

T 7/118

TESIS DOCTORAL

Aportaciones en la estimación no paramétrica de modelos de regresión aditivos

Departamento de Estadística e I.O.

UNIVERSIDAD DE GRANADA

UNIVERSIDAD DE GRANADA Facultad de Ciencias Fecha <u>05.02.02</u> ENTRADA NUM. <u>272</u>



Rocío Raya Miranda
Granada, 2002

BIBLIOTECA UNIVERSITARIA GRANADA N.º Documento <u>613670773</u> N.º Copia <u>115756713</u>

TESIS DOCTORAL

Rocío Raya Miranda

**APORTACIONES EN LA ESTIMACIÓN
NO PARAMÉTRICA DE MODELOS DE
REGRESIÓN ADITIVOS**

**Memoria presentada por Rocío Raya
Miranda para optar al grado de Doctor
por la Universidad de Granada**



Rocío Raya Miranda

**Vº Bº
Director de tesis**



Fdo. D. Andrés González Carmona

**Vº Bº
Director de tesis**



Fdo. Dª. María Dolores Martínez Miranda

**Departamento de Estadística e Investigación Operativa
Universidad de Granada**

Agradecimientos

En primer lugar, quiero expresar mi agradecimiento a los directores de esta memoria, Dr. D. Andrés González Carmona y Dra. D^a. María Dolores Martínez Miranda, por su interés y colaboración, que desde el principio han mostrado, en la realización de este trabajo.

No me puedo olvidar de mis compañeros y amigos, que me han ofrecido su apoyo y ánimos en incontables ocasiones. También al director del Departamento de Estadística e I.O. de la Universidad de Granada, Dr. D. Ramón Gutiérrez Jaímez, por haberme dado la posibilidad de realizar este trabajo.

Y por último, dar las gracias a mi familia, a Alfonso y a mis amigos, por darme fuerzas para seguir adelante cada día.

Índice general

Prólogo	1
1. Introducción	5
1.1. Regresión no paramétrica univariante y multivariante	5
1.2. Estimador polinomial local de grado p	7
1.2.1. Modelo univariante	7
1.2.2. Modelo multivariante	8
1.2.3. Estimador interno	9
1.3. Propiedades de los estimadores	10
1.3.1. Modelo univariante	10
1.3.2. Modelo multivariante	12
1.4. El problema de la dimensionalidad	14
2. Estudio de los modelos aditivos	17
2.1. Definición del modelo	17
2.2. Métodos de estimación	18
2.3. Algoritmo <i>backfitting</i>	19
2.3.1. Definición	19
2.3.2. Convergencia para suavizadores lineales generales	20
2.3.3. Suavizadores polinomiales locales	22
2.4. Método de integración marginal	27
2.4.1. Caso bivariente	27
2.4.2. Extensión al caso de D -variables	30
2.5. Estimador eficiente de Linton	35
2.5.1. Definición del estimador	35
2.5.2. Propiedades del estimador	38
2.6. Estimador de proyección- <i>backfitting</i>	39
2.6.1. Definición del estimador	39
2.6.2. Existencia de solución	43

3. Selección del ancho de banda	45
3.1. Introducción	45
3.2. Métodos para la selección del ancho de banda	46
3.2.1. Selector <i>plug-in</i>	46
3.2.2. Validación cruzada por mínimos cuadrados	48
3.3. Selectores del ancho de banda en los modelos aditivos	48
3.3.1. Selector <i>plug-in</i>	49
3.3.2. Validación cruzada generalizada	54
4. <i>SiZer Map</i> para modelos aditivos	57
4.1. Introducción	57
4.2. Definición de <i>SiZer Map</i>	58
4.2.1. Desarrollo de <i>SiZer Map</i>	59
4.2.2. Aproximación <i>binned</i>	59
4.3. Extensión a un modelo aditivo	61
4.3.1. Desarrollo de <i>SiZer Map</i> ($D = 2$)	61
4.3.2. Aproximación <i>binned</i>	63
4.4. Estudio de simulación	64
4.5. Conclusiones	69
5. Selector <i>bootstrap</i> del ancho de banda	73
5.1. Introducción	73
5.1.1. <i>Bootstrap</i> en regresión no paramétrica	74
5.2. Estimación <i>bootstrap</i> del error cuadrático medio en la estimación de un modelo aditivo	75
5.2.1. Elección <i>bootstrap</i> del ancho de banda	87
5.2.2. Expresión exacta del estimador <i>bootstrap</i>	88
5.3. Elección de los parámetros piloto	89
5.4. Estudio de simulación	99
5.4.1. Estudio del estimador <i>bootstrap</i> del error cuadrático medio . .	101
5.4.2. Modelos estimados con anchos de banda <i>bootstrap</i>	103
5.5. Conclusiones	108
A. Computación estadística	109
A.1. Funciones para la simulación de las observaciones	109
A.2. Funciones para la creación de gráficos <i>SiZer Map</i> en un modelo aditivo bivalente	111
A.3. Funciones para evaluar el selector <i>bootstrap</i>	129
B. Expresiones adicionales	145
C. Notación	149

ÍNDICE GENERAL

III

Bibliografía

151

Prólogo

La teoría y métodos de regresión no paramétrica se han desarrollado principalmente en las últimas décadas. Los modelos de tipo no paramétrico no imponen fuertes restricciones sobre la estructura de los datos, permitiendo una mayor flexibilidad y adaptabilidad a lo observado, a diferencia de aquellos clásicos modelos lineales o polinómicos los cuales, por otro lado, posibilitaban un estudio teórico más sencillo y un coste computacional mucho menor. Pero aún en el caso de optar por un modelo paramétrico, estas técnicas de regresión no paramétrica pueden servir de guía en la elección del mismo, presentándose como una potente herramienta de tipo exploratorio.

En esta memoria consideramos de partida el planteamiento de un problema de regresión multivariante, y estudiamos las generalizaciones a tal situación de suavizadores univariantes de interés. No obstante y aunque en general es posible la generalización de la mayoría de técnicas de suavizamiento univariantes, presenta mayores inconvenientes tanto desde el punto de vista teórico, como en la posterior aplicación práctica, dado que la mayor parte del *software* disponible que implementa tales técnicas admite hasta un máximo de tres o cuatro variables explicativas. Estos problemas son los que han motivado la búsqueda de otras técnicas de estimación, que siguiendo la línea no paramétrica permitan una reducción de la dimensionalidad en el tratamiento.

En los últimos años se han propuesto diferentes aproximaciones; entre ellas destacamos por su importancia las siguientes: *projection pursuit* (Friedman y Stuetzle, 1981); ACE¹ (Breiman y Friedman, 1985); los modelos aditivos y aditivos generalizados (Hastie y Tibshirani, 1986); y MARS² (Friedman, 1991). Entre ellas, centramos nuestro estudio en los modelos aditivos para estimar la función de regresión multivariante.

Propuestos originalmente por Friedman y Stuetzle (1981) los modelos aditivos son una técnica de regresión no paramétrica multivariante muy utilizada, en gran parte debido a la extensa discusión de la técnica que realizan Hastie y Tibshirani.

¹Método que estima un modelo aditivo con transformaciones en la variable de respuesta y en cada una de las variables explicativas. El nombre corresponde a las iniciales de su nombre en inglés *Alternating Conditional Expectation algorithm*.

²Iniciales de *Multivariate Adaptive Regression Splines*.

ni (1990) y a las posibilidades de ajuste mediante rutinas de S-Plus descritas por Chambers y Hastie (1992). Estos modelos resuelven los problemas que presentaban las versiones multivariantes de los suavizadores univariantes, al considerar la función de regresión como la suma de funciones suaves de cada variable explicativa. De este modo, se puede estimar cada una de las componentes por separado utilizando suavizadores univariantes, y la estimación global se define por la suma de todas las estimaciones. Frente a otras técnicas de regresión no paramétrica multivariante, los modelos aditivos pueden interpretar fácilmente los efectos individuales de las variables independientes sobre la variable dependiente, sin tener en cuenta que el número de variables sea elevado.

Los principales métodos de estimación de estos modelos son el algoritmo *back-fitting*, de Buja, Hastie y Tibshirani (1989) y el método de integración marginal propuesto por Linton y Nielsen (1995). A partir de estos métodos se han desarrollado varios procedimientos en un intento siempre de obtener mejores resultados en las estimaciones.

La generalización de las técnicas de suavizamiento univariantes al problema de regresión multivariante lleva asociado el hecho de tener que seleccionar una matriz de anchos de banda; problema que ha sido mucho menos considerado en la literatura que el problema en el caso univariante, en el que interviene un único ancho de banda. En la teoría de los modelos aditivos, en lugar de tener matrices de anchos de banda, tendremos un vector, esto es, un nivel de suavizado, en principio distinto, para cada variable explicativa. El problema central de la tesis lo va a constituir el problema de selección de los anchos de banda o parámetros de suavizamiento que determinen el buen comportamiento del modelo aditivo estimado. Nuestras aportaciones por tanto están asociadas a tal problema, y las desarrollamos en el siguiente contexto: Se considera un modelo aditivo bivalente estimado utilizando el algoritmo *backfitting* aplicado con suavizadores lineales locales en las componentes, según las expresiones de las estimaciones dadas por Opsomer y Ruppert (1997).

En la selección del parámetro ancho de banda diversos autores han utilizado la metodología *bootstrap*, sobre todo en la estimación de densidades, con resultados bastante satisfactorios. Recientemente, se ha propuesto un selector local *bootstrap* del estimador polinomial local en el contexto de regresión no paramétrica univariante.

Puesto que la estimación de los modelos aditivos conduce a estimaciones de tipo univariante, proponemos la metodología *bootstrap* como método de selección de los anchos de banda del modelo aditivo. Entonces, se define un nuevo selector de tipo local-variable, que se prueba que es válido tanto desde el punto de vista teórico como del práctico.

La herramienta gráfica *SiZer Map* ha sido desarrollada para el estudio de las características de funciones, tales como funciones de densidades o de regresión, y ha demostrado su gran potencial como herramienta para el análisis de datos, mostrando características de estos que de otra forma eran de difícil detección. No obstante,

además de servir como herramienta de tipo exploratorio en el análisis de los datos, es posible aplicar el potencial de *SiZer Map* al problema de decidir sobre el nivel adecuado de suavizamiento, valorando a su vez el comportamiento de selectores que habitualmente se utilizan en este contexto.

En este trabajo se propone el uso de *SiZer Map* con estas dos finalidades, como herramienta de análisis individual de las componentes aditivas y como ayuda al problema de decisión del nivel de suavizamiento adecuado, tanto en cada variable como para cada punto de estimación considerado.

La memoria se estructura en cinco capítulos y tres apéndices. El primer capítulo se dedica a la presentación de los problemas de regresión univariante y multivariante, en su forma más general, describiendo los principales métodos de estimación en ambos casos y viendo las dificultades que surgen al generalizar los métodos univariantes al caso multivariante. El segundo capítulo define y estudia las propiedades fundamentales de los métodos de estimación de los modelos aditivos, más importantes, destacando el algoritmo *backfitting*, puesto que será el método que utilizaremos en el resto de la memoria. El tercer capítulo describe algunos de los selectores del ancho de banda propuesto para el problema de regresión multivariante y recoge los métodos, existentes hasta el momento, para el caso de los modelos aditivos. En los capítulos cuarto y quinto se encuentran las aportaciones fundamentales de esta memoria, en el cuarto se adapta la definición de *SiZer Map* a un modelo aditivo bivariante y se presenta un ejemplo de aplicación a un modelo aditivo con diferentes tamaños muestrales y niveles de correlación entre las variables independientes; en el quinto se describe un nuevo selector para los anchos de banda de un modelo aditivo bivariante estimado utilizando el algoritmo *backfitting* con suavizadores lineales locales, empleando una estimación *bootstrap* del error cuadrático medio local. El selector *bootstrap* propuesto se estudia desde un punto de vista teórico, probando su validez (consistencia de la estimación *bootstrap*) y desde un punto de vista práctico, realizando un estudio de simulación que permite ilustrar su comportamiento finito-muestral. Finalmente se incluyen tres apéndices, el primero de ellos contiene la programación realizada en *Matlab* 5.10.421, incluyendo las funciones que han sido definidas para el desarrollo de *SiZer Map* y el cálculo del selector propuesto, así como para la obtención de los resultados del estudio de simulación; el segundo contiene algunas expresiones omitidas para conseguir mayor claridad en la expresión en las demostraciones del capítulo quinto; y el último recoge la notación empleada en esta memoria.

Las aportaciones fundamentales realizadas en este trabajo son las siguientes:

- Se realiza una descripción de los procedimientos propuestos hasta el momento para la selección de los parámetros ancho de banda que intervienen en un modelo aditivo.
- Se propone un nuevo selector del parámetro en el contexto asumido, exten-

diendo las recientes ideas propuestas en un problema de regresión univariante estimado con un suavizador lineal local por Martínez-Miranda (2000).

- Se realiza un estudio de sus propiedades teóricas bajo una perspectiva asintótica y se comprueba su validez desde el punto de vista práctico a través de un estudio de simulación.
- Se extiende la definición del gráfico *SiZer Map* a un modelo aditivo bivariante, permitiendo el estudio del comportamiento de las componentes y la elección del nivel de suavizamiento adecuado.
- Se muestra a través de un ejemplo la aplicación de *SiZer Map* a un modelo aditivo bivariante.
- Se realizan implementaciones en *Matlab* 5.10.241 para las diferentes aportaciones.

Las aportaciones al estudio gráfico de los modelos aditivos recogidas en esta memoria han sido presentadas y aceptadas en *XXVI Congreso Nacional de Estadística e I.O.*, SEIO 2001, y en *XV Conference on Computational Statistics*, COMPSTAT 2002.

Capítulo 1

Introducción

Para el estudio de una variable de respuesta, Y , a veces es suficiente la información de una única variable explicativa, X . Sin embargo, en la práctica, lo más habitual es que el comportamiento de Y no pueda explicarse con una sola variable, y sea necesario considerar un conjunto de ellas. Por ello, dentro del contexto de regresión no paramétrica vamos a comenzar con la distinción entre modelos univariantes y multivariantes.

1.1. Regresión no paramétrica univariante y multivariante

Consideremos un conjunto de variables aleatorias, independientes e idénticamente distribuidas, $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$, procedentes de una variable aleatoria bidimensional, (X, Y) . El modelo de **regresión no paramétrica univariante**¹ se define como

$$Y_i = m(X_i) + v^{1/2}(X_i)\varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

donde m es la función de regresión desconocida, es decir, $m(x) = E[Y | X = x]$ y $v(x) = \text{Var}(Y | X = x)$. Además, ε_i , son los errores de observación, que se suponen variables aleatorias independientes con media 0, varianza 1 y son independientes de X . A la función de densidad de X la denotamos por f . La única hipótesis que se hace es que las funciones, m y v , sean suaves, donde esta suavidad se entiende en términos de derivabilidad.

Entre los estimadores de regresión no paramétrica univariantes caben destacar los de tipo núcleo, como son el de Nadaraya-Watson [Nadaraya (1964), Watson (1964)] el de Priestley-Chao [Priestley y Chao (1972)] el de Clark (1977) y el de Gasser-Müller [Gasser y Müller (1979)] que han gozado de gran interés por parte

¹Hemos considerado el caso general heterocedástico.

de muchos autores, debido fundamentalmente a su formulación sencilla y facilidad de tratamiento teórico. Desde comienzos de los noventa, hemos de resaltar la notable evolución de los denominados estimadores de regresión polinomial local², que se presentan como extensiones de los sencillos estimadores de tipo núcleo y que los mejoran al responder y solucionar algunos de los problemas que estos últimos planteaban en las estimaciones en puntos fronterizos. Cleveland (1979) introduce la regresión localmente ponderada, cuya definición está muy ligada a la regresión polinomial local. Otros métodos a destacar son el suavizamiento por *splines* [Schoenberg (1964), Eubank (1988) y Wahba (1990)] y los métodos basados en desarrollos en serie ortogonales.

En el caso multivariante la situación que se plantea es la siguiente: Dado el conjunto de variables aleatorias $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ independientes e idénticamente distribuidas según una variable aleatoria $D + 1$ dimensional, (\mathbf{X}, Y) , con valores en \mathbb{R}^{D+1} , donde las variables aleatorias tienen función de densidad conjunta f con $\text{sop}(f) \subseteq \mathbb{R}^D$; la estructura del modelo de regresión no paramétrica multivariante tiene la misma formulación (1.1):

$$Y_i = m(X_i) + v^{1/2}(X_i) \varepsilon_i, \quad i = 1, \dots, n,$$

donde $v(\mathbf{x})$ es finita y ε_i son variables aleatorias independientes e idénticamente distribuidas con media cero, varianza unidad e independientes de las variables aleatorias, X_i .

En esta situación, los estimadores más comunes para $m(\mathbf{x})$ son versiones multivariantes de los estimadores núcleo de Nadaraya-Watson y Gasser-Müller, y de la técnica de suavizamiento por *splines*. Además de estos, el ajuste de polinomios locales mediante mínimos cuadrados ponderados es un método de estimación con buenas propiedades teóricas (Fan (1992)). Cleveland y Devlin (1988) tratan el caso ($D > 1$) y Ruppert y Wand (1994) desarrollaron la teoría asintótica para el estimador polinómico local, para D variables y distintos grados del polinomio.

Pasamos a describir en detalle el estimador polinomial local. En primer lugar, para el caso general de utilizar un polinomio de grado p , y a continuación para dos casos particulares, $p = 0$ y $p = 1$. Se presenta también un estimador núcleo, denominado estimador interno³. En lo que sigue denotaremos por $\hat{m}_p(\mathbf{x})$, al estimador polinómico local de grado p multivariante y por $m^{(\nu)}$ a la derivada de orden ν de la función m .

²Una formulación detallada y clara de tales estimadores se puede encontrar en Fan y Gijbels (1996).

³La importancia de introducir este estimador se debe a su relación con el estimador lineal local.

1.2. Estimador polinomial local de grado p

Dada la importancia de los estimadores polinomiales locales, vamos a ver cómo se definen para los dos modelos anteriores, univariante y multivariante.

1.2.1. Modelo univariante

Consideramos el siguiente problema de mínimos cuadrados ponderados:

$$\min_{\beta} \sum_{i=1}^n \{Y_i - \beta_0 - \beta_1 (X_i - x) - \dots - \beta_p (X_i - x)^p\}^2 K_h (X_i - x), \quad (1.2)$$

donde $\beta = (\beta_0, \dots, \beta_p)^T$ es un vector de coeficientes, h es el parámetro ancho de banda ($h > 0$); y $K_h(\cdot) = h^{-1}K(\cdot/h)$, siendo $K(\cdot)$ una función núcleo, que se supone función de densidad simétrica con soporte compacto.

Si denotamos por $\hat{\beta}_j$, $j = 0, \dots, p$, a las soluciones del problema (1.2), entonces, usando el desarrollo de Taylor, el estimador polinomial local vendrá dado por la primera de ellas, es decir,

$$\hat{m}_p(x; h) = \hat{\beta}_0,$$

y la curva completa, $\hat{m}_p(\cdot; h)$, se obtendrá variando x en el rango de estimación considerado.

La expresión matricial para la estimación de la función de regresión en cada punto, está dada por:

$$\hat{m}_p(x; h) = e_1^T (\mathbf{X}_x^T \mathbf{W}_{x,h} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{x,h} \mathbf{Y},$$

donde e_1 es un vector $(p+1) \times 1$, que tiene un 1 en la primera posición y ceros en el resto; $\mathbf{Y} = (Y_1, \dots, Y_n)^T$,

$$\mathbf{X}_x = \begin{pmatrix} 1 & X_1 - x & \dots & (X_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_n - x & \dots & (X_n - x)^p \end{pmatrix},$$

es la matriz de diseño $n \times (p+1)$ y

$$\mathbf{W}_{x,h} = \text{diag} \{K_h(X_1 - x), \dots, K_h(X_n - x)\},$$

es una matriz $(n \times n)$ de pesos, diagonal.

Casos particulares

- Cuando $p = 0$, se obtiene el **estimador de Nadaraya-Watson**:

$$\hat{m}_0(x; h) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i}{\sum_{i=1}^n K_h(X_i - x)}. \quad (1.3)$$

- Cuando $p = 1$, se obtiene el **estimador lineal local**:

$$\hat{m}_1(x; h) = n^{-1} \sum_{i=1}^n \frac{\{s_2(x; h) - s_1(x; h)(X_i - x)\} K_h(X_i - x) Y_i}{s_2(x; h) s_0(x; h) - s_1(x; h)^2}, \quad (1.4)$$

donde

$$s_r(x; h) = n^{-1} \sum_{i=1}^n (X_i - x)^r K_h(X_i - x), \quad r = 0, 1, 2.$$

1.2.2. Modelo multivariante

Las expresiones para el modelo multivariante no resultan tan triviales. Sea ahora el problema de mínimos cuadrados ponderados:

$$\min_{\beta} \sum_{i=1}^n \left\{ Y_i - \sum_{L=1}^p \sum_{l_1 + \dots + l_D = L} \beta_{l_1, \dots, l_D} \prod_{j=1}^D (X_j - x_j)^{l_j} \right\}^2 K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}),$$

donde $\beta = \{\beta_{l_1, \dots, l_D} : l_1 + \dots + l_D = L; L = 0, \dots, p\}$ es un vector de coeficientes, \mathbf{H} es una matriz de dimensión $D \times D$ simétrica, definida positiva; $K(\cdot)$ es una función núcleo no negativa D -variante y $K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2} \mathbf{u})$.

A la matriz \mathbf{H} se la denomina matriz ancho de banda, dado que es la extensión multivariante del parámetro ancho de banda univariante.

Si, de nuevo, denotamos por $\hat{\beta}_j$, $j = 0, \dots, p$, a las soluciones del problema anterior, entonces, usando el desarrollo de Taylor, el estimador polinómico local vendrá dado por la primera de ellas, esto es,

$$\hat{m}_p(\mathbf{x}) = \hat{\beta}_{0, \dots, 0}.$$

El problema en forma matricial se escribe como:

$$\min_{\beta} (\mathbf{Y} - \mathbf{X}_x \beta)^T \mathbf{W}_{x, \mathbf{H}} (\mathbf{Y} - \mathbf{X}_x \beta),$$

y su solución:

$$\hat{m}_p(\mathbf{x}) = e_1^T (\mathbf{X}_x^T \mathbf{W}_{x, \mathbf{H}} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{x, \mathbf{H}} \mathbf{Y},$$

donde e_1 es un vector $(pD + 1) \times 1$, con un 1 en la primera posición y 0 en el resto, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$,

$$\mathbf{X}_x = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T & \dots & ((\mathbf{X}_1 - \mathbf{x})^p)^T \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T & \dots & ((\mathbf{X}_n - \mathbf{x})^p)^T \end{pmatrix},$$

y

$$\mathbf{W}_{x,H} = \text{diag}\{K_H(\mathbf{X}_1 - \mathbf{x}), \dots, K_H(\mathbf{X}_n - \mathbf{x})\}.$$

Casos particulares

- $p = 0$, que coincide con la versión multivariante del estimador núcleo de Nadaraya-Watson:

$$\hat{m}_0(\mathbf{x}) = \frac{\sum_{i=1}^n K_H(\mathbf{X}_i - \mathbf{x}) Y_i}{\sum_{i=1}^n K_H(\mathbf{X}_i - \mathbf{x})}.$$

- $p = 1$, es el estimador lineal local multivariante (Ruppert y Wand, 1994):

$$\hat{m}_1(\mathbf{x}) = e_1^T (\mathbf{X}_{x,H}^T \mathbf{W}_{x,H} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{x,H} \mathbf{Y},$$

Para este caso no es posible dar expresiones sencillas, como las (1.3) y (1.4).

1.2.3. Estimador interno

Incluimos aquí un estimador tipo núcleo de la función de regresión, m , propuesto por Mack y Müller (1989), y estudiado en profundidad por Jones, Davies y Park (1994). Se define el **estimador interno** según la siguiente expresión:

$$\hat{m}_I(\mathbf{x}) = n^{-1} \sum_{i=1}^n \frac{K_h(\mathbf{X}_i - \mathbf{x}) Y_i}{f(\mathbf{X}_i)}.$$

En este caso, h es un vector $(D \times 1)$ de anchos de banda y $K_h(\mathbf{u}) = h^{-1} K(h^{-1}\mathbf{u})$.

La relación de este estimador con el estimador lineal local, en el caso univariante, es fácil de obtener, para ello se establecen las siguientes equivalencias asintóticas:

- $s_0(x) \sim f(x)$,
- $s_1(x) \sim -h\mu_2(K) f'(x)$, y
- $s_2(x) \sim \mu_2(K) f(x)$.

donde $\mu_2(K) = \int uu^T K(u) dt$ y $f'(x) = \frac{\partial f}{\partial x}$.

Obsérvese entonces que:

$$\begin{aligned} & (s_0(x) s_2(x) - s_1^2(x))^{-1} \times (s_2(x) - h^{-1}(X_i - x) s_1(x)) \sim \\ & \sim (\mu_2(K) f^2(x))^{-1} (\mu_2(K) f(x) + (X_i - x) \mu_2(K) f'(x)) = \\ & = \left(\frac{1}{f}\right)(x) + (X_i - x) \left(\frac{f'}{f^2}\right)(x) = \\ & = \left(\frac{1}{f}\right)(x) + (X_i - x) \left(\frac{1}{f}\right)'(x) \simeq \frac{1}{f(X_i)}. \end{aligned}$$

Con lo cual, los pesos que se introducen en (1.4) son asintóticamente equivalentes a $f^{-1}(X_i)$; y de esta forma, un estimador lineal local se puede aproximar asintóticamente por un estimador interno.

1.3. Propiedades de los estimadores

En general, como medida de error para estudiar el comportamiento del estimador de regresión, vamos a considerar el error cuadrático medio, MSE.

$$\text{MSE}[\hat{m}(x) | X] = E[\hat{m}(x) - m(x) | X]^2.$$

Dicho término se descompone en la suma de dos términos: un primer término que corresponde al cuadrado del sesgo del estimador y un segundo que corresponde a su varianza, en ambos casos sobre la distribución condicional de la variable de respuesta, Y , a la variable X .

$$\text{MSE}[\hat{m}(x) | X] = E[\hat{m}(x) - m(x) | X] + \text{Var}[\hat{m}(x) | X].$$

Siguiendo la misma estructura de las secciones anteriores, comenzamos estudiando las propiedades en el modelo de regresión univariante, en el caso general de un polinomio de grado p y en el caso particular $p = 1$. A continuación hacemos la generalización al modelo de regresión multivariante.

1.3.1. Modelo univariante

Introducimos alguna notación necesaria para esta sección:

Sea $\int u^j K(u) du = \mu_j(K)$, N_p una matriz $(p+1) \times (p+1)$, cuyo elemento (i, j) -ésimo es μ_{i+j-2} , y $M_p(u)$ la misma matriz que N_p , pero reemplazando la primera columna por $(1, u, \dots, u^p)^T$.

Se considera la función:

$$K_{(p)}(u) = \left\{ \frac{|M_p(u)|}{|N_p|} \right\} K(u).$$

Sea $\mathcal{D}_{x,h} = \{z : h^{-1}(x-z) \in \text{sop}(K)\}$ el soporte de $K_h(x-\cdot)$. Consideraremos x como un **punto interior**, si $\mathcal{D}_{x,h} \in \text{sop}(f)$. En otro caso, x se definirá como un **punto frontera**. Si x es un punto fijo en el interior de $\text{sop}(f)$, entonces x es un punto interior para todo n elevado. Sin embargo, conviene considerar una serie $x = \{x_n\}$ que converja a un punto, x_δ , de la frontera de $\text{sop}(f)$ tan rápido que x sea un punto frontera para todo n , es decir, $x = x_\delta + hc$, para un $c \in \text{sop}(K)$.

Para el estudio de las propiedades asintóticas del sesgo y la varianza condicionados de $\hat{m}_p(x)$ son necesarias las siguientes hipótesis:

- (i) La función núcleo, K , es acotada y tiene soporte compacto, tal que $\mu_2(K) \neq 0$. Además todos los momentos de orden impar de K son nulos.
- (ii) En un punto, x , de $\text{sop}(f)$, v , f' y $m^{(2)}$ son continuas. Además, $f(x) > 0$ y $v(x) > 0$.
- (iii) La función $m^{(p+2)}$ es continua en un entorno de x .
- (iv) $h \rightarrow 0$, $nh \rightarrow \infty$ cuando $n \rightarrow \infty$.

El siguiente teorema da el sesgo y la varianza (componentes del error cuadrático medio) del estimador \hat{m}_p en el caso univariante.

Teorema 1.3.1 *Si suponemos que x es un punto interior de $\text{sop}(f)$ y se cumplen las hipótesis (i)-(iv), entonces, si p es impar,*

$$\begin{aligned} E[\hat{m}_p(x) - m(x) \mid X_1, \dots, X_n] &= \\ &= \left\{ \int u^{p+1} K_{(p)}(u) du \right\} \left\{ \frac{m^{(p+1)}}{(p+1)!} h^{p+1} \right\} + o_P(h^{p+1}), \end{aligned}$$

y si p es par,

$$\begin{aligned} E[\hat{m}_p(x) - m(x) \mid X_1, \dots, X_n] &= \\ &= \left\{ \int u^{p+2} K_{(p)}(u) du \right\} \left\{ \frac{m^{(p+1)} f'(x)}{f(x)(p+1)!} + \frac{m^{(p+2)}}{(p+2)!} \right\} h^{p+2} + o_P(h^{p+2}). \end{aligned}$$

En ambos casos,

$$\begin{aligned} \text{Var}[\hat{m}_p(x) \mid X_1, \dots, X_n] &= \\ &= \left\{ \int K_{(p)}(u)^2 du \right\} \left\{ (nh)^{-1} v(x) / f(x) \right\} \{1 + o_P(1)\}. \end{aligned}$$

Caso particular ($p = 1$)

Las expresiones⁴ para el sesgo y la varianza en este caso son más sencillas, y son las siguientes:

$$E[\widehat{m}_1(x) - m(x) \mid X_1, \dots, X_n] = \frac{1}{2}h^2 m^{(2)}(x) \mu_2(K) + o_P(h^2)$$

y

$$\text{Var}[\widehat{m}_1(x) \mid X_1, \dots, X_n] = (nh)^{-1} \frac{R(K)}{f(x)} v(x) + o_P\{(nh)^{-1}\},$$

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

1.3.2. Modelo multivariante

Para el modelo de regresión multivariante sólo vamos a presentar las propiedades para el caso particular $p = 1$, ya que las expresiones para un p general no están dadas.

Denotamos por $D_g(\mathbf{x})$ el vector $D \times 1$ de derivadas parciales de primer orden y por $\mathcal{H}_g(\mathbf{x})$ la matriz Hessiana de dimensión $D \times D$ de una función D -variante g , suficientemente suave en \mathbf{x} . Sea $R(K) = \int K(\mathbf{u})^2 d\mathbf{u}$.

Las condiciones en este caso quedan de la siguiente forma:

- (i') El núcleo, K , es acotado y tiene soporte compacto. Además se verifica que $\int \mathbf{u}\mathbf{u}^T K(\mathbf{u}) d\mathbf{u} = \mu_2(K)\mathbf{I}$, donde $\mu_2(K) \neq 0$ es un escalar e \mathbf{I} es la matriz identidad (de dimensión $D \times D$); y todos los momentos impares de K son nulos, es decir, $\int u_1^{l_1} \cdots u_D^{l_D} K(\mathbf{u}) d\mathbf{u} = 0$, para todos los enteros no negativos, l_1, \dots, l_D , tales que su suma sea impar.
- (ii') Para un punto $\mathbf{x} \in \text{sop}(f)$, v es continua, f es continuamente derivable y todas las derivadas de segundo orden de m son continuas. Además, $f(\mathbf{x}) > 0$ y $v(\mathbf{x}) > 0$.
- (iii') $m^{(p+2)}$ es continua en un entorno de \mathbf{x} .
- (iv') La matriz \mathbf{H} es tal que $n^{-1}|\mathbf{H}|$ y cada uno de los elementos de \mathbf{H} tienden a cero cuando $n \rightarrow \infty$, siempre que \mathbf{H} sea una matriz simétrica y definida positiva. Además, existe una constante, L , tal que el número de condición⁵ de \mathbf{H} , es al menos L para todo n .

⁴Fan (1992).

⁵Cociente entre el mayor y el menor de los autovalores.

(v') Existe un conjunto convexo, \mathcal{C} , con interior no nulo y que contiene a \mathbf{x}_f tal que

$$\inf_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}) > 0.$$

Definimos una función:

$$K^*(\mathbf{u}; \mathbf{x}, \mathbf{H}) = e_1^T (\mathbf{X}_x^T \mathbf{W}_{\mathbf{x}, \mathbf{H}} \mathbf{X}_x)^{-1} \begin{bmatrix} 1 & (\mathbf{u} - \mathbf{x})^T \end{bmatrix}^T K_{\mathbf{H}}(\mathbf{u} - \mathbf{x}),$$

obsérvese que verifica:

$$\sum_{i=1}^n K^*(\mathbf{X}_i; \mathbf{x}, \mathbf{H}) = 1,$$

y

$$\sum_{i=1}^n K^*(\mathbf{X}_i; \mathbf{x}, \mathbf{H}) (\mathbf{X}_i - \mathbf{x}) = 0.$$

De este modo el estimador lineal local se puede escribir de la forma:

$$\hat{m}_1(\mathbf{x}) = \sum_{i=1}^n K^*(\mathbf{X}_i; \mathbf{x}, \mathbf{H}) Y_i.$$

El siguiente teorema da expresiones asintóticas del sesgo y de la varianza para el caso del estimador lineal local, $\hat{m}_1(\mathbf{x})$, para un punto, \mathbf{x} , interior.

Teorema 1.3.2 *Sea \mathbf{x} un punto interior de $\text{sop}(f)$. Si se verifican las hipótesis (i'), (ii') y (iv'), entonces,*

$$E[\hat{m}_1(\mathbf{x}) - m_1(\mathbf{x}) \mid X_1, \dots, X_n] = \frac{1}{2} \mu_2(\mathbf{K}) \text{tr}\{\mathbf{H}\mathcal{H}_m(\mathbf{x})\} + o_P(\text{tr}(\mathbf{H})),$$

y

$$\text{Var}[\hat{m}_1(\mathbf{x}) \mid X_1, \dots, X_n] = \{n^{-1} |\mathbf{H}|^{-1/2} R(\mathbf{K})/f(\mathbf{x})\} v(\mathbf{x}) \{1 + o_P(1)\}.$$

Si consideramos que el punto, \mathbf{x} , está en la frontera, las expresiones correspondientes a las componentes sesgo y varianza se recogen a continuación:

Teorema 1.3.3 *Sea $\mathbf{x} = \mathbf{x}_f + \mathbf{H}^{1/2}c$, donde $c \in \text{sop}(\mathbf{K})$. Entonces, bajo las condiciones (ii')-(iv'),*

$$\begin{aligned} E[\hat{m}_1(\mathbf{x}) - m_1(\mathbf{x}) \mid X_1, \dots, X_n] &= \\ &= \frac{e_1^T N_{\mathbf{x}}^{-1}}{2} \int_{D_{\mathbf{x}}} \begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix} K(\mathbf{u}) \mathbf{u}^T \mathbf{H}^{1/2} \mathcal{H}_m(\mathbf{x}) \mathbf{H}^{1/2} \mathbf{u} d\mathbf{u} + o_P(\text{tr}(\mathbf{H})), \end{aligned}$$

y

$$\begin{aligned} \text{Var} [\hat{m}_1(\mathbf{x}) \mid X_1, \dots, X_n] &= \\ &= \{n^{-1} |\mathbf{H}|^{-1/2} e_1^T N_{\mathbf{x}}^{-1} T_{\mathbf{x}} N_{\mathbf{x}}^{-1} e_1 / f(\mathbf{x})\} v(\mathbf{x}) \{1 + o_P(1)\}, \end{aligned}$$

donde

$$\begin{aligned} N_{\mathbf{x}} &= \int_{D_{\mathbf{x}}} \begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{u} \end{bmatrix} K(\mathbf{u}) \, d\mathbf{u}, \\ T_{\mathbf{x}} &= \int_{D_{\mathbf{x}}} \begin{bmatrix} 1 \\ \mathbf{u} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{u} \end{bmatrix} K^2(\mathbf{u}) \, d\mathbf{u}. \end{aligned}$$

1.4. El problema de la dimensionalidad

La generalización al caso multidimensional de la mayoría de técnicas de suavizado es posible, pero existen varios problemas asociados a los suavizadores D -dimensionales:

1. **La maldición de la dimensionalidad**⁶. Este problema se refiere al hecho de que cuando estamos estimando considerando un entorno con un número fijo de datos, y tenemos una superficie de gran dimensión, dicho entorno puede ser demasiado grande como para ser llamado local; hecho que produce grandes sesgos.
2. La falta de interpretabilidad, ya que será difícil visualizarlos gráficamente. No es posible representar superficies para $D > 2$.
3. El excesivo coste computacional de las versiones multivariantes, que requieren un gran número de operaciones (más de $O(n)$). Esto hace que en la práctica los suavizadores multidimensionales sólo se apliquen hasta dimensiones 2 o 3.

Los problemas anteriores llevan a plantear modelos alternativos que eviten estos inconvenientes, como los que tratamos en este trabajo y que reciben el nombre de **modelos aditivos**.

Los modelos aditivos se presentan como una herramienta útil para el análisis de datos. Estos modelos mantienen una importante característica de interpretación de los modelos lineales, al tener representada cada variable de forma separada; así la naturaleza de los efectos de una variable sobre la variable de respuesta no depende de los valores de las otras variables.

⁶Del inglés *curse of dimensionality*, Bellman (1961).

Los modelos aditivos fueron formulados por Friedman y Stuetzle (1981)⁷ y constituían el centro del algoritmo ACE de Breiman y Friedman (1985). A partir de entonces, estos y otros modelos relacionados son objeto de estudio para muchos autores, como por ejemplo, Wahba (1986) Engle, Granger, Rice y Weiss (1986) Burman (1985) Buja, Hastie y Tibshirani (1989) y Hastie y Tibshirani (1990).

Además de los conocidos datos de regresión, los modelos aditivos pueden aplicarse a otro tipo de datos. En Hastie y Tibshirani (1990) encontramos una discusión de la técnica aplicada a datos binomiales, binarios, de supervivencia y de estudios de control de casos.

⁷La superficie de regresión la consideraban como la suma de combinaciones lineales de funciones suaves.



Capítulo 2

Estudio de los modelos aditivos

2.1. Definición del modelo

Partimos del modelo de regresión múltiple heterocedástico¹:

$$\mathbf{Y} = m(\mathbf{X}) + v^{1/2}(\mathbf{X})\varepsilon,$$

donde la función $v(\mathbf{x}) = \text{Var}[Y | \mathbf{X} = \mathbf{x}]$ es finita, y los residuos, ε , son variables independientes, idénticamente distribuidas, con media cero, varianza uno e independientes del vector de variables, $\mathbf{X} = (X_1, \dots, X_D)^T$. El modelo aditivo supone que la función, m , se puede escribir como la suma de funciones univariantes de las variables independientes, que se suponen suaves; es decir,

$$m(\mathbf{x}) = E[Y | \mathbf{X} = (x_1, \dots, x_D)^T] = \alpha + m_1(x_1) + \dots + m_D(x_D). \quad (2.1)$$

Sea $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$, un conjunto de variables aleatorias independientes e idénticamente distribuidas, con valores en \mathbb{R}^{D+1} , y supongamos que dichos datos han sido generados mediante el modelo:

$$Y_i = \alpha + m_1(X_{1i}) + \dots + m_D(X_{Di}) + \varepsilon_i, \quad i = 1, \dots, n. \quad (2.2)$$

Para asegurar la identificabilidad de las componentes aditivas, $m_d(\cdot)$, suponemos que $E[m_d(X_{di})] = 0$, $d = 1, \dots, D$.

Sean $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{X}_i = (X_{1i}, \dots, X_{Di})^T$ y $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$. Además, las columnas de \mathbf{X} las denotamos por $\mathbf{X}^d = (X_{d1}, \dots, X_{dn})^T$ y los vectores de las funciones aditivas en las observaciones serán $\mathbf{m}_d = (m_d(X_{d1}), \dots, m_d(X_{dn}))^T$. Para una constante a , \mathbf{a} es un vector de n valores $(a, \dots, a)^T$. Si representamos la matriz de suavizamiento con respecto a la d -ésima variable, de dimensión $n \times n$, como \mathbf{S}_d ,

¹En principio, consideramos el modelo heterocedástico, más general; aunque será tratado solo en algunos casos.

definida para algún método de suavizamiento, las funciones aditivas, m_d , se pueden estimar de forma no paramétrica resolviendo el siguiente sistema de ecuaciones normales:

$$\begin{bmatrix} \mathbf{I} & \mathbf{S}_1 & \cdots & \mathbf{S}_1 \\ \mathbf{S}_2 & \mathbf{I} & \cdots & \mathbf{S}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_D & \mathbf{S}_D & \cdots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_D \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_D \end{bmatrix} \mathbf{Y}. \quad (2.3)$$

2.2. Métodos de estimación

Para estimar el modelo (2.2) se han propuesto diferentes métodos en los últimos años. Entre ellos, destacamos el algoritmo *backfitting* y el método de integración marginal, como los más importantes, y a partir de estos se han propuesto distintas modificaciones, todas ellas con el fin de mejorar las estimaciones obtenidas.

El algoritmo *backfitting*, propuesto por Buja, Hastie y Tibshirani (1989) proporciona un método iterativo para estimar las componentes unidimensionales. En Hastie y Tibshirani (1987, 1990) se pueden encontrar diferentes aplicaciones prácticas que demuestran los buenos resultados que proporciona el método. Sin embargo, el hecho de ser un proceso iterativo dificulta el estudio de las propiedades asintóticas de las estimaciones. Años más tarde, Opsomer y Ruppert (1997, 1998) y Opsomer (2000) dan las expresiones para el sesgo y la varianza en el caso particular de utilizar suavizadores polinomiales locales.

La dificultad del estudio de las propiedades del algoritmo *backfitting* es lo que lleva a otros autores a plantearse métodos de estimación alternativos. Así, Linton y Nielsen (1995) proponen el método de integración marginal para un modelo aditivo bivalente, este es un método que estima directamente cada una de las componentes univariantes y permite el estudio de sus propiedades, y fue generalizado para un modelo multivalente ($D > 2$) por Hengartner (1996) y Kim, Linton y Hengartner (1997). Fan, Härdle y Mammen (1998) extienden la idea del método de integración marginal de Linton y Nielsen (1995) incluyendo dos novedades: en primer lugar, introducen una función de ponderación, W , para obtener unas estimaciones eficientes y, en segundo lugar, permiten mayor flexibilidad en el modelo, ya que se puede utilizar el método con variables discretas.

Linton (1997) y Kim, Linton y Hengartner (1999) definen estimadores en dos etapas, que combinan los dos métodos que hemos mencionado. El procedimiento consiste en calcular unas estimaciones iniciales utilizando el método de integración marginal y a continuación utilizar estos valores para aplicar una iteración del algoritmo *backfitting*. De esta forma se obtienen mejores resultados que cuando se aplica sólo uno de los métodos individualmente.

Por último Mammen, Linton y Nielsen (1999) proponen un nuevo estimador

para la función de regresión aditiva, basado en una interpretación del estimador de Nadaraya-Watson como una proyección sobre el espacio de Hilbert.

A continuación pasamos a describir en detalle los más importantes: el algoritmo *backfitting*, el método de integración marginal y el algoritmo de proyección *backfitting*. En cada uno de ellos se da la definición del método en general y se estudian sus propiedades para algunos casos particulares. Nos detendremos más en el caso del algoritmo *backfitting* aplicado con suavizadores lineales locales, ya que será el método que utilizemos en nuestras aportaciones. Más información sobre otros métodos se puede encontrar en Raya Miranda, R. (2001).

2.3. Algoritmo *backfitting*

El algoritmo *backfitting*, como hemos mencionado anteriormente, fue propuesto por Buja, Hastie y Tibshirani (1989) y es un método iterativo para resolver el conjunto de ecuaciones normales dado en (2.3).

Consideramos el modelo (2.1), de donde:

$$E \left[Y - \alpha - \sum_{j \neq k} m_j(X_j) \mid X_k \right] = m_k(X_k), \quad k = 1, \dots, D;$$

esta expresión sugiere un algoritmo iterativo para el cálculo de cada una de las funciones m_1, \dots, m_D . La idea es la siguiente: Si α y las funciones m_j , ($j \neq k$), son conocidas, la función m_k se puede estimar mediante un suavizador univariante sobre el conjunto de observaciones $\{(X_{ki}, Y_i); i = 1, \dots, n\}$. Nótese además que para el suavizamiento se puede utilizar cualquier técnica de regresión no paramétrica univariante.

No obstante, para que las estimaciones, \hat{m}_k , obtenidas aplicando un suavizador dado, S_k , verifiquen la condición de identificabilidad² del modelo en cada etapa, será necesario centrar las estimaciones aplicando el siguiente cambio a los resultados:

$$\hat{m}_k^* = \hat{m}_k - n^{-1} \sum_{i=1}^n \hat{m}_k(X_{ki}). \quad (2.4)$$

2.3.1. Definición

Siguiendo las ideas antes expuestas, se define el algoritmo *backfitting* para la estimación de las funciones, m_1, \dots, m_D , a través de los siguientes pasos:

² $E[\hat{m}_d(X_d)] = 0, \quad d = 1, \dots, D.$

Algoritmo *backfitting*

Paso 1. Inicialización:

$$\hat{\alpha} = n^{-1} \sum_{i=1}^n Y_i \quad \text{y} \quad \hat{m}_k = m_k^{[0]}, \quad k = 1, \dots, D.$$

Paso 2. Para cada $k = 1, \dots, D$ obtener

$$\hat{m}_k = S_k \left(Y - \hat{\alpha} - \sum_{j \neq k} \hat{m}_j(X_j) \mid X_k \right),$$

y calcular \hat{m}_k^* según la expresión (2.4).

Paso 3. Repetir el paso 2 hasta que se obtenga el criterio de convergencia³ deseado.

El algoritmo involucra la elección de unas funciones iniciales, $m_1^{[0]}, \dots, m_D^{[0]}$, de las que en general no se tiene conocimiento a priori y, por ejemplo, se pueden utilizar las funciones de regresión lineales de Y sobre cada una de las variables X_k , $k = 1, \dots, D$. En cambio, si el algoritmo *backfitting* está incluido dentro de otro procedimiento mayor, este es el que proporciona las funciones iniciales. La estimación $\hat{\alpha} = n^{-1} \sum_{i=1}^n Y_i$, surge de forma natural, ya que $\alpha = E[Y]$.

A partir de ahora vamos a considerar sin pérdida de generalidad las observaciones, Y_i , centradas con respecto a su media, de modo que nos olvidaremos del término constante, α .

Nótese que la descripción general del algoritmo *backfitting* puede dar la impresión de que es necesario usar el mismo suavizador para cada una de las variables, no obstante, esto se hace así solo para facilitar la presentación, ya que de hecho este método permite que se utilice un suavizador diferente para cada variable, e incluso suavizadores que se apliquen a varias variables o a transformaciones de variables, sin que esto afecte para obtener la convergencia deseada.

2.3.2. Convergencia para suavizadores lineales generales

El algoritmo *backfitting* es uno de los métodos más utilizados para resolver el sistema (2.3) que consiste en nD ecuaciones con nD incógnitas. De forma teórica, y

³La convergencia se consigue cuando la diferencia entre la solución de una etapa con la anterior es menor que una cantidad prefijada.

siempre que exista la inversa de \mathbf{M} , es posible escribir la solución de los estimadores directamente según la siguiente expresión:

$$\begin{bmatrix} \hat{m}_1 \\ \hat{m}_2 \\ \vdots \\ \hat{m}_D \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{S}_1 & \cdots & \mathbf{S}_1 \\ \mathbf{S}_2 & \mathbf{I} & \cdots & \mathbf{S}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_D & \mathbf{S}_D & \cdots & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_D \end{bmatrix} \mathbf{Y} \equiv \mathbf{M}^{-1} \mathbf{C} \mathbf{Y},$$

Con lo que los estimadores del modelo aditivo son además suavizadores lineales en las observaciones, Y_i . Sea la matriz suavizadora aditiva, \mathbf{W}_d , definida como:

$$\mathbf{W}_d = \mathbf{E}_d \mathbf{M}^{-1} \mathbf{C},$$

donde \mathbf{E}_d es una matriz por bloques, de dimensión $n \times nD$, que tiene una matriz identidad en el bloque d -ésimo y ceros en el resto. Por tanto, la estimación para la componente d -ésima se puede escribir como $\hat{m}_d = \mathbf{W}_d \mathbf{Y}$.

Sea $\mathbf{W}_M = \mathbf{W}_1 + \cdots + \mathbf{W}_D$, la matriz de suavizamiento aditiva para la función completa, ($m = m_1 + \cdots + m_D$), entonces:

$$\hat{m} = \sum_{d=1}^D \hat{m}_d = \mathbf{W}_M \mathbf{Y}.$$

Además, sea $\mathbf{W}_M^{[-d]}$, la matriz de suavizamiento para la función $(D-1)$ -variante, $m_{(-d)}$, $m_{(-d)} = \sum_{k=1, k \neq d}^D m_k$, que se puede considerar como el suavizador de un modelo aditivo en el que los datos se han generado por el siguiente modelo:

$$Y'_i = m_1(X_{1i}) + \cdots + m_{d-1}(X_{d-1,i}) + m_{d+1}(X_{d+1,i}) + \cdots + m_D(X_{Di}) + \varepsilon_i,$$

$i = 1, \dots, n$.

Son muchos los autores que han estudiado las condiciones necesarias para la convergencia del sistema (2.3); el colorario 4.3 de Buja, Hastie y Tibshirani (1989) establece que, para un modelo aditivo bivalente, las ecuaciones de estimación son consistentes si:

$$\|\mathbf{S}_1 \mathbf{S}_2\| < 1, \quad (2.5)$$

para una norma matricial $\|\cdot\|$. Y las situaciones en las que $\|\mathbf{S}_1 \mathbf{S}_2\| = 1$ serán casos que son denominados de *concurvity*, situación análoga a la colinealidad en los modelos lineales. Hemos de señalar que la condición (2.5) no la satisface cualquier técnica de suavizamiento a no ser que las matrices suavizadoras sean centradas, es decir, que las matrices \mathbf{S}_d sean reemplazadas por $\mathbf{S}_d^* = (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) \mathbf{S}_d$.

Hastie y Tibshirani (1990) dan las expresiones explícitas para \mathbf{W}_1 y \mathbf{W}_2 cuando $D = 2$; que vienen expresadas como sigue:

$$\begin{aligned} \mathbf{W}_1 &= \mathbf{I} - (\mathbf{I} - \mathbf{S}_1 \mathbf{S}_2)^{-1} (\mathbf{I} - \mathbf{S}_1) \\ \mathbf{W}_2 &= \mathbf{I} - (\mathbf{I} - \mathbf{S}_2 \mathbf{S}_1)^{-1} (\mathbf{I} - \mathbf{S}_2), \end{aligned} \quad (2.6)$$

siempre que las inversas involucradas en ambas expresiones existan.

El lema 2.3.1, dado por Opsomer (2000) y expuesto a continuación, generaliza la condición (2.5) para el caso de D -variables y proporciona las condiciones que deben verificar las matrices suavizadoras para garantizar que los estimadores *backfitting* existan y sean únicos. La condición es totalmente general y no depende del tipo de suavizador que se utilice en el algoritmo. En el lema también se generaliza la expresión (2.6) al modelo aditivo D -variante.

Lema 2.3.1 *Un modelo aditivo D -variante estimado mediante el algoritmo *backfitting* con matrices de suavizamiento S_1, \dots, S_D , converge a una solución única, si*

$$\| S_d W_M^{[-d]} \| < 1,$$

para $d \in (1, \dots, D)$ y $\| \cdot \|$ una norma matricial. En este caso el suavizador aditivo con respecto a la variable d -ésima tiene la siguiente expresión:

$$W_d = I - \left(I - S_d W_M^{[-d]} \right)^{-1} (I - S_d).$$

Con esto se demuestra que la unicidad de los estimadores para suavizadores generales depende del comportamiento conjunto entre las matrices de suavizamiento individuales y el suavizador aditivo para el resto de variables, pero no proporciona una forma práctica de evaluar la existencia y unicidad de los estimadores, ya que no existe una expresión explícita para $W_M^{[-d]}$.

2.3.3. Suavizadores polinomiales locales

Cuando se estima el modelo aditivo mediante el algoritmo *backfitting*, hemos visto que el estudio teórico es complicado porque los estimadores se definen como la solución de un proceso iterativo. Sin embargo, Opsomer y Ruppert (1997, 1998) y Opsomer (2000) estudian las propiedades asintóticas de los estimadores *backfitting* en el caso en el que los suavizadores que se utilizan sean estimadores polinómicos locales. Dicho estudio hace posible la comparación de este método con los siguientes que estudiaremos en el capítulo.

Notación

Introducimos en primer lugar la notación necesaria:

Sean $S_d = (s_{d,x_{d1}}, \dots, s_{d,x_{dn}})^T$, donde s_{d,x_d}^T , $d = 1, \dots, D$ representan los núcleos equivalentes para la regresión polinomial local de la variable d -ésima en x_d , y su expresión es la siguiente:

$$s_{d,x_d}^T = e_1^T \left(X_{x_d}^{dT} W_{x_d,h_d} X_{x_d}^d \right)^{-1} X_{x_d}^{dT} W_{x_d,h_d},$$

donde e_i es un vector con un uno en la posición i -ésima y ceros en el resto, la matriz

$$W_{x_d, h_d} = \text{diag} \left\{ (1/h_d) K((X_{d1} - x_d)/h_d), \dots, (1/h_d) K((X_{dn} - x_d)/h_d) \right\},$$

para K una función núcleo, h_d el parámetro ancho de banda, y

$$X_{x_d}^d = \begin{bmatrix} 1 & (X_{d1} - x_d) & \cdots & (X_{d1} - x_d)^{p_d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_{dn} - x_d) & \cdots & (X_{dn} - x_d)^{p_d} \end{bmatrix},$$

donde p_d es el orden del polinomio local que se utiliza para ajustar la componente m_d (para el estudio de las propiedades los autores se apoyan en el trabajo de Ruppert y Wand (1994)). Sea $f(\mathbf{x})$, que representa la función de densidad de \mathbf{X}_i , y $f_d(x_d)$ la función de densidad marginal de la variable d -ésima. Para la función núcleo, K , $\mu_j(K) = \int u^j K(u) du$ es el momento de orden j y $R(K) = \int K(u)^2 du$. Por último, D_d^ν es el operador derivada ν -ésima con respecto a la variable d -ésima,

$$D_d^\nu \mathbf{m}_d = \begin{bmatrix} \frac{\partial^\nu m_d(X_{d1})}{\partial x_d^\nu} \\ \vdots \\ \frac{\partial^\nu m_d(X_{dn})}{\partial x_d^\nu} \end{bmatrix},$$

y

$$E \left[m_k^{(\nu)}(X_{ki}) \mid \mathbf{X}^d \right] = \begin{bmatrix} E \left[m_k^{(\nu)}(X_{k1}) \mid X_{d1} \right] \\ \vdots \\ E \left[m_k^{(\nu)}(X_{kn}) \mid X_{dn} \right] \end{bmatrix}.$$

Propiedades de los estimadores

El siguiente conjunto de hipótesis son necesarias para garantizar la existencia de los estimadores:

(O.I) La función núcleo, K , es acotada, continua, tiene soporte compacto y la primera derivada tiene un número finito de cambios de signo sobre su soporte. Además, $\mu_{p_d+1}(K) \neq 0$, para todo d .

(O.II) Las funciones de densidad f y f_d , $d = 1, \dots, D$ son continuas, acotadas, tienen soporte compacto y la primera derivada de cada una de ellas, tiene un número finito de cambios de signo sobre su soporte. Además, $f_d(x_d) > 0$ para todo $x_d \in \text{sop}(f_d)$.

(O.III) $h_d \rightarrow 0$ y $nh_d/\log(n) \rightarrow \infty$, para todo $d = 1, \dots, D$, cuando $n \rightarrow \infty$.

A continuación presentamos el sesgo y la varianza asintóticos en puntos interiores de soporte de f , para el caso en que el grado del polinomio es impar y $D = 2$, para simplificar la exposición. Las propiedades en los puntos de la frontera, grado par del polinomio y en el caso bivalente se pueden encontrar en Opsomer y Ruppert (1997). Y en Opsomer (2000) es posible encontrar las expresiones para el caso general de D variables.

Para el caso particular bivalente añadimos una restricción a la (O.II) dada para el caso general.

(O.II') Las funciones de densidad f y f_d , $d = 1, \dots, D$, son continuas, acotadas, tienen soporte compacto y la primera derivada de cada una de ellas, tiene un número finito de cambios de signo sobre su soporte. Además, $f_d(x_d) > 0$ para todo $x_d \in \text{sop}(f)$ y,

$$\sup_{x_1, x_2} \left| \frac{f(x_1, x_2)}{f_1(x_1) f_2(x_2)} - 1 \right| < 1.$$

Los dos resultados siguientes prueban que las matrices inversas que aparecen en las expresiones de W_1 y W_2 , dadas en (2.6), están bien definidas para polinomios locales de grado p_1 y p_2 , cuando $n \rightarrow \infty$. Los resultados⁴ solo prueban la existencia del estimador \hat{m}_1 , pero análogamente se pueden obtener para \hat{m}_2 y \hat{m} .

Lema 2.3.2 *Bajo las hipótesis (O.I), (O.II') y (O.III) son válidas las siguientes aproximaciones asintóticas, para todos los elementos de las matrices,*

$$\begin{aligned} S_1^* &= S_1 - \mathbf{1}\mathbf{1}^T/n + o(\mathbf{1}\mathbf{1}^T/n) \\ S_1^*S_2^* &= T_{12}^* + o(\mathbf{1}\mathbf{1}^T/n), \end{aligned}$$

donde T_{12}^* es una matriz cuyo elemento (ij) -ésimo es:

$$[T_{12}^*]_{ij} = \frac{1}{n} \frac{f(X_{1i}, X_{2j})}{f_1(X_{1i}) f_2(X_{2j})} - \frac{1}{n}.$$

Lema 2.3.3 *Bajo las hipótesis (O.I), (O.II') y (O.III), la matriz $(I - T_{12}^*)$ es invertible para todo n , y*

$$P[\exists N \text{ / } (I - S_1^*S_2^*) \text{ sea invertible } \forall n \geq N] = 1.$$

⁴Las demostraciones se pueden encontrar en Opsomer y Ruppert (1997).

Cuando $(\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1}$ existe, entonces la siguiente aproximación es válida sobre todos los elementos de la matriz:

$$(\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} = (\mathbf{I} - \mathbf{T}_{12}^*)^{-1} + o(\mathbf{1}\mathbf{1}^T/n) = \mathbf{I} + O(\mathbf{1}\mathbf{1}^T/n).$$

Sean \mathbf{t}_i^T y \mathbf{v}_j , la i -ésima fila y la j -ésima columna, respectivamente, de la matriz $(\mathbf{I} - \mathbf{T}_{12}^*)^{-1}$, y la siguiente hipótesis adicional.

(O.IV) Las derivadas, $(p_1 + 1)$ -ésima de \mathbf{m}_1 y $(p_2 + 1)$ -ésima de \mathbf{m}_2 existen, son continuas y están acotadas.

Teorema 2.3.4 *Bajo las hipótesis (O.I), (O.II'), (O.III) y (O.IV), para el conjunto de observaciones $\{(X_{1i}, X_{2i}), i = 1, \dots, n\}$, y p_1 y p_2 impares; el sesgo y la varianza condicionados de $\hat{\alpha}$, $\hat{m}_1(X_{1i})$ y $\hat{m}_2(X_{1i}, X_{2i})$ se aproximan por⁵:*

$$E[\hat{\alpha} - \alpha \mid \mathbf{X}_1, \mathbf{X}_2] = \alpha + O_P\left(\frac{1}{\sqrt{n}}\right),$$

$$\text{Var}[\hat{\alpha}] = \frac{\sigma^2}{n},$$

$$E[\hat{m}_1(X_{1i}) - m_1(X_{1i}) \mid \mathbf{X}_1, \mathbf{X}_2] =$$

$$= \frac{1}{(p_1 + 1)!} h_1^{p_1+1} \mu_{p_1+1}(K_{(p_1)}, X_{1i}) m_1^{(p_1+1)}(X_{1i}) +$$

$$+ \frac{1}{(p_1 + 1)!} h_1^{p_1+1} \mu_{p_1+1}(K_{(p_1)}) \left\{ (\mathbf{t}_i^T - \mathbf{e}_i^T) D^{p_1+1} \mathbf{m}_1 - E[m_1^{(p_1+1)}(X_{1i})] \right\} -$$

$$- \frac{1}{(p_2 + 1)!} h_2^{p_2+1} \mu_{p_2+1}(K_{(p_2)}) \left\{ \mathbf{t}_i^T E[m_2^{(p_2+1)}(X_{2i}) \mid \mathbf{X}_1] - E[m_2^{(p_2+1)}(X_{2i})] \right\} +$$

$$+ O_P\left(\frac{1}{\sqrt{n}}\right) + o_P(h_1^{p_1+1} + h_2^{p_2+1}),$$

$$\text{Var}[\hat{m}_1(X_{1i}) \mid \mathbf{X}_1, \mathbf{X}_2] = \sigma^2 \frac{R(K_{(p_1)}, X_{1i})}{nh_1} f_1(X_{1i})^{-1} + o_P\left(\frac{1}{nh_1}\right),$$

⁵Los resultados para $\hat{m}_2(X_{2i})$ se obtienen de forma análoga.

$$\begin{aligned}
E[\widehat{m}(X_{1i}, X_{2i}) - m(X_{1i}, X_{2i}) \mid \mathbf{X}_1, \mathbf{X}_2] &= \\
&= \frac{1}{(p_1 + 1)!} h_1^{p_1+1} (\mu_{p_1+1}(K_{(p_1)}, X_{1i}) m_1^{(p_1+1)}(X_{1i}) + \\
&+ \mu_{p_1+1}(K_{(p_1)}) \left\{ (\mathbf{t}_i^T - \mathbf{e}_i^T) D^{p_1+1} \mathbf{m}_1 - \mathbf{v}_i^T E[m_1^{(p_1+1)}(X_{1i}) \mid \mathbf{X}_2] \right\}) + \\
&+ \frac{1}{(p_2 + 1)!} h_2^{p_2+1} (\mu_{p_2+1}(K_{(p_2)}, X_{2i}) m_2^{(p_2+1)}(X_{2i}) + \\
&+ \mu_{p_2+1}(K_{(p_2)}) \left\{ (\mathbf{t}_i^T - \mathbf{e}_i^T) D^{p_2+1} \mathbf{m}_2 - \mathbf{v}_i^T E[m_2^{(p_2+1)}(X_{2i}) \mid \mathbf{X}_1] \right\}) + \\
&+ o_P(h_1^{p_1+1} + h_2^{p_2+1}), \\
\text{Var}[\widehat{m}(X_{1i}, X_{2i}) \mid \mathbf{X}_1, \mathbf{X}_2] &= \\
&= \sigma^2 \left\{ \frac{R(K_{(p_1)}, X_{1i})}{nh_1} f_1(X_{1i})^{-1} + \frac{R(K_{(p_2)}, X_{2i})}{nh_2} f_2(X_{2i})^{-1} \right\} + \\
&+ o_P\left(\frac{1}{nh_1} + \frac{1}{nh_2}\right). \tag{2.7}
\end{aligned}$$

Corolario 2.3.5 Cuando la observación (X_{1i}, X_{2i}) está en el interior de $\text{sop}(f)$, el sesgo y la varianza condicionados de $\widehat{m}_1(X_{1i})$ y $\widehat{m}(X_{1i}, X_{2i})$ se aproximan por:

$$\begin{aligned}
E[\widehat{m}_1(X_{1i}) - m_1(X_{1i}) \mid \mathbf{X}_1, \mathbf{X}_2] &= \\
&= \frac{1}{(p_1 + 1)!} h_1^{p_1+1} \mu_{p_1+1}(K_{(p_1)}) \left\{ \mathbf{t}_i^T D^{p_1+1} \mathbf{m}_1 - E[m_1^{(p_1+1)}(X_{1i})] \right\} - \\
&- \frac{1}{(p_2 + 1)!} h_2^{p_2+1} \mu_{p_2+1}(K_{(p_2)}) \left\{ \mathbf{t}_i^T E[m_2^{(p_2+1)}(X_{2i}) \mid \mathbf{X}_1] - E[m_2^{(p_2+1)}(X_{2i})] \right\} + \\
&+ O_P\left(\frac{1}{\sqrt{n}}\right) + o_P(h_1^{p_1+1} + h_2^{p_2+1}),
\end{aligned}$$

$$\text{Var}[\widehat{m}_1(X_{1i}) \mid \mathbf{X}_1, \mathbf{X}_2] = \sigma^2 \frac{R(K_{(p_1)})}{nh_1} f_1(X_{1i})^{-1} + o_P\left(\frac{1}{nh_1}\right),$$

$$\begin{aligned}
E[\widehat{m}(X_{1i}, X_{2i}) - m(X_{1i}, X_{2i}) \mid \mathbf{X}_1, \mathbf{X}_2] &= \\
&= \frac{1}{(p_1 + 1)!} h_1^{p_1+1} \mu_{p_1+1}(K_{(p_1)}) \left\{ \mathbf{t}_i^T D^{p_1+1} \mathbf{m}_1 - \mathbf{v}_i^T E[m_1^{(p_1+1)}(X_{1i}) \mid \mathbf{X}_2] \right\} + \\
&+ \frac{1}{(p_2 + 1)!} h_2^{p_2+1} \mu_{p_2+1}(K_{(p_2)}) \left\{ \mathbf{v}_i^T D^{p_2+1} \mathbf{m}_2 - \mathbf{t}_i^T E[m_2^{(p_2+1)}(X_{2i}) \mid \mathbf{X}_1] \right\} + \\
&+ o_P(h_1^{p_1+1} + h_2^{p_2+1}),
\end{aligned}$$

$$\begin{aligned}
\text{Var}[\widehat{m}(X_{1i}, X_{2i}) \mid \mathbf{X}_1, \mathbf{X}_2] &= \\
&= \sigma^2 \left\{ \frac{R(K_{(p_1)})}{nh_1} f_1(X_{1i})^{-1} + \frac{R(K_{(p_2)})}{nh_2} f_2(X_{2i})^{-1} \right\} + o_P\left(\frac{1}{nh_1} + \frac{1}{nh_2}\right).
\end{aligned}$$

Al igual que para el caso general de D -variables, si se considera el caso en que las variables son independientes, entonces se obtiene una versión simplificada para el sesgo. Estas expresiones se pueden encontrar en Opsomer y Ruppert (1997).

Corolario 2.3.6 *Para un modelo con D variables independientes, el sesgo de \widehat{m}_d , en un punto interior, se puede aproximar por*

$$\begin{aligned} E[\widehat{m}_d(X_{di}) - m_d(X_{di}) | \mathbf{X}] &= \\ &= \frac{\mu_{p_d+1}(K_{(p_d)})}{(p_d + 1)!} h_d^{p_d+1} \left(m_d^{(p_d+1)}(X_{di}) - E[m_d^{(p_d+1)}(X_{di})] \right) + \\ &+ O_p\left(\frac{1}{\sqrt{n}}\right) + o_p(h_d^{p_d+1}). \end{aligned}$$

2.4. Método de integración marginal

Linton y Nielsen (1995) introducen un método que, a diferencia del algoritmo *backfitting* cuando se define de forma general, permite derivar directamente las propiedades estadísticas de los estimadores resultantes. Este hecho facilita el estudio del error cuadrático medio, y su aplicación para la definición de selectores del ancho de banda. Este método será posteriormente utilizado para obtener las estimaciones iniciales necesarias para la aplicación del algoritmo *backfitting* por Linton (1997) y Kim, Linton y Hengarnert (1999), que obtienen así mejores resultados para las propiedades de los estimadores.

Se presenta el método para el caso bivalente con la definición de los estimadores y el estudio de sus propiedades. A continuación se exponen las dificultades que encontraron dichos autores para extender este método al caso de D -variables y la modificación del método que propone Hengartner (1996) que se puede aplicar al caso de D -variables.

2.4.1. Caso bivalente

Definición del estimador

Sea $m(x_1, x_2)$ una función de regresión bivalente, y consideramos un modelo aditivo sin término constante, es decir,

$$m(x_1, x_2) = m_1(x_1) + m_2(x_2). \quad (2.8)$$

El método consiste en realizar una estimación inicial⁶ de la función bidimensional,

⁶A partir de ahora la denominaremos estimación piloto.

e integrar dicha estimación con respecto a cada una de las variables, para obtener las estimaciones individuales de las funciones aditivas.

Sea Q una función de pesos determinística, tal que $\int dQ(x_2) = 1$, y q la función de densidad de Q respecto de alguna medida discreta, o una medida de Lebesgue. Consideramos la siguiente expresión:

$$\eta_1(x_1) = \int m(x_1, x_2) dQ(x_2).$$

En el modelo aditivo será,

$$\eta_1(x_1) = m_1(x_1) + c_1,$$

donde

$$c_1 = \int m_2(x_2) dQ(x_2).$$

Entonces, $\eta_1(x_1)$ será la estimación de $m_1(x_1)$.

La estimación de la función de regresión de Y sobre X_1 y X_2 , $m(x_1, x_2)$, se calcula mediante el estimador lineal local⁷ con una función núcleo producto; es decir,

$$\hat{m}^{pi}(x_1, x_2) = e_1^T (\mathbf{X}_x^T \mathbf{W}_{x,h} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{x,h} \mathbf{Y} = \sum_{i=1}^n w_i(x_1, x_2) Y_i,$$

con $\mathbf{h} = (h_1, h_2)^T$ y $K_{h_j}(\cdot) = h_j^{-1} K(h_j^{-1} \cdot)$, $j = 1, 2$. La función K es una función de densidad univariante, derivable y simétrica respecto de cero. En Wand y Jones (1993) se puede encontrar una discusión sobre la utilización de diferentes anchos de banda para cada variable.

Entonces, la estimación de $\eta_1(x_1)$ se define como:

$$\hat{\eta}_1(x_1) = \int \hat{m}^{pi}(x_1, x_2) dQ(x_2) = \sum_{i=1}^n w_i(x_1) Y_i, \quad (2.9)$$

donde,

$$w_i(x_1) = \int w_i(x_1, x_2) dQ(x_2).$$

Y finalmente, para la otra componente se hace un desarrollo análogo.

⁷Fan (1992).

Propiedades del estimador

Consideremos un modelo de regresión bivalente y homocedástico, donde $\text{Var}(\varepsilon_i) = \sigma^2$. Sean $F_1(x_1)$ y $F_2(x_2)$, las funciones de distribución marginales de la variable explicativa $\mathbf{X} = (X_1, X_2)^T$. Se supone que f tiene soporte compacto en $A_1 \times A_2 = A \subseteq \mathbb{R}^2$, y que el soporte de Q está contenido en A_2 .

Basándose en los resultados para el sesgo y la varianza de Ruppert y Wand (1994), el teorema 2.4.1 de Linton y Nielsen (1995), expuesto a continuación, proporciona los dos primeros momentos del estimador de integración marginal y su distribución asintótica. Para ello se establecen las siguientes hipótesis:

(LN.1) m posee dos derivadas continuas en cada variable.

(LN.2) f es continuamente derivable.

(LN.3) $h_1, h_2 \rightarrow 0$ y $nh_1h_2^2 \rightarrow \infty$, cuando $n \rightarrow \infty$.

Teorema 2.4.1 *Bajo (LN.1)-(LN.3), condicionando sobre $\{X_{1i}, X_{2i}\}_{i=1}^n$, se tiene que:*

$$(nh_1)^{1/2} \{\hat{\eta}_1(x_1) - E[\hat{\eta}_1(x_1)]\} \rightarrow \mathcal{N}(0, v_1^2(x_1)),$$

donde

$$v_1^2(x_1) = R(K) \sigma^2 \int f^{-1}(x_1, x_2) q^2(x_2) dx_2,$$

y,

$$\begin{aligned} E[\hat{\eta}_1(x_1)] - \eta_1(x_1) &= \\ &= \frac{1}{2} \mu_2(K) \left\{ h_1^2 \int \frac{\partial^2 m(x_1, x_2)}{\partial x_1^2} dQ(x_2) + h_2^2 \int \frac{\partial^2 m(x_1, x_2)}{\partial x_2^2} dQ(x_2) \right\} + \\ &+ o(h_1^2 + h_2^2). \end{aligned} \quad (2.10)$$

Cuando $h_1, h_2 = O(n^{-1/5})$, $\hat{\eta}_1(x_1)$ converge con razón $n^{2/5}$. Se consigue una reducción en el sesgo asintótico cuando infrasuavizamos con respecto a x_2 , es decir, tomando $h_2 = o(n^{-1/5})$, en cuyo caso (2.10) depende solo de las derivadas con respecto a x_1 ; es decir, el sesgo será

$$\frac{1}{2} h_1^2 \mu_2(K) \eta_1''(x_1).$$

En cuanto a la elección de Q , se trata de elegir una función de ponderación de mínima varianza, Q_{MV} , con función de densidad q_{MV} , que minimice:

$$V = \int v_1^2(x_1) f_1(x_1) dx_1.$$

Dicha función resulta ser:

$$q_{MV}(x_2) = \tau^{-1}(x_2) \left(\int \tau^{-1}(x_2) dx_2 \right)^{-1},$$

donde

$$\tau(x_2) = \int f^{-1}(x_1, x_2) f_1(x_1) dx_1,$$

y tiene asociado el siguiente valor mínimo de la función,

$$V = R(K) \sigma^2 \left(\int \tau^{-1}(x_2) dx_2 \right)^{-1}.$$

Otra forma será elegir Q con error cuadrático medio integrado óptimo, pero en general este dependerá de f y de m , si bien, para el caso especial en que m es aditiva y $h_2 = o(n^{-1/5})$, el sesgo no depende de Q y entonces, q_{MV} tiene error cuadrático medio integrado óptimo.

En la práctica, se recomienda el uso de la función de distribución empírica, $F_{2n}(x_2)$, que converge rápidamente a $F_2(X_2)$; además, los resultados del teorema siguen siendo válidos. Cuando X_1 y X_2 son independientes, se verifica que $Q_{MV}(X_2) = F_2(X_2)$, lo que justifica esta elección.

2.4.2. Extensión al caso de D -variables

Si suponemos que x_2 es $(D - 1)$ -dimensional, el procedimiento para estimar la función univariante de x_1 sería el mismo, excepto que se utilizarán un estimador núcleo y una función de pesos, Q , $(D - 1)$ -dimensional. En esta situación, $\eta_1(x_1)$ estima consistentemente a $m_1(x_1)$ sin imponer ninguna estructura sobre la función $m_2(x_2)$, $(D - 1)$ -dimensional. Y aplicando el procedimiento a cada una de las componentes se obtendrían todas las funciones univariantes.

No obstante, el método de integración marginal de Linton y Nielsen (1995) verifica el teorema central del límite para funciones de regresión con dos derivadas continuas, pero el resultado sólo es válido para D muy pequeño, porque para valores elevados, el estimador lineal local sufre la maldición de la dimensionalidad, que se ha comentado en el capítulo 1, además de su complejidad computacional.

Hengartner (1996) establece los siguientes resultados: bajo ciertas condiciones y utilizando como estimador piloto el suavizador de regresión de Nadaraya-Watson

multivariante se obtiene una estimación asintóticamente normal de las componentes aditivas unidimensionales. Además, si se utiliza como estimador piloto el suavizador de integración interno, entonces se obtienen estimadores que son asintóticamente normales con razón de convergencia óptima, y se calculan con $O(n^2 + nD)$ operaciones⁸.

Definición y propiedades del estimador

Sea $\{(\mathbf{X}_i, Y_i), i = 1, \dots, n\}$, una muestra aleatoria independiente, idénticamente distribuida con función de densidad conjunta, $f(\mathbf{x}, Y) = f(Y|\mathbf{x})f(\mathbf{x})$. Sea $q_2(\mathbf{x}_2)$, una función de densidad sobre $x_2 \otimes \dots \otimes x_D$. La siguiente expresión define la función de regresión marginal para la primera componente, con respecto a $q_2(\mathbf{x}_2)$,

$$\eta_1(x_1) = \int_{\mathbb{R}^{D-1}} m(\mathbf{x}) q_2(\mathbf{x}_2) d\mathbf{x}_2,$$

donde $\mathbf{x} = (x_1, \mathbf{x}_2) \in \mathbb{R}^D$. Cuando la función de regresión $m(\mathbf{x})$ es aditiva, entonces $\eta_1(x_1)$ es igual a $m_1(x_1)$ más una constante.

Linton y Nielsen (1995) establecían que el método de integración marginal reduce la varianza de cada una de las componentes, pero no el sesgo; Hengartner (1996) y Kim, Linton y Hengartner (1997) resuelven este problema considerando estimadores centrados en cada componente. La diferencia entre ambos trabajos está en el estimador piloto considerado. Hengartner (1996) considera varios estimadores piloto: el estimador de Nadaraya-Watson, un estimador interno y el estimador lineal local. Y Kim, Linton y Hengartner (1997) utilizan un estimador interno. Presentamos los resultados que se obtienen en ambos trabajos.

En primer lugar, consideramos el suavizador de regresión multivariante de Nadaraya-Watson, utilizado por Hengartner (1996):

$$\hat{m}_{NW}^{pi}(\mathbf{x}) = \frac{n^{-1} \sum_{i=1}^n \mathbf{K}_h(\mathbf{X}_i - \mathbf{x}) Y_i}{n^{-1} \sum_{i=1}^n \mathbf{K}_h(\mathbf{X}_i - \mathbf{x})} \equiv \frac{\hat{r}(\mathbf{x})}{\hat{f}(\mathbf{x})},$$

Y sean las siguientes hipótesis:

(H.1) La función de regresión multivariante, $m(\mathbf{x})$, tiene s derivadas continuas con respecto a x_1 , y la varianza condicionada, $\sigma^2(\mathbf{x}) = \text{Var}[Y | \mathbf{X} = \mathbf{x}]$, es finita y continua Lipschitz.

(H.2) La función de densidad conjunta de las variables, $f(\mathbf{x})$, está estrictamente acotada por cero e infinito, y es continua Lipschitz sobre su soporte compacto. La función de densidad condicionada verifica $f(\mathbf{x}_2 | x_1) > 0$, para todo $(x_1, \mathbf{x}_2) \in \text{sop}(f(\mathbf{x}))$. Y existe la función de densidad marginal, $f_1(x_1)$.

⁸Esta cantidad resulta menor que el número de cálculos necesarios para realizar k iteraciones del algoritmo *backfitting*.

(H.3) La función de integración, $q_2(\mathbf{x}_2) = \prod_{j \geq 2} q_j(x_j)$, está acotada, es continua Lipschitz y el soporte de $q(\mathbf{x}) = \prod_{j=1}^D q_j(x_j)$ está contenido en el soporte de $f(\mathbf{x})$.

(H.4) El núcleo de suavizamiento multivariante, \mathbf{K} , es el producto de D funciones núcleo univariantes $K_j(\cdot)$, cada una con soporte compacto, acotada, continua Lipschitz y $R(K_j) = \int K_j^2(t) dt < \infty$. Las funciones $K_j(t)$ son de orden r_j , es decir,

$$\int t^k K_j(t) dt = 0 \text{ para } k = 1, \dots, r_j - 1,$$

$$\int t^{r_j} K_j(t) dt < \infty; \text{ y } r_1 = s.$$

(H.5) $h_1 = n^{-1/(2s+1)}$, $h_j \rightarrow 0$ y $n \prod_{j=1}^D h_j \rightarrow \infty$, cuando $n \rightarrow \infty$.

(H.6) La función de densidad condicionada, $f(\mathbf{x}_2 | x_1)$, tiene s derivadas continuas en x_1 para todo $(x_1, \mathbf{x}_2) \in \text{sop}(q(\mathbf{x})) \otimes \varepsilon$, para un $\varepsilon > 0$.

(H.7) Sea $r_0 = \lfloor (D-1)/2 \rfloor + 1$ y se supone que

$$\forall \mathbf{r} \in \Lambda(r_0) = \left\{ (r_2, \dots, r_D) : \sum_{j=2}^D r_j = r_0 \right\},$$

las derivadas parciales

$$\frac{\partial^{r_0}}{\partial^{r_2} x_2 \dots \partial^{r_D} x_D} \frac{q(\mathbf{x}_2)}{f(\mathbf{x}_2 | x_1)},$$

existen y son continuas en x_2, \dots, x_D .

(H.8) Para $j = 2, \dots, D$, el orden del núcleo $K_j(t)$ es $r_0 = \lfloor (D-1)/2 \rfloor + 1$ y los anchos de banda son $h_2 = \dots = h_D = n^{-s/(2r_0(s+1))}$.

Entonces, se define el estimador de integración para la función de regresión mar-

ginal, $\eta_1(x_1)$, por la siguiente expresión:

$$\begin{aligned}\hat{\eta}_1(x_1) &= \int_{\mathbb{R}^{D-1}} \frac{1}{n} \sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x}) Y_i \frac{q_2(\mathbf{x}_2)}{\hat{f}(x_1, \mathbf{x}_2)} d\mathbf{x}_2 = \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K_1\left(\frac{X_{1i} - x_1}{h_1}\right) \frac{Y_i}{\hat{f}(x_1)} \times \\ &\quad \times \int_{\mathbb{R}^{D-1}} \left\{ \prod_{j=2}^D \frac{1}{h_j} K_j\left(\frac{X_{ji} - x_j}{h_j}\right) \right\} \frac{q_2(\mathbf{x}_2)}{\hat{f}(\mathbf{x}_2 | x_1)} d\mathbf{x}_2 = \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K_1\left(\frac{X_{1i} - x_1}{h_1}\right) \frac{\hat{\xi}_i(x_1)}{\hat{f}(x_1)},\end{aligned}$$

donde

$$\hat{\xi}_i(x_1) = Y_i \int_{\mathbb{R}^{D-1}} \left\{ \prod_{j=2}^D \frac{1}{h_j} K_j\left(\frac{X_{ji} - x_j}{h_j}\right) \right\} \frac{q_2(\mathbf{x}_2)}{\hat{f}(\mathbf{x}_2 | x_1)} d\mathbf{x}_2.$$

Teorema 2.4.2 Si se considera el suavizador de regresión de Nadaraya-Watson multivariante como estimador piloto de la función de regresión, y se estima la función

$$\begin{aligned}\eta_1(X_1) &= \int_{\mathbb{R}^{D-1}} m(x_1, \mathbf{x}_2) q_2(\mathbf{x}_2) d\mathbf{x}_2 - \\ &\quad - \int_{\mathbb{R}^D} m(x_1, \mathbf{x}_2) q_2(\mathbf{x}_2) q_1(x_1) d\mathbf{x}_2 dx_1,\end{aligned}$$

mediante

$$\begin{aligned}\hat{\eta}_1(x_1) &= \int_{\mathbb{R}^{D-1}} \hat{m}_{NW}^{pi}(x_1, \mathbf{x}_2) q_2(\mathbf{x}_2) d\mathbf{x}_2 - \\ &\quad - \int_{\mathbb{R}^D} \hat{m}_{NW}^{pi}(x_1, \mathbf{x}_2) q_2(\mathbf{x}_2) q_1(x_1) d\mathbf{x}_2 dx_1,\end{aligned}$$

bajo las hipótesis (H.1)-(H.8), el estimador de integración verifica:

$$\sqrt{nh_1} \{\hat{\eta}_1(x_1) - \eta_1(x_1)\} \rightarrow \mathcal{N}(b_1(x_1), v_1^2(x_1)),$$

donde

$$b_1(x_1) = \frac{\mu_s(K)}{f(x_1)} \frac{\partial^s}{\partial z^s} \left\{ f(z) \int_{\mathbb{R}^{D-1}} m(z, \mathbf{z}_2) \frac{f(\mathbf{z}_2 | z)}{f(\mathbf{z}_2 | x_1)} q_2(\mathbf{z}_2) d\mathbf{z}_2 \right\} \Big|_{z=x_1},$$

y

$$v^2(x_1) = \frac{\text{Var}_\infty \left[\hat{\xi}_i(x_1) \mid x_1 = z_1 \right]}{f(x_1)} R(K_j),$$

siendo

$$\begin{aligned} \text{Var}_\infty \left[\widehat{\xi}_i(x_1) \mid x_1 = z_1 \right] &= \\ &= \int_{\mathbb{R}^{D-1}} \left(\sigma^2(z_1, \mathbf{z}_2) + m^2(z_1, \mathbf{z}_2) \right) \frac{q_2(\mathbf{z}_2)}{f(\mathbf{z}_2 \mid x_1)} \frac{f(\mathbf{x}_2 \mid z_1)}{f(\mathbf{z}_2 \mid x_1)} q_2(\mathbf{z}_2) d\mathbf{z}_2 - \\ &\quad - \left(\int_{\mathbb{R}^{D-1}} m(z_1, \mathbf{z}_2) \frac{f(\mathbf{x}_2 \mid z_1)}{f(\mathbf{z}_2 \mid x_1)} q_2(\mathbf{z}_2) d\mathbf{z}_2 \right)^2. \end{aligned}$$

El sesgo que se obtiene en este caso depende de la función de densidad de las variables, lo que requiere imponer condiciones de suavidad también sobre la función de densidad. Sin embargo, en el caso de utilizar como estimador piloto el suavizador interno, el sesgo que se obtiene no depende de esta función. Esta solución es la que se propone a continuación, aunque dado que el número de operaciones que requiere este estimador es muy elevado, no es muy recomendado en la práctica.

Se considera ahora el caso de utilizar como estimador piloto el suavizador de regresión interno, dado por Kim, Linton y Hengartner (1997), para el que se obtiene la siguiente expresión:

$$\widehat{m}_I^{pi}(\mathbf{x}) = \frac{1}{nh_1 h_2^{D-1}} \sum_{i=1}^n \frac{K_1\left(\frac{X_{1i}-x_1}{h_1}\right) K_2\left(\frac{\mathbf{X}_{2i}-\mathbf{x}_2}{h_2}\right) Y_i}{\widehat{f}(X_{1i}, \mathbf{X}_{2i})},$$

donde K_1 es una función núcleo univariante, K_2 es una función núcleo de dimensión $(D-1)$, ambas simétricas respecto de cero; h_1 y h_2 son dos anchos de banda escalares, y,

$$\widehat{f}(\mathbf{x}) = \frac{1}{nh_1 h_2^{D-1}} \sum_{i=1}^n K_1\left(\frac{X_{1i}-x_1}{h_1}\right) K_2\left(\frac{\mathbf{X}_{2i}-\mathbf{x}_2}{h_2}\right),$$

es un estimador de la función de densidad, $f(\mathbf{x})$. Entonces, el estimador de $\eta_1(X_{1i})$ se define de la siguiente forma:

$$\begin{aligned} \widehat{\eta}_1(X_{1i}) &= n^{-1} \sum_{j=1}^n \widehat{m}_I^{pi}(X_{1i}, \mathbf{X}_{2j}) = \\ &= \frac{1}{nh_1} \sum_{j=1}^n K_1\left(\frac{X_{1i}-X_{1j}}{h_1}\right) \frac{\widehat{f}_2(\mathbf{X}_{2j})}{\widehat{f}(X_{1i}, \mathbf{X}_{2j})} Y_{ij}, \end{aligned}$$

siempre que K_2 sea una función núcleo simétrica respecto de cero.

Las propiedades asintóticas de este estimador son similares a las del estimador de integración marginal, es decir, converge a una variable normal con razón de convergencia $n^{2/5}$. Este resultado se recoge en el siguiente teorema, y considerando el siguiente conjunto de hipótesis:

(KLH.1) K_1 y K_2 son núcleos simétricos con respecto de cero y de órdenes r_j , $j = 1, 2$ respectivamente, es decir, $\int u^l K_j(u) du = 0$, $l = 1, \dots, r_j - 1$, donde $r_1 = 2$ y $r_2 \geq 2$. Ambas funciones núcleo tienen soporte compacto, están acotadas y son continuas Lipschitz.

(KLH.2) h_1 es del orden de $n^{-1/5}$, $n^{2/5} h_2^r \rightarrow 0$, y $nh_1 h_2^{D-1} / \log n \rightarrow \infty$, cuando $n \rightarrow \infty$.

(KLH.3) Las funciones m y f tienen $r \geq D/2$ derivadas continuas en cada una de las variables de \mathbf{x}_2 y son dos veces derivables con respecto a x_1 . La función de densidad condicionada, $f_{1|2}(x_1 | \mathbf{x}_2)$, está acotada por cero e infinito, sobre su soporte compacto. La varianza condicionada, $\sigma^2(\mathbf{x}) = \text{Var}[Y | \mathbf{X} = \mathbf{x}]$, es continua Lipschitz, acotada por cero e infinito.

Teorema 2.4.3 *Bajo las hipótesis (KLH.1)-(KLH.3), y si $\lim_{n \rightarrow \infty} nh_1^5 = a > 0$, entonces,*

$$(nh_1)^{1/2} \{ \hat{\eta}_1(x_1) - \eta_1(x_1) \} \rightarrow \mathcal{N}(b_1(x_1), v_1^2(x_1)),$$

donde

$$b_1(x_1) = \sqrt{a} \frac{\mu_2(K_1)}{2} \left\{ m_1''(x_1) - \int \frac{m(x_1, \mathbf{x}_2) f_2(\mathbf{x}_2)}{f(x_1, \mathbf{x}_2)} \frac{\partial^2 f(x_1, \mathbf{x}_2)}{\partial x_1^2} d\mathbf{x}_2 \right\},$$

y

$$v_1^2(x_1) = R(K_1) \int \frac{f_2^2(\mathbf{x}_2)}{f(\mathbf{x})} \sigma^2(\mathbf{x}) d\mathbf{x}_2.$$

2.5. Estimador eficiente de Linton

Linton (1997) define un procedimiento de estimación para un modelo aditivo, que consiste en usar el método de integración marginal de Linton y Nielsen (1995) para obtener unos valores iniciales, que se utilizarán para aplicar un ciclo del algoritmo *backfitting* de Hastie y Tibshirani (1990). Este método es eficiente, en el sentido de que tiene menor error cuadrático medio que el método de integración marginal aplicado de forma individual; en concreto, esta técnica reduce el sesgo.

2.5.1. Definición del estimador

Se define el método para un modelo aditivo bivalente (para facilitar la presentación), es decir,

$$m(\mathbf{x}) = \alpha + m_1(x_1) + m_2(x_2).$$

En primer lugar, se aplica el método de integración marginal, por lo que se considera un estimador piloto bivalente de la función de regresión, $m(\mathbf{x})$, que al igual que en Linton y Nielsen (1995), será el estimador lineal local⁹. Entonces el problema que se plantea es:

Minimizar $\hat{\theta}_0$ y $\hat{\theta}_1$ en la expresión:

$$\sum_{i=1}^n K_{1,h_1} \left(\frac{X_{1i} - x_1}{h_1} \right) K_{2,h_2} \left(\frac{X_{2i} - x_2}{h_2} \right) \times \\ \times (Y_i - \theta_0 - \theta_1 (X_{1i} - x_1) - \theta_2 (X_{2i} - x_2))^2,$$

donde h_1 y h_2 son los anchos de banda escalares, K_1 y K_2 son funciones núcleo, y $\hat{m}^{pi}(x) = \hat{\theta}_0(x)$.

A continuación se estiman las componentes univariantes utilizando la estimación piloto obtenida por el estimador lineal local.

La expresión para $m_1(x_1)$ será:

$$\tilde{m}_1(x_1) = n^{-1} \sum_{i=1}^n \hat{m}^{pi}(x_1, X_{2i}) - \tilde{\alpha}, \quad (2.11)$$

donde $\tilde{\alpha} = n^{-1} \sum_{i=1}^n Y_i$, y

$$\tilde{m}(x) = \tilde{\alpha} + \tilde{m}_1(x_1) + \tilde{m}_2(x_2). \quad (2.12)$$

Bajo condiciones adecuadas (Linton y Nielsen (1995), Linton y Härdle (1996)), y suponiendo que $h_1 = h_2 = O(n^{-1/5})$, se verifica:

$$n^{2/5} \{ \tilde{m}_1(x_1) - m_1(x_1) \} \rightarrow \mathcal{N}(b_1(x_1), v_1^2(x_1)),$$

donde

$$b_1(x_1) = \frac{1}{2} \mu_2(K_1) m_1''(x_1),$$

y

$$v_1^2(x_1) = R(K_1) \sigma^2 \int f^{-1}(x) f_2^2(x_2) dx_2.$$

La estimación de la función de regresión, $\tilde{m}(x)$, es también asintóticamente normal con razón de convergencia $n^{2/5}$.

El procedimiento definido en (2.11) y (2.12) recibe el nombre de **método de integración empírico**.

⁹Fan (1992).

La segunda etapa del método se basa en la siguiente idea: suponemos que se conocen α y $m_2(\cdot)$. Entonces, se puede estimar $m_1(x_1)$ por un suavizador de regresión unidimensional, aplicado a los errores parciales que se obtienen de la siguiente expresión,

$$U_{1i} = Y_i - \alpha - m_2(X_{2i}),$$

sobre X_{1i} . Sean $\hat{\phi}_0$ y $\hat{\phi}_1$ los valores que minimizan:

$$\sum_{i=1}^n J_g \left(\frac{X_{1i} - x_1}{g} \right) \{U_{1i} - \phi_0 - \phi_1(X_{1i} - x_1)\}^2,$$

donde J es una función núcleo, g es el ancho de banda, y $\hat{m}_1^{opt}(x_1) = \hat{\phi}_0(x_1)$. Este estimador es lineal en la variable dependiente, es decir,

$$\hat{m}_1^{opt}(x_1) = \sum_{i=1}^n W_i(x_1; g) U_{1i}, \quad (2.13)$$

para $\{W_i(x_1; g)\}_{i=1}^n$, un conjunto de pesos que dependen del diseño. Bajo las condiciones clásicas de regularidad¹⁰, que incluyen $g = O(n^{-1/5})$, se cumple que:

$$n^{2/5} \{ \hat{m}_1^{opt}(x_1) - m_1(x_1) \} \rightarrow \mathcal{N} \left(b_1^{opt}(x_1), (v_1^{opt})^2(x_1) \right),$$

donde

$$b_1^{opt}(x_1) = \frac{1}{2} \mu_2(J) m_1''(x_1),$$

y

$$(v_1^{opt})^2(x_1) = \frac{R(J) \sigma^2}{f_1(x_1)}.$$

Aunque el sesgo de este procedimiento coincide con el sesgo obtenido en el método de integración marginal empírico, $\hat{m}_1^{opt}(x_1)$ mejora cualquier versión del método de integración marginal en términos de la varianza. A partir de ahora, un estimador que tenga este sesgo y varianza se llamará **estimador óptimo**.

Entonces, se ha establecido que el estimador $\hat{m}_1^{opt}(x_1)$ es más eficiente que $\tilde{m}_1(x_1)$ y la diferencia puede ser bastante elevada cuando el diseño presenta correlaciones altas o cuando los puntos en los que se está estimando están en la frontera. Por este motivo, Linton (1997) propone una modificación de $\tilde{m}_1(x_1)$ para obtener un resultado asintóticamente equivalente a las estimaciones $\hat{m}_1^{opt}(x_1)$.

Para ello se definen los residuos parciales,

$$\tilde{U}_{1i} = Y_i - \tilde{\alpha} - \tilde{m}_2(X_{2i}),$$

¹⁰Härdle (1990).

y se suavizan sobre X_{1i} . Entonces,

$$\tilde{m}_1(x_1) = \sum_{i=1}^n W_i(x_1; g) \tilde{U}_{1i},$$

donde $W_i(x_1; g)$ están definidos igual que en (2.13). Entonces, $\tilde{m}_1(x_1)$ será la estimación revisada de $m_1(x_1)$.

2.5.2. Propiedades del estimador

El siguiente conjunto de hipótesis es necesario para el teorema que se expone a continuación:

(L.1) Las funciones m_1, m_2 y todas las funciones de densidad tienen dos derivadas continuas.

(L.2) Las densidades f_1, f_2 y f tienen soporte compacto, están acotadas, son continuas Lipschitz y están acotadas inferiormente por cero.

(L.3) Los errores, $Y - m(\mathbf{X})$, condicionados sobre \mathbf{X} , son independientes e idénticamente distribuidos con varianza finita y positiva, σ^2 .

(L.4) Las funciones núcleo, K_1, K_2 y J , son densidades acotadas, con soporte compacto, continuas Lipschitz y simétricas respecto de cero.

(L.5) Los anchos de banda verifican, $nh_1^5, nh_2^5 \rightarrow 0$, y $nh_1h_2 \rightarrow \infty$, cuando:

$$0 < \liminf g^5 n \leq \limsup g^5 n < \infty.$$

Teorema 2.5.1 *Bajo las hipótesis (L.1)-(L.5), y si se verifica $h_1, h_2 = o(n^{-1/5})$ y $g = O(n^{-1/5})$, entonces,*

$$n^{2/5} \{ \tilde{m}_1(x_1) - \hat{m}_1^{opt}(x_1) \} \rightarrow 0,$$

en probabilidad.

2.6. Estimador de proyección-*backfitting*

Este estimador se define como la mejor aproximación del estimador de Nadaraya-Watson sobre el espacio de funciones aditivas. Además, el estimador se puede interpretar como un estimador *backfitting* definido por una solución iterada de las ecuaciones empíricas.

Se presenta también la distribución asintótica y un resultado de convergencia uniforme para los estimadores. Este procedimiento tiene el mismo sesgo y varianza que el estimador óptimo, basado en el conocimiento de las otras componentes; y en este sentido mejora el método analizado en Opsomer y Ruppert (1997). Las condiciones que se imponen son más débiles que las que se imponen en dicho trabajo y, en particular, no restringen la dependencia entre las variables.

2.6.1. Definición del estimador

En este apartado se trata una estimación *backfitting*, para suavizadores de regresión de Nadaraya-Watson, que utiliza como función núcleo multivariante una función producto, es decir,

$$\hat{m}_{NW}(\mathbf{x}) = \frac{\sum_{i=1}^n \prod_{j=1}^D K_h(X_{ji} - x_j) Y_i}{\sum_{i=1}^n \prod_{j=1}^D K_h(X_{ji} - x_j)} = \frac{\hat{r}(\mathbf{x})}{\hat{f}(\mathbf{x})}. \quad (2.14)$$

Sea un modelo aditivo de dimensión D . El algoritmo *backfitting* de Hastie y Tibshirani (1990) surge como solución de la versión empírica del siguiente conjunto de ecuaciones:

$$\begin{aligned} m_1(x_1) &= E[Y | X_1 = x_1] - \alpha - E[m_2(X_2) | X_1 = x_1] - \\ &\quad - \cdots - E[m_D(X_D) | X_1 = x_1], \\ &\quad \vdots \\ m_D(x_D) &= E[Y | X_D = x_D] - \alpha - E[m_2(X_2) | X_D = x_D] - \\ &\quad - \cdots - E[m_{D-1}(X_{D-1}) | X_D = x_D]. \end{aligned}$$

Cuando tenemos información muestral, las esperanzas condicionadas son reemplazadas por suavizadores unidimensionales, \hat{m}_j , y se itera el proceso a partir de unos valores iniciales para m_j .

Sea $\hat{f}(\mathbf{x}) = n^{-1} \sum_{i=1}^n \prod_{j=1}^D K_h(X_{ji} - x_j)$, el estimador núcleo multidimensional de la función de densidad, y sea $\hat{m}_{NW}(\mathbf{x})$ el estimador de Nadaraya-Watson multidimensional, definido como en (2.14). Entonces se definen las estimaciones de proyección empíricas, $\{\hat{m}_j, j = 1, \dots, D, \text{ y } \hat{\alpha}\}$, como los mínimos del siguiente criterio:

$$\|\hat{m}_{NW} - \bar{m}\|_{\hat{f}}^2 = \int \{\hat{m}_{NW}(\mathbf{x}) - \hat{\alpha} - \hat{m}_1(x_1) - \cdots - \hat{m}_D(x_D)\}^2 \hat{f}(\mathbf{x}) dx. \quad (2.15)$$

La minimización se hace sobre todas las funciones, $\tilde{m}(\mathbf{x})$, $\tilde{m}(\mathbf{x}) = \tilde{\alpha} + \tilde{m}_1(x_1) + \dots + \tilde{m}_D(x_D)$, con $\int \tilde{m}_j(x_j) \hat{f}_j(x_j) dx_j = 0$, y donde $\hat{f}_j(x_j)$ es la función de densidad marginal estimada de $\hat{f}(\mathbf{x})$:

$$\hat{f}_j(x_j) = n^{-1} \sum_{i=1}^n K_h(X_{ji} - x_j).$$

El mínimo de (2.15), existe si la estimación de la función de densidad, $\hat{f}(\mathbf{x})$, es no negativa. La solución de (2.15),

$$\tilde{m}(\mathbf{x}) = \tilde{\alpha} + \tilde{m}_1(x_1) + \dots + \tilde{m}_D(x_D),$$

es la proyección en el espacio $L_2(\hat{f})$ de \hat{m} sobre el subespacio de funciones aditivas, $\{m \in L_2(\hat{f}) : m(\mathbf{x}) = \alpha + m_1(x_1) + \dots + m_D(x_D)\}$.

La solución de (2.15) está caracterizada por el siguiente sistema de ecuaciones ($j = 1, \dots, D$):

$$\begin{aligned} \tilde{m}_j(x_j) &= \int \hat{m}_{NW}(\mathbf{x}) \frac{\hat{f}(\mathbf{x})}{\hat{f}_j(x_j)} d\mathbf{x}_{(-j)} - \tilde{\alpha} - \sum_{k \neq j} \int \tilde{m}_k(x_k) \frac{\hat{f}(\mathbf{x})}{\hat{f}_j(x_j)} d\mathbf{x}_{(-j)}, \quad (2.16) \\ 0 &= \int \tilde{m}_j(x_j) \hat{f}_j(x_j) dx_j. \end{aligned}$$

Donde en la primera ecuación:

$$\int \hat{m}_{NW}(\mathbf{x}) \frac{\hat{f}(\mathbf{x})}{\hat{f}_j(x_j)} d\mathbf{x}_{(-j)} = \frac{n^{-1} \sum_{i=1}^n K_h(X_{ji} - x_j) Y_i}{\hat{f}_j(x_j)} \equiv \hat{m}_{j,NW}(x_j),$$

dado que

$$\int \prod_{l \neq j} K_h(X_{li} - x_l) d\mathbf{x}_{(-j)} = 1.$$

Entonces, $\hat{m}_{j,NW}(x_j)$ es el correspondiente estimador univariante de Nadaraya-Watson de la componente $m_j(x_j)$. Y,

$$\tilde{\alpha} = \int \hat{m}_{NW}(\mathbf{x}) \hat{f}(\mathbf{x}) d\mathbf{x} = n^{-1} \sum_{i=1}^n Y_i.$$

Además,

$$\tilde{\alpha} = \int \hat{m}_{j,NW}(x_j) \hat{f}_j(x_j) dx_j, \text{ para } j = 1, \dots, D. \quad (2.17)$$

Se define $\tilde{m}_j(x_j)$, $j = 1, \dots, D$, que recibe el nombre de **estimador backfitting**, como la solución del sistema de ecuaciones ($j = 1, \dots, D$):

$$\begin{aligned} \tilde{m}_j(x_j) &= \hat{m}_{j,NW}(x_j) - \tilde{\alpha} - \sum_{k \neq j} \int \tilde{m}_k(x_k) \frac{\hat{f}(x)}{\hat{f}_j(x_j)} dx_{(-j)}, \\ 0 &= \int \tilde{m}_j(x_j) \hat{f}_j(x_j) dx_j. \end{aligned}$$

con $\tilde{\alpha}$ dado en (2.17).

En este proceso, la hipótesis de que las estimaciones multivariantes de la función de densidad y de la función de regresión existen para todo \mathbf{x} , no es necesaria, ya que la expresión (2.16) se puede escribir de la siguiente forma:

$$\tilde{m}_j(x_j) = \hat{m}_{j,NW}(x_j) - \tilde{\alpha} - \sum_{k \neq j} \int \tilde{m}_k(x_k) \frac{\hat{f}_{jk}(x_j, x_k)}{\hat{f}_j(x_j)} dx_k, \quad (2.18)$$

donde

$$\hat{f}_{jk}(x_j, x_k) = n^{-1} \sum_{i=1}^n K_h(X_{ji} - x_j) K_h(X_{ki} - x_k),$$

es la función de densidad marginal bidimensional de la estimación núcleo de la función de densidad D -dimensional, $\hat{f}(\mathbf{x})$. En esta expresión solo aparecen funciones marginales unidimensionales y bidimensionales, y por eso este procedimiento resuelve el problema de la maldición de la dimensionalidad.

Aún no se ha indicado ninguna condición sobre el soporte de \mathbf{X} . Se considera ahora una generalización del método que trata el efecto frontera que se produce cuando las funciones de densidad tienen soporte compacto. En tal caso no será necesario que se verifique (2.17), es decir, $\int \hat{m}_{j,NW}(x_j) \hat{f}_j(x_j) dx_j$ dependerá de j . Además, \hat{f}_j no tiene que ser una función de probabilidad, y tampoco tiene que ser la función marginal de \hat{f}_{jk} , es decir, puede que no se verifique que si $j \neq k$:

$$\hat{f}_j(x_j) = \int \hat{f}_{jk}(x_j, x_k) dx_k. \quad (2.19)$$

Esta situación puede ocurrir con las estimaciones núcleo de funciones de densidad con soporte compacto. Para este caso más general se trata de encontrar una modificación apropiada de (2.18).

Se expresa ahora de la siguiente forma:

$$\tilde{m}_j(x_j) = \hat{m}_{j,NW}(x_j) - \tilde{\alpha}_j - \sum_{k \neq j} \int \tilde{m}_k(x_k) \frac{\hat{f}_{jk}(x_j, x_k)}{\hat{f}_j(x_j)} dx_k, \quad (2.20)$$

donde $\tilde{\alpha}_j$ se elige de modo que se cumpla:

$$\int \tilde{m}_j(x_j) \hat{f}_j(x_j) dx_j = 0, \quad \forall j.$$

Bajo las hipótesis (2.17), (2.19) y $\int \hat{f}_j(x_j) dx_j = 1$, se obtiene (2.18). En general, (2.20) se puede expresar como:

$$\begin{aligned} \tilde{m}_j(x_j) = & \hat{m}_{j,NW}(x_j) - \tilde{\alpha}_j - \\ & - \sum_{k \neq j} \int \tilde{m}_k(x_k) \left(\frac{\hat{f}_{jk}(x_j, x_k)}{\hat{f}_j(x_j)} - \hat{f}_{k,[j+]}(x_k) \right) dx_k, \end{aligned} \quad (2.21)$$

donde,

$$\tilde{\alpha}_j = \frac{\int \hat{m}_j(x_j) \hat{f}_j(x_j) dx_j}{\int \hat{f}_j(x_j) dx_j},$$

y para $k \neq j$,

$$\hat{f}_{k,[j+]}(x_k) = \frac{\int \hat{f}_{jk}(x_j, x_k) dx_j}{\int \hat{f}_j(x_j) dx_j}.$$

En la práctica, este algoritmo *backfitting* consiste en seguir el siguiente proceso:

Se comienza con unos valores iniciales arbitrarios, $\tilde{m}_j^{[0]}$, para \tilde{m}_j ; como por ejemplo, $\tilde{m}_j^{[0]} = \hat{m}_{j,NW}$ o los obtenidos utilizando el estimador de integración marginal de Linton y Nielsen (1995). En el paso j -ésimo de la iteración r -ésima se aplica la siguiente fórmula:

$$\begin{aligned} \tilde{m}_j^{[r]}(x_j) = & \hat{m}_j(x_j) - \tilde{\alpha}_j - \\ & - \sum_{k < j} \int \tilde{m}_k^{[r]}(x_k) \left(\frac{\hat{f}_{jk}(x_j, x_k)}{\hat{f}_j(x_j)} - \hat{f}_{k,[j+]}(x_k) \right) dx_k - \\ & - \sum_{k > j} \int \tilde{m}_k^{[r-1]}(x_k) \left(\frac{\hat{f}_{jk}(x_j, x_k)}{\hat{f}_j(x_j)} - \hat{f}_{k,[j+]}(x_k) \right) dx_k, \end{aligned} \quad (2.22)$$

y este proceso se repite hasta que se verifique el criterio de convergencia deseado.

Un procedimiento análogo se puede hacer considerando en lugar del estimador de Nadaraya-Watson, el estimador lineal local, y calculando la mejor aproximación de este estimador sobre el espacio de funciones aditivas. No presentamos aquí los resultados por ser similares a los anteriores. Más detalles se pueden encontrar en Mammen, Linton y Nielsen (1999).

2.6.2. Existencia de solución

Se consideran las estimaciones, \tilde{m}_j , que están definidas en (2.21), donde \hat{m}_j , \hat{f}_{jk} y \hat{f}_j son estimaciones conocidas. En el siguiente teorema, bajo condiciones adecuadas¹¹, con probabilidad que tiende a 1, existe una solución \tilde{m}_j de (2.21), única y que se calcula utilizando el algoritmo *backfitting*.

Teorema 2.6.1 *Cuando $n \rightarrow \infty$, existe solución única \tilde{m}_j para (2.21), con probabilidad uno. Además, existen constantes, γ y c , tales que se verifica la siguiente desigualdad:*

$$\int \left\{ \tilde{m}_j^{[r]}(x_j) - \tilde{m}_j(x_j) \right\}^2 f_j(x_j) dx_j \leq \\ \leq c\gamma^{2r} \left\{ 1 + \sum_{j=1}^D \int \left(\tilde{m}_j^{[0]}(x_j) \right)^2 f_j(x_j) dx_j \right\},$$

con probabilidad uno. Aquí, $\tilde{m}_j^{[0]}(x_j)$, $j = 1, \dots, D$, son las funciones iniciales del algoritmo *backfitting*. Para $r > 0$, las funciones, $\tilde{m}_1^{[r]}(x_1), \dots, \tilde{m}_D^{[r]}(x_D)$, están definidas en (2.22).

Se demuestra que los términos del sesgo de las estimaciones *backfitting* se calculan por las proyecciones del sesgo teórico de la estimación D -dimensional de Nadaraya-Watson.

¹¹Ver Mammen, Linton y Nielsen (1999).



Capítulo 3

Selección del ancho de banda

3.1. Introducción

En la mayoría de los procedimientos de regresión no paramétrica intervienen parámetros que hay que determinar de forma óptima con el fin de obtener buenas estimaciones. Entre ellos están la función núcleo, K , y el ancho de banda, h .

La elección de la función núcleo resulta un problema de poca dificultad resuelto habitualmente utilizando elecciones que infieran buenas propiedades al estudio; no obstante la selección de h es crucial para el buen comportamiento del suavizador, \hat{m} .

El problema de selección del ancho de banda ha sido muy desarrollado en la literatura. En las últimas décadas se han propuesto muchas técnicas para la selección del parámetro de suavizamiento en diferentes contextos, fundamentalmente dentro de la estimación no paramétrica de densidades, y otras específicamente para la regresión no paramétrica.

La selección de una matriz de anchos de banda en el caso multivariante, \mathbf{H} , a partir de los datos, ha sido mucho menos considerado en la literatura que el problema en el caso univariante, en el que interviene un único ancho de banda. Sin embargo, muchas de las ideas del problema en el caso univariante se pueden extender al caso multivariante.

En general, la matriz \mathbf{H} , simétrica de dimensión $D \times D$, tiene $D(D+1)/2$ elementos diferentes lo cual, aún incluso cuando D no es muy elevado, supone la elección de un número considerable de parámetros de suavizamiento.

Algunos casos particulares de la matriz \mathbf{H} simplifican de alguna forma este problema. Sea \mathcal{F} la clase de matrices $D \times D$, simétricas, definidas positivas; se obtiene una simplificación imponiendo la restricción $\mathbf{H} \in \mathcal{D}$, donde $\mathcal{D} \subseteq \mathcal{F}$ es la subclase de matrices $D \times D$, diagonales, definidas positivas. Obsérvese que suponer una matriz de anchos, $\mathbf{H} \in \mathcal{D}$, $\mathbf{H} = \text{diag}\{h_1^2, \dots, h_D^2\}$, significa considerar un ancho de banda independiente para cada una de las variables.

Una matriz aún más simple se obtiene si consideramos la restricción $\mathbf{H} \in \mathcal{S}$, donde $\mathcal{S} = \{h^2\mathbf{I} : h > 0\}$, lo que supondrá elegir un único ancho de banda para todas las variables.

En el caso de los modelos aditivos el problema se puede plantear en los dos últimos contextos analizados, es decir, considerando un ancho de banda diferente para cada una de las componentes aditivas o un único parámetro común para todas ellas.

3.2. Métodos para la selección del ancho de banda

Un método que utilice el conjunto de observaciones para obtener un ancho de banda, \hat{h} , recibe el nombre de **selector de ancho de banda**.

Introducimos a continuación dos procedimientos muy utilizados, en el contexto de regresión no paramétrica. Un primer procedimiento de tipo *plug-in* que parte de alguna expresión teórica del error cuadrático medio y propone estimadores de los parámetros desconocidos que intervengan, obteniendo así el ancho de banda óptimo como el mínimo de dicha expresión estimada. Y un segundo método que se basa en la minimización de alguna función que aproxime, a partir de los datos, el error del estimador.

Para cada uno de ellos establecemos, en primer lugar, la definición en el caso univariante y a continuación, proporcionamos las directrices para extenderlo al caso multivariante.

3.2.1. Selector *plug-in*

Se utiliza como estimador de regresión no paramétrica, el estimador lineal local.

Por simplicidad suponemos que los errores son homocedásticos con varianza σ^2 , y que los X_i proceden de densidades con soporte compacto en $[0, 1]$.

Consideramos el criterio de error, MISE, para el estimador lineal local, $\hat{m}_1(x; h)$, dado por la siguiente expresión:

$$\text{MISE}[\hat{m}_1(\cdot; h) | \mathbf{X}] = \mathbb{E} \left[\int \{\hat{m}_1(x; h) - m(x)\}^2 f(x) dx | \mathbf{X} \right].$$

Según este criterio, el ancho de banda asintóticamente óptimo¹ será:

$$h_{\text{AMISE}} = \left(\frac{R(K) \sigma^2}{n \mu_2(K)^2 \theta_{2,2}} \right)^{1/5},$$

¹Wand y Jones (1995), p. 138.

donde $\theta_{2,2}$ es un caso particular de la función:

$$\theta_{r,s} = \int m^{(r)} m^{(s)} f(x) dx.$$

La expresión dada para el ancho de banda depende de $\theta_{2,2}$ y σ^2 , cantidades desconocidas, que es necesario estimar. Consideramos como estimador para $\theta_{2,2}$ el siguiente:

$$\hat{\theta}_{2,2}(g) = n^{-1} \sum_{i=1}^n \left(\hat{m}_3^{(2)}(X_i; g) \right)^2,$$

donde g es un ancho de banda (diferente de h), y para σ^2 ,

$$\hat{\sigma}^2(k) = v^{-1} \sum_{i=1}^n \{Y_i - \hat{m}(X_i; k)\}^2,$$

siendo k otro ancho de banda, y donde

$$v = n - 2 \sum_{i=1}^n w_{ii} + \sum_{i=1}^n \sum_{j=1}^n w_{ij}^2,$$

y

$$w_{ij} = e_1^T (\mathbf{X}_{X_i}^T W_{X_i,k} \mathbf{X}_{X_i})^{-1} \mathbf{X}_{X_i}^T W_{X_i,k} e_j.$$

Entonces, se define el selector *plug-in* del ancho de banda como:

$$\hat{h}_{\text{DPI}} = \left(\frac{R(K) \hat{\sigma}^2(k)}{n \mu_2(K)^2 \hat{\theta}_{2,2}(g)} \right)^{1/5}.$$

Sin embargo, esta fórmula requiere la elección de anchos de banda preliminares, g y k . Para elegir estos parámetros se obtiene el error cuadrático medio de cada uno de los estimadores, y se elegirán \hat{g} y \hat{k} , que minimicen la aproximación asintótica de los MSE de cada estimador.

Este método se puede extender al caso multivariante utilizando la misma idea, y considerando ahora las expresiones para el sesgo y la varianza dadas en el teorema 1.3.2, con las que obtener el criterio de error, MISE. La siguiente ecuación determina la elección óptima de la matriz ancho de banda.

Si la matriz Hessiana de m , $\mathcal{H}(\mathbf{x})$, es definida positiva o negativa, existe una única solución dada por:

$$\mathbf{H}\mathbf{H}^T = \left\{ \frac{R(\mathbf{K}) \sigma^2 |\mathcal{H}^*|^{1/2}}{\mu_2^2(\mathbf{K}) n f'(\mathbf{x})} \right\}^{2/(D+4)} (\mathcal{H}^*)^{-1}, \quad (3.1)$$

con

$$\mathcal{H}^* = \begin{cases} \mathcal{H} & \text{para } \mathcal{H} \text{ definida positiva} \\ -\mathcal{H} & \text{para } \mathcal{H} \text{ definida negativa} \end{cases}$$

Se elegirá la matriz $\hat{\mathbf{H}}$, ancho de banda, que verifique la ecuación (3.1).

3.2.2. Validación cruzada por mínimos cuadrados

Sea $\hat{m}(x; h)$, el estimador no paramétrico considerado para la función de regresión univariante. Para cada i ($i = 1, \dots, n$), denotaremos por $\hat{m}_{(-i)}(x; h)$, al estimador no paramétrico calculado usando todas las observaciones excepto la i -ésima. La validación del modelo se puede hacer examinando el error de predicción,

$$Y_i - \hat{m}_{(-i)}(X_i; h).$$

La técnica de validación cruzada por mínimos cuadrados promedia los errores de predicción al cuadrado, esto es

$$CV(h) = n^{-1} \sum_{i=1}^n \{Y_i - \hat{m}_{(-i)}(X_i; h)\}^2,$$

lo cual constituye una medida global de la efectividad de $\hat{m}(\cdot; h)$ como estimador de $m(\cdot)$ sobre los puntos de la muestra considerada.

Entonces, se define el selector del ancho de banda por validación cruzada como

$$\hat{h}_{CV} = \arg \min_h CV(h).$$

La extensión de este método al caso multivariante, se hace de forma natural. Se considera entonces, el siguiente criterio:

$$CV(\mathbf{H}) = n^{-1} \sum_{i=1}^n \{Y_i - \hat{m}_{(-i)}(\mathbf{X}_i, \mathbf{H})\}^2,$$

y el selector de la matriz ancho de banda por validación cruzada será:

$$\hat{\mathbf{H}}_{CV} = \arg \min_{\mathbf{H} \in \mathcal{F}} CV(\mathbf{H}).$$

Se podría minimizar $CV(\mathbf{H})$ también sobre $\mathbf{H} \in \mathcal{D}$ o sobre $\mathbf{H} \in \mathcal{S}$, para obtener selectores de anchos de banda que pertenezcan a las clases menores.

3.3. Selectores del ancho de banda en los modelos aditivos

La mayoría de los selectores propuestos para los modelos aditivos están basados en un método de validación cruzada o en una de sus aproximaciones; a pesar de su carácter intuitivo y su simplicidad esta aproximación tiene dos inconvenientes:

el primero, se refiere a las propiedades de los estimadores del ancho de banda, en concreto a la lenta razón de convergencia que presentan. El segundo inconveniente, está en que, desde un punto de vista práctico, este selector del ancho de banda tiene un intenso cálculo computacional.

Por esta razón, Opsomer y Ruppert (1998) y Opsomer (2000) proponen utilizar un selector *plug-in* del ancho de banda para los modelos aditivos, que se puede aplicar cuando se utiliza el algoritmo *backfitting* de Hastie y Tibshirani (1990). En el capítulo anterior se han proporcionado el sesgo y la varianza asintóticos para un modelo aditivo estimado utilizando el algoritmo *backfitting*; se trata entonces de establecer el sistema teórico necesario para desarrollar este método.

3.3.1. Selector *plug-in*

En primer lugar se presenta el selector para un modelo aditivo bivalente ajustado por regresión lineal local. El ajuste de este modelo aditivo requiere la elección de los anchos de banda para m_1 y m_2 , que denotamos por h_1 y h_2 , respectivamente. La elección óptima para estos parámetros será el mínimo del error cuadrático medio promediado (MASE), dado en la siguiente expresión:

$$\text{MASE}[h_1, h_2 | X_1, X_2] = \frac{1}{n} \sum_{i=1}^n \text{E} [\widehat{m}(X_{1i}, X_{2i}) - m(X_{1i}, X_{2i}) | X_1, X_2]^2.$$

A partir de los resultados teóricos que se presentan en Opsomer y Ruppert (1997) y que se recogen en el teorema 2.7, se obtiene que cuando el modelo aditivo se estima mediante regresión lineal local, la aproximación asintótica de MASE, denotada por AMASE es:

$$\begin{aligned} \text{AMASE}[h_1, h_2 | X_1, X_2] &= \frac{\mu_2(K)^2}{4} \{h_1^4 \theta_{1,1} + h_1^2 h_2^2 \theta_{1,2} + h_2^4 \theta_{2,2}\} + \\ &+ \sigma^2 R(K) \left\{ \frac{b_1 - a_1}{nh_1} + \frac{b_2 - a_2}{nh_2} \right\}, \end{aligned}$$

donde (a_i, b_i) , $i = 1, 2$ es el rango de X_i , y

$$\begin{aligned} \theta_{1,1} &= \frac{1}{n} \sum_{i=1}^n \left\{ \mathbf{t}_i^T D^2 \mathbf{m}_1 - \mathbf{v}_i^T \text{E} \left[m_1^{(2)}(X_{1i}) | X_2 \right] \right\}^2, \\ \theta_{2,2} &= \frac{1}{n} \sum_{i=1}^n \left\{ \mathbf{v}_i^T D^2 \mathbf{m}_2 - \mathbf{t}_i^T \text{E} \left[m_2^{(2)}(X_{2i}) | X_1 \right] \right\}^2, \\ \theta_{1,2} &= \frac{1}{n} \sum_{i=1}^n \left\{ \mathbf{t}_i^T D^2 \mathbf{m}_1 - \mathbf{v}_i^T \text{E} \left[m_1^{(2)}(X_{1i}) | X_2 \right] \right\} \times \\ &\quad \times \left\{ \mathbf{v}_i^T D^2 \mathbf{m}_2 - \mathbf{t}_i^T \text{E} \left[m_2^{(2)}(X_{2i}) | X_1 \right] \right\}. \end{aligned}$$

Los valores de los parámetros que minimizan esta función se denotarán por $h_{1,AMASE}$ y $h_{2,AMASE}$. La función MASE es más simple que la MSE, pero la minimización de esta expresión no resuelve el problema dado porque depende de $\theta_{1,1}$, $\theta_{2,2}$, $\theta_{1,2}$ y σ^2 , cantidades desconocidas. Entonces, se definen los selectores *plug-in*, \hat{h}_1 y \hat{h}_2 , para estos parámetros como el resultado de realizar el siguiente proceso en tres etapas:

Paso 1. Estimar $\theta_{1,1}$, $\theta_{2,2}$, $\theta_{1,2}$ y σ^2 por $\hat{\theta}_{1,1}$, $\hat{\theta}_{2,2}$, $\hat{\theta}_{1,2}$ y $\hat{\sigma}^2$.

Paso 2. Utilizar $\hat{\theta}_{1,1}$, $\hat{\theta}_{2,2}$, $\hat{\theta}_{1,2}$ y $\hat{\sigma}^2$ para calcular MASE.

Paso 3. Encontrar los valores $h_1, h_2 > 0$ que minimizan MASE.

Si para calcular las estimaciones en el paso 1 se utiliza la regresión no paramétrica, entonces se necesitan nuevos parámetros de suavizamiento para cada una de las estimaciones. Por tanto, también habrá que calcular el error cuadrático medio de estos estimadores. De modo que, en general, no hay expresiones explícitas para \hat{h}_1 y \hat{h}_2 .

Cuando X_1 y X_2 son independientes, el problema de selección del ancho de banda se simplifica, ya que en la expresión dada para AMASE, $\theta_{1,2} = 0$, y tendremos que

$$\theta_{1,1} = \frac{1}{n} \sum_{i=1}^n \left\{ m_1^{(2)}(X_{1i}) - E \left[m_1^{(2)}(X_{1i}) \right] \right\}^2,$$

y

$$\theta_{2,2} = \frac{1}{n} \sum_{i=1}^n \left\{ m_2^{(2)}(X_{2i}) - E \left[m_2^{(2)}(X_{2i}) \right] \right\}^2.$$

Bajo la hipótesis de independencia, los mínimos de AMASE tienen las siguientes expresiones:

$$h_{1,AMASE} = \left(\frac{\sigma^2 R(K) (b_{x_1} - a_{x_1})}{n \mu_2(K)^2 \theta_{1,1}} \right)^{1/5},$$

y

$$h_{2,AMASE} = \left(\frac{\sigma^2 R(K) (b_{x_2} - a_{x_2})}{n \mu_2(K)^2 \theta_{2,2}} \right)^{1/5}.$$

Estimación de funciones de regresión

Se trata de encontrar estimadores para funciones como la siguiente:

$$\theta_{1,1}(r, s) = \frac{1}{n} \sum_{i=1}^n \left\{ \mathbf{t}_i^T D^r \mathbf{m}_1 - \mathbf{v}_i^T E \left[m_1^{(r)}(X_{1i}) | X_2 \right] \right\} \times \\ \times \left\{ \mathbf{t}_i^T D^s \mathbf{m}_1 - \mathbf{v}_i^T E \left[m_1^{(s)}(X_{1i}) | X_2 \right] \right\},$$

y de forma similar para $\theta_{2,2}(r, s)$ y $\theta_{1,2}(r, s)$, con $r, s \geq 1$, y $r + s$ par. Obsérvese que, $\theta_{k,l}(2, 2) \equiv \theta_{k,l}$, los parámetros que aparecen en la sección anterior.

Sean $s_{1,x_1}^{(r)}$ y $s_{2,x_2}^{(r)}$, los suavizadores polinómicos locales para la r -ésima derivada en las observaciones x_1 y x_2 . Para x_1 , este suavizador tiene la siguiente expresión:

$$\left(s_{1,x_1}^{(r)} \right)^T = r! e_{r+1}^T \left(\mathbf{X}_{x_1}^T \mathbf{W}_{x_1, h_1} \mathbf{X}_{x_1} \right)^{-1} \mathbf{X}_{x_1}^T \mathbf{W}_{x_1, h_1},$$

donde e_{r+1}^T es un vector de dimensión $(p+1) \times 1$ con un 1 en la coordenada $r+1$ y 0, en el resto. Una expresión similar se tiene para $s_{2,x_2}^{(r)}$. Sean $S_1^{(r)}$ y $S_2^{(r)}$ las matrices cuyas filas son los suavizadores derivados en las observaciones de X_1 y X_2 .

Se definen los vectores de estimadores de regresión polinómicos locales de las funciones derivadas, $m_1^{(r)}$ y $m_2^{(r)}$, con respecto a x_1 y x_2 :

$$\hat{\mathbf{m}}_1^{(r)} = S_1^{(r)} (\mathbf{Y} - \hat{\mathbf{m}}_2),$$

y

$$\hat{\mathbf{m}}_2^{(r)} = S_2^{(r)} (\mathbf{Y} - \hat{\mathbf{m}}_1),$$

donde $\hat{\mathbf{m}}_1$ y $\hat{\mathbf{m}}_2$ se estimarán mediante un modelo aditivo con polinomios locales de grado p .

Sea $\mathbf{T}_{21}^* = (\mathbf{T}_{12}^*)^T$, entonces $\theta_{1,1}(r, s)$ se puede expresar del siguiente modo:

$$\theta_{1,1}(r, s) = \frac{1}{n} \text{tr} \left\{ \left((\mathbf{I} - \mathbf{T}_{12}^*)^{-1} D^r \mathbf{m}_1 - (\mathbf{I} - \mathbf{T}_{21}^*)^{-1} E \left[m_1^{(r)}(X_{1i}) | X_2 \right] \right) \times \right. \\ \left. \times \left((\mathbf{I} - \mathbf{T}_{12}^*)^{-1} D^s \mathbf{m}_1 - (\mathbf{I} - \mathbf{T}_{21}^*)^{-1} E \left[m_1^{(s)}(X_{1i}) | X_2 \right] \right) \right\}.$$

Utilizando las siguientes aproximaciones,

$$\mathbf{S}_1^* \mathbf{S}_2^* \approx \mathbf{T}_{12}^*, \\ E[\mathbf{S}_2 \mathbf{m}_1] \approx E[m_1(X_{1i}) | X_2] \text{ y} \\ E[\mathbf{W}_1^{(r)} \mathbf{m}_1] \approx D^r \mathbf{m}_1.$$

Se define el estimador para $\theta_{1,1}(r, s)$:

$$\hat{\theta}_{1,1}(r, s) = \frac{1}{n} \text{tr} \left\{ \left((\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} - (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} \mathbf{S}_2^* \right) \mathbf{W}_1^{(r)*} \mathbf{Y} \mathbf{Y}^T \times \right. \\ \left. \times \mathbf{W}_1^{(r)*T} \left((\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} - \mathbf{S}_2^{*T} (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-T} \right) \right\}.$$

La expresión para $\hat{\theta}_{2,2}(r, s)$ es completamente análoga, y para $\theta_{1,2}(r, s)$:

$$\hat{\theta}_{1,2}(r, s) = \frac{1}{n} \text{tr} \left\{ \left((\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} - (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} \mathbf{S}_2^* \right) \mathbf{W}_1^{(r)*} \mathbf{Y} \mathbf{Y}^T \times \right. \\ \left. \times \mathbf{W}_2^{(r)*T} \left((\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} - \mathbf{S}_1^{*T} (\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-T} \right) \right\}.$$

Para el cálculo de estos estimadores se usarán, g_1 y g_2 , anchos de banda correspondientes a m_1 y m_2 , para $\mathbf{S}_1^{(r)}$, $\mathbf{S}_2^{(r)}$, \mathbf{S}_1^* y \mathbf{S}_2^* .

Estos parámetros, g_1 y g_2 , se eligen tales que minimicen AMSE de los estimadores en los que intervienen. Estas expresiones se pueden encontrar en Opsomer y Ruppert (1998).

Estimación de la varianza

El parámetro σ^2 se estimará por

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \{Y_i - \hat{m}(X_{1i}, X_{2i})\}^2. \quad (3.2)$$

El cálculo de $\hat{\sigma}^2$ requiere la selección de dos parámetros anchos de banda, k_1 y k_2 . Como antes, k_1 y k_2 se elegirán minimizando AMSE de $\hat{\sigma}^2$. Las expresiones se pueden encontrar en Opsomer y Ruppert (1998).

Extensión para D variables independientes

La extensión del método al caso en que hay D variables requiere, como se indicó en la definición del método de estimación, la independencia de las variables. Esta hipótesis permite utilizar expresiones mucho más simples para los estimadores.

Consideramos el caso en que se ajusta el modelo mediante regresión polinómica local de grado impar. Por el teorema 2.3.6, MASE para el modelo aditivo con D variables independientes se puede aproximar por:

$$\text{MASE}[\mathbf{h}|X_1, \dots, X_D] \simeq \sum_{d=1}^D \frac{\mu_{p_d+1}(K_{(p_d)})^2}{(p_d+1)!} h_d^{2p_d+2} \theta_d(p_d) + \\ + \sigma^2 \sum_{d=1}^D \text{R}(K_{(p_d)}) \frac{b_d - a_d}{nh_d} + o_P \left\{ \sum_{d=1}^D \left(h_d^{2p_d+2} + \frac{1}{nh_d} \right) \right\}.$$

Y los valores de h_d que minimizan esta función son:

$$h_{d,AMASE} = \left(\frac{(p_d + 1) p_d! (R(K_{(p_d)})) \sigma^2 (b_d - a_d)}{2n \mu_{p_d} (K_{(p_d)})^2 \theta_d (p_d + 1)} \right)^{1/(2p_d+3)},$$

Para estimar $\theta_d (p_d + 1)$ y σ^2 es necesario encontrar los anchos de banda óptimos para ambos estimadores.

Opsomer y Ruppert (1998) desarrollan la expresión de AMSE de $\hat{\theta}_d$ utilizando un modelo aditivo con polinomios locales de grado $p + 2$. El mínimo de esta expresión se alcanza en:

$$g_{d,AMSE} = \left(C_\theta (p_d) \frac{(p_d + 3)! R(K_{(p_d+1,p_d+2)}) \sigma^2 (b_d - a_d)}{2n |\theta_d (p_d + 1, p_d + 3)| \mu_{p_d+3} (K_{(p_d+1,p_d+2)})} \right)^{1/(2p_d+5)},$$

para $d = 1, \dots, D$, donde

$$C_\theta (p_d) = \begin{cases} 1 & \text{si } \theta_d (p_d + 1, p_d + 3) < 0 \\ \frac{2p_d+3}{2} & \text{si } \theta_d (p_d + 1, p_d + 3) > 0 \end{cases}.$$

Y para $\hat{\sigma}^2$, se utiliza un modelo aditivo utilizando polinomios locales de grado p . Los valores que minimizan AMSE de $\hat{\sigma}^2$, son los siguientes:

$$k_{d,AMSE} = \left(C_\sigma (p_d) \frac{(p_d + 1)!^2 |R(K_{(p_d)}) - 2K_{(p_d)}(0)| \sigma^2 (b_d - a_d)}{2n \theta_d (p_d + 1, p_d + 1) \mu_{p_d+1} (K_{(p_d)})^2} \right)^{1/(2p_d+3)},$$

para $d = 1, \dots, D$, donde

$$C_\sigma (p_d) = \begin{cases} 1 & \text{si } R(K_{(p_d)}) - 2K_{(p_d)}(0) < 0 \\ \frac{2p_d+3}{2} & \text{si } R(K_{(p_d)}) - 2K_{(p_d)}(0) > 0 \end{cases}.$$

Se puede resumir el procedimiento a seguir para estimar un modelo aditivo de la siguiente forma:

Dado un conjunto de datos siguiendo un modelo del tipo:

$$Y_i = \alpha + m_1 (X_{1i}) + \dots + m_D (X_{Di}) + \varepsilon_i, \quad i = 1, \dots, n.$$

Paso 1. Estimar σ_p^2 y $\theta (p + 1, p + 3)$ por regresión polinómica local. Se obtiene así $\hat{\theta} (p + 1, p + 3)$ y $\hat{\sigma}_p^2$.

Paso 2. Estimar $\theta (p + 1, p + 1)$ mediante un modelo aditivo con \hat{g} , donde \hat{g} es el ancho de banda que minimiza AMSE de $\hat{\theta} (p + 1, p + 3)$. Se obtiene $\hat{\theta} (p + 1, p + 1)$.

Paso 3. Estimar σ^2 mediante un modelo aditivo con \hat{k} , donde \hat{k} es el ancho de banda que minimiza AMSE de $\hat{\sigma}_p^2$. Obteniéndose $\hat{\sigma}^2$.

Paso 4. Ajustar un modelo aditivo con \hat{h}_d . Para obtener $\hat{\alpha}, \hat{m}_1, \dots, \hat{m}_D$.

Entonces, el modelo ajustado será:

$$\hat{m}(\mathbf{x}) = \hat{\alpha} + \hat{m}_1(x_1) + \dots + \hat{m}_D(x_D).$$

3.3.2. Validación cruzada generalizada

El método es una extensión de la definición para el caso univariante del criterio de validación cruzada generalizado, originalmente propuesto por Craven y Wahba (1979) introduciendo una modificación en el método de Newton utilizado para minimizar dicho criterio en el contexto de regresión no paramétrica lineal local. Este procedimiento fue propuesto por Kauermann y Opsomer y se extiende además para los modelos aditivos generalizados.

La generalización directa del criterio de validación cruzada generalizado (GCV) a un modelo aditivo multivariante homocedástico con D variables sería la siguiente:

$$\text{GCV}(\mathbf{h}) = \frac{(\mathbf{Y} - \hat{\mathbf{m}})^T (\mathbf{Y} - \hat{\mathbf{m}})}{n(1 - \mathbf{W}_M/n)^2}.$$

Sin embargo, la matriz \mathbf{W}_M no es fácil de calcular, dado que requiere invertir la matriz \mathbf{M} de dimensión $(nD) \times (nD)$. Esto motiva que los autores propongan la selección de \mathbf{h} que minimice una aproximación de este criterio, dada en la siguiente expresión:

$$\text{GCV}(\mathbf{h}) = \frac{(\mathbf{Y} - \hat{\mathbf{m}})^T (\mathbf{Y} - \hat{\mathbf{m}})}{n(1 - \sum_d \text{tr}(\mathbf{S}_d^*)/n)^2}.$$

De esta forma definen el siguiente algoritmo:

i Sea $\mathbf{h}_0 = (h_{01}, \dots, h_{0D})$ un vector de anchos de banda inicial y $\mathbf{h}_t = \mathbf{h}_0$. Elegir $\tilde{\mathbf{h}}_t = (\tilde{h}_{t1}, \dots, \tilde{h}_{tD})$, tal que $\tilde{h}_{td} = o(h_{td})$, para $d = 1, \dots, D$ (por ejemplo, $\tilde{h}_{td} = h_{td}^{5/7}$).

ii Para $t = 1, 2, \dots$ calcular la actualización, \mathbf{h}_{t+1} , mediante:

$$\mathbf{h}_{t+1} = \mathbf{h}_t - \left[\frac{\partial^2 \text{E}[\text{GCV}(\mathbf{h}_t)]}{(\partial \mathbf{h})(\partial \mathbf{h})^T} \right]^{-1} \left(\frac{\partial \text{GCV}}{\partial \mathbf{h}} \right). \quad (3.3)$$

iii Repetir el paso ii hasta que el cambio en el criterio de validación cruzada, $GCV(\mathbf{h})$, sea lo suficiente pequeño.

Las aproximaciones que son necesarias para el cálculo de (3.3) se pueden encontrar en el trabajo de Kauermann y Opsomer.

Capítulo 4

SiZer Map para modelos aditivos

4.1. Introducción

Uno de los inconvenientes que se presentan en los problemas de regresión multidimensionales, y que ya comentábamos en el capítulo 1, es la dificultad de representar gráficamente superficies de dimensión mayor que 2. Los modelos aditivos, sin embargo, permiten que cada variable se represente de forma individual, mostrando por separado su influencia sobre la variable dependiente.

Los métodos de suavizamiento aplicados a la estimación de curvas se han presentado por muchos autores como una herramienta útil para identificar características de los datos. Algunos ejemplos sobre esto se pueden encontrar en Eubank (1988), Fan y Gijbels (1996), Härdle (1990) y Wand y Jones (1995).

Chaudhuri y Marron (1999) proponen una herramienta de tipo gráfico, denominada *SiZer*¹ *Map*, que permite descubrir las características que los datos ponen de relieve, en la estimación de curvas. Esta herramienta tiene dos importantes propiedades: En primer lugar, agiliza el proceso de decisión para un analista experimentado de "qué características están presentes", a la vez que resuelve, de alguna manera, la situación para aquellos puntos dudosos; y segundo, permite, incluso a los analistas no experimentados, hacer inferencia sobre características desconocidas de las curvas.

La formulación y desarrollo de *SiZer Map* está dada tan sólo para curvas univariantes, la extensión a mayores dimensiones supone una gran complejidad, si bien según estos autores goza de un notable interés.

En el contexto de la regresión no paramétrica, si se aplica *SiZer Map* a partir de las versiones multivariantes de los suavizadores, como los de tipo polinómico local, o los sencillos suavizadores tipo núcleo como el de Nadaraya-Watson, la dificultad parte desde la propia representación gráfica de *SiZer Map*, siendo necesario el estudio de la magnitud del gradiente de la superficie estimada, así como una visión

¹ *SiZer* procede del acrónimo en inglés de "Significant ZERo crossings of derivatives".

tridimensional de las distintas regiones que *SiZer Map* pone de manifiesto.

Como apuntábamos antes los modelos aditivos en la estimación de las superficies de regresión multidimensionales permiten tratar cada una de las covariables separadamente, de modo que abordar la aplicación de *SiZer Map* tomando como base dichos modelos permitirá una simplicidad mayor que la que ofrecerían los suavizadores multivariantes a los que nos referíamos antes.

4.2. Definición de *SiZer Map*

Se trata de un gráfico en el que por medio de colores o tonalidades de grises (para versiones en blanco y negro) se muestran, para distintos niveles de suavizamiento (distintas elecciones del ancho de banda) el crecimiento o decrecimiento significativo de la curva objeto de estudio.

SiZer Map permitirá comprobar la significación de características en la superficie, tales como picos y valles sobre una familia de suavizadores $\{\hat{m}(x; h) : h \in [h_{min}, h_{max}]\}$, basándose en intervalos de confianza para la derivada, $m'(x; h)$, en el espacio definido por el parámetro de suavizado, h . Sobre el espacio de localización definido por la posición de x y h , el gráfico muestra el comportamiento de la función con distintas tonalidades, que representan: color azul (negro, en versiones en blanco y negro) donde la derivada es significativamente positiva; rojo (gris oscuro) para las zonas en que la derivada es significativamente negativa, y morado (gris claro) si la derivada no es significativamente distinta de cero. Las zonas con la tonalidad más clara de gris marcan valores demasiado pequeños del ancho de banda, que hacen que el tamaño muestral efectivo sea insuficiente, esto es, el número de observaciones en cada entorno local imposibilita el cálculo de las estimaciones.

Los intervalos de confianza para la derivada se expresan de la forma:

$$\hat{m}'(x; h) \pm q \sqrt{\widehat{\text{Var}}(\hat{m}'(x; h))},$$

donde el cuantil, q , se calcula desde aproximaciones normales, o bien utilizando técnicas *bootstrap* (más detalles sobre este aspecto se pueden encontrar en Chaudhuri y Marron (1999)).

El rango para el ancho de banda, $[h_{min}, h_{max}]$, se puede definir de varias formas, eligiendo un rango más o menos amplio. Lo usual será elegir un rango bastante amplio que permita captar las características más interesantes en cada punto sobre la familia de suavizadores.

Es necesario, por tanto, dar expresiones para $\hat{m}'(x; h)$ y $\widehat{\text{Var}}(\hat{m}'(x; h))$, que permitan el cálculo de dichos intervalos. Se resumen dichas expresiones en la siguiente sección.

4.2.1. Desarrollo de *SiZer Map*

Consideramos, en primer lugar, la estimación de la primera derivada de m en un punto x_0 , que viene dada por el siguiente problema de mínimos cuadrados ponderados²:

$$\hat{m}'(x_0; h) = \min_{\beta_0, \beta_1} \sum_{i=1}^n \{Y_i - \beta_0 - \beta_1 (X_i - x_0)\}^2 K_h(X_i - x_0).$$

La expresión matricial de la solución es la siguiente:

$$\hat{m}'(x_0; h) = e_2^T (\mathbf{X}_{x_0} \mathbf{W}_{x_0, h} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0} \mathbf{W}_{x_0, h} \mathbf{Y}, \quad (4.1)$$

con $e_2 = (0, 1)^T$.

La estimación propuesta para $\widehat{\text{Var}}(\hat{m}'(x_0; h))$ se basa en el hecho de que el estimador de la derivada es la suma ponderada de Y_i , con lo cual se obtiene la siguiente expresión para la varianza condicionada:

$$\text{Var}(\hat{m}'(x_0; h) | X_1, \dots, X_n) = \sum_{i=1}^n \sigma^2(Y_i | X_i) (w(x_0, X_i; h))^2.$$

Y para estimar $\sigma^2(Y_i | X_i)$ se suavizan los residuos mediante la siguiente expresión:

$$\hat{\sigma}^2(Y|X) = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2 K_h(X_i - x_0)}{\sum_{i=1}^n K_h(X_i - x_0)},$$

donde $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i; h)$.

Como se puede deducir, la creación del mapa implica el cálculo repetido de muchas estimaciones por lo que es muy importante introducir métodos computacionales que permitan agilizar el proceso. Una técnica que se adapta muy bien a esta situación es el denominado método *binnig*³, cuya idea principal es la de reducir el número de evaluaciones núcleo, basándose en el hecho de que muchas de estas evaluaciones son muy similares.

En la siguiente sección incluimos las expresiones básicas para la aplicación del método *binnig* en la creación de *SiZer Map*.

4.2.2. Aproximación *binned*

Para aplicar el método *binning* primero hay que modificar las observaciones de la siguiente forma: Dado un conjunto de observaciones $\{(X_i, Y_i), i = 1, \dots, n\}$ se define

²Los autores utilizan el suavizador lineal local por la simplicidad de sus expresiones para el cálculo de la derivada de la función m .

³Fan y Marron (1994) explican en detalle la aplicación de esta herramienta en la estimación no paramétrica de funciones de densidad y de regresión. La aproximación se denomina *binned*.

una red de puntos equiespaciados: x_1, \dots, x_g . A continuación, cada observación, X_i , se reemplaza por el punto más próximo en la red, $x_{j(i)}$. Esta modificación del conjunto original genera los siguientes índices:

$$I_j \equiv \{i : X_i \rightarrow x_j\}, \quad j = 1, \dots, g.$$

Entonces, los datos *binned* resumen a los originales mediante

$$\{(x_j, \bar{Y}_j, c_j), \quad j = 1, \dots, g\},$$

siendo $\bar{Y}_j = \frac{1}{c_j} \sum_{i \in I_j} Y_i$ y $c_j = \#I_j$, es decir, el número de observaciones en I_j .

A partir de aquí, una aproximación *binned* eficiente del estimador lineal local de la primera derivada de m en x_j es la siguiente:

$$\hat{m}'(x_j, h) \approx \frac{\bar{T}_1(x_j) - \bar{T}_0(x_j) \bar{X}(x_j)}{\bar{S}_2(x_j) - \bar{S}_1(x_j)^2 / \bar{S}_0(x_j)},$$

donde

$$\begin{aligned} \bar{S}_l(x_j) &= \sum_{j'=1}^g k_{l,j-j'} c_{j'}, & \bar{T}_l(x_j) &= \sum_{j'=1}^g k_{l,j-j'} Y_{j'}^\Sigma, & \bar{X}(x_j) &= \bar{S}_1(x_j) / \bar{S}_0(x_j), \\ Y_{j'}^\Sigma &= \sum_{i \in I_{j'}} Y_i, & k_{l,j} &= K_h(j\Delta) (j\Delta)^l & \text{y } \Delta &= x_j - x_{j-1}. \end{aligned}$$

En este procedimiento hay que señalar que el uso de la función núcleo gaussiana asegura que $\bar{S}_0(x_j)$ sea distinto de cero.

Para la obtención de la estimación de $\hat{\sigma}^2(Y|X = x_j)$, utilizando cálculos similares a los empleados en regresión lineal local, se tiene que

$$\hat{\sigma}^2(Y|X = x_j) \approx (1 - \hat{\rho}(x_j)^2) \hat{\sigma}^2(x_j),$$

donde

$$\begin{aligned} \hat{\sigma}^2(x_j) &= \frac{\bar{U}_0(x_j)}{\bar{S}_0(x_j)} - \left(\frac{\bar{T}_0(x_j)}{\bar{S}_0(x_j)} \right)^2, \\ \hat{\rho}(x_j)^2 &= (\hat{m}'(x_j; h))^2 \left(\frac{\bar{S}_0(x_j) \bar{S}_2(x_j) - \bar{S}_0(x_j)^2}{\hat{\sigma}^2(x_j) \bar{S}_0(x_j)^2} \right) \end{aligned}$$

y

$$\bar{U}_0(x_j) = \sum_{j'=1}^g k_{j-j'} Y_{j'}^{2\Sigma},$$

siendo $Y_j^{2\Sigma} = \sum_{i \in I_j} Y_i^2$. Entonces, la aproximación para la varianza condicionada será

$$\text{Var}(\widehat{m}'(x_j; h) | X_1, \dots, X_n) \approx \frac{\overline{V}_2(x_j) - 2\overline{V}_1(x_j)\overline{X}(x_j) + \overline{V}_0(x_j)\overline{X}(x_j)^2}{(\overline{S}_2(x_j) - \overline{S}_1(x_j)^2 / \overline{S}_0(x_j))^2},$$

donde

$$\overline{V}_l(x_j) = \sum_{j'=1}^g k_{l,j-j'} c_{j'} \widehat{\sigma}^2(Y|X = x_{j'}).$$

4.3. Extensión a un modelo aditivo

Para un modelo aditivo, *SiZer Map* constará de tantos gráficos como sea la dimensión del modelo, y permitirá comprobar la significación de características en la superficie unidimensional de cada componente. En esta línea se considera una familia de suavizadores de tipo aditivo (como, por ejemplo, el estimador lineal local realizado en Opsomer (2000)) $\{\widehat{m}(x; h_1, \dots, h_D) : h_d \in [h_{d,\min}, h_{d,\max}], d = 1, \dots, D\}$, sobre la que se definen intervalos de confianza para las derivadas de las componentes, $\widehat{m}'_d(x_d; h_d)$, $d = 1, \dots, D$. Así el gráfico d -ésimo muestra el comportamiento de la función para la componente m_d a través de distintas tonalidades, que tendrán el mismo significado que en la definición de *SiZer Map* en el caso univariante.

Los intervalos de confianza para la derivada de la d -ésima componente se expresan de la siguiente forma:

$$\widehat{m}'_d(x_d; h_d) \pm \sqrt{\widehat{\text{Var}}(\widehat{m}'_d(x_d; h_d))},$$

donde el cuantil, q , de nuevo se calcula desde aproximaciones normales, o bien utilizando técnicas *bootstrap*.

Llegados a este punto, es necesario tener expresiones tanto para $\widehat{m}'_d(x_d; h_d)$ como para $\widehat{\text{Var}}(\widehat{m}'_d(x_d; h_d))$. En la siguiente sección se presentan dichas expresiones para un modelo aditivo bivalente, más simple, siendo inmediatas las expresiones para el caso D -variante.

4.3.1. Desarrollo de *SiZer Map* ($D = 2$)

Según lo señalado antes, desarrollamos ahora *SiZer Map* para la primera componente, m_1 ; siendo el caso de m_2 análogo por lo que aquí lo omitimos.

Consideramos la estimación de la primera derivada de m_1 en un punto x_0 , que vendrá dada por⁴

$$\widehat{m}'_1(x_0) = s_{x_0}^{(1)}(Y - \widehat{m}_2),$$

⁴Opsomer y Ruppert, 1997.

siendo $s_{x_0}^{(1)} = e_2^T (\mathbf{X}_{x_0}^T \mathbf{W}_{x_0, h_1} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^T \mathbf{W}_{x_0, h_1}$, según la expresión dada en (4.1).

Sustituyendo la expresión de \hat{m}_2 , dada en (2.6), $\hat{m}'_1(x_0) = \mathbf{W}_{1, x_0}^* \mathbf{Y}$, donde,

$$\mathbf{W}_{1, x_0}^* = s_{x_0}^{(1)} (\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} (\mathbf{I} - \mathbf{S}_1^*).$$

De aquí se tiene que

$$\text{Var}(\hat{m}'_1(x_0)) = \mathbf{W}_{1, x_0}^{*T} \Sigma \mathbf{W}_{1, x_0}^*,$$

siendo $\Sigma = \text{diag}(\varepsilon_i^2)$.

Dado que los residuos, ε_i , son desconocidos los estimamos por

$$\hat{\varepsilon}_i = Y_i - \hat{m}(X_i),$$

y sustituyendo estas estimaciones en la matriz Σ , la varianza anterior se estimaría por

$$\widehat{\text{Var}}(\hat{m}'_1(x_0)) = \mathbf{W}_{1, x_0}^{*T} \widehat{\Sigma} \mathbf{W}_{1, x_0}^*.$$

Con esto los intervalos representados en *SiZer Map* se construyen, para la primera componente, como:

$$\hat{m}'_1(x_0) \pm q \sqrt{\widehat{\text{Var}}(\hat{m}'_1(x_0))}.$$

Según apuntábamos antes, para el cálculo del cuantil, q , son posibles varias aproximaciones. Chaudhuri y Marron (1999) proponen tres estimaciones, una de ellas basada en técnicas *bootstrap* que añade un elevado coste computacional a la producción del mapa. En nuestro trabajo utilizamos dos aproximaciones gaussianas para dicho cuantil que pasamos a definir.

La aproximación más simple, que denotamos por q_1 , viene dada por:

$$q_1 = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right),$$

donde α es el nivel de significación asociado al intervalo de confianza.

La segunda aproximación, que denotamos por q_2 , también está basada en la aproximación normal, pero en lugar de definirse como una constante para todo el espacio de localización que construye *SiZer Map*, se define variable en la red de anchos de banda considerada, $q_2 = q_2(h_1)$. En su definición interviene el denominado **tamaño muestral efectivo**, ESS, dado por:

$$\text{ESS}(x, h_1) = \frac{\sum_{i=1}^n K_{h_1}(x - X_i)}{K_{h_1}(0)}$$

de forma que,

$$q_2(h_1) = \Phi^{-1} \left(1 + (1 - \alpha)^{\frac{1}{m(h_1)}} \right),$$

donde

$$m(h_1) = \frac{n}{L^{-1} \sum_{l=1}^L \text{ESS}(x_l, h_1)},$$

y x_l , ($1 \leq l \leq L$), son la red de valores de X que definen el espacio de localización junto con los anchos de banda.

4.3.2. Aproximación *binned*

Para aplicar el método *binning* hay que modificar las observaciones de la siguiente forma: Dado un conjunto de observaciones $\{(X_i, Z_i, Y_i), i = 1, \dots, n\}$ se definen dos redes de puntos equiespaciados, x_1, \dots, x_{g_1} y z_1, \dots, z_{g_2} . Expondremos aquí cómo se actúa para X y análogamente se hará con Z . Cada observación, X_i , se reemplaza por el punto más próximo en la red, $x_{j(i)}$. Esta modificación del conjunto original genera los siguientes índices:

$$I_{1,j} \equiv \{i : X_i \rightarrow x_j\}, \quad j = 1, \dots, g_1.$$

Entonces, los datos *binned* resumen a los originales mediante

$$\{(x_j, \bar{Y}_{1,j}, c_{1,j}), \quad j = 1, \dots, g_1\},$$

siendo $\bar{Y}_{1,j} = \frac{1}{c_{1,j}} \sum_{i \in I_{1,j}} (Y_i - \hat{m}_2(Z_i))$ y $c_{1,j} = \#I_{1,j}$.

A partir de aquí, una aproximación *binned* eficiente del estimador lineal local de la derivada de m_1 , $\hat{m}'_1(x_j, h_1)$, es la siguiente:

$$\hat{m}'_1(x_j, h_1) \approx \frac{\bar{T}_{1,1}(x_j) - \bar{T}_{1,0}(x_j) \bar{X}_1(x_j)}{\bar{S}_{1,2}(x_j) - \bar{S}_{1,1}(x_j)^2 / \bar{S}_{1,0}(x_j)},$$

donde

$$\begin{aligned} \bar{S}_{1,l}(x_j) &= \sum_{j'=1}^{g_1} k_{1l,j-j'} c_{1,j'}, & \bar{T}_{1,l}(x_j) &= \sum_{j'=1}^{g_1} k_{1l,j-j'} Y_{1,j'}^\Sigma, \\ \bar{X}_1(x_j) &= \bar{S}_{1,1}(x_j) / \bar{S}_{1,0}(x_j), & Y_{1,j'}^\Sigma &= \sum_{i \in I_{1,j'}} (Y_i - \hat{m}_2(Z_i)), \\ k_{1l,j} &= K_{h_1}(j\Delta_1)(j\Delta_1)^l \quad \text{y} \quad \Delta_1 = x_j - x_{j-1}. \end{aligned}$$

En el caso de $\hat{\sigma}_1^2((Y - \hat{m}_2) | (X, Z) = (x_j, z_j))$, utilizando cálculos similares a los empleados en regresión lineal local, se tiene que

$$\hat{\sigma}_1^2(Y - \hat{m}_2 | (X, Z) = (x_j, z_j)) \approx (1 - \hat{\rho}_1(x_j)^2) \hat{\sigma}_1^2(x_j),$$

donde

$$\begin{aligned} \hat{\sigma}_1^2(x_j) &= \frac{\bar{U}_{1,0}(x_j)}{\bar{S}_{1,0}(x_j)} - \left(\frac{\bar{T}_{1,0}(x_j)}{\bar{S}_{1,0}(x_j)} \right)^2, \\ \hat{\rho}_1(x_j)^2 &= (\hat{m}'_1(x_j; h_1))^2 \left(\frac{\bar{S}_{1,0}(x_j) \bar{S}_{1,2}(x_j) - \bar{S}_{1,0}(x_j)^2}{\hat{\sigma}_1^2(x_j) \bar{S}_{1,0}(x_j)^2} \right) \end{aligned}$$

y

$$\bar{U}_{1,0}(x_j) = \sum_{j'=1}^{g_1} k_{1,j-j'} Y_{1,j'}^{2\Sigma},$$

siendo $Y_{1,j'}^{2\Sigma} = \sum_{i \in I_{1,j'}} (Y_i - \hat{m}_2)^2$. Entonces, la aproximación para la varianza condicionada será

$$\text{Var}(\hat{m}'_1(x_j; h_1) | X, Z) \approx \frac{\bar{V}_{1,2}(x_j) - 2\bar{V}_{1,1}(x_j)\bar{X}_1(x_j) + \bar{V}_{1,0}(x_j)\bar{X}_1(x_j)^2}{(\bar{S}_{1,2}(x_j) - \bar{S}_{1,1}(x_j)^2 / \bar{S}_{1,0}(x_j))^2},$$

donde

$$\bar{V}_{1,l}(x_j) = \sum_{j'=1}^{g_1} k_{1l,j-j'} c_{1,j'} \hat{\sigma}_1^2 ((Y - \hat{m}_2) | (X, Z) = (x_{j'}, z_{j'})).$$

4.4. Estudio de simulación

El objeto de este estudio es, además de ilustrar el funcionamiento de *SiZer Map* para los modelos aditivos, el de permitir una comparación y valoración del selector *plug-in* de Opsomer y Ruppert (1998) descrito en el capítulo 3.

En la simulación llevada a cabo hemos considerado las funciones empleadas por Opsomer y Ruppert (1998) para la exploración del comportamiento del selector *plug-in*, que son las siguientes:

$$m_1(x) = 1 - 6x + 36x^2 - 53x^3 + 22x^5,$$

y

$$m_2(z) = \text{sen}(5\pi z).$$

Así, el modelo aditivo considerado se define como

$$Y = \alpha + m_1(X) + m_2(Z) + \varepsilon$$

donde los residuos, ε , se han generado según una distribución $\mathcal{N}(0; 0.5)$, y el vector de covariables, (X, Z) , según la siguiente distribución normal bivalente,

$$(X, Z) \rightarrow \mathcal{N}\left(\left(\begin{array}{c} 1/2 \\ 1/2 \end{array}\right), 1/9 \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right)\right),$$

para los siguientes valores de ρ : -0.5, -0.25, 0, 0.25 y 0.5. Se consideraron, además, dos tamaños de muestra, $n = 200$ y 500 . El núcleo utilizado para el estimador lineal local ha sido el gaussiano, definido como $K(x) = (2\pi)^{(-1/2)} \exp(-x^2/2)$.

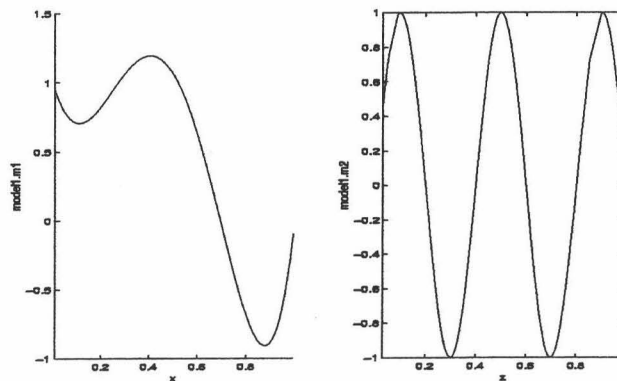


Figura 4.1: Componentes para el modelo 1.

Se trata de un modelo bivalente, ajustado mediante regresión lineal local, con distintos niveles de correlación entre las variables. Las funciones para cada componente están representadas en la figura 4.1.

Para cada modelo simulado, combinando un tamaño de muestra y un nivel de correlación, se han generado dos tipos de gráficos, el mapa tal y como se ha descrito en este capítulo, y un gráfico que representa las estimaciones de las componentes para distintos valores del ancho de banda. Este último tipo de gráfico permite comparar el comportamiento de las distintas elecciones del suavizado en la estimación de las funciones componentes; siendo el parámetro idéntico para ambas componentes.

La interpretación de los gráficos se hace más precisa si prestamos atención al primero de los casos considerado. En la figura 4.2, se representan dos gráficos para cada componente del modelo, por columnas. El primero de ellos, se ha denominado *family plot*⁵. Las curvas de color azul son las asociadas a distintas elecciones del grado de suavizamiento y la curva de color negro la que emplea el parámetro elegido según el criterio *plug-in* de Opsomer y Ruppert. Predeterminadamente se utiliza una red de once anchos de banda, aunque este número se ha elegido arbitrariamente. Las curvas que presentan más oscilaciones o cambios corresponden a los anchos de banda menores y estas se van suavizando conforme el ancho de banda utilizado es mayor.

El segundo gráfico, *SiZer Map*, ya ha sido definido a lo largo del capítulo. En horizontal se representa el rango de variación de la variable y en vertical aparecen los distintos anchos de banda en escala logarítmica. La línea continua representa el ancho de banda que proporciona el criterio *plug-in* de Opsomer y Ruppert (1998). En este caso, predeterminadamente, la red que se utiliza es de 21 anchos de banda. En la parte inferior del gráfico, correspondiente a los anchos de banda menores, nos

⁵Este nombre se ha adoptado del inglés, por ser más corto, que su posible en español, gráfico de familia de suavizadores.

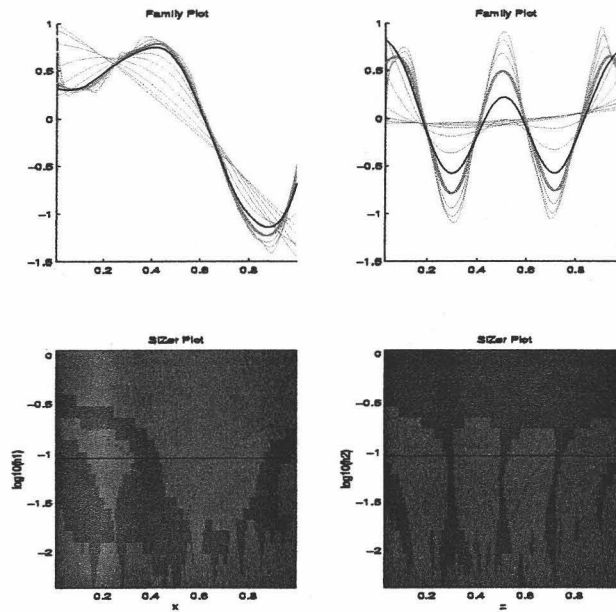


Figura 4.2: *Family plot* y *SiZER Map* para una muestra de tamaño $n = 200$ y variables incorreladas, $\rho = 0$.

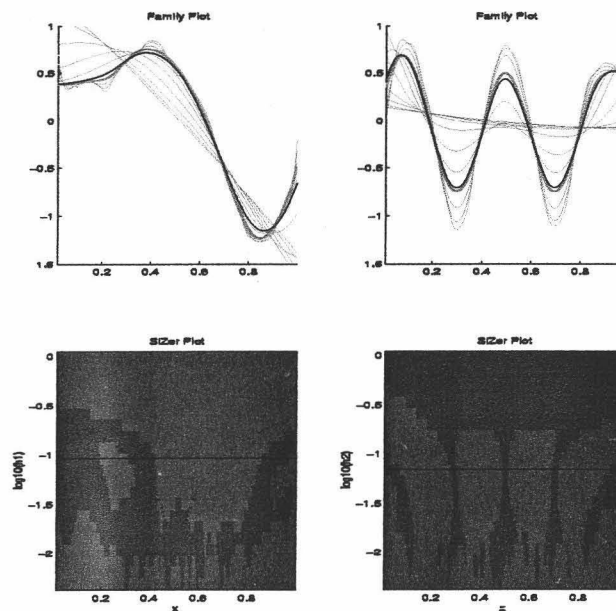


Figura 4.3: *Family plot* y *SiZER Map* para una muestra de tamaño $n = 200$ y nivel de correlación $\rho = -0.5$.

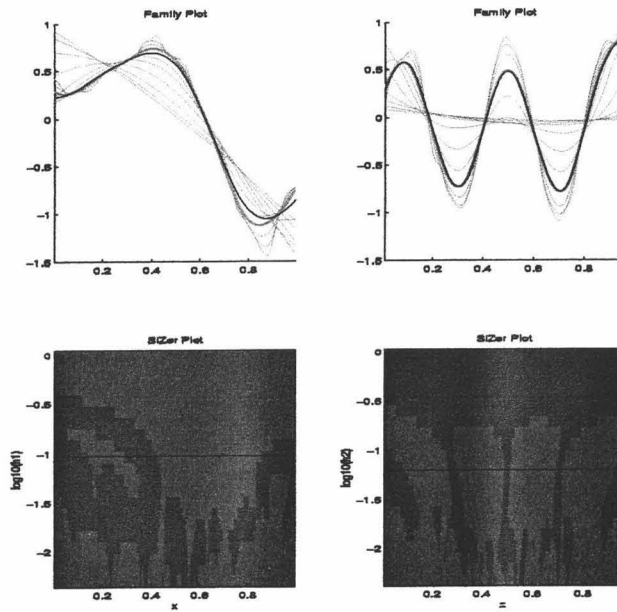


Figura 4.4: *Family plot* y *SiZer Map* para una muestra de tamaño $n = 200$ y nivel de correlación $\rho = -0.25$.

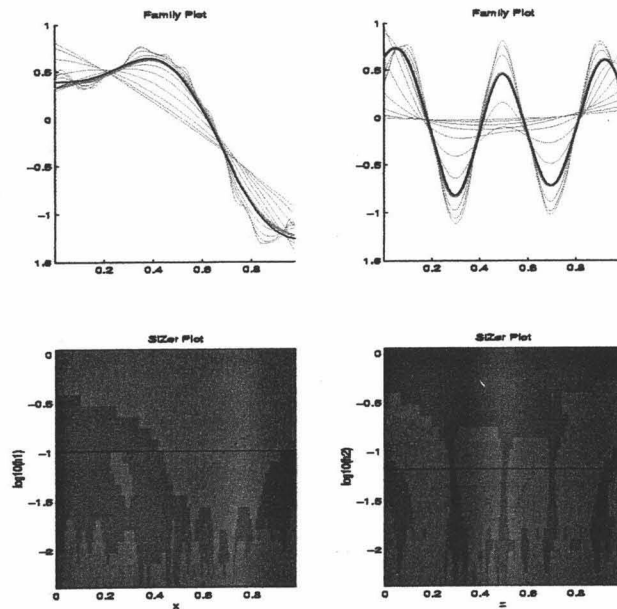


Figura 4.5: *Family plot* y *SiZer Map* para una muestra de tamaño $n = 200$ y nivel de correlación $\rho = 0.25$.

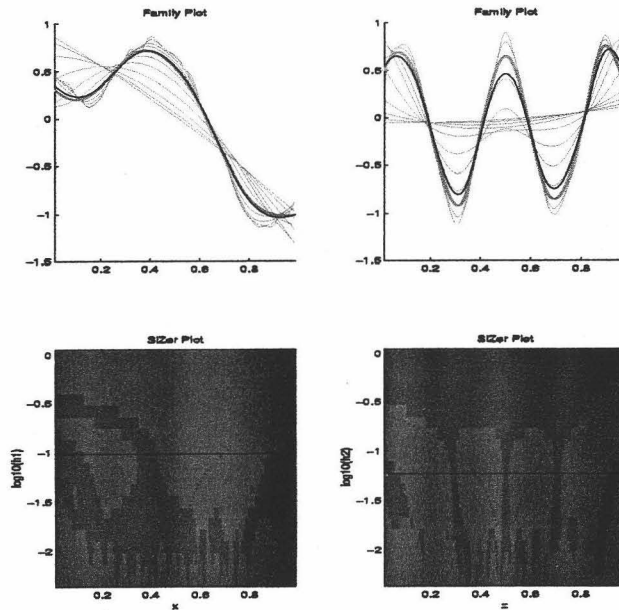


Figura 4.6: *Family plot* y *SiZer Map* para una muestra de tamaño $n = 200$ y nivel de correlación $\rho = 0.5$.

encontramos con zonas grises. Esto quiere decir que en el entorno que proporciona el ancho de banda correspondiente no se dispone de observaciones suficientes para calcular las estimaciones.

En las figuras 4.2, 4.3, 4.4, 4.5 y 4.6, que corresponden a muestras de tamaño $n=200$ y los diferentes niveles de correlación considerados, se puede observar como el ancho de banda *plug-in* no se adapta perfectamente en ninguna de las dos componentes.

Para la primera componente es destacable el comportamiento de las estimaciones en las fronteras o extremos, donde estas se alejan más de la verdadera función. Sin embargo, se puede observar que para anchos de banda menores se consiguen mejoras en estas zonas.

En la segunda componente se puede decir, sin embargo, que en general el ancho de banda *plug-in* sobresuaviza la función, aunque también es evidente el mal comportamiento de este selector en la frontera.

En las figuras 4.7, 4.8, 4.9, 4.10 y 4.11 se representan las muestras de tamaño $n = 500$ y los diferentes niveles de correlación considerados. En este caso se produce una notable mejoría del comportamiento del selector *plug-in* en las estimaciones de la frontera en ambas componentes, más destacable en la segunda componente.

Una situación que ocurre, en general, en todos los casos es que el valor para el selector es menor para la segunda componente y de forma más pronunciada en los

casos de $n=500$.

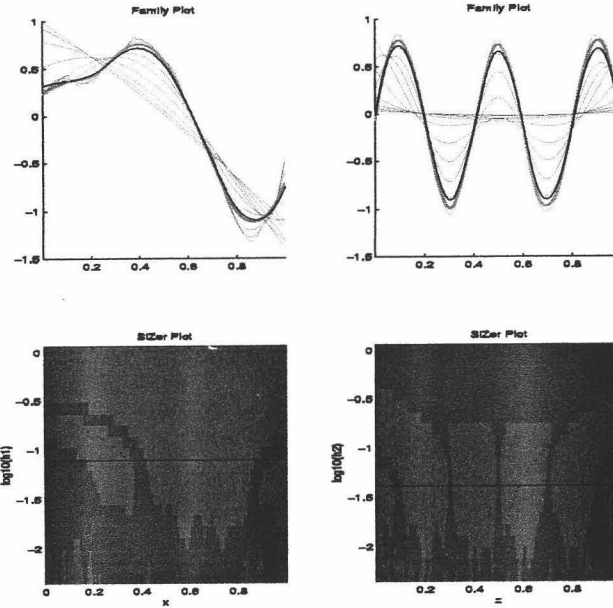


Figura 4.7: *Family plot* y *SiZer Map* para una muestra de tamaño $n = 500$ y variables incorreladas, $\rho = 0$.

4.5. Conclusiones

Del estudio desarrollado se desprende que no es suficiente con que el ancho de banda elegido sea diferente para cada variable, y parece que una posible solución sería considerar un procedimiento de regresión local con parámetros de ancho de banda locales. La solución de este nuevo problema fue desarrollada para modelos de regresión lineales locales univariantes por Fan y Gijbels (1995).

La solución que nosotros hemos considerado ha sido el desarrollo de un ancho de banda *bootstrap* local en cada variable. Esta metodología solo ha sido desarrollada por Martínez-Miranda (2000) para modelos de regresión univariantes estimados utilizando suavizadores lineales locales y en el siguiente capítulo se extiende para un modelo aditivo bivalente.

Para futuros trabajos se pretende la extensión a un modelo aditivo de cualquier dimensión y el desarrollo de un gráfico tipo *SiZer Map* que permita el estudio de estas nuevas metodologías de selección del ancho de banda en dimensiones superiores a dos.

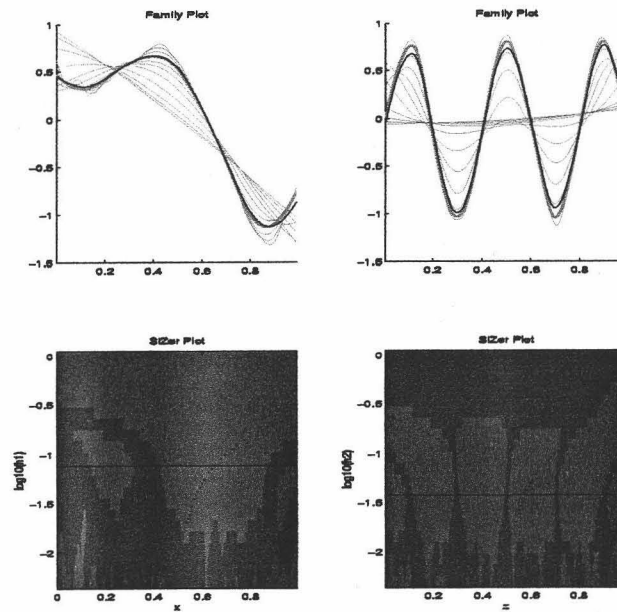


Figura 4.8: *Family plot* y *SiZER Map* para una muestra de tamaño $n = 500$ y nivel de correlación $\rho = -0.5$.

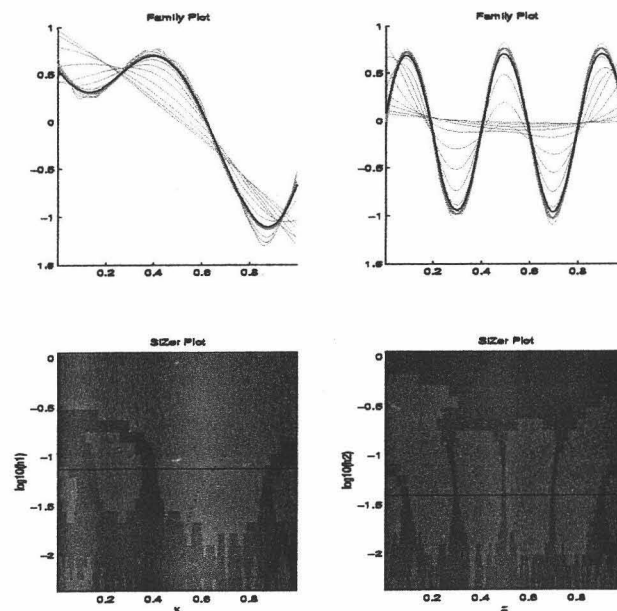


Figura 4.9: *Family plot* y *SiZER Map* para una muestra de tamaño $n = 500$ y nivel de correlación $\rho = -0.25$.

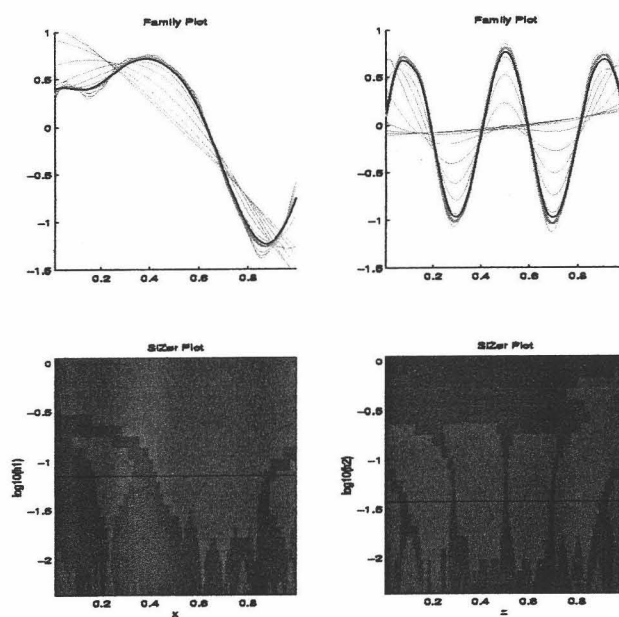


Figura 4.10: *Family plot* y *SiZer Map* para una muestra de tamaño $n = 500$ y nivel de correlación $\rho = 0.25$.

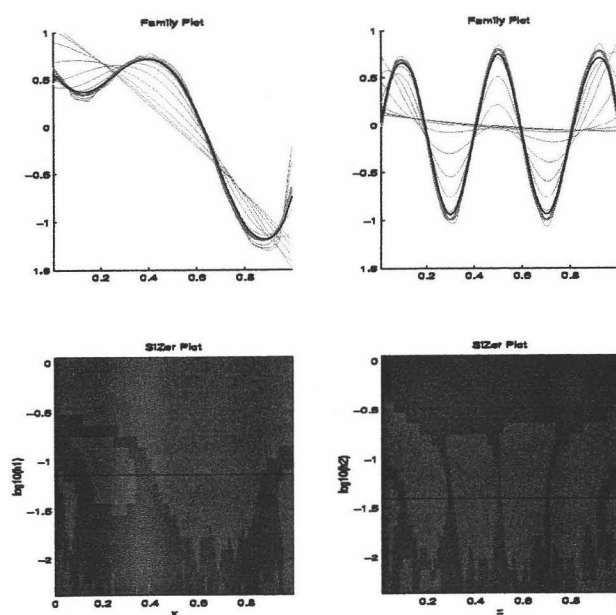


Figura 4.11: *Family plot* y *SiZer Map* para una muestra de tamaño $n = 500$ y nivel de correlación $\rho = 0.5$.

Capítulo 5

Selector *bootstrap* del ancho de banda

5.1. Introducción

Desde que la metodología *bootstrap* fuera introducida por Efron en 1979 ha sido objeto de estudio por muchos autores y aplicada en diferentes contextos como es el caso general de la estimación no paramétrica de curvas y en particular el de la regresión no paramétrica que nos ocupa.

Para los modelos de regresión no paramétrica heterocedásticos ha sido fundamental el denominado *wild bootstrap* que fue introducido por Wu en 1986, en el contexto particular de modelos de regresión paramétricos lineales. Este tipo de remuestreo también ha sido utilizado por otros autores en diferentes situaciones; así Cao-Abad (1991) y Härdle, Huet y Jolivet (1995) lo usan para aproximar distribuciones a la estimación de tipo Nadaraya-Watson de la función de regresión, Neumann y Polzehl (1998) para calcular intervalos de confianza, o Härdle y Mammen (1993) y Stute, González-Manteiga y Presedo-Quindimil (1998) que lo utilizan para realizar contrastes de hipótesis acerca de dicha función de regresión.

Como ya hemos mencionado en otros capítulos, una cuestión fundamental en el problema de regresión no paramétrica es la elección del parámetro ancho de banda, que ha sido muy estudiado en la literatura. En el caso unidimensional cabe citar los selectores para dicho parámetro del tipo *plug-in* (Ruppert, Sheather y Wand (1995), Fan y Gijbels (1995) y Ruppert (1997) entre otros) y validación cruzada (Härdle, Hall y Marron (1998)). En el caso multivariante debido a la complejidad del problema existen menos trabajos sobre este problema (Ruppert (1997) y Yang y Tschernig (1999)). Y en cuanto a los modelos aditivos, el método *plug-in* de Opsomer y Ruppert (1998) y el de validación cruzada de Opsomer (2000), ambos definidos en el capítulo 3, son las soluciones más relevantes ofrecidas para la selección del parámetro ancho de banda.

En los últimos años han surgido trabajos que utilizan el método *bootstrap* para la selección del parámetro de suavizado; en este sentido encontramos los trabajos de Härdle y Bowman (1988) para el estimador de Nadaraya-Watson, Hall (1990) para un estimador sencillo, tipo núcleo y más recientemente el trabajo de Martínez-Miranda (2000) para el estimador lineal local.

5.1.1. *Bootstrap* en regresión no paramétrica

En el problema de regresión, tanto si se considera desde un punto de vista clásico como desde una perspectiva no paramétrica, surgen dos formas de realizar el remuestreo: sobre los pares de observaciones o bien sobre los residuos de la regresión.

Aplicar la filosofía *bootstrap* sobre los pares, (\mathbf{X}_i, Y_i) , supone diseñar muestras *bootstrap* del tipo

$$\{(\mathbf{X}_{i_1}, Y_{i_1}), (\mathbf{X}_{i_2}, Y_{i_2}), \dots, (\mathbf{X}_{i_m}, Y_{i_m})\},$$

siendo $\{i_1, i_2, \dots, i_m\}$ una muestra aleatoria de enteros entre 1 y n .

El planteamiento sobre los residuos, ε_i , es totalmente diferente. En primer lugar, se obtiene una estimación inicial de la función de regresión desconocida, \hat{m} , y se estima cada ε_i de la siguiente manera

$$\hat{\varepsilon}_i = Y_i - \hat{m}(\mathbf{X}_i).$$

A partir de dichas estimaciones se diseñan los residuos *bootstrap*, ε_i^* , siguiendo algún método de muestreo especificado. Con este procedimiento se obtienen muestras *bootstrap* de la forma:

$$\{(\mathbf{X}_1, Y_1^*), (\mathbf{X}_2, Y_2^*), \dots, (\mathbf{X}_n, Y_n^*)\},$$

donde $Y_i^* = \hat{m}(\mathbf{X}_i) + \varepsilon_i^*$. En variantes elementales del *bootstrap* (*naive bootstrap*) se consideran versiones centradas de $\hat{\varepsilon}_i$, y se obtienen los residuos *bootstrap* mediante muestreo con reposición y probabilidades iguales.

Otra versión de la metodología *bootstrap* es la denominada *wild bootstrap*, que aplicada sobre los residuos, $\hat{\varepsilon}_i$, genera los residuos *bootstrap*, ε_i^* , a partir de una distribución degenerada en dos puntos, con media 0, varianza igual al cuadrado del residuo estimado, y momento de tercer orden igual al cubo del residuo estimado.

En las siguientes secciones se describe esta metodología *wild bootstrap* aplicada a la selección del ancho de banda en un modelo aditivo bivalente estimado con suavizadores lineales locales. Así, en primer lugar definiremos una estimación *bootstrap* del error cuadrático medio puntual del estimador aditivo y obtendremos el selector *bootstrap* del ancho de banda minimizando en cada punto dicha estimación. Obsérvese que de este modo el selector obtenido proporciona un parámetro de tipo local, esto es, una función.

5.2. Estimación *bootstrap* del error cuadrático medio en la estimación de un modelo aditivo

Partimos del modelo aditivo de regresión no paramétrica, formulado con anterioridad en esta memoria,

$$Y_i = \alpha + m_1(X_{1i}) + \cdots + m_D(X_{Di}) + \varepsilon_i,$$

donde los residuos, ε_i , se suponen de media 0 y varianza $\sigma^2(\mathbf{X}_i)$, esto es, una situación de heterocedasticidad.

La generación de muestras en este contexto de regresión no paramétrica multi-dimensional se lleva a cabo siguiendo los siguientes pasos:

i) Se construyen los residuos, $\hat{\varepsilon}_i = Y_i - \hat{m}_{0;\mathbf{g}}(\mathbf{X}_i)$, en donde $\hat{m}_{0;\mathbf{g}}$ es un estimador piloto de m y $\mathbf{g} = (g_1, \dots, g_D)^T$ es un vector de parámetros anchos de banda.

ii) Se generan los residuos *bootstrap*, ε_i^* , $i = 1, \dots, n$, verificando las siguientes condiciones

$$E^*[\varepsilon_i^*] = 0, \quad E^*[\varepsilon_i^{*2}] = \hat{\varepsilon}_i^2, \quad E^*[\varepsilon_i^{*3}] = \hat{\varepsilon}_i^3, \dots$$

iii) Se construye la muestra *bootstrap* $\{(\mathbf{X}_i^T, Y_i^*)\}$, donde Y_i^* tiene la siguiente definición: $Y_i^* = \hat{m}_{0;\mathbf{g}}(\mathbf{X}_i) + \varepsilon_i^*$, $i = 1, \dots, n$.

iv) Se calcula el estimador *bootstrap* de la función de regresión, $\hat{m}_{\mathbf{h}}^*$, utilizando como estimador el algoritmo *backfitting* con suavizadores lineales locales sobre la muestra *bootstrap* definida en el paso anterior.

El error cuadrático medio (MSE) en una observación \mathbf{x} ,

$$\text{MSE}_{\mathbf{h}}(\mathbf{x}) = E_{Y|\mathbf{X}}[\{\hat{m}_{\mathbf{h}}(\mathbf{x}) - m(\mathbf{x})\}^2],$$

con $\mathbf{h} = (h_1, \dots, h_D)^T$, se aproximará por el error cuadrático medio *bootstrap*:

$$\text{MSE}_{\mathbf{h},\mathbf{g}}^*(\mathbf{x}) = E^*[(\hat{m}_{\mathbf{h}}^*(\mathbf{x}) - \hat{m}_{0;\mathbf{g}}(\mathbf{x}))^2], \quad (5.1)$$

siendo E^* , el operador esperanza condicionado sobre la distribución en el remuestreo.

El siguiente teorema (dado para el caso particular $D = 2$) demuestra la consistencia de esta aproximación bajo las siguientes hipótesis:

H1. La función núcleo, K , es una función de densidad simétrica, acotada y con soporte compacto, tal que $\int K(u)u du = 0$, $\int K(u)u^2 du = \mu_2(K) \neq 0$ y $\int K(u)u^4 du < \infty$.

- H2. Las sucesiones de anchos de banda $\{h_1\}$ y $\{h_2\}$ verifican: $n^{-1}h_1^{-1} \rightarrow 0$ y $n^{-1}h_2^{-1} \rightarrow 0$, cuando $n \rightarrow \infty$, y $h_1^{-1}h_2 = O(1)$. Las mismas propiedades se dan para las sucesiones $\{g_1\}$ y $\{g_2\}$.
- H3. f se supone de soporte compacto y para cada \mathbf{x} en el soporte de f , $\sigma^2(\mathbf{x})$ es estrictamente positiva, y f , σ^2 y la función de regresión, m , se suponen continuamente diferenciables.
- H4. La función $\mu_4(\mathbf{x}) = E[Y^4 | \mathbf{X} = \mathbf{x}]$ está uniformemente acotada o es continuamente diferenciable y acotada.
- H5. $h_1g_1^{-1} \rightarrow 0$ y $h_2g_2^{-1} \rightarrow 0$ cuando $n \rightarrow \infty$.

Teorema 5.2.1 *Bajo las hipótesis H1-H5, la aproximación bootstrap del error MSE, (5.1), para un estimador aditivo, es consistente en el interior de soporte de f , es decir,*

$$\text{MSE}_{h,g}^*(\mathbf{x}) - \text{MSE}_h(\mathbf{x}) \rightarrow 0$$

en probabilidad.

Demostración: Consideramos la descomposición de MSE en términos del sesgo y la varianza,

$$\text{MSE}_h(\mathbf{x}) = \{E_{Y|\mathbf{X}}[\hat{m}_h(\mathbf{x})] - m(\mathbf{x})\}^2 + \text{Var}_{Y|\mathbf{X}}(\hat{m}_h(\mathbf{x})) = B_h(\mathbf{x})^2 + V_h(\mathbf{x}),$$

y, de forma análoga, para la aproximación *bootstrap* de MSE

$$\begin{aligned} \text{MSE}_{h,g}^*(\mathbf{x}) &= \{E_{Y|\mathbf{X}}^*[\hat{m}_h^*(\mathbf{x})] - \hat{m}_g(\mathbf{x})\}^2 + \text{Var}_{Y|\mathbf{X}}^*(\hat{m}_h^*(\mathbf{x})) = \\ &= B_{h,g}^*(\mathbf{x})^2 + V_{h,g}^*(\mathbf{x}), \end{aligned}$$

Entonces, para demostrar la consistencia de MSE^* , probamos la consistencia para cada una de sus componentes, sesgo y varianza *bootstrap*, por separado.

Empezamos, en primer lugar, con el sesgo. Haciendo cálculos se obtiene la siguiente expresión

$$\begin{aligned} B_{h,g}^*(\mathbf{x}) - B_h(\mathbf{x}) &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) w_i(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - Y_i) + \\ &\quad + \sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h}) w_j(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - Y_j) = \\ &= T_{1,1} + T_{1,2}, \end{aligned}$$

donde los pesos están definidos de la siguiente forma:

$$w_i(\mathbf{x}, \mathbf{h}) = s_{1,x_1}^T (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} (\mathbf{I} - \mathbf{S}_2^*) + s_{2,x_2}^T (\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} (\mathbf{I} - \mathbf{S}_1^*),$$

y asintóticamente se expresan

$$w_i(\mathbf{x}, \mathbf{h}) = n^{-1} f_1^{-1}(x_1) K_{h_1}(X_{i1} - x_1) + n^{-1} f_2^{-1}(x_2) K_{h_2}(X_{i2} - x_2). \quad (5.2)$$

Se trata de probar que la diferencia entre los sesgos es $o_p(1)$, o lo que es lo mismo, $T_{1,1} = T_{1,2} = o_p(1)$.

Por la desigualdad de Markov se tiene que

$$T_{1,1} = E[T_{1,1}] + O_p\left(\sqrt{\text{Var}(T_{1,1})}\right),$$

así que pasamos al cálculo de la esperanza y la varianza de $T_{1,1}$.

$$\begin{aligned} E[T_{1,1}] &= E\left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) w_i(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - Y_i)\right] = \\ &= E_{\mathbf{X}}\left[E_{Y|\mathbf{X}}\left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) w_i(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - Y_i)\right]\right] = 0 \end{aligned} \quad (5.3)$$

dado que $E_{Y|\mathbf{X}}[Y_i] = m(\mathbf{X}_i)$, y donde $E_{\mathbf{X}}$ denota la esperanza condicionada a la distribución de \mathbf{X} y $E_{Y|\mathbf{X}}$ es la esperanza condicionada a $Y | \mathbf{X}$.

$$\text{Var}(T_{1,1}) = \text{Var}_{\mathbf{X}}(E_{Y|\mathbf{X}}[T_{1,1}]) + E_{\mathbf{X}}[\text{Var}_{Y|\mathbf{X}}(T_{1,1})] = E_{\mathbf{X}}[\text{Var}_{Y|\mathbf{X}}(T_{1,1})]$$

puesto que el primer término se anula, igual que en (5.3). Entonces,

$$\begin{aligned} \text{Var}_{Y|\mathbf{X}}(T_{1,1}) &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g})^2 \text{Var}_{Y|\mathbf{X}}(m(\mathbf{X}_i) - Y_i) = \\ &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g})^2 \sigma^2(\mathbf{X}_i) \end{aligned}$$

Utilizando la expresión asintótica de los pesos, (5.2),

$$\begin{aligned} T_{1,1} &= \sum_{i=1}^n (n^{-1} f_1^{-1}(x_1) K_{h_1}(X_{i1} - x_1) + n^{-1} f_2^{-1}(x_2) K_{h_2}(X_{i2} - x_2)) \times \\ &\quad \times \{n^{-1} f_1^{-1}(X_{i1}) K_{g_1}(0) + n^{-1} f_2^{-1}(X_{i2}) K_{g_2}(0)\} (m(\mathbf{X}_i) - Y_i) \end{aligned}$$

la varianza sería

$$\begin{aligned} \text{Var}_{Y|\mathbf{X}}(T_{1,1}) &= \sum_{i=1}^n \left(n^{-1} f_1^{-1}(x_1) K_{h_1}(X_{1i} - x_1) + n^{-1} f_2^{-1}(x_2) K_{h_2}(X_{2i} - x_2) \right)^2 \times \\ &\quad \times \left\{ n^{-1} f_1^{-1}(X_{1i}) K_{g_1}(0) + n^{-1} f_2^{-1}(X_{2i}) K_{g_2}(0) \right\}^2 \sigma^2(\mathbf{X}_i) \end{aligned}$$

y si se aproxima la suma de Riemann por la integral, la esperanza queda de la siguiente forma:

$$\begin{aligned} E_{\mathbf{X}} [\text{Var}_{Y|\mathbf{X}}(T_{1,1})] &= n^{-3} \int_{\mathbf{x}_i} \left(f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2) \right)^2 \times \\ &\quad \times \left\{ f_1^{-1}(x_{1i}) g_1^{-1} K(0) + f_2^{-1}(x_{2i}) g_2^{-1} K(0) \right\}^2 \sigma^2(\mathbf{x}_i) f(\mathbf{x}_i) d\mathbf{x}_i \end{aligned}$$

y usando el desarrollo de Taylor en f y σ^2 , la expresión de la varianza será:

$$\begin{aligned} &n^{-3} K(0)^2 \int_{\mathbf{u}} \left(f_1^{-1}(x_1) h_1^{-1} K(u_1) + f_2^{-1}(x_2) h_2^{-1} K(u_2) \right)^2 \times \\ &\quad \times \left\{ f_1^{-1}(x_1 + h_1 u_1) g_1^{-1} + f_2^{-1}(x_2 + h_2 u_2) g_2^{-1} \right\}^2 \times \\ &\quad \times \sigma^2(x_1 + h_1 u_1, x_2 + h_2 u_2) f(x_1 + h_1 u_1, x_2 + h_2 u_2) h_1 h_2 d\mathbf{u} = \\ &= n^{-3} K(0)^2 \left(f_1^{-1}(x_1) + g_1 g_2^{-1} f_2^{-1}(x_2) \right)^2 \sigma^2(\mathbf{x}) f(\mathbf{x}) h_1 h_2 g_1^{-2} \times \\ &\quad \times \int_{\mathbf{u}} \left(f_1^{-1}(x_1) h_1^{-1} K(u_1) + f_2^{-1}(x_2) h_2^{-1} K(u_2) \right)^2 d\mathbf{u} = \\ &= n^{-3} K(0)^2 \left(f_1^{-1}(x_1) + g_1 g_2^{-1} f_2^{-1}(x_2) \right)^2 \sigma^2(\mathbf{x}) f(\mathbf{x}) h_1 h_2 g_1^{-2} \times \\ &\quad \times \left\{ h_1^{-2} f_1^{-2}(x_1) R(K) + h_2^{-2} f_2^{-2}(x_2) R(K) + 2h_1^{-1} h_2^{-1} f_1^{-1}(x_1) f_2^{-1}(x_2) \right\}, \end{aligned}$$

donde, por estar acotadas K , f y σ^2 , tenemos que:

$$\text{Var}(T_{1,1}) = O(n^{-3} g_1^{-2})$$

Finalmente

$$T_{1,1} = 0 + O_p\left(n^{-3/2} g_1^{-1/2}\right) = o_p(1).$$

Seguimos con el segundo término, $T_{1,2}$. De nuevo por la desigualdad de Markov

$$T_{1,2} = E[T_{1,2}] + O_p\left(\sqrt{\text{Var}(T_{1,2})}\right)$$

$$\begin{aligned} E [T_{1,2}] &= E \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h}) w_j(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - Y_j) \right] = \\ &= E_{\mathbf{X}} \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h}) w_j(\mathbf{X}_i, \mathbf{g}) E_{Y|\mathbf{X}} [m(\mathbf{X}_i) - Y_j] \right] = \\ &= E \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h}) w_j(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - m(\mathbf{X}_j)) \right] \end{aligned}$$

Aplicando la expresión asintótica de los pesos, (5.2), podemos llegar a la siguiente expresión:

$$\begin{aligned} &\int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2)) f(\mathbf{x}_i) \times \\ &\times \int_{\mathbf{x}_j} [(f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1j}) + f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2j})) f(\mathbf{x}_j) \times \\ &\times (m(\mathbf{x}_i) - m(\mathbf{x}_j))] d\mathbf{x}_j d\mathbf{x}_i \end{aligned}$$

Haciendo el desarrollo de Taylor sobre m y f en la segunda integral se obtiene

$$\begin{aligned} &\int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2)) f(\mathbf{x}_i) \times \\ &\times \{f_1^{-1}(x_1) g_1^{-1} + f_2^{-1}(x_2) g_2^{-1}\} o(1) f(\mathbf{x}_i) g_1 g_2 d\mathbf{x}_i \end{aligned}$$

y aplicando de nuevo el desarrollo de Taylor sobre f se obtiene

$$\begin{aligned} E [T_{1,2}] &= o(1) f(\mathbf{x})^2 (f_1^{-1}(x_1) h_2^{-1} + f_2^{-1}(x_2) h_1^{-1}) (f_1^{-1}(x_1) g_2^{-1} + f_2^{-1}(x_2) g_1^{-1}) = \\ &= o(1). \end{aligned}$$

Ahora la varianza:

$$\text{Var}(T_{1,2}) = \text{Var}_{\mathbf{X}} (E_{Y|\mathbf{X}} [T_{1,2}]) + E_{\mathbf{X}} [\text{Var}_{Y|\mathbf{X}} (T_{1,2})] = A_1 + A_2$$

$$\begin{aligned} A_1 &= \text{Var}_{\mathbf{X}} \left(E_{Y|\mathbf{X}} \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h}) w_j(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - Y_j) \right] \right) = \\ &= \text{Var}_{\mathbf{X}} \left(\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h}) w_j(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - m(\mathbf{X}_j)) \right) \end{aligned}$$

La expresión a la que calculamos la varianza tiene media $o(1)$ según hemos visto anteriormente, por tanto, solo tenemos que calcular el momento de orden 2,

$$\begin{aligned} A_1 &= \sum_{i \neq j} E_{\mathbf{X}} [w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g})^2 (m(\mathbf{X}_i) - m(\mathbf{X}_j))^2] = \\ &= n^{-2} \int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^2 f(\mathbf{x}_i) \times \\ &\quad \times \int_{\mathbf{x}_j} [(f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1j}) + f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2j}))^2 f(\mathbf{x}_j) \times \\ &\quad \times (m(\mathbf{x}_i) - m(\mathbf{x}_j))^2 d\mathbf{x}_j] d\mathbf{x}_i, \end{aligned}$$

Ahora, aplicando el desarrollo de Taylor sobre f y m en las dos integrales llegamos a

$$\begin{aligned} A_1 &= n^{-2} (f_1^{-1}(x_1) h_1^{-1} h_2 R(K) + f_2^{-1}(x_2) h_1 h_2^{-1} R(K) + 2f_1^{-1}(x_1) f_2^{-1}(x_2)) \times \\ &\quad \times (f_1^{-1}(x_1) g_1^{-1} g_2 R(K) + f_2^{-1}(x_2) g_1 g_2^{-1} R(K) + 2f_1^{-1}(x_1) f_2^{-1}(x_2)) o(1) f(\mathbf{x}_i)^2 = \\ &= O(n^{-2}). \end{aligned}$$

$$\begin{aligned} A_2 &= E_{\mathbf{X}} \left[\text{Var}_{Y|\mathbf{X}} \left(\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h}) w_j(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_i) - Y_j) \right) \right] = \\ &= E_{\mathbf{X}} \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g})^2 \sigma^2(\mathbf{X}_i) \right] \end{aligned}$$

Aplicando la expresión asintótica de los pesos, (5.2), y aproximando la suma de Riemann por la integral

$$\begin{aligned} A_2 &= n^{-2} \int_{\mathbf{x}_i} (f_1^{-1}(x_{1i}) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_{2i}) K_{h_2}(x_{2i} - x_2))^2 f(\mathbf{x}_i) \times \\ &\quad \times \int_{\mathbf{x}_j} [(f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1j}) + f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2j}))^2 f(\mathbf{x}_j) \sigma^2(\mathbf{x}_j) d\mathbf{x}_j] d\mathbf{x}_i \end{aligned}$$

y siguiendo el mismo procedimiento que el realizado para calcular $\text{Var}(T_{1,1})$ se demuestra que $A_2 = O(n^{-2})$.

Por tanto, $\text{Var}(T_{1,2}) = O(n^{-2})$, luego

$$T_{1,2} = o(1) + O_p(n^{-1/2}) = o_p(1).$$

Entonces hemos conseguido probar lo que buscábamos, es decir,

$$B_{\mathbf{h}, \mathbf{g}}^*(\mathbf{x}) - B_{\mathbf{h}}(\mathbf{x}) = T_{1,1} + T_{1,2} = o(1).$$

Continuamos ahora con el término de la varianza:

$$\begin{aligned} V_{h,g}^*(\mathbf{x}) - V_h(\mathbf{x}) &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 (\varepsilon_i^2 - \sigma^2(\mathbf{X}_i)) + \\ &\quad + \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 (\widehat{m}_g(\mathbf{X}_i) - m(\mathbf{X}_i))^2 - \\ &\quad - 2 \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \varepsilon_i (\widehat{m}_g(\mathbf{X}_i) - m(\mathbf{X}_i)) = \\ &= \vartheta_1 + \vartheta_2 - 2\vartheta_3. \end{aligned}$$

Desarrollamos cada uno de los términos ϑ_i , $i = 1, \dots, 3$, por separado.

El primer término tiene esperanza y varianza:

$$E[\vartheta_1] = E_{\mathbf{X}} [E_{Y|\mathbf{X}}[\vartheta_1]] = E_{\mathbf{X}} \left[E_{Y|\mathbf{X}} \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 (\varepsilon_i^2 - \sigma^2(\mathbf{X}_i)) \right] \right] = 0 \quad (5.4)$$

puesto que $E_{Y|\mathbf{X}}[\varepsilon^2] = \sigma^2(\mathbf{X}_i)$.

$$\text{Var}(\vartheta_1) = \text{Var}_{\mathbf{X}} (E_{Y|\mathbf{X}}[\vartheta_1]) + E_{\mathbf{X}} [\text{Var}_{Y|\mathbf{X}}(\vartheta_1)] = E_{\mathbf{X}} [\text{Var}_{Y|\mathbf{X}}(\vartheta_1)]$$

teniendo en cuenta (5.4).

Por la representación asintótica de los pesos, (5.2), el primer término es equivalente a

$$\vartheta_1 \approx n^{-2} \sum_{i=1}^n (f_1^{-1}(x_1) K_{h_1}(X_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(X_{2i} - x_2))^2 (\varepsilon_i^2 - \sigma^2(\mathbf{X}_i))$$

entonces,

$$\begin{aligned} \text{Var}(\vartheta_1) &= \\ &= E_{\mathbf{X}} \left[n^{-4} \sum_{i=1}^n (f_1^{-1}(x_1) K_{h_1}(X_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(X_{2i} - x_2))^4 \mu_4(\mathbf{X}_i) \right]. \end{aligned}$$

Si suponemos que $\mu_4(\mathbf{X}_i)$ es continuamente derivable, entonces, aplicando el desarrollo de Taylor en μ_4 y f es fácil comprobar que $\text{Var}(\vartheta_1) = o_p(1)$, y si $\mu_4(\mathbf{X}_i)$ se supone uniformemente acotada llegamos a la misma conclusión. Por tanto,

$$\vartheta_1 = o_p(1).$$

El segundo término,

$$\vartheta_2 = \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \left(\sum_{j=1}^n w_j(\mathbf{X}_i, \mathbf{g}) Y_j - m(\mathbf{X}_i) \right)^2 = \vartheta_{2,1} + \vartheta_{2,2},$$

donde

$$\begin{aligned} \vartheta_{2,1} &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g})^2 (Y_i - m(\mathbf{X}_i))^2 + \\ &+ \sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g})^2 (Y_j - m(\mathbf{X}_i))^2 = A_1 + A_2 \end{aligned}$$

y

$$\vartheta_{2,2} = \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \sum_{\substack{j \neq i \\ l, j \neq i}} w_j(\mathbf{X}_i, \mathbf{g}) w_l(\mathbf{X}_i, \mathbf{g}) (Y_j - m(\mathbf{X}_i)) (Y_l - m(\mathbf{X}_i)).$$

Por desarrollos análogos realizados durante la demostración se llega a

$$\begin{aligned} E[A_1] &= E \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g})^2 \sigma^2(\mathbf{X}_i) \right] = \\ &= n^{-3} \int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^2 \times \\ &\quad \times \{f_1^{-1}(x_{1i}) g_1^{-1} K(0) + f_2^{-1}(x_{2i}) g_2^{-1} K(0)\}^2 \sigma^2(\mathbf{x}_i) f(\mathbf{x}_i) d\mathbf{x}_i = \\ &= n^{-3} K(0)^2 \sigma^2(\mathbf{x}) f(\mathbf{x}) (f_1^{-1}(x_1) g_1^{-1} + f_2^{-1}(x_2) g_2^{-1})^2 \times \\ &\quad \times \{f_1^{-2}(x_1) h_1^{-1} h_2 R(K) + f_2^{-2}(x_2) h_1 h_2^{-1} R(K) + 2f_1^{-1}(x_1) f_2^{-1}(x_2)\} = \\ &= o(1). \end{aligned}$$

y

$$\begin{aligned} \text{Var}(A_1) &= E_{\mathbf{X}} \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^4 w_i(\mathbf{X}_i, \mathbf{g})^4 \mu_4(\mathbf{X}_i) \right] + \\ &+ \text{Var}_{\mathbf{X}} \left(\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g})^2 \sigma^2(\mathbf{X}_i) \right) = A_{1,1} + A_{2,2}. \end{aligned}$$

Ahora,

$$\begin{aligned} A_{1,1} &= n^{-7} K(0)^4 \int_{\mathbf{u}} (f_1^{-1}(x_1) h_1^{-1} K(u_1) + f_2^{-1}(x_2) h_2^{-1} K(u_2))^4 \times \\ &\quad \times (f_1^{-1}(x_{1i}) g_1^{-1} + f_2^{-1}(x_{2i}) g_2^{-1})^4 \mu_4(\mathbf{x}_i) f(\mathbf{x}_i) h_1 h_2 d\mathbf{u} = \\ &= n^{-7} K(0)^4 C(\mu_4) f(\mathbf{x}) (f_1^{-1}(x_1) g_1^{-1} + f_2^{-1}(x_2) g_2^{-1})^4 \times \\ &\quad \times \int_{\mathbf{u}} (f_1^{-1}(x_1) h_1^{-1} K(u_1) + f_2^{-1}(x_2) h_2^{-1} K(u_2))^4 d\mathbf{u} = \\ &= O\left(n^{-7} (g_1^{-1} + g_2^{-1})^4 (h_1^{-1} + h_2^{-1})^4\right), \end{aligned}$$

$$\begin{aligned} A_{1,2} &= \sum_{i=1}^n E_{\mathbf{X}} [w_i(\mathbf{x}, \mathbf{h})^4 w_i(\mathbf{X}_i, \mathbf{g})^4 \sigma^4(\mathbf{X}_i)] - \\ &\quad - \sum_{i=1}^n E_{\mathbf{X}} [w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g})^2 \sigma^2(\mathbf{X}_i)]^2 = \\ &= O\left(n^{-7} (h_1^{-1} + h_2^{-1})^4 (g_1^{-1} + g_2^{-1})^4\right), \end{aligned}$$

luego $\text{Var}(A_1) = O\left(n^{-7} (g_1^{-1} + g_2^{-1})^4 (h_1^{-1} + h_2^{-1})^4\right)$, con lo que

$$A_1 = o(1) + O_p\left(\sqrt{O\left(n^{-7} (g_1^{-1} + g_2^{-1})^4 (h_1^{-1} + h_2^{-1})^4\right)}\right) = o_p(1).$$

Ahora, veamos qué ocurre con A_2 :

$$\begin{aligned} E[A_2] &= E_{\mathbf{X}} \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g})^2 E_{Y|\mathbf{X}} [(\varepsilon_j + m(\mathbf{X}_j) - m(\mathbf{X}_i))^2] \right] = \\ &= \sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g})^2 \sigma^2(\mathbf{X}_j) = \\ &= O\left(n^{-3} (g_1^{-1} + g_2^{-1})^2\right) = o(1) \end{aligned}$$

y

$$\text{Var}(A_2) = E_{\mathbf{X}} [\text{Var}_{Y|\mathbf{X}}(A_2)] + \text{Var}_{\mathbf{X}}(E_{Y|\mathbf{X}}[A_2]) = B_1 + B_2,$$

donde

$$\begin{aligned} B_1 &= E_{\mathbf{X}} \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^4 w_j(\mathbf{X}_i, \mathbf{g})^4 \text{Var}_{Y_i|\mathbf{X}_i}((Y_j - m(\mathbf{X}_i))^2) \right] \approx \\ &\approx \sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^4 w_j(\mathbf{X}_i, \mathbf{g})^4 (\sigma^4(\mathbf{X}_j) + \sigma^2(\mathbf{X}_j) (m(\mathbf{X}_j) - m(\mathbf{X}_i))^2) = \\ &= O\left(n^{-7} (h_1^{-1} + h_2^{-1})^4 (g_1^{-1} + g_2^{-1})^4 h_1 h_2\right) = O\left(n^{-7} (h_1^{-1} + h_2^{-1})^3\right) \end{aligned}$$

y

$$\begin{aligned} B_2 &= \text{Var}_{\mathbf{X}} \left(\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g})^2 \sigma^2(\mathbf{X}_j) \right) = \\ &= O\left(n^{-7} (h_1^{-1} + h_2^{-1})^4 (g_1^{-1} + g_2^{-1})^4 h_1 h_2\right) = O\left(n^{-7} (h_1^{-1} + h_2^{-1})^3\right) \end{aligned}$$

con lo que

$$A_2 = o(1) + O_p\left(n^{-7/2} (h_1^{-1} + h_2^{-1})^{3/2}\right) = o_p(1).$$

Seguimos con $\vartheta_{2,2}$, $E[\vartheta_{2,2}]$ será:

$$E_{\mathbf{X}} \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \sum_{\substack{j \neq l \\ l, j \neq i}} w_j(\mathbf{X}_i, \mathbf{g}) w_l(\mathbf{X}_i, \mathbf{g}) E_{Y|\mathbf{X}} [(Y_j - m(\mathbf{X}_i)) (Y_l - m(\mathbf{X}_i))] \right] = o(1),$$

y

$$\text{Var}(\vartheta_{2,2}) = E_{\mathbf{X}} [\text{Var}_{Y|\mathbf{X}}(\vartheta_{2,2})] + \text{Var}_{\mathbf{X}}(E_{Y|\mathbf{X}}[\vartheta_{2,2}]) = B_1 + B_2$$

Es fácil ver que cada uno de los términos es $O(n^{-5}(h_1^{-1} + h_2^{-1})^4)$. Para B_1 tenemos que:

$$\begin{aligned} E_{\mathbf{X}} \left[\text{Var}_{Y|\mathbf{X}} \left(\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \sum_{\substack{j \neq l \\ l, j \neq i}} w_j(\mathbf{X}_i, \mathbf{g}) w_l(\mathbf{X}_i, \mathbf{g}) (Y_j - m(\mathbf{X}_i)) (Y_l - m(\mathbf{X}_i)) \right) \right] &= \\ &= n^3 \int_{\mathbf{x}_i} (n^{-1} f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + n^{-1} f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^4 f(\mathbf{x}_i) \times \\ &\quad \times \int_{\mathbf{x}_j} (n^{-1} f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1j}) + n^{-1} f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2j}))^2 f(\mathbf{x}_j) \times \\ &\quad \times \int_{\mathbf{x}_l} (n^{-1} f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1l}) + n^{-1} f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2l}))^2 f(\mathbf{x}_l) \times \\ &\quad \times \sigma^2(\mathbf{x}_j) \sigma^2(\mathbf{x}_l) d\mathbf{x}_l d\mathbf{x}_j d\mathbf{x}_i = \\ &= O(n^3 n^{-4} n^{-2} n^{-2} h_1^2 h_2^2 (h_1^{-1} + h_2^{-1})^4 (g_1^{-1} + g_2^{-1})^2 (g_1^{-1} + g_2^{-1})^2) = \\ &= O(n^{-5} (h_1^{-1} + h_2^{-1})^4) \end{aligned}$$

y B_2 :

$$\begin{aligned} \text{Var}_{\mathbf{X}} \left(\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \sum_{\substack{j \neq i \\ l, j \neq i}} w_j(\mathbf{X}_i, \mathbf{g}) w_l(\mathbf{X}_i, \mathbf{g}) (m(\mathbf{X}_j) - m(\mathbf{X}_i)) (m(\mathbf{X}_l) - m(\mathbf{X}_i)) \right) &\approx \\ \approx n^3 \int_{\mathbf{x}_i} (n^{-1} f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + n^{-1} f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^4 f(\mathbf{x}_i) \times \\ \times \int_{\mathbf{x}_j} (n^{-1} f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1j}) + n^{-1} f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2j}))^2 f(\mathbf{x}_j) \times \\ \times \int_{\mathbf{x}_l} (n^{-1} f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1l}) + n^{-1} f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2l}))^2 f(\mathbf{x}_l) \times \\ \times (m(\mathbf{x}_j) - m(\mathbf{x}_i)) (m(\mathbf{x}_l) - m(\mathbf{x}_i)) d\mathbf{x}_j d\mathbf{x}_l d\mathbf{x}_i = \\ = O(n^3 n^{-4} n^{-2} n^{-2} h_1^2 h_2^2 (h_1^{-1} + h_2^{-1})^4 (g_1^{-1} + g_2^{-1})^2 (g_1^{-1} + g_2^{-1})^2) = \\ = O(n^{-5} (h_1^{-1} + h_2^{-1})^4) \end{aligned}$$

luego, $\text{Var}(\vartheta_{2,2}) = O(n^{-5} (h_1^{-1} + h_2^{-1})^4)$, con lo que

$$\vartheta_{2,2} = o(1) + O_p(n^{-5/2} (h_1^{-1} + h_2^{-1})^2) = o_p(1).$$

Y esto lleva a que, por tanto, $\vartheta_2 = o_p(1)$.

Por último, el término ϑ_3 :

$$\begin{aligned} \vartheta_3 &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \varepsilon_i w_i(\mathbf{X}_i, \mathbf{g}) (Y_i - m(\mathbf{X}_i)) + \\ &+ \sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 \varepsilon_i w_j(\mathbf{X}_i, \mathbf{g}) (Y_j - m(\mathbf{X}_i)) = \vartheta_{3,1} + \vartheta_{3,2}, \end{aligned}$$

$$\begin{aligned} E[\vartheta_{3,1}] &= E_{\mathbf{X}} \left[E_{Y|\mathbf{X}} \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \varepsilon_i w_i(\mathbf{X}_i, \mathbf{g}) (Y_i - m(\mathbf{X}_i)) \right] \right] = \\ &= E_{\mathbf{X}} \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g}) \sigma^2(\mathbf{X}_i) \right] \end{aligned}$$

Ahora utilizando la expresión asintótica de los pesos, (5.2), y el desarrollo de Taylor sobre σ^2 y f como en pasos anteriores se llega a

$$\begin{aligned} E[\vartheta_{3,1}] &= n^{-2} K(0) \{f_1^{-1}(x_1) g_2 + f_2^{-1}(x_2) g_1\} \sigma^2(\mathbf{x}) \times \\ &\times \{f_1^{-1}(x_1) h_1^{-2} R(K) + f_2^{-1}(x_2) h_2^{-2} R(K) + 2f_1^{-1}(x_1) f_2^{-1}(x_2) h_1^{-1} h_2^{-1}\} = \\ &= O(n^{-2} h^{-1}) = o(1) \end{aligned}$$

$$\begin{aligned} \text{Var}(\vartheta_{3,1}) &= \text{Var}_{\mathbf{X}} (\text{E}_{Y|\mathbf{X}} [\vartheta_{3,1}]) + \text{E}_{\mathbf{X}} [\text{Var}_{Y|\mathbf{X}} (\vartheta_{3,1})] = \\ &= \text{Var}_{\mathbf{X}} \left(\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 w_i(\mathbf{X}_i, \mathbf{g}) \sigma^2(\mathbf{X}_i) \right) + \\ &\quad + \text{E}_{\mathbf{X}} \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^4 w_i(\mathbf{X}_i, \mathbf{g})^2 \mu_4(\mathbf{X}_i) \right] = A_1 + A_2, \end{aligned}$$

entonces, si sustituimos la expresión de los pesos por una aproximación asintótica y la suma de Riemann por la integral, obtenemos

$$\begin{aligned} A_1 &= n^{-5} \int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^4 \times \\ &\quad \times \{f_1^{-1}(x_1) g_1^{-1} K(0) + f_2^{-1}(x_2) g_2^{-1} K(0)\}^2 \sigma^2(\mathbf{x}_i) f(\mathbf{x}_i) d\mathbf{x}_i = \\ &= O(n^{-5} (h_1^{-4} + h_2^{-4})) \end{aligned}$$

y

$$\begin{aligned} A_2 &= n^{-5} \int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^4 \times \\ &\quad \times \{f_1^{-1}(x_1) g_1^{-1} K(0) + f_2^{-1}(x_2) g_2^{-1} K(0)\}^2 \mu_4(\mathbf{x}_i) f(\mathbf{x}_i) d\mathbf{x}_i = \\ &= O(n^{-5} (h_1^{-4} + h_2^{-4})). \end{aligned}$$

Luego,

$$\vartheta_{3,1} = o(1) + O_p(n^{-5/2} (h_1^{-4} + h_2^{-4})^{1/2}) = o_p(1).$$

Para el término $\vartheta_{3,2}$:

$$\text{E}[\vartheta_{3,2}] = \text{E}_{\mathbf{X}} \left[\text{E}_{Y|\mathbf{X}} \left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g}) \varepsilon_i (Y_j - m(\mathbf{X}_i)) \right] \right] = 0$$

ya que $\text{E}_{Y|\mathbf{X}}[\varepsilon_i \varepsilon_j] = 0$. Entonces, teniendo en cuenta que esta esperanza se anula, tenemos

$$\text{Var}(\vartheta_{3,2}) = \text{E}_{\mathbf{X}} \left[\text{Var}_{Y|\mathbf{X}} \left(\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g}) \varepsilon_i (Y_j - m(\mathbf{X}_i)) \right) \right]$$

y

$$\begin{aligned} \text{Var}_{Y|\mathbf{X}} \left(\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g}) \varepsilon_i (Y_j - m(\mathbf{X}_i)) \right) &\approx \\ &\approx \sum_{i \neq j} \text{Var}_{Y|\mathbf{X}} (w_i(\mathbf{x}, \mathbf{h})^2 w_j(\mathbf{X}_i, \mathbf{g}) \varepsilon_i (\varepsilon_j + m(\mathbf{X}_j) - m(\mathbf{X}_i))) = \\ &= \sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^4 w_j(\mathbf{X}_i, \mathbf{g})^2 \text{Var}_{Y|\mathbf{X}} (\varepsilon_i (\varepsilon_j + m(\mathbf{X}_j) - m(\mathbf{X}_i))). \end{aligned}$$

Ahora

$$\begin{aligned}
 \text{Var}_{Y|X}(\varepsilon_i(\varepsilon_j + m(\mathbf{X}_j) - m(\mathbf{X}_i))) &= \\
 &= E_{Y|X}[\varepsilon_i^2(\varepsilon_j + m(\mathbf{X}_j) - m(\mathbf{X}_i))^2] - E_{Y|X}[\varepsilon_i(\varepsilon_j + m(\mathbf{X}_j) - m(\mathbf{X}_i))]^2 = \\
 &= E_{Y|X}[\varepsilon_i^2\varepsilon_j^2] + (m(\mathbf{X}_j) - m(\mathbf{X}_i))^2 E_{Y|X}[\varepsilon_i^2] + \\
 &\quad + 2(m(\mathbf{X}_j) - m(\mathbf{X}_i)) E_{Y|X}[\varepsilon_i^2\varepsilon_j] = \\
 &= \sigma^2(\mathbf{X}_i)\sigma^2(\mathbf{X}_j) + (m(\mathbf{X}_j) - m(\mathbf{X}_i))^2\sigma^2(\mathbf{X}_i).
 \end{aligned}$$

Entonces, por cálculos similares a los ya empleados anteriormente,

$$\begin{aligned}
 \text{Var}(\vartheta_{3,2}) &= \\
 &= E_{\mathbf{X}}\left[\sum_{i \neq j} w_i(\mathbf{x}, \mathbf{h})^4 w_j(\mathbf{X}_i, \mathbf{g})^2 (\sigma^2(\mathbf{X}_i)\sigma^2(\mathbf{X}_j) + (m(\mathbf{X}_j) - m(\mathbf{X}_i))^2\sigma^2(\mathbf{X}_i))\right] = \\
 &= n^{-4} \int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^4 \times \\
 &\quad \times \int_{\mathbf{x}_j} \left\{ (f_1^{-1}(x_{1i}) K_{g_1}(x_{1i} - x_{1j}) + f_2^{-1}(x_{2i}) K_{g_2}(x_{2i} - x_{2j}))^2 \times \right. \\
 &\quad \left. \times \sigma^2(\mathbf{x}_i)\sigma^2(\mathbf{x}_j) + (m(\mathbf{x}_j) - m(\mathbf{x}_i))^2\sigma^2(\mathbf{x}_i) f(\mathbf{x}_j) d\mathbf{x}_j \right\} f(\mathbf{x}_i) d\mathbf{x}_i = \\
 &= n^{-4} \int_{\mathbf{x}_i} (f_1^{-1}(x_1) K_{h_1}(x_{1i} - x_1) + f_2^{-1}(x_2) K_{h_2}(x_{2i} - x_2))^4 \times \\
 &\quad \times \left\{ f_1^{-1}(x_1) g_1^{-2} R(K) + f_2^{-1}(x_2) g_2^{-2} R(K) + 2f_1^{-1}(x_1) f_2^{-1}(x_2) g_1^{-1} g_2^{-1} \right\} \times \\
 &\quad \times h_1 h_2 \sigma^4(\mathbf{x}_i) f(\mathbf{x}_i) d\mathbf{x}_i = \\
 &= O\left(n^{-4} (h_1^{-1} + h_2^{-1})^2\right),
 \end{aligned}$$

luego, $\vartheta_{3,2} = 0 + O_p(n^{-2}(h_1^{-1} + h_2^{-1}))$ y $\vartheta_3 = o_p(1)$.

□

Nota: La generalización del resultado a D variables se obtiene siguiendo la misma línea de demostración que el teorema, si bien la notación se hace más complicada.

5.2.1. Elección *bootstrap* del ancho de banda

Según el procedimiento descrito, se propone como criterio de selección del vector de parámetros ancho de banda, la estimación *bootstrap* del error cuadrático medio. Entonces, definimos la selección *bootstrap* de dicho vector como la función

$$\mathbf{h}_{boot}^*(\mathbf{x}) = \arg \min_{\mathbf{h}} \text{MSE}_{\mathbf{h}}^*(\mathbf{x})$$

que difiere, por tanto, en cada punto de estimación considerado, \mathbf{x} .

5.2.2. Expresión exacta del estimador *bootstrap*

Calculando la esperanza bajo la distribución en el remuestreo, E^* , obtendremos la expresión exacta de la estimación *bootstrap* del error cuadrático medio, $MSE_{h,g}^*(\mathbf{x})$.

Consideramos una descomposición del error cuadrático medio *bootstrap* en suma de una componente de sesgo y otra de varianza, de la siguiente forma

$$MSE_{h,g}^*(\mathbf{x}) = (B_{h,g}^*(\mathbf{x}))^2 + \text{Var}_{h,g}^*(\mathbf{x}),$$

donde los términos de sesgo y varianza se definen como

$$B_{h,g}^*(\mathbf{x}) = E^* [\hat{m}_h^*(\mathbf{x})] - \hat{m}_g(\mathbf{x})$$

$$V_{h,g}^*(\mathbf{x}) = \text{Var}^* (\hat{m}_h^*(\mathbf{x})).$$

Comenzamos por el cálculo del término de sesgo:

$$\begin{aligned} E^* [\hat{m}_h^*(\mathbf{x})] &= E^* \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) Y_i^* \right] = \\ &= E^* \left[\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) (\hat{m}_g(\mathbf{X}_i) + \varepsilon_i^*) \right] = \\ &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) \hat{m}_g(\mathbf{X}_i), \end{aligned}$$

de modo que

$$B_{h,g}^*(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) \hat{m}_g(\mathbf{X}_i) - \hat{m}_g(\mathbf{x}).$$

Para el término de varianza

$$\begin{aligned} \text{Var}^* (\hat{m}_h^*(\mathbf{x})) &= \text{Var}^* \left(\sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) Y_i^* \right) = \\ &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \text{Var}^* (Y_i^*) = \\ &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 \hat{\varepsilon}_i, \end{aligned}$$

entonces

$$V_{h,g}^*(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 (Y_i - \hat{m}_g(\mathbf{X}_i))^2.$$

La expresión final de la estimación *bootstrap* que se propone para el error cuadrático medio será

$$\text{MSE}_{\mathbf{h},\mathbf{g}}^*(\mathbf{x}) = \left\{ \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) \widehat{m}_{\mathbf{g}}(\mathbf{X}_i) - \widehat{m}_{\mathbf{g}}(\mathbf{x}) \right\}^2 + \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h})^2 (Y_i - \widehat{m}_{\mathbf{g}}(\mathbf{X}_i))^2.$$

Es posible obtener versiones matriciales de tales expresiones. Para ello, consideramos la siguiente notación:

$$\widehat{M}_{\mathbf{g}} = \begin{pmatrix} \widehat{\alpha} + \widehat{m}_{1:g_1}(X_{11}) + \widehat{m}_{2:g_2}(X_{21}) \\ \vdots \\ \widehat{\alpha} + \widehat{m}_{1:g_1}(X_{1n}) + \widehat{m}_{2:g_2}(X_{2n}) \end{pmatrix}$$

$$\Sigma_{\mathbf{g}} = \text{diag}(\widehat{\varepsilon}_i).$$

Con esta notación, la definida en el capítulo 2 y usando las expresiones de los estimadores obtenidas cuando se aplica el algoritmo *backfitting* con suavizadores lineales locales, se obtienen las siguientes expresiones matriciales del sesgo y la varianza, respectivamente:

$$\begin{aligned} B_{\mathbf{h},\mathbf{g}}^*(\mathbf{x}) &= \\ &= n^{-1} (\mathbf{1} + s_{1,x_1,h_1}^T (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} (\mathbf{I} - \mathbf{S}_2^*) + s_{2,x_2,h_2}^T (\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} (\mathbf{I} - \mathbf{S}_1^*)) \widehat{M}_{\mathbf{g}} - \\ &\quad - n^{-1} (\mathbf{1} + s_{1,x_1,g_1}^T (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} (\mathbf{I} - \mathbf{S}_2^*) + s_{2,x_2,g_2}^T (\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} (\mathbf{I} - \mathbf{S}_1^*)) \mathbf{Y} \quad (5.5) \end{aligned}$$

$$\begin{aligned} V_{\mathbf{h},\mathbf{g}}^*(\mathbf{x}) &= \\ &= n^{-2} (\mathbf{1} + s_{1,x_1,h_1}^T (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} (\mathbf{I} - \mathbf{S}_2^*) + s_{2,x_2,h_2}^T (\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} (\mathbf{I} - \mathbf{S}_1^*)) \Sigma_{\mathbf{g}} \times \\ &\quad \times (\mathbf{1} + s_{1,x_1,h_1}^T (\mathbf{I} - \mathbf{S}_2^* \mathbf{S}_1^*)^{-1} (\mathbf{I} - \mathbf{S}_2^*) + s_{2,x_2,h_2}^T (\mathbf{I} - \mathbf{S}_1^* \mathbf{S}_2^*)^{-1} (\mathbf{I} - \mathbf{S}_1^*)). \quad (5.6) \end{aligned}$$

5.3. Elección de los parámetros piloto

El procedimiento de estimación *bootstrap* que acabamos de describir, supone la elección de un vector de parámetros ancho de banda piloto, \mathbf{g} .

En esta sección describiremos la forma de elegir este ancho de banda piloto, considerando en los desarrollos el caso que nos ocupa de un modelo aditivo bivalente estimado utilizando estimadores lineales locales.

En la búsqueda de un criterio para decidir qué parámetro es el más adecuado, consideramos el sugerido por Härdle y Marron (1991) quienes en el contexto de regresión no paramétrica, obtienen bandas de confianza aplicando la metodología *bootstrap* sobre los residuos de tipo *wild bootstrap* para el estimador tipo núcleo de Nadaraya-Watson. Entonces definimos el vector de anchos de banda piloto óptimo como aquel que minimice la distancia en media cuadrática entre el sesgo *bootstrap* y el sesgo teórico, esto es,

$$\mathbf{g}_{opt} = \arg \min E_{Y|X} \left[\{B_{h,g}^*(\mathbf{x}) - B_h(\mathbf{x})\}^2 \right].$$

El siguiente lema da una aproximación asintótica para la anterior esperanza condicionada que denotamos por $L(\mathbf{h}, \mathbf{g})$, cuya expresión resulta compleja, no siendo fácil su minimización. Este problema queda abierto para posteriores trabajos. En las simulaciones, debido a que la implementación de estos resultados presentaba una gran dificultad se ha usado el criterio de validación cruzada para elegir estos parámetros.

Para la formulación del resultado necesitamos además de las hipótesis establecidas en el teorema 5.2 la siguiente:

H.6 m_1, m_2, f_1 y f_2 , son cuatro veces diferenciables; y K es dos veces continuamente diferenciable.

Lema 5.3.1 *Bajo las hipótesis H.1-H.6, el criterio $L(\mathbf{h}, \mathbf{g})$ tiene la siguiente aproximación asintótica:*

$$\begin{aligned} L(\mathbf{h}, \mathbf{g}) \approx & (B_1(g_1^2(h_1^2 + h_2^2)) + B_2(g_2^2(h_1^2 + h_2^2)))^2 + \\ & + \frac{1}{4}n^{-1}f_1^{-1}(x_1)f_2^{-1}(x_2)g_1g_2\sigma(x) \int [(b-a)\mu_2(K)(f_2^2(x_2)h_1^2g_1^{-3}D^2K(u_1) + \\ & + f_1^2(x_1)h_2^2g_2^{-3}D^2K(u_2)) + (b-a)^2f_1(x_1)f_2(x_2)(g_1^{-2}DK(u_1) + \\ & + g_2^{-2}DK(u_2)) + \frac{(b-a)^3}{3}(f_1^2(x_1)g_1^{-3}D^2K(u_1) + f_2^2(x_2)g_2^{-3}D^2K(u_2))]^2 du \end{aligned}$$

siendo $B_1(g_1^2(h_1^2 + h_2^2))$ y $B_2(g_2^2(h_1^2 + h_2^2))$ las definidas en (5.8) y (5.9), respectivamente.

Demostración. Consideramos la descomposición habitual en sesgo y varianza, $L(\mathbf{h}, \mathbf{g}) = B^2 + V$, donde

$$B = E_{Y|X} [B_{h,g}^*(\mathbf{x}) - B_h(\mathbf{x})]$$

y

$$V = \text{Var}_{Y|X} (B_{h,g}^*(\mathbf{x})).$$

El término de sesgo se puede aproximar por:

$$B \approx f_1^{-1}(x_1) f_2^{-1}(x_2) \times \left(\int (f_2(x_2) K_{g_1}(x_1 - s_1) + f_1(x_1) K_{g_2}(x_2 - s_2)) U_h(s) ds - U_h(\mathbf{x}) \right) \quad (5.7)$$

donde $U_h(\mathbf{x})$ está definido como:

$$\int (f_2(x_2) K_{h_1}(x_1 - s_1) + f_1(x_1) K_{h_2}(x_2 - s_2)) (m(s) - m(\mathbf{x})) f_1(s_1) f_2(s_2) ds.$$

Para llegar a esta expresión hay que hacer los siguientes cálculos:

$$B = E_{Y|\mathbf{X}} [B_{h,g}^*(\mathbf{x}) - B_h(\mathbf{x})],$$

donde

$$B_{h,g}^*(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) \sum_{j=1}^n w_j(\mathbf{X}_i, \mathbf{g}) Y_j - \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{g}) Y_i$$

y

$$B_h = \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) (Y_i - m(\mathbf{x})).$$

Entonces,

$$\begin{aligned} B &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) \sum_{j=1}^n w_j(\mathbf{X}_i, \mathbf{g}) - \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{g}) m(\mathbf{X}_i) - \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) m(\mathbf{x}) = \\ &= M_1 - M_2, \end{aligned}$$

no obstante, utilizando aproximaciones asintóticas de los pesos y aproximando la suma de Riemann por la integral, el término M_2 se expresa como sigue:

$$\begin{aligned} &n^{-1} f_1^{-1}(x_1) f_2^{-1}(x_2) \sum_{i=1}^n (f_2(x_2) K_{h_1}(X_{1i} - x_1) + f_1(x_1) K_{h_2}(X_{2i} - x_2)) \times \\ &\quad \times (m(\mathbf{X}_i) - m(\mathbf{x})) \approx \\ &\approx f_1^{-1}(x_1) f_2^{-1}(x_2) \int (f_2(x_2) K_{h_1}(X_{1i} - x_1) + f_1(x_1) K_{h_2}(X_{2i} - x_2)) \times \\ &\quad \times (m(s) - m(\mathbf{x})) f_1(s_1) f_2(s_2) ds \approx f_1^{-1}(x_1) f_2^{-1}(x_2) U_h(\mathbf{x}). \end{aligned}$$

De igual forma, en el primer término, M_1 , si aproximamos asintóticamente los pesos y las sumas de Riemann por integrales tenemos que

$$\begin{aligned} \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{g}) &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{g}) \sum_{j=1}^n w_j(\mathbf{X}_i, \mathbf{h}) m(\mathbf{X}_i) \approx \\ &\approx f_1^{-1}(x_1) f_2^{-1}(x_2) \int (f_2(x_2) K_{g_1}(s_1 - x_1) + f_1(x_1) K_{g_2}(s_2 - x_2)) \times \\ &\quad \times \int (f_2(s_2) K_{h_1}(t_1 - s_1) + f_1(s_1) K_{h_2}(t_2 - s_2)) f_1(t_1) f_2(t_2) m(\mathbf{s}) ds dt, \end{aligned}$$

puesto que $\sum_{j=1}^n w_j(\mathbf{X}_i, \mathbf{h}) = 1$. Para la primera parte del término M_1 ,

$$\begin{aligned} \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) \left(\sum_{j=1}^n w_j(\mathbf{X}_i, \mathbf{g}) m(\mathbf{X}_j) \right) &\approx \\ &\approx f_1^{-1}(x_1) f_2^{-1}(x_2) \int (f_2(x_2) K_{g_1}(s_1 - x_1) + f_1(x_1) K_{g_2}(s_2 - x_2)) \times \\ &\quad \times \int (f_2(s_2) K_{h_1}(t_1 - s_1) + f_1(s_1) K_{h_2}(t_2 - s_2)) f_1(t_1) f_2(t_2) m(\mathbf{t}) ds dt \end{aligned}$$

y sustituyendo tenemos que

$$\begin{aligned} &f_1^{-1}(x_1) f_2^{-1}(x_2) \int (f_2(x_2) K_{g_1}(s_1 - x_1) + f_1(x_1) K_{g_2}(s_2 - x_2)) \times \\ &\quad \times \int (f_2(s_2) K_{h_1}(t_1 - s_1) + f_1(s_1) K_{h_2}(t_2 - s_2)) f_1(t_1) f_2(t_2) (m(\mathbf{t}) - m(\mathbf{s})) dt ds = \\ &= f_1^{-1}(x_1) f_2^{-1}(x_2) \int (f_2(x_2) K_{g_1}(s_1 - x_1) + f_1(x_1) K_{g_2}(s_2 - x_2)) U_{\mathbf{h}}(\mathbf{s}) ds = M_1. \end{aligned}$$

Entonces, con las expresiones dadas para B_1 y B_2 , llegamos a la aproximación del sesgo dada en (5.7).

Sean $D_i m_i(x_i)$ la primera derivada parcial de la función m_i respecto a la i -ésima coordenada y $D_i^j m_i(x_i)$ la derivada j -ésima de la función m_i respecto a la i -ésima variable.

Pues bien, si integramos y aplicamos el desarrollo de Taylor a f y m , obtenemos

la siguiente expresión desarrollada para $U_h(\mathbf{x})$

$$\begin{aligned}
U_h(\mathbf{x}) &= \int_{\mathbf{u}} (f_2(x_2) K_{h_1}(u_1) + f_1(x_1) K_{h_2}(u_2)) \sum_{i=1}^2 (D_i m_i(x_i) u_i + \frac{1}{2} D_i^2 m_i(x_i) u_i^2) \times \\
&\quad \times (f_1(x_1) + D_1 f_1(x_1) u_1) (f_2(x_2) + D_2 f_2(x_2) u_2) d\mathbf{u} = \\
&= \frac{(b-a)^2}{2} f_1(x_1) f_2(x_2) \sum_{i=1}^2 \left(D_i m_i(x_i) f_i(x_i) + \frac{(b-a)^2}{2} \frac{1}{2} D_i^2 m_i(x_i) D_i f_i(x_i) \right) + \\
&\quad + \frac{(b-a)^3}{3} f_1(x_1) f_2(x_2) \sum_{i=1}^2 \left(D_i m_i(x_i) D_i f_i(x_i) + \frac{1}{2} D_i^2 m_i(x_i) \right) + \\
&\quad + (b-a) \left[f_2(x_2) \int_{u_1} K_{h_1}(u_1) (D_1 m_1(x_1) D_1 f_1(x_1) f_2(x_2) + \right. \\
&\quad + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1)) u_1^2 du_1 + f_1(x_1) \int_{u_2} K_{h_2}(u_2) (D_2 m_2(x_2) D_2 f_2(x_2) f_1(x_1) + \\
&\quad + \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2)) u_2^2 du_2 \left. \right] + \\
&\quad + \frac{(b-a)^2}{2} \left[D_2 f_2(x_2) f_2(x_2) \int_{u_1} K_{h_1}(u_1) (D_1 m_1(x_1) D_1 f_1(x_1) + \right. \\
&\quad + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1)) u_1^2 du_1 + D_1 f_1(x_1) f_1(x_1) \times \\
&\quad \left. \times \int_{u_2} K_{h_2}(u_2) (D_2 m_2(x_2) D_2 f_2(x_2) + \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2)) u_2^2 du_2 \right]
\end{aligned}$$

A continuación, calculamos las dos primeras derivadas parciales de $U_h(\mathbf{x})$:

$$\begin{aligned}
\frac{DU_h(\mathbf{x})}{\partial x_1} &= A_1 + (b-a) \left[f_2(x_2) \int_{u_1} K_{h_1}(u_1) \{ D_1^2 m_1(x_1) D_1 f_1(x_1) f_2(x_2) + \right. \\
&\quad + D_1 m_1(x_1) D_1^2 f_1(x_1) f_2(x_2) + \frac{1}{2} D_1^3 m_1(x_1) f_1(x_1) + \\
&\quad \left. + \frac{1}{2} D_1^2 m_1(x_1) D_1 f_1(x_1) \} u_1^2 du_1 + \right.
\end{aligned}$$

$$\begin{aligned}
& + D_1 f_1(x_1) \int_{u_2} K_{h_2}(u_2) \{D_2 m_2(x_2) D_2 f_2(x_2) f_1(x_1) + \\
& + \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2)\} u_2^2 du_2 + \\
& + f_1(x_1) \int_{u_2} K_{h_2}(u_2) D_1 f_1(x_1) D_2 m_2(x_2) D_2 f_2(x_2) u_2^2 du_2 \Big] + \\
& + \frac{(b-a)^2}{2} \left[D_2 f_2(x_2) f_2(x_2) \int_{u_1} K_{h_1}(u_1) \left\{ \frac{3}{2} D_1^2 m_1(x_1) D_1 f_1(x_1) + \right. \right. \\
& + D_1 m_1(x_1) D_1^2 f_1(x_1) + \left. \left. + \frac{1}{2} D_1^3 m_1(x_1) f_1(x_1) \right\} u_1^2 du_1 + \right. \\
& + \left. \left(D_1^2 f_1(x_1) f_1(x_1) + (D_1^2 f_1(x_1))^2 \right) \times \right. \\
& \left. \times \int_{u_2} K_{h_2}(u_2) \left\{ D_2 m_2(x_2) D_2 f_2(x_2) + \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2) \right\} u_2^2 du_2 \right],
\end{aligned}$$

$$\begin{aligned}
\frac{DU_h(\mathbf{x})}{\partial x_2} &= A_2 + (b-a) \left[f_1(x_1) \int_{u_2} K_{h_2}(u_2) \{D_2^2 m_2(x_2) D_2 f_2(x_2) f_1(x_1) + \right. \\
& + D_2 m_2(x_2) D_2^2 f_2(x_2) f_1(x_1) + \frac{1}{2} D_2^3 m_2(x_2) f_2(x_2) + \\
& + \frac{1}{2} D_2^2 m_2(x_2) D_2 f_2(x_2)\} u_2^2 du_2 + \\
& + D_2 f_2(x_2) \int_{u_1} K_{h_1}(u_1) \{D_1 m_1(x_1) D_1 f_1(x_1) f_2(x_2) + \\
& + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1)\} u_1^2 du_1 + \\
& + f_2(x_2) \int_{u_1} K_{h_1}(u_1) D_2 f_2(x_2) D_1 m_1(x_1) D_1 f_1(x_1) u_1^2 du_1 \Big] + \\
& + \frac{(b-a)^2}{2} \left[D_1 f_1(x_1) f_1(x_1) \int_{u_2} K_{h_2}(u_2) \left\{ \frac{3}{2} D_2^2 m_2(x_2) D_2 f_2(x_2) + \right. \right. \\
& + D_2 m_2(x_2) D_2^2 f_2(x_2) + \left. \left. + \frac{1}{2} D_2^3 m_2(x_2) f_2(x_2) \right\} u_2^2 du_2 + \right. \\
& + \left. \left(D_2^2 f_2(x_2) f_2(x_2) + (D_2^2 f_2(x_2))^2 \right) \times \right. \\
& \left. \times \int_{u_1} K_{h_1}(u_1) \left\{ D_1 m_1(x_1) D_1 f_1(x_1) + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1) \right\} u_1^2 du_1 \right],
\end{aligned}$$

$$\begin{aligned}
\frac{DU_h(\mathbf{x})}{\partial x_1^2} = & A_3 + (b-a) \left[f_2(x_2) \int_{u_1} K_{h_1}(u_1) \{ D_1^3 m_1(x_1) D_1 f_1(x_1) f_2(x_2) + \right. \\
& + 2D_1^2 m_1(x_1) D_1^2 f_1(x_1) f_2(x_2) + D_1 m_1(x_1) D_1^3 f_1(x_1) f_2(x_2) + \\
& + D_1^3 m_1(x_1) D_1 f_1(x_1) + \frac{1}{2} (D_1^4 m_1(x_1) f_1(x_1) + D_1^2 m_1(x_1) D_1^2 f_1(x_1)) \} u_1^2 du_1 + \\
& + (b-a) \int_{u_2} K_{h_2}(u_2) \{ D_1^2 f_1(x_1) (D_2 m_2(x_2) D_2 f_2(x_2) f_1(x_1) + \\
& + \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2)) + 2D_2 m_2(x_2) D_2 f_2(x_2) (D_1 f_1(x_1))^2 + \\
& + f_1(x_1) D_1^2 f_1(x_1) D_2 m_2(x_2) D_2 f_2(x_2) \} u_2^2 du_2 + \\
& + \frac{(b-a)^2}{2} \left[D_2 f_2(x_2) f_2(x_2) \int_{u_1} K_{h_1}(u_1) \{ 2D_1^3 m_1(x_1) D_1 f_1(x_1) + \right. \\
& + \frac{5}{2} D_1^2 m_1(x_1) D_1^2 f_1(x_1) + D_1 m_1(x_1) D_1^3 f_1(x_1) + \frac{1}{2} D_1^4 m_1(x_1) f_1(x_1) \} u_1^2 du_1 + \\
& + (D_1^3 f_1(x_1) f_1(x_1) + 3D_1^2 f_1(x_1) D_1 f_1(x_1)) \int_{u_2} K_{h_2}(u_2) \{ D_2 m_2(x_2) D_2 f_2(x_2) + \\
& \left. + \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2) \} u_2^2 du_2 \right],
\end{aligned}$$

$$\begin{aligned}
\frac{DU_h(\mathbf{x})}{\partial x_2^2} = & A_4 + (b-a) \left[f_1(x_1) \int_{u_2} K_{h_2}(u_2) \{ D_2^3 m_2(x_2) D_2 f_2(x_2) f_1(x_1) + \right. \\
& + 2D_2^2 m_2(x_2) D_2^2 f_2(x_2) f_1(x_1) + D_2 m_2(x_2) D_2^3 f_2(x_2) f_1(x_1) + \\
& + D_2^3 m_2(x_2) D_2 f_2(x_2) + \frac{1}{2} (D_2^4 m_2(x_2) f_2(x_2) + D_2^2 m_2(x_2) D_2^2 f_2(x_2)) \} u_2^2 du_2 + \\
& + (b-a) \int_{u_1} K_{h_1}(u_1) \{ D_2^2 f_2(x_2) (D_1 m_1(x_1) D_1 f_1(x_1) f_2(x_2) + \\
& + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1)) + 2D_1 m_1(x_1) D_1 f_1(x_1) (D_2 f_2(x_2))^2 + \\
& + f_2(x_2) D_2^2 f_2(x_2) D_1 m_1(x_1) D_1 f_1(x_1) \} u_1^2 du_1 + \\
& + \frac{(b-a)^2}{2} \left[D_1 f_1(x_1) f_1(x_1) \int_{u_2} K_{h_2}(u_2) \{ 2D_2^3 m_2(x_2) D_2 f_2(x_2) + \right. \\
& + \frac{5}{2} D_2^2 m_2(x_2) D_2^2 f_2(x_2) + D_2 m_2(x_2) D_2^3 f_2(x_2) + \frac{1}{2} D_2^4 m_2(x_2) f_2(x_2) \} u_2^2 du_2 + \\
& + (D_2^3 f_2(x_2) f_2(x_2) + 3D_2^2 f_2(x_2) D_2 f_2(x_2)) \int_{u_1} K_{h_1}(u_1) \{ D_1 m_1(x_1) D_1 f_1(x_1) + \\
& \left. + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1) \} u_1^2 du_1 \right],
\end{aligned}$$

donde A_1, \dots, A_4 son constantes¹, que no dependen de \mathbf{h} .

Si en (5.7) hacemos el cambio de variable $x_i - s_i = u_i$, $i = 1, 2$, y consideramos el desarrollo de Taylor

$$U_{\mathbf{h}}(\mathbf{x} + \mathbf{u}) - U_{\mathbf{h}}(\mathbf{x}) = DU_{\mathbf{h}}(\mathbf{x}) \mathbf{u} + \frac{1}{2} \mathbf{u}^T D^2 U_{\mathbf{h}}(\mathbf{x}) \mathbf{u}$$

se tiene que

$$\begin{aligned} B &\approx f_1^{-1}(x_1) \int K_{g_1}(u_1) \left\{ \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_2} u_2 + \frac{1}{2} \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_1^2} u_1^2 + \frac{1}{2} \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_2^2} u_2^2 \right\} du + \\ &+ f_2^{-1}(x_2) \int K_{g_2}(u_2) \left\{ \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_1} u_1 + \frac{1}{2} \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_1^2} u_1^2 + \frac{1}{2} \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_2^2} u_2^2 \right\} du = \\ &= \frac{1}{2} f_1^{-1}(x_1) \left[(b-a)^2 \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_2} + g_1^2 \mu_2(K) \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_1^2} + \frac{(b-a)^3}{3} \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_2^2} \right] + \\ &+ \frac{1}{2} f_2^{-1}(x_2) \left[(b-a)^2 \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_2} + \frac{(b-a)^3}{3} \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_1^2} + g_2^2 \mu_2(K) \frac{DU_{\mathbf{h}}(\mathbf{x})}{\partial x_2^2} \right], \end{aligned}$$

donde lo que interesa es la parte que depende a la vez de \mathbf{g} y de \mathbf{h} , por lo que nos queda la siguiente expresión

$$\begin{aligned} &\frac{1}{2} f_1^{-1}(x_1) g_1^2 \mu_2(K) \{ h_1^2 \mu_2(K) [(b-a) f_2(x_2) \{ D_1^3 m_1(x_1) D_1 f_1(x_1) f_2(x_2) + \\ &+ 2D_1^2 m_1(x_1) D_1^2 f_1(x_1) f_2(x_2) + D_1 m_1(x_1) D_1^3 f_1(x_1) f_2(x_2) + \\ &+ D_1^3 m_1(x_1) D_1 f_1(x_1) + \frac{1}{2} (D_1^4 m_1(x_1) f_1(x_1) + D_1^2 m_1(x_1) D_1^2 f_1(x_1)) \} + \\ &+ \frac{(b-a)^2}{2} D_2 f_2(x_2) f_2(x_2) \{ 2D_1^3 m_1(x_1) D_1 f_1(x_1) + \frac{5}{2} D_1^2 m_1(x_1) D_1^2 f_1(x_1) + \\ &+ D_1 m_1(x_1) D_1^3 f_1(x_1) + \frac{1}{2} D_1^4 m_1(x_1) f_1(x_1) \}] + \\ &+ h_2^2 \mu_2(K) [(b-a) D_1^2 f_1(x_1) \{ (D_2 m_2(x_2) D_2 f_2(x_2) f_1(x_1) + \\ &+ \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2)) + 2D_2 m_2(x_2) D_2 f_2(x_2) (D_1 f_1(x_1))^2 + \\ &+ f_1(x_1) D_1^2 f_1(x_1) D_2 m_2(x_2) D_2 f_2(x_2) \} + \\ &+ (D_1^3 f_1(x_1) f_1(x_1) + 3D_1^2 f_1(x_1) D_1 f_1(x_1)) \times \\ &\times (D_2 m_2(x_2) D_2 f_2(x_2) + \frac{1}{2} D_2^2 m_2(x_2) f_2(x_2))] \} + \end{aligned} \quad (5.8)$$

¹Las expresiones de estas constantes vienen recogidas en el apéndice B.

$$\begin{aligned}
& + \frac{1}{2} f_2^{-1}(x_2) g_2^2 \mu_2(K) \{ h_2^2 \mu_2(K) [(b-a) f_1(x_1) \{ D_2^3 m_2(x_2) D_2 f_2(x_2) f_1(x_1) + \\
& + 2D_2^2 m_2(x_2) D_2^2 f_2(x_2) f_1(x_1) + D_2 m_2(x_2) D_2^3 f_2(x_2) f_1(x_1) + \\
& + D_2^3 m_2(x_2) D_2 f_2(x_2) + \frac{1}{2} (D_2^4 m_2(x_2) f_2(x_2) + D_2^2 m_2(x_2) D_2^2 f_2(x_2)) \} + \\
& + \frac{(b-a)^2}{2} D_1 f_1(x_1) f_1(x_1) \{ 2D_2^3 m_2(x_2) D_2 f_2(x_2) + \frac{5}{2} D_2^2 m_2(x_2) D_2^2 f_2(x_2) + \\
& + D_2 m_2(x_2) D_2^3 f_2(x_2) + \frac{1}{2} D_2^4 m_2(x_2) f_2(x_2) \}] + \\
& + h_1^2 \mu_2(K) [(b-a) D_2^2 f_2(x_2) \{ (D_1 m_1(x_1) D_1 f_1(x_1) f_2(x_2) + \\
& + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1)) + 2D_1 m_1(x_1) D_1 f_1(x_1) (D_2 f_2(x_2))^2 + \\
& + f_2(x_2) D_2^2 f_2(x_2) D_1 m_1(x_1) D_1 f_1(x_1) \} + \\
& + (D_2^3 f_2(x_2) f_2(x_2) + 3D_2^2 f_2(x_2) D_2 f_2(x_2)) \times \\
& \times (D_1 m_1(x_1) D_1 f_1(x_1) + \frac{1}{2} D_1^2 m_1(x_1) f_1(x_1))] \} + \\
& = B_1 (g_1^2 (h_1^2 + h_2^2)) + B_2 (g_2^2 (h_1^2 + h_2^2)). \tag{5.9}
\end{aligned}$$

Consideramos ahora la varianza. De nuevo usando las representaciones lineales de los pesos

$$\begin{aligned}
B_{h,g}^*(\mathbf{x}) &= \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{h}) \sum_{j=1}^n w_j(\mathbf{X}_i, \mathbf{g}) Y_j - \sum_{i=1}^n w_i(\mathbf{x}, \mathbf{g}) Y_i \\
B_{h,g}^*(\mathbf{x}) &= \sum_{i=1}^n Y_i \left(\sum_{j=1}^n w_j(\mathbf{x}, \mathbf{h}) w_i(\mathbf{X}_j, \mathbf{g}) - w_i(\mathbf{x}, \mathbf{g}) \right) = \sum_{i=1}^n \tilde{w}_i(\mathbf{x}) Y_i.
\end{aligned}$$

Entonces, $V = \sum_{i=1}^n (\tilde{w}_i(\mathbf{x}))^2 \sigma(\mathbf{X}_i)$, y aproximando las sumas de Riemann por integrales se tiene que

$$\begin{aligned}
\tilde{w}_i(\mathbf{x}) &\approx \int (f_2(x_2) K_{h_1}(x_1 - x_{1j}) + f_1(x_1) K_{h_2}(x_2 - x_{2j})) \times \\
&\quad \times (f_2(x_{2j}) K_{g_1}(x_{1j} - x_{1i}) + f_1(x_{1j}) K_{g_2}(x_{2j} - x_{2i})) \, d\mathbf{x}_j - \\
&\quad - (f_2(x_2) K_{g_1}(x_1 - x_{1i}) + f_1(x_1) K_{g_2}(x_2 - x_{2i})) = \\
&= \int (f_2(x_2) K_{h_1}(x_1 - x_{1j}) + f_1(x_1) K_{h_2}(x_2 - x_{2j})) \times \\
&\quad \times \{ (f_2(x_{2j}) K_{g_1}(x_{1j} - x_{1i}) + f_1(x_{1j}) K_{g_2}(x_{2j} - x_{2i})) - \\
&\quad - (f_2(x_2) K_{g_1}(x_1 - x_{1i}) + f_1(x_1) K_{g_2}(x_2 - x_{2i})) \} \, d\mathbf{x}_j =
\end{aligned}$$

$$= \int (f_2(x_2) K_{h_1}(u_1) + f_1(x_1) K_{h_2}(u_2)) \times \\ \times \{ (f_2(x_2) K_{g_1}(x_1 + u_1 - x_{1i}) + f_1(x_1) K_{g_2}(x_2 + u_2 - x_{2i})) - \\ - (f_2(x_2) K_{g_1}(x_1 - x_{1i}) + f_1(x_1) K_{g_2}(x_2 - x_{2i})) \} du$$

$$V \approx n^{-1} f_1^{-2}(x_1) f_2^{-2}(x_2) \int \sigma(s) \int [(f_2(x_2) K_{h_1}(u_1) + f_1(x_1) K_{h_2}(u_2)) \times \\ \times \{ (f_2(x_2) K_{g_1}(x_1 + u_1 - s_1) + f_1(x_1) K_{g_2}(x_2 + u_2 - s_2)) - \\ - (f_2(x_2) K_{g_1}(x_1 - s_1) + f_1(x_1) K_{g_2}(x_2 - s_2)) \} du]^2 f(s) ds$$

Consideramos el desarrollo de Taylor, $i = 1, 2$

$$K_{g_i}(x_i + u_i - s_i) - K_{g_i}(x_i - s_i) = g_i^{-2} DK(g_i^{-1}(x_i - s_i)) u_i + \\ + \frac{1}{2} g_i^{-3} D^2 K(g_i^{-1}(x_i - s_i)) u_i^2$$

y sustituyendo en V tenemos

$$V \approx \frac{1}{4} n^{-1} f_1^{-2}(x_1) f_2^{-2}(x_2) \int \sigma(s) [(b-a) \mu_2(K) \times \\ \times \{ f_2^2(x_2) h_1^2 g_1^{-3} D^2 K(g_1^{-1}(x_1 - s_1)) + f_1^2(x_1) h_2^2 g_2^{-3} D^2 K(g_2^{-1}(x_2 - s_2)) \} + \\ + (b-a)^2 f_1(x_1) f_2(x_2) \times \\ \times \{ g_1^{-2} DK(g_1^{-1}(x_1 - s_1)) + g_2^{-2} DK(g_2^{-1}(x_2 - s_2)) \} + \\ + \frac{(b-a)^3}{3} \{ f_1^2(x_1) g_1^{-3} D^2 K(g_1^{-1}(x_1 - s_1)) + \\ + f_2^2(x_2) g_2^{-3} D^2 K(g_2^{-1}(x_2 - s_2)) \}]^2 f(s) ds,$$

y desarrollando por Taylor las funciones σ y f se tiene por continuidad que

$$V \approx \frac{1}{4} n^{-1} f_1^{-1}(x_1) f_2^{-1}(x_2) g_1 g_2 \sigma(x) \int [(b-a) \mu_2(K) \{ f_2^2(x_2) h_1^2 g_1^{-3} D^2 K(u_1) + \\ + f_1^2(x_1) h_2^2 g_2^{-3} D^2 K(u_2) \} + (b-a)^2 f_1(x_1) f_2(x_2) \{ g_1^{-2} DK(u_1) + \\ + g_2^{-2} DK(u_2) \} + \frac{(b-a)^3}{3} \{ f_1^2(x_1) g_1^{-3} D^2 K(u_1) + \\ + f_2^2(x_2) g_2^{-3} D^2 K(u_2) \}]^2 du.$$

□

5.4. Estudio de simulación

En el estudio de simulación realizado hemos considerado tres modelos teóricos (el primero de ellos también utilizado en el capítulo anterior).

Las funciones que definen cada uno de los modelos aditivos considerados son las siguientes:

Modelo 1

$$m_1(x_1) = 1 - 6x_1 + 36x_1^2 - 53x_1^3 + 22x_1^5,$$

y

$$m_2(x_2) = \text{sen}(5\pi x_2).$$

Donde los residuos se han generado según una distribución $\mathcal{N}(0, 0.5)$, y el vector de covariables, (X_1, X_2) , según la siguiente distribución normal bivalente,

$$(X_1, X_2) \rightarrow \mathcal{N}\left(\left(\begin{array}{c} 1/2 \\ 1/2 \end{array}\right), 1/9 \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right)\right).$$

Modelo 2

$$m_1(x_1) = \frac{1}{2}\text{sen}(\pi x_1)$$

y

$$m_2(x_2) = \text{sen}(4\pi x_2).$$

En este caso los residuos se han generado según una distribución $\mathcal{N}(0, 0.25)$, y el vector de covariables, (X_1, X_2) , según una uniforme, $(X_1, X_2) \rightarrow U(0, 1)$.

Modelo 3

$$m_1(x_1) = x_1^2$$

y

$$m_2(x_2) = x_2^2.$$

De nuevo los residuos se han generado según una distribución $\mathcal{N}(0, 0.25)$, y el vector de covariables, (X_1, X_2) , según una uniforme.

Las funciones para cada componente y modelo están representadas en las figuras 5.1, 5.2 y 5.3.

El núcleo utilizado para el estimador lineal local ha sido el gaussiano, definido como $K(x) = (2\pi)^{(-1/2)}\exp(-x^2/2)$.

Para todos los modelos se ha considerado un tamaño de muestra de $n = 100$, y el número de réplicas considerado ha sido, también, de 100, con objeto de eliminar el efecto de la muestra.

El estimador lineal local se ha evaluado sobre una red equiespaciada de 20×20 puntos, para los tres modelos, dentro del intervalo de definición.

En esta situación hemos evaluado en todos los casos el comportamiento del selector *bootstrap* del ancho de banda con una elección particular del parámetro piloto.

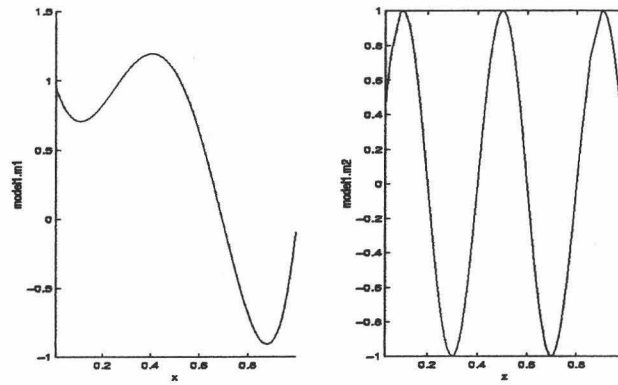


Figura 5.1: Componentes para el modelo 1.

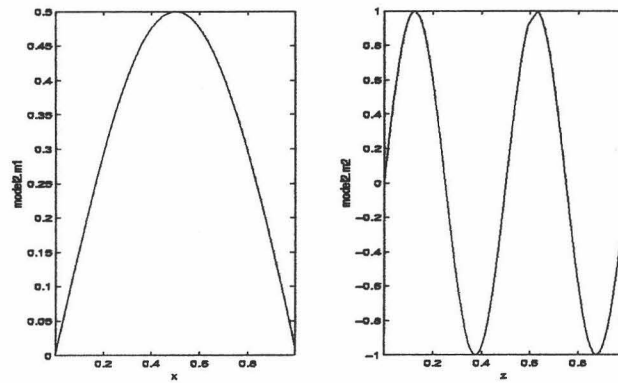


Figura 5.2: Componentes para el modelo 2.

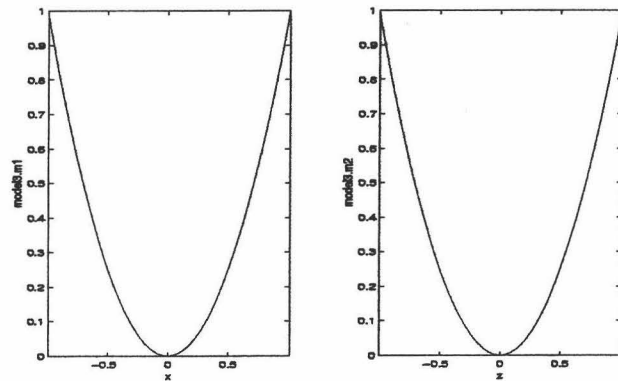


Figura 5.3: Componentes para el modelo 3.

5.4.1. Estudio del estimador *bootstrap* del error cuadrático medio

En esta sección estudiamos el comportamiento en muestras finitas del error cuadrático medio teórico, $MSE_h(\mathbf{x})$, y de su estimador *bootstrap*, $MSE_{h,g}^*(\mathbf{x})$. Estas medidas son locales, y se trata de ver la proximidad de ambas medidas en cada punto bidimensional.

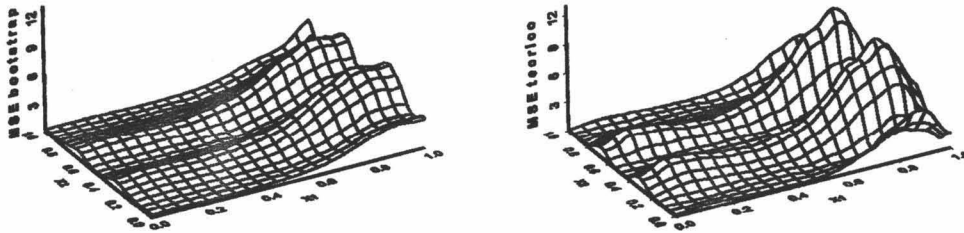


Figura 5.4: MSE *bootstrap* (a la izquierda) y teórico (a la derecha) para el modelo 1.

En las figuras 5.4, 5.5 y 5.6 se representan gráficamente las curvas $MSE_h^*(\mathbf{x})$ (a la izquierda de la figura) y $MSE_{h,g}(\mathbf{x})$ (a la derecha) para los tres modelos de prueba considerados. En el cálculo de la estimación *bootstrap* del error cuadrático medio, MSE^* , se ha utilizado un ancho de banda piloto seleccionado por un método sencillo de validación cruzada.

Sobre tales gráficos se pueden hacer los siguientes comentarios:

1. Tanto para el primer modelo como para el segundo se observa que las curvas para MSE *bootstrap* son más suaves que las curvas de MSE teóricas. Esta diferencia, en el caso del modelo 1, es más evidente en un entorno del punto 0.1 y en un entorno de 0.9, ambos para la variable x_1 ; ya que en estos puntos es donde se producen cambios bruscos en la curvatura de esta función. En el segundo modelo la diferencia es mayor, en general, en los valores de la segunda variable.

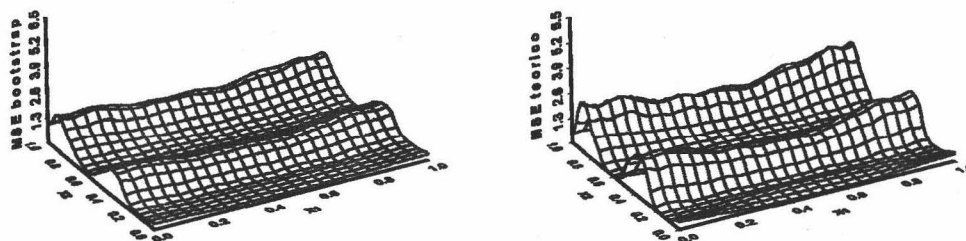


Figura 5.5: MSE *bootstrap* (a la izquierda) y teórico (a la derecha) para el modelo 2.

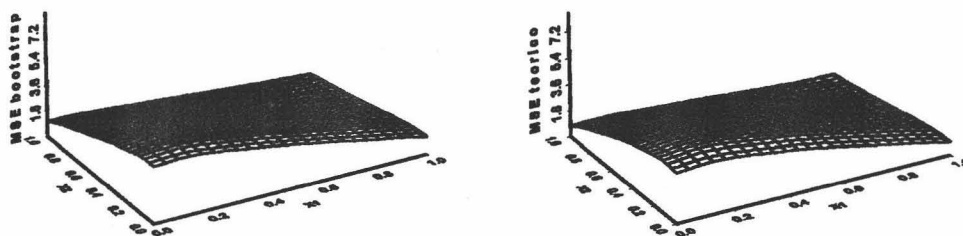


Figura 5.6: MSE *bootstrap* (a la izquierda) y teórico (a la derecha) para el modelo 3.

2. Sin embargo, se observa cómo, para el tercer modelo, debido a la simplicidad de las funciones, las curvas son prácticamente idénticas.

5.4.2. Modelos estimados con anchos de banda *bootstrap*

Se han calculado las curvas de regresión estimadas, utilizando el algoritmo *backfitting* con suavizadores lineales locales, para cada una de las réplicas simuladas de los tres modelos de prueba. En cada estimación se ha utilizado el ancho de banda local seleccionado según el procedimiento *wild bootstrap* sobre los residuos, descrito en este capítulo, y el ancho de banda local teórico.

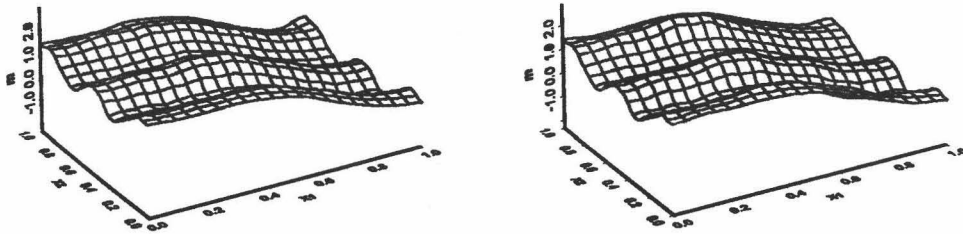


Figura 5.7: Función de regresión estimada con el ancho de banda *bootstrap* (a la izquierda) y estimada con el ancho de banda teórica (a la derecha) para el modelo 1.

En las figuras 5.7, 5.8 y 5.9 se representan gráficamente las curvas $\hat{m}_h^*(\mathbf{x})$, con h el ancho de banda local *bootstrap* (a la izquierda de la figura) y $\hat{m}_h(\mathbf{x})$, con h el ancho de banda local teórico (a la derecha) para los tres modelos de prueba considerados.

Sobre estos gráficos podemos hacer las mismas consideraciones que en la sección anterior, y aunque las diferencias en los dos primeros modelos no son tan evidentes, sí están presentes. De todas formas, resulta importante la buena descripción de la curva de regresión teórica en todos los casos.

Observamos por último las diferencias entre los anchos de banda *bootstrap* y teóricos, para comprender mejor las diferencias entre las estimaciones.

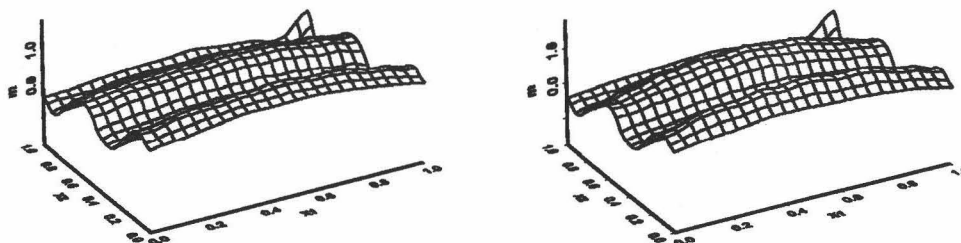


Figura 5.8: Función de regresión estimada con el ancho de banda *bootstrap* (a la izquierda) y estimada con el ancho de banda teórica (a la derecha) para el modelo 2.

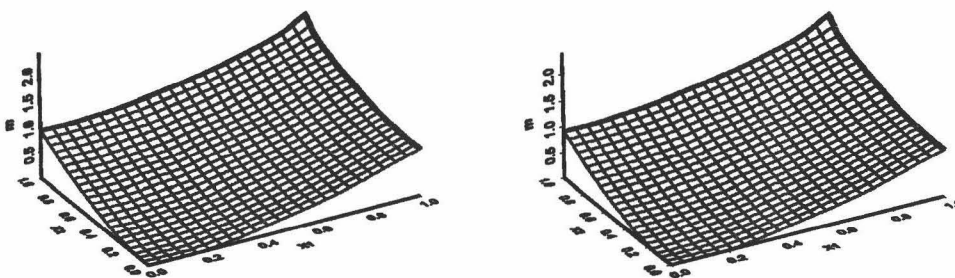


Figura 5.9: Función de regresión estimada con el ancho de banda *bootstrap* (a la izquierda) y estimada con el ancho de banda teórica (a la derecha) para el modelo 3.

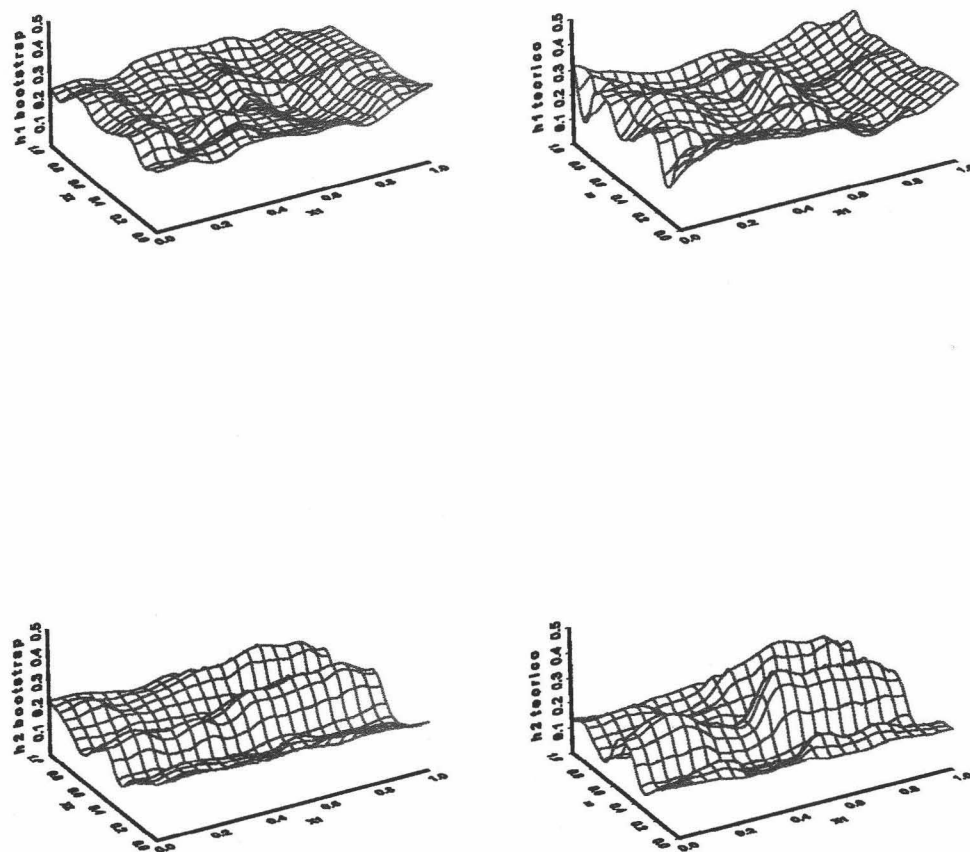


Figura 5.10: Ancho de banda *bootstrap* (a la izquierda) y teórico (a la derecha) para cada componente del modelo 1. Cada fila representa el ancho de banda para cada componente.

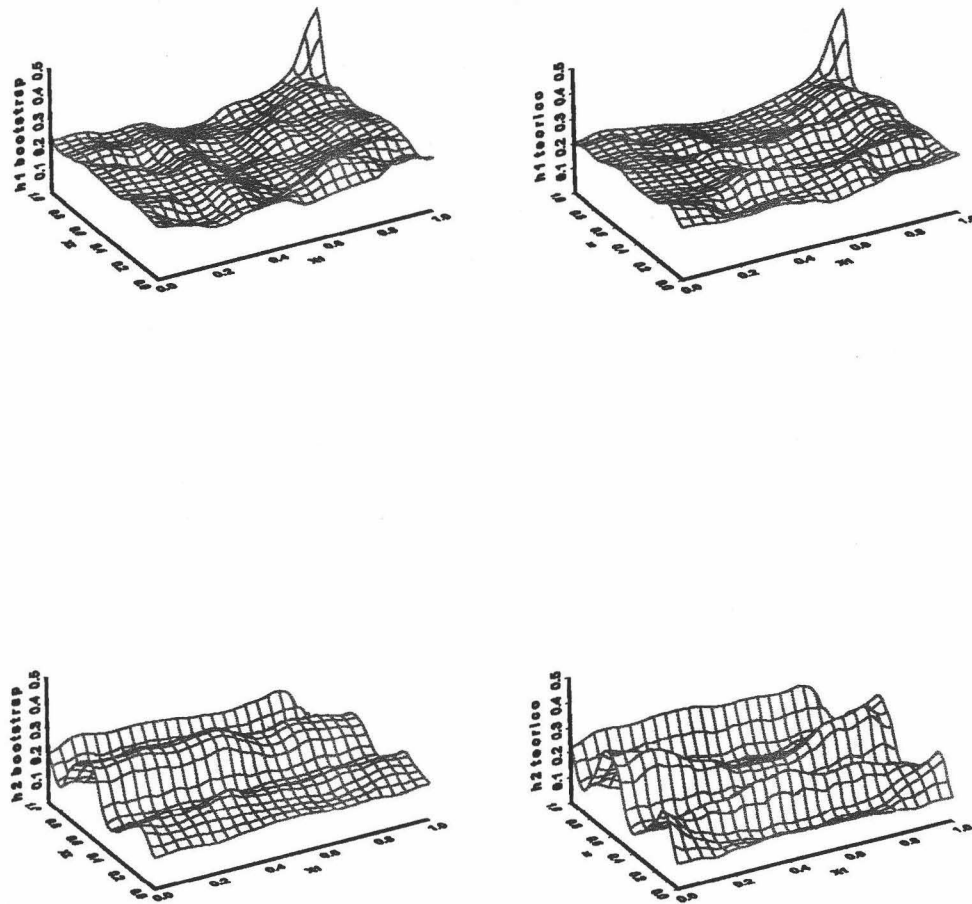


Figura 5.11: Ancho de banda *bootstrap* (a la izquierda) y teórico (a la derecha) para cada componente del modelo 2. Cada fila representa el ancho de banda para cada componente.

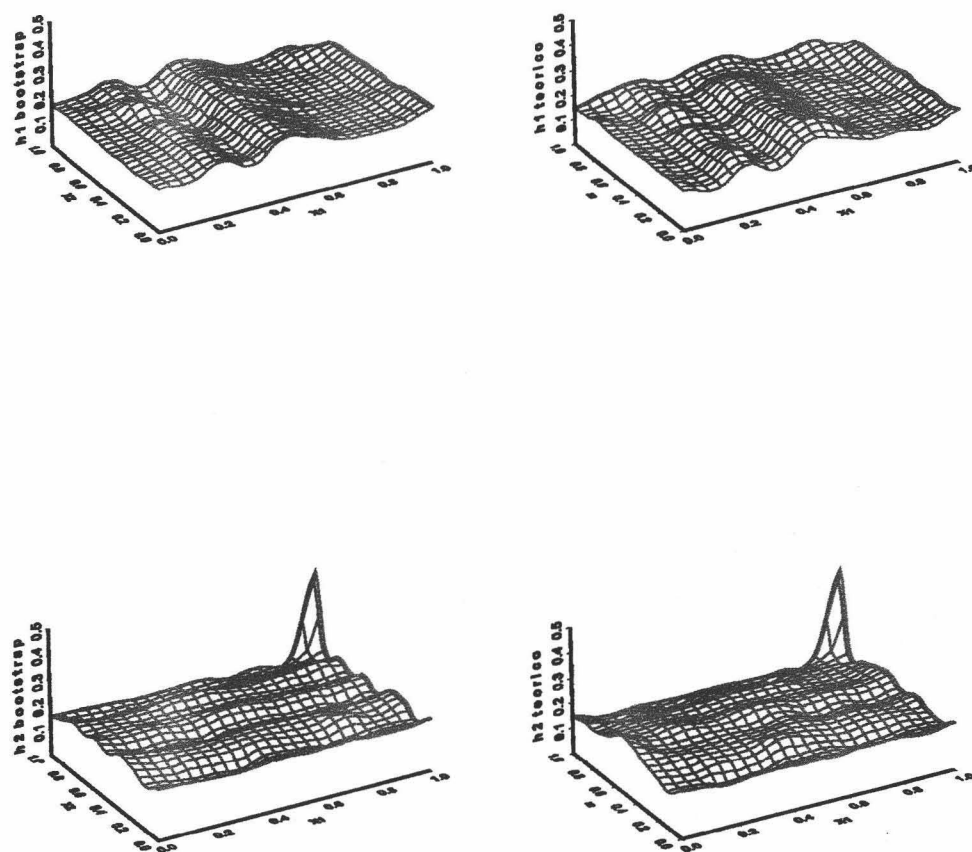


Figura 5.12: Ancho de banda *bootstrap* (a la izquierda) y teórico (a la derecha) para cada componente del modelo 3. Cada fila representa el ancho de banda para cada componente.

En las figuras 5.10, 5.11 y 5.12 se representan, por filas, el ancho de banda *bootstrap* y teórico, para cada componente.

Es evidente, a la vista de los gráficos, la necesidad de utilizar un ancho de banda local, sobre todo para los dos primeros modelos que presentan mayores dificultades. La mayor suavidad que presentan los anchos de banda locales *bootstrap* con respecto a los teóricos, en este caso, se explica teniendo presente el hecho de que puntos próximos deberían describir valores parecidos del ancho de banda local.

Es claro que para el tercer modelo los anchos de banda son muy poco variables debido a que ni la función ni ninguna de las componentes presentan cambios bruscos.

5.5. Conclusiones

El estudio de simulación realizado muestra el buen comportamiento, en la práctica, del selector *bootstrap* del ancho de banda. En cuanto a los modelos estimados utilizando el algoritmo *backfitting* con suavizadores lineales locales y anchos de banda locales *bootstrap*, describen bastante bien las estructuras subyacentes.

Por tanto, para futuros trabajos se pretende generalizar esta técnica a modelos de dimensión mayor que 2 y con polinomios de grado superior, no teniendo que ser el grado del polinomio el mismo para todas las componentes, sin que esto añada demasiada complejidad al problema.

Apéndice A

Computación estadística

A continuación se presentan en varias secciones las funciones que se han definido en *Matlab*, versión 5.2.0.3084, para realizar las simulaciones. Entre ellas las hay propias, otras que son adaptaciones de las definidas por Chaudhuri y Marron (1999) para realizar los gráficos *SiZer Map* y las definidas por Opsomer y Ruppert (1998) para implementar su método *plug-in* de selección del ancho de banda en un modelo aditivo bivalente.

A.1. Funciones para la simulación de las observaciones

```
function [x,y]=modell(n,mu1,mu2,sigma1,sigma2,ro,filedatos)
% Esta función crea una muestra de tamaño n según el modelo 1
% Los argumentos son:
% n (tamaño de la muestra)
% mu1 (media de la primera variable)
% mu2 (media de la segunda variable)
% sigma1 (varianza de la primera variable)
% sigma2 (varianza de la segunda variable)
% ro (nivel de correlación entre las variables)
x=[], y=[], m1=[], m2=[], er=[], cov=0;
cov=sqrt(sigma1)*sqrt(sigma2)*ro;
i=1;
for i=1:n ;
    rand('state',sum(100*clock));
    x(i,:)=mvnrnd([mu1 mu2],[sigma1 cov; cov sigma2],1);
    while (any(x(i,:)<0)|any(x(i,:)>1));
        x(i,:)=mvnrnd([mu1 mu2],[sigma1 cov; cov sigma2],1);
    end
end
```

```

    end;
end;
m1=1-6*x(:,1)+36*(x(:,1).^2)-53*(x(:,1).^3)+22*(x(:,1).^5);
m2=sin(5*pi*x(:,2));
er=randn(n,1);
er=0.25*er;
y=m1+m2+er;
fi=fopen(filedatos,'w');
if (fi==0);
    disp('Error al abrir el fichero de resultados');
    stop;
else;
    for i=1:n;
        datos(i,:)=[x(i,1) x(i,2) y(i)];
        fprintf(fi,'%f %f %f \n',datos(i,:));
    end;
end;
fclose(fi);

```

```

function [x,y]=model2(n,filedatos)
% Esta función crea una muestra de tamaño n para el modelo 2
% Los argumentos son:
% n (tamaño de la muestra)
% filedatos (fichero donde se grabarán los resultados)
x=[], y=[], m1=[], m2=[], er=[];
x(:,1)=rand(n,1);
rand('state',sum(100*clock));
x(:,2)=rand(n,1);
m1=0.5*sin(pi*x(:,1));
m2=sin(4*pi*x(:,2));
er=randn(n,1);
er=0.25*er;
y=m1+m2+er;
[fi]=fopen(filedatos,'w');
if (fi==0);
    disp('Error al abrir el fichero donde se van a guardar los datos');
    stop;
else;
    for i=1:n;
        data(i,:)=[x(i,1),x(i,2),y(i)];
        fprintf(fi,'%f %f %f \n',data(i,:));
    end;
end;

```

```

    end;
    st=fclose(fi);
end;

function [data]=model3(n,filedatos)
% Esta función crea una muestra de tamaño n para el modelo 3
% Los argumentos son:
% n    (tamaño de la muestra)
% filedatos  (nombre del fichero donde se grabarán los datos)
x=[], y=[], m1=[], m2=[], er=[], data=[];
x(:,1)=rand(n,1);
rand('state',sum(100*clock));
x(:,2)=rand(n,1);
i=1;
for i=1:n/2;
    x(i,:)=-x(i,:);
end;
m1=x(:,1).^2;
m2=x(:,2).^2;
er=randn(n,1);
er=0.25*er;
y=m1+m2+er;
[fi]=fopen(filedatos,'w');
if (fi==0);
    disp('Error al abrir el fichero donde se van a guardar los datos');
    stop;
else;
    for i=1:n;
        data(i,:)=[x(i,1),x(i,2),y(i),m1(i),m2(i)];
        fprintf(fi,'%f %f %f %f %f \n',data(i,:));
    end;
    st=fclose(fi);
end;
end;

```

A.2. Funciones para la creación de gráficos *SiZer* *Map* en un modelo aditivo bivalente

```

function [bindat,bincent]=binr(data,vgridp)
% Calcula las cantidades binned asociadas a un conjunto de datos
% Los argumentos son:

```

```

% data (la matriz de datos, que debe ser de dimensión (n, 2))
% vgridp (un vector de parámetros para construir un grid de puntos)
xdat=data(:,1);
ydat=data(:,2);
n=length(xdat);
if nargin == 1;
    lend=min(xdat);
    rend=max(xdat);
    nbin=401;
else;
if length(vgridp)==1;
    lend=min(xdat);
    rend=max(xdat);
    nbin=401;
elseif length(vgridp)==2;
    lend=vgridp(1);
    rend=vgridp(2);
    nbin=401;
else;
    lend=vgridp(1);
    rend=vgridp(2);
    nbin=vgridp(3);
end;
end;
if lend<rend;
    bxdat=zeros(nbin,1);
    bydat=zeros(nbin,1);
    loflag=((xdat-lend)<(10^(-10)*(rend-lend)));
    numlo=sum(loflag);
    if numlo>0;
        if numlo==n;
            disp('!!!Atención, problemas con gplbinr: todos los datos por debajo del
                rango binned!!!');
        end;
        bxdat(1)=bxdat(1)+numlo;
        bydat(1)=bydat(1)+sum(ydat(loflag));
    end;
    hiflag=((xdat-rend)>-(10^(-10)*(rend-lend)));
    numhi=sum(hiflag);
    if numhi>0;
        if numhi==n;

```

```

disp('!!!Atención, problemas con gplbinr: todos los datos por encima del
      rango binned!!!');
end;
bxdat(nbin)=bxdat(nbin)+numhi;
bydat(nbin)=bydat(nbin)+sum(ydat(hiflag));
end;
end;
iflag=(~loflag)&(~hiflag);
numi=sum(iflag);
if numi>0;
    ixdat=xdat(iflag);
    iydat=ydat(iflag);
    isixdat=((nbin-1)*(ixdat-lend)./(rend-lend))+1;
    vibinl=floor(isixdat);
    vwt=isixdat-vibinl;
    for idati=1:numi;
        bxdat(vibinl(idati))=bxdat(vibinl(idati))+(1-vwt(idati));
        bxdat(vibinl(idati)+1)=bxdat(vibinl(idati)+1)+vwt(idati);
    end;
    for idati=1:numi;
        bydat(vibinl(idati))=bydat(vibinl(idati))+(1-vwt(idati)).*iydat(idati);
        bydat(vibinl(idati)+1)=bydat(vibinl(idati)+1)+vwt(idati).*iydat(idati);
    end;
end;
bindat=[bxdat bydat];
bincent=linspace(lend,rend,nbin)';

```

```

function makeplot=gaditivos(data,icolor,vrange,nhp,psfname)
% Gaditivos, hace una sólo página con dos tipos de gráficos, un primer gráfico que
% representa las estimaciones de las componentes para distintos valores del ancho
% de banda y el SiZer Map
% Se puede utilizar con 1, 2, 3, 4 ó 5 argumentos.
% Los argumentos son:
% data (matriz de datos ( $n \times 3$ ),
%      X en las dos primeras columnas e Y en la tercera)
% icolor (- 0 Utiliza una escala de grises
%        1 (o no especificado) utiliza rojo, azul y morado)
% vrange (- 0 (o no especificadp) utiliza el mínimo y el máximo de los datos
%        2 vector: utiliza minx como vrange(1) y maxx como vrange(2)
%        3 vector: utiliza minx como vrange(1) y maxx como vrange(2)
%        y el número de puntos del grid como vrange(3))

```

```
% nhp (- 0 utiliza un valor por defecto de 11
%      en otro caso, el valor indica el número de anchos de banda diferentes
%      considerados)
% psfname (- nombre para el fichero postscript)
%      (automáticamente añadirá ".ps")
%      (será un fichero postscript a color cuando icolor = 1)
if nargin==1;
    iicolor=1;
else;
    iicolor=icolor;
end;
if nargin<=2;
    ivrange=0;
else;
    ivrange=vrangle;
end;
    if nargin<=3;
        inhp=11;
    else;
        if nhp==0;
            inhp=11;
        else;
            inhp=nhp;
        end;
    end;
end;
if nargin<=4;
    ipsfname=[];
else;
    ipsfname=psfname;
end;
xdat=[data(:,1) data(:,2)];
ydat=data(:,3);
n=length(xdat);
disp('Working on family');
if length(ivrange)==1;
    mind=min(xdat);
    maxd=max(xdat);
    ngrid=100;
else;
    mind=[vrangle(1) vrangle(3)];
    maxd=[vrangle(2) vrangle(4)];
```



```

    if length(ivrange)==5;
        ngrid=vrange(5);
    else;
        ngrid=100;
    end;
end;
centd=mean([mind;maxd]);
range=maxd-mind;
binw=range./(ngrid-1);
hmin=2.*binw;
hmax=range;
[yhat,hcent,m,ybar]=addfit(xdat,ydat,[1 1]);
i=1;
for i=1:2;
    vh(:,i)=logspace(log10(hmin(i)),log10(hmax(i)),inhp);
end;
if hcent(1)<min(vh(:,1))|hcent(1)>max(vh(:,1));
    disp('!!!Error desde gaditivos: la base de datos de h excede este rango!!!');
    hcflag(1)=0;
else;
    hcflag(1)=1;
end;
if hcent(2)<min(vh(:,2))|hcent(2)>max(vh(:,2));
    disp('!!!Error desde gaditivos: la base de datos de h excede este rango!!!');
    hcflag(2)=0;
else;
    hcflag(2)=1;
end;
ycen=ydat-ybar;
coeflin(1,:)=polyfit(xdat(:,1),ycen,1);
coeflin(2,:)=polyfit(xdat(:,2),ycen,1);
mstart(:,1)=polyval(coeflin(1,:),xdat(:,1));
mstart(:,2)=polyval(coeflin(2,:),xdat(:,2));
inh=1;
for inh=1:inhp;
    Countmax=30, step=1, count=1;
    mnew=mstart;
    while(step>0.0001&count<Countmax);
        mold=mnew;
        [msmoo(:,1),xgrid1]=gestimates([xdat(:,1) ycen-mold(:,2)],vh(inh,1), ...
            [mind(1);maxd(1);ngrid]);
    end;
end;

```

```

[msmoo(:,2),xgrid2]=gestimates([xdat(:,2) ycen-mold(:,1)],vh(inh,2), ...
                               [mind(2);maxd(2);ngrid]);
msmoo=msmoo-ones(ngrid,1)*sum(msmoo)/ngrid;
mnew(:,1)=interp1(xgrid1,msmoo(:,1),xdat(:,1),'cubic');
mnew(:,2)=interp1(xgrid2,msmoo(:,2),xdat(:,2),'cubic');
step=sum(sum(mnew-mold).^2)/n;
count=count+1;
end;
bottom(1)=min(min(msmoo(:,1)));
top(1)=max(max(msmoo(:,1)));
range(1)=top(1)-bottom(1);
bottom(1)=bottom(1)-0.05*range(1);
top(1)=top(1)+0.05*range(1);
subplot(2,2,1), hold on;
plot(xgrid1,msmoo(:,1),'-k');
vax=axis;
vax([1,2])=[mind(1),maxd(1)];
axis(vax);
title('Family Plot');
bottom(2)=min(min(msmoo(:,2)));
top(2)=max(max(msmoo(:,2)));
range(2)=top(2)-bottom(2);
bottom(2)=bottom(2)-0.05*range(2);
top(2)=top(2)+0.05*range(2);
subplot(2,2,2), hold on;
plot(xgrid2,msmoo(:,2),'-k');
vax=axis;
vax([1,2])=[mind(2),maxd(2)];
axis(vax);
title('Family Plot');
end;
Countmax=30, step=1, count=1;
mnew=mstart;
while (step>0.0001&count<Countmax);
    mold=mnew;
    [msmoo(:,1),xgrid1]=gestimates([xdat(:,1) ycen-mold(:,2)],hcent(1), ...
                                   [mind(1);maxd(1);ngrid]);
    [msmoo(:,2),xgrid2]=gestimates([xdat(:,2) ycen-mold(:,1)],hcent(2), ...
                                   [mind(2);maxd(2);ngrid]);
    msmoo=msmoo-ones(ngrid,1)*sum(msmoo)/ngrid;
    mnew(:,1)=interp1(xgrid1,msmoo(:,1),xdat(:,1),'cubic');

```

```

    mnew(:,2)=interp1(xgrid2,msmoo(:,2),xdat(:,2),'cubic');
    step=sum(sum(mnew-mold).^2)/n;
    count=count+1;
end;
i=1, j=1;
for i=1:inhp
    if hcent(1)<=vh(i,1)
        j=j+1;
    end;
end;
j=j+1;
i=1, k=1;
for i=1:inhp
    if hcent(2)<=vh(i,2)
        k=k+1;
    end;
end;
k=k+1;
bottom(1)=min(min(msmoo(:,1)));
top(1)=max(max(msmoo(:,1)));
range(1)=top(1)-bottom(1);
bottom(1)=bottom(1)-0.05*range(1);
top(1)=top(1)+0.05*range(1);
subplot(2,2,1), hold on;
plot(xgrid1,msmoo(:,1),'-k');
vcurvh=get(gca,'Children');
set(vcurvh(j),'LineWidth',2);
hold off;
bottom(2)=min(min(msmoo(:,2)));
top(2)=max(max(msmoo(:,2)));
range(2)=top(2)-bottom(2);
bottom(2)=bottom(2)-0.05*range(2);
top(2)=top(2)+0.05*range(2);
subplot(2,2,2), hold on;
plot(xgrid2,msmoo(:,2),'-k');
vcurvh = get(gca,'Children');
set(vcurvh(k),'LineWidth',2);
hold off;
disp('Working on SiZer');
gsizer([xdat(:,1) xdat(:,2) ydat]);
subplot(2,2,3);

```

```

title('SiZer Plot');
hold on;
if hcflag(1)==1;
    plot([mind(1);maxd(1)],ones(2,1)*log10(hcent(1)),'-k');
end;
plot(centd(1)+2.*vh(:,1),log10(vh(:,1)),':k');
plot(centd(1)-2.*vh(:,1),log10(vh(:,1)),':k');
subplot(2,2,4);
title('SiZer Plot');
hold on;
if hcflag(2)==1;
    plot([mind(2);maxd(2)],ones(2,1)*log10(hcent(2)),'-k');
end;
plot(centd(2)+2.*vh(:,2),log10(vh(:,2)),':k');
plot(centd(2)-2.*vh(:,2),log10(vh(:,2)),':k');
if icolor~=0;
    cocomap=[0,0,1;0.35,0.35,0.35;0.5,0,0.5;1,0,0];
    colormap(cocomap);
end;
if length(ipsfname)>0;
    orient tall;
    if icolor~=0;
        eval(['print -dpsc \matlab\bin\' ipsfname '.ps']);
    else;
        eval(['print -dps \matlab\bin\' ipsfname '.ps']);
    end;
end;
end;

function [npr,xgrid,mker] = gestimates(data,vh,vxgrid)
% Hace regresión polinomial local utilizando técnicas binned.
% Los argumentos son:
% data (matriz  $n \times 2$  de observaciones)
% vh (vector de anchos de banda)
% vxgrid (vector of parámetros para el grid)
if nargin==1;
    ivh=0;
else;
    ivh=vh;
end;
if nargin<=2;
    ivxgrid=0;

```

```

else;
    ivxgrid=vxgrid;
end;
if length(ivxgrid)>3;
    disp('!!! Error !!!') ;
end;
nbin=401;
lend=min(data(:,1));
rend=max(data(:,1));
if (length(ivxgrid)==2)|(length(ivxgrid)==3);
    lend=ivxgrid(1);
    rend=ivxgrid(2);
end;
if length(ivxgrid)==3;
    nbin=ivxgrid(3);
end;
if lend>rend;
    disp('!!! Error en gpnpr !!!');
    bincts=[];
else;
    bincts=binr(data,[lend,rend,nbin]);
end;
xgrid=linspace(lend,rend,nbin)';
npr=[];
for ih=1:length(ivh);
    h=ivh(ih);
    delta=(rend-lend)/(nbin-1);
    k=nbin-1;
    arg=linspace(0,k*delta/h,k+1)';
    kvec0=exp(-(arg.^2)/2)/sqrt(2*pi);
    arg=arg*h;
    kvec1=kvec0.*arg;
    kvec2=kvec1.*arg;
    kvec0=[flipud(kvec0(2:k+1));kvec0];
    s0=conv(bincts(:,1),kvec0);
    s0=s0(k+1:k+nbin);
    sy0=conv(bincts(:,2),kvec0);
    sy0=sy0(k+1:k+nbin);
    kvec1=[-flipud(kvec1(2:k+1));kvec1];
    s1=conv(bincts(:,1),kvec1);
    s1=s1(k+1:k+nbin);

```

```

sy1=conv(bincts(:,2),kvec1);
sy1=sy1(k+1:k+nbin);
kvec2=[flipud(kvec2(2:k+1));kvec2];
s2=conv(bincts(:,1),kvec2);
s2=s2(k+1:k+nbin);
denom=s2.*s0-s1.*s1;
vflag0d=(denom/max([denom;eps]))<=eps;
sflag0d=sum(vflag0d);
if sflag0d>0;
    disp('!!! Error !!!');
    denom(vflag0d)=ones(sflag0d,1);
end;
nprh=(s2.*sy0-s1.*sy1)./denom;
if sflag0d>0;
    vflagbf=bincts(:,1)>0;
    txbdatt=xgrid(vflagbf);
    tybdatt=bincts(vflagbf,2)./bincts(vflagbf,1);
    if sflag0d==nbin;
        nprh=interp1s(txbdatt,tybdatt,xgrid);
    else;
        if vflag0d(1)==1;
            [temp,ifge]=max(1-vflag0d);
            [temp,ifpbag]=max(vflagbf(ifge:nbin));
            ifpbag=ifpbag+ifge-1;
            flag=txbdatt<xgrid(ifpbag);
            if sum(flag)>0;
                vx=[txbdatt(flag);xgrid(ifpbag)];
                vy=[tybdatt(flag);nprh(ifpbag)];
            else;
                vx=xgrid(ifpbag);
                vy=nprh(ifpbag);
            end;
            nprh(1:(ifpbag-1))=interp1s(vx,vy,xgrid(1:(ifpbag-1)));
            vflag0d(1:(ifpbag-1))=zeros(ifpbag-1,1);
        end;
        if vflag0d(nbin)==1;
            [temp,ilge]=max(flipud(1-vflag0d));
            ilge=nbin+1-ilge;
            [temp,ilpbbg]=max(flipud(vflagbf(1:ilge)));
            ilpbbg=ilge+1-ilpbbg;
            flag=txbdatt>xgrid(ilpbbg);
        end;
    end;
end;

```

```

    if sum(flag)>0;
        vx=[xgrid(ilpbbg);txbdat(flag)];
        vy=[nprh(ilpbbg);tybdat(flag)];
    else;
        vx=xgrid(ilpbbg);
        vy=nprh(ilpbbg);
    end;
    nprh((ilpbbg+1):nbin)=interp1s(txbdat,tybdat,xgrid((ilpbbg+1):nbin));
    vflag0d((ilpbbg+1):nbin)=zeros(nbin-ilpbbg,1);
end;
while sum(vflag0d)>0;
    [temp,ind0d]=max(vflag0d);
    [temp,inge]=max(1-vflag0d(ind0d:nbin));
    inge=inge+ind0d-1;
    [temp,inpbag]=max(vflagbf(inge:nbin));
    inpbag=inpbag+inge-1;
    [temp,ilge]=max(flipud(1-vflag0d(1:ind0d)));
    ilge=ind0d+1-ilge;
    [temp,ilpbbg]=max(flipud(vflagbf(1:ilge)));
    ilpbbg=ilge+1-ilpbbg;
    flag=(txbdat>xgrid(ilpbbg)).*(txbdat<xgrid(inpbag));
    if sum(flag)>0;
        vx=[xgrid(ilpbbg);txbdat(flag);xgrid(inpbag)];
        vy=[nprh(ilpbbg);tybdat(flag);nprh(inpbag)];
    else;
        vx=[xgrid(ilpbbg);xgrid(inpbag)];
        vy=[nprh(ilpbbg);nprh(inpbag)];
    end;
    nprh((ilpbbg+1):(inpbag-1))=interp1(vx,vy,...
        xgrid((ilpbbg+1):(inpbag-1)));
    vflag0d((ilpbbg+1):(inpbag-1))=zeros(inpbag-ilpbbg-1,1);
end;
end;
end;
npr=[npr nprh];
end;
if nargout==3;
    cent=mean([lend;rend]);
    mih=vec2mat(1./ivh',nbin);
    mker=vec2mat((xgrid-cent),length(ivh)).* mih;
    mker=exp(-mker.^2/2).*mih/sqrt(2*pi);
end;

```

```

    mker=0.05*mker;
end;

function [mapout,xgrid] = gsizer(data,vxgp,vhgp,eptflag,alpha,simflag)
% Función que genera el gráfico SiZer Map de un modelo aditivo bivalente
% Los argumentos son:
% data   (matriz  $n \times (D + 1)$  de observaciones)
% vxgp   (vector de parámetros para el grid de x)
% vhgp   (vector de parámetros para el grid de h)
if nargin==1;
    ivxgp=0;
else;
    ivxgp=vxgp;
end;
if nargin<=2;
    ivhgp=0;
else;
    ivhgp=vhgp;
end;
if nargin<=3;
    ieptflag=0;
else;
    ieptflag=eptflag;
end;
if nargin<=4;
    ialpha=0.05;
else;
    ialpha=alpha;
end ;
if nargin<=5;
    isimflag=1;
else;
    isimflag=simflag;
end ;
xdat=[data(:,1) data(:,2)];
ydat=data(:,3);
n=length(xdat);
if ivxgp==0;
    ivxgp=[min(xdat),max(xdat),401];
end;
left=[ivxgp(1) ivxgp(2)];

```



```

right=[ivxgp(3) ivxgp(4)];
ngrid=ivxgp(5);
range=right-left;
binw=range/(ngrid-1);
if ivhgp==0;
    ivhgp=[2*binw,range,21];
end;
hmin=[ivhgp(1) ivhgp(2)];
hmax=[ivhgp(3) ivhgp(4)];
nh=ivhgp(5);
i=1;
for i=1:2;
    vh(:,i)=logspace(log10(hmin(i)),log10(hmax(i)),nh)';
end;
ybar=sum(ydat)/n;
ycen=ydat-ybar;
coffin(1,:)=polyfit(xdat(:,1),ycen,1);
coffin(2,:)=polyfit(xdat(:,2),ycen,1);
mstart(:,1)=polyval(coffin(1,:),xdat(:,1));
mstart(:,2)=polyval(coffin(2,:),xdat(:,2));
mdsurf1=[], mdsurf2=[], mesurf1=[], mesurf2=[], mvsurf1=[], mvsurf2=[], vgq1=[],
vgq2=[];
inh=1;
for inh=1:nh;
    Countmax=30;
    step=1, count=1;
    mnew=mstart;
    while (step>0.0001 & count<Countmax);
        mold=mnew;
        [msmoo(:,1),xgrid1]=gestimates([xdat(:,1) ycen-mold(:,2)],vh(inh,1),...
            [ivxgp(1);ivxgp(3);ngrid]);
        [msmoo(:,2),xgrid2]=gestimates([xdat(:,2) ycen-mold(:,1)],vh(inh,2),...
            [ivxgp(2);ivxgp(4);ngrid]);
        msmoo=msmoo-ones(ngrid,1)*sum(msmoo)/ngrid;
        mnew(:,1)=interp1(xgrid1,msmoo(:,1),xdat(:,1),'cubic');
        mnew(:,2)=interp1(xgrid2,msmoo(:,2),xdat(:,2),'cubic');
        step=sum(sum(mnew-mold).^2)/n;
        count=count+1;
    end;
    bincts1=binr([xdat(:,1) (ycen-mnew(:,2))],[ivxgp(1) ivxgp(3) ivxgp(5)]);
    bincts12=binr([xdat(:,1) (ycen-mnew(:,2)).^2],[ivxgp(1) ivxgp(3) ivxgp(5)]);

```

```

bincts2=binr([xdat(:,2) (ycen-mnew(:,1))],[ivxgp(2) ivxgp(4) ivxgp(5)]);
bincts22=binr([xdat(:,2) (ycen-mnew(:,1)).^2],[ivxgp(2) ivxgp(4) ivxgp(5)]);
delta=(right-left)/(ngrid-1);
k=ngrid-1;
arg1=linspace(0,k*delta(1)/vh(inh,1),k+1)';
kvec1=exp(-(arg1.^2)/2);
kvec1=[flipud(kvec1(2:k+1));kvec1];
arg2=linspace(0,k*delta(2)/vh(inh,2),k+1)';
kvec2=exp(-(arg2.^2)/2);
kvec2=[flipud(kvec2(2:k+1));kvec2];
ve1=conv(bincts1(:,1),kvec1);
ve1=ve1(k+1:k+ngrid);
ve2=conv(bincts2(:,1),kvec2);
ve2=ve2(k+1:k+ngrid);
flag1=(ve1<1);
flag2=(ve2<1);
ve1(flag1)=ones(sum(flag1),1);
ve2(flag2)=ones(sum(flag2),1);
s11=conv(bincts1(:,1).* xgrid1,kvec1);
s11=s11(k+1:k+ngrid);
s12=conv(bincts1(:,1).*xgrid1.^2,kvec1);
s12=s12(k+1:k+ngrid);
t10=conv(bincts1(:,2),kvec1);
t10=t10(k+1:k+ngrid);
s21=conv(bincts2(:,1).* xgrid2,kvec2);
s21=s21(k+1:k+ngrid);
s22=conv(bincts2(:,1).* xgrid2.^2,kvec2);
s22=s22(k+1:k+ngrid);
t20=conv(bincts2(:,2),kvec2);
t20=t20(k+1:k+ngrid);
xbar1=conv(bincts1(:,1).* xgrid1,kvec1);
xbar1=xbar1(k+1:k+ngrid);
xbar2=conv(bincts2(:,1).*xgrid2,kvec2);
xbar2=xbar2(k+1:k+ngrid);
xbar1=xbar1./ve1;
xbar2=xbar2./ve2;
t11=conv(bincts1(:,2).* xgrid1,kvec1);
t11=t11(k+1:k+ngrid);
t21=conv(bincts2(:,2).*xgrid2,kvec2);
t21=t21(k+1:k+ngrid);
numerd1=t11-t10.*xbar1;

```

```

numerd2=t21-t20.*xbar2;
denomd1=s12-2*s11.*xbar1+ve1.*xbar1.^2;
denomd2=s22-2*s21.*xbar2+ve2.*xbar2.^2;
flag12=denomd1<(10^(-10)*mean(denomd1));
flag22=denomd2<(10^(-10)*mean(denomd2));
ve1(flag12)=ones(sum(flag12),1);
ve2(flag22)=ones(sum(flag22),1);
flag1=flag1-flag12;
flag2=flag2-flag22;
denomd1(flag1)=ones(sum(flag1),1);
denomd2(flag2)=ones(sum(flag2),1);
mhat1=t10./ve1;
vd1=numerd1./denomd1;
mhat2=t20./ve2;
vd2=numerd2./denomd2;
sig12=conv(bincts12(:,2),kvec1);
sig12=sig12(k+1:k+ngrid);
sig12=sig12./ve1-mhat1.^2;
sig22=conv(bincts22(:,2),kvec2);
sig22=sig22(k+1:k+ngrid);
sig22=sig22./ve2-mhat2.^2;
flag12=sig12<(10^(-10)*mean(sig12));
ve1(flag12)=ones(sum(flag12),1);
flag22=sig22<(10^(-10)*mean(sig22));
ve2(flag22)=ones(sum(flag22),1);
flag1=flag1-flag12;
flag2=flag2-flag22;
sig12(flag1)=ones(sum(flag1),1);
sig22(flag2)=ones(sum(flag2),1);
rho12=vd1.^2.*(denomd1./(sig12.*ve1));
sig2res1=(1-rho12).*sig12;
rho22=vd2.^2.*(denomd2./(sig22.*ve2));
sig2res2=(1-rho22).*sig22;
u10=conv(bincts1(:,1).*sig2res1,kvec1.^2);
u10=u10(k+1:k+ngrid);
u11=conv(bincts1(:,1).*sig2res1.*xgrid1,kvec1.^2);
u11=u11(k+1:k+ngrid);
u12=conv(bincts1(:,1).*sig2res1.*xgrid1.^2,kvec1.^2);
u12=u12(k+1:k+ngrid);
vv1=u12-2*u11.*xbar1+u10.*xbar1.^2;
vv1=vv1./denomd1.^2;

```

```

u20=conv(bincts2(:,1).*sig2res2,kvec2.^2);
u20=u20(k+1:k+ngrid);
u21=conv(bincts2(:,1).*sig2res2.*xgrid2,kvec2.^2);
u21=u21(k+1:k+ngrid);
u22=conv(bincts2(:,1).*sig2res2.*xgrid2.^2,kvec2.^2);
u22=u22(k+1:k+ngrid);
vv2=u22-2*u21.*xbar2+u20.*xbar2.^2;
vv2=vv2./denomd2.^2;
flag1=(ve1>=5);
flag2=(ve2>=5);
if sum(flag1)>0;
    if isimflag==0;
        gquant1=-phiinv((1-ialpha)/2);
    else;
        nxbar1=mean(ve1(flag1));
        numind1=n/nxbar1;
        beta1=(1-ialpha)^(1/numind1);
        gquant1=-phiinv((1-beta1)/2);
    end;
else;
    gquant1=inf;
end;
if sum(flag2)>0;
    if isimflag==0;
        gquant2=-phiinv((1-ialpha)/2);
    else;
        nxbar2=mean(ve2(flag2));
        numind2=n/nxbar2;
        beta2=(1-ialpha)^(1/numind2);
        gquant2=-phiinv((1-beta2)/2);
    end;
else;
    gquant2=inf;
end;
mdsurf1=[mdsurf1 vd1];
mdsurf2=[mdsurf2 vd2];
mesurf1=[mesurf1 ve1];
mesurf2=[mesurf2 ve2];
mvsurf1=[mvsurf1 vv1];
mvsurf2=[mvsurf2 vv2];
vgq1=[vgq1 gquant1];

```

```

    vgq2=[vgq2 gquant2];
end;
if length(vgq1)>1;
    mloci1=mdsurf1-vec2mat(vgq1,ngrid).*sqrt(mvsurf1);
    mhici1=mdsurf1+vec2mat(vgq1,ngrid).*sqrt(mvsurf1);
else;
    mloci1=mdsurf1-vgq1*sqrt(mvsurf1);
    mhici1=mdsurf1+vgq1*sqrt(mvsurf1);
end;
if length(vgq2)>1;
    mloci2=mdsurf2-vec2mat(vgq2,ngrid).*sqrt(mvsurf2);
    mhici2=mdsurf2+vec2mat(vgq2,ngrid).*sqrt(mvsurf2);
else;
    mloci2=mdsurf2-vgq2*sqrt(mvsurf2);
    mhici2=mdsurf2+vgq2*sqrt(mvsurf2);
end;
mapout1=3*ones(size(mloci1));
mapout2=3*ones(size(mloci2));
flag1=(mloci1>0);
flag2=(mloci2>0);
ssflag1=sum(sum(flag1));
ssflag2=sum(sum(flag2));
if ssflag1>0;
    mapout1(flag1)=ones(ssflag1,1);
end;
if ssflag2>0;
    mapout2(flag2)=ones(ssflag2,1);
end;
flag1=(mhici1<0);
flag2=(mhici2<0);
ssflag1=sum(sum(flag1));
ssflag2=sum(sum(flag2));
if ssflag1>0;
    mapout1(flag1)=4*ones(ssflag1,1);
end;
if ssflag2>0;
    mapout2(flag2)=4*ones(ssflag2,1);
end;
flag1=(mesurf1<=5);
ssflag1=sum(sum(flag1));
if ssflag1>0;

```

```

    mapout1(flag1)=2*ones(ssflag1,1);
end;
flag2=(mesurf2<=5);
ssflag2=sum(sum(flag2));
if ssflag2>0;
    mapout2(flag2)=2*ones(ssflag2,1);
end;
mapout1=mapout1';
mapout2=mapout2';
if nargout==0;
    comap=[.2, .2, .2; .35, .35, .35; .5, .5, .5; .8, .8, .8];
    subplot(2,2,3);
    image([left(1),right(1)],[log10(hmin(1)),log10(hmax(1))],mapout1);
    set(gca,'YDir','normal');
    colormap(comap);
    xlabel('x');
    ylabel('log10(h1)');
    subplot(2,2,4);
    image([left(2),right(2)],[log10(hmin(2)),log10(hmax(2))],mapout2);
    set(gca,'YDir','normal');
    colormap(comap);
    xlabel('z');
    ylabel('log10(h2)');
end;

```

Para implementar el método *plug-in* de selección del ancho de banda de Opsomer y Ruppert (1998) se han utilizado las funciones definidas por los propios autores, que se pueden encontrar en la página web de Opsomer. La implementación se hace utilizando la siguiente función:

```
[yhat,h,m,ybar,variance]=addfit(x,y,p)
```

Esta función calcula el ajuste del siguiente modelo aditivo,

$$y = \alpha + m_1(x_1) + m_2(x_2) + \epsilon_i,$$

usando polinomios locales de grado $p = p_1, p_2$. Los argumentos de entrada son:

x matriz de variables explicativas ($n \times 2$)

y vector de observaciones de la variable respuesta ($n \times 1$)

p orden de los polinomios locales¹ para las funciones aditivas (1×2 ó escalar)

Y la salida que produce tiene los siguientes elementos:

\hat{y} vector estimado en las observaciones ($n \times 1$)

h vector de anchos de banda para las funciones aditivas (1×2)

m matriz de estimaciones para las funciones aditivas en las observaciones ($n \times 2$)

\bar{y} estimación de α (escalar)

variance varianza estimada de los errores (escalar)

A.3. Funciones para evaluar el selector *bootstrap*

```
function bucle1(r,n,C)
```

```
% Realiza r repeticiones del experimento para el modelo 1
```

```
% Los argumentos son:
```

```
% r (número de veces que se repetirá el experimento)
```

```
% n (tamaño de la muestra)
```

```
% C (constante que se utiliza para crear la red de anchos de banda)
```

```
xgrid=[];
```

```
xgrid=linspace(0,1,10);
```

```
xgrid=xgrid';
```

```
redx=[];
```

```
i=1, j=1, ncross=1;
```

```
for i=1:10;
```

```
    for j=1:10;
```

```
        redx(ncross,:)=xgrid(i) xgrid(j)];
```

```
        ncross=ncross+1;
```

```
    end;
```

```
end;
```

```
m1_x=1-6*redx(:,1)+36*(redx(:,1).^2)-53*(redx(:,1).^3)+22*(redx(:,1).^5);
```

```
m2_x=sin(5*pi*redx(:,2));
```

```
m_x=m1_x+m2_x;
```

```
i=1;
```

```
for i=1:r;
```

```
    filedatos=strcat('c:\Model1\m1',num2str(i),'.txt');
```

```
    fileresul=strcat('c:\Model1\rm1',num2str(i),'.txt');
```

¹La función está limitada a ajustar polinomios locales de grado 1.

```

    replica1(n,filedatos,fileresul,C,redx,m_x);
end;

```

```

function bucle2(r,n,C)
% Realiza r repeticiones del experimento para el modelo 2
% Los argumentos son:
% r (número de veces que se repetirá el experimento)
% n (tamaño de la muestra)
% C (constante que se utiliza para crear la red de anchos de banda)
xgrid=[];
xgrid=linspace(0,1,10);
xgrid=xgrid';
redx=[], i=1, j=1, ncross=1;
for i=1:10;
    for j=1:10;
        redx(ncross,:)= [xgrid(i) xgrid(j)];
        ncross=ncross+1;
    end;
end;
m1_x=0.5*sin(pi*redx(:,1));
m2_x=sin(4*pi*redx(:,2));
m_x=m1_x+m2_x;
i=1;
for i=1:r;
    filedatos=strcat('c:\Model2\m2',num2str(i),'.txt');
    fileresul=strcat('c:\Model2\rm2',num2str(i),'.txt');
    replica2(n,filedatos,fileresul,C,redx,m_x);
end;

```

```

function bucle3(r,n,C)
% Realiza r repeticiones del experimento para el modelo 3
% Los argumentos son:
% r (número de veces que se repetirá el experimento)
% n (tamaño de la muestra)
% C (constante que se utiliza para crear la red de anchos de banda)
xgrid=[];
xgrid=linspace(0,1,10);
xgrid=xgrid';
m1_x=xgrid.^2;
m2_x=xgrid.^2;
i=1, j=1, step=1;

```



```

for i=1:10;
    for j=1:10;
        m_x(step,:)=m1_x(i,:)+m2_x(j,:);
        step=step+1;
    end;
end;
i=1;
for i=1:r;
    filedatos=strcat('c:\Model3\datosm3',num2str(i),'.txt');
    fileresul=strcat('c:\Model3\resulm3',num2str(i),'.txt');
    replica3(n,filedatos,fileresul,C,xgrid,m_x);
end;

function [mse_boot_min,h_boot,mse_teo_min,h_teo]=comparamse(x,y,...
    redvh,z,g,m_x,yhat,m1hat,m2hat,ybar,W1,W2)
% Los argumentos son:
% x (matriz de observaciones)
% y (vector de respuesta)
% redvh (red de parámetros ancho de banda)
% z (punto de estimación)
% g (ancho de banda bootstrap)
% m_x (valor real de las componentes aditivas)
% yhat (estimación de la función de regresión)
% m1hat, m2hat (estimación de las componentes aditivas)
% ybar (media del vector de respuesta)
% W1, W2 (matriz W correspondiente a la estimación de cada componente)
nh=length(redvh);
i=1;
for i=1:nh;
    if (isnan(W1(:,i))|(isnan(W2(:,i))))
        mse_boot=NaN;
        mse_teo=NaN;
    else;
        [mse_boot,mse_teo]=estimamse(x,y,z,redvh(i,:),g,W1(:,i),W2(:,i),m_x,...
            yhat,m1hat,m2hat,ybar);
    end;
    resul(i,:)=[mse_boot,mse_teo,redvh(i,:)];
end;
[mse_boot_min,pos1]=min(resul(:,1));
h_boot=[resul(pos1,3),resul(pos1,4)];
[mse_teo_min,pos2]=min(resul(:,2));

```

```
h_teo=[resul(pos2,3),resul(pos2,4)];
```

```
function [mhat_x]=estimaz(x,y,ybar,m1hat,m2hat,yhat,z,h)
% Calcula la estimación de la función de regresión en un nuevo punto
% Los argumentos son:
% x (matriz de observaciones)
% y (vector de respuesta)
% ybar (media del vector y)
% m1hat, m2hat (las estimaciones de la función para las variables)
% yhat (las estimaciones de la función de regresión)
% z el valor donde queremos estimar
% h el ancho de banda que vamos a utilizar
ycen=y-ybar;
mhat_x=[];
s1_x1=defineS(x(:,1),z(:,1),h(:,1));
s2_x2=defineS(x(:,2),z(:,2),h(:,2));
m1hat_x1=s1_x1*(ycen-m2hat);
m2hat_x2=s2_x2*(ycen-m1hat);
mhat_x=ybar+m1hat_x1+m2hat_x2;
```

```
function [matrizpesos,tam]=defnepesos(x,z,h)
% Calcula la matriz de pesos de la función núcleo
% Se ha calculado usando la función núcleo gaussiana
% Los argumentos son:
% x (el vector de observaciones de la variable explicativa)
% z (el punto de estimación considerado)
% h (el ancho de banda)
fil=length(x);
P=zeros(fil,fil);
Kh=zeros(fil,1);
tamentorno=0;
for i=1:fil;
    d=((x(i,:)-z)/h);
    if (abs(d)<1);
        tamentorno=tamentorno+1;
    end;
    Kh(i,:)=(kgaus(d)/h);
    P(i,i)=Kh(i,:);
end;
matrizpesos=P;
tam=tamentorno;
```

```

function [S]=defineS(x,z,h,p,r)
% Calcula cada elemento de la matriz S, suavizador lineal local
% Los argumentos son:
% x (vector de observaciones de la variable explicativa)
% z (punto de estimación)
% h (ancho de banda)
% p (orden del polinomio que se ajusta)
% r (escalar, 0, 1, 2, ... si lo que queremos calcular son el valor
% de las estimaciones, o las estimaciones de la primera derivada,...)
fil=length(x);
S=zeros(1,fil);
col=p+1;
e=[], matrizX=[], matrizW=[], trasX=[], inversa=[];
e=zeros(1,col);
if (r==0);
    e(1)=1;
else;
    e(r+1)=1;
end;
matrizX=defineX(x,z,p);
tam=0;
[matrizpesos,tam]=definepesos(x,z,h);
control=0;
for i=1:fil;
    if (matrizpesos(i,i)>0.001);
        control=1;
    end;
end;
if (tam>=p+1 & control~=0)
    trasX=matrizX';
    determinante=det((trasX*matrizpesos)*matrizX);
    if (abs(determinante)>0.001);
        inversa=inv((trasX*matrizpesos)*matrizX);
        matrizS=inversa*(trasX*matrizpesos);
        S=(e*matrizS);
    else;
        for i=1:fil;
            S(i)=NaN;
        end;
    end;
end;
else;

```

```

    for i=1:fil;
        S(i)=NaN;
    end;
end;

```

```

function [Sast]=defineSast(S)
% El argumento es:
% S (matriz de suavizamiento)
Sast=[];
[fil,col]=size(S);
I=eye(fil);
uno=ones(fil,1);
i=1, control=0;
for i=1:fil;
    for j=1:col;
        if S(i,j)==NaN;
            control=control+1;
        end;
    end;
end;
if control==0;
    Sast=(I-(uno*uno')./fil)*S;
else;
    for i=1:fil;
        for j=1:col;
            Sast(i,j)=NaN;
        end;
    end;
end;

```

```

function [matrizX]=defineX(x,z,p)
% Define la matriz de diseño para cada variable
% Los argumentos son:
% x (vector de observaciones de la variable explicativa)
% z (punto de estimación)
% p (orden del polinomio que se ajusta)
fil=length(x);
col=p+1;
matrizX=ones(fil,col);
for i=1:fil;
    for j=2:col;

```

```

    matrizX(i,j)=(x(i,:)-z)^(j-1);
end;
end;

```

```

function [Wast_x_h]=defineWast(x,y,z,h,W1,W2)
n=length(x);
uno=ones(1,n);
s1_x1=defineS(x(:,1),z(:,1),h(:,1));
s2_x2=defineS(x(:,2),z(:,2),h(:,2));
W1ast=s1_x1*W1;
W2ast=s2_x2*W2;
Wast_x_h=(uno./n)+W1ast+W2ast;

```

```

function [yhat,m1hat,m2hat,ybar] = estimam(x,y,h)
% Los argumentos son:
% x (matriz de observaciones)
% y (vectores de respuesta)
% h (parámetro ancho de banda)
n=length(x);
mind=min(x);
maxd=max(x);
ngrid=401;
ybar=sum(y)/n;
ycen=y-ybar;
coefin(1,:)=polyfit(x(:,1),ycen,1);
coefin(2,:)=polyfit(x(:,2),ycen,1);
mstart(:,1)=polyval(coefin(1,:),x(:,1));
mstart(:,2)=polyval(coefin(2,:),x(:,2));
Countmax=30, step=1, count=1;
mnew=mstart;
while (step>0.0001& count<Countmax);
    mold=mnew;
    [msmoo(:,1),xgrid1]=gestimates([x(:,1) ycen-mold(:,2)],h(:,1),...
        [mind(1);maxd(1);ngrid]);
    [msmoo(:,2),xgrid2]=gestimates([x(:,2) ycen-mold(:,1)],h(:,2),...
        [mind(2);maxd(2);ngrid]);
    msmoo=msmoo-ones(ngrid,1)*sum(msmoo)/ngrid;
    mnew(:,1)=interp1(xgrid1,msmoo(:,1),x(:,1),'cubic');
    mnew(:,2)=interp1(xgrid2,msmoo(:,2),x(:,2),'cubic');
    step=sum(sum(mnew-mold).^2)/n;
    count=count+1;
end;

```

```
end ;
yhat=ybar+mnew(:,1)+mnew(:,2);
m1hat=mnew(:,1);
m2hat=mnew(:,2);
```

```
function [mse_boot,mse_teo]=estimaMSE(x,y,z,h,g,W1,W2,m_x,yhat,...
    m1hat,m2hat,ybar)
[yhath]=estimam(x,y,h);
[mhat_x]=estimaz(x,y,ybar,m1hat,m2hat,yhat,z,g);
[Wast_x_h]=defineWast(x,y,z,h,W1,W2);
[mse_boot]=(Wast_x_h*yhat-mhat_x).^2+Wast_x_h*(y-yhat)*(y-yhat)*Wast_x_h';
[mse_teo]=(Wast_x_h*y-m_x).^2+Wast_x_h*(y-yhath)*(y-yhath)*Wast_x_h';
```

```
function fundereplicas(r,n,filefinal)
% Los argumentos son:
% r (número de replicas)
% n (tamaño de muestra)
% filefinal (fichero donde se grabarán los resultados finales)
fid=fopen(filefinal,'w');
if (fid==0);
    disp('Error al abrir el fichero donde se grabarán los resultados);
    stop;
else;
    l=1;
    for l=1:r;
        fileresul=strcat('c:\Model3\rm3',num2str(l),'.txt');
        fi=fopen(fileresul,'r');
        if (fi==0);
            disp('Error al abrir un fichero de datos');
            stop;
        else;
            i=1;
            for i=1:n;
                position(i)=ftell(fi);
                control=0;
                temp=[];
                temp=fscanf(fi,'%f',12);
                iposition(i)=ftell(fi);
                if iposition(i)==position(i)+1;
                    fseek(fi,-1,0);
                    j=1;
```

```

        for j=1:12;
            fscanf(fi,'%s',1);
        end;
    else;
        temp=temp';
        fprintf(fid,'%f %f %f %f %f %f %f %f %f %f %f %f\n',temp);
    end;
    clear temp;
    i=i+1;
end;
end;
fclose(fi);
end;
fclose('all');
end; function [K]=kgaus(x)

```

% Calcula pesos según una función núcleo gaussiana

% Argumento:

% x (vector de observaciones)

K=0;

cons=1/sqrt(2*pi);

valor=(-(x^2))/2;

K=cons*exp(valor);

K;

function [S1ast,S2ast,W1,W2]=makeW(x,h,n)

% Los argumentos son:

% x (matriz de observaciones)

% h (parámetros ancho de banda)

% n (tamaño de muestra)

I=eye(n);

i=1;

for i=1:n;

 S1(i,:)=defineS(x(:,1),x(i,1),h(:,1));

 S2(i,:)=defineS(x(:,2),x(i,2),h(:,2));

end;

S1ast=defineSast(S1);

S2ast=defineSast(S2);

if (isnan(S1ast(1,1))|(isnan(S2ast(1,1))));

 for i=1:n;

 for j=1:n;

```

        W1(i,j)=NaN;
        W2(i,j)=NaN;
    end;
end;
else;
    prod1=I-S1ast*S2ast;
    prod2=I-S2ast*S1ast;
    determinante1=det(prod1);
    determinante2=det(prod2);
    if (abs(determinante1)>0.001)& (abs(determinante2)>0.001)
        W1=(inv(prod1))*(I-S1ast);
        W2=(inv(prod2))*(I-S2ast);
    else;
        for i=1:n;
            for j=1:n;
                W1(i,j)=NaN;
                W2(i,j)=NaN;
            end;
        end;
    end;
end;
end;

```

```

function replica1(n,filedatos,fileresul,C,redx,m_x)
% Realiza una repetición completa
% Los argumentos son:
% n (tamaño de la muestra)
% filedatos (nombre del fichero donde se guarda la muestra)
% fileresul (nombre del fichero donde se guardan los resultados)
% C (constante para crear la red de anchos de banda)
% redx (conjunto de valores de x donde se evaluará la función de regresión)
% m_x (evaluación de la función de regresión en redx)
S1ast=[], S2ast=[], W1=[], W2=[], yhatcv=[];
[x,y]=modell1(n,0.5,0.5,1/9,1/9,0,filedatos);
mind=min(x);
maxd=max(x);
range=maxd-mind;
hmin=range/n;
hmax=range/2;
i=1;
vh(1,:)=hmin;
while (vh(i,1)<hmax(:,1));

```



```

    hnuevo=(C^i)*hmin(:,1);
    i=i+1;
    if (hnuevo<hmax(:,1));
        vh(i,1)=hnuevo;
    end;
    if (hnuevo>=hmax(:,1));
        vh(i,1)=hmax(:,1);
    end;
end;
i=1;
while (vh(i,2)<hmax(:,2));
    hnuevo=(C^i)*hmin(:,2);
    i=i+1;
    if (hnuevo<hmax(:,2));
        vh(i,2)=hnuevo;
    end;
    if (hnuevo>=hmax(:,2));
        vh(i,2)=hmax(:,2);
    end;
end;
nh=length(vh);
redvh=[];
i=1, j=1, nhcross=1;
for i=1:nh;
    for j=1:nh;
        redvh(nhcross,:)= [vh(i,1) vh(j,2)];
        nhcross=nhcross+1;
    end;
end;
i=1;
for i=1:(nhcross-1);
    [S1ast(:, :, i), S2ast(:, :, i), W1(:, :, i), W2(:, :, i)] = makeW(x, redvh(i, :), n);
end;
j=1;
CV=[];
for j=1:(nhcross-1);
    if (isnan(S1ast(1,1,j)) || isnan(S2ast(1,1,j)));
        CV(j)=NaN;
    else;
        [yhatcv(:,j)] = estimam(x,y,redvh(j,:));
        for k=1:n;

```

```

        CV(j)=0;
        CV(j)=CV(j)+((y(k)-yhatcv(k,j))/(1-1/n-S1ast(k,k,j)-S2ast(k,k,j))).^2;
    end;
    CV(j)=CV(j)/n;
end;
end;
[Cvmin,pos]=min(CV);
g=redvh(pos,:);
[yhat,m1hat,m2hat,ybar]=estimam(x,y,g);
fi=fopen('fileresul','w');
if (fi==0);
    disp('Error al abrir el fichero de resultados');
    stop;
else;
    i=1;
    resul=[];
    for i=1:n;
        mse_boot_min=0, mse_teo_min=0, mhat_xgrid_boot=0;
        h_boot=[], h_teo=[];
        [mse_boot_min,h_boot,mse_teo_min,h_teo]=comparamse(x,y,redvh,redx(i,:),...
            C,g,m_x(i,:),yhat,m1hat,m2hat,ybar,W1,W2);
        mhat_xgrid_boot(i,:)=estimaz(x,y,ybar,m1hat,m2hat,yhat,redx(i,:),h_boot);
        mhat_xgrid_teo(i,:)=estimaz(x,y,ybar,m1hat,m2hat,yhat,redx(i,:),h_teo);
        resul(i,:)=[mse_boot_min,h_boot(:,1),h_boot(:,2),mse_teo_min,h_teo(:,1),...
            h_teo(:,2),redx(i,1),redx(i,2),mhat_xgrid_boot(i,:),...
            mhat_xgrid_teo(i,:),m_x(i,:),yhat(i,:)];
        fprintf(fi,'%f %f %f %f %f %f %f %f %f %f %f %f \n',resul(i,:));
    end;
    fclose(fi);
end;

```

```

function replica2(n,filedatos,fileresul,C,redx,m_x)
% Realiza una repetición completa
% Los argumentos son:
% n (tamaño de la muestra)
% filedatos (nombre del fichero donde se guarda la muestra)
% fileresul (nombre del fichero donde se guardan los resultados)
% C (constante para crear la red de anchos de banda)
% redx (conjunto de valores de x donde se evaluará la función de regresión)
% m_x (evaluación de la función de regresión en redx)
S1ast=[], S2ast=[], W1=[], W2=[], yhatcv=[];

```

```

[x,y]=model2(n,filedatos);
mind=min(x);
maxd=max(x);
range=maxd-mind;
hmin=range/n;
hmax=range/2;
i=1;
vh(1,:)=hmin;
while (vh(i,1)<hmax(:,1));
    hnuevo=(C^i)*hmin(:,1);
    i=i+1;
    if (hnuevo<hmax(:,1));
        vh(i,1)=hnuevo;
    end;
    if (hnuevo>=hmax(:,1));
        vh(i,1)=hmax(:,1);
    end;
end;
i=1;
while (vh(i,2)<hmax(:,2));
    hnuevo=(C^i)*hmin(:,2);
    i=i+1;
    if (hnuevo<hmax(:,2));
        vh(i,2)=hnuevo;
    end;
    if (hnuevo>=hmax(:,2));
        vh(i,2)=hmax(:,2);
    end;
end;
nh=length(vh);
redvh=[], i=1, j=1, nhcross=1;
for i=1:nh;
    for j=1:nh;
        redvh(nhcross,:)= [vh(i,1) vh(j,2)];
        nhcross=nhcross+1;
    end;
end;
i=1;
for i=1:(nhcross-1);
    [S1ast(:,i),S2ast(:,i),W1(:,i),W2(:,i)]=makeW(x,redvh(i,:),n);
end;

```

```

j=1, CV=[];
for j=1:(nhcross-1);
    if (isnan(S1ast(1,1,j))|(isnan(S2ast(1,1,j))));
        CV(j)=NaN;
    else;
        [yhatcv(:,j)]=estimam(x,y,redvh(j,:));
        for k=1:n;
            CV(j)=0;
            CV(j)=CV(j)+((y(k)-yhatcv(k,j))/(1-1/n-S1ast(k,k,j)-S2ast(k,k,j))).^2;
        end;
        CV(j)=CV(j)/n;
    end;
end;
[Cvmin,pos]=min(CV);
g=redvh(pos,:);
[yhat,m1hat,m2hat,ybar]=estimam(x,y,g);
fi=fopen('fileresul','w');
if (fi==0);
    disp('Error al abrir el fichero de resultados');
    stop;
else;
    i=1;
    resul=[];
    for i=1:n;
        mse_boot_min=0, mse_teo_min=0, mhat_xgrid_boot=0;
        h_boot=[], h_teo=[];
        [mse_boot_min,h_boot,mse_teo_min,h_teo]=comparamse(x,y,redvh,...
            redx(i,:),C,g,m_x(i,:),yhat,m1hat,m2hat,ybar,W1,W2);
        mhat_xgrid_boot(i,:)=estimaz(x,y,ybar,m1hat,m2hat,yhat,redx(i,:),h_boot);
        mhat_xgrid_teo(i,:)=estimaz(x,y,ybar,m1hat,m2hat,yhat,redx(i,:),h_teo);
        resul(i,:)=[mse_boot_min,h_boot(:,1),h_boot(:,2),mse_teo_min,h_teo(:,1),...
            h_teo(:,2),redx(i,1),redx(i,2),mhat_xgrid_boot(i,:),...
            mhat_xgrid_teo(i,:),m_x(i,:),yhat(i,:)];
        fprintf(fi,'%f %f %f %f %f %f %f %f %f %f %f %f \n',resul(i,:));
    end;
    fclose(fi);
end;

```

```

function replica3(n,filedatos,fileresul,C,xgrid,m_x)

```

```

% Realiza una repetición completa

```

```

% Los argumentos son:

```

```

% n (tamaño de la muestra)
% filedatos (nombre del fichero donde se guarda la muestra)
% fileresul (nombre del fichero donde se guardan los resultados)
% C (constante para crear la red de anchos de banda)
% redx (conjunto de valores de x donde se evaluará la función de regresión)
% m_x (evaluación de la función de regresión en redx)
W1(:, :, :) = [], W2(:, :, :) = [];
[x,y]=model3(n,filedatos);
[yh,g]=addfit(x,y,[1 1]);
[yhat,m1hat,m2hat,ybar]=estimam(x,y,g);
mind=min(x);
maxd=max(x);
range=maxd-mind;
hmin=range/n;
hmax=range/2;
i=1;
vh(1,:)=hmin;
while (vh(i,1)<hmax(:,1));
    hnuevo=(C^i)*hmin(:,1);
    i=i+1;
    if (hnuevo<hmax(:,1));
        vh(i,1)=hnuevo;
    end;
    if (hnuevo>=hmax(:,1));
        vh(i,1)=hmax(:,1);
    end;
end;
i=1;
while (vh(i,2)<hmax(:,2));
    hnuevo=(C^i)*hmin(:,2);
    i=i+1;
    if (hnuevo<hmax(:,2));
        vh(i,2)=hnuevo;
    end;
    if (hnuevo>=hmax(:,2));
        vh(i,2)=hmax(:,2);
    end;
end;
end;
nh=length(vh);
i=1, j=1, cross=1;
for i=1:nh;

```

```

for j=1:nh;
    [W1(:,:,cross),W2(:,:,cross)]=makeW(x,y,[vh(i,1),vh(j,2)],n);
    cross=cross+1;
end;
end;
fi=fopen('fileresul','w');
if (fi==0);
    disp('Error al abrir el fichero de resultados');
    stop;
else;
    i=1, j=1, step=1;
    resul=[];
    for i=1:10;
        for j=1:10;
            mse_boot_min=0, mse_teo_min=0, mhat_xgrid_boot=0;
            h_boot=[], h_teo=[];
            [mse_boot_min,h_boot,mse_teo_min,h_teo]=comparamse(x,y,vh,[xgrid(i,:)...
                xgrid(j,:)],C,g,m_x(step,:),yhat,m1hat,m2hat,ybar,W1,W2);
            mhat_xgrid_boot=estimaz(x,y,ybar,m1hat,m2hat,yhat,...
                [xgrid(i,:) xgrid(j:)],h_boot);
            resul(step,:)=[mse_boot_min,h_boot(:,1),h_boot(:,2),mse_teo_min,...
                h_teo(:,1),h_teo(:,2),xgrid(i,:),xgrid(j,:),mhat_xgrid_boot,m_x(step,:)];
            fprintf(fi,'%f %f %f %f %f %f %f %f %f %f %f \n',resul(step,:));
            step=step+1;
        end;
    end;
    fclose(fi);
end;
end;

```

Apéndice B

Expresiones adicionales

$$\begin{aligned} A_1 = & \frac{(b-a)^2}{2} D_1 f_1(x_1) f_2(x_2) \{ D_1 m_1(x_1) f_1(x_1) + D_2 m_2(x_2) f_2(x_2) + \\ & + \frac{(b-a)^2}{4} (D_1^2 m_1(x_1) D_1 f_1(x_1) + D_2^2 m_2(x_2) D_2 f_2(x_2)) \} + \\ & + \frac{(b-a)^2}{2} f_1(x_1) f_2(x_2) \{ D_1^2 m_1(x_1) f_1(x_1) + D_1 m_1(x_1) D_1 f_1(x_1) + \\ & + \frac{(b-a)^2}{4} (D_1^3 m_1(x_1) D_1 f_1(x_1) + D_1^2 m_1(x_1) D_1^2 f_1(x_1)) \} + \\ & + \frac{(b-a)^3}{3} D_1 f_1(x_1) f_2(x_2) \{ D_1 m_1(x_1) D_1 f_1(x_1) + D_2 m_2(x_2) D_2 f_2(x_2) + \\ & + \frac{1}{2} (D_1^2 m_1(x_1) + D_2^2 m_2(x_2)) \} + \\ & + \frac{(b-a)^3}{3} f_1(x_1) f_2(x_2) \{ D_1^2 m_1(x_1) D_1 f_1(x_1) + D_1 m_1(x_1) D_1^2 f_1(x_1) + \\ & + \frac{1}{2} D_1^3 m_1(x_1) \}, \end{aligned}$$

$$\begin{aligned} A_2 = & \frac{(b-a)^2}{2} D_2 f_2(x_2) f_1(x_1) \{ D_2 m_2(x_2) f_2(x_2) + D_1 m_1(x_1) f_1(x_1) + \\ & + \frac{(b-a)^2}{4} (D_2^2 m_2(x_2) D_2 f_2(x_2) + D_1^2 m_1(x_1) D_1 f_1(x_1)) \} + \\ & + \frac{(b-a)^2}{2} f_2(x_2) f_1(x_1) \{ D_2^2 m_2(x_2) f_2(x_2) + D_2 m_2(x_2) D_2 f_2(x_2) + \\ & + \frac{(b-a)^2}{4} (D_2^3 m_2(x_2) D_2 f_2(x_2) + D_2^2 m_2(x_2) D_2^2 f_2(x_2)) \} + \end{aligned}$$

$$\begin{aligned}
& + \frac{(b-a)^3}{3} D_2 f_2(x_2) f_1(x_1) \{D_2 m_2(x_2) D_2 f_2(x_2) + D_1 m_1(x_1) D_1 f_1(x_1) + \\
& + \frac{1}{2} (D_2^2 m_2(x_2) + D_1^2 m_1(x_1))\} + \\
& + \frac{(b-a)^3}{3} f_2(x_2) f_1(x_1) \{D_2^2 m_2(x_2) D_2 f_2(x_2) + D_2 m_2(x_2) D_2^2 f_2(x_2) + \\
& + \frac{1}{2} D_2^3 m_2(x_2)\},
\end{aligned}$$

$$\begin{aligned}
A_3 = & \frac{(b-a)^2}{2} [D_1^2 f_1(x_1) f_2(x_2) \{D_1 m_1(x_1) f_1(x_1) + D_2 m_2(x_2) f_2(x_2) + \\
& + \frac{(b-a)^2}{4} (D_1^2 m_1(x_1) D_1 f_1(x_1) + D_2^2 m_2(x_2) D_2 f_2(x_2))\} + \\
& + 2D_1 f_1(x_1) f_2(x_2) \{D_1^2 m_1(x_1) f_1(x_1) + D_1 m_1(x_1) D_1 f_1(x_1) + \\
& + \frac{(b-a)^2}{4} (D_1^3 m_1(x_1) D_1 f_1(x_1) + D_1^2 m_1(x_1) D_1^2 f_1(x_1))\} + \\
& + f_1(x_1) f_2(x_2) \{D_1^3 m_1(x_1) f_1(x_1) + 2D_1^2 m_1(x_1) D_1 f_1(x_1) + \\
& + D_1 m_1(x_1) D_1^2 f_1(x_1) + \frac{(b-a)^2}{4} (D_1^4 m_1(x_1) D_1 f_1(x_1) + \\
& + 2D_1^3 m_1(x_1) D_1^2 f_1(x_1) + D_1^2 m_1(x_1) D_1^3 f_1(x_1))\}] + \\
& + \frac{(b-a)^3}{3} [D_1^2 f_1(x_1) f_2(x_2) \{D_1 m_1(x_1) D_1 f_1(x_1) + D_2 m_2(x_2) D_2 f_2(x_2) + \\
& + \frac{1}{2} (D_1^2 m_1(x_1) + D_2^2 m_2(x_2))\} + \\
& + 2D_1 f_1(x_1) f_2(x_2) \{D_1^2 m_1(x_1) D_1 f_1(x_1) + D_1 m_1(x_1) D_1^2 f_1(x_1) + \\
& + \frac{1}{2} D_1^3 m_1(x_1)\} + f_1(x_1) f_2(x_2) \{D_1^3 m_1(x_1) D_1 f_1(x_1) + \\
& + 2D_1^2 m_1(x_1) D_1^2 f_1(x_1) + D_1 m_1(x_1) D_1^3 f_1(x_1) + \frac{1}{2} D_1^4 m_1(x_1)\}],
\end{aligned}$$

$$\begin{aligned}
A_4 = & \frac{(b-a)^2}{2} [D_2^2 f_2(x_2) f_1(x_1) \{D_2 m_2(x_2) f_2(x_2) + D_1 m_1(x_1) f_1(x_1) + \\
& + \frac{(b-a)^2}{4} (D_2^2 m_2(x_2) D_2 f_2(x_2) + D_1^2 m_1(x_1) D_1 f_1(x_1))\} + \\
& + 2D_2 f_2(x_2) f_1(x_1) \{D_2^2 m_2(x_2) f_2(x_2) + D_2 m_2(x_2) D_2 f_2(x_2) + \\
& + \frac{(b-a)^2}{4} (D_2^3 m_2(x_2) D_2 f_2(x_2) + D_2^2 m_2(x_2) D_2^2 f_2(x_2))\} +
\end{aligned}$$

$$\begin{aligned}
& + f_2(x_2) f_1(x_1) \{ D_2^3 m_2(x_2) f_2(x_2) + 2D_2^2 m_2(x_2) D_2 f_2(x_2) + \\
& + D_2 m_2(x_2) D_2^2 f_2(x_2) + \frac{(b-a)^2}{4} (D_2^4 m_2(x_2) D_2 f_2(x_2) + \\
& + 2D_2^3 m_2(x_2) D_2^2 f_2(x_2) + D_2^2 m_2(x_2) D_2^3 f_2(x_2)) \} + \\
& + \frac{(b-a)^3}{3} [D_2^2 f_2(x_2) f_1(x_1) \{ D_2 m_2(x_2) D_2 f_2(x_2) + D_1 m_1(x_1) D_1 f_1(x_1) + \\
& + \frac{1}{2} (D_2^2 m_2(x_2) + D_1^2 m_1(x_1)) \} + \\
& + 2D_2 f_2(x_2) f_1(x_1) \{ D_2^2 m_2(x_2) D_2 f_2(x_2) + D_2 m_2(x_2) D_2^2 f_2(x_2) + \\
& + \frac{1}{2} D_2^3 m_2(x_2) \} + f_1(x_1) f_2(x_2) \{ D_2^3 m_2(x_2) D_2 f_2(x_2) + \\
& + 2D_2^2 m_2(x_2) D_2^2 f_2(x_2) + D_2 m_2(x_2) D_2^3 f_2(x_2) + \frac{1}{2} D_2^4 m_2(x_2) \}].
\end{aligned}$$



Apéndice C

Notación

A continuación se incluye una lista con algunas aclaraciones sobre la notación y terminología empleada a lo largo de este trabajo.

- \mathbb{R}^D espacio real euclídeo de dimensión D .
- X variable de predicción en \mathbb{R} .
- $\mathbf{X} = (X_1, \dots, X_D)^T$ vector de variables de predicción en \mathbb{R}^D .
- Y variable de respuesta en \mathbb{R} .
- $f(\mathbf{x})$ función de densidad conjunta de \mathbf{X} .
- $f_j(x_j)$ función de densidad marginal de X_j .
- $f_{j|k}(x_j)$ función de densidad condicionada de $X_j|X_k$.
- $m(\mathbf{x})$ función de regresión de Y sobre \mathbf{X} .
- $m^{(\nu)}$ derivada ν -ésima de una función m . m' y m'' son la primera y segunda derivadas, respectivamente.
- σ^2 varianza de Y sobre \mathbf{X} . (Modelo homocedástico)
- $v(\mathbf{x})$ varianza condicionada de Y sobre \mathbf{X} . (Modelo heterocedástico)
- \mathbf{I} matriz identidad.
- A^T traspuesta de la matriz A .
- $\text{tr}(A)$ traza de la matriz A .
- $\mu_2(K) = \int uu^T K(u) du$ momento de segundo orden de K .

- $R(K) = \int K(u)^2 du.$
- $D_m(\mathbf{x})$ vector $D \times 1$ de derivadas parciales de primer orden.
- $\mathcal{H}_m(\mathbf{x})$ matriz Hessiana de m de dimensión $D \times D$ de una función D -variante.
- Sean a_n y b_n series de números reales.
 - $a_n = O(b_n)$ si $a_n/b_n \rightarrow c$, donde c es un constante, cuando $n \rightarrow \infty$.
 - $a_n = o(b_n)$ si $a_n/b_n \rightarrow 0$, cuando $n \rightarrow \infty$.
 - $a_n \sim b_n$ si $a_n/b_n = c + o(1)$, $c \neq 0$, cuando $n \rightarrow \infty$.
- Sean A_n y B_n series de variables aleatorias reales.
 - $A_n = O_P(B_n)$ si $\forall \delta > 0 \exists M, \exists N$ tales que $P[|A_n/B_n| > M] < \delta, \forall n > N$.
 - $A_n = o_P(B_n)$ si $\forall \delta > 0 \lim_n \rightarrow \infty P[|A_n/B_n| > \delta] = 0$.
 - $A_n \approx B_n$ si $A_n = B_n + O_P(B_n)$.

Bibliografía

- [1] **Bellman, R.E.** (1961) *Adaptive control processes*. Princeton University Press.
- [2] **Breiman, L. and Friedman, J.H.** (1985) Estimating optimal transformations for multiple regression and correlation (with discussion). *Journal of the American Statistical Association*, Vol. 80, pp. 580-619.
- [3] **Buja, A., Hastie, T.J. and Tibshirani, R.** (1989) Linear smoothers and additive models (with discussion). *The Annals of Statistics*, Vol. 17, No. 2, pp. 453-555.
- [4] **Burman, P.** (1985) *Estimation of generalized additive models*. Unpublished Ph.D. thesis, Rutgers University, New Brunswick, N.J.
- [5] **Cao-Abad, R.** (1991) Rate of convergence for the wild bootstrap in nonparametric regression. *The Annals of Statistics*, Vol. 19, No. 4, pp. 2226-2231.
- [6] **Chambers, J.M. and Hastie, T.J.** (1992) *Statistical models in S*. Pacific Grove, CA: Wadsworth and Brooks/Cole. (Eds.) (The white book)
- [7] **Chaudhuri, P. and Marron, J.S.** (1999) SiZer for exploration of Structures in Curves. *Journal of the American Statistical Association*, Vol. 94, No. 447, pp. 807-823.
- [8] **Clark, R.M.** (1977) Non-parametric estimation of a smooth regression function. *Journal of the Royal Statistical Society, Ser. B*, Vol. 39, pp. 107-113.
- [9] **Cleveland, W.S.** (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, Vol. 74, No. 368. Theory and Methods, pp. 829-836.
- [10] **Cleveland, W.S. and Devlin, S.J.** (1988) Locally weighed regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, Vol. 83, No. 403, pp. 596-610.

- [11] Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, Vol. 31, pp. 377-403.
- [12] Efron, B. (1979) Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, No. 1, pp. 1-26.
- [13] Engle, R.F., Granger, C.W.J., Rice, J.A. and Weiss, A. (1986) Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, Vol. 81, pp. 310-320.
- [14] Eubank, R.L. (1988) *Spline smoothing and nonparametric regression*. Marcel Dekker, New York.
- [15] Fan, J. (1992) Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, No. 87, pp. 998-1004.
- [16] Fan, J. and Gijbels, I. (1995) Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society, Ser. B*, Vol. 57, No. 2, pp. 371-394.
- [17] Fan, J. and Gijbels, I. (1996) *Local polynomial modelling and its applications*. Chapman & Hall.
- [18] Fan, J., Härdle, W. and Mammen, E. (1998) Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, Vol. 26, No. 3, pp. 943-971.
- [19] Fan, J. and Marron, J.S. (1994) Fast implementation of nonparametric curve estimators. *Journal of Computational and Graphical Statistics*, Vol. 3, No. 1, pp. 35-56.
- [20] Fiedman, J.H. (1991) Multivariate adaptive regression splines (with discussion). *The Annals of Statistics*, No. 19, pp. 1-141.
- [21] Friedman, J.H. and Stuetzle, W. (1981) Projection pursuit regression. *Journal of the American Statistical Association*, Vol. 76, No. 376, pp. 817-823.
- [22] Gasser, T. and Müller, H.G. (1979) Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics, No. 757, pp. 23-68. Springer-Verlag, New York.
- [23] Hall, P. (1990) Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, Vol. 32, pp. 177-203.

- [24] Härdle, W. (1990) *Applied nonparametric regression*. Econometric Monograph Series 19. Cambridge: Cambridge University Press.
- [25] Härdle, W. and Bowman, A.W. (1988) Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, Vol. 83, No. 401. Theory and Methods, pp. 102-110.
- [26] Härdle, W., Hall, P. and Marron, J.S. (1998) How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association*, Vol. 83, No. 401. Theory and Methods, pp. 86-95.
- [27] Härdle, W., Huet, S. and Jolivet, E. (1995) Better bootstrap confidence intervals for regression curve estimation. *Journal of the Royal Statistical Society*, Vol. 26, No. 4, pp. 287-306.
- [28] Härdle, W., Mammen, E. (1993) Comparing non parametric versus parametric regression fits. *The Annals of Statistics*, Vol. 21, pp. 1926-1947.
- [29] Härdle, W. and Marron, J.S. (1991) Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics*, Vol. 19, No. 2, pp. 778-796.
- [30] Hastie, T.J. and Tibshirani, R. (1986) Generalized additive models (with discussion). *Journal of the Royal Statistical Society*, No. 1, pp. 297-318.
- [31] Hastie, T.J. and Tibshirani, R. (1987) Generalized additive models: Some applications. *Journal of the American Statistical Association*, Vol. 82, No. 398, pp. 371-386.
- [32] Hastie, T.J. and Tibshirani, R. (1990) *Generalized additive models*. Washington, D.C.; Chapman & Hall.
- [33] Hengartner, N.W. (1996) Rate optimal estimation of additive regression via the integration method in the presence of many covariates. Preprint, Department of Statistics, Yale University.
- [34] Jones, M.C., Davies, S.J. and Park, B.U. (1994) Versions of kernel-type regression estimators. *Journal of the American Statistical Association*, Vol. 89, No. 427, pp. 825-832.
- [35] Kauermann, G. and Opsomer, J.D. A fast method for implementing generalized cross-validation in multi-dimensional nonparametric regression. *Submitted to Journal of Computational and Graphical Statistics*.

- [36] Kim, W., Linton, O.B. and Hengartner, N.W. (1997) A nimble method of estimating additive nonparametric regression. Electronic article, <http://www.stats.yale.edu>.
- [37] Kim, W., Linton, O.B. and Hengartner, N.W. (1999) A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. *Journal of Computational and Graphical Statistics*, Vol. 8, No. 2, pp. 278-297.
- [38] Linton, O.B. (1997) Efficient estimation of additive nonparametric regression models. *Biometrika*, Vol. 84, No. 2, pp. 469-473.
- [39] Linton, O.B. and Härdle, W. (1996) Estimating additive regression models with known links. *Biometrika*, No. 83, pp. 529-540.
- [40] Linton, O.B. and Nielsen, J.P. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, Vol. 82, No. 1, pp. 93-100.
- [41] Mack, Y.P. and Müller, H.G. (1989) Derivative estimation in nonparametric regression with random predictor variables. *Sankhyā*, Ser. A, No. 51, pp. 59-72.
- [42] Mammen, E., Linton, O.B. and Nielsen, J.P. (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, Vol. 27, No. 5, pp. 1443-1490.
- [43] Martínez-Miranda, M.D. (1996) *Técnicas de regresión no paramétrica*. Memoria de Licenciatura.
- [44] Martínez-Miranda, M.D. (2000) *Elección de los parámetros que intervienen en un modelo de regresión local*. Tesis doctoral.
- [45] Martínez-Miranda, M.D. y González-Carmona, A. (2000) Selección bootstrap de un ancho de banda variable para un modelo de regresión local. *Actas del XXV Congreso Nacional de Estadística e Investigación Operativa*. Vigo (España).
- [46] Martínez-Miranda, M.D. and González-Carmona, A. (2000) A bootstrap bandwidth selector in local polynomial regression. *International Seminar on Nonparametric Inference, ISNI 2000*. Santiago de Compostela (España).
- [47] Martínez-Miranda, M.D., González-Carmona, A. and González-Aguilera, S. (2000) Bootstrap choice of a variable bandwidth in local regression. *Proceedings in Computational Statistics. XIV COMPSTAT conference of the International Association for Statistical Computing*. Statistics Netherlands.

- [48] Nadaraya, E.A. (1964) On estimating regression. *Theory Probab. Appl*, No. 9, pp. 141-142.
- [49] Neumann, M. and Polzehl, J. (1998) Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, Vol. 9, No. 4, pp. 307-333.
- [50] Opsomer, J.D. (2000) Asymptotic properties of backfitting estimators. *Journal of Multivariate Analysis*, Vol. 73, pp. 166-179.
- [51] Opsomer, J.D. and Ruppert, D. (1997) Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, Vol. 25, No. 1, pp. 186-211.
- [52] Opsomer, J.D. and Ruppert, D. (1998) A fully automated bandwidth selection method for fitting additive models. *Journal of the American Statistical Association*, Vol. 93, No. 442, pp. 605-619.
- [53] Priestley, M.B. and Chao, M.T. (1972) Non-parametric function fitting. *Journal of the Royal Statistical Society, Ser. B*, Vol. 34, pp. 385-392.
- [54] Raya-Miranda, R. (2001) *Regresión no paramétrica multivariante: Modelos aditivos*. Trabajo de investigación.
- [55] Raya-Miranda, R., Martínez-Miranda, M.D. y González-Carmona, A. (2001) SiZer Map en la selección del ancho de banda para modelos aditivos. *Actas del XXVI Congreso Nacional de Estadística e Investigación Operativa*. Úbeda (España).
- [56] Raya-Miranda, R., Martínez-Miranda, M.D. and González-Carmona, A. (2002) Exploring the structure of regression surfaces by using SiZer Map for additive models. *Proceedings in Computational Statistics. XV COMPSTAT Conference of the International Association for Statistical Computing*. Statistics Netherlands.
- [57] Ruppert, D. (1997) Empirical-bias bandwidths for local polyomial nonparametric regression and density estimation. *Journal of the American Statistical Association*, Vol. 92, No. 439, pp. 1049-1062.
- [58] Ruppert, D., Sheather, S.J. and Wand, M.P. (1995) An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, Vol. 90, No. 432, pp. 257-1270. Theory and Methods.
- [59] Ruppert, D. and Wand, M.P. (1994) Multivariate locally weighted least squares regression. *The Annals of Statistics*, Vol. 22, No. 3, pp. 1346-1370.

- [60] **Schoenberg, I.** (1964) Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. U.S.A.* No. 52, pp. 947-950.
- [61] **Stute, W. González-Manteiga, W. and Presedo-Quíndimil, W.** (1998) Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, Vol. 93, pp. 441.
- [62] **Wahba, G.** (1986) Partial and interaction splines for semi-parametric estimation of functions of several variables. *Department of Statistics Technical report*. No. 784, University of Wisconsin, Madison.
- [63] **Wahba, G.** (1990) *Spline models for observational data*. SIAM, Philadelphia.
- [64] **Wand, M.P. and Jones, M.C.** (1995) *Kernel smoothing*. Chapman & Hall.
- [65] **Watson, G.S.** (1964) Smooth regression analysis. *Sankhya* Ser. A, No. 26, pp. 101-116.
- [66] **Wu, C.-F. J.** (1986) Bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, No. 4, pp. 1261-1350.
- [67] **Yang, L. and Tschernig, R.** (1999) Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society*, Ser. B, Stat Methodol. 61, No. 4, pp. 793-815.