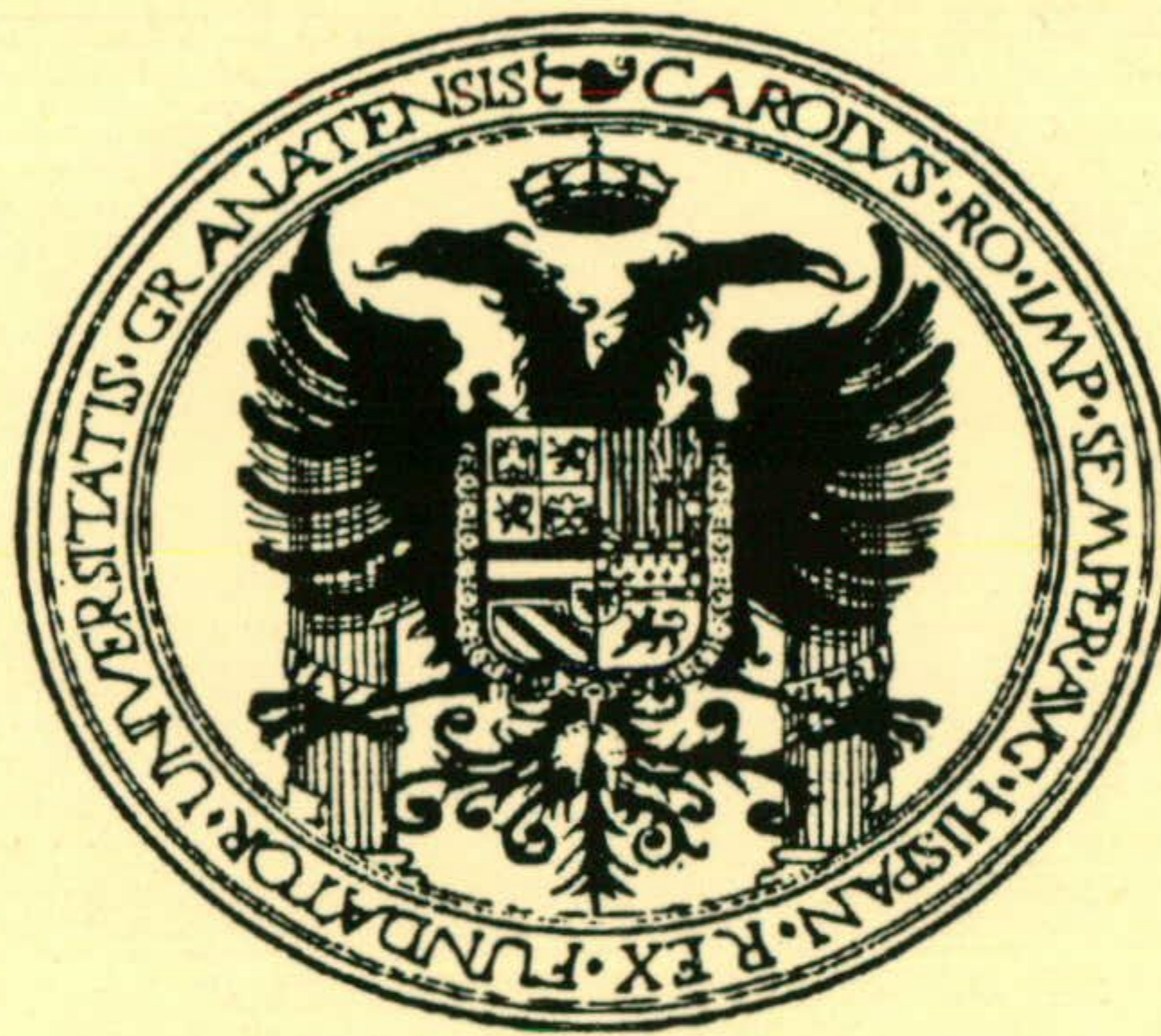


t 5/167

UNIVERSIDAD DE GRANADA



Departamento de Estadística e I. O.

TESIS DOCTORAL

---

---

*Elección de los parámetros que intervienen en  
un modelo de regresión local*

---

---

María Dolores Martínez Miranda  
Granada, Agosto de 2000

# UNIVERSIDAD DE GRANADA



UNIVERSIDAD DE GRANADA  
Facultad de Ciencias  
Fecha ...13/09/00.....  
ENTRADA NUM. 2615.....

Departamento de Estadística e I. O.

## TESIS DOCTORAL

---

*Elección de los parámetros que intervienen en  
un modelo de regresión local*

---

BIBLIOTECA UNIVERSITARIA  
GRANADA  
Nº Documento 613381507  
Nº Copia 15583594

UNIVERSIDAD DE GRANADA  
01 SET. 2000  
COMISION DE DOCTORADO

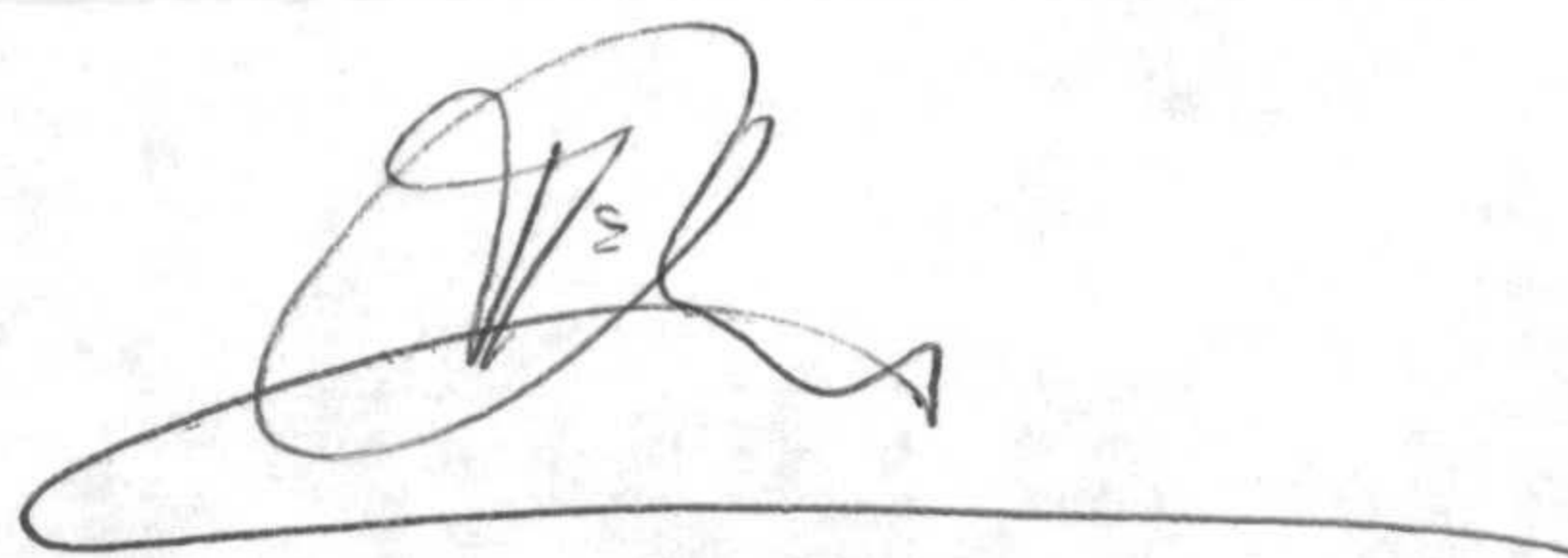
María Dolores Martínez Miranda  
Granada, Agosto de 2000

**TESIS DOCTORAL**

María Dolores Martínez Miranda

**ELECCIÓN DE LOS PARÁMETROS QUE INTERVIENEN EN UN  
MODELO DE REGRESIÓN LOCAL**

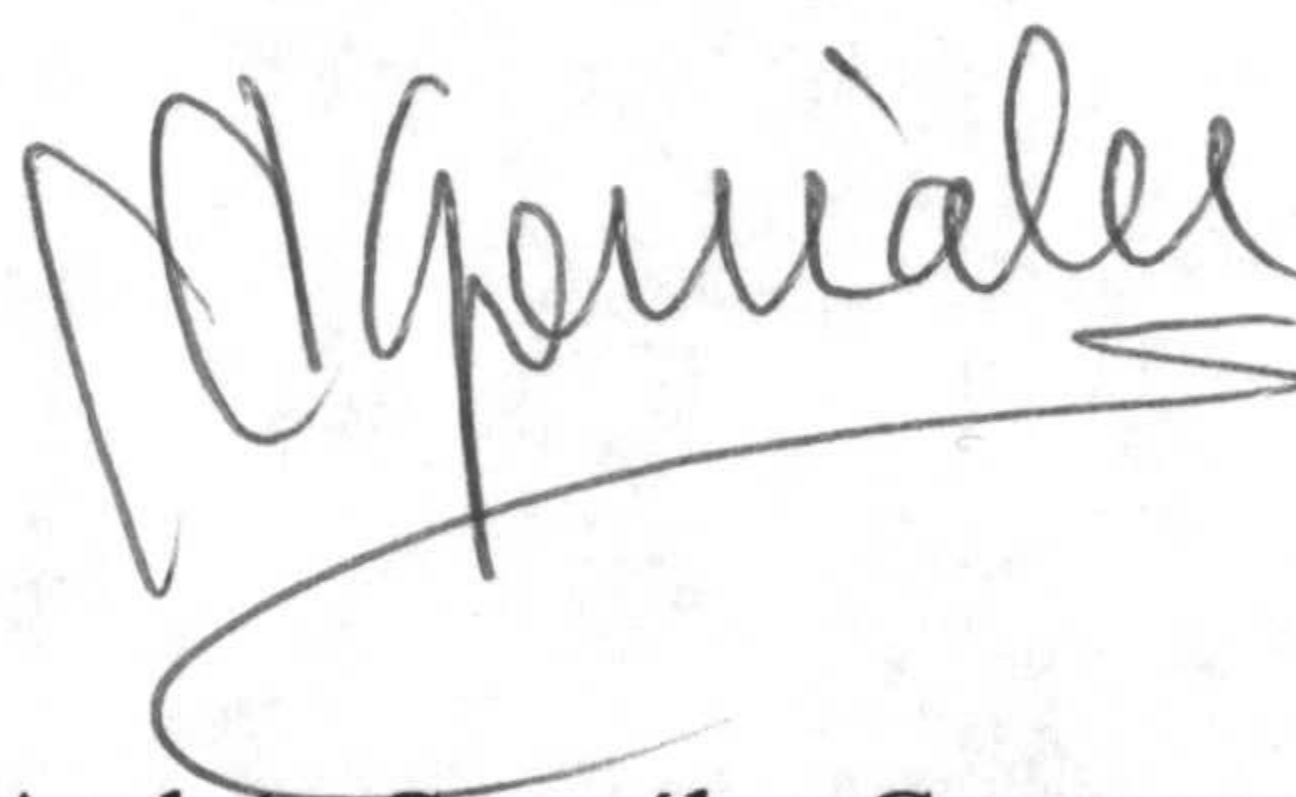
Memoria presentada para optar al  
grado de Doctor en Ciencias por  
María Dolores Martínez Miranda



María Dolores Martínez Miranda

Vº Bº

Director de Tesis



Prof. Dr. D. Andrés González Carmona

Departamento de Estadística e Investigación Operativa  
Universidad de Granada

*A mi padre y a mi madre, ejemplos  
para mí de trabajo, y a los que debo  
todo lo que soy.*

## Agradecimientos

Son muchas las personas que me han ayudado, apoyado y animado, durante el desarrollo de este trabajo, a todas y cada una de ellas mi más profundo y sincero agradecimiento:

De modo particular quiero agradecer al director de esta memoria, el Prof. Dr. D. Andrés González Carmona, la confianza que desde siempre ha depositado en mí, su ayuda y el apoyo que de él he recibido a lo largo de estos años de trabajo y durante la elaboración de esta memoria.

A mis compañeros de este departamento de Estadística e I.O. de la Universidad e Granada, y en especial a su director, el Prof. Dr. D. Ramón Gutiérrez Jaímez, al que agradezco la confianza y el interés mostrado, así como el proporcionarme los medios y facilidades necesarias para realizar mi trabajo.

A mi familia y a mis amigos, por estar siempre a mi lado...Gracias.

Y gracias a Dios porque me ha dado las fuerzas necesarias, y ha puesto en mi vida gente maravillosa que me empuja a luchar.

# Índice General

<b>Prólogo</b>	<b>1</b>
<b>1 Introducción</b>	<b>5</b>
1.1 Regresión paramétrica y regresión no paramétrica . . . . .	5
1.2 Orígenes de la regresión no paramétrica . . . . .	8
1.3 El modelo de regresión no paramétrica . . . . .	9
1.3.1 Modelo de diseño fijo . . . . .	10
1.3.2 Modelo de diseño aleatorio . . . . .	10
1.4 Criterios de error . . . . .	11
1.5 Técnicas de regresión no paramétrica . . . . .	12
1.5.1 Estimadores núcleo . . . . .	13
1.5.2 Regresión polinomial local . . . . .	15
1.5.3 Regresión localmente ponderada . . . . .	16
1.5.4 Suavizamiento por <i>splines</i> . . . . .	16
1.5.5 Métodos basados en desarrollos en serie ortogonal . . . . .	17
<b>2 Fundamentos de la regresión polinomial local</b>	<b>19</b>
2.1 Definición del estimador polinomial local . . . . .	19
2.1.1 Casos particulares: Estimador de Nadaraya-Watson y estima- dor lineal local. . . . .	21
2.2 Parámetros que intervienen en la definición . . . . .	22
2.3 Ventajas de la regresión polinomial local . . . . .	23
2.4 Aproximaciones asintóticas del sesgo y de la varianza del estimador .	24
2.4.1 Núcleos equivalentes . . . . .	26
2.4.2 Expresiones para el estimador lineal local . . . . .	28
2.5 Elección óptima del ancho de banda . . . . .	28
2.5.1 Elección óptima de un ancho de banda local . . . . .	28
2.5.2 Elección óptima de un ancho de banda global . . . . .	29
2.6 Elección del grado del ajuste polinomial . . . . .	30
2.6.1 Conveniencia de los ajustes de grado impar . . . . .	31

<b>3</b>	<b>Procedimientos para la selección del ancho de banda</b>	<b>33</b>
3.1	Introducción . . . . .	33
3.2	Una clasificación de los procedimientos de selección . . . . .	35
3.2.1	Criterios para la clasificación . . . . .	35
3.2.2	Elección global y elección local del ancho de banda . . . . .	35
3.2.3	Métodos <i>plug-in</i> . . . . .	37
3.2.4	Métodos de selección automática . . . . .	38
3.2.5	Procedimientos basados en <i>bootstrap</i> . . . . .	39
3.2.6	Métodos de selección automática frente a métodos <i>plug-in</i> . . . . .	40
3.3	El selector de Fan y Gijbels (1995b) . . . . .	42
3.3.1	Criterio de la suma de cuadrados residual . . . . .	44
3.3.2	Pasos para elegir un ancho de banda global . . . . .	44
3.3.3	Pasos para elegir un ancho de banda local . . . . .	45
3.4	Validación cruzada por mínimos cuadrados . . . . .	45
3.4.1	Un inconveniente: Correlación negativa . . . . .	46
3.4.2	Versión local del criterio de validación cruzada . . . . .	46
3.5	El selector de Ruppert (1997) . . . . .	47
<b>4</b>	<b>Elección <i>bootstrap</i> del ancho de banda</b>	<b>49</b>
4.1	La metodología <i>bootstrap</i> . . . . .	49
4.1.1	Validez y aplicabilidad de <i>bootstrap</i> . . . . .	50
4.2	<i>Bootstrap</i> en estimación de densidades . . . . .	51
4.3	<i>Bootstrap</i> en regresión no paramétrica . . . . .	52
4.3.1	Remuestreo sobre pares y remuestreo sobre residuos . . . . .	52
4.3.2	Antecedentes: Técnicas <i>bootstrap</i> desarrolladas . . . . .	53
4.4	<i>Bootstrap</i> sobre los pares . . . . .	53
4.4.1	Sobre la validez del <i>bootstrap</i> clásico . . . . .	53
4.4.2	Estimación <i>bootstrap</i> del error cuadrático medio . . . . .	55
4.4.3	Expresión exacta del estimador <i>bootstrap</i> . . . . .	56
4.5	<i>Bootstrap</i> clásico sobre los residuos . . . . .	58
4.5.1	Estimación <i>bootstrap</i> del error cuadrático medio . . . . .	58
4.5.2	Expresión exacta del estimador <i>bootstrap</i> . . . . .	59
4.5.3	Sobre el ancho de banda piloto . . . . .	61
4.6	<i>Wild bootstrap</i> sobre los residuos . . . . .	61
4.6.1	Estimación <i>bootstrap</i> del error cuadrático medio . . . . .	62
4.6.2	Elección <i>bootstrap</i> del ancho de banda . . . . .	63
4.6.3	Expresión exacta del estimador <i>bootstrap</i> . . . . .	63
4.6.4	Expresión asintótica del sesgo y la varianza <i>bootstrap</i> . . . . .	65
4.6.5	Elección del ancho de banda piloto . . . . .	68

---

<b>5</b>	<b>Comportamiento finito muestral del selector <i>bootstrap</i></b>	<b>75</b>
5.1	Introducción . . . . .	75
5.2	Descripción de las simulaciones . . . . .	76
5.3	Aspectos computacionales . . . . .	76
5.3.1	Minimización del estimador <i>bootstrap</i> . . . . .	78
5.3.2	Aproximación de las integrales . . . . .	78
5.3.3	Disminución del coste computacional . . . . .	79
5.4	Sobre el estimador <i>bootstrap</i> del error cuadrático medio . . . . .	79
5.5	Estimaciones <i>bootstrap</i> del sesgo y de la varianza . . . . .	84
5.6	Sobre la elección local del parámetro piloto . . . . .	88
5.7	El estimador polinomial local con ancho de banda <i>bootstrap</i> . . . . .	93
5.8	Conclusiones . . . . .	95
5.8.1	Versión suavizada del selector <i>bootstrap</i> . . . . .	95
<b>A</b>	<b>Gráficos adicionales</b>	<b>99</b>
<b>B</b>	<b>Computación estadística</b>	<b>113</b>
B.1	Obtención del selector <i>bootstrap</i> del ancho de banda . . . . .	113
B.2	Evaluación local de las componentes de sesgo y de varianza . . . . .	129
B.3	Análisis gráfico de los resultados . . . . .	133
	<b>Bibliografía</b>	<b>148</b>





# Índice de Figuras

5.1	Representación de las tres funciones de prueba objeto de la simulación.	77
5.2	Comparación del error cuadrático medio teórico (curva punteada) y <i>bootstrap</i> (curva continua), en los puntos de interés definidos para la función de prueba, $m_1$ . Cada fila corresponde a un tamaño muestral, cada columna a uno de los tres puntos de interés. . . . .	80
5.3	Comparación del error cuadrático medio teórico (curva punteada) y <i>bootstrap</i> (curva continua), en los puntos de interés definidos para la función de prueba, $m_2$ . Cada fila corresponde a un tamaño muestral, cada columna a uno de los tres puntos de interés. . . . .	81
5.4	Comparación del error cuadrático medio teórico (curva punteada) y <i>bootstrap</i> (curva continua), en los puntos de interés definidos para la función de prueba, $m_3$ . Cada fila corresponde a un tamaño muestral, cada columna a uno de los tres puntos de interés. . . . .	82
5.5	Sesgo y varianzas teóricos (curvas punteadas) y estimaciones <i>bootstrap</i> (curvas continuas), utilizando la elección teórica del parámetro piloto. Las curvas corresponden al punto, $x$ , $x = -0.5$ , y cada fila a un tamaño muestral: 50, 200 y 500. . . . .	85
5.6	Sesgo y varianzas teóricos (curvas punteadas) y estimaciones <i>bootstrap</i> (curvas continuas), utilizando la elección teórica del parámetro piloto. Las curvas corresponden al punto, $x$ , $x = 0$ , y cada fila a un tamaño muestral: 50, 200 y 500. . . . .	86
5.7	Sesgo y varianzas teóricos (curvas punteadas) y estimaciones <i>bootstrap</i> (curvas continuas), utilizando la elección teórica del parámetro piloto. Las curvas corresponden al punto, $x$ , $x = 1$ , y cada fila a un tamaño muestral: 50, 200 y 500. . . . .	87
5.8	Sesgo y varianzas teóricos (curvas punteadas); estimaciones <i>bootstrap</i> usando la elección teórica del parámetro piloto (curvas discontinuas), y utilizando otros posibles valores de dicho parámetro (curvas continuas). Las curvas corresponden al punto, $x$ , $x = -0.5$ , y a un tamaño muestral de 50. . . . .	89

- 5.9 Sesgo y varianza teóricos (curvas punteadas); estimaciones *bootstrap* usando la elección teórica del parámetro piloto (curvas discontinuas), y utilizando otros posibles valores de dicho parámetro (curvas continuas). Las curvas corresponden al punto,  $x, x = -0.5$ , y a un tamaño muestral de 200. . . . . 90
- 5.10 Sesgo y varianza teóricos (curvas punteadas); estimaciones *bootstrap* usando la elección teórica del parámetro piloto (curvas discontinuas), y utilizando otros posibles valores de dicho parámetro (curvas continuas). Las curvas corresponden al punto,  $x, x = -0.5$ , y a un tamaño muestral de 500. . . . . 91
- 5.11 Representación gráfica del ancho de banda piloto óptimo para las tres funciones de prueba. El tamaño muestral considerado es 200. . . . . 92
- 5.12 Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap*. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap*. Cada fila corresponde a una función de prueba. El tamaño muestral considerado es de 200. . . . . 94
- 5.13 Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizado con  $span = 5$ . El tamaño muestral considerado es 200. . . . . 97
- A.1 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden al punto  $x, x = -0.5$ , y al tamaño muestral 50. (Sólo aparece la curva *bootstrap* puesto que la curva teórica sale fuera del rango considerado.) . . . . . 100
- A.2 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden al punto,  $x, x = 0$ , y al tamaño muestral 50. . . . . 101
- A.3 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = 1$ ; y al tamaño muestral 50. . . . . 102
- A.4 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = -0.5$ ; y al tamaño muestral 200. . . . . 103

- A.5 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x$ ,  $x = 0$ ; y al tamaño muestral 200. . . . . 104
- A.6 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x$ ,  $x = 1$ ; y al tamaño muestral 200. . . . . 105
- A.7 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x$ ,  $x = -0.5$ ; y al tamaño muestral 500. . . . . 106
- A.8 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x$ ,  $x = 0$ ; y al tamaño muestral 500. . . . . 107
- A.9 Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x$ ,  $x = 1$ ; y al tamaño muestral 500. . . . . 108
- A.10 Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 2$ . El tamaño muestral considerado es 200. . . . . 109
- A.11 Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 3$ . El tamaño muestral considerado es 200. . . . . 110
- A.12 Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 4$ . El tamaño muestral considerado es 200. . . . . 111
- A.13 Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 10$ . El tamaño muestral considerado es 200. . . . . 112

# Prólogo

Es conocida la importancia que han adquirido en las últimas décadas los enfoques no paramétricos dentro de la Estadística. Las versiones no paramétricas de los procedimientos clásicos, resultan muy atractivas ya que no formulan hipótesis demasiado restrictivas sobre la estructura subyacente desconocida.

Dentro del campo de la Regresión, las técnicas de regresión no paramétrica permiten lograr una mejor adaptación a los datos disponibles, logrando estimaciones más próximas a la curva de regresión subyacente y, de nuevo, sin hacer fuertes hipótesis sobre la misma.

Esta memoria se enmarca dentro del campo de la regresión no paramétrica, buscando como objetivo fundamental la estimación de la función de regresión. El contexto que se ha asumido es de tipo univariante, esto es, una sola variable explicativa; y bajo una situación general de heterocedasticidad en el modelo de regresión.

Entre todas las técnicas de regresión no paramétrica que se han desarrollado en este campo, nos hemos centrado en la denominada *regresión polinomial local*, la cual ha cobrado un gran interés en los últimos años debido a su buen comportamiento y ventajas sobre otras que van quedando relegadas, como es la regresión por *estimadores tipo núcleo*, ya que presenta graves problemas de sesgo en la frontera. Y tomamos como problema central de esta tesis la elección del denominado *ancho de banda* o *parámetro de suavizamiento*, que determina el buen comportamiento del estimador polinomial local de la función de regresión.

El problema de la selección del *ancho de banda* ha sido ampliamente estudiado y ha originado una gran cantidad de literatura. Son muchas las técnicas desarrolladas para su selección, bajo distintos contextos y para las distintas técnicas de regresión no paramétrica. En esta memoria realizamos una clasificación de todas ellas y recogemos algunas de las más importantes desarrolladas recientemente para la *regresión polinomial local*.

Gran parte de los selectores del *ancho de banda* han supuesto generalizaciones de otros, diseñados en el campo de la *estimación no paramétrica de densidades*. El planteamiento del problema de estimación de una densidad resulta más sencillo de tratar que el de la regresión, dado que tan sólo interviene una variable aleatoria (unidimensional o multidimensional); pero plantear generalizaciones de metodologías desarrolladas en este contexto al de la regresión conlleva el problema adicional de

manejar los residuos del modelo, de modo que en general no resultan inmediatas, ni tan siquiera paralelas.

La metodología *bootstrap* ha sido empleada por diversos autores, para la elección del parámetro ancho de banda, en la estimación de densidades, obteniendo resultados bastante satisfactorios. Las generalizaciones a la regresión fueron inicialmente planteadas por Hall (1990), en casos sencillos, no obstante estas aproximaciones no resultan competitivas con otros selectores desarrollados aquí, y además no han sido generalizadas a estimadores relevantes en los últimos años, como son los polinomiales locales, que ocupan nuestra atención.

En este trabajo se propone la metodología *bootstrap* como vía de selección del ancho de banda, bajo tres perspectivas: La primera supone realizar un *bootstrap* sobre los pares de observaciones, a la que nos referimos como *bootstrap clásico sobre los pares*; y las dos restantes realizan un *bootstrap* sobre los residuos del modelo, según la metodología clásica, *bootstrap clásico sobre los residuos*, y según la variante denominada *wild bootstrap sobre los residuos*.

De las tres versiones propuestas, la heterocedasticidad supuesta en el modelo de regresión y algunas consideraciones hechas por Efron (1993) sobre el remuestreo de los pares, nos llevan a inclinarnos por la versión *wild bootstrap sobre los residuos*. Para dicha modalidad hacemos un estudio de sus propiedades teóricas, bajo un punto de vista asintótico, que validan el uso del procedimiento; además de un estudio de simulación que describe el comportamiento del selector *bootstrap* para muestras finitas. Los primeros resultados obtenidos resultan bastante satisfactorios, no obstante se han propuesto modificaciones posteriores, que lo hacen competitivo con otros, como los de Fan y Gijbels (1995), y Ruppert (1997), cuya bondad ha sido probada en la literatura, pero que presentan una mayor complejidad de cálculo y de tratamiento, que el aquí descrito.

La memoria se estructura en cinco capítulos y dos apéndices. El primer capítulo se dedica a generalidades sobre regresión no paramétrica, haciendo un breve recorrido por las técnicas de regresión no paramétrica habitualmente empleadas, además se establecen las hipótesis y el modelo de regresión bajo el que se describe todo el trabajo presentado. El segundo capítulo define y estudia las propiedades fundamentales del estimador polinomial local, planteando el problema de la elección de los parámetros que en él intervienen, esto es, el ancho de banda, el grado del ajuste polinomial local, y la función núcleo. La selección del primero, el ancho de banda, va a ocupar el resto de la memoria, de modo que el capítulo 3 hace una evaluación y una descripción del estado actual del problema, a la vez que propone una clasificación de todos los selectores del ancho de banda que han aparecido hasta el momento. En los capítulos 4 y 5 se encuentran las aportaciones fundamentales de esta memoria, en el primero se describe la selección propuesta para el ancho de banda del estimador polinomial local, empleando una estimación *bootstrap* del error cuadrático medio local; además de un estudio de las propiedades teóricas asintóticas; y en el

segundo se realiza un estudio de simulación que permite valorar el comportamiento de la metodología propuesta usando varias funciones de regresión, y considerando distintos tamaños muestrales. Finalmente se incluyen dos apéndices, el primero de los cuales contiene gráficos que asisten en el estudio finito-muestral del capítulo 5; y el segundo se dedica a la programación realizada en S Plus 2000 y R 1.1.x., incluyendo las funciones que, en dichos entornos de programación, han sido definidas para el cálculo del selector propuesto, así como para la obtención de los resultados del estudio de simulación.

Por tanto, podemos resumir las aportaciones realizadas en este trabajo en las siguientes:

- Se realiza una clasificación de todos los procedimientos que hasta hoy se han diseñado para la selección del parámetro *ancho de banda*.
- Se propone un nuevo selector del parámetro, generalizando la metodología *bootstrap* propuesta por Hall (1990) al contexto más general y complejo que adoptamos aquí; y además a diferencia de estos primeros intentos, el selector aquí propuesto es de tipo local consiguiendo una buena adaptación a los datos. Se obtiene la expresión exacta que conduce al *ancho de banda bootstrap* sin necesidad de realizar remuestreo, salvando inconvenientes habitualmente asociados a este tipo de metodologías, como es el excesivo coste computacional.
- Se realiza un estudio de sus propiedades teóricas bajo una perspectiva asintótica, obteniéndose una representación asintótica del selector que lleva a deducir que se trata de un estimador asintóticamente óptimo.
- Se investiga la elección de un parámetro piloto que interviene en la metodología *bootstrap* propuesta, de nuevo bajo una perspectiva asintótica, y se deduce la expresión teórica del parámetro piloto óptimo.
- Se lleva a cabo un estudio de simulación para valorar el comportamiento del selector propuesto en muestras finitas. En vista de lo obtenido se plantea una mejora del procedimiento suavizando el selector local, cuyos resultados resultan bastante satisfactorios.
- Se realizan implementaciones de la metodología propuesta, en los entornos de programación S Plus 2000 y R 1.1.x.

Del estudio realizado, destacamos como problemas abiertos, y líneas para trabajos futuros los siguientes:

- Extender los resultados teóricos, obtenidos para estimadores polinomiales locales de grado 1, al caso de un grado,  $p$ , genérico; y la consecuente validación finito-muestral extendiendo el estudio de simulación a tal situación general.

- Extender el selector *bootstrap* a modelos de regresión local multidimensionales.
- Comparar minuciosamente, a través de simulaciones, el selector propuesto con otros selectores relevantes y equiparables, como son los de Fan y Gijbels (1995b) y Ruppert (1997).
- Aplicar la metodología *bootstrap* propuesta para la obtención de un selector del *ancho de banda*, a la construcción de intervalos de confianza para la curva de regresión.
- Completar la definición de los selectores de tipo *bootstrap* propuestos, formulando y estudiando tanto desde el punto de vista teórico como del práctico, versiones de un *bootstrap suavizado* sobre los pares de observaciones, según se motiva en el capítulo 4.
- Mejorar las implementaciones informáticas realizadas incluyendo algoritmos diseñados para reducir el coste computacional de los cálculos del *estimador polinomial local* en la práctica, como son las técnicas *binning* y los algoritmos *updating*.

Distintas partes de esta memoria han sido presentadas y aceptadas en el XXV Congreso Nacional de Estadística e I.O., SEIO 2000; en el *International Seminar on Nonparametric Inference*, ISNI 2000; y en la *XIV Conference of the International Association for Statistical Computing*, COMPSTAT 2000.



# Capítulo 1

## Introducción

Dentro del amplio campo de técnicas encaminadas a la estimación de la función de regresión, comenzaremos marcando y definiendo el punto de bifurcación que da lugar a las metodologías de estudio denominadas: *Regresión paramétrica* o clásica, y *Regresión no paramétrica*.

### 1.1 Regresión paramétrica y regresión no paramétrica

Sea una variable de respuesta,  $Y$ , y una variable<sup>1</sup> explicativa,  $X$ . Consideremos un conjunto de datos  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ . Entonces la relación entre las variables se puede describir de la forma siguiente:

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \dots, n,$$

donde  $m$  es la función de regresión desconocida, esto es, la esperanza de  $Y$  condicionada a  $X$ , y  $\epsilon_i$  son los errores de observación (medición, manipulación, etc.), supuestos variables aleatorias independientes con media 0.

Un estudio de regresión bajo tal planteamiento, busca principalmente tres objetivos:

1. Explorar y representar la relación existente entre la variable explicativa,  $X$ , y la variable  $Y$ .
2. Dar predicciones del comportamiento de  $Y$  para valores de la variable  $X$  aún no observados.
3. Dar estimaciones de algunas propiedades de la función de regresión,  $m$ , como puede ser hacer un estudio de sus derivadas.

---

<sup>1</sup>Consideramos en este trabajo tan sólo el caso de modelos de regresión univariantes.

Una primera aproximación de la función,  $m$ , se puede obtener considerando el diagrama de dispersión para las observaciones,  $(X_i, Y_i)$ . La nube que describe tales puntos puede dar una idea, más o menos clara de la forma y de algunas de las características más notables que tiene la función,  $m$ . Y es aquí donde surgen dos puntos de vista diferentes para la resolución del problema de regresión, que da lugar a dos metodologías distintas, aunque muy ligadas entre sí. En concreto se puede seguir cualquiera de las dos formulaciones siguientes:

- Partiendo de consideraciones sobre la nube de puntos observada, o de alguna información previa sobre la forma de la función  $m$ , proponer una familia *paramétrica* de funciones,

$$\{m_\theta, \theta \in \Theta\}.$$

De este modo, la selección de la función dentro de la familia (que consistirá en estimar un número finito de parámetros desconocidos) se hará según algún criterio especificado que suele estar relacionado con alguna función de los residuos,  $\phi(Y_i - m_\theta(X_i))$ .

- Estimar la función,  $m$ , utilizando como única información los datos observados, pero sin suponer que dicha función posea una forma paramétrica definida. La única hipótesis que se hace sobre la función,  $m$ , es que sea *suave*, donde esta suavidad se entiende en términos de derivabilidad.

La primera de estas aproximaciones cae dentro de lo que se conoce como **Regresión paramétrica**; y el segundo enfoque conduce a la denominada **Regresión no paramétrica**.

En principio, la regresión no paramétrica constituye un procedimiento más general y más flexible, dado que formula pocas hipótesis sobre una función, de la que habitualmente, en una primera aproximación como es la recogida de algunos datos, se tiene casi un completo desconocimiento. De este modo, la regresión no paramétrica permite abordar los puntos 1 y 2, de forma bastante satisfactoria.

Generalmente un planteamiento de regresión no paramétrica supone que la función de regresión pertenece a alguna familia de funciones infinito dimensional, como puede ser la familia de funciones diferenciables con segunda derivada cuadrado integrable. Obviamente se trata de hipótesis mucho menos restrictivas que las impuestas en un planteamiento clásico de regresión paramétrica.

No obstante las ventajas de seguir un procedimiento de regresión paramétrica, donde se tienen garantizadas las buenas propiedades de la estimación resultante, son evidentes frente a un procedimiento de regresión no paramétrica, en el que, al hacer

muy pocas hipótesis sobre la función desconocida, se tendrá una mayor complejidad en la función estimada resultante.

Además obtener una función paramétrica como estimación de  $m$  será idóneo para cumplir el objetivo 2 de la regresión. Sin embargo, aún en tal caso, esto es, en el cumplimiento del objetivo 2, obtener estimaciones no paramétricas, que generalmente estarán próximas a los datos observados, puede servir incluso para, en un segundo paso, proponer un modelo paramétrico más adecuado.

Así, se puede resumir la utilidad de la regresión no paramétrica en los tres puntos siguientes:

- *Exploración de datos.* Aún en el caso en que nuestros objetivos se encaminen hacia la utilización de un modelo paramétrico, obtener y representar gráficamente la superficie de regresión no paramétrica ajustada a los datos, puede servir de guía en la difícil elección del modelo adecuado.
- *Diagnóstico no paramétrico de modelos paramétricos.* Varios trabajos en la literatura se han dirigido al estudio de técnicas de diagnóstico de un modelo paramétrico ajustado por técnicas clásicas usuales, basadas en estimaciones no paramétricas.
- *Obtención de una superficie de regresión no paramétrica.* Las estimaciones obtenidas directamente mediante técnicas de regresión no paramétricas, proporcionan una superficie de regresión que presenta muy buenas propiedades en cuanto a su definición intuitiva y su carácter suavizador del diagrama de dispersión.

Dos procedimientos habituales en regresión paramétrica son la regresión lineal, y en general la regresión polinomial. Una regresión lineal presenta, sin duda alguna, como característica más atractiva, su sencillez. Estimar una función lineal de una variable, de la que se tiene un conjunto de observaciones, queda reducido a la resolución de un sistema lineal de dos ecuaciones lineales con dos incógnitas. Y habitualmente ofrece una buena aproximación, ya que representa de una forma sencilla la tendencia general de la función desconocida,  $m$ .

Sin embargo, cuando los datos parecen mostrar que la función presenta cierta curvatura, considerar un modelo paramétrico lineal supondría obviar tales características y por tanto una pérdida de información. En tales casos se puede conseguir un mejor ajuste si se introduce un mayor número de parámetros, suponiendo modelos de tipo polinómico de grado dos o superior.

Aún así, si los datos presentan características diferentes, ofreciendo distintas zonas en el diagrama de dispersión, con distinta curvatura, intentar globalizar tal diversidad observada proponiendo una forma funcional ya sea lineal, o polinomial

de grado alto, será algo bastante complicado y a menudo supondrá cometer un error considerable.

Por tanto parecería conveniente eliminar en cierto modo ese carácter globalizador que conlleva una regresión paramétrica, atendiendo un poco más a las propiedades locales que presenta el diagrama de dispersión.

Otro de los inconvenientes que presenta la regresión polinomial es que las funciones polinomiales no son muy flexibles en la modelización de los problemas que hay en la práctica, ya que estas funciones tienen en todos los puntos derivadas de todos los órdenes.

Existen varias formas de reparar los inconvenientes de un ajuste paramétrico polinomial:

1. Aproximación por *splines*. Una primera posibilidad es permitir posibles discontinuidades en las derivadas, lo que lleva a considerar funciones *splines*.
2. Método de series ortogonales. Otra posibilidad es hacer el desarrollo en serie ortogonal de la función de regresión, elegir unos pocos subconjuntos útiles de las funciones base, y usarlos para aproximar la función de regresión.
3. Modelización local. Se trata de aplicar modelos de regresión polinomial de grado bajo localmente. De modo que para un punto de estimación dado, se define un entorno y se modeliza la función de regresión mediante un polinomio que se ajusta localmente en dicho entorno, según algún criterio definido.

La última de las aproximaciones propuestas cae dentro de la Regresión no paramétrica que hemos definido, y nos referimos a ella con el nombre de **regresión local**.

## 1.2 Orígenes de la regresión no paramétrica

La teoría y los métodos de suavizamiento o regresión no paramétrica, han sido desarrollados fundamentalmente en los últimos veinte años. El creciente interés por estas metodologías ha tenido dos razones principales: la primera, que los estadísticos se dieron cuenta de que planteamientos puramente paramétricos no aportaban la flexibilidad necesaria para la estimación de las curvas que aparecían en la práctica; y la segunda razón estaba en el avance de la informática y el desarrollo de un *hardware* que permitía calcular esos estimadores no paramétricos que resultaban muy atractivos por su flexibilidad y buen comportamiento.

No obstante, el suavizamiento mediante regresión no paramétrica, tiene una larga tradición, de hecho ha sido desarrollado desde comienzos del siglo XIX.

Durante cien años, las aportaciones metodológicas más importantes provenían de los investigadores que analizaban datos de tipo económico y empresarial. En

1857 el economista Engel construía la curva que hoy en día se conoce como *regresograma*. Habitualmente las aplicaciones se restringían al caso de una sola variable predictora, con valores igualmente espaciados; dado que casos más generales conllevaban un esfuerzo computacional no permitido entonces. A comienzos del siglo XX matemáticos como Henderson y Whittaker enriquecieron ampliamente la teoría y los métodos de suavizamiento. Henderson, en 1916, publicó resultados que aclaraban en gran medida las propiedades de la regresión local. La regresión local se introdujo en varias aplicaciones en series temporales dentro del entorno empresarial y la economía, convirtiéndose en una herramienta fundamental en dichos sectores.

En la literatura estadística de los años sesenta y principios de los setenta, la regresión local se aplicó a estudios de regresión más generales, esto es, con más de una variable predictora y valores no equiespaciados. Los primeros estimadores de regresión local o no paramétrica propuestos fueron los *suavizadores núcleo*, como los de Watson (1964) y Nadaraya (1964), que correspondían a ajustes polinomiales locales de grado cero. A partir de estos esfuerzos iniciales surgió una gran cantidad de literatura teórica centrada en la obtención de resultados de tipo asintótico, en los cuales el tamaño muestral tiende a infinito y el parámetro ancho de banda tiende a cero. En los años setenta, la regresión local con polinomios de grado mayor que cero fue introduciéndose en la literatura (Cleveland 1979); Fan demostraba la superioridad teórica y práctica de la regresión local con polinomios no constantes; en concreto, para grado uno, se obtienen estimadores de regresión que son óptimos en el sentido minimax, entre todos los estimadores que tienen una forma lineal en las observaciones.

En las últimas décadas, ha crecido enormemente la literatura a este respecto, refinando y perfeccionando los procedimientos de regresión local, y centrando la atención en la adecuada elección de los parámetros decisivos en el buen comportamiento de los estimadores de regresión resultantes, como son el ancho de banda o parámetro de suavizado, y el grado de los ajustes polinomiales locales. En este último punto destacan recientes trabajos de Fan y Gijbels en 1995.

### 1.3 El modelo de regresión no paramétrica

Consideramos un modelo de regresión univariante, esto es, la variable predictora es de tipo unidimensional.

A la hora de formular y plantear un problema de regresión, ya sea desde un punto de vista paramétrico o no paramétrico, es necesario distinguir los dos posibles contextos que se pueden considerar y a los que nos referimos por *diseño fijo* y *diseño aleatorio*.

Considerar un modelo de diseño fijo significa que el conjunto de datos de que se dispone son observaciones de la variable de respuesta,  $Y$ , tomadas en valores prefi-

dados de la variable explicativa; esto es, los valores en el eje de abscisas del diagrama de dispersión han sido fijados por el experimentador y diseñados a su conveniencia. Sin embargo, un diseño aleatorio considera que los datos observados corresponden a una muestra aleatoria desde una población  $(X, Y)$ . Obsérvese entonces que un diseño aleatorio supone un caso más general, si bien un diseño fijo en el que, habitualmente por simplicidad, se suponen los puntos de observación equiespaciados en un intervalo, es más sencillo y reduce considerablemente tanto los desarrollos teóricos como las implementaciones de tipo práctico.

A continuación formulamos el problema de regresión univariante en ambos contextos.

### 1.3.1 Modelo de diseño fijo

Sea  $\{x_1, \dots, x_n\}$  un conjunto de valores fijados de la variable,  $X$ , y sea  $\{Y_1, \dots, Y_n\}$  el resultado de observar la variable  $Y$  en dichos puntos<sup>2</sup>. Dichas observaciones se describen según un modelo del tipo:

$$Y_i = m(x_i) + \epsilon_i,$$

siendo  $m$  la función de regresión, y  $\epsilon_i$ , los errores del modelo, que son variables aleatorias independientes<sup>3</sup>, con media cero, y cuya varianza coincide con la varianza de  $Y_i$ , y la denotaremos  $\sigma^2(x_i)$ . Es habitual suponer que dichas varianzas son constantes de forma que,  $\sigma^2(x_i) = \sigma^2$ , considerando lo que se denomina un modelo de regresión *homocedástico*.

### 1.3.2 Modelo de diseño aleatorio

En este caso, las observaciones,  $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ , son variables aleatorias independientes e idénticamente distribuidas según la variable bidimensional  $(X, Y)$ . Entonces dichas observaciones se pueden describir según un modelo:

$$Y_i = m(X_i) + \epsilon_i,$$

donde,  $m$  es la función de regresión, esto es, la media condicional de  $Y$  dada  $X$ ; y  $\epsilon_i$  son variables aleatorias independientes, con media cero y varianza,  $\sigma^2(X_i)$ , que representa en este caso, la varianza condicional de  $Y$  dada  $X$ . De nuevo aquí, podemos asumir dos situaciones, la que acabamos de exponer en la que las varianzas

<sup>2</sup>Utilizaremos las mayúsculas para indicar una variable aleatoria y las minúsculas para indicar un valor fijo, no aleatorio.

<sup>3</sup>Si bien esta hipótesis se puede relajar en diversos contextos a una incorrelación; o incluso considerar el caso de errores dependientes, lo que daría lugar a una metodología de estudio contextualizada en un entorno que no es el que aquí nos ocupa.

de los errores se suponen distintas, esto es, una situación de *heterocedasticidad*; o bien simplificar el problema suponiendo que las varianzas son constantemente un valor  $\sigma^2$ , lo cual daría lugar a un modelo de regresión homocedástico.

El trabajo que aquí presentamos considera un modelo de regresión univariante, bajo un diseño aleatorio, en una situación general de heterocedasticidad.

## 1.4 Criterios de error

El estudio del comportamiento de un estimador de regresión requiere la especificación de un criterio de error apropiado. Para realizar un análisis tal, será de interés considerar los dos aspectos siguientes:

1. Por una parte, interesará medir la bondad del estimador en la estimación en cada punto. Las medidas de tal error se denominan **medidas del error local**.
2. Por otra parte, se podrá estudiar el error que se comete en la estimación de la curva completa, esto es, en todo el intervalo de estimación considerado, obteniendo las denominadas **medidas del error global**.

Entre las medidas locales se utilizan las siguientes:

- **Error cuadrático medio condicional (MSE):**

$$\text{MSE}(\hat{m}(x)) = E_{Y/X} [(\hat{m}(x) - m(x))^2]$$

siendo  $\hat{m}(\cdot)$  el estimador de regresión considerado,  $x$  un punto de estimación cualquiera; y  $E_{Y/X}$  denota la esperanza calculada en la distribución de la variable de respuesta,  $Y$ , condicionada a la variable explicativa,  $X$ .

- **Error cuadrático medio incondicional ( $\text{MSE}_I$ ).** Tomando esperanzas sobre la distribución de la variable  $X$ :

$$\text{MSE}_I(\hat{m}(x)) = E[\text{MSE}(x)].$$

Y como medidas globales:

- **Error cuadrático integrado medio condicional (MISE).** Integrando en el intervalo de estimación considerado:

$$\text{MISE}(\hat{m}) = E_{Y/X} \left[ \int (\hat{m}(x) - m(x))^2 dx \right]$$

- **Error cuadrático integrado medio incondicional (MISE<sub>I</sub>).** Tomando esperanza sobre la distribución de la variable  $X$ :

$$\text{MISE}_I(\hat{m}) = E[\text{MISE}(\hat{m})]$$

- **Error cuadrático integrado predictivo condicional (PISE):**

$$\text{PISE}(\hat{m}) = E_{Y/X} \left[ \int (\hat{m}(x) - m(x))^2 f(x) dx \right],$$

siendo  $f(\cdot)$  la función de densidad marginal de  $X$ .

- **Error cuadrático integrado predictivo incondicional (PISE<sub>I</sub>):**

$$\text{PISE}_I(\hat{m}) = E[\text{PISE}(\hat{m})]$$

- **Error cuadrático promediado sobre las observaciones condicional (ASE):**

$$\text{ASE}(\hat{m}) = \sum_{i=1}^n (\hat{m}(X_i) - m(X_i))^2 w(X_i),$$

para algunas ponderaciones  $w(X_i)$ .

- **Error cuadrático promediado medio (MASE).** Tomando esperanzas sobre la distribución de la variable explicativa,  $X$ :

$$\text{MASE}(\hat{m}) = E \left[ \sum_{i=1}^n (\hat{m}(X_i) - m(X_i))^2 w(X_i) \right].$$

De entre todas las formas presentadas de medir el error cometido en la estimación de la función de regresión, consideramos en nuestro trabajo dos medidas de error, una de tipo local, que será el error cuadrático medio condicional (MSE); y otra de tipo global, que será su versión integrada, esto es, el error cuadrático integrado medio (MISE). La razón que nos lleva a elegir tales medidas es, fundamentalmente, que constituyen las más utilizadas en la literatura y su tratamiento resulta más sencillo y general que algunas de las versiones incondicionales antes presentadas.

## 1.5 Técnicas de regresión no paramétrica

Las técnicas de regresión no paramétrica habitualmente buscan la representación de las características locales observadas en un diagrama de dispersión, de ahí que a veces a sus estimadores se les llame *suavizadores locales*. Además la forma más simple



de los estimadores de regresión no paramétrica vienen a ser versiones univariantes *locales* de estimadores usuales de localización de las observaciones, tales como la media o la media ponderada, la mediana, etc.

Así en términos generales, la idea que subyace en estas técnicas de regresión no paramétrica o de suavizamiento se puede resumir de la forma siguiente<sup>4</sup>:

*Si  $m$  se supone suave, entonces las observaciones,  $Y_i$ , en puntos próximos a  $x$ , deberían contener información acerca del valor de  $m$  en  $x$ . De este modo, sería posible emplear una media local de los puntos próximos a  $x$  para construir el estimador de  $m(x)$ .*

Siguiendo esta idea se han desarrollado diversas técnicas de regresión, con distintos matices, si bien, muy relacionadas entre sí. En Martínez-Miranda, M.D. (1996) se exponen varias de las más relevantes.

A continuación describimos las líneas generales que siguen algunas de las técnicas de regresión no paramétrica que han sido desarrolladas en la literatura.

### 1.5.1 Estimadores núcleo

Los estimadores núcleo<sup>5</sup> se definen como una media ponderada de las observaciones que caen dentro de un entorno del punto de estimación. Su nombre proviene del hecho de que las ponderaciones que se asignan a cada observación vienen definidas a través de una función peso denominada *función núcleo*.

En cada punto considerado, un estimador núcleo define un entorno centrado en dicho punto y cuya longitud viene determinada por la elección de un parámetro, denominado *ancho de banda*; y con las observaciones que caen en dicho entorno calcula una media ponderada que asignará mayor peso a las más próximas al punto de estimación.

Las distintas definiciones para las ponderaciones en el estimador núcleo, dan lugar a diversas modalidades de estimadores núcleo que han ido apareciendo en la literatura. No obstante en todos ellos aparecen como elementos comunes la presencia de una función núcleo y de un parámetro ancho de banda.

A continuación definimos algunos de los más relevantes, por orden cronológico, e introducimos alguna notación que será empleada:

Denotamos por  $K$  a la función núcleo, que se supone función de densidad simétrica; y por  $h$  al parámetro ancho de banda,  $h > 0$ . Definimos  $K_h(u) = h^{-1}K(u/h)$ .

#### 1. Estimador de Nadaraya-Watson (1964).

Sobre un diseño aleatorio, se define como:

<sup>4</sup>Eubank, R. (1988).

<sup>5</sup>Del inglés *kernel estimates*.

$$\hat{m}_h(x) = n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)}{\hat{f}_h(X_i)} Y_i, \quad (1.1)$$

donde  $\hat{f}_h$  es el estimador núcleo de Rosenblatt-Parzen de la densidad marginal de  $X$ , dado por:

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

## 2. Estimador de Priestley-Chao (1972).

Originalmente definido por estos autores sobre un diseño fijo equiespaciado donde se supone que  $0 \leq x_1 \leq \dots \leq x_n \leq 1$ , el estimador tiene la expresión siguiente:

$$\hat{m}_h(x) = \sum_{i=1}^n (x_i - x_{i-1}) K_h(x - x_i) Y_i.$$

## 3. Estimador de Clark (1977).

De nuevo sobre un diseño fijo como el planteado en la definición del anterior estimador de Priestley-Chao, se define el estimador de Clark de la forma siguiente:

$$\hat{m}_h(x) = \int l(u) K_h(x - u) du,$$

donde  $l(u)$  es la poligonal que une las observaciones.

## 4. Estimador de Gasser-Müller (1979).

Sobre un diseño aleatorio, se define el estimador como sigue:

$$\hat{m}_h(x) = \sum_{i=1}^n Y_i \int_{s_{i-1}}^{s_i} K_h(x - u) du,$$

donde  $s_0 = -\infty$ ,  $s_n = \infty$ ,  $X_{(i)} \leq s_i \leq X_{(i+1)}$ ,  $i = 1, \dots, n-1$  y  $X_{(i)}$  es el  $i$ -ésimo estadístico ordenado.

Obsérvese que el estimador de Nadaraya-Watson y el de Priestley-Chao se definen a partir de la evaluación de una función núcleo, por lo que reciben el nombre de estimadores de evaluación; mientras que los de Clark y de Gasser-Müller emplean la convolución con dicha función núcleo, por lo que se denominan estimadores de convolución.

### 1.5.2 Regresión polinomial local

La idea de la regresión polinomial local fue desarrollada sistemáticamente por primera vez por Stone (1977, 1980, 1982) y Cleveland (1979), no obstante estudios más exhaustivos de los ajustes polinomiales locales se pueden encontrar en Fan (1992, 1993), Fan y Gijbels (1992), y Ruppert y Wand (1994).

En nuestro trabajo consideramos como base esta técnica de regresión, por lo cual será definida y estudiada con detalle en los siguientes capítulos, no obstante ahora daremos una definición somera de tales estimadores.

Dada la hipótesis de suavidad sobre la función de regresión, es posible suponer, usando el teorema de Taylor, que dicha función puede aproximarse localmente por medio de funciones polinomiales. De este modo, se define el siguiente problema de mínimos cuadrados ponderados:

$$\min_{\beta_0, \dots, \beta_p} \sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j (X_i - x)^j \right\}^2 K_h(X_i - x). \quad (1.2)$$

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

$$\hat{m}_h(x) = \hat{\beta}_0;$$

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

Los estimadores núcleo antes definidos vienen a ser un caso particular de estimadores de regresión polinomial local, de hecho los ajustes locales de grado cero dan lugar al estimador núcleo de Nadaraya-Watson. No obstante los estimadores núcleo no muestran un comportamiento del todo satisfactorio, adoleciendo de problemas como son el efecto frontera, para lo cual son necesarias versiones corregidas de los mismos, y de grandes sesgos. Así, frente a la utilización de estimadores tipo núcleo para la estimación de la función de regresión, siempre es preferible aumentar la complejidad del modelo pasando a ajustes polinomiales locales de al menos grado uno. Es más, se demuestra que el estimador de Nadaraya-Watson tiene eficiencia minimax nula (Fan 1992, 1993), frente a los estimadores lineales locales que son óptimos en un sentido minimax.

Comparaciones entre ajustes de tipo constante, esto es, estimadores núcleo, y ajustes de al menos grado uno pueden encontrarse, con detalle, en Chu y Marron (1991), Fan (1992), y Hastie y Loader (1993).

### 1.5.3 Regresión localmente ponderada

La regresión localmente ponderada fue introducida por Cleveland (1979). Su definición está muy ligada a la de la anterior regresión polinomial local, si bien el estimador definido por Cleveland, define los entornos locales fijando el número de observaciones que contendrán, en lugar del ancho del entorno. Este tipo de estimadores se conoce con el nombre de estimadores por el entorno más próximo, y al número de observaciones que caen en cada entorno local, *span* o en general, parámetro de suavizado.

Esta técnica es una de las más ampliamente utilizadas y está disponible en el entorno S-Plus, en las funciones *loess* y *lowess*, esta última corresponde a una versión robusta del procedimiento.

Sin entrar en detalles describiremos la idea base de la versión robusta del procedimiento:

Se ajusta un polinomio de grado  $p$  localmente resolviendo un problema de mínimos cuadrados ponderados del tipo (1.2) anterior; pero ahora el ancho de banda,  $h$ , será variable en cada punto, si bien función de un parámetro fijo, el *span* antes definido. Se calculan los residuos y se les asignan pesos menores a los de mayor magnitud. Entonces se vuelve a ajustar un polinomio localmente, pero asignando ahora a las observaciones pesos definidos por el producto de los pesos del ajuste inicial y los pesos asignados a los residuos. De este modo las observaciones con grandes residuos se infraponderan en el segundo paso. El procedimiento se repite un número determinado de veces hasta obtener el estimador localmente ponderado robusto.

### 1.5.4 Suavizamiento por *splines*

El suavizamiento por *splines*, si bien se considera como una técnica de regresión no paramétrica, constituye una evolución de la inferencia paramétrica clásica, definiendo un puente que cubre el vacío entre los métodos paramétricos clásicos y los no paramétricos.

Los *splines* carecen de una forma paramétrica, no obstante, se pueden escribir como una combinación lineal de las funciones de una base, que usualmente admiten una representación polinomial.

El suavizamiento por *splines* se formula de la siguiente forma:

Dado el conjunto de datos de partida, se considera un problema de mínimos cuadrados penalizados cuya función objetivo se define como:

$$S(m) = \sum_{i=1}^n (Y_i - m(X_i))^2 + \lambda \int \{m^{(r)}(x)\}^2 dx,$$

donde  $\lambda$  es el parámetro de suavizamiento, controlando la cantidad de suavidad.

Entonces, si se minimiza dicha función objetivo sobre el conjunto de las funciones con derivada de orden  $r$  cuadrado integrable, se obtiene como solución el spline de suavizamiento, que es justamente un spline de orden  $r$  con nodos en los puntos  $X_i$ . Es habitual hacer la elección  $r = 2$ , obteniéndose el spline cúbico de suavizamiento.

El comportamiento de los *splines* en la práctica es bastante bueno, no obstante presentan graves problemas computacionales y su cálculo se hace muy complicado cuando se consideran contextos de regresión multivariantes. Esto hace que sea preferible optar por procedimientos más sencillos, como los expuestos antes de regresión núcleo o regresión polinomial local, no perdiendo demasiado en la bondad de los estimadores, y ganando por el contrario en simplicidad y facilidad de uso.

### 1.5.5 Métodos basados en desarrollos en serie ortogonal

Si la función de regresión,  $m$ , se supone perteneciente al espacio de las funciones cuadrado integrables, entonces dicha función admite un desarrollo en serie ortogonal del tipo

$$m(x) = \sum_{j=1}^{\infty} \theta_j q_j(x),$$

donde las funciones,  $q_i(x)$ , constituyen una base completa y ortonormal, y los coeficientes son  $\theta_j = \int q_j(x)m(x)dx$ .

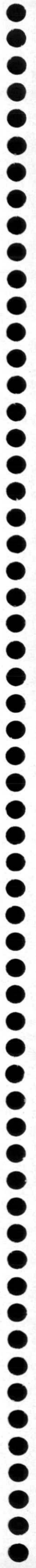
Un estimador natural de los coeficientes viene dado por:

$$\hat{\theta}_j = n^{-1} \sum_{i=1}^n q_j(X_i)Y_i,$$

de forma que se construye el estimador de  $m$  basado en las subseries truncadas:

$$\hat{m}_N(x) = \sum_{j=1}^N \hat{\theta}_j q_j(x),$$

siendo el número de coeficientes de Fourier considerados en la subserie,  $N$ , una función de  $n$  tendiendo a  $\infty$ , que constituye el parámetro de suavizamiento del estimador.



## Capítulo 2

# Fundamentos de la regresión polinomial local

### 2.1 Definición del estimador polinomial local

Sea  $\{(X_1, Y_1), (X_2, Y_2) \dots, (X_n, Y_n)\}$  una muestra aleatoria de una población  $(X, Y)$ . Supongamos que dichos datos han sido generados mediante el modelo

$$Y = m(X) + \sigma(X)\epsilon,$$

donde  $m(\cdot)$ , representa la función de regresión desconocida, y los residuos verifican  $E[\epsilon] = 0$ ,  $\text{Var}(\epsilon) = 1$ ; siendo  $X$  y  $\epsilon$  independientes.

En el modelo hemos supuesto una situación generalizada de heterocedasticidad donde  $\sigma^2(x)$ , representa la varianza condicional de  $Y$  dado que  $X = x$ . Bajo un diseño aleatorio como el planteado denotaremos por  $f(\cdot)$ , a la función de densidad marginal de  $X$ .

Supongamos que la función de regresión desconocida es suficientemente suave, en concreto supongamos que existe la derivada de orden  $(p + 1)$  en un punto,  $x_0$ . Entonces, mediante el teorema de Taylor, es posible aproximar  $m(x)$  localmente mediante un polinomio de orden  $p$ , en concreto se verifica el siguiente desarrollo de la función en un entorno de  $x_0$ :

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p.$$

Obsérvese que, según este desarrollo, la función de regresión desconocida,  $m$ , localmente, es una función polinómica.

La regresión polinomial local toma como base y motivación esta idea, y define un estimador de  $m(x_0)$  según un ajuste polinomial de grado  $p$ , usando tan sólo las observaciones más próximas, esto es, las que caen en un entorno de  $x_0$ . La selección

de los puntos más próximos se consigue ajustando el polinomio mediante un criterio de mínimos cuadrados ponderados.

Definimos por tanto un problema de mínimos cuadrados ponderados del tipo:

$$\text{Minimizar}_{\beta} \sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right\}^2 K_h(X_i - x_0), \quad (2.1)$$

donde  $\beta = (\beta_0, \dots, \beta_p)^T$ ,  $h$  es un parámetro positivo que controla el tamaño del entorno local, denominado **ancho de banda**; y las ponderaciones locales (dependen del punto de estimación considerado),  $K_h(\cdot) = h^{-1}K(\cdot/h)$ , vienen determinadas por una función,  $K$ , que en general supondremos<sup>1</sup> que es una densidad simétrica con soporte acotado, y que recibe el nombre de **función núcleo**.

Si denotamos por  $\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_p)^T$  a la solución del problema (2.1), según el desarrollo de Taylor antes expuesto, el estimador de la función de regresión en  $x_0$ , coincidiría con la primera componente,  $\hat{\beta}_0$ . Si denotamos por  $\hat{m}_h(x_0)$ , el estimador, entonces:

$$\hat{m}_h(x_0) = e_1^T \hat{\beta};$$

y el resto,  $\hat{\beta}_1, \dots, \hat{\beta}_p$ , corresponderían a las estimaciones de las derivadas de la función de regresión en el punto de estimación considerado. De este modo, si denotamos por  $\hat{m}_h^\nu(x_0)$  a la estimación de la derivada de orden  $\nu$  ( $\nu = 1, \dots, p$ ), cada una de ellas vendrá dada por:

$$\hat{m}_h^\nu(x_0) = e_\nu^T \hat{\beta},$$

siendo  $e_i$  el vector de orden  $p+1$ , cuyos elementos son todos cero excepto el  $i$ -ésimo, que es igual a 1 ( $i = 1, \dots, p+1$ ).

Por semejanza con la teoría paramétrica de regresión polinomial por mínimos cuadrados ponderados, se puede obtener fácilmente una expresión matricial para los coeficientes,  $\hat{\beta}_j$ , anteriores. No obstante hemos de introducir alguna notación más:

Sea

$$X_{x_0} = \begin{pmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{pmatrix},$$

$$y = (Y_1, \dots, Y_n)^T$$

También, sea  $W_{x_0, h}$  la matriz diagonal de orden  $n$ , dada por

$$W_{x_0, h} = \text{diag}\{K_h(X_i - x_0)\}.$$

<sup>1</sup>Esta hipótesis se relajará o se ampliará dependiendo de las necesidades técnicas de algunos de los resultados que obtendremos más adelante.



De este modo el problema (2.1) se puede escribir de la forma

$$\underset{\beta}{\text{Minimizar}} \quad (y - X_{x_0}\beta)^T W_{x_0,h} (y - X_{x_0}\beta), \quad (2.2)$$

y el vector solución será

$$\hat{\beta} = (X_{x_0}^T W_{x_0,h} X_{x_0})^{-1} X_{x_0}^T W_{x_0,h} y. \quad (2.3)$$

Con lo cual, la estimación de la función de regresión en cada punto  $x_0$ , vendrá dada por:

$$\hat{m}_h(x_0) = e_1^T (X_{x_0}^T W_{x_0,h} X_{x_0})^{-1} X_{x_0}^T W_{x_0,h} y. \quad (2.4)$$

### 2.1.1 Casos particulares: Estimador de Nadaraya-Watson y estimador lineal local.

Dentro de la clase de los estimadores polinomiales locales (que proporcionan las distintas elecciones de sus parámetros) consideramos los casos particulares de grados  $p = 0$  y  $p = 1$ .

- Si consideramos ajustes locales de tipo constante,  $p = 0$ , el problema de mínimos cuadrados ponderados que define el estimador en tal caso será de la forma:

$$\underset{\beta_0 \in \mathbb{R}}{\text{Minimizar}} \quad \sum_{i=1}^n \{Y_i - \beta_0\}^2 K_h(X_i - x_0),$$

de este modo, el estimador resultante será la media ponderada normalizada de las observaciones, esto es:

$$\hat{m}_h(x_0) = \frac{\sum_{i=1}^n K_h(X_i - x_0) Y_i}{\sum_{j=1}^n K_h(X_j - x_0)}$$

que coincide con el estimador núcleo de Nadaraya-Watson definido en (1.1). Por tanto, la clase de los estimadores polinomiales locales engloba a la clase de los estimadores núcleo, o más exactamente a uno de los estimadores núcleo más ampliamente utilizado, como es el estimador de Nadaraya-Watson.

- Si hacemos la elección  $p = 1$ , el estimador resultante se denomina *estimador lineal local*. Es posible desarrollar la expresión matricial (2.4) anterior y obtener la siguiente expresión para el estimador lineal local:

$$\hat{m}_h(x_0) = n^{-1} \frac{\{S_{n,2}(x_0; h) - S_{n,1}(x_0; h)(X_i - x_0)\} K_h(X_i - x_0)}{S_{n,2}(x_0; h)S_{n,0}(x_0; h) - S_{n,1}(x_0; h)^2} Y_i,$$

donde

$$S_{n,j} = n^{-1} \sum_{i=1}^n (X_i - x_0)^j K_h(X_i - x_0). \quad (2.5)$$

## 2.2 Parámetros que intervienen en la definición

Dentro de esta formulación, y de la propia solución propuesta al problema de regresión planteado, es necesario tener en cuenta algunos factores de crucial importancia:

- La elección del parámetro ancho de banda,  $h$ . Como ocurre en todas las técnicas de suavizamiento, dicha elección resulta decisiva en el buen funcionamiento del estimador resultante. Elegir un ancho de banda muy grande conllevaría un sobresuavizamiento de los datos, provocando un gran aumento del sesgo; por el contrario, un ancho de banda demasiado pequeño supondría un infrasuavizamiento y consecuentemente la función estimada sería muy "ruidosa", esto es, proporcionaría estimaciones con una alta variabilidad. Por tanto, elegir un ancho de banda apropiado supondrá un compromiso entre sesgo y varianza. El problema de la elección de dicho parámetro será estudiado en profundidad más adelante.
- La elección del orden del polinomio que se ajusta localmente. Los ajustes polinomiales locales se suelen considerar de grado fijo y no muy alto, siendo usuales las elecciones lineales o cuadráticas. Los ajustes cuadráticos son preferibles cuando los datos muestran una fuerte curvatura. Los polinomios locales de grado cero dan lugar a graves problemas de sesgo en la frontera<sup>2</sup>, siendo preferibles los ajustes de tipo lineal. De este modo, para un ancho de banda dado, un orden de polinomio alto producirá una reducción en el sesgo y un aumento en la varianza de las estimaciones resultantes, y por el contrario, un orden bajo dará lugar a estimaciones con grandes sesgos y sobresuavizadas. Trataremos este problema con más detalle en puntos posteriores.
- La elección de la función núcleo,  $K$ . Aunque el papel de dicha función no es determinante en la calidad del estimador resultante, se puede probar<sup>3</sup> que la

<sup>2</sup>Problema conocido en la literatura como *efecto frontera*.

<sup>3</sup>Cheng, Fan y Marron (1993).

mejor elección que se puede hacer es la conocida como *Núcleo de Epanechnikov*,

$$K(u) = \frac{3}{4}(1 - u^2)1_{\{|u| < 1\}},$$

dado que minimiza el error cuadrático medio (MSE) del estimador polinomial local resultante. No obstante, son habituales otras elecciones atendiendo a razones de tipo técnico, como es la situación que plantearemos en este trabajo; en concreto en las aplicaciones consideramos un núcleo del tipo bponderado, que se define como sigue:

$$K(u) = \frac{15}{16}(1 - u^2)^2 1_{\{|u| < 1\}}.$$

## 2.3 Ventajas de la regresión polinomial local

La regresión polinomial local presenta muchas ventajas, tanto teóricas como prácticas, sobre otros tipos de regresión no paramétrica como es el caso de los bien conocidos y ampliamente utilizados, estimadores núcleo.

- **Adaptación al tipo de diseño.** La regresión polinomial local no tiene problema para adaptarse a varios tipos de diseño, tanto fijos como aleatorios, además no sufre los efectos frontera de los que adolecen los estimadores núcleo antes mencionados. De hecho el estimador de Gasser-Müller definido en el capítulo introductorio no puede adaptarse a diseños de tipo aleatorio, incluso el comportamiento del estimador en tal caso es desastroso en la asignación de ponderaciones a cada observación. En Chu y Marron (1991) se pueden encontrar más detalles a este respecto. De forma similar, el estimador de Nadaraya-Watson no se adapta bien a diseños altamente *clusterizados* (donde  $|f'(x)/f(x)|$  es grande) ya que el sesgo de dicho estimador sufre un gran incremento<sup>4</sup>.
- **Ausencia del efecto frontera.** En la formulación general de las técnicas de suavizamiento, cuando se considera un punto de estimación,  $x_0$ , próximo a la frontera del intervalo de estimación, es posible que el entorno local definido por el ancho de banda,  $x_0 \pm h$ , sobrepase dicho intervalo. De ahí que muchas propiedades válidas (generalmente en cuanto a condiciones de simetría sobre los momentos) en los puntos interiores no se mantienen para puntos próximos a la frontera, lo que hace que las estimaciones resultantes presenten sesgos considerables. Si se consideran contextos multidimensionales, el problema se acentúa aún más. Esta situación, conocida bajo el nombre de efecto frontera,

<sup>4</sup>Fan, J. (1992).

se presenta en todos los estimadores núcleo presentados en el primer capítulo, y son muchos los intentos que han aparecido en la literatura para solventar en alguna medida dicho problema. No obstante, ninguno de los métodos propuestos resulta eficiente y de fácil aplicabilidad. Los estimadores de regresión polinomial local hacen una corrección automática de los efectos frontera. Un estudio detallado se puede encontrar en Ruppert y Wand (1994) o bien en Fan y Gijbels (1996); si bien aquí resaltamos el hecho de que el sesgo de los estimadores polinomiales locales en la frontera no sufre los incrementos antes mencionados, si el grado del ajuste polinomial que se realiza localmente es impar, por tanto las elecciones lineales o cúbicas son las más adecuadas.

- **Eficiencia minimax.** También los estimadores polinomiales locales tienen buenas propiedades de eficiencia minimax. Fan (1992) demuestra que los estimadores polinomiales locales tienen una alta eficiencia minimax lineal y algunos de ellos, como es el caso de ajustes lineales locales con la función núcleo de Epanechnikov, y con un ancho de banda del orden de  $n^{-1/5}$ , alcanzan un 100% de eficiencia (tanto en los puntos interiores como en los puntos frontera). El riesgo minimax lineal empleado se define como:

$$R_L(h, C_{p+1}) = \inf_{\hat{m} \text{ lineal}} \sup_{m \in C_{p+1}} E [\{\hat{m}(x_0) - m(x_0)\}^2]$$

donde  $x_0$  representa un punto arbitrario dentro del intervalo de estimación,  $C_{p+1}$  es la clase de las funciones con derivada de orden  $p+1$  acotada en un entorno de  $x_0$ , y  $\hat{m}$  es cualquier estimador de regresión de tipo lineal, esto es, admite una representación de la forma<sup>5</sup>:

$$\hat{m}(x) = \sum_{i=1}^n W_i(x; X_1, \dots, X_n) Y_i.$$

## 2.4 Aproximaciones asintóticas del sesgo y de la varianza del estimador

Como comentábamos en el capítulo introductorio, consideraremos dos medidas de error para estudiar el comportamiento del estimador de regresión: El error cuadrático medio, MSE, como medida local, y su versión integrada, MISE, como medida global.

Es conocida la descomposición del error cuadrático medio de la forma:

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

<sup>5</sup>Obsérvese que dentro de esta clase se incluyen todos los estimadores núcleo, los estimadores polinomiales locales y además los splines de suavizamiento y los estimadores en serie ortogonal.

donde, el primer término corresponde al cuadrado del sesgo del estimador, y el segundo a su varianza, en ambos casos sobre la distribución condicional de la variable de respuesta,  $Y$ , a la variable  $X$ .

Denotaremos el término de sesgo por  $B(\widehat{m}_h(x))$ , y el término de varianza por  $V(\widehat{m}_h(x))$ , eliminando por comodidad en la notación, el indicador de la distribución sobre la que se calculan, si bien entenderemos que siempre hacemos referencia a la condicional antes definida.

A continuación obtenemos las expresiones del sesgo y de la varianza del estimador polinomial local.

El sesgo y la varianza del estimador,  $\widehat{m}_h(x_0)$ , calculados directamente a partir de la definición del mismo, en (2.4), vienen dados por:

$$\begin{aligned} B(\widehat{m}_h(x_0)) &= e_1^T (X_{x_0}^T W_{x_0,h} X_{x_0})^{-1} X_{x_0}^T W_{x_0,h} m - m(x_0) \\ &= e_1^T (X_{x_0}^T W_{x_0,h} X_{x_0})^{-1} X_{x_0}^T W_{x_0,h} r, \end{aligned} \quad (2.6)$$

$$V(\widehat{m}_h(x_0)) = e_1^T (X_{x_0}^T W_{x_0,h} X_{x_0})^{-1} (X_{x_0}^T \Sigma X_{x_0}) (X_{x_0}^T W_{x_0,h} X_{x_0})^{-1} e_1,$$

donde  $m = (m(X_1), \dots, m(X_n))^T$ ,  $r = m - X_{x_0} \beta$ , corresponde al vector de los residuos del ajuste polinomial local; y  $\Sigma = \text{diag} \{K_h^2(X_i - x_0) \sigma^2(X_i)\}$ .

Obsérvese que tales expresiones dependen de los parámetros ancho de banda, el grado del ajuste local o la función núcleo, de forma bastante complicada, lo que hace que resulte muy difícil interpretar la influencia que tienen en el buen comportamiento del suavizador local. Por esto resulta de interés obtener expresiones para el sesgo y la varianza, válidas para tamaños muestrales grandes<sup>6</sup>, esto es, expresiones de tipo asintótico.

Para ello introducimos la siguiente notación:

$$\mu_j = \int u^j K(u) du \quad \text{y} \quad \nu_j = \int u^j K^2(u) du;$$

y sean las matrices y vectores

$$S = (\mu_{j+l})_{0 \leq j, l \leq p} \quad C_p = (\mu_{p+1}, \dots, \mu_{2p+1})^T$$

$$\tilde{S} = (\mu_{j+l+1})_{0 \leq j, l \leq p} \quad \tilde{C}_p = (\mu_{p+2}, \dots, \mu_{2p+2})^T$$

$$S^* = (\nu_{j+l})_{0 \leq j, l \leq p}$$

<sup>6</sup>Resultados deducidos por Ruppert y Wand (1994), otros desarrollos más sencillos adaptados al caso univariante pueden encontrarse en Fan y Gijbels (1996).

Entonces, si suponemos que  $f(x_0) > 0$  y que  $f$ ,  $m^{(p+1)}$  y  $\sigma^2$  son funciones continuas en un entorno de  $x_0$ , y que además  $h \rightarrow 0$  con  $nh \rightarrow \infty$ , entonces la varianza y el sesgo asintóticos condicionales están dados por las siguientes expresiones:

$$V(\widehat{m}_h(x_0)) = e_1^T S^{-1} S^* S^{-1} e_1 \frac{\sigma^2(x_0)}{f(x_0)nh} + o_P((nh)^{-1}), \quad (2.7)$$

y para el término de sesgo, si  $p$  es impar:

$$B(\widehat{m}_h(x_0)) = e_1^T S^{-1} c_p \frac{1}{(p+1)!} m^{(p+1)}(x_0) h^{p+1} + o_P(h^{p+1}), \quad (2.8)$$

y en el caso en que  $p$  sea par, suponiendo que  $f'$  y  $m^{(p+2)}$  son funciones continuas en un entorno de  $x_0$  y que  $nh^3 \rightarrow \infty$ , el sesgo está dado por:

$$B(\widehat{m}_h(x_0)) = e_1^T S^{-1} \tilde{c}_p \frac{1}{(p+2)!} \left\{ m^{(p+2)}(x_0) + (p+2) m^{(p+1)}(x_0) \frac{f'(x_0)}{f(x_0)} h^{p+2} \right\} + o_P(h^{p+2}). \quad (2.9)$$

Por simplicidad en la notación, en lo sucesivo escribiremos  $B_h(x_0)$  para referirnos al sesgo, y  $V_h(x_0)$ , para la varianza.

### 2.4.1 Núcleos equivalentes

Para simplificar las expresiones anteriores se introducen los denominados núcleos equivalentes. Con este objetivo, comenzamos introduciendo alguna notación adicional.

Sea  $S_n = X_{x_0}^T W_{x_0,h} X_{x_0}$  la matriz cuadrada  $(S_{n,j+l})_{0 \leq j,l \leq p}$ , para los  $S_{n,j}$  definidos en (2.5).

El estimador,  $\widehat{m}_h(x_0)$ , se puede escribir también de la forma:

$$\widehat{m}_h(x_0) = e_1^T S_n^{-1} X_{x_0}^T W_{x_0,h} y = \sum_{i=1}^n W^n \left( \frac{X_i - x_0}{h} \right) Y_i,$$

donde  $W^n(t) = e_1^T S_n^{-1} (1, th, \dots, (th)^p)^T K(t)/h$ . Se puede probar fácilmente (Fan y Gijbels, 1996) que los pesos,  $W^n$ , satisfacen las siguientes condiciones, en términos de momentos:

$$\sum_{i=1}^n (X_i - x_0)^j W^n \left( \frac{X_i - x_0}{h} \right) = 0 \quad 0 \leq j \leq p.$$

Se tiene la siguiente expresión asintótica para los  $S_{n,j}$  definidos en (2.5):

$$S_{n,j} = nh^j f(x_0) \mu_j \{1 + o_P(1)\},$$

de donde se obtiene de forma inmediata

$$S_n = nf(x_0) H X_{x_0} H \{1 + o_P(1)\},$$

donde  $H = \text{diag}(1, h, \dots, h^p)$ . Y si se sustituye en la definición de  $W^n$ , se tiene que:

$$W^n(t) = \frac{1}{nhf(x_0)} e_1^T S^{-1}(1, t, \dots, t^p)^T K(t) \{1 + o_P(1)\}.$$

De este modo, el estimador polinomial local queda de la forma

$$\hat{m}(x_0) = \frac{1}{nhf(x_0)} \sum_{i=1}^n K^* \left( \frac{X_i - x_0}{h} \right) Y_i \{1 + o_P(1)\},$$

incluyendo un núcleo,  $K^*$ , que denominaremos *núcleo equivalente*, y cuya definición es

$$K^*(t) = e_1^T S^{-1}(1, t, \dots, t^p)^T K(t) = \left( \sum_{l=0}^p S^{0l} t^l \right) K(t)$$

con  $S^{-1} = (S^{jl})_{0 \leq j, l \leq p}$ .

El núcleo equivalente satisface las siguientes condiciones en términos de momentos:

$$\int u^j K^*(u) du = \delta_{0,j} \quad 0 \leq j \leq p.$$

En el caso del núcleo de Epanechnikov, el núcleo equivalente,  $K^*$ , para algunos valores de  $p$  viene dado por<sup>7</sup>:

- Si  $p = 1$ ,  $K^*(t) = K(t)$ ,
- Para  $p = 3$ ,  $K^*(t) = (\mu_4 - \mu_2^2)^{-1} (\mu_4 - \mu_2 t^2) K(t)$ .

Las expresiones asintóticas del sesgo y de la varianza condicionales en términos de los núcleos equivalentes, vienen dadas por:

$$B_h(x_0) = \left\{ \int t^{p+1} K^*(t) dt \right\} \frac{1}{(p+1)!} m^{(p+1)}(x_0) h^{p+1} + o_P(h^{p+1}),$$

$$V_h(x_0) = \int K^{*2}(t) dt \frac{\sigma^2(x_0)}{f(x_0)nh} + o_P((nh)^{-1}).$$

<sup>7</sup>Fan, J. y Gijbels, I. (1996).

### 2.4.2 Expresiones para el estimador lineal local

Considerando el caso particular del estimador lineal local, elección  $p = 1$ , la expresión del sesgo queda como sigue:

$$\begin{aligned} B_h(x_0) &= \left\{ \int t^2 K^*(t) dt \right\} \frac{1}{2} m''(x_0) h^2 + o_P(h^2) = \\ &= \frac{1}{2} m''(x_0) h^2 \mu_2 + o_P(h^2). \end{aligned} \quad (2.10)$$

Y la expresión de la varianza es:

$$V_h(x_0) = R(K) \frac{\sigma^2(x_0)}{f(x_0)nh} + o_P((nh)^{-1}), \quad (2.11)$$

donde  $R(K) = \int K^2(t) dt$ .

## 2.5 Elección óptima del ancho de banda

Trataremos en este punto la elección del ancho de banda óptimo considerando tanto que dicho parámetro es un parámetro global, constante para todos los puntos de estimación; como que se trate de un parámetro local que varía en cada punto de estimación considerado.

En ambas situaciones entenderemos por elección óptima del ancho de banda, aquella que hace mínimo el error cuadrático medio, si se trata del parámetro local; y cuando se trate del parámetro global será la que hace mínimo el error cuadrático medio integrado.<sup>8</sup> Obviamente dichas elecciones serán teóricas dado que dependerán de parámetros desconocidos y por tanto este estudio no será concluyente como solución al problema de elegir el mejor ancho de banda. En el capítulo siguiente trataremos este problema, estudiando distintos métodos de selección que han ido apareciendo en la literatura, y cuyo objetivo es aproximar lo mejor posible las elecciones óptimas teóricas que aquí presentaremos.

### 2.5.1 Elección óptima de un ancho de banda local

Definimos por tanto el ancho de banda local óptimo como:

$$h_{\text{opt}}(x_0) = \underset{h}{\operatorname{argmin}} \operatorname{MSE}(\hat{m}_h(x_0)).$$

Dado que el error cuadrático medio del estimador admite la descomposición en suma del sesgo al cuadrado y de la varianza, si utilizamos las expresiones asintóticas

<sup>8</sup> Ambos errores condicionados a la muestra considerada.



de dichos términos dadas en (2.8), (2.9) y (2.7), es posible obtener la expresión asintótica del ancho de banda local óptimo.

De este modo, el parámetro óptimo asintóticamente viene dado por<sup>9</sup>:

$$h_{\text{opt}}(x_0) = C_p(K) \left[ \frac{\sigma^2(x_0)}{\{m^{(p+1)}(x_0)\}^2 f(x_0)} \right]^{1/(2p+3)} n^{-1/(2p+3)} \quad (2.12)$$

donde la constante,  $C_p(K)$ , tiene la expresión

$$C_p(K) = \left[ \frac{(p+1)!^2 \int K^{*2}(t) dt}{2(p+1) \left\{ \int t^{p+1} K^*(t) dt \right\}^2} \right]^{1/(2p+3)}$$

Y en el caso particular del estimador lineal local,

$$h_{\text{opt}}(x_0) = \left[ \frac{R(K)\sigma^2(x_0)}{\mu_2^2 \{m''(x_0)\}^2 f(x_0)} \right]^{1/5} n^{-1/5}.$$

### 2.5.2 Elección óptima de un ancho de banda global

Si en lugar de considerar el error cuadrático medio, que depende del punto donde se estima, consideramos su integral respecto al espacio del diseño, esto es, el error cuadrático medio integrado, MISE, se define el ancho de banda global óptimo como:

$$h_{\text{opt}} = \underset{h}{\operatorname{argmin}} \operatorname{MISE}(\hat{m}_h)$$

Ahora el parámetro resultante no depende del punto de estimación, sino que es constante para todo el rango de puntos de estimación considerado.

Utilizando las expresiones asintóticas del sesgo y la varianza es posible obtener la expresión asintótica del ancho de banda global óptimo:

$$h_{\text{opt}} = C_p(K) \left[ \frac{\int \{\sigma^2(x)/f(x)\} dx}{\int \{m^{(p+1)}(x)\}^2 dx} \right]^{1/(2p+3)} n^{-1/(2p+3)}. \quad (2.13)$$

Que en el caso de ajustes lineales locales quedaría como:

$$h_{\text{opt}} = \left[ \frac{R(K) \int \{\sigma^2(x)/f(x)\} dx}{\mu_2^2 \int (m''(x))^2 dx} \right]^{1/5} n^{-1/5}.$$

<sup>9</sup>Fan, Gijbels, Hu y Huang (1996).

## 2.6 Elección del grado del ajuste polinomial

El problema de elegir el grado del polinomio que se ajustará localmente constituye otra de las importantes elecciones a realizar, puesto que determinará, en buena medida, el buen comportamiento del estimador resultante.

No obstante ha sido un problema que, a diferencia de la elección del ancho de banda, se ha resuelto considerando casos sencillos, como son los ajustes constantes (estimador de Nadaraya-Watson), lineales (estimador lineal local) o en algunos casos ajustes de tipo cuadrático o a lo sumo cúbico.

Es fácil observar que considerar ajustes polinomiales locales de órdenes elevados, conduce a posibles reducciones en el sesgo, si bien también conlleva un aumento de la variabilidad. De este modo, razonando de forma similar a la elección del ancho de banda, la selección del orden del polinomio constituirá un compromiso entre sesgo y varianza.

Intuitivamente parece claro que en zonas donde la función de regresión presente poca curvatura o incluso una forma lineal, será suficiente considerar un ajuste de tipo constante o de tipo lineal; mientras que en zonas que correspondan a picos o valles, zonas de fuerte curvatura, será necesario considerar un ajuste de grado dos o tres para poder describir lo observado. Esta idea nos lleva a la conclusión de que será conveniente considerar el grado del ajuste polinomial como un parámetro local que depende del punto donde se estime en cada ocasión. Siguiendo esta línea hay un trabajo de Fan y Gijbels (1995a) que presenta un procedimiento de selección local del grado del ajuste polinomial.

No obstante, y como venimos comentando, el grado del polinomio a ajustar y el ancho de banda están muy ligados y su elección responde a consideraciones similares sobre la función de regresión subyacente. En el artículo de Fan y Gijbels antes mencionado se observa que si se trata el problema de la elección del grado del ajuste localmente, el problema del ancho de banda pasa a ocupar un segundo plano, obteniéndose resultados similares para varias elecciones arbitrarias de este último. De forma similar, si se trata la elección del ancho de banda localmente, ajustes sencillos de tipo lineal pueden describir bastante bien curvas de regresión con distintos comportamientos locales.

En este trabajo consideramos como objetivo principal la elección del ancho de banda, buscando localidad para lograr una mejor adaptación a los datos observados, de forma que el grado de ajuste pasa a ocupar un segundo plano; de hecho, supondremos habitualmente el caso del estimador lineal local que se comporta de forma bastante satisfactoria.

Sin embargo, es interesante advertir que, según resultados de Fan y Gijbels (1996), son preferibles las elecciones impares de  $p$ . Una breve descripción del razonamiento se incluye a continuación.

### 2.6.1 Conveniencia de los ajustes de grado impar

Supongamos que se ajusta localmente un polinomio de grado  $p$ . El sesgo de tal ajuste será de orden  $h^{p+1}$  si  $p$  es impar (según la expresión (2.9)), y de orden  $h^{p+2}$  si  $p$  es par, (2.8), de forma que según señalábamos antes, considerar polinomios de grado alto conlleva reducciones de sesgo. No obstante la pregunta es: ¿Qué pasará con la varianza si se incrementa el grado del polinomio local?

Para responder a esta cuestión consideremos de nuevo la expresión obtenida para la varianza asintótica del estimador:

$$V_h(x_0) = \int K^{*2}(t) dt \frac{\sigma^2(x_0)}{f(x_0)nh} + o_P((nh)^{-1}).$$

El orden de la varianza asintótica es por tanto,  $(nh)^{-1}$ , el cual no se ve afectado por el parámetro  $p$ . No obstante si se estudia con más detalle el término constante (se entiende cuando  $n \rightarrow \infty$ ) para distintas elecciones de la función núcleo,  $K$ , se observa<sup>10</sup> un incremento en la varianza a medida que aumenta el grado del ajuste local.

Es notorio ver que no existe incremento en la varianza entre un grado par y su consecutivo impar, de forma que los ajustes lineales serán preferibles a los ajustes constantes; y los ajustes cúbicos serán más convenientes que los ajustes cuadráticos, puesto que suponen una reducción en el sesgo sin pagar ningún precio en la variabilidad.

---

<sup>10</sup>Fan y Gijbels (1996).



## Capítulo 3

# Procedimientos para la selección del ancho de banda

### 3.1 Introducción

El ancho de banda, también denominado parámetro de suavizamiento, interviene de forma muy similar en todos los procedimientos de regresión no paramétrica que desde 1964 han aparecido en la literatura; y en todos ellos, elegirlo supone un problema de solución compleja.

En el caso del estimador polinomial local definido en (2.4), será el ancho de banda el que determine el tamaño del entorno local considerado en la nube de puntos (centrado en cada punto de estimación), cuyos puntos intervendrán en el ajuste polinomial local. De ahí que si se considera para el parámetro,  $h$ , el valor 0, el estimador resultante será un interpolador de los datos observados; por el contrario si se toma muy grande,  $h \rightarrow \infty$ , los entornos locales englobarán todas las observaciones, lo que dará lugar al ajuste global de un polinomio del grado considerado.

Es obvio que una curva interpoladora de los datos será la que más se aproxime a la nube de puntos observada reduciendo al máximo el sesgo de las estimaciones locales, a costa de un gran aumento en la variabilidad del estimador local resultante. Por el contrario, el ajuste polinomial global que produciría considerar un ancho de banda tendiendo a  $\infty$ , proporciona estimaciones que pagan el precio de la poca variabilidad en términos de elevados sesgos. De este modo la elección del ancho de banda apropiado supondrá un compromiso entre el sesgo y la varianza del estimador, eligiendo por tanto el grado de suavidad del estimador (de ahí que el ancho de banda también reciba el nombre de parámetro de suavizamiento) y la complejidad del modelo, que variará desde el modelo más sencillo, para  $h \rightarrow \infty$ , al modelo más complejo para el caso  $h = 0$ .

Según lo expuesto, el ancho de banda que logre el perfecto equilibrio entre sesgo y varianza será el óptimo para definir el estimador local. Sin embargo habría que

distinguir dos posibles alternativas, una sería considerar sesgos y varianzas locales, esto es, dependientes del punto de estimación considerado en cada ocasión; y otra sería considerar ambas medidas globales para todos los puntos de estimación. La primera implicará la minimización del error cuadrático medio (MSE), que se descompone en un término de sesgo y otro de varianza; y la segunda la minimización del error cuadrático integrado medio (MISE).

Los dos puntos de vista planteados dan lugar a dos líneas de trabajo en la literatura correspondiente a esta cuestión: El estimador polinomial local con ancho de banda fijo y el estimador polinomial local con ancho de banda variable.

Por tanto, en primer lugar, habría que decidir entre el criterio local y global para el ancho de banda, y seguidamente buscar el ancho de banda que haga mínimo el error cuadrático medio (si se busca local) o el error cuadrático medio integrado en el caso global. Este problema se resolvió a partir de las expresiones asintóticas de los errores obtenidas en la sección 2.4 del capítulo anterior, y alcanzan su mínimo para el valor del ancho de banda local dado en la anterior expresión (2.12) que recordamos era

$$h_{opt}(x_0) = C_p(K) \left[ \frac{\sigma^2(x_0)}{\{m^{(p+1)}(x_0)\}^2 f(x_0)} \right]^{1/(2p+3)} n^{-1/(2p+3)}$$

y para el ancho global, la expresión (2.13) dada por

$$h_{opt} = C_p(K) \left[ \frac{\int \{\sigma^2(x)/f(x)\} dx}{\int \{m^{(p+1)}(x)\}^2 dx} \right]^{1/(2p+3)} n^{-1/(2p+3)}.$$

No obstante, aunque pudiera parecer que el problema está resuelto, estas expresiones son tan sólo formulaciones teóricas que dependen de varias cantidades como la derivada de orden  $p+1$  de la función de regresión, o la función varianza residual,  $\sigma^2(\cdot)$ , que por lo general serán desconocidas.

Por tanto el problema en realidad empieza aquí, es necesario idear procedimientos que permitan la selección de un ancho de banda que optimice el comportamiento del suavizador local, partiendo de la consideración de que la solución ideal es tan sólo teórica y, por tanto, inaccesible en la práctica.

La selección del parámetro ancho de banda ha sido el problema que más literatura ha desarrollado dentro del contexto de la regresión no paramétrica. 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 el ámbito de la regresión no paramétrica. De ahí que nos planteemos como próximo objetivo, ofrecer unos criterios de clasificación que permitan un estudio ordenado y en profundidad del

amplio trabajo que se ha desarrollado y se sigue desarrollando, para la resolución del problema.

## 3.2 Una clasificación de los procedimientos de selección

### 3.2.1 Criterios para la clasificación

Consideraremos dos criterios generales de clasificación:

1. Atendiendo a la naturaleza del parámetro: Se clasifican los procedimientos de selección en **globales**, o **constantes**, si consideran el ancho de banda como un parámetro fijo y constante para todos los puntos de estimación; y **locales**, o **variables**, si tratan el ancho de banda como un valor variable que depende del punto donde se estime en cada ocasión.
2. Atendiendo a la metodología seguida en el procedimiento de selección: se dividen en métodos **plug-in**, métodos de **selección automática** y procedimientos basados en **bootstrap**.

De las características generales que comparten cada uno de estos grupos de procedimientos y de las posibles subclasificaciones que se puedan establecer, tratamos en los siguientes apartados.

### 3.2.2 Elección global y elección local del ancho de banda

Los primeros planteamientos dentro de la teoría de regresión no paramétrica, involucraban un parámetro ancho de banda o parámetro de suavizado, fijo y constante para todos los puntos de estimación considerados.

Esta concepción planteaba un inmediato problema práctico, ya que era posible que en algún punto de estimación considerado, el entorno allí definido por el ancho de banda diera lugar a una banda en el diagrama de dispersión que no contuviera ninguna observación, imposibilitando el cálculo del estimador local. Obviamente este hecho ocurrirá siempre que no se disponga de muchos datos observados y sobre todo, si dichos datos no están suficientemente esparcidos en el diagrama de dispersión. Obviamente en tal caso habría puntos en los que sería necesario un ancho de banda excesivamente grande, para poder abarcar algo de lo observado.

Este primer problema dio lugar a una variante del estimador no paramétrico, el denominado *nearest neighbourhood estimator*, que traducimos como estimador basado en los vecinos más próximos. En este caso se supone que el entorno considerado en cada punto de estimación no viene determinado por su tamaño, sino

por el porcentaje de observaciones que contiene. Desde esta perspectiva, Cleveland (1979) presenta un estimador análogo al que hemos adoptado en este trabajo, y que habitualmente se conoce con el acrónimo LOESS. El estimador LOESS no considera, por tanto, el parámetro ancho de banda para definir el estimador, sino lo que se denomina *span*, que se define como el porcentaje de observaciones que intervienen en cada estimación puntual.

Con la definición de un estimador tal, se resuelve el problema de encontrar entornos vacíos de observaciones. No obstante, aún con esta modificación, surgen otros problemas respecto de la utilización de un *span* fijo, que se formulan a continuación.

¿Qué ocurrirá cuando la función de regresión que pretendemos estimar presente zonas de fuerte curvatura (picos, valles, etc.)? En los puntos de estimación ubicados en tales zonas, sería deseable un ancho de banda pequeño que permitiera describir bien la zona sin correr el riesgo de suavizar demasiado y por tanto incrementar en exceso el sesgo de la estimación resultante. Por el contrario, en zonas de la función de regresión correspondientes a tramos lineales o constantes, un ancho de banda grande permitirá un fuerte suavizamiento en los puntos de dicha zona sin producir un aumento de sesgo y obteniendo una notable reducción en la variabilidad de la estimación.

Este último razonamiento lleva directamente a considerar deseable que el ancho de banda (o también el *span* en el caso LOESS) del estimador de regresión no paramétrico dependa del punto donde se realice la estimación. De este modo sería posible adaptarlo a la forma que la nube de puntos presente en cada zona, salvando además la dificultad de encontrar entornos vacíos, al tomarlo mayor allí donde hay pocas observaciones.

Un procedimiento de regresión local con un parámetro ancho de banda local, parece por tanto la mejor solución al problema de regresión planteado. No obstante esto supone un considerable incremento en la dificultad de tratamiento computacional, el coste en tiempo será mucho mayor. Además el tratamiento teórico también resulta más complicado, ya que no estaríamos hablando de un parámetro constante sino de una función.

En los últimos años los anchos de banda locales han alcanzado un notable interés, desarrollándose interesantes trabajos como el de Fan y Gijbels (1995b) al que posteriormente aludiremos.

En general un procedimiento de selección de anchos de banda locales, involucra una medida de error local, como el error cuadrático medio, MSE, asegurando el buen comportamiento del estimador en cada punto.

Por el contrario, los procedimientos encaminados a la selección de anchos de banda globales o constantes para todos los puntos de estimación, ponen mayor interés en optimizar el comportamiento de la curva de regresión estimada en todo el intervalo de valores considerado; y por tanto se basan en medidas de error global, como el error cuadrático integrado medio, MISE.



### 3.2.3 Métodos *plug-in*

Bajo este término se engloban varios procedimientos de selección desarrollados, tanto dentro de la teoría de estimación no paramétrica de densidades, como en el contexto propio de la regresión no paramétrica.

Las primeras formulaciones de tipo *plug-in* fueron introducidas por Woodroffe (1970) en estimación de densidades. Sin embargo hasta mediados de los ochenta no se proponen algoritmos específicos y se promociona fuertemente. Hoy en día la metodología *plug-in* es ampliamente utilizada y surgen continuamente mejoras.

La idea que subyace en este tipo de técnicas, está en partir de la expresión óptima teórica del ancho de banda y proponer estimadores de lo desconocido, que se incrustarán<sup>1</sup> en dicha expresión teórica. Otra posibilidad, dentro de esta misma filosofía, está en considerar alguna expresión teórica desarrollada del error cuadrático medio, o del error cuadrático integrado medio; y sustituir lo desconocido por estimaciones, obteniendo así el ancho de banda óptimo como el minimizador de dicha expresión estimada. Si las estimaciones que se incrustan en la expresión presentan buenas propiedades de consistencia, dotan al error estimado resultante de tales propiedades con respecto al error teórico.

No obstante, aunque estos procedimientos comparten una misma filosofía, podemos hacer una clasificación de los mismos en dos categorías:

- ***Plug-in* asintótico.** Parte de expresiones asintóticas teóricas del error (MSE o MISE) y estima lo desconocido; si bien también se pueden realizar las estimaciones directamente sobre la expresión del ancho de banda teórico óptimo que minimiza tales expresiones asintóticas. Uno de los inconvenientes que presenta este tipo de procedimientos es que para poder utilizar tales expresiones asintóticas es necesario realizar hipótesis bastante restrictivas sobre la función de regresión desconocida, lo cual limita su campo de aplicación en la práctica. Además, el hecho de ser expresiones válidas para tamaños muestrales grandes hace que en la práctica los procedimientos fallen notablemente para tamaños muestrales pequeños. Corresponden a este tipo, procedimientos como el desarrollado por Ruppert y otros (1995).
- ***Plug-in* no asintótico.** Debido al problema que presenta el *plug-in* asintótico en muestras pequeñas, algunos autores toman como punto de partida expresiones desarrolladas de los errores teóricos de naturaleza no asintótica, incrustando en ellas las estimaciones propuestas para los términos desconocidos. Dentro de este grupo de procedimientos destacamos el desarrollado por Fan y Gijbels (1995b).

Además, dependiendo de si utilizan iteraciones para llegar a la solución óptima, se puede distinguir entre:

<sup>1</sup>De ahí el nombre de métodos *plug-in*.



- \* *Plug-in iterativo.* Corresponde a procedimientos *plug-in*, asintóticos o no, que añaden una sucesión de iteraciones, con el fin de conseguir una progresiva mejora del ancho de banda. Como inconveniente fundamental destaca el consecuente aumento del coste computacional. Dentro de este grupo está la técnica de selección propuesta por Gasser y otros (1994).
- \* *Plug-in no iterativo o directo.* No hacen uso de iteraciones para la obtención del selector del parámetro.

### 3.2.4 Métodos de selección automática

Al contrario que en la filosofía *plug-in* antes descrita, los denominados métodos de selección automática no parten de expresiones teóricas del error exactas o asintóticas, sino que se basan en la minimización de alguna función que aproxime, a partir de los datos, el error del estimador, de hecho, la medida de error considerada aquí es la que definimos en el capítulo 1 como error cuadrático promediado sobre las observaciones, que se definía como

$$\text{ASE}(\hat{m}_h(x)) = n^{-1} \sum_{i=1}^n (\hat{m}_h(X_i) - m(X_i))^2 w(X_i).$$

De esta forma, si se sustituyen los valores desconocidos,  $m(X_i)$ , por los datos observados,  $Y_i$ , se obtendría como medida aproximada del error la denominada **suma residual de cuadrados**, dada por (eliminando por comodidad en la notación la función de ponderaciones,  $w(\cdot)$ )

$$\text{RSS}(h) = n^{-1} \sum_{i=1}^n (Y_i - \hat{m}_h(X_i))^2.$$

No obstante esta medida no es insesgada del error ASE, por lo que se proponen funciones de la medida, RSS, que sí sean insesgadas. En general se consideran correcciones del tipo

$$\xi(n^{-1}h^{-1})\text{RSS}(h),$$

donde el factor,  $\xi(n^{-1}h^{-1})$ , puede ser aleatorio o no. Distintas elecciones de tal factor de corrección, ofrecen diferentes criterios o métodos de selección del parámetro ancho de banda, que reciben el nombre de **métodos de selección automática**. Algunas de las elecciones más habituales son:

1. Validación cruzada por mínimos cuadrados:

Teniendo en cuenta que el estimador polinomial local es lineal en las observaciones, éste se puede escribir como  $\hat{m}_h(X_i) = \sum_{j=1}^n A_{ij}(h)Y_j$ , para alguna matriz  $(A_{ij}(h))_{i,j}$ . De modo que se define como factor la siguiente cantidad

$$\xi(n^{-1}h^{-1}) = (1 - A_{ii}(h))^2.$$

2. Validación cruzada generalizada:

$$\xi(n^{-1}h^{-1}) = (1 - n^{-1}h^{-1}K(0))^{-2}.$$

3. Criterio de información de Akaike:

$$\xi(n^{-1}h^{-1}) = (1 - n^{-1}h^{-1}K(0))^{-2}.$$

4. Error de predicción finito:

$$\xi(n^{-1}h^{-1}) = \frac{1 + n^{-1}h^{-1}K(0)}{1 - n^{-1}h^{-1}K(0)}.$$

5. Selector de Shibata:

$$\xi(n^{-1}h^{-1}) = 1 + 2n^{-1}h^{-1}K(0).$$

6. T de Rice:

$$\xi(n^{-1}h^{-1}) = (1 - 2n^{-1}h^{-1}K(0))^{-1}.$$

Se puede demostrar que todos los selectores automáticos anteriores resultan asintóticamente equivalentes<sup>2</sup> y por consiguiente su comportamiento será aproximadamente el mismo, para tamaños muestrales grandes.

Este tipo de procedimientos presenta la ventaja de su facilidad de uso y de implementación computacional; además no hace grandes hipótesis sobre la función de regresión desconocida, como es el caso de las aproximaciones *plug-in*. De hecho algunos de ellos, como el de validación cruzada por mínimos cuadrados, resultan ser los métodos más aplicados en la práctica.

Sin embargo también presentan importantes inconvenientes ya que, en general dan lugar a estimadores con una fuerte variabilidad.

Dentro de este grupo de selectores, recientemente han surgido modificaciones que intentan solventar en alguna medida sus dificultades, aprovechando sus buenas cualidades. Entre ellos destacamos el propuesto por Hart y Yi (1998).

### 3.2.5 Procedimientos basados en *bootstrap*

La metodología *bootstrap* tiene como propósito ganar información acerca de la distribución de un estimador. En regresión no paramétrica la metodología *bootstrap* es utilizada fundamentalmente para dos tareas: una es la de elegir el parámetro de suavizamiento o ancho de banda; y otra es la de construir intervalos de confianza

<sup>2</sup>Hall, Härdle y Marron (1988).

para la curva de regresión (un trabajo a destacar aquí es el de Härdle y Marron (1990)).

En la elección del parámetro ancho de banda se considera por lo general una medida del error global o local, y se utiliza la distribución *bootstrap* para obtener un estimador de dicho error, de forma que el parámetro se elegirá como su minimizador sobre el conjunto de los posibles anchos de banda.

Siguiendo esta línea dentro de la literatura correspondiente a regresión no paramétrica se puede encontrar poco, es Hall (1990) quien hace una aproximación de tales técnicas para el caso más simple de estimador núcleo y suponiendo diseños de tipo fijo y equiespaciado. También Härdle y Bowman (1988) emplean la metodología *bootstrap* en la elección del ancho de banda, no obstante este trabajo sigue unas pautas distintas a las que acabamos de exponer, usando las estimaciones *bootstrap* dentro de una metodología *plug-in*.

No obstante, para los estimadores polinomiales locales en el contexto de diseño aleatorio y heterocedasticidad que venimos considerando, no se han desarrollado técnicas *bootstrap* para la selección del ancho de banda. En el capítulo 4 proponemos una extensión de la metodología *bootstrap* a tal contexto, y probamos su validez a través de un estudio teórico de tipo asintótico; y en cuanto al comportamiento en muestras finitas realizamos un estudio de simulación en el capítulo 5.

### 3.2.6 Métodos de selección automática frente a métodos *plug-in*

Loader (1999) divide los selectores del ancho de banda que han surgido en la literatura (tanto en el contexto propio de la regresión no paramétrica como en el de estimación no paramétrica de densidades) en dos grupos: Los que denomina *selectores clásicos*, que corresponden a los que en esta memoria denominamos métodos de selección automática, y los de tipo *plug-in*. Este autor realiza una detallada comparación, con matices diferentes de los habitualmente empleados, entre ambas formas genéricas de selección. En esta sección recogemos algunos de los argumentos que utiliza así como varias de sus conclusiones; de hecho nos servirán como iniciativa para los análisis que proponemos en capítulos posteriores.

Los autores que proponen selectores del ancho de banda tipo *plug-in*, tanto en el contexto de regresión, que venimos tratando, como en el de estimación de densidades, atacan fuertemente a los métodos clásicos, como el de validación cruzada, por ofrecer selectores que presentan una fuerte variabilidad y unas razones de convergencia lentas, que además son sobradamente aventajadas por diversas variantes de los selectores del tipo *plug-in*. Análisis de este tipo se pueden encontrar en Park y Marron (1990), Ruppert, Sheather y Jones (1995) y Marron (1996); y suelen basarse en el estudio de ejemplos con datos reales, en simulaciones y en desarrollos teóricos de tipo asintótico.

No obstante y según las argumentaciones dadas por Loader (1999), el buen comportamiento de los selectores *plug-in* se ha ponderado en exceso, pasando por alto algunos detalles que hacen dudar de que su calidad esté garantizada en muchos casos. En concreto, las estimaciones tipo *plug-in* están sujetas a especificaciones arbitrarias de las estimaciones piloto, y son propensas al sobresuavizamiento cuando la curva de regresión subyacente es difícil.

De este modo, si cuando se utilizan métodos de selección automática para la selección del parámetro de suavizado, a menudo se obtienen estimaciones con gran variabilidad en presencia de no muchos datos, las estimaciones de tipo *plug-in* en la misma situación suavizan en exceso, haciendo verdaderamente difícil recoger la estructura real de los datos. Es por esto que, en muchos casos, el problema de selección no está tanto en elegir selectores de tipo *plug-in* frente a selectores clásicos y sencillos como validación cruzada, sino más bien en que el número de datos que se está utilizando es insuficiente.

Junto con el argumento de la excesiva variabilidad de los selectores clásicos del parámetro ancho de banda, se utilizan habitualmente razonamientos basados en cálculos de tipo asintótico para desacreditarlos. Sin embargo para la correcta comprensión de estos razonamientos habría que hacer notar algunos elementos, ya que cualquier comparación de tipo asintótico que pretenda obtener resultados significativos, debe plantear el siguiente procedimiento:

1. Formular un conjunto de hipótesis.
2. Someter el selector planteado, bajo las hipótesis anteriores, a las siguientes cuestiones:
  - (a) ¿Está próximo el ancho de banda elegido al óptimo buscado? Esto es, estudiar la razón de convergencia del selector al óptimo.
  - (b) ¿Es el estimador de regresión que involucra el ancho de banda elegido un buen estimador de la verdadera función de regresión? Esto es, estudiar la razón de convergencia del estimador que emplea el parámetro seleccionado, a la verdadera función, empleando para ello alguna función de pérdidas.

Habitualmente los razonamientos empleados para garantizar el buen comportamiento de los selectores *plug-in* se basan tan sólo en la primera de las dos cuestiones antes formuladas. Sin embargo no debemos olvidar que el objetivo que se persigue es, fundamentalmente, aproximar de forma eficiente el estimador de regresión a la función que estima.

Si atendemos, por tanto, a esta cuestión descubrimos que el problema no está sólo en el selector usado sino que también lo está, y en gran medida, en el tipo de estimador de regresión no paramétrica considerado de partida. De este modo, partir

de estimadores de regresión polinomiales locales de grado dos o superior, permitirá obtener una mejor descripción de los datos cuando la estructura que subyace en ellos es difícil de capturar.

Además puede resultar interesante la siguiente observación: Está claro según la expresión del ancho de banda óptimo dada en (2.13), que un buen selector tipo *plug-in* deberá emplear una buena estimación de  $\int m''(x)^2 dx$ , y esto equivale a una estimación localmente cuadrática. De ahí que los métodos *plug-in* estén haciendo uso implícito de estimaciones localmente cuadráticas para estimar la curvatura de la función de regresión; si bien esta información no la usan de forma eficiente, sino tan sólo para estimar el sesgo del estimador de regresión.

Por tanto, en la práctica es posible mejorar el comportamiento de los selectores empleando un estimador de regresión más eficiente como es el estimador polinomial local de grado dos.

No obstante, el problema que surge cuando se dispone de pocos datos sigue latente aún cuando se utilicen estos ajustes cuadráticos locales, de hecho en tal caso será necesario un tamaño muestral grande.

De ahí que para tamaños muestrales pequeños los selectores tipo *plug-in* sean inadecuados, debiendo sustituirse por selectores automáticos clásicos que involucren estimadores polinomiales locales de grado cero o a lo sumo uno.

### 3.3 El selector de Fan y Gijbels (1995b)

Fan y Gijbels (1995b) desarrollaron algunos procedimientos de selección que involucran las nociones *plug-in*. Uno de los más interesantes, dado que ha obtenido en general buenos resultados, es el que ellos denominan *selector del ancho de banda refinado*, que se presenta en dos versiones, una local y otra global (según la clasificación que hacíamos al comienzo del capítulo).

El procedimiento que proponen implica dos etapas para la selección del ancho de banda final. En una primera etapa se considera un ancho de banda piloto (basado también en la metodología *plug-in*) el cual intervendrá en las estimaciones que, en la segunda etapa, conducirán a una elección más refinada del parámetro.

Fan y Gijbels lo describen en general para la estimación de una derivada de la función de regresión y suponiendo el contexto general de heterocedasticidad en el modelo. A continuación incluimos una versión resumida de su propuesta adaptada al contexto en que hemos situado nuestro trabajo, esto es, en la estimación de la función de regresión (en principio no nos interesa la estimación de las derivadas) para un modelo de regresión polinomial local.

Partiendo de las expresiones para el sesgo y la varianzas condicionales exactas en

un punto,  $x$ :

$$\begin{aligned} B_h(x) &= e_1^T (X_x^T W_{x,h} X_x)^{-1} X_x^T W_{x,h} (m - X_x \beta^T) \\ &= (X_x^T W_{x,h} X_x)^{-1} X_x^T W_{x,h} r \\ V_h(x) &= \sigma^2(x) e_1^T (X_x^T W_{x,h} X_x)^{-1} (X_x^T W_{x,h}^2 X_x) (X_x^T W_{x,h} X_x)^{-1} e_1, \end{aligned}$$

consideramos la expresión resultante para el error cuadrático medio condicional,

$$\text{MSE}(\hat{m}_h(x)) = \{B_h(x)\}^2 + V_h(x). \quad (3.1)$$

Obsérvese que los únicos términos desconocidos que aparecen son  $r$  y  $\sigma^2(x)$ . Fan y Gijbels proponen el siguiente estimador de  $\sigma^2(x)$ , usando un ancho de banda piloto,  $g$ :

$$\hat{\sigma}_g^2(x) = \frac{\sum_{i=1}^n (Y_i - \hat{m}_g(X_i))^2 K_g(X_i - x)}{\text{traza}(W_{x,g}^* - W_{x,g}^* X_x^* (X_x^* W_{x,g}^{*2} X_x^*)^{-1} X_x^{*T} W_{x,g}^{*2})} \quad (3.2)$$

donde  $X_x^*$  y  $W_{x,g}^*$  se definen de forma similar a las ya introducidas  $X_x$  y  $W_{x,h}$ , pero asociadas a un ajuste polinomial local de grado  $p + a$  (donde  $a$  será un natural a elegir, ellos proponen  $a = 2$ ).

Y para estimar  $r$ ,  $r = m - X_x \beta^T$ , sugieren realizar primero un desarrollo de Taylor para obtener

$$m - X_x \beta^T \approx \begin{pmatrix} \beta_{p+1}(X_1 - x)^{p+1} + \dots + \beta_{p+a}(X_1 - x)^{p+a} \\ \vdots \\ \beta_{p+1}(X_n - x)^{p+1} + \dots + \beta_{p+a}(X_n - x)^{p+a} \end{pmatrix}, \quad (3.3)$$

y entonces estimar los coeficientes,  $\beta_{p+1}, \dots, \beta_{p+a}$ , mediante un ajuste polinomial local de grado  $p + a$  con el mismo ancho de banda piloto,  $g$ , usado en la estimación de  $\sigma^2(x)$ .

Insertando las estimaciones (3.2) y (3.3) en (3.1), se obtendría la estimación tipo *plug-in* propuesta para el error cuadrático medio del estimador local. Si se pretenden obtener versiones globales habría que integrar en todo el intervalo de estimación considerado.

En cuanto al ancho de banda piloto,  $g$ , Fan y Gijbels proponen para su elección, el denominado *criterio de la suma de cuadrados residual* que describimos a continuación.

### 3.3.1 Criterio de la suma de cuadrados residual

El *criterio de cuadrados residuales* (RSC) viene dado por la expresión

$$\text{RSC}(x; h) = \hat{\sigma}_h^2(x) \{1 + (p + 1)V_h\}, \quad (3.4)$$

donde  $\hat{\sigma}_h^2(x)$  es una estimación de  $\sigma^2(x)$  similar a la anterior,  $\hat{\sigma}_g^2(x)$ , pero para un ajuste lineal local con ancho de banda  $h$ ; y  $V_h$  es el primer elemento de la diagonal de  $(X_x^T W_{x,h} X_x)^{-1} (X_x^T W_{x,h}^2 X_x) (X_x^T W_{x,h} X_x)^{-1}$ .

Si se busca un ancho de banda global o fijo, se considerará la versión integrada (en el intervalo de estimación considerado  $(a, b)$ ) de RSC, esto es

$$\text{IRSC}(h) = \int_a^b \text{RSC}(x; h) dx, \quad (3.5)$$

y el ancho de banda global que hace mínima tal medida,  $h_{\text{IRSC}}$ , será el elegido según tal criterio.

Cuando se pretende conseguir un ancho de banda local, Fan y Gijbels proponen dividir el intervalo  $(a, b)$  en subintervalos,  $I_s$ , de modo que los ajustes polinomiales se hacen en cada uno de estos subintervalos. De este modo, se elegirá un ancho de banda,  $h_{\text{IRSC}}^s$ , para cada intervalo,  $I_s$ , como el minimizador de

$$\text{IRSC}(h_s) = \int_{I_s} \text{RSC}(x; h_s) dx. \quad (3.6)$$

### 3.3.2 Pasos para elegir un ancho de banda global

Resumimos ahora los pasos necesarios para obtener el selector de Fan y Gijbels del ancho de banda.

1. Obtener el ancho de banda,  $h_{\text{IRSC}}$ , como minimizador del criterio,  $\text{IRSC}(h)$ , dado en (3.5); y calcular el ancho de banda piloto como  $g = \text{adj}(K) h_{\text{IRSC}}$ . Aquí  $\text{adj}(K)$  es una constante que depende de la función núcleo considerada, su expresión explícita puede encontrarse en Fan y Gijbels (1995b), además de sus valores calculados para algunas de las situaciones más habituales.
2. Calcular las estimaciones,  $\hat{\sigma}_g^2(x)$ ,  $\hat{\beta}_{p+1}(x)$ ,  $\dots$ ,  $\hat{\beta}_{p+a}(x)$ , mediante un ajuste polinomial local de grado  $p + a$ , con ancho de banda,  $g$ .
3. Estimar los términos de sesgo y varianza sustituyendo las estimaciones calculadas en el paso anterior, y obtener la estimación del error cuadrático medio que denotamos  $\widehat{\text{MSE}}(x; h)$ .
4. Seleccionar el ancho de banda final,  $h_{\text{FG}}$ , como el minimizador de

$$\widehat{\text{MISE}}(h) = \int_a^b \widehat{\text{MSE}}(x; h) dx.$$



### 3.3.3 Pasos para elegir un ancho de banda local

En el caso en que el ancho de banda buscado sea variable en cada punto de estimación, los pasos serán similares a los descritos para el caso global con leves diferencias; como se expone a continuación:

1. Dividir el intervalo de estimación,  $(a, b)$ , en subintervalos,  $I_s$ , y obtener los anchos de banda,  $h_{IRSC}^s$ , como minimizadores de criterios  $IRSC(h_s)$  dados en (3.6). Entonces, suavizar la función ancho de banda resultante (que será una función de salto) promediando localmente, usando el mismo parámetro de suavizamiento que para la partición inicial, esto es, la longitud de  $I_s$ . Obtener la función ancho de banda suavizada piloto resultante, multiplicando por la constante  $adj(K)$ .
2. Calcular las estimaciones,  $\hat{\sigma}_g^2(x)$ ,  $\hat{\beta}_{p+1}(x), \dots, \hat{\beta}_{p+a}(x)$ , mediante un ajuste polinomial local de grado  $p+a$ , con el ancho de banda piloto obtenido en el paso anterior.
3. Estimar los términos de sesgo y varianza sustituyendo las estimaciones calculadas en el paso anterior, y obtener la estimación del error cuadrático medio,  $\widehat{MSE}(x; h)$ .
4. Para cada intervalo,  $I_s$ , seleccionar el ancho de banda que minimiza

$$\int_{I_s} \widehat{MSE}(x; h) dx,$$

y suavizar la función de salto resultante, usando de nuevo la longitud de los intervalos  $I_s$ .

## 3.4 Validación cruzada por mínimos cuadrados

Sea de nuevo  $\hat{m}_h(x)$ , el estimador no paramétrico considerado para la función de regresión. Para cada  $i$  ( $i = 1, \dots, n$ ), denotaremos por  $\hat{m}_{h,-i}(x)$ , 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,  $Y_i - \hat{m}_{h,-i}(X_i)$ .

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

$$CV(h) = n^{-1} \sum_{i=1}^n \{Y_i - \hat{m}_{h,-i}(X_i)\}^2, \quad (3.7)$$

lo cual constituye una medida global de la efectividad de  $\hat{m}_h(\cdot)$ , 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

$$\hat{h}_{CV} = \underset{h}{\operatorname{argmin}} \operatorname{CV}(h).$$

### 3.4.1 Un inconveniente: Correlación negativa

En Härdle, Hall y Marron (1988-*Rejoinder*) se detecta y comenta un inconveniente observado en los métodos de selección automática habitualmente utilizados, como éste de validación cruzada.

Cuando se generan varias muestras aleatorias a partir de un modelo de regresión no paramétrico, en algunas se observarán correlaciones en serie positivas, es decir, los puntos de la nube de puntos tenderán a estar por encima o por debajo de la verdadera curva de regresión, formando agrupaciones. En otras se detectarán correlaciones negativas, es decir, los puntos tenderán a alternarse alrededor de la curva de regresión. En la primera situación, es necesario un ancho de banda grande para suavizar, eliminando el efecto producido por las agrupaciones; de modo que el ancho de banda óptimo teórico será grande. No obstante, los métodos de selección automática en tal caso entenderán que hay alguna estructura fina presente y por tanto su selector será relativamente pequeño. Por otro lado, cuando las correlaciones son negativas, la curva puede ser recorrida perfectamente con un ancho de banda pequeño, por lo que el ancho de banda óptimo será pequeño. Sin embargo un método de selección automática no reconocerá ninguna estructura por lo que tenderá a elegir un ancho de banda grande.

Todo esto ha suscitado muchas críticas contra el criterio de validación cruzada, dado que generalmente produce una alta variabilidad. De este modo, son muchas las propuestas de mejora que han aparecido en los últimos años, una de ellas es el criterio denominado *one-sided cross-validation*, desarrollado por Hart y Yi (1998).

### 3.4.2 Versión local del criterio de validación cruzada

View (1991) propone, en el contexto de la estimación núcleo, una versión del criterio de validación cruzada que permite obtener el ancho de banda como una función dependiente del punto de estimación. La extensión al caso de la regresión polinomial local resulta sencilla por lo que la hacemos a continuación.

El método de validación cruzada local para elegir el ancho de banda,  $h(x)$ , en cada punto,  $x$ , consiste en minimizar un criterio del tipo

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

donde  $w_{n,x}$ , es una  $\delta$ -sucesión de funciones concentradas alrededor de  $x$ . En la práctica se suelen tomar distribuciones normales de media  $x$ , y desviación típica  $S_w$ :

$$w_{n,x}(u) = (2\pi)^{-1/2} S_w^{-1} \exp\{-(u-x)^2/2S_w^2\}.$$

Y la función ancho de banda se obtendrá como

$$h_{CV}(x) = \underset{h \in A_n}{\operatorname{argmin}} \quad CV(h; x)$$

calculando dicho mínimo sobre algún conjunto de anchos de banda,  $A_n$ .

Según las observaciones de View, en la elección de las funciones,  $w_{n,x}$ , el principal problema está en la estimación de la varianza,  $S_w^2$ . Las justificaciones del autor apuntan a que sería razonable intentar cuantificar la cantidad de variabilidad en el ancho de banda local y a tenor de lo obtenido, elegir  $S_w^2$ . Sin embargo a partir de varias simulaciones y ejemplos con datos reales<sup>3</sup>, se observa que este problema es mucho menos importante que el de la propia selección del ancho de banda.

### 3.5 El selector de Ruppert (1997)

Tomando como base el selector propuesto por Fan y Gijbels (1995b), Ruppert construye un procedimiento para la selección del ancho de banda de tipo *plug-in* no asintótico directo. Las argumentaciones de Ruppert muestran que el procedimiento de Fan y Gijbels puede resultar demasiado costoso y complicado si se plantea una versión multivariante del problema de regresión.

El procedimiento de selección propuesto por Ruppert tiene las siguientes características:

1. No utiliza expresiones asintóticas para el sesgo ni para la varianza, sino que el sesgo lo estima empíricamente utilizando una metodología similar al *bootstrap* (pero sin remuestreo), y para estimar la varianza utiliza la misma expresión que Fan y Gijbels (1995b).
2. No emplea en los cálculos ajustes polinomiales de grado superior al elegido para el estimador de la función de regresión.
3. El planteamiento lo hace para cualquier dimensión, si bien nosotros lo trataremos tan sólo en el caso univariante.

El procedimiento es el siguiente:

Sea una red de puntos de estimación,  $G_x = \{x_l : l \in L\}$ , para  $L$ , un conjunto de índices. Para cada  $x_l \in G_x$ , se estima  $\operatorname{MSE}(\hat{m}_h(x_l))$ , el error cuadrático medio del estimador en el punto  $x_l$ , tratando por separado el sesgo y la varianza.

<sup>3</sup>Que se pueden encontrar en View (1991)

1. **Estimación del sesgo:** Fijamos un punto,  $x_l$ , y sea un ancho de banda,  $h_0$ , de modo que estamos interesados en calcular el sesgo de  $\widehat{m}_{h_0}(x_l)$ . Sea  $J_b$ ,  $J_b > 1$ , un entero y sean  $h_0^1, \dots, h_0^{J_b}$  un entorno de  $h_0$ . Entonces se calcula cada  $\widehat{m}_{h_0^j}(x_l)$ ,  $j = 1, \dots, J_b$ . Después, para algún  $t$ ,  $t \geq 1$ , se ajusta la curva

$$\widehat{m}_h(x_l) \approx b_0(x_l) + b_{p+1}(x_l)h^{p+1} + \dots + b_{p+t}(x_l)h^{p+t}$$

a los datos  $\{(h_0^j, \widehat{m}_{h_0^j}(x_l)) : j = 1, \dots, J_b\}$ , por mínimos cuadrados. De este modo, el sesgo de  $\widehat{m}_{h_0}(x)$  se estimaría por

$$\widehat{B}_{h_0}(x_l) = \widehat{b}_{p+1}(x_l)h_0^{p+1} + \dots + \widehat{b}_{p+t}(x_l)h_0^{p+t},$$

donde  $\widehat{b}_i$  corresponden a las estimaciones obtenidas para los coeficientes  $b_i$  ( $i = 0, p+1, \dots, p+t$ ).

2. **Estimación de la varianza:** La estimación propuesta sigue la misma línea que la utilizada por Fan y Gijbels (1995b), que recogemos en la sección 3.3 de este capítulo; la única diferencia radica en la estimación propuesta para la varianza de los residuos,  $\sigma^2(\cdot)$ . De este modo, la estimación de la varianza se calcula como sigue:

$$\widehat{V}_{h_0}(x_l) = \widehat{\sigma}^2(x_l)e_1^T (X_x^T W_{x,h} X_x)^{-1} (X_x^T W_{x,h}^2 X_x) (X_x^T W_{x,h} X_x)^{-1} e_1$$

donde  $\widehat{\sigma}^2(\cdot)$  representa un estimador de la varianza residual. Ruppert propone para su cálculo, un estimador similar al propuesto por Ruppert, Wand y Hössjer (1995).

Por tanto, una estimación del error cuadrático medio (MSE) estaría dada por

$$\widehat{\text{MSE}}(x_l; h_0) = \left\{ \widehat{B}_{h_0}(x_l) \right\}^2 + \widehat{V}_{h_0}(x_l). \quad (3.8)$$

Para la elección del ancho de banda se considera una versión suavizada de dicha estimación que denotamos por  $\widehat{\text{SMSE}}(x_l; h_0, N_l)$ , y que se define como la media de (3.8) sobre los puntos de un entorno,  $N_l$ , de  $x_l$ . El ancho de banda local propuesto,  $\widehat{h}(x_l; N_l)$ , se define como el primer mínimo local de  $\widehat{\text{SMSE}}(x; h)$  sobre una red de anchos de banda.

Las consideraciones hechas por Ruppert (1997) llevan a considerar una versión suavizada del selector, usando para ello una función peso triangular, en los mismos entornos,  $N_l$ , utilizados para suavizar el estimador del error cuadrático medio.

## Capítulo 4

# Elección *bootstrap* del ancho de banda

### 4.1 La metodología *bootstrap*

Una de las tareas más importantes en Estadística es medir la precisión de los estimadores propuestos. Resulta indudable que la mejor forma de llevar a cabo este propósito consiste en utilizar la distribución muestral del estimador en cuestión. El problema está en que habitualmente tal distribución resulta muy difícil de obtener y, en bastantes situaciones, tan sólo es accesible una aproximación asintótica. De cualquier forma, en ambos casos es habitual que tanto la distribución muestral verdadera, como su aproximación asintótica, dependan de alguna característica o parámetro desconocidos, de forma que es necesario encontrar alguna función de distribución conocida que sea próxima, en algún sentido, a la distribución muestral del estimador. Una posibilidad la constituye el método *bootstrap*.

La metodología *bootstrap* fue introducida por primera vez por Efron (1979) y desde entonces, se han escrito una gran número de trabajos concernientes a ella. Se han intentado aplicar las ideas originales de Efron a una multitud de campos tales como la regresión, datos censurados, estimación no paramétrica de curvas, etc.

En un contexto general, las ideas *bootstrap* propuestas originalmente por Efron, se formulan en los términos siguientes:

Dada una muestra aleatoria,  $\mathbf{X} = (X_1, \dots, X_n)$ , de observaciones independientes de una población con función de distribución,  $\mathbf{F}$ , y una variable aleatoria,  $\mathbf{R} = R(\mathbf{X}, \mathbf{F})$ , de interés, se realizan los siguientes pasos:

1. Construir la función de distribución empírica,  $F_n$ .
2. Diseñar réplicas *bootstrap*,  $\mathbf{X}^* = (X_1^*, \dots, X_n^*)$ , como muestras aleatorias a partir de la distribución empírica,  $F_n$ .

3. Construir la versión *bootstrap* del estadístico,  $\mathbf{R}^* = R(\mathbf{X}^*, F_n)$ , y aproximar la distribución muestral de  $\mathbf{R}$ , mediante la distribución *bootstrap*, o distribución en el remuestreo de  $\mathbf{R}^*$ .

La metodología se puede extender fácilmente a otras situaciones donde se disponga de alguna información sobre la distribución. Tal es el caso en el que la distribución desconocida pertenece a alguna clase paramétrica de funciones,  $\{F_s/s \in S\}$ . Entonces, el papel de la distribución empírica,  $F_n$ , lo desempeña  $F_{\hat{s}}$ ; siendo  $\hat{s}$  un estimador de  $s$ . Esta variante se conoce como *bootstrap paramétrico*.

Dentro del ámbito no paramétrico está el caso en que la función de distribución subyacente es continua. Aquí una elección inteligente consiste en usar el denominado *bootstrap suave*<sup>1</sup>, que en lugar de elegir la distribución empírica,  $F_n$ , propone alguna versión suavizada de la misma, por ejemplo mediante una función núcleo,  $K$ , de modo que se considera una distribución,  $\hat{F} = F_n * K_h$ .

#### 4.1.1 Validez y aplicabilidad de *bootstrap*

Sea cual sea la modalidad de *bootstrap* que se pretenda utilizar, su aplicación en cada caso debe pasar por el estudio detenido de dos aspectos esenciales:

1. ¿Es asintóticamente correcta la aproximación *bootstrap*?
2. ¿Cómo se calculará la distribución *bootstrap* (distribución en el remuestreo) del estadístico?

Ambos aspectos son cruciales en la utilización de la metodología *bootstrap* ya que suponen, por una parte, la validez del método, y por otra, su aplicabilidad.

La solución a la primera cuestión suele darse probando que alguna distancia entre la distribución *bootstrap* de  $\mathbf{R}^*$  y la distribución muestral de  $\mathbf{R}$  tiende a cero a medida que  $n$  tiende a infinito. En cuanto al segundo punto, ocasionalmente es posible obtener teóricamente una expresión exacta de la distribución *bootstrap*, aunque lo más habitual suele ser que dicha distribución sea desconocida, de modo que una expresión exacta resulte inaccesible. En tales casos se suele optar por una aproximación de Monte Carlo. La idea subyacente consiste en diseñar un gran número de remuestras *bootstrap*,  $X^{*(1)}, X^{*(2)}, \dots, X^{*(B)}$ , y utilizar las  $B$  correspondientes realizaciones del estadístico,  $\mathbf{R}^{*(j)} = \mathbf{R}^*(X^{*(j)}, F_n)$ , para aproximar la distribución en el remuestreo de  $\mathbf{R}^*$ . Obviamente el inconveniente que plantea, a simple vista, esta aproximación, es que el incremento en  $B$  se paga en términos de un enorme coste computacional.

<sup>1</sup>Esta idea ha sido ampliamente desarrollada en el caso de estimación de densidades por Cao (1993).

## 4.2 *Bootstrap* en estimación de densidades

La metodología *bootstrap* ha sido ampliamente utilizada, en el contexto de la estimación de densidades, para la construcción de selectores del ancho de banda.

Habitualmente se ha planteado para la búsqueda de anchos de banda globales, buscando estimaciones del error cuadrático medio integrado,

$$\text{MISE}(\hat{f}_h) = E \left[ \int (\hat{f}_h(t) - f(t))^2 \right],$$

asociado al estimador no paramétrico,  $\hat{f}_h$ , de una densidad,  $f$ . La estimación *bootstrap* propuesta es de la forma

$$\text{MISE}^*(h) = E_* \left[ \int (\hat{f}_h^*(t) - \hat{f}_g(t))^2 \right], \quad (4.1)$$

donde  $E_*$  denota la esperanza sobre la muestra *bootstrap*,  $X_1^*, \dots, X_n^*$ ;  $g$ , es un ancho de banda piloto, y  $\hat{f}_h^*(t)$ , es el estimador no paramétrico de la densidad basado en la muestra *bootstrap*. El ancho de banda *bootstrap* se define como el minimizador del criterio  $\text{MISE}^*(h)$ .

Dependiendo de cómo se tome el ancho de banda piloto y de cómo se genere la muestra *bootstrap*, surgen distintos selectores *bootstrap*. En Cao, Cuevas y González-Manteiga (1994) se puede encontrar una revisión de tales procedimientos. En concreto se distingue entre:

1. *Bootstrap* suavizado sin ancho de banda piloto (Taylor, 1989). Las muestras *bootstrap* son diseñadas desde la densidad estimada  $\hat{f}_h(t)$  y el estimador *bootstrap* (4.1) se calcula tomando  $g = h$ .
2. *Bootstrap* suavizado con ancho de banda piloto (Faraway y Jhun, 1990). Las muestras *bootstrap* se diseñan desde  $\hat{f}_g(t)$ , y  $g$  se elige mediante validación cruzada por mínimos cuadrados. En este caso no es posible encontrar la expresión exacta de la estimación propuesta por lo que se aproxima usando varias muestras *bootstrap*. Dentro de esta misma línea, Cao, R. (1993) propone la elección de  $g$  buscando la proximidad entre las expresiones asintóticas de  $\text{MISE}(h)$  y  $\text{MISE}^*(h)$ ; en este caso sí se puede obtener la expresión exacta del estimador *bootstrap*.
3. *Bootstrap* no suavizado (Hall, 1990). En este caso las muestras *bootstrap* se diseñan desde la distribución empírica, y se consideran de tamaño  $m$ ,  $m < n$ , (razonamientos que aparecen en Hall, P. (1990) llevan a elegir  $m = n^{1/2}$ ) de modo que el ancho de banda *bootstrap* es de la forma  $h = g(m/n)^{1/5}$  siendo  $g$  el minimizador de  $\text{MISE}^*(\cdot)$ .

### 4.3 *Bootstrap* en regresión no paramétrica

#### 4.3.1 Remuestreo sobre pares y remuestreo sobre residuos

En el planteamiento habitual de la regresión, tanto desde un punto de vista clásico como desde una perspectiva no paramétrica, surgen dos posibilidades para el remuestreo: considerar un remuestreo sobre los pares de observaciones, o bien sobre los residuos de la regresión. Esto constituye un primer indicador de que la dificultad que conlleva el *bootstrap* en el contexto de regresión, es mayor que en la estimación de densidades, y por tanto las extensiones a la regresión de lo estudiado en estimación de densidades no son inmediatas.

Aplicar la filosofía *bootstrap* sobre los pares<sup>2</sup>,  $(X_i, Y_i)$ , supone diseñar muestras *bootstrap* del tipo

$$\{(X_{i_1}, Y_{i_1}), (X_{i_2}, Y_{i_2}), \dots, (X_{i_n}, Y_{i_n})\},$$

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

Si, por el contrario, se realiza sobre los residuos,  $\epsilon_i$ , el planteamiento es sustancialmente diferente. En general el procedimiento *bootstrap* consistiría en sustituir la función de distribución desconocida,  $F$ , de dichos residuos,  $\epsilon_i$ , por la empírica,  $F_n$ . Sin embargo, en un modelo de regresión, tales residuos son desconocidos, de modo que se plantea obtener una estimación inicial,  $\hat{m}$ , de la función de regresión desconocida y entonces estimar cada  $\epsilon_i$  de la forma

$$\hat{\epsilon}_i = Y_i - \hat{m}(X_i). \quad (4.2)$$

A continuación, y dado que tales estimaciones no tienen por qué tener media 0, se calculan las versiones centradas

$$\tilde{\epsilon}_i = \hat{\epsilon}_i - n^{-1} \sum_{i=1}^n \hat{\epsilon}_i.$$

Finalmente se obtendrían muestras *bootstrap* de la forma:

$$\{(X_1, Y_1^*), (X_2, Y_2^*), \dots, (X_n, Y_n^*)\},$$

donde<sup>3</sup>  $Y_i^* = \hat{m}(X_i) + \tilde{\epsilon}_i$ .

<sup>2</sup>Obsérvese que en este caso sólo es posible suponer que el problema de regresión se formula bajo un diseño aleatorio.

<sup>3</sup>En dicho caso es posible considerar tanto un diseño aleatorio como un diseño fijo.



Efron y Tibshirani (1993) razonan que el *bootstrap* sobre los pares supone una aproximación al *bootstrap* sobre los residuos, que mejora a medida que el tamaño muestral,  $n$ , crece.

### 4.3.2 Antecedentes: Técnicas *bootstrap* desarrolladas

En el contexto de la regresión no paramétrica la metodología *bootstrap* no ha sido muy utilizada como método de selección del parámetro ancho de banda. Trabajos en esta línea son los de Hall (1990), y Härdle y Bowman (1988).

Hall (1990) considera un *bootstrap* clásico no suavizado, a partir de los residuos, para estimadores núcleo sencillos y en un contexto de diseño fijo. Härdle y Bowman (1988) consideran, también para estimadores núcleo sobre un diseño fijo, un *bootstrap* clásico sobre los residuos, si bien la metodología seguida para estimar el error cuadrático medio no es propiamente la usual en los procedimientos *bootstrap*, sino que mezcla razonamientos de tipo asintótico.

No obstante la mayoría de las aplicaciones del *bootstrap* que se encuentran en el contexto de la regresión no paramétrica, están dirigidas a la obtención de bandas de confianza. Entre estos trabajos destacamos el de Härdle y Marron (1990), que consideran un *bootstrap* sobre los residuos del tipo *wild bootstrap* para el estimador tipo núcleo de Nadaraya-Watson en un contexto de diseño aleatorio; y el de Hall (1992) que considera un *bootstrap* sobre los residuos de tipo clásico no suavizado, para el mismo estimador núcleo, pero bajo un diseño fijo.

En las siguientes secciones nos proponemos extender algunos de estos procedimientos *bootstrap* al caso que constituye el punto de partida de esta memoria, esto es, el de los estimadores polinomiales locales en un diseño aleatorio para el caso de heterocedasticidad en el modelo.

No obstante, en todas las extensiones, nuestro interés estará en obtener estimadores *bootstrap* del error cuadrático medio, como posteriores selectores del parámetro ancho de banda.

La sección 4.4 define un selector del ancho de banda *bootstrap* basado en un remuestreo sobre los pares; la sección 4.5 considera un remuestreo sobre los residuos del tipo clásico no suavizado; y, finalmente, la sección 4.6 realiza sobre los residuos un *bootstrap* del tipo *wild bootstrap*.

## 4.4 *Bootstrap* sobre los pares

### 4.4.1 Sobre la validez del *bootstrap* clásico

Entre los distintos tipos de *bootstrap* que se han desarrollado en la literatura cuando se pretende llevar a cabo un remuestreo sobre los pares, en principio podríamos distinguir dos casos:

- *Bootstrap* clásico no suavizado, que selecciona muestras *bootstrap* con reemplazamiento a partir de la muestra de pares original.
- *Bootstrap* suavizado, donde la selección de las muestras *bootstrap* se hace utilizando una estimación de la densidad asociada a la población  $(X, Y)$ .

Comenzamos indicando la metodología a seguir para un *bootstrap* suavizado. Sea  $f(x, y)$  la densidad de la población,  $(X, Y)$ , asociada al modelo de regresión no paramétrico, de forma que cada par de observaciones de partida,  $(X_i, Y_i)$ , se distribuye según dicha densidad. Si nos planteamos generar muestras *bootstrap* una posibilidad sería recurrir a la distribución teórica de la población, no obstante, dado que será desconocida, podemos considerar un estimador de la misma, por ejemplo de tipo núcleo, que en este caso bidimensional sería:

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

donde  $K_h(u) = h^{-2}K(u/h)$ , para alguna función núcleo,  $K$ , definida sobre  $\mathbb{R}^2$ , y un ancho de banda  $h$ ,  $h > 0$ .

No obstante, este planteamiento conlleva la dificultad de elegir un ancho de banda para dicho estimador núcleo de la densidad poblacional, junto con la dificultad del tratamiento teórico al manejar densidades bivariantes.

Una aproximación más sencilla supondría utilizar un *bootstrap* de tipo clásico no suavizado, diseñando las muestras *bootstrap* con reemplazamiento a partir de la muestra original. No obstante hay algo importante que decir sobre la validez de esta metodología según los razonamientos hechos en el trabajo de Hall (1990).

Dadas las observaciones de partida,  $(X_1, Y_1), \dots, (X_n, Y_n)$ , diseñamos las muestras *bootstrap* del tipo  $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$  con reemplazamiento. Si definimos el estimador *bootstrap*,  $\hat{m}_h^*(x)$ , como el estimador polinomial local basado en los datos *bootstrap* generados, entonces el estimador *bootstrap* estándar del error cuadrático medio se construye de la forma siguiente:

$$\text{MSE}^*(x; h) = E_* [\hat{m}_h^*(x) - \hat{m}_h(x)], \quad (4.3)$$

donde es posible una descomposición en suma de un término de sesgo, al cuadrado,

$$B^*(x; h) = E_* [\hat{m}_h^*(x)] - \hat{m}_h(x)$$

más un término de varianza,

$$V^*(x; h) = \text{Var}_* (\hat{m}_h^*(x)).$$

Sin embargo, si calculamos  $E_* [\hat{m}_h^*(x)]$  considerando la distribución en el remuestreo, dada la linealidad del estimador polinomial local, es inmediato ver que el término de sesgo *bootstrap*,  $B^*(x; h)$ , se anula, con lo cual no sería una estimación válida para el sesgo del estimador y por tanto el estimador *bootstrap* (4.3) no sería admisible como estimador del error cuadrático medio.

Como solución a este problema se podría modificar la metodología *bootstrap* que acabamos de definir siguiendo cualquiera de las dos vías siguientes (según sugerencias de Hall, 1990):

- Considerar un estimador *bootstrap* modificado para el error cuadrático medio de la forma:

$$\text{MSE}^*(x; h) = E_* [\hat{m}_h^*(x) - \hat{m}_g(x)],$$

donde se incluye un parámetro ancho de banda piloto,  $g$ .

- Diseñar muestras *bootstrap* de tamaños,  $m$ , distintos al de la muestra de datos de partida.

En la siguiente sección se reformula el *bootstrap* clásico sobre los pares en nuestro contexto de regresión polinomial local. Mezclando las dos modificaciones que acabamos de citar, se obtiene una estimación válida para el error cuadrático medio.

#### 4.4.2 Estimación *bootstrap* del error cuadrático medio

Consideremos el conjunto de datos de partida,

$$H_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\},$$

que bajo un diseño aleatorio, supone una muestra aleatoria de una población,  $(X, Y)$ . El procedimiento para la construcción del estimador *bootstrap* del error cuadrático medio se describe en los siguientes pasos:

1. Se elige un ancho de banda piloto,  $g$ , y se construye el estimador polinomial local,  $\hat{m}_g$ .
2. Se diseñan muestras *bootstrap*:

$$H_m^* = \{(X_1^*, Y_1), (X_2^*, Y_2), \dots, (X_m^*, Y_m)\}$$

de tamaño  $m$ ,  $m \leq n$ , a partir de la muestra original con reemplazamiento.

3. Para cada  $h$ ,  $h > 0$ , se define el estimador *bootstrap*

$$\hat{m}_h^*(x) = \sum_{i=1}^m w_i^*(x; h) Y_i^*,$$

donde  $w_i^*(x; h)$ , son los pesos basados en la muestra *bootstrap* considerada.

4. Se define la estimación *bootstrap* del error cuadrático medio como:

$$\text{MSE}^*(x; h) = E_* [\{\widehat{m}_h^*(x) - \widehat{m}_g(x)\}^2],$$

donde  $E_*$ , denota la esperanza sobre las muestras *bootstrap*,  $H_m^*$ .

5. El selector *bootstrap* propuesto para el ancho de banda, estará dado por

$$h_{\text{boot}}^*(x) = \underset{h}{\text{argmin}} \text{MSE}^*(x; h);$$

A continuación obtendremos la expresión exacta del estimador *bootstrap* que acabamos de definir, calculando la esperanza sobre la distribución en el remuestreo.

#### 4.4.3 Expresión exacta del estimador *bootstrap*

Como es habitual, el error cuadrático medio admite una descomposición en suma del sesgo al cuadrado más un término de varianza, esto es,

$$\text{MSE}^*(x; h) = \{B_{h,g}^*(x)\}^2 + V_{h,g}^*(x),$$

donde el término de sesgo en este caso está dado por

$$B_{h,g}^*(x) = E_* [\widehat{m}_h^*(x)] - \widehat{m}_g(x) \quad (4.4)$$

y la varianza por

$$V_{h,g}^*(x) = \text{Var}_*(\widehat{m}_h^*(x)). \quad (4.5)$$

Comenzamos calculando el término de sesgo y en concreto probaremos que

$$B_{h,g}^*(x) = \frac{m}{n} \widehat{m}_h(x) - \widehat{m}_g(x). \quad (4.6)$$

Partimos de:

$$E_* [\widehat{m}_h^*(x)] = E_* \left[ \sum_{i=1}^m w_i^*(x; h) Y_i^* \right] = E_* \left[ \sum_{i=1}^n w_i(x; h) Y_i e_i \right],$$

donde  $e_i$ , representa una variable aleatoria definida como el número de veces que la observación,  $(X_i, Y_i)$ , aparece en la remuestra considerada; de modo que  $e_1, \dots, e_n$ ,

forman un conjunto de  $n$  variables aleatorias independientes e idénticamente distribuidas según  $B(m, \frac{1}{n})$ . Teniendo en cuenta que el operador  $E_*$ , tan sólo actúa sobre las variables  $e_i$ , se tiene trivialmente que:

$$E_* [\widehat{m}_h^*(x)] = \frac{m}{n} \sum_{i=1}^n w_i(x; h) Y_i = \frac{m}{n} \widehat{m}_h(x),$$

y junto con la definición (4.4) llegamos a la expresión deseada (4.6).

Ahora probamos que el término de varianza (4.5) admite la expresión:

$$V_{h,g}^*(x) = \frac{m}{n} \sum_{i=1}^n w_i^2(x; h) Y_i^2 - \frac{m}{n^2} (\widehat{m}_h(x))^2. \quad (4.7)$$

Aplicando la fórmula de Köning para el cálculo de la varianza,

$$\text{Var}_* (\widehat{m}_h^*(x)) = E_* \left[ \left( \sum_{i=1}^m w_i(x; h)^* Y_i^* \right)^2 \right] - \left( \frac{m}{n} \widehat{m}_h(x) \right)^2. \quad (4.8)$$

Ahora bien

$$\begin{aligned} E_* \left[ \left( \sum_{i=1}^m w_i^*(x; h) Y_i^* \right)^2 \right] &= E_* \left[ \left( \sum_{i=1}^n w_i(x; h) Y_i e_i \right)^2 \right] \\ &= \sum_{i=1}^n w_i(x; h)^2 Y_i^2 E_* [e_i^2] + \sum_{i \neq j}^n w_i(x; h) w_j(x; h) Y_i Y_j E_* [e_i e_j]. \end{aligned}$$

Si reconsideramos la definición dada anteriormente a las variables aleatorias  $e_i$ , resulta fácil ver que la variable aleatoria  $n$ -dimensional,  $(e_1, \dots, e_n)$ , se distribuye según una multinomial,  $M(m, \frac{1}{n}, \dots, \frac{1}{n})$ ; y como para tal distribución es conocido que  $E[e_i e_j] = \frac{m^2 - m}{n^2}$ , sustituyendo este valor y reagrupando términos se obtiene que

$$E_* \left[ \left( \sum_{i=1}^m w_i^*(x; h) Y_i^* \right)^2 \right] = \frac{m}{n} \sum_{i=1}^n w_i(x; h)^2 Y_i^2 + \frac{m^2 - m}{n^2} \left( \sum_{i=1}^n w_i(x; h) Y_i \right)^2.$$

Si sustituímos este resultado en (4.8) tenemos que

$$V_{h,g}^*(x) = \frac{m}{n} \sum_{i=1}^n w_i(x; h)^2 Y_i^2 + \left\{ \frac{m^2 - m}{n^2} - \frac{m^2}{n^2} \right\} (\widehat{m}_h(x))^2.$$

Finalmente, si unimos las expresiones (4.6) y (4.7) obtenemos la expresión buscada para la estimación *bootstrap* del error,

$$\text{MSE}^*(x; h) = \left( \frac{m}{n} \widehat{m}_h(x) - \widehat{m}_g(x) \right)^2 + \frac{m}{n} \sum_{i=1}^n w_i(x; h)^2 Y_i^2 - \frac{m}{n^2} (\widehat{m}_h(x))^2. \quad (4.9)$$

Y si integramos la expresión anterior se obtiene que la estimación *bootstrap* propuesta para el error global, MISE, viene dada por

$$\begin{aligned} \text{MISE}^*(h) &= \int \left\{ \frac{m}{n} \widehat{m}_h(x) - \widehat{m}_g(x) \right\}^2 dx + \\ &+ \frac{m}{n} \sum_{i=1}^n Y_i^2 \int w_i(x; h)^2 dx - \frac{m}{n^2} \int (\widehat{m}_h(x))^2 dx. \end{aligned} \quad (4.10)$$

## 4.5 *Bootstrap* clásico sobre los residuos

### 4.5.1 Estimación *bootstrap* del error cuadrático medio

Cuando se opta por un remuestreo sobre los residuos, según se planteaba en el punto (4.3.1) de este capítulo, puesto que los residuos,  $\epsilon_i$ , del modelo de regresión no son observables, es necesario estimarlos de acuerdo a una expresión del tipo

$$\widehat{\epsilon}_i = Y_i - \widehat{m}_g(X_i),$$

siendo  $\widehat{m}_g$  el estimador polinomial local con un *ancho de banda piloto*,  $g$ .

El procedimiento para la estimación del error cuadrático medio local se puede resumir en los siguientes pasos:

1. Elegir un ancho de banda piloto,  $g$ , y construir el estimador polinomial local,  $\widehat{m}_g$ .
2. Calcular las estimaciones de los residuos según la expresión:

$$\widehat{\epsilon}_i = Y_i - \widehat{m}_g(X_i);$$

y como tales estimaciones no tienen por qué tener media cero, considerar las versiones centradas:

$$\tilde{\epsilon}_i = \widehat{\epsilon}_i - n^{-1} \sum_{i=1}^n \widehat{\epsilon}_i.$$

3. Diseñar muestras *bootstrap* de los residuos centrados,  $\tilde{\epsilon}_i$ , mediante muestreo aleatorio con reemplazamiento de la forma:

$$\{\epsilon_1^*, \epsilon_2^*, \dots, \epsilon_n^*\},$$

con lo que se obtienen respuestas *bootstrap*  $Y_i^*$ ,  $Y_i^* = \hat{m}_g(X_i) + \epsilon_i^*$ , esto es, a partir de los datos de partida, generar conjuntos de datos *bootstrap* del tipo

$$H_n^* = \{(X_1, Y_1^*), (X_2, Y_2^*), \dots, (X_n, Y_n^*)\}.$$

4. Considerar para cada  $h$ ,  $h > 0$ , el estimador polinomial local basado en el conjunto de datos generado,  $H_n^*$ ,

$$\hat{m}_h^*(x) = \sum_{i=1}^n w_i(x; h) Y_i^*.$$

5. Obtener la estimación *bootstrap* del error cuadrático medio definida como:

$$\text{MSE}^*(x; h) = E_* [\{\hat{m}_h^*(x) - \hat{m}_g(x)\}^2],$$

de forma que el selector *bootstrap* propuesto para el ancho de banda estará dado por

$$h_{\text{boot}}^*(x) = \underset{h}{\text{argmin}} \text{MSE}^*(x; h).$$

A continuación calculamos la expresión que adopta la estimación propuesta para el error cuadrático medio bajo este remuestreo sobre los residuos.

#### 4.5.2 Expresión exacta del estimador *bootstrap*

De nuevo el error cuadrático medio admite una descomposición del tipo

$$\text{MSE}^*(x; h) = (B_{h,g}^*(x))^2 + V_{h,g}^*(x),$$

siendo

$$B_{h,g}^*(x) = E_* [\hat{m}_h^*(x)] - \hat{m}_g(x) \quad (4.11)$$

y la varianza

$$V_{h,g}^*(x) = \text{Var}_* (\widehat{m}_h^*(x)). \quad (4.12)$$

En lo que sigue probaremos que

$$B_{h,g}^*(x) = \sum_{j=1}^n Y_j \left( \sum_{i=1}^n w_i(x; h) w_j(X_i; g) - w_j(x; g) \right) \quad (4.13)$$

y

$$V_{h,g}^*(x) = n^{-1} \sum_{i=1}^n w_i(x; h)^2 \sum_{j=1}^n \tilde{\epsilon}_j^2. \quad (4.14)$$

Empezamos calculando:

$$\begin{aligned} E_* [\widehat{m}_h^*(x)] &= E_* \left[ \sum_{i=1}^n w_i(x; h) (\widehat{m}_g(X_i) + \epsilon_i^*) \right] \\ &= \sum_{i=1}^n w_i(x; h) (\widehat{m}_g(X_i) + E_* [\epsilon_i^*]), \end{aligned}$$

y dado que en la distribución *bootstrap*,

$$E_* [\epsilon_i^*] = \frac{1}{n} \sum_{j=1}^n \tilde{\epsilon}_j = 0,$$

tenemos que

$$E_* [\widehat{m}_h^*(x)] = \sum_{i=1}^n w_i(x; h) \widehat{m}_g(X_i) = \sum_{i=1}^n \sum_{j=1}^n w_i(x; h) w_j(X_i; g) Y_j,$$

obteniéndolo la expresión buscada, (4.13).

Probamos ahora la expresión dada para el término de varianza:

$$\begin{aligned} \text{Var}_* (\widehat{m}_h^*(x)) &= \text{Var}_* \left( \sum_{i=1}^n w_i(x; h) (\widehat{m}_g(X_i) - \epsilon_i^*) \right) \\ &= \sum_{i=1}^n w_i(x; h)^2 \text{Var}_* (\widehat{m}_g(X_i) - \epsilon_i^*) = \sum_{i=1}^n w_i(x; h)^2 \text{Var}_* (\epsilon_i^*) \\ &= \frac{1}{n} \sum_{i=1}^n w_i(x; h)^2 \sum_{j=1}^n \tilde{\epsilon}_j^2 \end{aligned}$$



que es justamente la expresión, (4.14), a la que pretendíamos llegar.

Y si integramos la estimación obtenida para el error cuadrático medio, obtenemos la estimación *bootstrap* del error global, MISE:

$$\begin{aligned} \text{MISE}^*(h) &= \int \left\{ \sum_{j=1}^n Y_j \left( \sum_{i=1}^n w_j(X_i; g) w_i(x; h) - w_j(x; g) \right) \right\}^2 dx \\ &\quad + n^{-1} \sum_{j=1}^n \tilde{\epsilon}_j^2 \sum_{i=1}^n \int w_i(x; h) dx. \end{aligned}$$

### 4.5.3 Sobre el ancho de banda piloto

Utilizar la metodología *bootstrap* descrita requiere el paso previo de la elección del ancho de banda piloto que involucra.

En los procedimientos de tipo *bootstrap* es habitual la aparición de un parámetro de este tipo, y generalmente interviene de forma notable en el buen funcionamiento de los mismos.

Muchos autores tratan dicho parámetro como una función de  $h$ , del tipo  $g = Ch^\alpha$ , y proponen elecciones de  $C$  y  $\alpha$ , buscando la proximidad entre el estimador *bootstrap* de la función y la propia función.

En nuestro trabajo tratamos la elección del parámetro piloto sin suponer dependencia de  $h$ , y definiendo un criterio de proximidad entre el error cuadrático medio teórico y su estimación *bootstrap*. El estudio bajo la perspectiva de *bootstrap* clásico sobre los residuos, es análogo al que surge cuando consideramos un *bootstrap* del tipo *wild bootstrap* sobre los residuos, que será el objeto de la siguiente sección. Por tanto dejamos por el momento el problema de selección del parámetro piloto, si bien los resultados que se obtengan en el *wild bootstrap* se aplicarán de forma directa a este *bootstrap* clásico.

## 4.6 Wild bootstrap sobre los residuos

Consideramos de nuevo un remuestreo sobre los residuos, pero en este caso generaremos los residuos *bootstrap* a partir de una distribución degenerada en dos puntos, que tenga media 0, varianza igual al cuadrado del residuo, y momento de tercer orden igual al cubo del residuo.

Esta versión de la metodología *bootstrap* fue aplicada por Härdle y Marron (1991) en el contexto de la regresión no paramétrica, si bien considerando estimadores núcleo del tipo Nadaraya-Watson. No obstante, estos autores persiguen una estima-

ción *bootstrap* del error cuadrático medio, que sirva para la construcción de intervalos de confianza sobre las estimaciones no paramétricas de la curva de regresión.

En esta memoria ampliamos el contexto considerado por estos autores, al caso en que se considera como estimadores de regresión los de tipo polinomial local; además nuestra intención es utilizar el estimador *bootstrap* resultante del error cuadrático medio, como selector del parámetro ancho de banda, obteniéndose del tipo local no asintótico y no iterativo.

A continuación desarrollamos ampliamente estas ideas.

#### 4.6.1 Estimación *bootstrap* del error cuadrático medio

Partimos del modelo de regresión no paramétrica bajo el diseño aleatorio, formulado al comienzo de esta memoria,

$$Y_i = m(X_i) + \epsilon_i,$$

donde los residuos,  $\epsilon_i$ , se suponen de media 0 y varianza  $\sigma^2(X_i)$ . De nuevo, y dado que dichos residuos serán desconocidos, consideramos las estimaciones,

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

siendo  $\hat{m}_g(X_i)$ , el estimador polinomial local para algún ancho de banda piloto,  $g$ . Estos residuos estimados,  $\hat{\epsilon}_i$ , serán los objetos del remuestreo que proponemos.

Según la metodología *wild bootstrap*, definimos nuevas variables aleatorias,  $\epsilon_i^*$ , con distribución,  $\hat{G}_i$ , dada por

$$\hat{G}_i = \gamma\delta_a + (1 - \gamma)\delta_b;$$

donde  $\delta_a$  y  $\delta_b$ , denotan medidas puntuales en  $a$  y  $b$  respectivamente; y los parámetros  $a$ ,  $b$  y  $\gamma$ , se eligen satisfaciendo las siguientes condiciones:

1.  $E[\epsilon_i^*] = 0$ ,
2.  $E[\epsilon_i^{*2}] = \hat{\epsilon}_i^2$ ,
3.  $E[\epsilon_i^{*3}] = \hat{\epsilon}_i^3$ .

Es sencillo ver que, consecuentemente, los valores que deben tener dichos parámetros son:

$$a = \hat{\epsilon}_i \frac{1 - \sqrt{5}}{2}, \quad b = \hat{\epsilon}_i \frac{1 + \sqrt{5}}{2}, \quad \gamma = \frac{5 + \sqrt{5}}{10}.$$

Después de dicho remuestreo se definen las nuevas observaciones *bootstrap* de la forma

$$Y_i^* = \hat{m}_g(X_i) + \epsilon_i^*.$$

Si ahora aplicamos el suavizador polinomial local al conjunto de datos *bootstrap*,  $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$ , se obtiene el estimador polinomial local *bootstrap*,

$$\widehat{m}_h^*(x) = \sum_{i=1}^n w_i(x; h) Y_i^*.$$

Finalmente se propone como estimación *bootstrap* del error cuadrático medio asociado al estimador polinomial local considerado,  $\widehat{m}_h(\cdot)$ , que denotamos por  $MSE^*$ , la definida como:

$$MSE^*(x; h) = E_* [(\widehat{m}_h^*(x) - \widehat{m}_g(x))^2], \quad (4.15)$$

siendo  $E_*$ , el operador esperanza sobre la distribución en el remuestreo, esto es la definida por las  $\widehat{G}_i$  anteriores.

Obsérvese que es posible obtener una versión integrada de la forma:

$$MISE^*(h) = \int MSE^*(x; h) dx,$$

teniendo así una estimación *bootstrap* del error cuadrático medio integrado MISE.

#### 4.6.2 Elección *bootstrap* del ancho de banda

Proponemos utilizar la estimación *bootstrap* del error cuadrático medio, obtenida según el procedimiento que acabamos de describir, como criterio de selección del parámetro ancho de banda. De este modo definimos la elección *bootstrap* del ancho de banda local como la función

$$h_{\text{boot}}^*(x) = \underset{h}{\operatorname{argmin}} \quad MSE^*(x; h)$$

que varía en cada punto de estimación considerado,  $x$ .

De forma análoga es posible elegir un ancho de banda global, constante en todo el rango de estimación considerado, utilizando la versión integrada de la estimación *bootstrap* del error como criterio de selección. Definimos por tanto, el ancho de banda *bootstrap* global como

$$h_{\text{boot}}^* = \underset{h}{\operatorname{argmin}} \quad MISE^*(h).$$

#### 4.6.3 Expresión exacta del estimador *bootstrap*

Nuestro objetivo es ahora calcular la expresión exacta que adopta la estimación *bootstrap* del error cuadrático medio,  $MSE^*(x; h)$ , calculando la esperanza bajo la distribución en el remuestreo,  $E_*$ .

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

$$\text{MSE}^*(x; h, g) = (\text{B}_{h,g}^*(x))^2 + \text{V}_{h,g}^*(x),$$

donde en la notación se ha incluido la dependencia existente del parámetro piloto,  $g$ , y los términos de sesgo y varianza se definen como

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

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

Procedemos ahora al cálculo de cada término, comenzando por el de sesgo:

$$\begin{aligned} \text{E}_* [\hat{m}_h^*(x)] &= \text{E}_* \left[ \sum_{i=1}^n w_i(x; h) Y_i^* \right] = \text{E}_* \left[ \sum_{i=1}^n w_i(x; h) (\hat{m}_g(X_i) + \epsilon_i^*) \right] \\ &= \sum_{i=1}^n w_i(x; h) \hat{m}_g(X_i) + \sum_{i=1}^n w_i(x; h) \text{E}_*[\epsilon_i^*] = \sum_{i=1}^n w_i(x; h) \hat{m}_g(X_i), \end{aligned}$$

con lo cual

$$\text{B}_{h,g}^*(x) = \sum_{i=1}^n w_i(x; h) \hat{m}_g(X_i) - \hat{m}_g(x).$$

Para el término de varianza

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

y por tanto

$$\text{V}_{h,g}^*(x) = \sum_{i=1}^n w_i(x; h)^2 (Y_i - \hat{m}_g(X_i))^2.$$

De este modo la expresión final de la estimación *bootstrap* propuesta para el error cuadrático medio sería

$$\text{MSE}^*(x, h; g) = \left\{ \sum_{i=1}^n w_i(x; h) \hat{m}_g(X_i) - \hat{m}_g(x) \right\}^2 + \sum_{i=1}^n w_i(x; h)^2 (Y_i - \hat{m}_g(X_i))^2.$$

Y su versión integrada

$$\begin{aligned} \text{MISE}^*(h; g) &= \int \left\{ \sum_{i=1}^n w_i(x; h) \widehat{m}_g(X_i) - \widehat{m}_g(x) \right\}^2 dx + \\ &+ \sum_{i=1}^n (Y_i - \widehat{m}_g(X_i))^2 \int w_i(x; h)^2 dx. \end{aligned}$$

Es posible obtener versiones matriciales de tales expresiones, similares a las habituales en los estimadores de regresión no paramétrica que estamos considerando. Con este propósito introducimos la siguiente notación:

$$\begin{aligned} \widehat{M}_g &= (\widehat{m}_g(X_1), \dots, \widehat{m}_g(X_n))^T \\ \Sigma_g &= \text{diag}(\widehat{\epsilon}_i). \end{aligned}$$

Con ella y con la definida al comienzo de esta memoria, siguiendo una metodología análoga a la empleada para obtener el sesgo y varianza del estimador, se obtienen las siguientes expresiones matriciales del sesgo y la varianza respectivamente:

$$\begin{aligned} B_{h,g}^*(x) &= e_1^T (X_x^T W_{x,h} X_x)^{-1} X_x^T W_{x,h} \widehat{M}_g - \\ &- e_1^T (X_x^T W_{x,g} X_x)^{-1} X_x^T W_{x,g} y \end{aligned} \quad (4.16)$$

$$V_{h,g}^*(x) = e_1^T (X_x^T W_{x,h} X_x)^{-1} X_x^T W_{x,h} \Sigma_g W_{x,h} X_x^T (X_x^T W_{x,h} X_x)^{-1} e_1. \quad (4.17)$$

#### 4.6.4 Expresión asintótica del sesgo y la varianza *bootstrap*

Supondremos por simplicidad el caso de ajustes polinomiales de grado 1, si bien siguiendo las mismas pautas, no resultará difícil generalizar a un grado  $p$ ,  $p > 1$ , arbitrario.

Formulemos las siguientes hipótesis:

1. En el punto de estimación considerado,  $x$ ,  $f(x) > 0$ ,
2. las funciones  $f(\cdot)$ ,  $m''(\cdot)$  y  $\sigma^2(\cdot)$ , son continuas en un entorno de  $x$ ,
3.  $h \rightarrow 0$  y  $nh \rightarrow \infty$  cuando  $n \rightarrow \infty$ .

**Teorema 4.6.1** *Bajo los supuestos anteriores, se verifica la siguiente expresión asintótica para la estimación bootstrap del error cuadrático medio:*

$$\begin{aligned} \text{MSE}^*(x; h, g) &= \frac{1}{4} h^4 (\widehat{m}_g''(x))^2 \mu_2^2 + (nh)^{-1} R(K) f(x)^{-1} (m(x) - \widehat{m}_g(x) + \epsilon_x)^2 \\ &+ o_P(h^4) + o_P((nh)^{-1}). \end{aligned}$$

**Demostración** En Wand y Jones podemos encontrar la siguiente aproximación asintótica:

$$n^{-1} (X_x^T W_{x,h} X_x)^{-1} = \begin{pmatrix} -\frac{1}{f(x)} + o_P(1) & f'(x)f(x)^2 + o_P(1) \\ -\frac{f'(x)}{f(x)^2} + o_P(1) & (h^2 \mu_2 f(x))^{-1} + o_P(h^{-2}) \end{pmatrix}.$$

Considerando el desarrollo de Taylor de la función  $\hat{m}_g$ , en cada  $X_i$ :

$$\hat{m}_g(X_i) = \hat{m}_g(x) + (X_i - x)\hat{m}'_g(x) + \frac{1}{2}(X_i - x)^2\hat{m}''_g(x) + \dots,$$

el vector  $\widehat{M}_g$  se expresa como:

$$\begin{aligned} \widehat{M}_g &= \begin{pmatrix} \hat{m}_g(x) + (X_1 - x)\hat{m}'_g(x) + \frac{1}{2}(X_1 - x)^2\hat{m}''_g(x) + \dots \\ \vdots \\ \hat{m}_g(x) + (X_n - x)\hat{m}'_g(x) + \frac{1}{2}(X_n - x)^2\hat{m}''_g(x) + \dots \end{pmatrix} \\ &= X_x \begin{pmatrix} \hat{m}_g(x) \\ \hat{m}'_g(x) \end{pmatrix} + \frac{1}{2}\hat{m}''_g(x) \begin{pmatrix} (X_1 - x)^2 \\ \vdots \\ (X_n - x)^2 \end{pmatrix} + \dots \end{aligned}$$

Sustituyendo en la expresión matricial del sesgo *bootstrap* (4.16) se tiene que:

$$\begin{aligned} B_{h,g}^*(x) &\approx e_1^T (X_x^T W_{x,h} X_x)^{-1} X_x^T W_{x,h} X_x \begin{pmatrix} \hat{m}_g(x) \\ \hat{m}'_g(x) \end{pmatrix} + \\ &\quad + \frac{1}{2}\hat{m}''_g(x) e_1^T (X_x^T W_{x,h} X_x)^{-1} X_x^T W_{x,h} \begin{pmatrix} (X_1 - x)^2 \\ \vdots \\ (X_n - x)^2 \end{pmatrix} - \hat{m}_g(x), \end{aligned}$$

y de nuevo usando cálculos estándar en la teoría de regresión polinomial local, tenemos la siguiente expresión asintótica para el sesgo *bootstrap*:

$$B_{h,g}^*(x) = \frac{1}{2}h^2\hat{m}''_g(x)\mu_2 + o_P(h^2).$$

Procedemos ahora con la varianza *bootstrap*, que matricialmente está definida como:

$$V_{h,g}(x) = e_1^T (X_x^T W_{x,h} X_x)^{-1} X_x^T W_{x,h} \Sigma_g W_{x,h} X_x^T (X_x^T W_{x,h} X_x)^{-1} e_1,$$

y consideramos el término:

$$\begin{aligned} n^{-1} X_x^T W_{x,h} \Sigma_g W_{x,h} X_x &= n^{-1} \sum_{i=1}^n K_h^2(X_i - x) \widehat{\epsilon}_i^2 \begin{pmatrix} 1 & X_i - x \\ X_i - x & (X_i - x)^2 \end{pmatrix} = \\ &= \begin{pmatrix} n^{-1} \sum_{i=1}^n K_h^2(X_i - x) \widehat{\epsilon}_i^2 & n^{-1} \sum_{i=1}^n (X_i - x) K_h^2(X_i - x) \widehat{\epsilon}_i^2 \\ n^{-1} \sum_{i=1}^n (X_i - x) K_h^2(X_i - x) \widehat{\epsilon}_i^2 & n^{-1} \sum_{i=1}^n (X_i - x)^2 K_h^2(X_i - x) \widehat{\epsilon}_i^2 \end{pmatrix} \end{aligned} \quad (4.18)$$

donde  $\widehat{\epsilon}_i^2 = (Y_i - \widehat{m}_g(X_i))^2$ . Buscamos ahora una expresión asintótica para estos residuos estimados que permita simplificar los elementos de la matriz anterior,

$$Y_i - \widehat{m}_g(X_i) = m(X_i) + \epsilon_i.$$

Si  $n$  es suficientemente grande y la varianza residual  $\sigma^2(\cdot)$  es continua, entonces en un entorno del punto de estimación considerado podemos hacer la siguiente aproximación:

$$m(X_i) - \widehat{m}_g(X_i) + \epsilon_i \approx m(x) - \widehat{m}_g(x) + \epsilon_x,$$

siendo  $\epsilon_x$ , el residuo en  $x$ , y por tanto de media 0 y varianza  $\sigma^2(x)$ .

Por tanto, la matriz (4.18) admite una expresión asintótica de la forma:

$$(m(x) - \widehat{m}_g(x) + \epsilon_x)^2 \begin{pmatrix} h^{-1} R(K) f(x) + o_P(h^{-1}) & o_P(1) \\ o_P(1) & h \mu_2(K^2) f(x) + o_P(h) \end{pmatrix}$$

y utilizando desarrollos anteriores, asintóticamente, se tiene que

$$V_{h,g}^*(x) = (nh)^{-1} R(K) f(x)^{-1} (m(x) - \widehat{m}_g(x) + \epsilon_x)^2 + o_P(n^{-1} h^{-1}).$$

□

**Corolario 4.6.2** *Bajo las mismas condiciones del teorema anterior, el selector bootstrap local del ancho de banda tiene la siguiente expresión asintótica:*

$$h_{boot}^*(x) = \left( \frac{R(K) (m(x) - \widehat{m}_g(x) + \epsilon_x)^2}{f(x) (\widehat{m}_g''(x))^2 \mu_2^2} \right)^{1/5} n^{-1/5} \quad (4.19)$$

**Corolario 4.6.3** *Bajo los supuestos hechos en el teorema, la expresión asintótica para el estimador bootstrap de error global, MISE, es*

$$\begin{aligned} MISE^*(h) &= \frac{1}{4} h^4 \mu_2^2 \int (\widehat{m}_g''(x))^2 f(x)^{-1} dx + \\ &+ (nh)^{-1} R(K) \int (m(x) - \widehat{m}_g(x) + \epsilon_x)^2 dx + o_P(h^4) + o_P((nh)^{-1}). \end{aligned}$$

y el selector *bootstrap global* del ancho de banda es asintóticamente:

$$h_{boot}^* = \left( \frac{R(K) \int (m(x) - \hat{m}_g(x) + \epsilon_x)^2}{\mu_2^2 \int (\hat{m}_g''(x))^2 f(x)^{-1} dx} \right)^{1/5} n^{-1/5}.$$

**Observación 4.6.4** Resultan elecciones del orden de  $n^{-1/5}$ , tanto en el caso local como en el caso global, por tanto se trata de una elección asintóticamente óptima del ancho de banda, del mismo orden que las elecciones óptimas teóricas dadas en (2.12) y (2.13).

#### 4.6.5 Elección del ancho de banda piloto

El procedimiento de estimación *bootstrap* anteriormente descrito, supone la elección de un parámetro ancho de banda piloto,  $g$ , que determinará las estimaciones de los residuos que intervienen en el remuestreo.

Del estudio con datos simulados que se presenta en el capítulo siguiente, se deduce que no es válida cualquier aproximación para dicho parámetro, de hecho, para algunas elecciones, la estimación *bootstrap* resultante no describe la estructura real subyacente.

En esta sección describiremos cómo elegir de forma adecuada este ancho de banda piloto, considerando en los desarrollos el caso del estimador polinomial local de grado uno.

En principio sería razonable hacer una selección rápida y sencilla basada, por ejemplo, en algún procedimiento como el de validación cruzada. No obstante tales elecciones no resultan adecuadas en la práctica. Probaremos que el ancho de banda piloto,  $g$ , debe ser de orden  $n^{-1/9}$ , en contraposición a elecciones como la mencionada validación cruzada que proporcionan órdenes  $n^{-1/5}$ .

En la búsqueda de un criterio para decidir qué parámetro,  $g$ , es el más adecuado, consideramos la expresión asintótica para la estimación *bootstrap* del error cuadrático medio,  $MSE^*(x; h, g)$  (introducimos de nuevo en la notación la dependencia existente de  $g$ ) que proporciona el teorema 4.6.1:

$$\begin{aligned} MSE^*(x; h, g) &= \frac{1}{4} h^4 (\hat{m}_g''(x))^2 \mu_2^2 + (nh)^{-1} R(K) f(x)^{-1} (m(x) - \hat{m}_g(x) + \epsilon_x)^2 + \\ &+ o_P(h^4) + o_P((nh)^{-1}). \end{aligned} \tag{4.20}$$



Si tomamos ahora la expresión asintótica del error cuadrático medio teórico (véase por ejemplo Wand y Jones, 1995):

$$\begin{aligned} \text{MSE}(x; h) &= \frac{1}{4}h^4 (m''(x))^2 \mu_2^2 + (nh)^{-1}R(K)f(x)^{-1}\sigma^2(x) + \\ &+ o_P(h^4) + o_P((nh)^{-1}), \end{aligned} \quad (4.21)$$

se puede observar que en tales expresiones, para tamaños muestrales grandes, el término de varianza es casi despreciable, dado que aparece multiplicado por  $(nh)^{-1}$ . De este modo y desde un punto de vista asintótico, es el término de sesgo el que adquiere mayor relevancia. No obstante, en las simulaciones realizadas en el capítulo siguiente se puede observar que, ciertamente, el parámetro piloto no es determinante en la estimación de la varianza, incluso para tamaños muestrales pequeños.

Como consecuencia y dado que el término de sesgo depende del ancho de banda piloto,  $g$ , tomamos como criterio para su elección, minimizar la distancia entre el sesgo *bootstrap* y el sesgo teórico. Esta idea ya fue empleada entre otros, por Cao, R. (1993) en el contexto de estimación de densidades y para un *bootstrap* suavizado; y en el contexto de regresión no paramétrica para estimadores tipo núcleo, por Härdle y Marron (1991). En ambos trabajos, al igual que en la presente memoria, se consideran las distancias en media cuadrática, esto es, buscamos minimizar:

$$E_{Y/X} [B_{h,g}^*(x) - B_h(x)]^2,$$

donde el sesgo teórico es

$$B_h(x) = \frac{1}{2}h^2\mu_2m''(x).$$

Estudiando las expresiones  $B_h(x)$  y  $B_{h,g}(x)$ , obtenemos que la proximidad entre ambas tan sólo depende de la proximidad entre  $m''(x)$  y  $\hat{m}_g''(x)$ , de forma que adoptamos una versión más simple del criterio anterior, que denotamos  $L(x; g)$ , y que definimos como:

$$L(x; g) = E_{Y/X} [\hat{m}_g''(x) - m''(x)]^2. \quad (4.22)$$

El ancho de banda piloto óptimo será el que minimice tal criterio y por tanto el que consiga una mayor proximidad entre el sesgo *bootstrap* y el sesgo teórico. Denotaremos a dicha elección por  $g_{\text{opt}}(\cdot)$ , de modo que en cada punto de estimación,  $x$ , estará definida como:

$$g_{\text{opt}}(x) = \underset{g}{\text{argmin}} L(x; g). \quad (4.23)$$

También sería posible considerar una versión integrada del criterio,  $L(\cdot; g)$ , sobre todo el rango de estimación considerado, esto es, utilizar un criterio global que denotamos  $IL(g)$ , y que viene dado por:

$$IL(g) = \int E_{Y/X} [\hat{m}_g''(x) - m''(x)]^2 dx.$$

En tal caso definimos el ancho de banda piloto óptimo global como:

$$g_{\text{opt}} = \underset{g}{\operatorname{argmin}} IL(g). \quad (4.24)$$

El siguiente teorema da una expresión asintótica del criterio,  $L(x; g)$ , que acabamos de definir:

**Teorema 4.6.5** *Bajo los siguientes supuestos:*

1. *En el punto de estimación considerado,  $x$ ,  $f(x) > 0$ ,*
2. *Las funciones  $f(\cdot)$  y  $\sigma^2(\cdot)$ , son continuas en un entorno de  $x$ , y existen las derivadas  $f'$ ,  $f''$ ,  $K'$ ,  $K''$ ,  $m''$  y  $m^{(4)}$  y son continuas en un entorno de  $x$ ,*
3. *La función varianza,  $\sigma^2(\cdot)$ , es constante en un entorno de  $x$ , con valor  $\sigma^2(x)$ , y*
4.  *$h \rightarrow 0$  y  $nh \rightarrow \infty$ , cuando  $n \rightarrow \infty$ ,*

*el criterio  $L(x; g)$  tiene la siguiente aproximación asintótica:*

$$\begin{aligned} L(x; g) = & \frac{1}{4} (m^{(4)}(x))^2 g^4 \mu_2^2 + \\ & + \sigma^2(x) \left\{ n^{-1} g^{-1} R(K) \frac{2f''(x)f(x)^2 + 6f(x)^4 - 7f''(x)f'(x)^2 f(x)}{f(x)^5} \right. \\ & \left. + n^{-1} g^{-3} R(K') \frac{2f'(x)}{f(x)^3} + n^{-1} g^{-5} R(K'') f(x)^{-1} \right\} \\ & + o_P(g^2) + o_P(n^{-1}g^{-1}) + o_P(n^{-1}g^{-3}) + o_P(n^{-1}g^{-5}) \end{aligned} \quad (4.25)$$

**Demostración.** El criterio  $L(x; g)$  admite una descomposición en suma de un término de sesgo y otro de varianza de la forma

$$L(x; g) = \{E_{Y/X} [\hat{m}_g''(x)] - m''(x)\}^2 + \operatorname{Var}_{Y/X} (\hat{m}_g''(x)). \quad (4.26)$$

Comenzamos con el término de sesgo:

Dado que  $\hat{m}_g(x) = \sum_{i=1}^n w_i(x; g)Y_i$ , su derivada segunda es

$$\hat{m}_g''(x) = \sum_{i=1}^n w_i''(x; g)Y_i,$$

con lo que

$$E_{Y/X} [\hat{m}_g''(x)] = \sum_{i=1}^n w_i''(x; g)m(X_i).$$

Es posible obtener una expresión asintótica para los pesos<sup>4</sup>,  $w_i(x; g)$ , que resulte más sencilla de derivar:

$$w_i(x; g) = n^{-1}g^{-1}f(x)^{-1}K_g(X_i - x)(1 + o_P(1)),$$

entonces

$$\begin{aligned} w_i''(x) &= -n^{-1}g^{-1} \frac{f''(x)f(x) - 2f'(x)^2}{f(x)^3} K\left(\frac{X_i - x}{g}\right) + \\ &+ 2n^{-1}g^{-2} \frac{f'(x)}{f(x)^2} K'\left(\frac{X_i - x}{g}\right) + \\ &+ n^{-1}g^{-3}f(x)^{-1} K''\left(\frac{X_i - x}{g}\right), \end{aligned} \quad (4.27)$$

y por tanto:

$$\begin{aligned} E_{Y/X} [\hat{m}_g''(x)] &= \frac{f''(x)f(x) - 2f'(x)^2}{f(x)^3} \left( -n^{-1}g^{-1} \sum_{i=1}^n K\left(\frac{X_i - x}{g}\right) m(X_i) \right) + \\ &+ \frac{f'(x)}{f(x)^2} 2n^{-1}g^{-2} \sum_{i=1}^n K'\left(\frac{X_i - x}{g}\right) m(X_i) + \\ &+ f(x)^{-1}n^{-1}g^{-3} \sum_{i=1}^n K''\left(\frac{X_i - x}{g}\right) m(X_i). \end{aligned}$$

Si definimos

---

<sup>4</sup>Fan y Gijbels (1996), pp. 64.

$$S_{n,0}(x) = n^{-1}g^{-1} \sum_{i=1}^n K\left(\frac{X_i - x}{g}\right),$$

es fácil ver que la expresión anterior se puede reducir a:

$$E_{Y/X} [\hat{m}_g''(x)] = (f(x)^{-1}S_{n,0}(x))''.$$

Consideremos ahora el término  $S_{n,0}(x)$  y el desarrollo de Taylor de la función,  $m$ , en cada  $X_i$ , de forma que:

$$\begin{aligned} S_{n,0}(x) &= m(x)n^{-1}g^{-1} \sum_{i=1}^n K\left(\frac{X_i - x}{g}\right) + \\ &+ m'(x)n^{-1}g^{-1} \sum_{i=1}^n (X_i - x)K\left(\frac{X_i - x}{g}\right) + \\ &\frac{1}{2}m''(x)n^{-1}g^{-1} \sum_{i=1}^n (X_i - x)^2 K\left(\frac{X_i - x}{g}\right) \end{aligned}$$

y dado que

$$n^{-1}g^{-1} \sum_{i=1}^n K\left(\frac{X_i - x}{g}\right) = f(x) + o_P(1)$$

$$n^{-1}g^{-1} \sum_{i=1}^n (X_i - x)K\left(\frac{X_i - x}{g}\right) = o_P(g)$$

$$n^{-1}g^{-1} \sum_{i=1}^n (X_i - x)^2 K\left(\frac{X_i - x}{g}\right) = g^2 f(x)\mu_2 + o_P(g^2)$$

sustituyendo se obtiene:

$$S_{n,0}(x) = m(x)f(x) + \frac{1}{2}m''(x)g^2 f(x)\mu_2 + o_P(1) + o_P(g^2),$$

y por tanto

$$E_{Y/X} [\hat{m}_g''(x)] = m''(x) + \frac{1}{2}m^{(4)}(x)g^2\mu_2 + o_P(1) + o_P(g^2) \quad (4.28)$$

Si consideramos ahora el término de varianza y dado que suponemos homocedasticidad local en el modelo, se verifica que  $\sigma^2(X_i) \approx \sigma^2(x)$ , en un entorno de  $x$ , de forma que:

$$\text{Var}_{Y/X}(\hat{m}_g''(x)) = \sigma^2(x) \sum_{i=1}^n (w_i''(x; g))^2 \quad (4.29)$$

Y si volvemos a la expresión (4.27) y elevamos al cuadrado y desarrollamos todos los términos que aparecen, después de simplificar se llega a la expresión:

$$\begin{aligned} \sum_{i=1}^n (w_i''(x; g))^2 &= n^{-1}g^{-1}R(K) \frac{2f''(x)f(x)^2 + 6f(x)^4 - 7f''(x)f'(x)^2f(x)}{f(x)^5} + \\ &+ n^{-1}g^{-3}R(K') \frac{2f'(x)}{f(x)^3} + n^{-1}g^{-5}R(K'')f(x)^{-1} + \\ &+ o_P(n^{-1}g^{-1}) + o_P(n^{-1}g^{-3}) + o_P(n^{-1}g^{-5}) \end{aligned} \quad (4.30)$$

donde se han utilizado las siguientes aproximaciones:

$$n^{-2}g^{-2} \sum_{i=1}^n K^2 \left( \frac{X_i - x}{g} \right) = n^{-1}g^{-1}R(K)f(x) + o_P(n^{-1}g^{-1})$$

$$n^{-2}g^{-4} \sum_{i=1}^n K'^2 \left( \frac{X_i - x}{g} \right) = n^{-1}g^{-3}R(K')f(x) + o_P(n^{-1}g^{-3})$$

$$n^{-2}g^{-6} \sum_{i=1}^n K''^2 \left( \frac{X_i - x}{g} \right) = n^{-1}g^{-5}R(K'')f(x) + o_P(n^{-1}g^{-5})$$

$$n^{-2}g^{-3} \sum_{i=1}^n K \left( \frac{X_i - x}{g} \right) K' \left( \frac{X_i - x}{g} \right) = \frac{1}{2}n^{-1}g^{-1}R(K)f'(x) + o_P(n^{-1}g^{-1})$$

$$n^{-2}g^{-5} \sum_{i=1}^n K' \left( \frac{X_i - x}{g} \right) K'' \left( \frac{X_i - x}{g} \right) = \frac{1}{2}n^{-1}g^{-3}R(K')f'(x) + o_P(n^{-1}g^{-3})$$

$$n^{-2}g^{-4} \sum_{i=1}^n K \left( \frac{X_i - x}{g} \right) K'' \left( \frac{X_i - x}{g} \right) = \frac{1}{2}n^{-1}g^{-1}R(K)f''(x) -$$

$$-n^{-1}g^{-3}R(K')f(x) + o_P(n^{-1}g^{-1}) + o_P(n^{-1}g^{-3})$$

siendo  $R(K) = \int K^2(u)du$ . Finalmente, si sustituimos (4.30), (4.29) y (4.28) en (4.26), se obtiene la expresión del criterio,  $L(x; g)$ , buscada.  $\square$

**Corolario 4.6.6** *Bajo las hipótesis del teorema y si la densidad del diseño verifica que  $f'(x) = f''(x) = 0$ , entonces:*

$$L(x; g) = \frac{1}{4} (m^{(4)}(x))^2 g^4 \mu_2^2 + \sigma^2 n^{-1} g^{-5} R(K'') f(x)^{-1} + \\ + o_P(g^2) + o_P(n^{-1}g^{-1}) + o_P(n^{-1}g^{-3}) + o_P(n^{-1}g^{-5})$$

y el ancho de banda piloto óptimo viene dado por

$$g_{opt}(x) = \left( \frac{5\sigma^2(x)R(K'')}{f(x)\mu_2^2 (m^{(4)}(x))^2} \right)^{\frac{1}{9}} n^{-\frac{1}{9}}, \quad (4.31)$$

siempre que  $m^{(4)}(x) \neq 0$ .

**Corolario 4.6.7** *Bajo las hipótesis del teorema y del corolario 4.6.6, el ancho de banda piloto global óptimo, en el sentido de minimizar el criterio global,  $IL(g)$ , tiene la siguiente expresión:*

$$g_{opt} = n^{-\frac{1}{9}} \left( \frac{5R(K'') \int \{\sigma^2(x)/f(x)\} dx}{\mu_2^2 \int (m^{(4)}(x))^2 dx} \right)^{\frac{1}{9}},$$

siempre que  $\int (m^{(4)}(x)) dx < \infty$ .

**Observación 4.6.8** *El ancho de banda piloto que resulta del criterio propuesto es de orden  $n^{-1/9}$ , lo que implica un mayor suavizamiento que el que realizan otros selectores automáticos habitualmente utilizados como validación cruzada por mínimos cuadrados, cuyo orden es de  $n^{-1/5}$ .*

**Observación 4.6.9** *Dado que en la expresión del ancho de banda piloto óptimo local, la derivada de orden 4 de la función de regresión aparece en el denominador, se puede prever que en zonas donde la función de regresión sea lineal, el ancho de banda piloto tenderá a adoptar valores muy pequeños, por lo que sería deseable una corrección del criterio en tales casos. En las simulaciones realizadas en el capítulo siguiente se puede observar esta peculiaridad con más detalle.*

**Observación 4.6.10** *Para poder aplicar esta elección de parámetro piloto en la práctica, bastaría considerar en la expresión 4.31 un estimador eficiente de la derivada cuarta de la función de regresión.*

## Capítulo 5

# Comportamiento finito muestral del selector *bootstrap*

### 5.1 Introducción

En el capítulo anterior describíamos tres selectores del ancho de banda de un estimador polinomial local, siguiendo la metodología *bootstrap*.

De los selectores *bootstrap* propuestos destacamos la dificultad que conlleva el selector basado en los pares, dado que involucra la estimación de una densidad bidimensional, así como la determinación del tamaño apropiado para las muestras *bootstrap*. Es necesario un estudio teórico más profundo de este selector *bootstrap*, lo que será objeto de futuros trabajos.

En cuanto a los selectores basados en un remuestreo de los residuos, observamos que ofrecen una expresión muy similar para la estimación *bootstrap* del error cuadrático medio. En concreto la componente de sesgo es la misma en ambos casos, y tan sólo varía la componente de varianza.

En el contexto en que nos movemos suponemos una situación general de heterocedasticidad en el modelo; por tanto el *wild bootstrap* es recomendable frente a otras alternativas. Este hecho junto con el estudio realizado en el capítulo anterior realizábamos un estudio desde el punto de vista teórico y asintótico del estimador *bootstrap*, motiva que planteemos ahora un estudio del comportamiento finito muestral tan sólo para este selector *bootstrap*. No obstante, se pueden esperar resultados muy similares para el selector *bootstrap* clásico sobre los residuos.

Dedicamos este capítulo por tanto a investigar el comportamiento de la metodología *bootstrap* mediante *wild bootstrap* sobre los residuos, a través de varios ejemplos simulados.

## 5.2 Descripción de las simulaciones

En el estudio de simulación llevado a cabo hemos considerado tres funciones de prueba, también utilizadas por Fan y Gijbels (1995b) para valorar el comportamiento del selector local del ancho de banda allí propuesto.

Las funciones de prueba utilizadas han sido:

$$m_1(x) = x + 2 \exp(-16x^2)$$

$$m_2(x) = \sin(2x) + 2 \exp(-16x^2)$$

$$m_3(x) = 0.3 \exp(-4(x+1)^2) + 0.7 \exp(-16(x-1)^2)$$

todas ellas sobre un diseño aleatorio según una distribución uniforme en el intervalo  $(-2,2)$ . Dichas funciones están representadas en la figura (5.1).

Los tres modelos de regresión se han simulado con errores,  $\epsilon_i$ , normales de media 0 y desviación típica ( $\sigma(X_i) = \sigma$ ): 0.3, 0.4 y 0.1, para las funciones de prueba,  $m_1$ ,  $m_2$  y  $m_3$ , respectivamente.

El núcleo utilizado para el estimador polinomial local ha sido el bponderado<sup>1</sup>, definido como  $K(z) = \frac{15}{16}(1-z)^2 1_{\{|z|<1\}}$ .

Se han considerado tamaños muestrales 50, 200 y 500, para las tres funciones de prueba, y el número de réplicas considerado ha sido de 20 para cada función de prueba y cada tamaño muestral considerado, siendo dicho número suficiente para la fiabilidad de los resultados que aquí se presentan.

El estimador polinomial local se ha evaluado sobre una red equiespaciada de 101 puntos, para las tres funciones de prueba y los tres tamaños muestrales considerados, dentro del intervalo de definición. Y por simplicidad, dado el contexto en que se han obtenido los resultados teóricos, consideramos sólo el caso de ajustes lineales locales.

Hemos evaluado en todos los casos el comportamiento del selector *bootstrap* (basado en *wild bootstrap* sobre los residuos) del ancho de banda propuesto en el capítulo anterior, con la elección teórica local del parámetro piloto allí calculada.

## 5.3 Aspectos computacionales

Dentro de los entornos de programación S Plus 2000 y R 1.1.x., se han implementado los estimadores de regresión polinomial local así como el selector *bootstrap* del ancho

<sup>1</sup>La elección del núcleo bponderado ha sido debida, fundamentalmente, a que para él, existe  $R(K'') = \int K''^2(u)du$ , condición necesaria para poder utilizar la elección piloto del ancho de banda que involucra el procedimiento *bootstrap* que estamos contrastando.



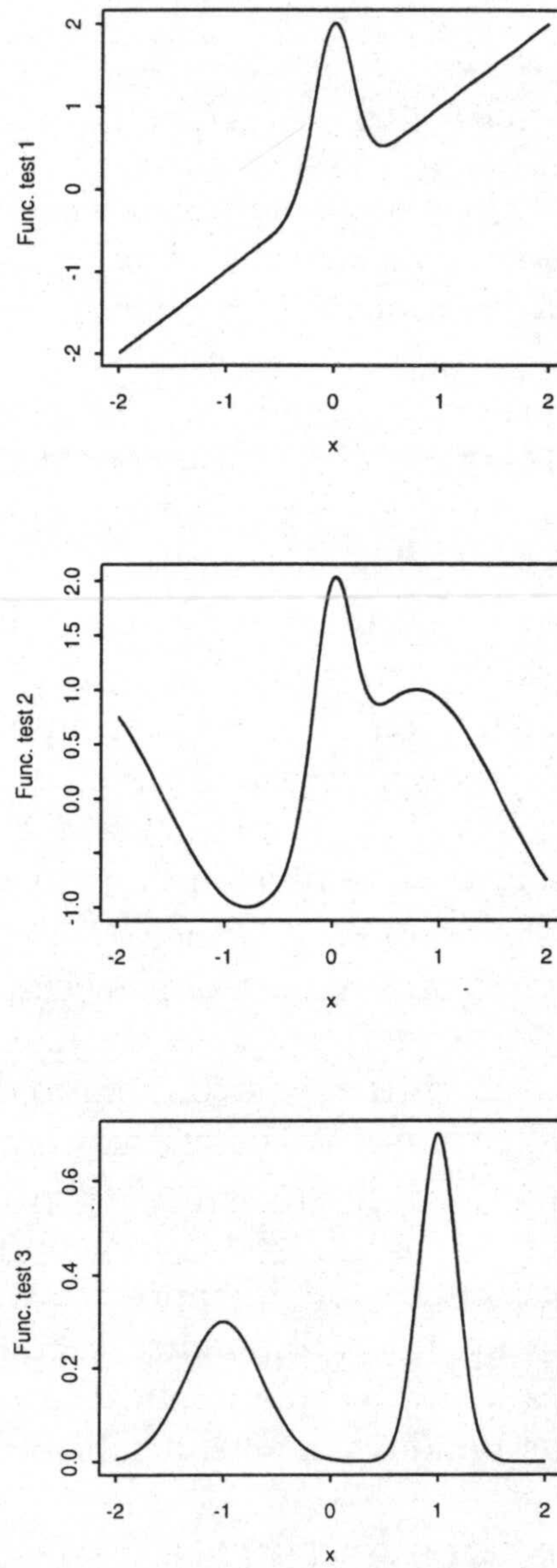


Figura 5.1: Representación de las tres funciones de prueba objeto de la simulación.

de banda propuesto. Los programas realizados se han utilizado para obtener todos los resultados que aparecen en este capítulo, y los incluimos en el apéndice B.

En este apartado comentamos algunos aspectos que han sido tenidos en cuenta en las implementaciones computacionales.

### 5.3.1 Minimización del estimador *bootstrap*

La metodología *bootstrap* propuesta para la selección del ancho de banda, involucra la resolución de un problema de minimización, esto es, el cálculo del ancho de banda,  $h$ , que hace mínima la estimación *bootstrap* propuesta para el error cuadrático medio,  $MSE^*(x; h)$ , en cada punto de estimación,  $x$ . La expresión obtenida para el estimador *bootstrap*, resulta complicada de minimizar en  $h$ , de modo que usar el método de Newton-Raphson para encontrar el mínimo es bastante difícil. Una posibilidad más sencilla, y viable en la práctica es comparar los valores de la función a minimizar en una red de anchos de banda.

En las implementaciones realizadas hemos definido redes de anchos de banda de tipo geométrico, es decir,  $\{h_j = C^j h_{min}, j = 0, 1, \dots\}$ , donde  $h_{min}$ , corresponde al primer punto de la red, y  $C$ , determina el número de puntos de la red, siendo  $[h_{min}, h_{max}]$ , el intervalo donde se realiza la minimización. Hemos hecho las siguientes elecciones:  $h_{min} = (b - a)/n$ ,  $h_{max} = (b - a)/2$ , y  $C = 1.1$ , siendo  $(a, b)$  el intervalo de estimación considerado para la función de regresión, y  $n$  el tamaño muestral.

Según las observaciones de Fan y Gijbels (1995b), sería posible reducir en gran medida el coste computacional si comenzamos evaluando la función a minimizar en los anchos de banda más pequeños y vamos aumentando hasta que el valor de la función crezca consecutivamente un cierto número de veces, evitando calcular el estimador polinomial local para anchos de banda grandes, lo que resulta muy costoso. No obstante, aunque la reducción en tiempo sería muy notable, conllevaría el problema de la posibilidad existente de que la función presente varios mínimos locales, de forma que con esta modificación, siempre se tendería a elegir anchos de banda menores y posiblemente no se correspondiesen con los verdaderos minimizadores. En nuestras aplicaciones hemos evaluado el estimador *bootstrap*,  $MSE^*(x; h)$ , sobre toda la red de anchos de banda, y hemos localizado el mínimo global de la función en ella.

### 5.3.2 Aproximación de las integrales

Las integrales que aparecen en las versiones globales del selector del ancho de banda y en la elección del parámetro piloto, se han implementado usando aproximaciones de Monte Carlo. Una integral genérica del tipo  $\int_a^b g(x)dx$  se puede aproximar de la forma

$$\frac{b-a}{N} \sum_{i=1}^N g(x_i),$$

siendo  $\{x_1, \dots, x_N\}$  una red de puntos en el intervalo  $(a, b)$ .

### 5.3.3 Disminución del coste computacional

Es posible incrementar la rapidez de las implementaciones hechas para los procedimientos aquí presentados. Una posibilidad que no ha sido aplicada y que disminuye notablemente el tiempo de cómputo, es la utilización de algoritmos computacionales rápidos, como las aproximaciones *binning* y *updating* (ver por ejemplo Fan y Marron, 1994).

Los programas que hemos realizado implementan los procedimientos de forma directa, y será objeto de futuros trabajos la incorporación de mejoras en el tratamiento computacional.

## 5.4 Sobre el estimador *bootstrap* del error cuadrático medio

En esta sección estudiamos el comportamiento en muestras finitas del error cuadrático medio teórico,  $MSE(x; h)$ , y de su estimador *bootstrap*,  $MSE^*(x; h)$ . Dado que se trata de medidas locales, que varían con el punto de estimación,  $x$ , es de interés estudiar la proximidad de ambas medidas en cada punto. Hemos estudiado el comportamiento puntual de dichas funciones y hemos seleccionado tres puntos de interés en cada función de prueba considerada, correspondientes a distintos comportamientos locales:

- Función de prueba  $m_1$ . El punto 0, donde la función presenta una fuerte curvatura, el punto -0.5, y el punto 1, donde la función presenta una forma lineal.
- Función de prueba  $m_2$ . Los puntos -1.5, -0.5 y 1.
- Función de prueba  $m_3$ . Los puntos de interés son, en este caso, -0.5, 0.25 y 1.

En las figuras 5.2, 5.3 y 5.4 se representa gráficamente las curvas  $MSE^*(x; h)$  (corresponde a la curva continua) y  $MSE(x; h)$  (curva punteada) para las tres funciones de prueba,  $m_1$ ,  $m_2$  y  $m_3$  respectivamente. En cada figura aparecen representadas dichas funciones para los tres tamaños muestrales considerados, correspondiendo la primera fila al tamaño muestral 50, la segunda al tamaño muestral 200 y la tercera

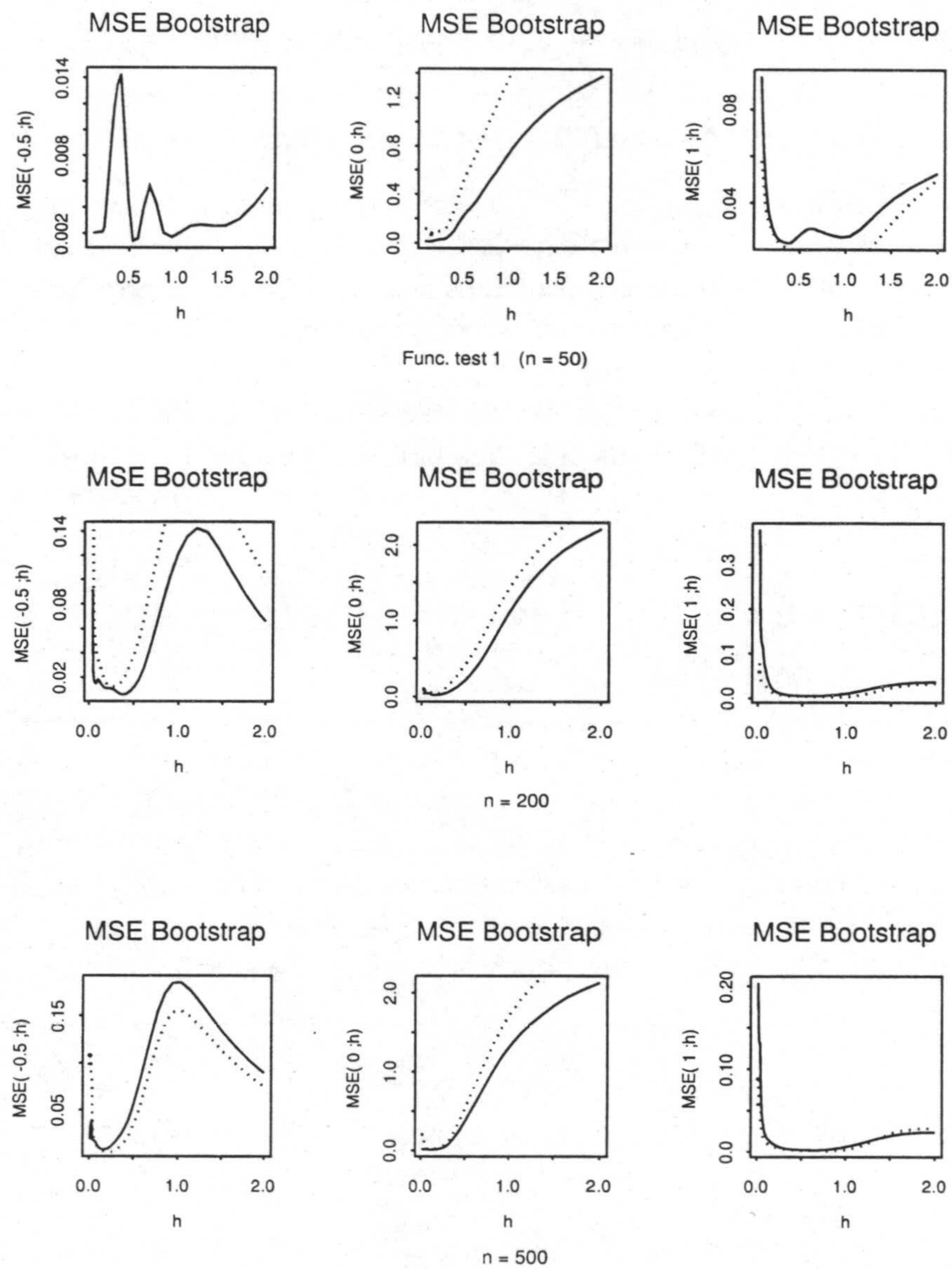


Figura 5.2: Comparación del error cuadrático medio teórico (curva punteada) y *bootstrap* (curva continua), en los puntos de interés definidos para la función de prueba,  $m_1$ . Cada fila corresponde a un tamaño muestral, cada columna a uno de los tres puntos de interés.

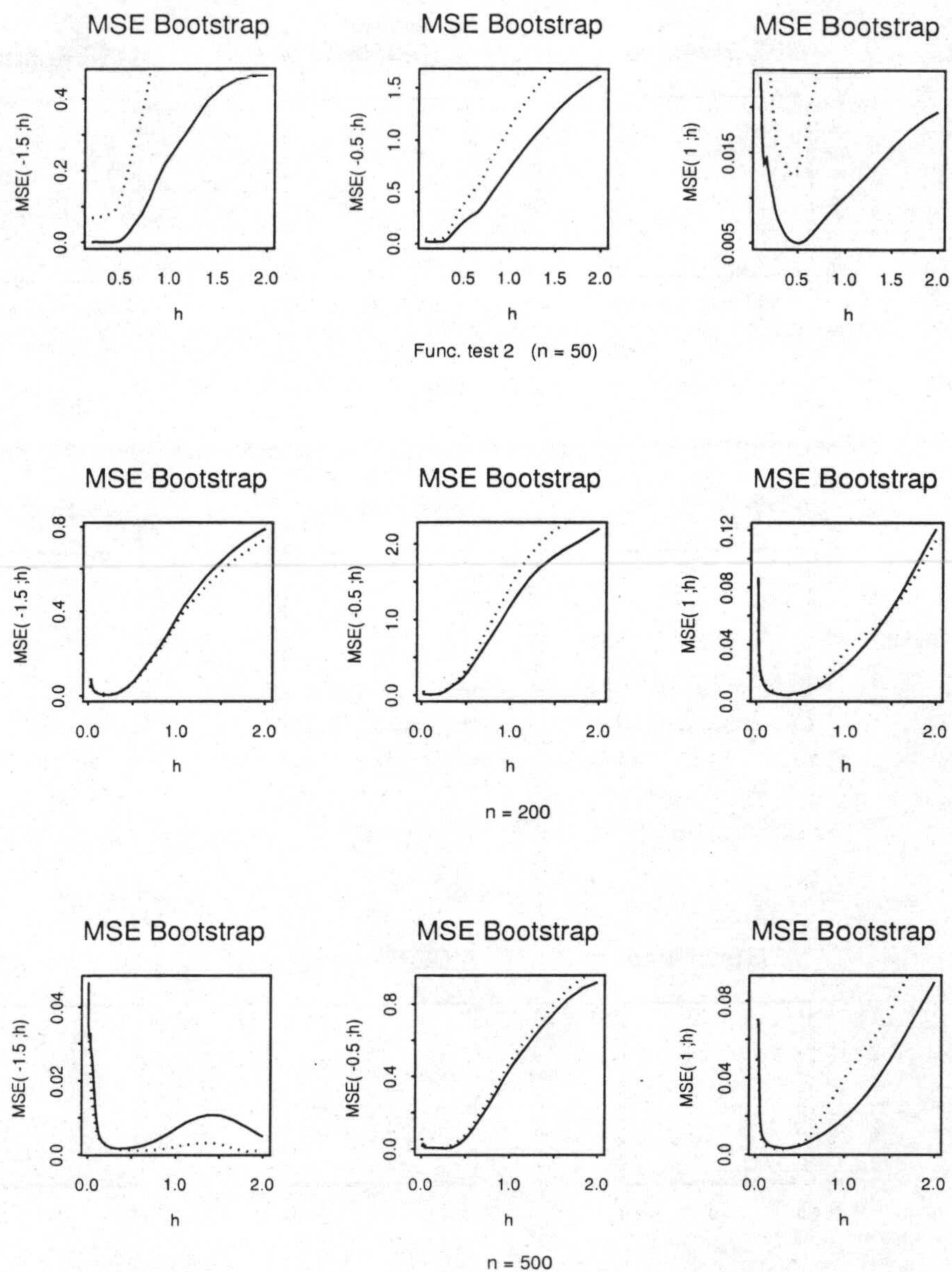


Figura 5.3: Comparación del error cuadrático medio teórico (curva punteada) y *bootstrap* (curva continua), en los puntos de interés definidos para la función de prueba,  $m_2$ . Cada fila corresponde a un tamaño muestral, cada columna a uno de los tres puntos de interés.

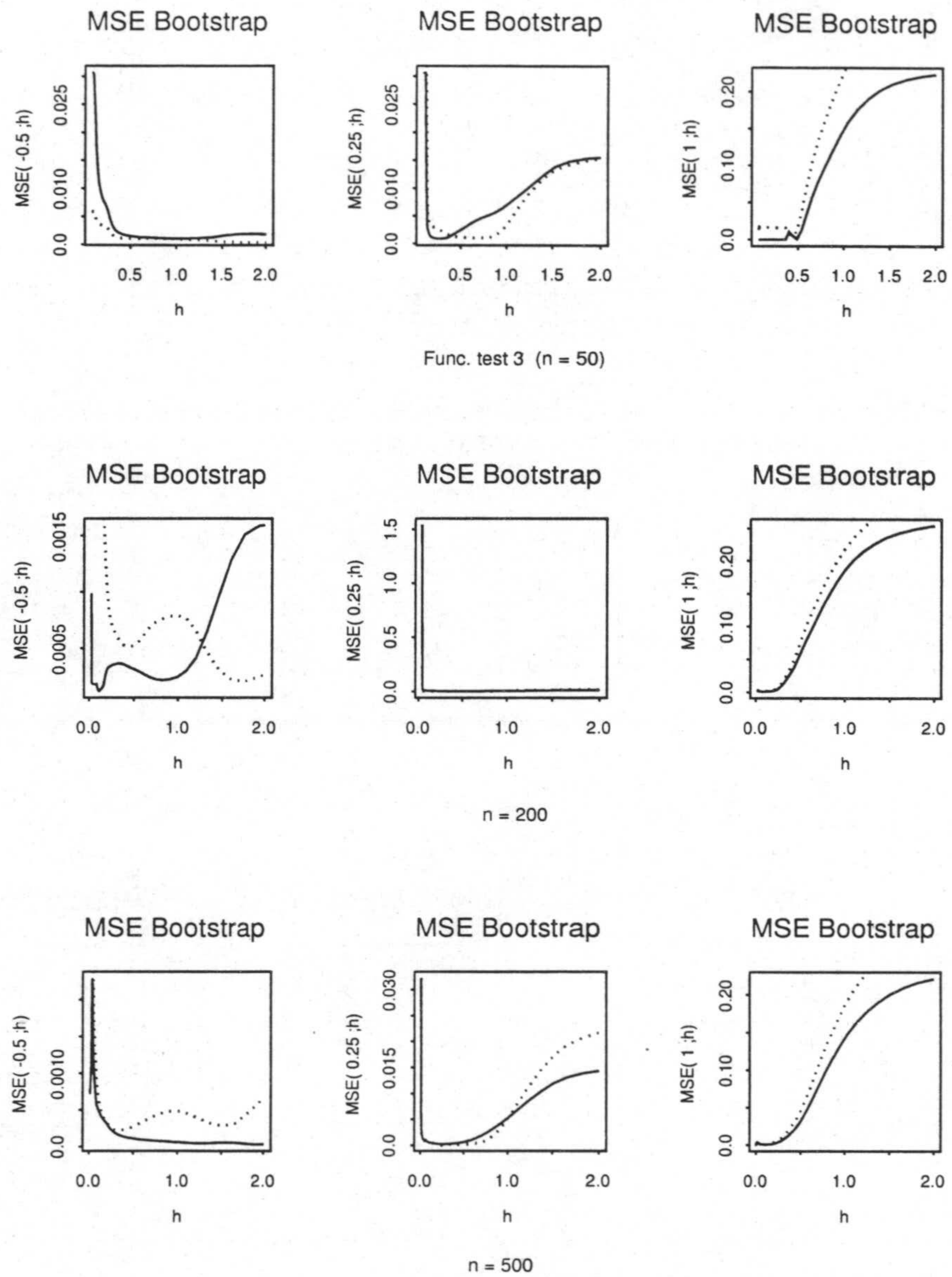


Figura 5.4: Comparación del error cuadrático medio teórico (curva punteada) y *bootstrap* (curva continua), en los puntos de interés definidos para la función de prueba,  $m_3$ . Cada fila corresponde a un tamaño muestral, cada columna a uno de los tres puntos de interés.

fila a muestras de 500 datos. Cada función MSE está evaluada en los puntos de interés antes descritos. En el cálculo de la estimación *bootstrap*,  $MSE^*$ , representada se ha utilizado la elección del ancho de banda piloto propuesta en el capítulo anterior.

Sobre tales gráficos se pueden hacer las siguientes observaciones:

1. Para la primera función de prueba, se observa que las curvas MSE *bootstrap* y MSE teórico, en los puntos 0 y 1, resultan bastante próximas mejorando la calidad de la aproximación a medida que crece el tamaño muestral. Resulta muy notable la dificultad que presenta la estimación en el punto -0.5, para muestras de tamaño 50, la estimación *bootstrap* resulta muy diferente de la teórica, de hecho la curva teórica sale fuera del rango considerado en el gráfico. Sin embargo en cuanto aumenta el tamaño de muestra, el problema se reduce notablemente, consiguiéndose una buena descripción del error cuadrático medio teórico para los tamaños muestrales 200 y 500.
2. Considerando la segunda función de prueba, en los tres puntos de interés localizados se puede observar un comportamiento del estimador *bootstrap*,  $MSE^*$ , parecido al descrito para la primera función de prueba, no observándose puntos conflictivos, como el -0.5. De nuevo resultados similares aparecen para la función de prueba  $m_3$ .

Para contrastar la bondad de nuestra elección del ancho de banda piloto frente a otras posibles elecciones, hemos considerado en los tres puntos de interés de la función de prueba  $m_1$  (el estudio para las otras dos funciones ofrece resultados muy similares) y para los tres tamaños muestrales, el estimador *bootstrap*  $MSE^*$  calculado sobre una red de posibles valores de  $g$ , definida por el conjunto  $\{0.08, 0.16, 0.32, 0.64, 1.28, 2\}$ .

Los gráficos correspondientes se incluyen en las figuras A.1, A.2 y A.3, para los tres puntos de interés y un tamaño muestral de 50; en las figuras A.4, A.5 y A.6, para un tamaño muestral de 200; y en las figuras A.7, A.8 y A.9 para el tamaño 500.

Una primera ojeada a estos gráficos revela algo ya comentado en el capítulo anterior, esto es, la importancia de la elección del parámetro piloto en la estimación *bootstrap*.

Resulta destacable el hecho de que para el punto -0.5, en muestras de tamaño 50, y al igual que ocurría con la elección teórica que hemos propuesto para el parámetro piloto, ningún otro valor de dicho parámetro permite aproximar la función MSE teórica, de hecho en todos los casos se omite la curva teórica pues no cae dentro del rango considerado en el gráfico. No obstante el problema de nuevo carece de importancia en cuanto se consideran muestras de tamaños más razonables: 200 y 500.

Para muestras de tamaño 200, si se observan los gráficos en cada punto, vemos que en el punto 0, entre los valores de  $g$  que mejor aproximan la curva MSE teórico, los valores 0.08 y 0.16, ofrecen resultados muy similares a los obtenidos con nuestra elección de  $g$  (en este caso el valor de  $g$  que resulta de dicha elección 0.1979). Algo similar ocurre en el punto 1, donde algunos valores de  $g$  ofrecen buenos resultados, pero en ningún caso superan a nuestra elección. Sin embargo en el punto de estimación -0.5, el parámetro piloto,  $g$ ,  $g = 0.32$ , ofrece una aproximación al MSE teórico algo mejor que la que ofrecía nuestra elección, lo cual puede ser un indicativo de que es posible mejorarla.

De forma similar, en muestras de tamaño 500, nuestra elección del parámetro piloto trabaja bastante bien frente a otras posibles elecciones, salvo quizás en el punto -0.5, donde se ve superada por el valor 0.16 (frente al que se elegiría según el criterio propuesto que, en tal caso, vale 0.09122).

## 5.5 Estimaciones *bootstrap* del sesgo y de la varianza

En esta sección estudiamos, a través de los ejemplos simulados, la proximidad entre las componentes del error cuadrático medio teórico y sus estimaciones *bootstrap*. De nuevo, al ser medidas locales, las comparamos para los puntos de interés antes señalados.

Hemos realizado el estudio para las tres funciones de prueba consideradas, no obstante los resultados son parecidos, por lo que tan sólo incluimos y comentamos los resultados para la primera función de prueba. Los gráficos aparecen en las figuras 5.5, 5.6 y 5.7, correspondientes a los puntos de interés, -0.5, 0 y 1, respectivamente. Cada una de estas figuras presenta los gráficos del sesgo *bootstrap* y la varianza *bootstrap* (líneas continuas) y el sesgo teórico y la varianza teórica (líneas punteadas), y cada fila corresponde a uno de los tres tamaños muestrales considerados.

Es de notable interés observar los gráficos correspondientes al punto -0.5, dado que en la sección anterior veíamos que en dicho punto y para el tamaño muestral 50, la estimación *bootstrap*,  $MSE^*$ , suponía una pésima aproximación al MSE teórico. En este caso es de interés detectar en qué componente radica el problema. Observando la figura 5.5 se advierte que para el tamaño muestral 50, tanto la componente de sesgo como la componente de varianza suponen una pésima aproximación de las correspondientes teóricas, y si observamos los siguientes tamaños muestrales, la componente *bootstrap* de varianza y la teórica resultan casi idénticas; y en cuanto a la componente de sesgo, la aproximación resulta bastante buena, mejorando para el tamaño muestral 500.

Si se observan las gráficas para los otros dos puntos de interés, se comprueba que la bondad de las aproximaciones es mayor a medida que aumenta el tamaño



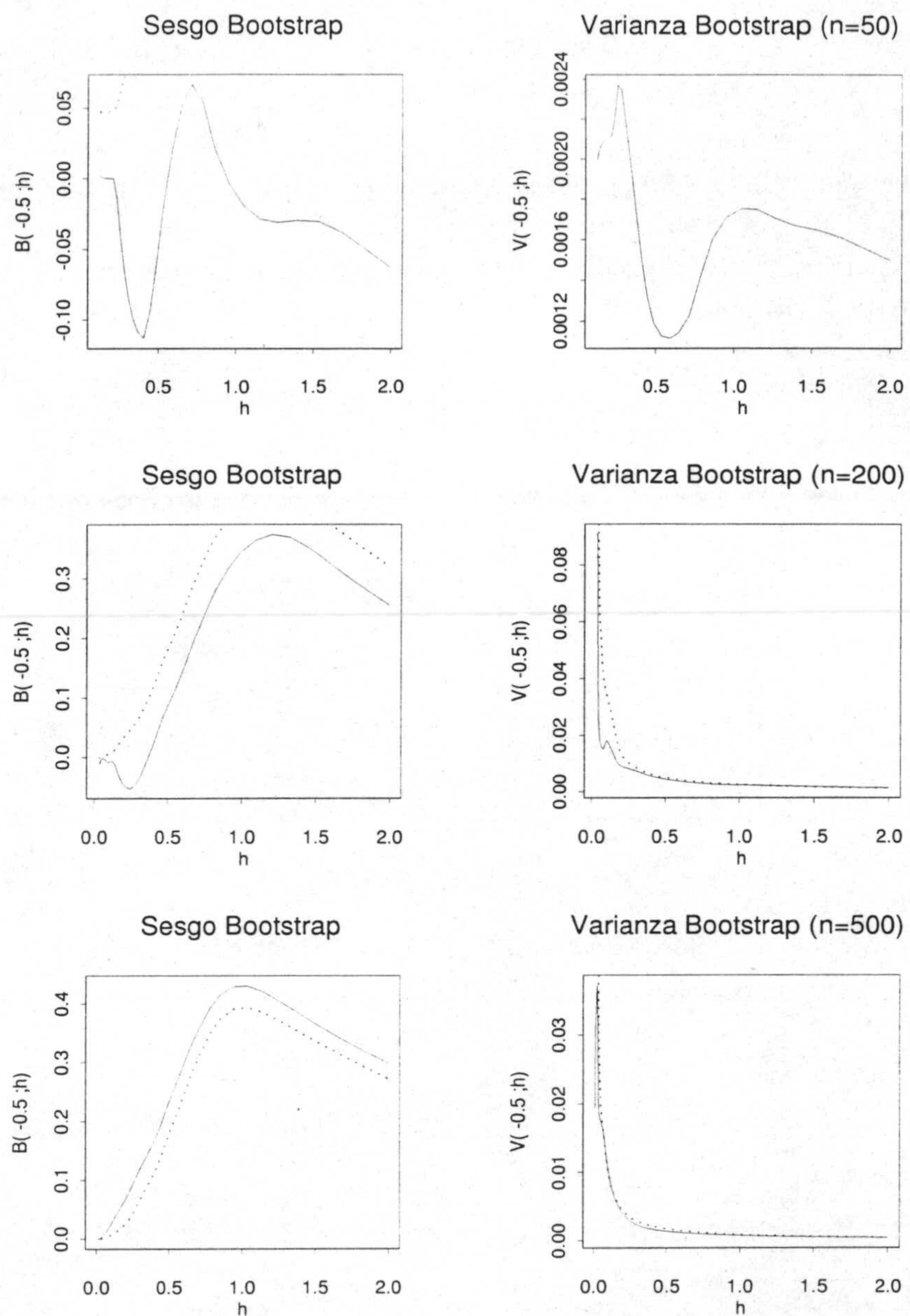


Figura 5.5: Sesgo y varianza teóricos (curvas punteadas) y estimaciones *bootstrap* (curvas continuas), utilizando la elección teórica del parámetro piloto. Las curvas corresponden al punto,  $x$ ,  $x = -0.5$ , y cada fila a un tamaño muestral: 50, 200 y 500.

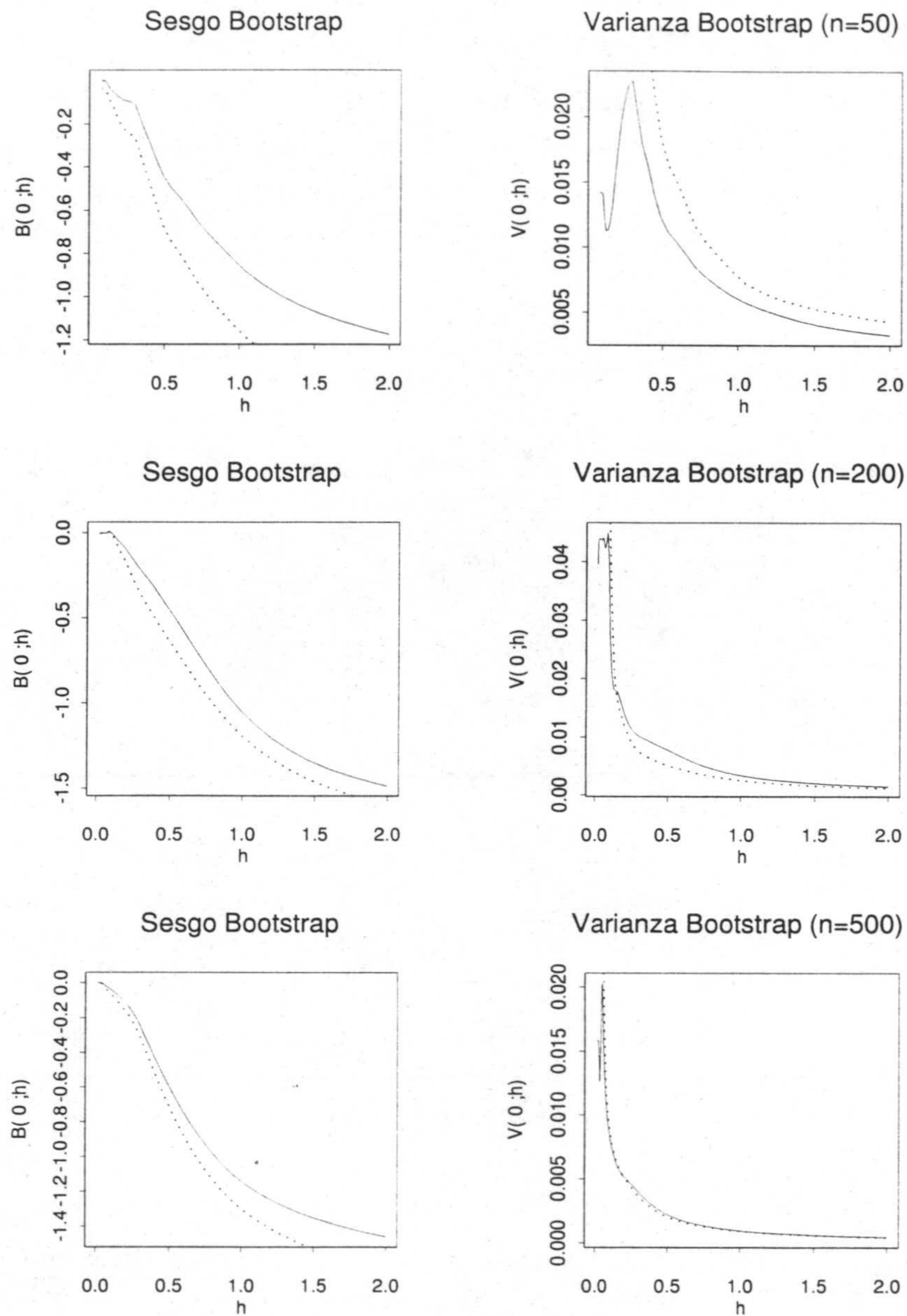


Figura 5.6: Sesgo y varianza teóricos (curvas punteadas) y estimaciones *bootstrap* (curvas continuas), utilizando la elección teórica del parámetro piloto. Las curvas corresponden al punto,  $x$ ,  $x = 0$ , y cada fila a un tamaño muestral: 50, 200 y 500.

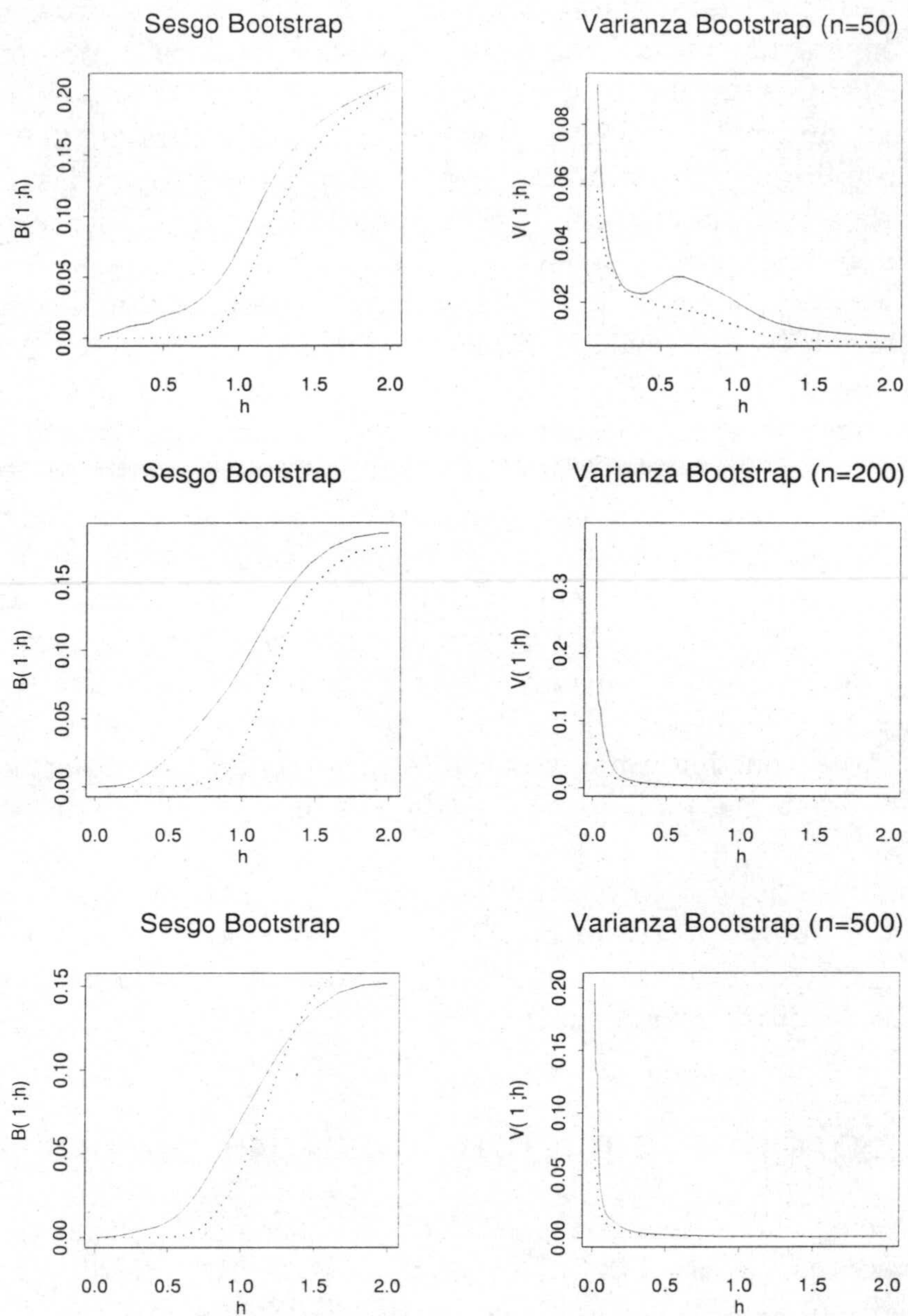


Figura 5.7: Sesgo y varianza teóricos (curvas punteadas) y estimaciones *bootstrap* (curvas continuas), utilizando la elección teórica del parámetro piloto. Las curvas corresponden al punto,  $x = 1$ , y cada fila a un tamaño muestral: 50, 200 y 500.

muestral; y para el tamaño muestral de 50, se observa un peor comportamiento, si bien no tan notable como en el punto  $x$ ,  $x = -0.5$ , antes señalado.

En todos los casos resulta notable la bondad de la aproximación *bootstrap* de la componente de varianza, que para tamaños muestrales no demasiado pequeños describe casi perfectamente la teórica.

Hemos contrastado además para el tamaño muestral 50, si otras elecciones del parámetro piloto,  $g$ , mejoran las aproximaciones *bootstrap* de las componentes. Concretamente resultaría interesante comprobarlo en la componente de varianza, puesto que en la elección propuesta en el capítulo anterior obviamos el criterio de proximidad en dicha componente, si bien era debido a que para tamaños muestrales grandes era despreciable (aparecía con el factor  $(nh)^{-1}$ ).

La figura 5.8 contiene los gráficos de las componentes de sesgo y varianza *bootstrap*, así como del error cuadrático medio, en el punto de estimación  $-0.5$ . Cada uno contiene la representación gráfica de tres curvas, la curva continua corresponde a dichas funciones usando uno de los valores de  $g$  que se están contrastando ( $g = 0.16, 0.32$  y  $0.64$ ), la curva punteada corresponde a las funciones teóricas; y la curva discontinua corresponde a las que se tenían con la elección propuesta de  $g$ . Obsérvese que el valor  $0.32$ , mejora algo la estimación *bootstrap* de la componente de varianza, por lo que se podría sugerir que para tamaños muestrales pequeños se tomara en cuenta la componente de varianza para la elección del parámetro piloto. No obstante no se pueden esperar notables mejorías en el comportamiento, además el problema radica más en la escasez de datos que en el criterio de estimación *bootstrap* propuesto.

Las figuras 5.9 y 5.10 ofrecen los mismos gráficos pero para los tamaños muestrales 200 y 500 respectivamente. En estos casos se observa que ningún valor de  $g$  mejora la aproximación *bootstrap* que involucra la elección propuesta en el capítulo anterior para dicho parámetro.

## 5.6 Sobre la elección local del parámetro piloto

En la sección 4.6.5 definíamos la elección óptima del ancho de banda piloto,  $g$ , en sus versiones local, (4.23), y global, (4.24). Del estudio del error cuadrático medio *bootstrap* y de sus componentes, realizado en el apartado anterior, para cada punto, resulta notable el hecho de que dicho parámetro determina la bondad de las estimaciones *bootstrap* propuestas en esta memoria.

La figura 5.11 representa, para un tamaño muestral de 200, la elección teórica local del parámetro piloto, para las funciones de prueba,  $m_1$ ,  $m_2$  y  $m_3$ .

Dicha gráfica revela la conveniencia de una elección local frente a una global, para el parámetro piloto, dado que las curvas no describen, en absoluto, formas constantes en el intervalo de estimación considerado, para ninguna de las funciones

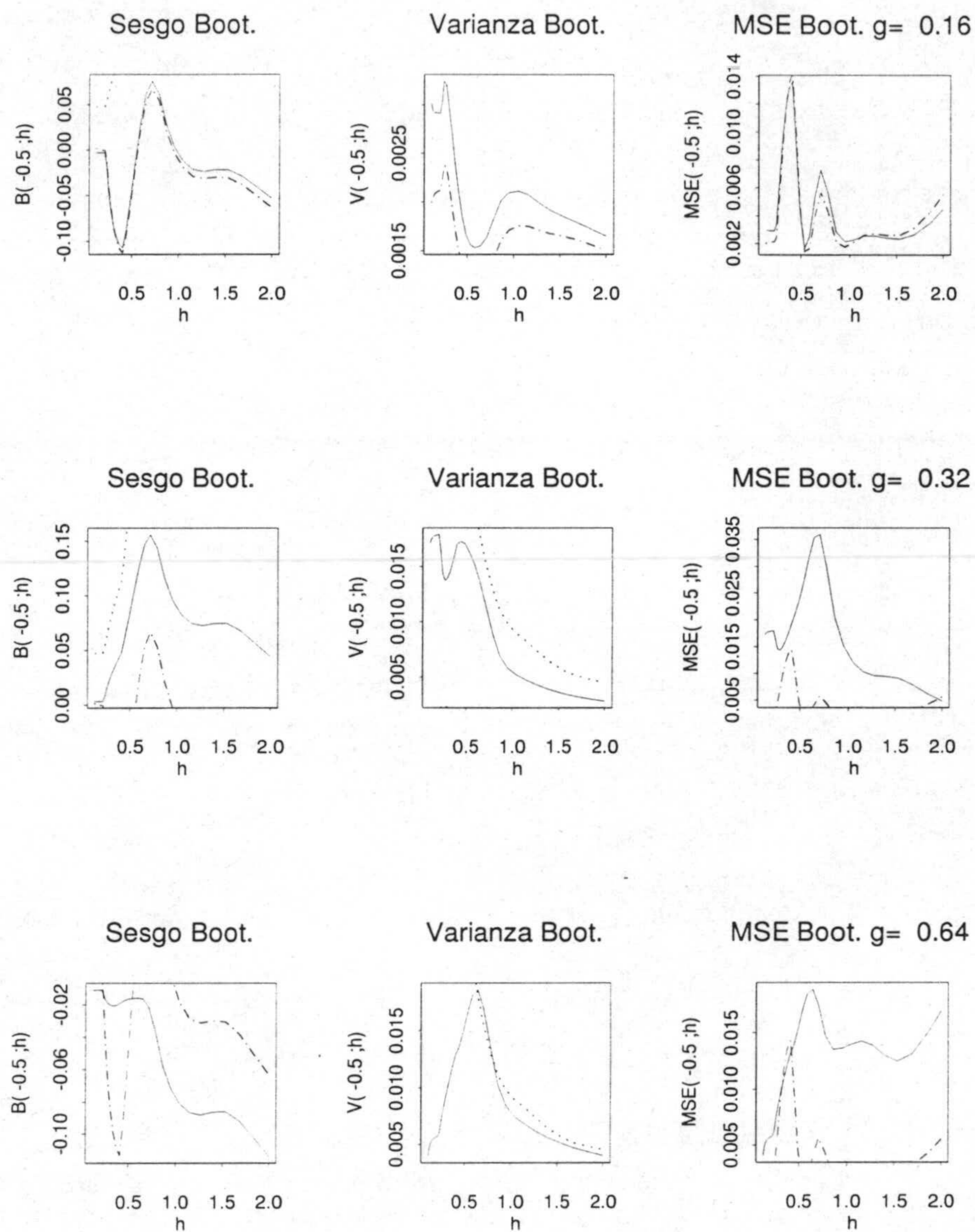


Figura 5.8: Sesgo y varianza teóricos (curvas punteadas); estimaciones *bootstrap* usando la elección teórica del parámetro piloto (curvas discontinuas), y utilizando otros posibles valores de dicho parámetro (curvas continuas). Las curvas corresponden al punto,  $x, x = -0.5$ , y a un tamaño muestral de 50.

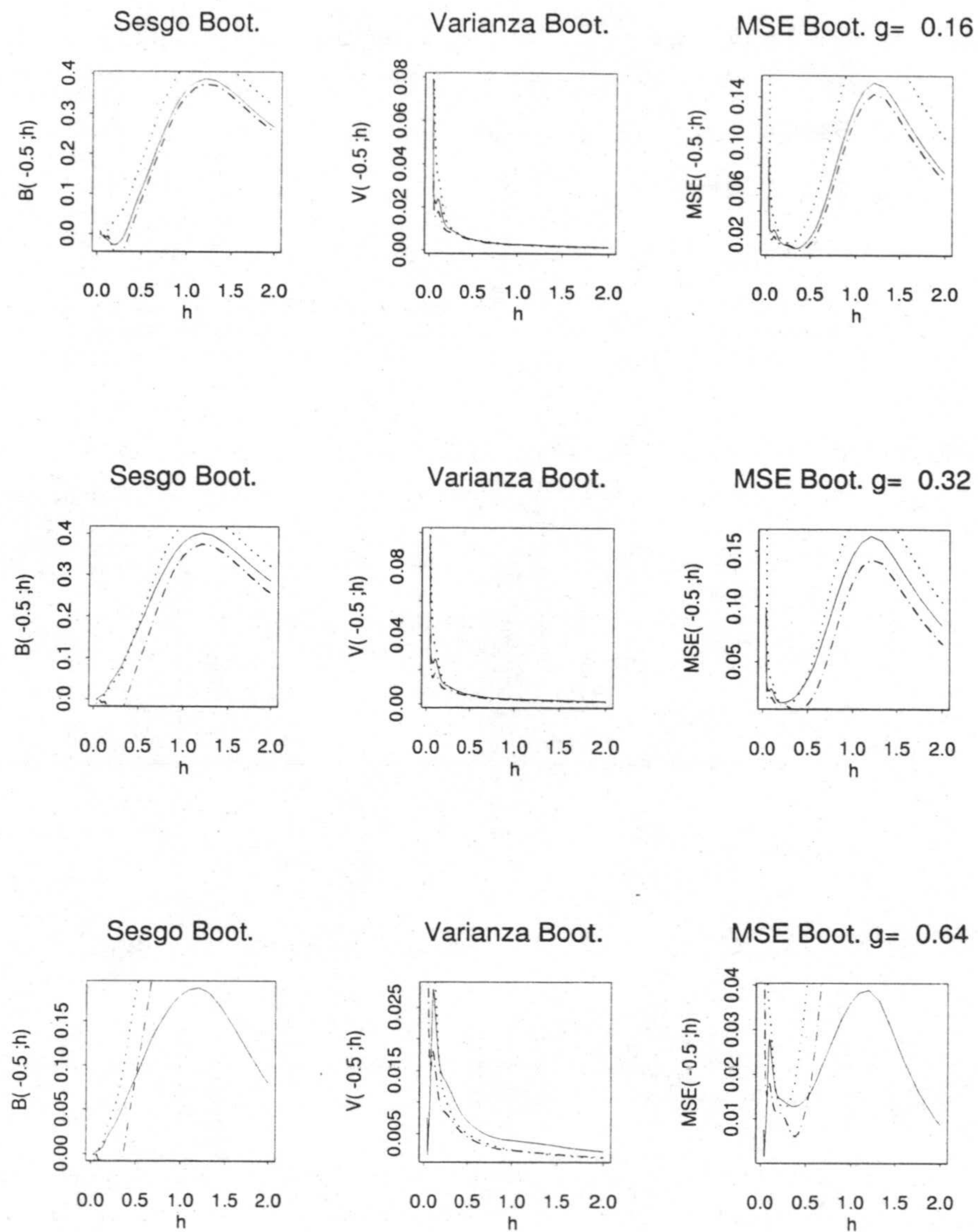


Figura 5.9: Sesgo y varianza teóricos (curvas punteadas); estimaciones *bootstrap* usando la elección teórica del parámetro piloto (curvas discontinuas), y utilizando otros posibles valores de dicho parámetro (curvas continuas). Las curvas corresponden al punto,  $x = -0.5$ , y a un tamaño muestral de 200.

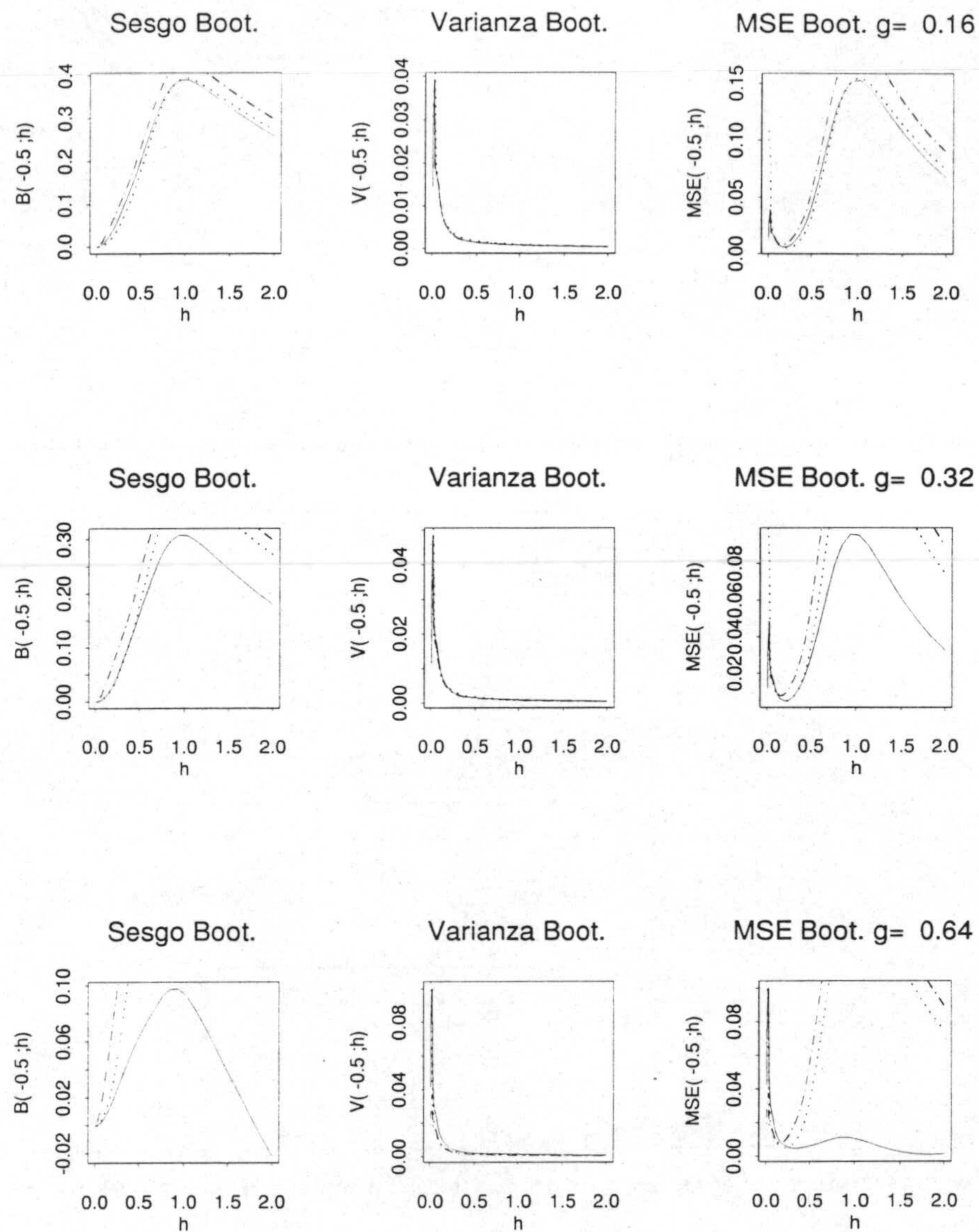


Figura 5.10: Sesgo y varianza teóricos (curvas punteadas); estimaciones *bootstrap* usando la elección teórica del parámetro piloto (curvas discontinuas), y utilizando otros posibles valores de dicho parámetro (curvas continuas). Las curvas corresponden al punto,  $x, x = -0.5$ , y a un tamaño muestral de 500.

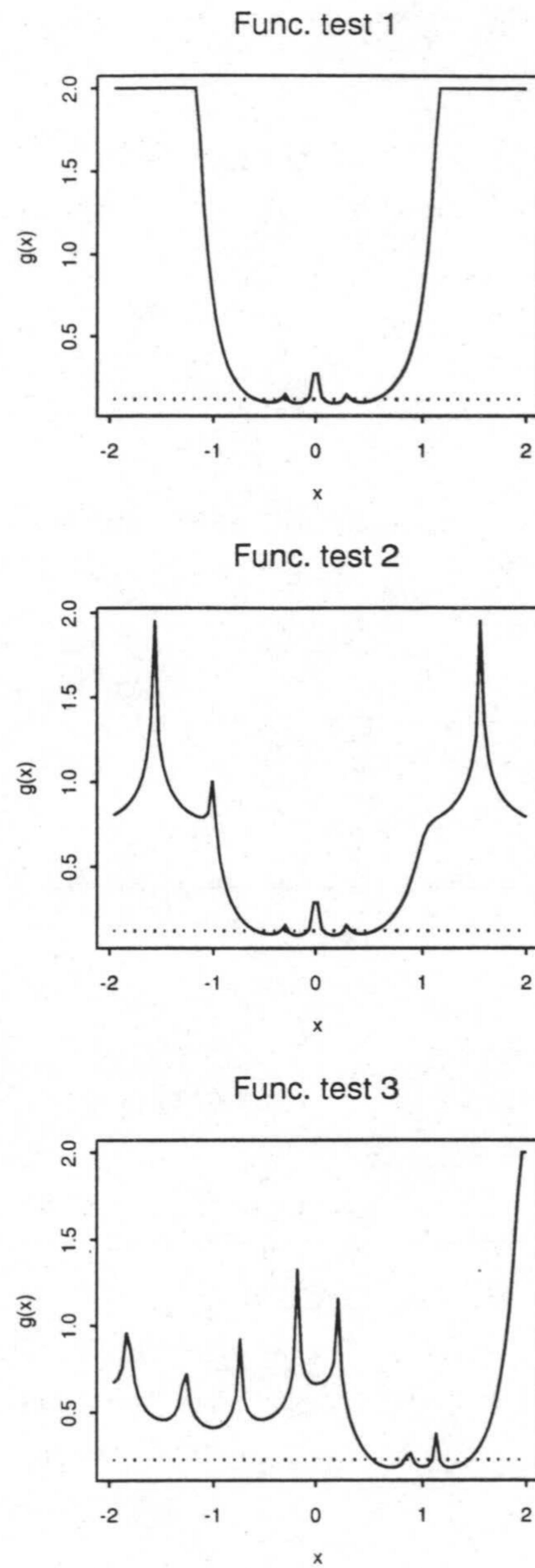


Figura 5.11: Representación gráfica del ancho de banda piloto óptimo para las tres funciones de prueba. El tamaño muestral considerado es 200.



de prueba. Se puede observar además que la elección global corresponde a un valor muy bajo en todos los casos, de modo que si volvemos a gráficos puntuales como los de la figura A.6, se observa que en puntos donde la función presenta formas casi lineales, los anchos piloto de magnitud pequeña repercuten en un empeoramiento de la estimación *bootstrap* del error cuadrático medio, de hecho, en tales puntos la elección óptima es bastante grande, apareciendo truncada en los gráficos y en los cálculos, por la cota superior,  $(b - a)/2$ .

## 5.7 El estimador polinomial local con ancho de banda *bootstrap*

Se han calculado las curvas de regresión estimadas, según el estimador polinomial local de grado 1, para cada una de las réplicas simuladas de las tres funciones de prueba. En cada estimación se ha utilizado el ancho de banda local seleccionado según el procedimiento *wild bootstrap* sobre los residuos, descrito en el capítulo 4, utilizando la elección local óptima del parámetro piloto.

La figura 5.12 representa tres de estas curvas representativas, junto con la curva de regresión teórica (curva continua), para el tamaño muestral intermedio de 200. Las curvas representativas han sido seleccionadas de la forma siguiente: Se han obtenido las 20 curvas de regresión estimadas usando los datos simulados, según se describe al comienzo de este capítulo, y se han ordenado según la magnitud de la suma residual de cuadrados correspondiente. Las curvas representativas que se han seleccionado han sido las correspondientes al 10%, 50% y 95%. Además de estas curvas representativas se han incluido gráficos del ancho de banda local *bootstrap* empleado en una de ellas (la correspondiente al 10%).

Si se analizan los gráficos correspondientes a las curvas estimadas, resulta notable la buena descripción de la curva de regresión teórica para las dos primeras funciones de prueba. La variabilidad de dichas estimaciones, no obstante, resulta algo mayor en la zona central del intervalo  $(0,1)$ , donde las funciones de prueba presentan una fuerte curvatura. De hecho, la función ancho de banda correspondiente a las estimaciones de la primera función de prueba, adopta valores pequeños en dicha zona buscando una mejor descripción de lo observado, esto es, reduciendo sesgos con un consecuente aumento de variabilidad.

Obsérvese además que el selector *bootstrap* del ancho de banda ofrece valores mayores de dicho parámetro para describir zonas donde la estructura subyacente tiende más a la linealidad.

Para la tercera función de prueba se observa una mayor variabilidad de las estimaciones obtenidas, sobre todo en el intervalo  $(-0.5,0.5)$ , donde la curva subyacente presenta un valle. En vista de lo obtenido, para esta función resultará más conveniente emplear ajustes polinomiales locales de grado dos o tres, en lugar de los de

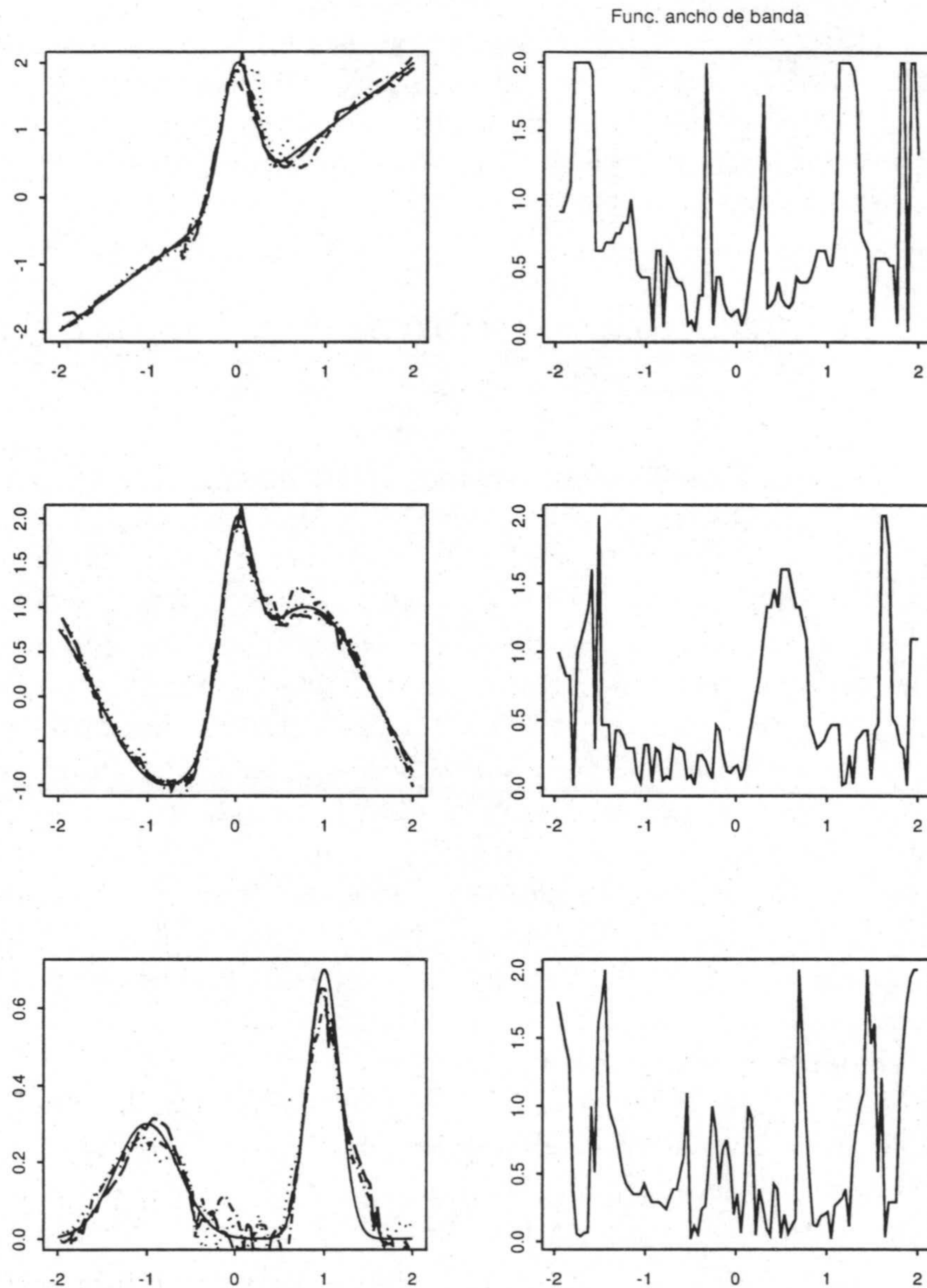


Figura 5.12: Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap*. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap*. Cada fila corresponde a una función de prueba. El tamaño muestral considerado es de 200.

tipo lineal.

Finalmente una ojeada a los gráficos de la función ancho de banda en las tres funciones de prueba revela la posible conveniencia de un suavizamiento de la misma, basado en el hecho de que puntos próximos deberían describir valores parecidos del ancho de banda local. Además se puede observar que, en zonas próximas a la frontera, los cambios bruscos en los valores se podrían resolver, en alguna medida, realizando este suavizado. Esta idea se desarrolla en el siguiente apartado, no obstante adelantamos que producen mejoras en las estimaciones.

## 5.8 Conclusiones

El estudio de simulación realizado revela un buen comportamiento del selector *bootstrap* del ancho de banda. De hecho la estimación según *wild bootstrap* sobre los residuos, propuesta para el error cuadrático medio, ofrece resultados bastante satisfactorios para tamaños muestrales moderados (200).

En cuanto a las curvas estimadas según el estimador polinomial local, involucrando el ancho de banda local *bootstrap*, describen bastante bien la estructura subyacente, salvo algunos problemas de excesiva variabilidad en zonas de fuerte curvatura. Algunas de estas estimaciones pueden mejorarse en gran medida empleando ajustes polinomiales de grado superior a uno (que es el único con el que hemos trabajado en las simulaciones) sin añadir demasiada complejidad al problema. Extensiones y mejoras de este tipo serán objeto de futuros trabajos.

A continuación y para concluir este capítulo definimos una versión suavizada del selector *bootstrap* del ancho de banda, e investigamos su comportamiento a través de las simulaciones aquí descritas.

### 5.8.1 Versión suavizada del selector *bootstrap*

Dado que la función ancho de banda *bootstrap* presenta una forma muy irregular, proponemos una versión suavizada de la misma utilizando una función núcleo de tipo triangular<sup>2</sup>.

Se define, por tanto, la función ancho de banda *bootstrap* suavizada como:

$$\tilde{h}_{\text{boot}}(x) = \sum_{l=1}^{n_{\text{grid}}} W(x_l; x) h_{\text{boot}}(x_l), \quad (5.1)$$

donde  $\{x_1, \dots, x_{n_{\text{grid}}}\}$ , es la red de puntos utilizada en el cálculo del ancho de banda local; y los pesos,  $W(x_l; x)$ , se definen como:

<sup>2</sup>Es conveniente usar esta función de pesos dado que asigna ponderaciones que dan lugar a valores que decrecen linealmente con la distancia al punto de estimación.

$$W(x_l; x) = d^{-1} K \left( \frac{x_l - x}{d} \right)$$

siendo  $K$ , la función núcleo triangular definida como

$$K(u) = 1 - |u| 1_{\{|u| \leq 1\}},$$

y  $d$ , el parámetro de suavizado, definido por  $d = \Delta \times span$ ; para un parámetro  $span$  que controla el número de puntos que contiene cada entorno local definido en el suavizado; y donde  $\Delta$  es la distancia entre puntos adyacentes en la red de  $ngrid$  puntos que se ha considerado.

En puntos próximos a la frontera es posible que la función ancho de banda sea muy inestable, según se podía apreciar en la figura 5.12, por lo que se considerará constante. Por tanto se calculan las versiones suavizadas  $\tilde{h}_{boot}(x_l)$  para  $l = span + 1, \dots, ngrid - span$ ; y se define en la frontera como:

$$\tilde{h}_{boot}(x_l) = \tilde{h}_{boot}(x_{span+1}) \quad \text{para } l = 1, \dots, span, \text{ y}$$

$$\tilde{h}_{boot}(x_l) = \tilde{h}_{boot}(x_{ngrid-span}) \quad \text{para } l = ngrid - span + 1, \dots, ngrid.$$

Se han calculado las estimaciones de las curvas de regresión simuladas, usando esta versión suavizada del ancho de banda local *bootstrap* y para los valores de  $span$ : 2, 3, 4, 5, y 10.

La figura 5.13 presenta las gráficas obtenidas utilizando esta versión suavizada del selector *bootstrap*, para un valor de  $span$  igual a 5. En ella se puede observar que las estimaciones de la curva obtenidas en el caso de las dos primeras funciones de prueba son muy próximas a la verdadera función simulada, corrigiéndose en gran medida la variabilidad existente en las estimaciones representadas en la figura 5.12 correspondiente al ancho de banda *bootstrap* sin suavizar. Se puede ver por ejemplo, que las curvas representativas correspondientes a  $m_2$ , aproximan mucho mejor dicha función en el intervalo (0,1).

En cuanto a la función de prueba  $m_3$ , el suavizado del ancho de banda con un valor de  $span$  igual a 5, corrige en gran medida la variabilidad resultante, no obstante la descripción de la estructura verdadera no es tan satisfactoria como en el caso de las otras dos funciones. En este caso concreto y como ya se comentaba antes, será deseable elevar el grado de los ajustes polinomiales locales a dos o tres.

En el apéndice A se recogen los gráficos correspondientes a otros niveles de suavizado en el ancho de banda *bootstrap*: Figuras A.10, A.11, A.12 y A.13.

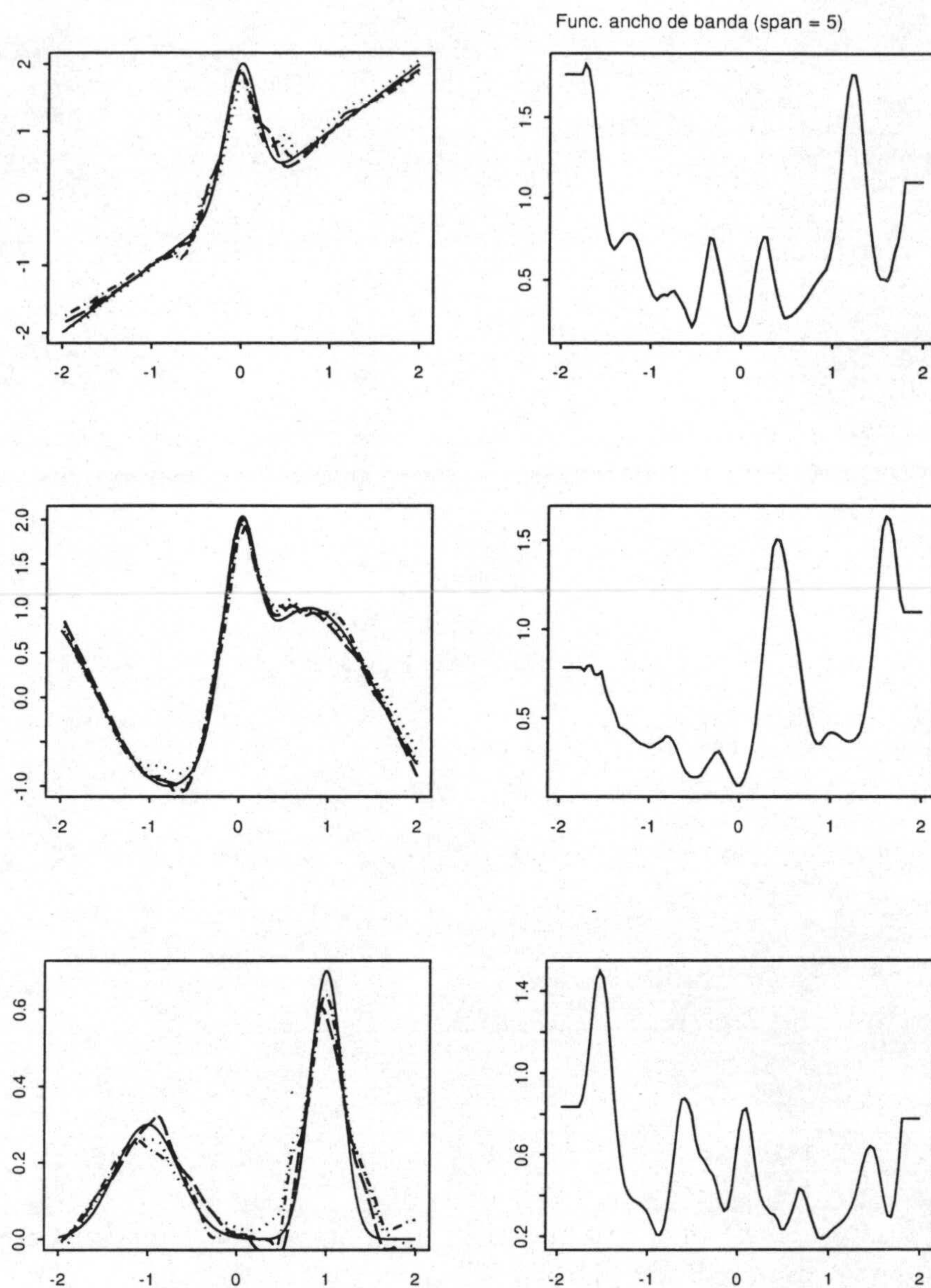


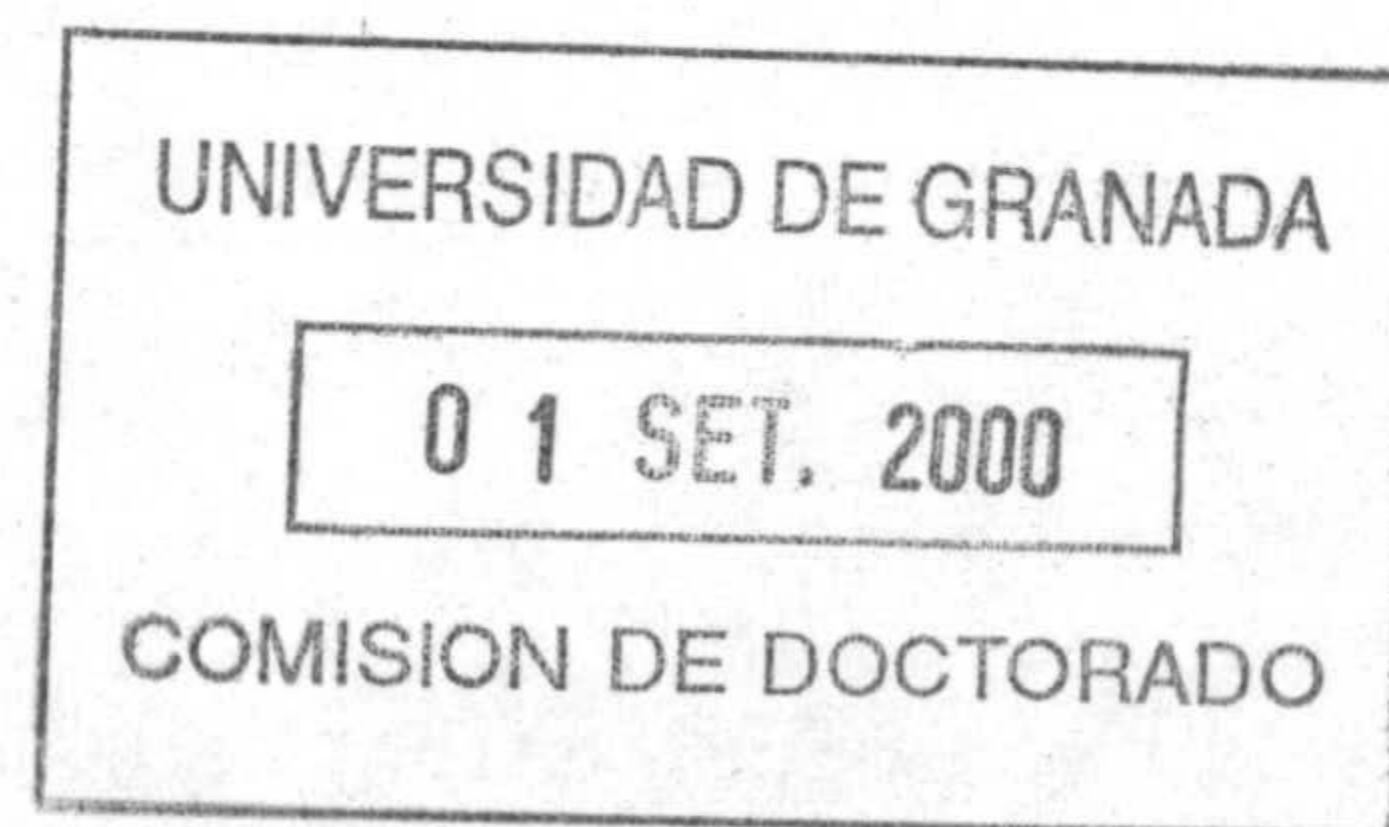
Figura 5.13: Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizado con  $span = 5$ . El tamaño muestral considerado es 200.



# Apéndice A

## Gráficos adicionales

Se incluyen a continuación los gráficos que complementan el estudio de simulación realizado en el capítulo 5.



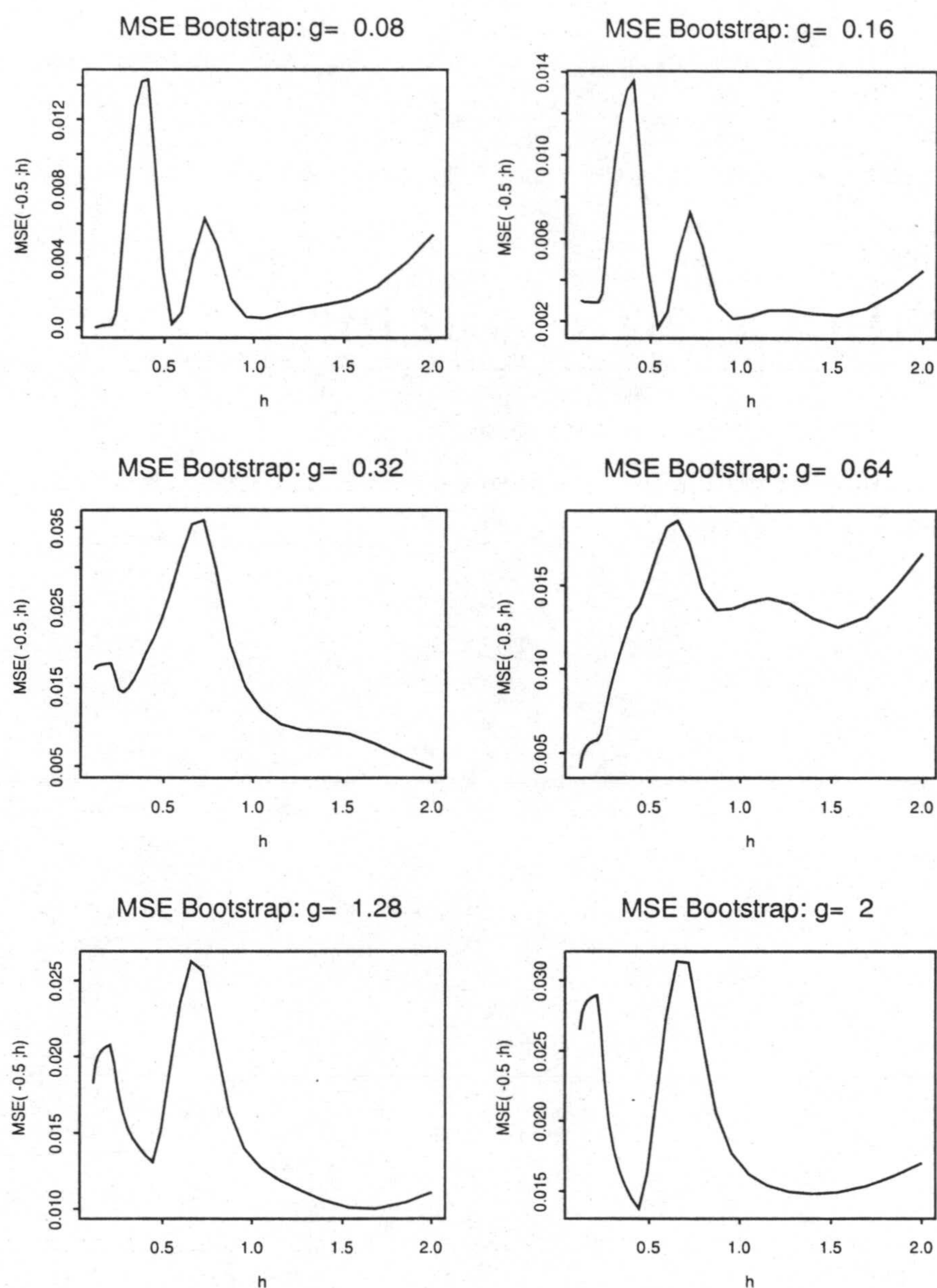


Figura A.1: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden al punto  $x$ ,  $x = -0.5$ , y al tamaño muestral 50. (Sólo aparece la curva *bootstrap* puesto que la curva teórica sale fuera del rango considerado.)



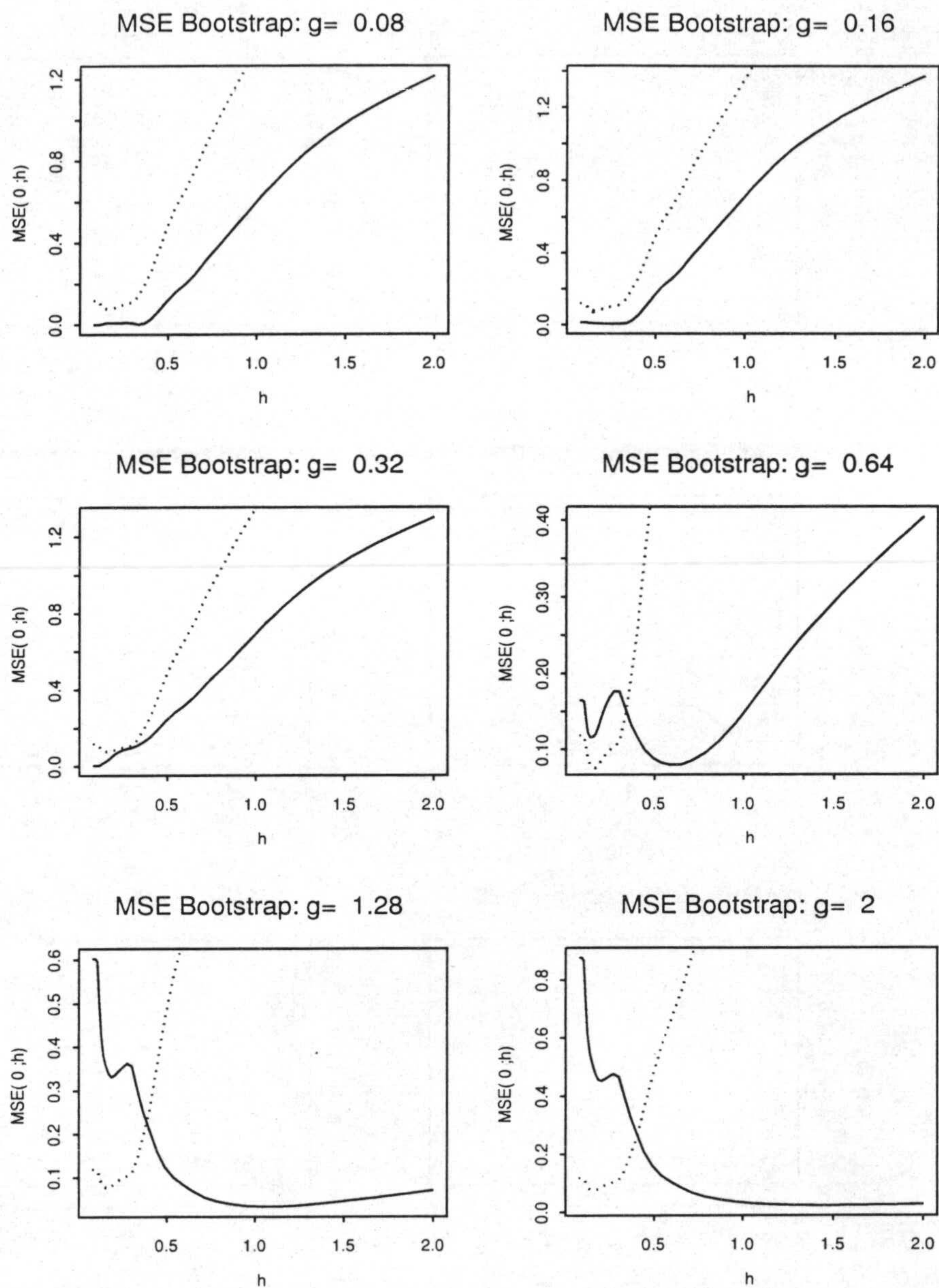


Figura A.2: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden al punto,  $x, x = 0$ , y al tamaño muestral 50.

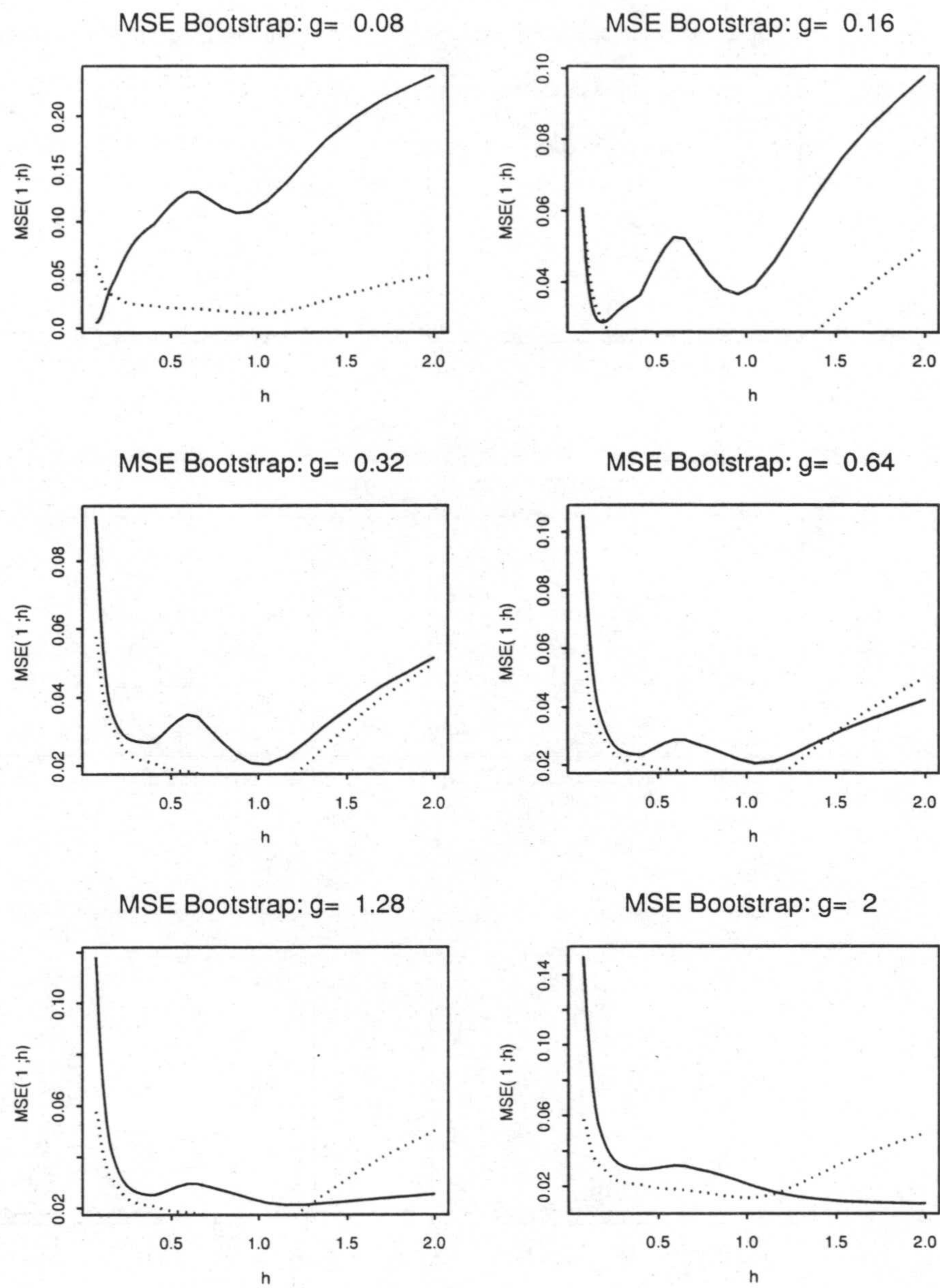


Figura A.3: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = 1$ ; y al tamaño muestral 50.

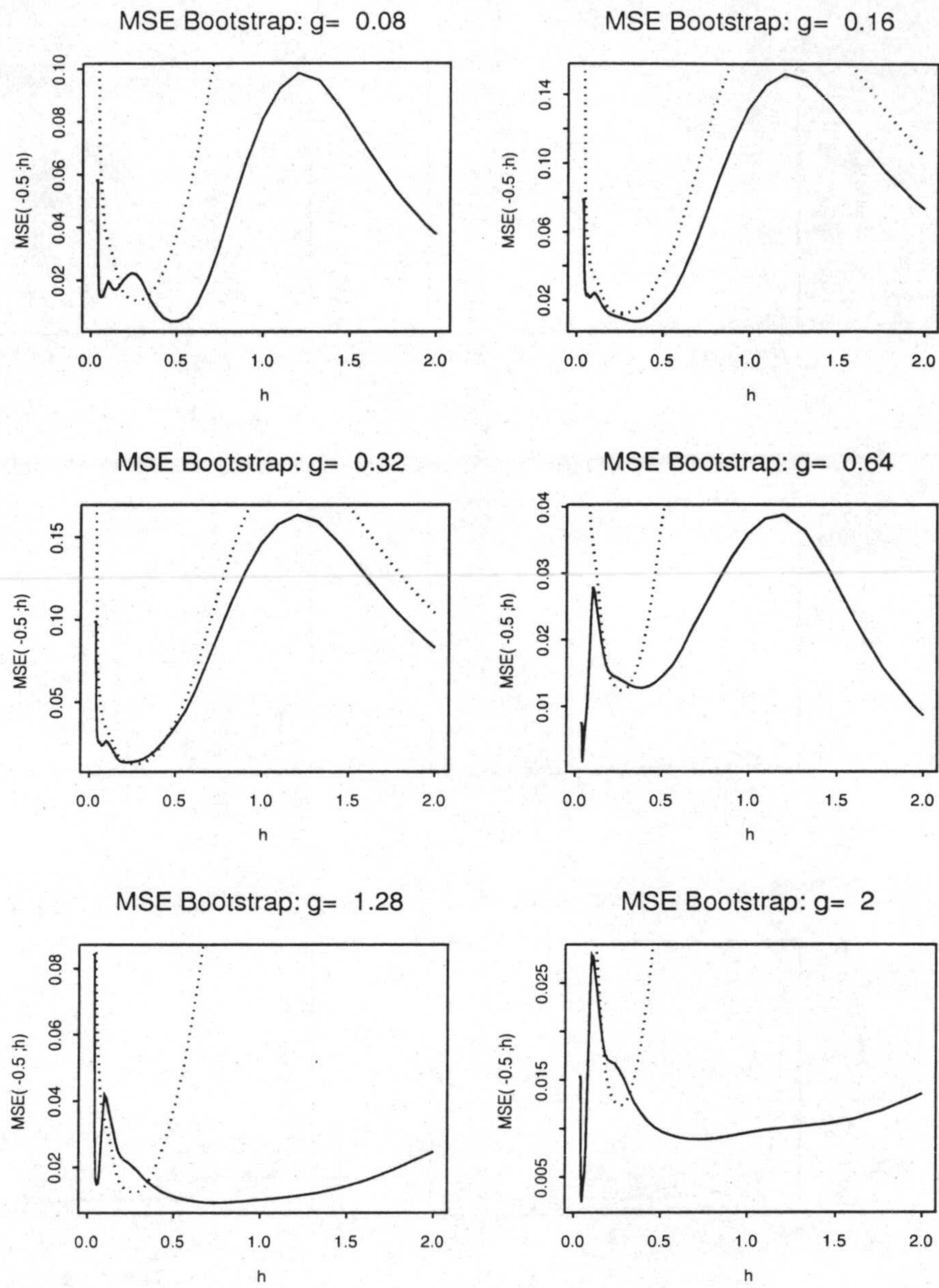


Figura A.4: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = -0.5$ ; y al tamaño muestral 200.

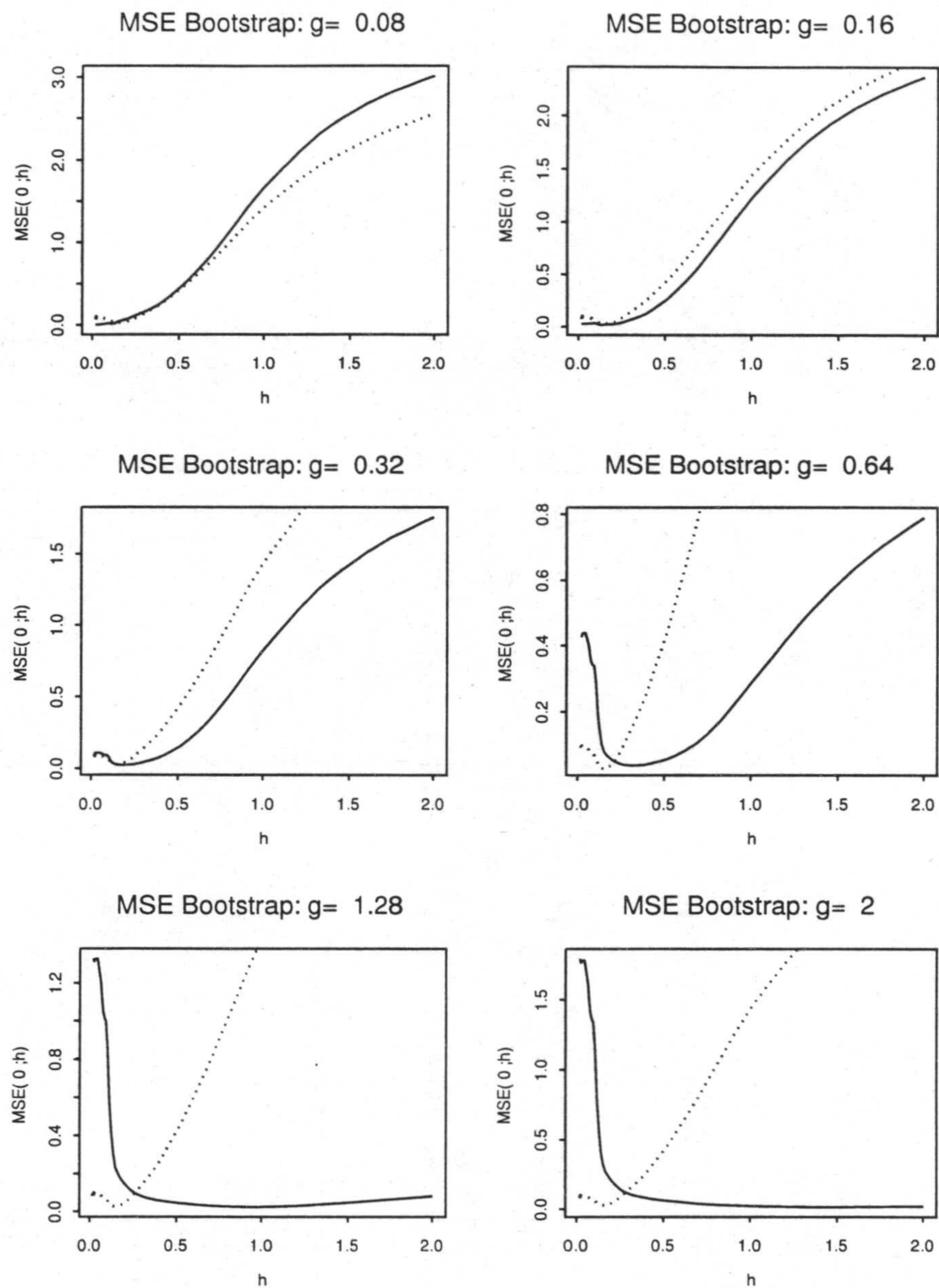


Figura A.5: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = 0$ ; y al tamaño muestral 200.

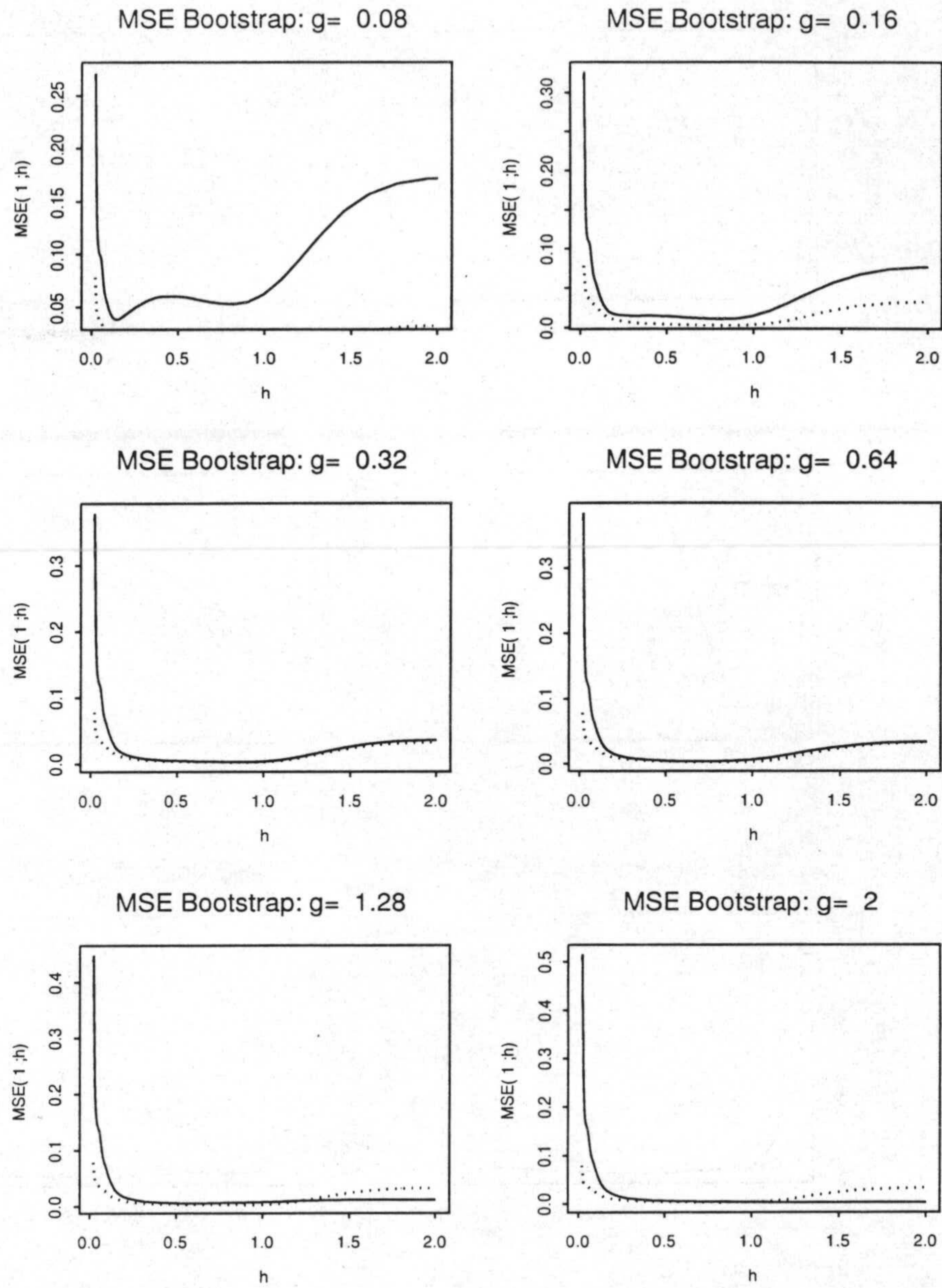


Figura A.6: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = 1$ ; y al tamaño muestral 200.

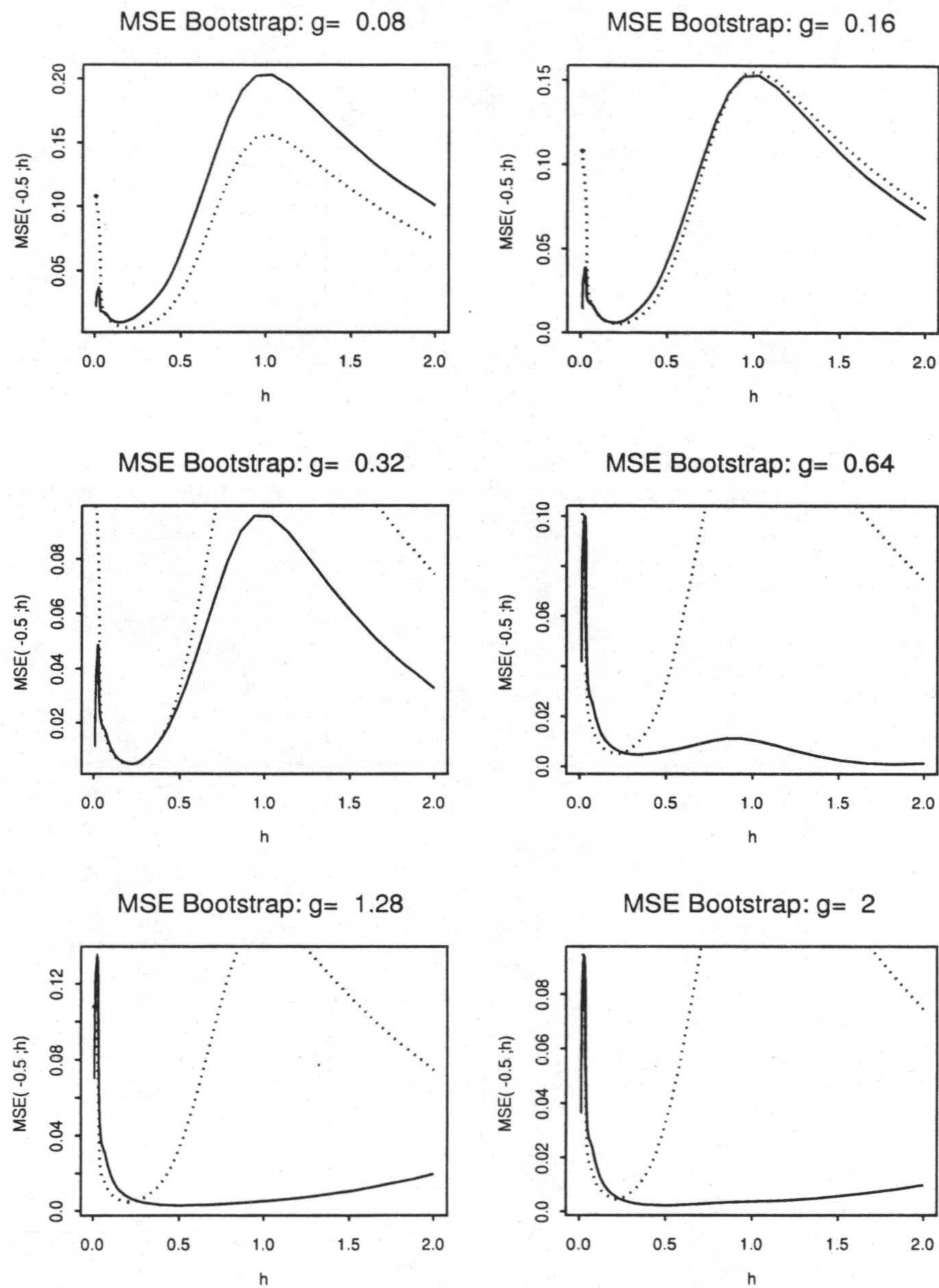


Figura A.7: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = -0.5$ ; y al tamaño muestral 500.

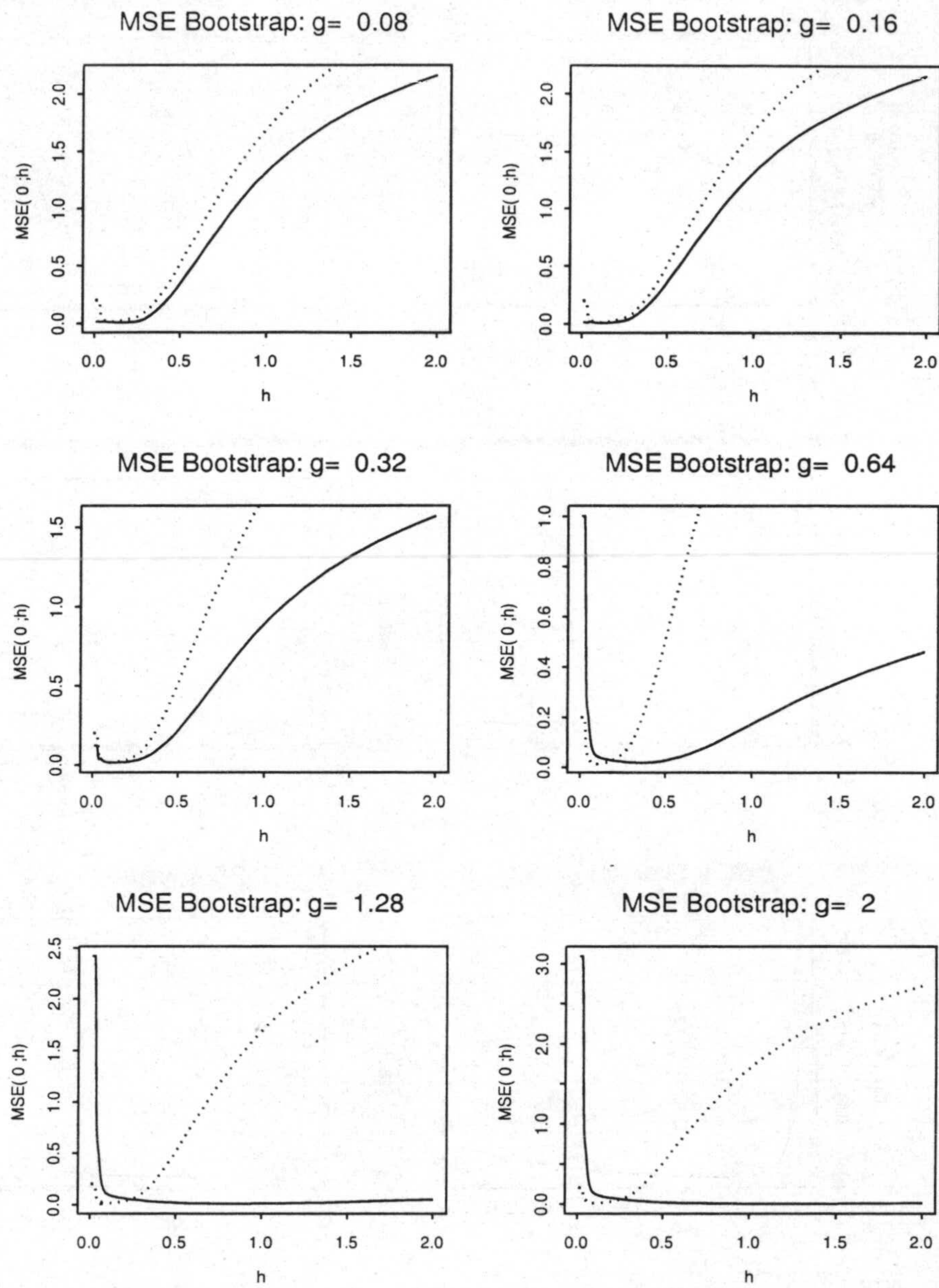


Figura A.8: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = 0$ ; y al tamaño muestral 500.

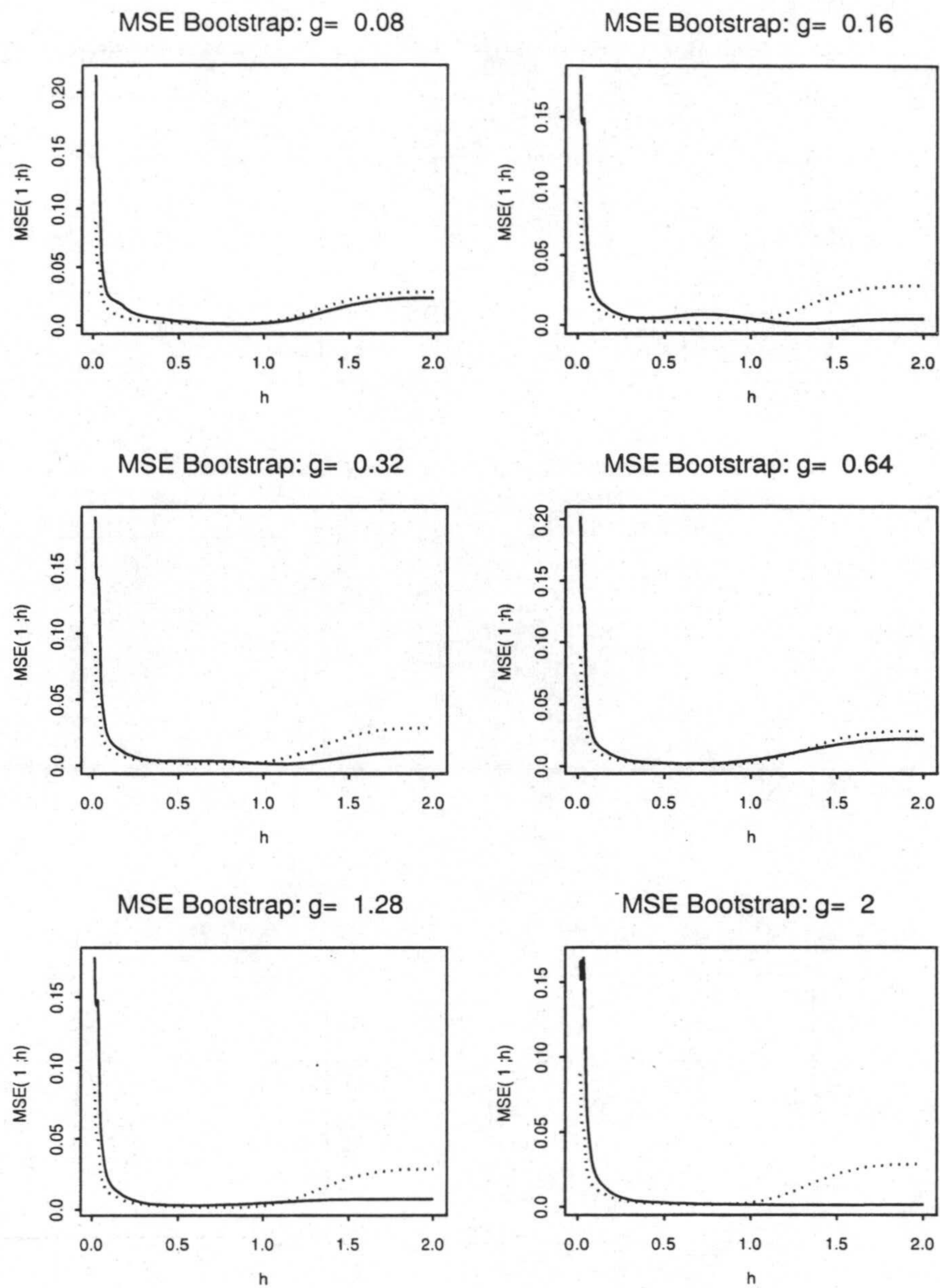


Figura A.9: Error cuadrático medio teórico (curva punteada) y estimación *bootstrap* (curva continua) usando varios parámetros piloto,  $g$ . Las curvas corresponden a la función de prueba 1 en el punto,  $x, x = 1$ ; y al tamaño muestral 500.



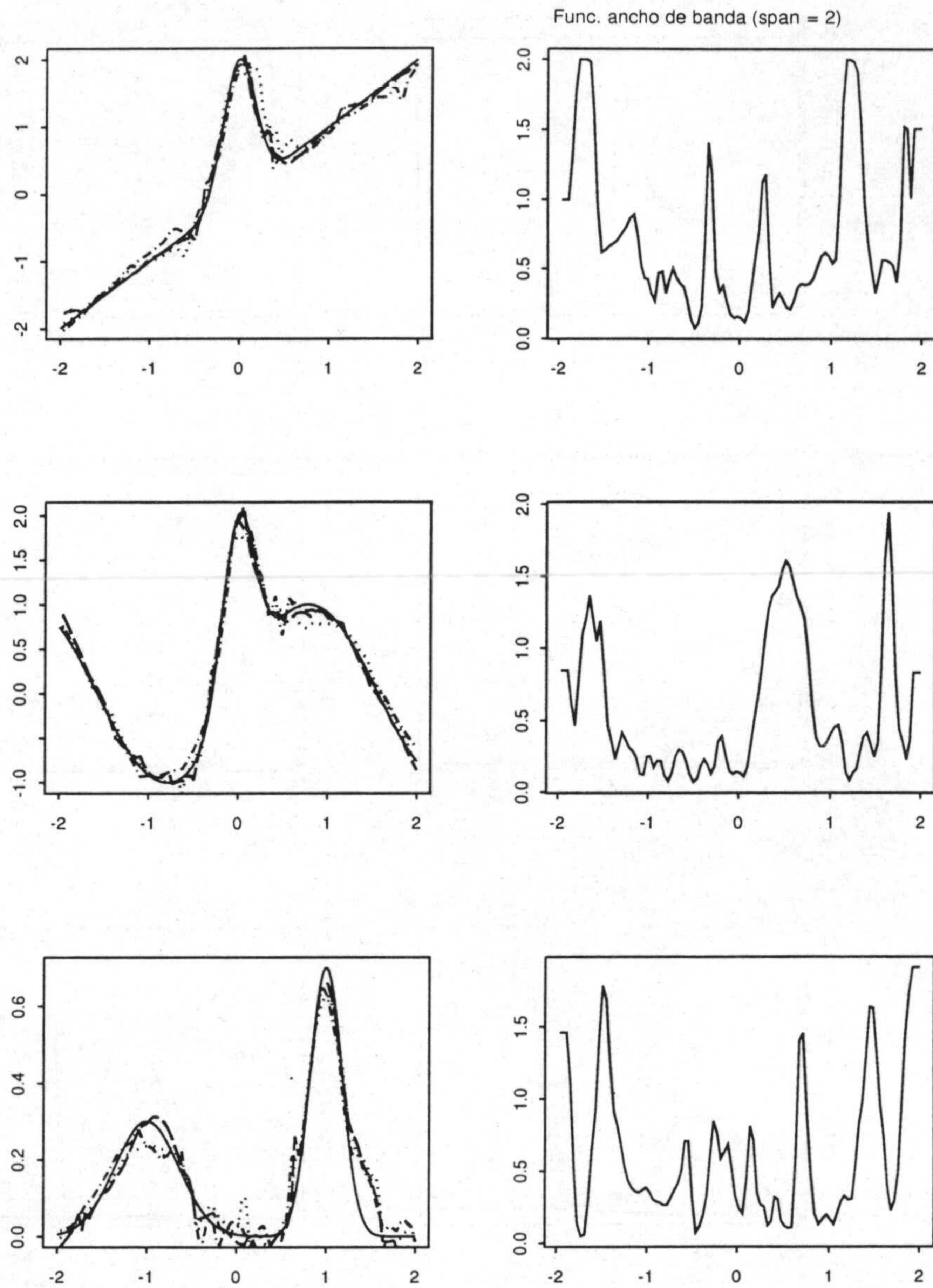


Figura A.10: Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 2$ . El tamaño muestral considerado es 200.

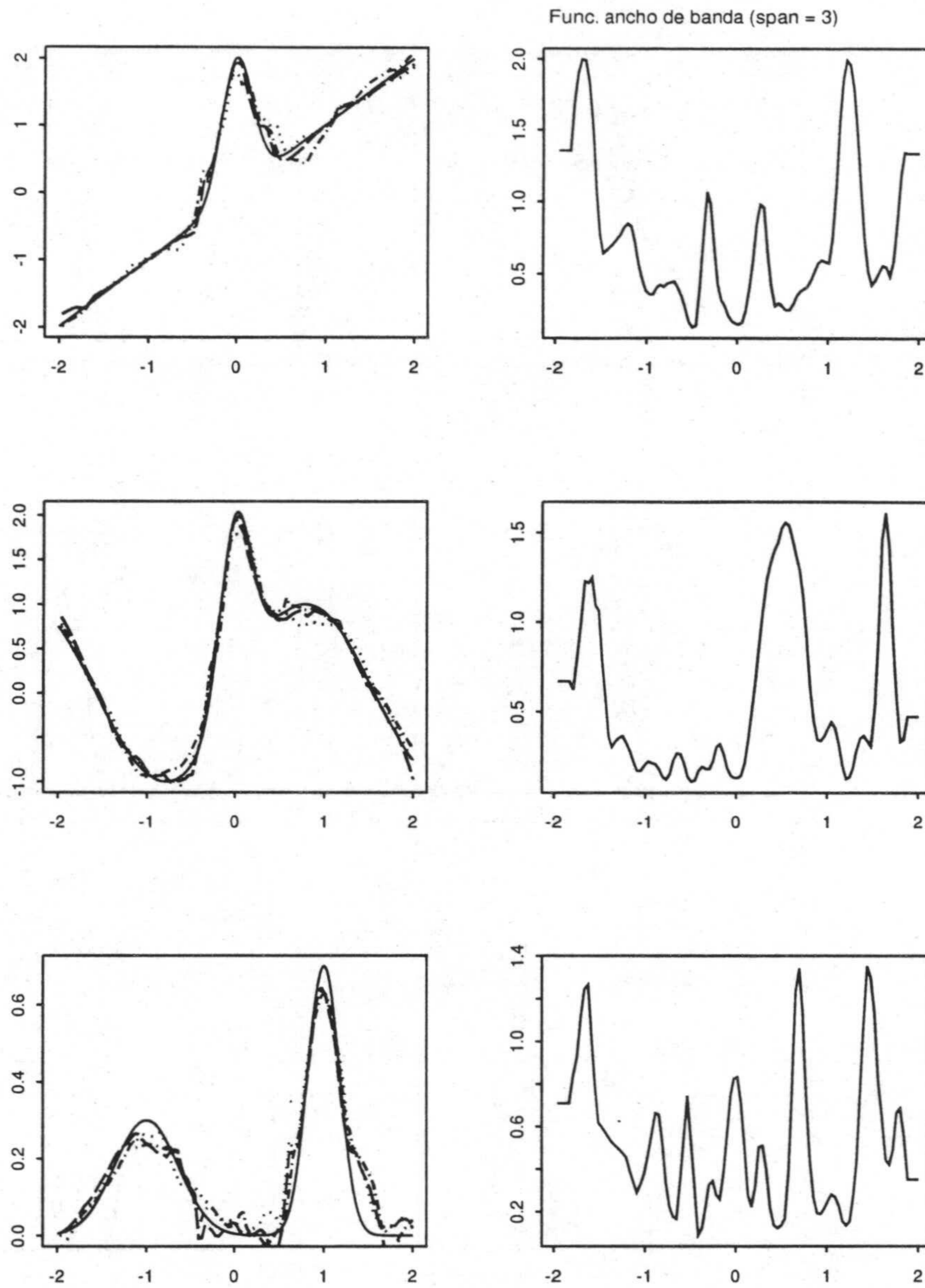


Figura A.11: Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 3$ . El tamaño muestral considerado es 200.

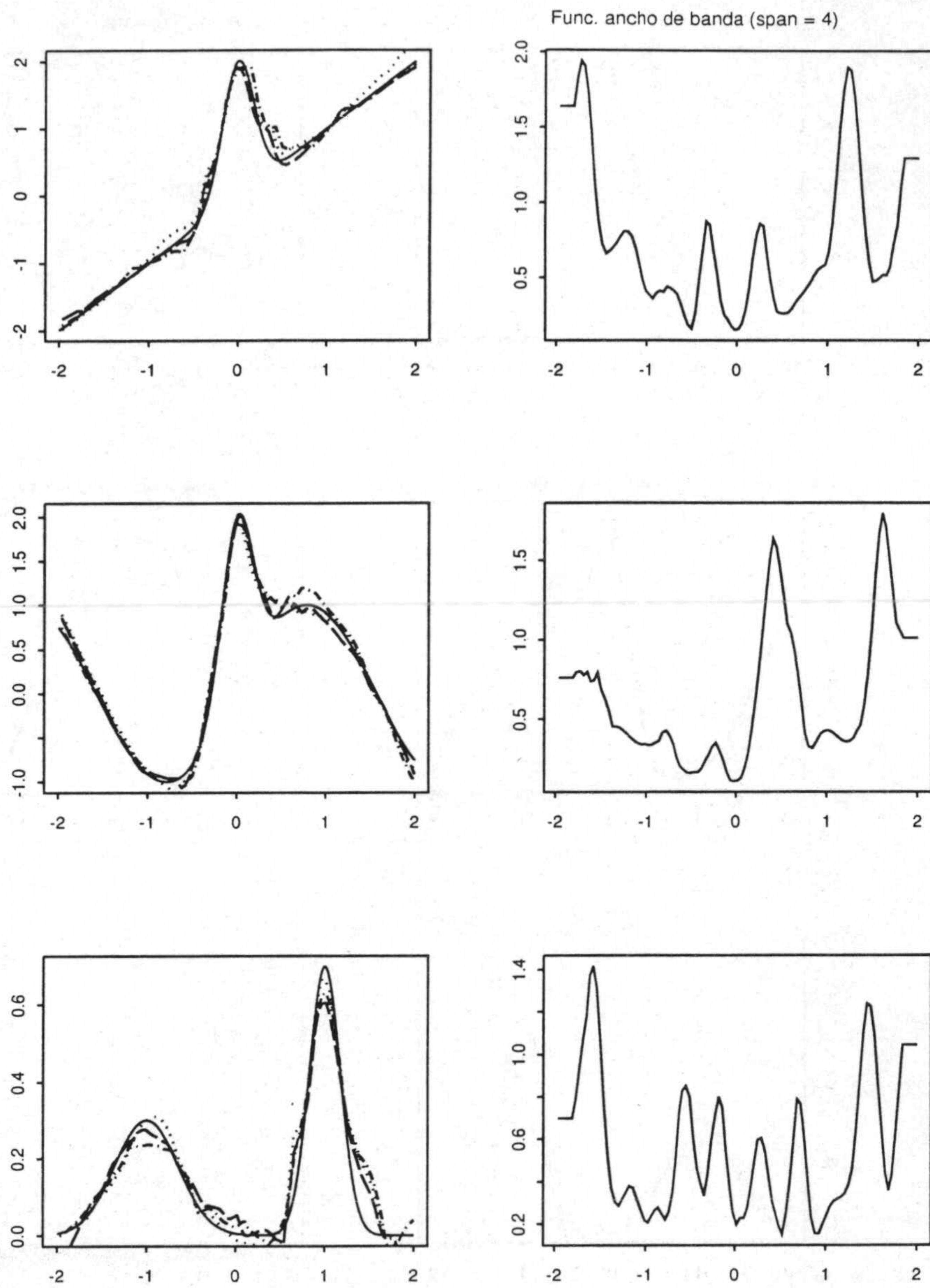


Figura A.12: Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 4$ . El tamaño muestral considerado es 200.

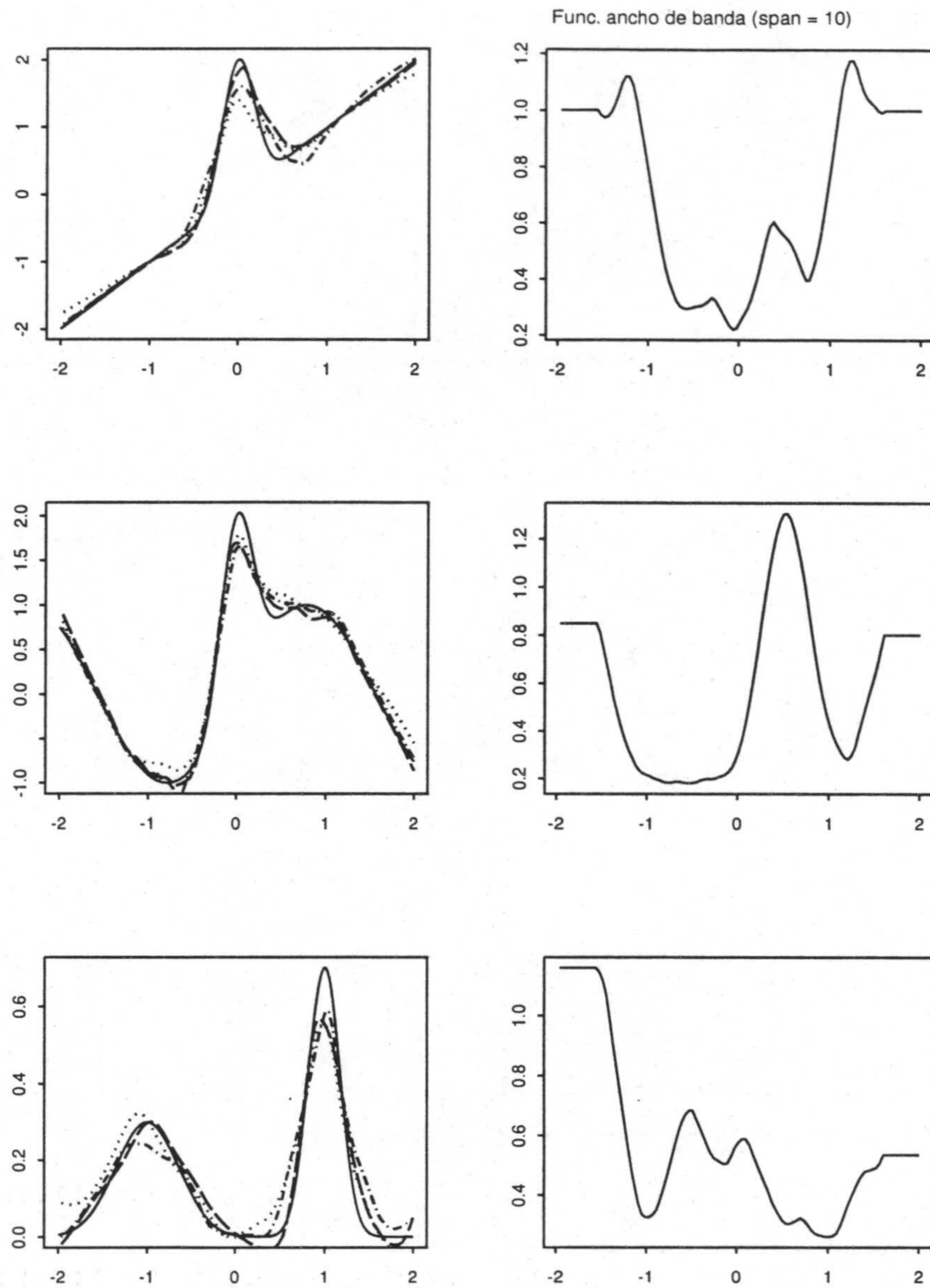


Figura A.13: Curvas representativas calculadas según el estimador polinomial local (con ajustes de grado uno) empleando un ancho de banda local *bootstrap* suavizado. Los gráficos de la derecha representan una de estas funciones ancho de banda local *bootstrap* suavizada con  $span = 10$ . El tamaño muestral considerado es 200.

## Apéndice B

# Computación estadística

### B.1 Obtención del selector *bootstrap* del ancho de banda

A continuación se incluyen sendas funciones que calculan el error cuadrático medio teórico, así como sus estimaciones *bootstrap*, según *wild bootstrap* y según *bootstrap* clásico, tal y como se ha propuesto en el capítulo 4. Y junto con ellas, funciones que minimizan dichos errores, teórico y estimado, obteniendo los anchos de banda locales óptimos, tanto teórico como *bootstrap*.

Las funciones aparecen por orden alfabético, y en los comentarios, incluyen una explicación de sus argumentos. Todas ellas funcionan tanto en S Plus 2000 como en R 1.1.x.

```
defineX <- function( x , z , p )
# Calcula la matriz de diseño X: matriz de dimensiones (n,p+1)
# Los argumentos son:
# x (el vector de observaciones de la variable explicativa)
# z (el punto de estimación considerado)
# p (el grado del ajuste polinomial local)
{
  fil <- length(x)
  col <- p + 1
  matrizX <- double()
  matrizX <- matrix(0, fil, col)
  for(i in 1:fil) {
    matrizX[i, 1] <- 1
```

```

}
for(j in 2:col) {
  matrizX[i, j] <- ((x[i] - z)^(j - 1))
}

```

matrizX

eW <- function( x , z , h )  
 # Calcula la matriz de pesos, se trata de una matriz diagonal  
 # que ha calculado usando la función núcleo bponderado  
 # 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)
matrizW <- double(fil)
matrizW <- matrix(0, fil, fil)
Kh <- double(fil)
tamentorno <- 0
for(i in 1:fil) {
  d <- ((x[i] - z)/h)
  if(abs(d) < 1) {
    tamentorno <- tamentorno + 1
  }
  Kh[i] <- (kbiwei(d)/h)
  matrizW[i, i] <- Kh[i]
}
result <- list(matriz = matrizW, tam = tamentorno)
result
}

```

```

det <- function(X)
# Calcula el determinante de una matriz X
{
  prod(eigen(X)$values)
}

```

```

eval.h.boot <- function(puntos, x, y, fboot, tam, inf, sup,
  ftest, C, p, sigma)
# Evalúa la función "h.boot.local" sobre una red de puntos.

```

```
{
  vectorh <- red.h(inf, sup, tam, C)
  boot <- apply(puntos, MARGIN = 2, FUN = h.boot.local,
              x = x, y = y, vectorh = vectorh,
              resultados = fboot, tam = tam, ftest = ftest,
              p = p, sigma = sigma, inf = inf, sup = sup,
              C = C)
}

eval.h.boot.clasico <- function(puntos,x, y, fboot, tam, inf,
                               sup, ftest, C, p, sigma)
# Similar a eval.h.boot pero para el remuestreo clásico
# sobre los residuos.
{
  vectorh <- red.h(inf, sup, tam, C)
  boot <- apply(puntos, MARGIN = 2, FUN = h.boot.local.clasico,
              x = x, y = y, vectorh = vectorh,
              resultados = fboot, tam = tam, ftest = ftest,
              p = p, sigma = sigma, inf = inf, sup = sup,
              C = C)
}

eval.h.teo <- function(puntos,x, y, m, fteo, tam, inf, sup,
                      ftest, C, p, sigma)
# Eval\''{u}a la función "h.teorico.local" sobre una red de puntos.
{
  eval <- apply(puntos, MARGIN = 2, FUN = h.teorico.local,
              x = x, y = y, m = m, fresultados = fteo,
              tam = tam, ftest = ftest, inf = inf,
              sup = sup, C = C, p = p, sigma = sigma)
}

h.boot.local <- function( ftest , x , y , resultados , tam , z ,
                          vectorh , p , sigma , inf , sup , C )
# Calcula mse.boot sobre una red de anchos de banda y minimiza
# en dicha red el error, obteniendo el h óptimo bootstrap.
# Los argumentos son:
# x , y (vectores de observaciones)
# z (punto de estimación considerado)
# ftest (función test considerada: 1,2 ó 3)
# resultados (nombre del fichero donde se escribirán los
```

```

#           resultados, en cada fila se escriben dos
#           valores, el h donde se alcanza el mínimo,
#           el valor del mínimo y el punto de estimación,
#           z, donde se ha calculado)
{
  longred <- length(vectorh)
  minimo <- 1000
  minh <- 0
  error <- double(longred)
  g <- piloto(ftest, tam, sigma, z, inf, sup)
  mgz <- polinomial.estimacion.piloto(x,y,z,g,p,C,inf,sup)
  mgxi <- double(tam)
  for(i in 1:tam) {
    mgxi[i] <- polinomial.estimacion.piloto(x, y, x[i], g, p, C,
                                           inf, sup)
  }
  for(i in 1:longred) {
    error[i] <- mse.boot(x, y, tam, z, mgz, vectorh[i], mgxi, p)
    if(!is.na(error[i]) && error[i] < minimo) {
      minimo <- error[i]
      minh <- vectorh[i]
    }
  }
  linea <- c(minimo, minh, z)
  write(linea, file = resultados, ncolumns = 3, append = T)
}

h.boot.local.clasico <- function(ftest, x, y, resultados, tam, z,
                                vectorh, p, sigma, inf, sup, C)
# Similar a h.boot.local, pero para el bootstrap clásico sobre
# los residuos.
{
  longred <- length(vectorh)
  minimo <- 1000
  minh <- 0
  error <- double(longred)
  g <- piloto(ftest, tam, sigma, z, inf, sup)
  mgz <- polinomial.estimacion.piloto(x,y,z,g,p,C,inf,sup)
  mgxi <- double(tam)
  for(i in 1:tam) {
    mgxi[i] <- polinomial.estimacion.piloto(x, y, x[i], g,

```



```

                                p, C, inf, sup)
    }
    for(i in 1:longred) {
        error[i] <- mse.boot.clasico(x, y, tam, z, mgz,
                                    vectorh[i], mgxi, p)
        if(!is.na(error[i]) && error[i] < minimo) {
            minimo <- error[i]
            minh <- vectorh[i]
        }
    }
    linea <- c(minimo, minh, z)
    write(linea, file = resultados, ncolumns = 3, append = T)
}

h.teorico.local <- function(x, y, m, fresultados, tam, z,
                            ftest, inf, sup, C, p, sigma)
# Calcula mse.teorico sobre una red de anchos de banda y
# minimiza en dicha red el error, obteniendo el h teórico
# óptimo local.
# Los argumentos son:
# x, y (vectores de observaciones)
# m (vector de componentes m(Xi))
# fresultados (nombre del fichero donde se escriben los
# anchos de banda mínimos, el valor del MSE
# mínimo y los puntos z)
# tam (tamaño muestral)
# z (punto de estimación)
# ftest (función de prueba considerada)
# (inf,sup) es el intervalo de estimación
# C (constante que define la red geométrica de anchos)
# p (grado del ajuste polinomial local, se supone 1)
# sigma (desviación típica residual del modelo)
{
    vectorbeta <- obtener.beta(ftest, z)
    vectorh <- red.h(inf, sup, tam, C)
    longred <- length(vectorh)
    anchos <- matrix(0, 1, longred)
    for(i in 1:longred) {
        anchos[1, i] <- vectorh[i]
    }
    error <- apply(anchos, 2, FUN = mse.teorico, x = x, y = y,

```

```

        z = z, m = m, vectorbeta = vectorbeta, p = p,
        sigma = sigma)
#minimización
minimo <- 1000
for(i in 1:longred) {
    if(!is.na(error[i]) && error[i] < minimo) {
        minimo <- error[i]
        minh <- vectorh[i]
    }
}
linea <- c(minimo, minh,z)
write(linea, file = fresultados, ncolumns = 3, append = T)
}

leer.datos <- function(fichero, tam)
# Lee un fichero de datos y los almacena en una tabla
# Los argumentos son:
# fichero (nombre del fichero de entrada)
# tam (tamaño muestral)
{
    x <- double(tam)
    y <- double(tam)
    m <- double(tam)
    e <- double(tam)
    dm <- double(tam)
    nombres <- c("x", "y", "m", "e", "dm")
    tabla <- read.table(file = fichero, col.names = nombres)
    tabla
}

mse.boot <- function( x , y , tam , z , mgz , h , mgxi , p )
# Calcula la estimación bootstrap del error cuadrático medio
# según el wild bootstrap
# Los argumentos son:
# x, y (vectores de observaciones)
# z (punto de estimación)
# tam (el tamaño muestral)
# mgz (la estimación polinomial en el punto z usando el ancho
# de banda piloto g)
# h (ancho de banda)
# mgxi (igual que mgz, pero en cada observación Xi)

```

```
# p (grado del ajuste polinomial local)
{
  wh <- pesos.estimador(x, y, z, h, p)
  obs <- t(x)
  if(!is.na(wh[1])) {
    suma1 <- 0
    suma2 <- 0
    for(i in 1:tam) {
      suma1 <- suma1 + (wh[i] * mgxi[i])
      suma2 <- suma2 + ((wh[i]^2) * (y[i] - mgxi[i])^2)
    }
    sesgo <- suma1 - mgz
    var <- suma2
    mse <- sesgo^2 + var
  }
  else {
    mse <- NA
  }
  mse
}
```

```
mse.boot.clasico <- function(x, y, tam, z, mgz, h, mgxi, p)
# Calcula la estimación bootstrap del error cuadrático medio
# según un remuestreo clásico sobre los residuos,
# el sesgo igual que en mse.boot, cambia la varianza.
# Los argumentos son iguales que en mse.boot
{
  wh <- pesos.estimador(x, y, z, h, p)
  obs <- t(x)
  if(!is.na(wh[1])) {
    suma1 <- 0
    suma2 <- 0
    suma3 <- 0
    suma4<-0
    for(i in 1:tam) {
      suma1 <- suma1 + (wh[i] * mgxi[i])
      suma2 <- suma2 + (wh[i]^2)
      suma3 <- suma3 + (y[i]-mgxi[i])^2
      suma4 <- suma4 + (y[i]-mgxi[i])
    }
    med<-suma4 /tam
  }
```

```

        sesgo <- suma1 - mgz
        var <- (1/tam)*suma2*(suma3 - tam* med^2)
        mse <- sesgo^2 + var
    }
    else {
        mse <- NA
    }
    mse
}

mse.teorico <- function(x, y, z, m, vectorbeta, h, p, sigma)
# Calcula el error MSE teórico, según la expresión matricial.
# Se ha implementado tan sólo el caso p=1
# Los argumentos son:
# x, y (vectores de observaciones)
# z (punto de estimación)
# m (vector de componentes m(Xi))
# vectorbeta (vector (m(z), m'(z)))
# sigma (desviación típica residual del modelo)
{
    fil <- length(x)
    col <- p + 1
    matrizX <- defineX(x, z, p)
    r <- m - (matrizX %*% vectorbeta)
    e1 <- t(double(col))
    e1[1, 1] <- 1
    trasX <- t(matrizX)
    resw <- defineW(x, z, h)
    matrizW <- resw$matriz
    tamentorno <- resw$tam
    determinante <- det((trasX %*% matrizW) %*% matrizX)
    if(tamentorno >= p + 1 && abs(determinante) > 0.001) {
        inversa <- solve((trasX %*% matrizW) %*% matrizX)
        sesgo <- e1 %*% (inversa %*% trasX %*% matrizW %*% r)
    }
    varianza <- e1 %*% (inversa %*% trasX %*% (matrizW %*%
        matrizW) %*% matrizX %*% inversa) %*% t(e1)
        * sigma^2
    mse <- sesgo^2 + varianza
}
else {
    mse <- NA
}

```

```

    }
    mse
}

obtener.beta <- function(ftest, z)
# Calcula el vector de componentes (m(z), m'(z))
# Los argumentos son:
#   ftest (función de prueba considerada)
#   z (punto de estimación)
{
  vectorbeta <- double(2)
  if(ftest == 1) {
    m <- z + 2 * exp(-16 * z^2)
    dm <- 1 - 64 * z * exp(-16 * z^2)
  }
  if(ftest == 2) {
    m <- sin(2 * z) + 2 * exp(-16 * z^2)
    dm <- 2 * cos(2 * z) - 64 * z * exp(-16 * z^2)
  }
  if(ftest == 3) {
    m <- 0.3*exp(-4*(z+1)^2) + 0.7 * exp(-16*(z-1)^2)
    dm <- 0.3*(-8*(z+1))*exp(-4*(z+1)^2) + 0.7*(-32*(z-1))
      * exp(-16*(z-1)^2)
  }
  vectorbeta <- c(m, dm)
  vectorbeta
}

pesos.estimador <- function( x , y , z , h , p )
# Calcula un vector que define las ponderaciones que asigna
# el estimador polinomial local a cada observación Yi.
# Corresponden a los pesos wi (z;h) (i = 1,...,n)
# Los argumentos son:
#   x (el vector de observaciones de la variable X)
#   y (vector de observaciones de la variable Y)
#   z (el punto de estimación)
#   h (el ancho de banda)
#   p (el grado del ajuste polinomial local)
{
  fil <- length(x)
  w <- double(fil)

```

```

col <- p + 1
eprevio <- double(col)
e1 <- t(eprevio)
e1[1, 1] <- 1
derivada <- double(col)
matrizX <- defineX(x, z, p)
tamentorno <- 0
resW <- defineW(x, z, h)
matrizW <- resW$matriz
control <- 0
for(i in 1:fil) {
  if(matrizW[i, i] > 0.001) {
    control <- 1
  }
}
tamentorno <- resW$tam
if(tamentorno >= p + 1 && control != 0) {
  trasX <- t(matrizX)
  determinante <- det((trasX %*% matrizW) %*% matrizX)
  if(abs(determinante) > 0.001) {
    inversa <- solve((trasX %*% matrizW) %*% matrizX)
    matrizpesos <- inversa %*% (trasX %*% matrizW)
    w <- (e1 %*% matrizpesos)
  }
  else {
    for(i in 1:fil) {
      w[i] <- NA
    }
  }
}
else {
  for(i in 1:fil) {
    w[i] <- NA
  }
}
w
}

```

```

piloto <- function(ftest, tam, sigma, z, inf, sup)
# Calcula el ancho de banda piloto según la expresión teórica
# obtenida (en la estimación según el wild bootstrap)

```

```

# Los argumentos son:
# ftest (función de prueba considerada)
# tam (tamaño muestral)
# sigma (desviación típica residual del modelo)
# z (punto de estimación)
# (inf, sup) el intervalo de estimación de la función
{
  if(ftest == 1) {
    der4 <- 64 * exp(-16 * z^2) * (96 - 199680 * z^2 + 2097152
      * z^4)
  }
  if(ftest == 2) {
    der4 <- 16 * sin(2 * z) + 64 * exp(-16 * z^2) * (96 -
      199680 * z^2 + 2097152 * z^4)
  }
  if(ftest == 3) {
    der4 <- (exp(-4*(z+1)^2))*(1228.8*(z+1)^4 - 921.6*(z+1)^2
      +57.6) + +(exp(-16*(z-1)^2))*(68812.8*(z-1)^4 -
      137625.6*(z-1)^2 +2150.4)
  }
#para el núcleo bipoderado
  mu2 <- 1/7
  RK2 <- 45/2
  C2 <- (1/(sup - inf)) * (mu2^2) * (der4^2)
  C1 <- 5 * RK2 * (sigma^2)
  g <- tam^(-1/9) * (C1/C2)^(1/9)
  min <- (sup - inf)/tam
  max <- (sup - inf)/2
  if(g < min)
    g <- min
  if(g > max)
    g <- max
  g
}

polinomial.estimacion <- function( x , y , z , h , p )
# Usando los pesos calculados mediante la función
# "pesos.estimador", calcula el estimador polinomial local
# en un punto z
{
  w <- pesos.estimador(x, y, z, h, p)
}

```

```

    if(!is.na(w[1])) {
      m <- w %*% y
    }
    else {
      m <- NA
    }
    m
  }

polinomial.estimacion.piloto <- function(x, y, z, g, p, C,
                                       inf, sup)
# Variante de la función "polinomial.estimacion" para capaz de
# controlar que el ancho de banda que involucra (el piloto)
# no da lugar a un valor missing.
# Los argumentos son:
# x, y (vectores de observaciones)
# z (punto de estimación)
#
#           g (ancho de banda - piloto)
#           p (grado del ajuste polinomial)
#           C (define la red geométrica de anchos de banda)
#           (inf,sup) es el intervalo de estimación
{
  w <- pesos.estimador(x, y, z, g, p)
  i <- 0
  nuevog <- g
  while(is.na(w[1]) && nuevog < (sup - inf)/2) {
    i <- i + 1
    nuevog <- (C^i) * g
    w <- pesos.estimador(x, y, z, nuevog, p)
  }
  m <- w %*% y
  m
}

red.h <- function(inf, sup, tam, C)
# Construye una red de anchos de banda, de tipo geométrico.
# El mínimo es (rango/n) y el máximo (rango/2)
# Los argumentos son:
# (inf,sup) el intervalo de estimación
# tam (el tamaño muestral)
# C (constante que define la red geométrica)

```



```
{
  rango <- sup - inf
  hmin <- rango/tam
  hmax <- rango/2
  i <- 1
  h <- c(hmin)
  while(h[i] < hmax) {
    hnuevo <- (C^i) * hmin
    i <- i + 1
    if(hnuevo < hmax) {
      h <- c(h, hnuevo)
    }
    if(hnuevo >= hmax) {
      h <- c(h, hmax)
    }
  }
  h
}

simulacion.m1 <- function(tam, replicas, ficherosdatos, fboot,
                          fteo)
# Realiza la simulación para la primera función de prueba considerada.
# Está realizado para estimaciones lineales locales (p=1)
# Los argumentos son:
#   tam (tamaño muestral)
#   replicas (número de réplicas a realizar)
#   ficherosdatos (nombre del fichero que almacenará los datos
#                 simulados)
#   fboot (nombre del fichero que almacenará los anchos de banda
#         bootstrap)
#   fteo (nombre del fichero que almacenará los anchos de banda
#        teóricos)
{
  library(Matrix)
  ftest <- 1
  sigma<-0.4
  p <- 1
  inf <- -2
  sup <- 2
  C <- 1.1
  ngrid <- 101
```

```

x <- double(tam)
y <- double(tam)
m <- double(tam)
e <- double(tam)
dm <- double(tam)
linea <- double(5)
for(b in 1:replicas) {
  set.seed <- b
  x <- sort(runif(tam, inf, sup))
  m <- x + 2 * exp(-16 * x^2)
  dm <- 1 - 64 * x * exp(-16 * x^2)
  e <- rnorm(tam) * sigma
  y <- m + e
  for(i in 1:tam) {
    linea <- c(x[i], y[i], m[i], e[i], dm[i])
write(linea, file = ficherosdatos, ncolumns = 5,
      append = T)
  }
  #la red es de ngrid puntos
  #buscamos una red de puntos equiespaciados
  puntos <- double(ngrid)
  for(i in 1:ngrid) {
    puntos[i] <- inf + (sup - inf) * (i/ngrid)
  }
  puntos <- t(puntos)
  eval.h.boot(puntos,x, y, fboot, tam, inf, sup, ftest,
              C, p, sigma)
  eval.h.teo(puntos,x, y, m, fteo, tam, inf, sup, ftest,
             C, p, sigma)
}
}

simulacion.m2 <- function(tam, replicas, ficherosdatos, fboot,
                        fteo)
# Realiza la simulación para la segunda función de prueba.
# Considera sólo el caso de estimaciones lineales locales (p=1)
# Los argumentos se definen como en "simulacion.m1"
{
  library(Matrix)
  sigma<-0.3
  p<-1

```

```
fctest <- 2
inf <- -2
sup <- 2
C <- 1.1
ngrid <- 101
x <- double(tam)
y <- double(tam)
m <- double(tam)
e <- double(tam)
dm <- double(tam)
linea <- double(5)
for(b in 1:replicas) {
  set.seed <- b
  x <- sort(runif(tam, -2, 2))
  m <- sin(2 * x) + 2 * exp(-16 * x^2)
  dm <- 2 * cos(2 * x) - 64 * x * exp(-16 * x^2)
  e <- rnorm(tam) * 0.3
  y <- m + e
  for(i in 1:tam) {
    linea <- c(x[i], y[i], m[i], e[i], dm[i])
  write(linea, file = ficherosdatos, ncolumns = 5,
    append = T)
  }
  #la red es de ngrid puntos
  #buscamos una red de puntos equiespaciados
  puntos <- double(ngrid)
  for(i in 1:ngrid) {
    puntos[i] <- inf + (sup - inf) * (i/ngrid)
  }
  puntos <- t(puntos)
eval.h.boot(puntos,x, y, fboot, tam, inf, sup, fctest, C,
  p, sigma)
eval.h.teo(puntos,x, y, m, fteo, tam, inf, sup, fctest, C,
  p, sigma)
}
}

simulacion.m3 <- function(tam, replicas, ficherosdatos, fboot,
  fteo)
# Realiza la simulación para la tercera función de prueba.
# Considera sólo el caso de estimaciones lineales locales (p=1)
```

```
# Sus argumentos se definen como en "simulacion.m1"
{
  library(Matrix)
  sigma<-0.1
  p<-1
  ftest <- 2
  inf <- -2
  sup <- 2
  C <- 1.1
  ngrid <- 101
  x <- double(tam)
  y <- double(tam)
  m <- double(tam)
  e <- double(tam)
  dm <- double(tam)
  linea <- double(5)
  for(b in 1:replicas) {
    set.seed <- b
    x <- sort(runif(tam, -2, 2))
    m<- 0.3*exp(-4*(x+1)^2) + 0.7 * exp(-16*(x-1)^2)
    dm<- 0.3*(-8*(x+1))*exp(-4*(x+1)^2) + 0.7*(-32*(x-1)) *
      exp(-16*(x-1)^2)
    e <- rnorm(tam) * sigma
    y <- m + e
    for(i in 1:tam) {
      linea <- c(x[i], y[i], m[i], e[i], dm[i])
    write(linea, file = ficherosdatos, ncolumns = 5,
      append = T)
    }
    #la red es de ngrid puntos
    #buscamos una red de puntos equiespaciados
    puntos <- double(ngrid)
    for(i in 1:ngrid) {
      puntos[i] <- inf + (sup - inf) * (i/ngrid)
    }
    puntos <- t(puntos)
    eval.h.boot(puntos,x, y, fboot, tam, inf, sup, ftest,
      C, p, sigma)
    eval.h.teo(puntos,x, y, m, fteo, tam, inf, sup, ftest,
      C, p, sigma)
  }
}
```

```
}  
  
suaviza anchos <- function(h, x, span)  
# Suaviza la función ancho de banda bootstrap resultante  
# utilizando una función peso triangular.  
# Los argumentos son:  
# h (vector de anchos de banda bootstrap)  
# x (vector de puntos de estimación)  
# span (parámetro para el suavizamiento)  
{  
  ngrid <- length(x)  
  delta <- x[2] - x[1]  
  d <- delta * span  
  hsuave <- double(ngrid)  
  W <- double(ngrid)  
  for(i in 1:ngrid) {  
  }  
  for(i in (span + 1):(ngrid - span)) {  
    suma <- 0  
    for(l in 1:ngrid) {  
      u <- ((x[l] - x[i])/d)  
      W[l] <- triangular(u)  
      suma <- suma + W[l] * h[l]  
    }  
    hsuave[i] <- suma/delta  
  }  
  for(i in 1:span) {  
    hsuave[i] <- hsuave[span + 1]  
  }  
  for(i in (ngrid - span + 1):ngrid) {  
    hsuave[i] <- hsuave[ngrid - span]  
  }  
  hsuave  
}
```

## B.2 Evaluación local de las componentes de sesgo y de varianza

Incluimos en esta sección las funciones que permiten evaluar el error cuadrático medio, sobre una red de anchos de banda. Se consideran las versiones teórica y

*bootstrap* del error local, para tres puntos de interés de las funciones de prueba definidas en el capítulo 5.

De nuevo, las funciones aparecen por orden alfabético, y el significado de sus argumentos se incluye en los comentarios.

```
cambiar.z <- function(fteo, fbootvariosg, fbooteligeg, z1, z2,
                    z3, datos, tam, ftest, C, p, sigma, inf, sup)
# Realiza el cálculo y la comparación de los errores MSE para
# tres puntos zi.
# Se considera sólo el caso de estimación lineal local.
# Los parámetros están descritos en la función "comp.mse.z"
{
  p <- 1
  comp.mse.z(fteo, fbootvariosg, fbooteligeg, z1, datos, tam,
            ftest, sigma, C, p, inf, sup)
  comp.mse.z(fteo, fbootvariosg, fbooteligeg, z2, datos, tam,
            ftest, sigma, C, p, inf, sup)
  comp.mse.z(fteo, fbootvariosg, fbooteligeg, z3, datos, tam,
            ftest, sigma, C, p, inf, sup)
}
```

```
comp.mse.z <- function(fteo, fbootvariosg, fbooteligeg, z, datos,
                    tam, ftest, sigma, C, p, inf, sup)
# Calcula las funciones MSE teórico y MSE bootstrap para varios
# valores del parámetro g, y para la elección teórica propuesta.
# Además de las componentes sesgo y varianza.
# Los argumentos son:
#   fteo (nombre del fichero que almacenará los resultados
#         teóricos)
#   fbootvariosg (fichero que almacenará los resultados de
#                 las estimaciones bootstrap. Considera varios
#                 valores posibles para g)
#   fbooteligeg (fichero que almacenará los valores bootstrap.
#                 Considera la elección teórica de g)
#   z (punto de estimación)
#   datos (nombre del fichero que almacena los datos
#           utilizados en los cálculos)
#   tam (tamaño muestral)
#   ftest (función de prueba considerada)
```

```
# sigma (desviación típica residual del modelo)
# C (constante que define la red geométrica de anchos
#   de banda)
# (inf,sup) intervalo de estimación
{
  mseteo.z(fteo, z, datos, tam, ftest, sigma, C, p, inf, sup)
  mseboot.z.variosg(fbootvariosg, z, datos, tam, ftest, sigma,
                  C, p, inf, sup)
  mseboot.z.eligeg(fbooteligeg, z, datos, tam, ftest, sigma,
                  C, p, inf, sup)
}

mseboot.z.eligeg <- function(resultadosboot, z, datos, tam,
                             ftest, sigma, C, p, inf, sup)
# Calcula el error bootstrap y sus
# componentes de sesgo y de varianza como funciones del ancho
# de banda. Utiliza la elección teórica del parámetro piloto.
# Los argumentos están definidos en la función "comp.mse.z".
{
  datos <- leer.datos(datos, tam)
  x <- datos$x
  obs <- t(x)
  y <- datos$y
  vectorh <- red.h(inf, sup, tam, C)
  longred <- length(vectorh)
  error <- double(longred)
  b <- double(longred)
  v <- double(longred)
  g <- piloto(ftest, tam, sigma, z, inf, sup)
  mgxi <- double(tam)
  for(i in 1:tam) {
    mgxi[i] <- polinomial.estimacion.piloto(x, y, x[i], g, p,
                                           C, inf, sup)
  }
  mgz <- polinomial.estimacion.piloto(x, y, z, g, p, C, inf, sup)
  for(i in 1:longred) {
    res <- mse.boot.Vhg.comp(x, y, tam, z, mgz, vectorh[i], i,
                             mgxi, p)
    error[i] <- res$mse
    b[i] <- res$sesgo
    v[i] <- res$var
  }
}
```

```

        linea <- c(error[i], vectorh[i], b[i], v[i])
        write(linea, file=resultadosboot, ncolumns=4, append=T)
    }
}

mseboot.z.variosg <- function(resultadosboot, z, datos, tam,
                             ftest, sigma, C, p, inf, sup)
# Calcula el error cuadrático medio bootstrap, usando varios
# valores del parámetro piloto g; así como sus componentes de
# sesgo y de varianza. Se consideran como funciones del ancho
# de banda, y por tanto son evaluadas en una red anchos.
# Los argumentos están definidos en la función "comp.mse.z".
{
    datos <- leer.datos(datos, tam)
    x <- datos$x
    obs <- t(x)
    y <- datos$y
    vectorh <- red.h(inf, sup, tam, C)
    longred <- length(vectorh)
    error <- double(longred)
    b <- double(longred)
    v <- double(longred)
    mgxi <- double(tam)
    vectorg <- red.h(inf, sup, 50, 2)
    numerog <- length(vectorg)
    for(k in 1:numerog) {
        g <- vectorg[k]
        for(i in 1:tam) {
            mgxi[i] <- polinomial.estimacion.piloto(x, y, x[i], g,
            p, C, inf, sup)
        }
    }
    mgz <- polinomial.estimacion.piloto(x, y, z, g, p, C,
    inf, sup)

    for(i in 1:longred) {
        res <- mse.boot.Vhg.comp(x, y, tam, z, mgz, vectorh[i],
        i, mgxi, p)

        error[i] <- res$mse
        b[i] <- res$sesgo
        v[i] <- res$var
        linea <- c(error[i], vectorh[i], b[i], v[i])
        write(linea, file = resultadosboot, ncolumns = 4,

```



```

        append = T)
    }
}

mseteo.z <- function(resultadosteo, z, datos, tam, ftest, sigma,
                    C, p, inf, sup)
# Calcula el error MSE teórico y sus componentes de sesgo y
# de varianza como funciones del ancho de banda, siendo para ello,
# evaluadas en una red de anchos de banda.
# Los argumentos ya han sido definidos en la función "comp.mse.z"
{
  datos <- leer.datos(datos, tam)
  x <- datos$x
  obs <- t(x)
  y <- datos$y
  m <- datos$m
  vectorbeta <- obtener.beta(ftest, z)
  vectorh <- red.h(inf, sup, tam, C)
  longred <- length(vectorh)
  error <- double(longred)
  v <- double(longred)
  b <- double(longred)
  for(i in 1:longred) {
    res <- mse.teorico.comp(x, y, z, m, vectorbeta, vectorh[i],
                          p, sigma)

    error[i] <- res$mse
    v[i] <- res$var
    b[i] <- res$sesgo
    linea <- c(error[i], vectorh[i], b[i], v[i])
    write(linea, file = resultadosteo, ncolumns = 4, append=T)
  }
}

```

### B.3 Análisis gráfico de los resultados

En esta sección se incluyen las funciones necesarias para la obtención de los gráficos que se incluyen en esta memoria.

```
curvas.repres <- function(fichero, datos, tam, ftest, replicas)
```

```
# Usando los anchos de banda bootstrap obtenidos como resultado
# de las funciones "simulacion.m1", "simulacion.m2" y
# "simulacion.m3", esta función calcula y representa la
# estimación polinomial resultante (grado uno), para tres
# curvas representativas.
# Además crea un gráfico que representa la propia función
# ancho de banda bootstrap correspondiente a la curva del
# 10%).
# Los argumentos son:
# fichero (nombre del fichero que contiene los anchos de
# banda bootstrap para cada punto de estimación
# la red)
# datos (fichero que contiene los datos de la simulación,
# correspondientes a los anchos bootstrap)
# tam (tamaño de muestra en las simulaciones)
# ftest (función de prueba considerada)
# replicas (número de réplicas realizadas)
{
  per1<-0.10
  per2<-0.50
  per3<-0.95
  inf<--2
  sup<-2
  ngrid<-101
  p<-1
  #calculamos la función exacta
  abc <- double(1000)
  m <- double(1000)
  z<-double(1000)
  for(i in 1:1000)
  {
    z[i] <- inf + (sup - inf) * (i/1000)
    if (ftest==1)
    {
      m[i] <- z[i] + 2 * exp(-16 * z[i]^2)
    }
    if (ftest==2)
    {
      m[i] <- sin(2 * z[i]) + 2 * exp(-16 * z[i]^2)
    }
    if (ftest==3)
```

```
{
  m[i] <- 0.3 * exp(-4 * (z[i] + 1)^2) + 0.7 *
    exp(-16 * (z[i] - 1)^2)
}
}
h<-matrix(0,nrow=ngrid,ncol=replicas)
puntos<-matrix(0,nrow=ngrid,ncol=replicas)
mh <- matrix(0,nrow=ngrid,ncol=replicas)
teo<-double(ngrid)
res <- leer.anchos(fichero)
datos <- leer.datos(datos)
src<-double(replicas)
x<-double(tam)
y<-double(tam)
for (i in 1:replicas)
{
  for (k in 1:ngrid)
  {
    h[k,i]<-res$h[k+ (i-1)*ngrid]
    puntos[k,i]<-res$x[k+ (i-1)*ngrid]
  }
  for (k in 1:tam)
  {
    x[k] <- datos$x[k+ (i-1)*tam]
    y[k] <- datos$y[k+ (i-1)*tam]
  }
  for(k in 1:ngrid)
  {
    if(!is.na(h[k,i]))
    {
      mh[k,i] <- polinomial.estimacion(x, y, puntos[k,i],
        h[k,i],p)
    }
  }
}
#calcula la suma residual de cuadrados y selecciona tres
#curvas representativas
suma<-0
for (k in 1:ngrid)
{
  if (i==1)
  {
```

```

    if (ftest==1)
    {
        teo[k] <- puntos[k,i] + 2 * exp(-16*puntos[k,i]^2)
    }
    if (ftest==2)
    {
        teo[k] <- sin(2 * puntos[k,i]) + 2 * exp(-16 *
            puntos[k,i]^2)
    }
    if (ftest==3)
    {
        teo[k] <- 0.3 * exp(-4 * (puntos[k,i] + 1)^2)+0.7
            * exp(-16 * (puntos[k,i] - 1)^2)
    }
    }
    suma <- suma + (teo[k]-mh[k,i])^2
}
src[i]<-suma/ngrid
}
c1<-quantile(src,per1)
c2<-quantile(src,per2)
c3<-quantile(src,per3)
ord<-sort(src)
for (i in 1:(replicas-1))
{
    if ( ord[i]<=c1 && c1< ord[i+1]) {i1<-i}
    if ( ord[i]<=c2 && c2< ord[i+1]) {i2<-i}
    if ( ord[i]<=c3 && c3< ord[i+1]) {i3<-i}
}
for (i in 1:replicas)
{
    if (src[i]==ord[i1]) {ind1<-i}
    if (src[i]==ord[i2]) {ind2<-i}
    if (src[i]==ord[i3]) {ind3<-i}
}
plot(z,m, type = "l", lty = 1, lwd = 2,xlab="",ylab="")
lines(puntos[,ind1], mh[,ind1],type="l",lty=2,lwd=2)
lines(puntos[,ind2], mh[,ind2],type="l",lty=3,lwd=2)
lines(puntos[,ind3], mh[,ind3],type="l",lty=4,lwd=2)
plot(puntos[,ind1], h[,ind1], type="l", lty=4,lwd=1,
    xlab="", ylab="")

```

```
}

leer anchos <- function(fichero)
# Función que lee un fichero que contiene anchos de banda
# calculados con funciones anteriores. El argumento "fichero"
# corresponde al nombre del fichero a leer.
{
  tabla <- read.table(file = fichero)
  h<-tabla$V2
  x<-tabla$V3
  res<-list(h=h,x=x)
  res
}

leer.datos <- function(fichero)
# Función que lee un fichero de datos de los generados en
# las simulaciones. El argumento "fichero" corresponde al nombre
# del fichero de datos a leer.
{
  nombres <- c("x", "y", "m", "e", "dm")
  tabla <- read.table(file = fichero, col.names = nombres)
  tabla
}

leer.datos.graf <- function(teo, bootelec, bootvar, tam, inf,
                           sup, C, ftest, zetas)
# Leer los ficheros generados para el estudio del error
# MSE y sus componentes.
# Los argumentos:
  teo, bootelec, bootvar (ficheros que contienen los datos de
                        MSE y sus componentes para analizar)
# zetas (vector de puntos donde se ha evaluado el error MSE)
{
  numz <- length(zetas)
  teorico <- read.table(file = teo)
  elec <- read.table(file = bootelec)
  varios <- read.table(file = bootvar)
  red<-red.h(inf, sup, tam, C)
  fil <- length(red)
  mseteo <- matrix(0, nrow = fil, ncol = numz)
  mseelec <- matrix(0, nrow = fil, ncol = numz)
```

```

Bteo <- matrix(0, nrow = fil, ncol = numz)
Vteo <- matrix(0, nrow = fil, ncol = numz)
Belec <- matrix(0, nrow = fil, ncol = numz)
Velec <- matrix(0, nrow = fil, ncol = numz)
numg <- 6
dimension <- c(fil, numg, numz)
msevar <- array(0, dim = dimension)
Bvar <- array(0, dim = dimension)
Vvar <- array(0, dim = dimension)
t1 <- teorico$V1
t3 <- teorico$V3
t4 <- teorico$V4
e1 <- elec$V1
e3 <- elec$V3
e4 <- elec$V4
var1 <- varios$V1
var3 <- varios$V3
var4 <- varios$V4
for(j in 1:numz)
  for(i in 1:fil) {
    mseteo[i, j] <- t1[i + (j - 1) * fil]
    Bteo[i, j] <- t3[i + (j - 1) * fil]
    Vteo[i, j] <- t4[i + (j - 1) * fil]
    mseelec[i, j] <- e1[i + (j - 1) * fil]
    Belec[i, j] <- e3[i + (j - 1) * fil]
    Velec[i, j] <- e4[i + (j - 1) * fil]
    for(g in 1:numg) {
      msevar[i, g, j] <- var1[i + (g - 1) * fil + (
        numg * fil * (j - 1))]
      Bvar[i, g, j] <- var3[i + (g - 1) * fil + (
        numg * fil * (j - 1))]
      Vvar[i, g, j] <- var4[i + (g - 1) * fil + (
        numg * fil * (j - 1))]
    }
  }
res <- list(h = red, mseteo = mseteo, mseelec = mseelec,
  msevar = msevar, Bteo = Bteo, Vteo = Vteo, Belec = Belec,
  Velec = Velec, Bvar = Bvar, Vvar = Vvar)
}

pintar.componentes.m1 <- function(indz, teo50, teo200, teo500, 1boot50,

```

```
1boot200, 1boot500, 2boot50, 2boot200, 2boot500)
# Gráficos de las componentes de sesgo y de varianza.
{
zetas<-c(-0.5,0,1)
redg<-c(0.08,0.16,0.32,0.64,1.28,2)
numg<-length(redg)
numz <- length(zetas)
ftest<-1
C<- 1.1
inf<--2
sup<-2
tam<-50
dat <- leer.datos.graf(teo50, 1boot50, 2boot50, tam, inf, sup,
                      C, ftest, zetas)

h <- dat$h
Bteo50 <- dat$Bteo
Vteo50 <- dat$Vteo
mseteo50 <- dat$mseteo
Belec50 <- dat$Belec
Velec50 <- dat$Velec
mseelec50 <- dat$mseelec
Bvar50 <- dat$Bvar
Vvar50 <- dat$Vvar
msevar50 <- dat$msevar
tam<-200
dat <- leer.datos.graf(teo200, 1boot200, 2boot200, tam, inf,
                      sup, C, ftest, zetas)

Bteo200 <- dat$Bteo
Vteo200 <- dat$Vteo
mseteo200 <- dat$mseteo
Belec200 <- dat$Belec
Velec200 <- dat$Velec
mseelec200 <- dat$mseelec
Bvar200 <- dat$Bvar
Vvar200 <- dat$Vvar
msevar200 <- dat$msevar
tam<-500
dat <- leer.datos.graf(teo500, 1boot500, 2boot500, tam, inf,
                      sup, C, ftest, zetas)

Bteo500 <- dat$Bteo
Vteo500 <- dat$Vteo
```

```
mseteo500 <- dat$mseteo
Belec500 <- dat$Belec
Velec500 <- dat$Velec
mseelec500 <- dat$mseelec
Bvar500 <- dat$Bvar
Vvar500 <- dat$Vvar
msevar500 <- dat$msevar
par(mfrow=c(3,2),pin=c(3,2.3))
for(j in 1:numz)
{
  #para n=50
  #gráficos del sesgo teórico y bootstrap
  et1<-paste("B(", zetas[j],";h)")
  sesgo<- Belec50[, j]
  plot(h,sesgo, type = "l", lty=1, lwd=2, xlab="h", ylab=et1)
  titulo <- "Sesgo Bootstrap"
  title(main = titulo)
  sesgo<-Bteo50[,j]
  lines(h,sesgo, type="l",lty=2,lwd=2)
  #gráficos de la varianza teórica y bootstrap
  var<-Velec50[,j]
  et2<-paste("V(", zetas[j],";h)")
  plot(h,var, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
  titulo <- paste("Varianza Bootstrap (n=50)")
  title(main = titulo)
  var<-Vteo50[,j]
  lines(h,var, type="l",lty=2,lwd=2)
  #para n=200
  #gráficos del sesgo teórico y bootstrap
  et1<-paste("B(", zetas[j],";h)")
  sesgo<- Belec200[, j]
  plot(h,sesgo, type = "l", lty=1, lwd=2, xlab="h", ylab=et1)
  titulo <- "Sesgo Bootstrap"
  title(main = titulo)
  sesgo<-Bteo200[,j]
  lines(h,sesgo, type="l",lty=2,lwd=2)
  #gráficos de la varianza teórica y bootstrap
  var<-Velec200[,j]
  et2<-paste("V(", zetas[j],";h)")
  plot(h,var, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
  titulo <- paste("Varianza Bootstrap (n=200)")
```



```

        title(main = titulo)
        var<-Vteo200[,j]
        lines(h,var, type="l",lty=2,lwd=2)
        #para n=500
#gráficos del sesgo teórico y bootstrap con elección de g
        et1<-paste("B(", zetas[j],";h)")
        sesgo<- Belec500[, j]
        plot(h,sesgo, type = "l", lty=1, lwd=2, xlab="h", ylab=et1)
        titulo <- "Sesgo Bootstrap"
        title(main = titulo)
        sesgo<-Bteo500[,j]
        lines(h,sesgo, type="l",lty=2,lwd=2)
#gráficos de la varianza teórica y bootstrap con elección de g
        var<-Velec500[,j]
        et2<-paste("V(", zetas[j],";h)")
        plot(h,var, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
        titulo <- paste("Varianza Bootstrap (n=500)")
        title(main = titulo)
        var<-Vteo500[,j]
        lines(h,var, type="l",lty=2,lwd=2)
}
#gráficos del punto problema con otros g
#el punto problema es el de índice "indz"
j<-indz
par(mfrow=c(3,3),pin=c(2,2))
#para n=50
    for(g in 2:(numg-2))
        {
            #gráficos del sesgo
            et1<-paste("B(", zetas[j],";h)")
            sesgo<- Bvarios50[,g, j]
            plot(h,sesgo, type = "l", lty=1, lwd=2, xlab="h", ylab=et1)
            titulo <- paste("Sesgo Bootstrap g= ", redg[g])
            title(main = titulo)
            sesgo<-Bteo50[,j]
            lines(h,sesgo, type="l",lty=2,lwd=2)
            sesgo<-Belec50[,j]
            lines(h,sesgo, type="l",lty=3,lwd=1)
            #gráficos de la varianza
            var<-Vvarios50[,g,j]
            et2<-paste("V(", zetas[j],";h)")

```

```

plot(h,var, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
titulo <- paste("Varianza Bootstrap g= ", redg[g])
title(main = titulo)
var<-Vteo50[,j]
lines(h,var, type="l",lty=2,lwd=2)
var<-Velec50[,j]
lines(h,mse, type="l",lty=3,lwd=1)
#gráficos del MSE
mse<-msevarios50[,g,j]
et2<-paste("MSE(", zetas[j],";h)")
plot(h,mse, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
titulo <- paste("MSE Bootstrap g= ", redg[g])
title(main = titulo)
mse<-mseteo50[,j]
lines(h,mse, type="l",lty=2,lwd=2)
mse<-mseelec50[,j]
lines(h,mse, type="l",lty=3,lwd=1)
}
#para n=200
for(g in 2:(numg-2))
{
et1<-paste("B(", zetas[j],";h)")
sesgo<- Bvarios200[,g, j]
plot(h,sesgo, type = "l", lty=1, lwd=2, xlab="h", ylab=et1)
titulo <- paste("Sesgo Bootstrap g= ", redg[g])
title(main = titulo)
sesgo<-Bteo200[,j]
lines(h,sesgo, type="l",lty=2,lwd=2)
sesgo<-Belec200[,j]
lines(h,sesgo, type="l",lty=3,lwd=1)
#gráficos de la varianza
var<-Vvarios200[,g,j]
et2<-paste("V(", zetas[j],";h)")
plot(h,var, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
titulo <- paste("Varianza Bootstrap g= ", redg[g])
title(main = titulo)
var<-Vteo200[,j]
lines(h,var, type="l",lty=2,lwd=2)
var<-Velec200[,j]
lines(h,mse, type="l",lty=3,lwd=1)
#gráficos del MSE

```

```
mse<-msevarios200[,g,j]
et2<-paste("MSE(", zetas[j],";h)")
plot(h,mse, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
titulo <- paste("MSE Bootstrap g= ", redg[g])
title(main = titulo)
mse<-mseteo200[,j]
lines(h,mse, type="l",lty=2,lwd=2)
mse<-mseelec200[,j]
lines(h,mse, type="l",lty=3,lwd=1)
}
#para n=500
for(g in 2:(numg-2))
{
et1<-paste("B(", zetas[j],";h)")
sesgo<- Bvarios500[,g, j]
plot(h,sesgo, type = "l", lty=1, lwd=2, xlab="h", ylab=et1)
titulo <- paste("Sesgo Bootstrap g= ", redg[g])
title(main = titulo)
sesgo<-Bteo500[,j]
lines(h,sesgo, type="l",lty=2,lwd=2)
sesgo<-Belec500[,j]
lines(h,sesgo, type="l",lty=3,lwd=1)
var<-Vvarios500[,g,j]
et2<-paste("V(", zetas[j],";h)")
plot(h,var, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
titulo <- paste("Varianza Bootstrap g= ", redg[g])
title(main = titulo)
var<-Vteo500[,j]
lines(h,var, type="l",lty=2,lwd=2)
var<-Velec500[,j]
lines(h,mse, type="l",lty=3,lwd=1)
mse<-msevarios500[,g,j]
et2<-paste("MSE(", zetas[j],";h)")
plot(h,mse, type = "l", lty=1, lwd=2, xlab="h", ylab=et2)
titulo <- paste("MSE Bootstrap g= ", redg[g])
title(main = titulo)
mse<-mseteo500[,j]
lines(h,mse, type="l",lty=2,lwd=2)
mse<-mseelec500[,j]
lines(h,mse, type="l",lty=3,lwd=1)
}
```

```
}

pintar.mse.m1 <- function(teo50,teo200,teo500,1boot50,1boot200,
                        1boot500,2boot50,2boot200,2boot500)
# Gráficos del error cuadrático medio teórico y de la
# estimación bootstrap propuesta. (Función de prueba 1)
{
  inf<--2
  sup<-2
  C <- 1.1
  redg<-c(0.08,0.16,0.32,0.64,1.28,2)
  ftest<-1
  sigma<-0.3
  zetas<-c(-0.5,0,1)
  numz<-length(zetas)
  numg<-length(redg)
  #tamaño 50
  tam<-50
  dat<-leer.datos.graf(teo50, 1boot50, 2boot50, tam, inf, sup,
                      C, ftest, zetas)

  h<-dat$h
  mseteo<-dat$mseteo
  mseelec<-dat$mseelec
  msevar<-dat$msevar
  #gráficos de MSE para varios g, son gráficos 3 por 2
  par(mfrow=c(3,2),pin=c(3,2.3))
  for(j in 1:numz)
  {
    et<-paste("MSE(", zetas[j],";h)")
    for(g in 1:numg)
    {
      MSE<- msevar[, g, j]
      plot(h,MSE, type = "l", lty=1, lwd=2, xlab="h", ylab= et)
      titulo <- paste("MSE Bootstrap: g= ", redg[g])
      title(main = titulo)
      MSE<-mseteo[,j]
      lines(h,MSE,type="l",lty=2,lwd=2)
    }
  }
}
par(mfrow=c(3,3),pin=c(1.5,1.5))
for(j in 1:numz)
```

```
{
  et<-paste("MSE(", zetas[j],";h)")
  MSE<- mseelec[, j]
  pil<-piloto(ftest, tam, sigma, zetas[j], inf, sup)
  plot(h,MSE, type = "l", lty=1, lwd=2, xlab="h", ylab= et)
  titulo <-"MSE Bootstrap"
  title(main = titulo)
  MSE<-mseteo[,j]
  lines(h,MSE,type="l",lty=2,lwd=2)
}

#tamaño 200
tam<-200
dat<-leer.datos.graf(teo200, 1boot200, 2boot200, tam, inf,
                    sup, C, ftest, zetas)

h<-dat$h
mseteo<-dat$mseteo
mseelec<-dat$mseelec
msevar<-dat$msevar
for(j in 1:numz)
{
  et<-paste("MSE(", zetas[j],";h)")
  MSE<- mseelec[, j]
  pil<-piloto(ftest, tam, sigma, zetas[j], inf, sup)
  plot(h,MSE, type = "l", lty=1, lwd=2, xlab="h", ylab= et)
  titulo <- "MSE Bootstrap"
  title(main = titulo)
  MSE<-mseteo[,j]
  lines(h,MSE,type="l",lty=2,lwd=2)
}

#tamaño 500
tam<-500
dat<-leer.datos.graf(teo500, 1boot500, 2boot500, tam, inf,
                    sup, C, ftest, zetas)

h<-dat$h
mseteo<-dat$mseteo
mseelec<-dat$mseelec
msevar<-dat$msevar
for(j in 1:numz)
{
  et<-paste("MSE(", zetas[j],";h)")
  MSE<- mseelec[, j]
```

```

    pil<-piloto(ftest, tam, sigma, zetas[j], inf, sup)
    plot(h,MSE, type = "l", lty=1, lwd=2, xlab="h", ylab= et)
    titulo <- "MSE Bootstrap"
    title(main = titulo)
    MSE<-mseteo[,j]
    lines(h,MSE,type="l",lty=2,lwd=2)
  }
#gráficos como los primeros pero para los tamaños 200 y 500
#tamaño 200
  tam<-200
  dat<-leer.datos.graf(teo200, 1boot200, 2boot200, tam, inf,
sup, C, ftest, zetas)
  h<-dat$h
  mseteo<-dat$mseteo
  mseelec<-dat$mseelec
  msevar<-dat$msevar
#gráficos de MSE para varios g, son gráficos 3 por 2
  par(mfrow=c(3,2),pin=c(3,2.3))
  for(j in 1:numz)
  {
    et<-paste("MSE(", zetas[j],";h)")
    for(g in 1:numg)
    {
      MSE<- msevar[, g, j]
      plot(h,MSE, type = "l", lty=1, lwd=2, xlab="h", ylab= et)
      titulo <- paste("MSE Bootstrap: g= ", redg[g])
      title(main = titulo)
      MSE<-mseteo[,j]
      lines(h,MSE,type="l",lty=2,lwd=2)
    }
  }
}
#tamaño 500
  tam<-500
  dat<-leer.datos.graf(teo500, 1boot500, 2boot500, tam, inf,
sup, C, ftest, zetas)
  h<-dat$h
  mseteo<-dat$mseteo
  mseelec<-dat$mseelec
  msevar<-dat$msevar
  par(mfrow=c(3,2),pin=c(3,2.3))
  for(j in 1:numz)

```

```
{
  et<-paste("MSE(", zetas[j],";h)")
  for(g in 1:numg)
  {
    MSE<- msevar[, g, j]
    plot(h,MSE, type = "l", lty=1, lwd=2, xlab="h", ylab=et)
    titulo <- paste("MSE Bootstrap: g= ", redg[g])
    title(main = titulo)
    MSE<-mseteo[,j]
    lines(h,MSE,type="l",lty=2,lwd=2)
  }
}
```





## Bibliografía

- [1] **Altman, N.S.** (1992) An Introduction to Kernel and Nearest-Neighbourhood Nonparametric Regression. *The American Statistician*, Vol. 46, No. 3, pp. 175-185.
- [2] **Benedetti, J.K.** (1977) On the Nonparametric Estimation of Regression Functions. *Journal of the Royal Statistical Society, Ser. B*, Vol. 39, pp. 248-253.
- [3] **Brockman, M., Gasser, T. and Herrmann, E.** (1993) Locally Adaptive Bandwidth Choice for the Kernel Regression Estimators. *Journal of the American Statistical Association*, Vol. 88, No. 424. Theory and Methods, pp. 1302-1309.
- [4] **Cao, R.** (1993) Bootstrapping the Mean Integrated Squared Error. *Journal of Multivariate Analysis*, Vol. 45, No. 1, pp. 137-160.
- [5] **Cao, R.** (1999) An overview of bootstrap methods for estimating and predicting in time series. *Test*, Vol. 8, No. 1, pp. 96-116.
- [6] **Cao, R.; Cuevas, A. and Manteiga, W.G.** (1994) A comparative study of several smoothing methods in density estimation. *Computational Statistics and Data Analysis*, 17, pp. 153-176.
- [7] **Cheng, M.Y., Fan, J. and Marron, J.S.** (1993) Minimax efficiency of local polynomial fit estimators at boundaries. *Institute of Statistics Mimeo Series #2098*, University of North Carolina at Chapel Hill.
- [8] **Chiu, S. T.** (1991) Bandwidth Selection for Kernel Density Estimation. *The Annals of Statistics*, 19, No. 4, pp. 1883-1905.
- [9] **Chu, C.K. and Marron, J.S.** (1991) Choosing a Kernel Regression Estimator. *Statistical Science*, Vol. 6, No. 4, pp. 404-436.
- [10] **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.

- [11] **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.
- [12] **Cleveland, W.S. and Devlin, S.** (1988) Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, Vol. 83, No. 403. Theory and Methods, pp. 596-610.
- [13] **Eubank, R.L.** (1988) *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.
- [14] **Efron, B. and Tibshirani, R.J.** (1993) *An Introduction to the Bootstrap*. Ed. Chapman and Hall.
- [15] **Eubank, R.L. and Spiegelman, C.H.** (1990) Testing the Goodness of Fit of a Linear Model Via Nonparametric Regression Techniques. *Journal of the American Statistical Association*, Vol. 85, No. 410. Theory and Methods, pp. 387-392.
- [16] **Fan, J.** (1992) Design-adaptive Nonparametric Regression. *Journal of the American Statistical Association*, Vol. 87, No. 420, Theory and Methods, pp. 998-1004.
- [17] **Fan, J.** (1993) Local linear regression smoothers and their minimax efficiency. *The Annals of Statistics*, Vol. 21, pp. 600-610.
- [18] **Fan, J., Farmen, M. and Gijbels, I.** (1998) Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society B*, 60, No. 3, pp. 591-608.
- [19] **Fan, J. and Gijbels, I.** (1992) Variable Bandwidth and local lineal regression smoothers. *The Annals of Statistics*, Vol. 20, No. 4, pp. 2008-2036.
- [20] **Fan, J. and Gijbels, I.** (1995a) Adaptive Order Polynomial Fitting: Bandwidth Robustification and Bias Reduction. *Journal of the Computational and Graphical Statistics*, Vol. 4, No. 3, pp. 213-227.
- [21] **Fan, J. and Gijbels, I.** (1995b) 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.
- [22] **Fan, J., Gijbels, I. and Hu, T.-C. and Huang, L.-S.** (1996) An asymptotic study of variable bandwidth selection for local polynomial regression with application to density estimation. *Statistica Sinica*, Vol. 6, No. 1.

- [23] **Fan, J., Hall, P., Martin, M.A. and Patil, P.** (1996) On Local Smoothing of Nonparametric Curve Estimators. *Journal of the American Statistical Association*, Vol. 91, No. 433, pp. 258-266.
- [24] **Fan, J. and Marron, J.S.** (1994) Fast Implementations of Nonparametric Curve Estimators. *Journal of Computational and Graphical Statistics*, Vol. 3, No. 1, pp. 35-56.
- [25] **Farmen, M. and Marron, J.S.** (1999) An assessment of finite sample performance of adaptive methods in density estimation. *Computational Statistics and Data Analysis*, 30, pp. 143-168.
- [26] **Gasser, T., Kneip, A. and Köhler, W.** (1991) A flexible and fast method for automatic Smoothing. *Journal of the American Statistical Association*, Vol. 86, No. 415. Theory and Methods, pp. 643-652.
- [27] **González-Manteiga, W., Sánchez-Sellero, C. and Wand, M.P.** (1996) Accuracy of binned kernel functional approximations. *Computational Statistics and Data Analysis*, 22, pp. 1-16.
- [28] **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.
- [29] **Hall, P. and Johnstone, I.** (1992) Empirical Functionals and Efficient Smoothing Parameter Selection. *Journal of the Royal Statistic Society*, 54, No. 2, pp. 475-530.
- [30] **Hall, P., Marron, J.S. and Park, B.U.** (1992) Smoothed cross-validation. *Probability Theory and Related Fields*, Vol. 92, pp. 1-20.
- [31] **Härdle, W.** (1990) *Applied nonparametric regression*. Cambridge University Press, Boston.
- [32] **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.
- [33] **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.
- [34] **Härdle, W. and Marron, J.S.** (1995) Fast and simple scatterplot smoothing. *Computational Statistics and Data Analysis*, 20, pp. 1-17.

- [35] Härdle, W., Hall, P. and Marron, J.S. (1988) 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.
- [36] Härdle, W., Hall, P. and Marron, J.S. (1992) Regression Smoothing Parameters That Are Not Far From Their Optimum. *Journal of the American Statistical Association*, Vol. 87, No. 417. Theory and Methods, pp. 227-233.
- [37] Hart, J.D. and Yi, S. (1998) One-Sided Cross-Validation. *Journal of the American Statistical Association*, Vol. 93, No. 442. Theory and Methods, pp. 620-631.
- [38] Hastie, T. and Loader, C. (1993) Local Regression: Automatic Kernel Carpentry. *Statistical Science*, Vol. 8, No. 2, pp.120-143.
- [39] Herrmann, E. and Wand, M.P. (1995) A bandwidth selector for bivariate kernel regression. *Journal of the Royal Statistical Society, B*, 57, No. 1, pp. 171-180.
- [40] Jones, M.C., Marron, J.S. and Park, B.U. (1991) A simple root  $n$  bandwidth selector. *The Annals of Statistics*, 19, No. 4, pp. 1919-1932.
- [41] Jones, M.C., Marron, J.S. and Sheather, S.J. (1996) Progress in Data-Based Bandwidth Selection for Kernel Density Estimation. *Computational Statistics*, 11, pp. 337-381.
- [42] Kim, W.C., Park, B.U. and Marron, J.S. (1994) Asymptotically best bandwidth selectors in kernel density estimation. *Statistics and Probability Letters*, 19, pp. 119-127.
- [43] Lee, T. and Solo, V. (1999) Bandwidth Selection for Local Linear Regression: A Simulation Study. *Computational Statistics*.
- [44] Loader, C.R. (1996) *Local Regression and Likelihood*. Electronic book, <http://cm.bell-labs.com/stat/project/locfit/>
- [45] Loader, C.R. (1999) Bandwidth Selection: Classical or Plug-In? *The Annals of Statistics*, 27, No. 2, pp 415-438.
- [46] Magee, L. (1998) Nonlocal Behavior in Polynomial Regressions. *American Statistical Association*, Vol. 52, No. 1, pp. 20-22.
- [47] Martínez-Miranda, M.D. (1996) *Técnicas de Regresión No Paramétrica*. Memoria de Licenciatura.

- [48] **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).
- [49] **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).
- [50] **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*. Utrecht (The Netherlands).
- [51] **Marron, J.S.** (1992) Discussion of: "The Performance of Six Popular Bandwidth Selection Methods on Some Real Data Sets" by Sheather. *Computational Statistics*, 2, pp. 272-273.
- [52] **Marron, J.S.** (1996) A personal view of smoothing and statistics. In *Statistical Theory and Computational Aspects of Smoothing* (W. Härdle and M.G. Schimek eds.) 1-9. Physica, Heidelberg.
- [53] **Nychka, D.** (1995) Splines as Local Smoothers. *The Annals of Statistics*, Vol. 23, No. 4, pp. 1175-1197.
- [54] **Park, B.U. and Marron, J.S.** (1990) Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85, No. 409. Theory and Methods, pp. 66-72.
- [55] **Park, B.U. and Turlach, B.A.** (1992) Practical Performance of Several Data Driven Bandwidth Selectors. *Computational Statistics*, 7, pp. 251-270.
- [56] **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.
- [57] **Ruppert, D.** (1997) Empirical-Bias Bandwidths for Local Polynomial 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. Sección de Teoría y Métodos.

- [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] **Schucany, W.R.** (1995) Adaptive Bandwidth Choice for Kernel Regression. *Journal of the American Statistical Association*, 90, No. 430. Theory and Methods, pp. 535-540.
- [61] **Scott, D.W. and Terrell, G.R.** (1987) Biased and Unbiased Cross-validation in Density Estimation. *Journal of the American Statistical Association*, 82, No. 400. Theory and Methods, pp. 1131-1146.
- [62] **Seifert, B., Brockmann, M., Engel, J. and Gasser, T.** (1994) Fast Algorithms for Nonparametric Curve Estimation. *Journal of Computational and Graphical Statistics*, Vol. 3, No. 2, pp. 192-213.
- [63] **Sheater, S.J. and Jones, M.C.** (1991) A Reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society B*, 53, No. 3, pp. 683-690.
- [64] **Sheater, S.J.** (1992) The Performance of Six Popular Bandwidth Selection Methods on Some Real Data Sets. *Computational Statistics*, 7, pp. 225-250.
- [65] **Silverman, B.W.** (1984) Spline Smoothing: The Equivalent Variable Kernel Method. *The Annals of Statistics*, Vol. 12, No. 3, pp. 898-916.
- [66] **Stone, C.J.** (1974) Cross-validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society, Ser. B*, Vol. 36, pp. 111-147.
- [67] **Stone, C.J.** (1977) Consistent nonparametric regression. *The Annals of Statistics*, Vol. 5, No. 2, pp. 595-645.
- [68] **Stone, C.J.** (1980) Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, Vol. 8, No. 4, pp. 1348-1360.
- [69] **Stone, C.J.** (1982) Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, Vol. 10, No. 4, pp. 1040-1053.
- [70] **Vieu, Ph.** (1991) Nonparametric Regression: Optimal Local Bandwidth Choice. *Journal of the Royal Statistical Society, B*, 53, No. 2, pp. 453-464.
- [71] **Wand, M.P.** (1994) Fast Computation of Multivariate Kernel Estimators. *Journal of Computational and Graphical Statistics*, Vol. 3, No. 4, pp. 433-445.
- [72] **Wand, M.P. and Jones, M.C.** (1993) Comparison of smoothing parameterizations in bivariate Kernel density estimation. *Journal of the American Statistical Association*, Vol. 88, No. 422. Theory and Methods, pp. 520-528.

- 
- [73] **Wand, M.P. and Jones, M.C.** (1994) Multivariate Plug-in Bandwidth Selection. *Computational Statistics*, 9. pp. 97-116.
- [74] **Wand, M.P. and Jones, M.C.** (1995) *Kernel Smoothing*. Monographs on Statistics and Applied Probability 60. Ed. Chapman and Hall.
- [75] **Werthenbach, C. and Herrmann, E.** (1998) A Fast and Stable Updating Algorithm for Bivariate Nonparametric Curve Estimation. *Journal of Computational and Graphical Statistics*, Vol. 7, No. 1, pp. 61-76.



Biblioteca Universitaria de Granada



01066670