACIÓN DEL TEOREMA 10.4. Por simplicidad sean $\theta = (\beta, \eta)$ y $\theta_0 = (\beta_0, \eta_0)$. A partir de la Proposición 10.3 y del Lema 10.9 basta mostrar que para cualquier $\epsilon > 0$,

$$\inf_{(\beta, \eta) \in \mathcal{A}_\epsilon} M(\beta, \eta, \sigma_0) > M(\beta_0, \eta_0, \sigma_0),$$

donde $\mathcal{A}_\epsilon = \{(\beta, \eta) \in \mathcal{L}_1([0, 1]) \times \mathcal{L}_1([0, 1]) \mid \|\beta - \beta_0\|_{\mathcal{L}_1} + \|\eta - \eta_0\|_{\mathcal{L}_1} \leq L, d(\theta, \theta_0) \geq \epsilon\}$ y $d(\theta, \theta_0) = \|\beta - \beta_0\|_\infty + \|\eta - \eta_0\|_\infty$.

Al igual que en la demostración del Lema 10.9, sean $(\beta_k, \eta_k) \in \mathcal{A}_\epsilon$, $k \in \mathbb{N}$, tales que $M_k = M(\beta_k, \eta_k, \sigma_0) \rightarrow \inf_{(\beta, \eta) \in \mathcal{A}_\epsilon} M(\beta, \eta, \sigma_0)$. Denotemos por $\nu_k = \|\beta_k - \beta_0\|_{\mathcal{L}_1} + \|\eta_k - \eta_0\|_{\mathcal{L}_1}$. Como ν_k está acotada, tenemos que $\|\beta_k - \beta_0\|_\infty$ es acotada, más aun, $\beta_k - \beta_0$ es equicontinua pues su primer derivada es Lipschitz con constante de Lipschitz uniformemente acotada. Por lo tanto, usando el Teorema de Arzelá-Ascoli, tenemos que existe una subsucesión $\{k_j^{(1)}\}$ tal que $\beta_{k_j^{(1)}} - \beta_0$ converge uniformemente a una función continua f . En forma análoga, obtenemos que existe una subsucesión de $\{k_j^{(1)}\}$ que indicaremos $\{k_j\}$ tal que $g_j = \eta_{k_j} - \eta_0$ converge uniformemente a una función continua g . Sean $\tilde{\beta} = f + \beta_0$ y $\tilde{\eta} = g + \eta_0$ los límites uniformes de β_{k_j} y η_{k_j} respectivamente. Si indicamos por $f_j = \beta_{k_j} - \beta_0$ tenemos que $\|\beta_{k_j} - \tilde{\beta}\|_\infty + \|\eta_{k_j} - \tilde{\eta}\|_\infty \rightarrow 0$, $\lim_j \|f_j\|_\infty = \|f\|_\infty$ y $\lim_j \|g_j\|_\infty = \|g\|_\infty$. Como $(\beta_k, \eta_k) \in \mathcal{A}_\epsilon$, tenemos que $\|f_j\|_\infty + \|g_j\|_\infty \geq \epsilon$, de manera que $d(\tilde{\theta}, \theta_0) \geq \epsilon$ con $\tilde{\theta} = (\tilde{\beta}, \tilde{\eta})$. Como ρ es una función acotada continua, el Teorema de Convergencia Mayorada permite concluir que $M_{k_j} \rightarrow M(\tilde{\beta}, \tilde{\eta}, \sigma_0)$, lo que implica que $\inf_{(\beta, \eta) \in \mathcal{A}_\epsilon} M(\beta, \eta, \sigma_0) = M(\tilde{\beta}, \tilde{\eta}, \sigma_0)$. El Lema 10.1 junto con el hecho que $d(\tilde{\theta}, \theta_0) \geq \epsilon$ implican que $M(\tilde{\beta}, \tilde{\eta}, \sigma_0) > M(\beta_0, \eta_0, \sigma_0)$ lo que concluye la demostración. \blacksquare

10.3.5. Demostración del Teorema 10.6

Recordemos que definimos $r(\beta, \eta) = y - \langle X, \beta \rangle - \eta(t)$ y que supusimos que $\sigma_0 = S(\beta_0, \eta_0)$ donde $S(\beta_0, \eta_0)$ es la solución de

$$\mathbb{E}\rho_{c_1} \left(\frac{r(\beta_0, \eta_0)}{S(\beta_0, \eta_0)} \right) = b,$$

o sea, que $\mathbb{E}\rho_{c_1}(\epsilon) = b$. Además, los estimadores de escala $\hat{\sigma} = s_n(\hat{\beta}_{\text{INI}}, \hat{\eta}_{\text{INI}})$ satisfacen

$$\frac{1}{n - (p_1 + p_2)} \sum_{i=1}^n \rho_{c_1} \left(\frac{y_i - \langle X_i, \hat{\beta}_{\text{INI}} \rangle - \hat{\eta}_{\text{INI}}(t_i)}{s_n(\hat{\beta}_{\text{INI}}, \hat{\eta}_{\text{INI}})} \right) = b.$$

Para aliviar la notación escribimos ρ en vez de ρ_{c_1} y $\hat{\beta}$ y $\hat{\eta}$ en lugar de $\hat{\beta}_{\text{INI}}$ y $\hat{\eta}_{\text{INI}}$.

Mostraremos que para cualquier $\delta > 0$, con probabilidad 1, existe $n_0 \geq 1$ tal que para $n \geq n_0$, tenemos que $|\hat{\sigma} - \sigma_0| \leq \delta$.

A partir del Lema 10.2, tenemos que

$$\sup_{\sigma > 0, \beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} \left| \frac{1}{n - p_1 - p_2} \sum_{i=1}^n \left[\rho \left(\frac{y_i - \langle X_i, \beta \rangle - \eta(z_i)}{\sigma} \right) - M(\beta, \eta, \sigma) \right] \right| \xrightarrow{a.s.} 0.$$

Por lo tanto, existe un conjunto de probabilidad cero \mathcal{N}_1 tal que para cualquier $\omega \notin \mathcal{N}_1$,

$$\sup_{\sigma > 0, \beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} \left| \frac{1}{n - p_1 - p_2} \sum_{i=1}^n \left[\rho \left(\frac{y_i - \langle X_i, \beta \rangle - \eta(z_i)}{\sigma} \right) - M(\beta, \eta, \sigma) \right] \right| \rightarrow 0 \quad (10.12)$$

se cumple. Por otro lado, por la ley de los grandes números

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{\sigma_0 \epsilon}{\sigma_0 + \delta} \right) \xrightarrow{a.s.} \mathbb{E}\rho \left(\frac{\sigma_0 \epsilon}{\sigma_0 + \delta} \right) < \mathbb{E}\rho(\epsilon) = b.$$

Como $(n - p_1 - p_2)/n \rightarrow 1$, tenemos que

$$\frac{1}{n - p_1 - p_2} \sum_{i=1}^n \rho \left(\frac{\sigma_0 \epsilon}{\sigma_0 + \delta} \right) \xrightarrow{a.s.} \mathbb{E}\rho \left(\frac{\sigma_0 \epsilon}{\sigma_0 + \delta} \right).$$

Luego, existe un conjunto de probabilidad nula \mathcal{N}_2 tal que, para cualquier $\omega \notin \mathcal{N}_2$,

$$A_n(\delta) = \frac{1}{n - p_1 - p_2} \sum_{i=1}^n \rho \left(\frac{\sigma_0 \epsilon}{\sigma_0 + \delta} \right) \rightarrow \mathbb{E}\rho \left(\frac{\sigma_0 \epsilon}{\sigma_0 + \delta} \right). \quad (10.13)$$

Finalmente, considerando que $\mathbb{E}(\|X\|) < \infty$, por la ley fuerte de los grande números y usando nuevamente que $(n - p_1 - p_2)/n \rightarrow 1$, obtenemos que existe un conjunto de probabilidad nula \mathcal{N}_3 tal que

$$\frac{1}{n - p_1 - p_2} \sum_{i=1}^n \|X_i\| \rightarrow \mathbb{E}(\|X\|), \quad (10.14)$$

para cualquier $\omega \notin \mathcal{N}_3$. Fijemos $\omega \notin \mathcal{N}_1 \cup \mathcal{N}_2 \cup \mathcal{N}_3$.

Comenzamos mostrando que existe n_0 tal que $\hat{\sigma} \leq \sigma_0 + \delta$ para $n \geq n_0$. Usando **FC5** y **FC7** y Schumaker (1991), existen $\tilde{\beta} \in \mathcal{M}_{p_1}^{(1)}$ y $\tilde{\eta} \in \mathcal{M}_{p_2}^{(2)}$ tales que

$$\|\tilde{\beta} - \beta_0\|_\infty = O(n^{-r\nu}) \quad \|\tilde{\eta} - \eta_0\|_\infty = O(n^{-r\nu}). \quad (10.15)$$

Usando un desarrollo de Taylor de orden uno resulta

$$\begin{aligned} \frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{y_i - \langle X_i, \tilde{\beta} \rangle - \tilde{\eta}(t_i)}{\sigma_0 + \delta} \right) &= \frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{\sigma_0 \epsilon_i + \langle X_i, \beta_0 - \tilde{\beta} \rangle + (\eta_0 - \tilde{\eta})(t_i)}{\sigma_0 + \delta} \right) \\ &= \frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{\sigma_0 \epsilon_i}{\sigma_0 + \delta} \right) + R_n = A_n(\delta) + R_n, \end{aligned}$$

donde

$$R_n = \frac{1}{n-p_1-p_2} \sum_{i=1}^n \psi \left(\frac{\sigma_0 \epsilon_i + \xi_i}{\sigma_0 + \delta} \right) \frac{\langle X_i, \beta_0 - \tilde{\beta} \rangle + (\eta_0 - \tilde{\eta})(t_i)}{\sigma_0 + \delta}$$

y ξ_i son puntos intermedios entre 0 y $\langle X_i, \beta_0 - \tilde{\beta} \rangle + (\eta_0 - \tilde{\eta})(t_i)$. De (10.13) obtenemos inmediatamente que

$$A_n(\delta) \rightarrow \mathbb{E} \rho \left(\frac{\sigma_0 \epsilon}{\sigma_0 + \delta} \right) = b_1 < \mathbb{E} \rho \left(\frac{\epsilon}{\sigma_0} \right) = b.$$

Además, la cota

$$|R_n| \leq \|\psi\|_\infty \frac{1}{\sigma_0 + \delta} \left(\|\eta_0 - \tilde{\eta}\|_\infty + \|\beta_0 - \tilde{\beta}\|_\infty \frac{1}{n-p_1-p_2} \sum_{i=1}^n \|X_i\| \right)$$

junto con (10.14) y (10.15) implican que $|R_n| \rightarrow 0$. Por lo tanto, obtenemos que

$$\frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{y_i - \langle X_i, \tilde{\beta} \rangle - \tilde{\eta}(t_i)}{\sigma_0 + \delta} \right) \rightarrow b_1.$$

Sea $\delta_1 > 0$ tal que $b_1 + \delta_1 < b$, entonces existe $n_0 \in \mathbb{N}$ tal que para $n \geq n_0$,

$$\frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{y_i - \langle X_i, \tilde{\beta} \rangle - \tilde{\eta}(t_i)}{\sigma_0 + \delta} \right) < b_1 + \delta_1 < b. \quad (10.16)$$

Notando que

$$\frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{y_i - \langle X_i, \tilde{\beta} \rangle - \tilde{\eta}(t_i)}{s_n(\tilde{\beta}, \tilde{\eta})} \right) = b,$$

a partir de (10.16) y usando que ρ es no decreciente, obtenemos que $s_n(\tilde{\beta}, \tilde{\eta}) < \sigma_0 + \delta$. Por otra parte, sabemos que $\hat{\sigma} = \min_{\beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} s_n(\beta, \eta)$ y que $\hat{\beta} \in \mathcal{M}_{p_1}^{(1)}$ y $\hat{\eta} \in \mathcal{M}_{p_2}^{(2)}$, luego $\hat{\sigma} \leq s_n(\tilde{\beta}, \tilde{\eta})$, lo que nos permite concluir que para $n \geq n_0$, $\hat{\sigma} = s_n(\hat{\beta}, \hat{\eta}) \leq \sigma_0 + \delta$.

Falta mostrar que existe $n_1 \in \mathbb{N}$ tal que, para cualquier $n \geq n_1$, $\hat{\sigma} \geq \sigma_0 - \delta$.

El hecho de que ρ sea no decreciente junto con **FC1** implican que $M(\beta_0, \eta_0, \sigma_0 - \delta) > M(\beta_0, \eta_0, \sigma_0) = b$ (ver el Lema 3 de Salibián–Barrera, 2006). Sea $\delta_2 > 0$ tal que $M(\beta_0, \eta_0, \sigma_0 - \delta) = b_2 = b + \delta_2$.

Usando que se cumple (10.12), obtenemos que existe $n_1 \in \mathbb{N}$ tal que para todo $n \geq n_1$,

$$\sup_{\sigma > 0, \beta \in \mathcal{M}_{p_1}^{(1)}, \eta \in \mathcal{M}_{p_2}^{(2)}} \left| \frac{1}{n-p_1-p_2} \sum_{i=1}^n \left[\rho \left(\frac{y_i - \langle X_i, \beta \rangle - \eta(z_i)}{\sigma} \right) - M(\beta, \eta, \sigma) \right] \right| < \frac{\delta_2}{2}.$$

Luego,

$$\left| \frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{y_i - \langle X_i, \hat{\beta} \rangle - \hat{\eta}(z_i)}{\hat{\sigma}} \right) - \frac{n}{n-p_1-p_2} M(\hat{\beta}, \hat{\eta}, \hat{\sigma}) \right| < \frac{\delta_2}{2},$$

lo que implica que

$$\frac{n}{n-p_1-p_2} M(\widehat{\beta}, \widehat{\eta}, \widehat{\sigma}) < \frac{1}{n-p_1-p_2} \sum_{i=1}^n \rho \left(\frac{y_i - \langle X_i, \widehat{\beta} \rangle - \widehat{\eta}(z_i)}{\widehat{\sigma}} \right) + \frac{\delta_2}{2} = b + \frac{\delta_2}{2}.$$

Por otro lado, usando nuevamente que $(n-p_1-p_2)/n \rightarrow 1$ y que ρ es acotada obtenemos que

$$\left| \frac{n}{n-p_1-p_2} M(\widehat{\beta}, \widehat{\eta}, \widehat{\sigma}) - M(\widehat{\beta}, \widehat{\eta}, \widehat{\sigma}) \right| \leq \left| \frac{n}{n-p_1-p_2} - 1 \right| \rightarrow 0$$

de manera que, sin pérdida de generalidad, podemos suponer que para cualquier $n \geq n_1$,

$$M(\widehat{\beta}, \widehat{\eta}, \widehat{\sigma}) < \frac{n}{n-p_1-p_2} M(\widehat{\beta}, \widehat{\eta}, \widehat{\sigma}) + \frac{\delta_2}{2} < b + \delta_2.$$

El Lema 10.1 implica que $M(\beta_0, \eta_0, \widehat{\sigma}) \leq M(\widehat{\beta}, \widehat{\eta}, \widehat{\sigma})$, entonces

$$M(\beta_0, \eta_0, \widehat{\sigma}) < b + \delta_2 = M(\beta_0, \eta_0, \sigma_0 - \delta),$$

de donde deducimos que $\widehat{\sigma} \geq \sigma_0 - \delta$ para $n \geq n_1$, lo que concluye la demostración.

10.3.6. Demostración del Teorema 10.7

Para demostrar el Teorema 10.7 necesitamos un lema auxiliar para calcular algunos números de entropía. Recordemos que $\Theta = \mathcal{L}_r([0, 1]) \times \mathcal{L}_r([0, 1])$, $\Theta_n = \mathcal{M}_{p_1}^{(1)} \times \mathcal{M}_{p_2}^{(2)} \cap \{\theta = (\beta, \eta) \in \Theta : \|\beta - \beta_0\|_\infty + \|\eta - \eta_0\|_\infty \leq \epsilon_0\}$ y $\pi^2(\theta, \theta_0) = \mathbb{E}[\langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z)]^2$. Observemos que, salvo un conjunto de probabilidad nula, $\widehat{\theta} = (\widehat{\beta}, \widehat{\eta}) \in \Theta_n$ para n suficientemente grande.

Lema 10.10. *Dados $\mathbf{b}_0 \in \mathbb{R}^{p_1}$ y $\mathbf{a}_0 \in \mathbb{R}^{p_2}$, sean $\widetilde{\beta}_0(t) = \sum_{j=1}^{p_1} b_{j,0} B_j^{(1)}(t)$ y $\widetilde{\eta}_0(z) = \sum_{j=1}^{p_2} a_{j,0} B_j^{(2)}(z)$, $\widetilde{\theta}_0 = (\widetilde{\beta}_0, \widetilde{\eta}_0) \in \mathcal{M}_{p_1}^{(1)} \times \mathcal{M}_{p_2}^{(2)}$. Sea ρ una función acotada que cumple **FC4**. Definamos la clase de funciones*

$$\begin{aligned} \mathcal{G}_{n,\delta,\widetilde{\theta}_0} &= \{f_{\theta,\sigma} = V_{\theta,\sigma} - V_{\widetilde{\theta}_0,\sigma} : d(\theta, \widetilde{\theta}_0) \leq \delta, \theta \in \Theta_n, \sigma \in \mathcal{V}\} \\ &= \{V_{\theta,\sigma} - V_{\widetilde{\theta}_0,\sigma} : \theta \in \Theta_{n,\delta}, \sigma \in \mathcal{V} = [\sigma_1, \sigma_2]\}, \end{aligned}$$

con $\sigma_1 = \sigma_0/2$, $\sigma_2 = 2\sigma_0$, $d(\theta, \theta_0) = \|\beta - \beta_0\|_\infty + \|\eta - \eta_0\|_\infty$ y

$$V_{\theta,\sigma} = \rho \left(\frac{y - \langle X, \beta \rangle + \eta(z)}{\sigma} \right)$$

para $\theta = (\beta, \eta)$. Si $\mathbb{E}\|X\|^2 < \infty$ existe una constante $C_2 > 0$ independiente de n , $\widetilde{\theta}_0$ y ϵ tal que

$$N_{[]}(\epsilon, \mathcal{G}_{n,\delta,\widetilde{\theta}_0}, L_2(P)) \leq C_2 \left(\frac{c}{\epsilon} \right)^{p_1+p_2+1}.$$

DEMOSTRACIÓN. La demostración sigue pasos análogos a los utilizados para probar el Lema 3.17. Como $\mathbb{E}g^2(z) \leq \|g\|_\infty^2$, los argumentos considerados en Shen y Wong (1994) al analizar el Caso 3 de la página 596 implican que el logaritmo del número de cubrimiento corchete de la clase

$$\mathcal{H}_{c,\widetilde{\beta}_0} = \left\{ \beta(t) = \sum_{j=1}^{p_1} b_j B_j^{(1)}(t), \mathbf{b} \in \mathbb{R}^{p_1}, \|\beta - \widetilde{\beta}_0\| \leq c \right\}$$

puede acotarse por

$$\log N_{[\cdot]}(\epsilon, \mathcal{H}_{c, \tilde{\beta}_0}, L_2(P)) \leq C_1 p_1 \log(c/\epsilon) \quad (10.17)$$

para $0 < \epsilon < c$ y la constante C_1 es independiente de c , $\tilde{\beta}_0$ y n . Similarmente, el logaritmo del número de cubrimiento corchete de la clase

$$\mathcal{H}_{c, \tilde{\eta}_0} = \left\{ \eta(z) = \sum_{j=1}^{p_2} a_j B_j^{(2)}(z), \mathbf{a} \in \mathbb{R}^{p_2}, \|\eta - \tilde{\eta}_0\|_\infty \leq c \right\}$$

puede acotarse por

$$\log N_{[\cdot]}(\epsilon, \mathcal{H}_{c, \tilde{\eta}_0}, L_2(P)) \leq C_2 p_2 \log(c/\epsilon) \quad (10.18)$$

Por (10.17) y (10.18), para cualquier $0 < \epsilon < c_1$, $\mathcal{H}_{c, \tilde{\beta}_0}$ puede cubrirse con un número finito $M_1(\epsilon) \leq (c/\epsilon)^{C_1 p_1}$ de ϵ -corchetes $\{[\beta_{j,L}, \beta_{j,U}], 1 \leq j \leq M_1(\epsilon)\}$ mientras que $\mathcal{H}_{c, \tilde{\eta}_0}$ puede cubrirse por un número finito $M_2(\epsilon) \leq (c/\epsilon)^{C_2 p_2}$ de ϵ -corchetes $\{[\eta_{j,L}, \eta_{j,U}], 1 \leq j \leq M_2(\epsilon)\}$. Por otro lado, el conjunto $\mathcal{V} = [\sigma_1, \sigma_2]$ puede cubrirse con $M_3(\epsilon) \leq C_3(1/\epsilon)$ bolas de radio ϵ y centro $\sigma^{(s)}$, $1 \leq s \leq M_3(\epsilon)$.

Como ψ es acotada si $\sigma \in [\sigma_1, \sigma_2]$,

$$\left| \frac{\partial}{\partial u} \rho \left(\frac{y-u}{\sigma} \right) \right| \leq \frac{\|\psi\|_\infty}{\sigma} \leq 2 \frac{\|\psi\|_\infty}{\sigma_0},$$

Dada $f_{\theta, \sigma} \in \mathcal{G}_{n, \delta, \tilde{\theta}_0}$, sea j , m y s tales que β pertenece al ϵ -corchete $[\beta_{j,L}, \beta_{j,U}]$, η pertenece al ϵ -corchete $[\eta_{m,L}, \eta_{m,U}]$ y $|\sigma - \sigma^{(s)}| < \epsilon$. Denotemos

$$\begin{aligned} f_{j,m,s}(y, X, z) &= \rho \left(\frac{y - \langle X, \beta_{j,U} \rangle + \eta_{m,U}(z)}{\sigma^{(s)}} \right) - \rho \left(\frac{y - \langle X, \beta_0 \rangle + \eta_0(z)}{\sigma^{(s)}} \right) \\ f_{j,m}(y, X, z) &= \rho \left(\frac{y - \langle X, \beta_{j,U} \rangle + \eta_{m,U}(z)}{\sigma} \right) - \rho \left(\frac{y - \langle X, \beta_0 \rangle + \eta_0(z)}{\sigma} \right) \end{aligned}$$

Usando un desarrollo de Taylor de orden uno y el hecho de que $\zeta(u) = u\psi(u)$ está acotada, obtenemos que

$$\begin{aligned} |f_{\theta, \sigma} - f_{j,m,s}| &\leq |f_{\theta, \sigma} - f_{j,m}| + |f_{j,m} - f_{j,m,s}| \\ &\leq \|\psi\|_\infty \{ \|X\| \|\beta - \beta_{j,U}\| + |\eta(z) - \eta_{m,U}(z)| \} + 2 \frac{\|\zeta\|_\infty}{\sigma_0} |\sigma - \sigma^{(s)}| \\ &\leq \epsilon \left(\|\psi\|_\infty \|X\| + 2 \frac{\|\zeta\|_\infty}{\sigma_0} \right) + \|\psi\|_\infty (\eta_{m,L}(z) - \eta_{m,U}(z)) \\ &\leq \epsilon \left(\|\psi\|_\infty \|X\| + 2 \frac{\|\zeta\|_\infty}{\sigma_0} + \|\psi\|_\infty \right), \end{aligned}$$

donde las últimas desigualdades siguen del hecho que $\eta_{m,L} \leq \eta(z) \leq \eta_{m,U}$, $\|\eta_{j,L}(z) - \eta_{j,U}(z)\|_\infty \leq \epsilon$, $0 \leq \beta_{j,U}(t) - \beta(t) \leq \beta_{j,U}(t) - \beta_{j,L}(t)$, con lo cual $\int_0^1 [\beta_{j,U}(t) - \beta(t)]^2 dt \leq \int_0^1 [\beta_{j,U}(t) - \beta_{j,L}(t)]^2 dt < \epsilon$ y $|\sigma - \sigma^{(s)}| < \epsilon$. Definamos las funciones

$$\begin{aligned} \varphi_{j,m,s}^{(U)}(y, X, z) &= f_{j,m,s}(y, X, z) + \epsilon \left(\|\psi\|_\infty \|X\| + 2 \frac{\|\zeta\|_\infty}{\sigma_0} + \|\psi\|_\infty \right) \\ \varphi_{j,m,s}^{(L)}(y, \mathbf{x}, tX, z) &= f_{j,m,s}(y, X, z) - \epsilon \left(\|\psi\|_\infty \|X\| + 2 \frac{\|\zeta\|_\infty}{\sigma_0} + \|\psi\|_\infty \right). \end{aligned}$$

Observemos que $\varphi_{j,m,s}^{(L)} \leq f_{\theta,\sigma} \leq \varphi_{j,m,s}^{(U)}$. Luego, considerando que $\mathbb{E}\|X\|^2 < \infty$, obtenemos que

$$\begin{aligned} \|\varphi_{j,m,s}^{(U)} - \varphi_{j,m,s}^{(L)}\|_2 &\leq 2 \left\{ \epsilon \left(\|\psi\|_\infty (\mathbb{E}\|X\|^2)^{1/2} + 2 \frac{\|\zeta\|_\infty}{\sigma_0} \right) + \|\psi\|_\infty \right\} \\ &\leq 2\epsilon \left(\|\psi\|_\infty (\mathbb{E}\|X\|^2)^{1/2} + 2 \frac{\|\zeta\|_\infty}{\sigma_0} + \|\psi\|_\infty \right) = A_1 \epsilon. \end{aligned}$$

Por lo tanto, el número total de corchetes de tamaño $A_1\epsilon$ necesarios para cubrir $\mathcal{G}_{n,\delta,\tilde{\theta}_0}$ está acotado por $\prod_{i=1}^3 M_i(\epsilon) \leq A(\delta/\epsilon)^{C(p_1+p_2+1)}$ con $C = \max(C_1, C_2, 1)$, lo que concluye la demostración. ■

La demostración del Teorema 10.7 utiliza argumentos semejantes a los considerados en la demostración del Teorema 3.11.

DEMOSTRACIÓN DEL TEOREMA 10.7. Como en la demostración de la Proposición 10.3, sean $\tilde{\beta} \in \mathcal{M}_{p_1}^{(1)}$ y $\tilde{\eta} \in \mathcal{M}_{p_2}^{(2)}$ tales que $\|\tilde{\beta} - \beta_0\|_\infty = O(n^{-r\nu})$, $\|\tilde{\eta} - \eta_0\|_\infty = O(n^{-r\nu})$ y llamemos $\theta_{0,n} = (\tilde{\beta}, \tilde{\eta})$. Más aun, indiquemos por $\mathbf{b}_{0,n} \in \mathbb{R}^{p_1}$ y $\mathbf{a}_{0,n} \in \mathbb{R}^{p_2}$ a los vectores tales que $\tilde{\beta}(t) = \mathbf{b}_{0,n}^\top \mathbf{B}^{(1)}(t)$ y $\tilde{\eta}(z) = \mathbf{a}_{0,n}^\top \mathbf{B}^{(2)}(z)$ donde $\mathbf{B}^{(1)}(t) = (B_1^{(1)}, \dots, B_{p_1}^{(1)}(t))^\top$ y $\mathbf{B}^{(2)}(z) = (B_1^{(2)}(z), \dots, B_{p_2}^{(2)}(z))^\top$.

Para obtener la tasa de convergencia de nuestro estimador $\hat{\theta} = (\hat{\beta}, \hat{\eta})$ aplicaremos el Teorema 3.4.1 de van der Vaart y Wellner (1996). Siguiendo su notación llamemos $L_n(\theta) = -M(\theta, \hat{\sigma})$ y $\mathbb{L}_n(\theta) = -M_n(\theta, \hat{\sigma})$ y para $\theta \in \Theta_n$ sea $d_n(\theta, \theta_0) = \pi(\theta, \theta_0)$. La función L_n es aleatoria debido al estimador del parámetro de escala $\hat{\sigma}$.

Sea $\delta_n = A \left(\|\beta_0 - \tilde{\beta}\|_\infty + \|\eta_0 - \tilde{\eta}\|_\infty \right)$, donde $A = 4 \sqrt{(C_0 (\mathbb{E}\|X\|^2 + 1) + A_0)/C_0}$ con $A_0 = 4 \|\psi'\|_\infty (\mathbb{E}\|X\|^2 + 1)$ y C_0 está dada en **FC9**.

Teniendo en cuenta que

$$|(M_n(\theta, \hat{\sigma}) - M(\theta, \hat{\sigma})) - (M_n(\theta_{0,n}, \hat{\sigma}) - M(\theta_{0,n}, \hat{\sigma}))| = |(\mathbb{L}_n - L_n)(\theta) - (\mathbb{L}_n - L_n)(\theta_{0,n})|$$

para usar el Teorema 3.4.1 de van der Vaart y Wellner (1996), mostraremos que existe una función ϕ_n tal que $\phi_n(\delta)/\delta^{\nu_1}$ es decreciente en (δ_n, ∞) para algún $\nu_1 < 2$ y que para cualquier $\delta > \delta_n$,

$$\sup_{\theta \in \Theta_{n,\delta}} L_n(\theta) - L_n(\theta_{0,n}) = \sup_{\theta \in \Theta_{n,\delta}} M(\theta_{0,n}, \hat{\sigma}) - M(\theta, \hat{\sigma}) \lesssim -\delta^2 \quad (10.19)$$

$$\mathbb{E}^* \sup_{\theta \in \Theta_{n,\delta}} \sqrt{n} |(M_n(\theta, \hat{\sigma}) - M(\theta, \hat{\sigma})) - (M_n(\theta_{0,n}, \hat{\sigma}) - M(\theta_{0,n}, \hat{\sigma}))| \lesssim \phi_n(\delta) \quad (10.20)$$

$$d_n(\hat{\theta}, \theta_{0,n}) \xrightarrow{p} 0, \quad (10.21)$$

donde el símbolo \lesssim significa *menor o igual salvo por una constante multiplicativa*, \mathbb{E}^* es la esperanza exterior y $\Theta_{n,\delta} = \{\theta \in \Theta_n : \delta/2 < d_n(\theta, \theta_{0,n}) \leq \delta\}$.

El hecho de que $d_n(\theta, \theta_{0,n}) \leq d_n(\theta, \theta_0) + d_n(\theta_{0,n}, \theta_0)$ implica que

$$d_n^2(\theta, \theta_{0,n}) \leq 2 \left\{ d_n^2(\theta, \theta_0) + (\mathbb{E}\|X\|^2 + 1) \left(\|\tilde{\beta} - \beta_0\|_\infty^2 + \|\tilde{\eta} - \eta_0\|_\infty^2 \right) \right\}. \quad (10.22)$$

Usando (10.22), $d_n^2(\hat{\theta}, \theta_0) \xrightarrow{a.s.} 0$, $\|\tilde{\beta} - \beta_0\|_\infty = O(n^{-r\nu})$, $\|\tilde{\eta} - \eta_0\|_\infty = O(n^{-r\nu})$, deducimos que (10.21) se cumple.

El supuesto **FC9** y $\hat{\sigma} \xrightarrow{a.s.} \sigma_0$ implican que, salvo un conjunto de probabilidad nula, si $n \geq n_0$, para todo $\theta \in \Theta_n$, $M(\theta, \hat{\sigma}) - M(\theta_0, \hat{\sigma}) \geq C_0 \pi^2(\theta, \theta_0)$. Además, como los errores tienen distribución simétrica y son independientes de las covariables, obtenemos que

$$\mathbb{E} \left[\psi \left(\frac{y - \langle X, \beta_0 \rangle + \eta_0(z)}{\sigma} \right) \left(\langle X, \tilde{\beta} - \beta_0 \rangle + \tilde{\eta}(z) - \eta_0(z) \right) \right] = 0,$$

de donde se deduce que

$$\begin{aligned} M(\theta_{0,n}, \sigma) - M(\theta_0, \sigma) &= \mathbb{E} \left[\psi \left(\frac{y - \langle X, \beta_0 \rangle + \eta_0(z)}{\sigma} \right) \left(\langle X, \tilde{\beta} - \beta_0 \rangle + \tilde{\eta}(z) - \eta_0(z) \right) \right] \\ &\quad + \frac{1}{2} \mathbb{E} \left[\psi' \left(\frac{\xi}{\sigma} \right) \left(\langle X, \tilde{\beta} - \beta_0 \rangle + \tilde{\eta}(z) - \eta_0(z) \right)^2 \right] \\ &= \frac{1}{2} \mathbb{E} \left[\psi' \left(\frac{\xi}{\sigma} \right) \left(\langle X, \tilde{\beta} - \beta_0 \rangle + \tilde{\eta}(z) - \eta_0(z) \right)^2 \right] \\ &\leq \frac{1}{2} \|\psi'\|_\infty \mathbb{E} \left(\langle X, \tilde{\beta} - \beta_0 \rangle + \tilde{\eta}(z) - \eta_0(z) \right)^2 \\ &\leq \frac{1}{2} \|\psi'\|_\infty 4 \left(\mathbb{E} \langle X, \tilde{\beta} - \beta_0 \rangle^2 + \mathbb{E} (\tilde{\eta}(z) - \eta_0(z))^2 \right) \\ &\leq 2 \|\psi'\|_\infty \left(\mathbb{E} \|X\|^2 \|\tilde{\beta} - \beta_0\|_\infty^2 + \|\tilde{\eta} - \eta_0\|_\infty^2 \right) \\ &\leq A_0 \left[\|\tilde{\beta} - \beta_0\|_\infty^2 + \|\tilde{\eta} - \eta_0\|_\infty^2 \right] = O(n^{-2r\nu}), \end{aligned}$$

donde $A_0 = 4 \|\psi'\|_\infty (\mathbb{E} \|X\|^2 + 1)$ y ξ es un punto intermedio entre $y - \langle X, \tilde{\beta} \rangle + \tilde{\eta}(z)$ e $y - \langle X, \beta_0 \rangle + \eta_0(z)$. Usando (10.22) y como $\delta/2 < d_n(\theta, \theta_{0,n})$, deducimos que

$$\begin{aligned} M(\theta, \hat{\sigma}) - M(\theta_{0,n}, \hat{\sigma}) &\geq C_0 d_n^2(\theta, \theta_0) - A_0 \left[\|\tilde{\beta} - \beta_0\|_\infty^2 + \|\tilde{\eta} - \eta_0\|_\infty^2 \right] \\ &\geq \frac{C_0}{2} d_n^2(\theta, \theta_{0,n}) - (C_0 (\mathbb{E} \|X\|^2 + 1) + A_0) \left(\|\tilde{\beta} - \beta_0\|_\infty^2 + \|\tilde{\eta} - \eta_0\|_\infty^2 \right) \\ &\geq \frac{C_0}{2} d_n^2(\theta, \theta_{0,n}) - (C_0 (\mathbb{E} \|X\|^2 + 1) + A_0) \left(\|\beta_0 - \tilde{\beta}\|_\infty + \|\eta_0 - \tilde{\eta}\|_\infty \right)^2 \\ &\geq \frac{C_0}{8} \delta^2 - \frac{1}{A^2} (C_0 (\mathbb{E} \|X\|^2 + 1) + A_0) \delta_n^2 = \frac{C_0}{8} \delta^2 - \frac{C_0}{16} \delta_n^2 \geq \frac{C_0}{16} \delta^2, \end{aligned}$$

lo que prueba (10.19).

Ahora debemos hallar $\phi_n(\delta)$ tal que $\phi_n(\delta)/\delta$ sea decreciente en δ y se cumpla (10.20). De la consistencia de $\hat{\sigma}$, con probabilidad uno, para n suficientemente grande

$$\begin{aligned} \sqrt{n} |(L_n(\theta, \hat{\sigma}) - L(\theta, \hat{\sigma})) - (L_n(\theta_{0,n}, \hat{\sigma}) - L(\theta_{0,n}, \hat{\sigma}))| &\leq \\ &\sup_{\sigma \in \mathcal{V}} \sqrt{n} |(L_n(\theta, \sigma) - L(\theta, \sigma)) - (L_n(\theta_{0,n}, \sigma) - L(\theta_{0,n}, \sigma))|, \end{aligned}$$

donde $\mathcal{V} = [\sigma_0/2, 2\sigma_0]$. Definamos la clase de funciones

$$\mathcal{F}_{n,\delta} = \left\{ V_{\theta,\sigma} - V_{\theta_{0,n},\sigma} : \frac{\delta}{2} \leq d_n(\theta, \theta_{0,n}) \leq \delta, \theta \in \Theta_n, \sigma \in \mathcal{V} \right\} = \{V_{\theta,\sigma} - V_{\theta_{0,n},\sigma} : \theta \in \Theta_{n,\delta}, a \in \mathcal{V}\},$$

con

$$V_{\theta,\sigma} = \rho \left(\frac{y - \langle X, \beta \rangle + \eta(z)}{\sigma} \right)$$

para $\theta = (\beta, \eta)$. La desigualdad (10.20) involucra un proceso empírico indexado en $\mathcal{F}_{n,\delta}$ pues

$$\mathbb{E}^* \sup_{\theta \in \Theta_{n,\delta}} \sqrt{n} |(L_n(\theta, \hat{\sigma}) - L(\theta, \hat{\sigma})) - (L_n(\theta_{0,n}, \hat{\sigma}) - L(\theta_{0,n}, \hat{\sigma}))| \leq \mathbb{E}^* \sup_{f \in \mathcal{F}_{n,\delta}} \sqrt{n} |(P_n - P)f|.$$

Para cualquier $f \in \mathcal{F}_{n,\delta}$, tenemos que $\|f\|_\infty \leq A_1 = 2\|\rho\|_\infty$. Más aun, si $A_2 = 2\|\psi\|_\infty/\sigma_0$ usando que $\sigma \in \mathcal{V}$, la desigualdad

$$|V_{\theta,\sigma} - V_{\theta_{0,n},\sigma}| = \left| \rho \left(\frac{y - \langle X, \beta \rangle + \eta(z)}{\sigma} \right) - \rho \left(\frac{y - \langle X, \tilde{\beta} \rangle + \tilde{\eta}(z)}{\sigma} \right) \right| \leq 2\|\psi\|_\infty \left| \frac{\langle X, \beta - \tilde{\beta} \rangle + \eta(z) - \tilde{\eta}(z)}{\sigma_0} \right|,$$

y el hecho de que $d_n(\theta, \theta_{0,n}) \leq \delta$, tenemos que

$$Pf^2 \leq \frac{4\|\psi\|_\infty^2}{\sigma_0^2} \mathbb{E} \left(\left[\langle X, \beta - \tilde{\beta} \rangle + \eta(z) - \tilde{\eta}(z) \right]^2 \right) = A_2^2 d_n^2(\theta, \theta_{0,n}) \leq A_2^2 \delta^2.$$

Por lo tanto, el Lema 3.4.2 de van der Vaart y Wellner (1996) permite concluir que

$$\mathbb{E}^* \sup_{f \in \mathcal{F}_{n,\delta}} \sqrt{n} |(P_n - P)f| \leq J_{[]} \left(A_2^{1/2} \delta, \mathcal{F}_{n,\delta}, L_2(P) \right) \left(1 + A_1 \frac{J_{[]} (A_2^{1/2} \delta, \mathcal{F}_{n,\delta}, L_2(P))}{A_2 \delta^2 \sqrt{n}} \right),$$

donde $J_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon$ es la integral corchete de la clase \mathcal{F} .

Observemos que cualquier $\theta = (\beta, \eta) \in \Theta_{n,\delta}$, β puede escribirse como $\beta = \mathbf{b}^T \mathbf{x}$, mientras $\eta = \mathbf{a}^T \mathbf{B}$ para algún $\mathbf{b} \in \mathbb{R}^{p_1}$ y $\mathbf{a} \in \mathbb{R}^{p_2}$, entonces

$$d_n^2(\theta, \theta_{0,n}) = \mathbb{E} \left(\left[\mathbf{x}^T (\mathbf{b} - \mathbf{b}_{0,n}) + (\mathbf{a} - \mathbf{a}_{0,n})^T \mathbf{B}(z) \right]^2 \right).$$

Luego, $\mathcal{F}_{n,\delta} \subset \mathcal{G}_{n,c,\lambda_n}$ con $c = \delta$ y la cota dada en el Lema 10.10 permite obtener

$$N_{[]}(\epsilon, \mathcal{F}_{n,\delta}, L_2(P)) \leq C_2 \left(\frac{\delta}{\epsilon} \right)^{p_1 + p_2 + 1},$$

de donde se tiene que

$$J_{[]} (A_2^{1/2} \delta, \mathcal{F}_{n,\delta}, L_2(P)) \lesssim \delta \sqrt{p_1 + p_2 + 1}.$$

Si llamamos $q_n = p_1 + p_2 + 1$, obtenemos que para alguna constante A_3 independiente de n y δ ,

$$\mathbb{E}^* \sup_{\theta \in \Theta_{n,\delta}} |\mathbb{G}_n V_{\theta_{0,n},\kappa_0} - \mathbb{G}_n V_{\theta,\kappa_0}| \leq A_3 \left[\delta q_n^{1/2} + \frac{q_n}{\sqrt{n}} \right].$$

Eligiendo

$$\phi_n(\delta) = \delta q_n^{1/2} + \frac{q_n}{\sqrt{n}},$$

tenemos que $\phi_n(\delta)/\delta$ es decreciente en δ lo que prueba (10.20).

Sea $\gamma_n = O(n^{\min(r\nu, (1-\nu)/2)})$, entonces $\gamma_n \lesssim \delta_n^{-1}$, donde $\delta_n = A \left\{ \|\beta_0 - \tilde{\beta}\|_\infty + \|\eta_0 - \tilde{\eta}\|_\infty \right\} = O(n^{-r\nu})$ como hemos definido previamente. Tenemos que mostrar que $\gamma_n^2 \phi_n(1/\gamma_n) \lesssim \sqrt{n}$. Notemos que

$$\gamma_n^2 \phi_n \left(\frac{1}{\gamma_n} \right) = \gamma_n q_n^{1/2} + \gamma_n^2 \frac{q_n}{\sqrt{n}} = \sqrt{n} a_n (1 + a_n),$$

donde $a_n = \gamma_n q_n^{1/2} / \sqrt{n}$. Para derivar que $\gamma_n^2 \phi_n(1/\gamma_n) \lesssim \sqrt{n}$, basta con mostrar que $a_n = O(1)$, lo que sigue de $q_n = O(n^\nu)$ y $\gamma_n = O(n^\varsigma)$ con $\varsigma = \min(r\nu, (1 - \nu)/2)$.

Finalmente, la condición $\mathbb{L}_n(\hat{\theta}) \geq \mathbb{L}_n(\theta_{0,n}) - O_{\mathbb{P}}(\gamma_n^{-2})$ requerida por el Teorema 3.4.1 de van der Vaart y Wellner (1996) se verifica trivialmente porque $\hat{\theta}$ minimiza $M_n(\theta, \hat{\sigma})$. Luego, tenemos que $\gamma_n^2 d_n^2(\theta_{0,n}, \hat{\theta}) = O_{\mathbb{P}}(1)$.

Por otro lado $d_n(\theta_{0,n}, \theta_0) \leq \|\tilde{\beta} - \eta_0\|_{\infty} (\mathbb{E}\|X\|^2)^{1/2} + \|\tilde{\eta} - \eta_0\|_{\infty} = O(n^{-r\nu}) \leq \gamma_n$, que junto a $\gamma_n^2 d_n^2(\theta_{0,n}, \hat{\theta}) = O_{\mathbb{P}}(1)$ y la desigualdad triangular $\gamma_n^2 d_n^2(\theta_0, \hat{\theta}) = O_{\mathbb{P}}(1)$ lo que concluye la demostración. \blacksquare

10.3.7. Demostración del Lema 10.8

Como en la demostración del Teorema 10.4 tenemos que

$$\begin{aligned} M(\theta, \sigma) - M(\theta_0, \sigma) &= \mathbb{E} \left[\rho \left(\frac{y - \langle X, \beta \rangle + \eta(z)}{\sigma} \right) - \rho \left(\frac{y - \langle X, \beta_0 \rangle + \eta_0(z)}{\sigma} \right) \right] \\ &= \mathbb{E} \left[\rho \left(\frac{\sigma_0 \epsilon - \langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z)}{\sigma} \right) - \rho \left(\frac{\sigma_0 \epsilon}{\sigma} \right) \right] \\ &= \mathbb{E} \left[\psi \left(\frac{\sigma_0 \epsilon}{\sigma} \right) (\langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z)) \right] \\ &\quad + \frac{1}{2} \mathbb{E} \left[\psi' \left(\frac{\sigma_0 \epsilon + \xi}{\sigma} \right) (\langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z))^2 \right] \\ &= \frac{1}{2} \mathbb{E} \left[\psi' \left(\frac{\sigma_0 \epsilon + \xi}{\sigma} \right) (\langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z))^2 \right], \end{aligned}$$

donde ξ es un punto intermedio entre $g(X, z) = \langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z)$ y 0. Notemos que $|g(X, z)| \leq \|X\| \|\beta - \beta_0\|_{\infty} + \|\eta - \eta_0\|_{\infty}$, entonces si $\|\beta - \beta_0\|_{\infty} + \|\eta - \eta_0\|_{\infty} < \epsilon_0$, tenemos que $|\xi| \leq (C + 1)\epsilon_0$ con probabilidad 1.

El hecho de que $\varphi = \mathbb{E}\psi'(\epsilon) > 0$ y la continuidad de ψ' implican que para δ suficientemente chico

$$\inf_{\sigma > 0, |\sigma - \sigma_0| < \delta, |a| < \delta} \mathbb{E}\psi' \left(\frac{\sigma_0 \epsilon + a}{\sigma} \right) > \frac{\varphi}{2} > 0,$$

Luego, si $\mathcal{V} = \{\sigma > 0 : |\sigma - \sigma_0| < \delta\}$ y $\epsilon_0 = \delta/(C + 1)$, tenemos que

$$\begin{aligned} M(\theta, \sigma) - M(\theta_0, \sigma) &= \frac{1}{2} \mathbb{E} \left[\psi' \left(\frac{\sigma_0 \epsilon + \xi}{\sigma} \right) (\langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z))^2 \right] \\ &> \frac{\varphi}{2} \mathbb{E} \left[(\langle X, \beta - \beta_0 \rangle + \eta(z) - \eta_0(z))^2 \right] = \frac{\varphi}{2} \pi(\theta, \theta_0) \end{aligned}$$

lo que concluye la demostración.

Capítulo 11

Estudio numérico y análisis de un conjunto de datos reales

11.1. Estudio de Monte Carlo

En esta sección reportamos los resultados obtenidos en un estudio de Monte Carlo diseñado para investigar, en muestras finitas, el desempeño de los estimadores basados en B-splines definidos en la Sección 9.2 y su estabilidad frente a observaciones atípicas.

En todas las tablas, el procedimiento clásico basado en cuadrados mínimos se indica como CL, mientras que los M-estimadores definidos en Huang *et al.* (2015) y los MM-estimadores introducidos en esta tesis a través de las ecuaciones (9.4) y (9.5), se indican por M y MM, respectivamente. Los M-estimadores se calcularon usando la función de Huber con constante $c = 1.345$, mientras que para los MM-estimadores, las funciones ρ_{c_1} y ρ_{c_2} se tomaron como la función bicuadrada de Tukey con constantes de calibración $c_1 = 1.54764$ ($b = 1/2$) y $c_2 = 4.685$, tomando 5000 submuestras para calcular el S-estimador. Para el cálculo de los MM-estimadores, se utilizó la función `lmrob` del paquete `robustbase`. Los estimadores de β_0 y η_0 se aproximaron usando splines cúbicos con nodos equiespaciados. En la Sección 11.1.2, se reportan los resultados de un estudio preliminar en el que analizamos, para muestras sin contaminar, el comportamiento de los criterios para seleccionar la dimensión de las bases, definidos en la Sección 9.4. En la Sección 11.1.3 se analiza el comportamiento de los estimadores para muestras con y sin contaminar, para evaluar sus propiedades de eficiencia y robustez.

Teniendo en cuenta que el modelo con el que generamos los datos fue construido utilizando una función η_0 monótona, se presentan los resultados obtenidos mediante el procedimiento de monotonización de Neumeyer (2007) indicados en las Tablas por $\widehat{\eta}_{\text{MOD}}$.

Para evaluar el desempeño de cada estimador consideramos las medidas resumen denotadas como Sesgo^2 y $\text{MISE} = \text{VAR} + \text{Sesgo}^2$ que son aproximaciones del cuadrado del sesgo integrado y el error cuadrático medio integrado, respectivamente. Dichas medidas fueron calculadas en grillas equiespaciadas de 100 puntos sobre \mathcal{T} al evaluar los estimadores de β_0 y sobre \mathcal{Z} para los estimadores de η_0 . A continuación, damos su definición.

Dada una función $\gamma : \mathcal{I} \rightarrow \mathbb{R}$, consideramos una grilla de $M = 100$ puntos equiespaciados $t_1 \leq \dots \leq t_M$ sobre \mathcal{I} . Sea $\hat{\gamma}_j$, $1 \leq j \leq n_R$, la estimación de γ obtenida en la j -ésima replicación. Definimos

$$\text{Sesgo}^2(\hat{\gamma}) = \frac{1}{M} \sum_{\ell=1}^M \left(\frac{1}{n_R} \sum_{j=1}^{n_R} \hat{\gamma}_j(t_\ell) - \gamma(t_\ell) \right)^2 \quad \text{y} \quad \text{MISE}(\hat{\gamma}) = \frac{1}{n_R} \sum_{j=1}^{n_R} \left[\frac{1}{M} \sum_{\ell=1}^M (\hat{\gamma}_j(t_\ell) - \gamma(t_\ell))^2 \right]$$

que dan aproximaciones numéricas de $\int_{\mathcal{I}} \left(1/n_R \sum_{j=1}^{n_R} \hat{\gamma}_j(t) - \gamma(t) \right)^2 dt$ y $\sum_{j=1}^{n_R} \int_{\mathcal{I}} (\hat{\gamma}_j(t) - \gamma(t))^2 dt/n_R$, respectivamente. Tanto Sesgo^2 como MISE pueden verse dominados por errores numéricos cerca del límite del soporte de los datos. Para atenuar este efecto, como en He y Shi (1998), consideramos el desempeño de los estimadores en el interior del espacio de diseño tomando las medidas

$$\begin{aligned} \text{Sesgo}_{\text{TR}}^2(\hat{\gamma}) &= \frac{1}{M-2q} \sum_{\ell=q+1}^{M-q} \left(\frac{1}{n_R} \sum_{j=1}^{n_R} \hat{\gamma}_j(t_\ell) - \gamma(t_\ell) \right)^2 \\ \text{MISE}_{\text{TR}}(\hat{\gamma}) &= \frac{1}{n_R} \sum_{j=1}^{n_R} \left[\frac{1}{M-2q} \sum_{\ell=q+1}^{M-q} (\hat{\gamma}_j(t_\ell) - \gamma(t_\ell))^2 \right] \end{aligned}$$

En nuestro estudio excluimos el 10% de los puntos exteriores, esto es $q = \lceil M \times 0.05 \rceil$. Por lo tanto, calculamos las medidas resumen en el restante 90% interior. Todas las tablas que presentaremos en las siguientes secciones corresponden a medidas con un 10% de poda. Relegamos al apéndice del presente capítulo las correspondientes versiones de las tablas sin poda alguna.

11.1.1. Esquemas de simulación

Los siguientes modelos se consideraron para generar los datos. En el primero de ellos, la función β_0 se eligió como una combinación lineal de las autofunciones del operador de covarianza de X . En el segundo modelo, β_0 es un polinomio. En todos los casos, se realizaron $n_R = 500$, replicaciones generando muestras de tamaño $n = 300$.

Las muestras de datos fueron generadas a partir del siguiente modelo.

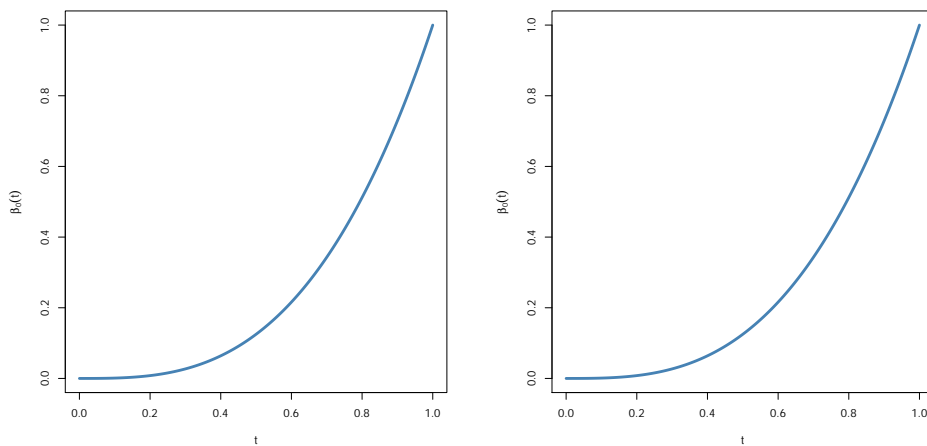
$$y_i = \langle \beta_0, X_i \rangle + \eta_0(z_i) + \epsilon_i \quad (11.1)$$

con $\mathcal{T} = [0, 1]$, $z_i \sim \mathcal{U}(-1, 1)$, de modo que $\mathcal{Z} = [-1, 1]$. La componente no paramétrica se eligió monótona $\eta_0(z) = 3 \arctan(10(z - 0.5))$. Por otra parte, $X_i(t)$ es un proceso con autofunciones

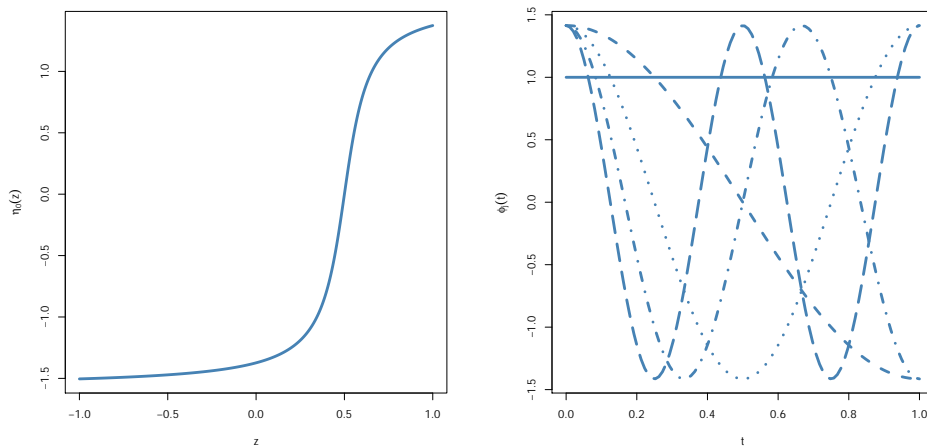
$\phi_j(t)$ donde ϕ_j son los elementos de la base de Fourier $\phi_1(t) \equiv 1$ y $\phi_j(t) = \sqrt{2} \cos((j-1)\pi t)$, $j \geq 2$ y con *scores* ξ_{ij} independientes e idénticamente distribuidas. Para muestras sin contaminar, que indicaremos C_0 , $\xi_{ij} \sim N(0, j^{-2})$ y $\epsilon_i \sim N(0, 1)$. Teniendo en cuenta que $\text{VAR}(\xi_{ij}) \leq 1/2500$, cuando $j > 2500$, se aproximó numéricamente el proceso mediante los primeros $k = 50$ términos.

En el primer modelo, llamado **Modelo 1**, $\beta_0(t) = \sum_{j=1}^{50} \beta_{j,0} \phi_j$, con $\beta_{1,0} = 0.3$ y $\beta_{j,0} = 4(-1)^{j+1} j^{-2}$, $j \geq 2$. Este es el modelo considerado en Qingguo (2015) y Huang *et al.* (2015). Mientras que en el segundo modelo, llamado **Modelo 2**, $\beta_0(t) = t^3$.

La Figura 11.1 da una representación gráfica de los parámetros β_0 y η_0 de ambos modelos, así como de los primeros elementos de la base de autofunciones del proceso $X(t)$.



(a) Pendiente funcional β_0 para el Modelo 1. (b) Pendiente funcional β_0 para el Modelo 2.



(c) Componente no paramétrica η_0 . (d) Primeros cinco elementos ϕ_j de la base de autofunciones.

Figura 11.1: Parámetros verdaderos de la simulación.

Para cada modelo, consideramos dos esquemas de contaminación donde esperamos que el primero afecte principalmente a los estimadores de la función de regresión no paramétrica η_0 , mientras que

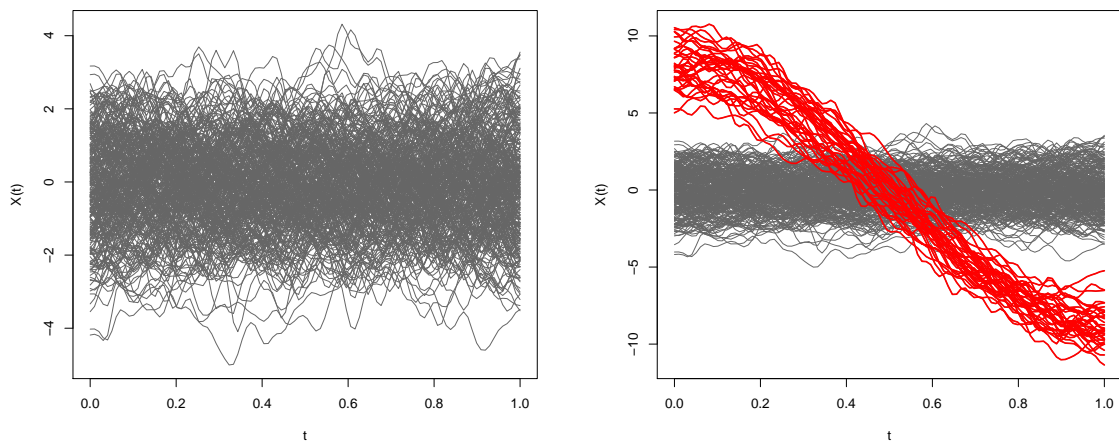
el segundo debería afectar a los estimadores de la función de regresión lineal β_0 .

- C_1 : $\epsilon_i^{(c)} \sim 0.9 N(0, 1) + 0.1 N(\mu, 0.25)$, con $\mu = 12$, que corresponde al modelo de errores groseros (*gross-errors model*).
- C_2 : En este caso, contaminamos la covariables X_i como sigue. $X_i^{(c)} = \sum_{j=1}^{50} \xi_{ij}^{(c)} \phi_j(t)$, con $\xi_{ij}^{(c)} \sim N(0, j^{-2})$ para $j \neq 2$ y $\xi_{i2}^{(c)} \sim 0.9 N(0, 0.25) + 0.1 N(\mu/2, 0.25)$. Para obtener puntos con alta palanca dañinos también contaminamos el error cada vez que las covariables eran contaminadas. Concretamente se generaron variables $v_i \sim Bin(1, 0.10)$ de modo tal que
 - Si $v_i = 0$, $\epsilon_i^{(c)} = \epsilon_i$ y $X_i^{(c)} = X_i$.
 - Si $v_i = 1$, $\epsilon_i^{(c)} \sim N(\mu, 0.25)$ y $X_i^{(c)} = \sum_{j=1}^{50} \xi_{ij}^{(c)} \phi_j(t)$, con $\xi_{ij}^{(c)} \sim N(0, j^{-2})$ para $j \neq 2$ y $\xi_{i2}^{(c)} \sim N(\mu/2, 0.25)$.

Las respuestas se generaron utilizando el modelo $y_i^{(c)} = \langle \beta_0, X_i^{(c)} \rangle + \eta_0(z_i) + \epsilon_i^{(c)}$. Como antes, se tomó $\mu = 12$.

De esta forma, para C_1 se obtuvieron errores groseros en las respuestas (outliers verticales) y para C_2 se obtuvieron valores grandes de las respuestas asociados a puntos de alta palanca.

La Figura 11.2 muestra las trayectorias generadas en una de las replicaciones para ilustrar el efecto producido por la contaminación $C_{2,\mu}$.



(a) Trayectorias $X_i(t)$ bajo C_0 .

(b) Trayectorias $X_i(t)$ bajo C_2 .

Figura 11.2: A la izquierda, trayectorias $X_i(t)$ para muestras sin contaminación. A la derecha, las trayectorias contaminadas para C_2 (en rojo se muestran las observaciones contaminadas).

11.1.2. Sobre los criterios de selección de las dimensiones de las bases

A partir de los modelos descritos en la sección precedente, calculamos para cada una de las muestras todos los criterios detallados en la sección 9.4. Las dimensiones de las bases variaron en los rangos $4 \leq p_1 \leq 13$ y $4 \leq p_2 \leq 13$ ya que se utilizaron splines cúbicos. La Tabla 11.1 muestra, para cada criterio de selección, el desempeño de los estimadores medido a través del $\text{Sesgo}_{\text{TR}}^2$ y MISE_{TR} cuando se consideran muestras sin contaminación (bajo C_0) generadas a partir del Modelo 1. Para los criterios BIC_α y BIC_α^* tomamos dos valores de α : $\alpha = 1$ que corresponde al criterio usual en el caso del estimador de mínimos cuadrados y $\alpha = 0.5$.

Como puede observarse en la Tabla 11.1 y en la Figura 11.3, los valores de MISE_{TR} de $\hat{\eta}$ y $\hat{\eta}_{\text{MOD}}$ son más estables para los distintos criterios que los obtenidos para los estimadores $\hat{\beta}$ del parámetro de regresión funcional. Para estos últimos, los valores más pequeños corresponden a los criterios BIC_1 y BIC_1^* . Si bien ambos penalizan por igual a los parámetros p_1 y p_2 , el criterio BIC_1^* pesa en mayor medida al valor de la función de pérdida. Cabe destacar sin embargo, que los valores de sesgo cuadrado para los estimadores clásicos de la componente no paramétrica es mucho menor cuando se utiliza el criterio de Akaike (AIC o AIC_1), aunque esa reducción no se refleja en una reducción del MISE_{TR} . Por otra parte, al estimar el parámetro de regresión funcional β_0 , los menores valores de $\text{Sesgo}_{\text{TR}}^2$ se obtienen con el criterio $BIC_{0.5}$ en particular, al considerar el MM-estimador. Un hecho destacable es que el M-estimador da origen a valores de MISE_{TR} menores a los del estimador clásico.

Cabe mencionar que la pérdida de eficiencia del MM-estimador respecto del estimador de mínimos cuadrados es menor al considerar el criterio BIC_1 que cualquiera de los otros criterios salvo para el HIC que da valores muy grandes de error cuadrático medio. Por otra parte, para el criterio $BIC_{0.5}^*$ el MISE_{TR} del MM-estimador del parámetro de regresión funcional es el triple que para el estimador de mínimos cuadrados, mientras que en el caso del criterio $BIC_{0.5}$ aumenta solo una tercera parte, produciendo una pérdida de eficiencia del 28%. Análogamente, cuando $\alpha = 1$, el MM-estimador utilizando el criterio BIC_α tiene mejor eficiencia respecto del estimador clásico el criterio que cuando se utiliza BIC_α^* . Como se mencionó anteriormente, el estimador de η_0 presenta un comportamiento más estable para los distintos criterios y la pérdida de eficiencia del MM-estimador es del 5% al utilizar el criterio BIC_α . Cabe mencionar que al tomar $\alpha = 0.5$ el MISE_{TR} del estimador de β_0 es al menos 3.3 veces mas grande que cuando consideramos $\alpha = 1$, mientras que los valores de $\text{MISE}_{\text{TR}}(\hat{\eta})$ usando BIC_1 sólo son un 25% más grandes. Por esta razón, en la Sección 11.1.3 se reporta el comportamiento de los estimadores bajo contaminación cuando se utiliza el criterio BIC_1 para elegir la dimensión de las bases de B-splines.

Conclusiones similares se obtienen para el Modelo 2, como se deduce de los resultados dados en

la Tabla 11.2 y Figura 11.4.

Estimador	Criterio	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		Sesgo _{TR} ²	MISE _{TR}	Sesgo _{TR} ²	MISE _{TR}	Sesgo _{TR} ²	MISE _{TR}
CL	<i>AIC</i>	0.0019	0.4945	0.0038	0.0451	0.0059	0.0305
	<i>AIC</i> [*]	0.0019	0.4945	0.0038	0.0451	0.0059	0.0305
	<i>BIC</i> _{0.5}	0.0015	0.2195	0.0064	0.0473	0.0079	0.0330
	<i>BIC</i> ₁	0.0018	0.0859	0.0198	0.0621	0.0190	0.0448
	<i>BIC</i> _{0.5} [*]	0.0015	0.2195	0.0064	0.0473	0.0079	0.0330
	<i>BIC</i> ₁ [*]	0.0018	0.0859	0.0198	0.0621	0.0190	0.0448
	<i>HIC</i>	0.0047	1.5175	0.0062	0.0477	0.0077	0.0335
M	<i>AIC</i>	0.0012	0.3791	0.0045	0.0471	0.0066	0.0324
	<i>AIC</i> [*]	0.0017	0.0947	0.0157	0.0580	0.0154	0.0414
	<i>BIC</i> _{0.5}	0.0013	0.1944	0.0081	0.0500	0.0091	0.0352
	<i>BIC</i> ₁	0.0017	0.0863	0.0248	0.0687	0.0231	0.0498
	<i>BIC</i> _{0.5} [*]	0.0017	0.0833	0.0271	0.0713	0.0249	0.0519
	<i>BIC</i> ₁ [*]	0.0016	0.0795	0.0764	0.0989	0.0586	0.0741
	<i>HIC</i>	0.0032	1.4050	0.0083	0.0507	0.0093	0.0358
MM	<i>AIC</i>	0.0023	0.6798	0.0041	0.0477	0.0063	0.0323
	<i>AIC</i> [*]	0.0024	0.9959	0.0058	0.0504	0.0076	0.0345
	<i>BIC</i> _{0.5}	0.0012	0.2940	0.0064	0.0497	0.0079	0.0344
	<i>BIC</i> ₁	0.0018	0.0876	0.0187	0.0633	0.0181	0.0454
	<i>BIC</i> _{0.5} [*]	0.0019	0.6485	0.0075	0.0523	0.0089	0.0367
	<i>BIC</i> ₁ [*]	0.0019	0.1314	0.0197	0.0667	0.0189	0.0482
	<i>HIC</i>	0.0048	1.6940	0.0062	0.0503	0.0078	0.0349

Tabla 11.1: Medidas resumen podadas bajo C_0 para todos los criterios para el Modelo 1.

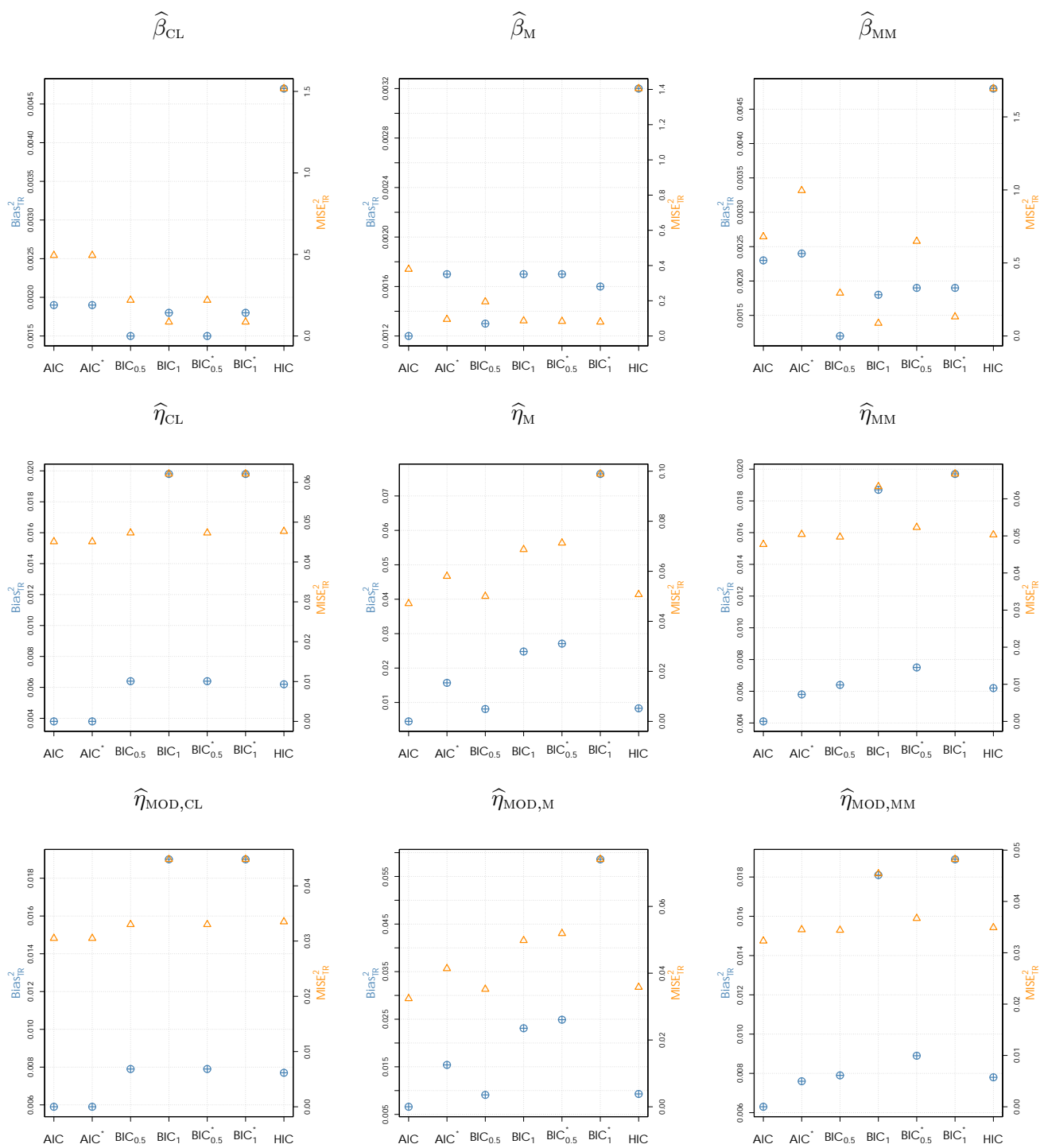


Figura 11.3: Medidas resumen podadas para todos los criterios considerados bajo C_0 para el Modelo 1.

Estimador	Criterio	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		Sesgo $^2_{\text{TR}}$	MISE $_{\text{TR}}$	Sesgo $^2_{\text{TR}}$	MISE $_{\text{TR}}$	Sesgo $^2_{\text{TR}}$	MISE $_{\text{TR}}$
CL	<i>AIC</i>	0.0011	0.5003	0.0038	0.0452	0.0060	0.0304
	<i>AIC*</i>	0.0011	0.5003	0.0038	0.0452	0.0060	0.0304
	<i>BIC</i>	0.0004	0.2169	0.0063	0.0472	0.0077	0.0328
	<i>BIC</i> ₁	0.0001	0.0843	0.0195	0.0620	0.0185	0.0443
	<i>BIC*</i>	0.0004	0.2169	0.0063	0.0472	0.0077	0.0328
	<i>BIC*</i> ₁	0.0001	0.0843	0.0195	0.0620	0.0185	0.0443
	<i>HIC</i>	0.0060	1.5257	0.0062	0.0476	0.0076	0.0333
M	<i>AIC</i>	0.0007	0.5176	0.4958	0.5180	0.4839	0.5045
	<i>AIC*</i>	0.0004	0.2808	0.4975	0.5210	0.4853	0.5070
	<i>BIC</i>	0.0004	0.2682	0.4975	0.5209	0.4853	0.5070
	<i>BIC</i> ₁	0.0001	0.1419	0.5071	0.5342	0.4927	0.5179
	<i>BIC*</i>	0.0001	0.1675	0.5008	0.5263	0.4878	0.5114
	<i>BIC*</i> ₁	0.0002	0.1303	0.5180	0.5454	0.5017	0.5268
	<i>HIC</i>	0.0060	1.9632	0.4975	0.5212	0.4851	0.5071
MM	<i>AIC</i>	0.0015	0.6646	0.0040	0.0473	0.0061	0.0321
	<i>AIC*</i>	0.0015	0.9797	0.0057	0.0502	0.0074	0.0342
	<i>BIC</i>	0.0006	0.2924	0.0064	0.0496	0.0078	0.0343
	<i>BIC</i> ₁	0.0001	0.0865	0.0179	0.0624	0.0173	0.0446
	<i>BIC*</i>	0.0010	0.6283	0.0073	0.0526	0.0087	0.0367
	<i>BIC*</i> ₁	0.0002	0.1220	0.0185	0.0655	0.0178	0.0473
	<i>HIC</i>	0.0041	1.7205	0.0061	0.0505	0.0076	0.0348

Tabla 11.2: Medidas resumen podadas bajo C_0 para todos los criterios para el Modelo 2.

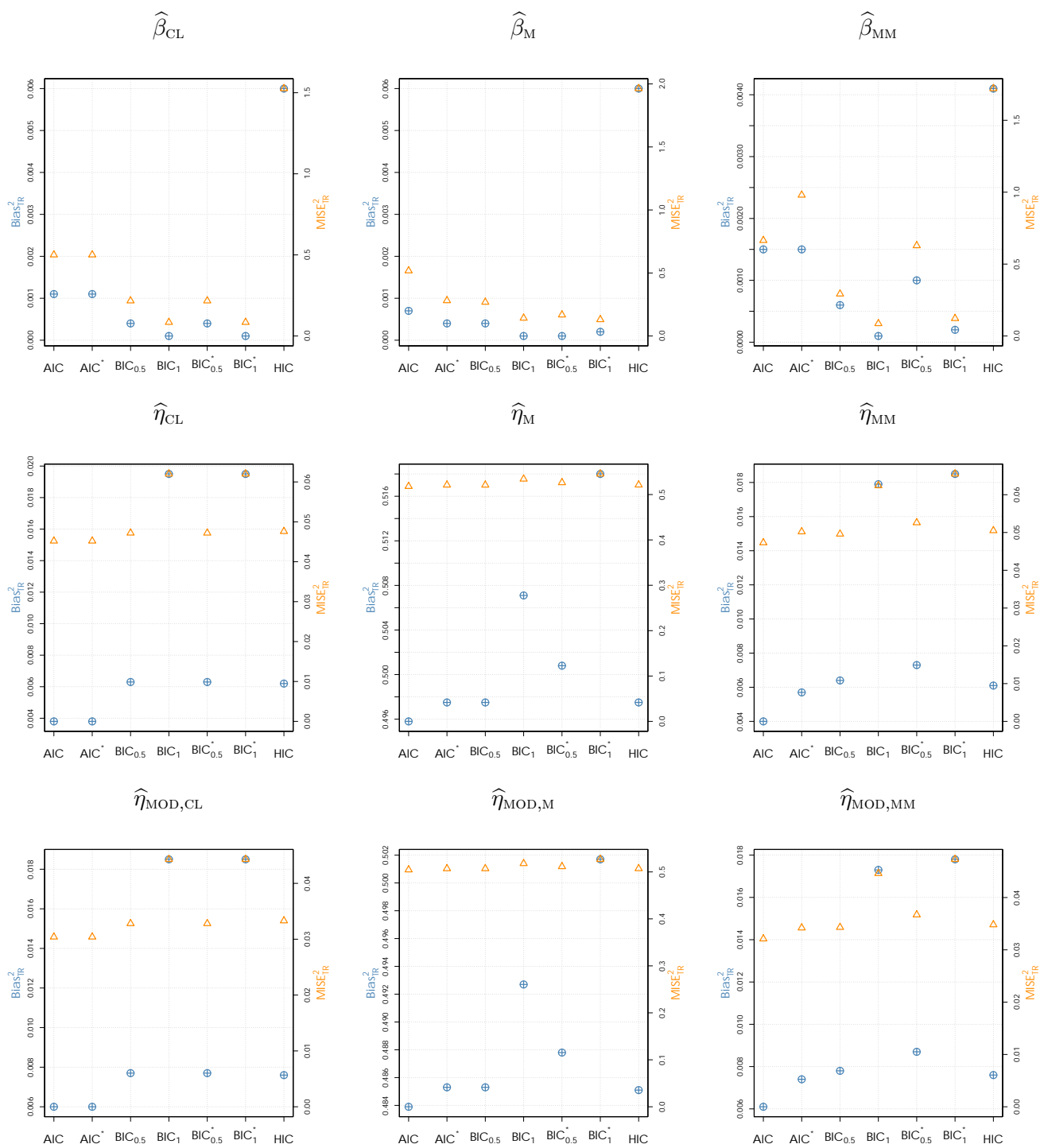


Figura 11.4: Medidas resumen podadas para todos los criterios considerados bajo C_0 para el Modelo 2.

11.1.3. Resultados para muestras contaminadas

En base a las conclusiones dadas en la Sección 11.1.2, se reportan los resultados obtenidos cuando los parámetros de suavizado $p_1 = k_{n,\beta}$ y $p_2 = k_{n,\eta}$ fueron elegidos mediante el criterio BIC_1 dado en (9.15), es decir, las dimensiones de las bases variaron en los rangos $4 \leq p_1 \leq 13$ y $4 \leq p_2 \leq 13$ y se eligieron aquellos valores que minimizaban

$$BIC_1(p_1, p_2) = \log \left(\hat{\sigma}^2 \frac{1}{n} \sum_{i=1}^n \rho \left(\frac{r_{i,p_1,p_2}}{\hat{\sigma}} \right) \right) + \frac{\log n}{n} (p_1 + p_2).$$

Las Tablas 11.3 y 11.4 reportan, para el Modelo 1 y el Modelo 2, respectivamente, las medidas resumen $\text{Sesgo}_{\text{TR}}^2$ y MISE_{TR} de los estimadores clásicos, el M-estimador y el MM-estimador indicados por CL, M y MM, respectivamente.

	Estimador	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		$\text{Sesgo}_{\text{TR}}^2$	MISE_{TR}	$\text{Sesgo}_{\text{TR}}^2$	MISE_{TR}	$\text{Sesgo}_{\text{TR}}^2$	MISE_{TR}
C_0	CL	0.0018	0.0859	0.0198	0.0621	0.0190	0.0448
	M	0.0017	0.0863	0.0248	0.0687	0.0231	0.0498
	MM	0.0018	0.0876	0.0187	0.0633	0.0181	0.0454
C_1	CL	0.0038	1.1129	1.3964	1.7289	1.4633	1.6551
	M	0.0022	0.2559	0.2836	0.4176	0.2864	0.3757
	MM	0.0013	0.1260	0.0237	0.0733	0.0223	0.0526
C_2	CL	2.7805	2.9177	0.0516	0.1108	0.0446	0.0807
	M	2.8108	2.9438	0.0673	0.1178	0.0551	0.0872
	MM	0.0763	0.8097	0.0325	0.1078	0.0338	0.0846

Tabla 11.3: Medidas resumen podadas para los estimadores $\hat{\beta}$, $\hat{\eta}$ y $\hat{\eta}_{\text{MOD}}$ de β_0 y η_0 , bajo el Modelo 1, calculados sin restricciones usando B-splines y cuando el estimador de η_0 es calculado a partir de $\hat{\eta}$ usando la modificación monótona descrita en la Sección 9.2.2.

Como es de esperar, para las muestras sin contaminar el comportamiento de todos los estimadores es similar, mientras que para ambos escenarios de contaminación, el estimador clásico muestra su sensibilidad frente a datos atípicos. En particular, bajo C_1 el MISE_{TR} de ambas componentes aumenta drásticamente con un fuerte incremento del sesgo cuadrado en la estimación de la componente no paramétrica η_0 . Por otro lado, bajo C_2 el efecto es mayor para la estimación de la pendiente funcional nuevamente debido a un gran aumento del sesgo cuadrado. Vale la pena mencionar que bajo C_0 , los valores de $\text{Sesgo}_{\text{TR}}^2(\hat{\beta})$ son mucho menores para el Modelo 2 que para el Modelo 1 ya que para el primero la función β_0 se aproxima mejor por splines por ser un polinomio.

En contraste, el M-estimador definido en Huang *et al.* (2015) triplica el MISE_{TR} bajo C_1 respec-

	Estimador	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		Sesgo $^2_{\text{TR}}$	MISE $_{\text{TR}}$	Sesgo $^2_{\text{TR}}$	MISE $_{\text{TR}}$	Sesgo $^2_{\text{TR}}$	MISE $_{\text{TR}}$
C_0	CL	0.0001	0.0843	0.0195	0.0620	0.0185	0.0443
	M	0.0001	0.1419	0.5071	0.5342	0.4927	0.5179
	MM	0.0001	0.0865	0.0179	0.0624	0.0173	0.0446
C_1	CL	0.0018	1.0967	1.4071	1.7394	1.4745	1.6665
	M	0.0004	0.2506	0.2843	0.4175	0.2874	0.3760
	MM	0.0004	0.1220	0.0227	0.0724	0.0214	0.0518
C_2	CL	2.7536	2.8855	0.0503	0.1101	0.0436	0.0799
	M	2.7865	2.9285	0.0658	0.1174	0.0539	0.0864
	MM	0.0758	0.8297	0.0316	0.1076	0.0329	0.0845

Tabla 11.4: Medidas resumen podadas para los estimadores $\hat{\beta}$, $\hat{\eta}$ y $\hat{\eta}_{\text{MOD}}$ de β_0 y η_0 , bajo el Modelo 2, calculados sin restricciones usando B-splines y cuando el estimador de η_0 es calculado a partir de $\hat{\eta}$ usando la modificación monótona descrita en la Sección 9.2.2.

to del obtenido bajo C_0 cuando se estima el parámetro de regresión funcional y lo multiplica por seis al estimar la componente no paramétrica. Sin embargo, como esperábamos a partir del diseño de los esquemas de contaminación, bajo C_2 el sesgo cuadrado de $\hat{\beta}$ aumenta considerablemente comprometiendo la estimación y dando origen a estimaciones poco fiables. El impacto de los datos atípicos en las covariables X no es atenuado por el M-estimador por lo que se arruinan las estimaciones $\hat{\beta}$.

Los MM-estimadores que constituyen la propuesta de esta tesis tiene un desempeño notoriamente mejor que los anteriormente mencionados bajo ambas contaminaciones. En particular, bajo C_1 los valores de Sesgo $^2_{\text{TR}}$ y MISE $_{\text{TR}}$ del M-estimador de η_0 son al menos 5 veces mayores que los MM-estimador, aun después de haber realizado la modificación monótona descrita en la Sección 9.2.2. Para el estimador de β_0 , la diferencia entre el M-estimador y el MM-estimador es menor aunque se observa una mejoría en términos del Sesgo $^2_{\text{TR}}$. Si bien el MISE $_{\text{TR}}$ del MM-estimador de β_0 se multiplica por nueve bajo C_2 con respecto al obtenido para muestras sin contaminar, su valor es la tercera parte del que arrojan los estimadores clásicos y el M-estimador quienes por otra parte dan origen a estimaciones muy sesgadas en base a los valores reportados de Sesgo $^2_{\text{TR}}$. En este sentido, el MM-estimador aumenta su variabilidad más que su sesgo, aunque el efecto en el sesgo parece ser mayor bajo el Modelo 2.

En todos los casos, cuando el estimador de η_0 es calculado a partir de $\hat{\eta}$ usando la modificación monótona descrita en la Sección 9.2.2, el MISE $_{\text{TR}}$ disminuye respecto del obtenido para el estimador sin restricciones.

11.2. Aplicación a un conjunto de datos reales: TECATOR

En esta sección, retomaremos el conjunto de datos reales TECATOR comentado ya en la Sección 8. Recordemos que el conjunto de datos TECATOR¹ corresponde a un estudio de control de calidad en la industria alimenticia donde se estudiaron 215 muestras de carne. Por cada unidad se observa el contenido de grasa, de proteínas y de agua y se registra una curva espectrométrica, \mathcal{X} , correspondiente a la absorbancia medida en 100 longitudes de onda. Intentamos predecir el contenido de grasa, que será nuestra respuesta y , utilizando como variables explicativas el contenido de agua, que indicaremos v , el de proteínas (z) y alguna característica de la curva espectrométrica que indicaremos por X .

Aneiros-Pérez y Vieu (2006) recomiendan usar un modelo donde (z, v) aparezca como covariable en la regresión lineal, mientras que la curva espectrométrica \mathcal{X} aparece modelada con una componente no paramétrica $y = \beta_{10} z + \beta_{20} v + \eta_0(\mathcal{X}) + \sigma_0 \epsilon$ y la cercanía entre curvas se mide a través de la distancia L^2 de las derivadas segundas.

Huang *et al.* (2015) estudiaron varios modelos donde X es la derivada segunda de la curva espectrométrica \mathcal{X} . En todos sus modelos, X aparece como covariable en una componente de regresión lineal funcional mientras que las componentes de (z, v) aparecen mediante un término no paramétrico aditivo $y = \langle X, \beta_0 \rangle + \eta_{0,1}(z) + \eta_{0,2}(v) + \sigma_0 \epsilon$, o bien, mediante un término de interacción entre contenido de los contenidos de proteína y agua $y = \gamma_0 + \langle X, \beta_0 \rangle + v \eta_0(z) + \sigma_0 \epsilon$. Para nuestro análisis nos inclinamos por la segunda variante, es decir, consideraremos el siguiente modelo

$$y = \gamma_0 + \langle X, \beta_0 \rangle + v \eta_0(z) + \sigma_0 \epsilon. \quad (11.2)$$

El objetivo de nuestro análisis no es alcanzar un estudio completo de este conjunto de datos, sino que deseamos ilustrar como nuestro procedimiento puede ser utilizado para detectar datos anómalos y proveer estimaciones confiables. Los MM-estimadores se calcularon como descrito en la Sección 11.1. Las dimensiones de las bases se eligieron utilizando el criterio BIC_1 .

Antes de elegir nuestras muestras de entrenamiento y de testeo, queremos discutir el comportamiento de las trayectorias para detectar posibles datos atípicos entre ellas. Para ello consideramos el boxplot funcional introducido en Sun y Genton (2011). La Figura 11.5 da el boxplot funcional de la derivada segunda de la absorbancia. Los datos detectados como outliers por este procedimiento se muestran en rojo en el gráfico y corresponden a las curvas identificadas como 7, 34, 35, 43, 44, 45, 129, 140, 172, 186 y 215.

Dividimos la muestra en un conjunto de entrenamiento correspondiendo las primeras 155 observaciones, con índices en \mathcal{I} , y uno de testeo con las 60 restantes observaciones, con índices en \mathcal{J} .

¹Disponible en la librería de R `fda.usc` o en el sitio <http://lib.stat.cmu.edu/datasets/tecator>.

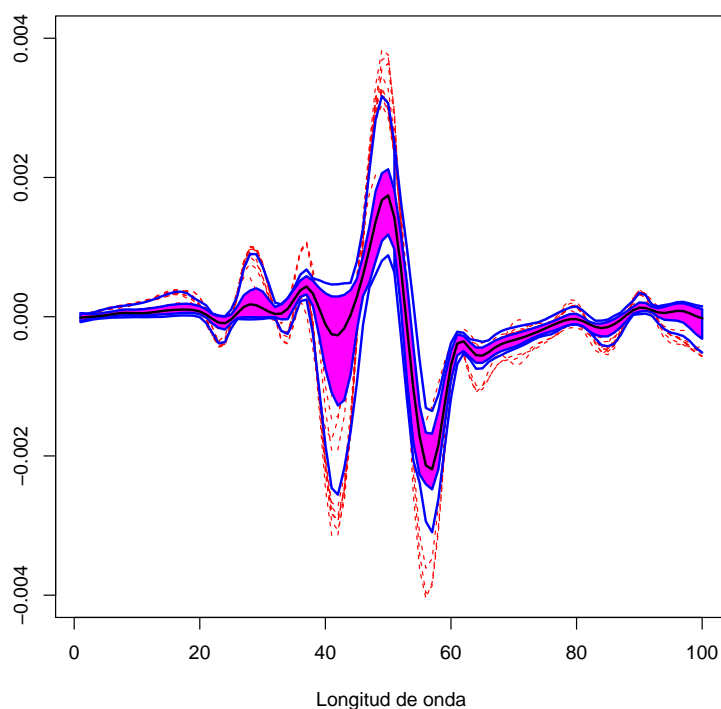


Figura 11.5: Boxplot funcional de la derivada segunda de la absorbancia, Sun y Genton (2011). Se advierte la presencia de datos atípicos en la covariable funcional.

Dada $s_{\mathcal{J}} = \text{MAD}_{j \in \mathcal{J}}(y_j)$, como en Boente y Vahnovan (2017), definimos como medidas de bondad de ajuste a la media y mediana de los errores cuadráticos de predicción sobre el conjunto de testeo, definidas respectivamente por

$$MSPE = \frac{1}{n_{\mathcal{J}}} \sum_{j \in \mathcal{J}} \frac{(y_i - \hat{y}_i)^2}{s_{\mathcal{J}}^2} \quad \text{y} \quad MedPE = \frac{\text{mediana}(y_i - \hat{y}_i)^2}{s_{\mathcal{J}}^2}.$$

La Tabla 11.5 reporta la media y mediana de los errores cuadráticos de predicción para el ajuste clásico y robusto.

	CL	MM
$MSPE$	0.00252	0.00504
$MedSPE$	0.00095	0.00084

Tabla 11.5: Errores de predicción para el ajuste clásico y robusto.

Las diferencias que se observan entre la media y la mediana sugieren la presencia de posibles observaciones atípicas ya sea en el conjunto de entrenamiento o en el de testeo. Esto explica también el mejor ajuste obtenido por el método robusto al tomar la mediana del error cuadrático de predicción $MedSPE$. Algunos datos atípicos en el conjunto de entrenamiento podrían influenciar la

estimación de β_0 y η_0 al usar los estimadores clásicos, así como valores atípicos de las respuestas en la muestra de testeo agranda los valores del $MSPE$ del estimador robusto, este hecho es análogo a lo que ocurre en regresión con el efecto de los así llamados outliers verticales.

Para evaluar la capacidad de predicción de datos no atípicos, como en Boente y Vahnovan (2017), calculamos sobre el conjunto de testeo el error de predicción de la muestra *limpia*. Si indicamos a las observaciones atípicas como

$$\gamma_i = \begin{cases} 1 & \text{si } (y_i, X_i, v_i, z_i) \text{ corresponde a un dato atípico detectado} \\ 0 & \text{si no,} \end{cases}$$

definimos

$$MSPE_{\text{CLEAN}} = \frac{1}{n_{\mathcal{J}} - \sum_{i \in \mathcal{J}} \gamma_i} \sum_{j \in \mathcal{J}} (1 - \gamma_j) \frac{(y_j - \hat{y}_j)^2}{s_{\mathcal{J}}^2}$$

Para detectar las observaciones atípicas en ambos conjuntos analizamos los residuos del ajuste robusto. Las Figuras 11.6(a) y 11.6(b) muestran los boxplots de los residuos en el conjunto de entrenamiento y testeo, respectivamente. Para el conjunto de entrenamiento, las observaciones 28, 29, 31, 34, 86, 89, 122, 140, 141 y 143 se detectan como posibles datos atípicos. Para el conjunto de testeo hay cinco residuos con valores atípicos correspondientes a las observaciones 177, 180, 181, 184 y 185.

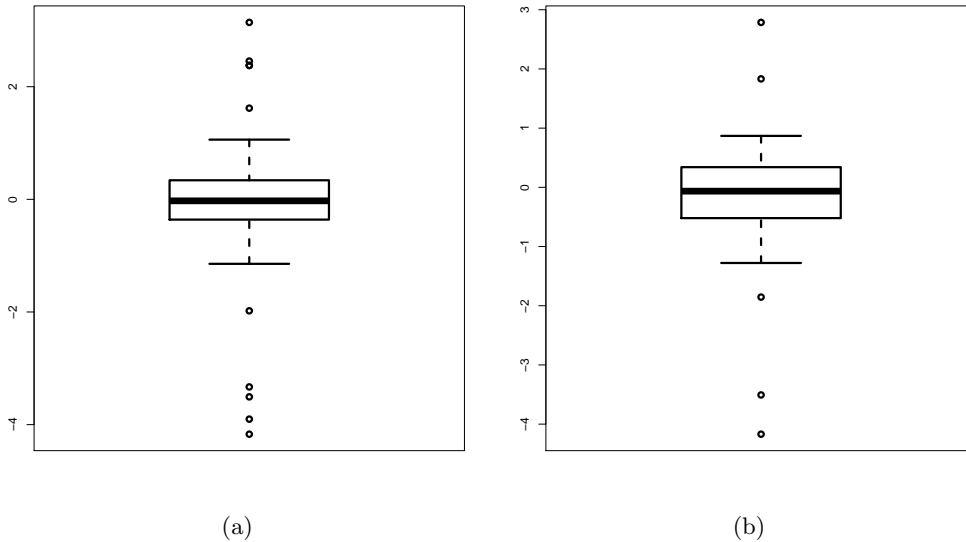


Figura 11.6: Boxplot de los residuos en el conjunto de entrenamiento (a) y de testeo (b)

El $MSPE_{\text{CLEAN}}$ fue calculado eliminando, por lo tanto, las 5 observaciones atípicas del conjunto de testeo detectadas por el boxplot de los residuos. Los valores del $MSPE_{\text{CLEAN}}$ cuando se utiliza el estimador clásico y el MM-estimador se reportan en la Tabla 11.6 junto con los resultados obtenidos al aplicar el procedimiento clásico al conjunto de entrenamiento sin las 10 observaciones identificadas como atípicas. Denotamos a este último procedimiento CL^{-out} .

	CL	CL ^{-out}	MM
$MSPE$	0.00252	0.00512	0.00504
$MedSPE$	0.00095	0.00087	0.00084
$MSPE_{CLEAN}$	0.00128	0.00151	0.00149

Tabla 11.6: Efectos de los 14 datos atípicos en el conjunto de entrenamiento y test.

Como puede observarse, los resultados obtenidos muestran que, como ha sido ampliamente discutido, el estimador de mínimos cuadrados trata de dar un compromiso entre observaciones atípicas y no atípicas. Después de haber eliminado los datos que se detectaron como atípicos, el procedimiento clásico lleva a resultados similares a los obtenidos con el MM-estimador, lo que confirma la utilidad de los procedimientos robustos tanto para proveer métodos de inferencia confiables como para identificar potenciales datos atípicos.

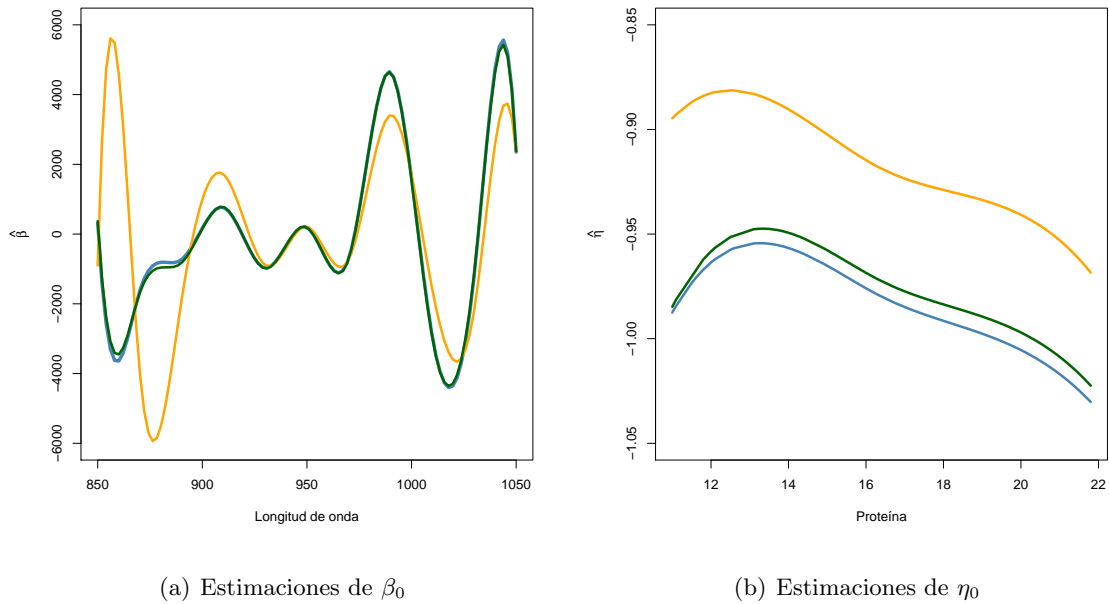


Figura 11.7: Estimaciones para β_0 y η_0 . Las líneas azul, naranja y verde oscuro corresponden al ajuste robusto, clásico y al ajuste clásico sin los datos atípicos del conjunto de entrenamiento, respectivamente.

Las Figuras 11.7(a) y 11.7(b) muestran, respectivamente, las estimaciones de η_0 y β_0 obtenidas mediante el procedimiento clásico (en líneas color naranja), el robusto (en azul) y el clásico sin los datos atípicos (en verde oscuro). En ambos casos, la estimación robusta no se ve afectada por los datos atípicos del conjunto de entrenamiento y da resultados similares a los del estimador clásico una vez que los datos atípicos son eliminados de la muestra. En particular, la estimación clásica de β_0 se ve seriamente afectada para valores pequeños de longitudes de onda, mientras que la de η_0 muestra un corrimiento respecto de las estimaciones obtenidas mediante el MM-estimador y el

estimador de mínimos cuadrados basado en la muestra sin datos atípicos.

11.3. Apéndice: Tablas sin poda.

Las Tablas 11.7, 11.8, 11.9 y 11.10, que reportamos a continuación, contienen las medidas resumen sin poda correspondientes a las Tablas 11.1, 11.2, 11.3 y 11.4, respectivamente, presentadas en las Secciones 11.1.2 y 11.1.3. Las conclusiones que se pueden extraer de las mismas son análogas a las descritas para las medidas $\text{Sesgo}_{\text{TR}}^2$ y MISE_{TR} .

Estimador	Criterio	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		Sesgo ²	MISE	Sesgo ²	MISE	Sesgo ²	MISE
CL	<i>AIC</i>	0.0110	0.8305	0.0038	0.0620	0.0106	0.0416
	<i>AIC</i> *	0.0110	0.8305	0.0038	0.0620	0.0106	0.0416
	<i>BIC</i> _{0.5}	0.0104	0.3814	0.0065	0.0636	0.0121	0.0436
	<i>BIC</i> ₁	0.0122	0.1515	0.0201	0.0783	0.0225	0.0536
	<i>BIC</i> _{0.5} *	0.0104	0.3814	0.0065	0.0636	0.0121	0.0436
	<i>BIC</i> ₁ *	0.0122	0.1515	0.0201	0.0783	0.0225	0.0536
	<i>HIC</i>	0.0193	2.4368	0.0064	0.0649	0.0120	0.0444
M	<i>AIC</i>	0.0089	0.6292	0.0046	0.0646	0.0113	0.0439
	<i>AIC</i> *	0.0118	0.1660	0.0159	0.0745	0.0192	0.0512
	<i>BIC</i> _{0.5}	0.0093	0.3380	0.0083	0.0663	0.0133	0.0460
	<i>BIC</i> ₁	0.0119	0.1530	0.0254	0.0855	0.0266	0.0587
	<i>BIC</i> _{0.5} *	0.0117	0.1458	0.0279	0.0884	0.0285	0.0607
	<i>BIC</i> ₁ *	0.0112	0.1386	0.0856	0.1169	0.0620	0.0801
	<i>HIC</i>	0.0128	2.2454	0.0085	0.0676	0.0136	0.0468
MM	<i>AIC</i>	0.0108	1.1058	0.0041	0.0659	0.0113	0.0443
	<i>AIC</i> *	0.0113	1.6008	0.0059	0.0683	0.0124	0.0462
	<i>BIC</i> _{0.5}	0.0089	0.4874	0.0065	0.0670	0.0124	0.0459
	<i>BIC</i> ₁	0.0122	0.1529	0.0189	0.0803	0.0219	0.0551
	<i>BIC</i> _{0.5} *	0.0111	1.0764	0.0076	0.0700	0.0134	0.0480
	<i>BIC</i> ₁ *	0.0112	0.2251	0.0201	0.0844	0.0228	0.0578
	<i>HIC</i>	0.0169	2.7118	0.0063	0.0681	0.0125	0.0468

Tabla 11.7: Medidas resumen sin podar bajo C_0 para todos los criterios para el Modelo 1.

Estimador	Criterio	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		Sesgo ²	MISE	Sesgo ²	MISE	Sesgo ²	MISE
CL	<i>AIC</i>	0.0019	0.8322	0.0038	0.0626	0.0106	0.0416
	<i>AIC*</i>	0.0019	0.8322	0.0038	0.0626	0.0106	0.0416
	<i>BIC</i> _{0.5}	0.0006	0.3685	0.0064	0.0639	0.0120	0.0435
	<i>BIC</i> ₁	0.0003	0.1389	0.0198	0.0785	0.0221	0.0533
	<i>BIC*</i> _{0.5}	0.0006	0.3685	0.0064	0.0639	0.0120	0.0435
	<i>BIC*</i> ₁	0.0003	0.1389	0.0198	0.0785	0.0221	0.0533
	<i>HIC</i>	0.0124	2.4311	0.0063	0.0652	0.0119	0.0443
M	<i>AIC</i>	0.0008	0.8150	0.6484	0.6817	0.6312	0.6602
	<i>AIC*</i>	0.0005	0.4497	0.6475	0.6829	0.6294	0.6602
	<i>BIC</i> _{0.5}	0.0005	0.4262	0.6475	0.6828	0.6294	0.6602
	<i>BIC</i> ₁	0.0002	0.2287	0.6457	0.6877	0.6225	0.6588
	<i>BIC*</i> _{0.5}	0.0002	0.2745	0.6463	0.6851	0.6263	0.6600
	<i>BIC*</i> ₁	0.0002	0.2098	0.6476	0.6898	0.6194	0.6555
	<i>HIC</i>	0.0100	3.2260	0.6474	0.6831	0.6291	0.6601
MM	<i>AIC</i>	0.0034	1.0768	0.0040	0.0662	0.0111	0.0441
	<i>AIC*</i>	0.0034	1.5625	0.0058	0.0685	0.0122	0.0459
	<i>BIC</i> _{0.5}	0.0007	0.4732	0.0065	0.0671	0.0124	0.0457
	<i>BIC</i> ₁	0.0003	0.1406	0.0181	0.0797	0.0211	0.0543
	<i>BIC*</i> _{0.5}	0.0015	1.0435	0.0074	0.0706	0.0132	0.0481
	<i>BIC*</i> ₁	0.0003	0.2043	0.0189	0.0835	0.0217	0.0569
	<i>HIC</i>	0.0088	2.7495	0.0062	0.0688	0.0124	0.0466

Tabla 11.8: Medidas resumen sin podar bajo C_0 para todos los criterios para el Modelo 2.

	Estimador	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		Sesgo ²	MISE	Sesgo ²	MISE	Sesgo ²	MISE
C_0	CL	0.0122	0.1515	0.0201	0.0783	0.0225	0.0536
	M	0.0119	0.1530	0.0254	0.0855	0.0266	0.0587
	MM	0.0122	0.1529	0.0189	0.0803	0.0219	0.0551
C_1	CL	0.0186	1.8468	1.5961	2.0783	1.6827	1.9357
	M	0.0132	0.4304	0.3189	0.5390	0.3402	0.4746
	MM	0.0093	0.2169	0.0244	0.0922	0.0265	0.0628
C_2	CL	3.4921	3.7270	0.0556	0.1375	0.0501	0.0931
	M	3.5309	3.7600	0.0739	0.1451	0.0605	0.0983
	MM	0.1127	1.1543	0.0327	0.1334	0.0403	0.1014

Tabla 11.9: Medidas resumen sin podar para los estimadores $\hat{\beta}$, $\hat{\eta}$ y $\hat{\eta}_{\text{MOD}}$ de β_0 y η_0 , bajo el Modelo 1, calculados sin restricciones usando B-splines y cuando el estimador de η_0 es calculado a partir de $\hat{\eta}$ usando la modificación monótona descrita en la Sección 9.2.2.

	Estimador	$\hat{\beta}$		$\hat{\eta}$		$\hat{\eta}_{\text{MOD}}$	
		Sesgo ²	MISE	Sesgo ²	MISE	Sesgo ²	MISE
C_0	CL	0.0003	0.1389	0.0198	0.0785	0.0221	0.0533
	M	0.0002	0.2287	0.6457	0.6877	0.6225	0.6588
	MM	0.0003	0.1406	0.0181	0.0797	0.0211	0.0543
C_1	CL	0.0028	1.8065	1.6087	2.0884	1.6958	1.9474
	M	0.0007	0.4114	0.3200	0.5388	0.3417	0.4743
	MM	0.0005	0.2022	0.0234	0.0917	0.0255	0.0621
C_2	CL	3.3690	3.5973	0.0542	0.1371	0.0491	0.0923
	M	3.4067	3.6500	0.0723	0.1451	0.0593	0.0978
	MM	0.0929	1.1589	0.0317	0.1340	0.0395	0.1018

Tabla 11.10: Medidas resumen sin podar para los estimadores $\hat{\beta}$, $\hat{\eta}$ y $\hat{\eta}_{\text{MOD}}$ de β_0 y η_0 , bajo el Modelo 2, calculados sin restricciones usando B-splines y cuando el estimador de η_0 es calculado a partir de $\hat{\eta}$ usando la modificación monótona descrita en la Sección 9.2.2.

Capítulo 12

Consideraciones finales

Desde su introducción por Engle *et al.* (1986), el modelo parcialmente lineal ha sido extensamente estudiado en la literatura estadística. El creciente interés por el desarrollo de técnicas basadas en datos funcionales ha suscitado varias propuestas que combinan enfoques tanto paramétricos como no paramétricos con regresión funcional resultando en modelos parcialmente lineales funcionales. Entre otros modelos, la propuesta que estudiamos considera un modelo con una componente no paramétrica η_0 y un término de regresión funcional con parámetro, de dimensión infinita, β_0 .

Uno de los enfoques tradicionales para estimar estas funciones consiste en desarrollar ambas componentes en ciertas bases y realizar inferencia sobre los coeficientes. Los B-splines han sido considerados para aproximar tanto la pendiente funcional β_0 como la componente no paramétrica η_0 . Una ventaja de los B-splines es que proveen un procedimiento de estimación que puede extenderse a la situación en que hay restricciones de monotonía en la componente no paramétrica imponiendo restricciones de no decrecimiento en los coeficientes. Para superar la sensibilidad del procedimiento clásico basado en funciones de pérdida no convexas hemos introducido una familia de estimadores robustos para ambas componentes del modelo. Mediante una monotonización del estimador robusto de B-splines sin restricciones de η_0 se estiman ambas componentes sin necesidad de recalcularse el estimador de β_0 . Una alternativa consiste en desarrollar el término funcional en la base de autofunciones del operador de covarianza. En esta línea, presentamos una propuesta robusta de componentes principales esféricas, el estudio detallado de sus propiedades asintóticas y del comportamiento para muestras finitas serán objeto de estudio futuro. Asimismo, evaluamos diferentes criterios para la selección del modelo.

Obtuvimos resultados de consistencia y tasas de convergencia tanto para los estimadores sin restricciones como para aquellos obtenidos imponiendo restricciones de orden a los coeficientes. Resultados para la versión monotonizada se deducen de los anteriores. El comportamiento inadecuado

del método clásico cuando hay datos atípicos en la muestra se confirma en nuestro estudio de simulación. El procedimiento robusto da origen estimadores más confiables conduciendo prácticamente a los mismos resultados para muestras con y sin datos atípicos.

Por último, ilustramos mediante un conjunto de datos reales la utilidad de los procedimientos robustos tanto para proveer métodos de inferencia confiables como para identificar potenciales datos atípicos.

Pablo Vena

Graciela Boente

Matías Salibián-Barrera

Bibliografía

- [1] Aït Sahalia, Y. (1995). *The delta method for nonparametric kernel functionals*. PhD. dissertation, University of Chicago.
- [2] Akaike, H. (1974). A new look at the statistical model identification. In Selected Papers of Hirotugu Akaike (pp. 215-222). Springer, New York, NY.
- [3] Álvarez, E. & Yohai, J. (2012). M-estimators for isotonic regression. *Journal of Statistical Planning and Inference*, **142**, 2241-2284.
- [4] Aneiros-Pérez G. & Vieu P. (2006). Semi-functional partial linear regression. *Statistics and Probability Letters*, **76**, 1102–1110.
- [5] Ansley, C. & Wecker, W. (1983). Extension and examples of the signal extraction approach to regression. In *Applied Time Series Analysis of Economic Data*, 181–192.
- [6] Bali, J. L., Boente, G., Tyler, D. E., & Wang, J. L. (2011). Robust functional principal components: A projection-pursuit approach. *Annals of Statistics*, **39**(6), 2852-2882.
- [7] Bhattacharya, P. K. & Zhao, P. L. (1997). Semiparametric inference in a partial linear model. *Annals of Statistics*, **25**, 244–262.
- [8] Bianco, A. & Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *Journal of Statistical Planning and Inference*, **122**, 229–252.
- [9] Bianco, A, Boente, G. & Rodrigues, I. (2013a). Resistant estimators in Poisson and Gamma models with missing responses and an application to outlier detection. *Journal of Multivariate Analysis*, **114**, 209-226.
- [10] Bianco, A, Boente, G. & Rodrigues, I. (2013b) Robust tests in generalized linear models with missing responses. *Computational Statistics and Data Analysis*, **65**, 80-97.
- [11] Bianco, A., García Ben, M. & Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *Canadian Journal of Statistics*, **33**, 511-528.

- [12] Birke, M., & Dette, H. (2007). Testing strict monotonicity in nonparametric regression. *Mathematical Methods of Statistics*, **16**(2), 110-123.
- [13] Boente, G. & Fraiman, R. (1988). Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, **29**, 180-198.
- [14] Boente, G., He, X. & Zhou, J. (2006). Robust estimates in generalized partially linear models. *Annals of Statistics*, **34**, 2856-2878.
- [15] Boente, G. & Martínez, A. (2017). Marginal integration M -estimators for additive models. *TEST*, **26**, 231-260.
- [16] Boente, G. & Rodríguez, D. (2010). Robust inference in generalized partially linear models. *Computational Statistics and Data Analysis*, **54**, 2942-2966.
- [17] Boente, G., Rodríguez, D., & Sued, M. (2019). The spatial sign covariance operator: Asymptotic results and applications. *Journal of Multivariate Analysis*, **170**, 115-128.
- [18] Boente, G., & Salibian-Barrera, M. (2015). S -estimators for functional principal component analysis. *Journal of the American Statistical Association*, **110**(511), 1100-1111.
- [19] Boente, G., Barrera, M. S., & Tyler, D. E. (2014). A characterization of elliptical distributions and some optimality properties of principal components for functional data. *Journal of Multivariate Analysis*, **131**, 254-264.
- [20] Boente, G. & Vahnovan, A. (2017). Robust estimators in semi-functional partial linear regression models. *Journal of Multivariate Analysis*, **154**, 59-84.
- [21] Cantoni, E. & Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, **96**, 1022-1030.
- [22] Cardot, H., Cénac, P., & Zitt, P. A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, **19**(1), 18-43.
- [23] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Annals of Statistics*, **16**, 136-146.
- [24] Chen, H. & Chen, K. (1991). Selection of the splined variables and convergence rates in a partial spline model. *Canadian Journal of Statistics*, **19**, 323-339.
- [25] Chen, H. & Shiau, J. (1991). A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning and Inference*, **25**, 187-201.

- [26] Chen, H. & Shiau, J. (1994). Data-driven efficient estimates for partially linear models. *Annals of Statistics*, **22**, 211–237.
- [27] Croux, C. & Haesbroeck, G. (2002). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics and Data Analysis*, **44**, 273–295.
- [28] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, **147**, 1–23.
- [29] Denby, L. (1986). Smooth regression functions. *Statistical Research Report*, **26**, AT and T Bell Laboratories, Murray Hill.
- [30] Dette, H., Neumeier, N. & Pilz, K.F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, **12**, 469–490.
- [31] Du, J., Sun, Z. & Xie, T. (2013). M-estimation for the partially linear regression model under monotonic constraints. *Statistics and Probability Letters*, **83**, 1353–1363.
- [32] Engle, R. F., Granger, C. W., Rice, J., & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, **81**(394), 310–320.
- [33] Fan, Y.Q. & Li, Q. (1999). Root- n -consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics*, **11**, 251–269.
- [34] Ferraty, F. & Romain, Y. (2010). *The Oxford Handbook of Functional Data Analysis*, Oxford University Press, Oxford.
- [35] Ferraty, F., & Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17(4), 545–564.
- [36] Ferraty, F. & Vieu, Ph. (2006). *Nonparametric Functional data analysis: Theory and Practice*. Springer Series in Statistics, Springer, New York.
- [37] Gao, J. & Liang, H. (1995). Asymptotic normality of pseudo-LS estimator for partly linear autoregression models. *Statistics and Probability Letters*, **23**, 27–34.
- [38] Gao, J. & Zhao, L. (1993). Adaptive estimation in partly linear regression models. *Science in China, Series A*, **1**, 14–27.
- [39] Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, 95(3), 587–600.

- [40] Goia, A. & Vieu, P. (2016). An introduction to recent advances in high/infinite-dimensional statistics. *Journal of Multivariate Analysis*, **146**, 1–6.
- [41] Green, P., Jennison, C. and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B*, **47**, 299–315.
- [42] Hall, P. & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, **35**, 70–91.
- [43] Härdle, W., Liang, H. and Gao, J. (2000). *Partially linear models*. Springer-Verlag.
- [44] Härdle, W. & Tsybakov, A. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Annals of Statistics*, **16**, 120–135.
- [45] He, X. & Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, **3**, 299–308.
- [46] He, X. & Shi, P. (1996). Bivariate tensor-product B-spline in a partly linear model. *Journal of Multivariate Analysis*, **58**, 162–181.
- [47] He, X. & Shi, P. (1998). Monotone B-Spline smoothing. *Journal of the American Statistical Association*, **93**, 643–650.
- [48] He, X., Zhu, Z. and Fung, W. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, **89**, 579–590.
- [49] Heckman, N. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society, Series B*, **48**, 244–248.
- [50] Heritier, S., Cantoni, E., Copt, S., & Victoria-Feser, M. P. (2009). *Robust Methods in Biostatistics* (Vol. 825). John Wiley & Sons.
- [51] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York.
- [52] Hsing, T. & Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley, New York.
- [53] Huang, J. (2002). A note on estimating a partly linear model under monotonicity constraints. *Journal of Statistical Planning and Inference*, **107**, 343–351.
- [54] Huang, L., Wang, H., Cui, H. and Wang, S. (2015). Sieve M -estimator for a semi-functional linear model. *Science China, Mathematics*, **58**, 2421–2434.

- [55] Kolmogorov, A. N. & Tikhomirov, V. M. (1959). ϵ -entropy and ϵ -capacity of sets in function spaces. *Uspekhi Matematicheskikh Nauk*, **14**, 3-86. (In Russian.) [Published in English in (1961) American Mathematical Society Translation (2) 17 277-364.]
- [56] Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag, New York.
- [57] Lee, S., Shin, H., & Billor, N. (2013). M-type smoothing spline estimators for principal functions. *Computational Statistics & Data Analysis*, **66**, 89-100.
- [58] Lian H. (2011). Partial functional linear regression. *Journal of Nonparametric Statistics*, **23**, 115-128.
- [59] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T., Cohen, K. L., ... & Fan, J. (1999). Robust principal component analysis for functional data. *Test*, 8(1), 1-73.
- [60] Lu, M., Zhang, Y. & Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, **94**, 705-718.
- [61] Lu, M. (2010). Spline-based sieve maximum likelihood estimation in the partly linear model under monotonicity constraints. *Journal of Multivariate Analysis*, **101**, 2528-2542.
- [62] Lu, M. (2015). Spline estimation of generalised monotonic regression. *Journal of Nonparametric Statistics*, **27**, 19-39.
- [63] Mallows, C. (1974). On some topics in robustness. *Memorandum, Bell Laboratories, Murray Hill, N.J.*
- [64] Manchester, L. (1996). Empirical influence for robust smoothing. *Australian Journal of Statistics* **38**, 275-296.
- [65] Maronna, M. & Yohai, V. (2013). Robust functional linear regression based on splines. *Computational Statistics and Data Analysis*, **65**, 46-55.
- [66] Maronna, R., Martin, R., Yohai, V. & Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)*. Wiley, New York.
- [67] Neumeyer, N. (2007). A note on uniform consistency of monotone function estimators. *Statistics and Probability Letters*, **77**, 693-703.
- [68] Qingguo, T. (2015). Estimation for semi-functional linear regression. *Statistics*, **49**, 1262-1278.
- [69] Ramsay, J. (1988). Monotone regression splines in action. *Statistical Science*, **3**, 425-441.

- [70] Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*, 2nd edition. Springer, New York.
- [71] Rice, J. (1986). Convergence rates for partially splined models. *Statistics and Probability Letters*, **4**, 203–208.
- [72] Robinson, P. (1988). Root-n-consistent Semiparametric regression. *Econometrica*, **56**, 931–954.
- [73] Ronchetti, E. (1985). Robust model selection in regression. *Statistics and Probability Letters*, **3**, 21–23.
- [74] Salibian-Barrera, M., & Yohai, V. J. (2006). A fast algorithm for S-regression estimates. *Journal of computational and Graphical Statistics*, 15(2), 414-427.
- [75] Sawant, P., Billor, N., & Shin, H. (2012). Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, **27**(1), 83-102.
- [76] Schumaker, L. (1981). *Spline Functions: Basic Theory*, Wiley, New York.
- [77] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
- [78] Shen, X., & Wong, W. H. (1994). Convergence rate of sieve estimates. *Annals of Statistics*, 580-615.
- [79] Smucler, E. (2016). *Estimadores robustos para el modelo de regresión lineal con datos de alta dimensión*. Tesis doctoral Universidad de Buenos Aires. Disponible en <http://cms.dm.uba.ar/academico/carreras/doctorado/Tesis%20Smucler.pdf>
- [80] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, **50**, 413–436.
- [81] Stefanski, L. A., Carroll, R. J., & Ruppert, D. (1986). Optimally hounded score functions for generalized linear models with applications to logistic regression. *Biometrika*, **73**(2), 413-424.
- [82] Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, 590-606.
- [83] Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2), 316-334.
- [84] Sun,Z., Zhang,Z. & Du,J. (2012). Semiparametric analysis of isotonic errors-in-variables regression models with missing response. *Communications in Statistics: Theory and Methods*, **41**, 2034–2060.

- [85] Tamine, J. (2002). Smoothed influence function: another view at robust nonparametric regression. *Discussion paper 62* Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.
- [86] Tharmaratnam, K. & Claeskens, G. (2013). A comparison of robust versions of the AIC based on M -, S - and MM -estimators. *Statistics*, **47**, 216–235.
- [87] Van de Geer, S. (2000). *Empirical Processes in M -Estimation*, Cambridge University Press.
- [88] van der Vaart, A. (1998). *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- [89] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer-Verlag, New York.
- [90] Yee, T. and Wild, C. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society, Series B*, **58**, 481–493.
- [91] Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, **15** 642–656.
- [92] Zhang, Y., Hua, L. E. I., & Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics*, **37**(2), 338–354.
- [93] Zhou, J. & Chen M. (2012). Spline estimators for semi-functional linear model. *Statistics and Probability Letters*, **82**, 505–513.