



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

Tesis de Licenciatura

Una propuesta de estimación robusta en modelos parcialmente lineales
con errores en las variables

Paula Mercedes Spano

Directora: Dra. Ana M. Bianco

Fecha

Una propuesta de estimación robusta en modelos parcialmente lineales con errores en las variables

En el modelo parcialmente lineal suponemos que observamos vectores aleatorios independientes $(Y_i, \mathbf{X}'_i, T_i)'$, $i = 1, \dots, n$, $Y_i \in \mathbb{R}$, $T_i \in \mathbb{R}$, $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})' \in \mathbb{R}^p$ tales que $Y_i = \boldsymbol{\beta}'\mathbf{X}_i + g(T_i) + \varepsilon_i$ para cada i , donde los errores ε_i son independientes, idénticamente distribuidos e independientes de $(\mathbf{X}'_i, T_i)'$. En esta tesis estamos interesados en la estimación del parámetro de regresión $\boldsymbol{\beta}$ y de la función de regresión g cuando las covariables \mathbf{X}_i son observadas con error. Es decir, en lugar de observar \mathbf{X}_i , observamos $\mathbf{V}_i = \mathbf{X}_i + \mathbf{U}_i$ para $i = 1, \dots, n$ siendo los errores \mathbf{U}_i independientes, idénticamente distribuidos e independientes de $(Y_i, \mathbf{X}'_i, T_i, \varepsilon_i)'$. Sin embargo, los métodos existentes de estimación para este modelo basados en mínimos cuadrados pueden verse seriamente afectados por la presencia de datos atípicos. El objetivo de esta tesis es introducir estimadores robustos de las componentes paramétrica y no paramétrica bajo condiciones generales. Probaremos la consistencia de los estimadores propuestos. Un estudio de simulación permitirá comparar la performance de estos estimadores con su versión clásica.

Palabras Claves: Modelos parcialmente lineales; Robustez; Errores en las Variables.

Agradecimientos

Índice general

1. Introducción	1
2. Estimación no paramétrica de la densidad	4
2.1. Histograma	4
2.1.1. Construcción	4
2.1.2. Sesgo	5
2.1.3. Varianza	6
2.1.4. Error Cuadrático Medio y Error Cuadrático Medio Integrado	6
2.1.5. Ventana Óptima	7
2.1.6. Histograma Medio Corrido	8
2.2. Estimadores de Núcleos	8
2.2.1. Motivación y Derivación	8
2.2.2. Sesgo	9
2.2.3. Varianza	10
2.2.4. Error Cuadrático Medio y Error Cuadrático Medio Integrado	10
2.2.5. Selección del Parámetro de Suavizado	11
3. Regresión No Paramétrica	14
3.1. Regresión Univariada basada en Núcleos	14
3.1.1. Introducción	14
3.1.2. Regresión basada en núcleos	15
3.1.3. Propiedades Estadísticas	16
3.2. Selección del Parámetro de Suavizado	17
3.2.1. Una mirada más de cerca al Error Cuadrático Promediado	18
3.2.2. Convalidación Cruzada	19

3.2.3.	Funciones Penalizadoras	19
4.	Robustez	21
4.1.	El Modelo de Posición	21
4.1.1.	M-estimadores de Posición	22
4.1.2.	M-estimadores Redescendientes	24
4.2.	Estimadores de Dispersión	24
4.3.	M-estimadores de Escala	25
4.4.	La Función de Influencia	26
4.5.	El Punto de Ruptura	27
4.6.	Robustez en el Análisis Multivariado	28
4.6.1.	Punto de Ruptura de Estimadores Multivariados	28
4.6.2.	M-estimadores	29
4.6.3.	Estimadores basados en una escala robusta	31
5.	Regresión Lineal con Errores-en-las-Variables	33
5.1.	Modelo	33
5.2.	Alternativas robustas	34
5.3.	Regresión Ortogonal Pesada Robusta en el Modelo Errores-en-las-Variables	36
5.3.1.	Función de Influencia y Propiedades Asintóticas	38
6.	Estimación Robusta en el Modelo Parcialmente Lineal con Errores-en-las-Variables	41
6.1.	Introducción	41
6.2.	Modelo y Enfoque Clásico	42
6.3.	Propuesta	43
6.4.	Consistencia	44
7.	Estudio de Monte Carlo	48
7.1.	Conclusiones	50
7.2.	Cuadros	52
7.3.	Gráficos	55

Capítulo 1

Introducción

Los modelos paramétricos son los modelos más difundidos en Estadística y se basan en el supuesto de que las observaciones siguen una distribución conocida, determinada por uno o más parámetros. Estos parámetros son en general el objeto de interés de la inferencia y tanto puede interesar estimarlos como realizar intervalos de confianza o tests de hipótesis. Estos modelos son sólo una aproximación a la realidad y podría ocurrir que los supuestos que subyacen bajo el modelo sean relativamente fuertes en tanto, eventualmente, ningún modelo paramétrico podría generar los datos. Asimismo, muchos de los procedimientos tradicionalmente usados en Estadística son muy sensibles a las suposiciones realizadas y podrían conducir a conclusiones erróneas si éstas no se satisficieran.

En las últimas décadas se ha desarrollado el interés por otro enfoque más flexible que es el que ofrecen los modelos no paramétricos, en particular en el ámbito de regresión, que es el de nuestro interés. En regresión el objetivo es ajustar una curva que exprese la relación entre una variable de respuesta y una o varias covariables. En los modelos de regresión no paramétrica no se propone a priori una relación paramétrica específica entre la respuesta y las covariables, mientras que en el enfoque paramétrico suponemos una forma funcional específica predeterminada que es descrita en términos de parámetros a estimar. El enfoque no paramétrico permite abordar relaciones complejas entre las variables, capturando características inusuales o inesperadas de la relación entre las mismas. Con los métodos paramétricos podemos arribar a inferencias de mayor precisión, aunque posiblemente incorrectas si el modelo no se cumple, en cambio con el enfoque no paramétrico obtendremos mayor estabilidad aunque menos precisión.

En un modelo de regresión no paramétrico asumimos que observamos vectores aleatorios $(Y_i, \mathbf{X}_i)'$, $i = 1, \dots, n$, independientes e idénticamente distribuidos tales que $Y_i \in \mathbb{R}$, $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})' \in \mathbb{R}^p$, que cumplen

$$Y_i = m(\mathbf{X}_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (1.1)$$

siendo ε_i errores independientes e independientes de \mathbf{X}_i .

Los métodos usados en un modelo de regresión no paramétrico como el dado en (1.1) están basados en técnicas de suavizado multivariadas en tanto el dominio de la función de regresión m es multidimensional. Estas técnicas padecen la conocida *maldición de la dimensionalidad*, que esencialmente consiste en que al crecer la dimensión p de las covariables la vecindad entorno al punto \mathbf{x} de interés se hace cada vez más vacía, esto se traduce en una menor velocidad de convergencia de

los estimadores de m .

Este problema llevó a buscar otros modelos alternativos tales como los aditivos, que fueron introducidos por Hastie y Tibshirani (1990) como generalización del modelo de regresión lineal conservando su sencilla interpretación. Una estrategia intermedia es la que ofrecen los modelos semiparamétricos o parcialmente lineales, que estudiaremos en esta tesis, que combinan la facilidad de la interpretación de los modelos lineales con la flexibilidad de los modelos no paramétricos.

En los modelos parcialmente lineales suponemos que se observa una muestra de tamaño n tal que $(Y_i, \mathbf{X}_i', T_i)'$, con $i = 1, \dots, n$, son observaciones independientes, $Y_i \in \mathbb{R}$, $T_i \in \mathbb{R}$, $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})' \in \mathbb{R}^p$ y

$$Y_i = \beta' \mathbf{X}_i + g(T_i) + \varepsilon_i, \quad 1 \leq i \leq n,$$

donde los errores ε_i son independientes, idénticamente distribuidos e independientes de $(\mathbf{X}_i', T_i)'$.

Los modelos parcialmente lineales son más flexibles que el modelo lineal clásico ya que tienen una componente paramétrica y otra no paramétrica. Pueden ser una alternativa adecuada cuando se sospecha que la respuesta tiene una relación lineal con las covariables \mathbf{X} , pero que está relacionada no linealmente con la covariable T . Éste podría ser el caso cuando el investigador conoce más de la dependencia de Y con \mathbf{X} que de la relación entre Y y la variable predictora T .

Estos modelos han sido estudiados, entre otros, por: Ansley y Wecker (1983), Green et al. (1985), Denby (1986), Heckman (1986), Engle et al. (1986), Rice (1986), Chen (1988), Robinson (1988), Speckman (1988), Chen y Chen (1991), Chen y Shiau (1991, 1994), Gao (1992), Gao y Zhao (1993), Gao y Liang (1995), He y Shi (1996) y Yee y Wild (1996) que investigaron algunos resultados asintóticos para el caso en el que el suavizado involucra splines, núcleos o vecinos más cercanos.

Sin embargo, es bien sabido que en regresión lineal como en regresión no paramétrica los estimadores de mínimos cuadrados pueden ser seriamente afectados por datos atípicos. Bianco y Boente (2004) propusieron métodos robustos para la estimación del parámetro de regresión y la función de regresión y estudiaron algunas de sus propiedades asintóticas.

En nuestro caso estamos interesados en la estimación del parámetro desconocido β y de la componente no paramétrica g cuando las covariables \mathbf{X}_i son observadas con error. Es decir, en lugar de observar \mathbf{X}_i , observamos

$$\mathbf{V}_i = \mathbf{X}_i + \mathbf{U}_i \quad 1 \leq i \leq n,$$

siendo los errores \mathbf{U}_i independientes, idénticamente distribuidos e independientes de $(Y_i, \mathbf{X}_i', T_i, \varepsilon_i)'$. En el modelo clásico se supone además que $E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma_\varepsilon$ y además que $E(\mathbf{U}_i) = \mathbf{0}$, $\Sigma_{\mathbf{U}_i} = \sigma_{\mathbf{U}} \mathbf{I}_p$.

El estimador de verosimilitud semiparamétrico de Severini y Staniswalis (1994) implementado ignorando que las \mathbf{X}_i 's son observadas con error, lleva a un estimador de β que no es consistente. Liang, Härdle y Carroll (1999) propusieron un estimador consistente de β basado en mínimos cuadrados en el que introducen un **término de atenuación** que depende de la matriz de covarianza de los errores \mathbf{U}_i , que asumen conocida. Sin embargo, las propuestas existentes son sensibles a la presencia de datos anómalos en tanto se basan en mínimos cuadrados. El propósito de esta tesis es introducir estimadores robustos consistentes de las componentes paramétrica y no paramétrica de un modelo parcialmente lineal con errores en las variables bajo condiciones generales.

Si bien en esta tesis centramos nuestra atención en el problema de regresión, se incluyó el Capítulo 2, en el que tratamos un tópico relacionado, que es el de estimación no paramétrica de

la densidad, con el fin de introducir algunos conceptos y técnicas que son comunes al problema de estimación de la función de regresión. En dicho capítulo consideramos primero al histograma y luego una alternativa basada en núcleos a fin de superar las propiedades indeseables que presenta el histograma.

En el Capítulo 3 estudiamos la regresión no paramétrica para el caso en que la covariable es una variable aleatoria unidimensional, es decir un modelo univariado con diseño aleatorio. Vemos el método de regresión basado en núcleos, algunas propiedades estadísticas, como así también la selección del parámetro de suavizado, del cual depende el estimador obtenido.

Con el fin de obtener estimadores robustos presentamos en el Capítulo 4 algunos de ellos para posición y escala introduciendo diferentes medidas para estudiar su sensibilidad a la presencia de datos atípicos, así como su generalización al caso multivariado.

En el Capítulo 5 introducimos la regresión lineal para el modelo de errores-en-las-variables. Estudiamos con más detalle la alternativa robusta propuesta por Ruiz-Gazen y Fekri (2004) por ser la que aplicamos en el estudio de simulación.

En el Capítulo 6 presentamos el problema de la estimación robusta en el modelo parcialmente lineal con errores-en-las-variables, que es el tema de interés de esta tesis. Se propone un procedimiento de tres pasos y estudiamos la consistencia de los estimadores obtenidos. Finalmente, en el Capítulo 7 realizamos un estudio de Monte Carlo para analizar la performance de los mismos en muestras pequeñas tanto contaminadas como sin contaminar.

Capítulo 2

Estimación no paramétrica de la densidad

Éste es un capítulo que nos servirá para introducir la teoría de estimación no paramétrica de una densidad. Más detalles sobre el contenido de este capítulo pueden hallarse en Härdle, Müller, Sperlich y Werwatz (2004).

Sean X una variable aleatoria continua y f su función de densidad. Esta última nos dice “cómo está distribuida la X ” y de ella podemos calcular la media y la varianza de X (si existen) y, a partir de ella, la probabilidad de que X tome valores en un cierto intervalo. Así, la función de densidad es muy útil para caracterizar la distribución de la variable aleatoria X , pero en general es desconocida.

2.1. Histograma

Supongamos que tenemos n observaciones independientes X_1, \dots, X_n de X y que, por lo tanto, todas provienen de la misma distribución llamada $f(x)$. Nuestro objetivo es estimar f en un cierto valor x a partir de datos *i.i.d.* (independientes e idénticamente distribuidos). El estimador de densidad más usado es el “*histograma*”.

2.1.1. Construcción

Supongamos que tenemos una muestra aleatoria X_1, \dots, X_n de alguna distribución continua desconocida. Elegimos un origen x_0 y dividimos la recta real en intervalos de tamaño h y definimos el *intervalo* o *clase* $B_j = [x_0 + (j - 1)h, x_0 + jh)$, $j \in \mathbb{Z}$.

Luego, el histograma está dado por

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n \sum_j I(X_i \in B_j) I(x \in B_j) \quad (2.1)$$

donde

$$I(X_i \in B_j) = \begin{cases} 1 & \text{si } X_i \in B_j \\ 0 & \text{caso contrario.} \end{cases}$$

Notemos que la fórmula (2.1) da un estimador de f para todo valor x . Si llamamos m_j al centro del intervalo B_j , el histograma le asigna a cada valor $x \in B_j = [m_j - \frac{h}{2}, m_j + \frac{h}{2})$ el mismo estimador para f , que llamaremos $\hat{f}_h(m_j)$.

Si aproximamos el área debajo de la curva de la función de densidad entre $m_j - \frac{h}{2}$ y $m_j + \frac{h}{2}$ por una barra con altura $f(m_j)$ y ancho h , tenemos que:

$$P\left(X \in \left[m_j - \frac{h}{2}, m_j + \frac{h}{2}\right)\right) = \int_{m_j - \frac{h}{2}}^{m_j + \frac{h}{2}} f(u) du \approx f(m_j)h \quad (2.2)$$

Un estimador natural para esta probabilidad es la frecuencia relativa de las observaciones en este intervalo

$$P\left(X \in \left[m_j - \frac{h}{2}, m_j + \frac{h}{2}\right)\right) \approx \frac{1}{n} \# \left\{ X_i \in \left[m_j - \frac{h}{2}, m_j + \frac{h}{2}\right)\right\}. \quad (2.3)$$

Combinando (2.2) y (2.3) obtenemos

$$\hat{f}_h(m_j) = \frac{1}{nh} \# \left\{ X_i \in \left[m_j - \frac{h}{2}, m_j + \frac{h}{2}\right)\right\}.$$

La forma del histograma depende del origen x_0 y de la ventana h del intervalo o clase. La suavidad del estimador dado por el histograma depende de la elección de la ventana, a medida que la ventana crece el histograma será más suave. Sin embargo, sin un criterio determinado es muy difícil decidir cuál de ellas dará el grado *óptimo* de suavidad.

2.1.2. Sesgo

Supongamos que el origen es $x_0 = 0$ y que queremos estimar la densidad en algún punto $x \in B_j = [(j-1)h, jh)$. El estimador asignado por el histograma es

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n I(X_i \in B_j), \quad (2.4)$$

donde $I(X_i \in B_j)$ tiene distribución Bernoulli y por lo tanto,

$$E\left(\hat{f}_h(x)\right) = \frac{1}{h} \int_{(j-1)h}^{jh} f(u) du.$$

En general, éste no es un estimador insesgado de $f(x)$. En efecto

$$\text{Sesgo}\left(\hat{f}_h(x)\right) = E\left(\hat{f}_h(x) - f(x)\right) = \frac{1}{h} \int_{(j-1)h}^{jh} f(u) du - f(x).$$

La expresión exacta del sesgo dependerá de la verdadera densidad $f(x)$, que es desconocida, pero mediante un desarrollo de Taylor de primer orden alrededor del punto medio del intervalo $m_j = (j - \frac{1}{2})h$ podemos aproximarlos de la siguiente forma

$$\begin{aligned} \text{Sesgo}(\widehat{f}_h(x)) &= \frac{1}{h} \int_{(j-1)h}^{jh} (f(u) - f(x)) du \approx \frac{1}{h} \int_{(j-1)h}^{jh} f'(m_j)(u - x) du \\ &\approx f'(m_j)(m_j - x). \end{aligned}$$

Por lo tanto,

$$E(\widehat{f}_h(x) - f(x)) \approx f'(m_j)(m_j - x), \quad (2.5)$$

que se anula cuando $x = m_j$.

2.1.3. Varianza

A partir de (2.4) tenemos que

$$\text{Var}(\widehat{f}_h(x)) = \text{Var}\left(\frac{1}{nh} \sum_{i=1}^n I(X_i \in B_j)\right).$$

Dado que las X_i son variables aleatorias i.i.d. obtenemos que

$$\text{Var}(\widehat{f}_h(x)) = \frac{1}{n^2 h^2} \sum_{i=1}^n \text{Var}(I(X_i \in B_j)) = \frac{1}{n^2 h^2} n \text{Var}(I(X_i \in B_j)).$$

Teniendo en cuenta que las variables $I(X_i \in B_j)$ son Bernoulli es posible deducir que la varianza del histograma es

$$\text{Var}(\widehat{f}_h(x)) = \frac{n}{n^2 h^2} \int_{B_j} f(u) du \left(1 - \int_{B_j} f(u) du\right).$$

Puede verse que esta expresión puede ser aproximada mediante la fórmula

$$\text{Var}(\widehat{f}_h(x)) \approx \frac{1}{nh} f(x).$$

Observemos que aumentando h , la varianza se reduce, pero mirando (2.5) vemos que ocurre lo contrario para el sesgo. Por lo tanto, debemos encontrar un valor de h que nos dé el compromiso óptimo entre reducción de varianza y de sesgo.

2.1.4. Error Cuadrático Medio y Error Cuadrático Medio Integrado

Consideremos el error cuadrático medio (MSE) del histograma

$$MSE(\widehat{f}_h(x)) = E\left(\left\{\widehat{f}_h(x) - f(x)\right\}^2\right),$$

que puede ser escrito como

$$MSE\left(\widehat{f}_h(x)\right) = Var\left(\widehat{f}_h(x)\right) + \left[Sesgo\left(\widehat{f}_h(x)\right) \right]^2.$$

Se puede mostrar para el error cuadrático medio del histograma que

$$MSE\left(\widehat{f}_h(x)\right) = \frac{1}{nh}f(x) + f'\left(\left(j - \frac{1}{2}\right)h\right)^2 \left[\left(j - \frac{1}{2}\right)h - x\right]^2 + o(h) + o\left(\frac{1}{nh}\right), \quad (2.6)$$

donde $o(h)$ y $o\left(\frac{1}{nh}\right)$ denotan términos que son de orden menor que h y $\frac{1}{nh}$, respectivamente.

A partir de (2.6) podemos concluir que el MSE de $\widehat{f}_h(x)$ tiende a cero cuando $h \rightarrow 0$ y $nh \rightarrow \infty$, es decir que es necesario que la longitud del intervalo converja a 0, pero no tan rápido. Dado que la convergencia del error cuadrático medio implica la convergencia en probabilidad, $\widehat{f}_h(x)$ es un estimador consistente de $f(x)$.

La utilidad de la fórmula del MSE es dudosa en la práctica ya que depende de la densidad desconocida f , por esta razón se usa una medida global de la precisión del estimador. La más usada es el *error cuadrático medio integrado* ($MISE$) definida por

$$MISE(\widehat{f}_h) = \int_{-\infty}^{\infty} MSE\left(\widehat{f}_h(x)\right) dx.$$

Dado x , usando (2.6), para h tendiendo a 0 podemos realizar la siguiente aproximación

$$MISE(\widehat{f}_h) \approx \frac{1}{nh} + \frac{h^2}{12} \int (f'(x))^2 dx.$$

Dada una función g a valores reales, notemos el cuadrado de la norma L_2 de g como $\|g\|_2^2 = \int (g(x))^2 dx$. Luego, definimos el $MISE$ asintótico ($AMISE$) como

$$AMISE(\widehat{f}_h) = \frac{1}{nh} + \frac{h^2}{12} \|f'\|_2^2. \quad (2.7)$$

2.1.5. Ventana Óptima

Ahora podemos definir un criterio para la selección de la ventana óptima: elegir la ventana h que minimice al $AMISE$. Derivando (2.7) respecto de h e igualando a cero, obtenemos que

$$h_0 = \left(\frac{6}{n \|f'\|_2^2} \right)^{1/3} \sim n^{-1/3}$$

es la ventana óptima. Sin embargo, notemos que en esta expresión f' es desconocida.

Por ejemplo, si suponemos que f es la densidad de la $N(0, 1)$, se puede mostrar que

$$\|f'\|_2^2 = \frac{1}{4\sqrt{\pi}},$$

y en ese caso

$$h_0 = \left(\frac{24\sqrt{\pi}}{n} \right)^{1/3} \approx 3.5n^{-1/3}.$$

2.1.6. Histograma Medio Corrido

Como ya observamos la ventana h no es el único parámetro del que dependen la forma y apariencia del histograma. También dependen de la elección del origen. Una solución natural podría ser calcular histogramas usando la misma ventana, pero diferentes orígenes y luego promediarlos. Esta idea da lugar al *histograma promediado o Averaged Sshifted Histogram (ASH)* que no depende del origen y parece corresponder a un h más pequeño que el de los histogramas usados para su construcción.

Consideremos un histograma con origen $x_0 = 0$ e intervalos $B_j = [(j-1)h, jh)$, $j \in \mathbb{Z}$, es decir

$$\dots B_1 = [0, h), \quad B_2 = [h, 2h), \quad B_3 = [2h, 3h), \dots$$

Generemos $M-1$ intervalos nuevos corriendo a cada B_j hacia la derecha en lh/M :

$$B_{jl} = [(j-1 + l/M)h, (j + l/M)h), \quad l \in \{1, \dots, M-1\}.$$

Si calculamos un histograma para cada uno de los M conjuntos de intervalos, tendremos M estimadores distintos para f en cada x

$$\hat{f}_{h,l}(x) = \frac{1}{nh} \sum_{i=1}^n \left\{ \sum_j I(X_i \in B_{jl}) I(x \in B_{jl}) \right\}.$$

El *ASH* se obtiene promediando estos estimadores

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{Mh} \sum_{l=0}^{M-1} \sum_j I(X_i \in B_{jl}) I(x \in B_{jl}) \right\}.$$

Cuando $M \rightarrow \infty$, el *ASH* deja de depender del origen.

2.2. Estimadores de Núcleos

2.2.1. Motivación y Derivación

Notemos que, aún cuando el *ASH* parece resolver el problema de la elección del origen, el histograma tiene algunas propiedades indeseables:

- El histograma le asigna a cada x en $[m_j - \frac{h}{2}, m_j + \frac{h}{2})$ el mismo estimador para f .
- El histograma no es una función continua, no es derivable en los saltos y tiene derivada cero en varios lugares. Esto conduce a la apariencia especialmete inadecuada si queremos estimar una función de densidad continua y suave.

Una alternativa fue presentada por Rosenblatt (1956) al introducir el método de *estimación de la densidad basada en núcleos*.

Un método razonable para estimar $f(x)$ es basarse en un intervalo centrado en x . Para eso consideramos intervalos de la forma $[x - h, x + h)$. Así podemos escribir

$$\hat{f}_h(x) = \frac{1}{2hn} \#\{X_i \in [x - h, x + h)\}.$$

Esta fórmula puede ser reescrita definiendo una función de peso, llamada función de núcleo uniforme, definida por

$$K(u) = \frac{1}{2} I(|u| \leq 1),$$

y tomando $u = (x - X_i)/h$. Luego, obtenemos

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} I\left(\left|\frac{x - X_i}{h}\right| \leq 1\right).$$

Parecería razonable darle más peso a las contribuciones de observaciones que estén más cerca de x que a aquellas que estén más lejos. Para eso podemos elegir distintos tipos de funciones núcleo K . Como por ejemplo el núcleo “normal” que está dado por

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right). \quad (2.8)$$

La siguiente es una forma general de definir el estimador de la densidad f basado en núcleos. Sea X_1, X_2, \dots, X_n una muestra aleatoria donde X_i tiene densidad f . Definimos el estimador de densidad por el método de núcleos de f como

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

donde

$$K_h(\cdot) = \frac{1}{h} K(\cdot/h),$$

siendo $K(\cdot)$ alguna función núcleo.

Al igual que en el histograma, h controla la suavidad del estimador y la elección de la ventana es un problema crucial.

Las funciones núcleo K usadas son, en general, funciones de densidad. Supondremos que $K : \mathbb{R} \rightarrow \mathbb{R}$ es una función par, tal que $\int K(u) du = 1$ y $K \geq 0$. En este caso, el estimador basado en núcleos también es una función de densidad. Además, \hat{f}_h heredará todas las propiedades de continuidad y diferenciabilidad de la K .

Por un lado trataremos de encontrar una ventana h óptima, pero dado un valor h no garantiza el mismo grado de suavidad si usamos diferentes funciones núcleo.

2.2.2. Sesgo

El sesgo del estimador de núcleo de la densidad es

$$\text{Sesgo}(\hat{f}_h(x)) = E(\hat{f}_h(x) - f(x))$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^n E \{K_h(x - X_i)\} - f(x) \\
&= E \{K_h(x - X_1)\} - f(x).
\end{aligned}$$

Mediante un desarrollo de Taylor de segundo orden y usando la simetría del núcleo K , se puede probar que, cuando h tiende a 0, el sesgo del estimador de la densidad basado en núcleos es

$$Sesgo(\widehat{f}_h(x)) = \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2), \quad (2.9)$$

donde $\mu_2(K) = \int s^2 K(s) ds$.

Como observando (2.9) vemos que el sesgo es proporcional a h^2 , debemos elegir un h pequeño para que se reduzca el sesgo. Además, el $Sesgo(\widehat{f}_h(x))$ depende de $f''(x)$.

2.2.3. Varianza

La varianza del estimador puede calcularse como

$$\begin{aligned}
Var(\widehat{f}_h(x)) &= Var\left\{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i)\right\} \\
&= \frac{1}{n^2} \sum_{i=1}^n Var\{K_h(x - X_i)\} \\
&= \frac{1}{n} Var\{K_h(x - X_1)\}.
\end{aligned}$$

Puede verse que si $nh \rightarrow \infty$, entonces

$$Var(\widehat{f}_h(x)) = \frac{1}{nh} \|K\|_2^2 f(x) + o\left(\frac{1}{nh}\right). \quad (2.10)$$

Como la varianza es proporcional a $(nh)^{-1}$, tenemos que elegir h lo más grande posible para hacerla disminuir.

2.2.4. Error Cuadrático Medio y Error Cuadrático Medio Integrado

Como observamos si aumentamos h la varianza disminuye, mientras que el sesgo aumenta. Ocurre lo contrario si disminuimos el valor de h . Por lo tanto, la minimización del MSE representa un compromiso entre varianza y sesgo.

El comportamiento asintótico del MSE nos da una forma de determinar si el estimador de la densidad basado en núcleos es consistente. De (2.9) y (2.10) obtenemos que

$$MSE(\widehat{f}_h(x)) = \frac{h^4}{4} [f''(x)]^2 \mu_2(K)^2 + \frac{1}{nh} \|K\|_2^2 f(x) + o(h^4) + o\left(\frac{1}{nh}\right). \quad (2.11)$$

De (2.11) concluimos que el MSE tiende a cero cuando $h \rightarrow 0$ y $nh \rightarrow \infty$. Por lo tanto, el estimador de la densidad basado en núcleos es consistente. Como en el caso del histograma depende de f y en este caso también de f'' , que son desconocidas.

Respecto del Error Cuadrático Medio Integrado, obtenemos que si $h \rightarrow 0$ y $nh \rightarrow \infty$, se verifica

$$MISE(\hat{f}_h) = \int MSE(\hat{f}_h(x)) dx = \frac{1}{nh} \|K\|_2^2 + \frac{h^4}{4} [\mu_2(K)]^2 \|f\|_2^2 + o\left(\frac{1}{nh}\right) + o(h^4),$$

Luego, el Error Cuadrático Medio Integrado Asintótico es

$$AMISE(\hat{f}_h) = \int MSE(\hat{f}_h(x)) dx = \frac{1}{nh} \|K\|_2^2 + \frac{h^4}{4} [\mu_2(K)]^2 \|f\|_2^2.$$

Por último, la ventana óptima determinada con el criterio de minimizar el $AMISE$ resulta

$$h_{opt} = \left(\frac{\|K\|_2^2}{\|f\|_2^2 \{\mu_2(K)\}^2 n} \right)^{-1/5} \sim n^{-1/5}. \quad (2.12)$$

2.2.5. Selección del Parámetro de Suavizado

En las siguientes dos subsecciones introduciremos dos de los métodos más usados de selección de ventana, el método plug-in y el método de convalidación cruzada.

La Regla Básica de Silverman

El método plug-in debe su nombre a su principio subyacente: si se tiene una expresión que incluye a un parámetro desconocido, reemplazarlo con un estimador del mismo.

Tomando (2.12) como ejemplo y suponiendo que $f \sim N(\mu, \sigma^2)$, tenemos que

$$\|f''\|_2^2 = \sigma^{-5} \frac{3}{8\sqrt{\pi}} \approx 0.212\sigma^{-5}.$$

Reemplazamos a σ por un estimador $\hat{\sigma}$ tal como

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Resta elegir una función de núcleo K . Si elegimos el núcleo normal dado por (2.8) obtenemos la siguiente “regla básica” para la ventana \hat{h}_{rb}

$$\hat{h}_{rb} = \left(\frac{4\hat{\sigma}^5}{3n} \right)^{1/5} \approx 1.06\hat{\sigma}n^{-1/5}. \quad (2.13)$$

Si X está distribuida normalmente, la ventana \hat{h}_{rb} dada en (2.13) es la ventana óptima. Si no, dará una ventana que no está muy lejos de la óptima si la distribución de X no es muy diferente de la normal.

Un problema de esta regla básica para la ventana es su sensibilidad a la presencia de datos atípicos, debido al estimador utilizado para el σ . Podemos obtener un estimador más robusto usando el rango intercuartil

$$R = X_{[0.75n]} - X_{[0.25n]}.$$

Suponiendo que $X \sim N(\mu, \sigma^2)$ y $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$,

$$R = \sigma(Z_{[0.75n]} - Z_{[0.25n]}) \approx 1.34\sigma$$

y así

$$\hat{\sigma} = \frac{R}{1.34}.$$

Reemplazando en (2.13) resulta

$$\hat{h}_{rb} = 1.06 \frac{R}{1.34} n^{-1/5} \approx 0.79 R n^{-1/5}. \quad (2.14)$$

Combinando (2.13) y (2.14) obtenemos una “regla básica mejor”

$$\hat{h}_{rb} = 1.06 \min \left\{ \hat{\sigma}, \frac{R}{1.34} \right\} n^{-1/5}.$$

Si la verdadera densidad se aleja sustancialmente de la forma de la distribución normal podría resultar engañoso usar ventanas con esta regla básica.

Convalidación Cruzada

Enfocaremos nuestra atención en el criterio de convalidación cruzada. Para definirlo comencemos por considerar una medida de distancia alternativa entre \hat{f} y f llamada *error cuadrático integrado (ISE)*:

$$ISE(h) = ISE(\hat{f}_h) = \int [\hat{f}_h - f](x)^2 dx = \int [\hat{f}_h(x) - f(x)]^2 dx.$$

Nuestro objetivo es encontrar el valor de h que minimice al *ISE*. Es fácil ver que

$$ISE(h) - \int f^2(x) dx = \int \hat{f}_h^2(x) dx - 2E[\hat{f}_h(X)].$$

Podemos estimar la esperanza por

$$E[\widehat{f}_h(X)] = \frac{1}{n} \sum_{i=1}^n \hat{f}_{h,-i}(X_i),$$

donde

$$\hat{f}_{h,-i}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n K_h(x - X_j)$$

es el estimador de núcleo de $f(x)$ basado en todas las observaciones excepto la i -ésima (*estimador leave-one-out*). Luego, reemplazando la esperanza por su estimador obtenemos el llamado *criterio de convalidación cruzada*

$$CV(h) = \int \hat{f}_h^2(x) dx - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_h(X_i - X_j).$$

Se puede probar (Härdle, 1991, p. 230) que

$$\int \widehat{f}_h^2(x) dx = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K * K \left(\frac{X_j - X_i}{h} \right),$$

donde $K * K(u)$ es la convolución de K , es decir $K * K(u) = \int K(u-v)K(v) dv$. Así obtenemos el siguiente criterio para minimizar con respecto a h :

$$CV(h) = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K * K \left(\frac{X_j - X_i}{h} \right) - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_h(X_i - X_j).$$

De esta manera encontramos una forma de elegir una ventana basada en un criterio razonable sin tener que hacer suposiciones sobre la familia a la cual pertenece la densidad.

Denotemos por \widehat{h}_{cv} a la ventana que minimiza al criterio y supongamos que la densidad f es una función acotada. Stone (1984) probó que esta ventana es asintóticamente óptima en el siguiente sentido

$$\frac{ISE(\widehat{h}_{cv})}{\min_h ISE(h)} \xrightarrow{c.s.} 1,$$

donde c.s. indica convergencia con probabilidad 1 o casi segura.

Capítulo 3

Regresión No Paramétrica

3.1. Regresión Univariada basada en Núcleos

Un importante estudio es la relación que puede haber entre dos variables, X e Y . En el análisis de regresión el objetivo es ajustar una curva que exprese la relación entre una variable de respuesta Y y una o varias covariables o variables independientes X , tratando de explicar la variación de Y en términos de X . En este sentido, buscamos una relación de la forma

$$Y = m(X) + \varepsilon,$$

donde $m(\cdot)$ es una función en el sentido matemático y ε es el error aleatorio que en general representa al error de medición y suele asumirse que tiene media nula. En el enfoque tradicional paramétrico se imponen restricciones a m asumiendo una forma específica de la misma (lineal, cuadrática o logarítmica, por ejemplo) que está determinada por parámetros a estimar. En el enfoque no paramétrico la función de regresión m será estimada de acuerdo a la evidencia empírica que sugieran los datos, sin hacer suposiciones de forma, excepto las que involucren la suavidad de la curva.

3.1.1. Introducción

Sean X e Y dos variables aleatorias con función de densidad conjunta $f(x, y)$. La esperanza condicional de Y dado $X = x$ está definida como

$$E(Y|X = x) = \int y f(y|x) dy = \int y \frac{f(x, y)}{f_X(x)} dy = \frac{\int y f(x, y) dy}{f_X(x)} = \frac{r(x)}{f_X(x)} = m(x), \quad (3.1)$$

donde $f(y|x)$ es la función de densidad condicional de Y dado $X = x$, y $f_X(x)$ es la densidad marginal de X .

Notemos que la $E(Y|X = x)$ es, solamente, función de x . Por eso, podemos abreviar este término con $m(x)$. Si variamos x , tenemos un conjunto de esperanzas condicionales. A esta relación de x , $m(x)$, la llamamos *función de esperanza condicional* y frecuentemente se denota como $E(Y|X)$. Esta nos dice cómo están relacionadas “en promedio” Y y X . Por lo tanto, será de interés estimar a m .

De aquí en adelante supondremos que X e Y son variables aleatorias con función de densidad conjunta $f(x, y)$. Por lo que tendremos observaciones de la forma (X_i, Y_i) , $i = 1, \dots, n$ y podemos ver a los n pares (X_i, Y_i) como vectores aleatorios independientes e idénticamente distribuidos. Este esquema muestral corresponde al *diseño aleatorio*.

3.1.2. Regresión basada en núcleos

En (3.1) los valores desconocidos son $f(x, y)$ y $f_X(x)$. Podemos usar la estimación basada en núcleos para estimar estas funciones de densidad. La estimación de $f_X(x)$ es directa. Para estimar $f(x, y)$, podemos usar un estimador basado en núcleos multiplicativos

$$\hat{f}_{h,g}(x, y) = \frac{1}{n} \sum_{i=1}^n K_h \left(\frac{x - X_i}{h} \right) K_g \left(\frac{y - Y_i}{g} \right).$$

Por lo tanto, realizando un cambio de variables y usando la paridad de K , podemos estimar el numerador de (3.1) con

$$\begin{aligned} \hat{r}_h(x) &= \int y \hat{f}_{h,g}(x, y) dy = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K \left(\frac{x - X_i}{h} \right) \int \frac{y}{g} K \left(\frac{y - Y_i}{g} \right) dy \\ &= \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \int (sg + Y_i) K(s) ds \\ &= \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i, \end{aligned}$$

Esto nos conduce al estimador de Nadaraya-Watson (1964)

$$\hat{m}_h(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{n^{-1} \sum_{j=1}^n K_h(x - X_j)}. \quad (3.2)$$

Es interesante notar que si $K(u) = (1/2)I_{[-1,1]}(u)$, el estimador resultante de la función de regresión es el promedio de las respuestas y_i correspondientes a valores x_i de la variable independiente próximos al punto x , es decir, tales que $|x_i - x| \leq h$.

Tengamos en cuenta que:

- Reescribiendo (3.2) como

$$\hat{m}_h(x) = \frac{1}{n} \sum_{i=1}^n \left(\frac{K_h(x - X_i)}{n^{-1} \sum_{j=1}^n K_h(x - X_j)} \right) Y_i = \frac{1}{n} \sum_{i=1}^n W_{hi}(x) Y_i,$$

podemos ver al estimador de Nadaraya-Watson como un promedio pesado de las variables de respuesta Y_i (notemos que $\sum_{i=1}^n W_{hi}(x) = 1$).

- La ventana h determina el grado de suavidad de \hat{m}_h . Si $h \rightarrow 0$, luego $W_{hi}(x) \rightarrow 1$ si $x = X_i$ y no está definida en otro caso. Por lo tanto, en una observación X_i , $\hat{m}_h(X_i)$ converge a Y_i , es decir que tenemos una interpolación de los datos. Por otro lado, si $h \rightarrow \infty$ luego $W_{hi}(x) \rightarrow 1$ para todo x , y $\hat{m}_h(X_i) \rightarrow \bar{Y}$, es decir que el estimador es una función constante que le asigna a cada x la media muestral de Y . Este razonamiento muestra la importancia de la elección de la ventana en el proceso de estimación.
- Si el denominador de $W_{hi}(x)$ es igual a cero, el numerador también es igual a cero, y entonces la estimación no está definida. Esto puede suceder en regiones con escasa cantidad de datos.

3.1.3. Propiedades Estadísticas

En esta sección enunciaremos los resultados de consistencia de los estimadores de núcleo.

La consistencia débil puntual del estimador de Nadaraya-Watson está dada por el siguiente teorema, cuya demostración puede verse en Härdle (1990).

Teorema 3.1.1. *Supongamos que estamos en el modelo univariado con diseño aleatorio y que se cumplen las siguientes condiciones de regularidad*

A1. $\int |K(u)|du < \infty$

A2. $uK(u) \rightarrow 0$ cuando $|u| \rightarrow \infty$

A3. $E(Y^2) < \infty$.

A4. $h \rightarrow 0$ y que $nh \rightarrow \infty$

Si x es tal que $f_X(x) > 0$ y, además, es un punto de continuidad de $m(x)$, de $f_X(x)$ y de $\sigma^2(x) = \text{Var}(Y|X = x)$, entonces

$$\frac{1}{n} \sum_{i=1}^n W_{hi}(x) Y_i = \hat{m}_h(x) \xrightarrow{P} m(x).$$

Nos gustaría conocer la tasa de convergencia del estimador, para eso deberíamos estudiar el *MSE*.

Podemos escribir al estimador de Nadaraya-Watson de la siguiente manera

$$\hat{m}_h(x) = \frac{\hat{r}_h(x)}{\hat{f}_h(x)},$$

y así podemos obtener

$$\hat{m}_h(x) - m(x) = \frac{\hat{r}_h(x) - m(x)\hat{f}_h(x)}{f_X(x)} + [\hat{m}_h(x) - m(x)] \frac{f_X(x) - \hat{f}_h(x)}{f_X(x)}.$$

Se puede mostrar que el primer término es el principal en la distribución de $\hat{m}_h(x) - m(x)$, mientras que el segundo puede ser despreciado. Por lo tanto, podemos aproximar al *MSE* calculando

$$E \left[\frac{\hat{r}_h(x) - m(x)\hat{f}_h(x)}{f_X(x)} \right]^2.$$

El siguiente teorema nos da una aproximación para el MSE .

Teorema 3.1.2. *Supongamos que estamos en el modelo univariado con diseño aleatorio y que se cumplen las condiciones **A1** a **A4** del Teorema 3.1.1. Si, además, x es tal que $f_X(x) > 0$ y x es un punto de continuidad de $m'(x)$, $m''(x)$, $f_X(x)$, $f'_X(x)$, y de $\sigma^2(x) = \text{Var}(Y|X = x)$, entonces*

$$MSE[\hat{m}_h(x)] \approx \underbrace{\frac{1}{nh} \frac{\sigma^2(x)}{f_X(x)} \|K\|_2^2}_{\text{varianza}} + \underbrace{\frac{h^4}{4} \left[m''(x) + 2 \frac{m'(x)f'_X(x)}{f_X(x)} \right]^2}_{\text{sesgo}} \mu_2^2(K), \quad (3.3)$$

donde $\mu_2^2(K) = \int u^2 K(u) du$.

En la ecuación (3.3), la mayor parte de los componentes son constantes con respecto a n y h , por lo tanto podemos reescribirla como

$$AMSE(n, h) = \frac{1}{nh} C_1 + h^4 C_2$$

donde C_1 y C_2 denotan a valores constantes.

Si minimizamos esta expresión con respecto a h obtenemos que la ventana óptima es $h_{opt} \sim n^{-1/5}$. Si introducimos la ventana $h \sim n^{-1/5}$ en (3.3), veremos que el $AMSE$ es del orden $O(n^{-4/5})$, una tasa de convergencia menor que la del estimador de mínimos cuadrados en regresión lineal, pero es la misma que para la estimación de la función de densidad.

3.2. Selección del Parámetro de Suavizado

En esta sección nos enfocaremos en encontrar una forma de elegir el parámetro de suavizado para los estimadores de regresión basados en núcleos, que llamaremos la ventana h . Como ya fue mencionado el parámetro de suavizado h cumple un rol fundamental en la determinación de la forma de la curva estimada. Debemos encontrar un valor de h que nos brinde un buen compromiso entre sobresuavizar y subsuavizar los datos.

Dado que queremos que el estimador esté cerca de la verdadera curva, como antes, utilizaremos medidas de esta distancia

- El *error cuadrático medio*

$$MSE(x, h) = MSE[\hat{m}_h(x)] = E\{[\hat{m}_h(x) - m(x)]^2\}$$

mide el desvío del estimador \hat{m}_h de m en un sólo punto x .

- El *error cuadrático integrado*

$$ISE(h) = ISE(\hat{m}_h) = \int_{-\infty}^{+\infty} [\hat{m}_h(x) - m(x)]^2 w(x) f(x) dx$$

es una medida global de la discrepancia, por lo tanto es la indicada si estamos interesados en saber si estimamos bien a m en todo x . La función de peso $w(x)$ se introduce a fin de dar distinto peso a diferentes regiones. Por ejemplo, a una región rala (o sea con pocas observaciones) podríamos desear darle una menor ponderación como para reducir la varianza en dicha región. $ISE(h)$ todavía sigue siendo una variable aleatoria.

- El *error cuadrático integrado medio*

$$MISE(h) = MISE[\widehat{m}_h] = E[ISE(h)]$$

no es una variable aleatoria, pues la esperanza está tomada con respecto a X e Y .

- El *error cuadrático promediado*

$$ASE(h) = ASE[\widehat{m}_h] = \frac{1}{n} \sum_{j=1}^n [\widehat{m}_h(X_j) - m(X_j)]^2 w(X_j)$$

es una aproximación discreta al ISE y, como ella, es una variable aleatoria y una medida global de la discrepancia.

- El *error cuadrático promediado medio*

$$MASE(h) = MASE[\widehat{m}_h] = E[ASE(h) | X_1 = x_1, \dots, X_n = x_n]$$

es la esperanza condicional del ASE , donde la esperanza está tomada con respecto a la distribución conjunta de Y_1, \dots, Y_n . Si vemos a X_1, \dots, X_n como variables aleatorias, el $MASE$ es una variable aleatoria.

Por simplicidad, nos restringiremos a la selección de la ventana para el estimador de Nadaraya-Watson. Para este estimador ha sido probado (Marron y Härdle, 1986) que el ASE , el ISE y el $MISE$ llevan, asintóticamente, al mismo nivel de suavidad. Por lo tanto, usaremos el criterio que es más fácil de calcular y manipular: el $ASE(h)$ discreto.

3.2.1. Una mirada más de cerca al Error Cuadrático Promediado

Queremos encontrar la ventana h que minimice al $ASE(h)$. El $ASE(h)$ cumple

$$ASE(h) = \frac{1}{n} \sum_{i=1}^n m^2(X_i)w(X_i) + \frac{1}{n} \sum_{i=1}^n \widehat{m}_h^2(X_i)w(X_i) - 2\frac{1}{n} \sum_{i=1}^n m(X_i)\widehat{m}_h(X_i)w(X_i).$$

Como el ASE es una variable aleatoria, podemos calcular su esperanza condicional:

$$\begin{aligned} MASE(h) &= E[ASE(h) | X_1 = x_1, \dots, X_n = x_n] \\ &= \frac{1}{n} E\{[\widehat{m}_h(X_i) - m(X_i)]^2 | X_1 = x_1, \dots, X_n = x_n\} w(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \underbrace{\text{Var}[\widehat{m}_h(X_i) | X_1 = x_1, \dots, X_n = x_n]}_{v(h)} \right. \\ &\quad \left. + \underbrace{\text{Sesgo}^2[\widehat{m}_h(X_i) | X_1 = x_1, \dots, X_n = x_n]}_{b^2(h)} \right\} w(X_i), \end{aligned}$$

donde el cuadrado del sesgo es

$$b^2(h) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{n} \sum_{j=1}^n \frac{K_h(X_i - X_j)}{\hat{f}_h(X_i)} m(X_j) - m(X_i) \right]^2 w(X_i)$$

y la varianza

$$v(h) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{n^2} \sum_{j=1}^n \left[\frac{K_h(X_i - X_j)}{\hat{f}_h(X_i)} \right]^2 \sigma(X_j)^2 \right\} w(X_i).$$

Como el $ASE(h)$ involucra a $m(\cdot)$, que es la función que queremos estimar, tenemos que reemplazar al ASE con una aproximación que se pueda calcular con los datos. Una manera simple de reemplazar a $m(\cdot)$ podría ser usar las observaciones de Y , es decir

$$p(h) = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_h(X_i)]^2 w(X_i),$$

llamado el *estimador resustituído*, y es, esencialmente, una suma pesada de residuos cuadrados (RSS). Sin embargo, existe un problema con este método ya que Y_i es usado en $\hat{m}_h(X_i)$ para predecirse a sí mismo. Además se puede ver que $p(h)$ no es un estimador insesgado del $ASE(h)$, pero será usado de diferentes maneras en las siguientes dos secciones.

3.2.2. Convalidación Cruzada

El método de convalidación cruzada resuelve el problema de que la Y_i es usada en $\hat{m}_h(X_i)$ para predecirse a sí misma, empleando el estimador *leave-one-out*, como se hizo previamente, en el que para la estimación $\hat{m}_h(\cdot)$ en X_i la i -ésima observación no es considerada

$$\hat{m}_{h,-i}(X_i) = \frac{\sum_{j \neq i} K_h(X_i - X_j) Y_j}{\sum_{j \neq i} K_h(X_i - X_j)}.$$

Esto lleva a la *función de convalidación cruzada*

$$CV(h) = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_{h,-i}(X_i)]^2 w(X_i).$$

Se puede ver que, en promedio, minimizar a la función $CV(h)$ es equivalente a minimizar al $ASE(h)$. Luego, podemos concluir que elegiremos a aquella ventana que minimice a $CV(h)$.

3.2.3. Funciones Penalizadoras

El método de la *función penalizadora* corrige el sesgo presente en $p(h)$ multiplicándolo por un factor de corrección que penaliza a aquellos h muy pequeños. La “versión corregida” de $p(h)$ se puede escribir como

$$G(h) = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{m}_h(X_i)]^2 \Psi \left(\frac{1}{n} W_{hi}(X_i) \right) w(X_i),$$

con función penalizadora Ψ . Si elegimos una función penalizadora Ψ con desarrollo de Taylor de primer orden $\Psi(u) = 1 + 2u + O(u^2)$ para $u \rightarrow 0$, se puede probar que $G(h)$ es aproximadamente igual al $ASE(h)$. Además, si denotamos por \hat{h} a la ventana que minimiza a $G(h)$ y \hat{h}_0 a aquella que minimiza al $ASE(h)$, entonces para $n \rightarrow \infty$

$$\frac{ASE(\hat{h})}{ASE(\hat{h}_0)} \xrightarrow{P} 1 \quad \text{y} \quad \frac{\hat{h}}{\hat{h}_0} \xrightarrow{P} 1.$$

Por lo tanto, elegir la ventana que minimiza a $G(h)$ es otra regla “buena” para la elección de la misma en la estimación por regresión basada en núcleos.

Observemos que si consideramos $\Psi(u) = (1 - u)^{-2}$, $CV(h) = G(h)$. En consecuencia, el método de convalidación cruzada es equivalente al concepto de funciones penalizadoras y tiene las mismas propiedades asintóticas.

Capítulo 4

Robustez

La mayoría de los procedimientos estadísticos clásicos de estimación e inferencia están basados en supuestos fuertes, tales como errores normales u observaciones equidistribuidas, etc. Bajo estos supuestos se pueden deducir procedimientos óptimos. Tal es el caso del modelo de regresión en el que el procedimiento óptimo es el de mínimos cuadrados. En general, para modelos paramétricos, los procedimientos óptimos clásicos son los basados en máxima verosimilitud. Una desventaja de estos métodos es que son sensibles al apartamiento a alguna de estas hipótesis en las que están basados. Por ejemplo, pequeñas desviaciones a la normalidad manifestadas por algunas pocas desviaciones atípicas podrían invalidar las conclusiones basadas en estos métodos.

El objetivo de los métodos robustos es desarrollar estimadores que tengan un “buen” comportamiento en un “entorno” del modelo. Los métodos robustos son resistentes a estos apartamientos y, a la vez, son eficientes cuando los modelos son válidos.

Nos enfocaremos primero en el estudio de los estimadores de posición y dispersión y luego los generalizaremos al caso multivariado. Maronna, Martin y Yohai (2006) dan una revisión muy completa sobre los tópicos que desarrollamos en este capítulo.

4.1. El Modelo de Posición

Supongamos que el resultado de cada observación X_i depende de un valor verdadero μ de un parámetro desconocido y también de algún proceso de error aleatorio.

La simple elección de errores aditivos corresponde al *modelo de posición*:

$$X_i = \mu + U_i, \quad i = 1, \dots, n, \quad (4.1)$$

donde los errores U_1, \dots, U_n son variables aleatorias.

Si las observaciones son replicaciones independientes del mismo experimento bajo las mismas condiciones, se puede suponer que X_1, \dots, X_n son independientes con función de distribución común

$$F(x) = F_0(x - \mu), \quad (4.2)$$

donde F_0 es la distribución común de los errores.

Buscaremos estimadores $\hat{\mu}$ de μ tales que $\hat{\mu}$ se aproxime al verdadero valor μ en algún sentido. Una manera de medir esta aproximación es con el error cuadrático medio. Si los datos fueran exactamente normales, la media muestral sería un estimador óptimo: es el estimador de máxima verosimilitud y minimiza el error cuadrático medio entre todos los estimadores insesgados y entre los equivariantes. El objetivo es buscar estimadores que sean casi tan buenos como la media cuando F es exactamente normal, pero que también sean “buenos” en algún sentido cuando F es aproximadamente normal. Diremos que F es una *distribución normal contaminada* si

$$F = (1 - \varepsilon)G + \varepsilon H \quad (4.3)$$

donde $G = N(\mu, \sigma^2)$ y H puede ser cualquier distribución.

Si usamos la mediana como estimador, se puede ver que el beneficio de la robustez obtenido se paga con un aumento en la varianza (“una pérdida en eficiencia”) bajo la distribución normal. Por esta razón desarrollaremos estimadores que combinen la baja varianza de la media bajo normalidad con la robustez de la mediana bajo contaminación. Sólo trataremos con distribuciones simétricas.

4.1.1. M-estimadores de Posición

Presentaremos una familia general de estimadores que contienen a la media y a la mediana como casos especiales.

Consideremos el modelo de posición dado en (4.1). Supongamos que F_0 tiene densidad $f_0 = F_0'$. Luego la densidad conjunta de las observaciones (la *función de verosimilitud*) es

$$L(x_1, \dots, x_n; \mu) = \prod_{i=1}^n f_0(x_i - \mu).$$

El *estimador de máxima verosimilitud* (MLE) de μ es el valor

$$\hat{\mu} = \hat{\mu}(X_1, \dots, X_n) = \arg \max_{\mu} L(X_1, \dots, X_n; \mu).$$

Si f_0 es positiva en todos lados

$$\hat{\mu} = \arg \min_{\mu} \sum_{i=1}^n \rho(X_i - \mu), \quad (4.4)$$

donde

$$\rho(t) = -\log(f_0(t)).$$

Si ρ , generalmente llamada función de pérdida, es derivable, derivando (4.4) obtenemos

$$\sum_{i=1}^n (X_i - \hat{\mu}) = 0 \quad (4.5)$$

con $\hat{\mu} = \rho'$. Notemos que si f_0 es simétrica, ρ es par y por lo tanto ρ' es impar.

Si $\rho(x) = x^2/2$, tenemos que la solución es $\hat{\mu} = \bar{X}$. Para $\rho(x) = |x|$, se puede probar que cualquier mediana de X es una solución de (4.4).

Luego, dada una función ρ , podemos definir a un *M-estimador de posición* como una solución de (4.4).

La función ρ será elegida para asegurar que el estimador sea:

- “cercanamente óptimo” cuando F_0 es exactamente normal,
- “cercanamente óptimo” cuando F_0 es aproximadamente normal (como por ejemplo una normal contaminada).

Si suponemos que ψ , habitualmente llamada función de score, es monótona creciente, con $(-\infty) < 0 < (\infty)$, se puede probar que (4.5) y por lo tanto (4.4) siempre tiene solución. Si ψ es continua y estrictamente creciente, la solución es única.

Es fácil ver que los M-estimadores son equivariantes por traslaciones.

Un tipo de funciones ρ y ψ con propiedades importantes es la familia de *funciones de Huber*:

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

con derivada $\psi_k(x)$ donde

$$\psi_k(x) = \begin{cases} x & \text{si } |x| \leq k \\ \text{sgn}(x)k & \text{si } |x| > k. \end{cases} \tag{4.6}$$

Para evaluar el rendimiento de los M-estimadores, es necesario calcular su distribución.

Supongamos que ψ es estrictamente creciente. Dada F una distribución, definimos $\mu_0 = \mu_0(F)$ como la solución de

$$E_F[\psi(X - \mu_0)] = 0.$$

Bajo condiciones de regularidad, se puede probar que $\hat{\mu}$ es *consistente* para μ_0 , es decir que si $n \rightarrow \infty$, entonces $\hat{\mu} \xrightarrow{P} \mu_0$ y también que la distribución asintótica de $\hat{\mu}$ es aproximadamente $N(\mu_0, \frac{\nu}{n})$ donde la *varianza asintótica* es

$$\nu = \frac{E_F[(X - \mu_0)^2]}{E_F[\psi'(X - \mu_0)]^2}. \tag{4.7}$$

Notemos que bajo el modelo (4.2) ν no depende de μ_0 , es decir,

$$\nu = \frac{E_{F_0}[(X)^2]}{E_{F_0}[\psi'(X)]^2}.$$

Luego, podemos definir la *eficiencia asintótica* de $\hat{\mu}$ como el cociente

$$Eff(\hat{\mu}) = \frac{\nu_0}{\nu},$$

donde ν_0 es la varianza asintótica del estimador de máxima verosimilitud, y mide cuán cerca está $\hat{\mu}$ del óptimo. La función de score ψ_k de Huber definida en (4.6) es uno de los pocos casos en los que se puede calcular analíticamente la varianza asintótica bajo normalidad. Así el valor de k se elige para asegurar una eficiencia asintótica dada, bajo la distribución normal.

4.1.2. M-estimadores Redescendientes

Una elección popular para las funciones ρ y ψ es la familia de funciones *bicuada*, dada por:

$$\rho_k(x) = \begin{cases} 1 - [1 - (x/k)^2]^3 & \text{si } |x| \leq k \\ 1 & \text{si } |x| < k \end{cases}$$

con derivada $\rho'_k(x) = 6(x/k)^2$ donde

$$\psi_k(x) = x \left[1 - \left(\frac{x}{k} \right)^2 \right]^2 I(|x| \leq k).$$

Los estimadores definidos como solución de (4.5) con ρ monótona son llamados “M-estimadores monótonos”, mientras que aquellos definidos por (4.4) cuando ρ no es monótona, como en el caso de la bicuada, serán llamados “M-estimadores redescendientes”. Estos últimos resultan ser más robustos ante la presencia de grandes datos atípicos.

Llameremos ρ -función a una función ρ tal que:

- R1.** $\rho(x)$ es una función creciente del $|x|$
- R2.** $\rho(0) = 0$
- R3.** $\rho(x)$ es estrictamente creciente para $x > 0$ y tal que $\rho(x) < \rho(\infty)$
- R4.** Si ρ es acotada, supondremos también que $\rho(\infty) = 1$.

Luego, una ψ -función denotará a una función ψ que será la derivada de una ρ -función, lo cual implica en particular que:

- Ψ1.** ψ es impar y $\psi(x) \geq 0$ para $x \geq 0$.

4.2. Estimadores de Dispersión

La manera tradicional de medir la variabilidad de un conjunto de datos $\mathbf{x} = \{X_1, \dots, X_n\}$ es con el desvío estándar (*SD*)

$$SD(\mathbf{x}) = \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{1/2}.$$

Para toda constante c el SD es invariante por traslaciones y por cambios de escala, es decir

$$SD(\mathbf{x} + c) = SD(\mathbf{x}), \quad SD(c\mathbf{x}) = |c|SD(\mathbf{x}). \quad (4.8)$$

Luego, si un estimador satisface (4.8), diremos que es un *estimador de dispersión*. Una opción menos sensible a la presencia de datos atípicos es el *desvío medio absoluto* (*MD*):

$$MD(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|, \quad (4.9)$$

Una alternativa robusta es reemplazar la media por la mediana, obteniendo el *desvío mediano absoluto sobre la mediana* (*MAD*):

$$MAD(\mathbf{x}) = Med(|\mathbf{x} - Med(\mathbf{x})|), \quad (4.10)$$

que claramente satisface (4.8).

De la misma manera que en (4.9) y (4.10), definimos los desvíos medio y mediano absolutos de una variable aleatoria X como

$$MD(X) = E(|X - E(X)|)$$

y

$$MAD(X) = Med(|X - Med(X)|).$$

Notemos que si $X \sim N(\mu, \sigma^2)$, $SD(X) = \sigma$ por definición, mientras que:

$$MD(X) = c_1\sigma, \quad MAD(X) = c_2\sigma,$$

donde

$$c_1 = 2\varphi(0) \text{ y } c_2 = \Phi^{-1}(0.75),$$

con φ y Φ las funciones de densidad y de distribución de una variable aleatoria $N(0, 1)$, respectivamente.

Si queremos que un estimador de dispersión “mida lo mismo” que el SD bajo normalidad, deberíamos normalizar la MAD dividiéndola por $c_2 \approx 0.675$, obteniendo la “ MAD normalizada” ($MADN$)

$$MADN(X) = \frac{MAD(X)}{0.675}. \quad (4.11)$$

4.3. M-estimadores de Escala

Consideremos observaciones X_i que satisfacen el modelo *multiplicativo*

$$X_i = \sigma U_i, \quad (4.12)$$

donde los U_i 's son i.i.d. con densidad f_0 y $\sigma > 0$ es el parámetro desconocido. La distribución de las X_i 's constituye una *familia de escala*, con densidad

$$\frac{1}{\sigma} f_0\left(\frac{x}{\sigma}\right).$$

El estimador de máxima verosimilitud de σ en (4.12) es

$$\hat{\sigma} = \arg \max_{\sigma} \frac{1}{\sigma^n} \prod_{i=1}^n f_0\left(\frac{X_i}{\sigma}\right).$$

Tomando logaritmo y derivando respecto a σ obtenemos que

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{X_i}{\hat{\sigma}}\right) = 1,$$

donde $\rho(t) = t - (t)$, con $(t) = -f'_0(t)/f_0(t)$. Luego, un estimador que satisface una ecuación de la forma

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{X_i}{\hat{\sigma}} \right) = \delta, \quad (4.13)$$

donde ρ es una ρ -función y δ es una constante positiva, será llamado un *M-estimador de escala*. Notemos que para que (4.13) tenga solución debemos pedir que $0 < \delta < \rho(\infty)$. Por lo tanto, si ρ es acotada supondremos sin pérdida de generalidad que

$$\rho(\infty) = 1, \quad \delta \in (0, 1).$$

Se puede probar que los M-estimadores de escala son equivariantes en el sentido que $\hat{\sigma}(c\mathbf{x}) = c\hat{\sigma}(\mathbf{x})$ para todo $c > 0$, y si ρ es par

$$\hat{\sigma}(c\mathbf{x}) = |c|\hat{\sigma}(\mathbf{x})$$

para todo c . Para n grande, la sucesión de estimadores (4.13) converge a la solución de

$$E \left[\rho \left(\frac{X}{\sigma} \right) \right] = \delta,$$

si es única.

4.4. La Función de Influencia

Estudiaremos el comportamiento asintótico de los estimadores cuando el tamaño de la muestra tiende a infinito. Consideremos un estimador $\hat{\theta}_n = \hat{\theta}_n(\mathbf{x})$ dependiendo de una muestra $\mathbf{x} = \{X_1, \dots, X_n\}$ de tamaño n de variables aleatorias i.i.d. con distribución F . Luego definimos $\hat{\theta}_\infty = \hat{\theta}_\infty(F)$ al “valor asintótico” del estimador en F , es decir

$$\hat{\theta}_n \xrightarrow{P} \hat{\theta}_\infty(F).$$

Si $\hat{\theta}_n = \bar{X}$, entonces $\hat{\theta}_\infty(F) = E_F(X_1)$; si $\hat{\theta}_n$ es un M-estimador que es solución de (4.5) con monótona, entonces $\hat{\theta}_\infty(F)$ es solución de $E_F[(X_1 - \theta)] = 0$.

Un estimador $\hat{\theta}$ de los parámetros de una familia paramétrica F_θ será consistente si

$$\hat{\theta}_\infty(F_\theta) = \theta.$$

Estamos interesados en el comportamiento de $\hat{\theta}_\infty(F)$ cuando F pertenece a un “entorno” de una distribución F_0 . Para ello se define un *entorno de contaminación* a

$$\mathcal{F}(F, \varepsilon) = \{(1 - \varepsilon)F + \varepsilon G : G \in \mathcal{G}\},$$

donde \mathcal{G} es un conjunto adecuado de distribuciones.

La función de influencia (*IF*), también llamada curva de influencia, de un estimador, fue introducida por Hampel (1974). Es una aproximación del comportamiento de $\hat{\theta}_\infty$ cuando la muestra

contiene una pequeña fracción ε de datos atípicos idénticos, y se define como

$$\begin{aligned} IF_{\hat{\theta}}(x_0, F) &= \lim_{\varepsilon \rightarrow 0^+} \frac{\hat{\theta}_{\infty}((1-\varepsilon)F + \varepsilon\delta_{x_0}) - \hat{\theta}_{\infty}(F)}{\varepsilon} \\ &= \left. \frac{\partial}{\partial \varepsilon} \hat{\theta}_{\infty}((1-\varepsilon)F + \varepsilon\delta_{x_0}) \right|_{\varepsilon \rightarrow 0^+}, \end{aligned}$$

donde δ_{x_0} es la distribución de masa puntual en x_0 .

Así, si ε es pequeño el *sesgo* definido por $\hat{\theta}_{\infty}((1-\varepsilon)F + \varepsilon\delta_{x_0}) - \hat{\theta}_{\infty}(F)$ puede ser aproximado por $\varepsilon IF_{\hat{\theta}}(x_0, F)$.

Para el caso de un M-estimador de posición $\hat{\mu}$ con dispersión conocida σ , donde el valor asintótico cumple que

$$E_F \left[\left(\frac{X - \hat{\mu}_{\infty}}{\sigma} \right) \right] = 0,$$

la IF es

$$IF_{\hat{\mu}}(x_0, F) = \sigma \frac{((x_0 - \hat{\mu}_{\infty})/\sigma)}{E_F['((X - \hat{\mu}_{\infty})/\sigma)]}.$$

Sea ahora $\hat{\mu}$ un M-estimador de posición con dispersión estimada previamente por $\hat{\sigma}$. Si F es simétrica, la IF es

$$IF_{\hat{\mu}}(x_0, F) = \hat{\sigma}_{\infty} \frac{((x_0 - \hat{\mu}_{\infty})/\hat{\sigma}_{\infty})}{E_F['((X - \hat{\mu}_{\infty})/\hat{\sigma}_{\infty})]}.$$

Bajo ciertas condiciones de regularidad, se puede ver que para un M-estimador de posición la varianza asintótica ν satisface

$$\nu = E_F[IF(X, F)^2].$$

4.5. El Punto de Ruptura

En línea general, el *punto de ruptura* (BP) de un estimador $\hat{\theta}$ del parámetro θ es la mayor cantidad de contaminación (proporción de observaciones atípicas) que pueden contener los datos, de manera tal que $\hat{\theta}$ siga dando alguna información sobre θ .

Sea $\theta \in \Theta$, Θ espacio paramétrico. Para que el estimador $\hat{\theta}$ dé alguna información sobre θ , la contaminación no debería llevar a $\hat{\theta}$ al infinito o a la frontera de Θ .

Luego, definimos el BP *de contaminación asintótico* del estimador $\hat{\theta}$ en F , denotado por $\varepsilon^*(\hat{\theta}, F)$, al valor más grande $\varepsilon^* \in (0, 1)$ tal que para todo $\varepsilon < \varepsilon^*$, $\hat{\theta}_{\infty}((1-\varepsilon)F + \varepsilon G)$ se mantiene acotado en función de G , y también lejos de la frontera de Θ .

Veamos por ejemplo, qué ocurre en el caso de M-estimadores en el modelo de posición con escala conocida. Supongamos que es monótona, pero no necesariamente impar y que

$$k_1 = -(-\infty), \quad k_2 = (\infty)$$

son finitos. Luego, se puede mostrar que

$$\varepsilon^* = \frac{\min(k_1, k_2)}{k_1 + k_2}.$$

Por lo tanto, si k es impar, $k_1 = k_2$ y se alcanza la cota $\varepsilon^* = 1/2$. Los estimadores redescendentes también alcanzan la cota $\varepsilon^* = 1/2$.

Se puede mostrar que los estimadores de dispersión SD y MAD tienen como punto de ruptura a 0 y $1/2$, respectivamente.

4.6. Robustez en el Análisis Multivariado

En el caso del análisis multivariado observamos vectores aleatorios $\mathbf{X} = (X_1, \dots, X_p)' \in \mathbb{R}^p$ con una distribución F en \mathbb{R}^p . En el enfoque clásico describimos la posición del vector aleatorio \mathbf{X} por la esperanza $\boldsymbol{\mu} = E(\mathbf{X}) = (E(X_1), \dots, E(X_p))'$ y la dispersión por la matriz de covarianza

$$\mathbf{Var}(\mathbf{X}) = E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})').$$

La matriz de covarianza $\mathbf{Var}(\mathbf{X})$ es simétrica y semidefinida positiva y satisface que para cada vector constante \mathbf{a} y cada matriz constante \mathbf{A}

$$E(\mathbf{A}\mathbf{X} + \mathbf{a}) = \mathbf{A}E(\mathbf{X}) + \mathbf{a}, \quad \mathbf{Var}(\mathbf{A}\mathbf{X} + \mathbf{a}) = \mathbf{A}\mathbf{Var}(\mathbf{X})\mathbf{A}'.$$

Los métodos clásicos de estimación multivariada están basados en el supuesto de que se observa una muestra aleatoria de observaciones independientes e idénticamente distribuidas $\underline{\mathbf{X}} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ donde cada \mathbf{X}_i tiene distribución normal multivariada $N_p(\boldsymbol{\mu}, \Sigma)$ con densidad

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (4.14)$$

donde $\Sigma = \mathbf{Var}(\mathbf{X})$ y $|\Sigma|$ es el determinante de Σ . Bajo la distribución normal (4.14), los estimadores de máxima verosimilitud de $\boldsymbol{\mu}$ y Σ para una muestra $\underline{\mathbf{X}}$ son respectivamente la media y la matriz de covarianza muestral

$$\bar{\underline{\mathbf{X}}} = ave(\underline{\mathbf{X}}) = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \quad \text{y} \quad \mathbf{Var}(\underline{\mathbf{X}}) = ave[(\underline{\mathbf{X}} - \bar{\underline{\mathbf{X}}})(\underline{\mathbf{X}} - \bar{\underline{\mathbf{X}}})'].$$

Sean $(\hat{\boldsymbol{\mu}}(\underline{\mathbf{X}}), \hat{\Sigma}(\underline{\mathbf{X}}))$ estimadores de posición y dispersión correspondientes a la muestra $\underline{\mathbf{X}} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$. Diremos que los estimadores son afín equivariantes si

$$\hat{\boldsymbol{\mu}}(\mathbf{A}\underline{\mathbf{X}} + \mathbf{b}) = \mathbf{A}\hat{\boldsymbol{\mu}}(\underline{\mathbf{X}}) + \mathbf{b}, \quad \hat{\Sigma}(\mathbf{A}\underline{\mathbf{X}} + \mathbf{a}) = \mathbf{A}\hat{\Sigma}\mathbf{A}'.$$

La media y la matriz de covarianza muestrales poseen esta propiedad.

Al igual que en el caso univariado, una pequeña cantidad de datos atípicos puede alterar completamente a la media y a la matriz de covarianza muestrales.

4.6.1. Punto de Ruptura de Estimadores Multivariados

Para el vector de posición, el espacio paramétrico es un espacio finito-dimensional Euclídeo, y por lo tanto pediremos que $\hat{\boldsymbol{\mu}}$ permanezca en un conjunto acotado. Diremos que “ $\hat{\Sigma}$ es acotada y está acotada lejos del borde” si sus autovalores están acotados lejos del cero y del infinito.

Para $0 \leq m \leq n$ llamemos Z_m al conjunto de “muestras” $\underline{\mathbf{Z}} = \{\mathbf{Z}_1, \dots, \mathbf{Z}_n\}$ tales que $\#\{\mathbf{Z}_i = \mathbf{X}_i\} = m$, y sean $\hat{\boldsymbol{\mu}}(\underline{\mathbf{Z}})$ y $\hat{\boldsymbol{\Sigma}}(\underline{\mathbf{Z}})$ los estimadores de posición y dispersión basados en la muestra $\underline{\mathbf{Z}}$.

El punto de ruptura finito (FBP) de $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ está definido como $\varepsilon^* = m^*/n$ donde m^* es el m más grande tal que existen constantes finitas positivas a , b , y c tales que

$$\|\hat{\boldsymbol{\mu}}(\underline{\mathbf{Z}})\| \leq a \text{ y } b \leq \lambda_p(\hat{\boldsymbol{\Sigma}}(\underline{\mathbf{Z}})) \leq \lambda_1(\hat{\boldsymbol{\Sigma}}(\underline{\mathbf{Z}})) \leq c$$

para todo $\underline{\mathbf{Z}} \in Z_m$, donde $\lambda_1(\hat{\boldsymbol{\Sigma}}(\underline{\mathbf{Z}})) \geq \dots \geq \lambda_p(\hat{\boldsymbol{\Sigma}}(\underline{\mathbf{Z}}))$ son los autovalores en orden descendiente de $\hat{\boldsymbol{\Sigma}}(\underline{\mathbf{Z}})$.

El BP asintótico de $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ es el mayor $\varepsilon^* \in (0, 1)$ para el cual existen constantes positivas a , b y c tales que

$$\begin{aligned} \|\hat{\boldsymbol{\mu}}_\infty((1-\varepsilon)F + \varepsilon G)\| &\leq a \\ b \leq \lambda_p(\hat{\boldsymbol{\Sigma}}((1-\varepsilon)F + \varepsilon G)) &\leq \lambda_1(\hat{\boldsymbol{\Sigma}}((1-\varepsilon)F + \varepsilon G)) \leq c \end{aligned}$$

vale para todo G .

4.6.2. M-estimadores

En el caso multivariado definiremos estimadores equivariantes robustos de posición y dispersión *simultáneamente*.

La densidad de la distribución normal multivariada dada en (4.14) puede escribirse como

$$f(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{|\Sigma|}} h(d(\mathbf{x}, \boldsymbol{\mu}, \Sigma)), \quad (4.15)$$

donde $h(s) = c \exp(-s/2)$ con $c = (2\pi)^{-p/2}$ y $d(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$. En general, para cualquier elección de h positiva tal que f satisfaga (4.15) e integre uno, se dice que f es *elípticamente simétrica* (“elíptica”).

Sea $\mathbf{X}_1, \dots, \mathbf{X}_n$ una muestra aleatoria i.i.d. con densidad f con la forma (4.15) en la que h es asumida positiva en todo punto. Notemos que la función de verosimilitud para calcular los estimadores de $\boldsymbol{\mu}$ y Σ es

$$L(\boldsymbol{\mu}, \Sigma) = \frac{1}{|\Sigma|^{n/2}} \prod_{i=1}^n h(d(\mathbf{x}_i, \boldsymbol{\mu}, \Sigma)),$$

y maximizar a $L(\boldsymbol{\mu}, \Sigma)$ es equivalente a minimizar

$$-2 \log(L(\boldsymbol{\mu}, \Sigma)) = n \log(|\hat{\boldsymbol{\Sigma}}|) + \sum_{i=1}^n \rho(d_i),$$

donde

$$\rho(s) = -2 \log(h(s)) \text{ y } d_i = d(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}).$$

Derivando con respecto a $\boldsymbol{\mu}$ y Σ obtenemos el siguiente sistema de ecuaciones

$$\sum_{i=1}^n W(d_i)(\mathbf{X}_i - \hat{\boldsymbol{\mu}}) = \mathbf{0}$$

$$\frac{1}{n} \sum_{i=1}^n W(d_i)(\mathbf{X}_i - \hat{\boldsymbol{\mu}})(\mathbf{X}_i - \hat{\boldsymbol{\mu}})' = \hat{\Sigma},$$

con $W(s) = \rho'(s)$. Bajo la distribución normal, tenemos que $W \equiv 1$.

En general, definimos a los M-estimadores como solución de

$$\sum_{i=1}^n W_1(d_i)(\mathbf{X}_i - \hat{\boldsymbol{\mu}}) = \mathbf{0} \quad (4.16)$$

$$\frac{1}{n} \sum_{i=1}^n W_2(d_i)(\mathbf{X}_i - \hat{\boldsymbol{\mu}})(\mathbf{X}_i - \hat{\boldsymbol{\mu}})' = \hat{\Sigma}, \quad (4.17)$$

donde las funciones W_1 y W_2 no son necesariamente iguales. Notemos que por (4.17) podemos interpretar a $\hat{\Sigma}$ como una matriz de covarianza pesada, y por (4.16) podemos expresar a $\hat{\boldsymbol{\mu}}$ como un promedio pesado

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^n W_1(d_i)\mathbf{X}_i}{\sum_{i=1}^n W_1(d_i)}, \quad (4.18)$$

con los pesos dependiendo de una medida de *atipicidad* d_i .

La existencia y unicidad fueron tratadas por Maronna (1976) y más en general por Tatsuoka y Tyler (2000). Para la unicidad de las soluciones de (4.16) y (4.17) se pide que $dW_2(d)$ sea una función no decreciente de d , en tal caso llamaremos a estos M-estimadores de posición y dispersión *monótonos*, si no los llamaremos *redescendientes*.

Está probado en el Capítulo 8 de Huber (1981) que si las \mathbf{X}_i son i.i.d. con distribución F , bajo supuestos generales cuando $n \rightarrow \infty$, los M-estimadores monótonos definidos como alguna solución $\hat{\boldsymbol{\mu}}$ y $\hat{\Sigma}$ de (4.16) y (4.17) convergen en probabilidad a la solución $(\hat{\boldsymbol{\mu}}_\infty, \hat{\Sigma}_\infty)$ de

$$E[W_1(d)(\mathbf{X} - \hat{\boldsymbol{\mu}}_\infty)] = \mathbf{0}$$

$$E[W_2(d)(\mathbf{X} - \hat{\boldsymbol{\mu}}_\infty)(\mathbf{X} - \hat{\boldsymbol{\mu}}_\infty)'] = \hat{\Sigma}_\infty,$$

donde $d = d(\mathbf{X}, \hat{\boldsymbol{\mu}}_\infty, \hat{\Sigma}_\infty)$. Es fácil probar que los M-estimadores son afín equivariantes.

Punto de Ruptura

Es intuitivamente claro que para la robustez de los estimadores necesitaremos que ningún término domine las sumas (4.16) y (4.17), y para conseguir esto supondremos

$$W_1(d)\sqrt{d} \text{ y } W_2(d)d \text{ acotadas para } d \geq 0. \quad (4.19)$$

Sea $K = \sup_d W_2(d)d$. Si suponemos Σ conocido, se puede probar que el BP asintótico de $\hat{\boldsymbol{\mu}}$ es $1/2$. Por otro lado, si $\boldsymbol{\mu}$ es conocido, Maronna (1976), y la distribución subyacente F_0 le da masa cero a cualquier hiperplano, el BP asintótico de un M-estimador *monótono* de Σ con W_2 satisfaciendo (4.19) es

$$\varepsilon^* = \min\left(\frac{1}{K}, 1 - \frac{p}{K}\right).$$

El FBP es similar pero los detalles están desarrollados en Tyler (1990). Diremos que una muestra está en *posición general* si ningún hiperplano contiene más de p puntos. Davies (1987) mostró que el máximo FBP de cualquier estimador equivariante para una muestra en posición general es $m_{\text{máx}}^*/n$ con

$$m_{\text{máx}}^* = \left\lceil \frac{n-p}{2} \right\rceil. \quad (4.20)$$

4.6.3. Estimadores basados en una escala robusta

Definiremos estimadores de posición y dispersión que hagan que las medidas d_i sean “pequeñas”. Para este fin buscaremos $\hat{\boldsymbol{\mu}}$ y $\hat{\Sigma}$ que minimicen alguna medida de la “magnitud” de $d(\mathbf{X}, \hat{\boldsymbol{\mu}}, \hat{\Sigma})$. Llamemos S_p al conjunto de todas las matrices de $p \times p$ simétricas y definidas positivas. Dado un conjunto de datos $\underline{\mathbf{X}}$ llamemos $\mathbf{d}(\underline{\mathbf{X}}, \hat{\boldsymbol{\mu}}, \hat{\Sigma})$ al vector de elementos $d(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma})$, $i = 1, \dots, n$, y sea $\hat{\sigma}$ un estimador de escala robusto. Luego, definimos los estimadores $\hat{\boldsymbol{\mu}}$ y $\hat{\Sigma}$ por

$$\hat{\sigma} \left(\mathbf{d}(\underline{\mathbf{X}}, \hat{\boldsymbol{\mu}}, \hat{\Sigma}) \right) = \text{mín} \quad \text{con} \quad \hat{\boldsymbol{\mu}} \in \mathbb{R}^p, \quad \hat{\Sigma} \in S_p, \quad |\hat{\Sigma}| = 1. \quad (4.21)$$

Es fácil probar que los estimadores definidos por (4.21) son equivariantes.

El estimador de mínimo volumen de elipsoide

El estimador de mínimo volumen de elipsoide (MVE) se obtiene considerando en (4.21) a $\hat{\sigma}$ como la mediana muestral. El nombre proviene del hecho que entre todos los elipsoides $\{\mathbf{x} : d(\mathbf{x}, \boldsymbol{\mu}, \Sigma) \leq 1\}$ que contienen al menos la mitad de las observaciones, el que corresponde al estimador MVE es aquel que tiene mínimo volumen, es decir mínimo $|\Sigma|$. La velocidad de convergencia del MVE es solamente $n^{-1/3}$, y por lo tanto es muy ineficiente (Davies, 1992).

S-estimadores

Para superar la ineficiencia del estimador MVE consideramos una nueva clase de estimadores llamados S-estimadores (Davies, 1987), definidos por (4.21) tomando para $\hat{\sigma}$ un M-estimador de escala definido por

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{d_i}{\hat{\sigma}} \right) = \delta \quad (4.22)$$

donde ρ es una ρ -función acotada y suave. Se puede probar que un S-estimador $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ es un M-estimador en el sentido que para algún $\tilde{\boldsymbol{\mu}}$ y $\tilde{\Sigma}$ con $|\tilde{\Sigma}| = 1$ y $\hat{\sigma} = \hat{\sigma} \left(\mathbf{d}(\underline{\mathbf{X}}, \tilde{\boldsymbol{\mu}}, \tilde{\Sigma}) \right)$

$$\sum_{i=1}^n \rho \left(\frac{d(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma})}{\hat{\sigma}} \right) \leq \sum_{i=1}^n \rho \left(\frac{d(\mathbf{X}_i, \tilde{\boldsymbol{\mu}}, \tilde{\Sigma})}{\hat{\sigma}} \right).$$

Si ρ es derivable, la solución de (4.21) debe satisfacer el siguiente sistema de ecuaciones de estimación

$$\sum_{i=1}^n W \left(\frac{d_i}{\hat{\sigma}} \right) (\mathbf{X}_i - \hat{\boldsymbol{\mu}}) = \mathbf{0},$$

$$\frac{1}{n} \sum_{i=1}^n W\left(\frac{d_i}{\hat{\sigma}}\right) (\mathbf{X}_i - \hat{\boldsymbol{\mu}})(\mathbf{X}_i - \hat{\boldsymbol{\mu}})' = c\hat{\boldsymbol{\Sigma}},$$

donde

$$W = \rho' \text{ y } \hat{\sigma} = \hat{\sigma}(d_1, \dots, d_n),$$

y c es una constante escalar tal que $|\hat{\boldsymbol{\Sigma}}| = 1$. Si ρ es acotada, como es usual, éste no es un M-estimador monótono, por lo tanto las ecuaciones de estimación sólo llevarán a mínimos *locales* de $\hat{\sigma}$. Se puede demostrar (Davies, 1987) que si ρ es derivable, para los S-estimadores $\sqrt{n}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_\infty, \hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}}_\infty)$ tiende a una normal multivariada. El máximo FBP dado en (4.20) es alcanzado por los S-estimadores si consideramos en (4.22)

$$n\delta = m_{\text{máx}}^* = \left\lfloor \frac{n-p}{2} \right\rfloor.$$

Definimos el S-estimador multivariado *bicadrado* como el que tiene la escala dada por (4.22) con

$$\rho(t) = \min\{1, 1 - (1-t)^3\},$$

y función de peso

$$W(t) = 3(1-t)^2 I(t \leq 1).$$

El estimador de mínimo determinante de covarianza (MCD)

Otra posibilidad es usar una escala podada para $\hat{\sigma}$ en vez de una M-escala. Sean $d_{(1)} \leq \dots \leq d_{(n)}$ los estadísticos de orden de las distancias $d_i = d(\mathbf{X}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, y para $1 \leq h < n$ definimos la escala podada de las distancias como

$$\hat{\sigma} = \sum_{i=1}^h d_{(i)}.$$

Así, un estimador $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ definido por (4.21) con esta escala podada es llamado un estimador de *mínimo determinante de covarianza* (MCD). Su nombre se debe a que para cada elipsoide $\{\mathbf{x} : d(\mathbf{x}, \mathbf{t}, \mathbf{V}) \leq 1\}$ que contenga al menos h observaciones, se calcula la matriz de covarianza \mathbf{C} de las observaciones que están dentro del elipsoide. Si $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ es un estimador MCD, el elipsoide con $\mathbf{t} = \hat{\boldsymbol{\mu}}$ y \mathbf{V} igual a un múltiplo escalar de $\hat{\boldsymbol{\Sigma}}$ minimiza al $|\mathbf{C}|$.

El máximo BP del estimador MCD es alcanzado si tomamos $h = n - m_{\text{máx}}^*$ con $m_{\text{máx}}^*$ definido en (4.20).

Capítulo 5

Regresión Lineal con Errores-en-las-Variables

5.1. Modelo

Todo problema estadístico puede volverse más complejo si las variables en términos de las cuales está definido el modelo no son observables. El término errores-en-las-variables (EV) se refiere al hecho de que una o más variables independientes o covariables han sido observadas con error. Modelos estadísticos que tratan este tipo de problemas datan del siglo XIX y existe una vasta bibliografía sobre el tema. Nosotros enfocaremos nuestra atención en el problema de regresión lineal.

En el modelo de regresión lineal simple con errores-en-las-variables consideramos vectores $\mathbb{Z}_i = (X_i, W_i)'$, $i = 1, \dots, n$ que no son observables y que están relacionadas linealmente mediante la ecuación

$$W_i = \alpha + \beta X_i. \quad (5.1)$$

Estas variables están expuestas a errores de medición y sólo podemos observar

$$Y_i = W_i + \varepsilon_i = \alpha + \beta X_i + \varepsilon_i \quad (5.2)$$

$$V_i = X_i + U_i$$

con $(U_i, \varepsilon_i)'$ independientes e idénticamente distribuidos con media cero y matriz de covarianza Σ diagonal dada por

$$\begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}. \quad (5.3)$$

Si las variables X_i y W_i son aleatorias el modelo se llama *estructural*, mientras que si ambas son determinísticas se denomina *funcional*.

En la práctica un investigador podría despreciar el hecho de que las variables predictoras estén sujeta a un error de medición y estimar los parámetros del modelo mediante Cuadrados Mínimos Ordinarios (OLS). Sin embargo, ésto conduciría a problemas pues, es bien sabido, que en el modelo EV, los estimadores de Cuadrados Mínimos Ordinarios son sesgados y no son consistentes.

Si todas las variables presentes son Gaussianas el modelo está subespecificado. Es frecuente el uso del supuesto de que el cociente $\sigma_\varepsilon/\sigma_u$ es conocido, por ejemplo, sin pérdida de generalidad, puede suponerse que es 1. Bajo estos supuestos los estimadores de máxima verosimilitud de α y β están dados por el método de regresión ortogonal que minimiza la distancia ortogonal de los puntos a la recta de regresión en lugar de la distancia vertical usada en mínimos cuadrados ordinarios.

En el modelo EV multivariado tenemos $\mathbf{Z}_i = (\mathbf{V}'_i, Y_i)'$, $i = 1, \dots, n$, n vectores observables donde \mathbf{V}_i es de dimensión $p \times 1$. Suponemos que:

$$\mathbf{Z}_i = \mathbb{Z}_i + \boldsymbol{\epsilon}_i = \begin{pmatrix} \mathbf{X}_i \\ W_i \end{pmatrix} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n$$

donde $\boldsymbol{\epsilon}_i = (\mathbf{U}'_i, \varepsilon_i)'$ y $W_i = \alpha + \boldsymbol{\beta}'\mathbf{X}_i$ con $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ y las \mathbb{Z}_i son vectores independientes e idénticamente distribuidos con media $\boldsymbol{\mu} = (\boldsymbol{\mu}'_X, \mu_W)'$ y matriz de covarianza

$$\begin{pmatrix} \Sigma_X & \Sigma_X \boldsymbol{\beta} \\ \boldsymbol{\beta}' \Sigma_X \boldsymbol{\beta} \end{pmatrix}.$$

Es usual suponer, además, que los errores $\boldsymbol{\epsilon}_i$ son independientes e idénticamente distribuidos con media $\mathbf{0}$ y matriz de covarianza $\sigma^2 \mathbf{I}_{p+1}$ y que \mathbb{Z}_i y $\boldsymbol{\epsilon}_i$ son independientes. Bajo el supuesto de normalidad, nuevamente el procedimiento de máxima verosimilitud resulta el método de regresión ortogonal. El estimador de regresión ortogonal se obtiene de minimizar

$$\sum_{i=1}^n \|\mathbf{Z}_i - \mathbb{Z}_i\|^2, \quad (5.4)$$

donde \mathbb{Z}_i pertenece a $D_{\alpha, \boldsymbol{\beta}} = \{(\mathbf{X}', W) : W = \alpha + \boldsymbol{\beta}'\mathbf{X}\}$. El modelo multivariado EV se puede reescribir como

$$\mathbf{Z}_i = \mathbb{Z}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\delta}'\mathbb{Z}_i = \alpha \quad i = 1, \dots, n,$$

con $\boldsymbol{\delta} = (\boldsymbol{\beta}', -1)'$. Por lo tanto, el estimador definido por la minimización de (5.4) puede hallarse buscando

$$\operatorname{argmin}_{\|\mathbf{d}\|=1, c} \sum_{i=1}^n (\mathbf{d}'\mathbf{Z}_i - c)^2. \quad (5.5)$$

Notemos que $(\mathbf{d}'\mathbf{Z}_i - c)^2$ es la distancia ortogonal del punto \mathbf{Z}_i al plano $H(\mathbf{d}, c) = \{\mathbf{s} : \mathbf{d}'\mathbf{s} = c\}$. Puede demostrarse que la solución $\hat{\mathbf{d}}$ corresponde a la componente principal de la matriz de covarianza muestral, asociada al autovalor más pequeño y que $\hat{c} = \hat{\mathbf{d}}'\bar{\mathbf{Z}}$, donde $\bar{\mathbf{Z}}$ denota la media muestral. Bajo condiciones de regularidad, se puede demostrar que este método conduce a estimadores consistentes y asintóticamente normales.

Sin embargo, los estimadores resultantes basados en la media y la matriz de covarianza muestrales son muy sensibles a la presencia de datos atípicos, lo que ha motivado la búsqueda de procedimientos robustos alternativos. En particular, Carroll y Gallo (1982) y Brown (1982) señalan la falta de robustez como un problema severo del método de regresión ortogonal.

5.2. Alternativas robustas

Carroll y Gallo (1982) proponen un procedimiento robusto, que resulta consistente y asintóticamente normal, que puede ser usado cuando las observaciones están replicadas exactamente dos

veces. Brown (1982) estudia la posibilidad de usar estimadores de regresión ortogonal ponderados, pero es necesario un estimador inicial de la pendiente para computar los pesos. Ketellapper y Ronner (1984) estudian la aplicación de métodos ordinarios robustos de regresión en el ámbito de los modelos EV.

Zamar (1989) propone M-estimadores robustos ortogonales y muestra que su comportamiento es superior al de los estimadores robustos de regresión ordinarios aplicados al problema ignorando la presencia de errores en las variables. Los estimadores introducidos por Zamar (1989) son la solución del problema de minimización dado por

$$\min_{\|\mathbf{d}\|=1, c} \sum_{i=1}^n \rho \left(\frac{\mathbf{d}'\mathbf{Z}_i - c}{S_n} \right),$$

donde ρ es una función de pérdida y S_n es un estimador de escala robusto de los residuos ortogonales. Es claro que si $\rho(t) = t$ y $S_n = 1$, los estimadores resultantes son la solución del problema (5.5). El rol de ρ es el de controlar la influencia de los residuos que son grandes en comparación de la escala S_n , por lo tanto es necesaria la estimación robusta de la escala de los residuos ortogonales, lo que plantea un problema adicional. Zamar (1989) define los S-estimadores de regresión ortogonales como sigue: dado un vector unitario \mathbf{d} y un escalar c , sea $S(\mathbf{d}, c)$ la M-escala de $\mathbf{d}'\mathbf{Z}_1, \dots, \mathbf{d}'\mathbf{Z}_n$, es decir, la solución de la ecuación

$$\frac{1}{n} \sum_{i=1}^n \chi \left(\frac{\mathbf{d}'\mathbf{Z}_i - c}{S} \right) = b,$$

donde χ es par, continua, no decreciente en $[0, \infty)$, $\chi(0) = 0$ y $\lim_{t \rightarrow \infty} \chi(t) = 1$. La constante b usualmente se toma como $E\chi(Z)$, siendo Z una variable con distribución normal estándar. Los estimadores de escala de regresión ortogonal $(\hat{\mathbf{d}}, \hat{c})$ se definen por la minimización de $S(\mathbf{d}, c)$ y finalmente, $S_n = S(\hat{\mathbf{d}}, \hat{c})$. Zamar (1989) estudia las propiedades asintóticas de los M-estimadores de regresión ortogonal.

Posteriormente, Zamar (1992) considera M-estimadores de regresión ortogonal desde el punto de vista del sesgo. Bajo condiciones de regularidad, caracteriza el sesgo sobre entornos de contaminación y computa las curvas de máximo sesgo. Deriva el M-estimador de regresión ortogonal más robusto desde el punto de vista del sesgo. Asimismo, calcula una cota superior fina para el punto de ruptura de un M-estimador de regresión ortogonal.

Cheng y Van Ness (1992) generalizan la propuesta de Zamar (1989) definiendo M-estimadores de regresión ortogonal generalizados, pero para el caso de regresión simple. Obtienen estimadores con curva de influencia acotada.

Por último, más recientemente, Ruiz-Gazen y Fekri (2004) proponen estimadores robustos de regresión ortogonal pesada. Estos estimadores están definidos por el criterio de mínimos cuadrados ponderados y son derivados de estimadores robustos multivariados de posición y dispersión. Dado que emplearemos estos estimadores en el estudio numérico del último capítulo, los describiremos en detalle en la sección siguiente.

5.3. Regresión Ortogonal Pesada Robusta en el Modelo Errores en las Variables

En el caso de regresión lineal simple con errores en las variables simple tenemos que $\mathbf{Z}_i = (V_i, Y_i)'$ y $\mathbb{Z}_i = (X_i, W_i)'$ cumplen el modelo dado por (5.1), (5.3) y (5.3). Éste es el modelo de efectos aleatorios con \mathbb{Z}_i perteneciente a la dirección afín $D_{\alpha,\beta} = \{(X, W) : W = \alpha + \beta X\}$.

Ruiz-Gazen y Fekri (2004) adaptaron la función objetivo dada en (5.4) introduciendo una función de peso $w : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ y de esta forma definieron una nueva clase de estimadores de α y β minimizando

$$\sum_{i=1}^n w \left(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2 \right) \|\mathbf{Z}_i - \mathbb{Z}_i\|^2, \quad (5.6)$$

donde $\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2 = (\mathbf{Z}_i - \boldsymbol{\mu}_n)' \Sigma_n^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}_n)$, y $\boldsymbol{\mu}_n$ y Σ_n son estimadores multivariados iniciales de posición y dispersión. Notemos que este criterio es invariante por cambios de traslaciones y por transformaciones ortogonales.

Si denotamos por $\Pi_{D_{\alpha,\beta}}$ a la proyección ortogonal en la dirección afín $D_{\alpha,\beta}$, tenemos que

$$\begin{aligned} & \sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2) \|\mathbf{Z}_i - \mathbb{Z}_i\|^2 = \\ & = \sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2) \|\mathbf{Z}_i - \Pi_{D_{\alpha,\beta}}(\mathbf{Z}_i)\|^2 + \sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2) \|\mathbb{Z}_i - \Pi_{D_{\alpha,\beta}}(\mathbf{Z}_i)\|^2. \end{aligned}$$

Entonces, (5.6) se minimiza tomando $\mathbb{Z}_i = \Pi_{D_{\alpha_n,\beta_n}}(\mathbf{Z}_i)$ con D_{α_n,β_n} que minimice a

$$\sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2) \|\mathbf{Z}_i - \Pi_{D_{\alpha,\beta}}(\mathbf{Z}_i)\|^2.$$

Sean

$$\begin{aligned} \boldsymbol{\mu}_{R,n} &= \frac{\sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2) \mathbf{Z}_i}{\sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2)} \\ \text{y } \Sigma_{R,n} &= \frac{\sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2) (\mathbf{Z}_i - \boldsymbol{\mu}_{R,n})(\mathbf{Z}_i - \boldsymbol{\mu}_{R,n})'}{\sum_{i=1}^n w(\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2)} \end{aligned}$$

las versiones reponderadas de un paso de los estimadores $\boldsymbol{\mu}_n$ y Σ_n respectivamente. El mínimo se alcanza en la dirección afín D_{α_n,β_n} que contiene a $\boldsymbol{\mu}_{R,n}$ y es paralela a la dirección generada por el autovector \mathbf{v}_n asociado al autovalor más grande de $\Sigma_{R,n}$.

Los estimadores de regresión ortogonal pesada para la pendiente y para el intercept están dados por:

$$\beta_n = \frac{v_n^{(2)}}{v_n^{(1)}} \quad \text{y} \quad \alpha_n = \boldsymbol{\mu}_{R,n}^{(2)} - \beta_n \boldsymbol{\mu}_{R,n}^{(1)}$$

donde $\mathbf{v}_n = (v_n^{(1)}, v_n^{(2)})'$ y $\boldsymbol{\mu}_{R,n} = (\boldsymbol{\mu}_{R,n}^{(1)}, \boldsymbol{\mu}_{R,n}^{(2)})'$.

La elección de $w \equiv 1$ corresponde a la regresión ortogonal clásica, llamada también Mínimos Cuadrados Totales (TLS). En el caso gaussiano, los estimadores β_n y α_n son los estimadores de máxima verosimilitud. Pero éstos son altamente sensibles a la presencia de datos atípicos. Por esta razón los autores introdujeron la función de peso w , decreciente, la cual asigna bajo peso a aquellas observaciones con $\|\mathbf{Z}_i - \boldsymbol{\mu}_n\|_{\Sigma_n^{-1}}^2$ alto. Los resultados teóricos obtenidos están dados bajo supuestos generales sobre la función de peso w . Consideraron tres estimadores multivariados robustos iniciales de posición y dispersión, los MCD, M- y S-estimadores vistos en el Capítulo 4, agregando a los subíndices de los estimadores $\boldsymbol{\mu}_n$, Σ_n , α_n y β_n las letras M , S y MCD de acuerdo al tipo de estimador inicial que se use.

Como vimos, los M-estimadores multivariados de posición y dispersión son la solución de las ecuaciones (4.17) y (4.18). Sean u_1 y u_2 dos funciones no negativas, los M-estimadores quedan definidos implícitamente por:

$$\boldsymbol{\mu}_{M,n} = \frac{\sum_{i=1}^n u_1(\|\mathbf{Z}_i - \boldsymbol{\mu}_{M,n}\|_{\Sigma_{M,n}^{-1}}) \mathbf{Z}_i}{\sum_{i=1}^n u_1(\|\mathbf{Z}_i - \boldsymbol{\mu}_{M,n}\|_{\Sigma_{M,n}^{-1}})}$$

$$\Sigma_{M,n} = \frac{1}{n} \sum_{i=1}^n u_2(\|\mathbf{Z}_i - \boldsymbol{\mu}_{M,n}\|_{\Sigma_{M,n}^{-1}}^2) (\mathbf{Z}_i - \boldsymbol{\mu}_{M,n})(\mathbf{Z}_i - \boldsymbol{\mu}_{M,n})'$$

Si suponemos que u_2 es monótona, Maronna (1976) mostró que el punto de ruptura de los M-estimadores multivariados es a lo sumo $\frac{1}{p+1}$. Por lo tanto, los M-estimadores son menos robustos a medida que aumenta la dimensión.

Los S-estimadores introducidos en el Capítulo 4 pueden definirse de la siguiente manera. Dada $\rho: \mathbb{R}^+ \rightarrow \mathbb{R}^+$, los S-estimadores $(\boldsymbol{\mu}_{S,n}, \Sigma_{S,n})$ minimizan el $\det(\mathbf{A})$ bajo la siguiente condición

$$\frac{1}{n} \sum_{i=1}^n \rho(\|\mathbf{Z}_i - \boldsymbol{\mu}\|_{\mathbf{A}^{-1}}) = b_0$$

entre todos los $(\boldsymbol{\mu}, \mathbf{A}) \in \mathbb{R}^p \times S_p$, S_p definido en el capítulo anterior. Generalmente, $b_0 = E_F[\rho(\|\mathbf{Z} - \boldsymbol{\mu}\|_{\mathbf{A}^{-1}})]$ donde F es la distribución de \mathbf{Z}_i y ρ es derivable. Para tener un punto de ruptura distinto de cero, ρ debe estar acotada y el BP está dado por el cociente de b_0 y el máximo de ρ (Lopuhaä y Rousseeauw, 1991). En contraste con los M-estimadores, los S-estimadores tienen alto punto de ruptura en cualquier dimensión. Las propiedades de éstos han sido investigadas por Davies (1987) y Lopuhaä (1989).

Finalmente, los estimadores MCD de posición y dispersión, definidos en la Sección 4.6.3, están dados por:

$$\boldsymbol{\mu}_{MCD,n} = \frac{1}{h_n} \sum_{j=1}^{h_n} \mathbf{Z}_{i_j},$$

$$\Sigma_{MCD,n} = \frac{c_\alpha}{h_n} \sum_{j=1}^{h_n} (\mathbf{Z}_{i_j} - \boldsymbol{\mu}_{MCD,n})(\mathbf{Z}_{i_j} - \boldsymbol{\mu}_{MCD,n})'$$

donde el subconjunto $\{\mathbf{Z}_{i_1}, \dots, \mathbf{Z}_{i_{h_n}}\}$ es de tamaño h_n ($1 \leq h_n \leq n$) y c_α es una constante definida para obtener un estimador consistente. Típicamente, el tamaño del conjunto iguala a $h_n = \lceil n(1-\alpha) \rceil$ con $\alpha = 0.25$ o $\alpha = 0.5$ y el punto de ruptura es igual a α . Para asegurar la consistencia del estimador es necesario que

$$c_\alpha = (1 - \alpha) \left(\frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2} + 1)} \int_0^{r_\alpha} r^{p+1} g(r^2) dr \right)^{-1}$$

donde r_α está determinado por

$$\frac{2\pi^{\frac{1}{2}}}{\Gamma(\frac{p}{2})} \int_0^{r_\alpha} r^{p-1} g(r^2) dr = 1 - \alpha.$$

Los estimadores α_n y β_n están calculados a partir de $\boldsymbol{\mu}_{R,n}$ y $\Sigma_{R,n}$, los cuales son estimadores reponderados de un paso con $\boldsymbol{\mu}_n$ y Σ_n como estimadores iniciales. Los estimadores reponderados de un paso han sido estudiados en profundidad en Lopuhaä y Rousseeuw (1991) y Lopuhaä (1999). Se ha probado que estos estimadores heredan las propiedades de robustez de los estimadores iniciales, en particular, se preserva el punto de ruptura.

Para el caso del modelo EV multivariado definido en la Sección 5.1, la dirección afín D_{α_n, β_n} contiene a $\boldsymbol{\mu}_{R,n}$ y es ortogonal a la dirección atravesada por el autovector $\mathbf{v} = (v_{1,n}, \dots, v_{p+1,n})'$ asociado al autovalor más pequeño de $\Sigma_{R,n}$. Luego, los estimadores ortogonales pesados robustos para $\boldsymbol{\beta}$ y α están dados por

$$\boldsymbol{\beta}_n = -\frac{1}{v_{p+1,n}}(v_{1,n}, \dots, v_{p,n})' \quad \text{y} \quad \alpha_n = \mu_{W,n} - \boldsymbol{\beta}_n' \boldsymbol{\mu}_{X,n}$$

5.3.1. Función de Influencia y Propiedades Asintóticas

Como ya vimos en el Capítulo 4 la función de influencia mide la sensibilidad del estimador ante la presencia de pequeñas proporciones de contaminación en la distribución. Las demostraciones de los resultados mencionados en esta sección pueden verse en Ruiz-Gazen y Fekri (2004).

Nos enfocaremos en el problema de estimar a los parámetros β y α del modelo cuando la distribución de \mathbf{Z}_i es elíptica con densidad dada por

$$f(\mathbf{z}) = |\Sigma|^{-\frac{1}{2}} g(\|\mathbf{z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2)$$

con $\boldsymbol{\mu} \in \mathbb{R}^p$, $\Sigma \in \mathcal{S}_p$ y $g: [0, \infty) \rightarrow [0, \infty)$ independiente de \mathbf{z} , $\boldsymbol{\mu}$ y Σ . Las funciones de influencia de los estimadores de la pendiente y del intercept dependen de las funciones de influencia de los estimadores de posición y dispersión.

Lopuhaä (1997) dio la función de influencia de los estimadores multivariados reponderados de un paso de posición y dispersión. A partir de estos resultados se derivaron las funciones de influencia de los estimadores de regresión ortogonal pesada de la pendiente y el intercept.

Las funciones de influencia $IF(\mathbf{z}; \boldsymbol{\mu}_{RC}, F)$ y $IF(\mathbf{z}; \Sigma_{RC}, F)$ de los estimadores reponderados están dadas, para cada $C = M, S$ y MCD , por:

$$IF(\mathbf{z}; \boldsymbol{\mu}_{RC}, F) = \frac{d_2}{d_1} IF(\mathbf{z}; \boldsymbol{\mu}_C, F) + \frac{1}{d_1} w(\|\mathbf{z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2)(\mathbf{z} - \boldsymbol{\mu}),$$

$$IF(\mathbf{z}; \Sigma_{RC}, F) = \frac{d_4}{d_1} IF(\mathbf{z}; \Sigma_C, F) + \frac{d_4}{2d_1} \text{tr}(IF(\mathbf{z}; \Sigma_C, F)) \Sigma \\ + \frac{1}{d_1} w(\|\mathbf{z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2) (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})' - \frac{d_3}{d_1} \Sigma.$$

donde

$$d_1 = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_0^{+\infty} w(r^2) g(r^2) r^{p-1} dr, \\ d_2 = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_0^{+\infty} w(r^2) \left[g(r^2) + \frac{2}{p} g'(r^2) r^2 \right] r^{p-1} dr, \\ d_3 = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_0^{+\infty} \frac{1}{p} w(r^2) g(r^2) r^{p+1} dr, \\ d_4 = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_0^{+\infty} w(r^2) \left[\frac{r^2}{p} g(r^2) + \frac{2r^4}{p(p+2)} g'(r^2) \right] r^{p-1} dr.$$

Finalmente, las funciones de influencia de los estimadores de la pendiente β_C y del intercept α_C están dadas por:

$$IF(\mathbf{z}; \beta_C, F) = \frac{1}{c_0} \frac{\gamma_C(\|\mathbf{z} - \boldsymbol{\mu}\|_{\Sigma^{-1}})}{(\lambda_1 - \lambda_2)v_1} \mathbf{v}'(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})' \boldsymbol{\delta}$$

$$IF(\mathbf{z}; \alpha_C, F) = \boldsymbol{\delta}' IF(\mathbf{z}; \boldsymbol{\mu}, F) - \mu^{(1)} IF(\mathbf{z}; \beta_C, F)$$

con $C = M, S, MCD$, $\mathbf{v} = (v_1, v_2)'$ es el autovector asociado al autovalor más grande de Σ , $\boldsymbol{\delta} = (-\beta, 1)'$,

$$\gamma_{TLS}(t) = 1, \\ \gamma_M(t) = \frac{d_4}{d_1} c_1 u_2(t^2) + \frac{1}{d_1} w(t^2), \\ \gamma_S(t) = \frac{d_4}{d_1} \frac{2}{a_2} \frac{\rho'(t)}{t} + \frac{1}{d_1} w(t^2) \text{ y} \\ \gamma_{MCD}(t) = -\frac{d_4}{d_1} \frac{1}{2a_5} I_{\{t^2 \leq q_\alpha\}} + \frac{1}{d_1} w(t^2)$$

donde c_0 es un factor de consistencia dado por $c_0 = \frac{d_3}{d_1}$.

Luego, las funciones de influencia de los estimadores de la pendiente y el intercept asociados al MCD, M- y S-estimadores están acotadas. La consistencia y la distribución asintótica de estos estimadores se dedujeron de resultados sobre los estimadores reponderados de posición y dispersión. Ruiz-Gazen y Fekri (2004) mostraron que $\boldsymbol{\mu}_{RC,n} \xrightarrow{c.s.} \boldsymbol{\mu}$ y que existe una constante c tal que $\Sigma_{RC,n} \xrightarrow{c.s.} c\Sigma$. También probaron que si w es acotada, continua en casi todo punto de $[0, \infty)$ y es de variación acotada, la distribución asintótica de los estimadores robustos de regresión ortogonal pesada de la pendiente y el intercept son respectivamente:

$$\sqrt{n}(\beta_{C,n} - \beta) \xrightarrow{D} N(0, ASV_C(\beta))$$

$$\sqrt{n}(\alpha_{C,n} - \alpha) \xrightarrow{D} N(0, ASV_C(\alpha))$$

donde

$$ASV_C(\beta) = \frac{(1 + \beta^2)^2}{8c^2} \frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)^2} E(\gamma_C^2(\|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}) \|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^4),$$

$$ASV_C(\alpha) = \frac{1 + \beta^2}{2} \lambda_2 E(\xi_C^2(\|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}) \|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2) + (\mu^{(1)})^2 ASV_C(\beta),$$

con

$$\begin{aligned} \xi_{TLS}(t) &= 1, \\ \xi_M(t) &= \frac{d_2}{d_1 c} u_1(t) + \frac{1}{d_1} w(t^2), \\ \xi_S(t) &= \frac{d_2}{d_1} \frac{\rho'(t)}{a_2 t} + \frac{1}{d_1} w(t^2), \text{ y} \\ \xi_{MCD}(t) &= -\frac{d_2}{d_1} \frac{1}{2a_5} I_{\{t^2 \leq q_\alpha\}} + \frac{1}{d_1} w(t^2). \end{aligned}$$

Finalmente dedujeron las funciones de influencia y las distribuciones asintóticas de la pendiente y el intercept en el modelo EV multivariado para $C = M, S$, y MCD .

- Las funciones de influencia están dadas respectivamente por:

$$IF(\mathbf{z}; \boldsymbol{\beta}_C, F) = \frac{\gamma_C(\|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}})}{a_{p+1} c} \mathbf{I} \boldsymbol{\beta} \sum_{j=1}^p \frac{1}{\lambda_j - \lambda_{p+1}} P_j(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})' \mathbf{a}$$

y

$$IF(\mathbf{z}; \alpha_C, F) = \frac{\xi_C(\|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}})}{a_{p+1}} \mathbf{a}'(\mathbf{z} - \boldsymbol{\mu}) - \boldsymbol{\mu}'_x IF(\mathbf{z}; \boldsymbol{\beta}_C, F),$$

que resultan estar acotadas.

- Si w es acotada, continua en casi todo punto de $[0, \infty)$ y es de variación acotada, $\sqrt{n}(\boldsymbol{\beta}_n - \boldsymbol{\beta})$ converge en distribución a una normal multivariada centrada con matriz de covarianza

$$ASV_C(\boldsymbol{\beta}) = E(\gamma_C^2(\|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}) \|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^4) \frac{1 + \|\boldsymbol{\beta}\|^2}{p(p+2)c^2} \sum_{j=1}^p \frac{\lambda_j \lambda_{p+1}}{(\lambda_j - \lambda_{p+1})^2} \mathbf{I} \boldsymbol{\beta} P_j \mathbf{I}' \boldsymbol{\beta}$$

y $\sqrt{n}(\alpha_n - \alpha)$ converge en distribución a una normal centrada con varianza

$$E(\xi_C^2(\|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}) \|\mathbf{Z} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2) \frac{\lambda_{p+1}(1 + \|\boldsymbol{\beta}\|^2)}{p} + \boldsymbol{\mu}'_x ASV_C(\boldsymbol{\beta}) \boldsymbol{\mu}_x$$

donde $\mathbf{I} \boldsymbol{\beta} = (\mathbf{I}_p, \boldsymbol{\beta})$ y P_j es la proyección ortogonal sobre el autoespacio asociado a λ_j .

Capítulo 6

Estimación Robusta en el Modelo Parcialmente Lineal con Errores en las Variables

6.1. Introducción

Como fue descrito en el Capítulo 1, en los modelos parcialmente lineales suponemos que se observa una muestra de tamaño n tal que $(Y_i, \mathbf{X}'_i, T_i)'$, con $i = 1, \dots, n$, son observaciones independientes, $Y_i \in \mathbb{R}$, $T_i \in \mathbb{R}$, $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})' \in \mathbb{R}^p$ y

$$Y_i = \boldsymbol{\beta}'\mathbf{X}_i + g(T_i) + \varepsilon_i, \quad 1 \leq i \leq n,$$

donde los errores ε_i son independientes, idénticamente distribuidos e independientes de $(\mathbf{X}'_i, T_i)'$. Estamos interesados en la estimación del parámetro desconocido $\boldsymbol{\beta}$ y de la componente no paramétrica g cuando las covariables \mathbf{X}_i son observadas con error, o sea, en lugar de observar \mathbf{X}_i , observamos

$$\mathbf{V}_i = \mathbf{X}_i + \mathbf{U}_i \quad 1 \leq i \leq n,$$

siendo los errores \mathbf{U}_i independientes, idénticamente distribuidos e independientes de $(Y_i, \mathbf{X}'_i, T_i, \varepsilon_i)'$. En el modelo clásico se suele suponer además que $E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma^2$ y además que $E(\mathbf{U}_i) = \mathbf{0}$, $\Sigma_{\mathbf{U}_i} = \sigma^2 \mathbf{I}_p$.

En estas condiciones, $(Y_i, \mathbf{X}'_i, T_i)'$, con $i = 1, \dots, n$, cumplen el modelo dado por

$$\begin{aligned} Y_i &= \boldsymbol{\beta}'\mathbf{X}_i + g(T_i) + \varepsilon_i \\ \mathbf{V}_i &= \mathbf{X}_i + \mathbf{U}_i. \end{aligned} \tag{6.1}$$

Como ya fue mencionado en la Introducción si se aplica el estimador de verosimilitud semiparamétrico de Severini y Staniswalis (1994) ignorando el hecho de que las \mathbf{X}_i 's son observadas con error, el estimador de $\boldsymbol{\beta}$ no es consistente. Liang, Härdle y Carroll (1999) propusieron un estimador consistente de $\boldsymbol{\beta}$ basado en mínimos cuadrados en el que introducen un **término de atenuación** que depende de la matriz de covarianza de los errores \mathbf{U}_i , la que asumen como conocida. Zhu y Cui

(2003) consideran el modelo (6.1) en el caso en que todas las covariables, incluyendo las T_i , son observadas con error y deducen las propiedades asintóticas de los estimadores propuestos.

Sin embargo, las propuestas existentes son sensibles a la presencia de datos anómalos en tanto se basan en mínimos cuadrados. El propósito de esta tesis es introducir estimadores robustos consistentes de las componentes paramétrica y no paramétrica de un modelo parcialmente lineal con errores en las variables bajo condiciones generales.

6.2. Modelo y Enfoque Clásico

Supongamos que $(Y, \mathbf{V}', T)'$ es un vector aleatorio con la misma distribución de $(Y_i, \mathbf{V}'_i, T_i)'$, esto es

$$\begin{aligned} Y &= \boldsymbol{\beta}'\mathbf{X} + g(T) + \varepsilon \\ \mathbf{V} &= \mathbf{X} + \mathbf{U}, \end{aligned} \quad (6.2)$$

donde \mathbf{U} es independiente de $(Y, \mathbf{X}', T)'$ y de ε .

Siguiendo el enfoque clásico, sean $\phi_0(t) = E(Y|T = t)$ y para cada coordenada j , $j = 1, \dots, p$, $\phi_j(t) = E(V_j|T = t) = E(X_j|T = t)$ y $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_p(t))'$.

Por un lado, de (6.2) se sigue que

$$\begin{aligned} Y &= \boldsymbol{\beta}'(\mathbf{V} - \mathbf{U}) + g(T) + \varepsilon \\ Y &= \boldsymbol{\beta}'\mathbf{V} + g(T) + (\varepsilon - \boldsymbol{\beta}'\mathbf{U}) \\ Y &= \boldsymbol{\beta}'\mathbf{V} + g(T) + \tau \end{aligned} \quad (6.3)$$

siendo $\tau = \varepsilon - \boldsymbol{\beta}'\mathbf{U}$. Luego, tomando esperanza condicional resulta

$$\begin{aligned} E(Y|T = t) &= \boldsymbol{\beta}'E(\mathbf{V}|T = t) + g(t) \\ \phi_0(t) &= \boldsymbol{\beta}'\boldsymbol{\phi}(t) + g(t) \end{aligned} \quad (6.4)$$

y por lo tanto

$$g(t) = \phi_0(t) - \boldsymbol{\beta}'\boldsymbol{\phi}(t).$$

De las ecuaciones (6.3) y (6.4) obtenemos que

$$Y - \phi_0(t) = \boldsymbol{\beta}'(\mathbf{V} - \boldsymbol{\phi}(t)) + \tau,$$

es decir

$$\begin{aligned} Y - \phi_0(t) &= \boldsymbol{\beta}'(\mathbf{X} - \boldsymbol{\phi}(t)) + \varepsilon \\ \mathbf{V} &= \mathbf{X} + \mathbf{U}. \end{aligned}$$

Para cada $T = t$, esto sugiere estimar previamente $\phi_0(t)$ y $\boldsymbol{\phi}(t)$, mediante $\widehat{\phi}_0(t)$ y $\widehat{\boldsymbol{\phi}}(t)$, e insertar estos estimadores en el proceso de estimación de $\boldsymbol{\beta}$. De esta manera combinaríamos estimadores no paramétricos para estimar $\phi_0(t)$ y $\boldsymbol{\phi}(t)$ con estimadores de regresión ortogonal para la estimación de $\boldsymbol{\beta}$.

Para el caso en que la matriz de covarianza de los errores \mathbf{U}_i , $\Sigma_{\mathbf{U}}$, es conocida, Liang, Härdle and Carroll (1999) propusieron un estimador consistente del parámetro de regresión basado en mínimos cuadrados en el que introducen un **término de atenuación** que depende de $\Sigma_{\mathbf{U}}$. En efecto, el estimador de β propuesto por estos autores queda definido por

$$\left[\sum_{i=1}^n (\mathbf{V}_i - \hat{\phi}(T_i)) (\mathbf{V}_i - \hat{\phi}(T_i))' - n\Sigma_{\mathbf{U}} \right]^{-1} \sum_{i=1}^n (\mathbf{V}_i - \hat{\phi}(T_i)) (\mathbf{Y}_i - \hat{\phi}_0(T_i)).$$

Sin embargo, estos estimadores, basados en técnicas clásicas, son muy sensibles a la presencia de datos atípicos. Una alternativa natural podría basarse en utilizar en cada paso del proceso de estimación un procedimiento menos sensible a datos anómalos, tal como fue propuesto en Bianco y Boente (2004) en el contexto del modelo parcialmente lineal. En efecto si los errores ε y \mathbf{U}_j tienen distribución simétrica, entonces $\phi_0(t)$ y $\phi(t)$ pueden pensarse como funcionales robustos condicionales de posición, tales como la mediana condicional, y el parámetro de regresión podría ser estimado mediante un procedimiento de regresión ortogonal robusto, como los mencionados en el capítulo anterior.

6.3. Propuesta

Considerando los aspectos mencionados previamente, proponemos el siguiente procedimiento de tres pasos:

- **Paso 1:** Estimar $\phi_j(t)$, $0 \leq j \leq p$, mediante un suavizado robusto, como la mediana local o un M-estimador local. Llamamos $\hat{\phi}_j(t)$, $0 \leq j \leq p$ y sea $\hat{\phi}(t) = (\hat{\phi}_1(t), \dots, \hat{\phi}_p(t))'$.
- **Paso 2:** Estimar β aplicando un método de regresión robusto para un modelo con errores–en–las–variables con variables observables

$$\tilde{\mathbf{V}}_i = \mathbf{V}_i - \hat{\phi}(T_i) \quad \text{y} \quad R_i = Y_i - \hat{\phi}_0(T_i).$$

Sea $\hat{\beta}$ el estimador obtenido.

- **Paso 3:** Definir el estimador de la función g como

$$\hat{g}(t) = \hat{\phi}_0(t) - \hat{\beta}' \hat{\phi}(t).$$

Como ya fue mencionado, podemos considerar para el **Paso 1** las medianas o M-estimadores locales. Definimos primero los pesos basados en núcleos como:

$$w_i(t) = \frac{K\left(\frac{T_i-t}{h}\right)}{\sum_{j=1}^n K\left(\frac{T_j-t}{h}\right)}.$$

Computamos las medianas locales $\widehat{\phi}_{0,med}(t)$ y $\widehat{\phi}_{j,med}(t)$ como la mediana de la función de distribución empírica condicional $\widehat{F}_0(y|T=t)$ y $\widehat{F}_j(v|T=t)$, respectivamente definidas como

$$\widehat{F}_0(y|T=t) = \sum_{i=1}^n w_i(t) I_{(-\infty, y]}(Y_i) \quad (6.5)$$

$$\widehat{F}_j(v|T=t) = \sum_{i=1}^n w_i(t) I_{(-\infty, v]}(V_{ij}) \quad 1 \leq j \leq p. \quad (6.6)$$

Notemos que $\widehat{F}_0(y|T=t)$ y $\widehat{F}_j(v|T=t)$ proveen de estimadores de la distribución de $Y|T=t$ y $V_j|T=t$, que se denotan $F_0(y|T=t)$ y $F_j(v|T=t)$, respectivamente.

Los M-estimadores locales, $\widehat{\phi}_{0,M}(t)$ y $\widehat{\phi}_{j,M}(t)$, son definidos como los M-estimadores de posición relacionados con $\widehat{F}_0(y|T=t)$ y $\widehat{F}_j(v|T=t)$, respectivamente. Luego, ellos son solución de

$$\sum_{i=1}^n w_i(t) \left(\frac{Y_i - \widehat{\phi}_{0,M}(t)}{\widehat{S}_0(t)} \right) = 0 \quad \sum_{i=1}^n w_i(t) \left(\frac{V_{ij} - \widehat{\phi}_{j,M}(t)}{\widehat{S}_j(t)} \right) = 0,$$

donde ψ es impar, creciente, acotada y continua y $\widehat{S}_0(t)$ y $\widehat{S}_j(t)$ son estimadores locales robustos de escala. Elecciones posibles para la función de score ψ son la función de score de Huber o de la bicuadrada, mientras que las escalas $\widehat{S}_0(t)$ y $\widehat{S}_j(t)$ pueden ser tomadas como la *MAD* local, es decir, la *MAD* con respecto a las distribuciones definidas en (6.5) y (6.6).

En cuanto a la elección de la función ψ en esta etapa de la estimación, tengamos en cuenta que si estamos tratando con una distribución condicional de colas pesadas, una elección natural como medida de posición en este contexto sería la mediana condicional. Esta correspondería a elegir $\psi(u) = \text{sg}(u)$. Cuando la distribución condicional es simétrica, la mediana condicional es el centro de la distribución y por lo tanto, cualquier función de score impar da lugar a un M-estimador consistente de la mediana condicional.

Como se describió en el **Paso 2** una vez calculados los estimadores $\widehat{\phi}_0(t)$ y $\widehat{\phi}(t)$ de $\phi_0(t)$ y $\phi(t)$ respectivamente, la estimación robusta del parámetro de regresión puede realizarse aplicando a los residuos $\widetilde{\mathbf{V}}_i = \mathbf{V}_i - \widehat{\phi}(T_i)$ y $R_i = Y_i - \widehat{\phi}_0(T_i)$ un método robusto para el problema de regresión lineal con errores en las variables. En particular, en las aplicaciones numéricas que se reportan en el siguiente capítulo se utilizaron los estimadores de regresión ortogonal pesada robusta introducidos por Ruiz-Gazen y Fekri (2004).

A fin de garantizar consistencia en el sentido de Fisher es necesario que $g(t) = \phi_0(t) - \beta' \phi(t)$. Si, por ejemplo, $V_j|T=t$ es simétrica alrededor de $\phi_j(t)$ y $Y|T=t$ es simétrica alrededor de $\phi_0(t) = \beta' \phi(t) - g(t)$, entonces los estimadores obtenidos en el **Paso 2** en efecto serán consistentes.

6.4. Consistencia

Sean $S_0(t)$ y $S_j(t)$ la *MAD* de la distribución condicional de $Y|T=t$ y de $V_j|T=t$ respectivamente, y definimos $\phi_j(t)$, $0 \leq j \leq p$ como las soluciones de

$$E \left[\left(\frac{Y - \phi_0(t)}{S_0(t)} \right) \middle| T = t \right] = 0, \quad E \left[\left(\frac{V_j - \phi_j(t)}{S_j(t)} \right) \middle| T = t \right] = 0, \quad 1 \leq j \leq p.$$

Necesitaremos las siguientes hipótesis que involucran a la función de score, a la distribución marginal de T , las distribuciones condicionales respecto de T , el núcleo y el orden de convergencia del parámetro de suavizado:

H1. es una función impar, estrictamente creciente, acotada y con derivada continua, tal que

$$a^{-1}(a) \leq a.$$

H2. $F_0(y|T = t)$ y $F_j(v|T = t)$, $1 \leq j \leq p$, son simétricas alrededor de $\phi_0(t)$ y $\phi_j(t)$ respectivamente.

H3. Para todo conjunto compacto $\mathbf{C} \subset \mathbb{R}$, la densidad de T , f_T está acotada sobre \mathbf{C} y

$$\inf_{t \in \mathbf{C}} f_T(t) > 0.$$

H4. $F_0(y|T = t)$ y $F_j(v|T = t)$, $1 \leq j \leq p$, son funciones continuas en t . Además satisfacen la siguiente condición de equicontinuidad para todo conjunto compacto $\mathbf{C} \subset \mathbb{R}$

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad / \quad |a - b| < \delta \quad \Rightarrow \quad \sup_{t \in \mathbf{C}} \max_{0 \leq j \leq p} (|F_j(a|T = t) - F_j(b|T = t)|) < \varepsilon.$$

H5. El núcleo $K : \mathbb{R} \rightarrow \mathbb{R}$ es una función no negativa y acotada tal que

$$\int K(a) da = 1, \quad \int |a|K(a) da < \infty, \quad |a|K(a) \rightarrow 0 \text{ si } |a| \rightarrow \infty$$

y satisface la condición de Lipschitz de orden uno.

H6. La sucesión $h = h_n$ es tal que

$$h_n \rightarrow 0, \quad nh_n \rightarrow \infty \text{ y } \frac{nh_n}{\log n} \rightarrow \infty.$$

El siguiente resultado es necesario para asegurar la consistencia tanto del parámetro de regresión como de la función g , cuando el suavizado está basado en medianas locales o M-estimadores locales con pesos basados en núcleos.

Proposition 6.4.1. *Supongamos que valen **H3** a **H6**. Luego, para todo conjunto compacto \mathbf{C} ,*

(a) *bajo **H1** y **H2**, tenemos que $\sup_{t \in \mathbf{C}} |\widehat{\phi}_{j,M}(t) - \phi_j(t)| \xrightarrow{c.s.} 0$, $0 \leq j \leq p$*

(b) *si, además, $F_j(v|T = t)$ tiene una única mediana en $\phi_j(t)$, para $0 \leq j \leq p$, tenemos que*

$$\sup_{t \in \mathbf{C}} |\widehat{\phi}_{j,med}(t) - \phi_j(t)| \xrightarrow{c.s.} 0, \quad 0 \leq j \leq p.$$

Antes de probar la consistencia del estimador robusto de regresión definido en la sección anterior, enunciaremos el siguiente lema auxiliar.

Lema 6.4.1. Sean $(R_i, \mathbf{S}'_i, T_i)' \in \mathbb{R}^{p+2}$, $1 \leq i \leq n$ vectores aleatorios independientes e idénticamente distribuidos sobre $(\Omega, \mathcal{A}, \mathcal{P})$ tales que $(R_i, \mathbf{S}'_i)'$ tiene distribución común P . Sean $\hat{\eta}_0(t)$ y $\hat{\boldsymbol{\eta}}(t) = (\hat{\eta}_1(t), \dots, \hat{\eta}_p(t))'$ funciones aleatorias tales que para todo conjunto compacto $\mathbf{C} \subset \mathbb{R}$

$$\sup_{t \in \mathbf{C}} |\hat{\eta}_j(t)| \xrightarrow{c.s.} 0, \quad 0 \leq j \leq p.$$

Sean P_n y Q_n las siguientes medidas empíricas sobre \mathbb{R}^{p+1}

$$P_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(R_i, \mathbf{S}_i) \quad Q_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(R_i + \hat{\eta}_0(T_i), \mathbf{S}_i + \hat{\boldsymbol{\eta}}(T_i)),$$

donde $A \subset \mathbb{R}^{p+1}$ es un conjunto Boreliano. Luego

(a) Para toda función continua y acotada $f : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ tenemos que

$$|E_{Q_n}(f) - E_{P_n}(f)| \xrightarrow{c.s.} 0$$

(b) $\Pi(Q_n, P) \xrightarrow{c.s.} 0$, con Π la distancia de Prohorov.

La demostración de los dos resultados anteriores puede verse en Bianco y Boente (2004). Haciendo una adaptación del Teorema 1 del mismo trabajo obtenemos el siguiente teorema que nos garantiza la consistencia del parámetro de regresión.

Teorema 6.4.1. Sean $(Y_i, \mathbf{V}'_i, T_i)', 1 \leq i \leq n$ vectores aleatorios independientes que satisfacen

$$\begin{aligned} Y_i &= \boldsymbol{\beta}' \mathbf{X}_i + g(T_i) + \varepsilon_i \\ \mathbf{V}_i &= \mathbf{X}_i + \mathbf{U}_i \end{aligned} \quad (6.7)$$

Sea P la distribución de $(R_i, \tilde{\mathbf{V}}'_i)' = (Y_i - \phi_0(T_i), \mathbf{V}'_i - \boldsymbol{\phi}(T_i))'$, donde $\phi_0(t)$ y $\boldsymbol{\phi}(t)$ están definidas en **H2**, con $\phi_0(t) = \boldsymbol{\beta}' \boldsymbol{\phi}(t) + g(t)$. Supongamos que $\hat{\phi}_j(t)$, $0 \leq j \leq p$ son estimadores de $\phi_j(t)$ tales que para todo conjunto compacto $\mathbf{C} \subset \mathbb{R}$

$$\sup_{t \in \mathbf{C}} |\hat{\phi}_j(t) - \phi_j(t)| \xrightarrow{c.s.} 0, \quad 0 \leq j \leq p. \quad (6.8)$$

Sea $\beta(H)$ un funcional de regresión para el modelo

$$\mathbf{Z} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b}'\mathbf{a} \end{pmatrix} + \begin{pmatrix} \mathbf{u} \\ \varepsilon \end{pmatrix} \quad (6.9)$$

donde $\mathbf{Z} \sim H$, \mathbf{u} es independiente de \mathbf{a} , y ε es independiente de \mathbf{a} y de \mathbf{u} . Supongamos que $\beta(H)$ es continua en P y que provee estimadores Fisher-consistentes. Si

$$\hat{P}_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(\hat{R}_i, \hat{\tilde{\mathbf{V}}}_i)$$

con $\hat{R}_i = Y_i - \hat{\phi}_0(T_i)$, $\hat{\tilde{\mathbf{V}}}'_i = \mathbf{V}'_i - \hat{\boldsymbol{\phi}}(T_i)'$ y $\hat{\boldsymbol{\beta}}_{ROB} = \beta(\hat{P}_n)$, entonces

$$\hat{\boldsymbol{\beta}}_{ROB} \xrightarrow{c.s.} \boldsymbol{\beta}.$$

Demostración: < Tenemos que

$$\widehat{P}_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(Y_i - \widehat{\phi}_0(T_i), \mathbf{V}_i - \widehat{\boldsymbol{\phi}}(T_i)).$$

Luego, por (6.8) aplicando el Lema 6.4.1 obtenemos que

$$|E_{\widehat{P}_n}(f) - E_{P_n}(f)| \xrightarrow{c.s.} 0,$$

siendo

$$P_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(Y_i - \phi_0(T_i), \mathbf{V}_i - \boldsymbol{\phi}(T_i))$$

y además

$$\Pi(\widehat{P}_n, P) \xrightarrow{c.s.} 0.$$

Como $\beta(H)$ es continua en P ,

$$\widehat{\boldsymbol{\beta}}_{ROB} = \beta(\widehat{P}_n) \xrightarrow{c.s.} \beta(P).$$

Por lo que sólo faltaría ver que $\beta(P) = \boldsymbol{\beta}$. Dado que $(Y_i, \mathbf{V}_i', T_i)'$, $1 \leq i \leq n$, satisfacen el modelo (6.7) y $g(t) = \phi_0(t) - \boldsymbol{\beta}'\boldsymbol{\phi}(t)$, entonces tenemos que

$$R_i = Y_i - \phi_0(T_i) = \boldsymbol{\beta}'(\mathbf{X}_i - \boldsymbol{\phi}(T_i)) + \varepsilon_i.$$

y además

$$\widetilde{\mathbf{V}}_i = \mathbf{V}_i - \boldsymbol{\phi}(T_i) = \mathbf{X}_i - \boldsymbol{\phi}(T_i) + \mathbf{U}_i.$$

Por lo tanto, $(R_i, \widetilde{\mathbf{V}}_i')$ satisfacen el modelo (6.9) con $\mathbf{b} = \boldsymbol{\beta}$ y en consecuencia, $\beta(P) = \boldsymbol{\beta}$, como queríamos probar. >

Corolario 6.4.1. Sean $(Y_i, \mathbf{V}_i', T_i)'$, $1 \leq i \leq n$, vectores aleatorios independientes que satisfacen (6.1) y supongamos que $\widehat{\phi}_j(t)$, $0 \leq j \leq p$, son estimadores de $\phi_j(t)$ tales que para todo conjunto compacto $\mathbf{C} \subset \mathbb{R}$

$$\sup_{t \in \mathbf{C}} |\widehat{\phi}_j(t) - \phi_j(t)| \xrightarrow{c.s.} 0, \quad 0 \leq j \leq p.$$

Bajo las condiciones del Teorema 6.4.1, los estimadores $\widehat{g}(t) = \widehat{\phi}_0(t) - \widehat{\boldsymbol{\beta}}_{ROB}'\widehat{\boldsymbol{\phi}}(t)$ de la función de regresión g son uniformemente consistentes sobre compactos.

Capítulo 7

Estudio de Monte Carlo

Hemos realizado un estudio de simulación con el fin de comparar el comportamiento de los nuevos estimadores propuestos con su versión clásica frente a distintos tipos de contaminación. Como ya hemos mencionado, Zhu y Cui (2003) propusieron un estimador clásico para el modelo parcialmente lineal con errores en todas las variables. El modelo que consideramos a continuación está basado en el que estos autores utilizaron en su trabajo y está dado por:

$$\begin{aligned} Y_i &= \boldsymbol{\beta}'\mathbf{X}_i + \sin\left(\frac{1}{2}\pi T_i\right) + \varepsilon_i, \quad 1 \leq i \leq n, \\ \mathbf{V}_i &= \mathbf{X}_i + \mathbf{U}_i, \quad 1 \leq i \leq n \end{aligned}$$

donde $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (1, 1)'$, $T_i = \frac{X_{i1} + X_{i2} + D_i}{\sqrt{6}/3}$, $(\mathbf{X}_i', D_i)' \sim N(\boldsymbol{\mu}, \Sigma)$ con $\boldsymbol{\mu} = (0, 0, 0)'$ y $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\varepsilon_i \sim N(0, \sigma^2)$ y $U_{ij} \sim N(0, \sigma^2)$ para $j = 1, 2$ con $\sigma = 0.3$.

El estudio de simulación se llevó a cabo generando 1000 muestras independientes de tamaño $n = 100$. Notaremos por C_0 al caso no contaminado, y por C_1 , C_2 y C_3 a las siguientes tres contaminaciones:

C_1 : $\varepsilon_1, \dots, \varepsilon_n$, son i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 10\sigma^2)$. De esta manera se agrandan los errores de las variables de respuesta.

C_2 : $\varepsilon_1, \dots, \varepsilon_n$, son i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 10\sigma^2)$ y además U_{1j}, \dots, U_{nj} son i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 10\sigma^2)$ para cada $j = 1, 2$. En esta contaminación se agrandan también a los errores de las covariables de regresión.

C_3 : $\varepsilon_1, \dots, \varepsilon_n$, son i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 10\sigma^2)$ y se reemplazaron además artificialmente 5 observaciones de las covariables de regresión, pero no las variables de respuesta, por el valor $(20, 20)'$ para valores equiespaciados de T . Esta última contaminación corresponde a introducir puntos de alto leverage.

Para el **Paso 1** del procedimiento se realizaron tres suavizados con pesos basados en un núcleo “normal”:

- suavizado clásico,
- suavizado basado en M-estimadores locales con función de score bicuadrada con constante $k = 4.685$,
- suavizado por medianas locales,

utilizándose para cada uno de éstos tres ventanas distintas $h = 0.8$, $h = 1$ y $h = 1.2$.

Luego del suavizado de las variables, en el **Paso 2** estimamos β mediante:

- mínimos cuadrados totales,
- regresión ortogonal pesada introducido por Ruiz-Gazen y Fekri (2004) con función de peso

$$w(x) = I_{[0,c]}(x), \quad c = \chi_{3,0.975}^2$$

con los siguientes estimadores multivariados iniciales de posición y dispersión respectivamente:

- 1) Estimadores MCD,
- 2) M-estimadores,
- 3) S-estimadores con ρ -función bicuadrada.

Finalmente, una vez obtenidos los distintos estimadores \hat{g} de g como se definió en el **Paso 3**, se midió su performance mediante dos medidas:

$$MSE(\hat{g}) = \frac{1}{n} \sum_{i=1}^n [\hat{g}(t_i) - g(t_i)]^2,$$

$$MedSE(\hat{g}) = Med([\hat{g}(t_i) - g(t_i)]^2).$$

En los siguientes cuadros presentamos medidas resúmenes de los resultados obtenidos tales como la media, la mediana, el error cuadrático medio (ECM) y el error cuadrático mediano (ECMed) de los estimadores de regresión de β_1 y β_2 . Notaremos por TLS al estimador obtenido por mínimos cuadrados totales luego de un suavizado clásico, por RMCD.med, RM.med y RS.med a aquellos obtenidos por regresión ortogonal pesada con MCD-, M- y S-estimadores multivariados iniciales, respectivamente, luego de un suavizado por medianas locales y finalmente, RM-CD.m, RM.m y RS.m notarán a aquellos estimadores obtenidos por regresión ortogonal pesada con MCD-, M- y S-estimadores multivariados iniciales, respectivamente, luego de un suavizado basado en M-estimadores locales.

Los Cuadros 7.1 y 7.2 contienen los valores correspondientes, a los estimadores de β_1 y β_2 , respectivamente, cuando el tamaño de la ventana es $h = 0.8$, los Cuadros 7.3 y 7.4 reflejan los valores para $h = 1$ y los Cuadros 7.5 y 7.6 para $h = 1.2$. La Figura 7.1 muestra los boxplots de los estimadores de β_1 y β_2 para los distintos valores de la ventana para el caso no contaminado.

Además hemos graficado los estimadores de las densidades de los $MSE(\hat{g})$ y de los $MedSE(\hat{g})$. Los gráficos en negro corresponden a las densidades en el caso no contaminado, mientras que aquéllos de color marrón corresponden a C_1 , aquéllos de color celeste a C_2 y los de color rosa a C_3 .

En las Figuras 7.2 y 7.3 se hallan los gráficos de las densidades de los $MSE(\hat{g})$ correspondientes al método clásico y a los métodos robustos, respectivamente, cuando la ventana es $h = 0.8$. Los estimadores de las densidades fueron calculados utilizando el núcleo normal con ventana 0.1 en la mayoría de los casos. Debido al rango diferente de la variación del método clásico utilizamos, en este caso, para la contaminación C_1 una ventana igual a 0.8, para la contaminación C_2 una ventana igual a 0.5 y para la C_3 una ventana igual a 10. En las Figuras 7.4 y 7.5 se encuentran los gráficos de las densidades de los $MSE(\hat{g})$ correspondientes al método clásico y a los métodos robustos, respectivamente, cuando la ventana es $h = 1$, mientras que en las Figuras 7.6 y 7.7 los correspondientes a la ventana $h = 1.2$. Para calcular los estimadores de las densidades se utilizó el núcleo normal y se usaron los mismos valores para las ventanas que en el caso $h = 0.8$.

Para el caso de los $MedSE(\hat{g})$ también sucede que el rango de la variación del método clásico es diferente ante la presencia de contaminación, teniendo que utilizar distintas ventanas para la estimación de las densidades respectivas. En las Figuras 7.8 y 7.9 se grafican las densidades de los $MedSE(\hat{g})$ correspondientes al método clásico y a los métodos robustos, respectivamente, cuando la ventana es $h = 0.8$. Los estimadores de las densidades fueron calculados utilizando el núcleo normal con ventana 0.5 para el método clásico bajo las contaminaciones C_1 y C_2 y 10 para la contaminación C_3 , mientras que para el resto de los métodos se utilizó una ventana igual a 0.1. En las Figuras 7.10 y 7.11 están los gráficos de las densidades de los $MedSE(\hat{g})$ correspondientes al método clásico y a los métodos robustos, respectivamente, cuando la ventana es $h = 1$. En este caso se usó para la estimación de las densidades el núcleo normal con las mismas ventanas que para $h = 0.8$, salvo para el método clásico que bajo la contaminación C_2 se usó una ventana de 0.3 y para la contaminación C_3 una de valor 5. Por último, en las Figuras 7.12 y 7.13 se hallan los gráficos de las densidades de los $MedSE(\hat{g})$ correspondientes al método clásico y a los métodos robustos, respectivamente, cuando la ventana es $h = 1.2$. Los respectivos estimadores de las densidades fueron calculados utilizando el núcleo normal con las mismas ventanas que para el caso $h = 1$.

7.1. Conclusiones

Como era de esperar, del estudio de Monte Carlo podemos concluir a grandes rasgos que el estimador clásico tiene un comportamiento muy pobre ante la presencia de contaminación, en cualquiera de las tres situaciones consideradas, mientras que los estimadores robustos logran superarlo.

Con respecto a la estimación de β por el método clásico, analizando los resultados que figuran en los cuadros para ambas coordenadas y para todas las ventanas consideradas, podemos ver que los estimadores correspondientes bajo la contaminación C_1 sobreestiman al valor verdadero de los parámetros duplicando por lo menos su verdadero valor y se observa un incremento de los errores cuadráticos medios y medianos con respecto al caso no contaminado. Cuando se contamina siguiendo el esquema C_2 el incremento de los errores cuadráticos es mayor que si se lo hace usando C_1 , mientras que los estimadores no se alejan en demasía de los valores reales. Para el caso de la contaminación C_3 los estimadores clásicos se alejan considerablemente de los parámetros verdaderos y los errores cuadráticos crecen de manera exagerada. Cabe observar que estos últimos valores se ven altamente afectados por algunas estimaciones muy alejadas obtenidas en algunas replicaciones. En resumen el error cuadrático medio aumenta por lo menos cien veces bajo estas contaminaciones.

Si la estimación de β se realiza mediante los métodos robustos, ésta no se ve afectada por la

presencia de contaminación, es decir los estimadores obtenidos se acercan a los valores verdaderos de los parámetros y los errores aumentan levemente, pero permaneciendo pequeños. Al mismo tiempo, en el caso no contaminado el comportamiento de los estimadores robustos es similar al de los estimadores clásicos. Ésto sucede para los tres valores utilizados para la ventana.

Si comparamos a los estimadores obtenidos para β luego de un suavizado por medianas locales (estimadores.med) con aquéllos obtenidos luego de un suavizado basado en M-estimadores locales (estimadores.m), podemos observar que si no hay contaminación o si contaminamos según C_1 , según C_2 o según C_3 , el sesgo de los estimadores basados en medianas locales es ligeramente menor que el sesgo de los basados en M-estimadores locales, mientras que los errores cuadráticos de los estimadores.m son menores que los errores cuadráticos de los estimadores.med. Por lo tanto, podríamos decir que los estimadores.m son mejores para las tres ventanas h consideradas.

Si miramos entre los estimadores basados en medianas locales veremos que, cualquiera sea la ventana utilizada, los estimadores obtenidos por el método de regresión ortogonal pesada utilizando M-estimadores como estimadores multivariados iniciales son los que tienen menores errores cuadráticos en la mayoría de los casos considerados entre C_0 , C_1 , C_2 y C_3 . Observemos que sucede lo mismo si miramos entre los estimadores.m, salvo que para la ventana $h = 1$ el estimador S.m de la primera coordenada también tiene sus errores menores en la mayoría de los los casos entre C_0 , C_1 , C_2 y C_3 .

Podríamos concluir que entre todos los distintos estimadores propuestos para estimar β , M.m es el más aconsejable.

Cabe destacar que no se observan grandes diferencias entre las distintas ventanas consideradas, sin embargo restaría realizar un estudio para la selección de una ventana óptima con un criterio que sea resistente a la presencia de outliers.

Con respecto a la estimación de g , cuando se estima mediante el método clásico, vemos en los gráficos de las densidades que hay un aumento del $MSE(\hat{g})$ y del $MedSE(\hat{g})$ ante la presencia de contaminación observándose un comportamiento similar en las tres ventanas consideradas. En los casos en que se estimó robustamente vemos que el comportamiento de las densidades no varía considerablemente entre las distintas ventanas y que tanto para los $MSE(\hat{g})$ como para los $MedSE(\hat{g})$ el estimador de la densidad que más se parece al estimador clásico cuando no hay contaminación son los estimadores .m. Luego, si miramos entre los distintos estimadores .m, vemos que son comparables y que tienen un comportamiento equivalente.

7.2. Cuadros

Cuadro 7.1: Estimación de la primera coordenada β_1 de β , con ventana $h = 0.8$

	TLS	RMCD.Med	RM.med	RS.med	RMCD.m	RM.m	RS.m	
Media	1.0162	1.0165	1.0136	1.0168	1.0227	1.0210	1.0205	C_0
Mediana	1.0129	1.0072	1.0097	1.0110	1.0211	1.0172	1.0163	
ECM	0.0073	0.0214	0.0143	0.0158	0.0136	0.0106	0.0114	
ECMed	0.0033	0.0086	0.0058	0.0061	0.0055	0.0043	0.0049	
Media	2.1898	1.0507	1.0576	1.0526	1.0769	1.0812	1.0780	C_1
Mediana	1.9946	1.0375	1.0399	1.0370	1.0616	1.0689	1.0633	
ECM	0.7488	0.0280	0.0181	0.0203	0.0204	0.0155	0.0167	
ECMed	0.2532	0.0093	0.0074	0.0080	0.0071	0.0065	0.0068	
Media	1.3715	1.0506	1.0662	1.0573	1.0696	1.0792	1.0723	C_2
Mediana	1.0449	1.0290	1.0465	1.0413	1.0559	1.0625	1.0557	
ECM	3.5659	0.0289	0.0245	0.0237	0.0218	0.0196	0.0192	
ECMed	0.4156	0.0128	0.0102	0.0111	0.0090	0.0074	0.0084	
Media	-0.6915	1.0527	1.0646	1.0585	1.0846	1.0938	1.0881	C_3
Mediana	-0.2773	1.0431	1.0517	1.0453	1.0737	1.0820	1.0830	
ECM	9250.23	0.0267	0.0200	0.0213	0.0191	0.0172	0.0172	
ECMed	22.453	0.0120	0.0091	0.0099	0.0081	0.0074	0.0070	

Cuadro 7.2: Estimación de la segunda coordenada β_2 de β , con ventana $h = 0.8$

	TLS	RMCD.Med	RM.med	RS.med	RMCD.m	RM.m	RS.m	
Media	1.0174	1.0191	1.0172	1.0188	1.0236	1.0214	1.0206	C_0
Mediana	1.0145	1.0100	1.0115	1.0123	1.0166	1.0170	1.0154	
ECM	0.0076	0.0226	0.0152	0.0171	0.0144	0.0107	0.0115	
ECMed	0.0032	0.0087	0.0066	0.0074	0.0068	0.0049	0.0052	
Media	2.1822	1.0529	1.0583	1.0532	1.0758	1.0812	1.0779	C_1
Mediana	2.0128	1.0403	1.0522	1.0477	1.0628	1.0730	1.0687	
ECM	0.6999	0.0299	0.0193	0.0208	0.0215	0.0169	0.0183	
ECMed	0.2644	0.0098	0.0075	0.0080	0.0078	0.0055	0.0066	
Media	1.4631	1.0535	1.0646	1.0563	1.0725	1.0812	1.0736	C_2
Mediana	1.0864	1.0401	1.0493	1.0424	1.0594	1.0671	1.0599	
ECM	6.6088	0.0282	0.0259	0.0242	0.0244	0.0230	0.0215	
ECMed	0.4551	0.0117	0.0098	0.0102	0.0096	0.0092	0.0086	
Media	0.7040	1.0469	1.0611	1.0531	1.0854	1.0950	1.0879	C_3
Mediana	0.3044	1.0345	1.0475	1.0390	1.0682	1.0809	1.0709	
ECM	9195.56	0.0262	0.0189	0.0215	0.0212	0.0174	0.0177	
ECMed	22.583	0.0090	0.0074	0.0079	0.0079	0.0069	0.0069	

Cuadro 7.3: Estimación de la primera coordenada β_1 de β , con ventana $h = 1$

	TLS	RMCD.Med	RM.med	RS.med	RMCD.m	RM.m	RS.m	
Media	1.0265	1.0283	1.0249	1.0258	1.0315	1.0316	1.0316	C_0
Mediana	1.0209	1.0175	1.0155	1.0153	1.0263	1.0280	1.0273	
ECM	0.0072	0.0200	0.0146	0.0158	0.0140	0.0108	0.0113	
ECMed	0.0031	0.0073	0.0058	0.0064	0.0055	0.0043	0.0046	
Media	2.1928	1.0586	1.0640	1.0594	1.0834	1.0869	1.0822	C_1
Mediana	1.9914	1.0421	1.0471	1.0461	1.0655	1.0710	1.0644	
ECM	0.7723	0.0254	0.0190	0.0203	0.0214	0.0165	0.0172	
ECMed	0.2712	0.0088	0.0075	0.0079	0.0079	0.0063	0.0067	
Media	1.2719	1.0525	1.0634	1.0540	1.0700	1.0789	1.0718	C_2
Mediana	1.0280	1.0387	1.0535	1.0406	1.0576	1.0681	1.0599	
ECM	10.922	0.0268	0.0215	0.0214	0.0223	0.0202	0.0191	
ECMed	0.3633	0.0109	0.0088	0.0094	0.0091	0.0080	0.0075	
Media	-23,912	1.0606	1.0747	1.0655	1.0905	1.0991	1.0912	C_3
Mediana	0,2017	1.0467	1.0640	1.0558	1.0799	1.0885	1.0806	
ECM	297465.6	0.0252	0.0195	0.0199	0.0203	0.0176	0.0168	
ECMed	608.80	0.0091	0.0076	0.0076	0.0079	0.0072	0.0067	

Cuadro 7.4: Estimación de la segunda coordenada β_2 de β , con ventana $h = 1$

	TLS	RMCD.Med	RM.med	RS.med	RMCD.m	RM.m	RS.m	
Media	1.0226	1.0258	1.0237	1.0238	1.0309	1.0295	1.0311	C_0
Mediana	1.0213	1.0209	1.0195	1.0170	1.0290	1.0296	1.0294	
ECM	0.0070	0.0202	0.0144	0.0158	0.0156	0.0110	0.0121	
ECMed	0.0029	0.0074	0.0053	0.0059	0.0063	0.0045	0.0047	
Media	2.1799	1.0540	1.0567	1.0529	1.0777	1.0781	1.0742	C_1
Mediana	2.0051	1.0443	1.0513	1.0476	1.0667	1.0710	1.0656	
ECM	0.7569	0.0248	0.0167	0.0185	0.0194	0.0153	0.0159	
ECMed	0.2588	0.0094	0.0072	0.0079	0.0080	0.0063	0.0069	
Media	1.3390	1.0478	1.0609	1.0519	1.0673	1.0743	1.0680	C_2
Mediana	1.0408	1.0404	1.0486	1.0376	1.0594	1.0604	1.0572	
ECM	3.8015	0.0236	0.0201	0.0198	0.0210	0.0197	0.0188	
ECMed	0.4137	0.0093	0.0083	0.0091	0.0081	0.0082	0.0075	
Media	23.912	1.0634	1.0777	1.0701	1.0922	1.1032	1.0967	C_3
Mediana	-0.2076	1.0479	1.0665	1.0566	1.0799	1.0924	1.0848	
ECM	296718.8	0.0254	0.0201	0.0201	0.0204	0.0181	0.0183	
ECMed	607.85	0.0105	0.0086	0.0086	0.0084	0.0072	0.0070	

Cuadro 7.5: Estimación de la primera coordenada β_1 de β , con ventana $h = 1.2$

	TLS	RMCD.Med	RM.med	RS.med	RMCD.m	RM.m	RS.m	
Media	1.0397	1.0359	1.0323	1.0337	1.0491	1.0444	1.0449	C_0
Mediana	1.0317	1.0259	1.0271	1.0266	1.0411	1.0356	1.0391	
ECM	0.0072	0.0208	0.0143	0.0149	0.0153	0.0114	0.0119	
ECMed	0.0031	0.0079	0.0050	0.00566	0.0056	0.0045	0.0045	
Media	2.2060	1.0639	1.0690	1.0642	1.0878	1.0944	1.0904	C_1
Mediana	2.0289	1.0555	1.0589	1.0539	1.0766	1.0847	1.0749	
ECM	0.7147	0.0236	0.0180	0.0190	0.0190	0.0161	0.0176	
ECMed	0.2680	0.0091	0.00695	0.0074	0.0066	0.0060	0.0062	
Media	1.2107	1.0666	1.0810	1.0711	1.0856	1.0961	1.0874	C_2
Mediana	1.0671	1.0491	1.0674	1.0546	1.0718	1.0836	1.0681	
ECM	5.7228	0.0244	0.0218	0.0208	0.0214	0.0186	0.0189	
ECMed	0.2838	0.0104	0.0081	0.0089	0.0081	0.0070	0.0072	
Media	-3.9393	1.0779	1.0874	1.0801	1.1055	1.1157	1.1090	C_3
Mediana	-0.3694	1.0598	1.0723	1.0659	1.0949	1.1006	1.0970	
ECM	8457.5	0.0259	0.0200	0.0204	0.0207	0.0184	0.0188	
ECMed	40.710	0.0090	0.0076	0.0077	0.0078	0.0074	0.0078	

Cuadro 7.6: Estimación de la segunda coordenada β_2 de β , con ventana $h = 1.2$

	TLS	RMCD.Med	RM.med	RS.med	RMCD.m	RM.m	RS.m	
Media	1.0384	1.0385	1.0341	1.0354	1.0449	1.0431	1.0425	C_0
Mediana	1.0351	1.0249	1.0265	1.0258	1.0337	1.0366	1.0366	
ECM	0.0078	0.0207	0.0149	0.0158	0.0162	0.0119	0.0126	
ECMed	0.0036	0.0074	0.0056	0.0057	0.0063	0.0047	0.0049	
Media	2.1923	1.0633	1.0667	1.0624	1.0876	1.0941	1.0896	C_1
Mediana	2.0246	1.0508	1.0563	1.0523	1.0771	1.0786	1.0751	
ECM	0.6608	0.0243	0.0180	0.0191	0.0193	0.0167	0.0177	
ECMed	0.2601	0.0088	0.0069	0.0081	0.0080	0.0076	0.0073	
Media	1.2538	1.0693	1.0813	1.0733	1.0842	1.0945	1.0883	C_2
Mediana	1.0818	1.0601	1.0689	1.0680	1.0737	1.0843	1.0788	
ECM	6.5031	0.0270	0.0213	0.0216	0.0212	0.0176	0.0184	
ECMed	0.3395	0.0101	0.00799	0.0080	0.0082	0.0074	0.0073	
Media	3.9360	1.0795	1.0894	1.0822	1.1060	1.1175	1.1093	C_3
Mediana	0.4205	1.0648	1.0747	1.0680	1.0909	1.1034	1.0958	
ECM	8440.06	0.0268	0.0223	0.0223	0.0235	0.0203	0.0201	
ECMed	40.631	0.0106	0.0085	0.0090	0.0087	0.0077	0.0077	

7.3. Gráficos

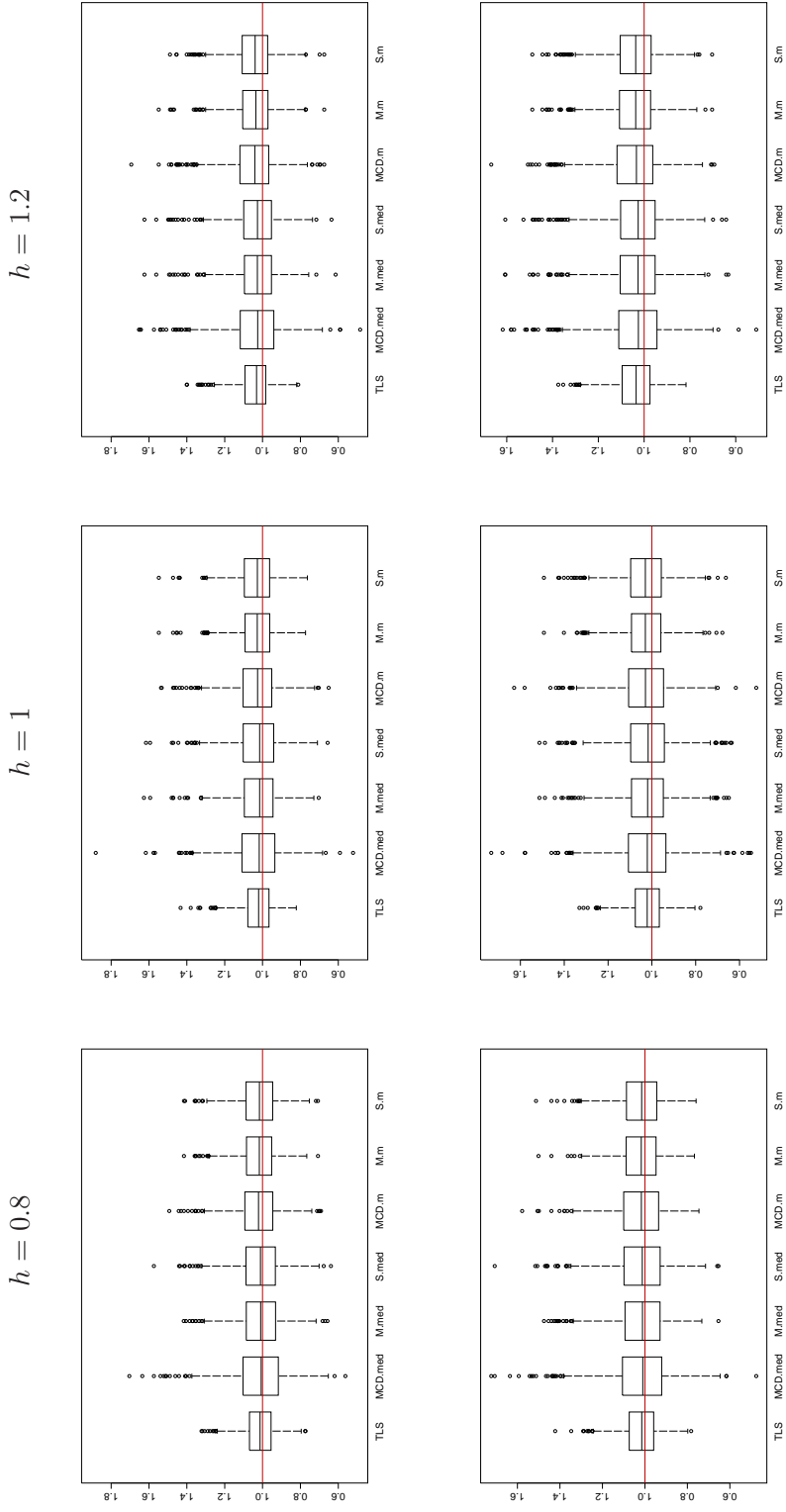


Figura 7.1: Boxplots de los estimadores de β_1 en la primera línea y de β_2 en la segunda línea para las ventanas $h = 0.8, 1$ y 1.2 bajo C_0

Figura 7.2: Densidades de los errores cuadráticos medios de \hat{g} para el método clásico con ventana $h = 0.8$. En negro se grafica la densidad para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

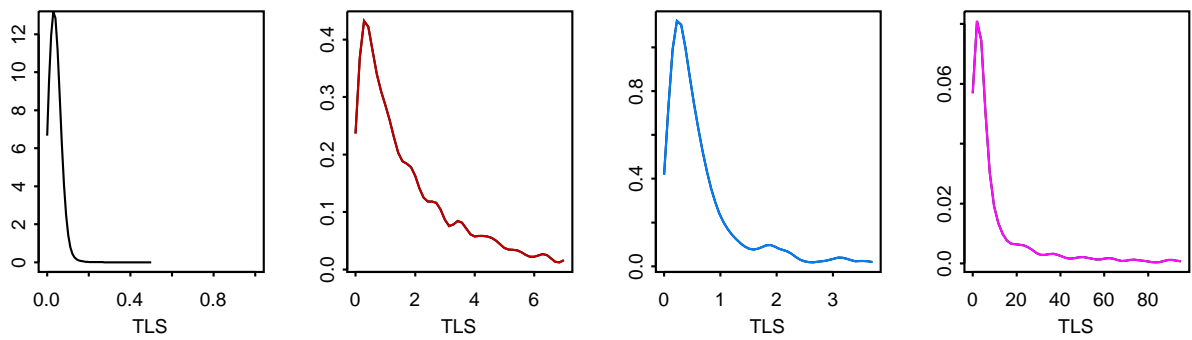


Figura 7.3: Densidades de los errores cuadráticos medios de \hat{g} para los métodos robustos con ventana $h = 0.8$. En negro se grafican las densidades para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

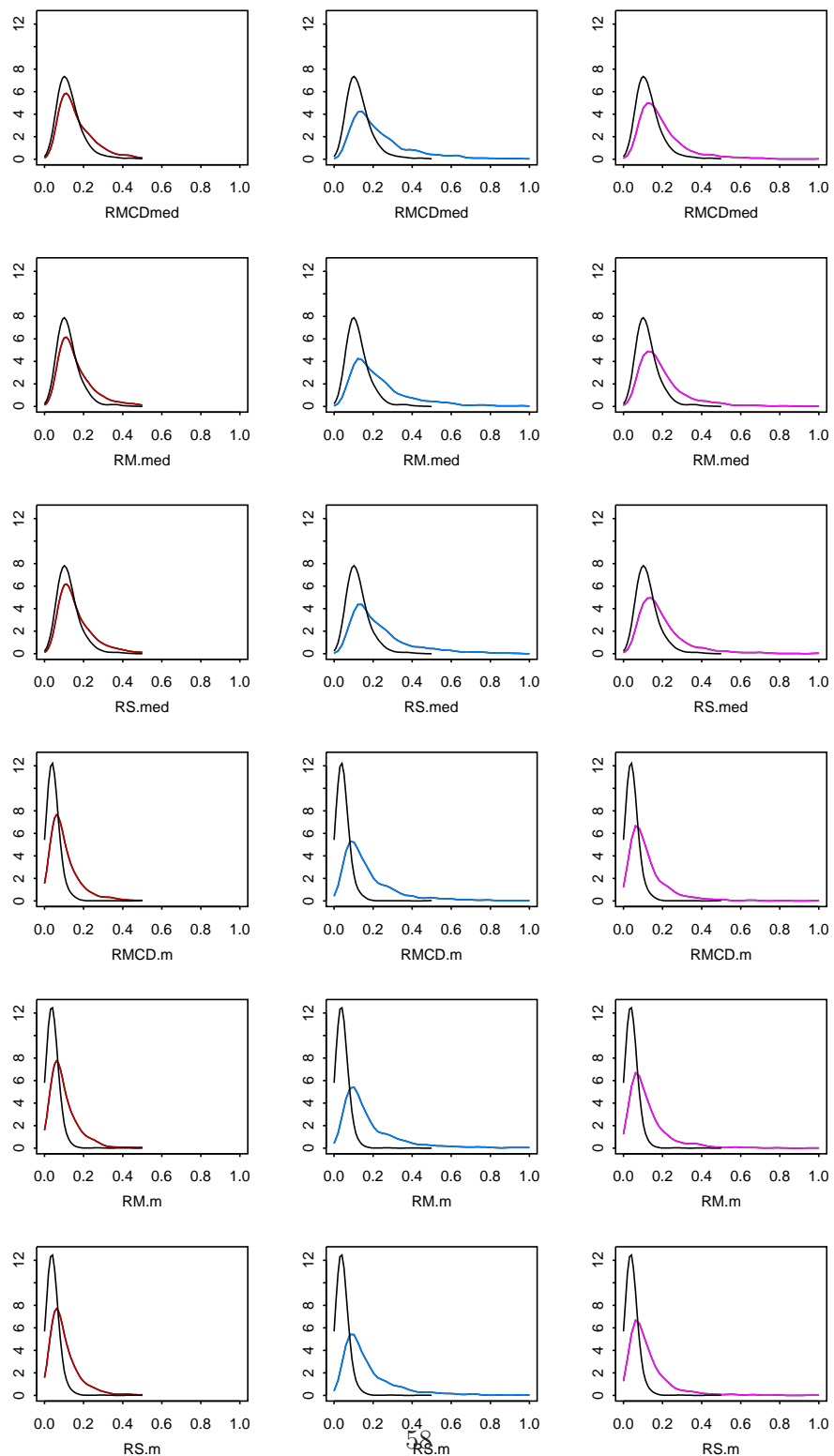


Figura 7.4: Densidades de los errores cuadráticos medios de \hat{g} para el método clásico con ventana $h = 1$. En negro se grafica la densidad para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

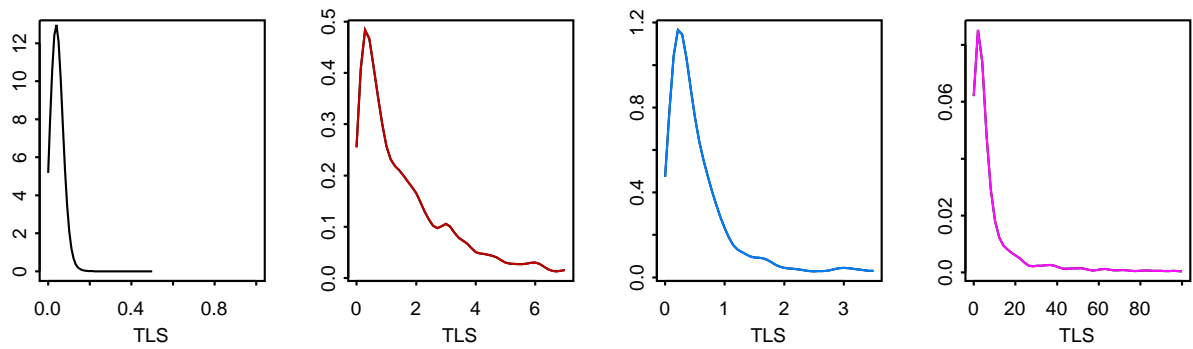


Figura 7.5: Densidades de los errores cuadráticos medios de \hat{g} para los métodos robustos con ventana $h = 1$. En negro se grafican las densidades para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

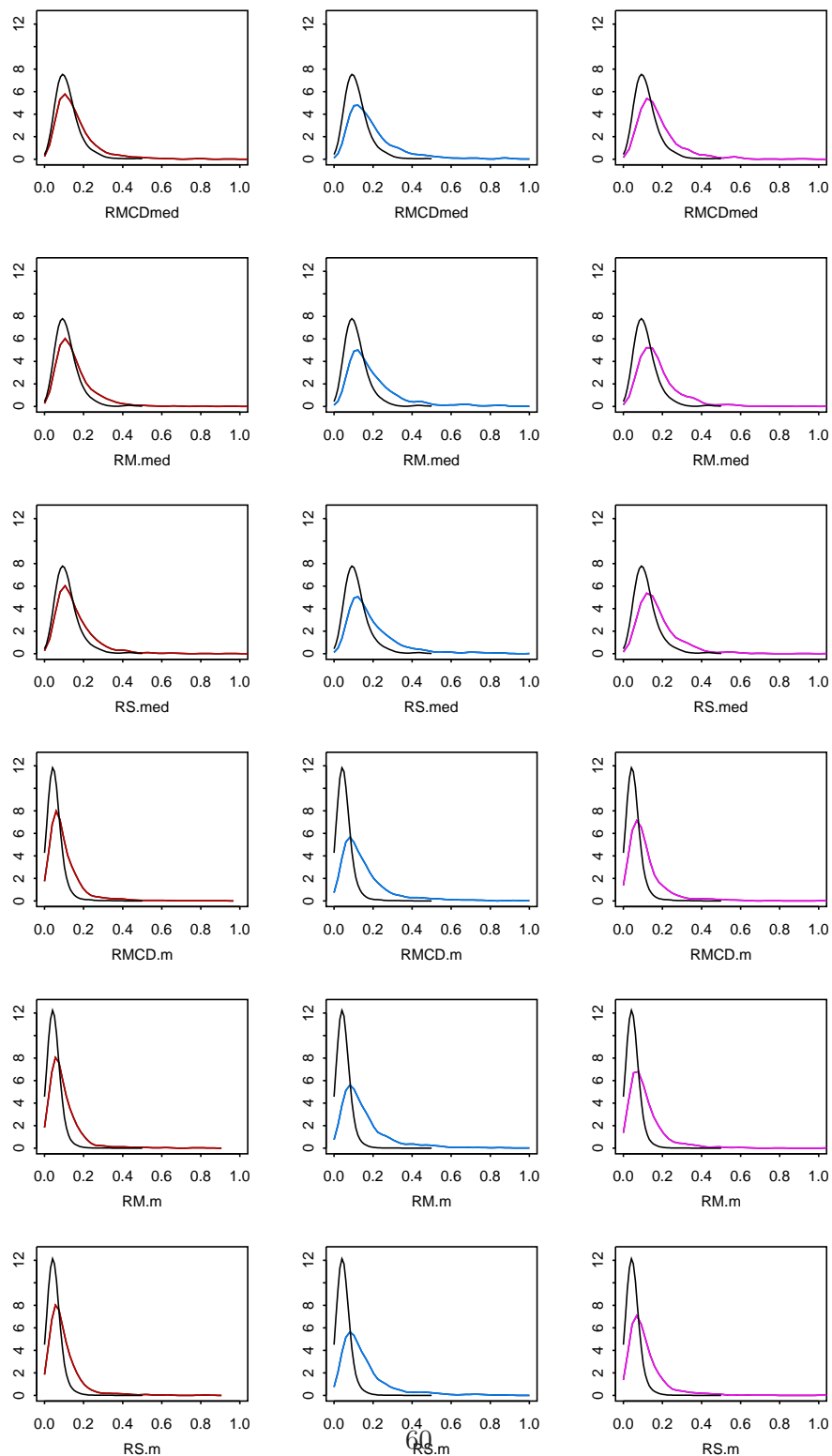


Figura 7.6: Densidades de los errores cuadráticos medios de \hat{g} para el método clásico con ventana $h = 1.2$. En negro se grafica la densidad para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

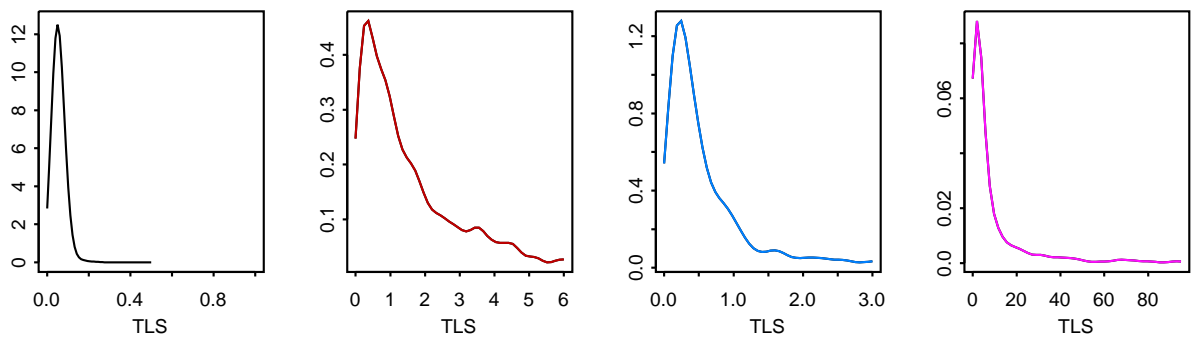


Figura 7.7: Densidades de los errores cuadráticos medios de \hat{g} para los métodos robustos con ventana $h = 1.2$. En negro se grafican las densidades para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

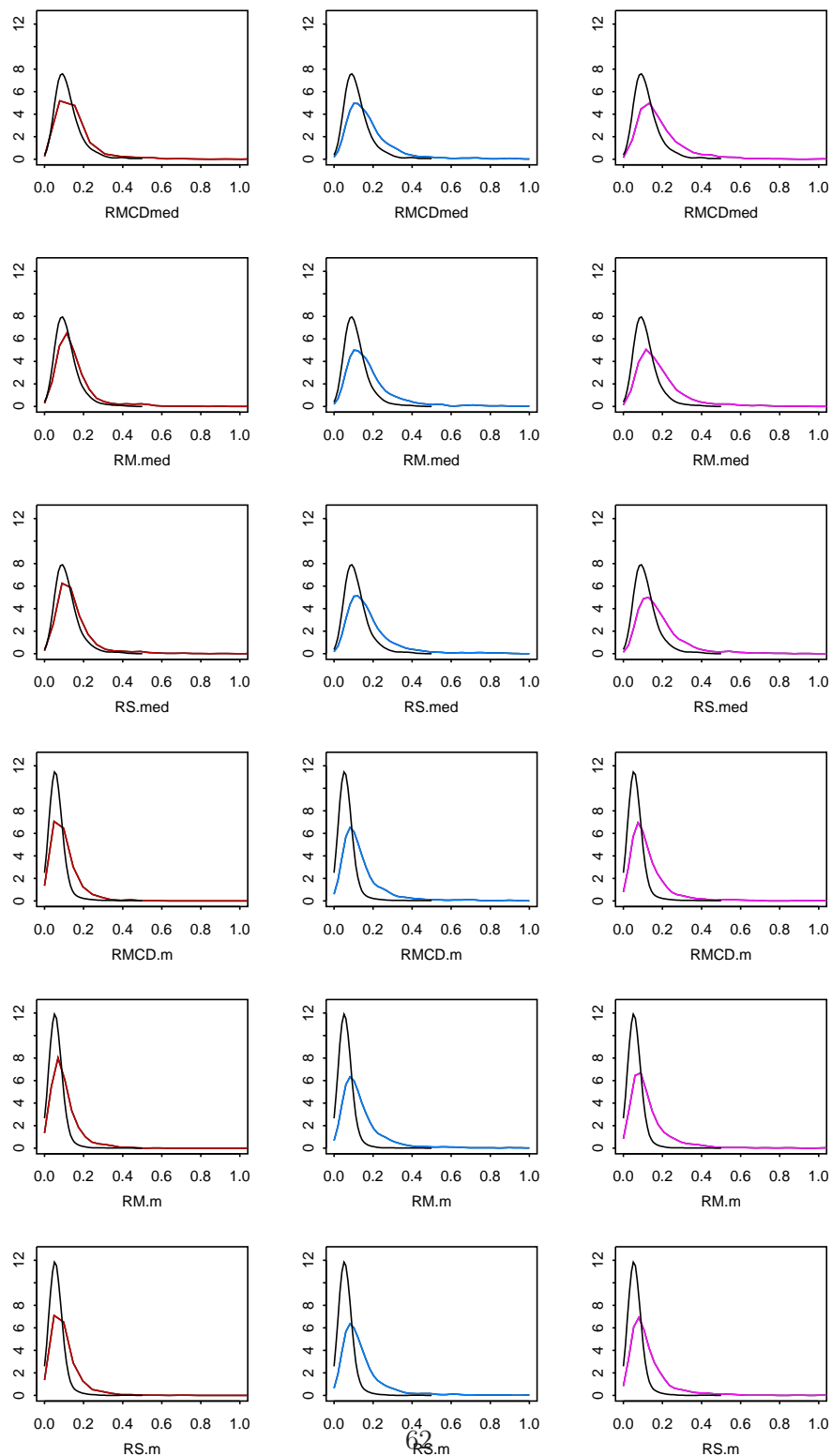


Figura 7.8: Densidades de los errores cuadráticos medianos de \hat{g} para el método clásico con ventana $h = 0.8$. En negro se grafica la densidad para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

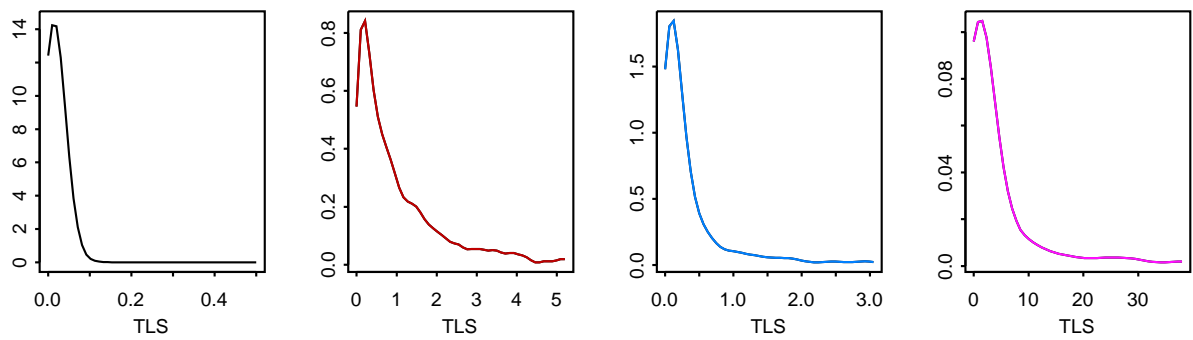


Figura 7.9: Densidades de los errores cuadráticos medianos de \hat{g} para los métodos robustos con ventana $h = 0.8$. En negro se grafican las densidades para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

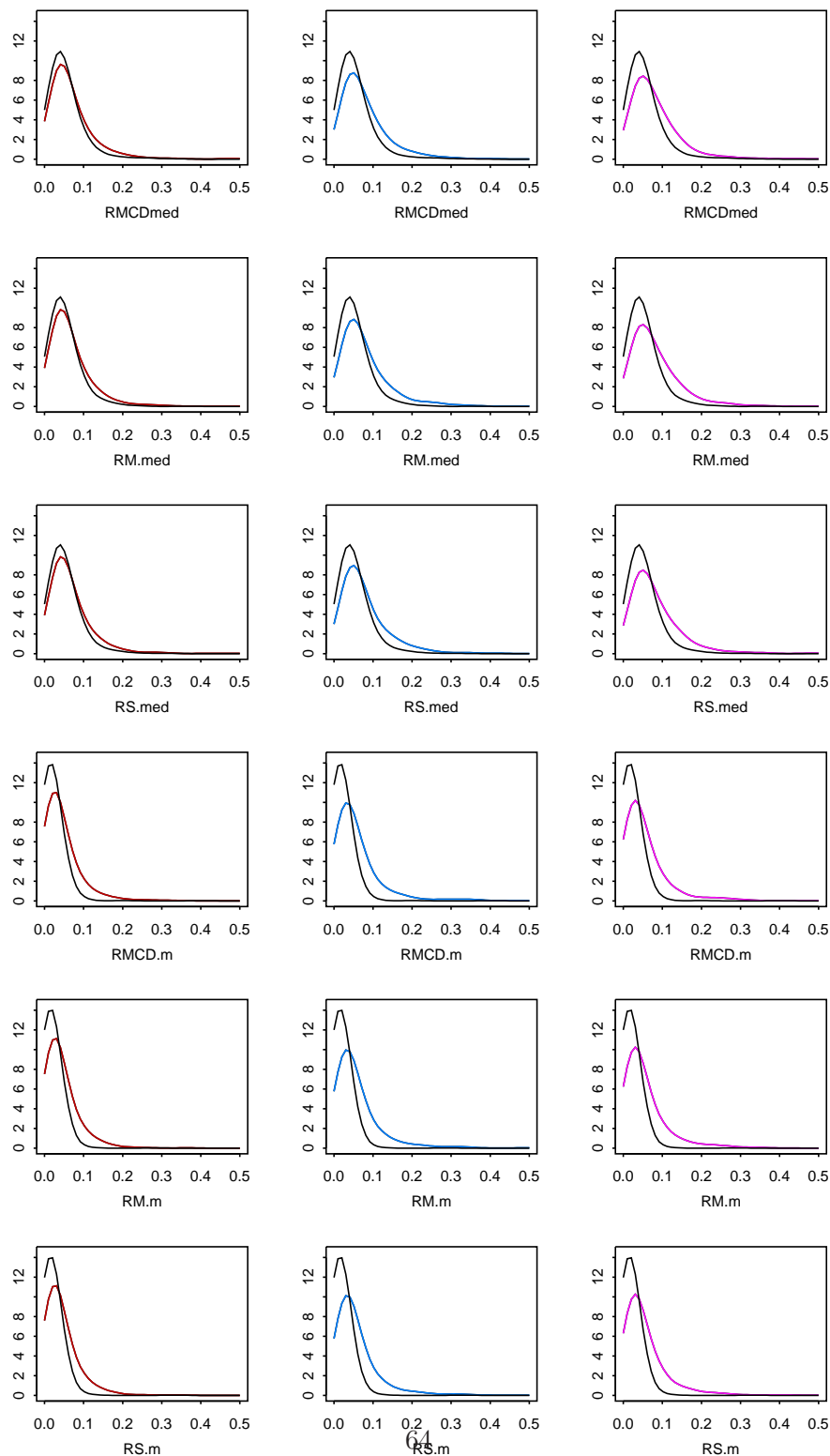


Figura 7.10: Densidades de los errores cuadráticos medianos de \hat{g} para el método clásico con ventana $h = 1$. En negro se grafica la densidad para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

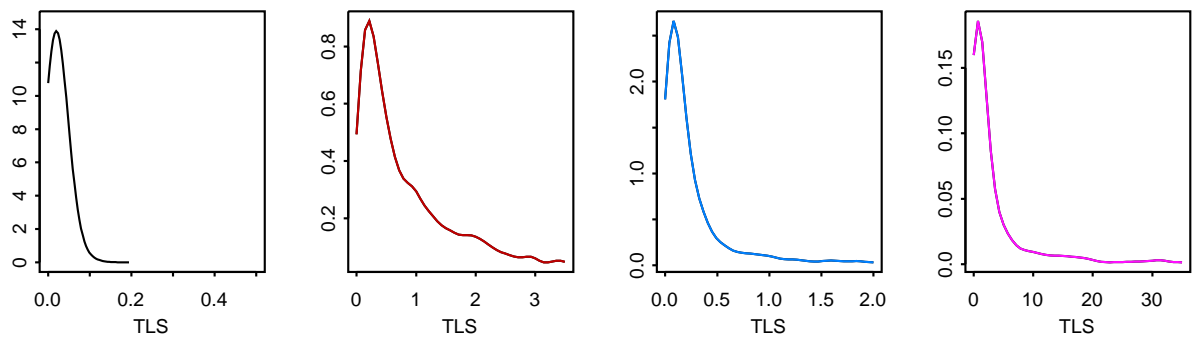


Figura 7.11: Densidades de los errores cuadráticos medianos de \hat{g} para los métodos robustos con ventana $h = 1$. En negro se grafican las densidades para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

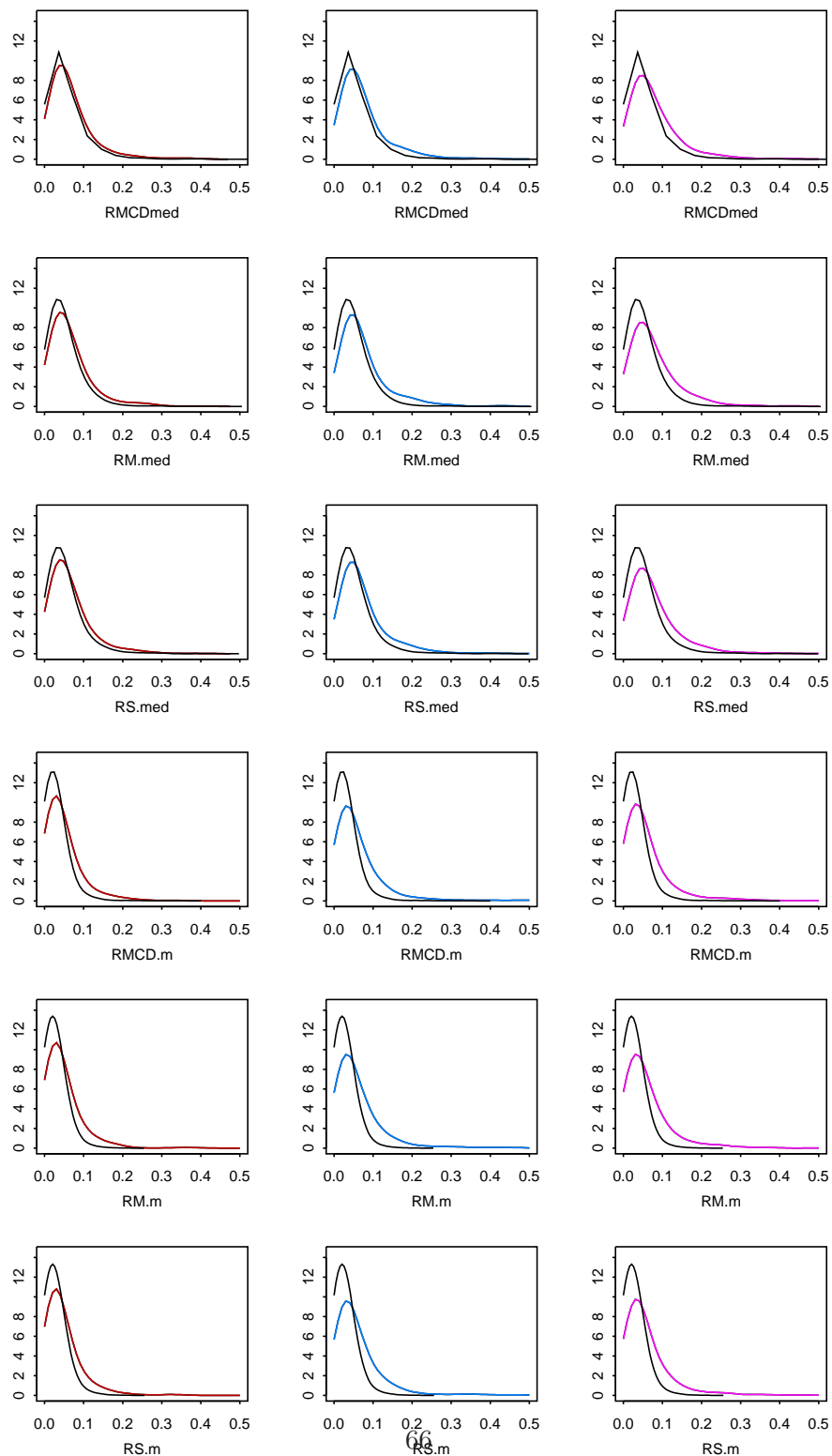


Figura 7.12: Densidades de los errores cuadráticos medianos de \hat{g} para el método clásico con ventana $h = 1.2$. En negro se grafica la densidad para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .

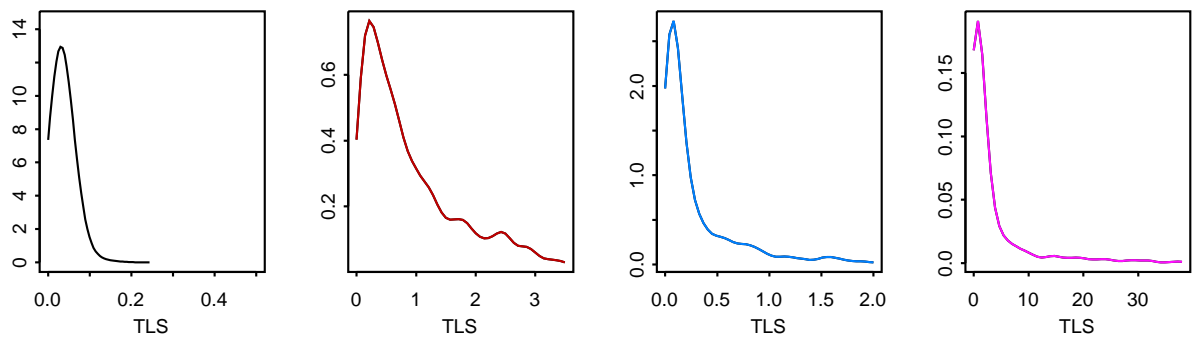
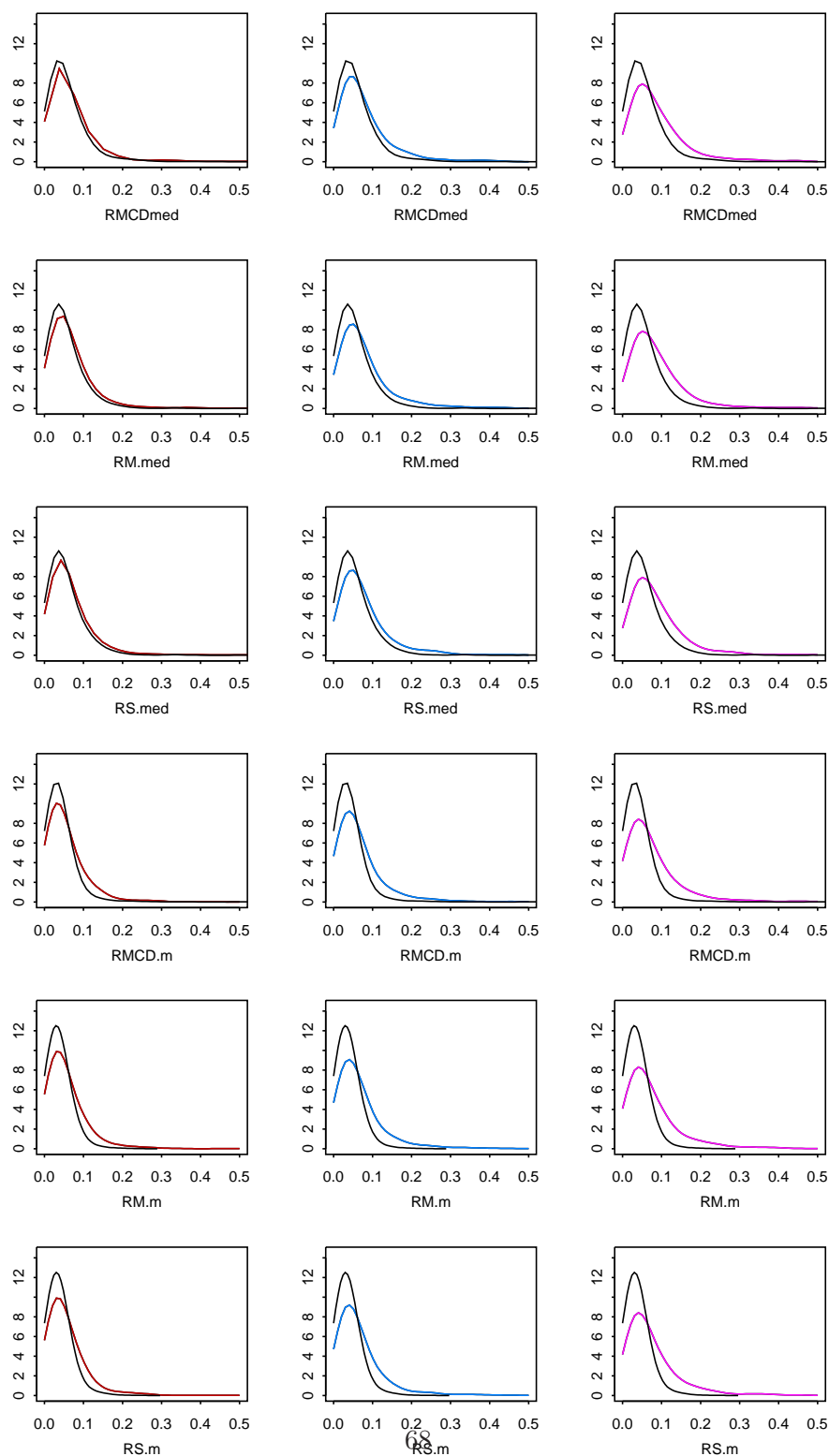


Figura 7.13: Densidades de los errores cuadráticos medianos de \hat{g} para los métodos robustos con ventana $h = 1.2$. En negro se grafican las densidades para C_0 , en marrón para C_1 , en celeste para C_2 y en rosa para C_3 .



Bibliografía

- [1] Ansley, C. y Wecker, W. (1983). Extension and examples of the signal extraction approach to regression. In *Applied Time Series Analysis of Economic Data*, 181-192.
- [2] Bianco, A. y Boente, G. (2004). Robust estimators in semiparametric partly linear regressions models. *Journal of Statistical Planning and Inference* **122**, 229–252.
- [3] Boente, G. y Fraiman, R. (1991). Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. *International Statistical Review*, **59**, 355–372.
- [4] Brown, M. (1982). Robust line estimation with errors in both variables. *American Statistical Association*, **77**, 71–79.
- [5] Carroll, R. J. y Gallo, P. P. (1982). Some aspects of robustness in functional errors-in-variables regression models. *Communications in Statistics, Series A*, **11**, 2573–2585.
- [6] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Annals of Statistics*, **16**, 136-146.
- [7] Chen, H. y Chen, K. (1991). Selection of the splined variables and convergence rates in a partial spline model. *Canadian Journal of Statistics*, **19**, 323-339.
- [8] Chen, H. y Shiau, J. (1991). A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning and Inference*, **25**, 187-201.
- [9] Chen, H. y Shiau, J. (1994). Data-driven efficient estimates for partially linear models. *Annals of Statistics*, **22**, 211-237.
- [10] Chen, H. y Van Ness, J. (1992). Generalized M-estimators for errors-in-variables regression. *Annals of Statistics*, **20**, 385–397.
- [11] Cui, H. y Zhu, L. (2003). A semi-parametric regression model with errors in the variables. *Scandinavian Journal of Statistics*, **30**, 429–442.
- [12] Davies, P. (1987). Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, **15**, 1269–1292.
- [13] Davies, P. (1992). The asymptotics of Rousseeuw’s minimum volume ellipsoid estimator. *The Annals of Statistics*, **20**, 1828–1843.
- [14] Denby, L. (1986). Smooth regression functions. *Statistical Research Report*, **26**, AT&T Bell Laboratories, Murray Hill.
- [15] Engle, R., Granger, C., Rice, J. y Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, **81**, 310-320.
- [16] Fekri, M. y Ruiz-Gazen, A. (2004). Robust weighted orthogonal regression in the errors-in-variables model. *Journal of Multivariate Analysis*, **88**, 89–108
- [17] Gao, J. (1992). *A large sample Theory in Semiparametric Regression Models*. Phd. Thesis, University of Science and Technology of China, Hefei, China.
- [18] Gao, J. y Liang, H. (1995). Asymptotic normality of pseudo-LS estimator for partly linear autoregression models. *Statistics and Probability Letters*, **23**, 27-34.
- [19] Gao, J. y Zhao, L. (1993). Adaptive estimation in partly linear regression models. *Science in China, Ser. A*, **1**, 14-27.

- [20] Green, P., Jennison, C. y Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Ser. B*, **47**, 299-315.
- [21] Hampel, F. (1974). The influence curve and its role in robust estimation. *The Annals of Statistics*, **69**, 383–393.
- [22] Härdle, W. (1990). *Applied Nonparametric Regression*. Econometric Society Monographs No.19, Cambridge University Press.
- [23] Härdle, W. (1991). *Smoothing Techniques, With Implementations in S*. Springer, New York.
- [24] Härdle, W., Müller, M., Sperlich, S. y Werwatz A. (2004). *Nonparametric and Semiparametric Models*. Springer - Verlag.
- [25] Hastie, T. y Tibshirani, R. (1990). *Generalized Additive Models*. Vol 43 de Monographs on Statistics y Applied Probability. Chapman y Hall.
- [26] He, X. y Shi, P. (1996). Bivariate tensor-product B-spline in a partly linear model. *Journal of Multivariate Analysis*, **58**, 162-181.
- [27] Heckman, N. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society, Ser. B*, **48**, 244-248.
- [28] Huber, P. (1981). *Robust Statistics*. New York: John Wiley and Sons, Inc.
- [29] Ketellaper, R. y Ronner, A. (1984). Are robust estimation methods useful in the structural errors-in-variables model. *Metrika*, **31**, 33–41.
- [30] Liang, H., Härdle, W. y Carroll, R. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics*, **27**, 1519–1535.
- [31] Lopuhaä, H. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics*, **17**, 1662–1683.
- [32] Lopuhaä, H. (1997). Asymptotic expansion of S-estimators and M-estimators of multivariate location and covariance. *Statistica Neerlandica*, **51**, 220–237.
- [33] Lopuhaä, H. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *The Annals of Statistics*, **27** (5), 1638–1665.
- [34] Lopuhaä, H. y Rousseeuw, P. (1991). Breakdown properties of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, **19**, 229–248.
- [35] Maronna, R. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, **4**, 51–67.
- [36] Maronna, R., Martin, R. y Yohai V. (2006). *Robust Statistics – Theory and Methods*. John Wiley and Sons Ltd.
- [37] Marron, J. y Härdle, W. (1986). Random approximations to some measures of accuracy in nonparametric curve estimation. *Journal of Multivariate Analysis*, **20**, 91–113.
- [38] Nadaraya, E. (1964). On estimating regression. *Theory of Probability and its Applications*, **10**, 186–190.
- [39] Rice, J. (1986). Convergence rates for partially splined models. *Statistics and Probability Letters*, **4**, 203-208.
- [40] Robinson, P. (1988). Root-n-consistent Semiparametric regression. *Econometrica*, **56**, 931-954.
- [41] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 832-837.
- [42] Severini, T. y Staniswalis, J. (1994). Quasi-Likelihood Estimation in Semiparametric Models. *Journal of the American Statistical Association*, **89**, 501-511
- [43] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Ser. B*, **50**, 413-436.
- [44] Stone, C. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, **12**, 1285–1297.
- [45] Totsuoka, K. y Tyler, D. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *The Annals of Statistics*, **28**, 1219–1243.
- [46] Tyler, D. (1990). Breakdown properties of the M-estimators of multivariate scatter. Technical Report, Department of Statistics, Rutgers University.

- [47] Yee, T. y Wild, C. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society, Ser. B*, **58**, 481-493.
- [48] Zamar, R. (1989). Robust estimation in the errors-in-variables model. *Biometrika*, **76**, 149-160.
- [49] Zamar, R. H. (1992). Bias robust estimation in orthogonal regression. *The Annals of Statistics*, **20**, 1875-1888.