
REGRESION NO-PARAMETRICA ROBUSTA

Contribución a la Teoría de la Estimación Robusta de Escala en
Modelos de Regresión No-paramétricos

por

Marcelo Ruiz

Licenciado en Matemática, Universidad Nacional de Río Cuarto, Argentina.

Magíster en Estadística, Universidad Nacional de Córdoba, Argentina.

TESIS PRESENTADA PARA OPTAR AL TITULO DE
Doctor en Matemáticas

en el

Departamento de Matemáticas

Facultad de Ciencias Físico Matemáticas y Naturales

UNIVERSIDAD NACIONAL DE SAN LUIS

ARGENTINA

16 de Septiembre de 2008

Asesora Científica:
Doctora Graciela Boente Boente
Instituto de Cálculo
Universidad Nacional de Buenos Aires
Argentina

Co-asesores Científicos:
Doctor Ruben Zamar
Statistics Department
University of British Columbia
Canada

Doctor Juan Cesco
Departamento de Matemáticas
Universidad Nacional de San Luis
Argentina

Contribución a la Teoría de la Estimación Robusta de Escala en Modelos de Regresión No-paramétricos

Cuando los datos utilizados para ajustar un modelo de regresión no-paramétrico homoscedástico están contaminados con valores atípicos, es necesario utilizar estimadores robustos de la escala para obtener estimadores robustos de la función de regresión. En esta tesis proponemos, para diseños fijos, M -estimadores de escala construidos en base a las diferencias sucesivas de las variables respuesta. Estos estimadores son una versión robusta del estimador propuesto por Rice (1984). Bajo condiciones de regularidad, obtenemos consistencia y normalidad asintótica de los M -estimadores de escala. Estos estimadores varían en términos de sus propiedades de robustez. Por esta razón, cuantificamos la robustez de un estimador por la vía del sesgo máximo y demostramos que los M -estimadores alcanzan el máximo punto de ruptura asintótico de $1/2$.

Los resultados previos se generalizan a modelos de regresión no-paramétricos heteroscedásticos, introduciendo la familia de los M -estimadores locales de escala. Establecemos también la consistencia y normalidad asintótica de esta familia. Estudiamos además el comportamiento asintótico de los M -estimadores locales de escala, cuando están basados en núcleos, para ventanas adaptivas.

Finalmente llevamos a cabo un estudio de simulación para comparar el desempeño de los M -estimadores, sobre muestras finitas, en relación a otros estimadores.

Palabras Claves: regresión no-paramétrica, diseño fijo, homoscedasticidad, heteroscedasticidad, robustez, M -estimadores, máximo sesgo asintótico, punto de ruptura asintótico.

Contribution to The Theory of Scale Robust Estimation in Nonparametric Regression Models

When the data used to fit a homoscedastic nonparametric regression model are contaminated with outliers, robust scale estimators are needed in order to obtain robust estimation of the regression function. In this thesis we develop, for fixed designs, M -estimators of scale constructed from consecutive differences of the responses. Our proposal robustifies the estimator considered by Rice (1984). Under regularity conditions, we derive consistency results and asymptotic normality of the M -estimators. The estimators in this family vary in terms of their robustness properties. We quantify the robustness of each estimator throughout the asymptotic maximum bias and we prove that M -estimators achieve the maximum asymptotic breakdown point of $1/2$.

The previous results are generalized to heteroscedastic nonparametric regression models, introducing a family of local scale M -estimators of scale. We also obtain consistency and asymptotic normality results for this family. We study the asymptotic behaviour of the local M -estimators of scale based on kernel weights using random bandwidths.

Finally, we conduct a simulation study to compare the finite sample performance of the M -estimators, with that of other competitors.

Keywords and phrases: nonparametric regression, fixed design, homoscedasticity, heteroscedasticity, robustness, M -estimators, maxbias, asymptotic breakdown point.

Agradecimientos

A todos aquellos que contribuyeron al proceso de aprender. Desde mis primeros maestros de la Escuela Primaria, hasta los actuales. Agradecimiento a la Escuela Pública Argentina, a los sujetos concretos que la construyen cotidianamente.

A Graciela Boente, Ruben Zamar, Jorge Adrover y Oscar Bustos por su compromiso y generosidad en el proceso (nunca acto) de construcción del conocimiento. A Juan Cesco por su predisposición permanente.

A mi viejo Iván, mi vieja Magel, mis hermanos Miriam y Néstor, mis sobrinos María José, Diego, Romina, Mauro, Esteban, Martina (la dulce), mis otros hermanos Tatano y Marité, por enseñarme el fuerte compromiso que implica el amor.

A mis otras familias, Marina, Jorge, Patricia, Susi, Elisa, Juan, Ignacio, Luis, Paola, Soe, Tomás, Juani y Agustina. Agradecimiento por su entrega y amor. Por compartir el abrazo y la utopía.

A mis compañeros del Departamento de Matemática de la U.N.R.C. Por su permanente apoyo.

A mis amigos que son muchos, compañeros de la tarea del transformar y transformarnos. No es del número de lo que hablamos sino de la cualidad del compartir.

A mis nuevos compañeros de la UBA, que con gran sutileza en los últimos días construyeron cariño y apoyatura: Dani, Juli, Lucas y Mariela.

A Germán (hermano), Patricia , Miguel, Carmen, Sandra por enseñarme lo esencial.

A Rho, quien me devolvió color al sentido. Sentido que sin color está vacío.

Índice general

1. Introducción	1
2. Elementos Generales de Teoría Asintótica	9
3. Estimación de la escala global	17
3.1. Preliminares	17
3.2. Teoría Asintótica de los M -estimadores de la Escala Global	25
3.2.1. Consistencia y Normalidad asintótica	25
3.3. Robustez de los M -estimadores de Escala	36
3.3.1. Máximo Sesgo Asintótico	38
3.3.2. Punto de Ruptura Asintótico	50
3.4. Resultados Monte Carlo	53
3.4.1. Análisis y Conclusiones	53
4. Estimación de la escala local	61
4.1. Preliminares	61
4.2. Teoría Asintótica de los M -estimadores locales de la función de Escala	66
4.3. Resultados Asintóticos para Estimadores de Núcleos con Ventana Adaptiva.	84
4.4. Resultados Monte Carlo	97
4.4.1. Problema y objetivos	97
4.4.2. Análisis y Conclusiones.	102
5. Conclusiones	106
A. Resultados Auxiliares	109

B. Cuadros y Gráficos	121
B.1. Correspondientes a los Resultados de la Simulación Monte Carlo para Modelos Homoscedásticos	121
B.2. Correspondientes a los Resultados de la Simulación Monte Carlo para Modelos Heteroscedásticos	129
Bibliografía	164

Índice de figuras

B.1. Datos simulados del modelo de contaminaciones independientes con $n = 100$ y contaminación simétrica $H(y) = \Phi(y/10)$ para diferentes valores de ϵ	123
B.2. Datos simulados del modelo de contaminaciones independientes con $n = 100$ y contaminación simétrica $H(y) = \Phi(y - 10)$ para diferentes valores de ϵ	124
B.3. Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ y en ausencia de contaminación ($\epsilon = 0$).	125
B.4. Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ y contaminación simétrica $H(y) = \Phi(y/10)$ según proporción de contaminación ϵ	126
B.5. Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ y contaminación asimétrica $H(y) = \Phi(y - 10)$ según proporción de contaminación ϵ	127
B.6. Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ según proporción ϵ de observaciones contaminadas intercaladas con $H(y) = \Delta_{10}$	128
B.7. Valores de las respuestas versus puntos del diseño del modelo $M1$ con funciones componentes $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.	136
B.8. Valores de las respuestas versus puntos del diseño del modelo $M2$ con funciones componentes $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.	137
B.9. Valores de las respuestas versus puntos del diseño del modelo $M3$ con funciones componentes $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.	138
B.10. Valores de las respuestas versus puntos del diseño del M.E.G. con funciones componentes $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = 1$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.	139
B.11. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	141

B.12.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	142
B.13.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	143
B.14.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	144
B.15.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	145
B.16.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	146
B.17.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	147
B.18.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	148
B.19.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	149
B.20.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo M.E.G. (con $g(x) = 1 + x$ y $\sigma(x) \equiv 1$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	150

B.21. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo M.E.G. (con $g(x) = 1 + x$ y $\sigma(x) \equiv 1$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	151
B.22. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo M1 (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	152
B.23. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo M1 (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	153
B.24. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo M1 (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	154
B.25. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo M2 (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	155
B.26. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo M2 (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	156
B.27. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo M2 (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	157
B.28. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo M3 (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	158
B.29. Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo M3 (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.	159

B.30.	Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$	160
B.31.	Modelo $M2$. (a) Gráfico de $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. (b), (c) y (d) corresponden a los gráficos de E.C.M.L. (x_i, h) versus x_i para $h = 0.04, 0.2$ y 0.5 , respectivamente con $\epsilon = 0$. La línea sólida de trazo fino corresponde al E.C.M.L. de $\hat{\sigma}_{R,n}$, la línea cortada (--) al de $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso al de $\hat{\sigma}_{BT,n}$	162
B.32.	Modelo $M2$. (a) Gráfico de $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. (b), (c) y (d) corresponden a los gráficos de E.C.M.L. (x_i, h) versus x_i para $h = 0.04, 0.2$ y 0.5 , respectivamente con contaminación simétrica y $\epsilon = 0.20$. La línea sólida de trazo fino corresponde al E.C.M.L. de $\hat{\sigma}_{R,n}$, la línea cortada (--) al de $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso al de $\hat{\sigma}_{BT,n}$	163

Índice de cuadros

B.1. Eficiencias $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ relativas a $\hat{\sigma}_{\text{R},n}$ con $\epsilon = 0$	121
B.2. Valores estimados del E.C.M. de $\hat{\sigma}_{\text{R},n}$, $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$. La contaminación es simétrica con $H(y) = \Phi(y/10)$	121
B.3. Valores estimados del E.C.M. de $\hat{\sigma}_{\text{R},n}$, $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$. La contaminación es asimétrica con $H(y) = \Phi(y - 10)$	122
B.4. Valores estimados de E.C.M. de los estimadores $\hat{\sigma}_{\text{R},n}$, $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ con contaminaciones intercaladas y $H(y) = \Delta_{10}(y)$	122
B.5. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.04$. Modelo $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Contaminación simétrica.	129
B.6. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Contaminación simétrica.	129
B.7. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Contaminación simétrica.	129
B.8. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.04$. Modelo $M2 : g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Contaminación simétrica.	130
B.9. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo $M2 : g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Contaminación simétrica.	130
B.10. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo $M2 : g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Contaminación simétrica.	130
B.11. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.04$. Modelo $M3 : g(x) = 1 + x$, $\sigma(x) = (1 + x)^2$. Contaminación simétrica.	131
B.12. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo $M3 : g(x) = 1 + x$, $\sigma(x) = (1 + x)^2$. Contaminación simétrica.	131
B.13. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo $M3 : g(x) = 1 + x$, $\sigma(x) = (1 + x)^2$. Contaminación simétrica.	131
B.14. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo M.E.G.: $g(x) = 1 + x$, $\sigma(x) \equiv 1$. Contaminación simétrica.	132

B.15. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo M.E.G.: $g(x) = 1 + x$, $\sigma(x) \equiv 1$. Contaminación simétrica.	132
B.16. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Modelo $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. $n = 100$. Contaminación asimétrica.	133
B.17. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Modelo $M2 : g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. $n = 100$. Contaminación asimétrica.	133
B.18. Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Modelo $M3 : g(x) = 1 + x$, $\sigma(x) = (1 + x)^2$. $n = 100$. Contaminación asimétrica.	134

Capítulo 1

Introducción

Consideremos el siguiente modelo de regresión no-paramétrico

$$Y_i = g(x_i) + U_i\sigma(x_i), \quad i = 1, \dots, n, \quad (1.1)$$

donde $\{Y_i\}_{i=1}^n$ son las respuestas observadas, $\{x_i\}_{i=1}^n$ son puntos de diseño fijo (valores reales, no aleatorios, en el intervalo $[0, 1]$), $g : [0, 1] \rightarrow \mathbb{R}$ es la función de regresión, desconocida, $\{U_i\}_{i=1}^n$ representan los errores y son v.a.'s i.i.d. (variables aleatorias independientes e idénticamente distribuidas) con distribución común F_0 . La función de escala, $\sigma : [0, 1] \rightarrow \mathbb{R}^+$, donde $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$, es también desconocida.

Tanto para la función de regresión como para la de escala no se supone ningún modelo paramétrico y el objetivo de este trabajo consiste en obtener estimadores robustos de la función de escala, que no dependan de una estimación previa de la función de regresión.

Vale la pena observar que, si el error U tiene media cero y varianza finita entonces, dado $x \in [0, 1]$, el modelo $Y = g(x) + U\sigma(x)$ satisface que $E(Y) = g(x)$ y $Var(Y) = \sigma^2(x)$. Por

esta razón, nuestro trabajo extiende a un contexto más general, que no requiere la existencia de momentos, las propuestas dadas previamente por varios autores.

Ya sea cuando la función de escala $\sigma(x) \equiv \sigma$ es constante (modelos homoscedásticos), o bien cuando dicha función no lo es (modelos heteroscedásticos), el problema de la estimación de $\sigma(x)$ es importante por diferentes razones.

Además de ser $\sigma(x)$ o $\sigma^2(x)$ un parámetro de interés por sí mismo, hallar estimadores de la función de escala es esencial para obtener estimadores robustos puntuales de la función de regresión (Buckley y Eagleson, 1988) como también intervalos de confianza para la misma. En el contexto heteroscedástico, surgen problemas reales donde la estimación de $\sigma(x)$ es importante ya sea, porque es intrínsecamente una función de interés, o bien porque en los datos se presentan violaciones severas al supuesto de homoscedasticidad. Un ejemplo de la primera situación se presenta en Ruppert *et al.* (1997). Dichos autores mencionan (y estudian) cómo en modelos de turbulencia además de la necesidad de estimar esperanzas condicionales tales como la velocidad media de las partículas, también son esenciales la varianza de la velocidad y sus derivadas. Un ejemplo de la segunda situación se da en Holst *et al.* (1996) (discutido también en Ruppert *et al.* , 1997). En este trabajo, utilizando regresión polinómica local, se evalúa la concentración de mercurio atómico en la atmósfera. La concentración es proporcional a la derivada de la función de regresión pero, debido a la clara heteroscedasticidad, se debe estimar la función de varianza (los autores lo hicieron paramétricamente) para obtener aberturas de ventana satisfactorias para la estimación de la derivada y, posteriormente, para estimar la varianza de la cantidad de contaminante en determinada área.

Otros campos donde la estimación de $\sigma^2(x)$ importa son, por ejemplo, control de calidad, ensayos inmunológicos, calibración y estimación de límites de detección (ver Hall y Carroll, 1989). Para una lista más exhaustiva de aplicaciones se puede consultar a Carrol y Ruppert (1988) y Brown y Levine (2007).

Haremos una breve reseña histórico–conceptual de las diferentes propuestas de resolución del problema de la estimación de la función de escala cuando es constante ($\sigma(x) \equiv \sigma$). Como se sostiene en Dette *et al.* (1998) la clase de estimadores de σ^2 que cubre la mayoría de las aproximaciones clásicas al problema de la estimación se basa en formas cuadráticas del vector de respuestas; es decir, los estimadores son de la forma $\hat{\sigma}_n^2 = \mathbf{Y}'\mathbf{D}\mathbf{Y}/tr(\mathbf{D})$ donde \mathbf{D} es una matriz conveniente, $\mathbf{Y} = (Y_1, \dots, Y_n)$ es el vector de respuestas y $tr(\mathbf{D})$ denota la traza de la matriz.

A su vez, dicha clase puede dividirse en dos grandes subclases. En una de ellas los estimadores son, o bien de *tipo spline*, o bien *basados en núcleos* donde la propuesta consiste en estimar σ^2 con una suma de cuadrados de residuos provenientes de un ajuste no–paramétrico. En esta dirección y en el contexto de los modelos homoscedásticos, se encuentran los trabajos de Wahba (1978), Hall y Carroll (1989), Hall y Marron (1990), Buckley *et al.* (1989) y Carter y Eagleson (1992).

El inconveniente de estos estimadores (Dette *et al.* 1998) es que dependen explícitamente de la elección de un parámetro de suavizado, lo que requiere conocer el cociente entre la variación de la función de regresión g , medida a través de $\int_0^1 [g''(t)]^2 dt$ y de σ^2 , con lo cual, desde un punto de vista práctico, su uso es limitado.

Otra de las subclases, que es la de nuestro interés, está compuesta por estimadores *basados en diferencias*. La primer propuesta de estimadores de este tipo fue dada por Rice (1984) quien propuso estimar σ^2 a través de

$$\hat{\sigma}_{R,n}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2.$$

Posteriormente Gasser et. al (1986) introdujeron la siguiente extensión a 3 puntos vecinos

$$\hat{\sigma}_{G,n}^2 = \frac{2}{3(n-2)} \sum_{i=2}^{n-1} \left(\frac{1}{2}Y_{i-1} - Y_i - \frac{1}{2}Y_{i+1} \right)^2,$$

y, finalmente, Hall et al. (1990) generalizaron los estimadores descritos anteriormente, construyendo los *estimadores basados en diferencias de orden r*, definidos como

$$\hat{\sigma}_{H,n}^2 = \frac{1}{(n-r)} \sum_{i=m_1+1}^{n-m_2} \left(\sum_{k=-m_1}^{m_2} d_k Y_{i+k} \right)^2,$$

donde $\{d_i\}_{i=-m_1}^{m_2}$ es una sucesión en \mathbb{R} tal que $\sum_i d_i = 0$ y $\sum_i d_i^2 = 1$, donde $d_{-m_1}, d_{m_2} \neq 0$, m_1 y m_2 son enteros no negativos y $r = m_1 + m_2$ denota el orden del estimador. Así el estimador de Rice (1984) es de orden 1 con $d_0 = 2^{-1/2}$ y $d_1 = -2^{-1/2}$, mientras que el de Gasser et al. (1986) es de orden 2 con $d_{-1} = -6^{-1/2}$, $d_0 = \left(\frac{2}{3}\right)^{1/2}$ y $d_1 = -6^{-1/2}$.

Los estimadores basados en diferencias no alcanzan la tasa de convergencia óptima, en el sentido del error cuadrático medio, que sí es alcanzada por algunas de las familias de la subclase anterior basada en residuos de un ajuste no-paramétrico previo. Sin embargo, entre

sus ventajas podemos mencionar que tienen menor sesgo para muestras de tamaño pequeño y sobre todo que, dada la elementalidad de su cálculo, requieren de un esfuerzo computacional mucho menor.

Los dispositivos de estimación de la escala global pueden extenderse a modelos donde la escala depende de los puntos del diseño (modelos de escala local). En este sentido recomendamos consultar los trabajos de Müller y Stadtmüller (1987) y, más recientemente, de Brown y Levine (2007).

Todas las familias de estimadores presentadas hasta ahora tienen un buen desempeño (en diferentes sentidos de optimalidad) bajo el supuesto de que los errores siguen el modelo central F_0 (sea éste normal o no), es decir $U_i \sim F_0$, $i = 1, \dots, n$. Sin embargo cuando una fracción de los errores sigue una distribución diferente a la del modelo central (introduciendo outliers o datos atípicos en los datos) dichos procedimientos se vuelven altamente sensibles provocando que la estimación se aleje considerablemente del verdadero valor del parámetro. Es por ello que es necesario desarrollar procedimientos que sean resistentes o robustos a la presencia de outliers.

En el contexto no-paramétrico los estimadores robustos de escala son, además, necesarios, por ejemplo, para computar estimadores robustos para la propia función de regresión; ver Härdle y Gasser (1984), Härdle y Tsybakov (1988), Boente y Fraiman (1989), y las referencias que allí se mencionan. Otras aplicaciones incluyen la detección de outliers como en el trabajo de Hannig y Lee (2006), la selección robusta de la abertura de ventana para mejorar la precisión de la estimación de la función de regresión; ver los trabajos de Boente et. al (1997), Leung,

Marriott y Wu (1993), Cantoni y Ronchetti (2001) y Leung (2005).

Boente *et al.* (1997) introducen, en el contexto de la selección robusta de la abertura de ventana el siguiente estimador de la escala global σ

$$\widehat{\sigma}_{\text{BFM},n} = \frac{q_{1/2}}{\sqrt{2}\Phi^{-1}(3/4)},$$

donde $q_{1/2}$ es el cuantil $\frac{1}{2}$ o mediana de las diferencias absolutas $|Y_{i+1} - Y_i|$, $i = 1, \dots, n$.

En la presente tesis proponemos una generalización de este estimador de la escala global, que denominaremos *M-estimador de escala basados en diferencias* definidos como solución de ecuaciones del tipo

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \chi \left(\frac{Y_{i+1} - Y_i}{as} \right) = b$$

donde $\chi : \mathbb{R} \longrightarrow \mathbb{R}_0^+ = [0, \infty)$ es una función de escores y a y b son constantes de ajuste cuyas propiedades serán especificadas luego.

Originariamente, los *M-estimadores de escala* (aunque en otro contexto) fueron introducidos por Huber (1964); ver también Martin y Zamar (1989).

Cuando la escala es local, para estimar $\sigma(x)$, con $x \in (0, 1)$, introducimos los *M-estimadores de la escala local basados en diferencias* definidos, análogamente, como soluciones de

la forma

$$\sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{Y_{i+1} - Y_i}{as} \right) = b$$

donde χ , a y b son como antes y $\{w_{n,i}(x)\}_{i=1}^n$ es una familia de funciones de pesos que ponderan más las observaciones correspondientes a los x_i 's cercanos a x , tales como los pesos basados en núcleos o en vecinos más cercanos.

Son varios los motivos por los cuales es importante contar con estimadores preliminares de escala en lugar de estimar conjuntamente la función de regresión y la de escala. Una de las razones es el aumento de la complejidad ya que se requiere hallar la solución de un sistema de ecuaciones no lineales. Una segunda razón, particularmente importante en el contexto heteroscedástico, es la posible pérdida de robustez de los estimadores de la función de regresión si se estima conjuntamente con la escala. Esta conjetura se basa en el hecho de que en el modelo de posición–escala $Y = \mu + \sigma U$ la estimación simultánea de posición y escala provoca que el estimador robusto de μ no alcance el punto de ruptura óptimo de $1/2$. Ver la sección 3.2.4 de Maronna *et al.* (2006).

A continuación, describimos el esquema del desarrollo de nuestro trabajo. En el Capítulo 2, introduciremos conceptos y resultados de teoría asintótica existente en la literatura, que serán necesarios para probar los resultados asintóticos de los M –estimadores de escala. En el Capítulo 3, definiremos formalmente los M –estimadores de escala basados en diferencias sucesivas cuando la función de escala es constante, es decir, el modelo es homoscedástico.

Enunciaremos y probaremos la consistencia y la normalidad asintótica de la sucesión de dichos estimadores. Mostraremos además que, asintóticamente, alcanzan el punto de ruptura óptimo de $1/2$ utilizando para ello la teoría del sesgo máximo. Realizaremos también un estudio de simulación para muestras finitas comparando los M -estimadores tanto con un estimador no robusto como con otros dos estimadores robustos. En el Capítulo 4, introduciremos los M -estimadores locales basados en diferencias sucesivas para estimar la función de escala en el caso de modelos heteroscedásticos. Probaremos consistencia y normalidad asintótica de los estimadores de esta familia. Estudiaremos el comportamiento asintótico de los M -estimadores locales basados en núcleos para sucesiones de aberturas de ventana adaptivas. Finalmente, llevaremos a cabo un estudio de simulación para muestras finitas. En el Capítulo 5 realizaremos una breve síntesis de las conclusiones arribadas. En el Apéndice A se demuestran resultados auxiliares. En el Apéndice B se muestran los cuadros y gráficos de los resultados correspondientes a las simulaciones Monte Carlo.

Capítulo 2

Elementos Generales de Teoría Asintótica

En este Capítulo introduciremos algunos elementos de teoría de probabilidad sobre espacios métricos que se utilizarán en las demostraciones sucesivas a lo largo de la tesis. La mayoría de las definiciones y demostraciones pueden ser halladas en los textos de Huber (1981), van der Vaart (1998) o Billingsley (1968, 1979) y no mencionaremos específicamente la fuente, salvo que nos interese destacar alguna particularidad. Para aquellos resultados que no se encuentren en los textos mencionados citaremos la bibliografía.

Definición 2.1 Sea \mathcal{P} la colección de las medidas en probabilidad sobre (S, \mathcal{S}) , donde S es un espacio métrico con métrica d y \mathcal{S} es la σ -álgebra de Borel asociada (o sea, aquella generada por los conjuntos abiertos del espacio). Una sucesión $\{P_n\}_{n \geq 1}$ de elementos de \mathcal{P} se dice que converge débilmente a $P \in \mathcal{S}$ y escribimos $P_n \Rightarrow P$ si

$$\lim_{n \rightarrow \infty} \int f dP_n = \int f dP$$

para toda función continua y acotada $f : S \rightarrow \mathbb{R}$.

A \mathcal{P} se la puede dotar de una topología, denominada *topología débil*, tomando como base de entornos de un punto $P \in \mathcal{P}$ a la colección

$$\left\{ Q \in \mathcal{P} : \left| \int_S f_i dQ - \int_S f_i dP \right| < \epsilon \quad 1 \leq i \leq k \right\}$$

donde f_i , $i = 1, \dots, k$ son funciones reales, continuas y acotadas sobre S y $\epsilon \in \mathbb{R}$ es positivo.

De este modo *la convergencia débil es la convergencia en esta topología*.

La topología débil, bajo ciertas condiciones sobre el espacio S es metrizable. La siguiente proposición da una de las posibles métricas o distancias.

Proposición 2.1 Sea $\pi : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}$ la función definida como

$$\pi(P, Q) = \inf \{ \epsilon > 0 : P(A) \leq Q(A^\epsilon) + \epsilon \quad \forall A \in \mathcal{S} \}$$

donde $A^\epsilon = \left\{ x \in S : \inf_{a \in A} d(x, a) \leq \epsilon \right\}$.

Entonces π es una métrica sobre \mathcal{P} y si, además, S es separable y completo entonces (\mathcal{P}, π) resulta también un espacio métrico, separable y completo y π metriza la topología débil. A π se le denomina *la métrica de Prohorov*.

La métrica de Prohorov no es la única que metriza la topología débil. Otra, por ejemplo, es la métrica acotada de Lipschitz.

Consideremos a \mathbb{R} con la métrica usual, $\mathcal{B}(\mathbb{R})$ la σ -álgebra de Borel asociada y \mathcal{D} , como en la Definición 3.1, la colección de funciones de distribución sobre \mathbb{R} . Utilizaremos la convención usual de denotar con la misma letra F tanto una medida en probabilidad como su función de distribución asociada. Definamos las siguientes dos colecciones de funciones:

$$\mathcal{C}_U(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} \text{ es una función uniformemente continua y acotada}\}$$

$$\mathcal{C}_L(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} \text{ es una función continua Lipschitz y acotada}\}$$

y, para cada $f \in \mathcal{C}_L(\mathbb{R})$, sea

$$\|f\|_L = \min \{k : |f(x) - f(y)| \leq k|x - y|, \forall x, y \in \mathbb{R}\}. \quad (2.1)$$

Enunciamos a continuación la siguiente versión de un lema bien conocido.

Proposición 2.2 *Lema de Pormanteau.*

Sean $\{F_n\}_{n \geq 1}$ y F medidas en probabilidad (o funciones de distribución) sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Son equivalentes:

(i) $F_n \Rightarrow F$;

(ii) $\lim_{n \rightarrow \infty} \int f dF_n = \int f dF$ para cada $f \in \mathcal{C}_U(\mathbb{R})$;

(iii) $\lim_{n \rightarrow \infty} \int f dF_n = \int f dF$ para cada $f \in \mathcal{C}_L(\mathbb{R})$;

(iv) $\liminf_{n \rightarrow \infty} F_n(G) \geq F(G)$ para cada conjunto abierto G ;

(v) $\limsup_{n \rightarrow \infty} F_n(T) \leq F(T)$ para cada conjunto cerrado T ;

(vi) $\lim_{n \rightarrow \infty} F_n(H) = F(H)$ para cada $H \in \mathcal{B}(\mathbb{R})$ tal que $F(\partial H) = 0$ donde ∂H es la frontera de H .

Observemos que, si consideramos $\{X_n\}_{n \geq 1}$ y X v.a.'s y $\{F_n\}_{n \geq 1}$ y F sus correspondientes funciones de distribución, el item (vi) del Lema de Pormanteau implica el concepto usual de convergencia en distribución, que denotaremos con $X_n \xrightarrow{d} X$; a saber, que $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, para cada punto de continuidad, x , de F .

Los resultados asintóticos para M -estimadores dependen de una propiedad clave del M -funcional de escala.

Proposición 2.3 *Asumiendo que la Hipótesis 1 es válida, el M -funcional de escala definido en (3.12), es continuo respecto de la topología débil.*

Dado que los M -estimadores están contruidos a partir de diferencias sucesivas de observaciones, los teoremas clásicos de convergencia basados en sumas de variables independientes no pueden ser aplicados (ya que las diferencias no serán independientes). Presentamos a continuación las definiciones y resultados que nos interesan sobre convergencia de sucesiones de variables dependientes.

Definición 2.2 *Dada una sucesión de v.a.'s $\{\xi_i\}_{i \geq 1}$, sea $\mathcal{F}_n^m = \sigma(\xi_i, n \leq i \leq m)$ la σ -álgebra generada por $\{\xi_i, n \leq i \leq m\}$. Se dice que la sucesión es m -dependiente si las σ -álgebras $\mathcal{F}_{j_1}^{k_1}, \dots, \mathcal{F}_{j_\ell}^{k_\ell}$ son independientes para $k_{i-1} + m < j_i$ para $i = 2, \dots, \ell$.*

Definición 2.3 Si \mathcal{A} y \mathcal{B} son dos σ -álgebras de eventos, definamos

$$\varphi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A) \neq 0} |P(B|A) - P(B)|.$$

Una sucesión de v.a.'s, $\{\xi_i\}_{i \geq 1}$, se dice φ -mezclante si $\lim_{n \rightarrow \infty} \varphi(n) = 0$ donde $\varphi(n) = \sup_{k \geq 1} \varphi(\mathcal{F}_1^k, \mathcal{F}_{k+n}^\infty)$, con \mathcal{F}_n^m , $1 \leq n, m \leq \infty$, como en la Definición 2.2.

Observación 2.1 Es claro que si una sucesión de v.a.'s $\{\xi_i\}_{i \geq 1}$ es m -dependiente entonces resulta φ -mezclante.

La siguiente proposición se encuentra enunciada, aunque no demostrada, en Billingsley (1979, págs. 75 y 257). En el Apéndice A damos la demostración.

Proposición 2.4 Ley Fuerte de los Grandes Números para v.a.'s m -dependientes.

Sea $\{\xi_i\}_{i \geq 1}$ una sucesión de v.a.'s m -dependientes, centradas y uniformemente acotadas. Entonces, $\sum_{i=1}^n \xi_i/n \xrightarrow{c.s.} 0$, donde $\xrightarrow{c.s.}$ denota la convergencia casi segura.

La siguiente proposición puede ser hallada en Pelligrad, M. y Utev, S. (1997, Teorema 2.2).

Proposición 2.5 Distribución asintótica de procesos φ -mezclantes.

Sea $\{a_{n,i}\}$ un arreglo triangular de números reales tales que

$$\sup_{n \geq 1} \sum_{i=1}^n a_{n,i}^2 < \infty \quad y \quad \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} |a_{n,i}| = 0, \quad (2.2)$$

y sea $\{\xi_i\}_{i \geq 1}$ una sucesión de v.a.'s centradas que verifique las siguientes condiciones:

$$\{\xi_i^2\}_{i \geq 1} \quad \text{es una familia uniformemente integrable y} \quad \text{Var} \left[\sum_{i=1}^n a_{n,i} \xi_i \right] = 1. \quad (2.3)$$

Si además $\{\xi_i\}_{i \geq 1}$ es una sucesión φ -mezclante, entonces

$$\sum_{i=1}^n a_{n,i} \xi_i \xrightarrow{d} N(0, 1). \quad (2.4)$$

Necesitamos introducir el concepto de ajuste que será utilizado para obtener resultados adicionales de teoría asintótica.

Definición 2.4 Sean \mathcal{P} y (S, \mathcal{S}) como en la Definición 2.1. Una familia, Π , de medidas en probabilidad se dice *ajustada* si para cada $\epsilon > 0$ existe un conjunto compacto, K , en S tal que $P(K) > 1 - \epsilon$, para toda $P \in \Pi$.

Consideremos $C = C[0, 1]$ el espacio de las funciones continuas sobre el $[0, 1]$, al cual se lo dota con la topología uniforme definiendo la distancia, d , entre dos puntos $x, y \in C$ como $d(x, y) = \sup_{t \in [0, 1]} |x(t) - y(t)|$. Sea \mathcal{C} la σ -álgebra de Borel asociada.

Sea $X : (\Omega, \mathcal{A}, P) \rightarrow C$ una aplicación, con lo cual, para cada $\omega \in \Omega$, $X(\omega) \in C$ y cuyo valor en $t \in [0, 1]$, se denotará con $X(t, \omega)$. Así, para un t fijo, $X(t, \cdot)$ denotará una variable aleatoria definida sobre Ω con valor $X(t, \omega)$, para cada $\omega \in \Omega$. Asumamos que X es un función medible a la que llamaremos elemento aleatorio de C . Se dice que una sucesión $\{X_n\}_{n \geq 1}$ de elementos aleatorios de C es ajustada si las medidas en probabilidad inducidas, $\{P_{X_n}\}_{n \geq 1}$, son

una sucesión ajustada.

Las proposiciones que siguen dan condiciones suficientes para determinar ajuste en el espacio C .

Proposición 2.6 *Una sucesión $\{X_n\}_{n \geq 1}$ en C es ajustada si y sólo si satisface las siguientes dos condiciones:*

(i) *La sucesión $\{X_n(0)\}_{n \geq 1}$ es ajustada.*

(ii) $\forall \epsilon, \eta > 0 \exists 0 < \delta < 1$ y $n_0 \in \mathbb{N}$:

$$P \left(\sup_{|t-s| < \delta} |X_n(t) - X_n(s)| \geq \epsilon \right) \leq \eta, \quad \forall n \geq n_0.$$

Proposición 2.7 *Una sucesión $\{X_n\}_{n \geq 1}$ en C es ajustada si satisface las siguientes dos condiciones:*

(i) *La sucesión $\{X_n(0)\}_{n \geq 1}$ es ajustada.*

(ii) *Existen constantes $\gamma \geq 0$, $\alpha > 1$ y una función, F , continua y no-decreciente sobre $[0, 1]$*

tal que $\forall t_1, t_2 \in [0, 1], n \in \mathbb{N}$ y $\lambda \in \mathbb{R}^+$:

$$P (|X_n(t_2) - X_n(t_1)| \geq \lambda) \leq \frac{1}{\lambda^\gamma} |F(t_2) - F(t_1)|^\alpha. \quad (2.5)$$

La siguiente condición de momentos implica (2.5)

$$E[|X_n(t_2) - X_n(t_1)|^\gamma] \leq |F(t_2) - F(t_1)|^\alpha. \quad (2.6)$$

Observación 2.2 *El intervalo $[0, 1]$ se puede reemplazar por cualquier otro intervalo $[r, s]$ con $r < s$.*

Finalmente introduzcamos el problema, de nuestro interés, al cual está vinculado la noción de ajuste. La convergencia débil en C no se deriva, en general, de la convergencia de las distribuciones finito-dimensionales. Para ser más precisos, dados $t_1, \dots, t_k \in [0, 1]$, consideremos la función $\pi_{t_1, \dots, t_k} : C \rightarrow \mathbb{R}^k$ definida como $\pi_{t_1, \dots, t_k}(x) = (x(t_1), \dots, x(t_k))$. La continuidad de π_{t_1, \dots, t_k} implica que si $P_n \Rightarrow P$ en (C, \mathcal{C}) , entonces las distribuciones finito-dimensionales $P_n \pi_{t_1, \dots, t_k}^{-1}$ convergen débilmente a $P \pi_{t_1, \dots, t_k}^{-1}$, cualesquiera sean t_1, \dots, t_k . La recíproca no es válida en el espacio C . La siguiente proposición da una condición suficiente para la convergencia débil en C .

Proposición 2.8 *Sean P_n y P medidas en probabilidad sobre (C, \mathcal{C}) . Si las distribuciones finito-dimensionales de P_n convergen débilmente a aquellas de P , y si la sucesión $\{P_n\}_{n \geq 1}$ es ajustada entonces $P_n \Rightarrow P$.*

Capítulo 3

Estimación de la escala global

3.1. Preliminares

Antes de presentar el problema específico objeto de este Capítulo, necesitamos introducir el concepto de entorno de contaminación (ver Huber, 1981, pág. 11).

Definición 3.1 *Sea \mathcal{D} la colección de las funciones de distribución definidas sobre la recta real. Dada $F_0 \in \mathcal{D}$ fija, el entorno de contaminación de F_0 , de tamaño $\epsilon \in (0, 1)$, se define como*

$$\mathcal{P}_\epsilon(F_0) = \{G : G(y) = (1 - \epsilon)F_0(y) + \epsilon H(y); H \in \mathcal{D}, y \in \mathbb{R}\}. \quad (3.1)$$

Observación 3.1 *Es importante mencionar que, si bien en la literatura $\mathcal{P}_\epsilon(F_0)$ es denominado entorno de contaminación, no constituye un entorno en un sentido topológico. Para dar entornos en el sentido topológico, se introducirá en la Subsección 2 la topología de la convergencia débil.*

Precisemos ahora el modelo sobre el cual vamos a trabajar en el presente capítulo.

Definición 3.2 Modelo de Regresión No-paramétrico con Escala Global

Consideremos el modelo

$$Y_i = g(x_i) + \sigma U_i, \quad i = 1, \dots, n. \quad (3.2)$$

donde $\{Y_i\}_{i=1}^n$ son las respuestas observadas, $\{x_i\}_{i=1}^n$ son los puntos del diseño (valores reales, no aleatorios, en el intervalo $[0, 1]$) tales que $0 \leq x_1 \leq \dots \leq x_n \leq 1$, $g : [0, 1] \rightarrow \mathbb{R}$ es la función de regresión, desconocida, $\{U_i\}_{i=1}^n$ representan los errores y son v.a.'s i.i.d. con distribución común $G \in \mathcal{P}_\epsilon(F_0)$, $\epsilon \in [0, 1/2)$ fijo y $F_0 \in \mathcal{D}$ es el modelo central. La escala $\sigma \in \mathbb{R}^+$ es desconocida.

Los métodos clásicos (no robustos) de estimación suponen que la totalidad de los errores provienen del modelo central F_0 . En el contexto de la estimación robusta se supone, en cambio, que la mayoría de las observaciones Y_i del modelo es de buena calidad (los errores se distribuyen según F_0) excepto una fracción, ϵ , que es posiblemente de mala calidad (los errores provienen de una distribución arbitraria). Esta noción intuitiva se formaliza con la suposición de que G pertenece a algún entorno de contaminación $\mathcal{P}_\epsilon(F_0)$.

Como se mencionó anteriormente, Rice (1984), propuso utilizar como estimador de σ^2 el definido como el promedio de los cuadrados de las diferencias sucesivas de las respuestas. Dado que este estimador pesa las diferencias utilizando una función no acotada (la función

cuadrática) resulta no robusto (ver Sección 3.3). Boente *et al.* (1997), en vistas de la selección robusta de la abertura de ventana, proponen reemplazar el promedio de los cuadrados por la mediana de los valores absolutos de las diferencias sucesivas obteniendo así una propuesta robusta de estimación.

Ambos estimadores pueden ser generalizados por medio de la siguiente definición. Antes de enunciarla introduzcamos la siguiente notación que será útil a lo largo de toda la tesis: dadas Z_1, \dots, Z_n , n v.a.'s i.i.d., definimos las diferencias sucesivas como

$$Z_i^* = Z_{i+1} - Z_i, \quad i = 1, \dots, n-1. \quad (3.3)$$

Definición 3.3 *Para el Modelo introducido en (3.2), definimos la familia de M -estimadores de escala basados en las diferencias sucesivas de las respuestas $\{Y_i^*\}_{i=1}^{n-1}$ con función de escores χ y constantes de ajuste a y b como*

$$\hat{\sigma}_{M,n} = \inf \left\{ s > 0 : \frac{1}{n-1} \sum_{i=1}^{n-1} \chi \left(\frac{Y_i^*}{as} \right) \leq b \right\}. \quad (3.4)$$

La función $\chi : \mathbb{R} \rightarrow \mathbb{R}_0^+$ es par, $\chi(0) = 0$, no decreciente en \mathbb{R}^+ , con $0 < \|\chi\|_\infty$ (siendo $\|\cdot\|_\infty$ la norma infinito). Las constantes $a \in (0, \infty)$ y $b \in (0, 1)$ satisfacen

$$E[\chi(Z_1)] = b \quad y \quad E \left[\chi \left(\frac{Z_1^*}{a} \right) \right] = b, \quad (3.5)$$

donde $Z_1^* = Z_2 - Z_1$, $\{Z_i\}_{i=1,2}$ v.a.'s i.i.d. con $Z_1 \sim F_0$.

Para abreviar, a $\hat{\sigma}_{M,n}$ le llamaremos simplemente M -estimador.

Hay que tener en cuenta que el ínfimo en (3.4) es necesario cuando la función de escores es discontinua. Sin embargo, para desarrollar la teoría asintótica y las propiedades de robustez asintótica en esta tesis vamos a hacer el siguiente supuesto.

Hipótesis 1 *La función de escores χ es continua, acotada, par, estrictamente creciente sobre el conjunto $\{x : \chi(x) < \|\chi\|_\infty\}$ con $\chi(0) = 0$. Sin pérdida de generalidad, supondremos que $\|\chi\|_\infty = 1$.*

Observación 3.2 *Bajo la Hipótesis 1, $\hat{\sigma}_{M,n}$ es la única solución positiva de la ecuación*

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \chi\left(\frac{Y_i^*}{as}\right) = b, \quad (3.6)$$

siempre que $q(\mathbf{D}) > b$, donde $\mathbf{D} = (Y_1^*, \dots, Y_{n-1}^*)$ y

$$q(\mathbf{D}) = \frac{1}{n-1} \#\{i : Y_i^* \neq 0; i = 1, \dots, n-1\}. \quad (3.7)$$

Observemos que $\hat{\sigma}_{M,n}$ es un miembro genérico de una familia, cuyos elementos particulares corresponden a diferentes elecciones de la función de escores χ y las constantes de ajuste a y b . Como se verá en la Sección 3.3 la elección de χ no es crucial para asegurar que $\hat{\sigma}_{M,n}$ alcance las propiedades de robustez asintóticas, aunque sí lo es la elección de la constante b .

Veremos también que, dada b , la constante a es elegida de tal modo que, cuando no hay contaminación en los datos (o sea, en el modelo central), el funcional (definido más abajo) asociado a la sucesión de M -estimadores $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$ resulte consistente Fisher.

A continuación, damos ejemplos de M -estimadores de escala para distintas elecciones de χ , b y a cuando el modelo central es $F_0 = \Phi$. Cada uno de los estimadores se halla eligiendo primero la función de escores, luego se determinan las constantes a y b que satisfacen (3.5) y el estimador resulta de la solución de la ecuación (3.6). En los ejemplos $I_A(\cdot)$ es la función indicadora sobre el conjunto A y, dado $t \in (0, 1)$, q_t denota el cuantil t de las diferencias absolutas $|Y_i^*|$, $i = 1, \dots, n-1$.

Ejemplo 3.1 Para $\chi(x) = x^2$, $b = 1$ y $a = \sqrt{2}$ la solución de (3.6) es el estimador de escala propuesto por Rice (1984)

$$\hat{\sigma}_{R,n} = \left(\frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_i^*)^2 \right)^{1/2}, \quad (3.8)$$

al cual le denominaremos “estimador de Rice”.

Ejemplo 3.2 Si $\chi(x) = I_{\{y: |y| > \Phi^{-1}(3/4)\}}(x)$, $b = 1/2$ y $a = \sqrt{2}$ entonces la solución de (3.6) es el estimador de escala propuesto por Boente et al. (1997)

$$\hat{\sigma}_{BFM,n} = \frac{q_{1/2}}{\sqrt{2}\Phi^{-1}(3/4)} \quad (3.9)$$

al que le llamaremos “estimador de Boente, Fraiman y Meloche”.

Ejemplo 3.3 Si $\chi(x) = I_{\{y: |y| > \Phi^{-1}(5/8)\}}(x)$, $b = 3/4$ y $a = \sqrt{2}$ entonces (3.6) nos provee el estimador de escala

$$\hat{\sigma}_{\text{PC},n} = \frac{q_{1/4}}{\sqrt{2}\Phi^{-1}(5/8)}, \quad (3.10)$$

que es una modificación del estimador del ejemplo anterior. Nos referiremos a él como “estimador primer cuartil”.

Ejemplo 3.4 Sea $c \in \mathbb{R}^+$ y sea

$$\chi_c(x) = \begin{cases} 3(x/c)^2 - 3(x/c)^4 + (x/c)^6 & \text{si } |x| \leq c \\ 1 & \text{si } |x| > c \end{cases} \quad (3.11)$$

la familia de función de escores introducida por Beaton y Tukey (1974). Eligiendo $\chi(x) = \chi_c(x)$, con $c = 0.70417$, $b = 3/4$ y $a = \sqrt{2}$ en (3.6), se obtiene el estimador

$$\hat{\sigma}_{\text{BT},n} = \inf \left\{ s > 0 : \frac{1}{n-1} \sum_{i=1}^{n-1} \chi_c \left(\frac{Y_i^*}{\sqrt{2}s} \right) \leq \frac{3}{4} \right\}.$$

al que llamaremos “ M -estimador con función de BT”.

Es inmediato comprobar que la sucesión de M -estimadores de escala son equivariantes por cambios de escala, es decir, satisfacen la siguiente propiedad.

Definición 3.4 Una sucesión de estimadores $\{T_n\}_{n \geq 1}$ se dice equivariante por transformaciones de escala si, para cada $\theta \in \mathbb{R}$ con $\theta \in \mathbb{R}^+$, para cada n -upla $(x_1, \dots, x_n) \in \mathbb{R}^n$ y para

cada $n \in \mathbb{N}$, se verifica que

$$T_n(\theta x_1, \dots, \theta x_n) = \theta T_n(x_1, \dots, x_n).$$

Como se probará en la próxima sección, bajo ciertas condiciones de regularidad, la sucesión de M -estimadores $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$ converge casi seguramente a cierto valor asintótico, el que puede ser definido como un funcional que opera sobre funciones de distribución.

Definición 3.5 *Dada una variable aleatoria X con función de distribución F , χ , a y b como en la Definición 3.3, al funcional $S : \mathcal{D} \rightarrow \mathbb{R}_0^+$ definido como*

$$S(F) = \inf \left\{ s > 0 : E \left[\chi \left(\frac{X}{as} \right) \right] \leq b \right\} \quad (3.12)$$

se le denomina M -funcional de escala (ver Huber, 1964).

Definimos entonces el M -funcional de escala correspondiente a $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$ como el valor $S(G_\sigma^*)$ donde G_σ^* es la función de distribución de la v.a. $\sigma U_1^* = \sigma(U_2 - U_1)$ donde $\{U_i\}_{i=1,2}$ son v.a.'s. i.i.d. con $U_1 \sim G$. Es decir,

$$S(G_\sigma^*) = \inf \left\{ s > 0 : E \left[\chi \left(\frac{\sigma U_1^*}{as} \right) \right] \leq b \right\}. \quad (3.13)$$

Para abreviar a $S(G_\sigma^*)$ le llamaremos simplemente M -funcional de escala.

Observación 3.3

a) Notemos que el M -funcional de escala correspondiente a $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$ no es otra cosa que el M -funcional de escala definido en (3.12) aplicado a un producto convolución; más específicamente a $(G \star G_-)(\cdot/\sigma)$ donde \star denota dicho producto y G_- es la función de distribución de $-U$ si $U \sim G$.

b) Si $F_n(y) = \frac{1}{n-1} \sum_{i=1}^{n-1} I_{(-\infty, y]}(Y_i^*)$ es la función de distribución empírica de la muestra de las diferencias Y_1^*, \dots, Y_{n-1}^* entonces por (3.4) y (3.12) obtenemos

$$S(F_n) = \widehat{\sigma}_{M,n}. \quad (3.14)$$

c) Análogamente a la Observación 3.2, bajo la Hipótesis 1, $S(G_\sigma^*)$ es la única solución positiva de la ecuación

$$E \left[\chi \left(\frac{\sigma U_1^*}{as} \right) \right] = b. \quad (3.15)$$

De este hecho junto con la condición (3.5) sobre las constantes de ajuste se deduce fácilmente que el funcional es consistente Fisher, es decir,

$$\forall \sigma > 0 : S(F_{0,\sigma}^*) = \sigma, \quad (3.16)$$

donde $F_{0,\sigma}^*$ es la distribución de σU_1^* cuando $U_1^* \sim F_0$. La ecuación (3.16) garantiza que, en el modelo central, asintóticamente $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$ estima la cantidad correcta.

Definamos ahora la noción de equivariancia por cambios de escala de un funcional.

Definición 3.6 Sea $T : \mathcal{D} \rightarrow \mathbb{R}$ un funcional. Se dice que T es equivariante por transformaciones de escala si para cada $F \in \mathcal{D}$ y para cada $\theta \in \mathbb{R}^+$, se verifica que

$$T(F_\theta) = \theta T(F),$$

donde F_θ denota la función de distribución de θZ siendo Z una v.a. tal que $Z \sim F$.

Es inmediato comprobar entonces que

$$\forall \sigma > 0 : S(G_\sigma^*) = \sigma S(G^*). \quad (3.17)$$

3.2. Teoría Asintótica de los M -estimadores de la Escala Global

3.2.1. Consistencia y Normalidad asintótica

En la presente subsección consideraremos $\{Y_i\}_{i=1}^n$ una sucesión de v.a.'s independientes satisfaciendo el modelo introducido en (3.2), $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$ la sucesión de M -estimadores de escala definida en (3.4), $S(\cdot)$ y G_σ^* como en la Definición 3.5. Las v.a.'s Y_i^* y U_i^* , $i = 1, \dots, n-1$ son como en (3.3), es decir, las diferencias $Y_{i+1} - Y_i$ y $U_{i+1} - U_i$ respectivamente; U_i , $i = 1, \dots, n$ denota la muestra aleatoria de los errores.

En las demostraciones a lo largo de toda la tesis, para simplificar, haremos la siguiente

convención: si $\{Z_n\}_{n \geq 1}$ y Z son variables aleatorias definidas sobre algún espacio de probabilidad (Ω, \mathcal{A}, P) y si $Z_n(\omega) \rightarrow Z(\omega)$ para todo $\omega \in \Omega$, esta convergencia en todo punto la denotaremos igual que la convergencia casi segura, es decir $Z_n \xrightarrow{c.s.} Z$.

Introduzcamos ahora las siguientes suposiciones adicionales:

Hipótesis 2 *Los puntos del diseño $\{x_i\}_{i=1}^n$ satisfacen $\lim_{n \rightarrow \infty} M_n = 0$, donde M_n es la sucesión*

$$M_n = \max_{1 \leq i \leq n-1} (x_{i+1} - x_i).$$

Hipótesis 3 *La función de regresión $g : [0, 1] \rightarrow \mathbb{R}$ es continua.*

Teorema 3.1 Consistencia Fuerte de los M -estimadores

Bajo las Hipótesis 1 a 3, la sucesión de M -estimadores es fuertemente consistente a $S(G_\sigma^)$; es decir,*

$$\hat{\sigma}_{M,n} \xrightarrow{c.s.} S(G_\sigma^*). \quad (3.18)$$

Demostración. Consideremos las siguientes funciones de distribución empíricas

$$F_n(y) = \frac{1}{n-1} \sum_{i=1}^{n-1} I_{(-\infty, y]}(Y_i^*) \text{ y } \tilde{F}_n(y) = \frac{1}{n-1} \sum_{i=1}^{n-1} I_{(-\infty, y]}(\sigma U_i^*).$$

Dado que el M -funcional $S(\cdot)$ es continuo respecto de la topología débil (Proposición 2.3), que la métrica de Prohorov, π , metriza la topología débil (Proposición 2.1) y que $S(F_n) = \hat{\sigma}_{M,n}$

(ver (3.14)), para probar (3.18) bastará con demostrar que

$$\pi(F_n, G_\sigma^*) \xrightarrow{c.s.} 0. \quad (3.19)$$

Para obtener (3.19) será suficiente mostrar que las dos convergencias casi seguras siguientes son válidas

$$\pi(F_n, \tilde{F}_n) \xrightarrow{c.s.} 0 \quad (3.20)$$

$$\pi(\tilde{F}_n, G_\sigma^*) \xrightarrow{c.s.} 0. \quad (3.21)$$

Por el Lema de Pormanteau (Proposición 2.2) para derivar (3.20) será equivalente demostrar que, para cada función $f \in \mathcal{C}_L(\mathbb{R})$, se verifica

$$\int f dF_n - \int f d\tilde{F}_n \xrightarrow{c.s.} 0. \quad (3.22)$$

Sean entonces $f \in \mathcal{C}_L(\mathbb{R})$ y $\epsilon > 0$. Como g es continua sobre el compacto $[0, 1]$ entonces es uniformemente continua y por ende, existe $\delta > 0$ tal que si $|x - y| < \delta$ entonces $|g(x) - g(y)| < \epsilon / \|f\|_L$. Además, por la Hipótesis 2, $M_n \rightarrow 0$ luego, existe n_0 tal que

$$\forall n \geq n_0, \forall i = 1, \dots, n-1 : |x_{i+1} - x_i| \leq M_n < \delta .$$

Con lo cual,

$$\begin{aligned}
 \left| \int f dF_n - \int f d\tilde{F}_n \right| &= \left| \frac{1}{n-1} \sum_{i=1}^{n-1} f(Y_i^*) - \frac{1}{n-1} \sum_{i=1}^{n-1} f(\sigma U_i^*) \right| \\
 &\leq \|f\|_L \frac{1}{n-1} \sum_{i=1}^{n-1} |Y_i^* - \sigma U_i^*| \\
 &= \|f\|_L \frac{1}{n-1} \sum_{i=1}^{n-1} |g(x_{i+1}) - g(x_i)| \\
 &< \|f\|_L \epsilon / (\|f\|_L) < \epsilon,
 \end{aligned}$$

para todo $n \geq n_0$, con lo que (3.20) queda demostrado.

Derivemos ahora (3.21) de manera análoga a como obtuvimos (3.20). Demostraremos que para cada $f \in \mathcal{C}_L(\mathbb{R})$ se cumple que

$$\int f d\tilde{F}_n \xrightarrow{c.s.} \int f dG_\sigma^*. \tag{3.23}$$

Por la definición de G_σ^* y como las U_i^* son v.a.'s idénticamente distribuidas, tenemos

$$\int f dG_\sigma^* = E[f(\sigma U_1^*)] = \frac{1}{n-1} \sum_{i=1}^{n-1} E[f(\sigma U_i^*)].$$

Teniendo en cuenta esta expresión y definiendo la sucesión de variables aleatorias $\xi_i =$

$f(\sigma U_i^*) - E[f(\sigma U_i^*)]$, podemos escribir

$$\begin{aligned} \int f d\tilde{F}_n - \int f dG_\sigma^* &= \frac{1}{n-1} \sum_{i=1}^{n-1} f(\sigma U_i^*) - E[f(\sigma U_1^*)] \\ &= \frac{1}{n-1} \sum_{i=1}^{n-1} \{f(\sigma U_i^*) - E[f(\sigma U_i^*)]\} \\ &= \frac{1}{n-1} \sum_{i=1}^{n-1} \xi_i. \end{aligned}$$

Notemos que $\{\xi_i\}_{i \geq 1}$ es una sucesión de v.a.'s uno-dependientes, con $E[\xi_i] = 0$ y uniformemente acotada por $2\|f\|_L$. En consecuencia, por la Ley Fuerte de los Grandes Números (Proposición 2.4), podemos concluir que $\sum_{i=1}^{n-1} \xi_i / (n-1) \xrightarrow{c.s.} 0$, lo que implica (3.23) y por ende (3.21). \square

Para demostrar la normalidad asintótica necesitamos introducir nuevas hipótesis.

Hipótesis 4 *La función de regresión $g : [0, 1] \rightarrow \mathbb{R}$ es continua Lipschitz con $\|g\|_L$ como en (2.1).*

Hipótesis 5 *La función de escores $\chi : \mathbb{R} \rightarrow \mathbb{R}$ es continua Lipschitz con $\|\chi\|_L$ como en (2.1).*

Hipótesis 6 *Los puntos del diseño $\{x_i\}_{i=1}^n$ satisfacen $M_n = \max_{1 \leq i \leq n-1} |x_{i+1} - x_i| = O(n^{-1})$.*

Recpr

Hipótesis 7 *La función de escores χ satisface las siguientes propiedades*

(i) χ tiene segunda derivada continua.

(ii) Las funciones $\chi_1, \chi_2 : \mathbb{R} \rightarrow \mathbb{R}$ definidas como $\chi_1(u) = u\chi'(u)$, $\chi_2(u) = u^2\chi''(u)$ están acotadas.

Teorema 3.2 Normalidad Asintótica de los M -estimadores

Sea $v = v(G_\sigma^*) = v_1/v_2^2$ con $v_i > 0$, $i = 1, 2$, definidos por

$$v_1 = v_1(G_\sigma^*) = \text{Var} \left[\chi \left(\frac{\sigma U_1^*}{aS(G_\sigma^*)} \right) \right] + 2 \text{Cov} \left[\chi \left(\frac{\sigma U_1^*}{aS(G_\sigma^*)} \right), \chi \left(\frac{\sigma U_2^*}{aS(G_\sigma^*)} \right) \right] \quad (3.24)$$

$$v_2 = v_2(G_\sigma^*) = E \left[\chi' \left(\frac{\sigma U_1^*}{aS(G_\sigma^*)} \right) \left(\frac{\sigma U_1^*}{a(S(G_\sigma^*))^2} \right) \right]. \quad (3.25)$$

Si se cumplen las Hipótesis 1, 4, 6 y 7, entonces

$$n^{1/2} (\hat{\sigma}_{M,n} - S(G_\sigma^*)) \xrightarrow{d} N(0, v). \quad (3.26)$$

Para la demostración del Teorema necesitamos el siguiente Lema cuya prueba se halla en el Apéndice A. Vale la pena mencionar que la Hipótesis 5, necesaria para el Lema, se encuentra ausente en el enunciado del Teorema de Normalidad Asintótica. Sin embargo, hay que tener en cuenta que es implicada por la Hipótesis 7.

Lema 3.1 Para cada $s > 0$, definamos las siguientes funciones

$$\begin{aligned}\lambda_{n,b}(s) &= \frac{1}{n-1} \sum_{i=1}^{n-1} \chi\left(\frac{Y_i^*}{as}\right) - b \\ \lambda_{n,b}^*(s) &= \frac{1}{n-1} \sum_{i=1}^{n-1} \chi\left(\frac{\sigma U_i^*}{as}\right) - b.\end{aligned}$$

Bajo las Hipótesis 4, 5 y 6

$$(n-1)^{1/2} \lambda_{n,b}(s) = (n-1)^{1/2} \lambda_{n,b}^*(s) + o_p(1) \quad (3.27)$$

Demostración del Teorema 3.2. Demostraremos (3.26) para $\sqrt{n-1}$ en lugar de \sqrt{n} sólo por una cuestión de coherencia con la notación introducida, pero es claro que ambos resultados son equivalentes.

Por simplicidad denotemos $S_\sigma = S(G_\sigma^*)$ y consideremos $\lambda_{n,b}$ y $\lambda_{n,b}^*$ como en el Lema 3.1. Un desarrollo de Taylor de primer orden de la función $\lambda_{n,b}$ alrededor de S_σ (y evaluando en $\hat{\sigma}_{M,n}$) nos permite obtener la siguiente igualdad:

$$\lambda_{n,b}(\hat{\sigma}_{M,n}) = \lambda_{n,b}(S_\sigma) + (\hat{\sigma}_{M,n} - S_\sigma) \lambda'_{n,b}(\hat{\sigma}_n) \quad (3.28)$$

con $\hat{\sigma}_n$ un punto intermedio, es decir,

$$\hat{\sigma}_n = (1 - \theta_n) S_\sigma + \theta_n \hat{\sigma}_{M,n}, \text{ para algún } \theta_n \in [0, 1]. \quad (3.29)$$

Por la Observación 3.2, $\lambda_{n,b}(\hat{\sigma}_{M,n}) = 0$, con lo cual (3.28) implica

$$(n-1)^{1/2}(\hat{\sigma}_{M,n} - S_\sigma) = (n-1)^{1/2}\lambda_{n,b}(S_\sigma)/(-\lambda'_{n,b}(\hat{\sigma}_n)).$$

Luego, si mostramos que

$$(n-1)^{1/2}\lambda_{n,b}(S_\sigma) \xrightarrow{d} N(0, v_1) \quad (3.30)$$

$$-\lambda'_{n,b}(\hat{\sigma}_n) \xrightarrow{p} v_2, \quad (3.31)$$

el Lema de Slutsky implicará (3.26).

De acuerdo al Lema 3.1, $(n-1)^{1/2}\lambda_{n,b}(S_\sigma) = (n-1)^{1/2}\lambda_{n,b}^*(S_\sigma) + o_p(1)$, con lo cual, sólo tenemos que estudiar la distribución asintótica de $(n-1)^{1/2}\lambda_{n,b}^*(S_\sigma)$. Notemos que

$$\begin{aligned} (n-1)^{1/2}\lambda_{n,b}^*(S_\sigma) &= (n-1)^{1/2} \left(\frac{1}{n-1} \sum_{i=1}^{n-1} \chi \left(\frac{\sigma U_i^*}{aS_\sigma} \right) - b \right) \\ &= V_n^{1/2} \sum_{i=1}^{n-1} a_{n-1,i} \xi_i \end{aligned} \quad (3.32)$$

donde $a_{n-1,i} = V_n^{-1/2}(n-1)^{-1/2}$ con

$$V_n = \text{Var} \left[(n-1)^{-1/2} \sum_{i=1}^{n-1} \xi_i \right] \quad \text{y} \quad \xi_i = \chi \left(\frac{\sigma U_i^*}{aS_\sigma} \right) - b.$$

Tengamos en cuenta que

$$\begin{aligned}
 V_n &= Var \left[(n-1)^{-1/2} \sum_{i=1}^{n-1} \xi_i \right] \\
 &= (n-1)^{-1} \sum_{i=1}^{n-1} Var \left[\chi \left(\frac{\sigma U_i^*}{aS_\sigma} \right) \right] + 2(n-1)^{-1} \sum_{i=1}^{n-2} Cov \left[\chi \left(\frac{\sigma U_i^*}{aS_\sigma} \right), \chi \left(\frac{\sigma U_{i+1}^*}{aS_\sigma} \right) \right] \\
 &= Var \left[\chi \left(\frac{\sigma U_1^*}{aS_\sigma} \right) \right] + 2 \left(\frac{n-2}{n-1} \right) Cov \left[\chi \left(\frac{\sigma U_1^*}{aS_\sigma} \right), \chi \left(\frac{\sigma U_2^*}{aS_\sigma} \right) \right]
 \end{aligned}$$

y, en consecuencia,

$$\lim_{n \rightarrow \infty} V_n = Var \left[\chi \left(\frac{\sigma U_1^*}{aS_\sigma} \right) \right] + 2Cov \left[\chi \left(\frac{\sigma U_1^*}{aS_\sigma} \right), \chi \left(\frac{\sigma U_2^*}{aS_\sigma} \right) \right] = v_1. \quad (3.33)$$

Claramente se satisfacen las condiciones de la Proposición 2.5 (Teorema de Pelligrad y Utev). En efecto, por (3.33) podemos suponer, sin pérdida de generalidad, que $\inf_{n>1} V_n > 0$.

Luego,

$$\sup_{n \geq 2} \sum_{i=1}^{n-1} a_{n-1,i}^2 = \sup_{n > 1} V_n^{-1} = 1 / \inf_{n > 1} V_n < \infty.$$

También, (3.33) implica que

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n-1} |a_{n-1,i}| = 0$$

y, de este modo, la condición (2.2) de la Proposición 2.5 se cumple. Por estar χ es acotada, la sucesión $\{\xi_i^2\}_{i \geq 1}$ es uniformemente integrable y, además es inmediato comprobar que

$Var \left[\sum_{i=1}^{n-1} a_{n-1,i} \xi_i \right] = 1$, cualquiera sea n ; por lo tanto, la condición (2.3) de dicha proposición es válida también. Finalmente, por (3.15) (ver Observación 3.3),

$$E [\xi_i] = E \left[\chi \left(\frac{\sigma U_i^*}{aS_\sigma} \right) \right] - b = 0,$$

es decir, la sucesión está centrada y como $\{\xi_i\}_{i \geq 1}$ es uno-dependiente (por serlo $\{U_i^*\}_{i \geq 1}$) resulta φ -mezclante. En consecuencia,

$$\sum_{i=1}^n a_{n,i} \xi_i \xrightarrow{d} N(0, 1). \quad (3.34)$$

Luego, de (3.32), (3.33) y (3.34), obtenemos (3.30).

Sea ahora $\eta(t) = (t/a)\chi'(t/a)$. Entonces, (3.31) puede reescribirse como

$$\frac{1}{\widehat{\sigma}_n} \frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{Y_i^*}{\widehat{\sigma}_n} \right) \xrightarrow{p} \frac{1}{S_\sigma} E \left[\eta \left(\frac{U_1^* \sigma}{S_\sigma} \right) \right].$$

Por el Teorema 3.1 y considerando (3.29), $\widehat{\sigma}_n \xrightarrow{p} S_\sigma$, con lo cual es suficiente mostrar que

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{Y_i^*}{\widehat{\sigma}_n} \right) \xrightarrow{p} E \left[\eta \left(\frac{U_1^* \sigma}{S_\sigma} \right) \right]. \quad (3.35)$$

Definamos, para $s \in \mathbb{R}^+$, $h(s) = \sum_{i=1}^{n-1} \eta(Y_i^*/s)$, entonces un desarrollo de Taylor de primer

orden de h alrededor de S_σ evaluado en $\hat{\sigma}_n$, nos permite obtener la siguiente expresión

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{Y_i^*}{\hat{\sigma}_n} \right) = \frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{Y_i^*}{S_\sigma} \right) + \frac{1}{\tilde{\sigma}_n} \frac{1}{n-1} \sum_{i=1}^{n-1} \eta' \left(\frac{Y_i^*}{\tilde{\sigma}_n} \right) \left(\frac{Y_i^*}{\tilde{\sigma}_n} \right) (S_\sigma - \hat{\sigma}_n)$$

con $\tilde{\sigma}_n = (1 - \gamma_n)S_\sigma + \gamma_n\hat{\sigma}_n$, y γ_n algún valor en el intervalo $[0, 1]$. Como $\tilde{\sigma}_n \xrightarrow{p} S_\sigma$ (ver (3.29)

y el Teorema 3.1) entonces para probar (3.35) bastará mostrar que

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \eta' \left(\frac{Y_i^*}{\tilde{\sigma}_n} \right) \left(\frac{Y_i^*}{\tilde{\sigma}_n} \right) = O_p(1) \quad (3.36)$$

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{Y_i^*}{S_\sigma} \right) \xrightarrow{c.s.} E \left[\eta \left(\frac{U_1^* \sigma}{S_\sigma} \right) \right]. \quad (3.37)$$

Observemos que (3.36) se deduce de la Hipótesis 7 ii) ya que $\eta'(t)t$ es una función acotada.

Por otra parte, la Hipótesis 7 implica que $\eta \in \mathcal{C}_L(\mathbb{R})$ ya que η posee primera derivada continua y acotada y además $g \in \mathcal{C}_L(\mathbb{R})$ por hipótesis, de donde

$$\begin{aligned} \left| \frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{Y_i^*}{S_\sigma} \right) - \frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{\sigma U_i^*}{S_\sigma} \right) \right| &\leq \frac{1}{n-1} \sum_{i=1}^{n-1} \left| \eta \left(\frac{Y_i^*}{S_\sigma} \right) - \eta \left(\frac{\sigma U_i^*}{S_\sigma} \right) \right| \\ &\leq \|\eta\|_L \frac{1}{S_\sigma} \frac{1}{n-1} \sum_{i=1}^{n-1} |Y_i^* - \sigma U_i^*| \\ &= \|\eta\|_L \frac{1}{S_\sigma} \frac{1}{n-1} \sum_{i=1}^{n-1} |g(x_{i+1}) - g(x_i)| \\ &\leq \|\eta\|_L \|g\|_L \frac{1}{S_\sigma} \frac{1}{n-1} \sum_{i=1}^{n-1} |x_{i+1} - x_i| \\ &\leq KM_n, \end{aligned}$$

siendo $K = \|\eta\|_L \|g\|_L / S_\sigma$. Como $M_n = O(n^{-1})$, deducimos que

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{Y_i^*}{S_\sigma} \right) - \frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{\sigma U_i^*}{S_\sigma} \right) \xrightarrow{c.s.} 0. \quad (3.38)$$

Una aplicación directa de la Ley Fuerte de los Grandes Números para v.a.'s m -dependientes (Proposición 2.4) permite probar que

$$\frac{1}{n-1} \sum_{i=1}^{n-1} \eta \left(\frac{\sigma U_i^*}{S_\sigma} \right) \xrightarrow{c.s.} E \left[\eta \left(\frac{U_1^* \sigma}{S_\sigma} \right) \right]. \quad (3.39)$$

De (3.38) y (3.39) se deduce (3.37), lo que concluye la demostración. \square

3.3. Robustez de los M -estimadores de Escala

En la presente sección desarrollaremos los aspectos cuantitativos de la robustez de los M -estimadores. Más específicamente utilizaremos las nociones de máximo sesgo asintótico para derivar el punto de ruptura asintótico de los M -estimadores de escala. El concepto de máximo sesgo asintótico fue introducido, para el modelo de posición, por Huber (1964). Martin y Zamar (1989 y 1993) realizaron extensiones para modelos de escala y posición-dispersión. Una exposición detallada de estos conceptos puede verse en el Capítulo 3 de Maronna *et al.* (2006).

A lo largo de esta sección utilizaremos la misma notación que la especificada en la Subsec-

ción 3.2.1. Introduzcamos la siguiente colección de funciones de distribución

$$\mathcal{D}_0 = \{F \in \mathcal{D} : F \text{ es absolutamente continua}\} \quad (3.40)$$

(donde \mathcal{D} es como en la Definición 3.1, la colección de las funciones de distribución definidas sobre la recta real) y las siguientes hipótesis adicionales sobre la distribución del modelo central y sobre la distribución contaminante.

Hipótesis 8 Sea $G = (1 - \epsilon)F_0 + \epsilon H \in \mathcal{P}_\epsilon(F_0)$ la distribución de los errores del modelo (3.2).

(i) La distribución del modelo central $F_0 \in \mathcal{D}_0$ tiene una densidad f_0 estrictamente positiva sobre \mathbb{R} .

(ii) La distribución contaminante es absolutamente continua, o sea, $H \in \mathcal{D}_0$.

Hipótesis 9 La distribución del modelo central $F_0 \in \mathcal{D}_0$ tiene densidad f_0 estrictamente positiva, unimodal y simétrica.

Observación 3.4 La restricción a funciones de distribución contaminantes absolutamente continuas simplifica las demostraciones de las propiedades de robustez asintóticas. Teniendo en cuenta que muchas funciones de distribución, incluyendo las masas puntuales, pueden ser bien aproximadas por distribuciones absolutamente continuas, la restricción no es seria a los fines prácticos.

3.3.1. Máximo Sesgo Asintótico

Cuando no hay contaminación en los datos, es decir $\epsilon = 0$, la función de distribución de los errores del Modelo de escala global (3.2) es F_0 . En este caso, deberíamos poder estimar sin sesgo asintótico el parámetro de escala σ , es decir S debe ser consistente Fisher, es decir, $S(F_{0,\sigma}^*) = \sigma$.

De todos modos, el Teorema 3.1 sugiere que $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$ es asintóticamente sesgado cuando los datos están contaminados, o sea, la sucesión converge a $S(G_\sigma^*)$ en vez de converger a σ y, por lo tanto, si $G \neq F_0$, entonces $S(G_\sigma^*) \neq S(F_{0,\sigma}^*) = \sigma$.

En la siguiente definición, formalizamos la noción de sesgo asintótico

Definición 3.7 *El sesgo asintótico de la sucesión de M -estimadores de escala, $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$, se define como $B(S(G_\sigma^*)) = (S(G_\sigma^*)/\sigma) - 1$.*

En la estimación de la escala, la distribución contaminante puede provocar dos tipos de desviaciones que ya han sido estudiadas en el modelo de posición-escala. En efecto, la distribución contaminante puede introducir en los datos contaminaciones con valores grandes (outliers) y, por lo tanto, el estimador tenderá a sobre-estimar, o bien, puede introducir datos cercanos al parámetro de posición (o sea con errores cercanos a cero) (inliers) con lo cual, el estimador tenderá a sub-estimar la escala. En términos asintóticos, tendremos un valor de sesgo asintótico positivo en la primera situación, mientras que en la segunda, el sesgo asintótico será negativo.

Una medida de sesgo asintótico más general se presenta a continuación.

Definición 3.8 *El sesgo asintótico generalizado de la sucesión de M -estimadores de escala,*

$\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$, se define como

$$B_g(S(G_\sigma^*)) = \begin{cases} L_1\left(\frac{S(G_\sigma^*)}{\sigma}\right), & \text{si } 0 < S(G_\sigma^*) \leq \sigma, \\ L_2\left(\frac{S(G_\sigma^*)}{\sigma}\right), & \text{si } \sigma < S(G_\sigma^*) < +\infty, \end{cases}$$

donde $L_1, L_2 : \mathbb{R} \longrightarrow \mathbb{R}_0^+$ son funciones continuas y monótonas tales que

$$L_1(1) = L_2(1) = 0 \quad \text{y} \quad \lim_{s \rightarrow 0^+} L_1(s) = \lim_{s \rightarrow +\infty} L_2(s) = +\infty;$$

L_1 y L_2 son denominadas funciones de penalización.

Cabe mencionar que, por la vía de las funciones de penalización, podemos pensar de manera diferente la sub o la sobre-estimación, dándonos un concepto más flexible que el del sesgo asintótico definido por B .

La robustez de un estimador está asociada no al comportamiento en una distribución en particular sino en el entorno de contaminación del modelo central. Las dos definiciones siguientes precisan esta idea.

Definición 3.9 *El máximo sesgo asintótico generalizado de la sucesión de M -estimadores,*

$\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$, en el entorno de contaminación $\mathcal{P}_\epsilon(F_0)$ se define como

$$\overline{B}_g(\epsilon) = \sup_{G \in \mathcal{P}_\epsilon(F_0)} B_g(S(G_\sigma^*)). \quad (3.41)$$

Definición 3.10 *La sucesión de M -estimadores, $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$, se dice que es asintóticamente robusta si existe $\epsilon \in (0, 1/2]$ tal que $\overline{B}_g(\epsilon) < \infty$.*

Los conceptos de sesgo introducidos anteriormente satisfacen la propiedad de invariancia definida a continuación.

Definición 3.11 *Sea $T : \mathcal{D} \rightarrow \mathbb{R}$ un funcional. Se dice que T es invariante por transformaciones de escala si, para cada $F \in \mathcal{D}$ y para cada $\theta \in \mathbb{R}$ con $\theta > 0$, se verifica que $T(F_\theta) = T(F)$, donde F_θ denota la función de distribución de θU siendo U una v.a. tal que $U \sim F$.*

Dado que, según (3.17), el M -funcional correspondiente a la sucesión de M -estimadores es equivariante a transformaciones de escala, los funcionales definidos por $T_1(G) = B(S(G_\sigma^*))$ y $T_2(G) = B_g(S(G_\sigma^*))$ resultan invariantes por transformaciones de escala. En consecuencia, la noción de máximo sesgo asintótico introducida en la Definición 3.9 no depende de σ , es decir,

$$\overline{B}_g(\epsilon) = \sup_{G \in \mathcal{P}_\epsilon(F_0)} B_g(S(G_\sigma^*)) = \sup_{G \in \mathcal{P}_\epsilon(F_0)} B_g(S(G^*)). \quad (3.42)$$

Observación 3.5 *La igualdad (3.42) permite suponer, sin pérdida de generalidad, que la escala es $\sigma = 1$. De este modo $\overline{B}_g(\epsilon)$ depende sólo de la fracción de contaminación ϵ . La curva obtenida graficando $\overline{B}_g(\epsilon)$ versus ϵ , puede ser utilizada para estudiar visualmente las propiedades de robustez asintóticas de $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$.*

Para derivar una expresión explícita de $\overline{B}_g(\epsilon)$ definamos

$$S_\sigma^+(\epsilon) = \sup_{G \in \mathcal{P}_\epsilon(F_0)} S(G_\sigma^*) \quad (3.43)$$

$$S_\sigma^-(\epsilon) = \inf_{G \in \mathcal{P}_\epsilon(F_0)} S(G_\sigma^*). \quad (3.44)$$

Por la equivariancia del M -funcional y por la monotonía de L_1 y L_2 se deduce que

$$\overline{B}_g(\epsilon) = \text{máx} \left\{ L_1 \left(\frac{S_\sigma^-(\epsilon)}{\sigma} \right), L_2 \left(\frac{S_\sigma^+(\epsilon)}{\sigma} \right) \right\} = \text{máx} \{ L_1(S^-(\epsilon)), L_2(S^+(\epsilon)) \}. \quad (3.45)$$

En lo que sigue vamos a hallar la expresión del máximo sesgo asintótico generalizado para la sucesión de M -estimadores. De acuerdo a (3.45), primero necesitamos hallar las expresiones de $S^-(\epsilon)$ y $S^+(\epsilon)$. A continuación presentamos los enunciados de las Proposiciones que caracterizan dichas expresiones y el Teorema que caracteriza el máximo sesgo asintótico generalizado.

Proposición 3.1 *Supongamos que se cumplen las Hipótesis 1, 8(ii) y 9. Sea $\epsilon \in (0, 1/2]$ fijo y definamos, para $b > \epsilon(2 - \epsilon)$, $s^+(\epsilon)$ como la solución de $\lambda_+(s) = 0$ donde*

$$\lambda_+(s) = (1 - \epsilon)^2 E \left[\chi \left(\frac{Z_1^*}{as} \right) \right] + \epsilon(2 - \epsilon) - b, \quad (3.46)$$

con $Z_1^* = Z_2 - Z_1$, siendo Z_1 y Z_2 v.a.'s i.i.d. tales que $Z_i \sim F_0$, $i = 1, 2$. Entonces, $S^+(\epsilon)$ introducido en (3.43) satisface $S^+(\epsilon) = s^+(\epsilon)$, si $b > \epsilon(2 - \epsilon)$, y $S^+(\epsilon) = +\infty$, en caso contrario.

Proposición 3.2 *Supongamos que se cumplen las Hipótesis 1 y 8. Sea $\epsilon \in (0, 1/2]$ fijo y definamos, para $b < 1 - \epsilon^2$, $s^-(\epsilon)$ como la solución de $\lambda_-(s) = 0$, donde*

$$\lambda_-(s) = (1 - \epsilon)^2 E \left[\chi \left(\frac{Z_1^*}{as} \right) \right] + 2\epsilon(1 - \epsilon) E \left[\chi \left(\frac{Z_1^*}{as} \right) \right] - b, \quad (3.47)$$

con $Z_1^* = Z_2 - Z_1$, siendo Z_1 y Z_2 v.a.'s i.i.d. tales que $Z_i \sim F_0$, $i = 1, 2$. Entonces $S^-(\epsilon)$ introducido en (3.44) satisface $S^-(\epsilon) = s^-(\epsilon)$, si $b < 1 - \epsilon^2$, y $S^-(\epsilon) = 0$, en caso contrario.

Observación 3.6 *Vale la pena observar que los Lemas 3.3 y 3.4, que se enuncian a continuación del Teorema 3.3, garantizan que $s^+(\epsilon)$ y $s^-(\epsilon)$ existen, están unívocamente definidos y son estrictamente positivos y finitos.*

Teorema 3.3 *Supongamos que se cumplen las Hipótesis 1, 8(ii) y 9. Sean $s^+(\epsilon)$ y $s^-(\epsilon)$ como en las Proposiciones 3.1 y 3.2 respectivamente. Son válidas las siguientes afirmaciones:*

(i) *Sea $b = 3/4$. Entonces $\overline{B}_g(\epsilon) = \max\{L_1(s^-(\epsilon)), L_2(s^+(\epsilon))\}$ si $\epsilon < 1/2$ y $\overline{B}_g(\epsilon) = +\infty$ si $\epsilon = 1/2$.*

(ii) *Sea $b \in (0, 3/4)$. Entonces $\overline{B}_g(\epsilon) = \max\{L_1(s^-(\epsilon)), L_2(s^+(\epsilon))\}$ si $\epsilon < 1 - \sqrt{1-b}$ y $\overline{B}_g(\epsilon) = +\infty$ si $1 - \sqrt{1-b} \leq \epsilon$.*

(iii) *Sea $b \in (3/4, 1)$. Entonces $\overline{B}_g(\epsilon) = \max\{L_1(s^-(\epsilon)), L_2(s^+(\epsilon))\}$ si $\epsilon < \sqrt{1-b}$ y $\overline{B}_g(\epsilon) = +\infty$ si $\sqrt{1-b} \leq \epsilon$.*

Antes de probar las proposiciones y el teorema enunciamos tres lemas auxiliares, cuyas

demostraciones se encuentran en el Apéndice A. El Lema 3.2 es necesario para demostrar los dos lemas siguientes.

Lema 3.2 *Sea $G \in \mathcal{P}_\epsilon(F_0)$ y, para $s > 0$, definamos*

$$\lambda_G(s) = E \left[\chi \left(\frac{\sigma U_1^*}{as} \right) \right] - b. \quad (3.48)$$

Bajo las Hipótesis 1 y 8 son válidas las siguientes afirmaciones

(i) *La función λ_G es continua, estrictamente decreciente y existen los siguientes límites*

$$\lim_{s \rightarrow +\infty} \lambda_G(s) = -b \text{ y } \lim_{s \rightarrow 0^+} \lambda_G(s) = 1 - b.$$

(ii) *La ecuación $\lambda_G(s) = 0$ admite como única solución al M -funcional de escala $S(G_\sigma^*)$.*

(iii) *Para cualquier $s > 0$, $\lambda_G(s)$ puede descomponerse de la siguiente forma*

$$\begin{aligned} \lambda_G(s) = & (1 - \epsilon)^2 E \left[\chi \left(\frac{\sigma(V_2 - V_1)}{as} \right) \right] + 2\epsilon(1 - \epsilon) E \left[\chi \left(\frac{\sigma(V_2 - W_1)}{as} \right) \right] \\ & + \epsilon^2 E \left[\chi \left(\frac{\sigma(W_2 - W_1)}{as} \right) \right] - b, \end{aligned}$$

donde $\{V_i\}_{i=1,2}$ son v.a.'s i.i.d. tal que $V_1 \sim F_0$; $\{W_i\}_{i=1,2}$ son v.a.'s i.i.d. con $W_1 \sim H$ y (V_i, W_i) , $i = 1, 2$, son vectores aleatorios independientes.

Lema 3.3 Sea $\epsilon \in (0, 1/2]$ fijo y, para cada $n \geq 1$, sean $\{U_{i,n}\}_{i=1,2}$ v.a.'s i.i.d. con $U_{1,n} \sim G_n$ donde $G_n = (1 - \epsilon)F_0 + \epsilon H_n \in \mathcal{P}_\epsilon(F_0)$ con $H_n(y) = \Phi(y/n)$. Además, si $s > 0$, sean $\lambda_{G_n}(s) = E[\chi(U_{1,n}^*/(as))] - b$ y $\lambda_+(s)$ como en (3.46).

Bajo las Hipótesis 1 y 8(i), son válidas las siguientes afirmaciones

(i) Para cualquier $s > 0$, $\lim_{n \rightarrow \infty} \lambda_{G_n}(s) = \lambda_+(s)$. Además, la sucesión $\{\lambda_{G_n}\}$ es estrictamente creciente.

(ii) La función λ_+ es continua, estrictamente decreciente y existen los siguientes límites

$$\lim_{s \rightarrow 0^+} \lambda_+(s) = 1 - b \quad \text{y} \quad \lim_{s \rightarrow +\infty} \lambda_+(s) = \epsilon(2 - \epsilon) - b.$$

(iii) Si $\epsilon(2 - \epsilon) < b$, la ecuación $\lambda_+(s) = 0$ tiene una única solución finita y positiva.

Lema 3.4 Sea $\epsilon \in (0, 1/2]$ fijo y, para cada $n \geq 1$, sean $\{U_{i,n}\}_{i=1,2}$ v.a.'s i.i.d. con $U_{1,n} \sim G_n$ donde $G_n = (1 - \epsilon)F_0 + \epsilon H_n \in \mathcal{P}_\epsilon(F_0)$ con $H_n(y) = \Phi(ny)$. Además, si $s > 0$, sea $\lambda_{G_n}(s) = E[\chi(U_{1,n}^*/(as))] - b$ y $\lambda_-(s)$ como en (3.47).

Bajo las Hipótesis 1 y Hipótesis 8(i), son válidas las siguientes afirmaciones

(i) Para cualquier $s > 0$, $\lim_{n \rightarrow \infty} \lambda_{G_n}(s) = \lambda_-(s)$. Además, la sucesión $\{\lambda_{G_n}\}_{n \geq 1}$ es estrictamente decreciente.

(ii) La función λ_- es continua, estrictamente decreciente y existen los siguientes límites

$$\lim_{s \rightarrow 0^+} \lambda_-(s) = 1 - \epsilon^2 - b \quad \text{y} \quad \lim_{s \rightarrow +\infty} \lambda_-(s) = -b.$$

(iii) Si $1 - \epsilon^2 > b$, la ecuación $\lambda_-(s) = 0$ tiene una única solución finita y positiva.

Demostración de la Proposición 3.1. Fijemos $\epsilon \in (0, 1/2]$ tal que $\epsilon(2 - \epsilon) < b$. Por (3.43), para probar que $S^+(\epsilon) = s^+(\epsilon)$, es suficiente demostrar que las siguientes afirmaciones son válidas

(a) $S(G^*) \leq s^+(\epsilon)$ para cualquier $G \in \mathcal{P}_\epsilon(F_0)$.

(b) Existe una sucesión $\{G_n\}_{n \geq 1} \subseteq \mathcal{P}_\epsilon(F_0)$ tal que $\lim_{n \rightarrow \infty} S(G_n^*) = s^+(\epsilon)$.

Probemos (a). Dada $G \in \mathcal{P}_\epsilon(F_0)$ fija, sea $\lambda_G(s) = E[\chi(U_1^*/(as))] - b$ como en (3.48). Si probamos la siguiente inclusión

$$\{s > 0 : s > s^+(\epsilon)\} \subseteq \{s > 0 : \lambda_G(s) \leq 0\}, \quad (3.49)$$

entonces, tomando ínfimo a ambos lados y utilizando la definición de $S(G^*)$ se obtiene (a). Sea $s > s^+(\epsilon)$, por (i) del Lema 3.2, λ_G es estrictamente creciente, con lo cual $\lambda_G(s) < \lambda_G(s^+(\epsilon))$.

Por lo tanto, para probar (3.49) basta con demostrar que $\lambda_G(s^+(\epsilon)) \leq 0$.

Utilizando (iii) del Lema 3.2 podemos escribir

$$\begin{aligned} \lambda_G(s^+(\epsilon)) &= (1 - \epsilon)^2 E \left[\chi \left(\frac{Z_2 - Z_1}{as^+(\epsilon)} \right) \right] + \epsilon(1 - \epsilon) E \left[\chi \left(\frac{Z_2 - W_1}{as^+(\epsilon)} \right) \right] \\ &\quad + \epsilon(1 - \epsilon) E \left[\chi \left(\frac{W_2 - Z_1}{as^+(\epsilon)} \right) \right] + \epsilon^2 E \left[\chi \left(\frac{W_2 - W_1}{as^+(\epsilon)} \right) \right] - b, \end{aligned}$$

con $\{Z_i\}_{i=1,2}$ v.a.'s i.i.d., $Z_1 \sim F_0$, $\{W_i\}_{i=1,2}$ v.a.'s i.i.d. con $W_1 \sim H$ y (Z_i, W_i) , $i = 1, 2$, vectores independientes. Teniendo en cuenta que χ es no decreciente, par, $\|\chi\|_\infty = 1$ y la definición de λ_+ obtenemos

$$\lambda_G(s^+(\epsilon)) \leq (1 - \epsilon)^2 E \left[\chi \left(\frac{Z_2 - Z_1}{as^+(\epsilon)} \right) \right] + \epsilon(2 - \epsilon) - b = \lambda_+(s^+(\epsilon)) = 0,$$

lo que concluye la demostración de (3.49).

Demostremos (b). Sea $\{G_n\}_{n \geq 1} \subseteq \mathcal{P}_\epsilon(F_0)$ con $G_n = (1 - \epsilon)F_0 + \epsilon H_n$ y $H_n(y) = \Phi(y/n)$. Por (i) del Lema 3.2, $\{\lambda_{G_n}\}_{n \geq 1}$ es una sucesión estrictamente creciente de funciones que converge a λ_+ , con lo cual para cada $n \geq 1$, $S(G_n^*) < s^+(\epsilon)$, con $s^+(\epsilon)$ la única solución finita y positiva de la ecuación $\lambda_+(s) = 0$ (ver (iii) del Lema 3.2). De este modo, la sucesión $\{S(G_n^*)\}_{n \geq 1}$, además de ser creciente, está acotada y por lo tanto existe $\lim_{n \rightarrow \infty} S(G_n^*)$. Supongamos por el absurdo que $\lim_{n \rightarrow \infty} S(G_n^*) = s_0 < s^+(\epsilon)$ y definamos $\delta = \lambda_+(s_0)$. Como λ_+ es estrictamente decreciente con un único cero en $s^+(\epsilon)$, por ser $s_0 < s^+(\epsilon)$ se obtiene $\lambda_+(s_0) = \delta > 0$.

Por otra parte, la función λ_{G_n} es estrictamente decreciente, $0 = \lambda_{G_n}(S(G_n^*)) > \lambda_{G_n}(s_0)$, con lo cual, $0 \geq \lim_{n \rightarrow \infty} \lambda_{G_n}(s_0) = \lambda_+(s_0) = \delta$, lo que es una contradicción ya que habíamos

establecido que $\lambda_+(s_0) = \delta > 0$.

Para completar la demostración de la Proposición resta mostrar que para $\epsilon \in (0, 1/2]$ tal que $\epsilon(2 - \epsilon) \geq b$, $S^+(\epsilon) = +\infty$. Para ello basta probar que existe $\{G_n\}_{n \geq 1} \subseteq \mathcal{P}_\epsilon(F_0)$ tal que $\lim_{n \rightarrow \infty} S(G_n^*) = +\infty$. Consideremos $G_n = (1 - \epsilon)F_0 + \epsilon H_n$ con $H_n(y) = \Phi(y/n)$. Supongamos, por el absurdo, que existe $K > 0$ tal que $S(G_n^*) \leq K$ para cualquier $n \geq 1$. Usando la monotonía de λ_{G_n} (ver (i) del Lema 3.3), tenemos que $\lambda_{G_n}(S(G_n^*)) \geq \lambda_{G_n}(K)$ para cualquier $n \geq 1$. Pero $\lambda_{G_n}(S(G_n^*)) = 0$, con lo cual $\lambda_{G_n}(K) \leq 0$ para todo $n \geq 1$ y en consecuencia, por (i) del Lema 3.3, $\lim_{n \rightarrow \infty} \lambda_{G_n}(K) = \lambda_+(K) \leq 0$. Al mismo tiempo, (ii) del Lema 3.3 implica que $\lambda_+(K) > \lim_{s \rightarrow +\infty} \lambda_+(s) = \epsilon(2 - \epsilon) - b \geq 0$, que es una contradicción ya que arriba establecimos que $\lambda_+(K) \leq 0$. \square

Demostración de la Proposición 3.2. Fijemos $\epsilon \in (0, 1/2]$ tal que $1 - \epsilon^2 > b$. Por (3.44), para probar que $S^-(\epsilon) = s^-(\epsilon)$, es suficiente mostrar que las siguientes afirmaciones son válidas

(a) $s^-(\epsilon) \leq S(G^*)$ para cualquier $G \in \mathcal{P}_\epsilon(F_0)$

(b) existe una sucesión de funciones de distribución $\{G_n\}_{n \geq 1} \subseteq \mathcal{P}_\epsilon(F_0)$ tales que

$$\lim_{n \rightarrow \infty} S(G_n^*) = s^-(\epsilon).$$

Demostremos (a). Con un análisis similar al hecho en la demostración de la proposición

anterior, bastará con demostrar que para $G \in \mathcal{P}_\epsilon(F_0)$ vale la siguiente inclusión

$$\{s > 0 : s < s^-(\epsilon)\} \subseteq \{s > 0 : \lambda_G(s) > 0\}, \quad (3.50)$$

para lo cual bastará mostrar que $\lambda_G(s^-(\epsilon)) \geq 0$. Por (iii) of Lemma 3.2

$$\begin{aligned} \lambda_G(s^-(\epsilon)) &= (1 - \epsilon)^2 E \left[\chi \left(\frac{Z_2 - Z_1}{as^-(\epsilon)} \right) \right] + 2\epsilon(1 - \epsilon) E \left[\chi \left(\frac{Z_2 - W_1}{as^-(\epsilon)} \right) \right] \\ &\quad + \epsilon^2 E \left[\chi \left(\frac{W_2 - W_1}{as^-(\epsilon)} \right) \right] - b, \end{aligned} \quad (3.51)$$

con $\{Z_i\}_{i=1,2}$ v.a.'s i.i.d. tal que $Z_1 \sim F_0$ y $\{W_i\}_{i=1,2}$ v.a.'s i.i.d. tal que $W_1 \sim H$.

Analicemos el segundo término de (3.51). Tengamos en cuenta que $Z_2 - W_1$ tiene una densidad dada por $g^*(x) = \int_{-\infty}^{+\infty} h(x)f_0(x-t)dt$, donde $h = H'$ y $f_0 = F_0'$. Usando ahora la simetría y unimodalidad de f_0 junto con la hipótesis de que χ es par y creciente sobre $\{x \in \mathbb{R}^+ : \chi(x) < \|\chi\|_\infty\}$ obtenemos

$$\begin{aligned} E \left[\chi \left(\frac{Z_2 - W_1}{as^-(\epsilon)} \right) \right] &= \int_{-\infty}^{+\infty} \chi \left(\frac{x}{as^-(\epsilon)} \right) g^*(x) dx \\ &= \int_{-\infty}^{\infty} h(t) \left[\int_{-\infty}^{+\infty} \chi \left(\frac{x}{as^-(\epsilon)} \right) f_0(x-t) dx \right] dt \\ &\geq \left[\int_{-\infty}^{+\infty} h(t) dt \right] \left[\int_{-\infty}^{\infty} \chi \left(\frac{x}{as^-(\epsilon)} \right) f_0(x) dx \right] \\ &= E \left[\chi \left(\frac{Z_2}{as^-(\epsilon)} \right) \right]. \end{aligned}$$

Considerando esta desigualdad y dado que el tercer término de (3.51) es claramente positivo,

obtenemos

$$\lambda_G(s^-(\epsilon)) \geq (1 - \epsilon)^2 E \left[\chi \left(\frac{Z_2 - Z_1}{as^-(\epsilon)} \right) \right] + 2\epsilon(1 - \epsilon) E \left[\chi \left(\frac{Z_2}{as^-(\epsilon)} \right) \right] - b = \lambda_-(s^-(\epsilon)) = 0.$$

Para demostrar (b) se procede de manera similar a la demostración de (b) en la Proposición 3.1, utilizando la sucesión con términos $G_n = (1 - \epsilon)F_0 + \epsilon H_n$ donde $H_n(y) = \Phi(ny)$ y aplicando el Lema 3.4. Análogamente, se prueba que, si $1 - \epsilon^2 \leq b$, entonces $\lim_{n \rightarrow \infty} S(G_n^*) = 0$. \square

Demostración del Teorema 3.3. La prueba del teorema es inmediata si tenemos en cuenta que dado $\epsilon \in (0, 1/2]$ sólo hay que resolver, respecto a ϵ , el siguiente sistema de desigualdades

$$0 < \epsilon \leq 1/2, \quad \epsilon(2 - \epsilon) < b \quad \text{y} \quad 1 - \epsilon^2 > b.$$

De aquí se deduce que los valores de ϵ deben satisfacer

$$\epsilon \in (0, 1/2], \quad \epsilon \in (-\infty, 1 - \sqrt{1 - b}) \cup (1 + \sqrt{1 - b}, +\infty) \quad \text{y} \quad \epsilon \in (-\sqrt{1 - b}, \sqrt{1 - b}).$$

En particular, si $b = 3/4$, entonces $\epsilon \in (0, 1/2)$ y si $b \in (0, 3/4)$, $\epsilon \in (0, 1 - \sqrt{1 - b})$; finalmente, si $b \in (3/4, 1)$, entonces $\epsilon \in (0, \sqrt{1 - b})$.

3.3.2. Punto de Ruptura Asintótico

Intuitivamente, lo que queremos introducir ahora es un concepto clásico en robustez, que nos dirá cuál es la menor cantidad de contaminación que el estimador no puede tolerar, en el sentido de que ya no nos siga proporcionando alguna información sobre el parámetro a estimar.

Definición 3.12 *Dada la sucesión de M -estimadores de escala, $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$, su punto de ruptura asintótico se define como*

$$\epsilon^* = \epsilon^*(\{\hat{\sigma}_{M,n}\}_{n \geq 1}) = \inf \{ \epsilon \in (0, 1/2] : \bar{B}_g(\epsilon) = \infty \}. \quad (3.52)$$

Equivalentemente, el punto de ruptura asintótico es la menor proporción de contaminación ϵ tal que el estimador deja de ser asintóticamente robusto.

Por las propiedades de las funciones de penalización, es inmediato comprobar que

$$\epsilon^* = \min(\epsilon^0, \epsilon^\infty), \quad (3.53)$$

donde $\epsilon^0 = \inf\{\epsilon \in (0, 1/2] : S^-(\epsilon) = 0\}$ y $\epsilon^\infty = \inf\{\epsilon \in (0, 1/2] : S^+(\epsilon) = +\infty\}$.

A ϵ^0 y ϵ^* se los denominan *punto de ruptura asintótico por implosión* y *punto de ruptura asintótico por explosión*, respectivamente, de la sucesión de M -estimadores. Estos conceptos dan cuenta de la siguiente situación. Si la cantidad de contaminación en los datos es demasiado grande, $\{\hat{\sigma}_{M,n}\}_{n \geq 1}$ puede sufrir dos tipos de puntos ruptura asintóticos: o bien puede explotar,

en el sentido de tomar valores arbitrariamente grandes o puede sufrir implosión, tomando valores arbitrariamente pequeños (ver Huber, 1981, pág. 110).

El siguiente teorema y su corolario presentan el principal resultado de esta sección, ya que establecen el punto de ruptura asintótico de $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$, el cual es una consecuencia directa del Teorema 3.3.

Teorema 3.4 *El punto de ruptura asintótico de la sucesión de los M -estimadores, $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$, ϵ^* definido en (3.52) satisface:*

(i) Si $b = 3/4$, entonces $\epsilon^* = 1/2$.

(ii) Si $b \in (0, 3/4)$, entonces $\epsilon^* = 1 - \sqrt{1 - b}$.

(iii) Si $b \in (3/4, 1)$, entonces $\epsilon^* = \sqrt{1 - b}$.

Corolario 3.1 *El máximo punto de ruptura asintótico, ϵ_{opt}^* , de la sucesión de los M -estimadores $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$ que puede ser alcanzado cuando la constante de ajuste b varía en el intervalo $(0, 1)$ es $\epsilon_{opt}^* = 1/2$ y se obtiene para $b = 3/4$.*

Hasta aquí hemos investigado las propiedades de robustez asintótica de la familia de los M -estimadores de escala, $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$ para una función de escores general que satisfaga la Hipótesis 1. Ahora bien, desde el punto de vista práctico, χ debe ser especificada.

Fijemos como modelo central $F_0 = \Phi$. Una elección particular es la clase de las funciones de Beaton-Tukey, $\{\chi_c\}_{c > 0}$, introducidas en (3.11) y es la que nosotros escogeremos. La constante c debe ser determinada de tal modo que $\{\widehat{\sigma}_{M,n}\}_{n \geq 1}$

- (i) alcance su punto de ruptura asintótico óptimo
- (ii) su valor límite $S(G^*)$, sea consistente Fisher cuando $G = F_0$, es decir cuando no hay contaminación.

El Corolario 3.1 nos dice que la elección de $b = 3/4$ permite alcanzar el máximo punto de ruptura asintótico posible de $1/2$. De acuerdo a las consideraciones de la consistencia Fisher debemos elegir c tal que $E[\chi_c(Z_1)] = 3/4$, donde $Z_1 \sim \Phi$. Esta ecuación produce un valor (aproximado) de $c = 0.70417$. Esta es la razón de haber elegido esas constantes en el Ejemplo 3.11. De este modo tendremos $\epsilon^*(\{\hat{\sigma}_{BT,n}\}_{n \geq 1}) = 1/2$. Es claro que el estimador de Rice $\hat{\sigma}_{R,n}$ (ver Ejemplo 3.1), como χ es no acotada, tiene un punto de ruptura asintótico igual a cero. Para los M -estimadores de Boente, Fraiman y Meloche $\hat{\sigma}_{BFM,n}$, y primer cuartil, $\hat{\sigma}_{PC,n}$, introducidos en los Ejemplos 3.2 y 3.3, respectivamente, no es posible deducir su punto de ruptura asintótico a partir de la teoría aquí construida ya que sus funciones de escores no son suaves (ni siquiera continuas). Pero, sí podemos utilizar el argumento heurístico de que estas funciones escores pueden ser bien aproximadas por funciones escores suaves. De este modo, si aplicamos el Teorema 3.4 con $b = 1/2$ para $\{\hat{\sigma}_{BFM,n}\}_{n \geq 1}$ y $b = 3/4$ para $\{\hat{\sigma}_{PC,n}\}_{n \geq 1}$ para garantizar la consistencia Fisher, entonces podemos conjeturar que los puntos de rupturas son, aproximadamente 0.29 y $1/2$ respectivamente. En síntesis, tenemos

$$\begin{cases} \epsilon^*(\{\hat{\sigma}_{R,n}\}_{n \geq 1}) = 0, & \epsilon^*(\{\hat{\sigma}_{BFM,n}\}_{n \geq 1}) \approx 0.29, \\ \epsilon^*(\{\hat{\sigma}_{PC,n}\}_{n \geq 1}) \approx 1/2, & \epsilon^*(\{\hat{\sigma}_{BT,n}\}_{n \geq 1}) = 1/2 \end{cases} \quad (3.54)$$

Los resultados de la Sección 3.4 van a confirmar (numéricamente) esta conjetura.

Observación 3.7 *Es muy importante aclarar que sólo estamos considerando el punto de ruptura asintótico bajo el modelo de contaminaciones independientes. En este contexto, nuestros M -estimadores alcanzar el máximo punto de ruptura asintótico de $1/2$. Este no es el caso cuando consideramos el contexto de muestras finitas, donde podemos construir configuraciones particulares de outliers que cambian el punto de ruptura. Por ejemplo, $\epsilon\%$ outliers intercalados van a contaminar $2\epsilon\%$ de las diferencias. De todos modos, estas configuraciones particulares tienen asintóticamente probabilidad cero de aparición bajo el modelo de contaminación independiente.*

3.4. Resultados Monte Carlo

3.4.1. Análisis y Conclusiones

Esta Sección contiene los resultados de un estudio de simulación diseñado con el objetivo de evaluar el comportamiento, sobre muestras finitas, de los estimadores $\hat{\sigma}_{R,n}$, $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ introducidos en los Ejemplos 3.1 a 3.4.

Los objetivos particulares del estudio son

- (a) investigar las propiedades de eficiencia de $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ relativas a $\hat{\sigma}_{R,n}$ en ausencia de contaminación,
- (b) comparar los errores cuadráticos medios (E.C.M.) estimados de los cuatro estimadores

en presencia de outliers,

- (c) estudiar si el incremento de dicho error puede atribuirse a la contribución del sesgo y/o de la varianza a través de los boxplots de los valores de $\log(\hat{\sigma}_n/\sigma)$, donde $\hat{\sigma}_n$ representa cualquiera de los estimadores mencionados arriba.

El modelo considerado en nuestro estudio de simulación es

$$Y_i = g(x_i) + \sigma U_i, \quad i = 1, \dots, n,$$

donde $\sigma = 1$, $g(x) = \text{sen}(4\pi x)$ y $x_i = (i - 1)/(n - 1)$, $1 \leq i \leq n$. Se generaron muestras de tamaños $n = 20, 50$ y 100 .

Para la generación de los datos consideramos la distribución de los errores en dos escenarios con distintos modos de producir contaminación.

- (I) *El modelo de contaminaciones independientes*, que es el modelo descrito en la Definición 3.2, corresponde a errores $\{U_i\}_{i=1}^n$ independientes, con distribución común $G = (1 - \epsilon)F_0 + \epsilon H$, donde $F_0 = \Phi$. Consideraremos dos distribuciones contaminantes

(I.a) $U_i \sim H(y) = \Phi(y/10)$ para modelar contaminaciones simétricas.

(I.b) $U_i \sim H(y) = \Phi(y - 10)$ para modelar contaminaciones asimétricas.

Para cada configuración particular, se considerarán distintos porcentajes de contaminación.

(II) *El modelo de contaminaciones intercaladas.* En este contexto, los errores son de la siguiente manera: $U_{2i-1} \sim H(y) = \Delta_{10}(y)$ para $i \in C = C_{n,\epsilon} = \{1, 2, \dots, \lfloor n\epsilon \rfloor\}$ y $U_i \sim F_0 = \Phi$ en caso contrario, donde Δ_{10} denota la masa puntual en 10 y $\lfloor a \rfloor$ indica la parte entera de a . Como en (I), se considerarán diferentes proporciones ϵ de contaminación.

De este modo tenemos diferentes configuraciones del modelo, que resulta de la combinación de

- (a) los diferentes tamaños muestrales n
- (b) los diferentes porcentajes ϵ de contaminación
- (c) los modelos de contaminaciones independientes (simétrica y asimétrica) y el modelo de contaminación intercalada.

Para cada configuración particular del modelo, realizamos $NR = 10000$ replicaciones. Las Figuras B.1 a B.6 y los Cuadros B.1 a B.4, que se encuentran en el Apéndice B, resumen los resultados obtenidos. Para no extender el análisis, los boxplots (Figuras B.3 a B.6) se han construido sólo para para muestras de tamaño $n = 100$ y algunos niveles de contaminación.

La Figura B.1 muestra los datos generados con $n = 100$ y $H(y) = \Phi(y/10)$ para diferentes valores de ϵ correspondientes a una de las replicaciones, mientras que la Figura B.2 provee el mismo esquema pero con $H(y) = \Phi(y - 10)$. Como esperamos ambas figuras revelan que a mayor tamaño de contaminación ϵ , mayor número de outliers hay en los datos. Cuando la

contaminación es simétrica, los outliers tienden a localizarse por encima y por debajo de la verdadera curva de regresión, produciendo un incremento en la variabilidad. Por otra parte, en el caso asimétrico los outliers se concentran solamente por encima de dicha curva, con lo cual se espera obtener sesgo en el estimador de la función de regresión.

Nos interesará evaluar la eficiencia relativa (E.R.) de los estimadores robustos $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ en relación al estimador no robusto $\hat{\sigma}_{\text{R},n}$ cuando $\epsilon = 0$. Si $\hat{\sigma}_n$ representa cualquiera de los estimadores robustos, evaluaremos la eficiencia de $\hat{\sigma}_n$ relativa a $\hat{\sigma}_{\text{R},n}$ a través de

$$\text{E.R.}(\hat{\sigma}_n, \hat{\sigma}_{\text{R},n}) = \frac{\widehat{\text{Var}}(\hat{\sigma}_{\text{R},n})}{\widehat{\text{Var}}(\hat{\sigma}_n)},$$

donde

$$\widehat{\text{Var}}(\hat{\sigma}_n) = \frac{1}{NR} \sum_{i=1}^{NR} \left(\hat{\sigma}_{n,i} - \overline{\hat{\sigma}_n} \right)^2. \quad (3.55)$$

Aquí, $\hat{\sigma}_{n,i}$ es el valor de $\hat{\sigma}_n$ correspondiente a la i -ésima muestra generada de la configuración del modelo de interés (i -ésima replicación) y $\overline{\hat{\sigma}_n} = (1/NR) \sum_{i=1}^{NR} \hat{\sigma}_{n,i}$ es la media sobre las replicaciones.

Antes de proseguir con el análisis de los resultados de nuestro estudio de Monte Carlo, recordemos las consideraciones que hicimos respecto a los puntos de ruptura asintóticos de los estimadores cuyos desempeños vamos a comparar. Más específicamente, en (3.54) habíamos concluido que

$$\begin{aligned} \epsilon^*(\{\hat{\sigma}_{\text{R},n}\}_{n \geq 1}) &= 0, & \epsilon^*(\{\hat{\sigma}_{\text{BFM},n}\}_{n \geq 1}) &\approx 0.29, \\ \epsilon^*(\{\hat{\sigma}_{\text{PC},n}\}_{n \geq 1}) &\approx 1/2, & \epsilon^*(\{\hat{\sigma}_{\text{BT},n}\}_{n \geq 1}) &= 1/2 \end{aligned}$$

Observemos que tanto $\hat{\sigma}_{\text{PC},n}$ como $\hat{\sigma}_{\text{BT},n}$ tienen casi el mismo punto de ruptura asintótico (recordar (3.54) y la discusión previa a esa ecuación), con lo cual resulta apropiado comparar las eficiencias relativas. Comparar la eficiencia relativa de $\hat{\sigma}_{\text{BFM},n}$ con la de $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ puede no ser apropiada ya que $\hat{\sigma}_{\text{BFM},n}$ tiene un punto de ruptura asintótico mucho menor (0.29 versus 0.5 de los anteriores) y por lo tanto, se espera que resulte más eficiente.

El Cuadro B.1 muestra los valores de eficiencia relativa. En este cuadro podemos observar que $\hat{\sigma}_{\text{BFM},n}$ alcanza una eficiencia mejor que $\hat{\sigma}_{\text{PC},n}$ y que $\hat{\sigma}_{\text{BT},n}$ a expensas de la pérdida de robustez asintótica, como se mencionó antes. Observemos que $\hat{\sigma}_{\text{BT},n}$ es ligeramente superior a $\hat{\sigma}_{\text{PC},n}$ en cuanto a su eficiencia relativa siendo comparable su punto de ruptura asintótico por lo que se concluye que $\hat{\sigma}_{\text{BT},n}$ parece ser el mejor candidato si se desea estimar robustamente la escala, aunque hay que tener en cuenta el problema de la baja eficiencia.

Comparemos ahora el comportamiento de los estimadores desde el punto de vista del error cuadrático medio bajo contaminación por outliers. Para cada configuración del modelo, estimamos los errores cuadráticos medios (E.C.M.) de los estimadores a través de

$$\text{E.C.M.}(\hat{\sigma}_n) = \frac{1}{NR} \sum_{i=1}^{NR} (\hat{\sigma}_{n,i} - \sigma)^2$$

donde $\hat{\sigma}_{n,i}$ está definido como en (3.55).

El Cuadro B.2 muestra los E.C.M. estimados de los cuatro estimadores de escala para el modelo de contaminaciones independientes con contaminación simétrica, mientras que el Cuadro B.3 corresponde a contaminaciones asimétricas. Por otra parte, el Cuadro B.4 exhibe

los resultados para el modelo de contaminaciones intercaladas. Del análisis de dichos cuadros podemos concluir lo siguiente

- (C1) Como era de esperar, independientemente del tamaño muestral y la cantidad de contaminación, $\hat{\sigma}_{R,n}$ tiene el peor desempeño desde el punto de vista del E.C.M. estimado en presencia de contaminación. Vale la pena mencionar que tanto para contaminación asimétrica como intercalada, cuando $\hat{\sigma}_{BFM,n}$ se quiebra da peores resultados que $\hat{\sigma}_{R,n}$. Esto se debe a que con un 40% de contaminación asimétrica, en algunas replicaciones, la cantidad de datos atípicos en la muestra es mayor al 50% y por lo tanto, la mediana se ve seriamente afectada perturbando, por lo tanto, al estimador de $\hat{\sigma}_{BFM,n}$. Este efecto es aún mayor en la contaminación intercalada.
- (C2) Para todos los tamaños muestrales y porcentajes de contaminación, el E.C.M. estimado de $\hat{\sigma}_{PC,n}$ y de $\hat{\sigma}_{BT,n}$ es ligeramente peor que el de $\hat{\sigma}_{BFM,n}$ cuando la cantidad de contaminación es pequeña, esto es para niveles de $\epsilon = 0.05$ o 0.10 , siendo el E.C.M. estimado de $\hat{\sigma}_{BT,n}$ el más parecido al de $\hat{\sigma}_{BFM,n}$. En cambio, cuando ϵ aumenta $\hat{\sigma}_{PC,n}$ y $\hat{\sigma}_{BT,n}$ se comportan mejor que $\hat{\sigma}_{BFM,n}$ para contaminaciones simétricas (Cuadro B.2) y sustancialmente mejor que $\hat{\sigma}_{BFM,n}$ para contaminaciones asimétricas e intercaladas (Cuadros B.3 y B.4), siendo los resultados entre ellos comparables.
- (C3) En el caso de contaminaciones intercaladas (Cuadro B.4) todos los estimadores dan claros signos de deterioro a partir de $\epsilon = 0.25$.

Las Figuras B.3 a B.6 presentan los boxplots de $\log(\hat{\sigma}_n/\sigma)$ cuando $n = 100$ y $\epsilon = 0, 0.10,$

0.20, 0.30 y 0.40. Estos gráficos permiten observar que

- (C4) Cuando no hay contaminación (Figura B.3), el estimador $\hat{\sigma}_{\text{BFM},n}$ tiene el mejor comportamiento ya que es el más parecido al estimador de Rice y presenta la menor pérdida de eficiencia como se observó anteriormente. Por otra parte, $\hat{\sigma}_{\text{PC},n}$ presenta las colas más pesadas obteniéndose en algunas replicaciones valores mucho menores que el valor verdadero (por ejemplo, valores inferiores a $e^{-0.5} \approx 0.60653$).
- (C5) Bajo contaminación simétrica, el estimador $\hat{\sigma}_{\text{R},n}$ se ve afectado tanto en sesgo como en variabilidad. Bajo contaminaciones asimétricas y más marcadamente en el esquema de contaminación intercalada, el sesgo tiene mayor impacto que la varianza sobre el E.C.M. estimado de $\hat{\sigma}_{\text{R},n}$, particularmente para niveles de contaminación del 20 % o superiores.
- (C6) En los esquemas de contaminación asimétrica e intercalada, el estimador $\hat{\sigma}_{\text{BFM},n}$ se quiebra para $\epsilon = 0.40$ y $\epsilon = 0.30, 0.35$, respectivamente. En ambos casos, el sesgo es importante siendo mayor en las contaminaciones intercaladas. Por otra parte, bajo contaminación asimétrica se produce un incremento en la variabilidad.
- (C7) Habíamos mencionado que para niveles bajos de contaminación $\hat{\sigma}_{\text{BFM},n}$ tiene un mejor desempeño que los otros estimadores robustos. En este sentido hay que observar que para un 10 % de contaminación los tres estimadores tienen un sesgo similar aunque la variabilidad de $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ es mayor que la de $\hat{\sigma}_{\text{BFM},n}$. Por otra parte, a partir de un 20 % de contaminación ya se observa sesgo en el estimador $\hat{\sigma}_{\text{BFM},n}$. Esto se debe a que a pesar de que el punto de ruptura asintótico es 0.29, para muestras finitas el punto

de ruptura efectivo puede ser mucho menor observándose sesgo para valores bastante menores que el punto de ruptura asintótico.

En resumen, de las discusiones anteriores podemos concluir que, desde un punto de vista práctico recomendamos $\hat{\sigma}_{PC,n}$ o $\hat{\sigma}_{BT,n}$, siendo el segundo mejor si se sospecha que la cantidad de outliers no supera al 20 %.

Capítulo 4

Estimación de la escala local

4.1. Preliminares

Como en el Capítulo 3, vamos a introducir formalmente el modelo que será objeto de estudio en este capítulo.

Definición 4.1 Modelo de Regresión No-paramétrico con Escala Local

Consideremos el modelo

$$Y_i = g(x_i) + \sigma(x_i)U_i, \quad i = 1, \dots, n. \quad (4.1)$$

donde $\{Y_i\}_{i=1}^n$ son las respuestas observadas, $\{x_i\}_{i=1}^n$ son puntos de diseño fijos (valores reales, no aleatorios, en el intervalo $[0, 1]$) tales que $0 \leq x_1 \leq \dots \leq x_n \leq 1$ y U_i , $1 \leq i \leq n$ son los errores aleatorios independientes e idénticamente distribuidos. Las funciones $g : [0, 1] \rightarrow \mathbb{R}$ y $\sigma : [0, 1] \rightarrow (0, \infty)$ son las funciones de regresión y de escala, respectivamente, ambas desconocidas. Supondremos que $U_i \sim G$ con $G \in \mathcal{P}_\epsilon(F_0)$, $\epsilon \in [0, 1/2)$ fijo, donde $F_0 \in \mathcal{D}$ es el

modelo central. $\mathcal{P}_\epsilon(F_0)$ y \mathcal{D} son como en la Definición 3.1.

A diferencia del problema planteado en el Capítulo 3, el problema ahora consiste en estimar $\sigma(x)$ para $x \in (0, 1)$ arbitrario. La escala dejó de ser global para convertirse en local, es decir, en una función no constante.

Definición 4.2 Para el Modelo introducido en la Definición 4.1 consideremos $\sigma(x)$ con $x \in (0, 1)$. Definimos la familia de M -estimadores de la escala $\sigma(x)$ basados en las diferencias sucesivas de las respuestas $\{Y_i^* = Y_{i+1} - Y_i\}_{i=1}^{n-1}$ con función de escores χ , funciones de peso $\{w_{n,i}(x)\}_{i=1}^{n-1}$ y constantes de ajuste a y b como

$$\hat{\sigma}_{M,n}(x) = \inf \left\{ s > 0 : \sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{Y_i^*}{as} \right) \leq b \right\}. \quad (4.2)$$

La función $\chi : \mathbb{R} \rightarrow \mathbb{R}_0^+$ es par, no decreciente en \mathbb{R}^+ , con $\chi(0) = 0$ y $0 < \|\chi\|_\infty$ como en la Definición 3.3. Las constantes $a \in (0, \infty)$ y $b \in (0, 1)$ satisfacen las condiciones ya enunciadas en (3.5), a saber,

$$E[\chi(Z_1)] = b \quad y \quad E \left[\chi \left(\frac{Z_1^*}{a} \right) \right] = b,$$

donde $Z_1^* = Z_2 - Z_1$ con Z_1 y Z_2 v.a.'s i.i.d. con distribución común F_0 . Para abreviar, a $\hat{\sigma}_{M,n}(x)$ lo llamaremos M -estimador local de escala.

Como funciones de escores se pueden considerar las introducidas en los Ejemplos 3.1 a 3.4.

En cuanto a las familias de pesos, típicos ejemplos de familias de pesos son los siguientes:

Ejemplo 4.1 En los ejemplos $\{h_n\}_{n \geq 1}$ es una sucesión de aberturas de ventana tal que $h_n \rightarrow 0$ y $nh_n \rightarrow \infty$. Sea K un núcleo, es decir un función $K : \mathbb{R} \rightarrow \mathbb{R}$ acotada, de cuadrado integrable, con $\int K(u)du = 1$ y además que $\lim_{u \rightarrow \infty} u^2 K(u) = 0$.

(i) Pesos basados en núcleos.

(a) Pesos de Nadaraya-Watson

$$w_{n,i}(x) = K\left(\frac{x - x_i}{h_n}\right) / \sum_{j=1}^{n-1} K\left(\frac{x - x_j}{h_n}\right), \quad i = 1, \dots, n$$

(b) Pesos de Rosenblatt

$$w_{n,i}(x) = (nh_n)^{-1} K\left(\frac{x - x_i}{h_n}\right), \quad i = 1, \dots, n$$

(c) Pesos de Priestley-Chao

$$w_{n,i}(x) = (h_n)^{-1} (x_{i+1} - x_i) K\left(\frac{x - x_j}{h_n}\right), \quad i = 1, \dots, n$$

(ii) Pesos basados en vecinos más cercanos.

Sea $\{k_n\}_{n \geq 1}$ una sucesión de enteros tal que $k_n \rightarrow \infty$ y asumamos que los puntos del diseño satisfacen $x_1 < x_2 < \dots < x_n$. Consideremos la sucesión $D_i(x) = |x_i - x|$ ordenada de menor a mayor, es decir $D^{(1)}(x) \leq \dots \leq D^{(n)}(x)$, y sea $D^{(k_n)}(x)$ la distancia

al k -ésimo vecino más cercano. Entonces, para $i = 1, \dots, n$

$$w_{n,i}(x) = \begin{cases} 1/k_n & \text{si } |x_i - x| \leq D^{(k_n)}(x) \\ 0 & \text{si } |x_i - x| > D^{(k_n)}(x). \end{cases} \quad (4.3)$$

(iii) Pesos de Gasser-Müller.

Sea K un núcleo y sea $\{s_i\}_{i=0}^n$ una sucesión de términos interpolantes tales que, $x_{i-1} \leq s_i \leq x_i$, $i = 1, \dots, n-1$, $x_0 = s_0 = 0$ y $s_n = 1$. Así,

$$w_{n,i}(x) = h_n^{-1} \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h_n}\right) du, \quad i = 1, \dots, n.$$

Vamos a realizar un conjunto de observaciones análogas a las que hicimos para los M -estimadores de la escala global.

Observación 4.1

(i) El ínfimo en (3.4) es necesario cuando la función de escores es discontinua. Para el desarrollo de la teoría asintótica de los M -estimadores de la escala local, vamos a suponer que se cumple la Hipótesis 1, que establecía que la función de escores χ es continua, acotada, par, estrictamente creciente sobre el conjunto $\{x : \chi(x) < \|\chi\|_\infty\}$ con $\chi(0) = 0$ y $\|\chi\|_\infty = 1$.

(ii) Bajo la Hipótesis 1, $\hat{\sigma}_{M,n}(x)$ es la única solución positiva de la ecuación

$$\sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{Y_{i+1} - Y_i}{a \hat{\sigma}_{M,n}(x)} \right) = b. \quad (4.4)$$

(iii) Las constantes a y b son elegidas de tal modo de asegurar que, en el modelo central F_0 , la sucesión $\{\hat{\sigma}_{M,n}(x)\}_{n \geq 1}$ sea asintóticamente consistente Fisher.

Definamos ahora el M -funcional local de escala al cual converge (como se demostrará más adelante) la sucesión de M -estimadores $\{\hat{\sigma}_{M,n}(x)\}_{n \geq 1}$.

Definición 4.3 Sea $S(\cdot)$ el M -funcional de escala introducido en la Definición 3.5. Definimos el M -funcional local de escala, correspondiente al M -estimador local $\hat{\sigma}_{M,n}(x)$, como el valor $S(G_x)$ donde G_x es la función de distribución de la v.a. $\sigma(x)U_1^* = \sigma(x)(U_2 - U_1)$ con U_1 y U_2 i.i.d. con distribución común G . Es decir,

$$S(G_x) = \inf \left\{ \sigma > 0 : E \left[\chi \left(\frac{\sigma(x)(U_2 - U_1)}{a\sigma} \right) \right] \leq b \right\}. \quad (4.5)$$

Observación 4.2

a) Dado $x \in (0, 1)$ fijo, si consideramos la función de distribución (condicional) empírica

$$F_n(y|x) = \sum_{i=1}^{n-1} w_{n,i}(x) I_{(-\infty, y]}(Y_i^*)$$

entonces por (4.2) y (4.5) obtenemos

$$S(F_n(\cdot|x)) = \hat{\sigma}_{M,n}(x). \quad (4.6)$$

b) Bajo la Hipótesis 1, $S(G_x)$ es la única solución positiva de la ecuación

$$E \left[\chi \left(\frac{\sigma(x)U_1^*}{aS(G_x)} \right) \right] = b. \quad (4.7)$$

De (4.7) y (3.5) se deduce fácilmente que el funcional es consistente Fisher, es decir

$$\forall x \in (0, 1) : S(F_{0x}) = \sigma(x). \quad (4.8)$$

4.2. Teoría Asintótica de los M -estimadores locales de la función de Escala

En la presente subsección consideraremos $\{Y_i\}_{i=1}^n$ una sucesión de v.a.'s independientes satisfaciendo el modelo (4.1), $\{\hat{\sigma}_{M,n}(x)\}_{n \geq 1}$ la sucesión de M -estimadores locales de la función de escala definida en (3.4), G_x y $S(G_x)$ como en la Definición 4.3. A continuación recordamos la notación introducida en (3.3) e introducimos notación adicional. Sean para $i = 1, \dots, n-1$

$$Y_i^* = Y_{i+1} - Y_i, \quad U_i^* = U_{i+1} - U_i \quad \text{y} \quad \tilde{U}_i = \sigma(x_{i+1})U_{i+1} - \sigma(x_i)U_i. \quad (4.9)$$

Consideremos dos hipótesis necesarias para establecer la consistencia fuerte de la sucesión de los M -estimadores.

Hipótesis 10 *La función de escala $\sigma : [0, 1] \rightarrow \mathbb{R}^+$ es continua.*

Hipótesis 11 *La sucesión de pesos $\{w_{n,i}(x)\}_{i=1}^{n-1}$ satisface*

$$(i) \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} w_{n,i}(x) = 1.$$

$$(ii) \text{ Existe } M > 0 \text{ tal que } \sum_{i=1}^{n-1} |w_{n,i}(x)| \leq M, \text{ para todo } n \geq 2.$$

$$(iii) \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} |w_{n,i}(x)| I_{\{|x_i - x| \geq a\}} = 0, \text{ para todo } a > 0.$$

$$(iv) \lim_{n \rightarrow \infty} w_n \log n = 0, \text{ donde } w_n = \max_{1 \leq i \leq n-1} |w_{n,i}(x)|.$$

Teorema 4.1 *Consistencia Fuerte de los M -estimadores.*

Bajo las Hipótesis 1, 2, 3, 10 y 11 para cada $x \in (0, 1)$ la sucesión de M -estimadores es fuertemente consistente a $S(G_x)$, es decir,

$$\hat{\sigma}_{M,n}(x) \xrightarrow{c.s.} S(G_x). \quad (4.10)$$

Para la prueba de la consistencia fuerte necesitaremos los dos lemas siguientes. La demostración del primero se puede hallar en Bosq (1996, pág. 22) y la del segundo en el Apéndice A.

Lema 4.1 Desigualdad de Hoeffding. Sean X_1, \dots, X_n v.a.'s independientes con media cero tales que $a_i \leq X_i \leq b_i$, $i = 1, \dots, n$ donde $a_1, b_1, \dots, a_n, b_n$ son constantes. Si $S_n = \sum_{i=1}^n X_i$ entonces, para cada $\epsilon > 0$, es válida la siguiente desigualdad

$$P(|S_n| \geq \epsilon) \leq 2 \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Lema 4.2 Sean Z_1, \dots, Z_n v.a.'s independientes con media cero y uniformemente acotadas con $\sup_{1 \leq i \leq n} |Z_i| \leq c$ y sea $\{w_{n,i}\}_{i=1}^{n-1}$ una sucesión de pesos.

(a) Si $\{w_{n,i}\}_{i=1}^{n-1}$ satisface la Hipótesis 11(ii) entonces, para cada $\epsilon > 0$,

$$P\left(\left|\sum_{i=1}^{n-1} w_{n,i} Z_i\right| \geq \epsilon\right) \leq 2 \exp\left(-\frac{\epsilon^2}{2c^2 M w_n}\right), \quad (4.11)$$

donde w_n es como en la Hipótesis 11.

(b) Si, además, la sucesión de pesos verifica la Hipótesis 11(iv), entonces

$$\sum_{i=1}^{n-1} w_{n,i} Z_i \rightarrow 0, \text{ completamente.} \quad (4.12)$$

Demostración del Teorema 4.1. Vamos a utilizar una estrategia similar a la de la demostración del Teorema 3.1. Fijemos $x \in (0, 1)$ y consideremos la función de distribución empírica condicional $F_n(y|x) = \sum_{i=1}^{n-1} w_{n,i}(x) I_{(-\infty, y]}(Y_i^*)$, que denotaremos por simplicidad $F_n(y)$, para cada

$y \in \mathbb{R}$, omitiendo la dependencia de x . Análogamente, definamos la función de distribución empírica (condicional) $\tilde{F}_n(y) = \sum_{i=1}^{n-1} w_{n,i}(x) I_{(-\infty, y]}(\tilde{U}_i)$. Como la Hipótesis 1 implica que S es continuo respecto de la topología débil, y teniendo en cuenta que $S(F_n) = \hat{\sigma}_{M,n}(x)$, para demostrar (4.10) será suficiente probar que el siguiente límite es válido

$$\pi(F_n, G_x) \xrightarrow{c.s.} 0, \quad (4.13)$$

donde π es la distancia de Prohorov. Para obtener (4.13), bastará demostrar que

$$\pi(F_n, \tilde{F}_n) \xrightarrow{c.s.} 0 \quad (4.14)$$

$$\pi(\tilde{F}_n, G_x) \xrightarrow{c.s.} 0. \quad (4.15)$$

Para mostrar (4.14), es suficiente probar que, para cualquier función f continua, Lipschitz y acotada ($f \in \mathcal{C}_L(\mathbb{R})$),

$$\int f(y) d(F_n - \tilde{F}_n)(y) \xrightarrow{c.s.} 0,$$

o, equivalentemente

$$\sum_{i=1}^{n-1} w_{n,i}(x) f(Y_i^*) - \sum_{i=1}^{n-1} w_{n,i}(x) f(\tilde{U}_i) \xrightarrow{c.s.} 0.$$

Por la Hipótesis 2, $M_n = \max_{1 \leq i \leq n-1} (x_{i+1} - x_i) \rightarrow 0$, luego, existe n_0 tal que $M_n < \delta$, para cada $n \geq n_0$. Además, por la continuidad de g , dado $\epsilon > 0$, existe $\delta > 0$ tal que si $|u - v| < \delta$ entonces $|g(u) - g(v)| < \epsilon / (M \|f\|_L)$ donde M está dado en la Hipótesis 11(ii). A partir de

estas afirmaciones, deducimos las siguientes desigualdades

$$\begin{aligned}
\left| \sum_{i=1}^{n-1} w_{n,i}(x) f(Y_i^*) - \sum_{i=1}^{n-1} w_{n,i}(x) f(\tilde{U}_i) \right| &\leq \sum_{i=1}^{n-1} |w_{n,i}(x)| |f(Y_i^*) - f(\tilde{U}_i)| \\
&\leq \|f\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)| |Y_i^* - \tilde{U}_i| \\
&= \|f\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)| |g(x_{i+1}) - g(x_i)| \\
&< \|f\|_L \epsilon / (M \|f\|_L) \sum_{i=1}^{n-1} |w_{n,i}(x)| < \epsilon
\end{aligned}$$

para todo $n \geq n_0$. Con lo cual, $\lim_{n \rightarrow \infty} \left| \sum_{i=1}^{n-1} w_{n,i}(x) f(Y_i^*) - \sum_{i=1}^{n-1} w_{n,i}(x) f(\tilde{U}_i) \right| = 0$, de donde se deduce (4.14).

Probemos ahora (4.15). Es suficiente mostrar que para cada $f \in \mathcal{C}_L(\mathbb{R})$

$$\sum_{i=1}^{n-1} w_{n,i}(x) f(\tilde{U}_i) - E \left[\sum_{i=1}^{n-1} w_{n,i}(x) f(\tilde{U}_i) \right] \xrightarrow{c.s.} 0, \quad (4.16)$$

y

$$\lim_{n \rightarrow \infty} E \left[\sum_{i=1}^{n-1} w_{n,i}(x) f(\tilde{U}_i) \right] = \int f dG_x = E[f(\sigma(x)U_1^*)]. \quad (4.17)$$

Definamos la siguiente sucesión de variables aleatorias $Z_i = f(\tilde{U}_i) - E[f(\tilde{U}_i)]$, para $i \geq 1$.

Luego (4.16) es equivalente a probar que $S_n = \sum_{i=1}^{n-1} w_{n,i}(x) Z_i \xrightarrow{c.s.} 0$. Los términos en S_n

son uno-dependientes pero pueden ser separados en dos conjuntos de variables aleatorias

independientes. En efecto, $S_n = S_{1,n} + S_{2,n}$ donde, para $j = 1, 2$, $S_{j,n} = \sum_{i \in I_{j,n}} w_{n,i}(x) Z_i$

con $I_{1,n} = \{1 < i \leq n-1 : i \text{ es par}\}$ y $I_{2,n} = \{1 < i \leq n-1 : i \text{ es impar}\}$. Apliquemos el Lema 4.2(a) a cada término. Sea $w_n^{(j)} = \max_{i \in I_{j,n}} |w_{n,i}(x)| \leq w_n$, como $\sum_{i \in I_{j,n}} |w_{n,i}(x)| \leq \sum_{i=1}^n |w_{n,i}(x)| \leq M$, obtenemos, para $j = 1, 2$,

$$P(|S_{j,n}| > \epsilon) \leq 2 \exp\left(-\frac{\epsilon^2}{2\|f\|_\infty^2 M w_n^{(j)}}\right) \leq 2 \exp\left(-\frac{\epsilon^2}{2\|f\|_\infty^2 M w_n}\right).$$

Por lo tanto,

$$P(|S_n| > 2\epsilon) \leq P(|S_{1,n}| > \epsilon) + P(|S_{2,n}| > \epsilon) \leq 4 \exp\left(-\frac{\epsilon^2}{2\|f\|_\infty^2 M w_n}\right),$$

de donde, por la Hipótesis 11(iv) $\sum_{n=1}^{\infty} P(|S_n| > 2\epsilon) < \infty$ y por lo tanto, $S_n \xrightarrow{c.s.} 0$.

Finalmente mostremos que se cumple (4.17).

$$\begin{aligned} E\left[\sum_{i=1}^{n-1} w_{n,i}(x) f(\tilde{U}_i)\right] - E[f(\sigma(x)U_1^*)] &= \sum_{i=1}^{n-1} w_{n,i}(x) E[f(\tilde{U}_i) - f(\sigma(x)U_1^*)] \\ &+ \left(\sum_{i=1}^{n-1} w_{n,i}(x) - 1\right) E[f(\sigma(x)U_1^*)]. \end{aligned}$$

Por la Hipótesis 11(i), el segundo término de la expresión de la derecha de la igualdad converge a 0. Sólo resta entonces probar que el primero también lo hace. Para ello definamos la siguiente

función $h(y) = E\left(f(\sigma(y)U_1^*)\right)$ y las sumatorias

$$\begin{aligned} T_{1,n} &= \sum_{i=1}^{n-1} w_{n,i}(x)(h(x_i) - h(x)) \\ T_{2,n} &= \sum_{i=1}^{n-1} w_{n,i}(x)E\left[f((\sigma(x_{i+1}) - \sigma(x_i))U_2 + \sigma(x_i)U_1^*) - f(\sigma(x_i)U_1^*)\right]; \end{aligned}$$

luego $\sum_{i=1}^{n-1} w_{n,i}(x)E\left[f(\tilde{U}_i) - f(\sigma(x)U_1^*)\right] = T_{1,n} + T_{2,n}$.

Analicemos el comportamiento de $T_{1,n}$. La continuidad de la función f y de la función de escala σ (Hipótesis 10) implican que h es continua en x . Por otra parte, h es acotada pues f lo es. En consecuencia, dado $\epsilon > 0$ existe $\delta > 0$ tal que si $|x - y| < \delta$ entonces $|h(x) - h(y)| < \epsilon/M$.

Por estas consideraciones y por la Hipótesis 11(ii) tenemos

$$\begin{aligned} |T_{1,n}| &\leq \sum_{i=1}^{n-1} |w_{n,i}(x)|I_{\{|x_i - x| \geq \delta\}} |h(x) - h(x_i)| + \sum_{i=1}^{n-1} |w_{n,i}(x)|I_{\{|x_i - x| < \delta\}} |h(x) - h(x_i)| \\ &\leq 2 \|h\|_\infty \sum_{i=1}^{n-1} |w_{n,i}(x)|I_{\{|x_i - x| \geq \delta\}} + \epsilon \end{aligned}$$

De donde, por la Hipótesis 11(iii), obtenemos que $\lim_{n \rightarrow \infty} |T_{1,n}| = 0$.

Mostremos ahora que $T_{2,n} \rightarrow 0$. Sean, para $1 \leq i \leq n-1$, $u_i = (\sigma(x_{i+1}) - \sigma(x_i))U_2 + \sigma(x_i)U_1^*$ y $v_i = \sigma(x_i)U_1^*$. Sea $\epsilon > 0$ y k tal que $P(|U_2| > k) \leq \epsilon/(4 \|f\|_\infty M)$. Luego, la Hipótesis 11(ii)

y el hecho que f es Lipschitz, implican que

$$\begin{aligned}
|T_{2,n}| &\leq \sum_{i=1}^{n-1} |w_{n,i}(x)| E |f(u_i) - f(v_i)| \\
&\leq \sum_{i=1}^{n-1} |w_{n,i}(x)| E \left[|f(u_i) - f(v_i)| I_{\{|U_2| \leq k\}} \right] + \sum_{i=1}^{n-1} |w_{n,i}(x)| E \left[|f(u_i) - f(v_i)| I_{\{|U_2| > k\}} \right] \\
&\leq \|f\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)| E \left[|\sigma(x_{i+1}) - \sigma(x_i)| |U_2| I_{\{|U_2| \leq k\}} \right] \\
&\quad + 2 \|f\|_\infty \sum_{i=1}^{n-1} |w_{n,i}(x)| P(|U_2| > k) \\
&\leq \|f\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)| E \left[|\sigma(x_{i+1}) - \sigma(x_i)| |U_2| I_{\{|U_2| \leq k\}} \right] + \epsilon/2.
\end{aligned}$$

Por la continuidad de la función de escala sobre $[0, 1]$ (continuidad uniforme ya que está definida sobre un compacto), existe $\delta > 0$ tal que si $|s - t| < \delta$ entonces $|\sigma(s) - \sigma(t)| < \epsilon/(2Mk \|f\|_L)$ y por la Hipótesis 2, existe n_0 tal que $M_n = \max_{1 \leq i \leq n-1} (x_{i+1} - x_i) < \delta$, para todo $n \geq n_0$, de donde

$$\|f\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)| E \left[|\sigma(x_{i+1}) - \sigma(x_i)| |U_2| I_{\{|U_2| \leq k\}} \right] \leq \epsilon/2$$

con lo cual, $\lim_{n \rightarrow \infty} |T_{2,n}| = 0$, lo que concluye la demostración. \square

Enunciaremos y demostraremos ahora el Teorema de distribución asintótica de la sucesión de M -estimadores locales de la función de escala. Primero introduciremos las siguientes hipótesis adicionales.

Hipótesis 12 *La función de escala $\sigma : [0, 1] \rightarrow \mathbb{R}^+$ cumple (a) o (b) donde*

(a) σ es Lipschitz-continua con $\|\sigma\|_L$ como en (2.1)

(b) σ es continua y $\lim_{n \rightarrow \infty} c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| |\sigma(x_{i+1}) - \sigma(x_i)| = 0$.

Hipótesis 13 Sea $c_n = \sum_{i=1}^{n-1} w_{n,i}^2(x)$. La sucesión de pesos verifica

(i) $\lim_{n \rightarrow \infty} c_n^{-1/2} w_n = 0$ con $w_n = \max_{1 \leq i \leq n-1} |w_{n,i}(x)|$.

(ii) $\lim_{n \rightarrow \infty} c_n^{-1/2} (\sum_{i=1}^{n-1} w_{n,i}(x) - 1) = 0$.

Hipótesis 14 La función de escores χ es tal que para cualquier $\alpha \neq 0$, $\beta \neq 0$,

$$\nu(\alpha, \beta) = E [|\chi'(\alpha U_2 + \beta U_1) U_2|] < \infty$$

donde $\{U_i\}_{i=1,2}$ son v.a.'s i.i.d. con $U_1 \sim G$.

Observación 4.3 La Hipótesis 14 es válida si, por ejemplo, χ' es acotada, $\chi'(u) = 0$ para $|u| > c$ y $E [|U_2| I_{|\alpha U_2 + \beta U_1| \leq c} I_{|\beta U_1| > 2c} I_{|\alpha U_2| > c}] < \infty$. Efectivamente, las siguientes acotaciones implican que $\nu(\alpha, \beta) < \infty$

$$\begin{aligned} \nu(\alpha, \beta) &= E [|\chi'(\alpha U_2 + \beta U_1) U_2|] \leq \|\chi'\|_\infty E [|U_2| I_{|\alpha U_2 + \beta U_1| \leq c}] \\ &\leq \|\chi'\|_\infty [E [|U_2| I_{|\alpha U_2 + \beta U_1| \leq c} I_{|\beta U_1| \leq 2c}] + E [|U_2| I_{|\alpha U_2 + \beta U_1| \leq c} I_{|\beta U_1| > 2c}]] \\ &\leq \|\chi'\|_\infty \left[\frac{2c}{|\beta|} + E (|U_2| I_{|\alpha U_2 + \beta U_1| \leq c} I_{|\beta U_1| > 2c} I_{|\alpha U_2| > c}) \right]. \end{aligned}$$

Por otra parte, si χ' es continua implica que $\sup_{(\alpha, \beta) \in \mathcal{K}_1 \times \mathcal{K}_2} \nu(\alpha, \beta) < \infty$ donde \mathcal{K}_i son conjuntos compactos en $\mathbb{R} - \{0\}$.

Cabe mencionar que la derivada de cualquier miembro, χ_c , de la familia de funciones de escores de Beaton-Tukey introducida en el Ejemplo 3.4 cumple claramente las condiciones especificadas arriba (posee soporte compacto $[-c, c]$ y derivada continua).

La condición de finitud de $E [|U_2| I_{|\alpha U_2 + \beta U_1| \leq c} I_{|\beta U_1| > 2c} I_{|\alpha U_2| > c}]$ no es restrictiva como pareciera ya que, por ejemplo, se cumple aún para la distribución de Cauchy, la cual carece de primer momento finito. En efecto, asumamos que $\{U_i\}_{i=1,2}$ son v.a.'s i.i.d. con $U_1 \sim \text{Cauchy}$. Hay sólo dos casos a considerar, o bien las constantes α y β tienen el mismo signo o tienen signos opuestos. Asumamos que ambas constantes son positivas (el otro caso por simetría tiene el mismo tratamiento). Un simple análisis muestra que para la región de integración $R = \{(x, y) : |\alpha y + \beta x| \leq c, |\beta x| > 2c, |\alpha y| > c\}$ existe k_0 una constante positiva tal que $R \subseteq S_1 \cup S_2$ donde $S_1 = \{(x, y) : -a - bx \leq y \leq a + bx; x \leq -k_0\}$ y $S_2 = \{(x, y) : -a - bx \leq y \leq a + bx; x \geq k_0\}$, $a = c/\alpha$ y $b = \beta/\alpha$. Luego, si f es la densidad conjunta del vector (U_1, U_2) entonces

$$E [|U_2| I_{|\alpha U_2 + \beta U_1| \leq c} I_{|\beta U_1| > 2c} I_{|\alpha U_2| > c}] \leq \int \int_{S_1} |y| f(x, y) dy dx + \int \int_{S_2} |y| f(x, y) dy dx.$$

Sólo es necesario comprobar que una de las integrales es finita. Así,

$$\begin{aligned} \int \int_{S_1} |y| f(x, y) dy dx &= \frac{1}{\pi^2} \int_{-\infty}^{-k_0} \int_{-a-bx}^{a-bx} \frac{y}{(1+x^2)(1+y^2)} dx dy \\ &= \frac{1}{2\pi^2} \int_{-\infty}^{-k_0} \frac{1}{1+x^2} \log \left(\frac{(a-bx)^2 + 1}{(-a-bx)^2 + 1} \right) dx \\ &\leq \gamma \operatorname{arc} \operatorname{tg}(-k_0) + \pi < \infty \end{aligned}$$

ya que $\gamma = \frac{1}{2\pi^2} \|h\|_\infty$ y $h(x) = \left| \log \left(\frac{(a-bx)^2+1}{(-a-bx)^2+1} \right) \right|$ es una función acotada.

Hipótesis 15 Para cada $x \in (0, 1)$ se verifican

$$(i) \quad \lim_{n \rightarrow \infty} c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) (\sigma(x_i) - \sigma(x)) = \beta_1$$

$$(ii) \quad \lim_{n \rightarrow \infty} c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| (\sigma(x_i) - \sigma(x))^2 = 0.$$

Teorema 4.2 Sea $x \in (0, 1)$ fijo y $c_n = \sum_{i=1}^{n-1} w_{n,i}^2(x)$ como en la Hipótesis 13. Supongamos que $c_n^{-1} \sum_{i=1}^{n-2} w_{n,i+1}(x) w_{n,i}(x) \rightarrow \beta > 0$ cuando $n \rightarrow \infty$ y que $v_i > 0$, $i = 1, 2$ donde

$$v_1 = v_1(G_x) = Var \left[\chi \left(\frac{\sigma(x)U_1^*}{aS(G_x)} \right) \right] + 2\beta Cov \left[\chi \left(\frac{\sigma(x)U_1^*}{aS(G_x)} \right), \chi \left(\frac{\sigma(x)U_2^*}{aS(G_x)} \right) \right] \quad (4.18)$$

$$v_2 = v_2(G_x) = E \left[\chi' \left(\frac{\sigma(x)U_1^*}{aS(G_x)} \right) \left(\frac{\sigma(x)U_1^*}{aS(G_x)} \right) \right]. \quad (4.19)$$

Sea $v = v(G_x) = v_1/v_2^2$. Supongamos que se cumplen además las Hipótesis 1, 4, 6, 7 y 11 a 15, entonces se tiene que

$$c_n^{-1/2} (\hat{\sigma}_{M,n}(x) - S(G_x)) \xrightarrow{d} N(S(G_x)\beta_1/\sigma(x), v). \quad (4.20)$$

Observación 4.4

(a) Si bien, en la demostración de este teorema, se utilizará el Teorema 3.1, la normalidad asintótica sigue valiendo si la convergencia casi segura de los M -estimadores es reemplazada por convergencia en probabilidad.

- (b) *Es importante mencionar también que si $\beta_1 = 0$ entonces no hay sesgo asintótico. Por otra parte, el sesgo asintótico depende de χ sólo a través del funcional $S(G_x)$. Por lo tanto, en el centro del entorno (o sea, si $G = F_0$) el sesgo asintótico del estimador es independiente de la función de escores. Es decir, el comportamiento asintótico depende de χ sólo a través de la varianza.*
- (c) *La dependencia inherente al hecho de haber introducido un estimador basado en diferencias induce la aparición del segundo sumando en v_1 . Por esta razón no se obtiene la expresión usual de la varianza de los M -estimadores de escala, a diferencia de lo que ocurre en regresión no-paramétrica en donde la eficiencia asintótica es la del M -estimador de posición.*
- (d) *Dado que en los estudios de simulación Monte Carlo tomaremos medidas basadas en $\log(\hat{\sigma}_{M,n}(x)/S(G_x))$, observemos que*

$$c_n^{-1/2} \log(\hat{\sigma}_{M,n}(x)/S(G_x)) \xrightarrow{d} N(\beta_1/S(G_x), v/S^2(G_x)).$$

Además, $v/S^2(G_x)$ es análoga a la medida recomendada en Huber (1981) para medir eficiencia de estimadores de escala.

Lema 4.3 *Para cada $s > 0$ y $x \in (0, 1)$, definamos*

$$\begin{aligned}\lambda_{n,b}(s, x) &= \sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{Y_i^*}{as} \right) - b \\ \lambda_{n,b}^*(s, x) &= \sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{\sigma(x)U_i^*}{as} \right) - b \\ \lambda_{1,n}(s, x) &= \sum_{i=1}^{n-1} w_{n,i}(x) \chi' \left(\frac{\sigma(x)U_i^*}{as} \right) \left(\frac{U_i^*}{as} \right) (\sigma(x_i) - \sigma(x))\end{aligned}$$

donde Y_i^* y U_i^* , $1 \leq i \leq n-1$ son como en (4.9). Entonces, bajo las Hipótesis 1, 4, 6, 7, 11(ii), 12, 13(i), 14 y 15, se tiene

$$c_n^{-1/2} \lambda_{n,b}(s, x) = c_n^{-1/2} \lambda_{n,b}^*(s, x) + c_n^{-1/2} \lambda_{1,n}(s, x) + o_p(1) \quad (4.21)$$

$$c_n^{-1/2} \lambda_{1,n}(s, x) = \beta_1 E \left[\chi' \left(\frac{\sigma(x)U_1^*}{as} \right) \left(\frac{U_1^*}{as} \right) \right] + o_p(1). \quad (4.22)$$

Demostracion del Teorema 4.2. Fijemos $x \in (0, 1)$ y para abreviar sea $S_x = S(G_x)$.

Consideremos $\lambda_{n,b}$ como en el Lema 4.3. Por (4.4), el M -estimador local $\hat{\sigma}_{M,n}(x)$ satisface $\lambda_{n,b}(\hat{\sigma}_{M,n}(x), x) = 0$. Un desarrollo de Taylor de primer orden de $\lambda_{n,b}(s, x)$ alrededor de $s = S_x$, evaluando en $\hat{\sigma}_{M,n}(x)$, implica

$$0 = \lambda_{n,b}(\hat{\sigma}_{M,n}(x), x) = \lambda_{n,b}(S_x, x) + (\hat{\sigma}_{M,n}(x) - S_x) \lambda'_{n,b}(\hat{\sigma}_{0,n}, x),$$

donde $\lambda'_{n,b}(s, x) = \frac{\partial}{\partial s} \lambda_{n,b}(s, x) = -\frac{1}{s} \sum_{i=1}^{n-1} w_{n,i}(x) \chi' \left(\frac{Y_i^*}{as} \right) \left(\frac{Y_i^*}{as} \right) y$

$$\widehat{\sigma}_{0,n} = (1 - \theta_n)S_x + \theta_n \widehat{\sigma}_{M,n}(x), \text{ con } \theta_n \text{ algún valor en } [0, 1]. \quad (4.23)$$

Luego, $c_n^{-1/2} (\widehat{\sigma}_{M,n}(x) - S_x) = -c_n^{-1/2} \lambda_{n,b}(S_x, x) / \lambda'_{n,b}(\widehat{\sigma}_{0,n}, x)$. Por lo tanto, es suficiente probar que

$$c_n^{-1/2} \lambda_{n,b}(S_x, x) \xrightarrow{d} N(S_x v_2 \beta_1 / \sigma(x), v_1) \quad (4.24)$$

$$-\lambda'_{n,b}(\widehat{\sigma}_{0,n}, x) \xrightarrow{p} v_2. \quad (4.25)$$

De acuerdo al Lema 4.3, $c_n^{-1/2} \lambda_{n,b}(S_x, x) = c_n^{-1/2} \lambda_{n,b}^*(S_x, x) + c_n^{-1/2} \lambda_{1,n}(S_x, x) + o_p(1)$ y

$$c_n^{-1/2} \lambda_{1,n}(S_x, x) \xrightarrow{p} \beta_1 E \left[\chi' \left(\frac{\sigma(x) U_1^*}{aS_x} \right) \left(\frac{U_1^*}{aS_x} \right) \right] = S_x v_2 \beta_1 / \sigma(x) \quad (4.26)$$

con lo cual para probar (4.24) bastará con demostrar

$$c_n^{-1/2} \lambda_{n,b}^*(S_x, x) \xrightarrow{d} N(0, v_1). \quad (4.27)$$

Reescribamos entonces $c_n^{-1/2} \lambda_{n,b}^*(S_x, x)$ de la siguiente manera

$$\begin{aligned}
c_n^{-1/2} \lambda_{n,b}^*(S_x, x) &= c_n^{-1/2} \left(\sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{\sigma(x) U_i^*}{aS_x} \right) - b \right) \\
&= V_n^{1/2} \sum_{i=1}^{n-1} a_{n,i} \xi_i + b c_n^{-1/2} \left(\sum_{i=1}^{n-1} w_{n,i}(x) - 1 \right) \\
&= B_{1,n} + B_{2,n}
\end{aligned}$$

con $a_{n,i} = V_n^{-1/2} c_n^{-1/2} w_{n,i}(x)$, $\xi_i = \chi(\sigma(x) U_i^* / (aS_x)) - b$ y $V_n = Var \left[\sum_{i=1}^{n-1} c_n^{-1/2} w_{n,i}(x) \xi_i \right]$. La Hipótesis 13(ii) implica que $B_{2,n} \rightarrow 0$. Por lo tanto, demostremos que

$$V_n^{1/2} \rightarrow v_1 \quad (4.28)$$

$$\sum_{i=1}^{n-1} a_{n,i} \xi_i \xrightarrow{d} N(0, 1), \quad (4.29)$$

de donde se deducirá (4.27). La convergencia en (4.28) es consecuencia de la siguiente igualdad y del hecho que $c_n^{-1} \sum_{i=1}^{n-2} w_{n,i}(x) w_{n,i+1}(x) \rightarrow \beta$

$$\begin{aligned}
V_n &= Var \left[\sum_{i=1}^{n-1} c_n^{-1/2} w_{n,i}(x) \xi_i \right] = c_n^{-1} \sum_{i=1}^{n-1} w_{n,i}^2(x) Var \left[\chi \left(\frac{\sigma(x) U_i^*}{aS_x} \right) \right] \\
&\quad + 2 c_n^{-1} \sum_{i=1}^{n-2} w_{n,i}(x) w_{n,i+1}(x) Cov \left[\chi \left(\frac{\sigma(x) U_i^*}{aS_x} \right), \chi \left(\frac{\sigma(x) U_{i+1}^*}{aS_x} \right) \right] \\
&= Var \left[\chi \left(\frac{\sigma(x) U_1^*}{aS_x} \right) \right] + 2Cov \left[\chi \left(\frac{\sigma(x) U_1^*}{aS_x} \right), \chi \left(\frac{\sigma(x) U_2^*}{aS_x} \right) \right] c_n^{-1} \sum_{i=1}^{n-2} w_{n,i}(x) w_{n,i+1}(x).
\end{aligned}$$

Para probar (4.29) utilizaremos la Proposición 2.5 para procesos φ -mezclantes. Teniendo en cuenta que $V_n \rightarrow v_1 > 0$, sin pérdida de generalidad podemos asumir que $\inf_{n>1} V_n > 0$. Por lo

tanto,

$$\sup_{n>1} \sum_{i=1}^n a_{ni}^2 = \sup_{n>1} V_n^{-1} c_n^{-1} \sum_{i=1}^n w_{n,i}^2(x) = 1 / \inf_{n>1} V_n < \infty .$$

Por otra parte, (4.28) y la Hipótesis 13(i) implican que

$$\max_{1 \leq i \leq n} |a_{n,i}| = V_n^{-1/2} \max_{1 \leq i \leq n} c_n^{-1/2} w_{n,i}(x) \rightarrow 0.$$

Como χ es acotada, la sucesión $\{\xi_i^2\}_{i \geq 1}$ es uniformemente integrable. Es claro que $\{\xi_i\}_{i \geq 1}$, por ser una sucesión de v.a.'s uno-dependiente, es φ -mezclante y además (ver Observación 4.2(b)), satisface $E[\xi_i] = E[\chi(\sigma(x)U_i^*/(aS_x) - b)] = 0$. Finalmente, por definición de la sucesión $\{a_{ni}\}$, $Var(\sum_{i=1}^n a_{ni}\xi_i) = 1$, de modo que se cumplen las hipótesis de la Proposición 2.5 y por ende se establece la validez de (4.29). Hemos demostrado entonces (4.24).

Probemos (4.25). Definamos $\eta(t) = (t/a)\chi'(t/a)$. Por el Teorema 4.1 y por (4.23), tenemos que $\hat{\sigma}_{0,n} \xrightarrow{p} S_x$. Por lo tanto, bastará demostrar que

$$\sum_{i=1}^{n-1} w_{n,i}(x) \eta\left(\frac{Y_i^*}{\hat{\sigma}_{0,n}}\right) \xrightarrow{p} E\left[\eta\left(\frac{U_1^* \sigma(x)}{S_x}\right)\right]. \quad (4.30)$$

Definamos, para $s \in \mathbb{R}^+$, $f(s) = \sum_{i=1}^{n-1} w_{n,i}(x) \eta(Y_i^*/s)$ y realicemos un desarrollo de Taylor de primer orden de f alrededor de S_x y evaluada en $\hat{\sigma}_{0,n}$. Obtenemos de esta forma

$$\begin{aligned} \sum_{i=1}^{n-1} w_{n,i}(x) \eta\left(\frac{Y_i^*}{\hat{\sigma}_{0,n}}\right) &= \sum_{i=1}^{n-1} w_{n,i}(x) \eta\left(\frac{Y_i^*}{S_x}\right) + \frac{1}{\hat{\sigma}_{0,n}^{(1)}} \sum_{i=1}^{n-1} w_{n,i}(x) \eta'\left(\frac{Y_i^*}{\hat{\sigma}_{0,n}^{(1)}}\right) \left(\frac{Y_i^*}{\hat{\sigma}_{0,n}^{(1)}}\right) (S_x - \hat{\sigma}_{0,n}) \\ &= A_{1,n} + A_{2,n}(S_x - \hat{\sigma}_{0,n}) \end{aligned}$$

con $\hat{\sigma}_{0,n}^{(1)} = (1 - \theta_n^{(1)})S_x + \theta_n^{(1)}\hat{\sigma}_{0,n}$ un punto intermedio y $\theta_n^{(1)}$ algún valor en el intervalo $[0, 1]$.

Por (4.23) y el Teorema 4.1, $\hat{\sigma}_{0,n}^{(1)} \xrightarrow{p} S_x$. Por otra parte, por la Hipótesis 7(ii), la función $h(t) = -a\eta'(t)t = [(t/a)\chi'(t/a) + (t/a)^2\chi''(t/a)]$ es acotada con lo cual $A_{2,n} = O_p(1)$ y entonces, para probar (4.30), bastará mostrar que

$$A_{1,n} = \sum_{i=1}^{n-1} w_{n,i}(x) \eta \left(\frac{Y_i^*}{S_x} \right) \xrightarrow{c.s.} E \left[\eta \left(\frac{U_1^* \sigma(x)}{S_x} \right) \right]. \quad (4.31)$$

Por la Hipótesis 7, $\eta \in \mathcal{C}_L(\mathbb{R})$, con lo cual (4.31) puede ser demostrado con las mismas técnicas que las utilizadas para demostrar (4.13) en el Teorema 4.1. \square

Observación 4.5 *Analizaremos las tasas de convergencias $c_n^{-1/2}$ (donde $c_n = \sum_{i=1}^{n-1} w_{n,i}^2(x)$) y el factor $\beta = \lim_{n \rightarrow \infty} c_n^{-1} \sum_{i=1}^{n-2} w_{n,i+1}(x)w_{n,i}(x)$ asociado a la covarianza de la distribución asintótica para algunos tipos de pesos introducidos en el Ejemplo 4.1. La siguiente notación será utilizada: dadas dos sucesiones de números reales $\{a_n\}_{n \geq 1}$ y $\{b_n\}_{n \geq 1}$, $a_n \sim b_n$ denota que ambas sucesiones tienen el mismo orden de convergencia.*

(a) *Pesos de Rosenblatt.*

Sea $w_{n,i}(x) = (nh_n)^{-1} K \left(\frac{x - x_i}{h_n} \right)$; entonces, como $c_n = \sum_{i=1}^{n-1} (nh_n)^{-2} K^2 \left(\frac{x - x_i}{h_n} \right)$ y dado que $(nh_n)^{-1} \sum_{i=1}^{n-1} K^2 \left(\frac{x - x_i}{h_n} \right) \rightarrow \int K^2(u) du$ tenemos que $c_n \sim (nh_n)^{-1}$. Así, el

orden de convergencia es $c_n^{-1/2} \sim (nh_n)^{1/2}$. Además

$$\beta = \lim_{n \rightarrow \infty} (nh_n)^{-1} \sum_{i=1}^{n-2} K\left(\frac{x-x_{i+1}}{h_n}\right) K\left(\frac{x-x_i}{h_n}\right) = \int K^2(u) du.$$

(b) Pesos de Priestley-Chao.

Consideremos $w_{n,i}(x) = h_n^{-1} (x_{i+1} - x_i) K\left(\frac{x-x_i}{h_n}\right)$ y asumamos que los puntos del diseño satisfacen $M_n = \max_{1 \leq i \leq n-1} (x_{i+1} - x_i) \sim n^{-1}$. Luego

$$c_n = \sum_{i=1}^{n-1} h_n^{-2} (x_{i+1} - x_i)^2 K^2\left(\frac{x-x_i}{h_n}\right) \sim (nh_n)^{-1} \left[(nh_n)^{-1} \sum_{i=1}^{n-1} K^2\left(\frac{x-x_i}{h_n}\right) \right],$$

lo que implica que $c_n^{-1/2} \sim (nh_n)^{1/2}$. Un análisis directo prueba que $\beta = \int K^2(u) du$.

(c) Pesos de Nadaraya-Watson.

Sea $w_{n,i}(x) = K\left(\frac{x-x_i}{h_n}\right) / s_n$ con $s_n = \sum_{j=1}^{n-1} K\left(\frac{x-x_j}{h_n}\right)$. Teniendo en cuenta que

$$\sum_{i=1}^{n-1} (nh_n)^{-2} K^2\left(\frac{x-x_i}{h_n}\right) \sim (nh_n)^{-1} \text{ y } (nh_n)^{-1} \sum_{j=1}^{n-1} K\left(\frac{x-x_j}{h_n}\right) \rightarrow \int K(u) du$$

se deduce que $c_n^{-1/2} \sim (nh_n)^{1/2}$ y que $\beta = \int K^2(u) du / \left(\int K(u) du \right)^2$.

(d) Pesos basados en vecinos más cercanos.

Sea $w_{n,i}(x) = 1/k_n$ si $|x_i - x| \leq D^{(k_n)}$ y 0 en caso contrario. Es inmediato probar que

$$c_n = k_n^{-1}, \text{ con lo cual } c_n^{-1/2} \sim k_n^{1/2} \text{ y } \beta = \lim_{n \rightarrow \infty} k_n k_n^{-2} (k_n - 2) = 1.$$

4.3. Resultados Asintóticos para Estimadores de Núcleos con Ventana Adaptiva.

Cuando los pesos de los M -estimadores de la escala local están basados en núcleos, las propiedades de los estimadores de la escala local dependen de la elección de la abertura de ventana h , es decir de la cantidad de suavizado de la estimación. Cuando el objetivo es estimar la función de regresión hay varias propuestas, desarrolladas en la literatura, para seleccionar la abertura de ventana. Por ejemplo, los métodos de convalidación cruzada de mínimos cuadrados son introducidos y discutidos por Clark (1975), Wahba y Wold (1975) entre otros. Aquellos del tipo plug-in pueden ser estudiados en Woodroffe (1970), Müller y Stadtmüller (1987). Una exposición detallada sobre distintas alternativas puede encontrarse también en Härdle (1990, Cap. 5) y en Härdle (2004, Cap. 4). Para una discusión sobre la sensibilidad a la presencia de valores atípicos de los selectores clásicos ver Leung *et al.* (1993), Wang y Scott (1994), Boente *et al.* (1997) y Cantoni y Ronchetti (2001).

El estudio de aberturas de ventanas adaptivas para la función de escala se encuentra menos desarrollado. Podría proponerse, como se ha hecho para la función de regresión, métodos del tipo de convalidación cruzada o plug-in para la función de escala. De todos modos, el objetivo de esta sección es estudiar las propiedades asintóticas de estimadores adaptivos. Para ser más precisos, cada método produce entonces un selector de ventana, o sea, una abertura de ventana $\hat{h}_n = \hat{h}_n(Y_1, \dots, Y_n)$ que depende de los datos y que es, por lo tanto, aleatoria. Determinaremos entonces, las propiedades asintóticas de los estimadores robustos de la escala local adaptados al selector, $\hat{\sigma}_{M,n}(x, \hat{h}_n)$, bajo el supuesto de que existe una sucesión no aleatoria de aberturas

de ventana $\{h_n\}_{n \geq 1}$ tal que tal que $\widehat{h}_n/h_n \xrightarrow{p} 1$.

Consideraremos a lo largo de la subsección los M -estimadores de la escala local, $\widehat{\sigma}_{M,n}(x, h)$, definidos en (4.2) con pesos de de Rosenblatt $w_{n,i}(x, h) = (nh)^{-1}K\left(\frac{x - x_i}{h}\right)$. Introduciremos las siguientes hipótesis adicionales

Hipótesis 16 *Existe una sucesión $\{h_n\}_{n \geq 1}$ en \mathbb{R}^+ tal que*

$$(i) \widehat{h}_n/h_n \xrightarrow{p} 1$$

$$(ii) \lim_{n \rightarrow \infty} nh_n = +\infty \text{ y } \lim_{n \rightarrow \infty} h_n = 0.$$

Hipótesis 17 *El núcleo K satisface*

$$(i) K : \mathbb{R} \longrightarrow \mathbb{R} \text{ es par, acotada y tal que } \int |K(u)|du < \infty, \int K^2(u)du < \infty \text{ y}$$

$$\lim_{u \rightarrow \infty} u^2 K(u) = 0.$$

$$(ii) \int K(u)du = 1.$$

(iii) K es continuamente diferenciable con derivada K' y $K_1(u) = uK'(u)$ es tal que K_1 y

$$K_1^2 \text{ cumplen (i).}$$

Observación 4.6 *Hay que tener en cuenta que las hipótesis anteriores implican la validez de la Hipótesis 11 para la sucesión de pesos $w_{n,i}(x, h_n)$.*

Teorema 4.3 *Sea $x \in (0,1)$ fijo. Bajo las Hipótesis 1, 3, 5, 6, 10,16, 17 y si además se cumplen (a) $\widehat{h}_n/h_n \xrightarrow{c.s.} 1$ y (b) $nh_n/\log(n) \rightarrow \infty$, entonces*

$$\widehat{\sigma}_{M,n}(x, \widehat{h}_n) \xrightarrow{c.s.} S(G_x). \quad (4.32)$$

Demostración del Teorema 4.3. Para abreviar denotemos $S_x = S(G_x)$ como antes. Consideremos la siguiente notación

$$\begin{aligned} \lambda_{n,b}(x, s, h) &= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \chi\left(\frac{Y_i^*}{as}\right) - b \\ \lambda(x, s) &= E\left[\chi\left(\frac{\sigma(x)U_1^*}{as}\right)\right] - b. \end{aligned}$$

Como estamos bajo las Hipótesis del Teorema 4.1 se deduce que

$$\lambda_n(x, s, h_n) \xrightarrow{c.s.} \lambda(x, s) = E\left[\chi\left(\frac{\sigma(x)U_1^*}{as}\right)\right] - b.$$

Supongamos que hemos demostrado la validez de

$$\lambda_n(x, s, \widehat{h}_n) - \lambda_n(x, s, h_n) \xrightarrow{c.s.} 0 \quad (4.33)$$

entonces, habremos probado que

$$\lambda_n(x, s, \widehat{h}_n) \xrightarrow{c.s.} \lambda(x, s). \quad (4.34)$$

De este modo, considerando que $\lambda(x, S_x) = 0$ y como χ es estrictamente creciente en $[0, \|\chi\|_\infty)$, dado $\epsilon > 0$, se sigue que $\lambda(x, S_x - \epsilon) < 0 < \lambda(x, S_x + \epsilon)$. Luego, como $\lambda_n(x, s, \hat{h}_n) \xrightarrow{c.s.} \lambda(x, s)$, entonces para n suficientemente grande, tenemos que $\lambda_n(x, S_x - \epsilon, \hat{h}_n) < 0 < \lambda_n(x, S_x + \epsilon, \hat{h}_n)$ *c.s.* implicando que $\hat{\sigma}_{M,n}(x, \hat{h}_n) \xrightarrow{c.s.} S(G_x)$ como queríamos.

Mostremos entonces que (4.33) es válido. Definamos $Z_i = \chi\left(\frac{\sigma(x)Y_i^*}{as}\right)$ y expresemos

$$\begin{aligned} \lambda_n(x, s, \hat{h}_n) - \lambda_n(x, s, h_n) &= n^{-1} \sum_{i=1}^n [\hat{h}_n^{-1} - h_n^{-1}] K\left(\frac{x - x_i}{h_n}\right) Z_i \\ &\quad + (n\hat{h}_n)^{-1} \sum_{i=1}^n \left\{ K\left(\frac{x - x_i}{\hat{h}_n}\right) - K\left(\frac{x - x_i}{h_n}\right) \right\} Z_i \\ &= S_{1,n} + S_{2,n}. \end{aligned}$$

Demostremos que

$$S_{1,n} \xrightarrow{c.s.} 0 \tag{4.35}$$

$$S_{2,n} \xrightarrow{c.s.} 0. \tag{4.36}$$

Reescribamos $S_{1,n} = (nh_n)^{-1} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) Z_i [h_n/\hat{h}_n - 1]$. Teniendo en cuenta que para cada $i \geq 1$, $|Z_i| \leq \|\chi\|_\infty$, que

$$(nh_n)^{-1} \sum_{i=1}^n \left| K\left(\frac{x - x_i}{h_n}\right) \right| \rightarrow \int |K(u)| du,$$

y la Hipótesis (a), se deduce (4.35).

Analicemos la convergencia de $\{S_{2,n}\}_{n \geq 1}$. Escribamos

$$\begin{aligned} S_{2,n} &= (n\widehat{h}_n)^{-1} \sum_{i=1}^n K' \left(\frac{x - x_i}{\xi_n} \right) \left[\left(\frac{x - x_i}{\xi_n^2} \right) \right] [h_n - \widehat{h}_n] Z_i \\ &= (nh_n)^{-1} \sum_{i=1}^n K_1 \left(\frac{x - x_i}{\xi_n} \right) \left[\frac{h_n}{\xi_n} \right] \left[\frac{h_n}{\widehat{h}_n} - 1 \right] Z_i \end{aligned}$$

donde $K_1(u) = uK'(u)$ y ξ_n es un punto intermedio entre el $\min(h_n, \widehat{h}_n)$ y el $\max(h_n, \widehat{h}_n)$.

Como $\widehat{h}_n/h_n \xrightarrow{c.s.} 1$, existe entonces un conjunto nulo N tal que, para cada $\omega \notin N$, se verifica $\frac{1}{2} < \widehat{h}_n/h_n < 2$, lo que implica que $\xi_n \in [h_m, h_M]$ con $h_m = \frac{1}{2}h_n$ y $h_M = 2h_n$. En el desarrollo siguiente nos restringimos a los puntos $\omega \notin N$.

$$\begin{aligned} |S_{2,n}| &= \left| (nh_n)^{-1} \sum_{i=1}^n K_1 \left(\frac{x - x_i}{\xi_n} \right) \left[\frac{h_n}{\xi_n} \right] \left[\frac{h_n}{\widehat{h}_n} - 1 \right] Z_i \right| \\ &\leq 2 \left| \frac{h_n}{\widehat{h}_n} - 1 \right| \|\chi\|_\infty (nh_n)^{-1} \sum_{i=1}^n \left| K_1 \left(\frac{x - x_i}{\xi_n} \right) \right| \end{aligned}$$

Bastaría entonces comprobar que, si $A_n = (nh_n)^{-1} \sum_{i=1}^n \left| K_1 \left(\frac{x - x_i}{\xi_n} \right) \right|$ entonces

$$\limsup |A_n| < \infty. \quad (4.37)$$

Reescribamos $A_n = \left(\frac{\xi_n}{h_n} \right) C_n$ con $C_n = \frac{1}{n\xi_n} \sum_{i=1}^n \left| K_1 \left(\frac{x - x_i}{\xi_n} \right) \right| = \left(\frac{\xi_n}{h_n} \right) C_n$ y, fijemos $\omega \notin N$.

Luego, como $\xi_n \in [h_m, h_M]$, entonces $\left(\frac{\xi_n}{h_n} \right) \in [1/2, 2]$ y, por la Hipótesis 16(ii), $\xi_n \rightarrow 0$ y

$n\xi_n \rightarrow \infty$, de donde $C_n \rightarrow \int |K(u)| du$. De este modo, hemos probado (4.37) finalizando la

demostración del Teorema. \square

A continuación vamos a hallar la distribución asintótica para el estimador basado en el selector de ventana. Más específicamente, demostraremos que el estimador local de la escala construido en base a $\{\widehat{h}_n\}_{n \geq 1}$ tiene la misma distribución asintótica que el basado en la sucesión de aberturas de ventana no aleatorias, $\{h_n\}_{n \geq 1}$. Antes haremos algunas consideraciones.

En primer lugar, tengamos en cuenta que si se cumplen las Hipótesis 1, 4, 6, 7, 12, 14, 15, 16(ii) y 17, entonces

$$(nh_n)^{1/2} [\widehat{\sigma}_{M,n}(x, h_n) - S(G_x)] \xrightarrow{d} N\left(\frac{S(G_x)}{\sigma(x)}\beta_1, v\right) \quad (4.38)$$

donde

$$\begin{aligned} v &= v_1/v_2^2 \\ v_1 &= Var\left[\chi\left(\frac{\sigma(x)U_1^*}{aS(G_x)}\right)\right] + 2\beta Cov\left[\chi\left(\frac{\sigma(x)U_1^*}{aS(G_x)}\right), \chi\left(\frac{\sigma(x)U_2^*}{aS(G_x)}\right)\right] \\ v_2 &= E\left[\chi'\left(\frac{\sigma(x)U_1^*}{aS(G_x)}\right)\left(\frac{\sigma(x)U_1^*}{aS(G_x)}\right)\right] \\ \beta &= \int K^2(u)du \text{ y } \beta_1 = \lim_{n \rightarrow \infty} \left\{ (nh_n)^{1/2} \frac{1}{nh_n} \sum_{i=1}^{n-1} K\left(\frac{x-x_i}{h_n}\right) (\sigma(x_i) - \sigma(x)) \right\}. \end{aligned}$$

En segundo lugar, observemos que si existe γ finito tal que $nh_n^5 \rightarrow \gamma^2$, $\sigma(x)$ tiene segunda derivada continua y $\int u^2 K(u)du < \infty$, entonces $\beta_1 = \gamma\sigma''(x) \int u^2 K(u)du$. Si, en cambio, $\int u^2 K(u) < \infty$, $nh_n^3 \rightarrow \gamma^2$ y la función de escala sólo tiene primera derivada continua, entonces $\beta_1 = 0$.

Teorema 4.4 Sea $x \in (0, 1)$ fijo. Bajo las las Hipótesis 1, 4, 6, 7, 12 y 14 a 17

$$(nh_n)^{1/2} \left[\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S(G_x) \right] \xrightarrow{d} N \left(\frac{S(G_x)}{\sigma(x)} \beta_1, v \right) \quad (4.39)$$

donde la media y la varianza asintótica son como en (4.38).

Demostración del Teorema 4.4. Definamos como antes $S_x = S(G_x)$ y

$$\begin{aligned} \lambda_{n,b}(x, s, h) &= (nh)^{-1} \sum_{i=1}^n K \left(\frac{x - x_i}{h} \right) \chi \left(\frac{Y_i^*}{as} \right) - b \\ S_n(h) &= (nh_n)^{1/2} \lambda_{n,b}(x, S_x, h) \\ \widetilde{\lambda}_{1n}(x, s, h) &= (nh)^{-1} S_x^{-1} \sum_{i=1}^n K \left(\frac{x - x_i}{h} \right) \chi_1 \left(\frac{Y_i^*}{as} \right) \end{aligned}$$

con $\chi_1(u) = u\chi'(u)$ como en la Hipótesis 7. Sea también $\chi_3(u) = \chi_1(u) + \chi_2(u)$, con $\chi_2(u) = u^2\chi''(u)$ (como en la Hipótesis 7).

Usando un desarrollo de Taylor de orden dos, obtenemos la siguiente igualdad

$$0 = (nh_n)^{1/2} \lambda_{n,b}(x, \widehat{\sigma}_{M,n}(x, \widehat{h}_n), \widehat{h}_n) = S_n(\widehat{h}_n) - (nh_n)^{1/2} \left(\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S_x \right) A_n(\widehat{h}_n)$$

donde

$$\begin{aligned} A_n(\widehat{h}_n) &= \widetilde{\lambda}_{1n}(x, S_x, \widehat{h}_n) - \left(\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S_x \right) B_n(\widehat{h}_n) \\ B_n(\widehat{h}_n) &= \xi_n^{-2} (n\widehat{h}_n)^{-1} \sum_{i=1}^n K \left(\frac{x - x_i}{\widehat{h}_n} \right) \chi_3 \left(\frac{Y_i^*}{a\xi_n} \right), \end{aligned}$$

y $\xi_n = \xi_n(x, \hat{h}_n)$ es un punto intermedio entre $\hat{\sigma}_{M,n}(x, \hat{h}_n)$ y S_x . Podemos entonces escribir

$$(nh_n)^{1/2} \left(\hat{\sigma}_{M,n}(x, \hat{h}_n) - S_x \right) = S_n(\hat{h}_n)/A_n(\hat{h}_n).$$

Por (4.24), $S_n(h_n) \xrightarrow{d} N(S_x v_2 \beta_1 / \sigma(x), v_1)$. Por ende, basta con demostrar que

$$A_n(\hat{h}_n) \xrightarrow{p} v_2 \tag{4.40}$$

$$S_n(\hat{h}_n) - S_n(h_n) \xrightarrow{p} 0. \tag{4.41}$$

Como $\tilde{\lambda}_{1n}(x, S_x, h_n) \xrightarrow{p} v_2$ (ver (4.25)) y $\left(\hat{\sigma}_{M,n}(x, \hat{h}_n) - S_x \right) \xrightarrow{p} 0$, entonces (4.40) será consecuencia de

$$\tilde{\lambda}_{1n}(x, S_x, h_n) - \tilde{\lambda}_{1n}(x, S_x, \hat{h}_n) \xrightarrow{p} 0 \tag{4.42}$$

$$B_n(\hat{h}_n) = O_p(1). \tag{4.43}$$

La demostración de (4.42) sigue un esquema análogo a la de (4.33). Tengamos en cuenta ahora

que $\xi_n = \xi_n(x, \hat{h}_n) \xrightarrow{p} S_x$ y,

$$\left| (n\hat{h}_n)^{-1} \sum_{i=1}^n K \left(\frac{x - x_i}{\hat{h}_n} \right) \chi_3 \left(\frac{Y_i^*}{a\xi_n} \right) \right| \leq \|\chi_3\|_\infty (n\hat{h}_n)^{-1} \sum_{i=1}^n \left| K \left(\frac{x - x_i}{\hat{h}_n} \right) \right|. \tag{4.44}$$

Nuevamente, de manera análoga a como se procedió para probar (4.33) se demuestra que

$$(n\widehat{h}_n)^{-1} \sum_{i=1}^n \left| K \left(\frac{x - x_i}{\widehat{h}_n} \right) \right| - (nh_n)^{-1} \sum_{i=1}^n \left| K \left(\frac{x - x_i}{h_n} \right) \right| \xrightarrow{p} 0.$$

Estas afirmaciones, junto con el hecho de que $(nh_n)^{-1} \sum_{i=1}^n \left| K \left(\frac{x - x_i}{h_n} \right) \right| \rightarrow \int |K(u)| du$ implican (4.43).

Demostremos (4.41). Como $\widehat{\tau}_n = \widehat{h}_n/h_n \xrightarrow{p} 1$, entonces $P(\widehat{\tau}_n \in [r, s]) \rightarrow 1$, para r y s constantes tales que $0 < r < 1 < s$. Definamos el siguiente proceso estocástico

$$V_n(\tau) = (nh_n)^{1/2} \lambda_{n,b}(x, S_x, \tau h_n) \quad (4.45)$$

con $\tau \in [r, s]$. Observemos que $V_n(\tau) = S_n(\widehat{h}_n)$ y $V_n(1) = S_n(h_n)$.

Supongamos que es válido

$$V_n \xrightarrow{d} V, \quad (4.46)$$

donde V es un elemento en $C[r, s]$ (ver el Capítulo 2).

Sean $\epsilon, \eta > 0$ arbitrarios. Como $\widehat{\tau}_n \xrightarrow{p} 1$, existe $\delta > 0$ y $n_0 \in \mathbb{N}$ tal que

$$P(|\widehat{\tau}_n - 1| > \delta) < \eta/2, \quad \forall n \geq n_0.$$

Por otro lado, por (4.46) y la Proposición (2.6), existe $n_1 \in \mathbb{N}$ tal que

$$P \left(\sup_{|\tau-1|<\delta} |V_n(\tau) - V_n(1)| > \epsilon \right) \leq \eta/2, \quad \forall n \geq n_1.$$

Estas desigualdades implican que $\forall n \geq \max(n_0, n_1)$

$$P(|V_n(\widehat{\tau}_n) - V_n(1)| > \epsilon) \leq P(|\widehat{\tau}_n - 1| > \delta) + P \left(\sup_{|\tau-1|<\delta} |V_n(\tau) - V_n(1)| > \epsilon \right) \leq \eta,$$

de donde se obtiene (4.41).

Probemos entonces (4.46). Definamos el siguiente proceso

$$U_n(\tau) = (nh_n)^{1/2} \{ \lambda_{n,b}(x, S_x, \tau h_n) - E[\lambda_{n,b}(x, S_x, \tau h_n)] \} \quad (4.47)$$

con lo cual

$$V_n(\tau) = U_n(\tau) + (nh_n)^{1/2} E[\lambda_{n,b}(x, S_x, \tau h_n)] \quad (4.48)$$

$$= U_n(\tau) + \gamma_n(\tau). \quad (4.49)$$

Es inmediato comprobar, utilizando las mismas técnicas del Lema 4.3, que

$$\sup_{\tau \in [r,s]} \left| (nh_n)^{1/2} E[\lambda_{n,b}(x, S_x, \tau h_n)] - \beta_1 \frac{S_x}{\sigma(x)} \frac{v_2}{\tau^{1/2}} \right| \rightarrow 0,$$

es decir que $\gamma_n(\tau) \rightarrow \gamma(\tau) = \beta_1 \frac{S_x}{\sigma(x)} \frac{v_2}{\tau^{1/2}}$ uniformemente en el intervalo $[r, s]$. Luego, para probar (4.46) basta ver que el proceso

$$U_n \xrightarrow{d} U \quad (4.50)$$

donde U es un proceso estocástico gaussiano en $C[r, s]$. Combinando las Proposiciones 2.7 y 2.8 y teniendo en cuenta la Observación 2.2 debemos probar que

- (i) Dados $\tau_1, \dots, \tau_k \in [r, s]$ arbitrarios, entonces $(U_{\tau_1}, \dots, U_{\tau_k})$ converge a una distribución normal multivariada $N(0, \Sigma)$.
- (ii) La sucesión $\{U_n(r)\}_{n \geq 1}$ es ajustada.
- (iii) Existen constantes $\gamma \geq 0$, $\alpha > 1$ y una función continua F no decreciente en $[r, s]$ tal que $E[|U_n(\tau_2) - U_n(\tau_1)|^\gamma] \leq |F(\tau_2) - F(\tau_1)|^\alpha$, para $0 < r < \tau_1 < \tau_2 < s < 1$.

Como es usual, para demostrar (i), consideremos $\mathbf{a} = (a_1, \dots, a_k) \in \mathbb{R}^k$ un vector arbitrario y demostremos que $W_n = \sum_{j=1}^k a_j U_n(\tau_j)$ converge en distribución a una normal. Para ello escribamos

$$U_n(\tau) = (nh_n)^{1/2} \frac{1}{n\tau h_n} \sum_{i=1}^n K\left(\frac{x - x_i}{\tau h_n}\right) Z_i$$

con $Z_i = \chi\left(\frac{Y_i^*}{aS_x}\right) - E\left[\chi\left(\frac{Y_i^*}{aS_x}\right)\right]$. La convergencia a la distribución normal de la sucesión

$\{W_n\}_{n \geq 1}$ es inmediata usando el Teorema 4.2 y teniendo en cuenta que

$$W_n = (nh_n)^{-1/2} \sum_{i=1}^n K^* \left(\frac{x - x_i}{h_n} \right) Z_i$$

con $K^*(u) = \sum_{j=1}^k \frac{a_j}{\tau_j} K(u/\tau_j)$ un nuevo núcleo.

La demostración de (ii) es trivial pues $U_n(r)$ converge en distribución.

Probemos ahora (iii). Será suficiente demostrar que $E[U_n(\tau_2) - U_n(\tau_1)]^2 \leq c(\tau_2 - \tau_1)^2$ con c una constante positiva. Como χ es acotada, existe una constante k_1 tal que $Var[Z_i] \leq k_1$ y $Cov[Z_i, Z_{i+1}] \leq k_1$ para todo $i \geq 1$. Luego

$$\begin{aligned} E[U_n(\tau_2) - U_n(\tau_1)]^2 &= Var[U_n(\tau_2) - U_n(\tau_1)] \\ &\leq H_{1,n} + H_{2,n} \end{aligned}$$

donde

$$\begin{aligned} H_{1,n} &= k_1 \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{1}{\tau_2} K \left(\frac{x - x_i}{\tau_2 h_n} \right) - \frac{1}{\tau_1} K \left(\frac{x - x_i}{\tau_1 h_n} \right) \right)^2 \\ H_{2,n} &= 2k_1 \frac{1}{nh_n} \sum_{i=1}^{n-1} \left(\frac{1}{\tau_2} K \left(\frac{x - x_i}{\tau_2 h_n} \right) - \frac{1}{\tau_1} K \left(\frac{x - x_i}{\tau_1 h_n} \right) \right) \times \\ &\quad \times \left(\frac{1}{\tau_2} K \left(\frac{x - x_{i+1}}{\tau_2 h_n} \right) - \frac{1}{\tau_1} K \left(\frac{x - x_{i+1}}{\tau_1 h_n} \right) \right). \end{aligned}$$

Acotemos $H_{1,n}$. Teniendo en cuenta que K es Lipschitz continua

$$\begin{aligned} H_{1,n} &\leq 2k_1 \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{1}{\tau_2} K \left(\frac{x-x_i}{\tau_2 h_n} \right) - \frac{1}{\tau_2} K \left(\frac{x-x_i}{\tau_1 h_n} \right) \right)^2 \\ &\quad + 2k_1 \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{1}{\tau_2} K \left(\frac{x-x_i}{\tau_2 h_n} \right) - \frac{1}{\tau_1} K \left(\frac{x-x_i}{\tau_1 h_n} \right) \right)^2 \\ &\leq T_{1,n} + T_{2,n}, \end{aligned}$$

siendo

$$\begin{aligned} T_{1,n} &= 2k_1 \frac{1}{\tau_2^2} \left(\frac{1}{\tau_2} - \frac{1}{\tau_1} \right)^2 \frac{1}{nh_n} \sum_{i=1}^n \left(K' \left(\frac{x-x_i}{\xi_i h_n} \right) \right)^2 \left(\frac{x-x_i}{h_n} \right)^2 \\ T_{2,n} &= 2k_1 \tau_1 \left(\frac{1}{\tau_2} - \frac{1}{\tau_1} \right)^2 \frac{1}{n\tau_1 h_n} \sum_{i=1}^n K^2 \left(\frac{x-x_i}{\tau_1 h_n} \right) \end{aligned}$$

con $\xi_i \in (\tau_1, \tau_2)$. Como $\tau_1, \tau_2 \in [r, s]$ y por las hipótesis sobre los puntos del diseño, como en el Teorema 3.1 de Boente *et al.* (1997) es fácil ver que, para todo $n \geq 1$

$$\frac{1}{nh_n} \sum_{i=1}^n \left(K' \left(\frac{x-x_i}{\xi_i h_n} \right) \right)^2 \left(\frac{x-x_i}{h_n} \right)^2 \leq B$$

con B una constante. Teniendo en cuenta también que, para todo $n \geq 1$,

$$\frac{1}{n\tau_1 h_n} \sum_{i=1}^n K^2 \left(\frac{x-x_i}{\tau_1 h_n} \right) \leq C$$

con C una constante, entonces

$$T_{1,n} \leq 2k_1 \frac{s^2}{r^4} B (\tau_2 - \tau_1)^2$$

$$T_{2,n} \leq 2k_1 \frac{s}{r^2} C (\tau_2 - \tau_1)^2$$

Así, $H_{1,n} \leq c_1 (\tau_2 - \tau_1)^2$ con $c_1 = 2k_1 \left(\frac{s^2}{r^4} B + \frac{s}{r^2} C \right)$. En forma análoga se prueba que existe una constante c_2 tal que $H_{2,n} \leq c_2 (\tau_2 - \tau_1)^2$. Luego (iii) se cumple con $\gamma = \alpha = 2$, $F(u) = c^{1/2}u$ y $c = c_1 + c_2$. \square

4.4. Resultados Monte Carlo

4.4.1. Problema y objetivos

I. Modelos y Estimadores

Consideremos el siguiente modelo

$$Y_i = g(x_i) + \sigma(x_i)U_i, \quad i = 1, \dots, n$$

introducido en la Definición 4.1, con puntos del diseño equi-espaciados, es decir $x_i = i/n + 1$, $i = 1, \dots, n$. Vamos a considerar tres *modelos heteroscedásticos* y un *modelo homoscedástico o de escala global* (M.E.G.) con las siguientes funciones componentes de regresión, $g(x)$, y de

escala, $\sigma(x)$

$g(x) = 2\text{sen}(4\pi x)$	$\sigma(x) = \exp(x)$	Modelo $M1$
$g(x) = 1 + x$	$\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$	Modelo $M2$
$g(x) = 1 + x$	$\sigma(x) = (1 + x)^2$	Modelo $M3$
$g(x) = 1 + x$	$\sigma(x) = 1$	M.E.G.

Para la distribución G , de los errores U_i , consideraremos como modelo central (F_0) a la normal estándar y dos tipos de contaminación (H)

$G = \epsilon N(0, 1) + (1 - \epsilon)\text{Cauchy}$	Contaminación Simétrica.
$G = \epsilon N(0, 1) + (1 - \epsilon)N(10, 1)$	Contaminación Asimétrica.

Para contaminaciones simétricas simularemos, para los cuatro modelos presentados, para tamaños de muestra $n = 100$ y $n = 200$, y proporciones de contaminación $\epsilon = 0, 0.10, 0.20$ y 0.30 . Para contaminaciones asimétricas sólo consideraremos los tres modelos heteroscedásticos con tamaños de muestra $n = 100$ (excluimos el tamaño $n = 200$ y el modelo homoscedástico ya que las conclusiones no varían).

Cuando nos referimos a modelos hablamos (sin hacerlo explícito) tanto de sus funciones de regresión y de escala como de la distribución asociada a los errores y el tamaño de la muestra.

Para estimar la escala $\sigma(x)$ en un punto $x \in (0, 1)$, compararemos tres estimadores basados

en la familia de pesos de Nadaraya-Watson con núcleo gaussiano, es decir,

$$w_{n,i}(x) = K\left(\frac{x_i - x}{h}\right) / \sum_{j=1}^{n-1} K\left(\frac{x_j - x}{h}\right), \quad i = 1, \dots, n$$

y $K(t) = (2\pi)^{-1/2} \exp(-\frac{1}{2}t^2)$.

En la estimación se compararán tres aberturas de ventana: $h = 0.04, 0.2$ y 0.5 , para todos los modelos excepto para el modelo homoscedástico para el cual se compararán las aberturas $h = 0.2$ y 0.5 ya que para este modelo sobre-suavizar provee mejores estimadores.

Consideremos la ecuación

$$\sum_{i=1}^{n-1} w_{n,i}(x) \chi\left(\frac{Y_{i+1} - Y_i}{as}\right) = b, \quad (4.51)$$

donde $a \in (0, \infty)$ y $b \in (0, 1)$ satisfacen las condiciones especificadas en (3.5) de tal modo que los funcionales asociados resulten consistentes Fisher en el modelo central.

Si hacemos $\chi(x) = x^2$, $a = \sqrt{2}$ y $b = 1$, la solución de (4.51) es

$$\hat{\sigma}_{R,n}(x) = \left[\sum_{i=1}^{n-1} w_{n,i}(x) \left(\frac{Y_{i+1} - Y_i}{\sqrt{2}} \right)^2 \right]^{1/2}, \quad (4.52)$$

al que llamaremos “*estimador de Rice de la escala local*”. Si utilizamos en (4.51) la familia χ_c de función de escores de Beaton-Tukey (BT) introducida en (3.11) con $c = 0.70417$, $a = \sqrt{2}$ y $b = 3/4$ obtenemos el “*M-estimador con función de BT de la escala local*”, al que denotaremos con $\hat{\sigma}_{BT,n}(x)$. Finalmente introducimos el “*estimador de Boente, Fraiman y Meloche de la escala*

local", $\hat{\sigma}_{\text{BFM},n}(x)$, definido como

$$\hat{\sigma}_{\text{BFM},n}(x) = \inf \left\{ \sigma > 0 : \sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{Y_{i+1} - Y_i}{a\sigma} \right) \leq b \right\}$$

con función de escores $\chi(y) = I_{\{u: |u| > \Phi^{-1}(3/4)\}}(y)$ y constantes de ajuste $a = \sqrt{2}$ y $b = 1/2$.

II. Esquema y objetivos de la simulación

A continuación, introducimos algunas medidas clásicas para evaluar el comportamiento de los estimadores. Para cada modelo y tamaño de muestra particulares haremos $NR = 1000$ réplicas independientes. Dada una réplica j (fija) y sus correspondientes variables respuestas $\{Y_i^{(j)}\}_{i=1}^n$, una medida de evaluación del desempeño de un estimador, $\hat{\sigma}$, es el estimador del Error Cuadrático Integrado (E.C.I.) definido como

$$\widehat{\text{E.C.I.}}_j(h) = \frac{1}{n} \sum_{i=1}^n (\sigma(x_i) - \hat{\sigma}^j(x_i))^2,$$

donde $\hat{\sigma}^j$ indica que el estimador está computado en base a la j -ésima muestra. Para medir cercanía en escala una medida más útil es el estimador del Error Cuadrático Integrado en escala logarítmica (E.C.I.L.)

$$\widehat{\text{E.C.I.L.}}_j(h) = \frac{1}{n} \sum_{i=1}^n \left[\log \left(\frac{\hat{\sigma}^j(x_i)}{\sigma(x_i)} \right) \right]^2.$$

Otra medida clásica de evaluación para un estimador $\hat{\sigma}$, en un punto x fijo y abertura de

ventana h , es el estimador del Error Cuadrático Medio (E.C.M.) definido como

$$\widehat{\text{E.C.M.}}(x, h) = \frac{1}{NR} \sum_{j=1}^n (\widehat{\sigma}^j(x) - \sigma(x))^2,$$

o, en escala log,

$$\widehat{\text{E.C.M.L.}}(x, h) = \frac{1}{NR} \sum_{j=1}^n \left[\log \left(\frac{\widehat{\sigma}^j(x)}{\sigma(x)} \right) \right]^2.$$

De ahora en más, por simplicidad, omitiremos el símbolo “ $\widehat{}$ ” en E.C.I.L. y en E.C.M.L. y también el término “estimador de”.

Es importante observar la siguiente relación entre los errores

$$\frac{1}{NR} \sum_{j=1}^{NR} \text{E.C.I.L.}_j(h) = \frac{1}{n} \sum_{i=1}^n \text{E.C.M.L.}(x_i, h)$$

y análogamente para el E.C.I. y el E.C.M.

Los cuadros y gráficos se encuentran en el Apéndice B. Los Cuadros B.5 a B.18 resumen los resultados de la simulación. Para cada uno de los modelos en estudio y para cada valor de abertura de ventana, los cuadros reportan la media y la mediana (esta última entre paréntesis) de los valores de $\{\text{E.C.I.L.}_j(h)\}_{j=1}^{NR}$ correspondientes a cada uno de los tres estimadores, para cada tamaño de muestra ($n = 100, 200$ para contaminaciones simétricas y $n = 100$ para asimétricas) y para cada proporción de contaminación.

Las Figuras B.7 a B.32 fueron ordenadas en tres secuencias o grupos. En la primera se-

cuencia, Figuras B.7 a B.10, para cada modelo se muestran los gráficos de los valores de las variables respuestas versus los puntos del diseño según nivel de contaminación simétrica, para una muestra dada. En un segundo grupo, Figuras B.11 a B.30, se exhiben las densidades (estimadas) del E.C.I.L.(h) para tamaños de muestra $n = 100$, según nivel de contaminación y ordenas por abertura de ventana (las correspondientes a $n = 200$, como el comportamiento es muy similar, se omitieron). Finalmente, las Figuras B.31 a B.32, comparan para $\epsilon = 0$ y 0.20 , respectivamente, el comportamiento del E.C.M.L. de los tres estimadores en el modelo $M2$ según abertura de ventana (el tamaño de muestra es $n = 100$).

La simulación tiene como objetivos los siguientes:

- a) Comparar el comportamiento de los tres estimadores en los diferentes modelos utilizando el E.C.I.L. para contaminaciones simétricas y asimétricas.
- b) Analizar la influencia de la abertura de ventana sobre el comportamiento de los estimadores, para ambos tipos de contaminación.
- c) Contrastar los resultados de la estimación en modelos heteroscedásticos frente a modelos homoscedásticos.

4.4.2. Análisis y Conclusiones.

A lo largo de las siguientes consideraciones cuando cuando hablamos de mejor (peor) desempeño de un estimador estamos refiriéndonos a su menor (mayor) E.C.I.L.. Análogamente cuando el criterio es el E.C.M.L.

En ausencia de contaminación, el estimador de Rice, $\hat{\sigma}_{R,n}$, tiene el mejor desempeño en los tres modelos $M1$, $M2$ y $M3$ cuando el valor de la abertura de ventana es el más pequeño ($h = 0.04$); para valores más grandes de h el estimador se deteriora.

En relación a los restantes estimadores hay que mencionar que, si $\epsilon = 0$, actúa mejor $\hat{\sigma}_{BFM,n}$ que $\hat{\sigma}_{BT,n}$ en los tres modelos, mostrando un comportamiento similar al que tenía en el modelo homoscedástico. Los menores valores medios de E.C.I.L. de estos estimadores se alcanzan al interior de cada modelo $M1$ y $M3$ para un valor intermedio de suavizado ($h = 0.2$) mientras que para $M2$ en el menor valor de ventana ($h = 0.04$).

Cuando se introduce contaminación, ya sea asimétrica o simétrica, $\hat{\sigma}_{R,n}$ se deteriora rápidamente a medida que aumenta ϵ , observándose tanto en los cuadros como en los gráficos de las densidades estimadas (aumentando la concentración de masa en la cola derecha de la distribución) del E.C.I.L. Los estimadores robustos $\hat{\sigma}_{BT,n}$ y $\hat{\sigma}_{BFM,n}$ tienen un mejor desempeño que $\hat{\sigma}_{R,n}$.

Analicemos primero el esquema con contaminación simétrica. En los modelos heteroscedásticos $M1$ y $M3$, para ambos estimadores, los menores valores del E.C.I.L. se alcanzan en el valor de ventana $h = 0.2$, mientras que en el $M2$ en $h = 0.04$. En todos los modelos tiene un mejor desempeño $\hat{\sigma}_{BFM,n}$ que $\hat{\sigma}_{BT,n}$. Esta descripción también se refleja en los gráficos de las densidades.

Es importante mencionar aquí la situación diferente según la abertura de ventana que se presenta para los modelos $M1$ y $M3$ en contraposición a $M2$. En primer lugar hay que

hacer notar que en este último modelo se encuentran, en general, los mayores valores del E.C.I.L.. Esto se explica a partir de la oscilación (presencia de mínimos y máximos locales) de la función de escala, lo que produce diferencias sucesivas en las respuestas mayores que en los otros modelos y, por lo tanto, se requieren valores menores de abertura de ventana.

En las Figuras B.31 y B.32 se muestran los gráficos de la función de escala del modelo $M2$ y los valores de E.C.M.L. para los tres estimadores versus los puntos del diseño x_i . Cuando $\epsilon = 0$ y la abertura de ventana es la menor ($h = 0.04$), $\hat{\sigma}_{R,n}$ es superior a los estimadores robustos y, entre éstos, $\hat{\sigma}_{\text{BFM},n}$ es mejor que $\hat{\sigma}_{\text{BT},n}$. Cuando el tamaño de la abertura aumenta el comportamiento de $\hat{\sigma}_{R,n}$ empeora. Cuando la proporción de contaminación es alta, $\epsilon = 0.20$, los estimadores robustos son claramente superiores al estimador de Rice para todas las aberturas de ventana y, para $h = 0.04$, $\hat{\sigma}_{\text{BFM},n}$ tiene un mejor desempeño que $\hat{\sigma}_{\text{BT},n}$ (confirmando el análisis realizado desde el punto de vista del E.C.I.L.). Observar que las zonas de los puntos del diseño que aportan mayores valores al E.C.M.L. son los entornos de los puntos críticos de la función de escala.

En relación al modelo homoscedástico (M.E.G.) el estimador $\hat{\sigma}_{R,n}$ es el mejor cuando $\epsilon = 0$ pero se quiebra en presencia de contaminación. Cuando los datos están contaminados, los estimadores robustos tienen mejor desempeño pero no aparecen diferencias sustanciales cuando variamos la abertura de ventana.

Cuando la contaminación es asimétrica, para el modelo $M2$ y $M3$ los menores valores del E.C.I.L. se alcanzan para el estimador $\hat{\sigma}_{\text{BFM},n}$ en las aberturas de ventana $h = 0.04$ y $h = 0.2$, respectivamente. Para el modelo $M1$, si $\epsilon = 0.10$, los menores valores del error se alcanzan

para $\hat{\sigma}_{\text{BT},n}$ en la abertura $h = 0.5$.

A partir de las consideraciones anteriores podemos establecer algunas conclusiones, las cuales no pretenden ser generalizaciones absolutas (tengamos en cuenta que están basadas sobre los modelos considerados) sino sólo sugerencias a tener en cuenta.

- (C1) El estimador de Rice de la escala local, $\hat{\sigma}_{\text{R},n}$, es claramente no robusto en presencia de contaminación, sea simétrica o asimétrica.
- (C2) Los estimadores robustos, en presencia de heteroscedasticidad son sensibles a la elección de la abertura de ventana. Cuando los datos sugieren que la función de escala presenta oscilaciones es mejor elegir una abertura de ventana pequeña mientras que si sugieren que la función de escala es monótona, es preferible un valor de suavizado intermedio. El estimador $\hat{\sigma}_{\text{BFM},n}$ es más robusto que $\hat{\sigma}_{\text{BT},n}$ a la presencia de ambos tipos de contaminación en casi todos los modelos (excepto en el modelo $M1$ con contaminación asimétrica).
- (C3) Cuando la escala es global no hay influencias sustanciales de la abertura de ventana sobre los estimadores. Por lo tanto, como era de esperar, serán preferibles los M-estimadores de escala global. Sin embargo, los estimadores de la escala local pueden ser utilizados para contrastar las hipótesis de homoscedasticidad.

Capítulo 5

Conclusiones

En esta sección, presentaremos una síntesis de los principales resultados de la tesis. La estimación robusta de la función de escala en modelos de regresión no-paramétricos tiene importancia en problemas donde, o bien en sí misma es un elemento intrínseco al fenómeno, o bien es necesaria para contar con estimadores robustos de la función de regresión.

Cuando la función de escala es constante, para modelos no-paramétricos con diseño fijo, propusimos una familia de estimadores robustos, los M -estimadores de escala basados en las diferencias sucesivas de las variables respuesta del modelo. Los miembros de esta familia son una expresión robusta del clásico estimador (no robusto) de Rice (1984) y una generalización de los estimadores de Boente-Fraiman-Meloche (Boente *et al.*, 1997). Para la familia de los M -estimadores:

- hemos probado consistencia y normalidad asintótica, usando herramientas de convergencia de funcionales sobre espacios métricos separables y completos. Demostramos que la tasa de convergencia alcanzada es óptima, del orden de \sqrt{n} .

- utilizando la teoría de máximo sesgo asintótico, hallamos una expresión del dicho sesgo y demostramos que alcanzan el punto de ruptura asintótico óptimo de $1/2$.
- realizamos estudios de simulación para determinar las propiedades de los miembros de la familia sobre muestras finitas, comparando nuestra propuesta con otras.

Para modelos no-paramétricos (con diseño fijo) en los cuales la función de escala no es constante, introdujimos la familia de los M -estimadores locales de la escala basados también en diferencias sucesivas (de las respuestas) y, definidos a través de familias generales de pesos. Obtuvimos los siguientes resultados

- A través de elementos teóricos mencionados en el párrafo anterior y de teoría asintótica de procesos φ -mezclantes, probamos consistencia y hallamos la distribución límite de la sucesión de estimadores.
- Demostramos, utilizando elementos teóricos de procesos estocásticos en el espacio de las funciones continuas (sobre un compacto de \mathbb{R}), que cuando los M -estimadores definidos a partir de núcleos y con aberturas de ventana adaptivas poseen el mismo comportamiento asintótico que cuando la ventana es determinística.
- Análogamente, realizamos estudios de simulación para muestras finitas.

La importancia de estos resultados es que permitirán a futuro obtener estadísticos para testear en forma robusta hipótesis paramétricas sobre la función de varianza, en particular, homoscedasticidad. Por otra parte, el estimador propuesto por Boente *et al.* (1997) no está in-

cluido en los resultados de distribución asintótica descriptos en esta tesis ya que corresponde, como la mediana, a una función de escores no derivable. Como en Boente y Fraiman (1995), se podría derivar el comportamiento asintótico del estimador $\hat{\sigma}_{\text{BFM},n}(x)$ siempre y cuando U_1^* tenga densidad (dicha distribución sería un caso límite de la dada). La ventaja de nuestra propuesta es que la distribuciones del entorno no se reducen a distribuciones absolutamente continuas, sino que permitimos cualquier distribución G para los errores que cumpla la Hipótesis 14 que, como hemos visto, es poco restrictiva obteniendo un comportamiento para muestras finitas análogo al del estimador propuesto por Boente *et al.* (1997).

Apéndice A

Resultados Auxiliares

Demostración de la Proposición 2.4. La m -dependencia implica que, para cada entero $1 \leq k \leq m + 1$, $\{\xi_{k+j(m+1)}\}_{j \geq 0}$ es una sucesión de variables aleatorias independientes con media cero y uniformemente acotadas con lo cual es posible aplicar la ley fuerte de los grandes números para variables aleatorias independientes (que abreviaremos por LFGNI). De este modo si para cada n y cada $k = 1, \dots, m + 1$ definimos

$$I_{k,n} = \{i : 1 \leq i \leq n, i = k + j(m + 1) \text{ para algún } j \in \mathbb{N}\}$$

$S_{k,n} = \sum_{i \in I_{k,n}} \xi_i / \#I_{k,n}$ entonces $S_n/n = \sum_{i=1}^n \xi_i/n = \sum_{k=1}^{m+1} \#I_{k,n} S_{k,n}/n$, de donde, por la LFGNI, para cada k , $S_{k,n} \xrightarrow{c.s.} 0$ cuando $n \rightarrow \infty$ y por ende, $S_n \xrightarrow{c.s.} 0$ ya que $\#I_{k,n}/n \rightarrow 1/(m + 1)$. \square

Demostración del Lema 3.1. Por las Hipótesis 4 y 5

$$\begin{aligned}
 |\lambda_{n,b}(s) - \lambda_{n,b}^*(s)| &= (n-1)^{1/2} \frac{1}{n-1} \sum_{i=1}^{n-1} \left| \chi\left(\frac{Y_i^*}{as}\right) - \chi\left(\frac{\sigma U_i^*}{as}\right) \right| \\
 &\leq (n-1)^{-1/2} (as)^{-1} \|\chi\|_L \sum_{i=1}^{n-1} |Y_i^* - \sigma U_i^*| \\
 &= (n-1)^{-1/2} (as)^{-1} \|\chi\|_L \sum_{i=1}^{n-1} |g(x_{i+1}) - g(x_i)| \\
 &\leq (n-1)^{-1/2} (as)^{-1} \|g\|_L \|\chi\|_L (n-1) M_n \\
 &\leq K n^{1/2} M_n
 \end{aligned}$$

con K una constante. Por la Hipótesis 6, $M_n = O(n^{-1})$, luego $n^{1/2} M_n = o(1)$. \square

Demostración del Lema 3.2: Probemos (i). Para simplificar, definamos $\chi_s(y) = \chi(\sigma y / (as))$.

Sea $s > 0$ arbitrario y sea $\{s_n\}_{n \geq 1}$ una sucesión de términos positivos tal que $\lim_{n \rightarrow \infty} s_n = s$.

Dado $y \in \mathbb{R}$, la continuidad de χ implica que $\lim_{n \rightarrow \infty} \chi_{s_n}(y) = \chi_s(y)$. Como χ está acotada, por el Teorema de la Convergencia Dominada de Lebesgue, se tiene que

$$\lim_{n \rightarrow \infty} \lambda_G(s_n) = \lim_{n \rightarrow \infty} \int \chi_{s_n}(y) dG^*(y) = \int \chi_s(y) dG^*(y) = \lambda_G(s),$$

resultando λ_G una función continua.

Por las propiedades de χ es inmediato que λ_G es monótona no-creciente. Para mostrar que λ_G es estrictamente decreciente, supongamos por el absurdo que existen $s_1 < s_2$ reales

positivos tales que $\lambda_G(s_1) = \lambda_G(s_2)$. Luego,

$$0 = \lambda_G(s_1) - \lambda_G(s_2) = \int_{-\infty}^{\infty} [\chi_{s_1}(y) - \chi_{s_2}(y)] g^*(y) dy, \quad (\text{A.1})$$

donde g^* es la densidad de G^* , dada por $g^*(y) = \int_{-\infty}^{+\infty} g(y-x)g(-x)dx$. Como $\chi_{s_1}(y) \geq \chi_{s_2}(y)$ para todo $y \in \mathbb{R}$ y dado que g^* es estrictamente positiva (ya que g lo es por la Hipótesis 8), (A.1) implica que

$$\chi\left(\frac{\sigma y}{as_1}\right) - \chi\left(\frac{\sigma y}{as_2}\right) = 0 \text{ para casi todo punto } y.$$

Pero esto contradice que χ sea estrictamente creciente sobre el conjunto $\{u \geq 0 : \chi(u) < 1\}$, de donde se deduce que λ_G es estrictamente creciente. Finalmente, la derivación de los límites resulta de una aplicación directa del Teorema de la Convergencia Dominada de Lebesgue utilizando las propiedades de la función χ .

Probemos (ii). Tengamos en cuenta que

$$S(G_\sigma^*) = \inf \left\{ s > 0 : E \left[\chi \left(\frac{\sigma U_1^*}{as} \right) \right] \leq b \right\} = \inf \{ s > 0 : \lambda_G(s) \leq 0 \}.$$

Dado que, por el item (i), λ_G es una función continua y estrictamente creciente tal que $\lim_{s \rightarrow \infty} \lambda_G(s) = -b$ y $\lim_{s \rightarrow 0^+} \lambda_G(s) = 1 - b$ con $0 < b < 1$, se concluye la validez de (ii).

Para probar (iii), sin pérdida de generalidad, para simplificar, asumamos que $\sigma = 1$. Sean $\{B_i\}_{i=1,2}$ v.a.'s i.i.d. tal que $B_1 \sim \text{Binomial}(1, \epsilon)$ e independientes de $\{V_i, W_i\}$, $i = 1, 2$. Luego

U_i se puede expresar como $U_i = (1 - B_i) V_i + B_i W_i$, $i = 1, 2$ y, de este modo,

$$\begin{aligned} E \left[\chi \left(\frac{U_2 - U_1}{as} \right) \right] &= E \left[\chi \left(\frac{(1 - B_2) V_2 + B_2 W_2 - (1 - B_1) V_1 - B_1 W_1}{as} \right) \right] \\ &= P(B_1 = 0, B_2 = 0) E \left[\chi \left(\frac{V_2 - V_1}{as} \right) \right] + P(B_1 = 1, B_2 = 0) E \left[\chi \left(\frac{V_2 - W_1}{as} \right) \right] \\ &\quad + P(B_1 = 0, B_2 = 1) E \left[\chi \left(\frac{W_2 - V_1}{as} \right) \right] + P(B_1 = 1, B_2 = 1) E \left[\chi \left(\frac{W_2 - W_1}{as} \right) \right]. \end{aligned}$$

La descomposición en (iii) se deduce de manera directa a partir de esta última expresión y teniendo en cuenta que las variables aleatorias $\{B_i\}_{i=1,2}$ son independientes y tienen distribución común Binomial(1, ϵ). \square

De aquí en más, sin pérdida de generalidad, asumiremos como se mencionó en la Observación 3.5 que el parámetro de escala es $\sigma = 1$.

Demostración del Lema 3.3. Demostremos (i). Fijemos $s > 0$ y utilicemos el Lema 3.2(iii) para expresar

$$\begin{aligned} \lambda_{G_n}(s) &= (1 - \epsilon^2) E \left[\chi \left(\frac{Z_2 - Z_1}{as} \right) \right] + 2\epsilon(1 - \epsilon) E \left[\chi \left(\frac{Z_2 - W_1^{(n)}}{as} \right) \right] \\ &\quad + \epsilon^2 E \left[\chi \left(\frac{W_2^{(n)} - W_1^{(n)}}{as} \right) \right] - b, \end{aligned} \tag{A.2}$$

con Z_1, Z_2 v.a.'s i.i.d., $Z_1 \sim F_0$, $\{W_i^{(n)}\}_{i=1,2}$ v.a.'s i.i.d. con $W_1^{(n)} \sim H_n$.

Analicemos la convergencia, cuando $n \rightarrow \infty$, del segundo y tercer término de (A.2). Notemos que la densidad de $Z_2 - W_1^{(n)}$ está dada por $g(x) = (1/n) \int \phi(x - y/n) f_0(y) dy$, donde

$\phi = \Phi'$ y $f_0 = F_0'$. En consecuencia,

$$E \left[\chi \left(\frac{Z_2 - W_1^{(n)}}{as} \right) \right] = \int \chi \left(\frac{x}{as} \right) g(x) dx = \int \int \chi \left(\frac{x}{as} \right) \frac{1}{n} \phi \left(\frac{x-y}{n} \right) f_0(y) dy dx$$

o, equivalentemente,

$$E \left[\chi \left(\frac{Z_2 - W_1^{(n)}}{as} \right) \right] = \int \int \chi \left(\frac{z + nu}{as} \right) f_0(z) \phi(u) dz du. \quad (\text{A.3})$$

Tomando límite, cuando n tiende a infinito, a ambos lados de esta igualdad y utilizando el Teorema de la Convergencia Dominada de Lebesgue y considerando que, por hipótesis, $\chi(\infty) = 1$ se obtiene fácilmente que el segundo término de la suma en (A.2) tiende a 1. Con un argumento similar podemos probar que el tercer término en (A.2) también tiende a 1. En consecuencia,

$$\lim_{n \rightarrow \infty} \lambda_{G_n}(s) = (1 - \epsilon^2) E \left[\chi \left(\frac{Z_2 - Z_1}{as} \right) \right] + 2\epsilon(1 - \epsilon) + \epsilon^2 - b = \lambda_+(s).$$

Resta demostrar que dado $s > 0$, para todo $n \geq 1$, $\lambda_{G_n}(s) < \lambda_{G_{n+1}}(s)$. Para ello, tengamos en cuenta la descomposición de λ_{G_n} hecha en (A.2). Consideremos Z_1, Z_2 como antes y $\{W_i^{(n+1)}\}_{i=1,2}$ v.a.'s i.i.d. con $W_1^{(n+1)} \sim H_{n+1}$, entonces, por (A.2),

$$\begin{aligned} \lambda_{G_{n+1}}(s) &= (1 - \epsilon^2) E \left[\chi \left(\frac{Z_2 - Z_1}{as} \right) \right] + 2\epsilon(1 - \epsilon) E \left[\chi \left(\frac{Z_2 - W_1^{(n+1)}}{as} \right) \right] \\ &\quad + \epsilon^2 E \left[\chi \left(\frac{W_2^{(n+1)} - W_1^{(n+1)}}{as} \right) \right] - b. \end{aligned} \quad (\text{A.4})$$

Utilizando el tipo de desarrollo hecho en (A.3), es inmediato probar, por las propiedades de la función de escores, que

$$E \left[\chi \left(\frac{Z_2 - W_1^{(n)}}{as} \right) \right] \leq E \left[\chi \left(\frac{Z_2 - W_1^{(n+1)}}{as} \right) \right]. \quad (\text{A.5})$$

Ahora, como para cada $k = n, n + 1$, $W_i^{(k)} \sim N(0, k^2)$, $i = 1, 2$ y son independientes, con un simple cambio de variables obtenemos

$$\begin{aligned} E \left[\chi \left(\frac{W_2^{(n)} - W_1^{(n)}}{as} \right) \right] &= \int \chi \left(\frac{x}{as} \right) \frac{1}{\sqrt{2n}} \phi \left(\frac{x}{\sqrt{2n}} \right) dx \\ &= \int \chi \left(\frac{u\sqrt{2n}}{as} \right) \phi(u) du \\ &< \int \chi \left(\frac{u\sqrt{2}(n+1)}{as} \right) \phi(u) du \\ &= E \left[\chi \left(\frac{W_2^{(n+1)} - W_1^{(n+1)}}{as} \right) \right]. \end{aligned} \quad (\text{A.6})$$

La desigualdad anterior es estricta por las propiedades de la función χ . Para probarlo, basta usar un razonamiento análogo al usado en el Lema 3.2 para demostrar que λ_G es estrictamente creciente. Por las descomposiciones de $\lambda_{G_n}(s)$ y $\lambda_{G_{n+1}}(s)$ dadas en (A.2) y (A.4), respectivamente, y las desigualdades (A.5) y (A.6) se concluye que $\lambda_{G_n}(s) < \lambda_{G_{n+1}}(s)$.

Probemos (ii) y (iii). Si escribimos

$$\begin{aligned}\lambda_+(s) &= (1 - \epsilon^2) \left\{ E \left[\chi \left(\frac{Z_2 - Z_1}{as} \right) \right] - b + b \right\} + \epsilon(2 - \epsilon) - b \\ &= (1 - \epsilon)^2 \{ \lambda_{F_0}(s) + b \} + \epsilon(2 - \epsilon) - b.\end{aligned}$$

entonces las afirmaciones de los ítems (ii) y (iii) se deducen de (i) y (ii) del Lema 3.2. \square

Demostración del Lema 3.4. Es completamente análoga a la del Lema anterior. \square

Demostración del Lema 4.2. Probemos (a). Sea $X_i = w_{n,i}Z_i$, entonces el Lema 4.1 implica que

$$P \left(\left| \sum_{i=1}^{n-1} w_{n,i} X_i \right| \geq \epsilon \right) \leq 2 \exp \left(- \frac{\epsilon^2}{2c^2 \sum_{i=1}^{n-1} w_{n,i}^2} \right). \quad (\text{A.7})$$

En consecuencia, (a) se sigue de la siguiente desigualdad

$$\sum_{i=1}^{n-1} w_{n,i}^2 \leq \sum_{i=1}^{n-1} |w_{n,i}| w_n \leq M w_n.$$

Demostremos (b). Utilizando el inciso anterior obtenemos

$$\sum_{n=1}^{\infty} P \left(\left| \sum_{i=1}^{n-1} w_{n,i} X_i \right| \geq \epsilon \right) \leq 2 \sum_{n=1}^{\infty} \exp(-\alpha h_n)$$

con $\alpha = \epsilon^2/(2c^2M)$ y $h_n = w_n^{-1}$. Por (iv) de la Hipótesis 11, $\lim_{n \rightarrow \infty} h_n/\log(n) = \infty$ y así obtenemos $\sum_{n=1}^{\infty} \exp(-\alpha h_n) < \infty$, concluyendo la prueba. \square

Demostración del Lema 4.3. Demostraremos (4.21) probando las siguientes dos igualdades

$$c_n^{-1/2}\lambda_{n,b}(s, x) = c_n^{-1/2}\tilde{\lambda}_{n,b}(s, x) + o_p(1) \quad (\text{A.8})$$

$$c_n^{-1/2}\tilde{\lambda}_{n,b}(s, x) = c_n^{-1/2}\lambda_{n,b}^*(s, x) + c_n^{-1/2}\lambda_{1,n}(s, x) + o_p(1), \quad (\text{A.9})$$

siendo $\tilde{\lambda}_{n,b}(s, x) = \sum_{i=1}^{n-1} w_{n,i}(x)\chi\left(\frac{\tilde{U}_i}{as}\right) - b$ y $\tilde{U}_i = \sigma(x_{i+1})U_{i+1} - \sigma(x_i)U_i$, $1 \leq i \leq n-1$ como en (4.9).

De las Hipótesis 4, 7(i) (que implica la continuidad Lipschitz de χ) y 11(ii) concluimos que

$$\begin{aligned} \left| c_n^{-1/2}\lambda_{n,b}(s, x) - c_n^{-1/2}\tilde{\lambda}_{n,b}(s, x) \right| &\leq c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| \left| \chi\left(\frac{Y_i^*}{as}\right) - \chi\left(\frac{\tilde{U}_i}{as}\right) \right| \\ &\leq (as)^{-1} c_n^{-1/2} \|\chi\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)| |Y_i^* - \tilde{U}_i| \\ &= (as)^{-1} c_n^{-1/2} \|\chi\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)| |g(x_{i+1}) - g(x_i)| \\ &\leq (as)^{-1} \|g\|_L \|\chi\|_L (nM_n) c_n^{-1/2} w_n. \end{aligned}$$

donde $w_n = \max_{1 \leq i \leq n-1} |w_{n,i}(x)|$ y $M_n = \max_{1 \leq i \leq n-1} (x_{i+1} - x_i)$. La igualdad (A.8) se deduce de la Hipótesis 6 sobre los puntos del diseño, $nM_n = O(1)$, y de la Hipótesis 13(i) ($c_n^{-1/2}w_n \rightarrow 0$).

Consideremos (A.9). Escribamos

$$c_n^{-1/2}\tilde{\lambda}_{n,b}(s, x) - c_n^{-1/2}\lambda_{n,b}^*(s, x) = H_n + T_n.$$

con

$$\begin{aligned}
 H_n &= c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \left[\chi \left(\frac{\sigma(x_{i+1})U_{i+1} - \sigma(x_i)U_i}{as} \right) - \chi \left(\frac{\sigma(x_i)U_i^*}{as} \right) \right] \\
 T_n &= c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \left[\chi \left(\frac{\sigma(x_i)U_i^*}{as} \right) - \chi \left(\frac{\sigma(x)U_i^*}{as} \right) \right]
 \end{aligned}$$

Podremos concluir la validez de (A.9) si demostramos que

$$H_n = o_p(1) \text{ y } T_n = c_n^{-1/2} \lambda_{1,n}(s, x) + o_p(1).$$

Por las Hipótesis 11(ii), 12 y 14 tenemos que

$$\begin{aligned}
 E[|H_n|] &\leq (as)^{-1} c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| E[|\chi'(\xi_i)U_2|] |\sigma(x_{i+1}) - \sigma(x_i)| \\
 &\leq (as)^{-1} c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| \nu(\alpha_i, \beta_i) |\sigma(x_{i+1}) - \sigma(x_i)| \\
 &\leq (as)^{-1} \sup_{1 \leq i \leq n-1} \nu(\alpha_i, \beta_i) c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| |\sigma(x_{i+1}) - \sigma(x_i)|
 \end{aligned}$$

donde $\xi_i = \alpha_i U_2 + \beta_i U_1$ con $\alpha_i = \tilde{\sigma}_i / (as)$, $\beta_i = -\sigma(x_i) / (as)$ y $\tilde{\sigma}_i$ es un punto intermedio entre $\sigma(x_{i+1})$ y $\sigma(x_i)$. Como σ es continua y estrictamente positiva en el $[0, 1]$, existen \mathcal{K}_j , $j = 1, 2$ conjuntos compactos de $\mathbb{R} - \{0\}$ tales que $\alpha_i \in \mathcal{K}_1$ y $\beta_i \in \mathcal{K}_2$ $1 \leq i \leq n - 1$. Por el análisis hecho en la Observación 4.3, $\sup_{1 \leq i \leq n-1} \nu(\alpha_i, \beta_i) < \sup_{(\alpha, \beta) \in \mathcal{K}_1 \times \mathcal{K}_2} \nu(\alpha, \beta) < \infty$. Luego, por las Hipótesis 6, 12 y 13(i), $E[|H_n|] \rightarrow 0$.

Analicemos ahora el comportamiento de T_n . Para ello, consideremos el siguiente desarrollo

de Taylor de orden 2:

$$\begin{aligned}
 T_n &= c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{\sigma(x_i) U_i^*}{as} \right) - c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \chi \left(\frac{\sigma(x) U_i^*}{as} \right) \\
 &= c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \chi' \left(\frac{\sigma(x) U_i^*}{as} \right) \left(\frac{U_i^*}{as} \right) (\sigma(x_i) - \sigma(x)) \\
 &\quad + c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \chi_2 \left(\frac{\tilde{\sigma}_i U_i^*}{as} \right) \frac{1}{\tilde{\sigma}_i^2} (\sigma(x_i) - \sigma(x))^2 \\
 &= c_n^{-1/2} \lambda_{1,n}(s, x) + c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \chi_2 \left(\frac{\tilde{\sigma}_i U_i^*}{as} \right) \frac{1}{\tilde{\sigma}_i^2} (\sigma(x_i) - \sigma(x))^2
 \end{aligned}$$

con $\chi_2(u) = u^2 \chi''(u)$ como en la Hipótesis 7 y $\tilde{\sigma}_i$ es un punto intermedio entre $\sigma(x_i)$ y $\sigma(x)$, $i = 1, \dots, n-1$. Así, por la Hipótesis 7(ii)

$$\begin{aligned}
 E [|T_n - c_n^{-1/2} \lambda_{1,n}(s, x)|] &\leq c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| E \left[\left| \chi_2 \left(\frac{\tilde{\sigma}_i U_i^*}{as} \right) \right| \frac{1}{\inf_{1 \leq i \leq n-1} \tilde{\sigma}_i^2} (\sigma(x_i) - \sigma(x))^2 \right] \\
 &\leq \frac{\|\chi_2\|}{\inf_{u \in (0,1)} \sigma^2(u)} c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| (\sigma(x_i) - \sigma(x))^2.
 \end{aligned}$$

Dado que σ es una función continua y estrictamente positiva sobre el intervalo $[0, 1]$, entonces $\inf_{u \in (0,1)} \sigma^2(u) > 0$. Por ende, la constante $\|\chi_2\| / \inf_{u \in (0,1)} \sigma^2(u)$ es finita. Luego, la desigualdad anterior y la Hipótesis 15(ii) implican que $E \left[|T_n - c_n^{-1/2} \lambda_{1,n}(s, x)| \right] \rightarrow 0$ y entonces $T_n - c_n^{-1/2} \lambda_{1,n}(s, x) = o_p(1)$.

Probemos ahora (4.22). Para abreviar escribamos $Z_i = \chi' \left(\frac{\sigma(x) U_i^*}{aS_x} \right) \left(\frac{U_i^*}{aS_x} \right)$ y $\xi_i = Z_i - E[Z_i]$. Observemos que $Z_i = \frac{1}{\sigma(x)} \chi_1 \left(\frac{\sigma(x) U_i^*}{aS_x} \right)$ donde $\chi_1(u) = u \chi'(u)$. Luego, por la Hipótesis 7(ii) χ_1 está acotada y por ende $\{\xi_i\}_{i \geq 1}$ es una sucesión de v.a.'s idénticamente

distribuidas, uno-dependientes y uniformemente acotadas. Observemos que $c_n^{-1/2} \lambda_{1,n}(s, x) =$

$H_n + T_n$ con

$$\begin{aligned} H_n &= c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) E[Z_i] (\sigma(x_i) - \sigma(x)) \\ T_n &= c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) (Z_i - E[Z_i]) (\sigma(x_i) - \sigma(x)) \\ &= c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x) \xi_i (\sigma(x_i) - \sigma(x)), \end{aligned}$$

entonces (4.22) quedará demostrada si probamos que

$$H_n \longrightarrow \beta_1 E[Z_1] \tag{A.10}$$

$$T_n \xrightarrow{p} 0. \tag{A.11}$$

La convergencia en (A.10) es inmediata por la Hipótesis 15(i). Demostremos ahora (A.11).

Como

$$\begin{aligned} \text{Var}[T_n] &\leq c_n^{-1} \sum_{i=1}^{n-1} w_{n,i}^2(x) E[\xi_i^2] (\sigma(x_i) - \sigma(x))^2 \\ &\quad + c_n^{-1} \sum_{i=1}^{n-2} |w_{n,i}(x)| |w_{n,i+1}(x)| E[|\xi_i, \xi_{i+1}|] (\sigma(x_i) - \sigma(x)) (\sigma(x_{i+1}) - \sigma(x)) \\ &= A_n + B_n, \end{aligned}$$

será suficiente mostrar que A_n y B_n tienden a cero. Tengamos en cuenta la siguiente acotación

$$\begin{aligned}
 A_n &= c_n^{-1} \sum_{i=1}^{n-1} w_{n,i}^2(x) E [\xi_i^2] (\sigma(x_i) - \sigma(x))^2 \\
 &\leq E [\xi_1^2] w_n c_n^{-1} \sum_{i=1}^{n-1} |w_{n,i}(x)| |\sigma(x_i) - \sigma(x)|^2 \\
 &= k [c_n^{-1/2} w_n] \left[c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| |\sigma(x_i) - \sigma(x)|^2 \right]
 \end{aligned}$$

con $k = E [\xi_1^2]$. Luego, por las Hipótesis 13(i) y 15(ii), $A_n \rightarrow 0$. Análogamente para B_n tenemos

$$\begin{aligned}
 B_n &= c_n^{-1} \sum_{i=1}^{n-2} |w_{n,i}(x)| |w_{n,i+1}(x)| E [|\xi_i, \xi_{i+1}|] (\sigma(x_i) - \sigma(x)) (\sigma(x_{i+1}) - \sigma(x)) \\
 &\leq k [c_n^{-1/2} w_n] \left[c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| |\sigma(x_i) - \sigma(x)| \right]
 \end{aligned}$$

con $k = E [|\xi_1, \xi_2|]$; luego $B_n \rightarrow 0$ por las mismas razones esgrimidas para A_n . De este modo

$Var [T_n] \rightarrow 0$. \square

Apéndice B

Cuadros y Gráficos

B.1. Correspondientes a los Resultados de la Simulación Monte Carlo para Modelos Homoscedásticos

n	E.R. $(\hat{\sigma}_{\text{BFM},n}, \hat{\sigma}_{\text{R},n})$	E.R. $(\hat{\sigma}_{\text{PC},n}, \hat{\sigma}_{\text{R},n})$	E.R. $(\hat{\sigma}_{\text{BT},n}, \hat{\sigma}_{\text{R},n})$
20	0.350	0.139	0.200
50	0.426	0.192	0.276
100	0.451	0.206	0.286
Asintótica	0.454	0.214	0.297

Cuadro B.1: Eficiencias $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$ relativas a $\hat{\sigma}_{\text{R},n}$ con $\epsilon = 0$.

	Estimador	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$	$\epsilon = 0.40$
$n = 20$	$\hat{\sigma}_{\text{R},n}$	3.012	6.385	13.805	21.660	29.930
	$\hat{\sigma}_{\text{BFM},n}$	0.273	0.549	2.349	7.346	15.860
	$\hat{\sigma}_{\text{PC},n}$	0.519	0.805	1.816	4.394	10.060
	$\hat{\sigma}_{\text{BT},n}$	0.352	0.578	1.449	3.720	8.544
$n = 50$	$\hat{\sigma}_{\text{R},n}$	2.505	5.661	13.027	20.900	29.220
	$\hat{\sigma}_{\text{BFM},n}$	0.072	0.163	0.730	3.181	10.060
	$\hat{\sigma}_{\text{PC},n}$	0.129	0.217	0.602	1.520	3.808
	$\hat{\sigma}_{\text{BT},n}$	0.092	0.168	0.526	1.455	3.867
$n = 100$	$\hat{\sigma}_{\text{R},n}$	2.319	5.452	12.78	20.770	29.020
	$\hat{\sigma}_{\text{BFM},n}$	0.038	0.100	0.485	2.106	7.781
	$\hat{\sigma}_{\text{PC},n}$	0.063	0.118	0.382	1.050	2.672
	$\hat{\sigma}_{\text{BT},n}$	0.047	0.098	0.356	1.055	2.849

Cuadro B.2: Valores estimados del E.C.M. de $\hat{\sigma}_{\text{R},n}$, $\hat{\sigma}_{\text{BFM},n}$, $\hat{\sigma}_{\text{PC},n}$ y $\hat{\sigma}_{\text{BT},n}$. La contaminación es simétrica con $H(y) = \Phi(y/10)$.

	Estimador	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$	$\epsilon = 0.40$
$n = 20$	$\hat{\sigma}_{R,n}$	2.564	5.281	10.32	14.120	16.440
	$\hat{\sigma}_{BFM,n}$	0.329	1.049	6.691	17.160	26.900
	$\hat{\sigma}_{PC,n}$	0.560	0.917	2.274	5.011	8.280
	$\hat{\sigma}_{BT,n}$	0.386	0.678	1.842	4.054	6.591
$n = 50$	$\hat{\sigma}_{R,n}$	2.160	4.816	9.882	13.720	16.090
	$\hat{\sigma}_{BFM,n}$	0.082	0.204	1.588	9.970	23.120
	$\hat{\sigma}_{PC,n}$	0.140	0.248	0.645	1.306	2.053
	$\hat{\sigma}_{BT,n}$	0.102	0.196	0.577	1.276	2.119
$n = 100$	$\hat{\sigma}_{R,n}$	2.037	4.703	9.778	13.670	16.050
	$\hat{\sigma}_{BFM,n}$	0.045	0.126	0.668	5.461	20.050
	$\hat{\sigma}_{PC,n}$	0.069	0.135	0.412	0.897	1.434
	$\hat{\sigma}_{BT,n}$	0.052	0.115	0.393	0.913	1.528

Cuadro B.3: Valores estimados del E.C.M. de $\hat{\sigma}_{R,n}$, $\hat{\sigma}_{BFM,n}$, $\hat{\sigma}_{PC,n}$ y $\hat{\sigma}_{BT,n}$. La contaminación es asimétrica con $H(y) = \Phi(y - 10)$.

	Estimador	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.25$	$\epsilon = 0.30$	$\epsilon = 0.35$
$n = 20$	$\hat{\sigma}_{R,n}$	0.723	4.029	11.550	15.340	19.660	24.320
	$\hat{\sigma}_{BFM,n}$	0.187	0.386	1.795	6.155	61.290	73.890
	$\hat{\sigma}_{PC,n}$	0.418	0.672	1.803	3.466	6.662	21.550
	$\hat{\sigma}_{BT,n}$	0.272	0.468	1.434	2.931	6.251	24.610
$n = 50$	$\hat{\sigma}_{R,n}$	1.012	4.777	12.220	15.360	20.290	23.660
	$\hat{\sigma}_{BFM,n}$	0.054	0.164	1.198	3.743	69.980	76.980
	$\hat{\sigma}_{PC,n}$	0.109	0.226	0.876	1.553	4.109	9.943
	$\hat{\sigma}_{BT,n}$	0.074	0.174	0.786	1.492	4.469	13.210
$n = 100$	$\hat{\sigma}_{R,n}$	1.813	5.032	12.450	16.400	20.460	24.610
	$\hat{\sigma}_{BFM,n}$	0.039	0.126	1.130	7.720	71.250	78.860
	$\hat{\sigma}_{PC,n}$	0.063	0.138	0.668	1.487	3.576	11.980
	$\hat{\sigma}_{BT,n}$	0.047	0.117	0.657	1.556	4.165	19.350

Cuadro B.4: Valores estimados de E.C.M. de los estimadores $\hat{\sigma}_{R,n}$, $\hat{\sigma}_{BFM,n}$, $\hat{\sigma}_{PC,n}$ y $\hat{\sigma}_{BT,n}$ con contaminaciones intercaladas y $H(y) = \Delta_{10}(y)$.

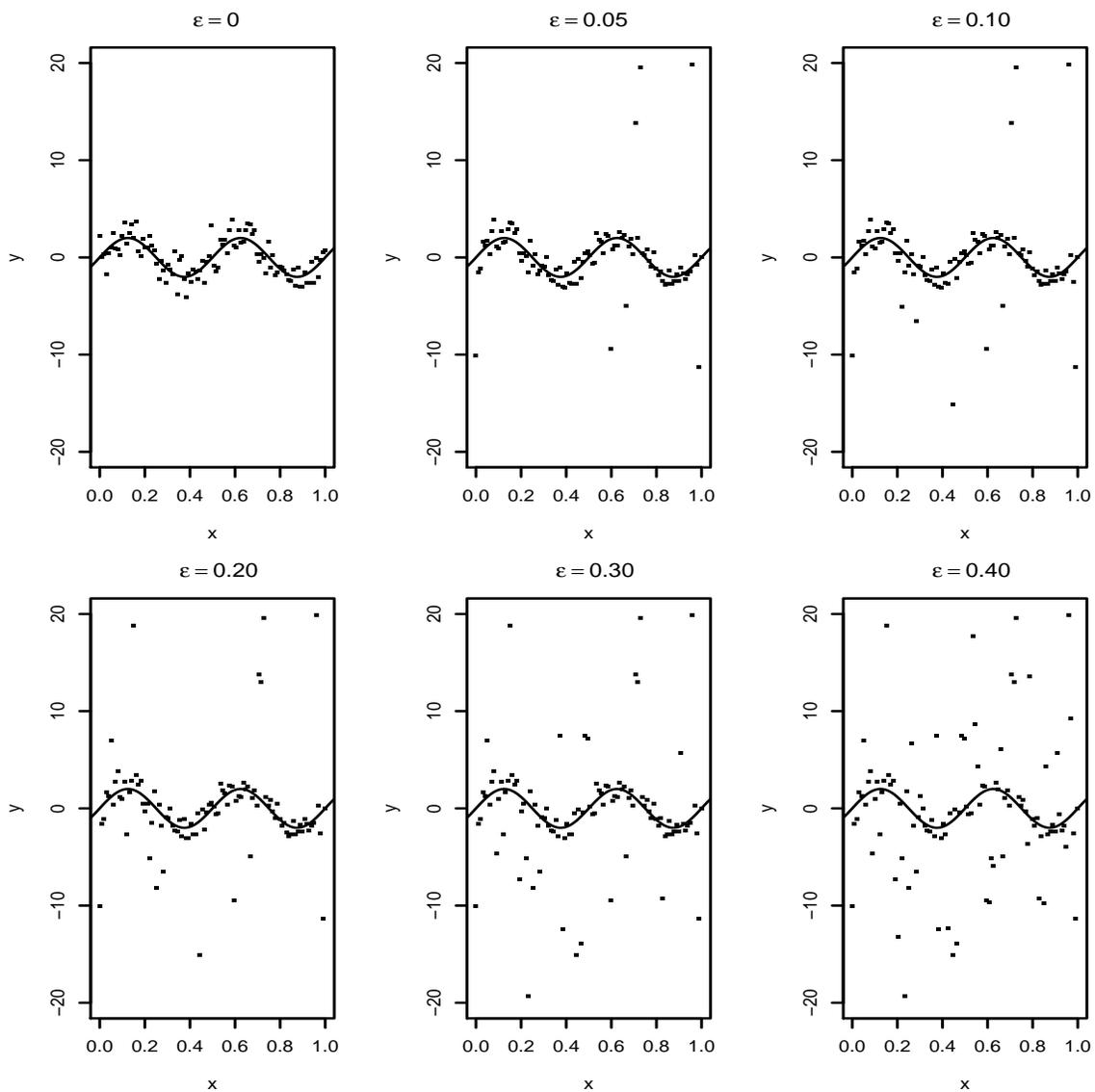


Figura B.1: Datos simulados del modelo de contaminaciones independientes con $n = 100$ y contaminación simétrica $H(y) = \Phi(y/10)$ para diferentes valores de ϵ .

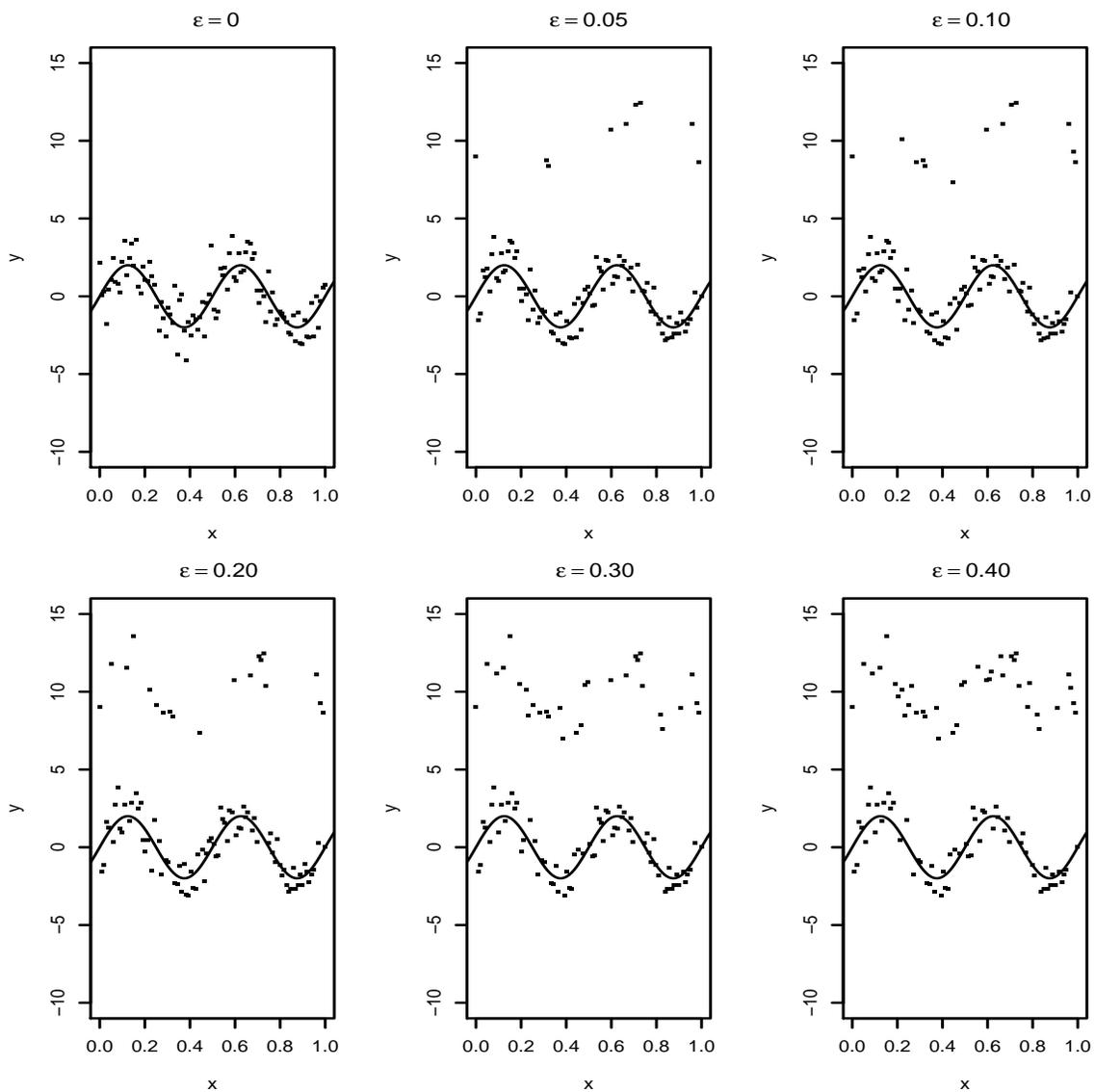


Figura B.2: Datos simulados del modelo de contaminaciones independientes con $n = 100$ y contaminación simétrica $H(y) = \Phi(y - 10)$ para diferentes valores de ϵ .

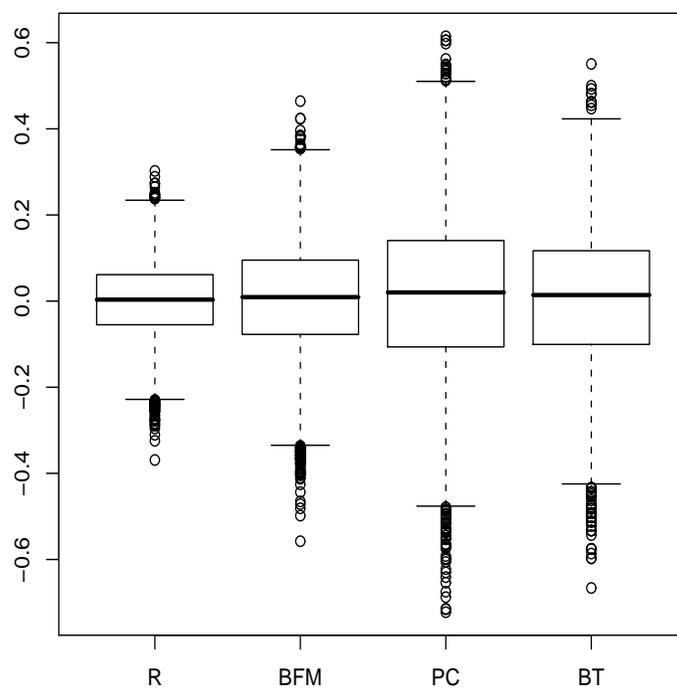


Figura B.3: Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ y en ausencia de contaminación ($\epsilon = 0$).

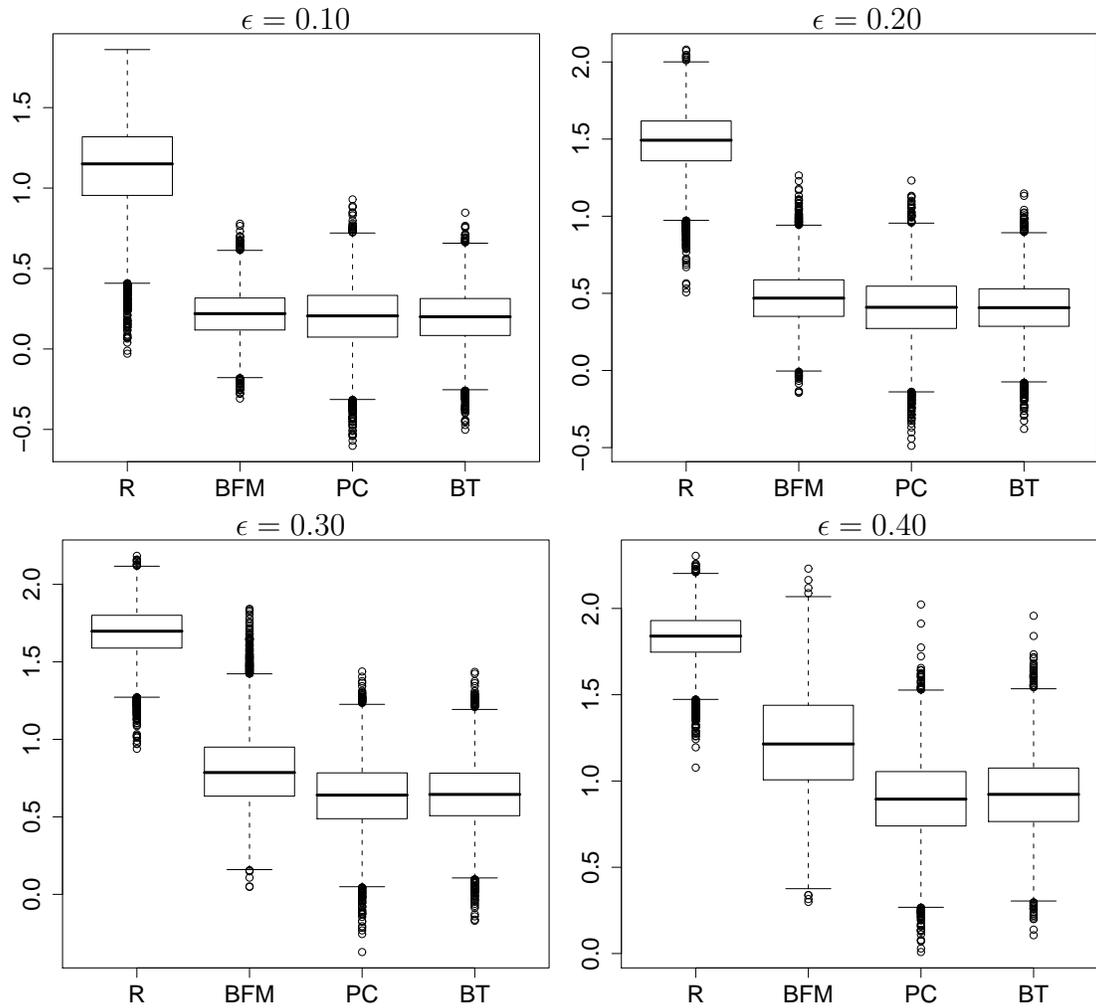


Figura B.4: Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ y contaminación simétrica $H(y) = \Phi(y/10)$ según proporción de contaminación ϵ .

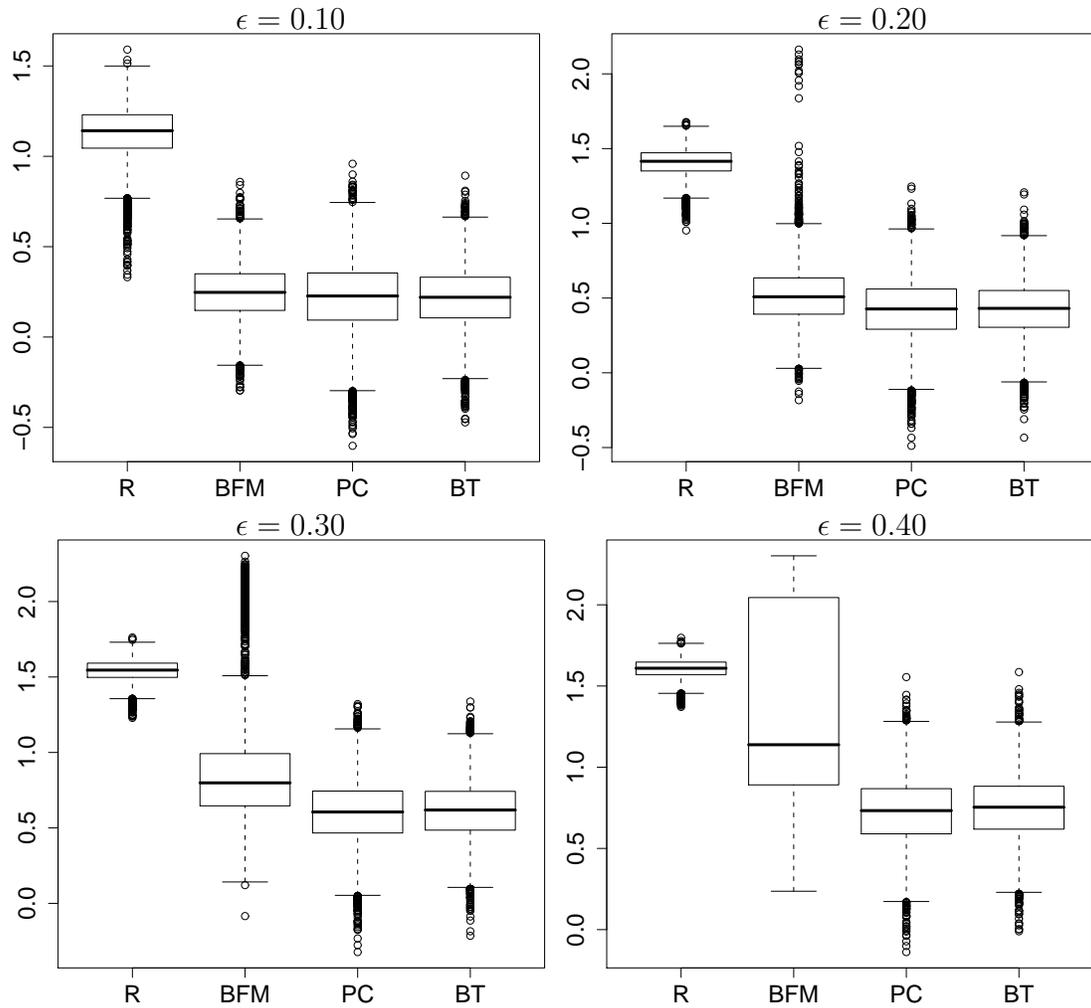


Figura B.5: Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ y contaminación asimétrica $H(y) = \Phi(y - 10)$ según proporción de contaminación ϵ .

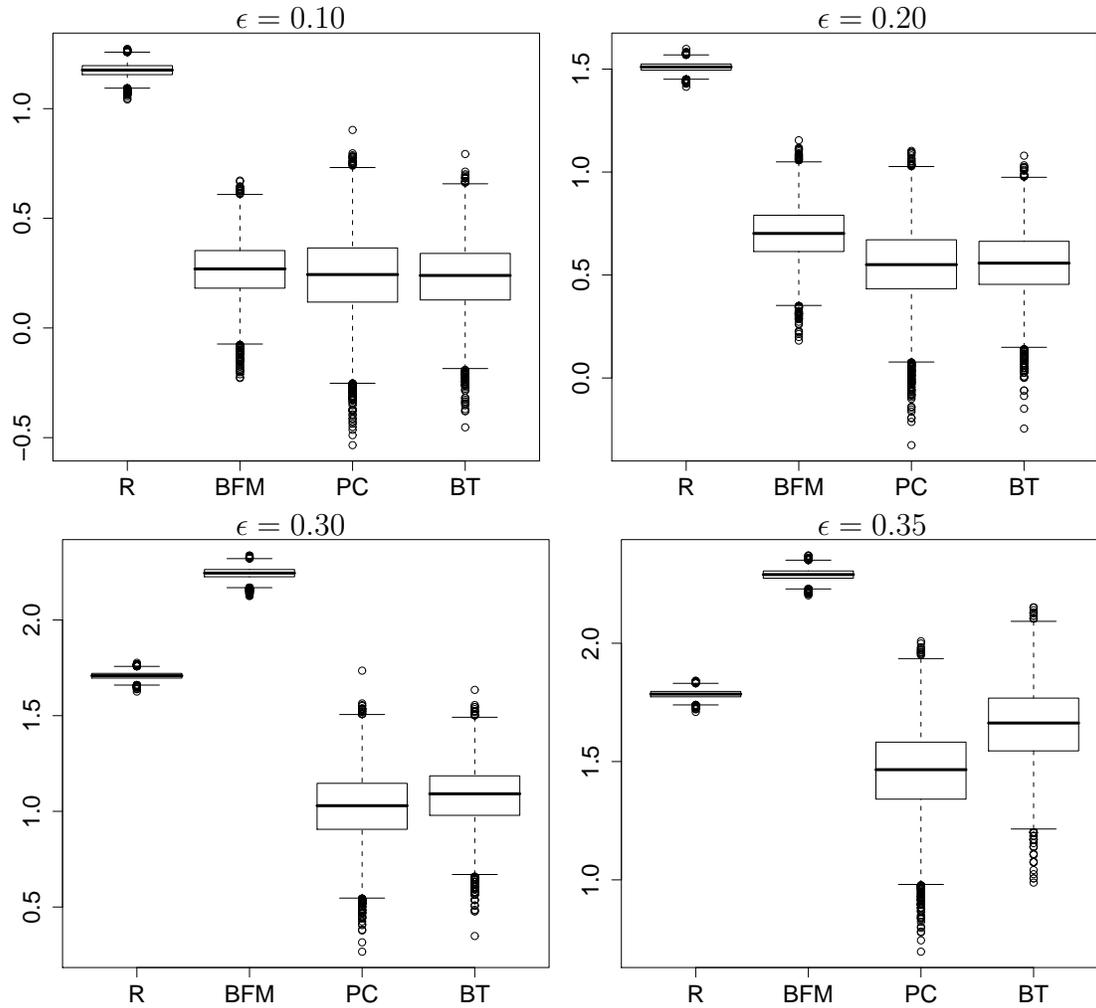


Figura B.6: Boxplots para los valores de $\log(\hat{\sigma}_{R,n})$, $\log(\hat{\sigma}_{BFM,n})$, $\log(\hat{\sigma}_{PC,n})$ y $\log(\hat{\sigma}_{BT,n})$ con $n = 100$ según proporción ϵ de observaciones contaminadas intercaladas con $H(y) = \Delta_{10}$.

B.2. Correspondientes a los Resultados de la Simulación Monte Carlo para Modelos Heteroscedásticos

Cuadros de Resúmenes Estadísticos del Error Cuadrático Integrado en escala logarítmica (E.C.I.L.).

Para modelos con contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.057 (0.052)	0.677 (0.210)	1.200 (0.636)	1.748 (1.058)
	$\hat{\sigma}_{BT,n}$	0.197 (0.176)	0.212 (0.189)	0.230 (0.212)	0.254 (0.235)
	$\hat{\sigma}_{BFM,n}$	0.126 (0.114)	0.135 (0.124)	0.018 (0.001)	0.021 (0.000)
$n = 200$	$\hat{\sigma}_{R,n}$	0.029 (0.027)	0.768 (0.358)	1.524 (0.927)	2.214 (1.608)
	$\hat{\sigma}_{BT,n}$	0.098 (0.089)	0.104 (0.098)	0.117 (0.109)	0.138 (0.130)
	$\hat{\sigma}_{BFM,n}$	0.063 (0.058)	0.070 (0.066)	0.084 (0.079)	0.107 (0.101)

Cuadro B.5: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.04$. Modelo $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.021 (0.016)	1.322 (0.199)	2.279 (0.862)	3.259 (1.525)
	$\hat{\sigma}_{BT,n}$	0.052 (0.038)	0.058 (0.043)	0.069 (0.051)	0.086 (0.065)
	$\hat{\sigma}_{BFM,n}$	0.037 (0.027)	0.042 (0.033)	0.056 (0.044)	0.078 (0.060)
$n = 200$	$\hat{\sigma}_{R,n}$	0.014 (0.011)	1.449 (0.449)	2.846 (1.338)	4.001 (2.366)
	$\hat{\sigma}_{BT,n}$	0.029 (0.020)	0.032 (0.024)	0.041 (0.033)	0.059 (0.049)
	$\hat{\sigma}_{BFM,n}$	0.021 (0.016)	0.025 (0.019)	0.037 (0.030)	0.058 (0.049)

Cuadro B.6: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.059 (0.055)	1.603 (0.247)	2.714 (1.005)	3.817 (1.742)
	$\hat{\sigma}_{BT,n}$	0.070 (0.062)	0.073 (0.065)	0.082 (0.068)	0.096 (0.079)
	$\hat{\sigma}_{BFM,n}$	0.061 (0.056)	0.064 (0.059)	0.077 (0.067)	0.097 (0.081)
$n = 200$	$\hat{\sigma}_{R,n}$	0.054 (0.052)	1.753 (0.520)	3.342 (1.488)	4.651 (2.664)
	$\hat{\sigma}_{BT,n}$	0.058 (0.052)	0.059 (0.054)	0.065 (0.060)	0.081 (0.072)
	$\hat{\sigma}_{BFM,n}$	0.052 (0.050)	0.054 (0.052)	0.064 (0.060)	0.085 (0.077)

Cuadro B.7: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.066 (0.061)	0.712 (0.239)	1.254 (0.687)	1.831 (1.135)
	$\hat{\sigma}_{BT,n}$	0.211 (0.184)	0.223 (0.198)	0.236 (0.212)	0.260 (0.241)
	$\hat{\sigma}_{BFM,n}$	0.124 (0.110)	0.124 (0.113)	0.129 (0.118)	0.139 (0.130)
$n = 200$	$\hat{\sigma}_{R,n}$	0.038 (0.035)	0.824 (0.407)	1.613 (1.006)	2.329 (1.718)
	$\hat{\sigma}_{BT,n}$	0.106 (0.098)	0.110 (0.102)	0.121 (0.113)	0.140 (0.133)
	$\hat{\sigma}_{BFM,n}$	0.068 (0.061)	0.069 (0.064)	0.076 (0.071)	0.068 (0.072)

Cuadro B.8: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.04$. Modelo $M2 : g(x) = 1 + x, \sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.326 (0.316)	1.723 (0.599)	2.740 (1.309)	3.821 (2.088)
	$\hat{\sigma}_{BT,n}$	0.284 (0.270)	0.276 (0.263)	0.274 (0.261)	0.275 (0.264)
	$\hat{\sigma}_{BFM,n}$	0.248 (0.244)	0.243 (0.239)	0.248 (0.244)	0.262 (0.253)
$n = 200$	$\hat{\sigma}_{R,n}$	0.321 (0.315)	1.987 (0.910)	3.509 (1.906)	4.708 (2.970)
	$\hat{\sigma}_{BT,n}$	0.264 (0.256)	0.253 (0.248)	0.248 (0.243)	0.250 (0.245)
	$\hat{\sigma}_{BFM,n}$	0.237 (0.235)	0.229 (0.228)	0.232 (0.230)	0.244 (0.240)

Cuadro B.9: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo $M2 : g(x) = 1 + x, \sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.402 (0.388)	1.999 (0.670)	3.170 (1.419)	4.375 (2.281)
	$\hat{\sigma}_{BT,n}$	0.382 (0.359)	0.370 (0.348)	0.365 (0.347)	0.366 (0.347)
	$\hat{\sigma}_{BFM,n}$	0.357 (0.346)	0.352 (0.342)	0.358 (0.345)	0.372 (0.352)
$n = 200$	$\hat{\sigma}_{R,n}$	0.398 (0.392)	2.293 (1.003)	4.018 (2.051)	5.351 (3.237)
	$\hat{\sigma}_{BT,n}$	0.370 (0.357)	0.356 (0.343)	0.347 (0.337)	0.346 (0.338)
	$\hat{\sigma}_{BFM,n}$	0.350 (0.342)	0.341 (0.336)	0.343 (0.337)	0.356 (0.346)

Cuadro B.10: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo $M2 : g(x) = 1 + x, \sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\widehat{\sigma}_{R,n}$	0.058 (0.053)	0.680 (0.211)	1.204 (0.640)	1.753 (1.062)
	$\widehat{\sigma}_{BT,n}$	0.203 (0.178)	0.216 (0.191)	0.231 (0.212)	0.256 (0.236)
	$\widehat{\sigma}_{BFM,n}$	0.115 (0.101)	0.117 (0.106)	0.125 (0.114)	0.138 (0.127)
$n = 200$	$\widehat{\sigma}_{R,n}$	0.029 (0.027)	0.770 (0.361)	1.527 (0.933)	2.219 (1.613)
	$\widehat{\sigma}_{BT,n}$	0.098 (0.089)	0.104 (0.100)	0.117 (0.110)	0.138 (0.129)
	$\widehat{\sigma}_{BFM,n}$	0.057 (0.052)	0.061 (0.056)	0.069 (0.064)	0.085 (0.079)

Cuadro B.11: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.04$. Modelo $M3 : g(x) = 1 + x, \sigma(x) = (1 + x)^2$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\widehat{\sigma}_{R,n}$	0.030 (0.025)	1.366 (0.226)	2.344 (0.885)	3.339 (1.596)
	$\widehat{\sigma}_{BT,n}$	0.057 (0.042)	0.062 (0.049)	0.073 (0.056)	0.088 (0.069)
	$\widehat{\sigma}_{BFM,n}$	0.041 (0.032)	0.045 (0.035)	0.057 (0.046)	0.075 (0.061)
$n = 200$	$\widehat{\sigma}_{R,n}$	0.023 (0.020)	1.498 (0.484)	2.914 (1.394)	4.090 (2.446)
	$\widehat{\sigma}_{BT,n}$	0.034 (0.026)	0.037 (0.028)	0.045 (0.037)	0.062 (0.053)
	$\widehat{\sigma}_{BFM,n}$	0.026 (0.021)	0.030 (0.024)	0.041 (0.034)	0.060 (0.052)

Cuadro B.12: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo $M3 : g(x) = 1 + x, \sigma(x) = (1 + x)^2$. Contaminación simétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\widehat{\sigma}_{R,n}$	0.112 (0.107)	1.711 (0.337)	2.852 (1.11)	3.974 (1.893)
	$\widehat{\sigma}_{BT,n}$	0.109 (0.102)	0.110 (0.102)	0.116 (0.104)	0.127 (0.111)
	$\widehat{\sigma}_{BFM,n}$	0.097 (0.094)	0.100 (0.097)	0.112 (0.104)	0.132 (0.119)
$n = 200$	$\widehat{\sigma}_{R,n}$	0.108 (0.105)	1.873 (0.611)	3.486 (1.628)	4.823 (2.848)
	$\widehat{\sigma}_{BT,n}$	0.096 (0.092)	0.095 (0.091)	0.010 (0.096)	0.113 (0.106)
	$\widehat{\sigma}_{BFM,n}$	0.089 (0.089)	0.021 (0.000)	0.101 (0.010)	0.121 (0.113)

Cuadro B.13: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo $M3 : g(x) = 1 + x, \sigma(x) = (1 + x)^2$. Contaminación simétrica.

Estimador		$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.014 (0.010)	1.277 (0.167)	2.207 (0.794)	3.177 (1.451)
	$\hat{\sigma}_{BT,n}$	0.047 (0.034)	0.053 (0.038)	0.065 (0.046)	0.082 (0.063)
	$\hat{\sigma}_{BFM,n}$	0.032 (0.024)	0.037 (0.027)	0.051 (0.037)	0.073 (0.051)
$n = 200$	$\hat{\sigma}_{R,n}$	0.007 (0.005)	1.393 (0.407)	2.784 (1.276)	3.912 (2.312)
	$\hat{\sigma}_{BT,n}$	0.023 (0.017)	0.027 (0.020)	0.037 (0.028)	0.056 (0.046)
	$\hat{\sigma}_{BFM,n}$	0.016 (0.011)	0.020 (0.015)	0.031 (0.024)	0.052 (0.043)

Cuadro B.14: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.2$. Modelo M.E.G.: $g(x) = 1 + x$, $\sigma(x) \equiv 1$. Contaminación simétrica.

Estimador		$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$n = 100$	$\hat{\sigma}_{R,n}$	0.008 (0.004)	1.478 (0.155)	2.555 (0.843)	3.641 (1.610)
	$\hat{\sigma}_{BT,n}$	0.027 (0.014)	0.033 (0.017)	0.044 (0.022)	0.060 (0.035)
	$\hat{\sigma}_{BFM,n}$	0.019 (0.010)	0.024 (0.012)	0.037 (0.021)	0.056 (0.033)
$n = 200$	$\hat{\sigma}_{R,n}$	0.004 (0.002)	1.616 (0.416)	3.196 (1.384)	4.459 (2.524)
	$\hat{\sigma}_{BT,n}$	0.014 (0.007)	0.017 (0.008)	0.027 (0.017)	0.044 (0.033)
	$\hat{\sigma}_{BFM,n}$	0.009 (0.005)	0.013 (0.007)	0.024 (0.016)	0.044 (0.034)

Cuadro B.15: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Abertura de ventana $h = 0.5$. Modelo M.E.G.: $g(x) = 1 + x$, $\sigma(x) \equiv 1$. Contaminación simétrica.

Para modelos con contaminación asimétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$h = 0.04$	$\hat{\sigma}_{R,n}$	0.057 (0.052)	1.194 (1.181)	1.909 (1.915)	2.331 (2.348)
	$\hat{\sigma}_{BT,n}$	0.197 (0.176)	0.285 (0.269)	0.494 (0.449)	0.780 (0.706)
	$\hat{\sigma}_{BFM,n}$	0.126 (0.114)	0.207 (0.197)	0.364 (0.356)	0.512 (0.517)
$h = 0.2$	$\hat{\sigma}_{R,n}$	0.021 (0.016)	1.362 (1.358)	2.075 (2.076)	2.476 (2.489)
	$\hat{\sigma}_{BT,n}$	0.052 (0.038)	0.105 (0.079)	0.251 (0.223)	0.452 (0.406)
	$\hat{\sigma}_{BFM,n}$	0.037 (0.027)	0.105 (0.085)	0.317 (0.286)	0.568 (0.563)
$h = 0.5$	$\hat{\sigma}_{R,n}$	0.059 (0.055)	1.506 (1.515)	2.233 (2.242)	2.641 (2.653)
	$\hat{\sigma}_{BT,n}$	0.070 (0.062)	0.056 (0.008)	0.250 (0.222)	0.443 (0.405)
	$\hat{\sigma}_{BFM,n}$	0.061 (0.056)	0.124 (0.103)	0.358 (0.317)	0.655 (0.713)

Cuadro B.16: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Modelo $M1$: $g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. $n = 100$. Contaminación asimétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$h = 0.04$	$\hat{\sigma}_{R,n}$	0.066 (0.061)	1.303 (1.294)	2.057 (2.068)	2.499 (2.516)
	$\hat{\sigma}_{BT,n}$	0.211 (0.184)	0.295 (0.281)	0.509 (0.465)	0.806 (0.736)
	$\hat{\sigma}_{BFM,n}$	0.124 (0.110)	0.154 (0.143)	0.249 (0.232)	0.344 (0.340)
$h = 0.2$	$\hat{\sigma}_{R,n}$	0.326 (0.316)	2.132 (2.163)	2.974 (3.019)	3.443 (3.467)
	$\hat{\sigma}_{BT,n}$	0.284 (0.270)	0.284 (0.267)	0.405 (0.363)	0.619 (0.562)
	$\hat{\sigma}_{BFM,n}$	0.248 (0.244)	0.287 (0.268)	0.447 (0.464)	0.547 (0.064)
$h = 0.5$	$\hat{\sigma}_{R,n}$	0.402 (0.388)	2.244 (2.271)	3.080 (3.122)	3.541 (3.561)
	$\hat{\sigma}_{BT,n}$	0.382 (0.359)	0.369 (0.348)	0.473 (0.430)	0.675 (0.620)
	$\hat{\sigma}_{BFM,n}$	0.357 (0.346)	0.404 (0.376)	0.534 (0.575)	0.572 (0.575)

Cuadro B.17: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Modelo $M2$: $g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. $n = 100$. Contaminación asimétrica.

	Estimador	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
$h = 0.04$	$\hat{\sigma}_{R,n}$	0.058 (0.053)	1.201 (1.188)	1.918 (1.924)	2.341 (2.359)
	$\hat{\sigma}_{BT,n}$	0.203 (0.178)	0.286 (0.269)	0.494 (0.456)	0.780 (0.703)
	$\hat{\sigma}_{BFM,n}$	0.115 (0.101)	0.154 (0.142)	0.243 (0.234)	0.327 (0.340)
$h = 0.2$	$\hat{\sigma}_{R,n}$	0.030 (0.025)	1.434 (1.429)	2.163 (2.164)	2.573 (2.581)
	$\hat{\sigma}_{BT,n}$	0.057 (0.042)	0.107 (0.081)	0.253 (0.225)	0.457 (0.410)
	$\hat{\sigma}_{BFM,n}$	0.041 (0.032)	0.100 (0.081)	0.249 (0.229)	0.391 (0.409)
$h = 0.5$	$\hat{\sigma}_{R,n}$	0.112 (0.107)	1.680 (1.680)	2.437 (2.450)	2.859 (2.873)
	$\hat{\sigma}_{BT,n}$	0.109 (0.102)	0.144 (0.123)	0.277 (0.246)	0.471 (0.428)
	$\hat{\sigma}_{BFM,n}$	0.097 (0.094)	0.162 (0.142)	0.352 (0.356)	0.486 (0.528)

Cuadro B.18: Media y mediana (entre paréntesis) del E.C.I.L. de los estimadores. Modelo $M3$: $g(x) = 1 + x$, $\sigma(x) = (1 + x)^2$. $n = 100$. Contaminación asimétrica.

*Gráficos de valores de las variables respuestas versus puntos del diseño.
Contaminación Simétrica.*

Modelo Heteroscedástico M1.

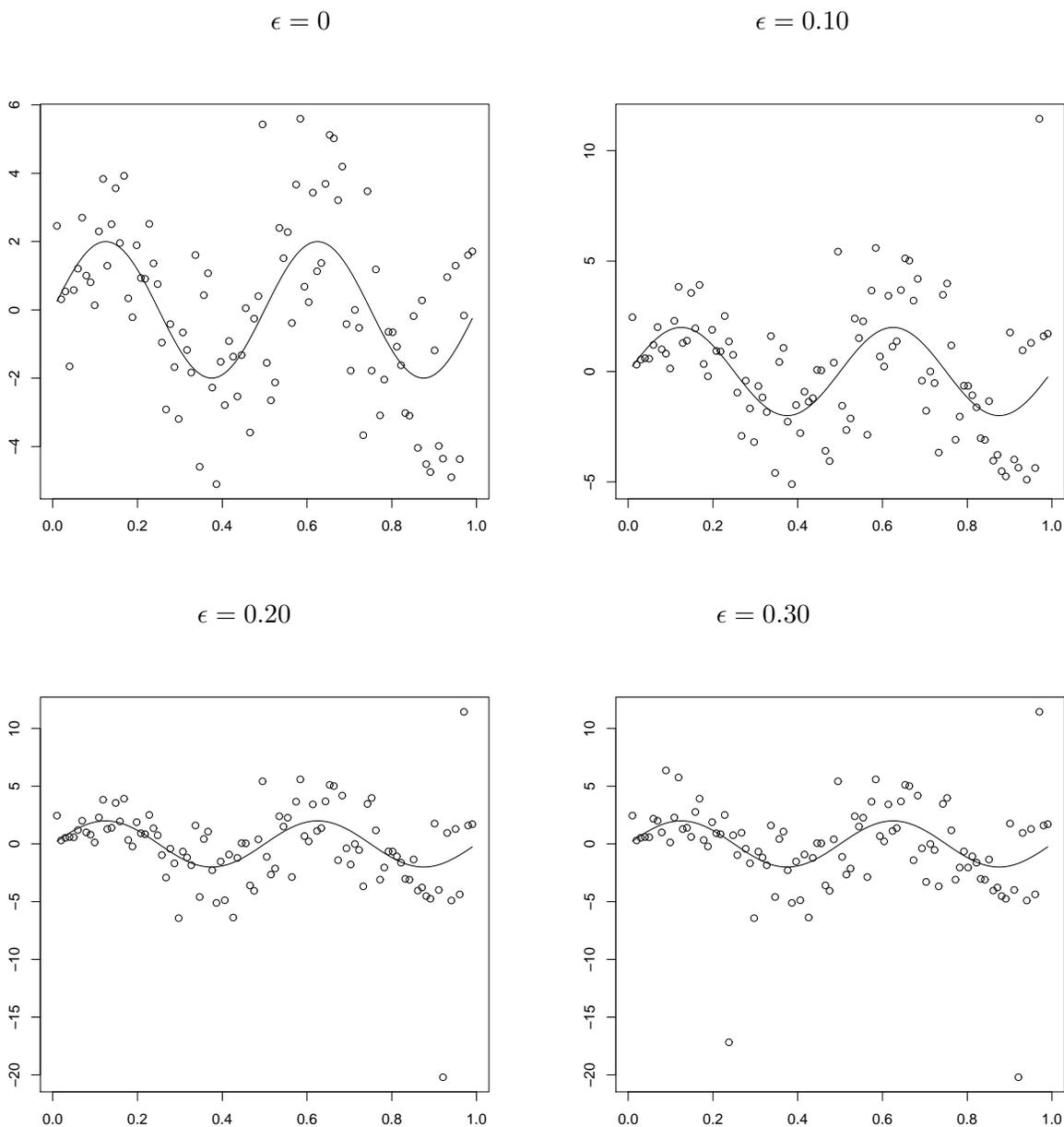


Figura B.7: Valores de las respuestas versus puntos del diseño del modelo $M1$ con funciones componentes $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.

Modelo Heteroscedástico M2.

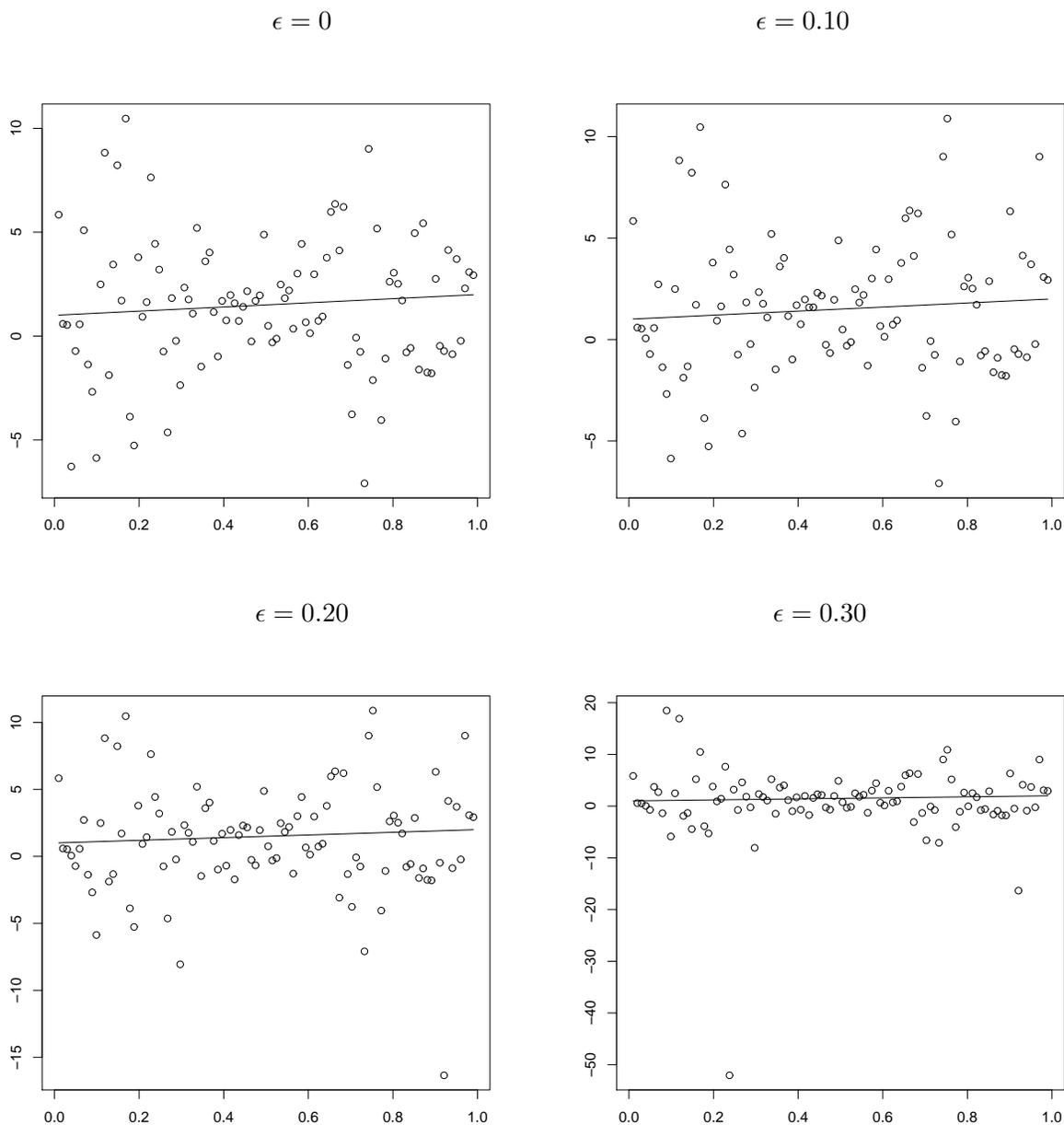


Figura B.8: Valores de las respuestas versus puntos del diseño del modelo $M2$ con funciones componentes $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.

Modelo Heteroscedástico M3.

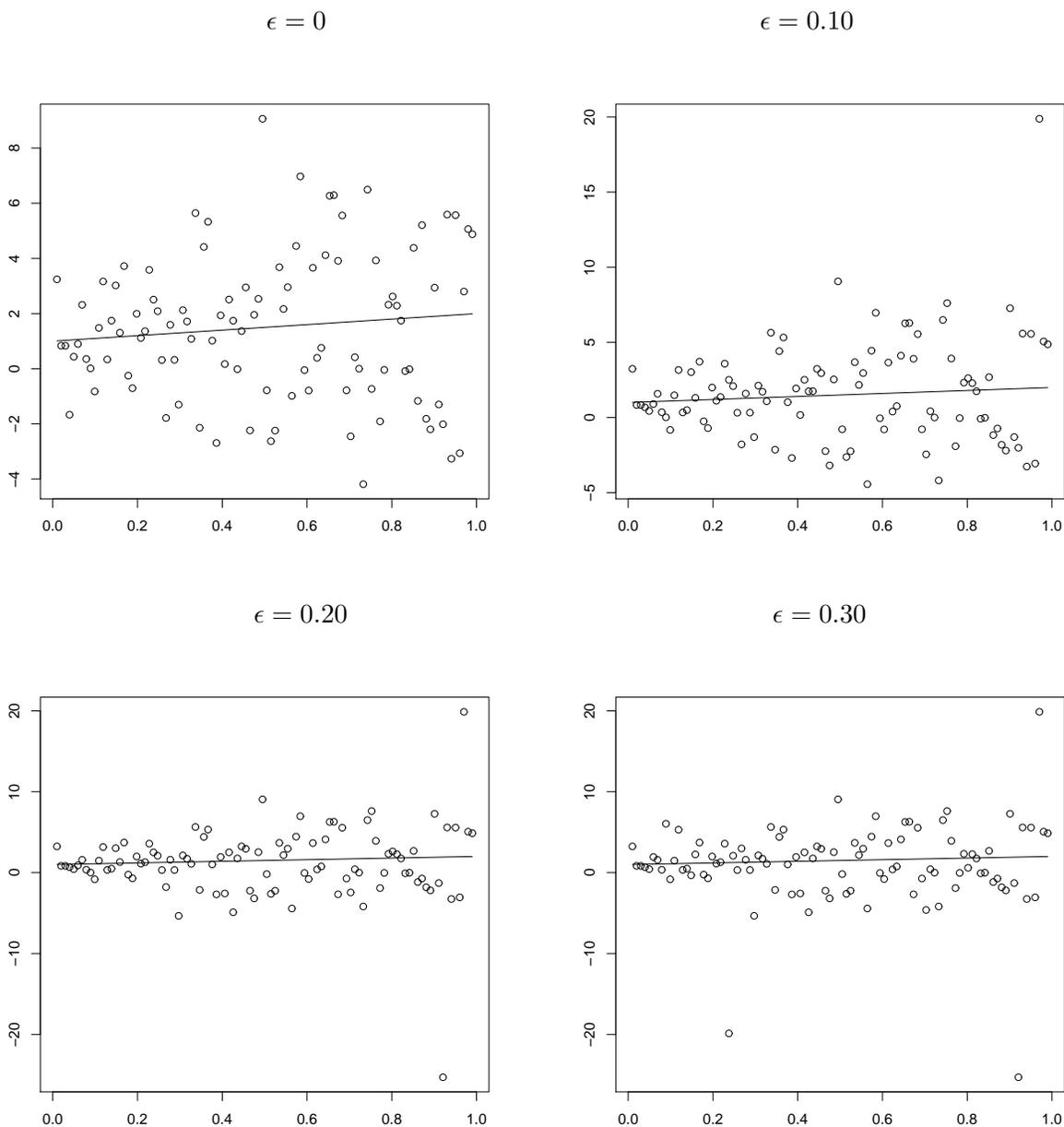


Figura B.9: Valores de las respuestas versus puntos del diseño del modelo $M3$ con funciones componentes $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.

Modelo de Escala Global.

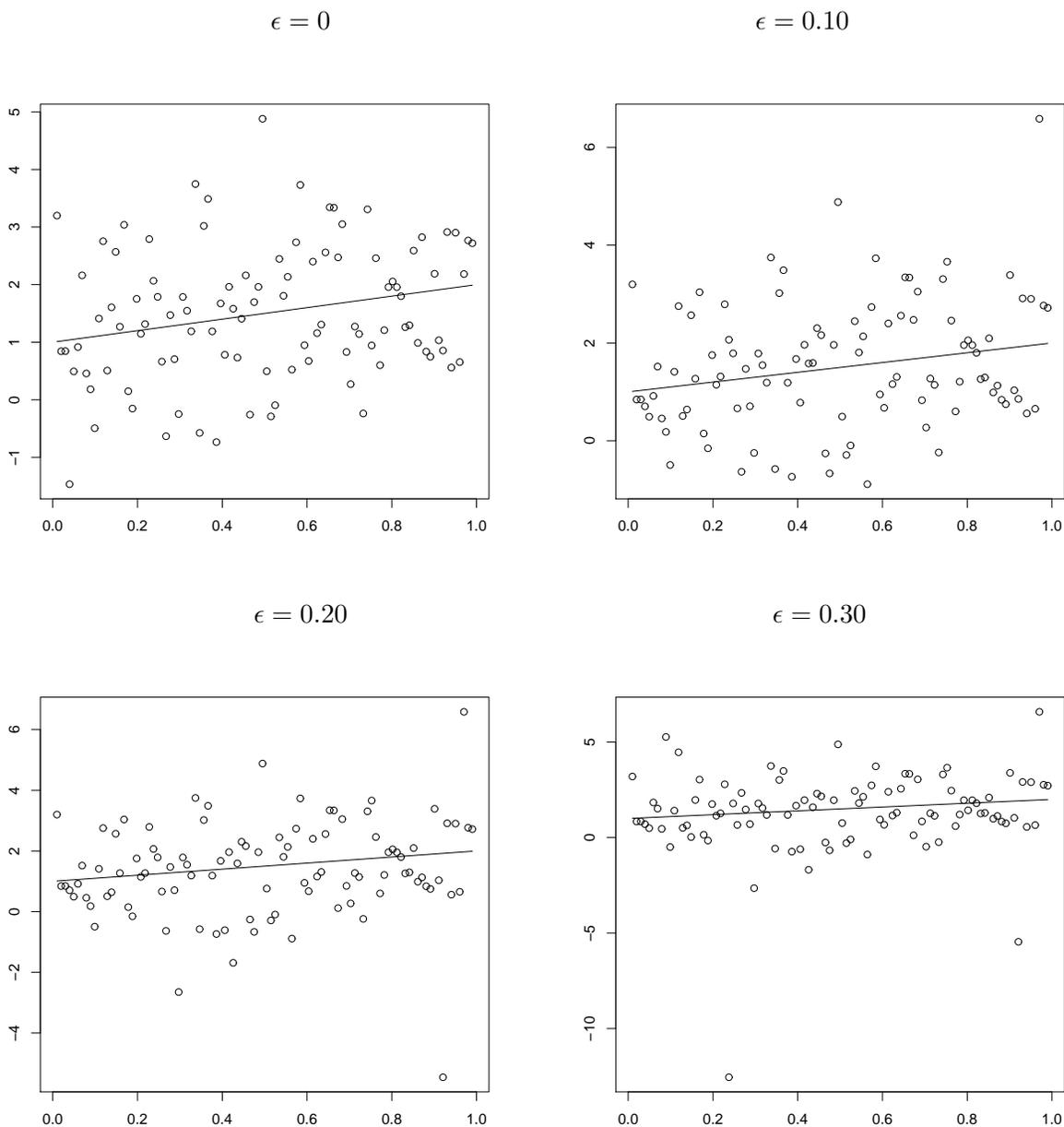


Figura B.10: Valores de las respuestas versus puntos del diseño del M.E.G. con funciones componentes $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = 1$. La curva corresponde al gráfico de la función g . Contaminación Simétrica.

*Gráficos de las densidades (estimadas) del Error Cuadrático Integrado en
escala logarítmica (E.C.I.L.).*

Gráficos de las densidades (estimadas) del Error Cuadrático Integrado en escala logarítmica para modelos con contaminación simétrica.

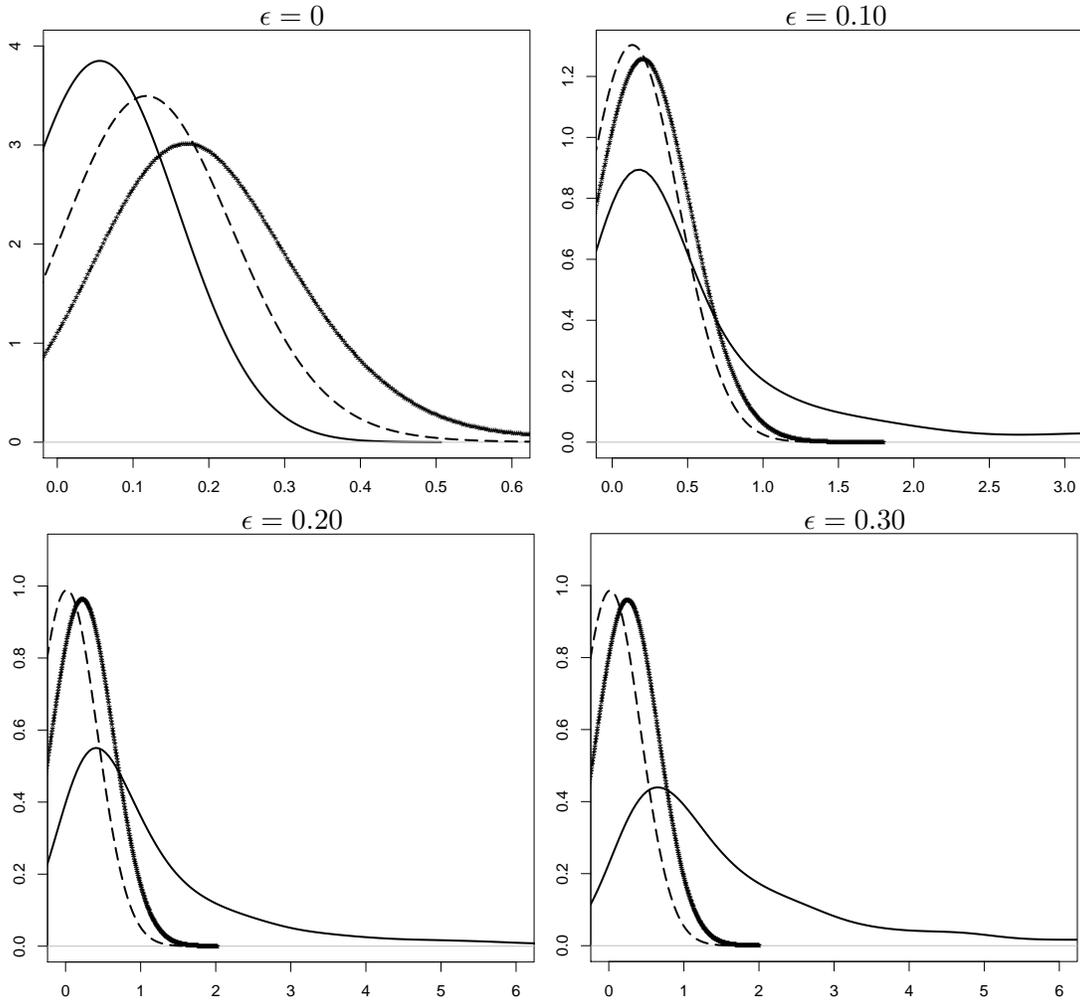


Figura B.11: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{\text{BFM},n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{\text{BT},n}$.

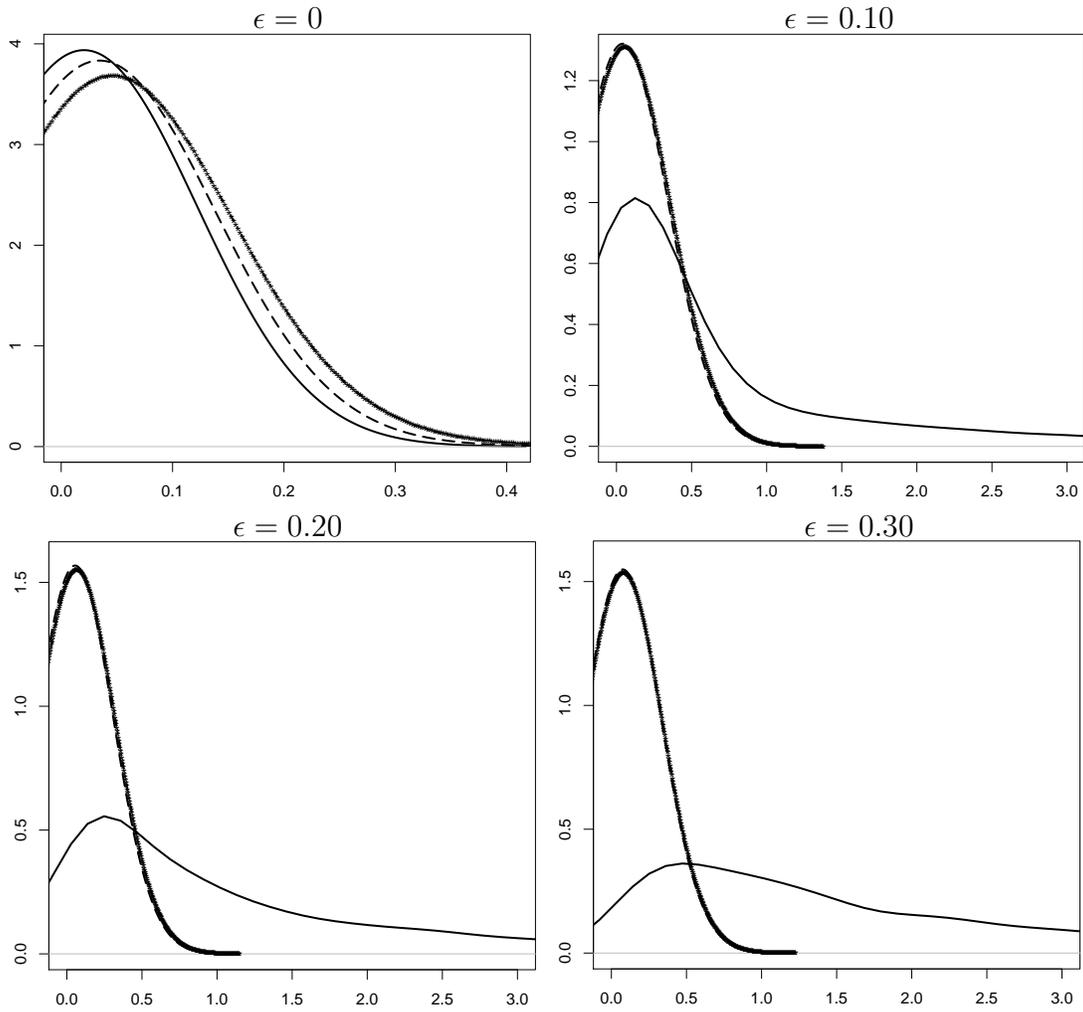


Figura B.12: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

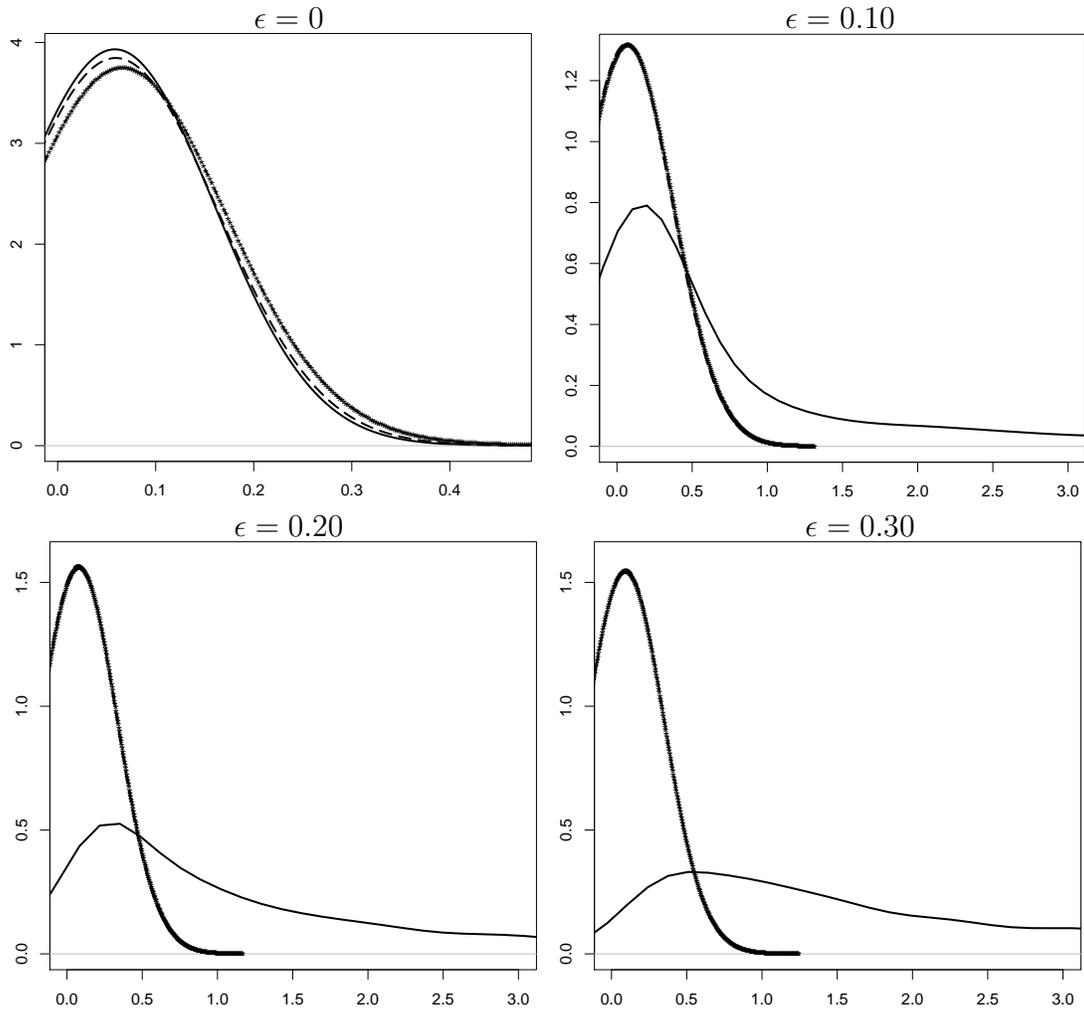


Figura B.13: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

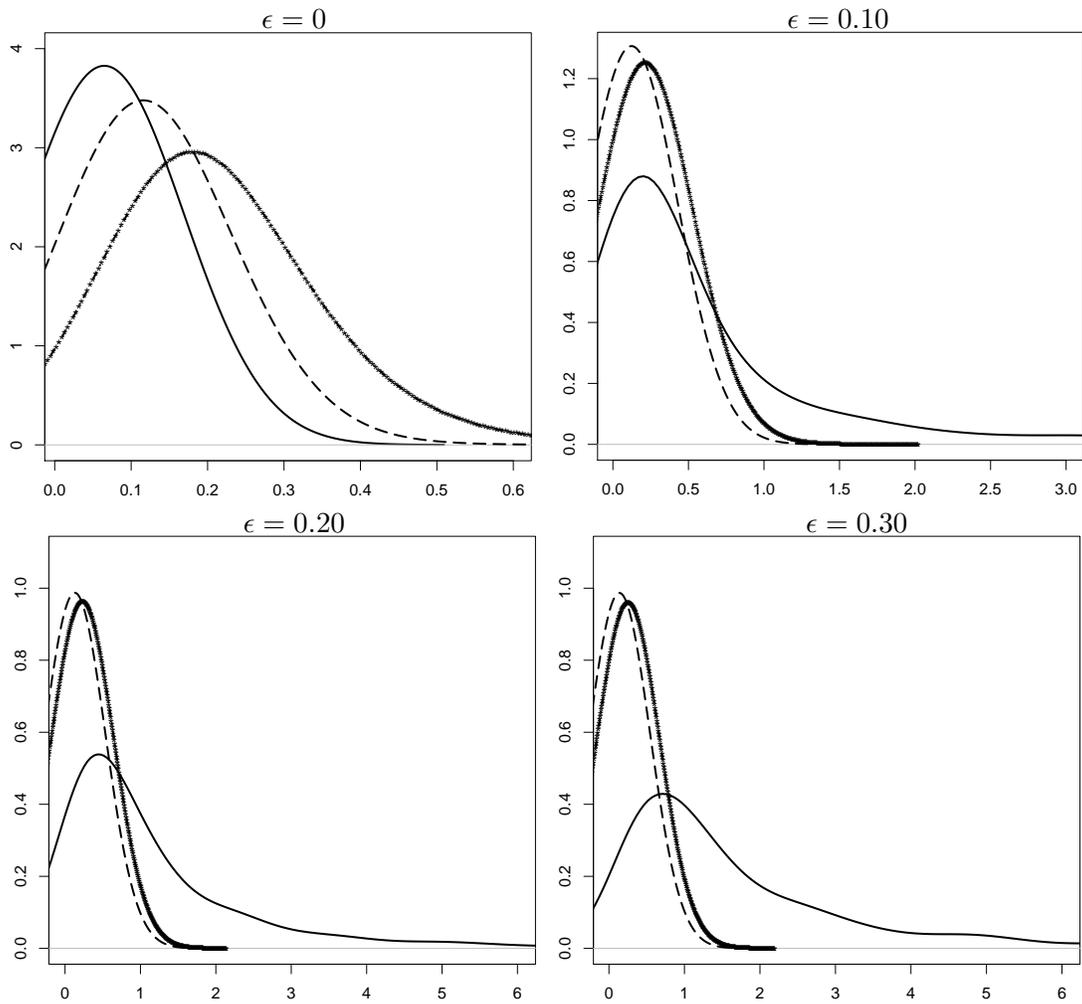


Figura B.14: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

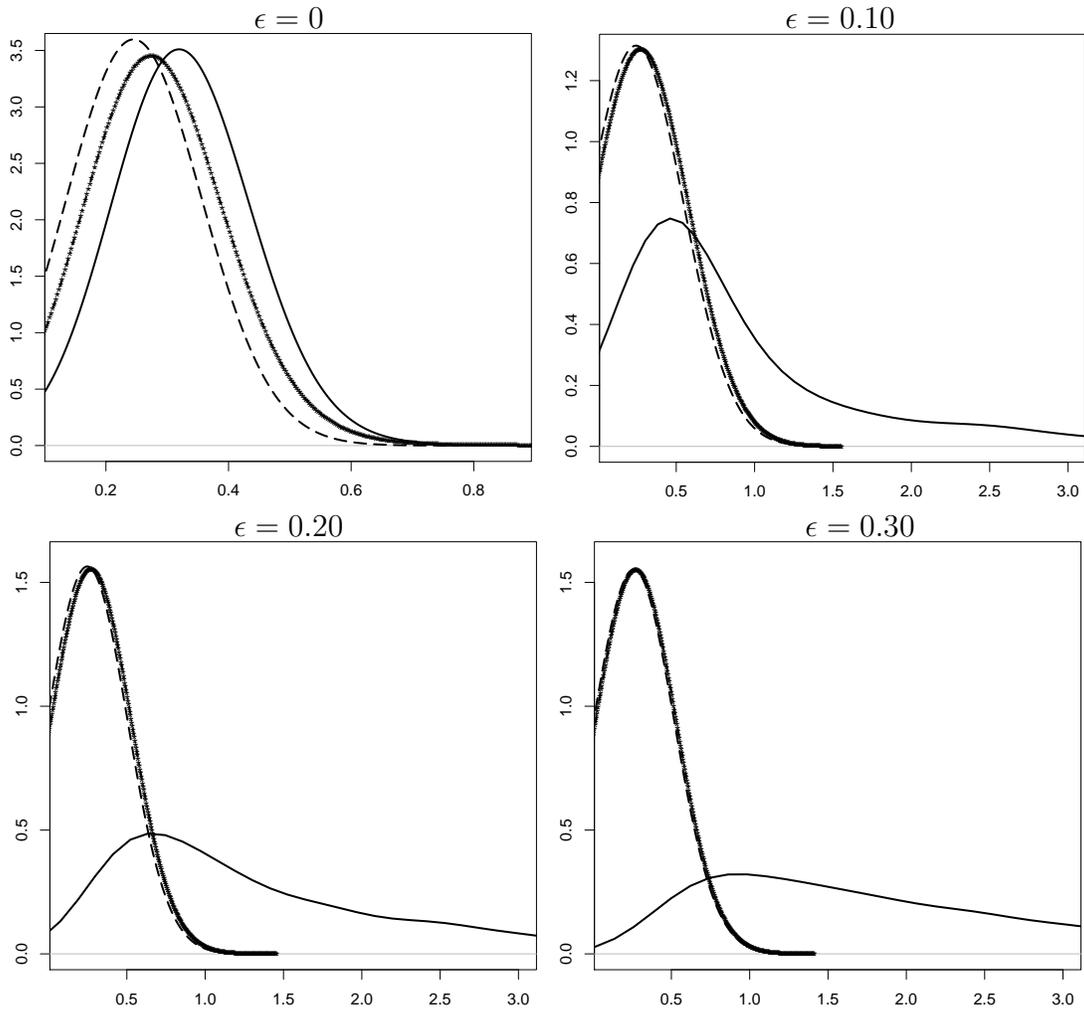


Figura B.15: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

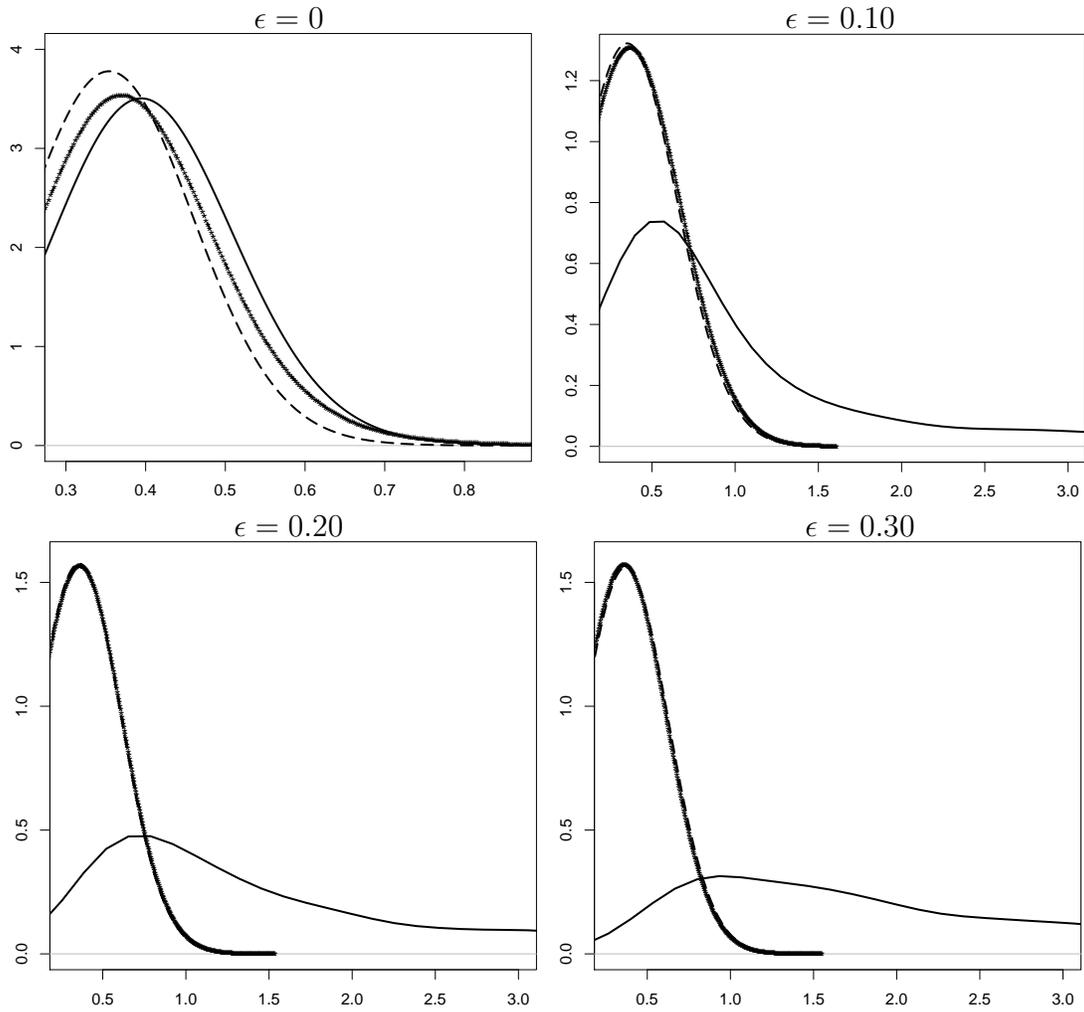


Figura B.16: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \sin(10x)]^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

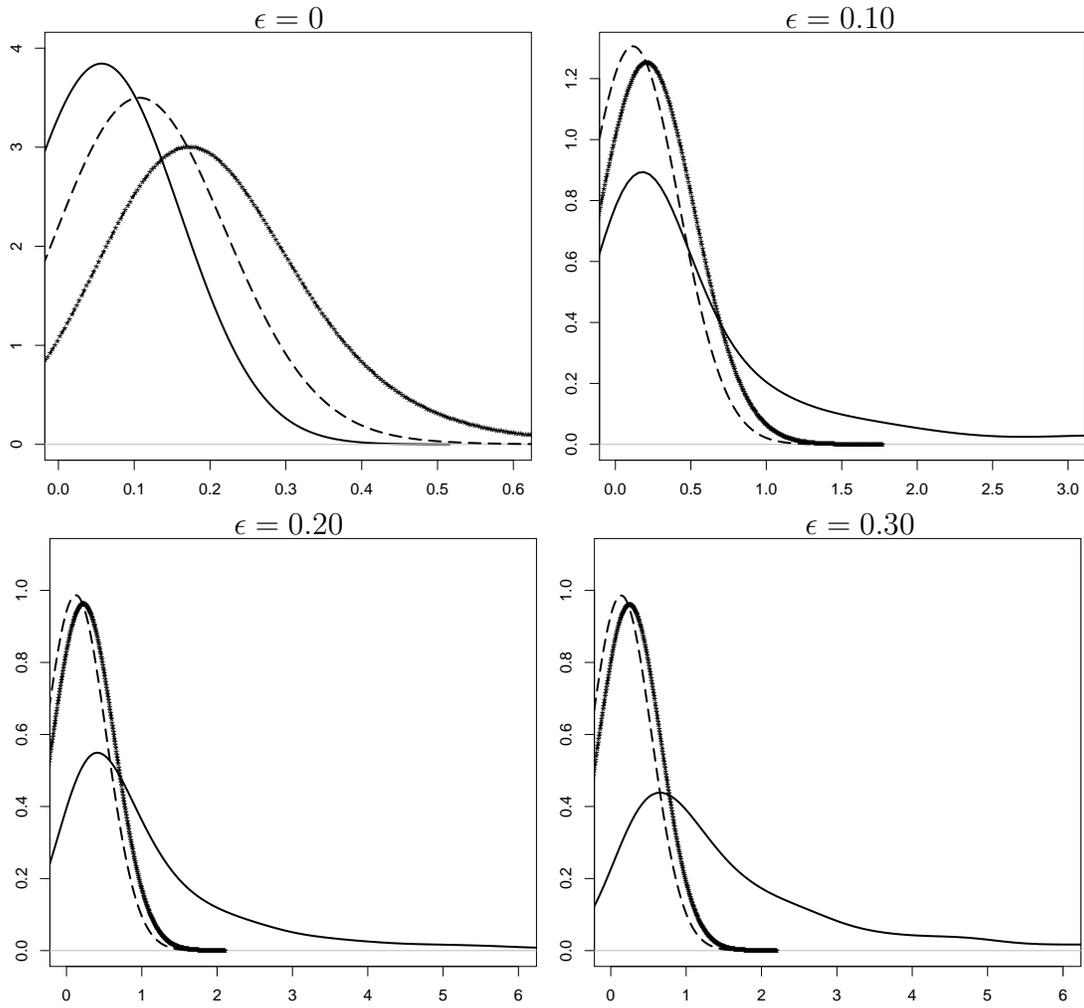


Figura B.17: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

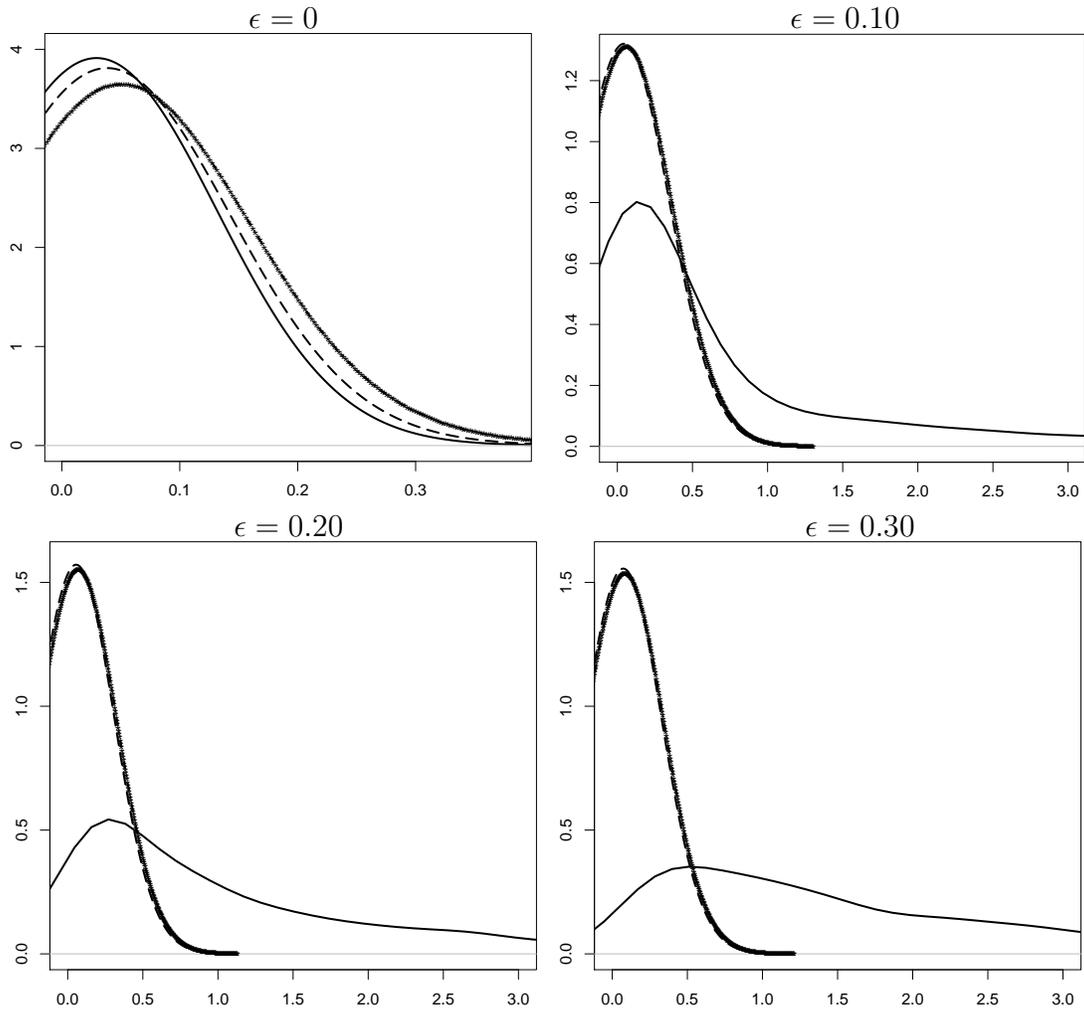


Figura B.18: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

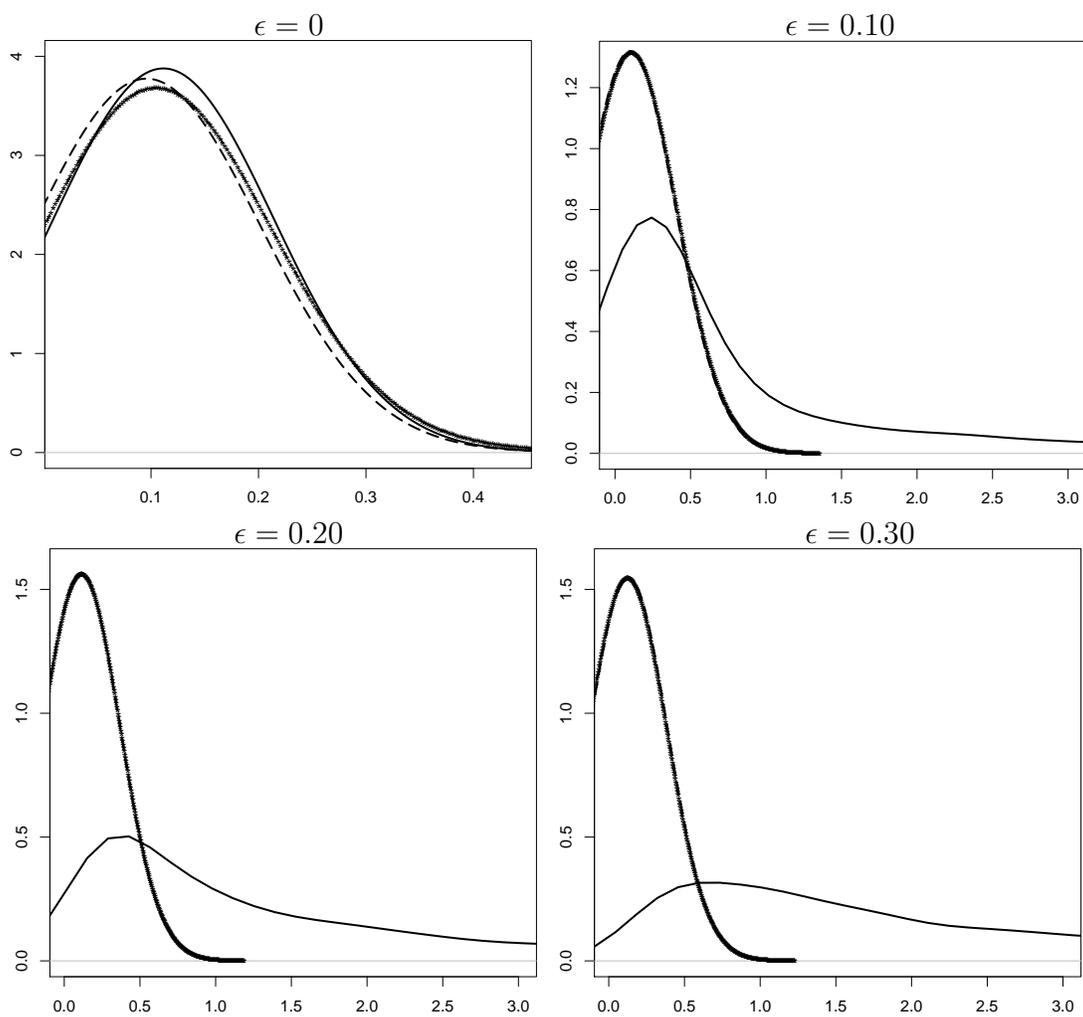


Figura B.19: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

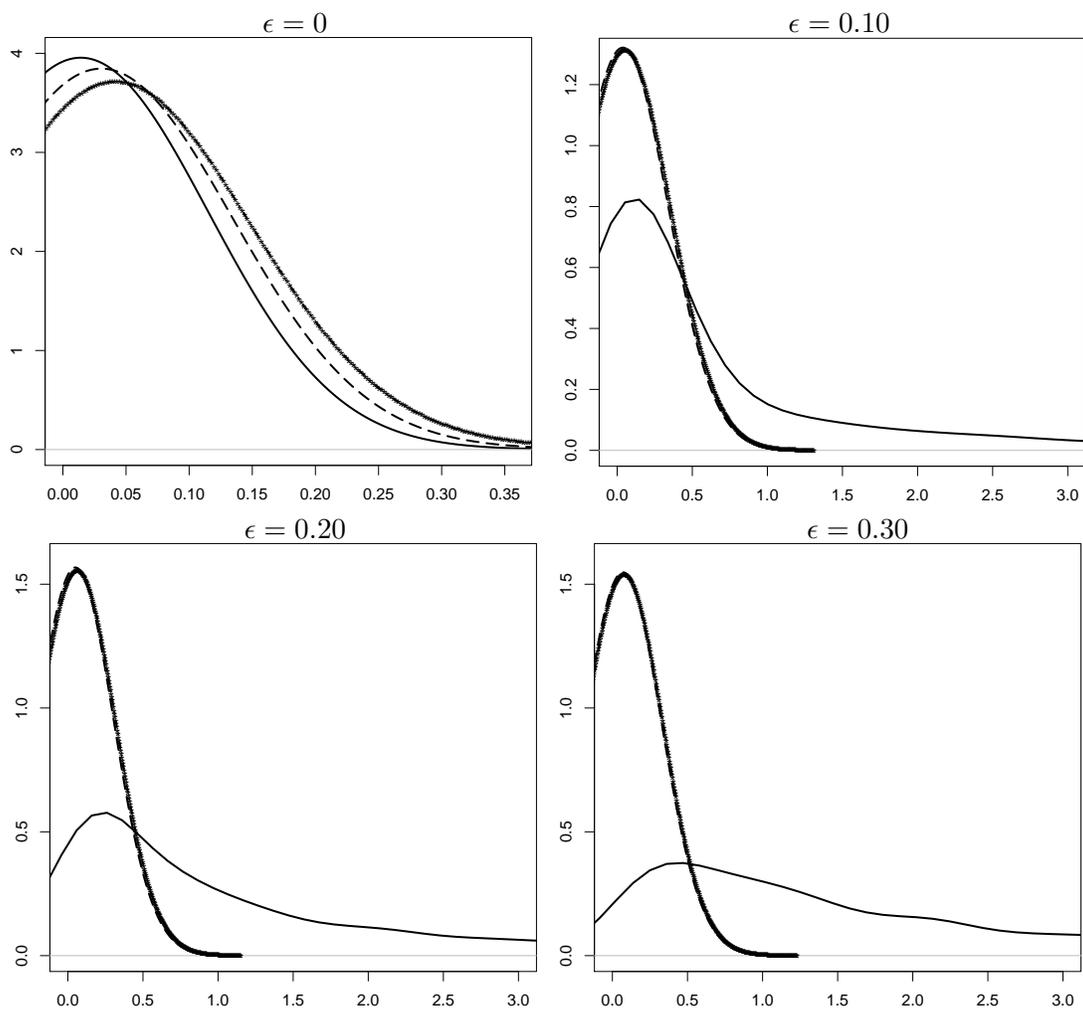


Figura B.20: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo M.E.G. (con $g(x) = 1 + x$ y $\sigma(x) \equiv 1$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

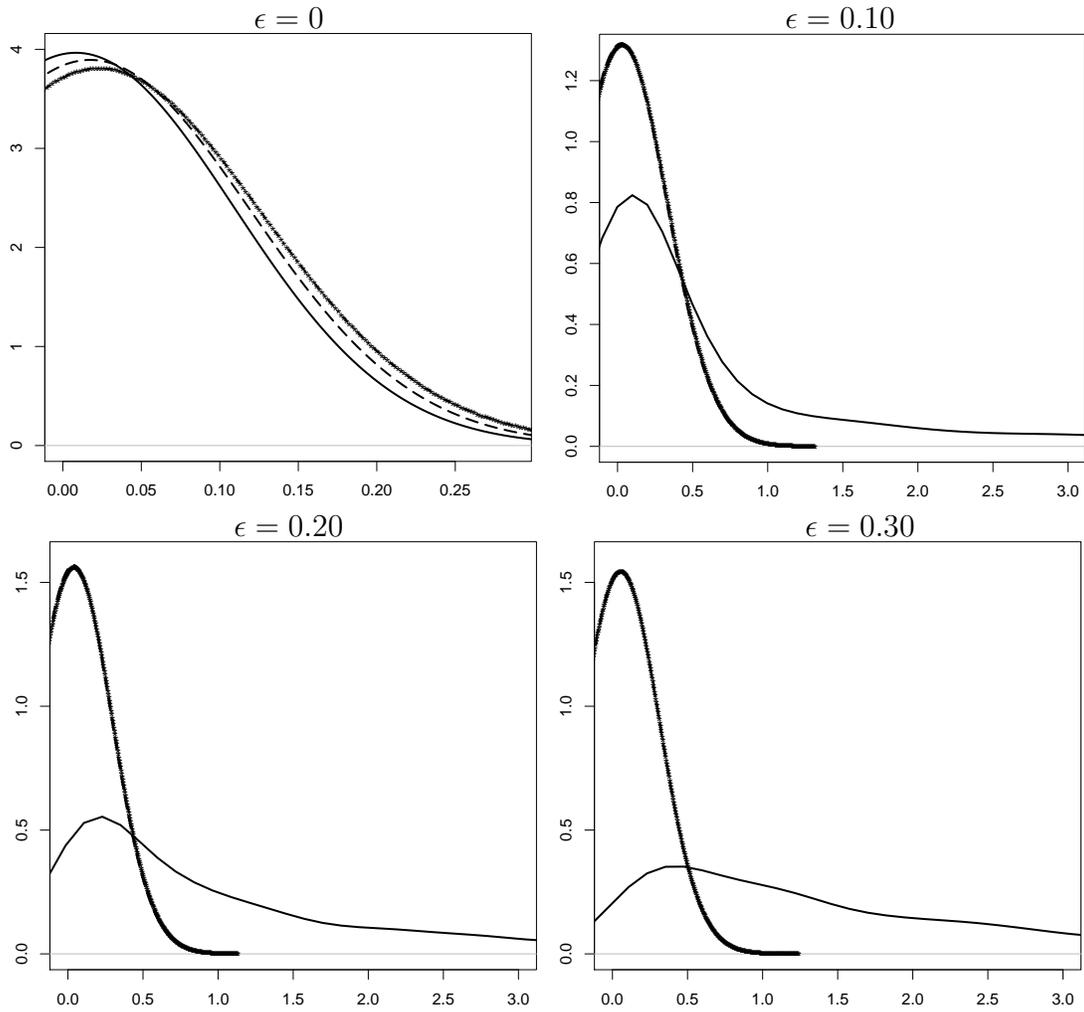


Figura B.21: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo M.E.G. (con $g(x) = 1 + x$ y $\sigma(x) \equiv 1$), $n = 100$ y contaminación simétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

Gráficos de las densidades (estimadas) del Error Cuadrático Integrado en escala logarítmica para modelos con contaminación asimétrica.

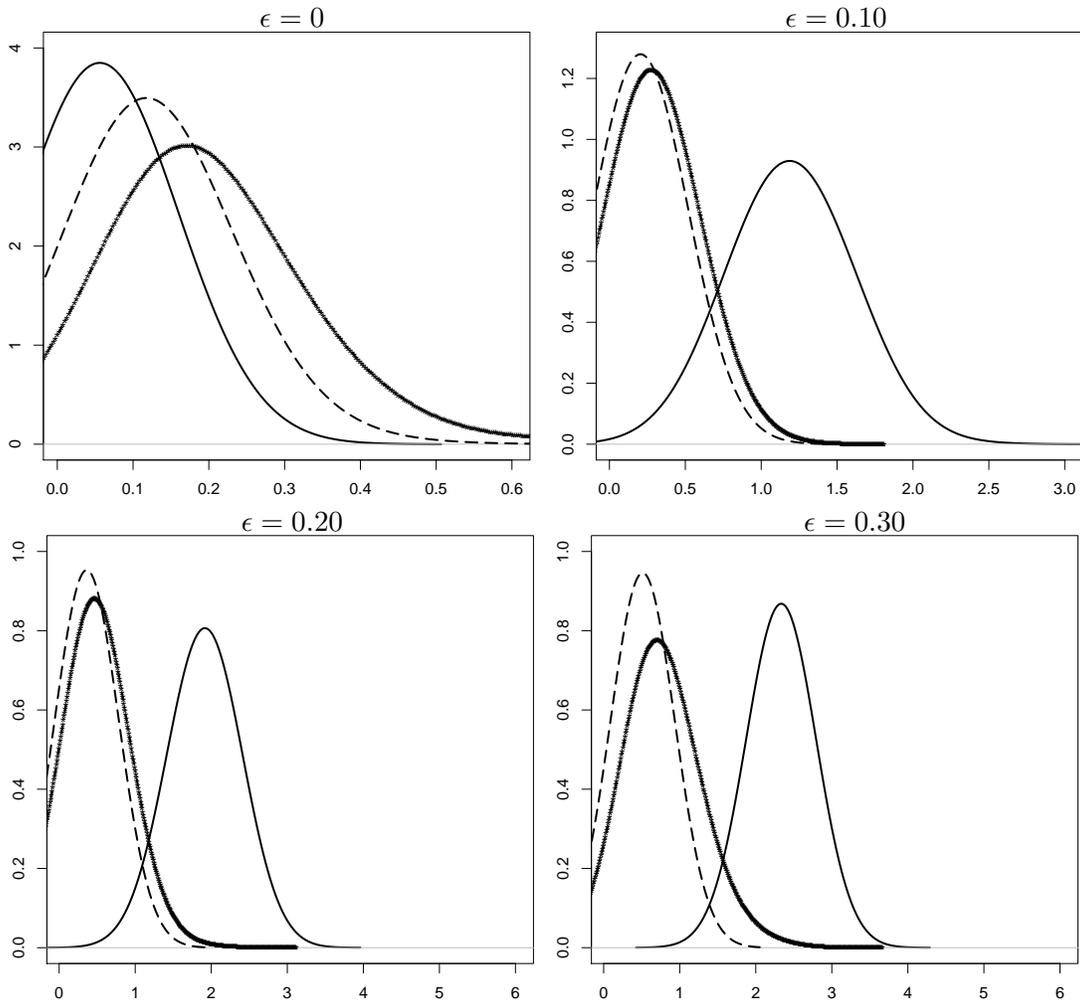


Figura B.22: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

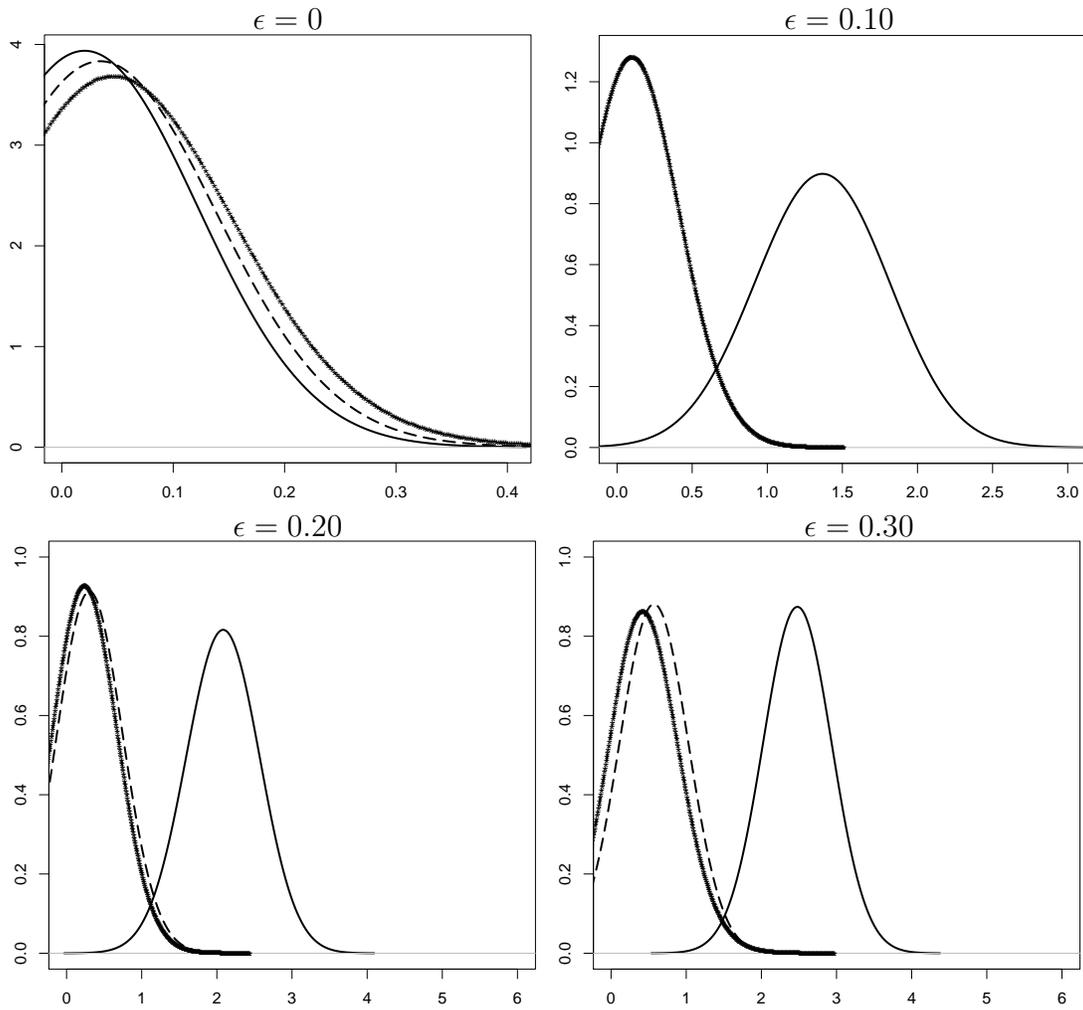


Figura B.23: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

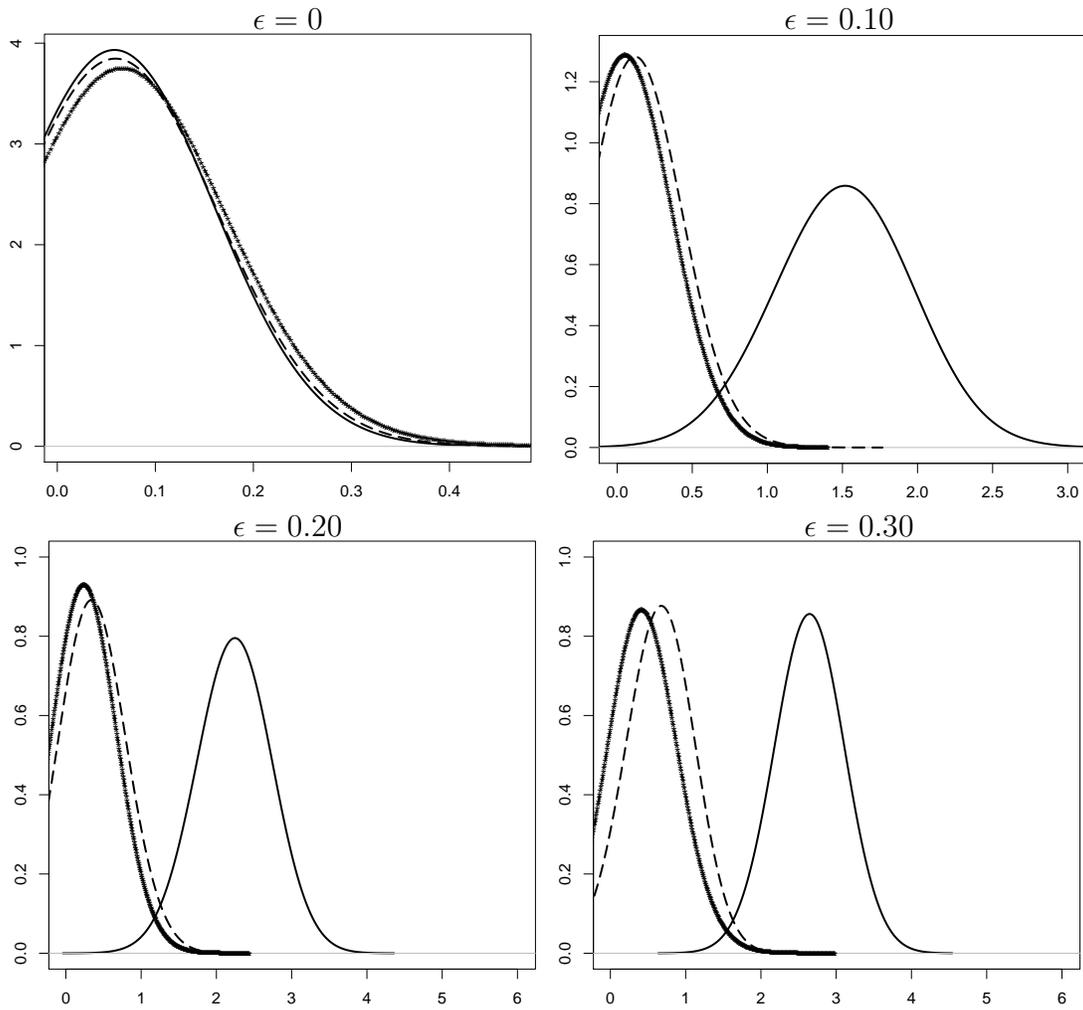


Figura B.24: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M1$ (con $g(x) = 2\text{sen}(4\pi x)$ y $\sigma(x) = \exp(x)$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

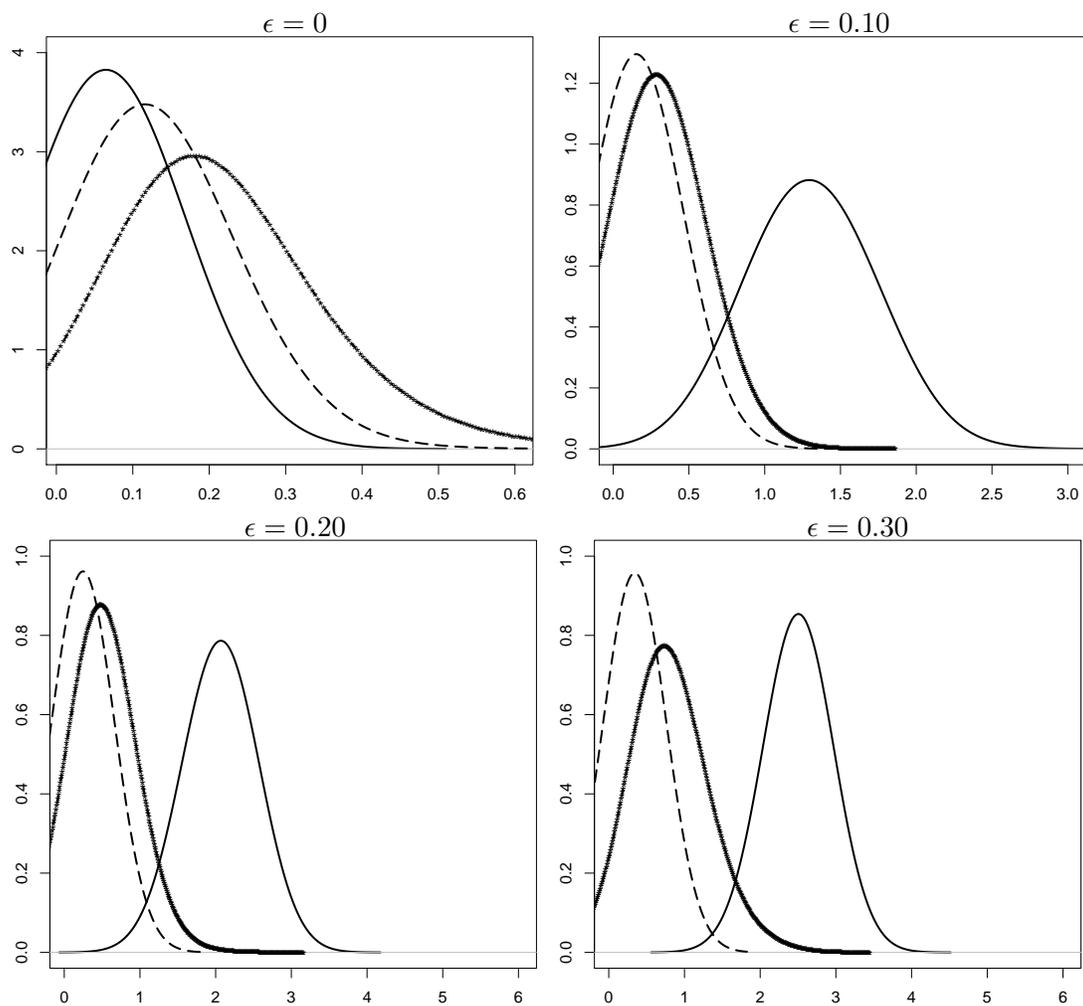


Figura B.25: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

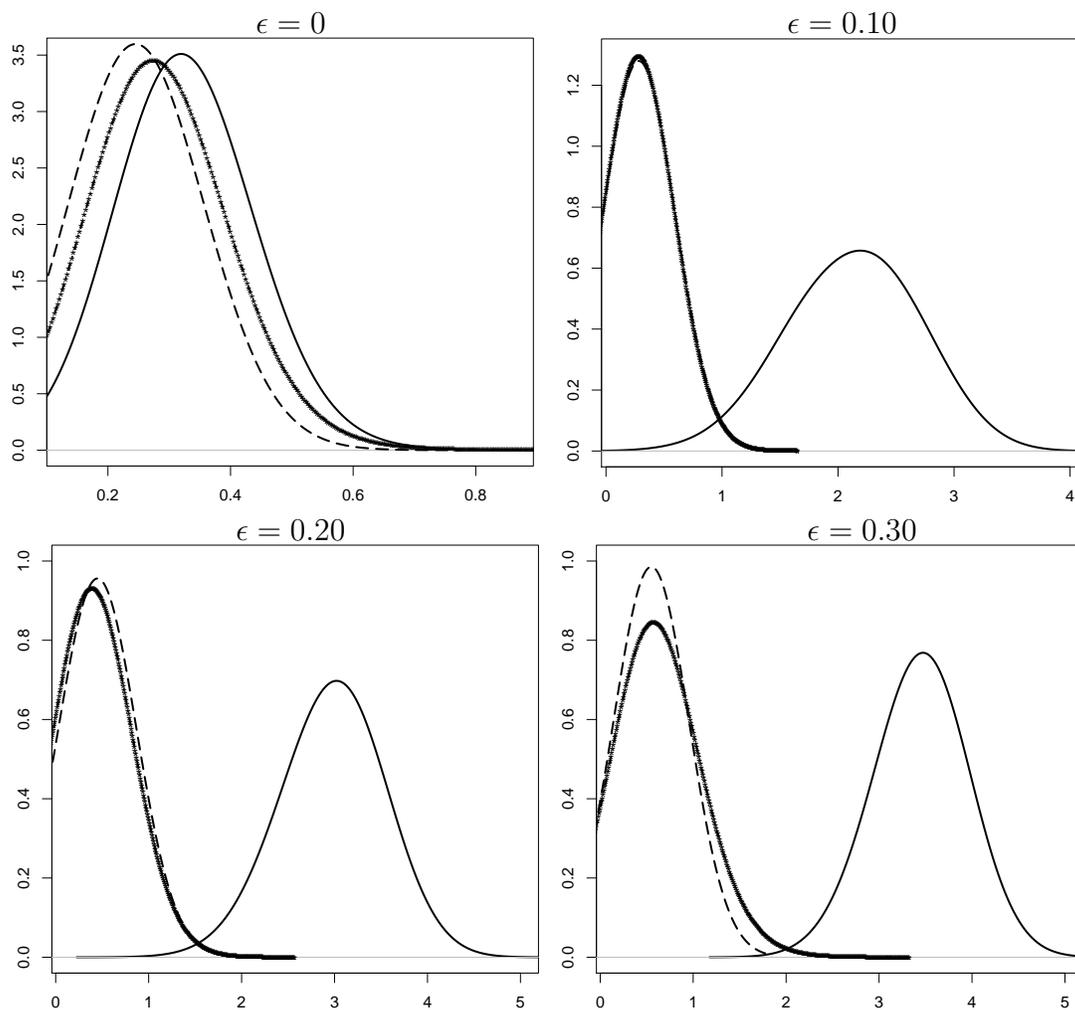


Figura B.26: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

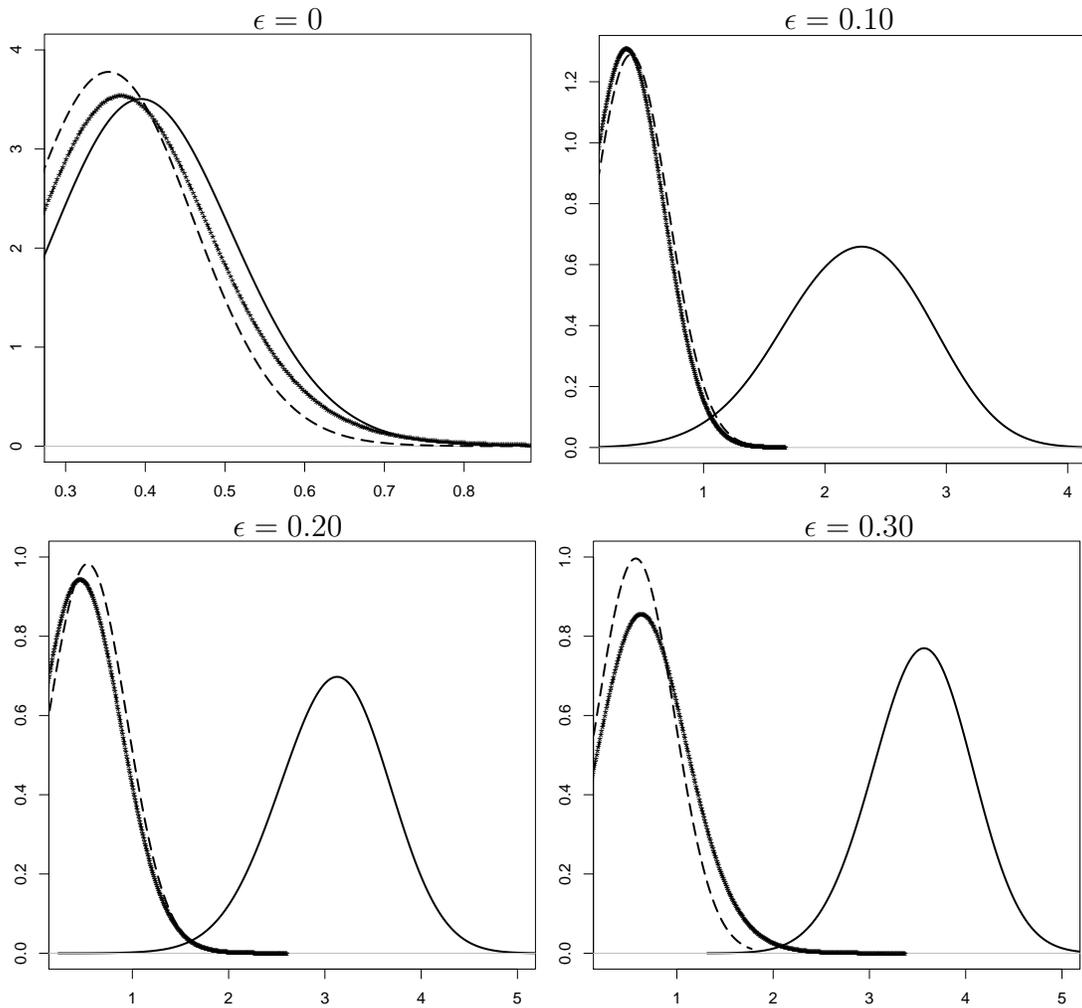


Figura B.27: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M2$ (con $g(x) = 1 + x$ y $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (--) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

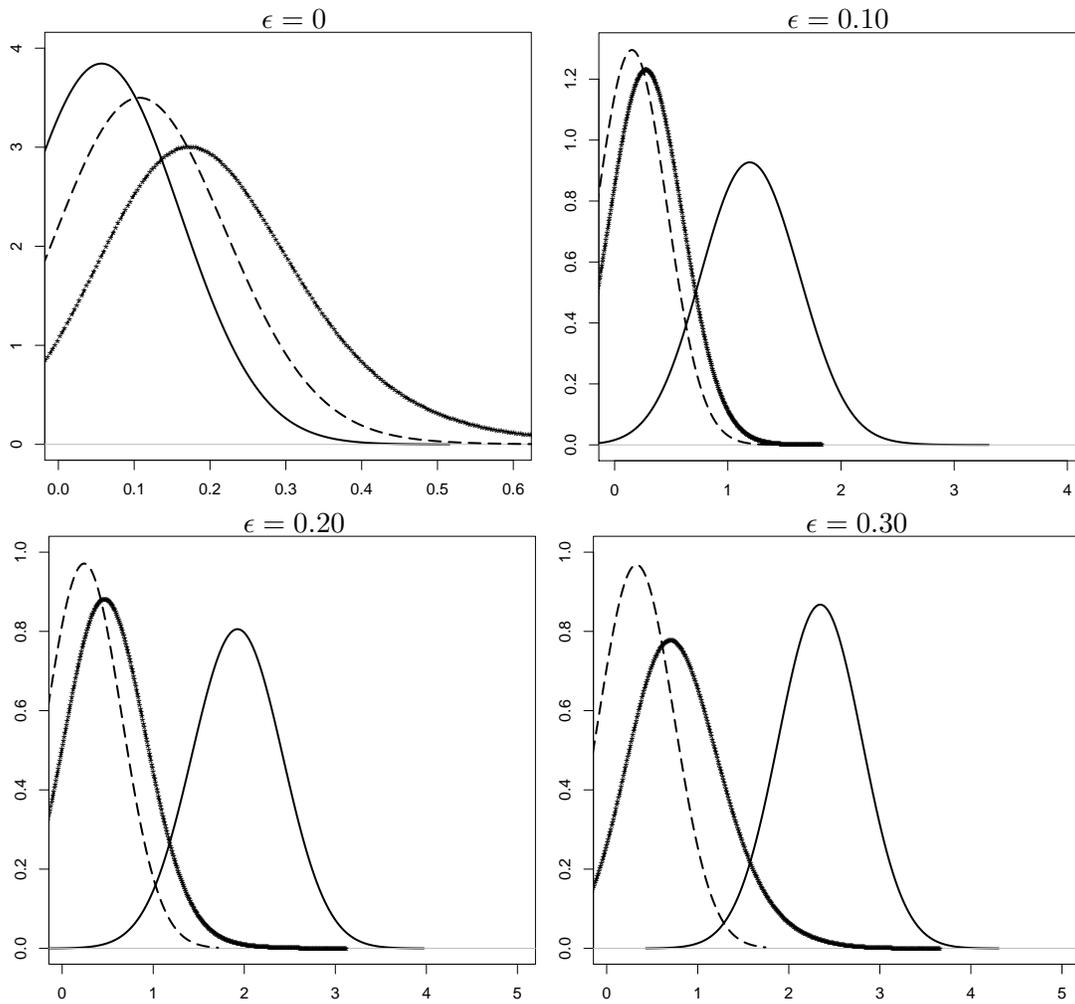


Figura B.28: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.04$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

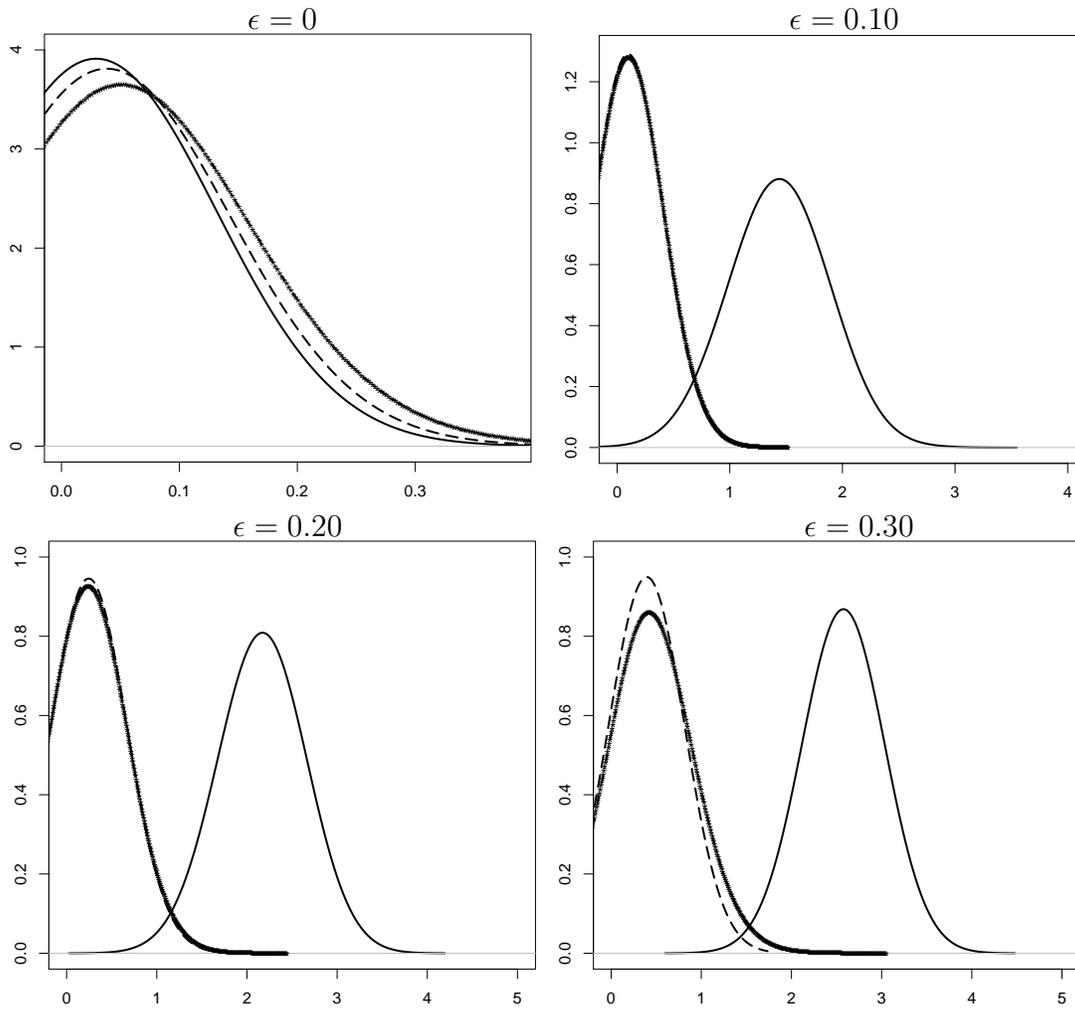


Figura B.29: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.2$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

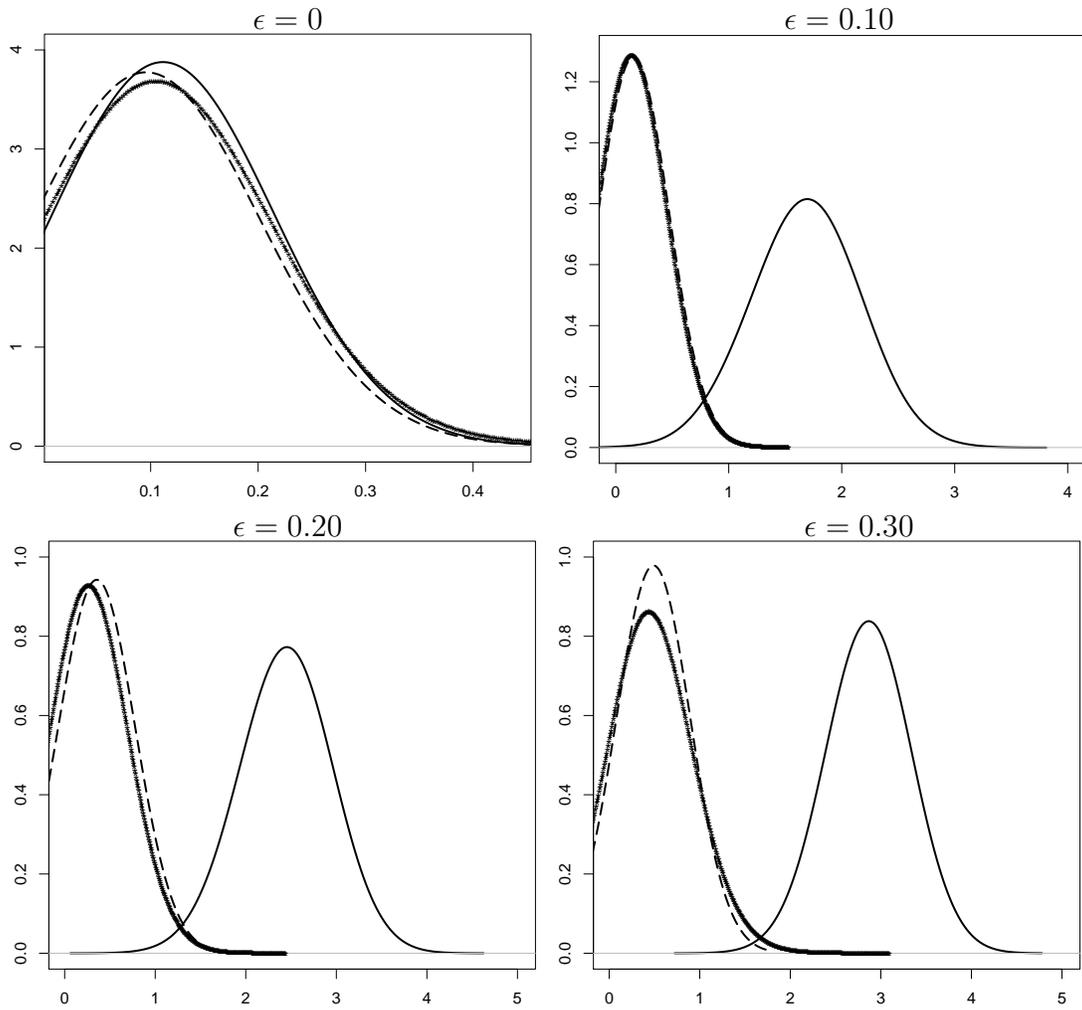


Figura B.30: Densidades estimadas del E.C.I.L. para la abertura de ventana $h = 0.5$ en el modelo $M3$ (con $g(x) = 1 + x$ y $\sigma(x) = (1 + x)^2$), $n = 100$ y contaminación asimétrica. La línea sólida de trazo fino corresponde a $\hat{\sigma}_{R,n}$, la línea cortada (---) a $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso a $\hat{\sigma}_{BT,n}$.

Gráficos del Error Cuadrático Medio en escala logarítmica (E.C.M.L.) para el modelo heteroscedástico M2.

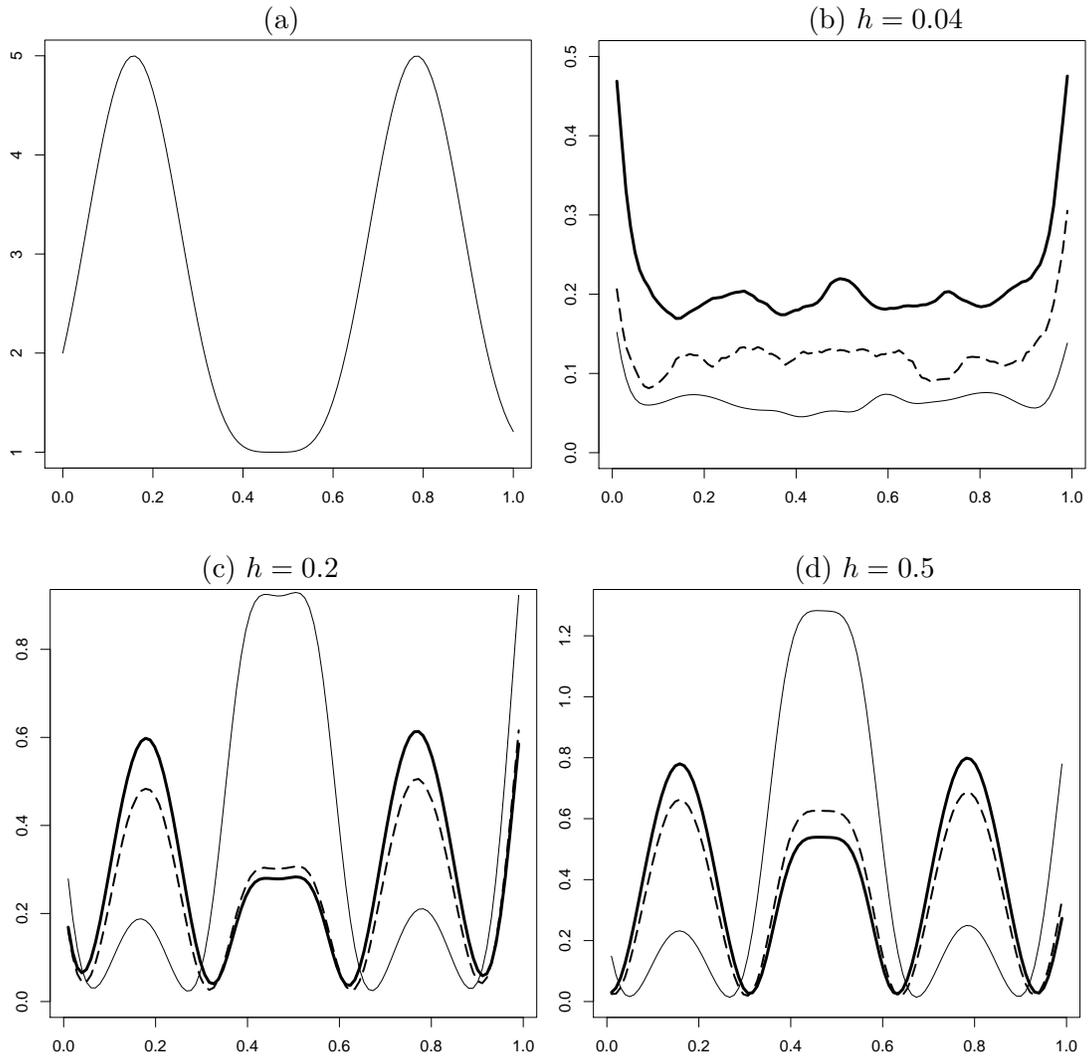


Figura B.31: Modelo $M2$. (a) Gráfico de $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. (b), (c) y (d) corresponden a los gráficos de E.C.M.L. (x_i, h) versus x_i para $h = 0.04, 0.2$ y 0.5 , respectivamente con $\epsilon = 0$. La línea sólida de trazo fino corresponde al E.C.M.L. de $\hat{\sigma}_{R,n}$, la línea cortada (---) al de $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso al de $\hat{\sigma}_{BT,n}$.

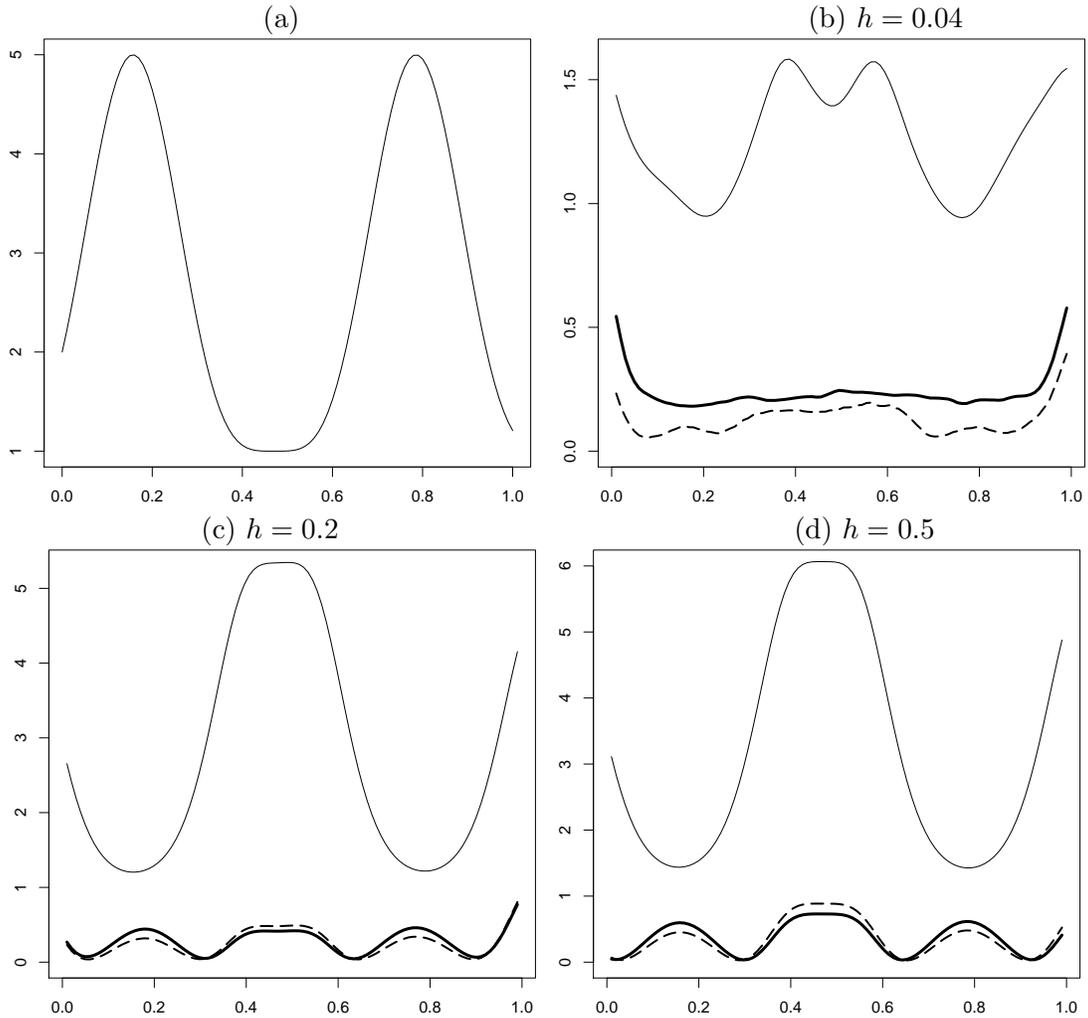


Figura B.32: Modelo $M2$. (a) Gráfico de $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. (b), (c) y (d) corresponden a los gráficos de E.C.M.L. (x_i, h) versus x_i para $h = 0.04, 0.2$ y 0.5 , respectivamente con contaminación simétrica y $\epsilon = 0.20$. La línea sólida de trazo fino corresponde al E.C.M.L. de $\hat{\sigma}_{R,n}$, la línea cortada (---) al de $\hat{\sigma}_{BFM,n}$ y la línea sólida de trazo grueso al de $\hat{\sigma}_{BT,n}$.

Bibliografía

- [1] Beaton, A. y Tukey, J. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*. **16**, 147–185.
- [2] Billingsley, P. (1979). *Probability and Measure*. John Wiley & Sons.
- [3] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons.
- [4] Boente, G. y Fraiman, R. (1989). Robust Nonparametric Regression Estimation for Dependent Observations. *The Annals of Statistics*. **17**, 1242–1256.
- [5] Boente, G. y Fraiman, R. (1995). Asymptotic distribution of data-driven smoothers in density and regression estimation under dependence. *The Canadian Journal of Statistics*. **23**, 383–397.
- [6] Boente, G., Fraiman, R. y Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *Journal of Statistical Planning and Inference*. **57**, 109–142.
- [7] Brown, L. y Levine, M. (2007). Variance Estimation in Nonparametric Regression via the Difference Sequence Method. *The Annals of Statistics*. **35**, 2219–2232.

- [8] Buckley, M., Eagleson, G. y Silverman, B. (1988). The Estimation of Residual Variance in Nonparametric Regression. *Biometrika*. **75**, 189–199.
- [9] Cantoni, E. y Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*. **11**, 141–146.
- [10] Carter, C. y Eagleson, G. (1992). A comparison of variance estimators in nonparametric regression. *Journal of the Royal Statistical Society, Series B*. **54**, 773–780.
- [11] Chung, K. (1974). *A course in Probability Theory*. Academic Press.
- [12] Clark, R. (1975). A calibration curve for radiocarbon dates. *Antiquity*. **49**, 251–266.
- [13] Dette, H., Munk, A. y Wagner, T. (1998). Estimating the variance in nonparametric regression - What is a reasonable choice?. *Journal of the Royal Statistics Society, Series B*. **60**, 751–764.
- [14] Gasser, T., Sroka, L. y Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*. **73**, 625–633.
- [15] Ghement, I., Ruiz, M. y Zamar, R. (2007). Robust Estimation of Error Scale in Nonparametric Regression Models. A aparecer en *Journal of Statistical Planning and Inference*.
- [16] Hall, P. y Carroll, R. (1989). Variance Function Estimation in Regression: The Effect of Estimating the Mean. *Journal of the Royal Statistical Society, Series B*. **51**, 3–14.
- [17] Hall, P. y Marron, J. (1990). On variance estimation in nonparametric regression. *Biometrika*. **77**, 415–419.

- [18] Hall, P., Kay, J. y Titterington, D. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*. **77**, 521–528.
- [19] Hampel, F., Ronchetti, E., Rousseeuw, P. y Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons.
- [20] Hannig, J. y Lee, T. (2006). Robust SiZer for Exploration of Regression Structures and Outlier Detection. *Journal of Computational and Graphical Statistics*. **15**, 101–117.
- [21] Härdle, W. y Gasser, T. (1984). Robust nonparametric function fitting. *Journal of the Royal Statistical Society, Series B*. **46**, 42–51.
- [22] Härdle, W. y Tsybakov, A. (1988). Robust Nonparametric Regression with Simultaneous Scale Curve Estimation. *The Annals of Statistics*. **25**, 443–456.
- [23] Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge. University Press.
- [24] Härdle, W., Müller, M., Sperlich, S. y Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer.
- [25] Holst, U., Hössjer, O., Björklund, C., Ragnarson, P. y Edner, H. (1996). Locally Weighted Least Squares Kernel Regression and Statistical Evaluation of LIDAR Measurements. *Environmetrics*. **7**, 401–406.
- [26] Huber, P. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*. **35**, 73–101.
- [27] Huber, P. (1981). *Robust Statistics*. John Wiley & Sons.

- [28] Leung, D. (2005). Cross-validation in nonparametric regression with outliers. *Annals of Statistics*. **33**, 2291–2310.
- [29] Leung, D., Marriott, F. y Wu, E. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics*. **2**, 333–339.
- [30] Maronna, R., Martin, D. y Yohai, V. (2006). *Robust Statistics: Theory and Methods*. John Wiley & Sons.
- [31] Martin, D. y Zamar, R. (1989). Asymptotically Min-Max Bias Robust M –Estimates of Scale for Positive Random Variables. *Journal of American Statistical Association*. **84**, 494–501.
- [32] Martin, D. y Zamar, R. (1993). Bias Robust Estimation of Scale. *The Annals of Statistics*. **21**, 494–501.
- [33] Müller, H. y Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *The Annals of Statistics*. **15**, 610–625.
- [34] Müller, H. y Stadtmüller, U. (1987). Variable bandwidth kernel estimators of regression. *The Annals of Statistics*. **15**, 182–201.
- [35] Pelligrad, M. y Utev, S. (1997). Bandwidth Choice for Nonparametric Regression. *The Annals of Statistics*. **25**, 443–456.

- [36] Salibian-Barrera, M. (2000). Contributions to the Theory of Robust Inference. *Ph.D. Thesis*. Department of Statistics, University of British Columbia, Vancouver, Canada. Disponible en <http://mathstat.carleton.ca/~matias>.
- [37] Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *The Annals of Statistics*. **12**, 1215–1230.
- [38] Roussas, G., Tran, L. e Ionanides, D.A. (1992). Fixed Design Regression for Time Series: Asymptotic Normality. *Journal of Multivariate Analysis*. **40**, 262–291.
- [39] Van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge. University Press.
- [40] Wahba, G. y Wold, S. (1975). A completely automatic french curve: fitting spline functions by cross validation. *Communications in Statistics - Theory and Methods*. **4**, 1–17.
- [41] Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B*. **40**, 364–372.
- [42] Wang F. y Scott D. (1994). The L1 method for robust nonparametric regression. *Journal of American Statistical Association*. **89**, 65–76.
- [43] Woodroffe, M. (1970). On choosing a delta sequence. *Annals of Mathematical Statistics*. **41**, 1665–1671.