Métodos robustos para seleccionar el parámetro de suavizado en modelos parcialmente lineales

Daniela Rodríguez

Directora: Dra. Graciela Boente

Universidad de Buenos Aires

Contenido

Capítulo 1

Introducción

La inferencia estadística comúnmente se focaliza sobre funciones de distribución que son puramente paramétricas o puramente noparamétricas. Un modelo paramétrico razonable produce inferencias precisas mientras que un modelo erróneo posiblemente conducirá a conclusiones equivocadas. Por otro lado, los modelos noparamétricos son asociados con alta estabilidad y menor precisión. Recientemente los modelos noparamétricos han ganado una importante atención en el estudio de fenómenos naturales con comportamiento de complejidad no lineal. Sea $(y_i, \mathbf{X}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}$ observaciones independientes idénticamente distribuidas (i.i.d) tal que $y_i \in \Re, t_i \in \Re$ y $\mathbf{X}_i \in \Re^p$ y

$$y_i = \gamma(\mathbf{X}_i, t_i) + \varepsilon_i \qquad 1 \leq i \leq n. \tag{1.1}$$

donde los errores ε_i son independientes e independientes de $(\mathbf{X}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}$.

El análisis del modelo (1.1) requiere de técnicas de suavizado multivariadas para la función γ. Por lo tanto este modelo encuentra en aplicaciones la dificultad conocida como "maldición de la dimensionalidad". En los últimos años, diversos autores han tratado el problema de reducción de la dimensión en los modelos de regresión noparamétrica. Para poder resolver el problema de vecindades vacías, se han dado varios métodos para estimar la función de regresión cuando las covariables permanecen en un espacio de dimensión alta. Hastie y Tibshirani (1990) introdujeron los modelos aditivos que generalizan los modelos lineales, resuelven el problema de "la maldición de la dimensión"y además son de fácil interpretación. Este nuevo planteo combina la flexibilidad de los modelos noparamétricos con la simple interpretación del modelo lineal estándar.

Durante la realización de este trabajo estudiaremos una estrategia intermedia, que es emplear un modelo semiparamétrico. Este modelo combina las ventajas de ambos métodos, paramétricos y noparamétricos. En este caso la función de regresión se descompone como $\gamma(\mathbf{X}, t) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + g(t)$, con lo cual (1.1) puede escribirse como

$$y_i = \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + g(t_i) + \varepsilon_i \qquad 1 \leq i \leq n. \tag{1.2}$$

Los modelos parcialmente lineales son más flexibles que los modelos lineales estándares ya que tienen una componente paramétrica y una noparamétrica. Estos modelos se convierten en una elección apropiada cuando sospechamos que la respuesta y depende linealmente de $\mathbf{X}$ pero es no lineal en t.

Engle (1986) introdujo el modelo (1.2) al estudiar el efecto del clima en la demanda de electricidad. Este modelo también fue aplicado en diversos problemas económicos (Daniel y Wood (1980)). Puede ser usado, por ejemplo para examinar el efecto del cambio de precio de un determinado producto sobre el volumen de ventas. Es de esperar que exista una relación entre el precio y el tiempo y en este caso, podemos suponer que $\mathbf{X} = \phi(t) + \eta$ donde ϕ es una función desconocida y η es una variable aleatoria. Este modelo fue tratado por diversos autores. Speckman (1988), Robinson (1988), Chen (1988), Rice (1986) fueron algunos de los primeros en trabajar en el tema.

Los estimadores estudiados para el modelo que consideramos requieren de la selección de una parámetro de suavizado h comúnmente llamado ventana. Linton (1995) obtuvo una expresión asintótica para la ventana óptima en el sentido que minimiza el error cuadrático medio del estimador del parámetro $\boldsymbol{\beta}$. Esta expresión depende de las funciones que estamos estimando y de parámetros que desconocemos, como $\boldsymbol{\beta}$ y los desvíos de los errores. Linton también propuso un estimador para la ventana óptima bajo el modelo (1.2) y Aneiros-Pérez y Quintela del Río (2002) lo extendieron para el caso de errores dependientes. Por otra parte, es conocido que la presencia de datos anómalos puede afectar seriamente los estimadores de la función g y del parámetro $\boldsymbol{\beta}$. Para evitar este problema, Bianco y Boente (2002) propusieron estimadores robustos para el parámetro y la función de regresión. Asimismo, la sensibilidad de los procedimientos clásicos de selección de ventana a datos atípicos ya fue observada en el modelo de regresión noparamétrica, o sea cuando $\boldsymbol{\beta} = 0$, por diversos autores entre ellos puede mencionarse a Leung, Marrot y Wu (1993); Wang y Scott (1994); Boente, Fraiman y Meloche (1997) y Cantoni y Roncheti (2001). En este trabajo, proponemos un estimador robusto del parámetro de suavizado, lo que resulta en estimadores adaptivos robustos de g y $\boldsymbol{\beta}$.

En el Capítulo 2 se tratará el modelo de interés, describiendo los procedimientos de estimación. Se dará un breve resúmen de los estimadores noparamétricos y se repasarán estimadores noparamétricos robustos para la función de regresión. En el Capítulo 3, se obtendrá la expresión de la ventana óptima para estimar el vector de parámetros del modelo. Para ello estandarizamos un estimador de dicho parámetro y obtenemos aproximaciones de sus dos primeros momentos, los cuales son usados para encontrar la expresión de la ventana óptima. En el Capítulo 4, se propondrá un estimador robusto del suavizador óptimo encontrado y mostraremos que aproxima en probabilidad a la ventana óptima. Finalmente en el Capítulo 5, se realizará un estudio de Monte Carlo para comparar para muestras pequeñas el comportamiento de las ventanas adaptivas y de los parámetros del modelo obtenidos a partir de ellas. En el Apéndice, se incluyen las demostraciones de algunos de los resultados.

Capítulo 2

Estimación del Modelo

2.1 Introducción

Supongamos que tenemos un conjunto de observaciones $\{(y_i, {\mathbf{X}_i}^{\mathrm{T}}, t_i)^{\mathrm{T}}\}_{i=1}^n$ de acuerdo al modelo

$$y_i = {\mathbf{X}_i}^{\mathrm{T}}\boldsymbol{\beta} + g(t_i) + \varepsilon_i \qquad 1 \leq i \leq n; \tag{2.1}$$

donde $\mathbf{X}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ son observaciones de un vector aleatorio, t_i son puntos de diseño fijos en $[0,1]$, $\boldsymbol{\beta}$ es un vector de parámetros desconocidos de dimensión $p, g(t)$ es una función escalar desconocida definida en $[0,1]$ y ε_i son errores independientes con media 0 en el caso clásico y simétricos en el caso robusto. Supondremos que $\varepsilon_i \sim F(./\sigma_\varepsilon)$ donde $\sigma_\varepsilon^2 = V(\varepsilon) = E(\varepsilon\varepsilon^{\mathrm{T}})$, si existe. Además, supondremos que existe una relación entre $\mathbf{X}_i$ y t_i. De este modo el modelo de interés puede escribirse como

$$\left\{ \begin{array}{rcll} y_i &=& {\mathbf{X}_i}^{\mathrm{T}}\boldsymbol{\beta} + g(t_i) + \varepsilon_i & \qquad 1 \leq i \leq n, \\ x_{ij} &=& \boldsymbol{\phi}_j(t_i) + \eta_{ij} & \qquad 1 \leq j \leq p, \end{array} \right. \tag{2.2}$$

con η_{ij} i.i.d, $\eta_{ij} \sim G(./\sigma_\eta)$ G simétrica. Tomando esperanza condicional respecto a t en la segunda ecuación de (2.2) obtenemos

$$\phi_j(t_i) = E(x_{ij}|t_i) \qquad 1 \leq j \leq p.$$

Llamemos $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))^{\mathrm{T}}$ y sea $\phi_0(t_i) = E(y_i|t_i)$ entonces

$$\phi_0(t_i) = \boldsymbol{\phi}^{\mathrm{T}}(t_i)\boldsymbol{\beta} + g(t_i). \tag{2.3}$$

Con lo cual, restando $\phi_0(t_i)$ a ambos lados de la primera ecuación de (2.2) el modelo se escribe

$$y_i - \phi_0(t_i) = (\mathbf{X}_i - \boldsymbol{\phi}(t_i))^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i. \tag{2.4}$$

Esto nos sugiere que si proponemos estimadores $\widehat{\phi}_0$ y $\widehat{\boldsymbol{\phi}}$ de ϕ_0 y $\boldsymbol{\phi}$ respectivamente, podemos obtener un estimador de $\boldsymbol{\beta}$. Por ejemplo, el estimador de mínimos cuadrados para $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_{LS}$ puede ser calculado como

$$\widehat{\boldsymbol{\beta}}_{LS} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} [(y_i - \widehat{\phi}_0(t_i)) - (\mathbf{X}_i - \widehat{\boldsymbol{\phi}}(t_i))^{\mathrm{T}} \boldsymbol{\beta}]^2. \tag{2.5}$$

Finalmente en base a (2.3), podemos estimar $g(t)$ de la siguiente forma,

$$\widehat{g}(t) = \widehat{\phi}_0(t) - \widehat{\boldsymbol{\phi}}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}.$$

2.2 Estimadores noparamétricos clásicos

En la sección anterior surgió la necesidad de obtener estimadores para

$$\begin{aligned} \phi_j(t_i) &= E(x_{ij}|t_i) \qquad 1 \leq j \leq p, \\ \phi_0(t_i) &= E(y_i|t_i). \end{aligned}$$

Una propuesta para este tipo de problemas son las técnicas de suavizados. Supongamos que tenemos una variable de respuesta o dependiente Y y una variable predictora o independiente T y buscamos describir la dependencia de la media de Y como función de T, es decir,

$$E(Y|T) = m(T).$$

Una técnica de suavizado es una herramienta que nos permite resumir la tendencia de la respuesta Y como función de una o más variables predictoras. Dado que no suponemos una forma rígida, como en el modelo lineal estándar, ese estimador es considerado como un estimador noparamétrico, donde se pretende que los datos sean los que describan el modelo. A continuación, daremos un repaso informal de algunos estimadores noparamétricos.

a) Promedio local

Trabaja dividiendo los valores de la variable predictora en regiones disjuntas y en cada una de estas regiones promedia los valores de la variable. Formalmente, elegimos puntos $c_0 < \ldots < c_K$ donde $c_0 = -\infty \quad y \quad c_K = +\infty$, y definimos $R_k = \{i : c_k \leq t_i < c_{k+1}\}; \quad k = 0, \ldots, K-1$. Si $t \in R_k$, definimos

$$\widehat{m}_n(t,h) = \frac{\sum\limits_{j \in R_k(t)} y_j}{\# R_k(t)}.$$

Observemos que la estimación resultante es discontinua y salvo que tengamos información para sospechar que la función de regresión sea discontinua en determinada región, no nos provee de una buena estimación.

b) Promedio Móvil

Este procedimiento está relacionado con el estimador anterior; la diferencia radica en que los intervalos pueden superponerse y se basan en la noción de vecinos más cercanos. Para cada punto t se promedian los valores de y asociados a un entorno de t, es decir,

$$\widehat{m}_n(t,h) = \frac{\sum\limits_{j \in N_k(t)} y_j}{\# N_k(t)}.$$

donde $N_k(t) = \{\max(i-k,1),\ldots,i-1,i,i+1,\ldots,\min(i+k,n)\}$, $N_k(t)$ es llamada vecindad. En este caso definimos una vecindad simétrica, pero podríamos tomar por ejemplo los r vecinos más cercanos a t.

Una generalización simple de este tipo de estimadores consiste en calcular un estimador de mínimos cuadrados ajustando un modelo de regresión simple en cada vecindad. Es decir,

$$\widehat{m}_n(t,h) = \widehat{\alpha}(t) + \widehat{\beta}(t)t.$$

Donde $\widehat{\alpha}(t)$ y $\widehat{\beta}(t)$ son los estimadores de mínimos cuadrados para el modelo de regresión simple $y_i = \alpha + \beta t_i + \varepsilon_i$ para los puntos de la vecindad $N_k(t)$.

c) Splines Cúbicos

En este caso, la estimación propuesta no es construida explícitamente, surge como solución a un problema de optimización. Donde se busca encontrar la función con derivada continua que minimice,

$$\sum_{i=1}^{n} \{y_i - m(t_i)\}^2 + \lambda \int_a^b \{m''(x)\}^2 dx.$$

El primer término mide la cercanía de los datos y el segundo penaliza la curvatura de la función. Se puede mostrar que esta ecuación tiene un único mínimo y que es un spline con nodos en los puntos t_i.

d) Suavizadores basados en núcleos.

Este tipo de estimadores, es definido como un conjunto de pesos locales para producir la estimación en cada punto t. En general los pesos usados decrecen de manera suave a medida que nos alejamos de t. La función de regresión es estimada como,

$$\widehat{m}_n(t,h) = \sum_{i=1}^{n} w_{ni}(t,h) y_i$$

donde

$$w_{ni}(t,h) = \frac{K\left(\frac{t-t_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{t-t_j}{h}\right)}. \tag{2.6}$$

$K(t)$ es llamada núcleo y es simétrica con $\int K(u)du = 1$. Algunos ejemplos posibles son:
Núcleo Normal:

$$K(t) = e^{-\frac{1}{2}t^2}.$$

Núcleo Epanechnicov:

$$K(t) = \begin{cases} \frac{3}{4}(1-t^2) & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}$$

Núcleo Tricúbico:

$$K(t) = \begin{cases} (1-|t|^3)^3 & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}$$

Todos estos ejemplos de estimadores noparamétricos son lineales es decir pueden escribirse como

$$\widehat{y} = Wy,$$

donde W es llamada matriz de suavizado y no depende de y.

Además de los pesos definidos por (2.6), Gasser y Müller (1979) consideraron para diseños fijos equiespaciados los pesos siguientes

$$w_{ni}(t,h) = \frac{1}{h}\int_{\frac{i-1}{n}}^{\frac{i}{n}} K(\frac{t-u}{h})du, \tag{2.7}$$

donde $h > 0$, $t_i = \frac{(i-0.5)}{n}$ y K es una función con soporte en $[-1,1]$. Como la función tiene soporte en $[0,1]$, si $t = qh \in [0,h)$ o $t = 1-qh \in (1-h,1]$, el soporte del núcleo para estimar en t con una ventana h no está contenido en el soporte de la función. Esto provoca un estimador sesgado, que se conoce como "efecto de frontera". Gasser y Müller (1984) propusieron una solución a este problema, introduciendo una modificación en el núcleo tomando núcleos $K_q(.)$ llamados núcleos de frontera con soporte en $[-1,q]$ para $t = qh \in [0.h)$ ($[-q,1]$ para $t = 1-qh \in (1-h,1]$).

Para el problema que nos interesa, los estimadores de ϕ_0 y de ϕ_j quedarán definidos por

$$\begin{aligned} \widehat{\phi}_0(t) &= \sum_{k=1}^{n} w_{nk}(t,h) y_k, \\ \widehat{\phi}_j(t) &= \sum_{k=1}^{n} w_{nk}(t,h) x_{kj} \qquad 1 \leq j \leq p. \end{aligned}$$

Ahora de acuerdo a lo expuesto en el comienzo del Capítulo podemos estimar $\boldsymbol{\beta}$ por un estimador de mínimos cuadrados. Para esto es conveniente expresar al modelo (2.1) en forma matricial;

$$Y = \mathbf{X}\boldsymbol{\beta} + \mathbf{G} + \varepsilon,$$

donde $Y = (y_1, \dots, y_n)^{\mathrm{T}}, \mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, con $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})^{\mathrm{T}}, \mathbf{G} = (g(t_1), \dots, g(t_n))^{\mathrm{T}}$ y $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$.
Entonces el estimador de $\boldsymbol{\beta}$ de mínimos cuadrados queda definido por

$$\widehat{\boldsymbol{\beta}} = (\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}\widetilde{Y},$$

con $\widetilde{Y} = (I - W)Y$, $\widetilde{\mathbf{X}} = (I - W)\mathbf{X}$ y W la matriz de suavizado donde $w_{ij} = w_{ni}(t_j, h)$.
En notación matricial también podemos escribir,

$$\mathbf{X} = \boldsymbol{\phi} + \boldsymbol{\eta}.$$

$\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n)$(donde $\boldsymbol{\phi}_i = (\phi_1(t_i), \dots, \phi_p(t_i))^{\mathrm{T}}$) y $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n)$ (donde $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{ip})^{\mathrm{T}}$).

2.3 Estimadores noparamétricos robustos

Es conocido que la presencia de datos atípicos puede afectar seriamente los estimadores obtenidos en la sección anterior. Por este motivo, Bianco y Boente (2002) propusieron estimadores robustos para el parámetro y la función de regresión en modelos parcialmente lineales. A continuación, daremos un breve resúmen de su propuesta. El procedimiento para obtener estimadores robustos en modelos parcialmente lineales puede ser descripto como un procedimiento en 3 pasos.

Paso 1: Estimar $\phi_0(t)$ y $\phi_j(t)$, $1 \leq j \leq p$ por estimadores robustos, como medianas locales o M-estimadores locales. Sean $\widehat{\phi}_{0,\mathrm{R}}$ y $\widehat{\phi}_{j,\mathrm{R}}$ los estimadores y llamemos $\widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t) = (\widehat{\phi}_{1,\mathrm{R}}(t), \dots, \widehat{\phi}_{p,\mathrm{R}}(t))^{\mathrm{T}}$.

Paso 2: Estimar el parámetro de regresión, utilizando un estimador de regresión robusto sobre los residuos $y_i - \widehat{\phi}_{0,\mathrm{R}}$ y $\mathbf{X}_i - \widehat{\boldsymbol{\phi}}_{\mathrm{R}}$. Llamemos $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ al estimador obtenido.

Paso 3: Calcular el estimador de la función de regresión g como $\widehat{g}_{\mathrm{R}}(t) = \widehat{\phi}_{0,\mathrm{R}}(t) - \widehat{\boldsymbol{\beta}}_{\mathrm{R}}^{\mathrm{T}} \widehat{\boldsymbol{\phi}}_{\mathrm{R}}(t)$.

En el paso 1, para calcular las medianas locales para estimar ϕ_0 y ϕ_j podemos tomar las medianas respecto a las funciones de distribución condicionales $\widehat{F}_0(y|T=t)$ y $\widehat{F}_j(y|T=t)$, respectivamente definidas como,

$$\begin{aligned}
\widehat{F}_0(y|T=t) &= \sum_{k=1}^{n} w_{nk}(t,h) I_{(-\infty,y]}(y_k) \\
\widehat{F}_j(y|T=t) &= \sum_{k=1}^{n} w_{nk}(t,h) I_{(-\infty,x]}(x_{kj}) \qquad 1 \leq j \leq p.
\end{aligned}$$

donde w_{ni} son los pesos definidos en (2.6) ó en (2.7).

Otra opción es usar M-estimadores locales, $\widehat{\phi}_{0,\mathrm{R}}$ y $\widehat{\phi}_{j,\mathrm{R}}$ definidos como solución de,

$$\sum_{k=1}^{n} w_{nk}(t,h) \psi \left(\frac{y_k - \widehat{\phi}_{0,\mathrm{R}}(t)}{\widehat{s}_0(t)} \right) = 0 \quad y \quad \sum_{k=1}^{n} w_{nk}(t,h) \psi \left(\frac{x_{kj} - \widehat{\phi}_{j,\mathrm{R}}(t)}{\widehat{s}_j(t)} \right) = 0,$$

respectivamente, donde w_{ni} son los pesos definidos en (2.6) ó en (2.7), ψ es una función impar, acotada y continua y $\widehat{s}_0(t)$ y $\widehat{s}_j(t)$ son estimadores de escala robustos. Posibles elecciones de ψ pueden ser la función de Huber o la función bicuadrada, mientras que para las escalas $\widehat{s}_0(t)$ y $\widehat{s}_j(t)$ pueden considerarse la mediana de las desviaciones absolutas respecto a la mediana (mad) (Huber 1981) con respecto a las distribuciones $\widehat{F}_0(y|T=t)$ y $\widehat{F}_j(y|T=t)$ definidas anteriormente.

En el paso 2, para la estimación robusta del parámetro de regresión han sido propuestos diversos métodos como M-estimadores (Huber 1981) resistente a outliers en los residuos pero no a puntos de alto leverage y los GM-estimadores (Mallows(1975), Krasker y Welsch (1982)) con punto de ruptura decreciente con la dimensión, pero resistentes a alto leverage. Entre estimadores de punto de ruptura $\frac{1}{2}$ podemos mencionar los S-estimadores (Rousseeuw y Yohai (1984)) los τ-estimadores y MM-estimadores (Yohai y Zamar (1988) y Yohai (1987)).

Capítulo 3

Selección y estimación de la ventana óptima

En este capítulo nos ocuparemos de como seleccionar la amplitud de la ventana requerida para la estimación del modelo. En forma intuitiva, podemos observar que ventanas grandes provocan un estimador con pequeña varianza pero de alto sesgo, y a la inversa si tomamos ventanas chicas. Si consideramos valores pequeños de h obtendremos un gráfico demasiado dentado, poco suave. En cambio valores altos de h provocarán un estimador demasiado llano que no capturará en forma correcta la tendencia deseada.

Para la selección del paramétro de suavizado en este tipo de estimadores se han considerado diversos métodos como validación cruzada y métodos plug-in.

En este Capítulo, obtendremos una expresión del parámetro de suavizado óptimo en el sentido de la ventana que minimiza el error cuadrático medio del estimador de mínimos cuadrados de $\boldsymbol{\beta}$. Para esto daremos una expresión asintótica del error cuadrático medio (MSE). Los resultados de este Capítulo se pueden encontrar en Linton (1995) y Aneiros-Pérez y Quintela del Río (2002) aunque algunas demostraciones han sido modificadas para adaptarlas a este contexto ya que Linton (1995) considera polinomios locales.

3.1 Aproximaciones Asintóticas: Ventana óptima

Para los resultados obtenidos en esta sección Linton (1995) y Aneiros-Pérez introdujeron las siguientes hipótesis:

H.1. The fixed regressor $\{t_i\}_{i=1}^n$ are equally spaced points: $t_i = \frac{i-0.5}{n}$.

H.2. The errores $\{\varepsilon_i : 1 \leq i \leq n\}$ is a sequence of i.i.d. random variables with $E(\varepsilon_i) = 0$ and $V(\varepsilon_i) = \sigma_\varepsilon^2 > 0$.

H.3. $\{\boldsymbol{\eta}_i : 1 \leq i \leq n\}$ is a sequence of i.i.d. random vectores with mean 0 and finite variance - covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}$.

H.4. $\{\varepsilon_i\}$ is independent of $\{\boldsymbol{\eta}_i\}$.

H.5. The exist integers $J \geq 2$ and $M \geq 4$ such that $E(|\varepsilon_1|^J) < \infty$ and $E(|\boldsymbol{\eta}_1|^M) < \infty$.

H.6. The functions $g(.), \phi_1(.), \ldots, \phi_p(.)$ have $\nu \geq 2$ contimuous derivatives on [0,1].

H.7. **a)** $K(.)$ is symmetric, even and Hölder continuous with support $[-1, 1]$.

b) $\int K(u)du = 1, \quad \int u^\nu K(u)du \neq 0$ with ν even and $\int u^j K(u)du = 0,\ 1 \leq j \leq \nu - 1$.

H.8. **a)** If $t = qh \in [0, h) \quad (t = 1-qh \in (1-h, 1]), K_q(.)$ has support $[-1, q] \quad ([-q, 1])$ and satisfies H.7 b) with bounded ν-th moment for $q \in [0, 1]$.

b) $K_q(.)$ is such that the asymptotic variance is uniformly bounded for $q \in [0, 1]$.

H.9. $h = h_n \in [an^{-\pi}, bn^{-\pi}]$ where $0 < a < b < \infty$ are constants and $\pi = \frac{2}{4\nu+1}$.

Consideremos el estadístico estandarizado $T = \sigma^{-1} c^{\mathrm{T}} n^{\frac{1}{2}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ donde $c \in \Re^p$ y σ^2 es la varianza asintótica de $c^{\mathrm{T}} n^{\frac{1}{2}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Mediante un desarrollo de Taylor si se cumplen H.1 a H.9 podemos descomponer a T de la siguiente manera,

$$T = T^* + n^{-\frac{3}{2}} R = T^* + R_T,$$

donde

$$\begin{aligned} T^* &= \sigma^{-1}(c^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_N - n^{-\frac{1}{2}} c^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_D \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_N + \\ &+ \ n^{-1} c^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_D \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_D \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_N). \end{aligned}$$

con

$$\begin{aligned} \mathbf{X}_N &= n^{-\frac{1}{2}} \widetilde{\mathbf{X}}^{\mathrm{T}} \widehat{\varepsilon} \\ \widehat{\varepsilon} &= \widetilde{Y} - \widetilde{\mathbf{X}} \boldsymbol{\beta} \\ \mathbf{X}_D &= n^{-\frac{1}{2}} (n^{-1} \widetilde{\mathbf{X}}^{\mathrm{T}} \widetilde{\mathbf{X}} - \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}) \\ R &= -\sigma c^{\mathrm{T}} (n^{-1} \widetilde{\mathbf{X}}^{\mathrm{T}} \widetilde{\mathbf{X}})^{-1} \mathbf{X}_D \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_D \boldsymbol{\Sigma}_{\eta_1}^{-1} \mathbf{X}_N. \end{aligned}$$

Si h satisface H.9 tenemos que $R_T = o_p(n^{-2\mu})$ con $\mu = \frac{4\nu-1}{2(4\nu+1)}$ (ver demostración del Teorema 3.1).

Linton (1995) observó que cuando $\sup_n E(T^2) < \infty$, $E(T^2) = E(T^{*2}) + o(n^{2\mu})$. Sin embargo, el problema reside en que T no tiene necesariamente los momentos uniformemente acotados. El siguiente Teorema nos provee de una aproximación de los momentos de T y T^* a través de una aproximación de sus distribuciones.

Teorema 3.1 *Assume that H.1 - H.9 hold, with* $\min(J,L) > M_1 = \max(6, \frac{16\mu}{3-4\mu})$ *where* $\mu = \frac{4\nu-1}{2(4\nu+1)}$. *Then, we have:*

a) T *and* T^* *have the same distribution to order* $n^{-2\mu}$,

b) $\sup_n(E(T^{*2})) < \infty$,

c) $E(T^*) = n^{\frac{1}{2}} h^{2\nu} A_1 + o(n^{-\mu}) \qquad V(T^*) = 1 + (nh)^{-1} A_2 + o(n^{-2\mu})$,

where

$$\begin{aligned}
A_1 &= \sigma^{-1} c^{\mathrm{T}} (\boldsymbol{\Sigma}_{\boldsymbol{\eta_1}})^{-1} \frac{\alpha_\nu^2(K)}{(\nu!)^2} \int_0^1 g^{(\nu)}(t) \phi^{(\nu)}(t) dt \\
A_2 &= \int K_*^2(u) du \\
\phi^{(\nu)}(t) &= (\phi_1^{(\nu)}(t), \ldots, \phi_p^{(\nu)}(t)) \\
\alpha_\nu(K) &= \int u^\nu K(u) du \\
K_*(u) &= K * K(u) - 2K(u) \\
\sigma^2 &= \sigma_\varepsilon^2 c^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta_1}^{-1} c.
\end{aligned}$$

En el Apéndice, se encuentra la demostración de este Teorema junto con los Lemas necesarios.

De acuerdo al Teorema 3.1 podemos encontrar una expresión asintótica para el error cuadrático medio de $\dfrac{c^{\mathrm{T}}\widehat{\boldsymbol{\beta}}}{\sigma}$,

$$\begin{aligned}
MSE(h) &= V\left(\frac{c^{\mathrm{T}}\widehat{\boldsymbol{\beta}}}{\sigma}\right) + [E\left(\frac{c^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma}\right)]^2 \\
&= n^{-1}(V(T) + (E(T))^2) \\
&= n^{-1}\{1 + (nh)^{-1}A_2 + o(n^{-2\mu}) + (n^{\frac{1}{2}} h^{2\nu} A_1 + o(n^{-\mu}))^2\}.
\end{aligned}$$

Luego,

$$MSE(h) \approx n^{-1}(1 + (nh)^{-1}A_2 + nh^{4\nu} {A_1}^2). \tag{3.1}$$

Derivando en h la expresión anterior tenemos que la ventana asintóticamente óptima h_{opt} es $h_{opt} = A_0 n^{-\pi}$, donde

$$A_0 = \left(\frac{A_2}{4\nu {A_1}^2} \right)^{\frac{1}{4\nu+1}}$$

$$A_0 = \left[\frac{\int K_*^2(u)du}{4\nu \left(\sigma^{-1} c^{\mathrm{T}} \boldsymbol{\Sigma}_{\eta_1}^{-1} \frac{\alpha_\nu^2(K)}{(\nu!)^2} \int_0^1 g^{(\nu)}(t) \phi^{(\nu)}(t) dt \right)^2} \right]^{\frac{1}{4\nu+1}}$$

y $\pi = \frac{2}{4\nu+1}$. Evaluando esta ventana en la aproximación (3.1) tenemos que el error óptimo asociado es

$$nMSE(h_{opt}) \approx 1 + n^{2\mu} A_2^{\frac{4\nu}{4\nu+1}} A_1^{\frac{2}{4\nu+1}} ((4\nu)^{\frac{1}{4\nu+1}} + (4\nu)^{\frac{-4\nu}{4\nu+1}}),$$

o sea el error cuadrático medio asintótico es de segundo orden $n^{2\mu}$.

3.2 Estimación de la ventana óptima

Obviamente esta ventana no puede ser calculada en la práctica ya que depende de cantidades que desconocemos como es el caso de σ_ε^2, $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}$ y las funciones $g^{(\nu)}(t)$ y ${\phi_0}^{(\nu)}(t)$. Una manera posible de estimar esta ventana óptima es mediante un método plug-in. Esto consiste en estimar consistentemente A_0 y luego usar $\widehat{A_0} n^{-\pi}$ para estimar a $\boldsymbol{\beta}$. En lo que resta del Capítulo discutiremos como estimar estas cantidades en el caso clásico.

Speckman (1988) propuso estimar $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}$ por $n^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}}$ que es un estimador consistente. Para estimar σ_ε^2 debemos proceder de la siguiente manera. Con una ventana inicial h_0 calculemos estimadores iniciales $\bar{\boldsymbol{\beta}}$ y $\bar{g}(t)$ para $\boldsymbol{\beta}$ y $g(t)$. Luego si $\widehat{\varepsilon}_i = y_i - \mathbf{X}_i^{\mathrm{T}} \bar{\boldsymbol{\beta}} - \bar{g}(t)$ podemos definir un estimador de σ_ε^2 como,

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n \widehat{\varepsilon}_i^2}{n},$$

que también resultará consistente. Por último, debemos estimar $g^{(\nu)}(t)$ y $\phi^{(\nu)}(t)$. Gasser y Müller (1984) propusieron estimadores para las derivadas de la función de regresión, a través de las derivadas de los pesos. Estos estimadores pueden escribirse como:

$$\widehat{g}^{(\nu)}(t) = \sum_{i=1}^n w_{ni}^{(\nu)}(t,h)(y_i - \mathbf{X}_i^{\mathrm{T}} \bar{\boldsymbol{\beta}})$$

$$\widehat{\phi}_j^{(\nu)}(t) \;=\; \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h) x_{ij} \quad 1 \leq i \leq p.$$

donde $w_{ni}^{(\nu)}(t,h)$ es la ν-ésima derivada de los pesos definidos en (2.6) ó en (2.7). De esta forma, podemos estimar de manera clásica la ventana asintóticamente óptima. Linton (1995) probó la convergencia de este estimador al parámetro de suavizado óptimo.

Capítulo 4

Propuesta robusta de estimación de la ventana óptima

En el Capítulo anterior se presentó la expresión del parámetro de suavizado óptimo y se dió una propuesta de estimación clásica para dicho parámetro. En esta parte del trabajo introduciremos un versión robusta del estimador de la ventana óptima. Por simplicidad supondremos $p = 1$ o sea $x_i \in \Re$. Como es sabido el parámetro de suavizado depende de σ_ε^2, σ_η^2, $g^{(\nu)}(t)$ y ${\phi_0}^{(\nu)}(t)$ luego debemos estimar de manera robusta dichas cantidades. Estimadores robustos para las derivadas de la función de regresión han sido estudiados por Härdle y Gasser (1985) cuando la escala es desconocida y por Boente (1989) en el caso heterocedástico. Consideremos primero el caso en el que $\nu = 2$, Boente, Fraiman y Meloche (1997) propusieron los siguientes estimadores de las derivadas de ϕ_0 y ϕ_1.

$$
\begin{aligned}
\widehat{\phi}_{0,\mathrm{R}}^{(2)}(t) &= \frac{\displaystyle\sum_{i=1}^{n} w_{ni}^{(2)}(t,h)\psi\left(\frac{y_i-\widehat{\phi}_{0,\mathrm{R}}(t)}{s_0}\right)s_0}{\displaystyle\sum_{i=1}^{n} w_{ni}(t,h)\psi'\left(\frac{y_i-\widehat{\phi}_{0,\mathrm{R}}(t)}{s_0}\right)}, \\
\widehat{\phi}_{1,\mathrm{R}}^{(2)}(t) &= \frac{\displaystyle\sum_{i=1}^{n} w_{ni}^{(2)}(t,h)\psi\left(\frac{x_i-\widehat{\phi}_{1,\mathrm{R}}(t)}{s_1}\right)s_1}{\displaystyle\sum_{i=1}^{n} w_{ni}(t,h)\psi'\left(\frac{x_i-\widehat{\phi}_{1,\mathrm{R}}(t)}{s_1}\right)}.
\end{aligned}
$$

donde $w_{ni}^{(2)}(t,h)$ es la derivada segunda de $w_{ni}(t,h)$ los pesos definidos en (2.6) ó en (2.7). Los estimadores de escala s_0 y s_1 puede ser un M-estimador de escala ó

$$
\begin{aligned}
s_0 &= \operatorname*{median}_{1\le i\le n-1} |y_i - y_{i-1}|/0.6754\sqrt{2}, \\
s_1 &= \operatorname*{median}_{1\le i\le n-1} |x_i - x_{i-1}|/0.6754\sqrt{2},
\end{aligned}
$$

donde hemos supuesto sin perdida de generalidad $0 \leq t_1 \leq \ldots \leq t_n \leq 1$.
En forma general, un estimador de la derivada de orden ν queda definido como,

$$\begin{aligned}
\widehat{\phi}^{(\nu)}_{0,\mathrm{R}}(t) &= \frac{\sum_{i=1}^{n} w^{(\nu)}_{ni}(t,h)\psi\left(\frac{y_i-\widehat{\phi}_{0,\mathrm{R}}(t)}{s_0}\right) - \sum_{\substack{3\leq j\leq \nu\\ j: impar}} \frac{1}{\nu! s_0^j}\lambda_{j,n}(t,\widehat{\phi}_{0,R}(t))H_j(\widehat{\phi}^{(1)}_{0,R}(t),\ldots,\widehat{\phi}^{(\nu-1)}_{0,R}(t))}{\lambda_{1,n}(t,\widehat{\phi}_{0,R}(t))} s_0, \\
\widehat{\phi}^{(\nu)}_{1,\mathrm{R}}(t) &= \frac{\sum_{i=1}^{n} w^{(\nu)}_{ni}(t,h)\psi\left(\frac{x_i-\widehat{\phi}_{1,\mathrm{R}}(t)}{s_1}\right) - \sum_{\substack{3\leq j\leq \nu\\ j: impar}} \frac{1}{\nu! s_1^j}\lambda^1_{j,n}(t,\widehat{\phi}_{1,R}(t))H_j(\widehat{\phi}^{(1)}_{1,R}(t),\ldots,\widehat{\phi}^{(\nu-1)}_{1,R}(t))}{\lambda^1_{1,n}(t,\widehat{\phi}_{1,R}(t))} s_1.
\end{aligned}$$

donde $w^{(\nu)}_{ni}(t,h)$ es la ν-ésima derivada de $w_{ni}(t,h)$ los pesos definidos en (2.6) ó en (2.7), s_0 y s_1 los estimadores de escala considerados anteriormente y

$$\begin{aligned}
\lambda_{j,n}(t,\phi(t)) &= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-t_i}{h}\right)\psi^{(j)}\left(\frac{y_i-\phi(t)}{s_0}\right), \\
\lambda^1_{j,n}(t,\phi(t)) &= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-t_i}{h}\right)\psi^{(j)}\left(\frac{x_i-\phi(t)}{s_1}\right) \quad \text{y} \\
H_j(\phi^{(1)}(t),\ldots,\phi^{(\nu-1)}(t)) &= \left.\{[\phi(u)-\phi(t)]^j\}^{(\nu)}\right|_{u=t}
\end{aligned}$$

Es fácil ver que $H_j(\phi^{(1)}(t),\ldots,\phi^{(\nu-1)}(t))$ solo depende de las derivadas de $\phi(t)$ hasta el orden $\nu-1$.

Lema 4.1 *Sea $v_j(u) = [\phi(u)-\phi(t)]^j$ entonces $v^{(k)}_j(t) = H_j(\phi^{(1)}(t),\ldots,\phi^{(k-1)}(t))$ para $j>1$.*

Demostración: Probaremos este resultado por inducción en j. Si $j=1$, $v^{(k)}_1(u) = \phi^{(k)}(u)$ entonces $v^{(k)}_1(t) = \phi^{(k)}(t)$. Si $j=2$,

$$\begin{aligned}
v^{(k)}_2(u) &= (v_1(u)v_1(u))^{(k)} = \sum_{r=0}^{k}\binom{k}{r} v^{(r)}_1(u)v^{(k-r)}_1(u) \\
&= 2\phi^{(k)}(u)v_1(u) + \sum_{r=1}^{k-1}\binom{k}{r}\phi^{(r)}_0(u)\phi^{(k-r)}(u)
\end{aligned}$$

entonces si evaluamos en t,

$$\begin{aligned}
v^{(k)}_2(t) &= \sum_{r=1}^{k-1}\binom{k}{r}\phi^{(r)}(t)\phi^{(k-r)}(t) \\
&= H_2(\phi^{(1)}(t),\ldots,\phi^{(k-1)}(t))
\end{aligned}$$

Ahora supongamos que el resultado vale para $j-1$ veamos que vale para j.

$$\begin{aligned} v_j^{(k)}(u) &= (v_{j-1}(u)v_1(u))^{(k)} = \sum_{r=0}^{k}\binom{k}{r} v_{j-1}^{(r)}(u)v_1^{(k-r)}(u) \\ &= v_1^{(k)}(u)v_{j-1}(u) + v_1(u)v_{j-1}^{(k)}(u) + \sum_{r=1}^{k-1}\binom{k}{r} v_{j-1}^{(r)}(u)\phi^{(k-r)}(u) \end{aligned}$$

entonces como $v_1(t)=0$ y $v_{j-1}(t)=0$

$$v_j^{(k)}(t) = \sum_{r=1}^{k-1}\binom{k}{r} v_{j-1}^{(r)}(t)\phi^{(k-r)}(t)$$

y por hipótesis inductiva $v_{j-1}^{(r)}(t) = H_{j-1}(\phi^{(1)}(t),\ldots,\phi^{(r-1)}(t))$ de este modo

$$v_j^{(k)}(t) = H_j(\phi^{(1)}(t),\ldots,\phi^{(k-1)}(t)) \quad \square$$

El Lema anterior nos provee de una forma recursiva para calcular H_j. Por ejemplo, en el caso $\nu = 3$ necesitamos

$$\begin{aligned} H_3(\phi^{(1)}(t),\ldots,\phi^{(3-1)}(t)) &= \sum_{r=1}^{3-1}\binom{3}{r} H_{3-1}(\phi^{(1)}(t),\ldots,\phi^{(r-1)}(t))\phi^{(3-r)}(t) \\ &= 3H_2(\phi(t))\phi^{(2)}(t) + 3H_2(\phi^{(1)}(t))\phi^{(1)}(t). \end{aligned}$$

Es fácil ver que $H_2(\phi(t)) = 0$ y $H_2(\phi^{(1)}(t)) = 2[\phi^{(1)}(t)]^2$ de esta forma

$$H_3(\phi^{(1)}(t),\phi^{(2)}(t)) = 6[\phi^{(1)}(t)]^3.$$

Entonces tenemos que si $\nu = 3$ podemos tomar como estimador robusto de la derivada tercera,

$$\begin{aligned} \widehat{\phi}_{0,\mathrm{R}}^{(3)}(t) &= \frac{\displaystyle\sum_{i=1}^{n} w_{ni}^{(3)}(t,h)\psi\left(\frac{y_i-\widehat{\phi}_{0,\mathrm{R}}(t)}{s_0}\right) - \frac{1}{s_0^3}\sum_{i=1}^{n} w_{ni}(t,h)\psi^{(3)}\left(\frac{y_i-\widehat{\phi}_{0,\mathrm{R}}(t)}{s_0}\right)[\widehat{\phi}_{0,\mathrm{R}}'(t)]^3}{\displaystyle\sum_{i=1}^{n} w_{ni}(t,h)\psi'\left(\frac{y_i-\widehat{\phi}_{0,\mathrm{R}}(t)}{s_0}\right)} s_0, \\ \widehat{\phi}_{1,\mathrm{R}}^{(3)}(t) &= \frac{\displaystyle\sum_{i=1}^{n} w_{ni}^{(3)}(t,h)\psi\left(\frac{x_i-\widehat{\phi}_{1,\mathrm{R}}(t)}{s_1}\right) - \frac{1}{s_1^3}\sum_{i=1}^{n} w_{ni}(t,h)\psi^{(3)}\left(\frac{x_i-\widehat{\phi}_{1,\mathrm{R}}(t)}{s_1}\right)[\widehat{\phi}_{1,\mathrm{R}}'(t)]^3}{\displaystyle\sum_{i=1}^{n} w_{ni}(t,h)\psi'\left(\frac{x_i-\widehat{\phi}_{1,\mathrm{R}}(t)}{s_1}\right)} s_1. \end{aligned}$$

De esta forma, por (2.3) podemos obtener un estimador robusto de $\widehat{g}^{(\nu)}$ como

$$\widehat{g}_{\mathrm{R}}^{(\nu)}(t) = \widehat{\phi}_{0,\mathrm{R}}^{(\nu)}(t) - \widehat{\phi}_{1,\mathrm{R}}^{(\nu)}(t)\widehat{\beta}. \tag{4.1}$$

Para estimar σ_ε^2 y σ_η^2, proponemos

$$\begin{aligned}\widehat{\sigma}_{\varepsilon,\mathrm{R}} &= \operatorname*{mad}_{1\le i\le n}(y_i - x_i\widehat{\beta} - \widehat{g}(t_i))/0.6754,\\ \widehat{\sigma}_{\eta,\mathrm{R}} &= \operatorname*{median}_{1\le i\le n-1}|x_i - x_{i-1}|/0.6754\sqrt{2} = s_1.\end{aligned}$$

donde $\widehat{g}$ y $\widehat{\beta}$ son estimadores calculados con una ventana inicial h_0. Entonces, el estimador robusto para h_{opt} propuesto sería,

$$\widehat{h}_R = \widehat{A}_{0,\mathrm{R}} n^{-\pi} \tag{4.2}$$

donde

$$\begin{aligned}\widehat{A}_{0,\mathrm{R}} &= \left(\frac{A_2}{4\nu\widehat{A}_{1,\mathrm{R}}^2}\right)^{\frac{1}{4\nu+1}}\\ \widehat{A}_{1,\mathrm{R}} &= (\widehat{\sigma}_{\varepsilon,\mathrm{R}}\widehat{\sigma}_{\eta,\mathrm{R}})^{-1}\frac{\alpha_\nu^2(K)}{(\nu!)^2}\int_{h_0}^{1-h_0}\widehat{g}_{\mathrm{R}}^{(\nu)}(t)\widehat{\phi}_{1,\mathrm{R}}^{(\nu)}(t)dt.\end{aligned}$$

4.1 Convergencia

El objetivo de esta sección es mostrar, que bajo ciertas condiciones, la ventana estimada satisface

$$\frac{\widehat{h}_{\mathrm{R}}}{h_{opt}} \xrightarrow{p} 1 \quad si \quad n\to\infty.$$

Boente, Fraiman y Meloche (1997) demostraron bajo ciertos supuestos que el estimador robusto de la derivada segunda de la función de regresión converge a la derivada segunda de la función de regresión.
Para probar la convergencia de la derivada de orden ν necesitaremos de las siguientes hipótesis:

A.1. $\{t_i\}_{i=1}^n$ are fixed of design points in $[0,1]$

A.2. $\{\varepsilon_i : 1\le i\le n\}$ is a sequence of i.i.d. random variables $\varepsilon_i \sim F(./\sigma_\varepsilon)$, F symmetric.

A.3. $\{\eta_i : 1\le i\le n\}$ is a sequence of i.i.d. random variables and $\eta\sim G(./\sigma_\eta)$ symmetric.

A.4. $\{\varepsilon_i\}$ is independent of $\{\boldsymbol{\eta}_i\}$.

A.5. The functions $g(.), \phi_1(.), \ldots, \phi_p(.)$ have $\nu\ge 2$ continuous derivatives on [0,1].

A.6. **a)**$K(.)$ is even, symmetric with ν continuous derivatives with support $[-1,1]$.

b)$\int K(u)du = 1, \quad \int u^\nu K(u)du \neq 0$ with ν even and $\int u^j K(u)du = 0$ for $1 \leq j \leq \nu - 1$.

c)$K^{(\nu)}$ is Lipschiz of order one.

A.7. $\sup_{t\in[h_0,1-2h_0]} |\widehat{\phi}^{(k)}_{j,\mathrm{R}}(t) - \phi_j^{(k)}(t)| \xrightarrow{a.s} 0 \quad if \quad n \to \infty$ for $j = 0, 1$ and $1 \leq k \leq \nu - 1$.

C.1. $\psi : R \to R$ is an odd, strictly increasing, bounded and continuous function such that $\lim_{t\to\infty} \psi(t) = a > 0$.

C.2. ψ is ν-th differentiable, its ν-th derivative $\psi^{(\nu)}$ bounded and Lipschitz of order one and $E(\psi'(\varepsilon_1)) \neq 0$.

C.3. $\lim_{n\to\infty} \dfrac{nh_0^{2\nu+1}}{\log n} = +\infty \qquad \lim_{n\to\infty} h_0 = 0.$

C.4. $\max_{1\leq i\leq n+1} ||t_i - t_{i-1}| - \dfrac{1}{n}| = O(n^{-\delta})$

C.5. The initial estimators $\widehat{g}(t)$ y $\widehat{\beta}$ cumplen $\sup_{t\in[0,1]} |\widehat{g}(t) - g(t)| \xrightarrow{a.s} 0$ if $n \to \infty$ and $\widehat{\beta} \xrightarrow{a.s} \beta$ if $n \to \infty$.

Teorema 4.2 *Under C1 - C5 and A1 - A7 hold, if* $n^{-\delta}h_0^{\nu+1} \to \infty$ *we have that*

$$\sup_{t\in[h_0,1-2h_0]} |\widehat{\phi}^{(\nu)}_{j,\mathrm{R}}(t) - \phi_j^{(\nu)}(t)| \xrightarrow{a.s} 0 \quad if \quad n \to \infty.$$

for $j = 0, 1$.

La demostración de se encuentra en el Apéndice.

Corolario 4.3 *Under the assumptions of Theorem 4.2 we have that,*

$$\frac{\widehat{h}_{\mathrm{R}}}{h_{opt}} \xrightarrow{p} 1 \quad if \quad n \to \infty.$$

PROOF: Using Theorem 4.2 we have that $\widehat{\phi}^{(\nu)}_{j,\mathrm{R}}(t) \xrightarrow{a.s} \phi_j^{(\nu)}(t)$ uniformly on $[h_0, 1-2h_0]$ for $j = 0, 1$. Then, by (4.1) and **C5** we have that,

$$\int_{h_0}^{1-h_0} \widehat{g}^{(\nu)}_{\mathrm{R}}(t)\widehat{\phi}^{(\nu)}_{1,\mathrm{R}}(t)dt - \int_{h_0}^{1-h_0} g^{(\nu)}(t)\phi_1^{(\nu)}(t)dt \xrightarrow{a.s} 0.$$

¿From **A.5** $g^{(\nu)}$ and $\phi_1^{(\nu)}$ are continuous an d boundeds on $[0, h_0]$ and $[1-h_0, 1]$. Then, $\int_0^{h_0} g^{(\nu)}(t)\phi_1^{(\nu)}(t)dt$ and $\int_{h_0}^{1-h_0} g^{(\nu)}(t)\phi_1^{(\nu)}(t)dt$ converge to 0 . This implies that

$$\int_{h_0}^{1-h_0} \widehat{g}_{\mathrm{R}}^{(\nu)}(t)\widehat{\phi}_{\mathrm{R}}^{(\nu)}(t)dt \xrightarrow{a.s} \int_0^1 g^{(\nu)}(t).\phi_1^{(\nu)}(t)dtw$$

It remains to show that

$$\widehat{\sigma}^2_{\varepsilon,\mathrm{R}} \xrightarrow{p} \sigma^2_{\varepsilon}, \tag{4.3}$$

$$\widehat{\sigma}^2_{\eta,\mathrm{R}} \xrightarrow{p} \sigma^2_{\eta}. \tag{4.4}$$

First, we prove (4.3). Write $\widehat{u}_i = y_i - x_i\widehat{\boldsymbol{\beta}} - \widehat{g}(t_i)$ and

$$\begin{aligned} \widehat{F}_n(y) &= \frac{1}{n}\sum_{i=1}^n I_{(-\infty,y]}(\widehat{u}_i) \\ F_n(y) &= \frac{1}{n}\sum_{i=1}^n I_{(-\infty,y]}(\varepsilon_i). \end{aligned}$$

Suppose that f is bounded and continuous then,

$$\begin{aligned} E_{\widehat{F}_n}(f) &= \frac{1}{n}\sum_{i=1}^n f(\widehat{u}_i) \\ E_{F_n}(f) &= \frac{1}{n}\sum_{i=1}^n f(\varepsilon_i). \end{aligned}$$

Let $\mathcal{C}_1 = \{|x| < C_1\}$ and $\mathcal{C}_2 = \{|y| < C_2\}$, such that if $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$, $P(\mathcal{C}) > 1 - \frac{\epsilon}{4\|f\|_\infty}$. Note that

$$\begin{aligned} |E_{\widehat{F}_n}(f) - E_{F_n}(f)| &\leq \frac{1}{n}\sum_{i=1}^n |f(\widehat{u}_i) - f(\varepsilon_i))| \\ &\leq \frac{1}{n}\sum_{i=1}^n |f(\widehat{u}_i) - f(\varepsilon_i))| I_{\mathcal{C}}(x_i, y_i) + \frac{1}{n}\sum_{i=1}^n |f(\widehat{u}_i) - f(\varepsilon_i))| I_{\mathcal{C}^c}(x_i, y_i) \\ &\leq S_1 + S_2. \end{aligned}$$

From,

$$|S_2| \leq 2\|f\|_\infty \frac{1}{n}\sum_{i=1}^n I_{\mathcal{C}^c}(x_i, y_i),$$

and the Strong Law of Large Numbers, we have that there exists a set $\mathcal{N}_1$ such that $P(\mathcal{N}_1) = 0$ and $\forall w \notin \mathcal{N}_1$ we obtain that

$$\frac{1}{n}\sum_{i=1}^n I_{\mathcal{C}^c}(x_i, y_i) \to P((x, y) \in \mathcal{C}^c).$$

then, $|S_2| < \frac{\epsilon}{2}$, if $w \notin \mathcal{N}_1$ and $n \geq n_2$.
Now, we study the other terms. By **C5**,

$$\sup_{t\in[0,1]} |\widehat{g}(t) - g(t)| \xrightarrow{a.s} 0 \quad and \quad |\widehat{\beta} - \beta| \xrightarrow{a.s} 0.$$

We get that $\exists \mathcal{N}_2 : P(\mathcal{N}_2) = 0 \quad \forall w \notin \mathcal{N}_2 \quad \sup_{t\in[0,1]} |\widehat{g}(t) - g(t)| \to 0$ and $\widehat{\beta} \to \beta$. Specially, if $w \notin \mathcal{N}_2$ and $n \geq n_2$ $\sup_{t\in[0,1]} |\widehat{g}(t) - g(t)| < 1$ and $|\widehat{\beta} - \beta| < 1$.
Since $x_i \in \mathcal{C}_1, y_i \in \mathcal{C}_2$ we will see that exist a compact set $\mathcal{T}$ large enough such that $y_i - x_i\widehat{\beta} - \widehat{g}(t_i)$ and $y_i - x_i\beta - g(t_i) \in \mathcal{T}$. Since f is uniformly continuous on $\mathcal{T}$ exists $\delta > 0$: if $u_1, u_2 \in \mathcal{C}, |u_1 - u_2| < \delta \Rightarrow |f(u_1) - f(u_2)| < \frac{\epsilon}{2}$.
Then

$$|y_i - x_i\beta - g(t_i)| \leq C_2 + C_1|\beta| + \|f\|_\infty.$$

Therefore, if $w \notin \mathcal{N}_2$ y $n \geq n_2$

$$|y_i - x_i\widehat{\beta} - \widehat{g}(t_i)| \leq C_2 + C_1(|\beta| + 1) + \|f\|_\infty + 1 = C_3.$$

If $\mathcal{T} = \{u : |u| \leq C_3\}$ we have that $\widehat{u}_i, \varepsilon_i \in \mathcal{T}$. From **C6** if $n \geq n_3$, $|\widehat{u}_i - \varepsilon_i| < \delta$ for $i \in \mathcal{J} = \{i : (x_i, y_i) \in \mathcal{C}\}$, because $|x_i| < C_1$, then $|S_1| < \frac{\epsilon}{2}$. Since $|E_{\widehat{F}_n}(f) - E_{F_n}(f)| < \epsilon$ if $w \notin \mathcal{N}_1 \cup \mathcal{N}_2$ y $n \geq N = \max(n_1, n_2, n_3)$.
If Π stands for the Prohorov distance. Then $\Pi(\widehat{F}_n, F_n) \to 0$ and since $\Pi(F_n, F) \to 0$ implies $\Pi(\widehat{F}_n, F) \to 0$. Now, (4.3) follow the fact that the MAD is a continuous function.

Now, we see (4.4), since $x_i = \phi_1(t_i) + \eta_i$ then,

$$\begin{aligned}
x_i - x_{i-1} &= \phi_1(t_i) - \phi_1(t_{i-1}) + \eta_i - \eta_{i-1} \\
x_i - x_{i-1} &= \phi_1'(\xi_{in})(t_i - t_{i-1}) + \eta_i - \eta_{i-1} \\
x_i - x_{i-1} &= \phi_1'(\xi_{in})(\frac{1}{n} + O(\frac{1}{n^\delta})) + \eta_i - \eta_{i-1} \\
x_i - x_{i-1} &= \phi_1'(\xi_{in})\frac{1}{n^\delta}O(1) + \eta_i - \eta_{i-1}.
\end{aligned}$$

by **C4**. Then,

$$\begin{aligned}
(x_i - x_{i-1})^2 &= (\phi_1'(\xi_{in}))^2\frac{1}{n^{2\delta}}O(1) + (\eta_i - \eta_{i-1})^2 + \frac{2}{n^\delta}\phi_1'(\xi_{in})(\eta_i - \eta_{i-1}) \\
(x_i - x_{i-1})^2 &= (\eta_i - \eta_{i-1})^2 + R_{in}.
\end{aligned}$$

Therefore,

$$\underset{1\leq i\leq n-1}{\text{median}}(x_i - x_{i-1})^2 = \underset{1\leq i\leq n-1}{\text{median}}[(\eta_i - \eta_{i-1})^2 + R_{in}].$$

It sufficies to show that

$$\underset{1\leq i\leq n-1}{\text{median}}((\eta_i-\eta_{i-1})^2+R_{in})-\underset{1\leq i\leq n-1}{\text{median}}((\eta_i-\eta_{i-1})^2)\xrightarrow{a.s}0.$$

Denote by $z_i=(\eta_i-\eta_{i-1})^2$ and

$$\begin{aligned}Q_n(y) &= \frac{1}{n}\sum_{i=1}^{n}I_{(-\infty,y]}(z_i+R_{in}),\\ P_n(y) &= \frac{1}{n}\sum_{i=1}^{n}I_{(-\infty,y]}(z_i).\end{aligned}$$

Suppose that f is bounded and continuous,

$$\begin{aligned}E_{Q_n}(f) &= \frac{1}{n}\sum_{i=1}^{n}f(z_i+R_{in}),\\ E_{P_n}(f) &= \frac{1}{n}\sum_{i=1}^{n}f(z_i).\end{aligned}$$

Let $\mathcal{B}=\{|z|<B\}\subset\Re$ compact, such that $P(B)>1-\frac{\epsilon}{4\|f\|_\infty}$ then

$$\begin{aligned}|E_{Q_n}(f)-E_{P_n}(f)| &\leq \frac{1}{n}\sum_{i=1}^{n}|f(z_i+R_{in})-f(z_i))|\\ &\leq \frac{1}{n}\sum_{i=1}^{n}|f(z_i+R_{in})-f(z_i))|I_{\mathcal{B}}(z_i)+\frac{1}{n}\sum_{i=1}^{n}|f(z_i+R_{in})-f(z_i))|I_{\mathcal{B}^c}(z_i)\\ &\leq A_1+A_2.\end{aligned}$$

We start with A_2. Since f is bounded,

$$|A_2|\leq\|f\|_\infty\frac{1}{n}\sum_{i=1}^{n}I_{B^c}(z_i),$$

since from The Strong Law of the Large Numbers

$$\frac{1}{n}\sum_{i=1}^{n}I_{B^c}(z_i)\to P(z\in\mathcal{B}^c).$$

We have that there exists a set $\mathcal{N}_3$ such that $P(\mathcal{N}_3)=0$ and $\forall w\notin\mathcal{N}_3\quad\frac{1}{n}\sum\limits_{i=1}^{n}I_{B^c}(z_i)\to P(z\in\mathcal{B}^c)$. Then if $w\notin\mathcal{N}_3$ and $n\geq n_3\quad|A_2|<\frac{\epsilon}{2}$.
The fact that f is uniformly continuous on $\mathcal{B}_1=\{|z|<B+1\}$ then $\xi>0:z_1,z_2\in$

$\Re, \quad |z_1 - z_2| < \xi \Rightarrow |f(z_1) - f(z_2)| < \frac{\epsilon}{2}$.
Since

$$\begin{aligned} |R_{in}| &\leq \|\phi'\|_\infty^2 \frac{O(1)}{n^{2\delta}} + 2\frac{O(1)}{n^\delta}\|\phi'\|_\infty^2 B^{\frac{1}{2}} \\ &< \xi \qquad if \quad n \geq n_4. \end{aligned}$$

Then, if $z_i \in \mathcal{B}$, $z_i + R_{in} \in \mathcal{B}_1$ and $|R_{in}| < \xi$ entail that $|f(z_i) - f(z_i + R_{in})| < \frac{\epsilon}{2}$ and $|A_2| < \epsilon$. This completes the proof. $\square$

Observación 1 *Un resultado análogo se obtiene si estimamos la* $\int g^{(\nu)}\phi_1^{(\nu)}$ *mediante* $\frac{1}{n}\sum_{i=1}^{n}\widehat{g}^{(\nu)}(t_i)\widehat{\phi}_1^{(\nu)}(t_i)I_{(h_0,1-h_0)}(t_i)$ *que es el procedimiento usado en el estudio de simulación para evitar el cálculo de la integral numérica.*

Observación 2 *Los resultados obtenidos anteriormente pueden extenderse al caso en que* t_i *sean v.a y no puntos de diseños fijos, si* t_i *tienen soporte compacto, digamos* $\mathcal{J} = [0,1]$ *y densidad* f_T *tal que* $f_T(t) > 0$ *para todo* t *en el interior de* $\mathcal{J}$*.*

Capítulo 5

Simulación

5.1 Condiciones de la simulación

Este Capítulo contiene los resultados de un estudio de simulación diseñado para evaluar el procedimiento de estimación propuesto en este trabajo.
Los objetivos de este estudio son

- Comparar la performance de los estimadores ante contaminación y bajo muestras normales.
- Mostrar la dependencia del método de selección de ventana en la ventana inicial.

Como hemos visto para estimar la ventana óptima necesitamos definir estimadores iniciales del parámetro y de la función de regresión, para estimar la varianza del error, así como estimar la derivadas de la función g y ϕ. A continuación se describen, los estimadores utilizados en el estudio de simulación

Estimadores iniciales del parámetro y la función de regresión:

- En la propuesta clásica se estimó la función de regresión por un estimador de núcleos con núcleo normal y el coeficiente de regresión por mínimos cuadrados.
- Para el caso robusto, el método de suavizado empleado inicialmente fue medianas locales y para el parámetro de regresión se utilizó un GM-estimador con función de Huber 1.6.

Estimadores de la función de regresión y sus derivadas:

- En el caso clásico, se tomo el núcleo normal, con media cero y desvío estándar $\sigma = (\frac{0.25}{0.675}) = 0.37$.

- En la propuesta robusta, se utilizaron los estimadores expuestos en el Capítulo 4 donde ψ es la función bicuadrada de Tukey con constante 4.685, y los pesos fueron calculados igual que en el caso clásico.

Además de los dos estimadores anteriores, se consideraron estimadores clásicos basados en polinomios locales y una propuesta robusta asociada.

Estimadores de la función de regresión y sus derivadas basados en polinomios locales:

- En el caso clásico, tanto la función de regresión como sus derivadas pueden estimarse utilizando polinomios locales (ver Wand and Jones (1995)). Si tenemos el modelo $Y_i = m(t_i) + \epsilon_i$, y nos interesa estimar $m(t)$ y $m^{(2)}(t)$, podemos utilizar polinomios locales de orden 2. En ese caso, un estimador de $m(t)$ será $\widehat{\alpha}_o$ mientras que $m^{(2)}(t)$ puede estimarse por $\widehat{\alpha}_2$ donde $(\widehat{\alpha}_o, \ldots, \widehat{\alpha}_2)$ es el valor que minimiza

$$\sum_{k=1}^{n} w_{nk}(t,h) \left(Y_k - \alpha_o - \alpha_1(t_i - t) - \alpha_2(t_i - t)^2 \right)^2$$

  Estos estimadores se indican por LS en las tablas y figuras donde se aclara que se trata de estimadores basados en polinomios locales.

- Teniendo en cuenta que los puntos de diseño t_i son fijos o se encuentran en un compacto, una propuesta robusta se basa en un M–estimador local de regresión. Se considera la solución $(\widehat{\alpha}_o, \ldots, \widehat{\alpha}_2)$ del sistema

$$\begin{aligned} \sum_{k=1}^{n} w_{nk}(t,h)\psi \left(Y_k - \alpha_o - \alpha_1(t_i - t) - \alpha_2(t_i - t)^2 \right) &= 0 \\ \sum_{k=1}^{n} w_{nk}(t,h)\psi \left(Y_k - \alpha_o - \alpha_1(t_i - t) - \alpha_2(t_i - t)^2 \right)(t_i - t) &= 0 \\ \sum_{k=1}^{n} w_{nk}(t,h)\psi \left(Y_k - \alpha_o - \alpha_1(t_i - t) - \alpha_2(t_i - t)^2 \right)(t_i - t)^2 &= 0 \, . \end{aligned}$$

  A partir de ella, se definen los estimadores de $m(t)$ y $m^{(2)}(t)$ como $\widehat{\alpha}_o$ y $\widehat{\alpha}_2$, respectivamente. En este estudio de simulación, se tomó como función escores ψ la función de Tukey con constante 4.685 y el estimador inicial del proceso iterativo fue un M–estimador local con ψ la función de Huber con constante 1.345. Este procedimiento puede extenderse a derivadas de orden mayor.

Luego de aplicar las técnicas de suavizado, el comportamiento del estimador de mínimos cuadrados (LS) de β fue comparado con los siguientes estimadores robustos del parámetro de regresión:

- GM-estimador con funcion de Huber de constante 1.6 en los residuos $\widehat{r}_i = y_i - \widehat{\phi}_{0,\mathrm{R}}$ y con función de peso en las covariables $w_2(z) = W\left[\left(\frac{z-\mu_z}{\sigma_z}\right)^2\right]$ donde W es la función de Huber con constante $c = \chi_{1,0.975}$ y μ_z y σ_z indican la mediana y la MAD de $\widehat{z}_i = \mathbf{X}_i - \widehat{\boldsymbol{\phi}}_{\mathrm{R}}$. Estos estimadores se indicaran por GM.

- Least median of square introducido por Rousseeuw (1984) indicado por LMS

- Least trimmed con 33% de las observaciones podadas propuesto por Rousseeuw (1984) indicado por LTS

- Estimador de un paso con estimador inicial $\widehat{\boldsymbol{\beta}}_{\mathrm{I}}$ el LMS definido como

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{\mathrm{I}} + s_{\mathrm{I}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi_1' \left(\frac{\widehat{r}_i - \widehat{z}_i' \widehat{\boldsymbol{\beta}}_{\mathrm{I}}}{s_{\mathrm{I}}} \right) w_2\left(\|\widehat{z}_i\|\right) \widehat{z}_i^2 \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi_1 \left(\frac{\widehat{r}_i - \widehat{z}_i' \widehat{\boldsymbol{\beta}}_{\mathrm{I}}}{s_{\mathrm{I}}} \right) w_2\left(\|\widehat{z}_i\|\right) \widehat{z}_i \right\} .$$

  El estimador de un paso mejora el orden del estimador inicial y tiene el mismo comportamiento asintótico que el GM. Este estimador se indicará por OS.

Se generaron 1000 muestras de tamaño $n = 100$ del siguiente modelo.

$$\begin{aligned} y_i &= 1 + x_i + 10t_i^2 + \varepsilon_i \quad 1 \le i \le n \\ x_i &= \frac{5}{\log(5)} e^{t_i} + \eta_i \quad 1 \le i \le n. \end{aligned}$$

Es decir con nuestra notación $g(t) = 1 + 10t^2$, $\beta = 1$ y $\phi(t) = \frac{5}{\log(5)} e^t$. Este modelo corresponde al considerado por Linton (1995).

En el caso no contaminado, indicado por C_0, se tomaron $(\varepsilon_i, \boldsymbol{\eta}_i)$ i.i.d normales con media 0 y matriz de covarianza identidad mientras que t_i fueron generados siguiendo una distribución $U[0,1]$.

En el caso contaminado, indicado por C_1, $\varepsilon_i \sim N(0,1)(1-\delta) + \delta\mathcal{C}(0,1)$, donde $\delta = 0.1$ y $\mathcal{C}(0,\sigma)$ indica la distribución Cauchy centrada en 0 con escala σ. Esta contaminación corresponde a inflar la varianza de los errores y por lo tanto, afectará la varianza de los estimadores del parámetro de regresión.

Se consideraron las siguientes medidas resumen:

- para los estimadores del parámetro β, se consideró el error cuadrático medio, calculado sobre las 1000 replicaciones e indicado por $MSE(\widehat{\beta})$ en las tablas y figuras. Por otra parte, en las tablas 5.3 a 5.6 se presentan los valores medios y los desvíos estándar.

- para los estimadores de la ventana óptima h_{opt}, se presentan en las tablas 5.1 y 5.2 los valores medios y los desvíos estándar.

- el comportamiento de un estimador $\widehat{g}$ de g se midió usando dos medidas

$$\begin{aligned} \mathrm{MSE}(\widehat{g}) &= \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{g}(t_i)-g(t_i)\right]^2 \\ \mathrm{MedSE}(\widehat{g}) &= \mathrm{median}\left(\left[\widehat{g}(t_i)-g(t_i)\right]^2\right) \,. \end{aligned}$$

  Las Tablas 5.7, 5.8, 5.9 y 5.10 dan el valor medio de $\mathrm{MSE}(\widehat{g})$, mientras que las Tablas 5.11 y 5.11 muestran los valores medianos de $\mathrm{MedSE}(\widehat{g})$ sobre las 1000 replicaciones porque en este caso el valor medio y la mediana eran similares. Para el tamaño de muestra considerado, bajo el modelo dado y con errores normales, el valor de h_{opt} es 0.3581.

En las figuras **??**, **??**, **??** y **??** llamamos OPT al verdadero valor del error cuadrático medio calculado sobre los diferentes valores de ventanas iniciales considerados. Es decir, la expresión (3.1) para diferentes valores del parámetro h que para el modelo que estamos considerando quedaría de la siguiente manera,

$$MSE(h) = \frac{1}{1000} + \frac{1.67}{1000^2}h + (0.37)^2h^8$$

5.2 Resultados

5.3 Conclusiones

El estudio de simulación confirma el comportamiento inadecuado del estimador de mínimos cuadrados (LS) bajo la presencia de outliers y como esto repercute en la estimación del parámetro de suavizado y por lo tanto del error cuadrático medio de los estimadores de β. De hecho, en los gráficos **??** y **??** correspondientes al error cuadrático medio bajo contaminación, el comportamiento del estimador de mínimos cuadrados ($MSE(\widehat{\beta})$) se ve tan influenciado por la presencia de datos anómalos que su gráfico queda fuera de la escala considerada.
Las tablas 5.1 y 5.2 muestran que bajo contaminación el estimador robusto de la ventana óptima presenta mayor estabilidad, sobre todo en el caso de estimadores basados en polinomios locales.
La misma variabilidad del estimador a un paso y del GM-estimador del parámetro de regresión que muestran las tablas 5.3 a 5.6 reflejan el similar comportamiento asintótico entre ambos ya observado anteriormente. Las tablas 5.4 y 5.6 confirman como es de esperar el aumento de variabilidad del estimador de mínimos cuadrados bajo contaminación y a su vez muestran mejores resultados en sesgo de los estimadores basados en polinomios locales que de los estimadores basados en núcleos.
Las figuras **??**, **??**, **??** y **??** muestran como incide la elección de la ventana inicial en la estimación del parámetro óptimo tanto en el caso clásico como en la propuesta robusta. Observemos que para cualquiera de los estimadores considerados la ventana óptima tiende a ser la mayor del rango elegido, lo que es acorde con las aproximaciones que hemos obtenido. Recordemos que nuestras aproximaciones dependen de h a través de $n^{-2}h^{-1}$ como se mostró en (3.1).
Por otro lado, en cuanto a la estimación del parámetro de regresión se puede observar que bajo C_1, least trimmed (LMS) y least median of square (LTS) parecieran ser la propuesta que menos se ve afectada bajo la presencia de outliers. En particular, en el caso donde utilizamos estimadores basado en polinomios locales.

Capítulo 6

Conclusiones Generales

En esta tesis se abordó el problema de estimación robusta del parámetro de suavizado óptimo y se probaron resultados de consistencia a la ventana óptima de un procedimiento definido vía la estimación robusta de las derivadas basadas en núcleos. A partir de este resultado se podría obtener la consistencia del parámetro de regresión y de la función de regresión; que será objeto de estudios posteriores.

Por otra parte, en el estudio de simulación se introdujo otro procedimiento basado en la estimación robusta de polinomios locales y se comparó con el anterior bajo normalidad y bajo contaminación en los errores. Los estimadores robustos muestran su ventaja frente a los clásicos cuando existen datos atípicos y resultados comparables cuando el modelo normal se verifica. Esto justifica la propuesta realizada en esta tesis y es consistente con los resultados teóricos obtenidos.

Apéndice A

Apéndice

A.1 Proof of Theorem 3.1

Lema A.1 *Under H.1 - H.9 and assuming that* $\phi_0, g_0 \in \{g, \phi_1, \ldots, \phi_p\}$ *we have,*

$$n^{-1}\widetilde{g}_0^{\mathrm{T}}\widetilde{\phi}_0 = h^{2\nu}\frac{\alpha_\nu^2(K)}{(\nu!)^2}\int_0^1 g_0^{(\nu)}(t)\phi_0^{(\nu)}(t)dt + o(h^{2\nu})$$

where $\widetilde{g}_0 = (I-W)g_0$ *and* $g_0 = (g_0(t_1), \ldots, g_0(t_n))^{\mathrm{T}}$.

PROOF: Taking into account that $\max_i \max\{\widetilde{g}_0(t_i), \widetilde{\phi}_0(t_i)\} = O(h^\nu)$, (Gasser and Müller, 1984)

$$n^{-1}\widetilde{g}_0^{\mathrm{T}}\widetilde{\phi}_0 = n^{-1}\sum_{i=1}^n \widetilde{g}_0(t_i)\widetilde{\phi}_0(t_i) = h^{2\nu}\frac{\alpha_\nu^2(K)}{(\nu!)^2}\int_0^1 g_0^{(\nu)}(t)\phi_0^{(\nu)}(t)dt + o(h^{2\nu}).\square$$

Lema A.2 *Let* $\rho_{ij} = n^{-\frac{1}{2}}(\sum_{s=1}^n w_{si}w_{sj} - 2w_{ij})$ $i,j \in \{1,\ldots,n\}$ *where* $w_{ij} = (W)_{ij}$. *Assume that H.1, H.2 and H.7 - H.9 holds. Then;*

a) $\max_{ij}|w_{ij}| = O((nh)^{-1})$ *and* $|w_{ij}| = 0$ *if* $|t_i - t_j| > 2h$.

b) *If* $t_k \in [3h, 1-3h] \Rightarrow \sum_{i=1}^n w_{ik} = 1$.

c) $\max_{ij}|\rho_{ij}| = O(n^{-\frac{3}{2}}h^{-1})$ *and* $|\rho_{ij}| = 0$ *if* $|t_i - t_j| > 3h$.

d) $\sum_{i=1}^n \sum_{i\neq j} {\rho_{ij}}^2 = O((nh)^{-1})$.

e) $\sum_{i=1}^{n} \rho_{ii} = O(n^{-\frac{1}{2}} h^{-1})$.

f) $\sum_{i=1}^{n} {\rho_{ii}}^2 = O((nh)^{-2})$.

g) $\sum_{i=1}^{n} \sum_{i \neq j} nh{\rho_{ij}}^2 = \int K_*^2(u)du + o(1)$ *where* $K_*(u) = K * K(u) - 2K(u)$.

Proof: The first six points are a consequence of the definition of ρ_{ij} and the our assumptions. For (g),

$$\rho_{ij} = n^{-\frac{1}{2}} (\sum_{s=1}^{n} w_{si} w_{sj} - 2w_{ij}).$$

¿From the mean value theorem of integration,

$$w_{ij} = \frac{1}{nh} K \left(\frac{t_i - t_j}{h} \right) + O((nh)^{-2}) = \frac{1}{n} K_h(t_i - t_j) + O((nh)^{-2}).$$

Therefore, we can write

$$\begin{aligned} \rho_{ij} &= n^{-\frac{1}{2}} \left[\sum_{s=1}^{n} \frac{1}{n^2} K_h(t_i - t_s) K_h(t_j - t_s) - 2\frac{1}{n} K_h(t_i - t_j) \right] + O(n^{-\frac{5}{2}} h^{-2}) \\ n^{\frac{3}{2}} h \rho_{ij} &= h \left(\sum_{s=1}^{n} \frac{1}{n} K_h(t_i - t_s) K_h(t_j - t_s) - 2K_h(t_i - t_j) \right) + O((nh)^{-1}). \end{aligned}$$

Since $K_*(u) = K * K(u) - 2K(u)$, from the mean value theorem

$$n^{\frac{3}{2}} h \rho_{ij} = K_* \left(\frac{t_i - t_j}{h} \right) + O((nh)^{-1})$$

then $nh^2 \to \infty$

$$\sum_{i=1}^{n} \sum_{i \neq j} nh{\rho_{ij}}^2 = n^{-2} h^{-1} \sum_{i=1}^{n} \sum_{i \neq j} {K_*}^2 \left(\frac{t_i - t_j}{h} \right) + o(1).$$

Then, K is Hölder with bounded support and K_* satifies the same properties. We have that

$$n^{-2} h^{-1} \sum_{i=1}^{n} \sum_{i \neq j} {K_*}^2 \left(\frac{t_i - t_j}{h} \right) + o(1) = \int K_*^2(u)du + o(1)$$

(see Aneiros-Pérez (2001) page 167) which concludes the proof. □

Firstlt, we need to study $\mathbf{X}_N$ and $\mathbf{X}_D$. We remember that

$$\begin{aligned} \mathbf{X}_N &= n^{-\frac{1}{2}}\widetilde{\mathbf{X}}^{\mathrm{T}}\widehat{\varepsilon} \quad \text{where} \quad \widehat{\varepsilon} = \widetilde{Y} - \widetilde{\mathbf{X}}\boldsymbol{\beta} \\ \mathbf{X}_D &= n^{\frac{1}{2}}(n^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}} - \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}) \end{aligned}$$

where

$$\widetilde{\mathbf{X}} = (I-W)\mathbf{X} \qquad \widetilde{Y} = (I-W)Y \quad \text{and} \quad \text{we denote} \quad V_a = Wa.$$

We can write;

$$\begin{aligned} \widehat{\varepsilon} &= \varepsilon + \widetilde{g} - V_\varepsilon \\ \widetilde{\mathbf{X}} &= \boldsymbol{\eta} + \widetilde{\phi} - V_{\boldsymbol{\eta}}. \end{aligned}$$

Then,

$$\mathbf{X}_N = n^{\frac{1}{2}}(\boldsymbol{\eta} + \widetilde{\phi} - V_{\boldsymbol{\eta}})^{\mathrm{T}}(\varepsilon + \widetilde{g} - V_\varepsilon),$$

and similary

$$\begin{aligned} \mathbf{X}_D &= n^{\frac{1}{2}}(n^{-1}(\boldsymbol{\eta} + \widetilde{\phi} - V_{\boldsymbol{\eta}})^{\mathrm{T}}(\boldsymbol{\eta} + \widetilde{\phi} - V_{\boldsymbol{\eta}}) - \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}), \\ &= n^{-\frac{1}{2}}((\boldsymbol{\eta} + \widetilde{\phi} - V_{\boldsymbol{\eta}})^{\mathrm{T}}(\boldsymbol{\eta} + \widetilde{\phi} - V_{\boldsymbol{\eta}}) - n \quad \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}). \end{aligned}$$

then,

$$\mathbf{X}_N = \widetilde{\mathbf{X}}_N + b_N + L_N + Q_N \quad \textit{and} \quad \mathbf{X}_D = \widetilde{\mathbf{X}}_D + b_D + L_D + Q_D,$$

where

$$\begin{aligned} \widetilde{\mathbf{X}}_N &= n^{-\frac{1}{2}}\boldsymbol{\eta}^{\mathrm{T}}\varepsilon, \\ b_N &= n^{-\frac{1}{2}}\widetilde{\phi}^{\mathrm{T}}\widetilde{g}, \\ L_N &= n^{-\frac{1}{2}}(\boldsymbol{\eta} - V_{\boldsymbol{\eta}})^{\mathrm{T}}\widetilde{g} + n^{-\frac{1}{2}}\widetilde{\phi}^{\mathrm{T}}(\varepsilon - V_\varepsilon) + \sum_{i=1}^{n}\rho_{ii}\varepsilon_i\boldsymbol{\eta}_i = \sum_{i=1}^{3}L_{N_i}, \\ Q_N &= \sum_{i=1}^{n}\sum_{j\neq i}\rho_{ij}\boldsymbol{\eta}_i\varepsilon_j, \\ \widetilde{\mathbf{X}}_D &= n^{-\frac{1}{2}}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta} - n^{\frac{1}{2}}\boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}, \\ b_D &= n^{-\frac{1}{2}}\widetilde{\phi}^{\mathrm{T}}\widetilde{\phi} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}\sum_{i=1}^{n}\rho_{ii} = \sum_{i=1}^{2}b_{D_i}, \\ L_N &= n^{-\frac{1}{2}}(\boldsymbol{\eta} - V_{\boldsymbol{\eta}})^{\mathrm{T}}\widetilde{\phi} + n^{-\frac{1}{2}}\widetilde{\phi}^{\mathrm{T}}(\boldsymbol{\eta} - V_{\boldsymbol{\eta}}) + \sum_{i=1}^{n}\rho_{ii}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}} - \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}) = \sum_{i=1}^{3}L_{D_i}, \\ Q_N &= \sum_{i=1}^{n}\sum_{j\neq i}\rho_{ij}\boldsymbol{\eta}_i\boldsymbol{\eta}_j. \end{aligned}$$

Lema A.3 *Under the assumptions of Lemma A.1, we have that*

$$b_N = n^{\frac{1}{2}} h^{2\nu} \frac{\alpha_\nu^2(K)}{(\nu!)^2} \int_0^1 g^{(\nu)}(t)\phi^{(\nu)}(t)dt + o(n^{\frac{1}{2}} h^{2\nu}) \qquad and$$
$$b_{D_1} = n^{\frac{1}{2}} h^{2\nu} \frac{\alpha_\nu^2(K)}{(\nu!)^2} \int_0^1 \phi^{(\nu)}(t)(\phi^{(\nu)}(t))^{\mathrm{T}} dt + o(n^{\frac{1}{2}} h^{2\nu}).$$

PROOF: It follows immediately from Lemma A.1. □

Lema A.4 *Under the assumptions of Lemma A.1, we have that*

$$b_{D_2} = O(n^{-\frac{1}{2}} h^{-1}).$$

PROOF: Since $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_1}$ is positive definite and $b_{D_2} = \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1} \sum_{i=1}^n \rho_{ii}$, he order of b_{D_2} is the same as that of $\displaystyle\sum_{i=1}^n \rho_{ii}$, by Lemma A.2 this concludes the proof. □

Lema A.5 *Under the assumptions of Lemma A.1, we have that*

$$L_{D_1}, L_{D_2}, L_{N_1}, L_{N_2} = o_p(h^\nu) \quad and \quad L_{D_3}, L_{N_3} = O(n^{-1}h^{-1}).$$

PROOF: We start with L_{N_1}

$$L_{N_1} = n^{-\frac{1}{2}} (\boldsymbol{\eta} - V\boldsymbol{\eta})^{\mathrm{T}} \widetilde{g} = n^{-\frac{1}{2}} \sum_{j=1}^n b_j \boldsymbol{\eta}_j$$

where

$$b_j = \widetilde{g}(I-W)_j = \widetilde{g}_j - \sum_{k=1}^n \widetilde{g}_k w_{kj} = \widetilde{g}(t_j) - \sum_{k=1}^n \widetilde{g}(t_k) w_{kj}.$$

If, we denote $V(U)$ the variance of U we have

$$\|V(L_{N_1})\| \leq C n^{-1} \sum_{j=1}^n b_j^2. \tag{A.1}$$

Let $B_h = \{j : \quad t_j < h \quad o \quad t_j > 1-h\} \qquad \#(B_h) = O(nh)$. (A.2)
By Lemma A.2 a) and since $\max_i |\widetilde{g}(t_i)| = O(h^\nu)$ (Gasser and Müller 1984), satisfies

$$\begin{array}{llll} If \quad j \in B_h & \Rightarrow & b_j = O(h^\nu) \quad \text{uniformly} \quad \text{in} \quad j \\ If \quad j \notin B_h & \Rightarrow & \{w_{jk} \neq 0 \Rightarrow k \notin B_h\} \end{array}$$

Therefore, by **H.2,** Lemma A.2 (a) (b) and the asymptotic expression of $\widetilde{g}(t_i)$ we obtain

$$b_j = h^{\nu} \frac{\alpha_{\nu}(K)}{\nu!} \sum_{h \notin B_h} w_{kj}(g^{(\nu)}(t_k) - g^{(\nu)}(t_j)) + o(h^{\nu}).$$

But

$$\begin{aligned}\left| \sum_{h \notin B_h} w_{kj}(g^{(\nu)}(t_k) - g^{(\nu)}(t_j)) \right| &\leq \max_{\{k,j: w_{kj} \neq 0\}} |g^{(\nu)}(t_k) - g^{(\nu)}(t_j)| \sum_{k=1}^{n} |w_{kj}| \\ &\leq C \max_{\{|k-j| \leq nh\}} |g^{(\nu)}(t_k) - g^{(\nu)}(t_j)|.\end{aligned}$$

By A.2 (a) the last term is $o(1)$ because $g^{(\nu)}$ is a continuous function defined on a compact support, therefore $b_j = o(h^{\nu})$, we have obtained

$$b_j = \begin{cases} O(h^{\nu}) & \text{uniformly} \quad \text{in} \quad j \in B_h \\ o(h^{\nu}) & \text{uniformly} \quad \text{in} \quad j \notin B_h \end{cases}$$

and since $\#(B_h) = O(nh)$ we have that

$$n^{-1} \sum_{i=1}^{n} b_j^2 \leq n^{-1} \sum_{j \in B_{3h}}^{n} b_j^2 + n^{-1} \sum_{j \notin B_{3h}}^{n} b_j^2 = o(h^{2\nu}),$$

consequently, using (A.1) and Markov's inequality we obtain that $L_{N_1} = o_p(h^{\nu})$.

Now, we see $L_{N_2} = o_p(h^{(\nu)})$

$$L_{N_2} = n^{-\frac{1}{2}} \sum_{j=1}^{n} c_j \varepsilon_j \quad \text{where} \quad c_j = \widetilde{\phi}^{\mathrm{T}}(I - W)_j$$

then

$$\|V(L_{N_2})\| = n^{-1} \|V(\sum_{j=1}^{n} c_j \varepsilon_j)\| \leq n^{-1} \| \sum_{j=1}^{n} c_j c_j^{\mathrm{T}} \sigma_{\varepsilon}^2 \|.$$

A similar reasonimg than the one used for L_{N_1} leads us to obtain that

$$c_j c_j^{\mathrm{T}} = \begin{cases} O(h^{2\nu}) & if \quad j \in B_h \\ o(h^{\nu}) & if \quad j \notin B_h \end{cases}$$

therefore $\|V(L_{N_2})\| \leq o(h^{2\nu})$ thus, $L_{N_2} = o_p(h^{\nu})$.

For the term L_{N_3}

$$\|V(L_{N_3})\| = \|V(\sum_{i=1}^{n} \rho_{ii} \varepsilon_i \boldsymbol{\eta}_i)\| = \| \sum_{i=1}^{n} \rho_{ii}^2 V(\varepsilon_i \boldsymbol{\eta}_i)\| = \|\sigma_{\varepsilon}^2 \boldsymbol{\Sigma}_{\boldsymbol{\eta}_1} \sum_{i=1}^{n} \rho_{ii}^2\|.$$

Using Lemma A.2 (f), **H.2** and **H.3** we have that $\|V(L_{N_3})\| = O((nh)^{-2})$ then $L_{N_3} = O_p((nh)^{-1})$.
For the terms L_{D_i} we can apply the same arguments as those for L_{N_i}. This concludes this proof. $\square$

Lema A.6 *Let us suppose that* $J, L > 2M$ *for some positive integer M in H.5. Then*

$$\sup_n E(\|\mathbf{X}_N\|^{2M}) < \infty \quad \textit{and} \quad \sup_n E(\|\mathbf{X}_D\|^{2M}) < \infty$$

PROOF: Using $E|X+Y|^r \leq C(E|X| + E|Y|)^r$ we have

$$E(\|\mathbf{X}_N\|^{2M}) \leq C(E(\|\widetilde{\mathbf{X}}_N\|^{2M}) + \|b_N\|^{2M} + E(\|L_N\|^{2M}) + E(\|Q_N\|^{2M})).$$

Now,we study the four terms.

$$\|b_N\| = O(n^{\frac{1}{2}}h^{2\nu}) = O(n^{-\mu}) \Rightarrow \sup_n \|b_N\|^{2M} < \infty$$

by Lemma A.1 and assumption **H.9**.
Since $\sup_n E(\|\widetilde{\mathbf{X}}_N\|^{2M}) = \sup_n E(\|n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{\eta}_i \varepsilon_i\|^{2M})$, the uniform bounds to the moments of ε_i and $\boldsymbol{\eta}_i$ are sufficient to guarantee that $\sup_n E(\|\widetilde{\mathbf{X}}_N\|^{2M}) < \infty$. For the term with L_N consider

$$E(\|L_N\|^{2M}) \leq C_1(E(\|L_{N_1}\|^{2M}) + E(\|L_{N_2}\|^{2M}) + E(\|L_{N_3}\|^{2M}))$$

with $L_{N_1 i} = n^{-\frac{1}{2}} \sum_{j=1}^n b_j \boldsymbol{\eta}_{ji}$ where $b_j = O(h^\nu)$ uniformly in j and $\{\boldsymbol{\eta}_{ij}\}_j$ are i.i.d with mean 0. Applynig Whittle's inequality (1960) we have that

$$E(\|L_{N_1 i}\|^{2M}) \leq C_2 n^M \left(\sum_{i=1}^n b_j^2 \right)^M = O(h^{2\nu M}).$$

For L_{N_2}, we put in the form $L_{N_2 i} = n^{-\frac{1}{2}} \sum_{j=1}^n c_{ji}\varepsilon_j$ with $i \in \{1, \ldots, p\}$ and $c_{ij} = O(h^\nu)$ uniformly in i, j . Similar arguments apply to L_{N_1} , $L_{N_2} = O(h^{2\nu M})$.
Finally, we write $L_{N_3} = \sum_{i=1}^n \rho_{ii}\varepsilon_i \boldsymbol{\eta}_i$ then applying Lemma A.2 (f) we have $E(\|L_{N_3}\|^{2M}) \leq C_3 \left(\sum_{i=1}^n \rho_{ii}^2 \right)^M = O((nh)^{-2M})$.

For finish the proof, we consider the term $Q_N = \sum_{i=1}^{n} \sum_{j \neq i} \rho_{ij} \boldsymbol{\eta}_i \varepsilon_j$. Applying Lemma 5.3.4 by Aneiros-Pérez (2001, page 170)

$$E(\|Q_n\|^{2M}) \leq C(\max_{1 \leq i,j \leq n} |\rho_{ij}|)^{2M} n^{2M} ([2h+1])^M \leq C'(\max_{1 \leq i,j \leq n} |\rho_{ij}|)^{2M} n^{2M} h^M.$$

By Lemma A.2 (c) we have that

$$E(\|Q_n\|^{2M}) = O((nh)^{-M}). \tag{A.2}$$

This completes the proof. □

Proof of Theorem 3.1.

a) We use the following result of Sargan and Mikhail (1971)

$$\forall x, \xi \quad |P(T \leq x) - P(T^* \leq x)| \leq P(|R_T| > \xi) + P(|T^* - x| < \xi).$$

see Rothenberg (1984). The last term is $O(\xi)$. We therefore choose $\xi = O((n^{2\mu} \log n)^{-1})$ and show that

$$P\left(|R_T| > C \frac{n\sqrt{n}}{n^{2\mu} \log n}\right) = o(n^{-2\mu}).$$

Let $A = \{|R_T| > C \dfrac{n\sqrt{n}}{n^{2\mu} \log n}\}$ and $B = \{\|\frac{\sum \boldsymbol{\eta}_i \boldsymbol{\eta}_i^{\mathrm{T}}}{n}\| > C_2\}$. We have

$$P(A) \leq P(A \cap B) + P(B^c)$$

where by Markov's inequality.

$$P(B^c) = P(\|\mathbf{X}_D\| > C_3 \sqrt{n}) \leq C_4 E(\|\mathbf{X}_D\|^{2M}).$$

Therefore provided $\sup_n E(\|\mathbf{X}_D\|^2) \leq \infty$, $P(B^c) = o(n^{-2\mu})$. When

$$\|\frac{\sum \boldsymbol{\eta}_i \boldsymbol{\eta}_i^{\mathrm{T}}}{n}\| > C_2, \quad |R_T| \leq C_5 |\mathbf{X}_N| \|\mathbf{X}_D\|^3$$

therefore $A \cap B \subset \bar{A}$ where

$$\bar{A} = \{|\mathbf{X}_N| \|\mathbf{X}_D\|^3 > C_6 \frac{n\sqrt{n}}{n^{2\mu} \log n}\} \quad \textit{and} \quad P(A \cap B) \leq P(\bar{A}) \leq C_7 (\log n)^{\rho} \frac{E(|\mathbf{X}_N|^{\rho} \|\mathbf{X}_D\|^{3\rho})}{n^{(\frac{3}{2} - 2\mu)\rho}}$$

for any ρ. Using that

$$E(|\mathbf{X}_N^{\rho} \|\mathbf{X}_D\|^{3\rho}|) \leq (E(|\mathbf{X}_N|^{4\rho}))^{\frac{1}{4}} (E(|\mathbf{X}_D|^{4\rho}))^{\frac{3}{4}}$$

while $\sup_n E(\|\mathbf{X}_D\|^{4\rho}) < \infty$ and $\sup_n E(|\mathbf{X}_N|^{4\rho}) < \infty$ by lemma A.6 and $(\frac{3}{2}-2\mu)\rho > 2\mu$; we have that

$$P\left(|R_T| > C\frac{n\sqrt{n}}{n^{2\mu}\log n}\right) = o(n^{-2\mu}).$$

b) By Hölder's inequality, we obtain

$$E(T^{*2}) \leq C\{E(|\mathbf{X}_N|^2) + n^{-1}E^{\frac{1}{2}}(\|\mathbf{X}_D\|^4)E^{\frac{1}{2}}(|\mathbf{X}_N|^4) + n^{-2}E^{\frac{2}{3}}(\|\mathbf{X}_D\|^6)E^{\frac{1}{3}}(|\mathbf{X}_N|^6)\}.$$

Now, an appication of Lemma A.6 yields the statement of Theorem 3.1 (b) .
c) Remember the expression for T^*

$$T^* = \sigma^{-1}(c^{\mathrm{T}}\mathbf{\Sigma}_{\eta_1}^{-1}\mathbf{X}_N - n^{-\frac{1}{2}}c^{\mathrm{T}}\mathbf{\Sigma}_{\eta_1}^{-1}\mathbf{X}_D\mathbf{\Sigma}_{\eta_1}^{-1}\mathbf{X}_N + n^{-1}c^{\mathrm{T}}\mathbf{\Sigma}_{\eta_1}^{-1}\mathbf{X}_D\mathbf{\Sigma}_{\eta_1}^{-1}\mathbf{X}_D\mathbf{\Sigma}_{\eta_1}^{-1}\mathbf{X}_N). \tag{A.3}$$

First, we show that the last term of T^* is $o_p(n^{2\mu})$. To check this, we only have to prove that $\mathbf{X}_N$ and $\mathbf{X}_D$ are $O_p(1)$ (because $2\mu < 1$) wich is obtained from Lemma A.4 with $M = 1$. Furthermore, using the lemmas of the above secction together with the bound(A.2) considered for Q_N and Q_D with $M = 1$, we have that

$$\mathbf{X}_N = \widetilde{\mathbf{X}}_N + O_p(n^{-\mu}) \qquad y \qquad \mathbf{X}_D = \widetilde{\mathbf{X}}_D + b_{D_2} + o_p(n^{-\mu}) \tag{A.4}$$

Form (A.3) and (A.4) we can write

$$\begin{aligned} T^* &= \sigma^{-1}c^{\mathrm{T}}\mathbf{\Sigma}_{\eta_1}^{-1}\Big\{\widetilde{\mathbf{X}}_N + b_N + Q_N + L_{N_1} + L_{N_2} + \left(L_{N_3} - n^{-\frac{1}{2}}b_{D_2}^{\mathrm{T}}\mathbf{\Sigma}_{\eta_1}^{-1}\widetilde{\mathbf{X}}_N\right) \\ &\quad -n^{-\frac{1}{2}}\widetilde{\mathbf{X}}_D^{\mathrm{T}}\mathbf{\Sigma}_{\eta_1}^{-1}\widetilde{\mathbf{X}}_N\Big\} + o_p(n^{-2\mu}) \qquad\qquad (\mathrm{A.5}) \\ &= T^{**} + o_p(n^{-2\mu}). \end{aligned}$$

Because of the independence between $\boldsymbol{\eta}_i$ and ε_i $\forall i, j$ it is easy to see that, except for b_n that is nonrandom, the expectations of all the other terms of T^{**} are zero. Therefore, using Lemma A.3 we have

$$\begin{aligned} E(T^{**}) &= \sigma^{-1}c^{\mathrm{T}}\mathbf{\Sigma}_{\eta_1}^{-1}\left\{n^{\frac{1}{2}}h^{2\nu}\frac{\alpha_\nu^2(K)}{(\nu!)^2}\int_0^1 g^{(\nu)}(t)\widetilde{\phi}^{(\nu)}(t)dt\right\} + o(h^{-\mu}) \\ &= n^{\frac{1}{2}}h^{2\nu}A_1 + o(n^{-\mu}). \end{aligned}$$

Now, we are going to compute $V(T^{**})$. We need to compute the variances of each term in (A.5) and the covariances between them.

$$\begin{aligned} V(\widetilde{\mathbf{X}}_n) &= \sigma_\varepsilon{}^2\mathbf{\Sigma}_{\boldsymbol{\eta}_1}, \\ V(b_n) &= 0 \quad \text{because} \quad b_n \quad \text{is nonrandom} \\ V(L_{N_1}) &= o(n^{-2\mu}) \quad \text{and} \quad V(L_{N_2}) = o(n^{-2\mu}) \end{aligned}$$

and by the proof of Lemma A.5,

$$V\left(L_{N_3} - n^{-\frac{1}{2}} b_{D_2}^{\mathrm{T}} \mathbf{\Sigma}_{\eta_1}^{-1} \widetilde{\mathbf{X}}_N\right) = o(n^{-2\mu}).$$

This follows from Lemmas A.4, A.5 and the fact that $V(\widetilde{\mathbf{X}}_N) = O(1)$, we obtain

$$V\left(n^{-\frac{1}{2}} \widetilde{\mathbf{X}}_D^{\mathrm{T}} \mathbf{\Sigma}_{\eta_1}^{-1} \widetilde{\mathbf{X}}_N\right) = o(n^{-2\mu}).$$

On the other hand, using the Lemma A.2

$$\begin{aligned}
V(Q_N) &= V\left(\sum_{i=1}^{n} \sum_{j \neq i} \rho_{ij} \boldsymbol{\eta}_i \varepsilon_j\right) = \sum_{i=1}^{n} \sum_{j \neq i} {\rho_{ij}}^2 \mathbf{\Sigma}_{\boldsymbol{\eta_1}} {\sigma_\varepsilon}^2 \\
&= (nh)^{-1} \mathbf{\Sigma}_{\boldsymbol{\eta_1}} {\sigma_\varepsilon}^2 \left(\int {K_*}^2(u) du + o(1)\right) \\
&= (nh)^{-1} \mathbf{\Sigma}_{\boldsymbol{\eta_1}} {\sigma_\varepsilon}^2 \int {K_*}^2(u) du + o((nh)^{-1}).
\end{aligned}$$

The cross term are $o(n^{-2\mu})$ or zero. Then by (A.5)

$$\begin{aligned}
V(T^{**}) &= \sigma^{-2} c^{\mathrm{T}} \mathbf{\Sigma}_{\eta_1}^{-1} \left\{V(\widetilde{\mathbf{X}}_N) + V(Q_N)\right\} \mathbf{\Sigma}_{\eta_1}^{-1} c + o(n^{-2\mu}) \\
&= \sigma^{-2} c^{\mathrm{T}} \mathbf{\Sigma}_{\eta_1}^{-1} \left\{{\sigma_\varepsilon}^2 \mathbf{\Sigma}_{\boldsymbol{\eta_1}} + (nh)^{-1} \mathbf{\Sigma}_{\boldsymbol{\eta_1}} {\sigma_\varepsilon}^2 \int {K_*}^2(u) du + o((nh)^{-1})\right\} \\
&\quad \mathbf{\Sigma}_{\eta_1}^{-1} c + o(n^{-2\mu}) \\
&= \left\{1 + (nh)^{-1} \int K_*^2(u) du\right\} + o(n^{-2\mu}).
\end{aligned}$$

Now, using Cauchy-Schwarz's inequality, Lemma A.6 and the bounds obtained troughot its proof together with assumption **H.9** we have that $E(|R_T^*|) = o(n^{-2\mu})$ and $E({R_T^*}^2) = o(n^{-4\mu})$ and we get the desidred result. □

A.2 Demostración del Teorema 4.2

In order to prove Theorem 4.2, we will need the following lemmas.

Lema A.7 *Under C.4, C.5 and A.6, if* $n^{\delta-1} h_0^{\nu+1} \to \infty$, *we have that*

(a) $\displaystyle \sup_{[h_0, 1-2h_0]} \left| \frac{1}{n h_0^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t - t_i}{h_0}\right) \right| \to 0 \quad if \quad n \to \infty$

(b) $\displaystyle \sup_{[h_0, 1-2h_0]} \left| \frac{1}{n h_0^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t - t_i}{h_0}\right) (\phi_0(t_i) - \phi_0(t)) - \frac{1}{h_0^{\nu}} \int_{-1}^{1} K^{(\nu)}(u) \phi_0(t - u h_0) du \right| \to$ $0 \quad if \quad n \to \infty.$

PROOF:
(a) By **A.6** we have that

$$\int_0^1 K^{(\nu)}\left(\frac{t-u}{h_0}\right) du = 0 \qquad for \quad t \in [h_0, 1-h_0].$$

Now, let n_0 be such that for $n \geq n_0$, $nh_0 > C$ and $|t_{n-1}| \leq \frac{C}{n}$, then

$$\int_{t_n}^1 K^{(\nu)}\left(\frac{t-u}{h_0}\right) du = 0 \quad \forall n \geq n_0 \quad for \quad t \in [h_0, 1-2h_0].$$

Thus, if M denotes the Lipschitz constant for $K^{(\nu)}$ and ξ_i is an intermediate point $t_{i-1} \leq \xi_i \leq t_{i+1}$ and $t_0 = 0$,

$$\begin{aligned}
\left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^n K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\right| &= \left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^n K^{(\nu)}\left(\frac{t-t_i}{h_0}\right) - \frac{1}{h_0^{\nu+1}}\int_0^{t_n} K^{(\nu)}\left(\frac{t-u}{h_0}\right) du\right| \\
&= \left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^n K^{(\nu)}\left(\frac{t-t_i}{h_0}\right) - \frac{1}{h_0^{\nu+1}}\sum_{i=1}^n \int_{t_{i-1}}^{t_i} K^{(\nu)}\left(\frac{t-u}{h_0}\right) du\right| \\
&= \frac{1}{h_0^{\nu+1}}\left|\frac{1}{n}\sum_{i=1}^n \left\{K^{(\nu)}\left(\frac{t-t_i}{h_0}\right) - n(t_i - t_{i-1})K^{(\nu)}\left(\frac{t-\xi_i}{h_0}\right)\right\}\right| \\
&\leq \frac{1}{h_0^{\nu+1}}\frac{1}{n}\sum_{i=1}^n M\left|\frac{t_i-\xi_i}{h_0}\right| + \frac{1}{nh_0^{\nu+1}}\sum_{i=1}^n \left|K^{(\nu)}\left(\frac{t-\xi_i}{h_0}\right)\right| |1 - n(t_i - t_{i-1})| \\
&\leq \frac{M}{h_0^{\nu+2}}\sup_i |t_i - \xi_i| + ||K^{(\nu)}||_\infty \frac{1}{h_0^{\nu+1}}\sum_{i=1}^n |\frac{1}{n} - (t_i - t_{i-1})| \\
&\leq \frac{M}{h_0^{\nu+2}}\sup_i |t_i - t_{i-1}| + ||K^{(\nu)}||_\infty \frac{1}{h_0^{\nu+1}} nO(n^{-\delta}) \\
&\leq \frac{M}{h_0^{\nu+2}}\left(\frac{1}{n} + O(n^{-\delta})\right) + ||K^{(\nu)}||_\infty \frac{O(1)}{h_0^{\nu+1}n^{\delta-1}} \\
&\leq \frac{MC}{nh_0^{\nu+2}} + ||K^{(\nu)}||_\infty \frac{O(1)}{h_0^{\nu+1}n^{\delta-1}}.
\end{aligned}$$

and the result follows.
(b) As in (a) since $K^{(\nu)}$ has compact support on $[-1, 1]$,

$$\frac{1}{h_0^{\nu+1}}\int_{-1}^1 K^{(\nu)}(u)\phi_0(t - uh_0)du = \frac{1}{h_0^{\nu}}\int_0^1 K^{(\nu)}\left(\frac{t-u}{h_0}\right)\phi_0(u)du$$

Thus, since (a) implies that,

$$\left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^n K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\phi_0(t)\right| \to 0,$$

uniformly on $[h_0, 1-2h_0]$, it remains to show that

$$\left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\phi_0(t_i) - \frac{1}{h_0^{\nu+1}}\int_0^1 K^{(\nu)}\left(\frac{t-u}{h_0}\right)\phi_0(u)du\right| \to 0,$$

uniformly on $[h_0, 1-h_0]$. Now, let n_0 be such that $nh_0 > C$ y $1-t_n \leq \frac{C}{n}$. As in (a), for $n \geq n_0$ we have

$$\begin{aligned}
| \quad \frac{1}{nh_0^{\nu+1}} & \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\phi_0(t_i) - \frac{1}{h_0^{\nu+1}}\int_0^1 K^{(\nu)}\left(\frac{t-u}{h_0}\right)\phi_0(u)du| \\
&= \left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^{n}\left\{K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\phi_0(t_i) - n\int_{t_{i-1}}^{t_i} K^{(\nu)}\left(\frac{t-u}{h_0}\right)\phi_0(u)du\right\}\right| \\
&= \left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^{n}\left\{K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\phi_0(t_i) - n(t_i-t_{i-1})K^{(\nu)}\left(\frac{t-\xi_i}{h_0}\right)\phi_0(\xi_i)\right\}\right|.
\end{aligned}$$

Using the same argument as in (a), the proof of (b) is completed using the fact that $K^{(\nu)}$ and ϕ_0 are Lipschitz functions. □

Using similar arguments we have

Lema A.8 *Under C.4, C.5 and A.6, if $n^{\delta-1}h_0^{\nu+1} \to \infty$ we have that*

$$\sup_{t\in[h_0,1-h_0]}\left|\frac{1}{nh_0}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)^p\right| - \int_{-1}^{1}|K^{(\nu)}(u)|^p du\right| \to 0, \quad p=1,2.$$

Lema A.9 *Let Z_i be independent random variables $|Z_i| \leq A$ a.s and $E(Z_i)=0$, then under C.4, C.5 and A.6 we have*

$$\sup_{t\in[h_0,1-h_0]}\left|\frac{1}{nh_0^{\nu+1}}\sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)Z_i\right| \to 0 \quad a.s.$$

PROOF: Denote by $S_n(t) = \frac{1}{nh_0^{\nu+1}}\sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)Z_i$.

Bernstein's inequality implies that for some positive constant α, we have

$$\sup_{t\in[h_0,1-2h_0]} P(|S_n(t)| > \epsilon) \leq 2e^{-\alpha n h_0^{2\nu+1}}$$

for $n \geq n_0$, since

$$\sup_{t\in[h_0,1-2h_0]}\left|\frac{1}{nh_0}\sum_{i=1}^{n}\left[K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\right]^2 - \int_{-1}^{1}[K^{(\nu)}(u)]^2 du\right| \to 0 \qquad n\to\infty$$

Consider a net in [0,1], $0 = u_1 \leq u_2 \leq \ldots \leq u_l = 1$ such that $|u_i - u_{i-1}| \leq h_0^{\gamma}$. We then have $l_n = O(h_0^{-\gamma})$ with $\gamma \geq \nu + 1$, and

$$\sup_{t\in[h_0,1-2h_0]} |S_n(t)| \leq \max_{1\leq j\leq l_n} \sup_{|t-u_j|\leq h_0^{\gamma}} |S_n(t) - S_n(u_j)| + \max_{\substack{1\leq j\leq l_n \\ u_j\in[h_0,1-h_0]}} |S_n(u_j)|.$$

Since $K^{(\nu)}$ is Liptschitz with constant M, the first term on the right-hand side can be majorized by $AMh_0^{\gamma-(\nu+1)}$, which entails that for $n \geq n_1$,

$$\max_{1\leq j\leq l_n} \sup_{|t-u_j|\leq h_0^{\gamma}} |S_n(t) - S_n(u_j)| \leq \epsilon.$$

For $n \geq \max\{n_1, n_0\}$,

$$P(\sup_{t\in[h_0,1-2h_0]} |S_n(t)| > 2\epsilon) \leq P\left(\max_{\substack{1\leq j\leq l_n \\ u_j\in[h_0,1-2h_0]}} |S_n(u_j)| \geq \epsilon\right) \leq 2l_n e^{-\alpha n h_0^{2\nu+1}} \leq C h_0^{-\gamma} n^{-\alpha \boldsymbol{\beta}_n},$$

where $\boldsymbol{\beta}_n = nh_0^{2\nu+1}/\log n \to \infty$ by **C.3** . Since $nh^{2\nu+1} \to \infty$ taking $\gamma = 2\nu$ we have that $nh^{\gamma} \to \infty$ and for n large enough, $nh_0^{\gamma} > 1$ and $1 - \alpha\boldsymbol{\beta}_n < -2$, therefore

$$P(\sup_{t\in[h_0,1-2h_0]} |S_n(t)| > 2\epsilon) \leq C n^{1-\alpha\boldsymbol{\beta}_n} \leq C n^{-2},$$

proving that $\sum_{i=1}^{n} P(\sup_{t\in[h_0,1-h_0]} |S_n(t)| > 2\epsilon) < \infty$ and the proof is completed using the Borel- Cantelli Lemma. $\square$

Lema A.10 *Under C.1 - C.5 and A.1 - A.6, if* $n^{\delta-1}h_0^{\nu+1} \to \infty$ *as* $n \to \infty$, *we have that*

(a)

$$\sup_{t\in[h_0,1-h_0]} \left| \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\psi[y_i - \phi_0(t)] - \phi_0^{(\nu)}(t)\lambda_1(t,\phi_0(t)) \right.$$
$$\left. - \sum_{\substack{3\leq j\leq\nu \\ j:odd}} \lambda_j(t,\phi_0(t)) \frac{1}{\nu!} H_j(\phi_0^{(1)}(t), \ldots, \phi_0^{(\nu-1)}(t)) \right| \xrightarrow{p.p} 0$$

where

$$w_{ni}^{(\nu)}(t,h_0) = \frac{1}{nh_0^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right),$$

and H_j *depends on the derivatives of* ϕ_0 *up to order* $\nu - 1$.

(b) $\sup_{t\in[h_0,1-h_0]} |\lambda_{j,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t)) - \lambda_j(t,\phi_0(t))| \xrightarrow{a.s} 0$
where

$$\begin{aligned}
\lambda_{j,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t)) &= \frac{1}{nh_0}\sum_{i=1}^{n} K\left(\frac{t-t_i}{h_0}\right)\psi^{(j)}(y_i-\widehat{\phi}_{0,\mathrm{R}}(t)),\\
\lambda_j(t,\phi_0(t)) &= E(\psi^{(j)}(y-\phi_0(t))) = E(\psi^{(j)}(\varepsilon_1)).
\end{aligned}$$

PROOF:

(a) A Taylor expansion of order ν gives

$$\sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\psi[y_i-\phi_0(t)] = \sum_{j=0}^{\nu} S_{jn}(t) + R_n$$

where

$$\begin{aligned}
S_{jn} &= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{\psi^{(j)}(\varepsilon_i)}{j!}[\phi_0(t_i)-\phi_0(t)]^j\\
R_n &= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{1}{\nu!}[\psi^{(\nu)}(\varepsilon_i+\xi_i)-\psi^{(\nu)}(\varepsilon_i)][\phi_0(t_i)-\phi_0(t)]^\nu
\end{aligned}$$

with $|\xi_i| \leq |\phi_0(t_i)-\phi_0(t)|$.
If j is even, since ψ is odd so $E(\psi^{(j)}(\varepsilon_1)) = 0$, then from Lemma A.9, $\sup_{t\in[h_0,1-h_0]} |S_{jn}| \xrightarrow{a.s} 0$ for j even.
If we denote by C_1 the Lipschitz constant for $\psi^{(\nu)}$ and by C_2 that of ϕ_0, since $|\xi_i| \leq |\phi_0(t_i)-\phi_0(t)|$

$$\begin{aligned}
|R_n(t)| &\leq \frac{1}{nh_0^{\nu+1}}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\right|\frac{1}{\nu!}|\psi^{(\nu)}(\varepsilon_i+\xi_i)-\psi^{(\nu)}(\varepsilon_i)||t-t_i|^\nu\\
&\leq \frac{1}{nh_0^{\nu+1}}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\right|\frac{C_1C_2^{\nu+1}}{\nu!}|t-t_i|^{\nu+1}\\
&\leq \frac{C_1C_2^{\nu+1}}{\nu!}\frac{1}{nh_0}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\right| h_0.
\end{aligned}$$

since K has compact support on $[-1,1]$.
Thus, since Lemma A.8

$$\sup_{t\in[h_0,1-h_0]}\left|\frac{1}{nh_0}\sum_{i=1}^{n}\left|K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\right| - \int_{-1}^{1}|K^{(\nu)}(u)|du\right| \to 0,$$

we have that, for n large enough, $\sup_{t\in[h_0,1-h_0]} |R_n(t)| \leq Ch_0$ then $\sup_{t\in[h_0,1-h_0]} |R_n(t)| \to 0$. It remains to show that $\sup_{t\in[h_0,1-h_0]} |S_{1n}(t) - E\psi'(\varepsilon_1)\phi_0^{(\nu)}(t)| \to 0 \quad a.s.$ if $n \to \infty$. Lemma A.9 implies that

$$\sup_{t\in[h_0,1-h_0]} \left| \frac{1}{nh_0^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right) [\psi'(\varepsilon_i) - E\psi'(\varepsilon_1)][\phi_0(t_i) - \phi_0(t)] \right| \to 0 \quad a.s.$$

Therefore, it will be enough to show that

$$\sup_{t\in[h_0,1-h_0]} \left| \frac{1}{nh_0^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right) [\phi_0(t_i) - \phi_0(t)] - \phi_0^{(\nu)}(t) \right| \to 0. \tag{A.6}$$

According to Lemma A.7

$$\sup_{t\in[h_0,1-h_0]} \left| \frac{1}{nh_0^{\nu+1}} \sum_{i=1}^{n} K^{(\nu)}\left(\frac{t-t_i}{h_0}\right) [\phi_0(t_i) - \phi_0(t)] - \frac{1}{h_0^{\nu}} \int_{-1}^{1} K^{(\nu)}(u)\phi_0(t-uh_0)du \right| \to 0.$$

since

$$\frac{1}{h_0^{\nu}} \int_{-1}^{1} K^{(\nu)}(u)\phi_0(t-uh_0)du = \int_{-1}^{1} K(u)\phi_0^{(\nu)}(t-uh)du$$

and the result for j even follows form $\phi_0^{(\nu)}$ is Lipschitz and (A.6). It remains to show that if j is odd y $j \geq 3$. $\sup_{t\in[h_0,1-h_0]} |S_{jn}(t) - E\psi^{(j)}(\varepsilon_1)\frac{1}{\nu!}H_j(\phi_0^{(1)}(t), \ldots, \phi_0^{(\nu-1)}(t))| \to 0 \quad a.s.$ if $n \to \infty$. From Lemma A.9 we have that

$$\sup_{t\in[h_0,1-h_0]} \left| \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)[\psi^{(j)}(\varepsilon_i) - E\psi^{(j)}(\varepsilon_1)]\frac{1}{\nu!}[\phi_0(t_i) - \phi_0(t)]^j \right| \to 0 \quad a.s.$$

Therefore, it will be enough to show that,

$$\sup_{t\in[h_0,1-h_0]} \left| \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)[\phi_0(t_i) - \phi_0(t)]^j - H_j(\phi_0^{(1)}(t), \ldots, \phi_0^{(\nu-1)}(t)) \right| \to 0.$$

Let $v_j(u) = [\phi_0(u) - \phi_0(t)]^j$ then $v_j(t) = 0$ thus, $\sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)[\phi_0(t_i) - \phi_0(t)]^j = \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)[v_j(t_i) - v_j(t)]$ For the Lemma A.7 and by a similar argument as that S_{1n} we have that

$$\sup_{t\in[h_0,1-h_0]} \left| \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)[\phi_0(t_i) - \phi_0(t)]^j - \int_{0}^{1} K(u)v_j^{(\nu)}(t-uh) \right| \to 0.$$

But $\int_0^1 K(u)v_j^{(\nu)}(t-uh) = v_j^{(\nu)}(t)$ and Lemma 4.1 $v_j^{(\nu)}(t) = H_j(\phi_0^{(1)}(t), \ldots, \phi_0^{(\nu-1)}(t))$ thus concluding the proof of (a).

(b) As in Lemma A.9, straightforward calculation show that

$$\sup_{t\in[h_0,1-h_0]} |\lambda_{j,n}(t,\phi_0(t)) - E\psi^{(j)}(\varepsilon_1)| \to 0 \quad a.s.$$

Thus, in order to obtain (b), it remains to show that

$$\sup_{t\in[h_0,1-h_0]} |\lambda_{j,n}(t,\phi_0(t)) - \lambda_{j,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))| \to 0 \quad a.s.$$

A Taylor expansion of order one gives

$$\begin{aligned} |\lambda_{j,n} & (t,\widehat{\phi}_{0,\mathrm{R}}(t)) - \lambda_{j,n}(t,\phi_0(t))| \\ = & \left| \frac{1}{nh_0}\sum_{i=1}^n K\left(\frac{t-t_i}{h_0}\right)\psi^{(j+1)}(y_i-\xi_{in})(\widehat{\phi}_{0,\mathrm{R}}(t)-\phi_0(t))\right| \\ \leq & \|\psi^{(j+1)}\|_\infty \frac{1}{nh_0}\sum_{i=1}^n \left|K\left(\frac{t-t_i}{h_0}\right)\right| |\widehat{\phi}_{0,\mathrm{R}}(t)-\phi_0(t)|, \end{aligned}$$

which implies the desired result since

$$\sup_{t\in[h_0,1-h_0]} \left| \frac{1}{nh_0}\sum_{i=1}^n \left|K\left(\frac{t-t_i}{h_0}\right)\right| - \int_{-1}^1 |K(u)|du\right| \to 0.$$

and $\sup_{t\in[h_0,1-h_0]} |\widehat{\phi}_{0,\mathrm{R}}(t)-\phi_0(t)| \to 0$ by a similar argument as that Theorem 2 from Härdel and Luckhaus (1984). □

PROOF OF THEOREM 4.2

Note that

$$\begin{aligned} \sup_{t\in[h_0,1-h_0]} | & \widehat{\phi}^{(\nu)}_{0,\mathrm{R}}(t) - \phi_0^{(\nu)}(t)| \\ = & \sup_{t\in[h_0,1-h_0]} |\lambda_{1,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))|^{-1} \left|\sum_{i=1}^n w_{ni}^{(\nu)}(t,h_0)\psi[y_i-\widehat{\phi}_{0,\mathrm{R}}(t)] - \right. \\ - & \left. \sum_{\substack{3\leq j\leq \nu \\ j:odd}} \lambda_{j,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))\frac{1}{\nu!}H_j(\widehat{\phi}^{(1)}_{0,\mathrm{R}}(t),\ldots,\widehat{\phi}^{(\nu-1)}_{0,\mathrm{R}}(t)) - \phi_0^{(\nu)}(t)\lambda_{1,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))\right| \\ \leq & \{\inf_{t\in[h_0,1-h_0]} |\lambda_{1,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))|\}^{-1} \sup_{t\in[h_0,1-h_0]} \left|\sum_{i=1}^n w_{ni}^{(\nu)}(t,h_0)\psi[y_i-\widehat{\phi}_{0,\mathrm{R}}(t)]\right. \\ - & \left. \sum_{\substack{3\leq j\leq \nu \\ j:odd}} \lambda_{j,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))\frac{1}{\nu!}H_j(\widehat{\phi}^{(1)}_{0,\mathrm{R}}(t),\ldots,\widehat{\phi}^{(\nu-1)}_{0,\mathrm{R}}(t)) - \phi_0^{(\nu)}(t)\lambda_{1,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))\right|. \end{aligned}$$

Since $\inf_{t\in[h_0,1-h_0]} |\lambda_{1,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))| \geq |E\psi'(\varepsilon_1)| - \sup_{t\in[h_0,1-h_0]} |\lambda_{1,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t)) - E\psi'(\varepsilon_1)|$, from Lemma A.10(b) and C.2, it will be enough to show that

$$\begin{aligned}\sup_{t\in[h_0,1-h_0]} \quad | & \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\psi[y_i - \widehat{\phi}_{0,\mathrm{R}}(t)] - \sum_{\substack{3\leq j\leq \nu\\ j:odd}} \lambda_{j,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))\frac{1}{\nu!}H_j(\widehat{\phi}_{0,\mathrm{R}}^{(1)}(t),\ldots,\widehat{\phi}_{0,\mathrm{R}}^{(\nu-1)}(t)) \\ - & \ \phi_0^{(\nu)}(t)\lambda_{1,n}(t,\widehat{\phi}_{0,\mathrm{R}}(t))| \to 0 \quad a.s.\end{aligned}$$

According to Lemma A.10, it suffices to show that

$$\sup_{t\in[h_0,1-h_0]} \left|\sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)[\psi(y_i - \widehat{\phi}_{0,\mathrm{R}}(t)) - \psi(y_i - \phi_0(t))]\right| \to 0 \quad a.s. \qquad \text{and} \quad \text{(A.7)}$$

$$\sup_{t\in[h_0,1-h_0]} |H_j(\widehat{\phi}_{0,\mathrm{R}}^{(1)}(t),\ldots,\widehat{\phi}_{0,\mathrm{R}}^{(\nu-1)}(t)) - H_j(\phi_0^{(1)}(t),\ldots,\phi_0^{(\nu-1)}(t))| \to 0 \quad a.s. \quad \text{(A.8)}$$

In order to prove (A.8), as in Lemma A.10 (b), using a Taylor's expansion of order ν we have that

$$\sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\psi[y_i - \phi_0(t)] = \sum_{j=0}^{\nu} S_{jn}(t) + R_n(t) \quad \text{and}$$

$$\sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\psi[y_i - \widehat{\phi}_{0,\mathrm{R}}(t)] = \sum_{j=0}^{\nu} S_{jn}^*(t) + R_n^*(t)$$

where

$$\begin{aligned}
S_{jn}(t) &= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{\psi^{(j)}(\varepsilon_i)}{j!}[\phi_0(t_i) - \phi_0(t)]^j \\
R_n(t) &= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{1}{\nu!}[\psi^{(\nu)}(\varepsilon_i + \xi_i) - \psi^{(\nu)}(\varepsilon_i)][\phi_0(t_i) - \phi_0(t)]^\nu \\
S_{jn}^*(t) &= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{\psi^{(j)}(\varepsilon_i)}{j!}[\phi_0(t_i) - \widehat{\phi}_{0,\mathrm{R}}(t)]^j \\
R_n^*(t) &= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{1}{\nu!}[\psi^{(\nu)}(\varepsilon_i + \xi_i^*) - \psi^{(\nu)}(\varepsilon_i)][\phi_0(t_i) - \widehat{\phi}_{0,\mathrm{R}}(t)]^\nu
\end{aligned}$$

with $|\xi_i| \leq |\phi_0(t_i) - \phi_0(t)|$ and $|\xi_i^*| \leq |\phi_0(t_i) - \widehat{\phi}_{0,\mathrm{R}}(t)|$. Since $S_{0n}^*(t) = S_{0n}(t)$ and

$$\begin{aligned}
S_{jn}^*(t) &= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{\psi^{(j)}(\varepsilon_i)}{j!}[\phi_0(t_i) - \widehat{\phi}_{0,\mathrm{R}}(t)]^j \\
&= \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)\frac{\psi^{(j)}(\varepsilon_i)}{j!}[\phi_0(t_i) - \phi_0(t) + \phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)]^j
\end{aligned}$$

$$
\begin{aligned}
&= S_{jn}(t) + \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0) \frac{\psi^{(j)}(\varepsilon_i)}{\nu!} \sum_{k=0}^{j-1} \binom{j}{k} [\phi_0(t) - \phi_0(t)]^k [\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)]^{j-k} \\
&= S_{jn}(t) + T_j(t)
\end{aligned}
$$

then

$$
\left| \sum_{i=1}^{n} w_{ni}^{(\nu)}(t,h_0)[\psi(y_i - \widehat{\phi}_{0,\mathrm{R}}(t)) - \psi(y_i - \phi_0(t))] \right| \leq \sum_{j=1}^{\nu} |T_j(t)| + |R_n^*(t) - R_n(t)|.
$$

By Lemma A.7(a), el Lemma A.9, $\sup\limits_{t \in [h_0, 1-h_0]} |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|$ and ϕ_0 is bounded implies that

$$
\sup_{t \in [h_0, 1-h_0]} |T_j| \to 0 \quad p.p
$$

Thus, since $\sup\limits_{t \in [h_0, 1-h_0]} |R_n(t)| \to 0 \quad a.s.$, we obtain that it will be enough to show that

$$
\sup_{t \in [h_0, 1-h_0]} |R_n^*(t)| \to 0 \quad p.p.
$$

Denote by C_1 the Lipschitz constant for $\psi^{(\nu)}(t)$ and C_2 that of $\phi_0(t)$. Since K has compact support on [-1,1] and $|\xi_{i,n}^*| \leq |\phi_0(t_i) - \widehat{\phi}_{0,\mathrm{R}}(t)|$.

$$
\begin{aligned}
|R_n^*(t)| &\leq \frac{C_1}{\nu!} \sum_{i=1}^{n} |w_{ni}^{(\nu)}(t,h_0)| |\xi_{i,n}^*| |\phi_0(t_i) - \widehat{\phi}_{0,\mathrm{R}}(t)|^{\nu} \\
&\leq C_1 \sum_{i=1}^{n} |w_{ni}^{(\nu)}(t,h_0)| |\xi_{i,n}^*| \{ |\phi_0(t_i) - \phi_0(t)|^{\nu} + |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|^{\nu} \} \\
&\leq C_1 \left\{ \sum_{i=1}^{n} |w_{ni}^{(\nu)}(t,h_0)| |\phi_0(t_i) - \phi_0(t)|^{\nu+1} + \sum_{i=1}^{n} |w_{ni}^{(\nu)}(t,h_0)| |\phi_0(t_i) - \phi_0(t)| |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|^{\nu} \right. \\
&+ \left. \sum_{i=1}^{n} |w_{ni}^{(\nu)}(t,h_0)| |\phi_0(t_i) - \phi_0(t)|^{\nu} |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)| + \sum_{i=1}^{n} |w_{ni}^{(\nu)}(t,h_0)| |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|^{\nu+1} \right\} \\
&\leq C_1 \left[\sum_{i=1}^{n} |w_{ni}^{(\nu)}(t,h_0)| \right] [C_2^{\nu+1} h_0^{\nu+1} + |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|^{\nu} C_2 h_0 + C_2^{\nu} h_0^{\nu} |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)| \\
&+ |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|^{\nu+1}].
\end{aligned}
$$

From

$$
\sup_{t \in [h_0, 1-h_0]} \left| \frac{1}{h_0} \sum \left| K^{(\nu)} \left(\frac{t - t_i}{h_0} \right) \right| - \int_{-1}^{1} |K^{(\nu)}(u)| du \right| \to 0,
$$

we get that there exists $A > 0$ such that $\sup_{t\in[h_0,1-h_0]} \frac{1}{h_0} \sum \left| K^{(\nu)}\left(\frac{t-t_i}{h_0}\right)\right| \leq A$. Since,

$$\begin{aligned}\sup_{t\in[h_0,1-h_0]} |R_n^*(t)| &\leq C_1 A \frac{1}{h_0^\nu} \left\{ C_2^{\nu+1} h_0^{\nu+1} + \sup_{t\in[h_0,1-h_0]} |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|^\nu C_2 h_0 \right. \\ &+ \left. C_2^\nu h_0^\nu \sup_{t\in[h_0,1-h_0]} |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)| + \sup_{t\in[h_0,1-h_0]} |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)|^{\nu+1} \right\}.\end{aligned}$$

In a similar way as in Theorem 2 from Härdle and Luckhaus (1984), one can obtain for some $\epsilon_0 > 0$,

$$\sum_{n=1}^\infty P(\sup_{t\in[h_0,1-h_0]} |\phi_0(t) - \widehat{\phi}_{0,\mathrm{R}}(t)| > \epsilon_0 h_0) < \infty.$$

Therefore, if we denote by $A_n = \{ \sup_{t\in[h_0,1-h_0]} |\phi_0(t) - \widehat{\phi}_{\mathrm{R}}(t)| > \epsilon_0 h_0\}$, we have that

$$P(\sup_{t\in[h_0,1-h_0]} |R_n^*(t)| > \eta) < P(A_n) + P(\sup_{t\in[h_0,1-h_0]} |R_n^*| > \eta \cap A_n^c).$$

In A_n^c, $\sup_{t\in[h_0,1-h_0]} |R_n^*(t)| \leq C_1 \frac{1}{h_0^\nu}(C_2^{\nu+1} + C_2^\nu \epsilon + C_2 \epsilon^\nu + \epsilon^{\nu+1}) h_0^{\nu+1} < \eta$ for $n \geq n_0$ which entails that $P(\sup_{t\in[h_0,1-h_0]} |R_n^*| > \eta \cap A_n^c) = 0$ for $n \geq n_0$ thus concluding the proof of(A.7).

We get (A.8) using **A.7** and the uniform continuity of $H_j(.)$. $\square$

Bibliografía

[1] Aneiros-Pérez, G. (2001). Estimación de modelos parcialmente lineales con errores dependientes. *Tesis Doctoral, Universidad de Santiago de Compostela.*

[2] Aneiros-Pérez, G. (2002). On bandwidth selection in partial linear regression models under dependence. *Statist. And Prob. Letters*, **57**, 393-401

[3] Aneiros-Pérez, G. y Quintela del Río, G. (2002). Plug-in bandwidth choice in partial linear regression models with autoregressive errors.*J. Statist. Planning and Inference*, **57**, 23-48.

[4] Bianco, A. y Boente, G. (2002). Robust estimators in semiparametric partly linear regression models.

[5] Boente, G. (1989) Fixed carries robust estimates for the derivates of the regression function in heterocedastic nonparametric models. *Trabajos de Matemática*, **156**, Publicaciones del IAM-CONICET.

[6] Boente, G. y Fraiman, R. (1991). Strong uniform convergence rates for some robust equivariante nonparametric regression estimates for mixing processes. *Int. Statist. Rev.*, **59**, 355-372.

[7] Boente, G., Fraiman, R. y Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *J. Statist. Planning and Inference*, **57**, 109-142.

[8] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, **16**, 136-146

[9] Daniel, C. y Wood, G. (1980). *Fitting Equation to Data: Computer Analysis of Multifactor Data for Scientist and Engineers.*2nd Edition. Wiley, New York.

[10] Gasser, T. y Müller, H. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171-185.

[11] Gasser, T. y Härdle, W. (1985). On robust kernel estimation of derivatives of regression functions. *Scand. J. Statist.*, **12**, 233-240.

[12] Hastie, T.J.y Tibshirani, R.J. (1990). *Generalized Additive Models.*Monographs on Statist. Appl. Probab. No. 43. Chapman and Hall, London.

[13] Härdle, W., Liang H. y Gao J. (2000). *Partially Linear Models.* Phisica-Verhag

[14] Härdle, W. y Luckhaus, S. (1988). Uniform consistency of a class of regression function estimator. *Ann. Statist.*, **12**,612-623.

[15] Huber, P. (1981). *Robust Statistics.* Wiley, New York.

[16] Kraser, W. y Welsch, R. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.*, **77**, 595-604.

[17] Linton, O. (1995). Second Order Approximation in the partially linear regression model. *Econometrica*, **63**, 1079-112.

[18] Mallows, C. (1975). On some topics in robustness. *Technical Memorandum,*AT&T Bell Laboratories, Murray Hill.

[19] Rice, J. (1986). Convergence rates for partially splined models. *Statist. and Prob. Letters*, **4**, 203-208.

[20] Robinson, P. (1988). Root-n-consistent Semiparametric regression. *Econometrica,* **56**, 931-954

[21] Rothenberg, T. (1984). Approximate Normality of Generalized least square estimates. *Econometrica*, **52**, 811-825.

[22] Rousseeuw, P. J. (1984). Least median of squares regression. Journal of the American Statistical Association , **79**, 871-88.

[23] Rousseeuw, P. J. y Yohai, V. J.(1984). Robust regression by means of S-estimators. In *Robust and nonlinear time series*, eds. Franke, Hardle y Martin. *Lectures Notes in Statistics*, **29**, 256-272, Springer Verlag, New York.

[24] Rousseeuw, P. y Leroy, A. (1987). *Robust regression and Outlier Detection.* Wiley, New York.

[25] Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc.*, Ser. B, **50**, 413-436.

[26] Wand, M. y Jones, M. (1995). *Kernel Smoothing.* Chapman and Hall, London.

[27] Yohai, V. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.*, **15**, 642-656.

[28] Yohai, V. y Zamar, R. (1988). High breakdown estimates for regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.*, **83**, 406-413.