



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

Tesis de Licenciatura

Estimación Robusta en Modelos Parcialmente Lineales
monótonos

Sebastián Sosa

Directora: Dra. Daniela Rodríguez

Mayo 2016

Índice general

Introducción	7
1. Modelos parcialmente lineales	9
1.1. Introducción	9
1.2. Estimación clásica del modelo	10
1.2.1. Estimación de una densidad	11
1.2.2. Estimadores para ϕ y ϕ_0	14
1.2.3. Estimación de β	16
1.3. Estimación robusta en modelos parcialmente lineales	16
2. Modelos parcialmente lineales con restricciones de orden	19
2.1. Estimadores de regresión bajo restricción de monotonía	19
2.1.1. Estimación clásica	19
2.1.2. Estimación robusta	20
2.2. Estimadores para el modelo parcialmente lineal con restricciones de orden	21
2.2.1. Propuestas basadas en splines	22
2.2.2. Nuestra propuesta	22
3. Estudios de simulación	25
3.1. Resultados y conclusiones	26
4. Un ejemplo con datos reales	35
4.1. Resultados y conclusiones	36
A. Código	39

Introducción

A lo largo del tiempo la estadística ha ido ampliando sus horizontes hasta llegar a abarcar una gran variedad de problemas. Se han destinado muchos esfuerzos a desarrollar procedimientos estadísticos que resulten resistentes a pequeñas perturbaciones sobre los supuestos (métodos robustos), estas perturbaciones son las observaciones que denominamos datos atípicos o outliers. También se ha trabajado sobre procedimientos para ganar en estabilidad respecto a variaciones que puedan surgir respecto del modelo paramétrico asumido.

La inferencia estadística se focalizaba comúnmente sobre funciones de distribución que fuesen puramente paramétricas o puramente no paramétricas. Un modelo paramétrico con una elección correcta de parámetros, presenta como ventaja inferencias precisas, pero cuando el modelo resulta incorrecto, por ejemplo si se violan las hipótesis (aún sutilmente), el escenario cambia y podemos llegar a conclusiones equivocadas debido a sesgos propios del modelo. Por otro lado, los modelos no paramétricos son capaces de lograr mayor adaptabilidad, ayudando en el descubrimiento de estructuras ocultas, pero esta gran flexibilidad tiene su costo, una menor precisión, o sea las conclusiones serán menos específicas. Además con los modelos no paramétricos se introduce la maldición de la dimensión, nombre con el que se conoce al aumento del volumen del espacio donde yacen nuestros datos, haciendo que los mismos queden demasiado dispersos, generando inconvenientes tanto teóricos como computacionales.

En el último tiempo los modelos semiparamétricos ganaron su espacio al ofrecer un balance de las características presentadas por los modelos paramétricos y los no paramétricos. Como ejemplo de esto tenemos a Engle et al. (1986), ellos propusieron utilizar el siguiente modelo parcialmente lineal

$$Y = X^T \beta + g(t) + \epsilon,$$

donde Y , la variable de respuesta, es un variable real, X y t son las variables explicativas, X es un vector de dimensión p , t es una variable de dimensión 1, g una función de los reales en los reales y ϵ el error aleatorio, que desconocemos. Engle et

al. (1986) utilizaron este modelo para estudiar el efecto del clima en la demanda de electricidad.

En esta tesis estudiaremos un modelo parcialmente lineal que nos brinda una mayor flexibilidad que el modelo lineal estandar, pero sin perder su fácil interpretación. La flexibilidad se debe a que contamos con una componente paramétrica y otra no paramétrica, esta elección se justifica cuando pensamos que la respuesta Y depende linealmente de X , pero no así de t . Como una restricción adicional asumiremos que la función g es suave y monótona. Esta restricción surge naturalmente en muchas aplicaciones donde se conoce que g posee estas características y resulta de mucha utilidad obtener un estimador que también sea monótono. Un ejemplo donde esto sucede resulta ser: el incremento de la polución en el aire en función de la cantidad de autos que circulan en el lugar donde se realizan las mediciones. También encontraremos ejemplos en los procesos biológicos relacionados con el crecimiento.

Por otro lado, es conocido que monitorear los datos, identificar outliers, quitarlos y recién despues aplicar los procedimientos de inferencia estadística es complicado y muchas veces es preferible incluir las observaciones atípicas atenuando su influencia a traves de métodos robustos. Algo a tener en cuenta es que quitar outliers reduce el tamaño de la muestra, y esto puede afectar negativamente el cálculo de distribuciones, y además llevarnos a subestimar la varianza del modelo, debido a esto es que resulta preferible realizar un análisis estadístico de los datos con un enfoque robusto.

En esta tesis estudiaremos una propuesta de estimación para β y g , bajo un modelo parcialmente lineal, donde asumimos que la variable X (en \mathbb{R}^p) influye linealmente sobre Y , y la variable t (en el $[0, 1]$) lo hace no linealmente, pero de forma monótona.

El trabajo está organizado en capítulos. En el Capítulo 1, presentamos el modelo parcialmente lineal junto a una descripción de alguno de los posibles métodos de estimación existentes en la literatura tanto clásicos como robustos. El Capítulo 2 resume alguno de los métodos existentes para estimar tanto modelos de regresión como modelos parcialmente lineales bajo la restricción de monotonía y presentaremos una nueva propuesta de estimador robusto para estos modelos bajo restricciones de orden, es decir bajo el supuesto que la componente a ser modelada no paramétricamente sea monótona. El Capítulo 3, resume los resultados de un estudio de simulación diseñado para evaluar la robustez de nuestra propuesta para muestras finitas y su comparación con otras propuestas existentes en la literatura. Finalmente, en el Capítulo 4, presentamos un ejemplo de aplicación con datos reales.

Capítulo 1

Modelos parcialmente lineales

1.1. Introducción

Supongamos que tenemos un conjunto de observaciones (y_i, X_i^T, t_i) que corresponden al modelo

$$Y = X^T \beta + g(t) + \epsilon,$$

donde $X_i^T = (x_{i1}, \dots, x_{ip})$ son observaciones de un vector aleatorio en \mathbb{R}^p , t_i variables aleatorias que supondremos pertenecen al intervalo $[0, 1]$, β es un vector de parámetros desconocidos de dimensión p , g es una función escalar desconocida definida en $[0, 1]$ y ϵ_i son errores independientes con $E[\epsilon_i] = 0$ y $\sigma_\epsilon^2 < \infty$, en el caso clásico o simétricos respecto al 0 en el caso robusto. Además ϵ_i , t_i y X_i son independientes entre si. Supondremos que $\epsilon_i \sim F(. / \sigma_\epsilon)$ donde $\sigma_\epsilon^2 = V(\epsilon) = E(\epsilon \epsilon^T)$, si existe.

Además supongamos que existe una relación entre X_i y t_i , de forma tal que nuestro modelo se puede escribir:

$$\begin{cases} y_i = X_i^T \beta + g(t_i) + \epsilon_i & 1 \leq i \leq n \\ x_{ij} = \phi_j(t_i) + \eta_{ij} & 1 \leq j \leq p \end{cases} \quad (1.1)$$

con η_{ij} variables independientes, $E[\eta_{ij}] = 0$ y $\eta_{ij} \sim G(. / \sigma_\eta)$ G simétrica alrededor de 0, y η_{ij} independientes con t_i . Esto es bajo el supuesto clásico de existencia de esperanza condicional. Notemos que tomando esperanza condicional respecto a t , si existiese, en la segunda ecuación de (1.1) obtenemos

$$\phi_j(t_i) = E[x_{ij} | t_i] \quad 1 \leq i \leq p. \quad (1.2)$$

Si despejamos g de la primera ecuación de (1.1), tenemos

$$y_i - X_i^T \beta = g(t_i) + \epsilon_i, \quad (1.3)$$

luego tomando esperanza condicional respecto a t tenemos

$$E[Y - X\beta | t] = g(t).$$

por lo tanto si llamamos

$$Z = Y - X\beta,$$

entonces tenemos

$$Z = g(t) + \epsilon.$$

Llamemos $\phi(t) = (\phi_1(t), \dots, \phi_p(t))^T$ y sea $\phi_0(t_i) = E[y_i | t_i]$, notemos que por la primera ecuación de (1.1) β satisface la siguiente relación

$$\phi_0(t_i) = \phi^T(t_i)\beta + g(t_i) \quad 1 \leq i \leq n. \quad (1.4)$$

Finalmente, restando $\phi_0(t_i)$ a ambos lados de la primera ecuación de (1.1) reescribimos el modelo de la siguiente forma

$$\begin{cases} y_i - \phi_0(t_i) = (X_i - \phi(t_i))^T \beta + \epsilon_i & 1 \leq i \leq n \\ x_{ij} = \phi_j(t_i) + \eta_{ij} & 1 \leq j \leq p. \end{cases} \quad (1.5)$$

1.2. Estimación clásica del modelo

Ya presentamos nuestro modelo, además bajo el supuesto de una relación entre los X_i y t_i que nos permitió reescribir el modelo parcialmente lineal (1.1) de una forma equivalente (1.5). Esta reescritura nos sugiere una posible forma de estimar los parámetros y funciones desconocidas.

Supongamos que estimamos de alguna manera $\phi(t) = (\phi_1(t), \dots, \phi_p(t))^T$ y ϕ_0 por $\hat{\phi} = (\hat{\phi}_1(t), \dots, \hat{\phi}_p(t))^T$ y $\hat{\phi}_0$ respectivamente, entonces podemos definir unas pseudo observaciones $\tilde{y}_i = y_i - \hat{\phi}_0(t_i)$ y $\tilde{X}_i = X_i - \hat{\phi}(t_i)$. De esta forma podemos calcular un estimador de β a partir de las pseudo observaciones \tilde{y}_i y \tilde{X}_i .

Notemos que si reemplazamos en la primera ecuación de (1.5) las funciones ϕ y ϕ_0 por sus estimadores, obtenemos

$$Y - \hat{\phi}_0 = (X - \hat{\phi})\beta + \epsilon_i.$$

Aquí la única incógnita resulta β , por lo tanto el problema se reduce a estimar el parámetro de regresión de un modelo lineal. Del mismo podemos conseguir un estimador $\hat{\beta}$, por ejemplo a partir del estimador de mínimos cuadrados aplicado a las pseudo observaciones \tilde{y}_i y \tilde{X}_i .

Para completar el proceso de estimación será necesario obtener estimadores para ϕ y ϕ_0 , para ello recordemos que $\phi(t) = (\phi_1(t), \dots, \phi_p(t))^T$, entonces debemos estimar,

$$\begin{aligned} \phi_j(t_i) &= E[x_{ij} | t_i] \quad 1 \leq i \leq n \\ \phi_0(t_i) &= E[y_i | t_i]. \end{aligned}$$

Supongamos que existen la función de densidad conjunta de x_{ij} y t_i , que denotaremos con $f_{j,t}$, y $f_{j|t}$, la densidad condicional a t de x_{ij} , entonces podemos calcular $\phi_j(t_i)$ como

$$E[x_{ij} | t_i = \tau] = \int s f_{j|t}(s|t = \tau) ds = \int s \frac{f_{j,t}(s, \tau)}{f_t(\tau)} ds \quad (1.6)$$

donde f_t es la densidad de t . El problema surge debido a que no conocemos $f_{j,t}$ y f_t , por lo tanto debemos estimarlas. Una posibilidad es hacer una estimación plug-in, es decir reemplazar las funciones de densidad por estimadores.

En la próxima sección repasaremos uno de los métodos de estimación de la densidad más estudiados.

1.2.1. Estimación de una densidad

Sean (x_1, \dots, x_n) observaciones independientes de una variable aleatoria X con función de densidad f . Sea $x \in \mathbb{R}$, queremos estimar el valor de $f(x)$. El objetivo es estimar f a partir de la muestra. Como buscamos que sean los datos los que describan el modelo, y no hacer suposiciones innecesarias, nos limitaremos a estimadores no paramétricos.

Nuestra primera opción, la más simple y tal vez la más difundida, es el histograma. Para ello, hacemos una partición del conjunto donde viven las observaciones, en nuestro caso \mathbb{R} , o sea tomamos $p_0 \leq \min x_i, \max x_i \leq p_m$ y definimos los intervalos $I_j = (p_{j-1}, p_j]$ $j = 1, \dots, m$ donde $p_{j-1} < p_j$.

Llamemos n_j al número de observaciones que caen en el intervalo I_j y f_j a la frecuencia relativa de este intervalo, o sea la proporción de observaciones en I_j ,

$$n_j = \#\{x_i : x_i \in I_j\} = \sum_{i=1}^n \mathbb{1}_{I_j}(x_i)$$

$$f_j = \frac{n_j}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{I_j}(x_i),$$

donde

$$\mathbb{1}_{I_j}(x_i) = \begin{cases} 1 & \text{si } x_i \in I_j \quad j = 1, \dots, m \\ 0 & \text{si no.} \end{cases}$$

Sobre cada intervalo I_j dibujamos un rectángulo con base de longitud $p_j - p_{j-1}$ y la altura correspondiente h_j de manera tal que el área del rectángulo sea igual a f_j .

$$h_j = \frac{f_j}{(p_j - p_{j-1})}.$$

Entonces el estimador de nuestra función de densidad f en x será

$$\hat{f}(x) = \begin{cases} \frac{f_j}{p_j - p_{j-1}} & \text{si } x \in I_j \quad j = 1, \dots, m \\ 0 & \text{si no.} \end{cases}$$

Este estimador resulta discontinuo y podemos buscar mejorarlo si tenemos indicios para creer que nuestra función de densidad es continua. Consideremos X una variable aleatoria con función de densidad continua f en x , entonces

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} = \lim_{h \rightarrow 0} \frac{P(x-h < X < x+h)}{2h}.$$

Notemos que para estimar esta probabilidad, $P(x-h < X < x+h)$, podemos contar la cantidad de observaciones que caen dentro del intervalo y dividir por el número total de observaciones, estos nos da, para h suficientemente pequeño, un estimador

$$\hat{f}(x) = \frac{1}{2h} \frac{\#\{x_i : x_i \in (x-h, x+h)\}}{n}.$$

Observemos que si llamamos F_n a la distribución empírica, podemos reescribir nuestro estimador

$$\hat{f}(x) = \frac{P(X < x+h) - P(X < x-h)}{2h} = \frac{F_n(x+h) - F_n(x-h)}{2h}.$$

Este estimador puede escribirse a partir de una función de pesos w , de la siguiente forma

$$\hat{f}(x) = \sum_{i=1}^n \frac{1}{nh} w\left(\frac{x-x_i}{h}\right).$$

donde $w(x) = \frac{1}{2} \mathbb{1}_{|x| < 1}$. Esta función de pesos w cumple ciertas propiedades,

- $w \geq 0$.
- $\int w(x) dx = 1$.
- $w\left(\frac{x-x_i}{h}\right) = \frac{1}{2}$ si y solo si $x_i \in (x-h, x+h)$, $1 \leq i \leq n$.

A la función w se la conoce como Núcleo Uniforme ya que le otorga igual peso a todas las observaciones que se encuentran en un entorno definido y 0 a las que se encuentran fuera de este entorno.

De lo anterior podríamos preguntarnos ¿cómo le doy menos importancia a las observaciones que se encuentren mas lejos de x ? ¿puedo lograr que el estimador de la función de densidad sea más suave?

Preguntas similares a estas llevaron a investigar distintas funciones de pesos ó núcleos K , algunos ejemplos:

Epanechnikov

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

Biweight

$$K(u) = \begin{cases} \frac{15}{16}(1 - u^2)^2 & \text{si } |u| \leq 1 \\ 0 & \text{si } |u| > 1 \end{cases}$$

Normal o Gaussiano

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}u^2}$$

Tricúbico

$$K(u) = \begin{cases} \frac{70}{81}(1 - u^3)^3 & \text{si } |u| \leq 1 \\ 0 & \text{si } |u| > 1. \end{cases}$$

Por lo tanto, hemos obtenido un estimador de la densidad basado en núcleos, estos han resultado unos de los estimadores no paramétricos más estudiados. Nuestro estimador $\hat{f}(x)$ tiene la forma general

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right).$$

De lo estudiado se desprende que la elección del núcleo no es tan esencial como uno pensaría, si no que por lo general depende más de cuestiones computacionales o de suavidad que uno necesita. Por otro lado, la elección del parámetro h , parámetro de ventana o suavizado, es más importante; veamos a que se debe esta importancia.

Cuando estimamos la densidad por núcleos, estamos centrándonos en cada punto donde deseamos realizar la estimación y realizando un promedio ponderado por la distancia de las observaciones al punto en cuestión. A mayores distancias menos importancia de la observación sobre el promedio. El peso que asignamos a las observaciones queda determinado por la elección del núcleo y el valor asignado al parámetro de ventana o suavizado h .

Si la elección del valor de ventana es muy chico, obtendremos estimadores muy ruidosos o con mayor variabilidad, ya que la estimación realizada en cada punto contará con pocas observaciones. Mientras que un valor de ventana demasiado grande,

nos dará un estimador muy suave, que no logrará captar la esencia de la estructura local de la densidad y por lo tanto, tendremos estimadores sesgados.

Como la función núcleo cumple que $\int K(x)dx = 1$ y además, $K \geq 0$, el estimador \hat{f} también resulta una función de densidad, ya que fácilmente se obtiene $\hat{f}(x) \geq 0$ y $\int \hat{f}(x)dx = 1$.

1.2.2. Estimadores para ϕ y ϕ_0

Comenzamos el estudio de la estimación de una función de densidad motivados por la necesidad de encontrar estimadores para ϕ y ϕ_0 . Retomemos nuestro objetivo, de proveer estimadores para

$$\begin{aligned}\phi_j(t_i) &= \mathbb{E}[x_{ij} | t_i] \quad 1 \leq i \leq p \\ \phi_0(t_i) &= \mathbb{E}[y_i | t_i].\end{aligned}$$

Es decir, el objetivo consiste en estimar la función de regresión. Más precisamente sean X y T variables aleatorias con $X, T \in \mathbb{R}$ y (x_i, t_i) con $1 \leq i \leq n$ una muestra aleatoria con la misma distribución que (X, T) , deseamos estimar $\mathbb{E}[X | T = t]$. Desarrollemos entonces esta esperanza condicional

$$\mathbb{E}[X | T = t] = \int x f_{X|T}(x|t)dx = \int x \frac{f_{XT}(x, t)}{f_T(t)}dx = \frac{m(t)}{f_T(t)},$$

donde llamamos $m(t) = \int x f_{XT}(x, t)dx$.

Como vimos en la sección anterior, un estimador para $f_T(t)$ podría ser:

$$\hat{f}_T(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - T_i}{h}\right).$$

Ahora para $f_{XT}(x, t)$ podemos considerar la extensión natural al caso bivariado

$$\hat{f}_h(x, t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{x - X_i}{h}, \frac{t - T_i}{h}\right),$$

donde K es un núcleo bivariado, o sea $K : \mathbb{R}^2 \rightarrow \mathbb{R}$.

La utilización de la misma ventana h para las dos componentes no es obligatoria, se pueden utilizar distintas ventanas según convenga al problema. Además podemos elegir un núcleo multivariado con la propiedad de ser multiplicativo, de esta forma nos quedaría :

$$\hat{f}_h(x, t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{x - X_i}{h}, \frac{t - T_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \frac{1}{h} K\left(\frac{t - T_i}{h}\right).$$

Si además tenemos que K es tal que $\int sK(s)ds = 0$. Obtenemos,

$$\begin{aligned}\hat{m}_h(t) &= \int x \hat{f}_h(x, t) dx \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t-T_i}{h}\right) \int \frac{x}{h} K\left(\frac{x-X_i}{h}\right) dx \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t-T_i}{h}\right) \int (sh + X_i) K(s) ds \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) X_i.\end{aligned}$$

Luego, un estimador para la esperanza condicional $E[X | T = t]$ resulta:

$$\begin{aligned}\hat{E}[X | T = t] &= \frac{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) X_i}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right)} \\ &= \frac{\sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) X_i}{\sum_{i=1}^n K\left(\frac{t-T_i}{h}\right)}\end{aligned}$$

Así finalmente aplicando estas ideas podemos obtener estimadores para ϕ y ϕ_0 del siguiente modo,

$$\begin{aligned}\hat{\phi}_0(t) &= \frac{\sum_{i=1}^n K\left(\frac{t-t_i}{h}\right) y_i}{\sum_{k=1}^n K\left(\frac{t-t_k}{h}\right)} \\ \hat{\phi}_j(t) &= \frac{\sum_{i=1}^n K\left(\frac{t-t_i}{h}\right) x_{ij}}{\sum_{k=1}^n K\left(\frac{t-t_k}{h}\right)}.\end{aligned}$$

O reescribiendo los pesos

$$w_{ni}(t, h) = \frac{K\left(\frac{t-t_i}{h}\right)}{\sum_{k=1}^n K\left(\frac{t-t_k}{h}\right)},$$

tenemos los estimadores:

$$\begin{aligned}\hat{\phi}_0(t) &= \sum_{i=1}^n w_{ni}(t, h) y_i \\ \hat{\phi}_j(t) &= \sum_{i=1}^n w_{ni}(t, h) x_{ij}.\end{aligned}$$

1.2.3. Estimación de β

Como mencionamos anteriormente una vez obtenidos estimadores para ϕ y ϕ_0 , un estimador de β puede calcularse a partir de aplicar mínimos cuadrados a las pseudo observaciones $\tilde{y}_i = y_i - \hat{\phi}_0(t_i)$ y $\tilde{X}_i = X_i - \hat{\phi}(t_i)$. Mas precisamente, podríamos estimar β de la siguiente manera,

$$\hat{\beta}_{LS} = \underset{b}{\operatorname{argmin}} \sum_{i=1}^n [(y_i - \hat{\phi}_0(t_i)) - (X_i - \hat{\phi}(t_i))^T b]^2.$$

Un modo conveniente de escribir los estimadores obtenidos es a partir de su notación matricial. Denotamos

- $Y = (y_1, y_2, \dots, y_n)^T$
- $\mathbf{X} = (X_1, X_2, \dots, X_n)$
- $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$
- $G = (g(t_1), g(t_2), \dots, g(t_n))^T$
- $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$,

luego podemos expresar el modelo de manera matricial del siguiente modo

$$Y = \mathbf{X}\beta + G + \epsilon.$$

Ademas, si denotamos W la matriz de pesos donde $(W)_{ij} = w_{ij} = w_{ni}(t_j, h)$ tenemos que $\hat{\phi}_j(t_i) = (W\mathbf{X})_{ij}$ y $\hat{\phi}_0(t_i) = (WY)_i$. Por lo tanto, el estimador de mínimos cuadrados de β , $\hat{\beta}_{LS}$ puede calcularse como

$$\hat{\beta}_{LS} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{Y}$$

con $\tilde{Y} = (I - W)Y$, $\tilde{\mathbf{X}} = (I - W)\mathbf{X}$.

1.3. Estimación robusta en modelos parcialmente lineales

Como mencionamos en la introducción, es conocido que la presencia de datos atípicos puede afectar seriamente los estimadores obtenidos en la sección anterior. Por este motivo, Bianco y Boente (2004) propusieron estimadores robustos para el parámetro y la función de regresión en modelos parcialmente lineales. A continuación, daremos un breve resumen de la propuesta introducida por Bianco y Boente (2004), para estimar la parte correspondiente al parámetro de regresión.

En esta situación podrían no estar definidos ϕ_0 y ϕ , pues podría no existir la esperanza. Entonces se definen de la siguiente manera: dada ψ impar, acotada y continua, y $\eta(\tau, a) = E[\psi(\frac{Y-a}{\sigma(\tau)}) | t = \tau]$,

$$\phi_0(\tau) \text{ es tal que } \eta(\tau, \psi_0(\tau)) = 0,$$

donde $\sigma(\tau)$ es una medida de escala robusta, como por ejemplo el $mad(\tau)$, es decir el mad respecto a $F(Y|t = \tau)$.

De la misma manera para $\phi_j(\tau)$, o sea

$$\phi_j(t_i) \text{ es el } a \text{ tal que } E[\psi(\frac{x_{ij} - a}{\sigma(t_i)}) | t = t_i] = 0.$$

Este procedimiento podemos describirlo en 2 pasos:

Paso 1. Estimar $\phi(t)$ y $\phi_0(t)$ por estimadores robustos, en nuestro caso M-estimadores locales que denotaremos $\hat{\phi}_0(t)$ y $\hat{\phi}(t)$.

Paso 2. Estimar el parámetro de regresión usando un estimador de regresión robusto sobre los residuos $y_i - \hat{\phi}_0(t)$ y $X_i - \hat{\phi}$.

En el Paso 1 también podríamos haber usado las medianas locales para estimar ϕ_0 y ϕ . En ese caso, se podrían calcular tomando las medianas respecto a las funciones de distribución condicionales definidas como:

$$\begin{aligned} \hat{F}_0(y|T = t) &= \sum_{i=1}^n w_{ni}(t, h) \mathbb{1}_{(-\infty, y]}(y_i) \\ \hat{F}_j(x|T = t) &= \sum_{i=1}^n w_{ni}(t, h) \mathbb{1}_{(-\infty, y]}(x_{ij}) \quad 1 \leq j \leq p, \end{aligned}$$

donde los w_{ni} son los pesos definidos anteriormente por:

$$w_{ni}(t, h) = \frac{K(\frac{t-t_i}{h})}{\sum_{k=1}^n K(\frac{t-t_k}{h})}.$$

Otra opción sería utilizar M-estimadores locales, donde $\hat{\phi}_0$ y $\hat{\phi}_j$ quedan definidos como solución de

$$\begin{aligned} \sum_{i=1}^n w_{ni}(t, h) \psi\left(\frac{y_i - \hat{\phi}_0(t)}{\hat{s}_0(t)}\right) &= 0 \\ \sum_{i=1}^n w_{ni}(t, h) \psi\left(\frac{x_{ij} - \hat{\phi}_j(t)}{\hat{s}_j(t)}\right) &= 0 \quad 1 \leq j \leq p, \end{aligned}$$

con ψ una función impar, acotada y continua. Por otro lado, $\hat{s}_0(t)$ y $\hat{s}_j(t)$ son estimadores de escala robustos. Algunas posibles elecciones de ψ pueden ser la función de Huber, la función bicuadrada, entre otras, y para los estimadores de escala robustos podemos optar por la mediana de las desviaciones absolutas respecto a la mediana ($\widehat{\text{mad}}$) (Huber 1981) con respecto a las distribuciones condicionales $\hat{F}_0(y|T = t)$ y $\hat{F}_j(x|T = t)$.

Para el Paso 2, la estimación robusta del parámetro de regresión, han sido propuestos diversos métodos como M-estimadores (Huber 1981) resistente a puntos atípicos 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 (Yohai y Zamar, 1988) y MM-estimadores (Yohai, 1987).

Finalmente, podemos obtener un estimador robusto de g basándonos simplemente en los estimadores que ya calculamos. Para ello despejamos la función g de (1.4), por lo tanto

$$g(t_i) = \phi_0(t_i) - \phi^T(t_i) \beta.$$

Utilicemos el método plug-in para reemplazar por los respectivos estimadores robustos y así obtener un estimador robusto de g

$$\hat{g}(t_i) = \hat{\phi}_0(t_i) - \hat{\phi}^T(t_i) \hat{\beta}.$$

Capítulo 2

Modelos parcialmente lineales con restricciones de orden

En la sección anterior resumimos algunos de los procedimientos existentes para estimar en forma clásica y robusta, β y g , el parámetro lineal y la función de regresión, respectivamente en un modelo parcialmente lineal. Sin embargo, si tenemos la información adicional que la función g es monótona, estos estimadores no tienen en cuenta dicha restricción de monotonía. Como mencionamos en la Introducción, en esta Sección estudiaremos una propuesta que nos permite estimar de forma robusta la función g bajo esa restricción.

2.1. Estimadores de regresión bajo restricción de monotonía

2.1.1. Estimación clásica

Los primeros trabajos sobre la estimación de funciones bajo restricciones de orden se remontan a la década del 50. Grenander (1956) estudia estimadores de máxima verosimilitud de una función de densidad decreciente. Brunk (1958, 1970) estudió el problema de estimar la función de regresión monótona en una señal más ruido. Más precisamente, si consideramos el siguiente modelo de regresión

$$z = g(t) + \epsilon,$$

Brunk propuso a partir de una muestra (z_i, t_i) con $1 \leq i \leq n$ y $t_1 \leq \dots \leq t_n$, un estimador de la función de regresión g bajo el supuesto que $g(t_1) \leq \dots \leq g(t_n)$. Este estimador resulta de minimizar la siguiente distancia L^2

$$\sum_{i=1}^n (z_i - g(t_i))^2 \tag{2.1}$$

sobre la clase de funciones no decrecientes continuas a trozos. El estimador que obtenemos no es único, ya que si tenemos una función que minimiza 2.1 podemos construir más funciones que coincidan solamente en los puntos de evaluación y de esta forma también minimicen 2.1. Para obtener unicidad restringimos la clase de funciones al conjunto de funciones constantes continuas a trozos. Otra alternativa es suavizar el estimador resultante a través de alguno de los métodos usuales. El estimador propuesto por Brunk, tiene una expresión explícita dada por

$$\hat{g}(t) = \max_{\{s:t_s \leq t\}} \min_{\{j:t_j \geq t\}} \frac{1}{j-s-1} \sum_{k=s}^j z_k.$$

Para obtener una curva suave de este estimador varios autores propusieron aplicar métodos de suavizados, en algunos casos similares a los que vimos anteriormente. Por ejemplo, Mukerjee (1988) propuso aplicar el estimador de Nadaraya-Watson para obtener un estimador suave

$$\hat{g}_S(t) = \frac{\sum_{j=1}^n K\left(\frac{t_j-t}{h}\right) \hat{g}(t_j)}{\sum_{j=1}^n K\left(\frac{t_j-t}{h}\right)}, \quad (2.2)$$

donde K es un núcleo simétrico, con segundo momento y h es la ventana tal que $\lim_n h_n \rightarrow 0$ y n el tamaño de la muestra. Como consecuencia del suavizado se espera que $\hat{g}_S(t)$ resulte monótono para tamaños de muestras razonables.

2.1.2. Estimación robusta

Los estimadores descritos anteriormente pueden ser muy influenciados por observaciones extremas o puntos atípicos dando como resultado estimadores muy distintos a los deseados. Por lo tanto, es preferible contar con estimadores que sean menos influenciados por observaciones extremas y así poder aplicarlos en estas situaciones. Wang y Huang (2002), analizan la sensibilidad de estos estimadores basados en la distancia L^2 y proponen considerar la minimización respecto a la distancia L^1 , es decir considerar la minimización

$$\sum_{i=1}^n |z_i - g(t_i)| \quad (2.3)$$

bajo el problema restringido. Sin embargo, al igual que para los problemas de posición estos estimadores resultan poco eficientes. Por tal motivo en 2011, Alvarez y Yohai propusieron una familia de estimadores robustos para el modelo de regresión isotónicas que mantienen un compromiso entre eficiencia y robustez. A continuación ofrecemos un detalle simplificado de esta familia y la forma de calcularla.

Sean z_1, \dots, z_n variables aleatorias independientes correspondientes a los tiempos $t_1 \leq \dots \leq t_n$, del modelo

$$z_j = g(t_j) + \epsilon_j, \quad (2.4)$$

donde la tendencia es monótona no decreciente, o sea $g(t_1) \leq \dots \leq g(t_n)$. Alvarez y Yohai (2011), proponen un M-estimador robusto isotónico buscando la función g que minimiza la expresión

$$\sum_{j=1}^n \rho \left(\frac{z_j - g(t_j)}{\hat{\sigma}_n} \right), \quad (2.5)$$

donde, al igual que para el caso de la distancia L^2 , para obtener unicidad necesitamos que g esté restringida a la familia de funciones constantes no decrecientes continuas a trozos, ya que si no, tendríamos múltiples funciones que por coincidir en los puntos t_j también serían solución; $\hat{\sigma}_n$ es un estimador de escala que calculamos previamente. Demás está decir que necesitamos que este estimador sea robusto para que nuestro estimador final también lo sea. Una posible elección podría ser utilizar la MAD (median absolute deviation), la cual se define como $\hat{\sigma} = C \text{ mediana}\{|x_i - \hat{g}_{I,n}|, 1 \leq i \leq n\}$ y $\hat{g}_{I,n} = \text{mediana}\{x_i, 1 \leq i \leq n\}$ y la constante C se elige de forma que el estimador sea consistente para la normal.

Por otra parte, la función ρ debe satisfacer las siguientes propiedades:

- $\rho(x)$ es no decreciente en $|x|$
- $\rho(0) = 0$
- ρ es par
- $\rho(x)$ es estrictamente creciente para $x > 0$
- ρ tiene 2 derivadas continuas y $\psi = \rho'$ es acotada y monótona no decreciente

Alvarez y Yohai (2011), estudian las propiedades de robustez del estimador propuesto a través de la curva de influencia y el punto de ruptura, como así también sus propiedades asintóticas.

2.2. Estimadores para el modelo parcialmente lineal con restricciones de orden

En esta sección retomaremos el estudio de un modelo parcialmente lineal. Recordemos que tenemos un conjunto de observaciones (y_i, x_i, t_i) que corresponden al modelo

$$Y = X^T \beta + g(t) + \epsilon$$

donde $X_i^T = (x_{i1}, \dots, x_{ip})$ son observaciones de un vector aleatorio en \mathbb{R}^p , t_i variables aleatorias que supondremos ordenadas y pertenecientes al intervalo $[0, 1]$ y ϵ_i son errores independientes, e independientes de x_i y t_i . El objetivo es proveer estimadores para β y g pero bajo la restricción de monotonía.

2.2.1. Propuestas basadas en splines

En 2010, Lu propuso un procedimiento de estimación para β y g basado en splines sujeto a la misma restricción de monotonía. Sin embargo, este procedimiento es sensible a la presencia de datos atípicos. Por tal motivo recientemente Du, Sun y Xie (2013) generalizan los resultados obtenidos por Lu (2010) considerando una familia de M – estimadores basados en splines sujeto a dicha restricción.

Para ello, consideran $\mathcal{S}_n(\mathcal{T}_n, l)$ la clase de splines de orden $l > 1$ con nodos en \mathcal{T}_n donde \mathcal{T}_n es un partición de $[a, b]$, $a = t_1 = \dots = t_l < t_{l+1} < \dots < t_{m_n+l+1} = \dots = t_{m_n+2l} = b$. Si $s \in \mathcal{S}_n(\mathcal{T}_n, l)$, existe una base de B-splines $\{B_i : 1 \leq i \leq k_n = m_n + l\}$ tal que $s = \sum_{i=1}^{k_n} \lambda_i B_i$. Sea $\mathcal{M}_n(\mathcal{T}_n, l)$ la sub-clase de $\mathcal{S}_n(\mathcal{T}_n, l)$ definida por

$$\mathcal{M}_n(\mathcal{T}_n, l) = \left\{ \sum_{i=1}^{k_n} \lambda_i B_i : \lambda_1 \leq \dots \leq \lambda_{k_n} \quad \max_i |\lambda_i| < L_n \right\}.$$

Esta sub-clase resulta una colección monótona de splines de acuerdo al teorema 5.9 de Schumaker [19]. Por otro lado, sea ρ una función de tipo Huber, Du, et.al (2013) consideran la siguiente función

$$L_n(\beta, \lambda) = \sum_{i=1}^n \rho \left(y_i - X_i^T \beta - \sum_{j=1}^{k_n} \lambda_j B_j(t_i) \right), \quad (2.6)$$

luego definen estimadores robustos bajo el modelo con restricción de monotonía sobre g minimizando $L_n(\beta, \lambda)$ sujeto a $\lambda_1 \leq \dots \leq \lambda_{k_n}$. Sean $\hat{\beta}$ y $\hat{\lambda}$ los resultantes de esta minimización, entonces obtendremos $(\hat{\beta}, \hat{g}(\cdot)) = (\hat{\beta}, \sum_{i=1}^{k_n} \hat{\lambda}_i B_i(\cdot))$.

Como mencionamos anteriormente, esta propuesta generaliza el trabajo de Lu (2010), donde los estimadores se definen para el caso particular de ρ la función cuadrática.

2.2.2. Nuestra propuesta

En esta sección presentaremos una forma alternativa de estimar el modelo parcialmente lineal bajo restricciones de orden. La idea es combinar el procedimiento de estimación propuesto por Bianco y Boente (2004) con la propuesta de Alvarez y Yohai (2011). Para ello recordemos que el procedimiento de Bianco y Boente (2004) podía ser descripto en los siguientes pasos:

Paso 1. Estimar $\phi(t)$ y $\phi_0(t)$ por estimadores robustos, en nuestro caso M-estimadores locales que denotaremos $\hat{\phi}_0(t)$ y $\hat{\phi}(t)$.

Paso 2. Estimar el parámetro de regresión, que llamaremos $\hat{\beta}$ usando un estimador de regresión robusto sobre los residuos $y_i - \hat{\phi}_0(t)$ y $X_i - \hat{\phi}$.

Por lo tanto, una vez finalizado estos pasos bastará con estimar la función g bajo la restricción de monotonía. Recordemos que despejando de (1.4) y reemplazando β , ϕ y ϕ_0 por sus estimadores, conseguimos un estimador $\hat{g} = \hat{\phi}_0 - \hat{\phi} \hat{\beta}$, pero esto no nos asegura que \hat{g} sea monótona.

Con el fin de considerar esta restricción, notemos que de (1.3) g satisface,

$$y_i - X_i^T \beta = g(t_i) + \epsilon_i,$$

luego si consideramos las pseudo-observaciones $\hat{z}_j = y_i - X_i^T \hat{\beta}$ podemos aplicar el método propuesto por Alvarez y Yohai (2011) a las pseudo-observaciones (\hat{z}_i, t_i) obteniendo así un estimador de g que satisfaga la restricción deseada.

Capítulo 3

Estudios de simulación

En esta sección presentamos los resultados obtenidos al realizar un estudio de simulación diseñado para evaluar el comportamiento de los estimadores propuestos. Para ello comparamos nuestra propuesta con los estimadores clásicos estudiados en Lu (2010) y la propuesta robusta basada en splines introducida en Du, et al. (2013); ambas propuestas fueron descritas en la Sección 2.2.1. Consideramos un modelo con dos diferentes esquemas de contaminación que describiremos a continuación donde en cada caso generamos 1000 replicaciones con distintos tamaños de muestra, siendo $n = 100, 200$ y 500 .

Para la propuesta basada en splines, se utilizaron B-splines cúbicos con una base de dimensión 5 y como función ρ la elegida fue la bicuadrada de Tukey con constante 3.44.

Para evaluar el estimador no paramétrico utilizamos un estimador del error cuadrático medio integrado (MISE) definido de la siguiente manera:

$$MISE = \frac{1}{n} \sum_{s=1}^n [\hat{g}(t_s) - g(t_s)]^2.$$

Para los estimadores del parámetro de regresión, se consideraron estimadores del sesgo, el desvío y error cuadrático medio calculado sobre las 1000 replicaciones e indicado por Bias, SD y MSE en las tablas. En todos los casos se reporta el promedio sobre las 1000 replicaciones de estas medidas y en el caso del MISE también el valor mediano.

En el modelo considerado se generaron muestras del siguiente modo:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + g(t_i) + \epsilon_i \quad 1 \leq i \leq n, \quad (3.1)$$

donde $x_{i1}, x_{i2} \sim \mathcal{N}(0, 1)$ independientes e independientes de $t_i \sim \mathcal{U}[0, 1]$. Se tomó la siguiente función de regresión $g(t) = \sin(\frac{\pi t}{2})$ y los siguientes parámetros de regresión $\beta_1 = 0.5$ y $\beta_2 = 1$. El esquema de contaminación corresponde a las siguientes

alternativas de distribución de los errores, siempre ϵ_i independientes, idénticamente distribuídos e independientes de x_i y t_i , donde:

Caso 0: $\epsilon_i \sim \mathcal{N}(0, 0.5^2)$,

Caso 1: $\epsilon_i \sim 0.9\mathcal{N}(0, 0.5^2) + 0.1U(2, 4)$,

Caso 2: $\epsilon_i \sim \mathcal{N}(5, 0.5^2)$ si la observación t_i satisface $0.4 < t_i < 0.5$.

Nuestros estimadores emplean un procedimiento de suavizado basado en núcleos, como tal tomamos el de Epanechnikov y consideramos las siguientes ventanas $h = 0.15, 0.3, 0.45, 0.6$.

Los casos de contaminación fueron elegidos con el objetivo de poner a prueba la robustez de nuestra propuesta respecto a la componente no paramétrica. Esta elección radica en que nuestra estimación de los parámetros de regresión coincide con la propuesta realizada por Bianco y Boente (2002) cuyas propiedades de robustez frente a muestras pequeñas ya fue evaluada exitosamente en dicho trabajo.

Los algoritmos necesarios para el cálculo de nuestros estimadores fue programado en R y las funciones implementadas se encuentran en el apéndice A. También se desarrollaron algoritmos para las propuestas introducidas en Lu (2010) y Du, et al. (2013) con el fin de realizar las comparaciones. Estos algoritmos no fueron incluídos en el presente trabajo ya que se basaron en una modificación de algoritmos realizados por Pablo Vena, a quien le agradecemos su colaboración.

3.1. Resultados y conclusiones

En tablas y figuras notaremos con CL la propuesta introducida en Lu (2010), utilizaremos RB para la propuesta robusta estudiada en Du, et al. (2013) y finalmente notaremos con ROB_h nuestra propuesta calculada con ventana h .

		$n = 100$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	Bias	0.0003	0.0004	-0.0026	-0.0009	-0.0005	-0.0008
	SD	0.0497	0.0499	0.0678	0.0660	0.0657	0.0657
	MSE	0.0025	0.0025	0.0046	0.0044	0.0043	0.0043
C_1	Bias	0.0026	0.0004	-0.0033	-0.0007	0.0005	0.0003
	SD	0.1051	0.0603	0.0761	0.0728	0.0737	0.0746
	MSE	0.0110	0.0036	0.0058	0.0053	0.0054	0.0056
C_2	Bias	-0.0034	0.0014	-0.0041	-0.0028	-0.0006	-0.0015
	SD	0.1544	0.0876	0.0969	0.0836	0.0756	0.0760
	MSE	0.0238	0.0077	0.0094	0.0070	0.0057	0.0058
		$n = 200$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	Bias	0.0018	0.0018	-0.0004	0.0001	0.0001	0.0000
	SD	0.0351	0.0352	0.0456	0.0445	0.0444	0.0449
	MSE	0.0012	0.0012	0.0021	0.0020	0.0020	0.0020
C_1	Bias	0.0010	0.0013	-0.0031	-0.0014	-0.0008	-0.0006
	SD	0.0747	0.0418	0.0502	0.0495	0.0499	0.0509
	MSE	0.0056	0.0018	0.0025	0.0024	0.0025	0.0026
C_2	Bias	0.0010	0.0024	-0.0013	-0.0005	-0.0006	-0.0011
	SD	0.1102	0.0511	0.0652	0.0541	0.0511	0.0514
	MSE	0.0121	0.0026	0.0042	0.0029	0.0026	0.0026
		$n = 500$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	Bias	-0.0005	-0.0006	-0.0024	-0.0021	-0.0020	-0.0020
	SD	0.0230	0.0232	0.0290	0.0289	0.0290	0.0293
	MSE	0.0005	0.0005	0.0008	0.0008	0.0008	0.0009
C_1	Bias	-0.0008	-0.0008	-0.0028	-0.0023	-0.0022	-0.0020
	SD	0.0474	0.0304	0.0316	0.0314	0.0316	0.0322
	MSE	0.0022	0.0009	0.0010	0.0010	0.0010	0.0010
C_2	Bias	0.0012	0.0017	-0.0026	-0.0022	-0.0020	-0.0020
	SD	0.0696	0.0322	0.0370	0.0318	0.0307	0.0314
	MSE	0.0048	0.0010	0.0014	0.0010	0.0010	0.0010

Tabla 3.1: Resultados para el parámetro β_1

		$n = 100$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	Bias	0.0008	0.0006	-0.0065	-0.0038	-0.0032	-0.0030
	SD	0.0520	0.0524	0.0723	0.0699	0.0688	0.0690
	MSE	0.0027	0.0027	0.0053	0.0049	0.0047	0.0048
C_1	Bias	-0.0005	0.0011	-0.0128	-0.0070	-0.0060	-0.0055
	SD	0.1024	0.0655	0.0873	0.0819	0.0811	0.0820
	MSE	0.0105	0.0043	0.0078	0.0068	0.0066	0.0068
C_2	Bias	0.0021	-0.0033	-0.0061	-0.0075	-0.0047	-0.0040
	SD	0.1547	0.0911	0.1016	0.0843	0.0776	0.0783
	MSE	0.0239	0.0083	0.0104	0.0072	0.0060	0.0062
		$n = 200$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	Bias	0.0011	0.0013	0.0000	0.0014	0.0019	0.0022
	SD	0.0343	0.0346	0.0463	0.0459	0.0458	0.0460
	MSE	0.0012	0.0012	0.0021	0.0021	0.0021	0.0021
C_1	Bias	0.0007	0.0014	-0.0032	-0.0002	0.0010	0.0019
	SD	0.0730	0.0428	0.0523	0.0510	0.0513	0.0521
	MSE	0.0053	0.0018	0.0028	0.0026	0.0026	0.0027
C_2	Bias	0.0067	0.0008	-0.0002	0.0008	0.0016	0.0022
	SD	0.1095	0.0506	0.0630	0.0515	0.0492	0.0499
	MSE	0.0120	0.0026	0.0040	0.0026	0.0024	0.0025
		$n = 500$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	Bias	-0.0002	-0.0002	-0.0010	-0.0006	-0.0003	-0.0001
	SD	0.0238	0.0238	0.0290	0.0288	0.0289	0.0293
	MSE	0.0006	0.0006	0.0008	0.0008	0.0008	0.0009
C_1	Bias	0.0035	0.0008	-0.0012	-0.0002	0.0002	0.0005
	SD	0.0464	0.0292	0.0323	0.0321	0.0322	0.0327
	MSE	0.0022	0.0008	0.0010	0.0010	0.0010	0.0011
C_2	Bias	-0.0014	-0.0003	-0.0016	-0.0013	-0.0010	-0.0006
	SD	0.0678	0.0322	0.0390	0.0337	0.0322	0.0325
	MSE	0.0046	0.0010	0.0015	0.0011	0.0010	0.0011

Tabla 3.2: Resultados para el parámetro β_2

		$n = 100$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	MISE	0.007	2.7381	0.0206	0.0204	0.0203	0.0202
	MedMISE	0.0055	0.0056	0.0206	0.0204	0.0203	0.0202
C_1	MISE	0.1202	6.5024	0.0378	0.0373	0.0371	0.0372
	MedMISE	0.1054	0.0097	0.0378	0.0373	0.0371	0.0372
C_2	MISE	0.3624	19.6443	0.0223	0.0219	0.0215	0.0214
	MedMISE	0.3192	0.0071	0.0223	0.0219	0.0215	0.0214
		$n = 200$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	MISE	0.0038	3.9453	0.0126	0.0126	0.0125	0.0125
	MedMISE	0.0031	0.0031	0.0126	0.0126	0.0125	0.0125
C_1	MISE	0.1051	7.2652	0.0213	0.0211	0.021	0.021
	MedMISE	0.0976	0.0055	0.0213	0.0211	0.021	0.021
C_2	MISE	0.3468	11.3339	0.0136	0.0134	0.0134	0.0133
	MedMISE	0.3295	0.0038	0.0136	0.0134	0.0134	0.0133
		$n = 500$					
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
		CL	RB	ROB _{0.15}	ROB _{0.3}	ROB _{0.45}	ROB _{0.6}
C_0	MISE	0.0017	0.7717	0.0066	0.0066	0.0066	0.0066
	MedMISE	0.0014	0.0014	0.0066	0.0066	0.0066	0.0066
C_1	MISE	0.0974	16.2912	0.0104	0.0104	0.0104	0.0104
	MedMISE	0.0948	0.0034	0.0104	0.0104	0.0104	0.0104
C_2	MISE	0.3309	9.0213	0.0071	0.0071	0.007	0.007
	MedMISE	0.3277	0.0016	0.0071	0.0071	0.007	0.007

Tabla 3.3: Resultados para la componente no paramétrica g .

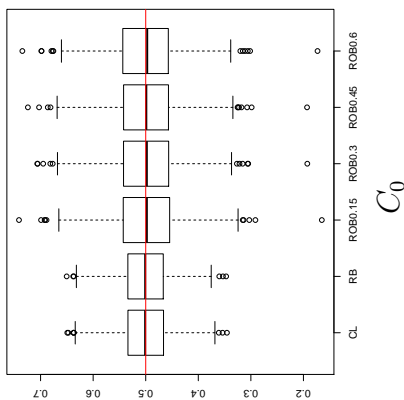
La Tabla 3.4 resume el número de veces donde el MISE del estimador de la componente no paramétrica introducido en Du, et al. (2013) resulta superior a 10. Tanto los estimadores clásicos definidos por Lu (2010) como nuestra propuesta no reporta ningún caso con MISE mayor a este valor. En los gráficos correspondientes al estimador de densidad del MISE que presentaremos a continuación hemos eliminado estos valores.

$n = 100$			$n = 200$			$n = 500$		
C_0	C_1	C_2	C_0	C_1	C_2	C_0	C_1	C_2
9	16	157	8	6	97	4	7	77

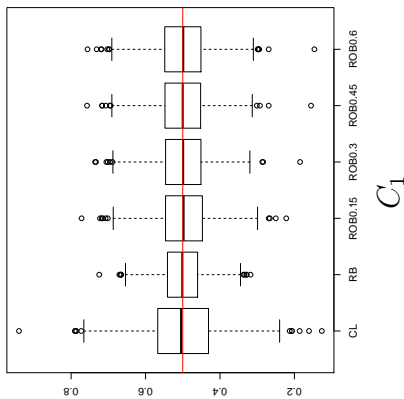
Tabla 3.4: Número de veces donde el MISE es mayor a 10 sobre las 1000 replicaciones.

$n = 100$

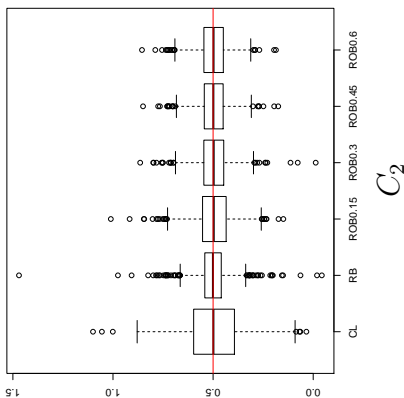
C_0



C_1

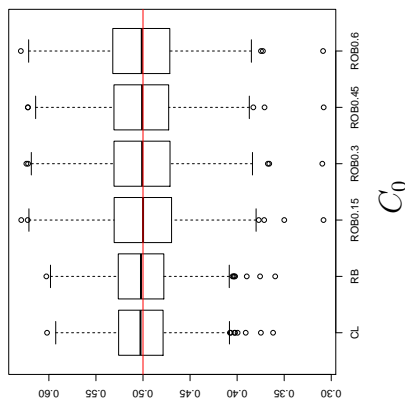


C_2

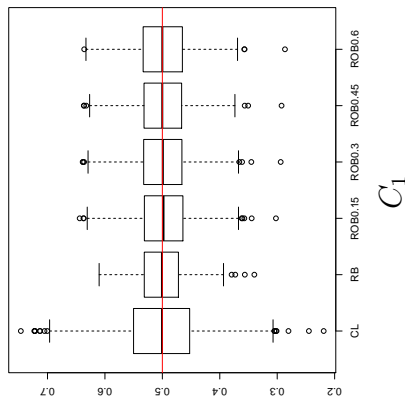


$n = 200$

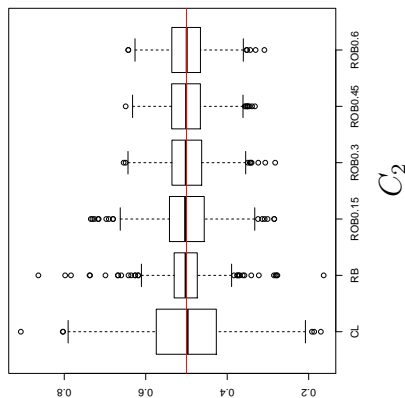
C_0



C_1

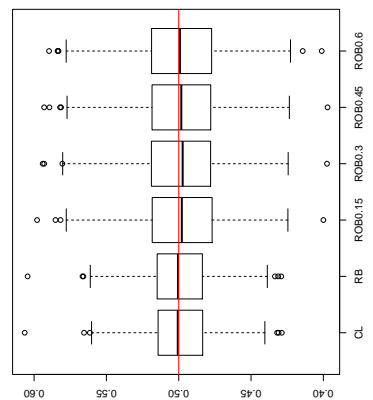


C_2

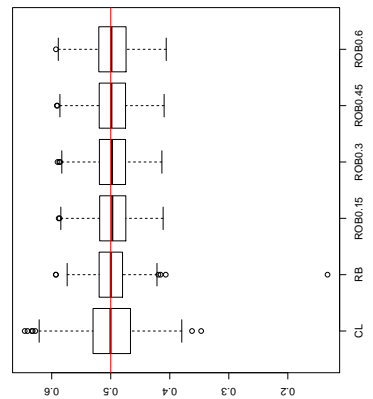


$n = 500$

C_0



C_1



C_2

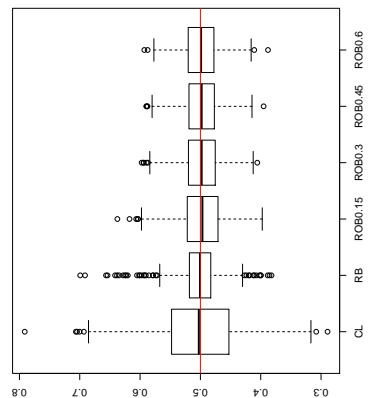
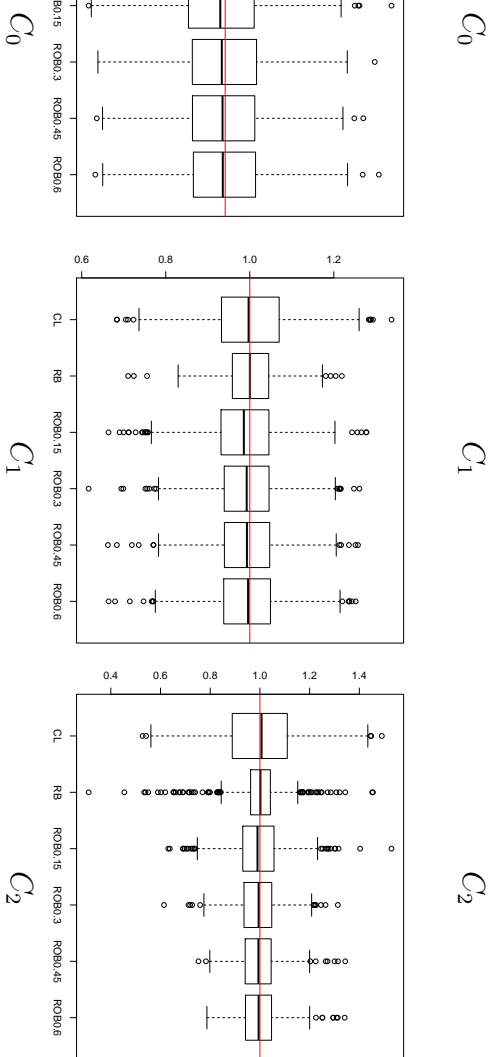


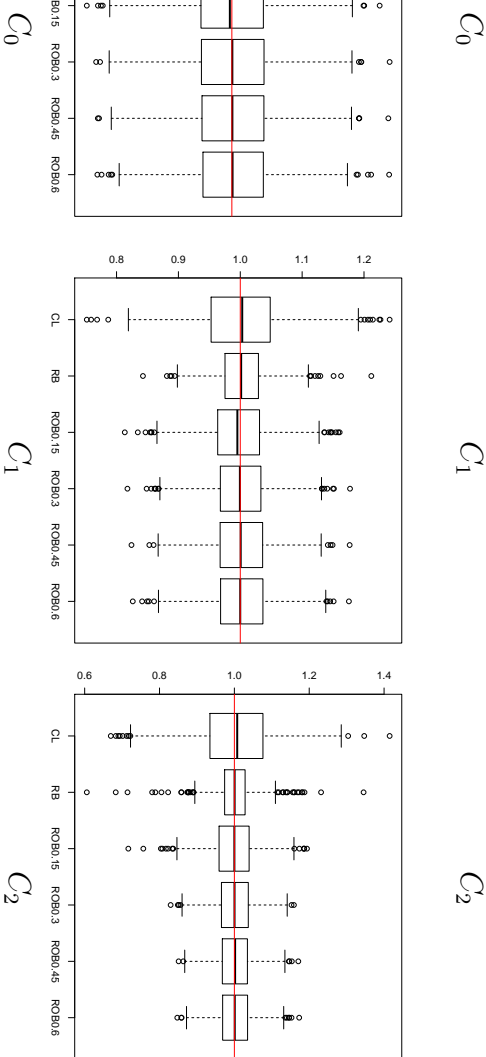
Figura 3.1: Resultados para el parámetro β_1

$n = 100$



Estimación de modelos parcialmente lineales isotónicos

$n = 200$



$n = 500$

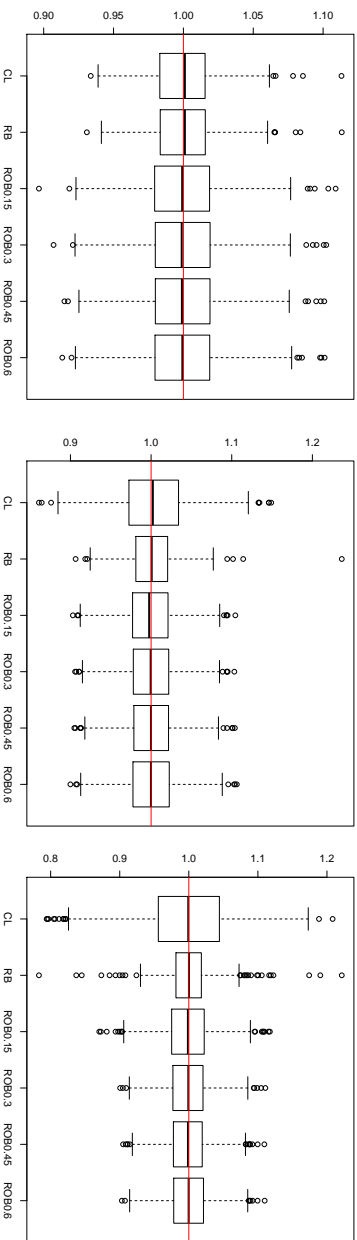


Figura 3.2: Resultados para el parámetro β_2

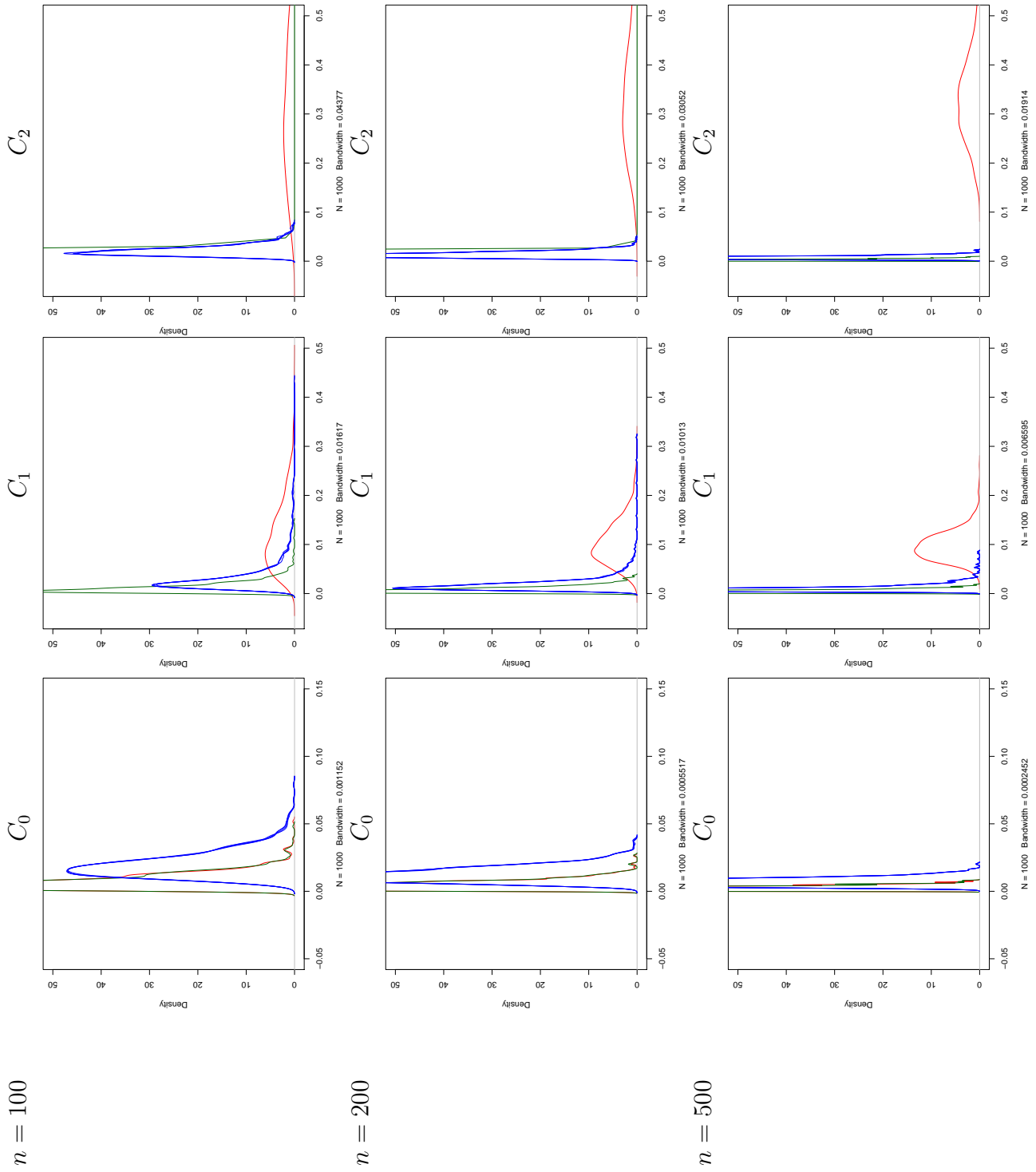


Figura 3.3: Resultados para la componente no paramétrica g .

El estudio de simulación nos aporta información valiosa sobre los distintos estimadores con los que nos estamos comparando. De las Tablas 3.1 y 3.2 observamos que los sesgos, desvios y MSEs de los estimadores propuestos para los parámetros de regresión son similares y el comportamiento se ajusta a lo esperado, a medida que aumentan las observaciones, mejoran.

Además notamos que nuestros estimadores no varían mucho con el cambio de ventana, mostrando un desempeño ligeramente peor que el del estimador clásico para el escenario sin contaminación e invirtiéndose esto en los escenarios con contaminación.

Cuando analizamos las Figuras 3.1 y 3.2, confirmamos lo dicho en el párrafo anterior, el estimador clásico es el adecuado cuando los datos se ajustan a las hipótesis, pero cuando nos alejamos de ese escenario, los estimadores robustos para los parámetros β_1 y β_2 son preferibles.

Hasta aquí no habría motivos aparentes para optar por nuestro estimador por sobre la versión robusta basada en splines, pero esto cambia cuando revisamos los resultados en la Tabla 3.3. Al observar el desempeño del estimador robusto basado en splines para la componente no paramétrica de g vemos que tiene peor rendimiento incluso comparado con el estimador clásico cuando existen outliers. Esto se debe a que el estimador no toma en cuenta la escala al no dividir por una escala robusta dentro de la función ρ . Por otra parte cuando evaluamos nuestro estimador según la componente no paramétrica g obtenemos un MISE comparable al del estimador clásico cuando no hay outliers, y lo mejoramos sustancialmente con outliers más fuertes.

Capítulo 4

Un ejemplo con datos reales

En esta sección aplicaremos el método propuesto para analizar la relación entre la polución del aire en una ruta, el número de autos y algunas variables meteorológicas. Es el mismo conjunto de datos utilizado en Lu (2010) y Du, et al. (2013). Las mediciones se realizaron en Alnabru, un barrio de Oslo, Noruega, entre Octubre de 2001 y Agosto de 2003 por la Administración de Carreteras Públicas de Noruega <http://www.vegvesen.no/>. Los datos fueron puestos a disposición por Magne Aldrin (2004).

Lu (2010) analizó el conjunto de datos usando un modelo de regresión parcialmente lineal monótono basado en cuadrados mínimos, mientras que Du et al. (2013) lo hicieron con M-estimadores, ambas propuestas fueron resumidas en la Sección 2.2.1.

Estos datos corresponden a una muestra de 500 observaciones cuyo objetivo, como mencionamos anteriormente, consistió en estudiar si la polución del aire en una ruta está relacionada con el nivel de tráfico y las condiciones meteorológicas. La variable de respuesta y , contiene el logaritmo de la concentración de NO_2 (partículas) cada hora. Los datos originales contienen 6 variables predictoras correspondientes a las siguientes medidas: logaritmo del número de autos por hora, temperatura en $^{\circ}C$ a 2 metros sobre el nivel del suelo, velocidad del viento (m/s), la diferencia de temperatura medida a 25 mts y 2 mts sobre el nivel del suelo $^{\circ}C$, dirección del viento (grados entre 0 y 360), hora del día, cantidad de días respecto al 1^{ro} de Octubre de 2001.

Al igual que en Lu (2010) y Du, et al. (2013), restringimos nuestro análisis a la variable de respuesta y (logaritmo de la concentración de NO_2 por hora), la variable predictora t correspondiente al logaritmo del número de autos por hora, x_1 la temperatura en $^{\circ}C$ a 2 metros sobre el nivel del suelo, x_2 la velocidad del viento en m/s y x_3 es la diferencia de temperatura medida a 25 mts y 2 mts sobre el nivel del suelo $^{\circ}C$.

Por lo tanto, utilizaremos el siguiente modelo

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + g(t) + \epsilon.$$

Como hay evidencia empírica de que a mayor nivel de tráfico la polución es mayor, o sea, a mayor número de autos se tiene una mayor concentración de NO_2 , es razonable asumir la monotonía y no linealidad de g y tiene sentido, entonces, calcular nuestros estimadores para el modelo parcialmente lineal bajo esta restricción.

Nuestros estimadores emplean un procedimiento de suavizado basado en núcleos, como núcleo tomamos el de Epanechnikov y consideramos las siguientes ventanas $h = 0.3, 0.5, 1, 1.5, 2, 3$.

4.1. Resultados y conclusiones

En tablas y figuras notaremos con CL la propuesta introducida en Lu (2010), utilizaremos RB para la propuesta robusta estudiada en Du, et al. (2013) y finalmente notaremos con ROB_h nuestra propuesta calculada con ventana h .

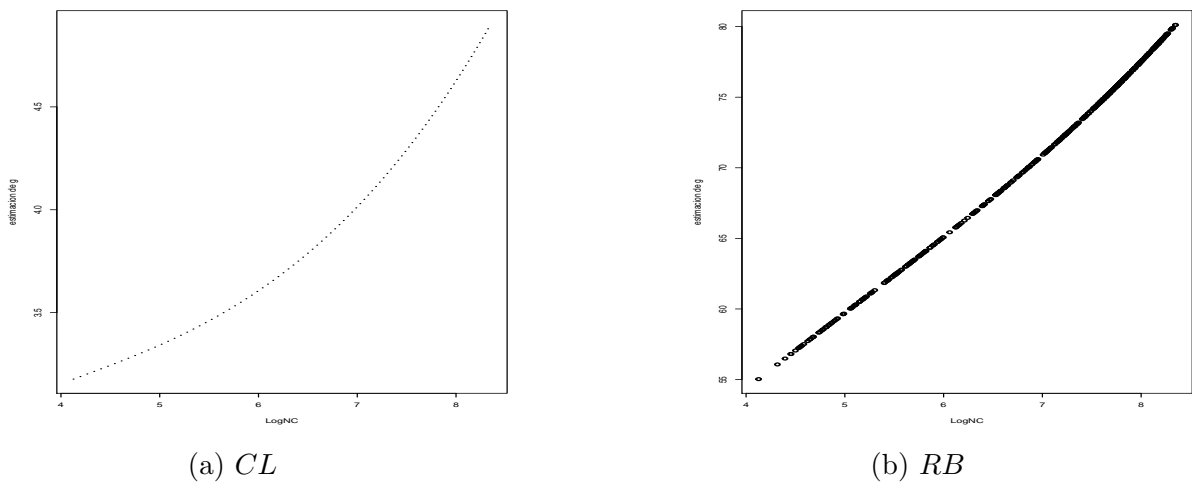


Figura 4.1: Resultados para la componente no paramétrica g sin contaminación para las propuestas clásicas (CL) y robusta (RB).

En la figura 4.1b podemos observar lo que ya habíamos mencionado, en el sentido de que el estimador robusto basado en splines parece estar fuera de escala y esto es efectivamente así, ya que no contempla la dispersión en la definición del estimador.

Por otra parte, observamos en la Figura 4.2 que nuestro estimador de g no varía significativamente al cambiar el parámetro suavizado, y hemos decidido continuar con el análisis utilizando $h = 1.5$.

Finalmente, introducimos un conjunto de outliers, contaminando la muestra de la siguiente manera. Hemos cambiado la variable de respuesta por el valor 7 para

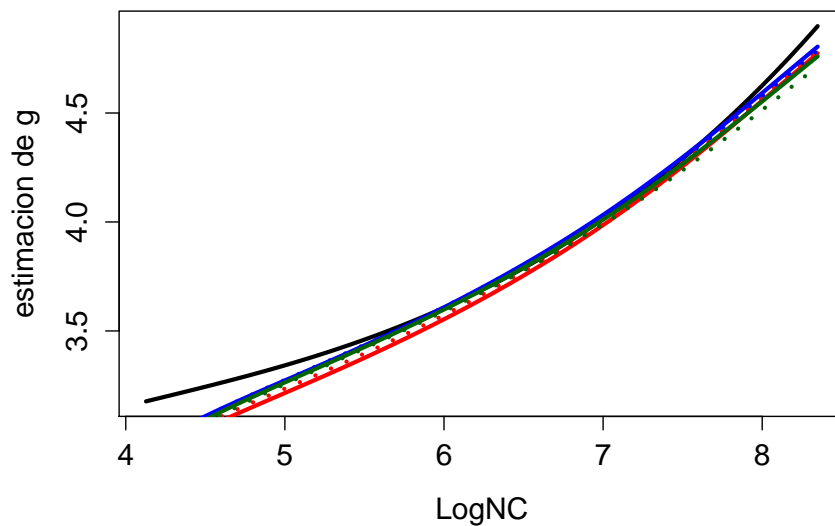


Figura 4.2: Resultados para la componente no paramétrica g sin contaminación para las propuestas clásicas CL (en negro) y robusta para diferentes ventanas $ROB_{0.3}$ (rojo solido), $ROB_{0.5}$ (rojo punteado), ROB_1 (azul solido), $ROB_{1.5}$ (punteado solido), ROB_2 (verde solido) y ROB_3 (verde punteado).

aquellos valores cuya variable t se encontraba en el intervalo $(4.5, 5)$, correspondiendo a un total de 31 observaciones.

Podemos observar en la Figura 4.4 el efecto devastador que una muestra contaminada por outliers, como la que se observa en la Figura 4.3, puede producir en la estimación de la componente no paramétrica g al utilizar estimadores no robustos.

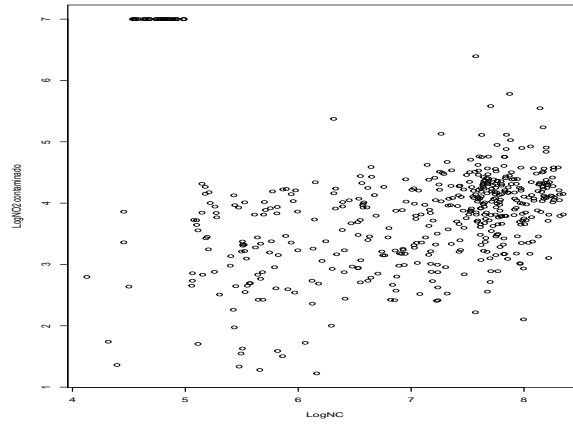


Figura 4.3: Datos contaminados.

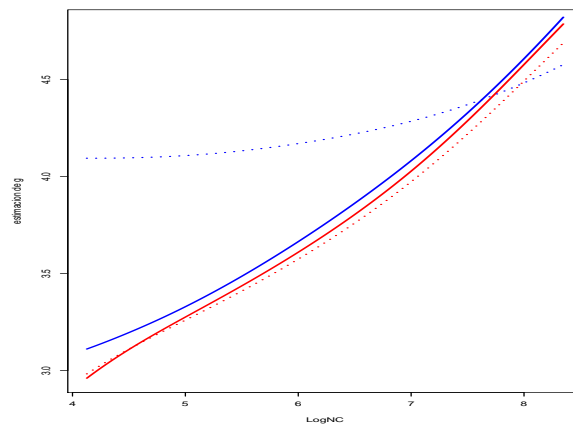


Figura 4.4: Resultados para la componente no paramétrica g con contaminación. $ROB_{1.5}$ (rojo), CL (azul), en líneas solidas denotamos el caso sin contaminar mientras que en línea punteada la estimación bajo contaminación.

Apéndice A

Código

```
library(robustbase) #usamos lmrob
library(Hmisc) #para usar wtd.quantile "Weighted Statistical Estimates"

monotonopl <- function( xx, yy, xt, ventana)
{
  #-----
  # Ordenamos las observaciones
  #-----
  tt_order <- order(xt)
  n <- dim(xx)[1]
  p <- dim(xx)[2]
  tt <- as.matrix(xt[tt_order],ncol=n)

  yy <- as.matrix(yy[tt_order],ncol=n)

  for (j in 1:p)
  {
    xx[,j] <- as.matrix(xx[,j][tt_order],ncol=n)
  }

  #-----
  # Calculo de Phi_{0} y Phi_{j}
  #-----

  phi0_sombrero <- as.matrix(etasombrero(yy,tt,ventana))

  phi_sombrero <- matrix(0,ncol=p,nrow=n)
```

```

for (j in 1:p)
{
  phi_sombrero[,j] <- etasombrero(as.vector(xx[,j]),tt,ventana)
}

#-----
# Calculo de betasombrero
#-----

Y <- yy - phi0_sombrero
X <- xx - phi_sombrero

pesosds <- pesos.dimp.DS(X)
modelo <- lmrob(Y ~ X-1, weights=pesosds)

betasombrero <- rep(0,p)
lista_coeficientes <- names(modelo$coefficients)

for (j in 1:p)
{
  betasombrero[j] <- modelo$coefficients[lista_coeficientes[j]]
}

#-----
# Calculo de gsombrero usando Zsombrero = g(tt) + error
#-----

Zsombrero <- yy
for (j in 1:p)
{
  Zsombrero <- Zsombrero -(xx[,j] * betasombrero[j])
}

# La funcion musombrero asume que todos los datos se encuentran ordenados
# de menor a mayor segun tt
gsombrero <- musombrero(Zsombrero,tt)

return(list(gestim =gsombrero, b = betasombrero))
}

```



```

etasombrero <- function(zz,tt,ventana)
{
  lz <- length(zz)
  eta <- rep(0,lz)
  medianalocal <- rep(0,lz)
  madlocal <- rep(0,lz)

  wt <- pesos_en_los_ts(tt,ventana)

  for(i in 1:lz)
  {
    medianalocal[i] <- wtd.quantile(zz, weights=wt[i,],type=c('i/n'),probs=c(.5)
    ,normwt=TRUE)
    madlocal[i] <- wtd.quantile(abs(zz-medianalocal[i]), weights=wt[i,]
    ,type=c('i/n'),probs=c(.5),normwt=TRUE)
  }

  for (i in 1:lz)
  {
    eta[i] <- optimize(objetivo_eta, c(-1000,1000), pesos_filai=wt[i,],zz
    , escala=madlocal[i])$minimum
  }
  eta
}

objetivo_eta <- function(aa,pesos_filai,zz,escala)
{
  sum(pesos_filai*rho.huber((zz-aa)/escala))
}

rho.huber <- function(x,k=1.345)
{
  ifelse(abs(x) <= k,(x*x)/2,k*abs(x)-(k*k)/2)
}

pesos_en_los_ts <- function(xt,ventana)
{
  lxt <- length(xt)
  wt <- matrix(0,lxt,lxt)

  for(i in 1:lxt )

```

```

    {
      punto <- xt[i]
      wt[,i] <- nucleo(punto,xt,ventana)
    }
    #divido por la suma
    sumas <- apply(wt,2,sum)
    wt <- wt/sumas
  }

#nucleo de epanechnikov
nucepan <- function(x)
{
  a <- 0.75*(1-x*x)
  kepan <- a*(abs(x)<=1)
  kepan
}

#me da el Kh
nucleo <- function(punto,xt,ventana)
{
  arg <- (punto-xt)/ventana
  nucleovec <- nucepan(arg)
  nucleovec
}

#-----
# Alvarez - Yohai
#-----
# La funcion musombrero asume que las observaciones se encuentran ordenadas
# de menor a mayor segun tt.

musombrero <- function(zz,tt)
{
  lz <- length(zz)
  mu.temp <- matrix(NA,nrow=lz,ncol=lz)

  #calculo de escala sigma_n inicial y un mu inicial
  m0 <- ltsReg(zz ~ 1)
  betaini <- coef(m0)
  escalaini <- m0$scale

```

```
#variables para simplificar las operaciones
unos <- as.matrix(rep(1,lz))
aa <- seq(1,lz,by=1)

for (ii in 1:lz)
{
  for (j in ii:lz)
  {
    pesosuv <- 1*(aa>=ii)*(aa<=j)
    zzuv <- zz [pesosuv>0]
    unosuv <- as.matrix (unos[pesosuv>0])
    modelo <- lmrob..M..fit(unosuv,zzuv, beta.initial = betaini,
    scale =escalaini,
    control = lmrob.control(tuning.psi = 3.44,psi = 'bisquare'))
    mu.temp[ii,j] <- coef(modelo) #$$coefficient
  }
}

#ahora calculo mu en cada tt[i].
mu <- rep(NA,lz)
aux <- as.matrix(mu.temp[1,1:lz])
minenv <- apply(aux,2,min)
mu[1] <- max(minenv)

aux <- t(as.matrix(mu.temp[1:lz,lz]))
minenv <- apply(aux,2,min)
mu[lz] <- max(minenv)

lz1 <- lz-1
for(i in 2:lz1)
{
  aux <- t(as.matrix(mu.temp[1:i,i:lz]))
  minenv <- apply(aux,2,min)
  mu[i] <- max(minenv)
}
mu
}
```

```
#-----  
#Funciones para calcular los pesos Donoho-stahel usa la libreria rrcov  
#para el S-estimador usa fastloc  
#-----  
  
library(rrcov)  
source("fastsloc.txt")  
  
pesos.dimp.DS <- function(x,percentil=0.975)  
{  
  p <- ncol(x)  
  DSestim <- CovSde(x)  
  media <- DSestim@center  
  cova <- DSestim@cov  
  norma2.x <- mahalanobis(x,media,cova)  
  
  w1 <- wbi(norma2.x,qchisq(percentil,p))  
  w1  
}  
  
wbi <- function(u,d)  
{  
  w <- 1*(abs(u)<=d)  
  v <- w*(1-(u/d)^2)^2  
  v  
}
```

Bibliografía

- [1] Aldrin, M., *ftp://rcom.univie.ac.at/mirrors/lib.stat.cmu.edu/datasets/NO2.dat*, magne.aldrin@nr.no, 2004.
- [2] Álvarez, E. and Yohai, V. *M-estimators for Isotonic Regression*, J. Statist. Plann. Inference, 142, 2351–2368, 2012.
- [3] Bianco, A. and Boente, G., *Robust estimators in semiparametric partly linear regression models*, J. Statist. Plann. Inference, 122, 229–252, 2004.
- [4] Brunk, H. *On the Estimation of Parameters restricted by Inequalities*, Annals of Mathematical Statistics, 29, 437–454, 1958.
- [5] Brunk, H. *Estimation of Isotonic Regression*, Nonparametric Techniques in Statistical Inference, 177–195, Cambridge Univ. Press., 1970.
- [6] Du, J., Sun, Z. and Xie, T. *M-estimation for the partially linear regression model under monotonic constraints*, Statistics and Probability Letters, 83, 1353–1363, 2013.
- [7] Engle, R., Granger, W., Rice, J. and Weiss, A. *Semiparametric Estimates of the Relation between Weather and Electricity Sales*, Journal of the American Statistical Association, 81, 310–319, 1986.
- [8] Fan, J. and Gijbels, I. *Local Polynomial Modelling and Its Applications*, Chapman & Hall, 1996.
- [9] Grenander, U. *On the theory of mortality measurement, Part II.*, Scandinavian Actuarial Journal, Issue 2, 125–153, 1956.
- [10] Huber, P. *Robust Statistics*. Wiley, New York, 1981.
- [11] Kraser, W. and Welsch, R. *Efficient bounded-influence regression estimation*. J. Amer. Statist. Assoc., 77, 595–604, 1982.

- [12] Lu, M. *Spline-based sieve maximum likelihood estimation in the partly linear model under monotonicity constraints*, Journal of Multivariate Analysis, 101,2528–2542, 2010.
- [13] Mallows, C. On some topics in robustness. *Technical Memorandum*, AT&T Bell Laboratories, Murray Hill ,1975.
- [14] Maronna, R.; Martin, D. and Yohai, V. *Robust Statistics: Theory and Methods*, John Wiley & Sons, 2006.
- [15] H. Mukerjee, *Monotone nonparametric regression*, Ann. Statist. 16, 741-750, 1988.
- [16] Pilz, K. and Dette, H., *A comparative study of monotone nonparametric kernel estimates*, Journal of Statistical Computation and Simulation, Volume 76, Issue 1, 2006.
- [17] Rousseeuw, P. J. Least median of squares regression. Journal of the American Statistical Association , 79, 871-88, 1984.
- [18] Rousseeuw, P. J. and Yohai, V. J. Robust regression by means of S-estimators. In *Robust and nonlinear time series*, eds. Franke, Hardle and Martin. *Lectures Notes in Statistics*, 29, 256-272, Springer Verlag, New York, 1984.
- [19] L. Schumaker, *Spline Functions: Basic Theory*, Wiley, New York, 1981.
- [20] Wang, Y., and Huang, J., *Limiting Distribution for Monotone Median Regression*. J. Statist. Plann. Inference, 108, 281–287, 2002.
- [21] Yohai, V. High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.*, 15, 642-656,1987.
- [22] Yohai, V. and Zamar, R. High breakdown estimates for regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.*, 83, 406-413,1988.