



UNIVERSIDADE DA CORUÑA
DEPARTAMENTO DE MATEMÁTICAS

**APLICACIÓN DE LA REGRESIÓN POLINÓMICA
LOCAL AL ANÁLISIS DISCRIMINANTE
Y ANÁLISIS CLUSTER DE SERIES DE TIEMPO**

Sonia Pértega Díaz

Diciembre 2007

Tesis Doctoral

UNIVERSIDADE DA CORUÑA
DEPARTAMENTO DE MATEMÁTICAS

APLICACIÓN DE LA REGRESIÓN POLINÓMICA
LOCAL AL ANÁLISIS DISCRIMINANTE
Y ANÁLISIS CLUSTER DE SERIES DE TIEMPO

Autor

Sonia Pértega Díaz

Director

José Antonio Vilar Fernández

MEMORIA PARA OPTAR Ó GRADO DE DOUTOR
NA UNIVERSIDADE DA CORUÑA

DICIEMBRE DE 2007

Realizado el acto público de defensa y mantenimiento de esta Tesis Doctoral el día 14 de Diciembre de 2007 en la Facultad de Informática de la Universidade da Coruña ante el tribunal formado por:

Presidente: Dr. D. Wenceslao González Manteiga
Vocales: Andrés M. Alonso Fernández
Pedro Saavedra Santana
José Ramón Fernández Bernárdez
Secretario: Juan Manuel Vilar Fernández

obtuvo la máxima calificación de SOBRESALIENTE CUM LAUDE, siendo director de la misma el Dr. D. José Antonio Vilar Fernández.

A mi madre.

Agradecimientos

*El que no considera lo que tiene como
la riqueza más grande, es desdichado,
aunque sea el dueño del mundo.*

Epicuro.

En primer lugar, desearía expresar mi más profundo agradecimiento al director de esta tesis, el profesor José Antonio Vilar Fernández, por su gran dedicación, interés y apoyo, sin el cual este trabajo nunca hubiese sido posible. Soy consciente del enorme esfuerzo que ha realizado, en más de una ocasión, para sacar este proyecto adelante.

Agraceder al resto de profesores del Departamento de Matemáticas de la Universidad de A Coruña, y en especial a sus directores durante estos años, los profesores Juan Vilar Fernández y Ricardo Cao Abad, el magnífico trato que me han dispensado en el Departamento, en el cual siempre me he sentido como en un segundo lugar de trabajo. Agradecerles también su inestimable ayuda y todos aquellos consejos que me han brindado durante el proceso de elaboración de esta tesis.

Gracias también a los profesores del Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela, con los que inicié mi formación en este área y que me ofrecieron la oportunidad de compartir con ellos parte de los resultados de esta Tesis.

Así mismo, quisiera agradecer el apoyo y la amistad de Elisa y Cris, y de todos aquellos que desde el Laboratorio 2.1 del Departamento de Matemáticas hicieron más amenas las tardes de trabajo que compartí con ellos.

A mis compañeros en el Hospital Juan Canalejo, por su ánimo y sus consejos. Gracias a todos ellos, sobre todo, por su amistad.

A María, Esperanza, Alicia(s) y Marta(s), por su amistad a lo largo de todos estos años. Me siento afortunada al saber que seguiréis ahí para compartir juntas todo lo que queda por venir, igual que lo hemos hecho hasta ahora.

A mi madre, porque sin duda es la persona de la que más he aprendido. Y a David, por haber aceptado compartir conmigo mis proyectos más importantes. A ellos, y a toda mi familia y amigos, en especial a Eladio y Mercedes, gracias por “soportar” el proceso de elaboración de esta tesis, por su ánimo y su cariño durante todo este tiempo.

Finalmente, quiero hacer constar que este trabajo ha sido subvencionado por los proyectos de investigación BFM2002-00265 (financiado por el Ministerio de Ciencia y Tecnología), PGIDIT03PXIC10505PN (financiado por la Xunta de Galicia), “Estimación no paramétrica de curvas bajo dependencia, censura o truncamiento. Aplicaciones termogravimétricas, biosanitarias y en sismología”, y por el proyecto MTM2005-00429, “Modelización, contrastes e inferencia no paramétrica: análisis de supervivencia, datos dependientes y aplicaciones”, financiado por el Ministerio de Ciencia y Tecnología.

Sonia Pértega Díaz

A Coruña, Octubre de 2007

Índice

Prólogo	1
1. Resultados preliminares	5
1.1. Estimación no paramétrica de la función de regresión: la regresión polinómica local.	5
1.1.1. Estimadores tipo núcleo de la función de regresión.	7
1.1.2. La regresión polinómica local ponderada.	8
1.1.3. La regresión polinómica local en el contexto de los modelos lineales generalizados.	14
1.2. Análisis espectral de series de tiempo.	20
1.2.1. La densidad espectral	20
1.2.2. El periodograma	23
1.2.3. Teoría espectral de procesos lineales	25
1.3. La suavización núcleo en la estimación de la densidad espectral.	27
2. Análisis discriminante de series de tiempo: Un enfoque no paramétrico	35
2.1. Introducción	35
2.2. Criterio discriminante basado en una medida general de disparidad espectral.	40
2.3. Un estadístico discriminante basado en técnicas de regresión polinómica local	43
2.3.1. Propiedades asintóticas	43
2.4. Análisis discriminante a partir de muestras de entrenamiento	68

2.4.1. Propiedades asintóticas	68
3. Comportamiento del procedimiento discriminante sobre muestras finitas	75
3.1. Estudio de simulación.	75
3.1.1. Procedimiento experimental.	76
3.1.2. Discriminación de procesos con errores gaussianos.	78
a) Procesos AR(1)	78
b) Procesos MA(1)	85
c) Procesos ARMA(p,q)	91
3.1.3. Discriminación de procesos con distribución marginal no gaussiana.	99
a) Procesos AR(1) con distribución marginal exponencial	100
b) Procesos AR(1) con distribución marginal doble exponencial	105
c) Series m -dependientes con distribución marginal gamma	108
d) Series m -dependientes con distribución marginal log-normal	111
3.1.4. Conclusiones	114
3.2. Aplicación a datos reales: clasificación de fonemas.	115
4. Análisis cluster de series temporales	125
4.1. Introducción	125
4.2. Algunas medidas de disparidad entre series de tiempo	132
4.3. Una vía no paramétrica para el análisis cluster de series de tiempo	137
4.4. Estudio de simulación	141
4.4.1. Procedimiento experimental	141
4.4.2. Clasificación de series como estacionarias o no estacionarias.	143
4.4.3. Clasificación de procesos ARMA	148
4.4.4. Clasificación de procesos no lineales	156

4.4.5. Conclusiones	161
4.5. Aplicación a datos reales: clasificación de registros electrocardiográficos. . .	163
5. Análisis de las series de cotización bursátil del sector bancario en España	173
5.1. Introducción	173
5.2. Presentación de las series de cotización bursátil	175
5.3. Análisis de las tendencias	177
5.4. Análisis cluster de las componentes aleatorias	191
5.5. Aplicación a datos simulados	207
5.6. Conclusiones	217
Apéndice: Resultados auxiliares	219
Bibliografía	221

Prólogo

Muchos de los problemas que se plantean al analizar series de tiempo se reducen a agrupar un conjunto más o menos grande de procesos en categorías similares (*análisis cluster*) o bien clasificar nuevas observaciones en una de dos o más categorías mutuamente excluyentes (*análisis discriminante*). Si bien ambas situaciones han sido exhaustivamente estudiadas desde el punto de vista de la teoría multivariante clásica, las características propias de las series temporales hacen que las soluciones desarrolladas para la clasificación de datos estáticos no siempre resulten adecuadas para abordar el proceso de clasificación de procesos estocásticos. Esencialmente, cualquier procedimiento cluster o discriminante se basa en la elección de una distancia adecuada para medir la discrepancia entre cada par de observaciones o entre una observación y cada una de las clases entre las que se discrimina. Así, al trabajar con series de tiempo, algunas distancias convencionales pueden no reflejar adecuadamente las similitudes o diferencias entre los diferentes procesos, al no tener en cuenta la estructura de dependencia subyacente. Por otra parte, con frecuencia se trabaja con series temporales de gran longitud, de modo que la alta dimensionalidad de las observaciones suele conllevar a problemas computacionales que dificultan las labores de clasificación.

La preponderancia de las series de tiempo en múltiples áreas de trabajo, como la sismología, economía, física o medicina, entre otras, hacen del análisis discriminante y del análisis cluster de series temporales problemas de gran interés teórico y práctico. Entre las diferentes soluciones que se han aportado en este campo, el análisis en el dominio de frecuencias o dominio espectral constituye una alternativa muy interesante, al permitir solventar el problema de la alta dimensionalidad inherente al trabajo con series de tiempo.

En esta memoria se presentan nuevos procedimientos, de corte no paramétrico, para abordar el análisis discriminante y cluster de series temporales en el ámbito espectral. La novedad de estos métodos radica en la utilización de estimadores tipo núcleo, basados en técnicas de regresión polinómica local, para la estimación de la densidad espectral de los procesos sujetos a clasificación. Dada la flexibilidad de las técnicas de corte no

paramétrico, cabe esperar que con estos procedimientos puedan igualarse los resultados de otros propuestos con anterioridad en la literatura, a la vez que se proponen técnicas que podrán aplicarse en contextos en los que no se pueden aplicar otros procedimientos que requieran de hipótesis más restrictivas.

La memoria consta de cinco capítulos, que se organizan tal y como se detalla a continuación. En el primero de ellos, se exponen algunas definiciones y resultados teóricos, ya conocidos, necesarios para el desarrollo de esta memoria. Se presentan los estimadores de la regresión de tipo polinómico local, sus propiedades asintóticas y su aplicación en el contexto de los modelos lineales generalizados. Tras una revisión de la teoría de representación espectral de procesos estacionarios, se presentan tres suavizadores no paramétricos diferentes que serán los utilizados a lo largo de esta memoria para estimar la densidad espectral de procesos estacionarios.

En el Capítulo 2 se propone un nuevo criterio discriminante para la clasificación de series de tiempo, basado en una medida de disparidad definida entre un estimador no paramétrico de la densidad espectral del proceso que se intenta clasificar y la densidad espectral de cada una de las clases de procesos entre las que se discrimina. Para la estimación del espectro se propone utilizar los tres estimadores tipo núcleo introducidos en el Capítulo 1. A continuación, se analizan las propiedades asintóticas de la regla discriminante propuesta, tanto en el caso de conocer la densidad teórica de las clases entre las que se discrimina, como en el caso de tener que estimarlas a partir de muestras de entrenamiento.

A continuación, en el Capítulo 3, se analiza el comportamiento del procedimiento discriminante propuesto para trabajar sobre muestras finitas. Para ello, se muestran los resultados de un estudio de simulación llevado a cabo con el objetivo de observar su comportamiento a la hora de clasificar entre distintas clases de procesos estocásticos. A continuación, se muestra una aplicación del procedimiento discriminante para la clasificación de un conjunto de datos reales, consistente en la identificación de cinco fonemas diferentes a partir de registros de voz digitalizados.

El Capítulo 4 está dedicado al análisis cluster de series de tiempo. Se proponen nuevas medidas, de corte no paramétrico, para abordar el análisis cluster de series temporales, algunas de las cuales extienden al contexto del análisis cluster el procedimiento discriminante que se describe en el Capítulo 2. Se proponen además otras medidas basadas en estadísticos de corte no paramétrico originalmente desarrollados para contrastar la igualdad del logaritmo del espectro de dos procesos estocásticos. Mediante un estudio de simulación, se compara el comportamiento de estas medidas, en diferentes contextos de clasificación, con el de otras distancias anteriormente propuestas, tanto en el dominio de tiempo como en el dominio espectral. Finalmente, se describen los resultados de la aplicación de los

procedimientos cluster propuestos a un conjunto de datos reales consistente en la clasificación de los registros electrocardiográficos de pacientes con diferentes tipos de arritmia, y pacientes sin ninguna patología coronaria.

Por último, en el Capítulo 5 se pretende mostrar la utilidad de las técnicas propuestas en la presente memoria para realizar, junto con otras técnicas de corte no paramétrico, un análisis completo del grado de similitud de un conjunto de series temporales. Para ello se utiliza un conjunto de datos reales, consistente en las series de cotización bursátil del sector bancario español durante un periodo de 2 años. Se analizan por separado las componentes deterministas y aleatorias de cada una de las series consideradas, y se utilizan los métodos aquí propuestos para realizar el análisis cluster de las series de residuos.

Finalmente, el Apéndice incluye algunos resultados teóricos utilizados para la elaboración de esta tesis.

Capítulo 1

Resultados preliminares

A lo largo de esta memoria se introducirán nuevos procedimientos, de corte no paramétrico, en el contexto del análisis discriminante y el análisis cluster de series de tiempo. Dichos procedimientos se basarán esencialmente en aproximaciones no paramétricas de las densidades espectrales de los procesos observados. Dadas las adecuadas propiedades que se conoce para los estimadores núcleo de tipo polinómico local, la estimación del espectro se realizará esencialmente empleando este tipo de técnicas.

En este capítulo se exponen algunos resultados teóricos, ya conocidos, que resultan de especial interés para el desarrollo de la presente memoria. La primera sección se centrará en los estimadores de la regresión de tipo polinómico local, sus propiedades asintóticas y su aplicación en el contexto de los modelos lineales generalizados. En la segunda sección se realizará una introducción a la representación espectral de procesos estacionarios. Finalmente se presentarán tres estimadores no paramétricos diferentes, propuestos por Fan y Kreutzberger (1998), para la estimación de la densidad espectral de un proceso estacionario. Dichos estimadores se obtienen suavizando el periodograma del proceso utilizando esencialmente técnicas de regresión polinómica local, y serán los utilizados en esta memoria para definir nuevos procedimientos discriminantes y de cluster en el contexto de series temporales.

1.1. Estimación no paramétrica de la función de regresión: la regresión polinómica local.

Una curva de regresión describe una relación general entre una variable aleatoria respuesta o dependiente, X , y una o más variables explicativas, regresoras o independientes, Z , de modo que, observado $Z = z$, la función de regresión $\varphi(z)$ proporciona el

valor esperado de X :

$$\varphi(z) = E(X/Z = z).$$

El análisis de regresión, cuyo objetivo es la obtención de una aproximación razonable a dicha curva, es de gran interés en la práctica. El conocer la forma de la función de regresión podría ayudar a predecir qué valores de X son esperados cuando se observan ciertos valores de la variable Z , o desenmascarar la existencia de algún tipo especial de dependencia estocástica entre ambas variables. El problema de la estimación de la función de regresión puede verse como un caso particular de un problema mucho más amplio, consistente en estimar una función $g(z), z \in \mathbb{R}^d$, asociada a una distribución, a partir de una muestra $\{z_1, z_2, \dots, z_n\}$. Casos particulares corresponden a la estimación de la función de densidad, la función de distribución, la función razón de fallo o la propia función de regresión.

El proceso de estimación de funciones de este tipo puede realizarse esencialmente de dos formas: paramétrica y no paramétricamente. En el primer caso, se supone de entrada que la curva de regresión teórica pertenece a una familia indexada por un parámetro finito-dimensional (por ejemplo, la familia de polinomios de grado r). Por esta vía el objetivo se centra en estimar los valores de aquellos parámetros que mejor “ajustan” la nube de datos muestrales. Lógicamente, la metodología paramétrica presenta el inconveniente de que el modelo seleccionado puede resultar demasiado restrictivo para los datos, pudiendo llevar así a estimaciones sesgadas. Por el contrario, el enfoque no paramétrico se caracteriza por no partir de la suposición inicial de que la curva de regresión deba pertenecer a una familia de funcionales dependientes de algún parámetro, de modo que la única información de la que se dispone es de la proporcionada por la propia muestra. Esta metodología ofrece así una herramienta mucho más flexible para analizar la relación de regresión desconocida. Dicha flexibilidad, junto con los importantes avances experimentados en el campo de la informática, han hecho que en los últimos años se produzca una gran proliferación de los estudios referentes a la **estimación no paramétrica de curvas**. Referencias clásicas en el supuesto de observaciones independientes son Silverman (1986) para la estimación de la función de densidad, Härdle (1990) para la regresión y Prakasa-Rao (1983) en varios contextos. Para el supuesto de que las observaciones verifiquen condiciones de dependencia, cabe destacar los trabajos de Vilar Fernández (1987), el de Györfi *et al.* (1989), y las referencias allí citadas. En el caso concreto de regresión no paramétrica, cumple citar los trabajos de Eubank (1988), Müller (1988), Härdle (1990), Hastie y Tibshirani (1990) y Wahba (1990).

El resto de la sección se estructurará como sigue. En primer lugar, se realizará una breve introducción de los estimadores tipo núcleo, como principales métodos de estimación

no paramétrica de la función de regresión. En segundo lugar, se describirán más detalladamente los estimadores polinómico locales que, por sus adecuadas propiedades teóricas y aplicadas, son hoy en día uno de los estimadores tipo núcleo más utilizados. Finalmente, se abordará la utilización de este tipo de estimadores en el contexto de los modelos lineales generalizados, exponiendo algunos resultados que resultarán de especial interés para los contenidos de la presente memoria.

1.1.1. Estimadores tipo núcleo de la función de regresión.

Dado $\{(z_i, X_i)\}_{i=1}^n$ un conjunto de n observaciones bidimensionales sobre cuya base se desea aproximar la respuesta media teórica de X dado $Z = z$, $\varphi(z)$, la relación de regresión puede modelizarse como:

$$X_i = \varphi(z_i) + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

donde $\varphi(\cdot)$ representa la función de regresión desconocida y ε el posible error de observación.

Los principales métodos de aproximación no paramétrica se obtienen promediando los valores X_i de aquellas observaciones cuyas abscisas z_i sean cercanas al punto z donde se quiere estimar la curva. Desde un punto de vista más formal, se proponen estimadores de la forma:

$$\hat{\varphi}(z) = \frac{1}{n} \sum_{i=1}^n W_{ni}(z) X_i, \quad (1.1)$$

donde $\{W_{ni}(z)\}_{i=1}^n$ denota una secuencia de pesos o coeficientes de ponderación que dependerán de las observaciones $\{(z_i, X_i)\}_{i=1}^n$. Dependiendo de cómo se tomen los pesos $\{W_{ni}(z)\}_{i=1}^n$ en (1.1) se obtendrán diferentes estimadores.

Un camino conceptualmente simple para representar los coeficientes de ponderación $\{W_{ni}(z)\}_{i=1}^n$ asociados a X_i en cada punto z es describir éstos en relación con la forma de una función de densidad centrada en z y adecuadamente reescalada mediante un parámetro que ajuste la distribución de los pesos en un determinado entorno de ese punto. Esta función se conoce con el nombre de **kernel** o **función núcleo**, y suele denotarse por K . Generalmente, se supone que el núcleo K es una función continua, acotada, simétrica y con integral igual a uno. Algunas de las funciones kernel más habituales son el núcleo de Epanechnikov, el kernel cuártico, el triangular, el gaussiano o el kernel uniforme.

El factor de reescalamiento de K , que dependerá a su vez del tamaño muestral n , determina el entorno donde ajustar K y por consiguiente su valor puede modificar notablemente los coeficientes de ponderación de cada X_i . A dicho factor se le denomina habitualmente **parámetro de suavizado** o **ventana** y se denotará por h_n o h indistintamente. El kernel

reparametrizado por h_n se denotará K_{h_n} :

$$K(u) \rightarrow K_{h_n}(u) = \frac{1}{h_n} K\left(\frac{u}{h_n}\right).$$

Entre los estimadores no paramétricos más utilizados, dentro de los de tipo núcleo, están:

- El **estimador de Nadaraya-Watson (N-W)** (Nadaraya (1964), Watson (1964)), con pesos:

$$W_{ni,N-W}(z) = \begin{cases} K_{h_n}(z - z_i)/\hat{d}_{h_n}(z) & \text{si } \hat{d}_{h_n}(z) > 0 \\ 0 & \text{si } \hat{d}_{h_n}(z) = 0 \end{cases}$$

donde $\hat{d}_{h_n}(z) = \frac{1}{n} \sum_{i=1}^n K_{h_n}(z - z_i)$ es el estimador kernel de la densidad de Rosenblatt-Parzen (Rosenblatt (1956) y Parzen (1962)).

- El **estimador de Gasser-Müller (G-M)** (Gasser y Müller (1979)), definido por la secuencia:

$$W_{ni,G-M}(z) = n \int_{s_{i-1}}^{s_i} K_{h_n}(z - u) du,$$

donde $s_0 = -\infty, s_n = \infty$ y $z_{i-1} \leq s_i \leq z_i$ para $i = 1, 2, \dots, n-1$ se eligen entre los datos y los pares $\{(z_i, X_i)\}_{i=1}^n$ se han ordenado previamente de manera creciente con respecto a la variable Z .

- El **estimador de Priestley-Chao (P-C)** (Priestley y Chao (1972)), definido por la sucesión de pesos:

$$W_{ni,P-C}(z) = n(z_i - z_{i-1})K_{h_n}(z_i - z), \text{ con } z_0 = 0,$$

donde los datos se suponen ordenados con respecto a la variable Z .

Durante años los estimadores no paramétricos de la función de regresión más utilizados han sido sin duda alguna los suavizadores de Nadaraya-Watson y de Gasser-Müller. Sin embargo más recientemente otro tipo de estimadores, los llamados de tipo polinómico local, han ganado en popularidad debido a las buenas propiedades que presentan.

1.1.2. La regresión polinómica local ponderada.

En la actualidad, la **regresión polinómica local** es una de las técnicas de regresión no paramétrica más utilizadas. Introducidos por Stone (1977), los estimadores de tipo polinómico local han sido posteriormente estudiados, en contextos con observaciones independientes, por Cleveland (1979), Lejeune (1985), Müller (1988), Cleveland y Devlin (1988)

y Fan (1992), y por Masry y Fan (1997) y Masry (1996a,1996b) para observaciones dependientes. Un estudio detallado de estos suavizadores y sus propiedades puede encontrarse en la monografía de Fan y Gijbels (1996).

La idea básica del estimador polinómico local es la siguiente: si se supone que existe la derivada de orden $(p + 1)$ de la función de regresión $\varphi(\cdot)$ en un punto $z_0 \in \mathbb{R}$, se podrá aproximar la función de regresión desconocida $\varphi(z)$ localmente por un polinomio de grado p en un entorno de z_0 . Aplicando el teorema de Taylor, se obtiene, en un punto z próximo a z_0 :

$$\varphi(z) \approx \varphi(z_0) + \varphi'(z_0)(z - z_0) + \dots + \frac{\varphi^{(p)}(z_0)}{p!}(z - z_0)^p. \quad (1.2)$$

Este polinomio puede ajustarse localmente utilizando la técnica de mínimos cuadrados locales ponderados. Es decir, minimizando la función:

$$\Psi(\beta) = \sum_{i=1}^n \left\{ X_i - \left(\sum_{j=0}^p \beta_j (z_i - z_0)^j \right) \right\}^2 K_{h_n}(z_i - z_0), \quad (1.3)$$

donde

$$\beta_j = \frac{\varphi^{(j)}(z_0)}{j!}, \quad j = 0, 1, \dots, p,$$

h_n es el parámetro de suavizado y $K_{h_n}(u) = (1/h_n)K(u/h_n)$ siendo K una función de densidad simétrica con soporte compacto (aunque dicha hipótesis puede debilitarse significativamente).

Denotando por $\widehat{\beta}_j$, $j = 0, \dots, p$ la solución del problema de mínimos cuadrados (1.3), resulta obvio por el desarrollo de Taylor en (1.2), que $\widehat{\varphi}_j(z_0) = j!\widehat{\beta}_j$ es un estimador de $\varphi^{(j)}(z_0)$, $j = 0, 1, \dots, p$. Para estimar la función $\varphi^{(j)}(\cdot)$ se resolverá el problema de mínimos cuadrados ponderados anterior para cada uno de los puntos z_0 en el intervalo de interés.

Es frecuente expresar el problema anterior en forma matricial. Para ello se utilizará la siguiente notación:

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & (z_1 - z_0) & \dots & (z_1 - z_0)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (z_n - z_0) & \dots & (z_n - z_0)^p \end{pmatrix} \in \mathcal{M}_{n \times (p+1)}$$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n \quad \text{y} \quad \widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta}_0 \\ \vdots \\ \widehat{\beta}_p \end{pmatrix} \in \mathbb{R}^{p+1}.$$

Sea además \mathbf{W}_n la matriz diagonal $n \times n$ dada por:

$$\mathbf{W}_n = \begin{pmatrix} K_{h_n}(z_1 - z_0) & 0 & \dots & 0 \\ 0 & K_{h_n}(z_2 - z_0) & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_{h_n}(z_n - z_0) \end{pmatrix}.$$

El problema de minimización (1.3) puede escribirse como:

$$\Psi(\boldsymbol{\beta}) = (\mathbf{X} - \mathbf{\Lambda}\boldsymbol{\beta})^t \mathbf{W}_n (\mathbf{X} - \mathbf{\Lambda}\boldsymbol{\beta}),$$

siendo $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^t$.

Utilizando la teoría de mínimos cuadrados ponderados se obtiene el estimador:

$$\hat{\boldsymbol{\beta}}(z_0) = (\mathbf{\Lambda}^t \mathbf{W}_n \mathbf{\Lambda})^{-1} \mathbf{\Lambda}^t \mathbf{W}_n \mathbf{X} = \mathbf{S}_n^{-1} \mathbf{T}_n,$$

donde se ha denotado

$$\mathbf{S}_n = \begin{pmatrix} s_0^n & s_1^n & \dots & s_p^n \\ s_1^n & s_2^n & \dots & s_{p+1}^n \\ \vdots & \vdots & \ddots & \vdots \\ s_p^n & s_{p+1}^n & \dots & s_{2p}^n \end{pmatrix} \in \mathcal{M}_{(p+1) \times (p+1)}$$

siendo

$$s_j^n = \sum_{i=1}^n K_{h_n}(z_i - z_0)(z_i - z_0)^j$$

y

$$\mathbf{T}_n = \begin{pmatrix} t_0^n \\ t_1^n \\ \vdots \\ t_p^n \end{pmatrix} \in \mathbb{R}^{p+1}$$

con

$$t_j^n = \sum_{i=1}^n K_{h_n}(z_i - z_0)(z_i - z_0)^j X_i.$$

Resulta obvio que para que podamos definir el estimador polinómico local es necesario que exista la matriz $\mathbf{S}_n^{-1} = (\mathbf{\Lambda}^t \mathbf{W}_n \mathbf{\Lambda})^{-1}$.

Frente a otros suavizadores tipo núcleo utilizados habitualmente, como el de Nadaraya-Watson o el de Gasser-Müller, los estimadores de tipo polinómico local poseen buenas propiedades tanto teóricas como aplicadas que los presentan como una técnica ventajosa.

En cualquier situación, igualan o mejoran los resultados obtenidos con los estimadores núcleo clásicos. Así mismo, se trata de un método que se adapta a diferentes diseños del modelo de regresión, como el diseño fijo o aleatorio, o a muestras con datos próximos o a los equiespaciados o con cluster de datos. A su vez, presenta un mejor comportamiento en los puntos frontera, siendo el sesgo del mismo orden que en los puntos interiores. Por último, posee interesantes propiedades de eficiencia minimax entre todos los estimadores lineales. Un extenso estudio de estos estimadores, mostrando su aplicación a una gran variedad de situaciones, puede encontrarse en la monografía de Fan y Gijbels (1996). A continuación se exponen algunos de los resultados que allí se presentan y que resultan de especial interés para el desarrollo de esta memoria.

En lo que sigue se empleará la siguiente notación. Se denotará por $\mu_j(K)$ y $\nu_j(K)$ los momentos de orden j de K y K^2 , respectivamente:

$$\mu_j(K) = \int u^j K(u) du, \quad \nu_j(K) = \int u^j K^2(u) du, \quad j = 0, 1, 2, \dots$$

Sean además las siguientes matrices:

$$\begin{aligned} \mathbf{S} &= (\mu_{i+j}(K))_{0 \leq i, j \leq p} & \boldsymbol{\mu} &= (\mu_{p+1}(K), \dots, \mu_{2p+1}(K))^t \\ \mathbf{S}^* &= (\nu_{i+j}(K))_{0 \leq i, j \leq p} & \hat{\boldsymbol{\mu}} &= (\mu_{p+2}(K), \dots, \mu_{2p+2}(K))^t \end{aligned}$$

El siguiente teorema establece el comportamiento asintótico de los estimadores polinómicos locales.

Teorema 1.1.1 (Ruppert y Wand (1994)) *Sea $d(z_0) > 0$ y $d(\cdot)$, $\varphi^{(p+1)}(\cdot)$ y $\sigma^2(\cdot)$ funciones continuas en un entorno de z_0 . Sea $h_n \rightarrow 0$ y $nh_n \rightarrow \infty$.*

La varianza asintótica del estimador polinómico local $\hat{\varphi}_j(z_0)$ viene dada por:

$$\text{Var}(\hat{\varphi}_j(z_0)) = \frac{j!^2 \sigma^2(z_0)}{d(z_0) n h_n^{2j+1}} V_j + o_P\left(\frac{1}{n h_n^{2j+1}}\right)$$

El sesgo asintótico de $\hat{\varphi}_j(z_0)$, para $p - j$ impar, es:

$$\text{Sesgo}(\hat{\varphi}_j(z_0)) = \frac{j!}{(p+1)!} \varphi^{(p+1)}(z_0) h_n^{p+1-j} B_j + o_P(h_n^{p+1-j})$$

En caso de que $p - j$ sea par, siendo $d'(\cdot)$ y $\varphi^{(p+2)}(\cdot)$ continuas en un entorno de z_0 y $nh_n^3 \rightarrow \infty$, el sesgo asintótico vendrá dado por:

$$\text{Sesgo}(\hat{\varphi}_j(z_0)) = \frac{j!}{(p+2)!} \left\{ \varphi^{(p+2)}(z_0) \right.$$

$$+(p+2)\varphi^{(p+1)}(z_0)\frac{d'(z_0)}{d(z_0)}\}h_n^{p+2-j}\widehat{B}_j + o_P(h_n^{p+2-j}),$$

siendo $\sigma^2(z_0)$ la varianza condicionada del modelo, $d(z_0)$ la función de densidad del diseño del modelo y donde los términos V_j , B_j y \widehat{B}_j denotan el j -ésimo elemento de la diagonal de la matriz $\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}$, el j -ésimo elemento de $\mathbf{S}^{-1}\boldsymbol{\mu}$ y el j -ésimo elemento de $\mathbf{S}^{-1}\widehat{\boldsymbol{\mu}}$, respectivamente.

Del teorema anterior pueden deducirse fácilmente las expresiones asintóticas del sesgo y la varianza del estimador lineal local. En particular, tomando $p = 1$ y $j = 0$ se obtiene que:

$$\text{Sesgo}(\widehat{\varphi}(z_0)) = \frac{1}{2}\varphi''(z_0)h_n^2\mu_2(K) + o_P(h_n^2)$$

y

$$\text{Var}(\widehat{\varphi}(x_0)) = \frac{\sigma^2(z_0)}{d(z_0)nh_n}\nu_0(K) + o_P((nh_n)^{-1})$$

También bajo hipótesis generales se obtiene la normalidad asintótica conjunta del vector de estimadores $\widehat{\boldsymbol{\beta}}(z_0)$:

$$\sqrt{nh_n}\left[\mathbf{H}_n\left(\widehat{\boldsymbol{\beta}}(z_0) - \boldsymbol{\beta}(z_0)\right) - \frac{\varphi^{(p+1)}(z_0)}{(p+1)!}h_n^{p+1}\mathbf{S}^{-1}\boldsymbol{\mu}\right] \rightarrow N_{(p+1)}\left(0, \frac{\sigma^2(z_0)}{d(z_0)}\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\right)$$

donde \mathbf{H}_n denota la matriz diagonal

$$\mathbf{H}_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \frac{1}{h_n} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{p!}{h_n^p} \end{pmatrix} \in \mathcal{M}_{(p+1)\times(p+1)}.$$

De los resultados anteriores se deduce que existe una clara diferencia entre los estimadores polinómicos locales con $p - j$ par o impar. Intuitivamente, si la función que se quiere estimar es compleja y se desea ajustar un polinomio de forma global, es obvio que se deberá utilizar un polinomio de grado alto. Sin embargo, en la suavización local esto no será necesario. De hecho, para estimar la función de regresión, se recomienda utilizar polinomios de grado 1 o en ocasiones de grado 3. En general, el ajuste de orden impar mejora el ajuste de orden par. Así, el estimador polinómico local lineal funciona mejor que el ajuste constante local, y el de orden 3 proporciona mejores resultados que el de orden 2, en el sentido de que al pasar de un grado par de polinomio al siguiente impar provoca una reducción en el sesgo del estimador sin perder nada en términos de varianza. Sin embargo, no hay comparación entre el ajuste lineal y el cúbico local. En general, si se desea estimar

$m^{(j)}(x)$ se recomienda utilizar un polinomio de grado $p = j + \nu$ con $\nu = 1$ ó 3 si j es impar y $\nu = 2$ si j es par.

La bondad de ajuste del estimador de la regresión depende, además del grado de los polinomios utilizados, de la elección del parámetro de suavizado y de la función Kernel empleada en la ponderación de las observaciones. Es conocido que la elección del Kernel no resulta crítica ni excesivamente relevante en términos de bondad de ajuste, como sí ocurre con la elección del parámetro ventana. Sin embargo, sí se ha demostrado, que independientemente de cuál sea el grado de los polinomios utilizados en el ajuste y de la ventana utilizada, el kernel de Epanechnikov resulta ser la función de ponderación óptima, en el sentido de minimizar el error cuadrático medio asintótico del estimador polinómico local resultante.

Mientras la elección del kernel resulta prácticamente irrelevante al objeto de minimizar el error cuadrático medio, la selección del parámetro de suavizado es crucial en este mismo sentido y la búsqueda de procedimientos encaminados a determinar la ventana óptima constituye sin duda uno de los problemas más serios y de mayor importancia en el contexto de la suavización de datos y, en particular, también en la regresión polinómica local. Por ello, es importante disponer de métodos de selección de banda que nos permitan calcular la ventana a utilizar de forma que en su cálculo sólo se utilice la información muestral. En este proceso, es importante tener en cuenta tanto el contexto del problema (diseño fijo o aleatorio, observaciones independientes o dependientes), como el tipo de banda a utilizar. Puede optarse por una ventana constante (también llamada ventana global) o bien por un parámetro ventana local o una banda variable. La utilización de una ventana local o variable permite introducir diferentes grados de suavización, de modo que pueda reducirse el sesgo cometido al estimar la función de regresión en los “picos”, así como la varianza de las estimaciones en las zonas en las que la curva sea más suave.

En el contexto de la regresión polinómica local los métodos de selección de banda más utilizados han sido los **métodos plug-in**. Estos se basan en aproximar el error cuadrático medio o el error cuadrático medio integrado asintótico o teórico, estimando aquellos valores que son desconocidos y calcular el valor de h que minimiza la expresión resultante.

Usando las expresiones asintóticas del sesgo y la varianza se obtiene que la banda óptima local, $h_{opt}(z_0)$, en el sentido de que minimiza el error cuadrático medio asintótico es:

$$h_{opt}^n(z_0) = C_{j,p}(K) \left[\frac{\sigma^2(z_0)}{\{\varphi^{(p+1)}(z_0)\}^2 d(z_0)} \right]^{1/(2p+3)} n^{-1/(2p+3)},$$

donde $C_{j,p}(K)$ es una constante que depende del kernel K de la forma:

$$C_{j,p}(K) = \left[\frac{(p+1)!^2(2j+1) \int K_j^2(t) dt}{2(p+j-1) \{ \int t^{(p+1)} K_j(t) dt \}^2} \right]^{1/(2p+3)}.$$

La Tabla 3.2 de Fan y Gijbels (1996) proporciona alguno de los valores de $C_{j,p}(K)$.

De igual forma, se obtiene una expresión para la ventana óptima asintótica global, minimizando el error cuadrático medio integrado asintótico:

$$h_{opt}^n = C_{j,p}(K) \left[\frac{\int \sigma^2(z) w(z) / d(z) dz}{\int \{ \varphi^{(p+1)}(z) \}^2 w(z) dz} \right]^{1/(2p+3)} n^{-1/(2p+3)}.$$

donde $w(z) \geq 0$ es una función peso incluida para evitar el efecto frontera.

Las expresiones anteriores para el parámetro de suavizado dependen de cantidades desconocidas como la varianza condicional, $\sigma^2(\cdot)$, y la función derivada, $\varphi^{(p+1)}(\cdot)$. En la práctica estas cantidades son reemplazadas por estimaciones de las mismas, obteniéndose la ventana plug-in asintótica local, $\hat{h}_{opt}^n(z_0)$, y la ventana plug-in asintótica global, \hat{h}_{opt}^n . Para estimar dichas cantidades existen varias alternativas. En Fan y Gijbels (1996) se propone estimar $\varphi^{(p+1)}(z)$ ajustando a la función de regresión, globalmente, un polinomio de grado $p+3$ y calcular a continuación la derivada $p+1$ de esta curva ajustada, $\check{\varphi}_{p+3}^{(p+1)}(z)$. Además, podrán estimarse los residuos como $\check{\varepsilon}_i = X_i - \check{\varphi}_{p+3}^{(p+1)}(z_i)$. Por otra parte, ésto mismo puede hacerse utilizando regresión polinómica local, para lo cual se necesita una ventana piloto g . Así, con esta ventana piloto se estima $\varphi^{(p+1)}(z)$ y se calculan los residuos estimados como $\tilde{\varepsilon}_i = X_i - \tilde{\varphi}_{p+3}^{(p+1)}(z_i)$, $i = 1, \dots, n$.

Referencias recientes sobre la regresión polinómica local son Fan (1992), Hastie y Loader (1993), Ruppert y Wand (1994), Fan, Heckman y Wand (1995) y Jones (1997), entre otras.

1.1.3. La regresión polinómica local en el contexto de los modelos lineales generalizados.

Aunque la regresión polinómica local se introduce originalmente en el contexto de modelos de regresión por mínimos cuadrados, su simplicidad, su sencilla interpretación y las excelentes propiedades estadísticas de este tipo de estimadores han propiciado su extensión a un amplio abanico de problemas estadísticos. Para el desarrollo de la presente memoria, resulta especialmente interesante su aplicación en el contexto de los **modelos lineales generalizados**.

Cuando se aborda un problema de regresión concreto, son muchas y muy distintas las situaciones que se pueden encontrar. Sin ir más allá, la variable respuesta Y puede

ser tanto discreta como continua. Asimismo, su varianza condicional puede o no depender de la función de regresión que se desea estimar. En la práctica, se dan así diferentes situaciones que requieren ser estudiadas mediante modelos que establezcan estructuras diferentes para la distribución condicional. Suele considerarse que esta distribución pertenece a la llamada familia exponencial, siendo ésta un conjunto de distribuciones entre las que se incluyen algunas de uso tan habitual como la distribución normal, binomial, Poisson o la distribución gamma. Bajo esta hipótesis, se asume que la función de regresión, una vez transformada convenientemente, puede expresarse como una función lineal de las covariables estudiadas. Los modelos resultantes se engloban bajo la denominación de **Modelos Lineales Generalizados**. Una explicación más detallada de estos modelos puede encontrarse en McCullagh y Nelder (1989). La aplicación de la teoría de la regresión polinómica local en este contexto permite obviar la hipótesis de linealidad de la función de regresión, y estimar su forma directamente a partir de los datos observados.

La densidad condicional de X dado $Z = z$ pertenece a la **familia exponencial** si es de la forma

$$\phi(x|z) = \exp \left\{ \frac{\theta(z)x - b(\theta(z))}{a(\phi)} + c(x, \phi) \right\} \quad (1.4)$$

para unas funciones conocidas $a(\cdot)$, $b(\cdot)$ y $c(\cdot, \cdot)$. El parámetro $\theta(\cdot)$ se denomina **parámetro canónico** y ϕ es el llamado **parámetro de dispersión**.

Bajo el modelo (1.4), puede demostrarse que

$$\varphi(z) = E(X|Z = z) = b'\{\theta(z)\} \quad (1.5)$$

y

$$\text{Var}(X|Z = z) = a(\phi)b''\{\theta(z)\}. \quad (1.6)$$

En el contexto de los modelos lineales generalizados, la función de regresión $\varphi(z)$ se modeliza linealmente a través de una transformación por una función g :

$$g(\varphi(z)) = \mathbf{z}^t \boldsymbol{\beta}, \quad (1.7)$$

con $\mathbf{z} = (z_1, \dots, z_n)^t$.

Si $g = (b')^{-1}$, g recibe el nombre de **link canónico**, ya que en ese caso $g(\varphi(z))$ es el parámetro canónico de la familia exponencial (1.4). Las expresiones (1.5), (1.6) y (1.7) caracterizan los modelos lineales generalizados. En estos modelos, el interés se centra en poder estimar el vector de parámetros $\boldsymbol{\beta}$ utilizando el método de máxima verosimilitud.

Ejemplo 1: Si la distribución condicional de X dado $Z = z$ sigue una distribución normal $N(\varphi(z), \sigma^2)$, la densidad condicional es de la forma

$$\begin{aligned}\phi(x|Z = z) &= \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \varphi(z)}{\sigma} \right)^2 \right\} \\ &= \exp \left\{ \frac{\varphi(z)x - \varphi^2(z)/2}{\sigma^2} - \frac{x^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right\}\end{aligned}$$

De modo que, en este caso, $\phi = \sigma^2$, $\theta(z) = \varphi(z)$, $a(\phi) = \phi$, $b(\varphi) = \varphi^2/2$ y $c(x, \phi) = -x^2/(2\sigma^2) - \log(\sqrt{2\pi}\sigma)$.

Ejemplo 2: Si la distribución condicional de X dado $Z = z$ es exponencial $\text{Exp}(\lambda(z)^{-1})$, la densidad condicional viene dada por

$$\begin{aligned}\phi(x|Z = z) &= \lambda(z)^{-1} \exp \{ -\lambda(z)^{-1}x \} = \exp \{ -\lambda(z)^{-1}x + \ln(\lambda(z)^{-1}) \} = \\ &= \exp \{ -\lambda(z)^{-1}x - \ln(\lambda(z)) \}.\end{aligned}$$

Bajo este supuesto, se tiene que $\phi = 1$, $\theta(z) = -\lambda(z)^{-1}$, $a(\phi) = 1$, $b(\varphi) = -\ln(-\varphi)$ y $c(x, \phi) = 0$.

En algunas situaciones, sin embargo, no puede garantizarse que se cumpla la relación lineal dada en (1.7). La utilización de las técnicas de regresión polinómica local en este contexto permiten estimar la función de regresión $\varphi(z)$ directamente a partir de la muestra sin asumir la hipótesis de linealidad. Puesto que estimar $\varphi(z)$ es equivalente a estimar $\theta(z) = (b')^{-1}(\varphi(z))$, se suele optar por estimar directamente el parámetro canónico $\theta(z)$.

Bajo los modelos lineales generalizados (1.4) y (1.7) se suele estimar $\theta(z)$ mediante el método de máxima verosimilitud. Dados $\{(z_i, X_i)\}_{i=1}^n$, la log-verosimilitud asociada viene dada por

$$\sum_{i=1}^n \left[\frac{\theta_i X_i - b(\theta_i)}{a(\phi)} + c(X_i, \phi) \right]$$

donde $\theta_i = \theta(z_i)$.

En muchas aplicaciones, sin embargo, tan sólo se tiene una idea de la función varianza condicional, que puede modelizarse según

$$\text{Var}(X|Z = z) = V(\varphi(z))$$

para alguna función $V(\cdot)$. En este caso, el método de mínimos cuadrados ponderados no resulta consistente. En su lugar, se puede reemplazar la log-verosimilitud condicional

por la llamada **función de cuasi-verosimilitud** (Wedderburn (1974)), $Q(\varphi(z), x)$, que satisface:

$$\frac{\partial}{\partial \mu} Q(\mu, x) = \frac{\mu - x}{V(\mu)}.$$

Los métodos de cuasi-verosimilitud presentan un comportamiento análogo a los métodos de máxima verosimilitud, y por lo tanto son un sustituto razonable cuando no se dispone de la función de verosimilitud. En particular, el método de log-verosimilitud coincide con el de cuasi-verosimilitud en el caso de la familia exponencial.

Para estimar $\eta(z) = g(\varphi(z))$, o alguna de sus derivadas, $\eta^{(\nu)}(\cdot)$, $\nu = 0, 1, 2, \dots, p$, $\eta(\cdot)$ puede aproximarse localmente por un polinomio de grado p :

$$\eta(u) = \beta_0 + \beta_1(u - z) + \dots + \beta_p(u - z)^p,$$

estando u en un entorno suficientemente próximo de z .

Utilizando los datos locales alrededor de z , se obtiene la cuasi-verosimilitud local:

$$\sum_{i=1}^n Q(g^{-1}(\beta_0 + \dots + \beta_p(z_i - z)^p), X_i) K_{h_n}(z_i - z). \quad (1.8)$$

Maximizando (1.8) con respecto a $\beta = (\beta_0, \dots, \beta_p)^t$ se obtiene el estimador de máxima cuasi-verosimilitud local para $\eta^{(\nu)}(z)$, $\nu = 0, \dots, p$:

$$\hat{\eta}_\nu(z) = \nu! \hat{\beta}_\nu(z).$$

La función de regresión $\varphi(z)$ puede entonces ser estimada mediante la inversa de la función link:

$$\hat{\varphi}(z) = g^{-1}(\hat{\eta}(z))$$

El método de cuasi-verosimilitud no admite una solución explícita al menos que se tome $p = 0$. El proceso de maximización debe afrontarse así mediante un método iterativo como el método de Newton-Raphson o alguna de sus variantes, no siendo en general necesarias demasiadas iteraciones para obtener una solución.

Un estudio detallado de las propiedades asintóticas de estos estimadores puede encontrarse en la monografía de Fan y Gijbels (1996) y en las referencias que allí se citan. A continuación se exponen de forma más escueta algunos resultados de especial interés en de esta memoria. En primer lugar, se impondrán algunas condiciones de regularidad con el fin de establecer algunos resultados referentes al comportamiento asintótico de los estimadores anteriores.

Sea $d(z)$ la densidad marginal de Z y $q_l(z, x) = \left(\frac{\partial^l}{\partial z^l}\right) Q(g^{-1}(\varphi(z)), x)$. Nótese que q_l es lineal en x , para cada z fijo, y que:

$$q_1(\eta(z), \varphi(z)) = 0,$$

$$q_2(\eta(z), \varphi(z)) = -\rho(z),$$

donde $\rho(z) = \left((g'(\varphi(z)))^2 V(\varphi(z))\right)^{-1}$.

Se asumirán las siguientes hipótesis:

Hipótesis (H1.1) La función $q_2(z, x) < 0$ para todo $z \in \mathbb{R}$ y x en el rango de la variable respuesta.

Hipótesis (H1.2) Las funciones $d(\cdot)$, $\eta^{(p+1)}(\cdot)$, $Var(X|Z = \cdot)$, $V(\cdot)$ y $g''(\cdot)$ son continuas.

Hipótesis (H1.3) $\rho(z) \neq 0$, $Var(X|Z = z) \neq 0$ y $g'(\varphi(z)) \neq 0$.

Hipótesis (H1.4) $E(X^4|Z = \cdot)$ está acotada en un entorno de z .

Hipótesis (H1.5) La función kernel $K(\cdot)$ es una función simétrica y no negativa de soporte compacto.

Teorema 1.1.2 (Fan, Heckman y Wand (1995)) *Sea $p - \nu$ impar. Bajo las hipótesis (H1.1)-(H1.5) se tiene que*

$$\sigma_\nu(z) \left\{ \hat{\eta}_\nu(z) - \eta^{(\nu)}(z) - b_\nu(z) + o(h_n^{p+1-\nu}) \right\} \rightarrow N(0, 1) \text{ si } h_n \rightarrow 0 \text{ y } nh_n \rightarrow \infty,$$

donde

$$b_\nu(z) = \left\{ \int t^{p+1} K(t) dt \right\} \frac{\nu!}{(p+1)!} \eta^{(p+1)}(z) h_n^{p+1-\nu}$$

y

$$\sigma_\nu^2(z) = \left(\int K^2(t) dt \right) \frac{\nu!^2 \{g'(\varphi(z))^2 Var(X|Z = z)\}}{d(z) n h_n^{2\nu+1}}.$$

Los resultados del teorema anterior pueden aplicarse fácilmente al caso en el que $\nu = 0$, $p = 1$. Más concretamente, denotando por $\hat{\eta}(z)$ el estimador de η basado en la cuasi-verosimilitud local con $p=1$, se tiene que:

$$\sigma(z) \left\{ \hat{\eta}(z) - \eta(z) - b(z) + o(h_n^2) \right\} \rightarrow N(0, 1),$$

donde

$$b(z) = \left\{ \int t^2 K(t) dt \right\} \frac{1}{2} \eta^2(z) h_n^2$$

y

$$\sigma^2(z) = \int K^2(t) dt \frac{\{g'(\varphi(z))^2 \text{Var}(X|Z=z)\}}{d(z)nh_n}.$$

Si se considera el estimador normalizado

$$\widehat{\boldsymbol{\beta}}^* = a_n^{-1} \left(\widehat{\beta}_0 - \eta(z), \dots, h_n^p \left(\widehat{\beta}_p - \eta^{(p)}(z)/p! \right) \right)^t, \quad \text{con } a_n = (nh_n)^{-1/2}$$

en la demostración del Teorema 1.1.2 se establece que $\widehat{\boldsymbol{\beta}}^*$ es de la forma

$$\widehat{\boldsymbol{\beta}}^* = -\mathbf{A}^{-1} \mathbf{W}_{n,z} + o_P(1),$$

donde

$$\mathbf{A} = \rho(z) d(z) (\mu_{i+j-2})_{i \leq i, j \leq p+1}$$

y

$$\mathbf{W}_{n,z} = a_n \sum_{i=1}^n q_1(\bar{\eta}(z, z_i), X_i) \mathbf{Z}_i K \left(\frac{z_i - z}{h_n} \right), \quad (1.9)$$

siendo

$$\mathbf{Z}_i = \left(1, \frac{z_i - z}{h_n}, \dots, \frac{(z_i - z)^p}{h_n^p} \right)^t$$

y

$$\bar{\eta}(z, z_i) = \sum_{j=0}^p \frac{\eta^{(j)}(z)}{j!} (z_i - z)^j.$$

Además, se demuestra la normalidad asintótica de $\widehat{\boldsymbol{\beta}}^*$ directamente a partir de la de $\mathbf{W}_{n,z}$:

$$\mathbf{W}_{n,z} - a_n^{-1} \rho(z) d(z) h_n^{p+1} \frac{\eta^{(p+1)}}{(p+1)!} \int u^{p+1} \mathbf{U} K(u) du \{1 + o(1)\} \rightarrow N(0, \mathbf{B}),$$

con

$$\mathbf{B} = \frac{d(z) \text{Var}(X|Z=z)}{(V(\varphi(z)g'(\varphi(z))))^2} \int \mathbf{U} \mathbf{U}^t K^2(u) du,$$

donde $\mathbf{U} = (1, u, \dots, u^p)^t$.

La distribución asintótica de $\widehat{\varphi}(\cdot) = g^{-1}(\widehat{\eta}(\cdot))$ se obtiene directamente del teorema anterior:

Corolario 1.1.1 *Bajo las condiciones del Teorema 1.1.2, si $nh_n^3 \rightarrow \infty$, el error $\widehat{\varphi}(z) - \varphi(z)$ tiene el mismo comportamiento asintótico que $\widehat{\eta}(z) - \eta(z)$, con el sesgo asintótico dividido por $g'(\varphi(z))$ y la varianza asintótica por $(g'(\varphi(z)))^2$.*

1.2. Análisis espectral de series de tiempo.

La representación espectral de un proceso estacionario $X = \{X(t), t \in \mathbb{Z}\}$ esencialmente descompone X en una suma de componentes sinusoidales con coeficientes aleatorios e incorrelados. La descomposición espectral es pues, en el ámbito de las series de tiempo, un concepto análogo a la representación de Fourier de las funciones determinísticas. El análisis de procesos estacionarios mediante su representación espectral se denomina habitualmente “análisis en el dominio de frecuencias” o “análisis espectral”. Resulta equivalente al análisis en el “dominio de tiempo” basado en la función de autocovarianzas, pero proporciona una forma diferente de analizar los procesos que puede resultar más interesante y útil en algunas aplicaciones.

A continuación se describirán algunos aspectos esenciales de la teoría espectral de procesos estacionarios, prestando especial atención a aquellos conceptos y resultados que más tarde serán de utilidad en el desarrollo de esta memoria. Un estudio más detallado de la teoría de análisis espectral puede encontrarse en referencias especializadas (Brillinger (1981), Priestley (1989), Brockwell y Davis (1996)).

1.2.1. La densidad espectral

Sea $X = \{X(t), t \in \mathbb{Z}\}$ un proceso estacionario de media cero y con función de autocovarianzas $\gamma(\cdot)$ absolutamente sumable, es decir:

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

Se define la **densidad espectral** del proceso $X = \{X(t), t \in \mathbb{Z}\}$ como la función $f(\cdot)$ dada por:

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\lambda h}, \quad -\infty < \lambda < \infty. \quad (1.10)$$

La sumabilidad de $|\gamma(\cdot)|$ garantiza que la serie que aparece en (1.10) converge absolutamente. Además, puesto que las funciones $\cos(\cdot)$ y $\sin(\cdot)$ tienen ambas periodo 2π , la función de densidad espectral es periódica de igual periodo, por lo que es suficiente definirla en el intervalo $(-\pi, \pi]$. En particular, la función de densidad espectral verifica las siguientes propiedades:

1. $f(\cdot)$ es par, es decir, $f(\lambda) = f(-\lambda)$ para todo $\lambda \in (-\pi, \pi]$.
2. $f(\lambda) \geq 0$ para todo $\lambda \in (-\pi, \pi]$.

3. La función de autocovarianzas del proceso X_t puede expresarse como:

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(h\lambda) f(\lambda) d\lambda, \forall h \in \mathbb{Z}. \quad (1.11)$$

Debe hacerse notar, sin embargo, que no toda función de autocovarianzas tiene asociada una densidad espectral. En general, existirá una función F en $(-\pi, \pi]$, continua a la derecha, no decreciente y acotada, con $F(-\pi) = 0$, tal que:

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda), \forall h \in \mathbb{Z}.$$

Así definida, F es la **función de distribución espectral** de $\gamma(\cdot)$. Si $F(\lambda)$ puede expresarse como $F(\lambda) = \int_{-\pi}^{\lambda} f(x) dx$ se dirá que la serie de tiempo tiene **espectro continuo** y que f es su **densidad espectral**. Si F es una distribución discreta, se dirá que la serie tiene **espectro discreto**.

En general, se puede demostrar que cualquier proceso estacionario es el resultado de la superposición de una cantidad infinita de componentes sinusoidales:

$$X(t) = \int_{(-\pi, \pi]} e^{it\lambda} dZ(\lambda), \quad (1.12)$$

donde $\{Z(\lambda), -\pi < \lambda \leq \pi\}$ es un proceso de valores complejos con incrementos incorrelados. La representación (1.12) de un proceso estacionario de media cero $X = \{X(t), t \in \mathbb{Z}\}$ se conoce como la **representación espectral** del proceso y es comparable a la representación espectral (1.11) de la función de autocovarianzas $\gamma(\cdot)$. Como consecuencia de esta expresión, puede deducirse que un “salto” en la función de distribución espectral (o, equivalentemente, un “pico” en la densidad espectral) en una frecuencia $\pm w$ indica la presencia en la serie de tiempo de un componente sinusoidal de frecuencia w (y periodo $2\pi/w$).

Ejemplo 1: Proceso ruido blanco normal.

Sea $X = \{X(t), t \in \mathbb{Z}\}$ un proceso ruido blanco normal con función de autocovarianzas $\gamma(0) = \sigma^2$ y $\gamma(h) = 0$ para todo $|h| > 0$, con $\sigma^2 > 0$. Para este proceso, su densidad espectral es constante igual a:

$$f(\lambda) = \frac{\sigma^2}{2\pi}, -\pi < \lambda \leq \pi. \quad (1.13)$$

En este caso, cada frecuencia en el espectro contribuye de igual forma a la varianza del proceso.

Ejemplo 2: Proceso AR(1) normal.

Sea $X = \{X(t), t \in \mathbb{Z}\}$ el proceso autorregresivo AR(1) dado por:

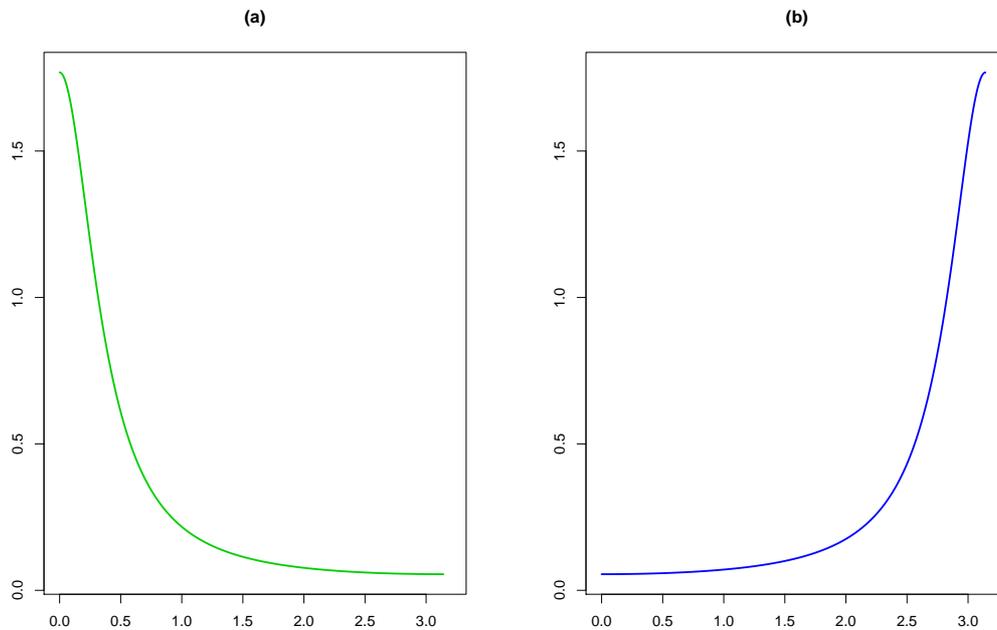
$$X(t) = \phi X(t-1) + \varepsilon(t),$$

donde $\varepsilon(t)$ es una secuencia de variables aleatorias i.i.d. según una distribución normal $N(0, \sigma^2)$. Entonces la densidad espectral de X viene dada por:

$$f(\lambda) = \frac{\sigma^2}{2\pi} (1 - 2\phi \cos \lambda + \phi^2)^{-1}, \quad -\pi < \lambda \leq \pi.$$

La Figura 1.1a muestra la densidad espectral de un proceso AR(1) para el caso en el que $\phi = 0,7$. En este caso, la densidad es mayor para frecuencias bajas y menor en las frecuencias más altas. Esto ocurre así puesto que para un proceso como éste, la función de autocovarianzas es positiva con un valor muy grande en el primer retardo, de modo que la serie resulta suave con pocas componentes de frecuencias altas. Por el contrario, para $\phi = -0,7$ la función de autocovarianzas presenta un valor negativo alto en el primer retardo, indicando que la serie fluctúa rápidamente en torno a su valor medio. En este caso, existirán componentes sinusoidales con frecuencias altas (Figura 1.1b).

Figura 1.1: Densidad espectral correspondientes a un proceso AR(1) con errores gaussianos y parámetro autorregresivo (a) $\phi = 0,7$ o (b) $\phi = -0,7$.



Ejemplo 3: Proceso ARMA(p,q) normal.

En general, para el caso en el que $X = \{X(t), t \in \mathbb{Z}\}$ sea un proceso ARMA(p,q) dado por

$$\phi(L)X(t) = \theta(L)\varepsilon(t), \quad \varepsilon(t) \approx N(0, \sigma^2)$$

su densidad espectral es igual a

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda < \pi.$$

1.2.2. El periodograma

En la práctica, resulta necesario obtener aproximaciones espectrales tanto de la función de autocovarianzas como del espectro de una serie. Dada $\mathbf{X}_n = (X_1, \dots, X_n)^t$ una realización parcial de un proceso estacionario de media cero, $X = \{X(t), t \in \mathbb{Z}\}$, la función de autocovarianzas muestral puede utilizarse como una estimación de $\gamma(\cdot)$, mientras que el periodograma $I_n(\cdot)$ resulta ser el análogo muestral de la densidad espectral $f(\cdot)$.

Sea $\mathbf{x} = (X_1, X_2, \dots, X_n)^t \in C^n$. Sean además $\lambda_k = \frac{2\pi k}{n}$, donde k recorre los enteros entre $-N$ y N , con $N = \lceil \frac{n-1}{2} \rceil$. Es decir:

$$\lambda_k = \frac{2\pi k}{n}, \quad k = -\left\lceil \frac{n-1}{2} \right\rceil, \dots, \left\lceil \frac{n-1}{2} \right\rceil. \tag{1.14}$$

Los valores en (1.14) reciben el nombre de **frecuencias de Fourier** asociadas al tamaño muestral n . De esta forma, los vectores:

$$\mathbf{e}_k = \frac{1}{\sqrt{n}} \left(e^{i\lambda_k}, e^{2i\lambda_k}, \dots, e^{ni\lambda_k} \right), \quad k = -\left\lceil \frac{n-1}{2} \right\rceil, \dots, \left\lceil \frac{n-1}{2} \right\rceil$$

forman una base en C^n , de modo que el vector $\mathbf{x} \in C^n$ puede expresarse como suma de n componentes:

$$x = \sum_{k=-\lceil \frac{n-1}{2} \rceil}^{\lceil \frac{n-1}{2} \rceil} a_k e_k,$$

donde

$$a_k = \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t e^{-it\lambda_k}. \tag{1.15}$$

La secuencia $\{a_k\}$ en (1.15) se conoce como la **transformada finita de Fourier** del proceso X . Sus adecuadas propiedades teóricas (proporciona variables prácticamente incorreladas) y la rapidez con la que puede calcularse mediante cualquiera de los algoritmos de transformación rápida de Fourier (Otnes y Enochson (1978)), hacen que desempeñe un papel fundamental en el análisis de series de tiempo.

Se define el **periodograma** de $\mathbf{X}_n = (X_1, \dots, X_n)^t$ como:

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t \exp(-i\lambda t) \right|^2, \quad \lambda \in [-\pi, \pi]. \quad (1.16)$$

Mediante cálculos sencillos se puede demostrar el siguiente resultado:

Proposición 1.2.1 (Priestley(1989)) *Dada $\mathbf{X}_n = (X_1, \dots, X_n)^t$ una realización de un proceso estacionario de media cero $X = \{X(t), t \in \mathbb{Z}\}$, y λ_k una de las frecuencias de Fourier, $\lambda_k = \frac{2\pi k}{n}$, en $(-\pi, \pi]$, $\lambda_k \neq 0$, entonces:*

$$I_n(\lambda_k) = \frac{1}{2\pi} \sum_{|h| < n} \hat{\gamma}(h) e^{-ih\lambda_k} \quad (1.17)$$

donde $\hat{\gamma}(h)$ es la función de autocovarianzas muestral asociada a \mathbf{X}_n .

La comparación de la expresión (1.17) con la definición (1.10) de la densidad espectral sugiere utilizar el periodograma $I_n(\lambda)$ como un estimador natural de $f(\lambda)$. Sin embargo, estudios detallados de sus propiedades revelan que no se trata de un estimador consistente. A continuación se expondrán algunos resultados en este sentido que resultarán de interés para el posterior desarrollo de esta memoria.

Teorema 1.2.1 (Priestley(1989)) *Sea $X = \{X(t), t \in \mathbb{Z}\}$ una serie de variables aleatorias independientes con cumulante de cuarto orden finito k_4 , entonces:*

$$\text{cov}(I_n(\lambda_1), I_n(\lambda_2)) = \frac{k_4}{4\pi^2 n} + \frac{\sigma_X^4}{2\pi n} \{F_n(\lambda_1 + \lambda_2) + F_n(\lambda_1 - \lambda_2)\}$$

donde $F_n(\lambda)$ es el núcleo de Fejer dado por:

$$F_n(\lambda) = \frac{1}{2\pi n} \frac{\text{sen}^2(\frac{N\lambda}{2})}{\text{sen}^2(\frac{\lambda}{2})}.$$

Como consecuencia del teorema anterior se deduce que, tomando $\lambda_1 = \lambda_2 = \lambda$:

$$\text{Var}(I_n(\lambda)) = \begin{cases} \frac{1}{4\pi^2} (\sigma_X^4 + \frac{k_4}{n}) + O(n^{-2}) & \lambda \neq 0, \pm\pi \\ \frac{1}{4\pi^2} (2\sigma_X^4 + \frac{k_4}{n}) & \lambda = 0, \pm\pi \end{cases}$$

De este modo, si X es un proceso normal (con lo cual $k_4 = 0$) y λ_k es una de las frecuencias de Fourier, se obtiene:

$$\text{Var}(I_n(\lambda_k)) = \begin{cases} \frac{1}{4\pi^2} \sigma_X^4 & \lambda \neq 0, \pm\pi \\ \frac{1}{4\pi^2} 2\sigma_X^4 & \lambda = 0, \pm\pi \end{cases}$$

Bajo estas mismas hipótesis, para $\lambda_1 \neq \pm\lambda_2$:

$$\text{Cov}(I_n(\lambda_1), I_n(\lambda_2)) = \begin{cases} 0 & \text{si } X \text{ es normal y } \lambda_1, \lambda_2 \text{ son múltiplos de } \frac{2\pi}{n} \\ O(n^{-2}) & \text{si } X \text{ es normal y } |\lambda_1 \pm \lambda_2| \gg \frac{2\pi}{n} \\ O(n^{-1}) & \text{si } X \text{ es no normal y } |\lambda_1 \pm \lambda_2| \gg \frac{2\pi}{n}, \\ & \text{o } \lambda_1, \lambda_2 \text{ son múltiplos de } \frac{2\pi}{n} \end{cases} \quad (1.18)$$

Es decir, incluso para procesos no normales, las ordenadas del periodograma son asintóticamente incorreladas si λ_1, λ_2 son múltiplos de $\frac{2\pi}{n}$ o están suficientemente espaciadas.

1.2.3. Teoría espectral de procesos lineales

El análisis espectral resulta especialmente útil en el estudio de procesos lineales. A continuación se abordará la transmisión de procesos estocásticos a través de filtros lineales, con el fin de mostrar la forma que toma la densidad espectral de un proceso lineal cualquiera.

Un proceso $X = \{X(t), t \in \mathbb{Z}\}$ es el “output” de un filtro lineal invariante $\psi = \{\psi_j, j = 0, \pm 1, \dots\}$ aplicado a un proceso “input” $Z = \{Z(t), t \in \mathbb{Z}\}$ si:

$$X(t) = \sum_{j=-\infty}^{\infty} \psi_j Z(t-j), \quad t = 0, \pm 1, \dots$$

Se dice entonces que el proceso $X = \{X(t), t \in \mathbb{Z}\}$ es un **proceso lineal**.

Proposición 1.2.2 (Priestley(1989)) *Sea $Z = \{Z(t), t \in \mathbb{Z}\}$ un proceso estacionario de media cero y densidad espectral $f_Z(\lambda)$. Sea $\psi = \{\psi_j, j = 0, \pm 1, \dots\}$ un filtro lineal invariante con $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Entonces el proceso lineal*

$$X(t) = \sum_{j=-\infty}^{\infty} \psi_j Z(t-j)$$

es un proceso estacionario de media cero y densidad espectral

$$f(\lambda) = \left| \psi(e^{-i\lambda}) \right|^2 f_Z(\lambda) = \psi(e^{-i\lambda}) \psi(e^{i\lambda}) f_Z(\lambda),$$

donde $\psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$.

La función $\psi(e^{-i\cdot})$ se denomina **función de transferencia del filtro**, y $|\psi(e^{-i\cdot})|^2$ **potencia de la función de transferencia**.

Del resultado anterior se deduce que si $X = \{X(t), t \in \mathbb{Z}\}$ es un proceso lineal Gaussiano dado por

$$X(t) = \sum_{j=-\infty}^{\infty} \psi_j Z(t-j),$$

con $\{Z(j), j \in \mathbb{Z}\}$ una secuencia de variables aleatorias i.i.d. $N(0, \sigma_Z^2)$, su densidad espectral es necesariamente de la forma

$$f(\lambda) = |\psi(\lambda)|^2 \frac{\sigma_Z^2}{2\pi}, \quad (1.19)$$

con

$$\psi(\lambda) = \sum_{j=-\infty}^{\infty} \psi_j \exp(-ij\lambda). \quad (1.20)$$

Del mismo modo, el siguiente teorema establece la expresión asintótica del periodograma de un proceso lineal.

Teorema 1.2.2 (Priestley(1989)) Sea $X = \{X(t), t \in \mathbb{Z}\}$ un proceso lineal general dado por

$$X(t) = \sum_{j=-\infty}^{\infty} \psi_j Z(t-j),$$

siendo $Z = \{Z(t), t \in \mathbb{Z}\}$ un proceso estacionario de variables aleatorias independientes con $E(Z(t)) = 0$, $E(Z^2(t)) = \sigma_Z^2$, $E(Z^4(t)) < \infty$ y $\sum_{j=-\infty}^{\infty} |\psi_j| |j|^\alpha < \infty$, $\alpha > 0$.

Entonces, se tiene que:

$$I_n(\lambda) = f(\lambda) \frac{2\pi}{\sigma_Z^2} I_{n,Z}(\lambda) + R_n(\lambda), \quad (1.21)$$

siendo $I_{n,Z}(\lambda)$ el periodograma asociado a $(Z_1, \dots, Z_n)^t$, y donde $E\{|R_n(\lambda)|^2\} = O(n^{-2\alpha})$ uniformemente en λ .

La expresión (1.21) permite obtener una expresión asintótica de $I_n(\lambda)$ directamente a partir de los resultados conocidos para $I_{n,Z}$. En particular, para un proceso lineal Gaussiano, con $\{Z(t)\}$ i.i.d. $N(0, \sigma_Z^2)$, se tiene para $\lambda_k = \frac{2\pi k}{n}$ que:

$$I_{n,Z}(\lambda_k) = \begin{cases} \frac{1}{4\pi} \sigma_Z^2 \chi_2^2 & k \neq 0, \frac{n}{2}, n \text{ par} \\ \frac{1}{4\pi} 2\sigma_Z^2 \chi_1^2 & k = 0, \frac{n}{2} \end{cases} \quad (1.22)$$

A partir de (1.21) y (1.22) se obtiene que:

$$I_n(\lambda_k) = \begin{cases} \frac{1}{2} f(\lambda_k) \chi_2^2 + R_{n,k} & k \neq 0, \frac{n}{2}, n \text{ par} \\ f(\lambda_k) \chi_1^2 + R_{n,k} & k = 0, \frac{n}{2} \end{cases} \quad (1.23)$$

De modo que, si se ignora $R_{n,k}$, sobre las frecuencias de Fourier se tiene que:

$$E(I_n(\lambda_k)) = f(\lambda_k), \quad \forall k = -N, \dots, N$$

$$\text{Var}(I_n(\lambda_k)) = \begin{cases} f^2(\lambda_k) & k \neq 0, \frac{n}{2} \\ 2f^2(\lambda_k) & k = 0, \frac{n}{2} \end{cases}$$

Puesto que $\text{Var}(I_n)$ no converge a cero, $I_n(\cdot)$ no es un estimador consistente de $f(\cdot)$.

De igual forma, puede establecerse la covarianza asintótica para las ordenadas del periodograma sobre las frecuencias de Fourier:

Teorema 1.2.3 (Priestley(1989)) Sea $X = \{X(t), t \in \mathbb{Z}\}$ un proceso lineal general como en el Teorema 1.2.2. Entonces:

$$\text{Cov}(I_n(\lambda_1), I_n(\lambda_2)) = \left\{ \frac{e}{n} + \frac{2\pi}{n} [F_n(\lambda_1 + \lambda_2) + F_n(\lambda_1 - \lambda_2)] \right\} f(\lambda_1)f(\lambda_2) + O(n^{-\alpha})$$

donde $e = \frac{k_4}{\sigma_Z^4} = E(Z_t^4) - 3$, F_n es el núcleo de Fejer y el término restante es de $O(n^{-\alpha})$ uniformemente en λ_1, λ_2 .

En particular, si X es un proceso lineal normal las ordenadas del periodograma sobre las frecuencias de Fourier satisfacen

$$\text{cov}(I_n(\lambda_t), I_n(\lambda_s)) = f(\lambda_t)f(\lambda_s)1_{\{\lambda_t=\lambda_s\}} + O(n^{-2}). \quad (1.24)$$

1.3. La suavización núcleo en la estimación de la densidad espectral.

La estimación de la densidad espectral $f(\cdot)$ de un proceso estacionario $X = \{X(t), t \in \mathbb{Z}\}$ puede resultar especialmente útil en el estudio de ciertas características de las series temporales. A partir de una realización de dicho proceso, $\mathbf{X}_n = (X_1, \dots, X_n)^t$, el periodograma $I_n(\cdot)$ resulta ser el análogo muestral de $f(\cdot)$. Sin embargo, ya se ha indicado que aunque éste es un estimador asintóticamente insesgado de la densidad espectral, resulta no ser un estimador consistente, lo que hace necesario estudiar otro de tipo de estimadores que mejoren sus propiedades.

Como en cualquier problema de estimación de curvas, la estimación de la densidad espectral correspondiente a un proceso estacionario $X = \{X(t), t \in \mathbb{Z}\}$ puede abordarse desde un punto de vista paramétrico o no paramétrico. En el primer caso, se asume

que la densidad espectral teórica pertenece a alguna familia indexada por un parámetro finito-dimensional como, por ejemplo, a la de los espectros de procesos ARMA(p,q). Los métodos resultantes pueden resultar muy útiles en algunas aplicaciones. Sin embargo, el coste computacional puede llegar a ser excesivo y no toda densidad espectral puede aproximarse de un modo eficiente por modelos ARMA(p,q). En el presente trabajo se utilizará el enfoque no paramétrico, recurriendo a estimadores consistentes de la densidad espectral $f(\cdot)$ obtenidos suavizando el periodograma usando esencialmente técnicas de regresión polinómica local. La estimación se realizará según tres métodos diferentes propuestos por Fan y Kreutzberger (1998), y de los que se incluirá a continuación una breve descripción.

Sea $X = \{X(t), t \in \mathbb{Z}\}$ un proceso lineal gaussiano dado por

$$X(t) = \sum_{j=-\infty}^{\infty} \Psi_j Z(t-j), \quad Z_j \sim iidN(0, \sigma^2)$$

y con densidad espectral

$$f(\lambda) = \frac{\sigma^2}{2\pi} |\Psi(\lambda)|^2, \quad \text{con } \Psi(\lambda) = \sum_{j=-\infty}^{\infty} \Psi_j e^{-ij\lambda}.$$

En primer lugar, y siguiendo a Fan y Kreutzberger (1998), se asumirán las siguientes hipótesis con el fin de establecer las propiedades asintóticas de los estimadores del espectro utilizados:

Hipótesis (H2.1) $X = \{X(t), t \in \mathbb{Z}\}$ es un proceso estacionario con $\sum_j |\Psi_j| j^2 < \infty$. Esta hipótesis garantiza que la densidad espectral tenga una segunda derivada acotada.

Hipótesis (H2.2) Su función de densidad espectral $f(\cdot)$ es positiva en $[0, \pi]$.

Hipótesis (H2.3) K es una función núcleo, función de densidad simétrica con soporte compacto.

Hipótesis (H2.4) $(\log(n))^4 h_n \rightarrow 0$ y $nh_n \rightarrow \infty$ cuando $n \rightarrow \infty$.

A continuación se describirán los tres estimadores del espectro propuestos por Fan y Kreutzberger (1998) y que serán utilizados en el desarrollo de la presente memoria.

a) El periodograma suavizado.

El primer método utilizado para obtener un estimador consistente de la densidad espectral $f(\lambda)$ consiste en construir el suavizador lineal local del periodograma basado en \mathbf{X}_n . Es decir, suavizar directamente los datos $\{\lambda_k, I_n(\lambda_k)\}$, donde $\lambda_k = 2\pi k/n$, con $k = -N, \dots, N$, $N = [(n-1)/2]$, denotan las frecuencias de Fourier, y

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t \exp(-i\lambda t) \right|^2, \quad \lambda \in [-\pi, \pi].$$

Si X es un proceso lineal Gaussiano, se ha visto en (1.23) que las coordenadas del periodograma evaluado sobre las frecuencias de Fourier, $I_n(\lambda_k)$, se distribuyen asintóticamente como una exponencial de media $f(\lambda_k)$ y son aproximadamente independientes. De un modo más formal, siguen el siguiente modelo de regresión heterocedástico:

$$I_k = I_n(\lambda_k) = f(\lambda_k)V_k + R_{n,k}, \quad (1.25)$$

donde $R_{n,k}$ denota un término asintóticamente nulo y las V_k 's son variables con una distribución exponencial estándar e independientes para todo $k \neq 0$. En particular, V_0 y $V_{[n/2]}$ (para n par) siguen una distribución χ_1^2 .

De acuerdo con la Hipótesis (H2.2), la densidad espectral del proceso X es positiva en $[-\pi, \pi]$, y bajo esta propiedad se demuestra que (Kooperberg, Stone y Truong (1995)):

$$\max_{-N \leq k \leq N} |R_{n,k}| = O_P(n^{-1/2} \log n). \quad (1.26)$$

La igualdad (1.26), junto con las Hipótesis (H2.3) y (H2.4) permiten cancelar, desde un punto de vista asintótico, el término $R_{n,k}$ en el modelo de regresión (1.25), resultando en un modelo de regresión estándar. De este modo, si se realiza un ajuste lineal local estándar aplicando mínimos cuadrados locales ponderados se obtiene un estimador de la densidad espectral de la forma:

$$\widehat{f}_{DLS}(\lambda) = \sum_{k=-N}^N \omega_k(\lambda) I_k,$$

siendo $\omega_k(\lambda)$ los pesos del ajuste lineal local, que dependerán de la función núcleo K y el ancho de banda h_n . A lo largo de la presente memoria, nos referiremos a \widehat{f}_{DLS} como el **periodograma suavizado**.

A partir de la teoría general de la regresión polinómica local se deduce que, bajo las condiciones (H2.1)-(H2.4), el estimador periodograma suavizado sigue asintóticamente una distribución normal según:

$$\sqrt{nh_n} \left\{ \widehat{f}_{DLS}(\lambda) - f(\lambda) - \frac{h_n^2}{2} f''(\lambda) \mu_2(K) + o(h_n^2) \right\} \rightarrow N(0, \nu_0(K) f^2(\lambda) \pi). \quad (1.27)$$

b) El log-periodograma suavizado.

El segundo enfoque consiste en suavizar el logaritmo del periodograma utilizando de nuevo la regresión lineal local ponderada. Aplicando una transformación logarítmica al modelo (1.25) se obtiene

$$Y_k = \log(I_n(\lambda_k)) = m(\lambda_k) + \varepsilon_k + r_k, \quad (1.28)$$

donde ahora la función media es $m(\lambda_k) = \log(f(\lambda_k))$, $\varepsilon_k = \log(V_k)$ son variables aleatorias i.i.d. con función de densidad $f_\varepsilon(x) = \exp\{-\exp(x) + x\}$, y $r_k = \log\{1 + R_{n,k}/f(\lambda_k)V_k\}$ denota un término asintóticamente nulo. Además, se sabe que $E(\varepsilon_k) = C_0 = -0,57721$, una constante de Euler, y $Var(\varepsilon_k) = \pi^2/6$.

De modo similar a como ocurría en el caso anterior, la igualdad (1.26), junto con las Hipótesis (H2.3) y (H2.4) permiten cancelar, desde el punto de vista asintótico, el término r_k en el modelo de regresión (1.28), dado que

$$\sum_{k=-N}^N w_k(\lambda)r_k = O_P(n^{-1/2} \log^2 n),$$

uniformemente en $\lambda \in [-\pi, \pi]$. Así, asintóticamente, el término r_k puede ignorarse en (1.28), resultando en un modelo de regresión estándar. De este modo, si se centra el nuevo modelo de regresión restando $C_0 = E(\varepsilon_k)$ en ambos lados de la expresión (1.28) y se desprecia el término r_k se obtiene el modelo aproximado

$$Y_k - C_0 = \log(I_n(\lambda_k)) - C_0 = m(\lambda_k) + (\varepsilon_k - C_0). \quad (1.29)$$

De nuevo puede aplicarse al modelo (1.29) el método de mínimos cuadrados para obtener el mejor ajuste lineal local, resultando el siguiente estimador del logaritmo del espectro

$$\widehat{m}_{LS}(\lambda) = \sum_{k=-N}^N \omega_k(\lambda) (Y_k - C_0), \quad (1.30)$$

donde, de nuevo, $\omega_k(\lambda)$ denota a los pesos del ajuste lineal local correspondiente.

Una nueva transformación de \widehat{m}_{LS} conduce al estimador de la densidad espectral

$$\widehat{f}_{LS}(\lambda) = \exp\{\widehat{m}_{LS}(\lambda)\}, \quad (1.31)$$

al que nos referiremos como **log-periodograma suavizado**.

Los logaritmos de las coordenadas del periodograma $\{I_n(\lambda_k)\}$ no son variables independientes ni idénticamente distribuidas según una distribución log-exponencial. Sin embargo, tal y como se ha apuntado antes, Fan y Kreutzberger (1995) han demostrado que \widehat{m}_{LS} se comporta asintóticamente como si los datos procediesen del modelo (1.29). Más concretamente, bajo las condiciones (H2.1)-(H2.4):

$$\widehat{m}_{LS}(\lambda) - m(\lambda) \sim N(h_n^2 m''(\lambda) \mu_2(K)/2, (\pi^2/6) \nu_0(K) \pi / (nh_n)). \quad (1.32)$$

Para el estimador log-periodograma suavizado \widehat{f}_{LS} se obtiene de (1.31), mediante Taylor:

$$\widehat{f}_{LS}(\lambda) - f(\lambda) = \exp(\widehat{m}_{LS}(\lambda)) - m(\lambda) =$$

$$\begin{aligned}
 &= \exp(m(\lambda))\{\widehat{m}_{LS}(\lambda) - m(\lambda)\} + O_P((\widehat{m}_{LS}(\lambda) - m(\lambda))^2) = \\
 &= \exp(m(\lambda))\{\widehat{m}_{LS}(\lambda) - m(\lambda)\} + O_P(h^4 + (nh_n)^{-1}). \tag{1.33}
 \end{aligned}$$

De donde

$$\sqrt{nh_n} \left\{ \widehat{f}_{LS}(\lambda) - f(\lambda) - \frac{h_n^2}{2} m''(\lambda) f(\lambda) \mu_2(K) + o(h_n^2) \right\} \rightarrow N \left(0, f^2(\lambda) \frac{\pi^2}{6} \nu_0(K) \pi \right). \tag{1.34}$$

De este modo, el parámetro ventana óptimo para la estimación de la densidad espectral, en el sentido de que minimiza el error cuadrático integrado, viene dado por:

$$h_{LS,OPT} = \left[\frac{\nu_0(K)(\pi^2/6)\pi}{\mu_2^2(K) \int_0^\pi \{m''(\lambda)\}^2 d\lambda} \right]^{1/5} n^{-1/5}. \tag{1.35}$$

c) El estimador de máxima verosimilitud local.

Un tercer estimador puede obtenerse suavizando el modelo de regresión dado en (1.28), utilizando el criterio de máxima verosimilitud local en lugar del criterio de mínimos cuadrados.

La función del logaritmo de verosimilitud local ponderada que ha de maximizarse viene dada, para cada λ , por

$$\mathcal{L}(a, b) = \sum_{k=-N}^N [-\exp\{Y_k - a - b(\lambda_k - \lambda)\} + Y_k - a - b(\lambda_k - \lambda)] K_{h_n}(\lambda_k - \lambda), \tag{1.36}$$

donde $K_{h_n}(\cdot) = h_n^{-1} K(h_n^{-1} \cdot)$.

De este modo, si \widehat{m}_{LK} denota el valor de a que maximiza (1.36), el cual puede hallarse por algún procedimiento numérico (e.g. el algoritmo de Newton-Raphson), la tercera propuesta que se utilizará para estimar el espectro por técnicas de ajuste lineal local será:

$$\widehat{f}_{LK}(\lambda) = \exp\{\widehat{m}_{LK}(\lambda)\}.$$

A lo largo de esta memoria, nos referiremos al estimador \widehat{f}_{LK} como **log-periodograma suavizado por máxima verosimilitud local**.

El siguiente resultado, debido a Fan y Kreutzberger (1998), muestra el comportamiento asintótico de dicho estimador.

Teorema 1.3.1 (Fan y Kreutzberger (1998)) *Si se verifican las hipótesis (H2.1)-(H2.4) se tiene, para cada $0 < \lambda < \pi$:*

$$\sqrt{nh_n} \left\{ \widehat{m}_{LK}(\lambda) - m(\lambda) - \frac{h_n^2}{2} m''(\lambda) \mu_2(K) + o(h_n^2) \right\} \rightarrow N(0, \nu_0(K) \pi)$$

Para el estimador de la densidad espectral, $\widehat{f}_{LK} = \exp\{\widehat{m}_{LK}(\lambda)\}$, se tiene que:

Corolario 1.3.1 (Fan y Kreutzberger (1998)) *Bajo las condiciones (H2.1)-(H2.4), se tiene, para cada $0 < \lambda < \pi$*

$$\sqrt{nh} \left\{ \widehat{f}_{LK}(\lambda) - f(\lambda) - h^2 m''(\lambda) f(\lambda) \mu_2(K)/2 + o(h^2) \right\} \rightarrow N(0, \nu_0(K) f^2(\lambda) \pi)$$

De los resultados anteriores se deduce que el estimador basado en la función del logaritmo de la verosimilitud ponderada tiene asintóticamente una varianza menor que el estimador log-periodograma suavizado, mientras que su sesgo asintótico es el mismo. Así mismo, el estimador basado en la función de verosimilitud tiene menor sesgo que el periodograma suavizado en las regiones en las que el logaritmo de la densidad espectral es convexa, no resultando comparables en otro caso. De cualquier modo, ambos estimadores presentan asintóticamente la misma varianza.

Nótese que, según se demuestra en Fan y Kreutzberger (1998), la función log-verosimilitud local (1.36) es también la log-verosimilitud local de la secuencia

$$\{\exp(Y'_k) = \exp(m(\lambda_k) + \varepsilon_k), k = 1, \dots, N\},$$

que son iid según una distribución exponencial. De un modo más detallado:

$$Y'_k = m(\lambda_k) + \varepsilon_k \Rightarrow Y''_k = \exp(Y'_k) = f(\lambda_k) V_k, \quad (1.37)$$

donde V_k es una secuencia de variables aleatorias i.i.d. $\text{Exp}(f(\lambda_k))$. Además, según (1.25), las variables Y''_k coinciden, salvo en un término asintóticamente despreciable, con las ordenadas del periodograma sobre cada una de las frecuencias de Fourier. Por lo tanto, se trata de un caso particular de la familia exponencial, donde el interés se centra en estimar $\eta(\lambda) = g(f(\lambda)) = \log(f(\lambda)) = m(\lambda)$, con la función de cuasi verosimilitud asociada dada por:

$$Q(\mu, y) = \log\left(\frac{y}{\mu}\right) - \frac{y}{\mu}.$$

El teorema anterior se deduce de modo inmediato del Teorema 1.1.2 con $\nu = 0$, $p = 1$ y densidad del diseño equiespaciada.

Denotando por

$$\widehat{\beta}^*_\lambda = a_n^{-1} \left(\widehat{a} - m(\lambda), h_n(\widehat{b} - m'(\lambda)) \right)^t$$

con $a_n = (nh_n)^{-1/2}$ y $(\widehat{a}, \widehat{b})$ maximizando (1.36), entonces

$$\widehat{\beta}^*_\lambda = -\mathbf{A}^{-1} \mathbf{W}_{n,\lambda} + o_P(1)$$

siendo $\mathbf{A} = -\frac{1}{2\pi} \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix}$ y

$$\mathbf{W}_{n,\lambda} = a_n \sum_{k=-N}^N q_1(\bar{m}(\lambda, \lambda_k), e^{Y'_k}) \mathbf{Z}_{k,\lambda} K \left(\frac{\lambda_k - \lambda}{h_n} \right)$$

con

$$\mathbf{Z}_{k,\lambda} = \left(1, \frac{\lambda_k - \lambda}{h_n} \right)^t, \quad k = -N, \dots, N,$$

$$\bar{m}(\lambda, \lambda_k) = m(\lambda) + m'(\lambda)(\lambda_k - \lambda),$$

y

$$q_1(\mu, y) = \left(\frac{\partial}{\partial \mu} \right) Q(g^{-1}(\mu), y) = ye^{-\mu} - 1.$$

Por lo tanto,

$$q_2(\mu, y) = \left(\frac{\partial^2}{\partial \lambda^2} \right) Q(g^{-1}(\mu), y) = -ye^{-\mu}$$

.

Además, se tiene que

$$\mathbf{W}_{n,\lambda} - a_n^{-1} \left\{ \frac{1}{2\pi} \frac{m_1''(\lambda)}{2} h_n^2 (\mu_2(K), 0)^t + o_P(h_n^3) \right\} \rightarrow N(0, \mathbf{B}_\lambda)$$

con

$$\mathbf{B}_\lambda = \frac{1}{2\pi} f_1^2(\lambda) \begin{pmatrix} \mu_0(K^2) & 0 \\ 0 & \mu_2(K^2) \end{pmatrix}.$$

Capítulo 2

Análisis discriminante de series de tiempo: Un enfoque no paramétrico

2.1. Introducción

En el contexto del *análisis discriminante*, también denominado *clasificación supervisada*, el objetivo se centra en definir una regla que permita asignar una nueva observación $\mathbf{y} = (y_1, \dots, y_n)^t$ a una de G categorías mutuamente excluyentes. Las soluciones tradicionalmente empleadas son de corte probabilístico y se basan en utilizar el criterio de Bayes y asignar la nueva observación \mathbf{y} a la categoría para la que se maximice su probabilidad a posteriori. Dicho criterio resulta óptimo, en el sentido de que minimiza la probabilidad global de mala clasificación (Anderson (1984)). Así, denotando por $f_g(\cdot)$, con $g = 1, \dots, G$, la densidad de probabilidad que define a la g -ésima población, y por π_g la probabilidad a priori de dicha categoría, la regla de clasificación de Bayes propone asignar la observación \mathbf{y} a la clase \hat{g} tal que:

$$\hat{g} = \arg \max_{g=1, \dots, G} \{f_g(\mathbf{y})\pi_g\}. \quad (2.1)$$

Para el caso en el que se discrimine entre $G = 2$ categorías, el criterio de Neyman-Pearson se basa en el cociente de verosimilitudes

$$\frac{f_1(\mathbf{y})}{f_2(\mathbf{y})}, \quad (2.2)$$

de forma que la observación \mathbf{y} se asignaría a G_1 si dicho cociente excede un cierto valor K y a G_2 en caso contrario. La ventaja de este criterio de clasificación es que resulta independiente de las probabilidades a priori y que, fijada la probabilidad de error de Tipo

I (equivalentemente, de Tipo II), permitiría minimizar la probabilidad de error del otro tipo sin más que fijar la constante K convenientemente. Es obvio, además, que cuando $K = \pi_2/\pi_1$ la regla discriminante coincide con el criterio de clasificación de Bayes.

Habitualmente se asume que, bajo cada clase, las observaciones siguen una distribución normal multivariante:

$$f_g(\mathbf{y}) = \Phi(\mathbf{y}|\mu_g, \Sigma_g).$$

Si la estructura de covarianzas para las diferentes clases entre las que se discrimina son iguales (es decir, $\Sigma = \Sigma_1 = \Sigma_2 = \dots = \Sigma_G$), el criterio de clasificación se reduce a la regla discriminante lineal de Fisher (*Linear Discriminant Analysis, LDA*). En dicho caso, la regla de clasificación se reduce a comprobar si una combinación lineal de las componentes de \mathbf{y} exceden o no un cierto punto de corte. En otro caso, si las matrices de covarianzas no se asumen iguales, el método resultante es el criterio de discriminación cuadrática (*Quadratic Discriminant Analysis, QDA*), de forma que el estadístico discriminante es una forma cuadrática en las componentes de \mathbf{y} .

En el caso más sencillo en el que se desee discriminar entre $G = 2$ poblaciones, dichas funciones discriminantes vendrán dadas como:

$$(LDA) \quad d_L(y) = (\mu_1 - \mu_2)^t \Sigma^{-1} \mathbf{y} - \frac{1}{2} \mu_1^t \Sigma^{-1} \mu_1 + \frac{1}{2} \mu_2^t \Sigma^{-1} \mu_2$$

$$(QDA) \quad d_Q(y) = \frac{1}{2} \mathbf{y}^t (\Sigma_2^{-1} - \Sigma_1^{-1}) \mathbf{y} + (\mu_1^t \Sigma_1^{-1} - \mu_2^t \Sigma_2^{-1}) \mathbf{y}$$

Obviamente, los métodos de clasificación LDA y QDA presentan ciertas limitaciones al asumir que cada una de las clases se define a partir de una distribución gaussiana. Generalmente, se desconoce tanto $f_g(\cdot)$ como π_g , $g = 1, \dots, G$ y la hipótesis de normalidad puede resultar bastante restrictiva.

En ocasiones se dispone de una serie de observaciones (conjunto de entrenamiento) de las que se conoce a priori el grupo de pertenencia. Si es factible asumir que el conjunto de entrenamiento constituye una muestra aleatoria del total de la población, las probabilidades a priori pueden estimarse fácilmente a partir de la frecuencia relativa de cada clase en la muestra, $\hat{\pi}_g = n_g/n$, donde n_g es el número de datos de la clase g observados en la muestra de entrenamiento. La estimación de las densidades condicionales $f_g(\cdot)$ resulta, en cambio, más compleja.

La solución adoptada con más frecuencia se basa en estimar las densidades condicionales mediante estimadores tipo núcleo (Hand (1982) y Silverman (1986)), que serán luego utilizados en la regla de clasificación (2.1). Sin embargo, son bien conocidas las limitaciones de los métodos tipo núcleo en el contexto multidimensional, así como la alta variabilidad

de las estimaciones que proporcionan para la cola de las densidades, que pueden jugar sin embargo un papel fundamental a la hora de discriminar entre dos poblaciones.

En este sentido se han propuesto recientemente otros métodos de clasificación que resultan más flexibles que los ya mencionados LDA o QDA y pueden solventar al mismo tiempo los problemas que surgen de utilizar los estimadores no paramétricos de las densidades condicionales. Algunos de estos métodos se basan en asumir que la densidad de cada una de las clases entre las que se discrimina es una mixtura de normales:

$$f_g(\mathbf{y}) = \sum_{j=1}^{G_j} \tau_{gj} \Phi(\mathbf{y} | \mu_{gj}, \Sigma_{gj}). \quad (2.3)$$

Bajo esta hipótesis, la aproximación de la densidad requiere estimar los vectores de medias μ_{gj} , las matrices de covarianzas Σ_{gj} y los pesos en las mixturas τ_{gj} para $j = 1, \dots, G_j$ y $g = 1, \dots, G$, lo que supone un número elevado de parámetros y dificulta la estimación en el caso de trabajar con tamaños muestrales reducidos. Siguiendo estas ideas, Hastie y Tibshirani (1996) introdujeron el análisis discriminante mediante mixtura de normales (*Mixture Discriminant Analysis, MDA*), que se basa en la hipótesis (2.3) exigiendo además dos condiciones que facilitan la estimación de los parámetros: la igualdad de las matrices de covarianzas de todas las componentes en cada mixtura (es decir, $\Sigma_{kj} = \Sigma$ para todo k, j) y que el número de componentes en la mixtura de gaussianas sea conocida a priori para cada una de las clases. Estas dos hipótesis han sido a su vez relajadas para dar lugar a métodos de clasificación más flexibles (ver Fraley y Raftery (2002)).

La extensión de cualquiera de estas técnicas de clasificación al ámbito de las series de tiempo resulta una cuestión de gran interés práctico. El problema de discriminar series de tiempo tiene de hecho importantes aplicaciones en un gran número de áreas de conocimiento incluyendo la física, economía, medicina, sismología, ingeniería o zoología, entre otras. Uno de los ejemplos que ha sido tratado de forma reiterada en la literatura es el problema de discriminar entre las ondas expansivas generadas por pequeñas explosiones nucleares y las procedentes de terremotos (Tojstheim (1975), Shumway y Blandford (1993) o Kakizawa, Shumway y Taniguchi (1998)). Otras aplicaciones consisten en la detección de señales dentro de un serie de ruido, para identificar la posición y velocidad de un objetivo (Davenport y Root (1958), Selin (1965), Helstrom (1968), Van Trees (1968)); la detección de patrones de voz (Welch y Wimpres (1961), Markel y Gray (1976), Wolf (1976)) o el procesamiento de imágenes (Huang, Schreiber y Tretiak (1971), Meisel (1972), Rosenfeld y Weszka (1976)). Otra importante aplicación en el campo de la medicina consiste en discriminar entre diferentes clases de registros electroencefalográficos. Este tipo de datos han sido utilizados para distinguir entre diferentes fases de sueño, discriminar entre niveles de anestesia suficientes o insuficientes para una intervención quirúrgica o en la predicción del

inicio de brotes epilépticos (Gevins *et al.* (1975), Gersch y Yonemoto (1977) y Gersch *et al.* (1979)). En Shumway (1982) puede encontrarse una revisión de los distintos ámbitos de aplicación del análisis discriminante de series temporales.

Es importante subrayar que las soluciones clásicas antes descritas pueden no resultar adecuadas en el caso de trabajar con series temporales. Tal y como se ha mencionado, para el caso de discriminar entre dos poblaciones, el estadístico discriminante basado en el cociente de verosimilitudes resulta óptimo en el sentido de minimizar las probabilidades de mala clasificación. De modo general, dicho estadístico se reduce a una forma cuadrática evaluada sobre la serie observada (o a una combinación lineal de sus observaciones en el caso de que las matrices de covarianzas de los procesos entre los que se discrimine sean iguales), definida a partir de la inversa de las matrices de covarianzas. Sin embargo, cuando se trabaja en el contexto de series temporales, la longitud n de las series suele ser grande, de modo que la alta dimensionalidad de las matrices con las que se debe trabajar puede conllevar graves problemas computacionales o incluso hacer imposible el obtener una solución numérica. De igual forma, el cómputo de las probabilidades de error resulta difícil debido a la complejidad de la distribución del estadístico discriminante. El enfoque del problema a través del dominio de frecuencias puede resultar una alternativa interesante para evitar estos problemas. La clave consiste en poder expresar el estadístico basado en la razón de verosimilitudes en términos de los correspondientes espectros, esto es, reemplazar la representación de cada proceso mediante su matriz de covarianzas por la correspondiente representación en términos de su densidad espectral, de forma que el problema de la dimensionalidad quedaría así solventado.

En los últimos años, se han propuesto diversas soluciones al análisis discriminante de series temporales, tanto en el dominio del tiempo como en el dominio de frecuencias (Shumway y Stoffer (2000), Taniguchi y Kakizawa (2000)). En el dominio temporal, el interés se centra en obtener alguna aproximación a la verdadera distribución del cociente de verosimilitudes. Algunas referencias en este sentido son Shumway (1982), donde se propone una aproximación gaussiana en función de la estructura de covarianzas de los procesos entre los que se discrimina, o Chan, Chinipardaz y Cox (1996), donde se proporciona una solución analítica particular para el caso en el que se discrimine entre dos procesos ARMA estacionarios. En este sentido, Dargahi-Noubary (1999) también propone un estadístico discriminante lineal que resulta eficiente para discriminar entre dos procesos gaussianos.

En el dominio de frecuencias, Capon (1965) y Ligget (1971) sugirieron ya la utilización de aproximaciones espectrales en el análisis discriminante de dos procesos gaussianos con igual estructura de covarianzas. Desde entonces, han sido varios los autores que han recurrido al dominio espectral para abordar el análisis discriminante de series de tiempo, generalmente bajo hipótesis de gaussianidad, como Shumway y Unger (1974), Dargahi-

Noubary y Laickock (1981), Shumway (1982), Alagon (1989) o Dargahi-Noubary (1992). Otros trabajos han discutido además la robustez de este tipo de métodos frente a la ausencia de normalidad (Zhang y Taniguchi (1994), Zhang y Taniguchi (1995) y Kakizawa (1996)) y explorado otras medidas de disparidad a partir de las que definir un criterio discriminante distinto (Kakizawa (1996), Krzysko y Wolynski (1997), Kakizawa, Shumway y Taniguchi (1998)).

Una tercera forma de realizar el análisis discriminante de series temporales es el enfoque funcional (Ferraty y Vieu (2000)). La idea consiste en estimar no paramétricamente la probabilidad a posteriori de que la curva observada pertenezca a cada una de las categorías entre las que se discrimina. Desde este punto de vista, cada observación se piensa como una función aleatoria, de modo que ha de definirse una semimétrica funcional específica para calcular el estimador no paramétrico. Ferraty y Vieu (2000) han reflejado el buen comportamiento de su método y su regularidad en diferentes situaciones.

En este capítulo se propondrá un nuevo criterio discriminante para la clasificación de series de tiempo, basado en la medida de disparidad espectral propuesta por Kakizawa (1996) (ver también Kakizawa, Shumway y Taniguchi (1998)). Dicho procedimiento consiste en medir la disparidad entre la densidad espectral bajo cada una de las categorías entre las que se discrimina y un estimador del espectro del proceso a clasificar obtenido a partir de los datos observados. De este modo, el proceso se asignará a la categoría con la que exista menor disparidad.

La aplicación de tal procedimiento exige estimar el espectro a partir de la serie observada, para lo cual se puede utilizar un enfoque bien paramétrico o bien no paramétrico. En el primer caso, se asume que la densidad espectral teórica tiene una forma paramétrica conocida (por ejemplo, la correspondiente a un proceso ARMA) que será estimada a partir de los datos. En este trabajo, se optará por el enfoque no paramétrico y se propondrá la utilización de estimadores basados en técnicas de regresión polinómica local en este contexto. En particular, se utilizarán para la estimación del espectro los estimadores no paramétricos propuestos en Fan y Kreutzberger (1998). Se comparará el comportamiento del estadístico discriminante obtenido al utilizar cada uno de los tres estimadores espectrales, y se demostrará su normalidad asintótica para el caso de discriminar entre procesos lineales gaussianos.

En la siguiente Sección se describirá el procedimiento discriminante basado en la medida general de disparidad espectral propuesta por Kakizawa (1996). A continuación, en la Sección 2.3, se estudiarán las propiedades teóricas de dicho criterio discriminante cuando la densidad espectral del proceso a clasificar se estima utilizando técnicas no paramétricas. En particular, se considerarán las tres vías de suavización local lineal apuntadas por Fan

y Kreutzberger (1998). Se estudiará la distribución asintótica del estadístico discriminante y se computarán las probabilidades de mala clasificación asociadas. Finalmente, en la Sección 2.4, se estudiarán las propiedades teóricas de dicho criterio discriminante en el caso de que se desconozca la expresión de las densidades espectrales teóricas de las clases entre las que se discrimina, y que éstas tengan que ser estimadas a partir de muestras de entrenamiento.

Los resultados expuestos en este capítulo están descritos en Vilar y Pértega (2004).

2.2. Criterio discriminante basado en una medida general de disparidad espectral.

Sea $\mathbf{X}_n = (X_1, \dots, X_n)^t$ una realización parcial de un proceso estacionario escalar con media cero $X = \{X_t, t \in \mathbb{Z}\}$. Se asume que el proceso X pertenece a una de k posibles categorías mutuamente excluyentes $\Pi_1, \Pi_2, \dots, \Pi_k$, donde cada $\Pi_j, j = 1, \dots, k$ especifica una estructura de covarianzas particular y conocida para X . En este contexto, se plantea el problema de clasificar X en una de dichas categorías a partir de la observación de \mathbf{X}_n .

Considérese inicialmente el caso más simple en el que $k = 2$. Una solución clásica consiste en asignar el proceso al grupo bajo el que se maximiza el logaritmo de la función de verosimilitud de \mathbf{X}_n , es decir, evaluar

$$D(\mathbf{X}_n) = D_1(\mathbf{X}_n) - D_2(\mathbf{X}_n) \quad (2.4)$$

con

$$D_j(\mathbf{X}_n) = n^{-1} \log(p_j(\mathbf{X}_n)), \quad j = 1, 2,$$

donde $p_j(\cdot)$ denota la densidad de probabilidad de \mathbf{X}_n bajo $\Pi_j, j = 1, 2$. De acuerdo con este criterio, se asignará \mathbf{X}_n a Π_1 si $D(\mathbf{X}_n) > 0$. En otro caso, \mathbf{X}_n se asignará a Π_2 .

Es bien sabido que el criterio (2.4) permite obtener un método óptimo de clasificación, en el sentido de que minimiza la probabilidad total de error (Anderson (1984)). Sin embargo, en el contexto de series temporales, el estadístico discriminante basado en (2.4) es, en general, bastante complicado de evaluar explícitamente. Sí, por ejemplo, X es un proceso Gaussiano, entonces D_j es una forma cuadrática dada por

$$D_j(\mathbf{X}_n) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_j| - \frac{1}{2} \mathbf{X}_n^t \Sigma_j^{-1} \mathbf{X}_n,$$

donde Σ_j denota la matriz $n \times n$ de covarianzas de \mathbf{X}_n bajo Π_j .

Por lo tanto, $D_j(\mathbf{X}_n)$ puede interpretarse como una especie de distancia, definida en términos de la estructura de covarianzas de X bajo Π_j , entre la j -ésima categoría y los

datos. Si el valor de n es grande, lo cual será habitual tratando con series de tiempo, $D_j(\mathbf{X}_n)$ puede resultar difícil de manejar, ya que requiere trabajar tanto con la inversa como con el determinante de una matriz de covarianzas $n \times n$. Un enfoque alternativo consiste en trabajar en el dominio de frecuencias en lugar de en el dominio temporal, es decir, obtener una medida de disparidad entre los datos y cada hipótesis en términos de la estructura espectral del proceso bajo Π_j , en lugar de hacerlo en términos de la estructura de covarianzas. En ese caso, el problema de la dimensionalidad podría ser solventado. Una discusión muy interesante sobre este punto puede encontrarse en Shumway y Unger (1974).

En esta línea, y para el caso de procesos normales, Ligget (1971) prueba que:

$$n^{-1} \log(p_j(\mathbf{X}_n)) = -\log(2\pi) - \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\log(f_j(\lambda)) + \frac{I_n(\lambda)}{f_j(\lambda)} \right] d\lambda + o_P(1)$$

donde $I_n(\lambda)$ denota el periodograma asociado a \mathbf{X}_n y $f_j(\cdot)$ la densidad espectral bajo la hipótesis Π_j , $j = 1, 2$.

De acuerdo con (2.4), una primera alternativa consiste entonces en considerar el estadístico discriminante

$$D^*(\mathbf{X}_n) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\log\left(\frac{f_2(\lambda)}{f_1(\lambda)}\right) + I_n(\lambda) \left(\frac{1}{f_2(\lambda)} - \frac{1}{f_1(\lambda)}\right) \right] d\lambda$$

como aproximación espectral a $D(\mathbf{X}_n)$ (Dargahi-Noubary y Laicoock (1981), Shumway (1982), Alagon (1989)). Zhang y Taniguchi (1994) han demostrado además la robustez de este criterio de clasificación para el caso de procesos no normales.

A partir del criterio anterior, Kakizawa, Shumway y Taniguchi (1998) proponen un criterio de clasificación más general basándose en los trabajos de Shumway y Unger (1974), Zhang y Taniguchi (1995) y Shumway (1996). En particular, se considera una medida clásica de disparidad entre dos densidades de probabilidad cualesquiera y que puede ser utilizada en este contexto: la información discriminante de Kullback-Leibler (Kullback y Leibler (1951), Kullback (1978)), definida por

$$I_n(p_1, p_2) = n^{-1} E_{p_1} \left\{ \log \left(\frac{p_1(\mathbf{X}_n)}{p_2(\mathbf{X}_n)} \right) \right\}$$

donde E_{p_1} denota la esperanza bajo p_1 . En el caso en el que $p_1(\cdot)$ y $p_2(\cdot)$ correspondan a dos distribuciones normales de media cero, la información discriminante de Kullback-Leibler toma la forma

$$I_n(p_1, p_2) = \frac{1}{2} n^{-1} \left(\text{tr}(\Sigma_1 \Sigma_2^{-1}) - \log \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) - n \right) \quad (2.5)$$

La medida (2.5) puede aproximarse asintóticamente por (Shumway y Unger (1974), Kakazos y Papantoni-Kakazos (1980)):

$$I(f_1, f_2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{f_1(\lambda)}{f_2(\lambda)} - \log \left(\frac{f_1(\lambda)}{f_2(\lambda)} \right) - 1 \right] d\lambda, \quad (2.6)$$

donde $f_1(\lambda)$ y $f_2(\lambda)$ son las densidades espectrales correspondientes.

En ocasiones se suele trabajar también con la tasa de J-divergencia, definida como

$$J_n(p_1, p_2) = I_n(p_1, p_2) + I_n(p_2, p_1),$$

que puede aproximarse asintóticamente, para el caso de procesos Gaussianos, por la expresión

$$J(f_1, f_2) = I(f_1, f_2) + I(f_2, f_1) \quad (2.7)$$

con $I(f_1, f_2)$, $I(f_2, f_1)$ definidas como en (2.6).

De modo análogo y siguiendo a Parzen (1990), Kakizawa, Shumway y Taniguchi (1998) proponen utilizar la información de Chernoff (Chernoff (1952), Renyi (1961)):

$$B_{\alpha, n}(p_1, p_2) = -\log E_{p_1} \left\{ \log \left(\frac{p_2(\mathbf{X}_n)}{p_1(\mathbf{X}_n)} \right)^\alpha \right\}, \quad 0 < \alpha < 1,$$

como una medida de disparidad entre dos densidades de probabilidad. Para el caso de procesos gaussianos de media cero, se tiene que

$$B_{\alpha, n}(p_1, p_2) = \frac{1}{2} \left[\log \left(\frac{|\alpha \Sigma_1 + (1 - \alpha) \Sigma_2|}{|\Sigma_1|} \right) - \alpha \log \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) \right]. \quad (2.8)$$

Asintóticamente, la medida (2.8) puede aproximarse por (Kakazos y Papantoni-Kakazos(1980)):

$$B_\alpha(f_1, f_2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\log \left(\frac{\alpha f_1(\lambda) + (1 - \alpha) f_2(\lambda)}{f_2(\lambda)} \right) - \alpha \log \left(\frac{f_1(\lambda)}{f_2(\lambda)} \right) \right] d\lambda. \quad (2.9)$$

Tanto $I(f_1, f_2)$ como $B_\alpha(f_1, f_2)$ pueden considerarse casos particulares de una medida de disparidad espectral más general dada por:

$$D_W(f, g) = \frac{1}{4\pi} \int_{-\pi}^{\pi} W(f(\lambda)g^{-1}(\lambda)) d\lambda, \quad (2.10)$$

donde $f(\cdot)$ y $g(\cdot)$ son dos densidades espectrales arbitrarias y $W(\cdot)$ es una función que satisface algunas propiedades de regularidad necesarias para asegurar que $D_W(f, g)$ es una cuasi-distancia (esto es, verifica los axiomas de una distancia con excepción del de simetría).

En Kakizawa (1996) se introducen diferentes medidas basadas en la forma (2.10) con el fin de construir un procedimiento general de discriminación espectral que constaría de los siguientes pasos:

1. Obtener una estimación adecuada de la densidad espectral, \hat{f}_n , en base al vector observado \mathbf{X}_n .

2. Evaluar la disparidad entre \mathbf{X}_n y cada una de las dos categorías entre las que se discrimina, Π_1 y Π_2 , mediante el cálculo de $D_W(\widehat{f}_n, f_1)$ y $D_W(\widehat{f}_n, f_2)$, respectivamente, donde f_j denota la densidad espectral del proceso bajo la j -ésima categoría, $j = 1, 2$.
3. Clasificar la serie observada \mathbf{X}_n en Π_1 o en Π_2 según que D_W sea mayor o menor igual que cero, respectivamente, siendo D_W el estadístico discriminante dado por

$$D_W = D_W(\widehat{f}_n, f_2) - D_W(\widehat{f}_n, f_1). \quad (2.11)$$

El criterio se generaliza para discriminar entre k categorías, con $k > 2$, sin mas que asignar \mathbf{X}_n a la categoría Π_j verificando $D_W(\widehat{f}_n, f_i) > D_W(\widehat{f}_n, f_j)$ para todo $i \neq j$.

2.3. Un estadístico discriminante basado en técnicas de regresión polinómica local

La conducta del criterio discriminante basado en el estadístico D_W dado en (2.11) está obviamente condicionada por el estimador empleado para aproximar las densidades espectrales. En esta sección se estudian las propiedades teóricas del criterio discriminante basado en D_W cuando la densidad espectral $f(\lambda)$ se estima utilizando técnicas no paramétricas. Específicamente se propone usar técnicas de regresión polinómica local, de modo que se consideran las tres vías de suavización local lineal apuntadas por Fan y Kreutzberger (1998) y que han sido previamente descritas en la Sección 1.3 de la presente memoria. El objetivo es probar que con los tres estimadores se obtiene la optimalidad asintótica del criterio discriminante y comparar la conducta asintótica de los tres.

2.3.1. Propiedades asintóticas

En orden a establecer las propiedades asintóticas del criterio discriminante propuesto, se requerirá asumir las siguientes hipótesis acerca del proceso X , la función de divergencia W y la ventana h_n y el núcleo K empleados para la aproximación local de los espectros.

Hipótesis (H3.1) El proceso estocástico $X = \{X_t, t \in \mathbb{Z}\}$ es un proceso lineal Gaussiano dado por

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},$$

con $\psi_0 = 1$, $\sum_{j=-\infty}^{\infty} |\psi_j| j^2 < \infty$ y $\{Z_j, j \in \mathbb{Z}\}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas según una $N(0, \sigma^2)$.

Hipótesis (H3.2) Si $f_1(\cdot)$ y $f_2(\cdot)$ denotan las densidades espectrales asociadas al proceso X bajo Π_1 y Π_2 , respectivamente, entonces se verifica:

$$f_1(\lambda) > 0, f_2(\lambda) > 0, \text{ para todo } \lambda \in [-\pi, \pi].$$

Hipótesis (H3.3) La función de divergencia $W(\cdot)$ es al menos tres veces continuamente diferenciable en $(0, \infty)$ con un único mínimo en 1 y tal que $W(1) = 0$.

Hipótesis (H3.4) La función núcleo $K(\cdot)$ es una función de densidad de probabilidad simétrica, acotada y con soporte compacto.

Hipótesis (H3.5) La sucesión de parámetros ventana h_n satisface:

$$(\log n)^4 h_n \rightarrow 0 \text{ y } nh_n^2 \rightarrow \infty \text{ cuando } n \rightarrow \infty.$$

La Hipótesis (H3.1) se requiere para poder asegurar la independencia de las ordenadas del periodograma del proceso de errores Z_t . Tal y como se precisó en (1.18) en la Sección 1.2.2, la independencia se alcanza en sentido asintótico incluso en el caso de que la serie de errores no sea gaussiana y, por tanto, la Hipótesis (H3.1) podría relajarse estableciendo algún tipo de restricción sobre los cumulantes de cuarto orden del proceso de errores. En dicho caso, la varianza asintótica del estadístico discriminante D_W incluiría un término adicional función de esos cumulantes (Kakizawa (1996)).

Por otro lado, de la Hipótesis (H3.1) se deduce además que X es un proceso estacionario, de media cero, varianza finita y densidad espectral dada por (ver (1.19) y (1.20) en la Sección 1.2.3):

$$f(\lambda) = |\psi(\lambda)|^2 \frac{\sigma^2}{2\pi},$$

con

$$\psi(\lambda) = \sum_{j=-\infty}^{\infty} \psi_j \exp(-ij\lambda).$$

De este modo, el requerimiento de sumabilidad $\sum_{j=-\infty}^{\infty} |\psi_j| j^2 < \infty$ garantiza que la densidad espectral del proceso tenga una segunda derivada acotada. Esta premisa es necesaria en los desarrollos asintóticos de los estimadores no paramétricos considerados en esta memoria

(Fan y Kreutzberger (1998)). Con otros estimadores espectrales diferentes la hipótesis de sumabilidad podría ser relajada.

Además, de acuerdo con (1.21) en el Capítulo de Resultados Preliminares, la hipótesis (H3.1) permite caracterizar el periodograma asociado a una realización parcial del proceso, $\mathbf{X}_n = (X_1, \dots, X_n)^t$, del modo que sigue:

$$I_n(\lambda) = f(\lambda) \frac{2\pi}{\sigma_Z^2} I_{n,Z}(\lambda) + R_n(\lambda),$$

con $E \{ |R_n(\lambda)|^2 \} = O(n^{-4})$ uniformemente en λ .

Aunque la densidad espectral de un proceso es siempre una función positiva, la Hipótesis (H3.2) especifica que las densidades espectrales bajo cada una de las hipótesis entre las que se discrimina deben ser estrictamente positivas. Este requerimiento es meramente técnico y tiene por objeto poder trabajar con los cocientes $f_1(\lambda)/f_2(\lambda)$ y $f_2(\lambda)/f_1(\lambda)$, $\lambda \in [-\pi, \pi]$, que aparecen en la definición del estadístico discriminante D_W . Nótese también que la convergencia de la serie $\sum_{j=-\infty}^{\infty} |\psi_j| j^2$ garantiza que $|\psi(\lambda)| < \infty$. Se tiene por tanto asegurada la existencia de dos constantes finitas C_1 y C_2 tales que, para todo $\lambda \in [-\pi, \pi]$, se verifica

$$0 < C_1 \leq f_1(\lambda), f_2(\lambda) \leq C_2. \tag{2.12}$$

La existencia de esta cota superior C_2 será necesaria para poder aplicar el teorema de la convergencia dominada en la demostración de los resultados que se enunciarán más adelante.

La Hipótesis (H3.3) se impone por cuestiones técnicas que surgirán en el desarrollo de las pruebas y, además, con el fin de asegurar que $D_W(\cdot, \cdot)$ sea una cuasi-distancia, es decir, que $D_W(f, g) \geq 0$ para cualesquiera dos densidades espectrales f y g , con $D_W(f, g) = 0$ si y sólo si $f(\lambda) = g(\lambda)$.

Existen múltiples posibles elecciones de W . Algunos ejemplos son:

$$W(x) = x - \log x - 1 \tag{2.13}$$

$$W(x) = \log(\alpha x + (1 - \alpha)) - \alpha \log x \tag{2.14}$$

$$W(x) = \frac{1}{2}(x - 1)^2$$

Como ya se señaló, $I(f, g)$ en (2.6) y $B_\alpha(f, g)$ en (2.9), son casos particulares de $D_W(f, g)$, que se obtienen considerando, respectivamente, las funciones de divergencia (2.13) y (2.14) en la relación anterior.

El resto de las hipótesis impuestas son condiciones de regularidad sobre la función núcleo y la sucesión de ventana. Se trata de hipótesis de regularidad, no en exceso restrictivas y que se precisan también en el trabajo de Fan and Kreutzberger (1998) para establecer las propiedades asintóticas básicas de los tres estimadores espectrales considerados y, en particular, su consistencia uniforme. En particular, para establecer la distribución asintótica del estadístico discriminante se necesita que nh_n^2 vaya a infinito con n (Fan and Kreutzberger requieren $nh_n \rightarrow \infty$) pero, en todo caso, no es una hipótesis restrictiva. Nótese que el orden de convergencia de la ventana óptima es $n^{-1/5}$ y satisface por tanto esta restricción. También es un resultado estándar que nh_n^{2j+1} debe tender a infinito para establecer la consistencia en media cuadrática del estimador por regresión polinómica local de la j -ésima derivada de la función de regresión.

Asumiendo las hipótesis anteriores, se demuestra el siguiente teorema en el que se establece el comportamiento asintótico del estadístico discriminante D_W dado en (2.11) cuando las densidades espectrales se estiman vía regresión polinómica local.

Previamente se introduce la siguiente notación para una presentación más concisa del teorema y de su demostración. En lo que sigue:

- (j, k) denotará el par $(1, 2)$ o el par $(2, 1)$,
- $Q_{j,k}(\lambda) = \frac{1}{f_k(\lambda)} W' \left(\frac{f_j(\lambda)}{f_k(\lambda)} \right)$, donde $W'(\cdot)$ denota la primera derivada de $W(\cdot)$,
- $m_j(\lambda) = \log(f_j(\lambda))$, $j = 1, 2$,
- $\mu_l(L) = \int \lambda^l L(\lambda) d\lambda$, para una función L arbitraria.

Teorema 2.3.1 *Supóngase que se verifican las hipótesis (H3.1), ..., (H3.5), y que $f_1(\lambda) \neq f_2(\lambda)$ en un conjunto de medida de Lebesgue positiva.*

Entonces, bajo Π_j , $j = 1, 2$,

$$D_W \xrightarrow{p} (-1)^{j+1} D_W(f_j, f_k) \quad (2.15)$$

y

$$\sqrt{n} \left\{ D_W + (-1)^j \mu_{W, \hat{f}_n}(j, k) \right\} \xrightarrow{d} N \left(0, V_{W, \hat{f}_n}^2(j, k) \right), \quad (2.16)$$

donde

$$\mu_{W, \hat{f}_n}(j, k) = D_W(f_j, f_k) + \frac{h_n^2}{8\pi} \mu_2(K) \int_{-\pi}^{\pi} Q_{j,k}(\lambda) \varphi_j(\lambda) d\lambda + o(h_n^2), \quad (2.17)$$

con

$$\varphi_j(\lambda) = \begin{cases} f_j''(\lambda) & \text{si } \hat{f}_n = \hat{f}_{DLS} \\ m_j''(\lambda) f_j(\lambda) & \text{si } \hat{f}_n = \hat{f}_{LS} \text{ o } \hat{f}_n = \hat{f}_{LK} \end{cases}$$

y

$$V_{W, \hat{f}_n}^2(j, k) = C \int_{-\pi}^{\pi} [Q_{j,k}(\lambda) f_j(\lambda)]^2 d\lambda + O(h_n^2), \quad (2.18)$$

con $C = (8\pi)^{-1}$ si $\hat{f}_n = \hat{f}_{DLS}$ o $\hat{f}_n = \hat{f}_{LK}$ y $C = (8\pi)^{-1} \pi^2/6$ si $\hat{f}_n = \hat{f}_{LS}$.

Demostración del Teorema 2.3.1

El Teorema 2.3.1 se establece recurriendo a las propiedades asintóticas de cada uno de los estimadores considerados y a las condiciones de regularidad impuestas sobre W y f . Se realizará la prueba del Teorema sólo en el caso en el que se suponga cierta Π_1 , siendo la demostración análoga bajo Π_2 .

a) Convergencia en probabilidad.

En lo que sigue \hat{f}_n denota cualquiera de los tres estimadores locales considerados.

De (2.10) y (2.11) se deduce que el estadístico discriminante D_W viene dado por

$$\begin{aligned} D_W &= D_W(\hat{f}_n, f_2) - D_W(\hat{f}_n, f_1) \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[W\left(\frac{\hat{f}_n(\lambda)}{f_2(\lambda)}\right) - W\left(\frac{\hat{f}_n(\lambda)}{f_1(\lambda)}\right) \right] d\lambda. \end{aligned} \quad (2.19)$$

Resultados previos en la literatura establecen la consistencia uniforme de cualquiera de los tres estimadores del espectro considerados, en particular:

$$\hat{f}_n(\lambda) - f_1(\lambda) = O_P(nh_n)^{-1/2}$$

uniformemente en λ bajo Π_1 . Este resultado puede verse, por ejemplo, en Masry (1996) o en Francisco-Fernández *et al.* (2003) para los estimadores por mínimos cuadrados \hat{f}_{DLS} y \hat{f}_{LS} y en Claeskens y Van Keilegom (2002) para el estimador por verosimilitud local \hat{f}_{LK} .

Por otro lado, y de acuerdo con la hipótesis (H3.3), $W(\cdot)$ es tres veces continuamente diferenciable en $(0, \infty)$, por lo que aplicando el Corolario 5.1.5 de Fuller (1976) se obtiene que:

$$W\left(\frac{\hat{f}_n(\lambda)}{f_2(\lambda)}\right) = W\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) + W'\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) \frac{\hat{f}_n(\lambda) - f_1(\lambda)}{f_2(\lambda)} + O_P(nh_n)^{-1} \quad (2.20)$$

y

$$W \left(\frac{\widehat{f}_n(\lambda)}{f_1(\lambda)} \right) = W(1) + W'(1) \frac{\widehat{f}_n(\lambda) - f_1(\lambda)}{f_1(\lambda)} + O_P(nh_n)^{-1}. \quad (2.21)$$

De acuerdo de nuevo con la Hipótesis (H3.3), W tiene un único mínimo cero en 1, de modo que la expresión (2.21) puede reescribirse de la forma

$$W \left(\frac{\widehat{f}_n(\lambda)}{f_1(\lambda)} \right) = O_P(nh_n)^{-1}. \quad (2.22)$$

Sustituyendo (2.20) y (2.22) en (2.19) se obtiene:

$$\begin{aligned} D_W &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[W \left(\frac{f_1(\lambda)}{f_2(\lambda)} \right) + W' \left(\frac{f_1(\lambda)}{f_2(\lambda)} \right) \frac{\widehat{f}_n(\lambda) - f_1(\lambda)}{f_2(\lambda)} + O_P(nh_n)^{-1} \right] d\lambda \\ &= D_W(f_1, f_2) + \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left(\widehat{f}_n(\lambda) - f_1(\lambda) \right) d\lambda + O_P(nh_n)^{-1} \end{aligned} \quad (2.23)$$

De este modo, se deduce que:

$$\begin{aligned} |D_W - D_W(f_1, f_2)| &\leq \frac{1}{4\pi} \int_{-\pi}^{\pi} |Q_{1,2}(\lambda)| |\widehat{f}_n(\lambda) - f_1(\lambda)| d\lambda + O_P(nh)^{-1} \leq \\ &\leq \frac{1}{4\pi} \int_{-\pi}^{\pi} |Q_{1,2}(\lambda)| \left(\max_{\lambda \in [-\pi, \pi]} |\widehat{f}_n(\lambda) - f_1(\lambda)| \right) d\lambda + O_P(nh)^{-1} \end{aligned}$$

Nótese que W' se supone una función continua y dos veces diferenciable en $(0, +\infty)$ y además según (2.12) existen constantes c y c' tales que:

$$0 < c \leq \frac{f_1(\lambda)}{f_2(\lambda)} \leq c' < \infty, \quad \forall \lambda \in [-\pi, \pi].$$

Ambas condiciones garantizan la existencia de la integral:

$$\int_{-\pi}^{\pi} |Q_{1,2}(\lambda)| d\lambda = \int_{-\pi}^{\pi} W' \left(\frac{f_1(\lambda)}{f_2(\lambda)} \right) \frac{1}{f_2(\lambda)} d\lambda$$

Las Hipótesis (H3.2) y (H3.3) permiten también asegurar que $Q_{1,2}(\lambda)$ está acotada en $[-\pi, \pi]$. Como además se tiene la consistencia uniforme de \widehat{f}_n a f_1 , para cualquiera de los tres estimadores, la aplicación directa del teorema de la convergencia dominada permite establecer la convergencia en probabilidad a cero de la integral en (2.23) y queda así probado (2.15) para $j = 1$.

b) Distribución asintótica.

Para obtener las expresiones asintóticas del error cuadrático medio y la normalidad asintótica del estadístico D_W se seguirá la línea de demostración clásica en el contexto de

estimación núcleo y, en particular, se tendrán en cuenta las propiedades asintóticas de los tres suavizadores lineales locales previamente establecidas por Fan y Kreutzberger (1998).

b.1) Demostración en el caso $\widehat{f}_n = \widehat{f}_{DLS}$.

En primer lugar, se obtendrán las expresiones asintóticas dadas en (2.17) y (2.18) para el sesgo y la varianza de D_W .

La igualdad (2.23) puede reescribirse como

$$D_W = D_W(f_1, f_2) + \Sigma_n + O_P(nh_n)^{-1} \tag{2.24}$$

donde se ha denotado

$$\Sigma_n = \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left(\widehat{f}_{DLS}(\lambda) - f_1(\lambda) \right) d\lambda.$$

El término Σ_n se fragmenta en dos sumandos como sigue:

$$\begin{aligned} \Sigma_n &= \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left\{ \widehat{f}_{DLS}(\lambda) - E\left(\widehat{f}_{DLS}(\lambda)\right) \right\} d\lambda \\ &\quad + \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left\{ E\left(\widehat{f}_{DLS}(\lambda)\right) - f_1(\lambda) \right\} d\lambda \\ &= V + B \end{aligned} \tag{2.25}$$

de modo que de (2.24) y (2.25) se concluye que B y V determinan el sesgo y la varianza de D_W respectivamente.

Recuérdese que en este caso \widehat{f}_{DLS} es el estimador lineal local correspondiente al modelo

$$I_k = I_n(\lambda_k) = f(\lambda_k)V_k + R_{n,k}$$

donde $R_{n,k}$ es un término asintóticamente despreciable y las V_k son variables aleatorias independientes con distribución exponencial.

Utilizando la notación matricial clásica en el contexto de la regresión polinómica local, se puede escribir:

$$\widehat{f}_{DLS}(\lambda) = \mathbf{e}_1^t (\mathbf{\Lambda}_\lambda^t \mathbf{W}_\lambda \mathbf{\Lambda}_\lambda)^{-1} \mathbf{\Lambda}_\lambda^t \mathbf{W}_\lambda \mathbf{I}_n \tag{2.26}$$

y, por lo tanto

$$\widehat{f}_{DLS}(\lambda) - E\left(\widehat{f}_{DLS}(\lambda)\right) = \mathbf{e}_1^t (\mathbf{\Lambda}_\lambda^t \mathbf{W}_\lambda \mathbf{\Lambda}_\lambda)^{-1} \mathbf{\Lambda}_\lambda^t \mathbf{W}_\lambda (\mathbf{I}_n - E(\mathbf{I}_n)), \tag{2.27}$$

donde $\mathbf{e}_1 = (1, 0)^t$, $\mathbf{\Lambda}_\theta, \theta \in [-\pi, \pi]$, es la matriz $n \times 2$ dada por

$$\mathbf{\Lambda}_\theta = \begin{pmatrix} 1 & (\lambda_{-N} - \theta) \\ \vdots & \vdots \\ 1 & (\lambda_N - \theta) \end{pmatrix}$$

siendo λ_j la j -ésima frecuencia de Fourier, $\lambda_j = 2\pi j/n$, y $N = [(n-1)/2]$; $\mathbf{W}_\theta = \text{diag}(w_{-N,\theta}, \dots, w_{0,\theta}, \dots, w_{N,\theta})$, con $w_{j,\theta} = h_n^{-1}K(h_n^{-1}(\lambda_j - \theta))$ y

$$\mathbf{I}_n = (I_n(\lambda_{-N}), \dots, I_n(0), \dots, I_n(\lambda_N))^t.$$

Sea $\mathbf{S}_{n,\theta}$ la matriz dada por

$$\mathbf{S}_{n,\theta} = \begin{pmatrix} s_{0,0}^n(\theta) & s_{0,1}^n(\theta) \\ s_{1,0}^n(\theta) & s_{1,1}^n(\theta) \end{pmatrix} = \frac{1}{n} \mathbf{\Lambda}_\theta^t \mathbf{W}_\theta \mathbf{\Lambda}_\theta,$$

donde el elemento $s_{i,j}^n(\theta) = s_{i+j}^n(\theta)$ viene dado por

$$s_l^n(\theta) = \frac{1}{n} \sum_{t=-N}^N (\lambda_t - \theta)^l K_{h_n}(\lambda_t - \theta), \text{ para } l = 0, 1, 2; \quad (2.28)$$

y sea $\mathbf{T}_{n,\theta}$ el vector

$$\mathbf{T}_{n,\theta} = (t_0^n(\theta), t_1^n(\theta))^t = \frac{1}{n} \mathbf{\Lambda}_\theta^t \mathbf{W}_\theta (\mathbf{I}_n - E(\mathbf{I}_n)),$$

de elementos

$$t_l^n(\theta) = \frac{1}{n} \sum_{t=-N}^N (\lambda_t - \theta)^l K_{h_n}(\lambda_t - \theta) (I_n(\lambda_t) - E(I_n(\lambda_t))), \quad l = 0, 1. \quad (2.29)$$

Se tiene entonces que (2.26) y (2.27) pueden escribirse:

$$\widehat{f}_{DLS}(\lambda) = \mathbf{e}_1^t n \mathbf{S}_{n,\lambda}^{-1} \mathbf{\Lambda}_\lambda^t \mathbf{W}_\lambda \mathbf{I}_n \quad (2.30)$$

$$\widehat{f}_{DLS}(\lambda) - E(\widehat{f}_{DLS}(\lambda)) = \mathbf{e}_1^t \mathbf{S}_{n,\lambda}^{-1} \mathbf{T}_{n,\lambda}. \quad (2.31)$$

Además, si \mathbf{H}_n denota la matriz $\text{diag}(1, h_n)$, un resultado estándar en el contexto de la regresión polinómica local es que

$$\lim_{n \rightarrow \infty} \mathbf{H}_n^{-1} \mathbf{S}_{n,\theta} \mathbf{H}_n^{-1} = \frac{1}{2\pi} \mathbf{S} = \frac{1}{2\pi} \begin{pmatrix} s_{0,0} & s_{0,1} \\ s_{1,0} & s_{1,1} \end{pmatrix}, \quad (2.32)$$

donde $s_{i,j} = \mu_{i+j}(K)$, para $i, j = 0, 1$. El resultado (2.32) es de hecho una de las formas del conocido Lema de Bochner (ver, por ejemplo, Parzen (1964)).

Paso 1. En primer lugar, nos centraremos en el término correspondiente al sesgo del estadístico discriminante. De (2.25) se deduce que:

$$\begin{aligned} E(\Sigma_n) &= E(V) + E(B) = E(B) = B = \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left\{ E(\widehat{f}_{DLS}(\lambda)) - f_1(\lambda) \right\} d\lambda \end{aligned}$$

A partir de (2.30) se tiene:

$$E\left(\widehat{f}_{DLS}(\lambda)\right) = f_1(\lambda) + \mathbf{e}_1^t n \mathbf{S}_{n,\lambda}^{-1} \mathbf{\Lambda}_\lambda^t \mathbf{W}_\lambda \left(E(\mathbf{I}_n) - \mathbf{\Lambda}_\lambda (f_1(\lambda), f_1'(\lambda))^t\right). \quad (2.33)$$

Puesto que $E(I_n(\lambda_k)) = f_1(\lambda_k), k = -N, \dots, N$ uniformemente en λ y, utilizando un desarrollo de Taylor:

$$f_1(\lambda_k) = f_1(\lambda) + f_1'(\lambda)(\lambda_k - \lambda) + \frac{f_1''(\lambda)}{2}(\lambda_k - \lambda)^2 + o_P((\lambda_k - \lambda)^2), \quad k = -N, \dots, N,$$

la expresión (2.33) puede escribirse

$$E\left(\widehat{f}_{DLS}(\lambda)\right) = f_1(\lambda) + \mathbf{e}_1^t n \mathbf{S}_{n,\lambda}^{-1} \mathbf{\Lambda}_\lambda^t \mathbf{W}_\lambda \left(\frac{1}{n} \frac{f_1''(\lambda)}{2} \mathbf{c}_n + o(nh_n^2)\right), \quad (2.34)$$

siendo $\mathbf{c}_n = (s_2^n(\lambda), s_3^n(\lambda))^t$, con $s_j^n(\lambda), j = 2, 3$ definidos como en (2.28). Utilizando (2.32) se obtiene de (2.34):

$$E\left(\widehat{f}_{DLS}(\lambda)\right) = f_1(\lambda) + \frac{h_n^2}{2} \mu_2(K) f_1''(\lambda) + o(h_n^2)$$

uniformemente en $\lambda \in [-\pi, \pi]$.

A partir de la expresión anterior, la Hipótesis (H3.5) y el teorema de la convergencia dominada conducen a:

$$\begin{aligned} E(\Sigma_n) &= \frac{1}{4\pi} \left[\int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1''(\lambda) d\lambda \right] \frac{1}{2} \mu_2(K) h_n^2 (1 + o(1)) = \\ &= \frac{1}{8\pi} \mu_2(K) h_n^2 \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1''(\lambda) d\lambda + o(h_n^2) \end{aligned} \quad (2.35)$$

Nótese que el término de la integral que aparece en (2.35) existe como consecuencia de la existencia y acotamiento de la segunda derivada de f_1 en $[-\pi, \pi]$. El resultado (2.17) queda así probado.

Paso 2. Se considera ahora el término correspondiente a la varianza. A partir de (2.24) y (2.25) se obtiene:

$$Var(\Sigma_n) = E(V^2) = \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') \Gamma_n(\lambda, \lambda') d\lambda d\lambda' \quad (2.36)$$

donde

$$\Gamma_n(\lambda, \lambda') = E\left\{ \left[\widehat{f}_{DLS}(\lambda) - E\left(\widehat{f}_{DLS}(\lambda)\right) \right] \left[\widehat{f}_{DLS}(\lambda') - E\left(\widehat{f}_{DLS}(\lambda')\right) \right] \right\}. \quad (2.37)$$

De (2.36), (2.37) y (2.31) se deduce:

$$Var(\Sigma_n) = \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') \mathbf{e}_1^t \mathbf{S}_{n,\lambda}^{-1} E\{\mathbf{T}_{n,\lambda} \mathbf{T}_{n,\lambda'}^t\} \mathbf{S}_{n,\lambda'}^{-1} \mathbf{e}_1 d\lambda d\lambda'.$$

Por tanto,

$$\begin{aligned} \text{Var}(\mathbf{H}_n \Sigma_n) &= \\ \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') \mathbf{e}_1^t \mathbf{H}_n \mathbf{S}_{n,\lambda}^{-1} \mathbf{H}_n \mathbf{V}_{n,\lambda,\lambda'} \mathbf{H}_n \mathbf{S}_{n,\lambda'}^{-1} \mathbf{H}_n \mathbf{e}_1 d\lambda d\lambda', \end{aligned} \quad (2.38)$$

donde $\mathbf{V}_{n,\lambda,\lambda'}$ es la matriz 2×2 definida por

$$\mathbf{V}_{n,\lambda,\lambda'} = \mathbf{H}_n^{-1} E \{ \mathbf{T}_{n,\lambda} \mathbf{T}_{n,\lambda'}^t \} \mathbf{H}_n^{-1} = \begin{pmatrix} v_{0,0}^n(\lambda, \lambda') & v_{0,1}^n(\lambda, \lambda') \\ v_{1,0}^n(\lambda, \lambda') & v_{1,1}^n(\lambda, \lambda') \end{pmatrix}, \quad (2.39)$$

de elementos

$$\begin{aligned} v_{i,j}^n(\lambda, \lambda') &= n^{-2} \sum_{t=-N}^N \sum_{s=-N}^N \left(\frac{\lambda_t - \lambda}{h_n} \right)^i \left(\frac{\lambda_s - \lambda'}{h_n} \right)^j \\ &\quad \times K_{h_n}(\lambda_t - \lambda) K_{h_n}(\lambda_s - \lambda') \gamma_n(\lambda_t, \lambda_s), \end{aligned} \quad (2.40)$$

donde

$$\begin{aligned} \gamma_n(\lambda_t, \lambda_s) &= E \{ (I_n(\lambda_t) - E(I_n(\lambda_t))) (I_n(\lambda_s) - E(I_n(\lambda_s))) \} = \\ &= \text{Cov}(I_n(\lambda_t), I_n(\lambda_s)). \end{aligned}$$

Al asumir que X es un proceso lineal Gaussiano de media cero, las ordenadas del periodograma sobre las frecuencias de Fourier son asintóticamente incorreladas y satisfacen (ver 1.24):

$$\gamma_n(\lambda_t, \lambda_s) = f_1(\lambda_t) f_1(\lambda_s) 1_{\{\lambda_t = \lambda_s\}} + o(n^{-1})$$

uniformemente cuando $n \rightarrow \infty$.

Así, el término dominante en (2.40) toma la forma

$$\begin{aligned} v_{i,j}^n(\lambda, \lambda') &= n^{-2} \sum_{t=-N}^N \left(\frac{\lambda_t - \lambda}{h_n} \right)^i \left(\frac{\lambda_t - \lambda'}{h_n} \right)^j \\ &\quad \times K_{h_n}(\lambda_t - \lambda) K_{h_n}(\lambda_t - \lambda') (f_1^2(\lambda_t) + o(1)), \end{aligned} \quad (2.41)$$

para $i, j \in \{0, 1\}$.

Considérese, para $i, j \in \{0, 1\}$, el término $I_{i,j}^n$ dado por

$$I_{i,j}^n = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') v_{i,j}^n(\lambda, \lambda') d\lambda d\lambda'.$$

De acuerdo con (2.41) se tiene que:

$$\begin{aligned}
 I_{i,j}^n &= n^{-2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') \left(\sum_{t=-N}^N \left(\frac{\lambda_t - \lambda}{h_n} \right)^i \left(\frac{\lambda_t - \lambda'}{h_n} \right)^j \right. \\
 &\quad \times K_{h_n}(\lambda_t - \lambda) K_{h_n}(\lambda_t - \lambda') (f_1^2(\lambda_t) + o(1)) \left. d\lambda d\lambda' \right) \\
 &= n^{-2} \left\{ \sum_{t=-N}^N f_1^2(\lambda_t) \left(\int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left(\frac{\lambda_t - \lambda}{h_n} \right)^i K_{h_n}(\lambda_t - \lambda) d\lambda \right) \right. \\
 &\quad \times \left. \left(\int_{-\pi}^{\pi} Q_{1,2}(\lambda') \left(\frac{\lambda_t - \lambda'}{h_n} \right)^j K_{h_n}(\lambda_t - \lambda') d\lambda' \right) \right\} (1 + o(1)).
 \end{aligned} \tag{2.42}$$

Realizando los cambios de variable $\lambda_t - \lambda = h_n u$ y $\lambda_t - \lambda' = h_n v$ y aproximando el sumatorio por una integral se sigue

$$\begin{aligned}
 2\pi n I_{i,j}^n &= \int_{-\pi}^{\pi} f_1^2(s) \left(\int_{h_n^{-1}(s-\pi)}^{h_n^{-1}(s+\pi)} Q_{1,2}(s - h_n u) u^i K(u) du \right) \\
 &\quad \times \left(\int_{h_n^{-1}(s-\pi)}^{h_n^{-1}(s+\pi)} Q_{1,2}(s - h_n v) v^j K(v) dv \right) ds (1 + o(1)).
 \end{aligned} \tag{2.43}$$

Las condiciones de regularidad impuestas sobre la función W garantizan que $Q_{1,2}(\cdot)$ es tres veces diferenciable con continuidad. Por tanto, un desarrollo de Taylor de segundo orden en s permite concluir que

$$\lim_{n \rightarrow \infty} \int_{h_n^{-1}(s-\pi)}^{h_n^{-1}(s+\pi)} Q_{1,2}(s - h_n u) u^i K(u) du = Q_{1,2}(s) \mu_i(K). \tag{2.44}$$

Los resultados asintóticos sobre los elementos de $\mathbf{H}_n^{-1} \mathbf{S}_{n,\lambda} \mathbf{H}_n^{-1}$ en (2.32) y sobre $I_{i,j}^n$ en (2.43) y (2.44) permiten aplicar el teorema de la convergencia dominada en (2.38) y obtener

$$8\pi n \text{Var}(\Sigma_n) = \int_{-\pi}^{\pi} f_1^2(u) Q_{1,2}^2(u) du + O(h_n^2) \tag{2.45}$$

estableciéndose así (2.18).

Paso 3. Para completar la demostración, se prueba a continuación la normalidad asintótica de D_W .

De acuerdo con la notación empleada, bajo Π_1 , se tiene

$$\begin{aligned}
 \sqrt{n} \left\{ D_W - \mu_{W, \hat{J}_{DLS}}(1, 2) \right\} &= \frac{\sqrt{n}}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) e_1^t \mathbf{S}_{n,\lambda}^{-1} \mathbf{T}_{n,\lambda} d\lambda (1 + o(1)) = \\
 &= \frac{\sqrt{n}}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) e_1^t \mathbf{H}_n^{-1} \mathbf{H}_n \mathbf{S}_{n,\lambda}^{-1} \mathbf{H}_n \mathbf{H}_n^{-1} \mathbf{T}_{n,\lambda} d\lambda (1 + o(1)).
 \end{aligned} \tag{2.46}$$

En consecuencia, la normalidad asintótica de $\sqrt{n} \left\{ D_W - \mu_{W, \hat{f}_{DLS}}(1, 2) \right\}$ se deduce de (2.32) y (2.46) una vez que se haya establecido la normalidad asintótica de la variable aleatoria

$$\theta_n = \sqrt{n} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) (c_0 \tilde{t}_0^n(\lambda) + c_1 \tilde{t}_1^n(\lambda)) d\lambda, \quad (2.47)$$

siendo c_0, c_1 constantes arbitrarias y, para $l = 0, 1$, $\tilde{t}_l^n(\lambda)$ dado por:

$$\tilde{t}_l^n(\lambda) = \frac{1}{n} \sum_{t=-N}^N \left(\frac{\lambda_t - \lambda}{h_n} \right)^l K_{h_n}(\lambda_t - \lambda) (I_n(\lambda_t) - E(I_n(\lambda_t))).$$

Nótese que θ_n puede escribirse como una suma de variables aleatorias de la forma

$$\theta_n = \frac{1}{\sqrt{n}} \sum_{t=-N}^N \eta_t, \quad (2.48)$$

donde

$$\eta_t = (I_n(\lambda_t) - E(I_n(\lambda_t))) \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \xi_{h_n}(\lambda_t - \lambda) d\lambda,$$

siendo $\xi_{h_n}(u) = \xi(u/h_n)/h_n$ y $\xi(u) = c_0 K(u) + c_1 u K(u)$.

Argumentos análogos a los empleados en (2.43) y (2.44) permiten demostrar que las variables η_t son tales que $Cov(\eta_t, \eta_s) = o(1)$ y

$$Var(\eta_t) = \left(Q_{1,2}(\lambda_t) f_1(\lambda_t) \int_{-\pi}^{\pi} \xi(u) du \right)^2 (1 + o(1)).$$

De este modo, aproximando de nuevo el sumatorio por una integral se obtiene

$$\begin{aligned} Var(\theta_n) &= \frac{1}{n} \sum_{t=-N}^N E(\eta_t^2) \\ &= \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} Q_{1,2}^2(u) f_1^2(u) du \right) \left(\int_{-\pi}^{\pi} \xi(u) du \right)^2 (1 + o(1)) \\ &= \sigma_{\theta}^2 (1 + o(1)). \end{aligned} \quad (2.49)$$

La normalidad asintótica de θ_n se establece finalmente mostrando que las variables $\{\eta_t\}$ satisfacen las condiciones estándar de Lindeberg-Feller para normalidad asintótica bajo condiciones de independencia. Dichas condiciones son

$$\frac{1}{n} \sum_{t=-N}^N E(\eta_t^2) \longrightarrow \sigma_{\theta}^2 = \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} \xi(u) du \right)^2 \int_{-\pi}^{\pi} Q_{1,2}^2(u) f_1^2(u) du \quad (2.50)$$

y

$$\frac{1}{n} \sum_{t=-N}^N E\left(\eta_t^2 1_{\{|\eta_t| \geq \varepsilon \sigma_{\theta} n^{1/2}\}}\right) \longrightarrow 0 \quad \text{para todo } \varepsilon > 0. \quad (2.51)$$

La primera condición se sigue inmediatamente de (2.48) y (2.49). Para probar la segunda condición se precisa usar un argumento de truncamiento debido a que las ordenadas del periodograma $I_n(\lambda_t)$ no están necesariamente acotadas.

Sea M un punto de truncamiento fijo y denotemos

$$I_n^M(\lambda_t) = I_n(\lambda_t)1_{\{|I_n(\lambda_t)| \leq M\}}.$$

En lo que sigue, se añadirá un superíndice M para indicar las cantidades computadas con $I_n^M(\lambda_t)$ en lugar de con $I_n(\lambda_t)$. Entonces $\theta_n = \theta_n^M + \bar{\theta}_n^M$, donde

$$\bar{\theta}_n^M = \frac{1}{\sqrt{n}} \sum_{t=-N}^N (\eta_t - \eta_t^M).$$

Puesto que la función núcleo K es una función acotada con soporte compacto (hipótesis H3.4), $\xi(\cdot)$ es acotada, luego existe una constante c tal que se verifica

$$|\eta_t^M| \leq ch_n^{-1}.$$

Por lo tanto, en base a la hipótesis H3.5, se concluye

$$\max_{-N \leq t \leq N} |\eta_t^M| \frac{1}{\sqrt{n}} \leq c \frac{1}{\sqrt{nh_n^2}} \rightarrow 0.$$

Es decir, cuando la longitud de la serie de tiempo, n , es grande, el conjunto $\{|\eta_t^M| \geq \varepsilon\sigma_{\theta, M}\}$ es un conjunto vacío y la condición de Lindeberg-Feller (2.51) se satisface para $\{\eta_t^M\}$, luego

$$\theta_n^M \rightarrow N(0, \sigma_{\theta, M}^2). \tag{2.52}$$

Para completar la prueba es ahora suficiente mostrar que, cuando primero $n \rightarrow \infty$ y después $M \rightarrow \infty$, se tiene

$$Var(\bar{\theta}_n^M) \rightarrow 0. \tag{2.53}$$

En efecto, si (2.53) es cierta, entonces

$$\varphi_{\theta_n}(t) \rightarrow \varphi_Z^{\sigma_\theta^2}(t) \text{ cuando } n \rightarrow \infty,$$

donde $\varphi_{\theta_n}(t)$ y $\varphi_Z^{\sigma_\theta^2}(t)$ denotan las funciones características de θ_n y de una variable aleatoria $N(0, \sigma_\theta^2)$ respectivamente. Obsérvese que:

$$\begin{aligned} & |E \exp(it\theta_n) - \exp(-t^2\sigma_\theta^2/2)| \leq \\ & \leq |E \exp(it\theta_n^M) (\exp(it\bar{\theta}_n^M) - 1)| \\ & + |E \exp(it\theta_n^M) - \exp(-t^2\sigma_{\theta, M}^2/2)| \\ & + |\exp(-t^2\sigma_{\theta, M}^2/2) - \exp(-t^2\sigma_\theta^2/2)|. \end{aligned} \tag{2.54}$$

El primer término está acotado por

$$E \left| \exp \left(it\bar{\theta}_n^M \right) - 1 \right| = O \left(\text{Var} \left(\bar{\theta}_n^M \right) \right),$$

y, por tanto, va a cero si (2.53) es cierto.

La convergencia a cero del segundo término se sigue inmediatamente de (2.52) y del teorema de Levi, para cualquier $M > 0$. Finalmente, el tercer término va a cero cuando $M \rightarrow \infty$ por el teorema de la convergencia dominada.

En consecuencia, sólo resta probar (2.53). Nótese que $\bar{\theta}_n^M$ y θ_n tienen la misma estructura. Por ello, razonando como en (2.49) se obtiene

$$\begin{aligned} \text{Var}(\bar{\theta}_n^M) &= \frac{1}{n} \sum_{t=-N}^N \bar{\eta}_t^{M^2} = \frac{1}{n} \sum_{t=-N}^N E \left((I_n(\lambda_t) - E(I_n(\lambda_t))) I(|I_n(\lambda_t)| > M) \right)^2 \times \\ &\times \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \xi_{h_n}(\lambda_t - \lambda) d\lambda. \end{aligned} \quad (2.55)$$

El primer término en la expresión anterior converge a 0 cuando $M \rightarrow \infty$, mientras que el segundo término está acotado, luego el teorema de la convergencia dominada permite concluir que $\text{Var}(\bar{\theta}_n^M)$ converge a 0, demostrándose así (2.53).

b.2) Demostración en el caso $\hat{f}_n = \hat{f}_{LS}$.

Razonando como antes, la igualdad (2.23) puede reescribirse como

$$D_W = D_W(f_1, f_2) + \Sigma_n + O_P(nh_n)^{-1}$$

donde ahora

$$\Sigma_n = \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left(\hat{f}_{LS}(\lambda) - f_1(\lambda) \right) d\lambda. \quad (2.56)$$

Utilizando los resultados asintóticos obtenidos para \hat{m}_{LS} en (1.32) se obtiene

$$\begin{aligned} \hat{f}_{LS}(\lambda) - f_1(\lambda) &= \exp \{m_1(\lambda)\} \{ \hat{m}_{LS}(\lambda) - m_1(\lambda) \} \\ &+ O_P(h_n^4 + (nh_n)^{-1}). \end{aligned} \quad (2.57)$$

Sustituyendo (2.57) en (2.56) se tiene que

$$D_W = D_W(f_1, f_2) + \Sigma_n^* + O_P(nh_n)^{-1} \quad (2.58)$$

con

$$\Sigma_n^* = \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) (\hat{m}_{LS}(\lambda) - m_1(\lambda)) d\lambda.$$

La expresión obtenida para Σ_n^* puede dividirse en dos términos de modo similar a como ocurría con Σ_n en (2.25):

$$\begin{aligned} \Sigma_n^* &= \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) \{\widehat{m}_{LS}(\lambda) - E(\widehat{m}_{LS}(\lambda))\} d\lambda \\ &+ \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) \{E(\widehat{m}_{LS}(\lambda)) - m_1(\lambda)\} d\lambda = V^* + B^*, \end{aligned} \quad (2.59)$$

de modo que B^* y V^* determinan el sesgo y la varianza de D_W , respectivamente.

Recuérdese que, en este caso, \widehat{m}_{LS} denota el estimador lineal local mínimo cuadrático correspondiente al modelo

$$Y_k = \log(I_n(\lambda_k)) = m_1(\lambda_k) + \varepsilon_k + r_k \quad (2.60)$$

una vez centrado restando $C_0 = E(\varepsilon_k) = -0,57721$ en ambos lados de la ecuación (2.60). Aquí $m_1(\lambda_k) = \log(f_1(\lambda_k))$, ε_k son variables aleatorias iid con función de densidad $f_\varepsilon(x) = \exp\{-\exp(x) + x\}$ y $r_k = \log[1 + R_{n,k}/\{f_1(\lambda_k)V_k\}]$ denota un término despreciable asintóticamente. Se sabe, además, que $Var(\varepsilon_k) = \frac{\pi^2}{6}$.

De hecho, el Teorema 6.4 en Fan y Gijbels (1996) establece que:

$$\widehat{m}_{LS}(\lambda) = \sum_{t=-N}^N K_{h_n}(\lambda_t - \lambda)(Y_k - r_k) + O\left(\frac{\log^2 n}{\sqrt{n}}\right) = \sum_{t=-N}^N K_{h_n}(\lambda_t - \lambda)Y'_k + O\left(\frac{\log^2 n}{\sqrt{n}}\right),$$

de modo que, en adelante, la demostración se continuará para el primer término en la expresión anterior, que se seguirá denotando por $\widehat{m}_{LS}(\lambda)$.

Recurriendo de nuevo a la notación matricial se puede escribir:

$$\widehat{m}_{LS}(\lambda) = e_1^t n \mathbf{S}_{n,\lambda}^{-1} \mathbf{A}_\lambda^t \mathbf{W}_\lambda \mathbf{Y}_n$$

$$\widehat{m}_{LS}(\lambda) - E(\widehat{m}_{LS}(\lambda)) = e_1^t \mathbf{S}_{n,\lambda}^{-1} \mathbf{T}_{n,\lambda}^*$$

donde la matriz $\mathbf{S}_{n,\lambda}$ se define como en (2.28), $\mathbf{Y}_n = (Y_{-N}, \dots, Y_0, \dots, Y_N)^t$ y $\mathbf{T}_{n,\theta}^* = (t_0^{*n}(\theta), t_1^{*n}(\theta))^t$, $\theta \in [-\pi, \pi]$ es ahora el vector de elementos

$$t_l^{*n}(\theta) = \frac{1}{n} \sum_{t=-N}^N (\lambda_t - \theta)^l K_{h_n}(\lambda_t - \theta) (Y_t - E(Y_t)), \quad l = 0, 1.$$

Paso 1. En primer lugar nos centraremos en el término correspondiente al sesgo del estadístico discriminante:

$$E(\Sigma_n^*) = B^* = \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) \{E(\widehat{m}_{LS}(\lambda)) - m_1(\lambda)\} d\lambda$$

En este caso, $\widehat{m}_{LS}(\lambda)$ es el estimador lineal local de $m_1(\lambda_k) = \log(f_1(\lambda_k))$. Procedimiento igual que en el caso anterior, se llega a la siguiente expresión del sesgo asintótico de \widehat{m}_{LS} bajo Π_1 :

$$E(\widehat{m}_{LS}(\lambda)) = m_1(\lambda) + \frac{h_n^2}{2} \mu_2(K) m_1''(\lambda) + o(h_n^2)$$

Puesto que bajo Π_1 , $\widehat{m}_{LS}(\lambda)$ converge uniformemente a $m_1(\lambda)$ en $\lambda \in [-\pi, \pi]$, utilizando la Hipótesis (H3.5) y el teorema de la convergencia dominada se obtiene:

$$\begin{aligned} E(\Sigma_n^*) &= \frac{1}{4\pi} \left[\int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) m_1''(\lambda) d\lambda \right] \frac{1}{2} \mu_2(K) h_n^2 (1 + o(1)) = \\ &= \frac{1}{8\pi} \mu_2(K) h_n^2 \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) m_1''(\lambda) d\lambda + o(h_n^2). \end{aligned} \quad (2.61)$$

De nuevo, el término de la integral que aparece en (2.61) existe como consecuencia de la existencia y acotamiento de la segunda derivada de f_1 en $[-\pi, \pi]$. El resultado (2.17) queda así probado para $\widehat{f}_n = \widehat{f}_{LS}$.

Paso 2. Consideraremos ahora el término correspondiente a la varianza. A partir de (2.58) y (2.59) se obtiene que

$$Var(\Sigma_n^*) = \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') f_1(\lambda) f_1(\lambda') \Gamma_n^*(\lambda, \lambda') d\lambda d\lambda' \quad (2.62)$$

donde

$$\Gamma_n^*(\lambda, \lambda') = E \{ [\widehat{m}_{LS}(\lambda) - E(\widehat{m}_{LS}(\lambda))] [\widehat{m}_{LS}(\lambda') - E(\widehat{m}_{LS}(\lambda'))] \}. \quad (2.63)$$

Recurriendo de nuevo a la notación matricial se puede escribir:

$$\widehat{m}_{LS}(\lambda) - E(\widehat{m}_{LS}(\lambda)) = \mathbf{e}_1^t \mathbf{S}_{n,\lambda}^{-1} \mathbf{T}_{n,\lambda}^*$$

donde la matriz $\mathbf{S}_{n,\lambda}$ se define como en (2.28), y $\mathbf{T}_{n,\theta}^* = (t_0^{*n}(\theta), t_1^{*n}(\theta))^t$, $\theta \in [-\pi, \pi]$ es ahora el vector de elementos

$$t_l^{*n}(\theta) = \frac{1}{n} \sum_{t=-N}^N (\lambda_t - \theta)^l K_{h_n}(\lambda_t - \theta) (Y_t - E(Y_t)), \quad l = 0, 1.$$

Utilizando esta notación, (2.62) puede reescribirse como:

$$Var(\Sigma_n^*) = \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') f_1(\lambda) f_1(\lambda') \mathbf{e}_1^t \mathbf{S}_{n,\lambda}^{-1} E \{ \mathbf{T}_{n,\lambda}^* \mathbf{T}_{n,\lambda'}^{*t} \} \mathbf{S}_{n,\lambda'}^{-1} \mathbf{e}_1 d\lambda d\lambda'.$$

Por tanto,

$$Var(\mathbf{H}_n \Sigma_n^*) =$$

$$= \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') f_1(\lambda) f_1(\lambda') e_1^t \mathbf{H}_n \mathbf{S}_{n,\lambda}^{-1} \mathbf{H}_n \mathbf{V}_{n,\lambda,\lambda'}^* \mathbf{H}_n \mathbf{S}_{n,\lambda'}^{-1} \mathbf{H}_n \mathbf{e}_1 d\lambda d\lambda', \quad (2.64)$$

siendo $\mathbf{V}_{n,\lambda,\lambda'}^*$ la matriz 2×2 definida de forma análoga a (2.39), con elementos

$$v_{i,j}^{*n}(\lambda, \lambda') = n^{-2} \sum_{t=-N}^N \sum_{s=-N}^N \left(\frac{\lambda_t - \lambda}{h_n} \right)^i \left(\frac{\lambda_s - \lambda'}{h_n} \right)^j K_{h_n}(\lambda_t - \lambda) K_{h_n}(\lambda_s - \lambda') \gamma_n^*(\lambda_t, \lambda_s),$$

donde ahora $\gamma_n^*(\lambda_t, \lambda_s) = E\{[Y_t - E(Y_t)][Y_s - E(Y_s)]\}$.

De acuerdo con el modelo (2.60):

$$\gamma_n^*(\lambda_t, \lambda_s) = Cov(\varepsilon_t \varepsilon_s) 1_{\{t=s\}} = \frac{\pi^2}{6} 1_{\{t=s\}},$$

de modo que

$$v_{i,j}^{*n}(\lambda, \lambda') = n^{-2} \sum_{t=-N}^N \left(\frac{\lambda_t - \lambda}{h_n} \right)^i \left(\frac{\lambda_t - \lambda'}{h_n} \right)^j K_{h_n}(\lambda_t - \lambda) K_{h_n}(\lambda_t - \lambda') \frac{\pi^2}{6},$$

para $i, j \in \{0, 1\}$.

Denotando por

$$I_{ij}^{*n} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') f_1(\lambda) f_1(\lambda') v_{i,j}^{*n}(\lambda, \lambda') d\lambda d\lambda', \text{ para } i, j \in \{0, 1\},$$

se puede proceder de modo análogo a como se hizo en (2.42) y (2.43), concluyéndose que

$$\begin{aligned} 2\pi n I_{i,j}^{*n} &= \frac{\pi^2}{6} \int_{-\pi}^{\pi} \left(\int_{h_n^{-1}(s-\pi)}^{h_n^{-1}(s+\pi)} Q_{1,2}(s - h_n u) f_1(s - h_n u) u^i K(u) du \right) \\ &\times \left(\int_{h_n^{-1}(s-\pi)}^{h_n^{-1}(s+\pi)} Q_{1,2}(s - h_n v) f_1(s - h_n v) v^j K(v) dv \right) ds. \end{aligned} \quad (2.65)$$

Utilizando razonamientos análogos al caso de trabajar con el estimador periodograma suavizado, \hat{f}_{DLS} , aplicando el teorema de la convergencia dominada en (2.64) se obtiene

$$8\pi n Var(\Sigma_n^*) = \frac{\pi^2}{6} \int_{-\pi}^{\pi} f_1^2(u) Q_{1,2}^2(u) du + O(h_n^2)$$

estableciéndose así (2.18).

Paso 3. Se probará a continuación la normalidad asintótica del estadístico discriminante D_W calculado a partir del estimador log-periodograma suavizado, \hat{f}_{LS} .

De acuerdo con la notación empleada, bajo Π_1 , se tiene

$$\sqrt{n} \left\{ D_W - \mu_{W, \hat{f}_{LS}}(1, 2) \right\} =$$

$$\begin{aligned}
 &= \frac{\sqrt{n}}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) e_1^t \mathbf{S}_{n,\lambda}^{-1} \mathbf{T}_{n,\lambda}^* d\lambda (1 + o(1)). \\
 &= \frac{\sqrt{n}}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) e_1^t \mathbf{H}_n^{-1} \mathbf{H}_n \mathbf{S}_{n,\lambda}^{-1} \mathbf{H}_n \mathbf{H}_n^{-1} \mathbf{T}_{n,\lambda}^* d\lambda (1 + o(1)). \tag{2.66}
 \end{aligned}$$

Así, la normalidad asintótica de $\sqrt{n} \{D_W - \mu_{W, \hat{f}_{LS}}(1, 2)\}$ se deduce de (2.32) y (2.66) una vez que se haya establecido la normalidad asintótica de la variable aleatoria

$$\theta_n^* = \sqrt{n} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) (c_0 \tilde{t}_0^{*n}(\lambda) + c_1 \tilde{t}_1^{*n}(\lambda)) d\lambda, \tag{2.67}$$

donde c_0 y c_1 denotan dos constantes arbitrarias y, para $l = 0, 1$, $\tilde{t}_l^{*n}(\lambda)$ se define de modo análogo a $\tilde{t}_l^n(\lambda)$ en (2.47):

$$\tilde{t}_l^{*n}(\lambda) = \frac{1}{n} \sum_{t=-N}^N \left(\frac{\lambda_t - \lambda}{h_n} \right)^l K_{h_n}(\lambda_t - \lambda) (Y_t - E(Y_t)).$$

De nuevo θ_n^* puede escribirse como una suma de variables aleatorias de la forma

$$\theta_n^* = \frac{1}{\sqrt{n}} \sum_{t=-N}^N \eta_t^* \tag{2.68}$$

donde ahora

$$\eta_t^* = (Y_t - E(Y_t)) \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) \xi_{h_n}(\lambda_t - \lambda) d\lambda,$$

con $\xi_{h_n}(u) = \xi(u/h_n)/h_n$ y $\xi(u) = c_0 K(u) + c_1 u K(u)$.

Utilizando argumentos similares a los empleados para (2.48) puede demostrarse que las variables η_t^* son asintóticamente independientes con varianza

$$Var(\eta_t^*) = \frac{\pi^2}{6} \left(Q_{1,2}(\lambda_t) f_1(\lambda_t) \int_{-\pi}^{\pi} \xi(u) du \right)^2 (1 + o(1)).$$

De este modo

$$\begin{aligned}
 Var(\theta_n^*) &= \frac{\pi^2}{6} \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} Q_{1,2}^2(u) f_1^2(u) du \right) \left(\int_{-\pi}^{\pi} \xi(u) du \right)^2 (1 + o(1)) = \\
 &= \sigma_{\theta^*}^2 (1 + o(1)). \tag{2.69}
 \end{aligned}$$

Una vez más, la normalidad asintótica de θ_n^* se establece tras comprobar que las variables η_t^* satisfacen las condiciones de Lindeberg-Feller para normalidad asintótica.

La primera condición,

$$\frac{1}{n} \sum_{t=-N}^N E(\eta_t^{*2}) \longrightarrow \sigma_{\theta^*}^2$$

se sigue inmediatamente de (2.68) y (2.69).

La segunda condición,

$$\frac{1}{n} \sum_{t=-N}^N E \left(\eta_t^{*2} 1_{\{|\eta_t^*| \geq \varepsilon \sigma_{\theta^*} n^{1/2}\}} \right) \longrightarrow 0 \quad \text{para todo } \varepsilon > 0$$

puede probarse utilizando un argumento de truncamiento completamente análogo al empleado para (2.51). Así, dado M un punto de truncamiento fijo y denotando

$$Y_t^M = Y_t 1_{\{|Y_t| \leq M\}},$$

se prueba que

$$\theta_n^{*M} \rightarrow N(0, \sigma_{\theta^*, M}^2). \tag{2.70}$$

Entonces se llega a que

$$\begin{aligned} & \left| E \exp(it\theta_n^*) - \exp(-t^2\sigma_{\theta^*}^2/2) \right| \leq \\ & \leq \left| E \exp(it\theta_n^{*M}) \left(\exp(it\bar{\theta}_n^{*M}) - 1 \right) \right| \\ & + \left| E \exp(it\theta_n^{*M}) - \exp(-t^2\sigma_{\theta^*, M}^2/2) \right| + \\ & \left| \exp(-t^2\sigma_{\theta^*, M}^2/2) - \exp(-t^2\sigma_{\theta^*}^2/2) \right| \end{aligned} \tag{2.71}$$

La convergencia a cero de cada uno de los tres términos en (2.71) puede probarse siguiendo argumentos análogos a como se hizo para (2.54), luego queda probada la normalidad asintótica del estadístico discriminante según (2.16).

b.3) Demostración en el caso $\widehat{f}_n = \widehat{f}_{LK}$.

Procediendo igual que en el caso anterior, bajo Π_1 se tiene:

$$\begin{aligned} D_W &= D_W(f_1, f_2) + \Sigma_n^{**} + O_P(h_n^4 + (nh_n)^{-1}) = \\ &= D_W(f_1, f_2) + \Sigma_n^{**} + O_P((nh_n)^{-1}) \end{aligned}$$

donde ahora

$$\Sigma_n^{**} = \frac{1}{4\pi} \left(\int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) (\widehat{m}_{LK}(\lambda) - m_1(\lambda)) d\lambda \right). \tag{2.72}$$

Según lo visto en la Sección 1.3 de la presente memoria, $\widehat{\beta}_\lambda = \left(\widehat{m}_{LK}(\lambda), \widehat{m}'_{LK}(\lambda) \right)^t$ maximiza la log-verosimilitud local asociada a la secuencia $\{Y_k = \log(I_n(\lambda_k))\}$, que coincide como se vió en (1.37) con la log-verosimilitud local de la secuencia $\{\exp(Y'_k)\}$ (donde $Y'_k = m_1(\lambda_k) + \varepsilon_k$), variables aleatorias independientes con distribución exponencial de media $f(\lambda_k)$. Este es un caso especial de los estudiados por Fan, Heckman y Wand (1995).

Denotando por

$$\widehat{\boldsymbol{\beta}}^*_{\lambda} = a_n^{-1} \left(\widehat{m}_{LK}(\lambda) - m_1(\lambda), h_n \left(\widehat{m}'_{LK}(\lambda) - m'_1(\lambda) \right) \right)^t \quad (2.73)$$

con $a_n = (nh_n)^{-1/2}$, se tiene que $\widehat{\boldsymbol{\beta}}^*_{\lambda}$ es de la forma (Fan, Heckman y Wand (1995)):

$$\widehat{\boldsymbol{\beta}}^*_{\lambda} = -\mathbf{A}^{-1} \mathbf{W}_{n,\lambda} + o_P(1) \quad (2.74)$$

uniformemente en $\lambda \in [-\pi, \pi]$ (Nielsen (2005)), siendo $\mathbf{A} = -\frac{1}{2\pi} \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix}$ y

$$\mathbf{W}_{n,\lambda} = a_n \sum_{k=-N}^N q_1(\overline{m}_1(\lambda, \lambda_k), e^{Y'_k}) \mathbf{Z}_{k,\lambda} K \left(\frac{\lambda_k - \lambda}{h_n} \right) \quad (2.75)$$

con

$$\begin{aligned} \mathbf{Z}_{k,\lambda} &= \left(1, \frac{\lambda_k - \lambda}{h_n} \right)^t, k = -N, \dots, N, \\ \overline{m}_1(\lambda, \lambda_k) &= m_1(\lambda) + m'_1(\lambda)(\lambda_k - \lambda), \end{aligned} \quad (2.76)$$

y

$$q_1(x, y) = \frac{\partial}{\partial x} Q(e^x, y) = ye^{-x} - 1.$$

En particular,

$$q_1(m_1(\lambda_k), e^{Y'_k}) = \frac{e^{Y'_k} - f_1(\lambda_k)}{f_1(\lambda_k)}, \quad (2.77)$$

y, denotando por $q_2(x, y) = \left(\frac{\partial^2}{\partial x^2} \right) Q(e^x, y) = -ye^{-x}$:

$$q_2(m_1(\lambda_k), e^{Y'_k}) = -\frac{e^{Y'_k}}{f_1(\lambda_k)}. \quad (2.78)$$

Además, se verifica

$$\mathbf{W}_{n,\lambda} - a_n^{-1} \left\{ \frac{1}{2\pi} \frac{m''_1(\lambda)}{2} h_n^2 (\mu_2(K), 0)^t + o_P(h_n^3) \right\} \rightarrow N(0, \mathbf{B}_{\lambda}) \quad (2.79)$$

con

$$\mathbf{B}_{\lambda} = \frac{1}{2\pi} f_1^2(\lambda) \begin{pmatrix} \mu_0(K^2) & 0 \\ 0 & \mu_2(K^2) \end{pmatrix}.$$

Sustituyendo (2.73) y (2.74) en (2.72) se llega a que

$$\begin{aligned} \Sigma_n^{**} &= \frac{1}{4\pi} \left(\int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) a_n e_1^t \widehat{\boldsymbol{\beta}}^*_{\lambda} d\lambda \right) = \\ &= \frac{-1}{4\pi} a_n \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) e_1^t \mathbf{A}^{-1} \mathbf{W}_{n,\lambda} d\lambda + o_P(nh_n)^{-1/2}. \end{aligned}$$

De modo análogo a como ocurría con Σ_n en (2.25) y Σ_n^* en (2.59), Σ_n^{**} puede expresarse como suma de dos componentes que determinan su sesgo y varianza, respectivamente:

$$\begin{aligned} \Sigma_n^{**} &= \frac{-1}{4\pi} a_n \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) e_1^t \mathbf{A}^{-1} (\mathbf{W}_{n,\lambda} - E(\mathbf{W}_{n,\lambda})) d\lambda + \\ &+ \left(\frac{-1}{4\pi} \right) a_n \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) e_1^t \mathbf{A}^{-1} E(\mathbf{W}_{n,\lambda}) d\lambda + o_P(nh_n)^{-1/2} = \\ &= V^{**} + B^{**} + o_P(nh_n)^{-1/2} \end{aligned} \quad (2.80)$$

Paso 1. La expresión asintótica del sesgo para el estadístico discriminante D_W se obtiene a partir de las expresiones (2.79) y (2.80):

$$\begin{aligned} E(\Sigma_n^{**}) &= B^{**} = \frac{-1}{4\pi} a_n \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) e_1^t \mathbf{A}^{-1} E(\mathbf{W}_{n,\lambda}) d\lambda = \\ &= \frac{a_n^2}{2} \sum_{k=-N}^N \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) E \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) K \left(\frac{\lambda_k - \lambda}{h_n} \right) \right\} d\lambda. \end{aligned} \quad (2.81)$$

Dado que $E \left\{ e^{Y'_k} \right\} = f_1(\lambda_k)$, para $k = -N, \dots, N$, se tiene que:

$$E \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) K \left(\frac{\lambda_k - \lambda}{h_n} \right) \right\} = \int_{-\pi}^{\pi} q_1(\bar{m}_1(\lambda, u), f_1(u)) K \left(\frac{u - \lambda}{h_n} \right) \frac{1}{2\pi} du.$$

Realizando el cambio de variable $u - \lambda = h_n v$, en la expresión anterior resulta:

$$\begin{aligned} E \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) K \left(\frac{\lambda_k - \lambda}{h_n} \right) \right\} &= \\ &= \int_{-\pi}^{\pi} q_1(\bar{m}_1(\lambda, \lambda + h_n v), f_1(\lambda + h_n v)) K(v) \frac{1}{2\pi} h_n dv. \end{aligned} \quad (2.82)$$

Se tendrá ahora en cuenta que, a partir de las expresiones obtenidas para $q_1(\cdot, \cdot)$ y $q_2(\cdot, \cdot)$:

$$q_1(m_1(\lambda), f_1(\lambda)) = f_1(\lambda) e^{-m_1(\lambda)} - 1 = 0 \quad (2.83)$$

y

$$q_2(m_1(\lambda), f_1(\lambda)) = -f_1(\lambda) e^{-m_1(\lambda)} = -1. \quad (2.84)$$

Obsérvese que el primer término del integrando en (2.82) verifica:

$$q_1(\bar{m}_1(\lambda, \lambda + h_n v), f_1(\lambda + h_n v)) = q_1(m_1(\lambda + h_n v) - m_1''(\lambda) \frac{h_n^2 v^2}{2} + o(h_n^2), f_1(\lambda + h_n v)),$$

de modo que un desarrollo de Taylor conduce a

$$q_1(\bar{m}_1(\lambda, \lambda + h_n v), f_1(\lambda + h_n v)) = q_1(m_1(\lambda + h_n v), f_1(\lambda + h_n v)) -$$

$$\begin{aligned} -q_2(m_1(\lambda + h_nv), f_1(\lambda + h_nv))m_1''(\lambda)\frac{h_n^2v^2}{2} + o(h_n^2) = \\ = m_1''(\lambda)\frac{h_n^2v^2}{2} + o(h_n^2), \end{aligned}$$

en virtud de lo establecido en (2.83) y (2.84).

En consecuencia, teniendo en cuenta que las variables $\{e^{Y'_k}\}$ son independientes e idénticamente distribuidas, volviendo a (2.81):

$$\begin{aligned} E(\Sigma_n^{**}) &= \frac{a_n^2}{2}n \int_{-\pi}^{\pi} Q_{1,2}(\lambda)f_1(\lambda)\frac{h_n}{2\pi}m_1''(\lambda)\frac{h_n^2}{2}\mu_2(K)d\lambda (1 + o(1)) = \\ &= \frac{1}{8\pi}\mu_2(K)h_n^2 \int_{-\pi}^{\pi} Q_{1,2}(\lambda)f_1(\lambda)m_1''(\lambda)d\lambda + o(h_n^2), \end{aligned}$$

y (2.17) queda así demostrado.

Paso 2. A continuación procederemos a calcular la componente relativa a la varianza del estadístico discriminante, $Var(\Sigma_n^{**})$. De acuerdo con (2.80):

$$\begin{aligned} Var(\Sigma_n^{**}) &= E(V^{**2}) = \\ &= \frac{a_n^2}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda)Q_{1,2}(\lambda')f_1(\lambda)f_1(\lambda')\gamma_n(\lambda, \lambda') d\lambda d\lambda' \end{aligned} \quad (2.85)$$

denotándose en este caso

$$\gamma_n(\lambda, \lambda') = \mathbf{e}_1^t \mathbf{A}^{-1} \Gamma_n^{**}(\lambda, \lambda') \mathbf{A}^{-1} \mathbf{e}_1,$$

con

$$\Gamma_n^{**}(\lambda, \lambda') = E \left\{ (\mathbf{W}_{n,\lambda} - E(\mathbf{W}_{n,\lambda})) (\mathbf{W}_{n,\lambda'} - E(\mathbf{W}_{n,\lambda'}))^t \right\}. \quad (2.86)$$

De acuerdo con la definición de $\mathbf{W}_{n,\lambda}$ en (2.75), se tiene que:

$$\begin{aligned} \gamma_n(\lambda, \lambda') &= a_n^2 \sum_{k=-N}^N \sum_{r=-N}^N 4\pi^2 Cov \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) K \left(\frac{\lambda_k - \lambda}{h_n} \right), \right. \\ &\quad \left. q_1(\bar{m}_1(\lambda', \lambda_r), e^{Y'_r}) K \left(\frac{\lambda_r - \lambda'}{h_n} \right) \right\}. \end{aligned}$$

Puesto que las variables $\{e^{Y'_k}, k = -N, \dots, N\}$ son independientes e idénticamente distribuidas y, además, $q_1(x, y)$ es una función lineal en y , entonces se tiene que

$$Cov \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) K \left(\frac{\lambda_k - \lambda}{h_n} \right), q_1(\bar{m}_1(\lambda', \lambda_r), e^{Y'_r}) K \left(\frac{\lambda_r - \lambda'}{h_n} \right) \right\} = 0$$

si $k \neq r$.

Por lo tanto:

$$\begin{aligned} \gamma_n(\lambda, \lambda') &= \\ &= 4\pi^2 a_n^2 \sum_{k=-N}^N Cov \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) K\left(\frac{\lambda_k - \lambda}{h_n}\right), q_1(\bar{m}_1(\lambda', \lambda_k), e^{Y'_k}) K\left(\frac{\lambda_k - \lambda'}{h_n}\right) \right\} = \\ &= 4\pi^2 a_n^2 \sum_{k=-N}^N E \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) q_1(\bar{m}_1(\lambda', \lambda_k), e^{Y'_k}) K\left(\frac{\lambda_k - \lambda}{h_n}\right) K\left(\frac{\lambda_k - \lambda'}{h_n}\right) \right\} - \\ &- 4\pi^2 a_n^2 \sum_{k=-N}^N E \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) K\left(\frac{\lambda_k - \lambda}{h_n}\right) \right\} E \left\{ q_1(\bar{m}_1(\lambda', \lambda_k), e^{Y'_k}) K\left(\frac{\lambda_k - \lambda'}{h_n}\right) \right\} \end{aligned}$$

De acuerdo con lo visto en el Paso 1, el segundo término en la expresión anterior es $O(h_n^4)$, luego

$$\begin{aligned} \gamma_n(\lambda, \lambda') &= 4\pi^2 a_n^2 \sum_{k=-N}^N E \left\{ q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) \times q_1(\bar{m}_1(\lambda', \lambda_k), e^{Y'_k}) \times \right. \\ &\quad \left. \times K\left(\frac{\lambda_k - \lambda}{h_n}\right) K\left(\frac{\lambda_k - \lambda'}{h_n}\right) \right\} + O(h_n^4) \end{aligned} \quad (2.87)$$

Por otro lado, teniendo en cuenta la definición de $\bar{m}_1(\cdot, \cdot)$ en (2.76) puede escribirse, para λ, λ' en un entorno de λ_k :

$$\begin{aligned} q_1(\bar{m}_1(\lambda, \lambda_k), e^{Y'_k}) &= q_1(m_1(\lambda_k), e^{Y'_k}) + q_2(m_1(\lambda_k), e^{Y'_k}) (\bar{m}_1(\lambda, \lambda_k) - m_1(\lambda_k)) + \\ &+ o\left((\bar{m}_1(\lambda, \lambda_k) - m_1(\lambda_k))^2\right) = q_1(m_1(\lambda_k), e^{Y'_k}) + o(h_n^2) \end{aligned} \quad (2.88)$$

De esta forma, sustituyendo (2.88) en (2.87) se tiene

$$\gamma_n(\lambda, \lambda') = 4\pi^2 a_n^2 \sum_{k=-N}^N E \left\{ q_1^2(m_1(\lambda_k), e^{Y'_k}) K\left(\frac{\lambda_k - \lambda}{h_n}\right) K\left(\frac{\lambda_k - \lambda'}{h_n}\right) \right\} (1 + o(1)). \quad (2.89)$$

Si tenemos en cuenta que, según (2.77)

$$q_1(m_1(\lambda_k), e^{Y'_k}) = \frac{e^{Y'_k} - f_1(\lambda_k)}{f_1(\lambda_k)},$$

entonces

$$E \left\{ q_1^2(m_1(\lambda_k), e^{Y'_k}) \right\} = E \left\{ \left(\frac{e^{Y'_k} - f_1(\lambda_k)}{f_1(\lambda_k)} \right)^2 \right\} = 1$$

De este modo, (2.89) resulta

$$\gamma_n(\lambda, \lambda') = 4\pi^2 a_n^2 n \int_{-\pi}^{\pi} K\left(\frac{u - \lambda}{h_n}\right) K\left(\frac{u - \lambda'}{h_n}\right) \frac{1}{2\pi} du \quad (2.90)$$

Sustituyendo (2.90) en (2.85) se tiene

$$\begin{aligned} \text{Var}(\Sigma_n^{**}) &= \frac{a_n^2}{16\pi^2} 4\pi^2 a_n^2 n \times \left(\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) Q_{1,2}(\lambda') f_1(\lambda) f_1(\lambda') \times \right. \\ &\quad \left. \times \left(\int_{-\pi}^{\pi} K\left(\frac{u-\lambda}{h_n}\right) K\left(\frac{u-\lambda'}{h_n}\right) \frac{1}{2\pi} du \right) d\lambda d\lambda' \right) = \\ &= \frac{a_n^4}{8\pi} n \int_{-\pi}^{\pi} \left(\int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) K\left(\frac{u-\lambda}{h_n}\right) d\lambda \right) \left(Q_{1,2}(\lambda') f_1(\lambda') K\left(\frac{u-\lambda'}{h_n}\right) d\lambda' \right) du. \end{aligned}$$

Realizando los cambios de variable $u - \lambda = h_n v$ y $u - \lambda' = h_n s$ se llega a que

$$\begin{aligned} \text{Var}(\Sigma_n^{**}) &= \frac{a_n^4}{8\pi} n \int_{-\pi}^{\pi} \left(\int_{-\pi}^{\pi} Q_{1,2}(u - h_n v) f_1(u - h_n v) K(v) h_n dv \right) \times \\ &\quad \times \left(\int_{-\pi}^{\pi} Q_{1,2}(u - h_n s) f_1(u - h_n s) K(s) h_n ds \right) du. \end{aligned} \quad (2.91)$$

Recurriendo a un desarrollo de Taylor se tiene, para cada una de las integrales

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} Q_{1,2}(u - h_n v) f_1(u - h_n v) K(v) dv = Q_{1,2}(u) f_1(u).$$

De forma que (2.91) resulta

$$\text{Var}(\Sigma_n^{**}) = \frac{1}{8\pi n} \int_{-\pi}^{\pi} Q_{1,2}^2(u) f_1^2(u) du,$$

quedando así establecido (2.18).

Paso 3. A continuación, pasará a probarse la normalidad asintótica del estadístico discriminante D_W . De (2.72) y (2.80) se deduce que

$$\begin{aligned} &\sqrt{n} \left\{ D_W - \mu_{W, \hat{f}_{LK}}(1, 2) \right\} = \\ &= \frac{-\sqrt{n} a_n}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) e_1^t \mathbf{A}^{-1} (\mathbf{W}_{n,\lambda} - E(\mathbf{W}_{n,\lambda})) d\lambda (1 + o_P(1)). \end{aligned} \quad (2.92)$$

Teniendo en cuenta las definiciones de \mathbf{A} y \mathbf{W}_n , para probar la normalidad asintótica de (2.92), basta con establecer la normalidad asintótica de la variable aleatoria

$$\theta_n^{**} = \frac{\sqrt{n} a_n^2}{2} \sum_{k=-N}^N \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1(\lambda) q_1 \left(\bar{m}_1(\lambda, \lambda_k), e^{Y_k'} \right) K\left(\frac{\lambda_k - \lambda}{h_n}\right) d\lambda,$$

suma de variables independientes e idénticamente distribuidas. Así, la normalidad asintótica de (2.92) se deduce sin más que aplicar razonamientos análogos a los empleados en las demostraciones anteriores.

c.q.d.

Al igual que con los resultados de Kakizawa(1996) y Zhang y Taniguchi (1994), el Teorema 2.3.1 establece que el estadístico discriminante D_W , basado en cualquiera de los tres estimadores lineales locales del espectro propuestos en Fan y Kreutzberger (1998), sigue asintóticamente una distribución normal. El centro de esta distribución es la verdadera disparidad existente entre las dos densidades espectrales f_1 y f_2 , $D_W(f_1, f_2)$, más un sesgo asintótico de orden h_n^2 que depende del sesgo del estimador espectral utilizado y de la función de divergencia W mediante el término $\int_{-\pi}^{\pi} Q_{j,k}(\lambda)\varphi_j(\lambda)d\lambda$.

Con respecto a la varianza asintótica del estadístico D_W , se puede llegar a consideraciones análogas. Así, la varianza asintótica de D_W cuando el estadístico discriminante se basa en el log-periodograma suavizado mediante mínimos cuadrados, \hat{f}_{LS} , resulta ser $\pi^2/6$ veces mayor que su varianza asintótica cuando en la construcción de D_W se emplean \hat{f}_{DLS} o \hat{f}_{LK} . Esto ya se observaba en el comportamiento asintótico de los tres estimadores espectrales y, de hecho, resultaba predecible dado que el método de máxima verosimilitud es más eficiente que el método de mínimos cuadrados ya que, según se recordará, la distribución de los términos ε_k en (1.28) no resulta normal. Por otro lado, la tasa de convergencia de orden $n^{-1/2}$ logra alcanzarse ya que el estadístico discriminante D_W se basa en integrar cada uno de los estimadores no paramétricos.

Una vez establecida la distribución asintótica del estadístico discriminante D_W según el Teorema 2.3.1, las probabilidades de mala clasificación asociadas a D_W pueden obtenerse directamente a partir de este resultado. Así, bajo las condiciones del Teorema 2.3.1, se concluye que

$$P_{D_W}(2|1) = P[D_W \leq 0|\Pi_1] = \Phi\left(-\sqrt{n}\frac{\mu_{W,\hat{f}_n}(1,2)}{V_{W,\hat{f}_n}(1,2)}\right) + o(1) \tag{2.93}$$

y

$$P_{D_W}(1|2) = P[D_W > 0|\Pi_2] = \Phi\left(-\sqrt{n}\frac{\mu_{W,\hat{f}_n}(2,1)}{V_{W,\hat{f}_n}(2,1)}\right) + o(1) \tag{2.94}$$

Cabe señalar que D_W tiende en probabilidad a $(-1)^{j+1}D_W(f_j, f_k)$ y, por otro lado, $D_W(f_j, f_k) > 0$, para $(j, k) = (1, 2), (2, 1)$, dado que $f_1(\lambda) \neq f_2(\lambda)$ en un conjunto de medida de Lebesgue positiva. Por lo tanto, las expresiones (2.93) y (2.94) permiten concluir que las probabilidades de mala clasificación asociadas al estadístico discriminante D_W tienden a cero, tal y como se establece en el siguiente teorema.

Teorema 2.3.2 *Bajo las Hipótesis (H3.1),..., (H3.5), si $f_1(\lambda) \neq f_2(\lambda)$ en un conjunto de medida de Lebesgue positiva, entonces,*

$$\lim_{n \rightarrow \infty} P_{D_W}(2|1) = \lim_{n \rightarrow \infty} P_{D_W}(1|2) = 0.$$

El Teorema 2.3.2 establece así que el estadístico discriminante D_W proporciona un criterio de clasificación consistente.

2.4. Análisis discriminante a partir de muestras de entrenamiento

Hasta el momento se ha trabajado bajo la hipótesis de que las densidades espectrales $f_1(\lambda)$ y $f_2(\lambda)$ asociadas a las poblaciones entre las que se discrimina, Π_1 y Π_2 , son conocidas. Sin embargo, en la práctica, pocas veces se da este supuesto y dichas densidades deben ser estimadas a partir de muestras de entrenamiento disponibles en cada una de las dos poblaciones entre las que se intenta discriminar.

Supongamos que se dispone de r_i realizaciones de longitud n_i

$$\mathbf{X}_{ij} = (X_{ij}(1), X_{ij}(2), \dots, X_{ij}(n_i))^t,$$

con $j = 1, \dots, r_i$, $i = 1, 2$ para estimar $f_i(\lambda)$, $i = 1, 2$. Entonces, la densidad espectral $f_i(\lambda)$ puede estimarse promediando los estimadores espectrales obtenidos a partir de cada una de las series disponibles de la i -ésima población:

$$\hat{f}_i(\lambda) = \frac{1}{r_i} \sum_{j=1}^{r_i} \hat{f}_{ij}(\lambda), \quad \lambda \in [-\pi, \pi],$$

utilizando para ello cualquiera de los tres estimadores lineales locales propuestos por Fan y Kreutzberger(1998).

Para clasificar \mathbf{X}_n se utilizará entonces el estadístico plug-in

$$\begin{aligned} \hat{D}_W &= D_W(\hat{f}_n, \hat{f}_2) - D_W(\hat{f}_n, \hat{f}_1) = \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(W \left(\frac{\hat{f}_n(\lambda)}{\hat{f}_2(\lambda)} \right) - W \left(\frac{\hat{f}_n(\lambda)}{\hat{f}_1(\lambda)} \right) \right) d\lambda, \end{aligned} \quad (2.95)$$

y se asignará X_n a Π_1 si $\hat{D}_W > 0$ y a Π_2 en caso contrario.

A continuación se estudiará el comportamiento asintótico del criterio discriminante resultante.

2.4.1. Propiedades asintóticas

Por simplicidad, se supondrá que la longitud de las series de entrenamiento en cada grupo $n_i = c_i n$ para algún $c_i > 0$, y que las series X_{1j} , ($j = 1, \dots, r_1$), X_{2j} , ($j = 1, \dots, r_2$)

y X_n son independientes entre sí. Así mismo, la densidad espectral de todas las series del mismo grupo se estimará utilizando el mismo parámetro ventana h_i . En orden a establecer las propiedades asintóticas del criterio propuesto se establecerán además las siguientes hipótesis:

Hipótesis (H3.1') Las series $\mathbf{X}_{ij} = (X_{ij}(1), X_{ij}(2), \dots, X_{ij}(n_i))^t$, $j = 1, \dots, r_i$ son realizaciones independientes de un proceso lineal gaussiano (para $i=1,2$).

Hipótesis (H3.5') La sucesión de parámetros ventana h_i , $i = 1, 2$ satisfacen

$$(\log n)^4 h_i \rightarrow 0 \text{ y } nh_i^2 \rightarrow \infty \text{ cuando } n \rightarrow \infty.$$

Manteniendo esta notación y la empleada en la sección anterior se demuestra el siguiente resultado:

Teorema 2.4.1 Si se verifican las mismas hipótesis del Teorema 2.3.1, y las condiciones (H3.1'), (H3.5'), entonces bajo Π_j se tiene que:

$$\widehat{D}_W \xrightarrow{p} (-1)^{j+1} D_W(f_j, f_k) \tag{2.96}$$

y

$$\sqrt{n} \left\{ \widehat{D}_W + (-1)^j \widehat{\mu}_{W, \widehat{f}}(j, k) \right\} \xrightarrow{d} N \left(0, \widehat{V}_{W, \widehat{f}}^2(j, k) \right), \tag{2.97}$$

donde

$$\begin{aligned} \widehat{\mu}_{W, \widehat{f}}(j, k) &= D_W(f_j, f_k) + \frac{h_n^2}{8\pi} \mu_2(K) \int_{-\pi}^{\pi} Q_{j,k}(\lambda) \varphi_j(\lambda) d\lambda - \\ &- \frac{h_k^2}{8\pi} \mu_2(K) \int_{-\pi}^{\pi} Q_{j,k}(\lambda) \frac{f_1(\lambda)}{f_2(\lambda)} \varphi_k(\lambda) d\lambda + o(h_n^2) + o(h_k^2), \end{aligned} \tag{2.98}$$

con

$$\varphi_l(\lambda) = \begin{cases} f_l''(\lambda) & \text{si } \widehat{f} = \widehat{f}_{DLS} \\ m_l''(\lambda) f_l(\lambda) & \text{si } \widehat{f} = \widehat{f}_{LS} \text{ ó } \widehat{f} = \widehat{f}_{LK} \end{cases}$$

y

$$\begin{aligned} \widehat{V}_{W, \widehat{f}}^2(j, k) &= C \int_{-\pi}^{\pi} [Q_{j,k}(\lambda) f_j(\lambda)]^2 d\lambda \\ &+ C \frac{1}{c_k r_k} \int_{-\pi}^{\pi} [Q_{j,k}(\lambda) f_j(\lambda)]^2 d\lambda + O(h_n^2) + O(h_k^2), \end{aligned} \tag{2.99}$$

con $C = (8\pi)^{-1}$ si $\widehat{f} = \widehat{f}_{DLS}$ o $\widehat{f} = \widehat{f}_{LK}$ y $C = (8\pi)^{-1} \pi^2/6$ si $\widehat{f} = \widehat{f}_{LS}$.

Demostración del Teorema 2.4.1

El Teorema 2.4.1 se establece recurriendo a las propiedades asintóticas de cada uno de los tres estimadores espectrales considerados, a las condiciones de regularidad impuestas

sobre W y f , y a los resultados alcanzados en la demostración del Teorema 2.3.1. Se realizará la demostración del Teorema sólo en el caso de que se suponga cierta Π_1 , siendo la demostración análoga bajo Π_2 .

a) Convergencia en probabilidad.

Puesto que para cada uno de los tres estimadores espectrales se demuestra su consistencia uniforme:

$$\widehat{f}_{ij}(\lambda) - f_i(\lambda) = O_P(nh_i)^{-1/2}, \quad i = 1, 2, \quad j = 1, \dots, r_i, \quad (2.100)$$

dicha propiedad se traslada a su promedio de modo inmediato:

$$\widehat{f}_i(\lambda) - f_i(\lambda) = O_P(nh_i)^{-1/2}, \quad i = 1, 2. \quad (2.101)$$

Teniendo en cuenta las condiciones de regularidad impuestas a $W(\cdot)$ y la consistencia uniforme de cada uno de los tres estimadores se obtiene que:

$$\begin{aligned} W\left(\frac{\widehat{f}_n(\lambda)}{\widehat{f}_2(\lambda)}\right) &= W\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) + W'\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) \frac{\widehat{f}_n(\lambda) - f_1(\lambda)}{f_2(\lambda)} - \\ &- W'\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) \frac{f_1(\lambda)}{f_2(\lambda)} \frac{\widehat{f}_2(\lambda) - f_2(\lambda)}{f_2(\lambda)} + O_P(nh_n)^{-1} + O_P(nh_2)^{-1}, \end{aligned} \quad (2.102)$$

y

$$W\left(\frac{\widehat{f}_n(\lambda)}{\widehat{f}_1(\lambda)}\right) = O_P(nh_n)^{-1} + O_P(nh_1)^{-1}. \quad (2.103)$$

Sustituyendo (2.102) y (2.103) en la expresión del estadístico plug-in (2.95), se tiene que

$$\begin{aligned} \widehat{D}_W &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[W\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) + W'\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) \frac{\widehat{f}_n(\lambda) - f_1(\lambda)}{f_2(\lambda)} - \right. \\ &- \left. W'\left(\frac{f_1(\lambda)}{f_2(\lambda)}\right) \frac{f_1(\lambda)}{f_2(\lambda)} \frac{\widehat{f}_2(\lambda) - f_2(\lambda)}{f_2(\lambda)} + O_P(nh_n)^{-1} + O_P(nh_1)^{-1} + O_P(nh_2)^{-1} \right] d\lambda = \\ &= D_W(f_1, f_2) + \Sigma_n - \widetilde{\Sigma}_n + O_P(nh_n)^{-1} + O_P(nh_1)^{-1} + O_P(nh_2)^{-1} \end{aligned} \quad (2.104)$$

donde Σ_n se define como en (2.25)

$$\Sigma_n = \frac{1}{4\pi} \int_{-\pi}^{\pi} Q_{1,2}(\lambda) \left(\widehat{f}_n(\lambda) - f_1(\lambda) \right) d\lambda$$

y $\widetilde{\Sigma}_n$ viene dada por

$$\widetilde{\Sigma}_n = \frac{1}{4\pi} \int_{-\pi}^{\pi} \widetilde{Q}_{1,2}(\lambda) \left(\widehat{f}_2(\lambda) - f_2(\lambda) \right) d\lambda,$$

donde se ha denotado

$$\tilde{Q}_{1,2} = \frac{f_1(\lambda)}{f_2(\lambda)} Q_{1,2}(\lambda).$$

Las condiciones (H3.2) y (H3.3) garantizan que tanto $Q_{1,2}(\lambda)$ como $\tilde{Q}_{1,2}(\lambda)$ están acotadas en $[-\pi, \pi]$. Este hecho, junto con la consistencia uniforme de \hat{f}_n a f_1 y de \hat{f}_2 a f_2 permiten utilizar el teorema de la convergencia dominada para establecer la convergencia en probabilidad a cero de Σ_n y $\tilde{\Sigma}_n$. Así, queda demostrada la convergencia en probabilidad de \hat{D}_W dada en (2.96).

b) Distribución asintótica.

Se obtendrán a continuación las expresiones asintóticas para el sesgo y la varianza del estadístico plug-in \hat{D}_W . La demostración se hará para el caso en el que los tres estimadores espectrales se obtengan suavizando directamente el periodograma mediante técnicas de suavización local. La demostración resulta análoga para el caso en el que se empleen cualquiera de los otros dos estimadores del espectro (log-periodograma suavizado o estimador de máxima verosimilitud), reemplazando convenientemente los términos obtenidos para el sesgo y la varianza del estadístico por aquellos obtenidos en la demostración del Teorema 2.3.1.

Paso 1. Para el caso de trabajar con el estimador periodograma suavizado, de la demostración del Teorema 2.3.1 se sabe que:

$$E(\Sigma_n) = \frac{1}{8\pi} \mu_2(K) h_n^2 \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1''(\lambda) d\lambda + o(h_n^2) \tag{2.105}$$

y

$$\begin{aligned} n \text{Var}(\Sigma_n) &= \frac{1}{8\pi} \int_{-\pi}^{\pi} f_1^2(u) Q_{1,2}^2(u) du + \\ &+ h_n^2 \mu_2(K) \int_{-\pi}^{\pi} f_1^2(u) Q_{1,2}(u) Q_{1,2}''(u) du + o(h_n^2) \end{aligned} \tag{2.106}$$

Paso 2. A continuación se obtendrán las expresiones asintóticas correspondientes al sesgo y la varianza de $\tilde{\Sigma}_n$.

De modo análogo a como se procedía con Σ_n en (2.25, se tiene que:

$$\begin{aligned} \tilde{\Sigma}_n &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) \left\{ \hat{f}_2(\lambda) - E(\hat{f}_2(\lambda)) \right\} d\lambda + \\ &+ \frac{1}{4\pi} \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) \left\{ E(\hat{f}_2(\lambda)) - f_2(\lambda) \right\} d\lambda = \tilde{V} + \tilde{B}. \end{aligned}$$

Utilizando de nuevo la teoría asintótica de la regresión lineal local y recurriendo al teorema de la convergencia dominada se llega a que

$$E(\tilde{\Sigma}_n) = \tilde{B} = \frac{1}{r_2} \sum_{j=1}^{r_2} \frac{1}{4\pi} \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) \left\{ E(\hat{f}_{2j}(\lambda)) - f_2(\lambda) \right\} d\lambda =$$

$$= \frac{1}{8\pi} \mu_2(K) h_2^2 \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) f_2''(\lambda) d\lambda + o(h_2^2). \quad (2.107)$$

Con respecto al término correspondiente a la varianza, teniendo en cuenta que

$$\hat{f}_2(\lambda) - E(\hat{f}_2(\lambda)) = \frac{1}{r_2} \sum_{j=1}^{r_2} \left\{ \hat{f}_{2j}(\lambda) - E(\hat{f}_{2j}(\lambda)) \right\}$$

se puede escribir

$$Var(\tilde{\Sigma}_n) = \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) \tilde{Q}_{1,2}(\lambda') \tilde{\Gamma}_n(\lambda, \lambda') d\lambda d\lambda'$$

donde

$$\begin{aligned} \tilde{\Gamma}_n(\lambda, \lambda') &= \frac{1}{r_2^2} \sum_{i=1}^{r_2} \sum_{j=1}^{r_2} E \left\{ \left(\hat{f}_{2i}(\lambda) - E(\hat{f}_{2i}(\lambda)) \right) \left(\hat{f}_{2j}(\lambda') - E(\hat{f}_{2j}(\lambda')) \right) \right\} \\ &= \frac{1}{r_2^2} \sum_{i=1}^{r_2} \sum_{j=1}^{r_2} cov \left(\hat{f}_{2i}(\lambda), \hat{f}_{2j}(\lambda') \right) \end{aligned}$$

Puesto que las series X_{2j} , $j = 1, \dots, r_2$ se consideran mutuamente independientes, se tiene que:

$$\tilde{\Gamma}_n(\lambda, \lambda') = \frac{1}{r_2^2} \sum_{j=1}^{r_2} cov \left(\hat{f}_{2j}(\lambda), \hat{f}_{2j}(\lambda') \right)$$

Con lo cual se llega a que

$$\begin{aligned} Var(\tilde{\Sigma}_n) &= \\ &= \frac{1}{r_2} \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) \tilde{Q}_{1,2}(\lambda') E \left\{ \left(\hat{f}_2(\lambda) - E(\hat{f}_2(\lambda)) \right) \left(\hat{f}_2(\lambda') - E(\hat{f}_2(\lambda')) \right) \right\} d\lambda d\lambda' \end{aligned}$$

donde a partir de aquí \hat{f}_2 denota al estimador espectral obtenido a partir de cualquiera de las series de entrenamiento disponibles de la población H_2 .

Procediendo igual que para (2.38) en la demostración del Teorema 2.2.1 se tiene que:

$$\begin{aligned} Var(\mathbf{H}_{n_2} \tilde{\Sigma}_n^*) &= \\ &= \frac{1}{r_2} \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) \tilde{Q}_{1,2}(\lambda') \mathbf{e}_1^t \mathbf{H}_{n_2} \mathbf{S}_{n_2, \lambda}^{-1} \mathbf{H}_{n_2} \mathbf{V}_{n_2, \lambda, \lambda'} \mathbf{H}_{n_2} \mathbf{S}_{n_2, \lambda'}^{-1} \mathbf{H}_{n_2} \mathbf{e}_1 d\lambda d\lambda', \end{aligned}$$

donde \mathbf{H}_{n_2} es la matriz 2×2

$$\mathbf{H}_{n_2} = \begin{pmatrix} 1 & 0 \\ 0 & h_{n_2} \end{pmatrix},$$

$\mathbf{S}_{n_2, \theta}$ la matriz de elementos $s_{i,j}^{n_2}(\theta) = s_{i+j}^{n_2}(\theta)$ dados por

$$s_l^{n_2}(\theta) = \frac{1}{n_2} \sum_{t=-N_2}^{N_2} \left(\frac{\lambda_t - \theta}{h_2} \right)^l K_{h_2}(\lambda_t - \theta), l = 0, 1, 2,$$

y $\mathbf{V}_{n_2, \lambda, \lambda'}$ la matriz

$$\mathbf{V}_{n_2, \lambda, \lambda'} = \mathbf{H}_{n_2}^{-1} E \{ \mathbf{T}_{n_2, \lambda} \mathbf{T}_{n_2, \lambda'}^t \} \mathbf{H}_{n_2}^{-1} = \begin{pmatrix} v_{0,0}^{n_2}(\lambda, \lambda') & v_{0,1}^{n_2}(\lambda, \lambda') \\ v_{1,0}^{n_2}(\lambda, \lambda') & v_{1,1}^{n_2}(\lambda, \lambda') \end{pmatrix}$$

de elementos

$$v_{i,j}^{n_2}(\lambda, \lambda') = n_2^{-2} \sum_{t=-N_2}^{N_2} \sum_{s=-N_2}^{N_2} \left(\frac{\lambda_t - \lambda}{h_2} \right)^i \left(\frac{\lambda_s - \lambda'}{h_2} \right)^j \times \\ \times K_{h_2}(\lambda_t - \lambda) K_{h_2}(\lambda_s - \lambda') \gamma_{n_2}(\lambda_t, \lambda_s),$$

con $\gamma_{n_2}(\lambda_t, \lambda_s) = E \{ (I_{n_2}(\lambda_t) - f_2(\lambda_t))(I_{n_2}(\lambda_s) - f_2(\lambda_s)) \} = cov(I_{n_2}(\lambda_t), I_{n_2}(\lambda_s))$.

donde ahora $I_{n_2}(\cdot)$ denota el periodograma asociado a cualquiera de las muestras de entrenamiento disponibles de la segunda población. Puesto que todas las series se consideran realizaciones independientes de procesos lineales Gaussianos de media cero se obtiene que

$$v_{i,j}^{n_2}(\lambda, \lambda') = \frac{1}{n_2^2} \sum_{t=-N_2}^{N_2} \left(\frac{\lambda_t - \lambda}{h_2} \right)^i \left(\frac{\lambda_t - \lambda'}{h_2} \right)^j \times \\ \times K_{h_2}(\lambda_t - \lambda) K_{h_2}(\lambda_t - \lambda') (f_2^2(\lambda_t) + o(1)).$$

Procediendo de modo análogo a como se procedía a partir de (2.38) se llega a que

$$8\pi n Var(\tilde{\Sigma}_n) = \frac{1}{r_2 c_2} \int_{-\pi}^{\pi} f_2^2(u) \tilde{Q}_{1,2}^2(u) du + \\ + h_2^2 \mu_2(K) \int_{-\pi}^{\pi} f_2^2(u) \tilde{Q}_{1,2}(u) \tilde{Q}_{1,2}''(u) du + o(h_2^2) \quad (2.108)$$

Paso 3. De acuerdo con (2.104), y puesto que series de las muestras de entrenamiento de cada una de las poblaciones Π_1 y Π_2 son mutuamente independientes se tiene que, a partir de las expresiones asintóticas obtenidas para el sesgo y la varianza de Σ_n y $\tilde{\Sigma}_n$:

$$E(\hat{D}_W) = D_W(f_1, f_2) + \frac{1}{8\pi} \mu_2(K) h_n^2 \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1''(\lambda) d\lambda - \\ - \frac{1}{8\pi} \mu_2(K) h_2^2 \int_{-\pi}^{\pi} \tilde{Q}_{1,2}(\lambda) f_2''(\lambda) d\lambda + o(h_n^2) + o(h_2^2)$$

y

$$nVar(\hat{D}_W) = \frac{1}{8\pi} \int_{-\pi}^{\pi} f_1^2(u) Q_{1,2}^2(u) du + \frac{1}{8\pi} h_n^2 \mu_2(K) \int_{-\pi}^{\pi} f_1^2(u) Q_{1,2}(u) Q_{1,2}''(u) du + \\ + \frac{1}{8\pi c_2 r_2} \int_{-\pi}^{\pi} f_2^2(u) \tilde{Q}_{1,2}^2(u) du + \frac{1}{8\pi c_2 r_2} h_2^2 \mu_2(K) \int_{-\pi}^{\pi} f_2^2(u) \tilde{Q}_{1,2}(u) \tilde{Q}_{1,2}''(u) du + \\ + o(h_2^2) + o(h_n^2).$$

De donde, realizando operaciones, se tiene que

$$E(\widehat{D}_W) = D_W(f_1, f_2) + \frac{1}{8\pi} \mu_2(K) h_n^2 \int_{-\pi}^{\pi} Q_{1,2}(\lambda) f_1''(\lambda) d\lambda - \\ - \frac{1}{8\pi} \mu_2(K) h_n^2 \int_{-\pi}^{\pi} \frac{f_1(\lambda)}{f_2(\lambda)} Q_{1,2}(\lambda) f_2''(\lambda) d\lambda + o(h_n^2) + o(h_n^2)$$

y

$$nVar(\widehat{D}_W) = \\ = \frac{1}{8\pi} \int_{-\pi}^{\pi} f_1^2(u) Q_{1,2}^2(u) du + \frac{1}{8\pi c_2} \int_{-\pi}^{\pi} \frac{1}{r_2} f_1^2(u) Q_{1,2}^2(u) du + O(h_n^2) + O(h_n^2).$$

Paso 4.

Para completar la demostración, quedaría por demostrar la normalidad asintótica del estadístico plug-in \widehat{D}_W . De acuerdo con la notación empleada, puede escribirse, bajo Π_1 :

$$\sqrt{n} \left\{ D_W - \widehat{\mu}_{W, \widehat{f}_{DLS}}(1, 2) \right\} = \frac{\sqrt{n}}{4\pi} \left\{ \int_{-\pi}^{\pi} Q_{1,2}(\lambda) e_1^t \mathbf{S}_{n,\lambda}^{-1} b m T_{n,\lambda} d\lambda - \right. \\ \left. - \frac{1}{r_2} \sum_{j=1}^{r_2} \int_{-\pi}^{\pi} \widetilde{Q}_{1,2}(\lambda) e_1^t \mathbf{S}_{n_2,\lambda}^{-1} \mathbf{T}_{n_2,\lambda}^j d\lambda \right\} (1 + o(1)),$$

donde $\mathbf{T}_{n_2,\lambda}^j$ es el vector 2x1 de elementos

$$t_{l,j}^{n_2}(\theta) = \frac{1}{n_2} \sum_{t=-N_2}^{N_2} \left(\frac{\lambda_t - \theta}{h_2} \right)^l K_{h_2}(\lambda_t - \theta) (I_{n_2,j}(\lambda_t) - f_2(\lambda_t)), l = 0, 1.$$

De este modo, la normalidad asintótica de $\sqrt{n} \left\{ D_W - \widehat{\mu}_{W, \widehat{f}_{DLS}}(1, 2) \right\}$ se deduce de modo inmediato una vez que se haya probado la normalidad asintótica de los vectores aleatorios

$$\sqrt{n} \int_{-\pi}^{\pi} \widetilde{Q}_{1,2}(\lambda) e_1^t \mathbf{T}_{n_2,\lambda}^j d\lambda, j = 1, 2, \dots, r_2$$

lo cual puede demostrarse siguiendo un razonamiento completamente análogo al utilizado en la demostración del Teorema 2.3.1 para (2.47).

Capítulo 3

Comportamiento del procedimiento discriminante sobre muestras finitas

3.1. Estudio de simulación.

En esta sección se presentarán los resultados de un estudio de simulación llevado a cabo para investigar el comportamiento del estadístico discriminante D_W propuesto en el Capítulo 2 a la hora de clasificar realizaciones de diferentes clases de procesos estocásticos. El estudio de simulación fue llevado a cabo con un doble propósito: en primer lugar, determinar las tasas de mala clasificación alcanzadas cuando se utiliza el estadístico D_W para discriminar entre diferentes clases de procesos y, en segundo lugar, comparar los resultados obtenidos con el estadístico D_W utilizando para la estimación de la densidad espectral cada uno de los tres suavizadores locales:

- El periodograma suavizado por mínimos cuadrados locales, \hat{f}_{DLS} .
- El logaritmo del periodograma suavizado por mínimos cuadrados locales, \hat{f}_{LS} .
- El suavizador del log-periodograma basado en máxima verosimilitud, \hat{f}_{LK} .

En primer lugar, se consideró el problema de discriminar entre procesos ARMA(p,q) con errores gaussianos. En un primer momento, las simulaciones se centraron en estudiar el comportamiento del estadístico D_W a la hora de discriminar únicamente entre dos poblaciones de procesos AR(1) o MA(1) con diferentes parámetros. A continuación se

procedió a examinar sus propiedades para discriminar entre un conjunto más grande de procesos de tipo ARMA(p,q), todos ellos con errores gaussianos. Finalmente, se estudiaron las propiedades de dicho estadístico en la discriminación de procesos con distribución marginal no normal y distintas estructuras de dependencia, simulando series a partir de procesos de diferente tipo:

- Procesos AR(1) con distribución marginal exponencial
- Procesos AR(1) con distribución marginal doble exponencial
- Series m-dependientes con distribución marginal gamma
- Series m-dependientes con distribución marginal lognormal

Estas series cubren un amplio espectro de los procesos estocásticos con los que se puede trabajar en la práctica, y constituyen un buen ejemplo para analizar el comportamiento del estadístico discriminante D_W a la hora de trabajar sobre muestras finitas en la clasificación de diferentes tipos de procesos.

Esta sección se estructura a lo largo de cuatro apartados. En primer lugar, se ofrece una descripción general del esquema seguido para el estudio de simulación que se ha llevado a cabo. A continuación, se procede a describir los resultados de las simulaciones para cada uno de los contextos analizados, según se discrimine entre procesos con errores gaussianos o entre procesos con distribución marginal no gaussiana. Finalmente, se realiza una discusión de los resultados alcanzados y se exponen las conclusiones que se han derivado del estudio de simulación.

3.1.1. Procedimiento experimental.

En todos los casos se siguió un mismo esquema para llevar a cabo las simulaciones, tal y como se detalla a continuación:

1. Especificación de los procesos a clasificar.

Para cada uno de los estudios de simulación llevados a cabo, se determinaron a priori las clases de procesos entre las que se pretendía discriminar. Se seleccionaron modelos que pudiesen dificultar las labores de clasificación, con índices de separabilidad bajos y/o previamente considerados en otros trabajos de temática similar. En general se seleccionaron $G \geq 2$ clases de procesos especificadas de acuerdo con los parámetros que se definirán en cada apartado de las simulaciones.

2. Generación de las series de tiempo.

Una vez determinados los modelos sujetos a discriminación, se obtuvieron s realizaciones de longitud n de cada uno de ellos, tomando n diferentes valores. En general, para cada ejemplo, se tomó $n = 200$ y $n = 400$ para valorar la influencia de la longitud de las series sobre los resultados obtenidos en la clasificación.

3. Clasificación de las series.

En el caso en el que se discriminase entre $G = 2$ clases de procesos, Π_1 y Π_2 , a continuación cada una de las series generadas se clasificó como perteneciente a una de las dos poblaciones de acuerdo con el estadístico discriminante:

$$D_W = D_W(\hat{f}_n, f_2) - D_W(\hat{f}_n, f_1),$$

tal y como se explicó en la Sección 2.2, y siendo \hat{f}_n un estimador de la densidad espectral correspondiente. Para el cálculo de D_W se utilizó cada uno de los tres estimadores no paramétricos del espectro propuestos por Fan y Kreutzberger (1998): \hat{f}_{DLS} , \hat{f}_{LS} y \hat{f}_{LK} , usando en todos los casos como función de divergencia la función:

$$W(x) = \log(\alpha x + (1 - \alpha)) - \alpha \log(x) \tag{3.1}$$

con $\alpha = 0,5$.

Para el cómputo de cualquiera de los tres estimadores espectrales se utilizó el núcleo de Epanechnikov, y el parámetro ventana h se determinó en cada caso minimizando la función de validación cruzada sobre un rango adecuado de valores de h . Para el cómputo del estimador del espectro basado en el criterio de máxima verosimilitud local se utilizó el algoritmo de Newton-Raphson.

En el caso en el que se discriminase entre $G > 2$ clases de procesos, $\Pi_j, j = 1, 2, \dots, G$, se generalizó el procedimiento anterior y cada una de las series generadas se asignó al modelo con el que mantenía la mínima discrepancia, evaluada de acuerdo con la medida de disparidad espectral

$$D_W(\hat{f}_n, f_j) = \int_{-\pi}^{\pi} W(\hat{f}_n(\lambda) f_j^{-1}(\lambda)) d\lambda, \quad j = 1, 2, \dots, G,$$

donde \hat{f}_n y W vienen dados como en el caso anterior.

4. Evaluación de los resultados de clasificación.

Para valorar los resultados de la clasificación, se calculó la proporción de series mal clasificadas para cada una de las poblaciones de origen. Finalmente, se promediaron estas tasas de error para obtener una tasa global de mala clasificación. Este calculo se realizó teniendo en cuenta los resultados obtenidos con el procedimiento discriminante descrito utilizando cada uno de los tres estimadores del espectro. De esta forma,

pudo valorarse la influencia del estimador empleado en los resultados alcanzados comparando las tasas globales de error correspondientes.

3.1.2. Discriminación de procesos con errores gaussianos.

a) Discriminación de procesos AR(1).

En primer lugar, se consideró el problema de discriminar entre $G = 2$ clases de procesos autorregresivos AR(1). Así, bajo Π_j , $j = 1, 2$, el proceso X sigue una estructura autorregresiva dada por

$$X_t = \phi_j X_{t-1} + \varepsilon_t, \quad |\phi_j| < 1, \quad (3.2)$$

donde $\{\varepsilon_t\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas según una normal $N(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon > 0$.

Es bien conocido que, si el proceso X sigue una estructura como la dada en (3.2), su densidad espectral vendrá determinada por la expresión

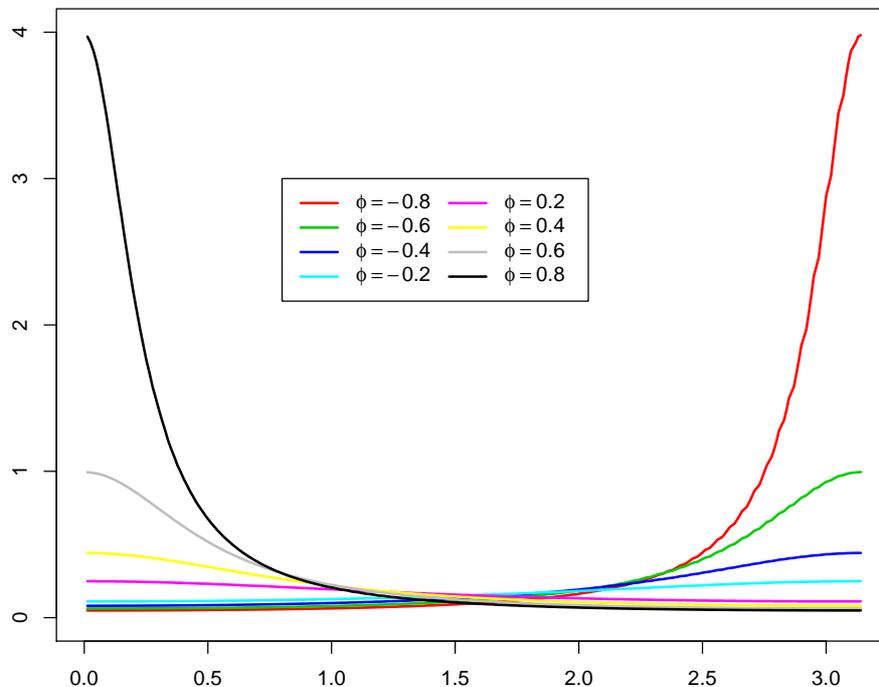
$$f_j(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi(1 + \phi_j^2 - 2\phi_j \cos \lambda)}.$$

En la Figura 3.1 se muestra así la densidad espectral correspondiente a procesos autorregresivos AR(1) según los diferentes valores del parámetro ϕ_j en (3.2).

Para el estudio de simulación, se generaron $s = 100$ series de longitud $n = 400$ a partir de cada una de las clases Π_j , ($j = 1, 2$), de acuerdo con la ecuación (3.2), para diferentes valores de los parámetros ϕ_1 y ϕ_2 y tomando siempre $\sigma_\varepsilon = 1$. En la Figura 3.2 se muestra un conjunto de realizaciones aleatorias de algunos de estos procesos. Como se puede observar, las series muestran mayor variabilidad en el caso de valores negativos del parámetro autorregresivo y de mayor valor absoluto, concentrándose en estos casos la densidad espectral sobre las frecuencias mayores, mientras que para el caso de procesos de parámetros positivos la potencia del espectro se concentra en las frecuencias de menor magnitud, dando lugar a realizaciones más “suaves”.

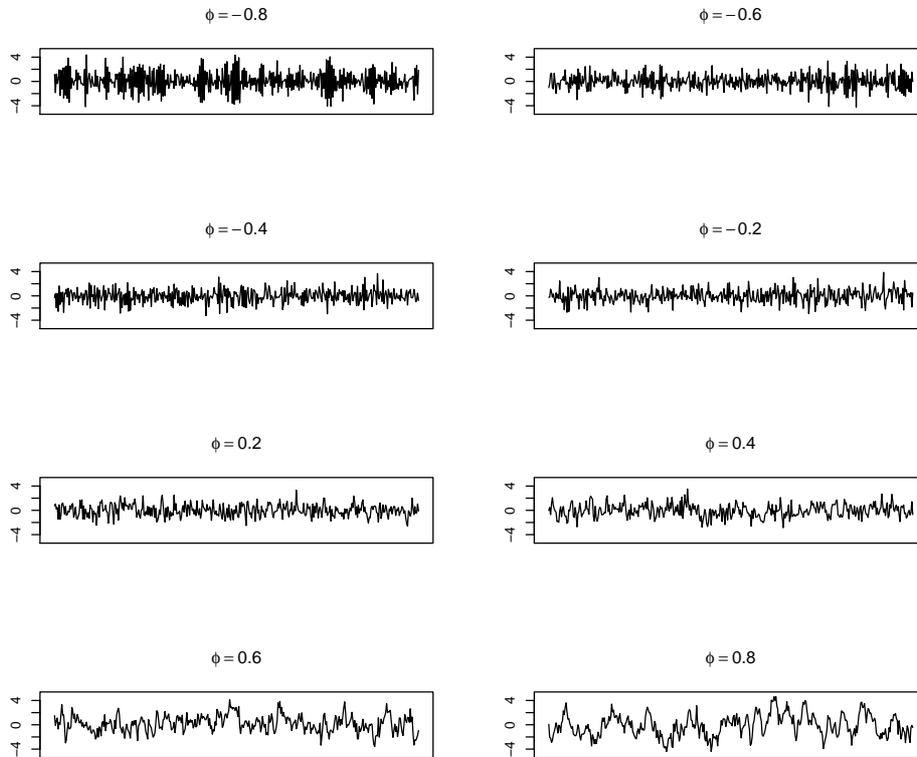
Cada una de las series generadas se clasificó como perteneciente a una de las dos poblaciones de acuerdo con el procedimiento descrito en la Sección 3.1.1. Se calculó la proporción de series mal clasificadas para cada una de las dos poblaciones y finalmente, ambas tasas de error se promediaron para obtener una tasa global de mala clasificación con cualquiera de los tres estimadores del espectro utilizados. Los resultados obtenidos se muestran en la Tabla 3.1.

Figura 3.1: Densidad espectral correspondiente a diferentes procesos AR(1) con errores gaussianos $N(0,1)$, según diferentes valores del parámetro autorregresivo ϕ en (3.2).



Tal y como se puede apreciar, los resultados muestran el buen comportamiento del estadístico discriminante D_W , con tasas de error que sólo en ocasiones superan el 20% y siempre cuando los procesos entre los que se discrimina vienen definidos por parámetros autorregresivos muy similares. Se observan mejores resultados a medida que la distancia entre los parámetros ϕ_1 y ϕ_2 aumenta. Esto es lo que cabía esperar, dado que la distancia entre estos valores determina la separabilidad entre las categorías Π_1 y Π_2 entre las que se discrimina. De hecho, una medida habitualmente empleada en el contexto del análisis discriminante para evaluar la separabilidad entre dos poblaciones (y por lo tanto la dificultad para discriminar entre ellas) es la tasa de J-divergencia, definida en (2.7). En la Tabla 3.2 se muestran los valores de la tasa de J-divergencia para el caso de discriminar entre dos clases de procesos autorregresivos AR(1) con errores gaussianos. Se observa cómo dicha tasa se hace mayor a medida que aumenta la distancia entre los parámetros autorregresivos que definen las dos poblaciones, pasando de un valor de 0,010 para el caso de discriminar entre dos procesos AR(1) de parámetros $\phi_1 = 0,1$ y $\phi_2 = 0,2$ a un valor

Figura 3.2: Realizaciones de procesos autorregresivos AR(1) con errores gaussianos $N(0,1)$ y diferentes valores del parámetro autorregresivo ϕ en (3.2).



de 1,930 para el caso de discriminar entre dos procesos AR(1) de parámetros $\phi_1 = 0,1$ y $\phi_2 = 0,9$. Asimismo, se objetiva que la separabilidad es menor entre procesos definidos por parámetros autorregresivos bajos que para parámetros autorregresivos altos. Así, por ejemplo, la tasa de J-divergencia para discriminar entre dos procesos AR(1) de parámetros $\phi_1 = 0,1$ y $\phi_2 = 0,2$ es de 0,010, frente a 0,023 para el caso de dos procesos AR(1) de parámetros $\phi_1 = 0,7$ y $\phi_2 = 0,8$. Esto se traduce en una tasa de mala clasificación más baja en este último caso, que trabajando con el estimador periodograma suavizado es de un 6,8 %, frente al 19,8 % de error que se obtiene con este mismo estimador para el caso de discriminar entre procesos autorregresivos AR(1) de parámetros $\phi_1 = 0,1$ y $\phi_2 = 0,2$. Conclusiones análogas se obtienen con los otros dos estimadores espectrales y otros valores estudiados de los parámetros autorregresivos.

Por otro lado, aunque los resultados alcanzados con cualquiera de los tres estimadores son buenos, parece que las propiedades asintóticas de cada uno de los suavizadores no paramétricos ejercen cierta influencia sobre el comportamiento del estadístico D_W . Así,

Tabla 3.1: Tasas de mala clasificación obtenidas al utilizar el estadístico D_W para discriminar entre dos clases de procesos AR(1) con errores gaussianos $N(0,1)$ y diferentes parámetros autorregresivo ϕ_1 y ϕ_2 . Longitud de las series: $n=400$.

$\phi_1 \downarrow$	$\phi_2 \rightarrow$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
0	\hat{f}_{DLS}	0,198	0,048						
	\hat{f}_{LS}	0,213	0,188						
	\hat{f}_{LK}	0,183	0,035						
0,1	\hat{f}_{DLS}		0,165	0,023					
	\hat{f}_{LS}	*	0,200	0,095					
	\hat{f}_{LK}		0,160	0,030					
0,2	\hat{f}_{DLS}			0,148	0,028				
	\hat{f}_{LS}		*	0,225	0,065				
	\hat{f}_{LK}			0,165	0,023				
0,3	\hat{f}_{DLS}				0,173	0,025			
	\hat{f}_{LS}			*	0,245	0,045			
	\hat{f}_{LK}				0,195	0,028			
0,4	\hat{f}_{DLS}					0,138	0,013		
	\hat{f}_{LS}				*	0,220	0,040		
	\hat{f}_{LK}					0,158	0,020		
0,5	\hat{f}_{DLS}						0,113	0,013	
	\hat{f}_{LS}					*	0,223	0,055	
	\hat{f}_{LK}						0,145	0,020	
0,6	\hat{f}_{DLS}							0,100	0,002
	\hat{f}_{LS}						*	0,175	0,037
	\hat{f}_{LK}							0,125	0,005
0,7	\hat{f}_{DLS}								0,068
	\hat{f}_{LS}							*	0,158
	\hat{f}_{LK}								0,125

el log-periodograma suavizado \hat{f}_{LS} resulta menos eficiente que los otros dos estimadores y por ello D_W muestra el peor comportamiento en todos los casos cuando se utiliza este estimador espectral. Sin embargo, los resultados alcanzados con los suavizadores \hat{f}_{DLS} y \hat{f}_{LK} son muy similares y no se aprecian diferencias reseñables entre ellos. De hecho, la mayor tasa de error fue sólo de un 19,8% con \hat{f}_{DLS} y de un 19,5% con \hat{f}_{LK} . En realidad, cabría esperar un peor comportamiento del estadístico D_W con \hat{f}_{LK} dado que tiene un sesgo asintótico mayor que \hat{f}_{DLS} en las regiones con pico del espectro.

A su vez, si se comparan los resultados alcanzados con cualquiera de los tres estimadores con las tasas de error reflejadas en otros trabajos al discriminar entre procesos AR(1) con errores gaussianos, puede concluirse que las tasas obtenidas resultan competitivas, si se tiene en cuenta además que en dichas investigaciones se trabajó con series de longitud mayor (Zhang y Taniguchi (1994), Zhang y Taniguchi (1995), Kakizawa (1996)). Así, por ejemplo, Zhang y Taniguchi (1994), utilizando una aproximación del cociente de verosimilitudes calculada a partir de los correspondientes periodogramas, obtienen una tasa de

Tabla 3.2: Separabilidad entre procesos AR(1) con errores gaussianos $N(0,1)$ y diferentes parámetros autorregresivos ϕ_1, ϕ_2 , evaluada mediante la tasa de J-divergencia.

$\phi_1 \downarrow$	ϕ_2							
	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,1	0,010	0,042	0,099	0,186	0,319	0,528	0,911	1,930
0,2		0,011	0,044	0,106	0,206	0,371	0,674	1,484
0,3			0,011	0,048	0,119	0,241	0,475	1,099
0,4				0,013	0,054	0,140	0,311	0,773
0,5					0,014	0,065	0,181	0,505
0,6						0,017	0,085	0,293
0,7							0,023	0,037
0,8								0,038

error del 15.9% al discriminar entre procesos AR(1) de parámetros $\phi_1 = 0,1$ y $\phi_2 = 0,2$, con series de longitud $n = 1024$. Dicha tasa es de un 6.7% cuando se discrimina entre procesos AR(1) de parámetros $\phi_1 = 0,1$ y $\phi_2 = 0,3$ y series de igual longitud. Estos porcentajes de error son igualados por el estadístico discriminante D_W trabajando con series de menor longitud ($n = 400$) obteniéndose, por ejemplo, con el estimador \hat{f}_{DLS} , tasas de error de un 16.5% y de un 2.3% en cada caso, respectivamente.

Conclusiones similares pueden extraerse si se comparan los resultados en la Tabla 3.1 con los alcanzados por Kakizawa (1996), utilizando la misma medida de disparidad que la empleada en esta memoria con otro tipo de estimadores espectrales. En este caso, los autores realizaron un estudio de simulación en el que se discriminó entre series de longitud $n = 512$ procedentes de dos clases de procesos autorregresivos AR(1) con errores gaussianos $N(0,1)$ y parámetros $\phi_1 = 0,6$ y $\phi_2 = 0,7$. Las tasas de error obtenidas, independientemente de la función de divergencia utilizada, se situaron en torno al 7.5%. Bajo este mismo supuesto se han alcanzado con el procedimiento discriminante propuesto en esta memoria tasas de error de un 10%, 17.5% y 12.5% según se utilizase el estimador \hat{f}_{DLS} , \hat{f}_{LS} o \hat{f}_{LK} , respectivamente. Se obtienen así tasas de error ligeramente superiores que pudiesen ser explicadas por la menor longitud de las series consideradas en esta simulación.

A este respecto, y con el fin de determinar la influencia de la longitud de las series en los resultados, a continuación se repitió el estudio de simulación con series de menor longitud. Las Tablas 3.3 y 3.4 muestran los resultados alcanzados al discriminar entre series de longitud $n=200$. Obviamente, el comportamiento de la regla discriminante es algo peor que el observado en los resultados anteriores, aunque pueden extraerse conclusiones

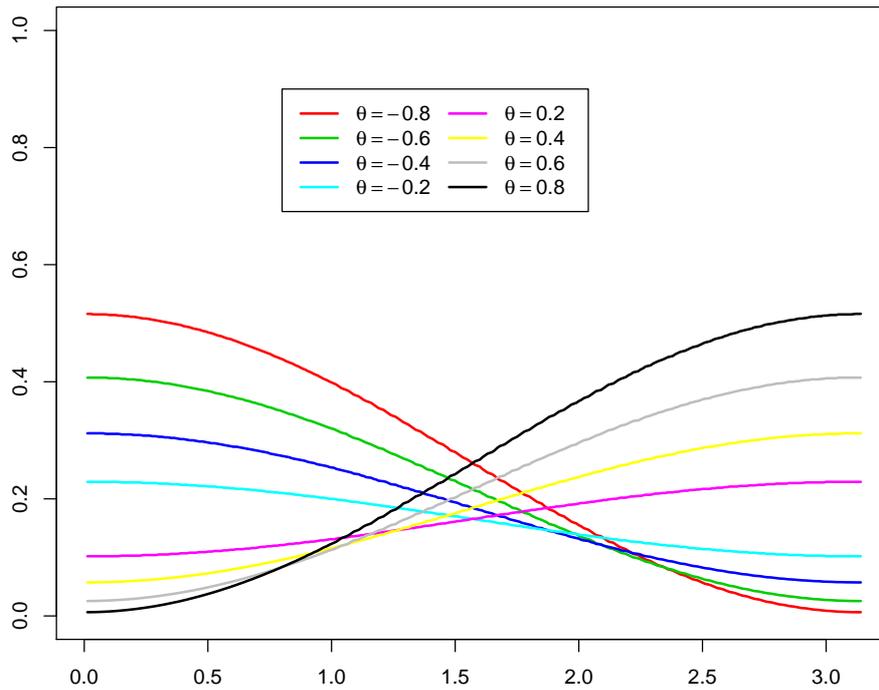
b) Discriminación de procesos MA(1).

A continuación se procedió a discriminar entre dos clases de procesos de medias móviles MA(1) para estudiar el funcionamiento del procedimiento discriminante propuesto en esta situación. En este caso, bajo Π_j , ($j = 1, 2$), el proceso X sigue una estructura de medias móviles dada por

$$X_t = \theta_j \varepsilon_{t-1} + \varepsilon_t, \quad |\theta_j| < 1, \quad (3.3)$$

donde $\{\varepsilon_t\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas según una normal $N(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon > 0$.

Figura 3.3: Densidad espectral correspondientes a diferentes procesos MA(1) con errores gaussianos $N(0,1)$ según diferentes valores del parámetro θ en (3.3).

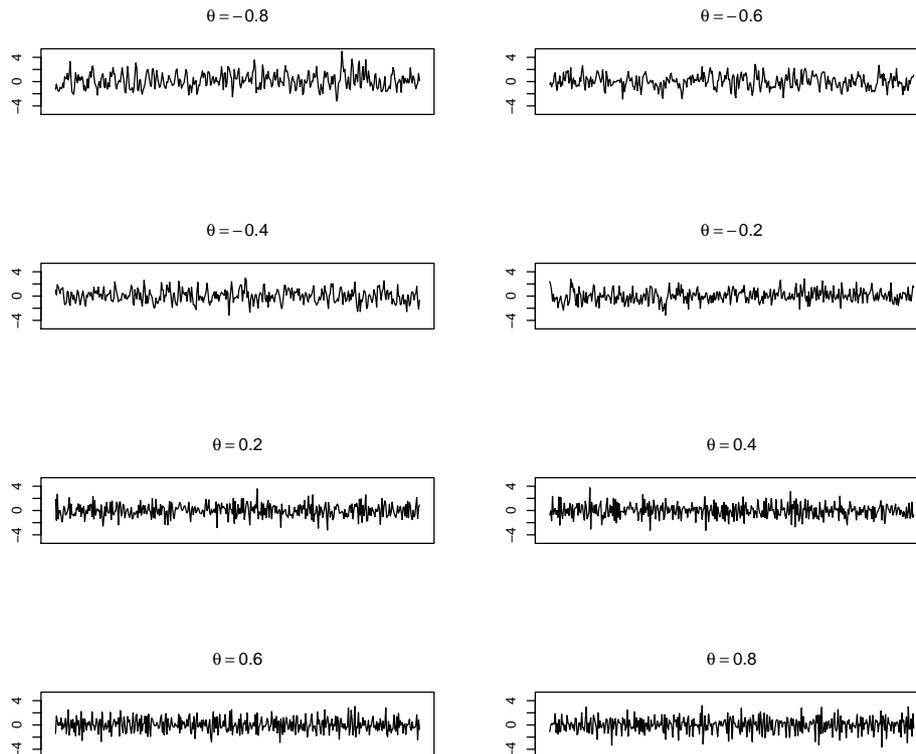


Si el proceso X sigue una estructura como la dada en (3.3), su densidad espectral vendrá determinada por la expresión

$$f_j(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi(1 + 2\theta_j \cos \lambda + \theta_j^2)}$$

según se indica en la Figura 3.3. En la Figura 3.4 se muestra un conjunto de realizaciones aleatorias de algunos de estos procesos. En este caso, las series que muestran una mayor variabilidad se corresponden con procesos MA(1) de parámetros positivos, concentrándose entonces la densidad espectral en las frecuencias de mayor magnitud.

Figura 3.4: Realizaciones de procesos autorregresivos MA(1) con errores gaussianos $N(0,1)$, correspondientes a diferentes valores del parámetro θ en (3.3).



Para el estudio de simulación, se procedió como en el caso anterior, generando $s = 100$ series de longitud $n = 400$ a partir de cada una de las clases Π_j , ($j = 1, 2$), de acuerdo ahora con la ecuación (3.3), para diferentes valores de los parámetros θ_1 y θ_2 y tomando siempre $\sigma_\varepsilon = 1$. Cada una de las series se clasificó de acuerdo con el estadístico discriminante $D_W = D_W(\hat{f}_n, f_2) - D_W(\hat{f}_n, f_1)$, utilizando cada uno de los tres estimadores espectrales: \hat{f}_{DLS} , \hat{f}_{LS} y \hat{f}_{LK} tal y como se indicó previamente. Los resultados obtenidos, en términos de la tasa global de mala clasificación, se muestran en la Tabla 3.5. A su vez, y tal y como se puede deducir de las expresiones teóricas de las densidades espectrales correspondientes, la tasa de J-divergencia para el caso de discriminar entre dos procesos MA(1) toma valores idénticos a los que se mostraron en la Tabla 3.2 para dos procesos AR(1), reemplazando

los parámetros autorregresivos ϕ_1 y ϕ_2 por idénticos valores para los correspondientes parámetros θ_1 y θ_2 .

Tabla 3.5: Tasas de mala clasificación obtenidas al utilizar el estadístico D_W para discriminar entre dos clases de procesos MA(1) con errores gaussianos $N(0,1)$ y diferentes parámetros θ_1 y θ_2 . Longitud de las series: $n=400$.

$\theta_1 \downarrow$	$\theta_2 \rightarrow$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
0	\hat{f}_{DLS}	0,185	0,035						
	\hat{f}_{LS}	0,217	0,080						
	\hat{f}_{LK}	0,195	0,025						
0,1	\hat{f}_{DLS}		0,205	0,025					
	\hat{f}_{LS}	*	0,282	0,060					
	\hat{f}_{LK}		0,222	0,025					
0,2	\hat{f}_{DLS}			0,142	0,040				
	\hat{f}_{LS}		*	0,227	0,075				
	\hat{f}_{LK}			0,157	0,045				
0,3	\hat{f}_{DLS}				0,162	0,040			
	\hat{f}_{LS}			*	0,235	0,070			
	\hat{f}_{LK}				0,180	0,045			
0,4	\hat{f}_{DLS}					0,182	0,060		
	\hat{f}_{LS}				*	0,222	0,065		
	\hat{f}_{LK}					0,162	0,035		
0,5	\hat{f}_{DLS}						0,185	0,110	
	\hat{f}_{LS}					*	0,227	0,055	
	\hat{f}_{LK}						0,202	0,035	
0,6	\hat{f}_{DLS}							0,192	0,125
	\hat{f}_{LS}						*	0,210	0,070
	\hat{f}_{LK}							0,165	0,030
0,7	\hat{f}_{DLS}								0,262
	\hat{f}_{LS}							*	0,240
	\hat{f}_{LK}								0,195

De modo general, se han obtenido tasas de error similares a las alcanzadas en el caso de discriminar entre dos clases de procesos autorregresivos AR(1), que superan el 20% en el caso de procesos entre los que exista muy poca separabilidad y que en otro caso se sitúan por debajo del 10%, objetivándose tasas en torno a un 2% o 4% incluso para clases de procesos entre los que la separabilidad sigue siendo escasa. Así mismo, sigue observándose la repercusión de las propiedades asintóticas de los suavizadores espectrales sobre los resultados alcanzados en las simulaciones, con tasas de mala clasificación más elevadas en el caso del estimador log-periodograma suavizado por mínimos cuadrados (\hat{f}_{LS}) y tasas similares en el caso de los otros dos estimadores del espectro. En particular, la mayor tasa de error fue de un 26,2% en el caso del estimador periodograma suavizado (\hat{f}_{DLS}) y de un 22,2% para el caso del estimador log-periodograma suavizado mediante máxima verosimilitud (\hat{f}_{LK}).

Una vez más, los resultados obtenidos con cualquiera de los tres suavizadores empeoran ligeramente al repetir el esquema de simulación con series de menor longitud. En las Tablas 3.6 y 3.7 se muestran las tasas de error obtenidas tras clasificar realizaciones aleatorias de longitud $n=200$. En este caso las máximas tasas de error, al igual que ocurría cuando se discriminaba entre procesos autorregresivos AR(1), se sitúan en torno al 30 %, siempre cuando se trata de discriminar entre procesos muy similares. La tasa de mala clasificación desciende a medida que aumenta la diferencia entre los parámetros que definen los modelos sujetos a discriminación, situándose en torno al 5 % cuando estos parámetros difieren en 0,2. De nuevo, se obtienen mejores resultados cuando la estimación del espectro se realiza mediante los estimadores periodograma suavizado, \hat{f}_{DLS} , y log-periodograma suavizado mediante máxima verosimilitud, \hat{f}_{LK} .

En las Tablas 3.6 y 3.7 se pueden comparar asimismo los resultados alcanzados por el procedimiento propuesto en esta memoria con los obtenidos en el trabajo de Chan, Chinipardaz y Cox (1996). Dichos autores proponen una solución analítica para estimar la distribución del estadístico discriminante cuadrático que resulta de aplicar el criterio de clasificación basado en el cociente de verosimilitudes, en caso de trabajar con procesos ARMA(p,q). Tal y como se puede observar, con dicho procedimiento se obtuvieron resultados similares, con tasas de error por encima del 20 % en caso de procesos definidos por coeficientes muy parecidos. Parece observarse un mejor comportamiento del procedimiento propuesto por Chan, Chinipardaz y Cox (1996) en los casos de procesos más separables, en términos de la tasa de J-divergencia (aquellos pares de procesos definidos por coeficientes de mayor valor absoluto), mientras que el criterio discriminante propuesto en esta memoria alcanza resultados ligeramente mejores en situaciones de menor separabilidad.

c) Discriminación de procesos ARMA(p,q).

A continuación se estudió el comportamiento del estadístico D_W a la hora de discriminar entre diferentes procesos de tipo ARMA(p,q). Para ello se llevó a cabo un experimento de simulación en el que se generaron realizaciones de longitudes $n=200$ y $n=400$ del siguiente modelo:

$$(1 - \phi B)(1 - \phi_s B^s) X_t = (1 - \theta_r B^r) \varepsilon_t, \quad (3.4)$$

donde $B^k X_t = X_{t-k}$ para $k \in \mathbb{Z}$ y ε_t sigue una distribución $N(0, \sigma_\varepsilon^2)$. El parámetro ϕ se consideró fijo $\phi = 0,8$ y $\sigma_\varepsilon = 1$, mientras que los parámetros s, r, ϕ_s y θ_r tomaron diferentes valores correspondientes a distintos procesos ARMA(p,q) según se indica a continuación:

- | | |
|--|---|
| 1. $s=0; r=0; \phi_s = 0; \theta_r = 0$ | 8. $s=0; r=1; \phi_s = 0; \theta_r = -0,3$ |
| 2. $s=1; r=0; \phi_s = -0,3; \theta_r = 0$ | 9. $s=0; r=3; \phi_s = 0; \theta_r = -0,3$ |
| 3. $s=3; r=0; \phi_s = -0,3; \theta_r = 0$ | 10. $s=0; r=6; \phi_s = 0; \theta_r = -0,3$ |
| 4. $s=6; r=0; \phi_s = -0,3; \theta_r = 0$ | 11. $s=0; r=1; \phi_s = 0; \theta_r = -0,5$ |
| 5. $s=1; r=0; \phi_s = -0,5; \theta_r = 0$ | 12. $s=0; r=3; \phi_s = 0; \theta_r = -0,5$ |
| 6. $s=3; r=0; \phi_s = -0,5; \theta_r = 0$ | 13. $s=0; r=6; \phi_s = 0; \theta_r = -0,5$ |
| 7. $s=6; r=0; \phi_s = -0,5; \theta_r = 0$ | |

Estos procesos han sido utilizados con anterioridad en otro trabajo (Paparoditis (2000)) para estudiar las propiedades de un nuevo test de bondad de ajuste basado en estimaciones núcleo de la densidad espectral. Constituyen un ejemplo adecuado para valorar el comportamiento del estadístico D_W a la hora de discriminar entre procesos autorregresivos (Procesos 1-7) y procesos autorregresivos y de medias móviles (Procesos 8-13), observando además la influencia sobre los resultados de la magnitud de las correlaciones con cada uno de los retardos o de las innovaciones correspondientes. Los procesos resultantes se muestran en la Tabla 3.8.

Es conocido que, para el caso de un proceso ARMA(p,q) como en (3.4) la densidad espectral correspondiente viene dada como:

$$f(\lambda) = \frac{\sigma_\varepsilon^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|}, \quad -\pi \leq \lambda \leq \pi \quad (3.5)$$

donde aquí $\phi(B) = (1 - \phi B)(1 - \phi_s B^s) = 1 - \phi_s B^s - \phi B + \phi \phi_s B^{s+1}$; $\theta(B) = 1 - \theta_r B^r$ y $\sigma_\varepsilon = 1$.

Para el ejemplo que nos ocupa, las densidades espectrales correspondientes a los 13 procesos estudiados son las que se muestran en la Figura 3.5. Así mismo, en la Figura 3.6

Tabla 3.8: Procesos ARMA(p,q) sujetos a discriminación.

	Parámetros AR	Parámetros MA
1. AR(1)	0.8	
2. AR(2)	(-0.5; 0.24)	
3. AR(4)	(0.8; 0; -0.3; 0.24)	
4. AR(7)	(0.8; 0; 0; 0; 0; -0.3; 0.24)	
5. AR(2)	(-0.3; 0.4)	
6. AR(4)	(0.8; 0; -0.5; 0.4)	
7. AR(7)	(0.8; 0; 0; 0; 0; -0.5; 0.4)	
8. ARMA(1,1)	0.8	0.3
9. ARMA(1,3)	0.8	(0; 0; 0.3)
10. ARMA(1,6)	0.8	(0; 0; 0; 0; 0; 0.3)
11. ARMA(1,1)	0.8	0.5
12. ARMA(1,3)	0.8	(0; 0; 0.5)
13. ARMA(1,6)	0.8	(0; 0; 0; 0; 0; 0.5)

se representa una realización aleatoria de cada uno de estos procesos. Tal y como se puede apreciar, se trata de procesos muy similares en su mayoría, lo cual debe tenerse en cuenta a la hora de valorar los resultados obtenidos en la clasificación.

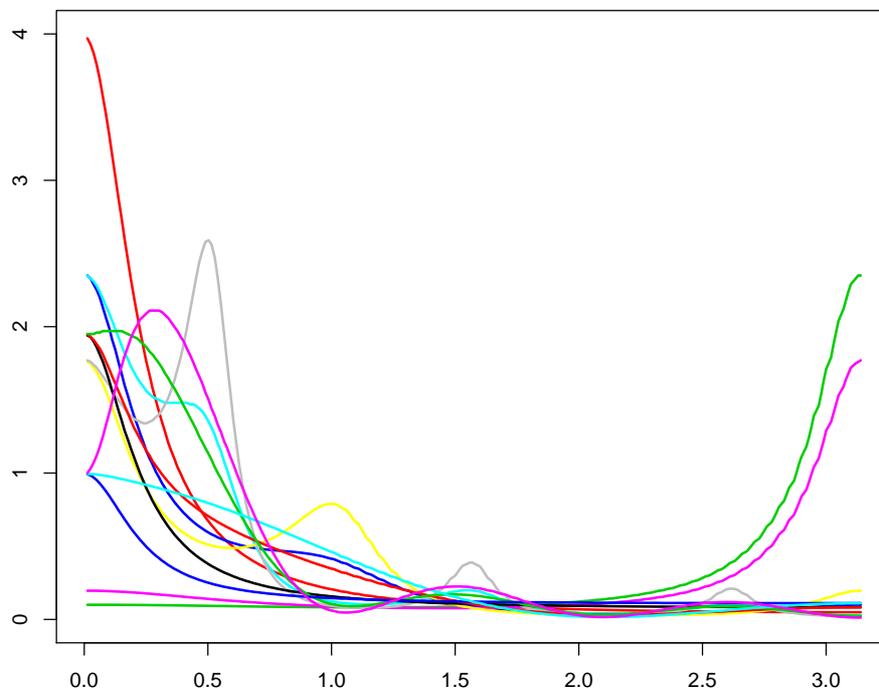
Para la simulación, se generaron $s = 100$ series de longitud $n=200$ y $n=400$ de cada uno de los 13 modelos considerados, de acuerdo con la ecuación (3.4). Cada una de las series generadas fue asignada al modelo con el que mantenía una mínima discrepancia, evaluada de acuerdo con la medida de disparidad espectral

$$D_W(\hat{f}_n, f_j) = \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\hat{f}_n(\lambda) f_j^{-1}(\lambda)) d\lambda, \quad j = 1, 2, \dots, 13, \quad (3.6)$$

donde $f_j, j = 1, 2, \dots, 13$ denota a la densidad espectral correspondiente a cada una de las clases de procesos consideradas y \hat{f}_n a un estimador de la verdadera densidad espectral de cada una de las series a clasificar.

Las Tablas 3.9 y 3.10 muestran los resultados obtenidos con el procedimiento discriminante descrito, una vez generadas $s = 100$ series de longitud $n=200$ de cada una de las clases de procesos consideradas. La tasa global de mala clasificación, para cada uno de los tres estimadores espectrales utilizados (\hat{f}_{DLS} , \hat{f}_{LS} , \hat{f}_{LK}), fue de un 24,8%, un 33,4% y un 27,1%, respectivamente. De nuevo se aprecia una tasa de error más elevada en el caso de trabajar con el estimador log-periodograma suavizado por mínimos cuadrados, que evidencia la influencia de las propiedades asintóticas de estos suavizadores sobre el

Figura 3.5: Densidades espectrales correspondientes a los 13 procesos ARMA(p,q) sujetos a discriminación.



comportamiento del estadístico discriminante. Por procesos, las probabilidades de error en cada clase oscilaron entre un 6% y un 56% para el estimador periodograma suavizado; entre un 6% y un 61% para el estimador log-periodograma suavizado por mínimos cuadrados y entre un 3% y un 53% para el log-periodograma suavizado mediante máxima verosimilitud. En todos los casos, las mayores tasas de error correspondieron a las realizaciones de los procesos 3, 4, 7, 9, 10 y 13, oscilando éstas entre el 31% y el 61%. Las tasas de mala clasificación más altas se observaron entre procesos que compartían una misma estructura de dependencia, con parámetros muy similares (como los procesos 4 y 7 o las series correspondientes a las clases 10 y 13), o entre procesos con distinto modelo generador pero una estructura de autocorrelación muy similar (procesos 3 y 9 o 4 y 10).

Figura 3.6: Realizaciones de longitud $n=400$ correspondientes a los 13 procesos $ARMA(p,q)$ sujetos a discriminación.

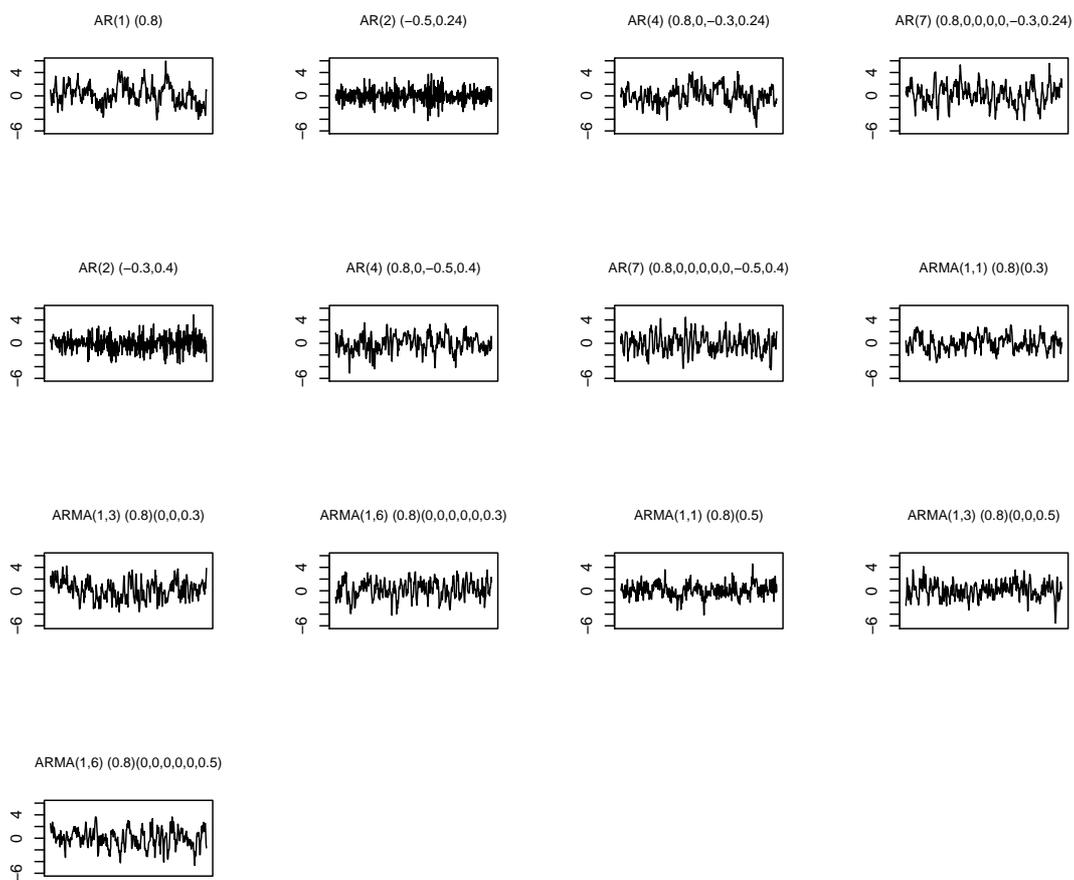


Tabla 3.9: Resultados obtenidos al clasificar series de longitud $n=200$ generadas del modelo $(1 - 0,8L)(1 - \phi_s L^s)X_t = (1 - \theta_r L^r)\varepsilon_t$. Los porcentajes muestran la proporción de series generadas del proceso de cada columna que fueron clasificadas como realizaciones de la clase de procesos de la fila correspondiente.

		Proceso							
		1	2	3	4	5	6	7	
		$\phi_s = 0$ $\theta_r = 0$ $s = 0$	$\phi_s = -0,3$ $\theta_r = 0$ $s = 1$	$s = 3$	$s = 6$	$\phi_s = -0,5$ $\theta_r = 0$ $s = 1$	$s = 3$	$s = 6$	
1	$\phi_s = 0$	\hat{f}_{DLS}	94%	0%	4%	8%	0%	0%	1%
	$\theta_r = 0$	\hat{f}_{LS}	94%	0%	9%	14%	0%	0%	8%
	$s = 0$	\hat{f}_{LK}	97%	0%	4%	9%	0%	0%	0%
2	$\phi_s = -0,3$	\hat{f}_{DLS}	0%	91%	0%	0%	7%	0%	0%
	$\theta_r = 0$	\hat{f}_{LS}	0%	83%	0%	0%	15%	0%	0%
	$s = 1$	\hat{f}_{LK}	0%	91%	0%	0%	0%	0%	0%
3	$\phi_s = -0,3$	\hat{f}_{DLS}	1%	0%	57%	0%	0%	8%	0%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	44%	0%	0%	21%	0%
	$s = 3$	\hat{f}_{LK}	0%	0%	47%	0%	0%	14%	0%
4	$\phi_s = -0,3$	\hat{f}_{DLS}	1%	0%	0%	44%	0%	0%	18%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	0%	39%	0%	0%	27%
	$s = 6$	\hat{f}_{LK}	0%	0%	0%	53%	0%	0%	35%
5	$\phi_s = -0,5$	\hat{f}_{DLS}	0%	9%	0%	0%	93%	0%	0%
	$\theta_r = 0$	\hat{f}_{LS}	0%	17%	0%	0%	85%	0%	0%
	$s = 1$	\hat{f}_{LK}	0%	9%	0%	0%	93%	0%	0%
6	$\phi_s = -0,5$	\hat{f}_{DLS}	0%	0%	3%	0%	0%	83%	0%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	1%	0%	0%	66%	0%
	$s = 3$	\hat{f}_{LK}	0%	0%	2%	0%	0%	73%	0%
7	$\phi_s = -0,5$	\hat{f}_{DLS}	0%	0%	0%	4%	0%	0%	60%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	0%	5%	0%	0%	46%
	$r = 6$	\hat{f}_{LK}	0%	0%	0%	4%	0%	0%	52%
8	$\phi_s = 0$	\hat{f}_{DLS}	2%	0%	0%	1%	0%	0%	1%
	$\theta_r = -0,3$	\hat{f}_{LS}	2%	0%	2%	1%	0%	0%	0%
	$r = 1$	\hat{f}_{LK}	0%	0%	0%	0%	0%	0%	0%
9	$\phi_s = 0$	\hat{f}_{DLS}	2%	0%	35%	2%	0%	3%	1%
	$\theta_r = -0,3$	\hat{f}_{LS}	3%	0%	38%	1%	0%	7%	1%
	$r = 3$	\hat{f}_{LK}	3%	0%	45%	0%	0%	4%	0%
10	$\phi_s = 0$	\hat{f}_{DLS}	0%	0%	0%	41%	0%	0%	14%
	$\theta_r = -0,3$	\hat{f}_{LS}	1%	0%	1%	40%	0%	0%	17%
	$r = 6$	\hat{f}_{LK}	0%	0%	0%	33%	0%	0%	9%
11	$\phi_s = 0$	\hat{f}_{DLS}	0%	0%	0%	0%	0%	0%	0%
	$\theta_r = -0,5$	\hat{f}_{LS}	0%	0%	0%	0%	0%	0%	0%
	$r = 1$	\hat{f}_{LK}	0%	0%	0%	0%	0%	0%	0%
12	$\phi_s = 0$	\hat{f}_{DLS}	0%	0%	1%	0%	0%	6%	0%
	$\theta_r = -0,5$	\hat{f}_{LS}	0%	0%	5%	0%	0%	6%	0%
	$r = 3$	\hat{f}_{LK}	0%	0%	2%	0%	0%	9%	0%
13	$\phi_s = 0$	\hat{f}_{DLS}	0%	0%	0%	0%	0%	0%	5%
	$\theta_r = -0,5$	\hat{f}_{LS}	0%	0%	0%	0%	0%	0%	9%
	$r = 6$	\hat{f}_{LK}	0%	0%	0%	1%	0%	0%	4%
Tasa global de error	\hat{f}_{DLS}	6%	9%	43%	56%	7%	17%	40%	
	\hat{f}_{LS}	6%	17%	56%	61%	15%	34%	54%	
	\hat{f}_{LK}	3%	9%	53%	47%	7%	27%	48%	

Tabla 3.10: Resultados obtenidos al clasificar series de longitud $n=200$ generadas del modelo $(1 - 0,8L)(1 - \phi_s L^s)X_t = (1 - \theta_r L^r)\varepsilon_t$. Los porcentajes muestran la proporción de series generadas del proceso de cada columna que fueron clasificadas como realizaciones de la clase de procesos de la fila correspondiente.

Proceso		Proceso						
		$\phi_s = 0$ $\theta_r = -0,3$			$\phi_s = 0$ $\theta_r = -0,5$			
		$r = 1$	$r = 3$	$r = 6$	$r = 1$	$r = 3$	$r = 6$	
1	$\phi_s = 0$	\hat{f}_{DLS}	5%	5%	6%	0%	0%	2%
	$\theta_r = 0$	\hat{f}_{LS}	8%	9%	12%	0%	0%	0%
	$s = 0$	\hat{f}_{LK}	2%	5%	9%	0%	0%	0%
2	$\phi_s = -0,3$	\hat{f}_{DLS}	0%	0%	0%	0%	0%	0%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	0%	0%	0%	0%
	$s = 1$	\hat{f}_{LK}	0%	0%	0%	0%	0%	0%
3	$\phi_s = -0,3$	\hat{f}_{DLS}	0%	23%	0%	0%	1%	0%
	$\theta_r = 0$	\hat{f}_{LS}	1%	16%	0%	0%	3%	0%
	$s = 3$	\hat{f}_{LK}	0%	19%	0%	0%	1%	0%
4	$\phi_s = -0,3$	\hat{f}_{DLS}	0%	0%	18%	0%	0%	1%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	18%	0%	0%	3%
	$s = 6$	\hat{f}_{LK}	0%	0%	21%	0%	0%	10%
5	$\phi_s = -0,5$	\hat{f}_{DLS}	0%	0%	0%	0%	0%	0%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	0%	0%	0%	0%
	$s = 1$	\hat{f}_{LK}	0%	0%	0%	0%	0%	0%
6	$\phi_s = -0,5$	\hat{f}_{DLS}	0%	0%	0%	0%	2%	0%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	0%	0%	5%	0%
	$s = 3$	\hat{f}_{LK}	0%	0%	0%	0%	3%	0%
7	$\phi_s = -0,5$	\hat{f}_{DLS}	0%	0%	1%	0%	0%	2%
	$\theta_r = 0$	\hat{f}_{LS}	0%	0%	1%	0%	0%	3%
	$r = 6$	\hat{f}_{LK}	0%	0%	0%	0%	0%	2%
8	$\phi_s = 0$	\hat{f}_{DLS}	86%	1%	2%	7%	0%	1%
	$\theta_r = -0,3$	\hat{f}_{LS}	74%	2%	4%	7%	0%	0%
	$r = 1$	\hat{f}_{LK}	87%	1%	2%	5%	0%	0%
9	$\phi_s = 0$	\hat{f}_{DLS}	0%	66%	1%	0%	17%	0%
	$\theta_r = -0,3$	\hat{f}_{LS}	0%	65%	2%	0%	22%	0%
	$r = 3$	\hat{f}_{LK}	0%	69%	1%	0%	20%	0%
10	$\phi_s = 0$	\hat{f}_{DLS}	0%	0%	69%	0%	0%	32%
	$\theta_r = -0,3$	\hat{f}_{LS}	0%	0%	61%	0%	0%	48%
	$r = 6$	\hat{f}_{LK}	0%	0%	64%	0%	0%	38%
11	$\phi_s = 0$	\hat{f}_{DLS}	9%	0%	0%	93%	0%	0%
	$\theta_r = -0,5$	\hat{f}_{LS}	17%	0%	0%	93%	0%	0%
	$r = 1$	\hat{f}_{LK}	11%	0%	0%	95%	0%	0%
12	$\phi_s = 0$	\hat{f}_{DLS}	0%	5%	0%	0%	80%	0%
	$\theta_r = -0,5$	\hat{f}_{LS}	0%	8%	0%	0%	70%	0%
	$r = 3$	\hat{f}_{LK}	0%	6%	0%	0%	76%	0%
13	$\phi_s = 0$	\hat{f}_{DLS}	0%	0%	3%	0%	0%	62%
	$\theta_r = -0,5$	\hat{f}_{LS}	0%	0%	2%	0%	0%	46%
	$r = 6$	\hat{f}_{LK}	0%	0%	3%	0%	0%	50%
Tasa global de error		\hat{f}_{DLS}	14%	34%	31%	7%	20%	38%
		\hat{f}_{LS}	26%	35%	39%	7%	30%	54%
		\hat{f}_{LK}	13%	31%	36%	5%	24%	50%

A continuación, se repitió el proceso de simulación con series de mayor longitud. La Tabla 3.11 muestra los resultados obtenidos utilizando el mismo esquema de simulación y series de longitud $n=400$. La probabilidad global de mala clasificación resultó ahora ser de un 14,02 % para el estimador periodograma suavizado, de un 20,92 % para el estimador log-periodograma suavizado y de un 15 % para el estimador log-periodograma suavizado mediante máxima verosimilitud, observándose de nuevo la influencia de la longitud de las series y el estimador espectral empleado sobre los resultados de la clasificación. Por procesos, las probabilidades de error en cada clase oscilaron entre un 0 % y un 43 % para el estimador periodograma suavizado; entre un 1 % y un 52 % para el estimador log-periodograma suavizado por mínimos cuadrados y entre un 0 % y un 35 % para el log-periodograma suavizado mediante máxima verosimilitud. En todos los casos, las mayores tasas de error correspondieron a las realizaciones de los procesos 3, 4, 7, 10 y 13, oscilando éstas entre el 23 % y el 52 %

Tabla 3.11: Tasas de error obtenidas al clasificar series de longitud $n=400$ generadas del modelo $(1 - 0,8L)(1 - \phi_s L^s)X_t = (1 - \theta_r L^r)\varepsilon_t$. Los porcentajes indican la proporción de series mal clasificadas dentro de cada clase de procesos según el estimador espectral utilizado.

Nº proceso		\widehat{f}_{DLS}	\widehat{f}_{LS}	\widehat{f}_{LK}
1	$\phi_s = 0; \theta_r = 0; s = 0$	0 %	6 %	0 %
2	$\phi_s = -0,3; \theta_r = 0; s = 1$	0 %	1 %	1 %
3	$\phi_s = -0,3; \theta_r = 0; s = 3$	32 %	37 %	35 %
4	$\phi_s = -0,3; \theta_r = 0; s = 6$	43 %	52 %	26 %
5	$\phi_s = -0,5; \theta_r = 0; s = 1$	4 %	4 %	4 %
6	$\phi_s = -0,5; \theta_r = 0; s = 3$	9 %	17 %	12 %
7	$\phi_s = -0,5; \theta_r = 0; s = 6$	23 %	45 %	33 %
8	$\phi_s = 0; \theta_r = -0,3; r = 1$	4 %	13 %	4 %
9	$\phi_s = 0; \theta_r = -0,3; r = 3$	11 %	14 %	12 %
10	$\phi_s = 0; \theta_r = -0,3; r = 6$	25 %	26 %	23 %
11	$\phi_s = 0; \theta_r = -0,5; r = 1$	3 %	3 %	2 %
12	$\phi_s = 0; \theta_r = -0,5; r = 3$	8 %	12 %	8 %
13	$\phi_s = 0; \theta_r = -0,5; r = 6$	32 %	42 %	35 %

Los resultados obtenidos muestran tasas de mala clasificación elevadas que evidencian la dificultad del problema de discriminación que ha sido propuesto. Tal y como se muestra en la Figura 3.5, la densidad espectral de los 13 procesos considerados resulta en algunos casos muy similar. Si se considera como índice de separabilidad entre dos poblaciones la

tasa de J-divergencia puede comprobarse que, desde este punto de vista, algunos pares de procesos son prácticamente idénticos (Tabla 3.12). Así, la separabilidad entre ellos oscila entre un valor de 0,008 (correspondiente a la distancia entre los procesos 3 y 9 y entre los procesos 4 y 10) y 5,840 (correspondiente a la separabilidad entre los procesos 2 y 13). De hecho, al trabajar con series de longitud $n=200$ y el estimador espectral \hat{f}_{DLS} , el 35 % de las series correspondientes al tercer proceso fueron clasificadas en el grupo 9. Lo mismo sucedió con el 41 % de las series del proceso 4, que fueron asignadas al grupo 10. Similares resultados se obtuvieron con series de mayor longitud y cualquiera de los otros estimadores espectrales.

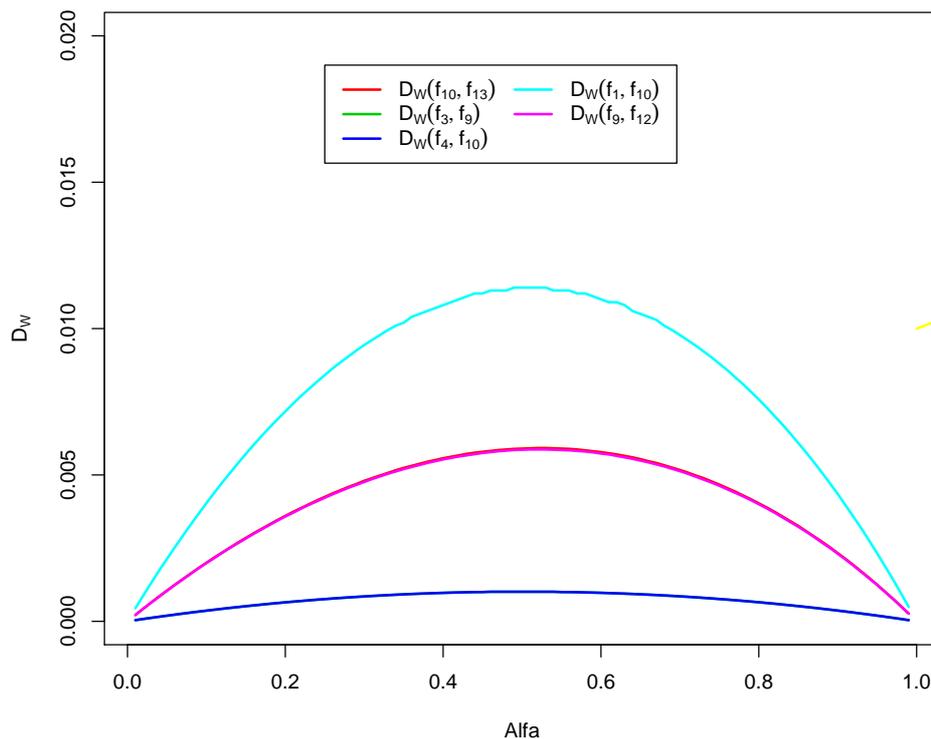
Tabla 3.12: Separabilidad entre los distintos pares de procesos ARMA(p,q) considerados, evaluada mediante la tasa de J-divergencia.

	2	3	4	5	6	7	8	9	10	11	12	13
1	3.761	0.095	0.094	2.377	0.293	0.292	0.094	0.094	0.094	0.290	0.290	0.292
2		2.744	4.331	0.048	2.573	5.249	1.913	2.762	4.442	1.173	2.596	5.840
3			0.229	1.762	0.049	0.466	0.188	0.008	0.234	0.360	0.072	0.488
4				2.876	0.518	0.049	0.199	0.167	0.008	0.408	0.313	0.073
5					1.730	3.600	1.173	1.786	2.958	0.720	1.749	4.037
6						0.840	0.395	0.074	0.539	0.566	0.065	0.938
7							0.415	0.353	0.073	0.659	0.476	0.065
8								0.189	0.199	0.048	0.398	0.418
9									0.170	0.363	0.048	0.368
10										0.410	0.316	0.049
11											0.583	0.663
12												0.507

Adicionalmente, cabe esperar que la elección del parámetro α en (3.1) pueda tener una gran repercusión sobre los resultados obtenidos. En este caso, la elección del parámetro óptimo obligaría a computar, para distintos valores de α , la distancia $D_W(f_i, f_j)$ entre las densidades espectrales de todos los posibles pares que se pudiesen formar con los 13 modelos, y seleccionar el valor de α que maximizase la distancia entre los grupos entre los que existiese una menor separabilidad. En una situación como esta, en la que se discrimina entre un número más o menos grande de procesos, puede ser extremadamente difícil determinar el valor del parámetro α óptimo, pues aquel que garantiza una mayor separabilidad entre dos clases de procesos puede comprometer la discriminación entre un par de procesos diferentes. Por esta razón, se procedió a evaluar a posteriori la idoneidad de la elección $\alpha = 0,5$ para afrontar este problema de discriminación en concreto. Para ello, se calculó la distancia, $D_W(f_i, f_j)$, para diferentes valores de α , entre aquellos pares de procesos en los que la probabilidad de mala clasificación había resultado más alta con cualquiera de los tres estimadores espectrales (Figura 3.7). En cualquier caso, no parece que el tomar $\alpha = 0,5$ haya resultado una mala elección, pues es en torno a este valor

donde se maximiza la distancia entre cada uno de estos pares de procesos. Nótese que, de todos modos, nos movemos en distancias de 0.001 a 0.012, por lo que resulta tan difícil discriminar entre estas series.

Figura 3.7: Distancia $D_W(f_i, f_j)$ entre algunos pares de los procesos ARMA(p,q) sujetos a discriminación, según diferentes valores del parámetro α en (3.1).



3.1.3. Discriminación de procesos con distribución marginal no gaussiana.

En este apartado se estudiaron las propiedades del estadístico discriminante propuesto en la discriminación de procesos no gaussianos. Para ello, se clasificaron realizaciones de procesos con distribución marginal no normal, manteniendo la estructura de dependencia de un proceso autorregresivo AR(1):

- Procesos con distribución marginal exponencial
- Procesos con distribución marginal doble exponencial

o con una estructura de dependencia distinta:

- Series m-dependientes con distribución marginal gamma
- Series m-dependientes con distribución marginal lognormal

Además de los procesos anteriores, se procedió también a discriminar entre procesos autorregresivos AR(1) con distribución marginal uniforme $U(-A, A)$, variando la distribución marginal según diferentes valores del parámetro A . En todos los casos el procedimiento propuesto logró clasificar correctamente el 100 % de las series generadas, por lo que no se mostrarán aquí los resultados obtenidos en dicha simulación.

a) Discriminación de procesos AR(1) con distribución marginal exponencial de parámetro λ .

En este apartado, se pretende evaluar la habilidad del estadístico propuesto para discriminar entre realizaciones de procesos autorregresivos AR(1) con distribución marginal exponencial. Se considera así, bajo Π_j , el siguiente modelo:

$$X_t = \phi_j X_{t-1} + \varepsilon_t, \quad |\phi_j| < 1, \quad (3.7)$$

donde $\{\varepsilon_t\}$ es una secuencia de variables aleatorias i.i.d. según

$$\varepsilon_t = \begin{cases} 0 & \text{con probabilidad } \phi_j \\ Y_t & \text{con probabilidad } 1 - \phi_j \end{cases}$$

siendo $\{Y_t\}$ variables aleatorias i.i.d. con distribución exponencial de parámetro λ_j , $j = 1, 2$.

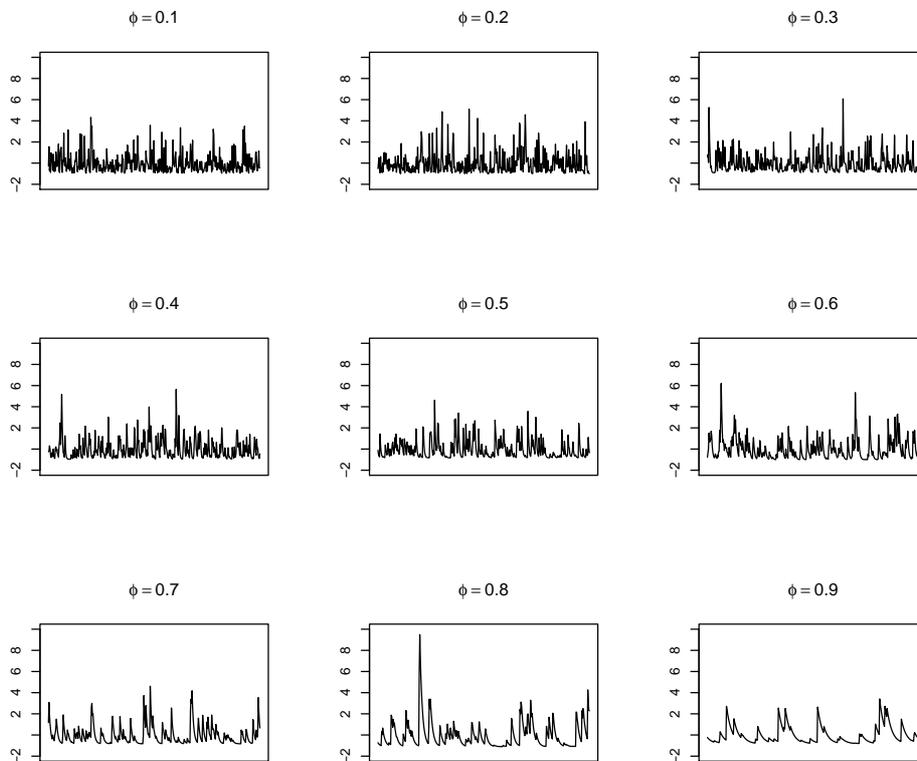
Se generaron series de longitud $n = 400$ a partir de Π_j , ($j = 1, 2$), de acuerdo con la ecuación (3.7), con valores del parámetro $\phi_1 = 0, 1, 0, 2, \dots, 0, 6$, $\phi_2 = \phi_1 + 0, 1 * k$, para $k = 1, 2, 3$, y tomando siempre $\lambda_1 = \lambda_2 = 1$, de forma que la distribución marginal correspondiese a la de una variable aleatoria exponencial estándar. En la Figura 3.8 se muestran realizaciones aleatorias de cada uno de los procesos utilizados en la discriminación, una vez desestacionalizadas en media.

La densidad espectral de cada uno de estos procesos vendrá dada por la ecuación:

$$f(\lambda) = \frac{1}{2\pi\lambda_j^2} \frac{1 - \phi_j^2}{1 - 2\phi_j \cos(\lambda) + \phi_j^2}, \quad \lambda \in (-\pi, \pi),$$

según se muestra en la Figura 3.9. Tal y como puede apreciarse, a medida que aumenta el valor del parámetro autorregresivo ϕ las frecuencias más bajas cercanas a cero presentan

Figura 3.8: Realizaciones de procesos autorregresivos AR(1) con distribución marginal exponencial estándar, según diferentes valores del parámetro ϕ en (3.7). Series desestacionalizadas en media.

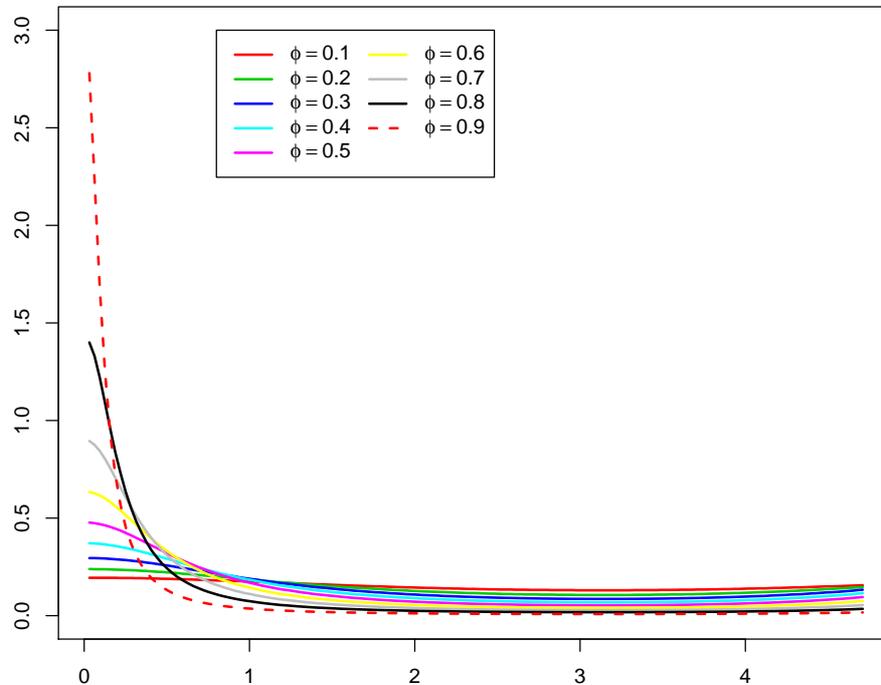


una potencia mayor, mientras que las frecuencias más altas tienden a mostrar potencias más bajas. Esta situación es típica en los casos de procesos autorregresivos AR(1), independiente de su distribución marginal, correspondiéndose con los valores de la función de autocorrelación en este tipo de procesos.

Cada una de las series generadas se clasificó como perteneciente a una de las dos poblaciones de acuerdo con el procedimiento descrito en la Sección 3.1.1. En este caso se generaron $s = 200$ series de cada clase, calculándose el porcentaje de series mal clasificadas para cada una de las dos poblaciones. Finalmente, ambas tasas de error se promediaron para obtener la tasa global de mala clasificación con cualquiera de los tres estimadores utilizados. Los resultados obtenidos se muestran en la Figura 3.10.

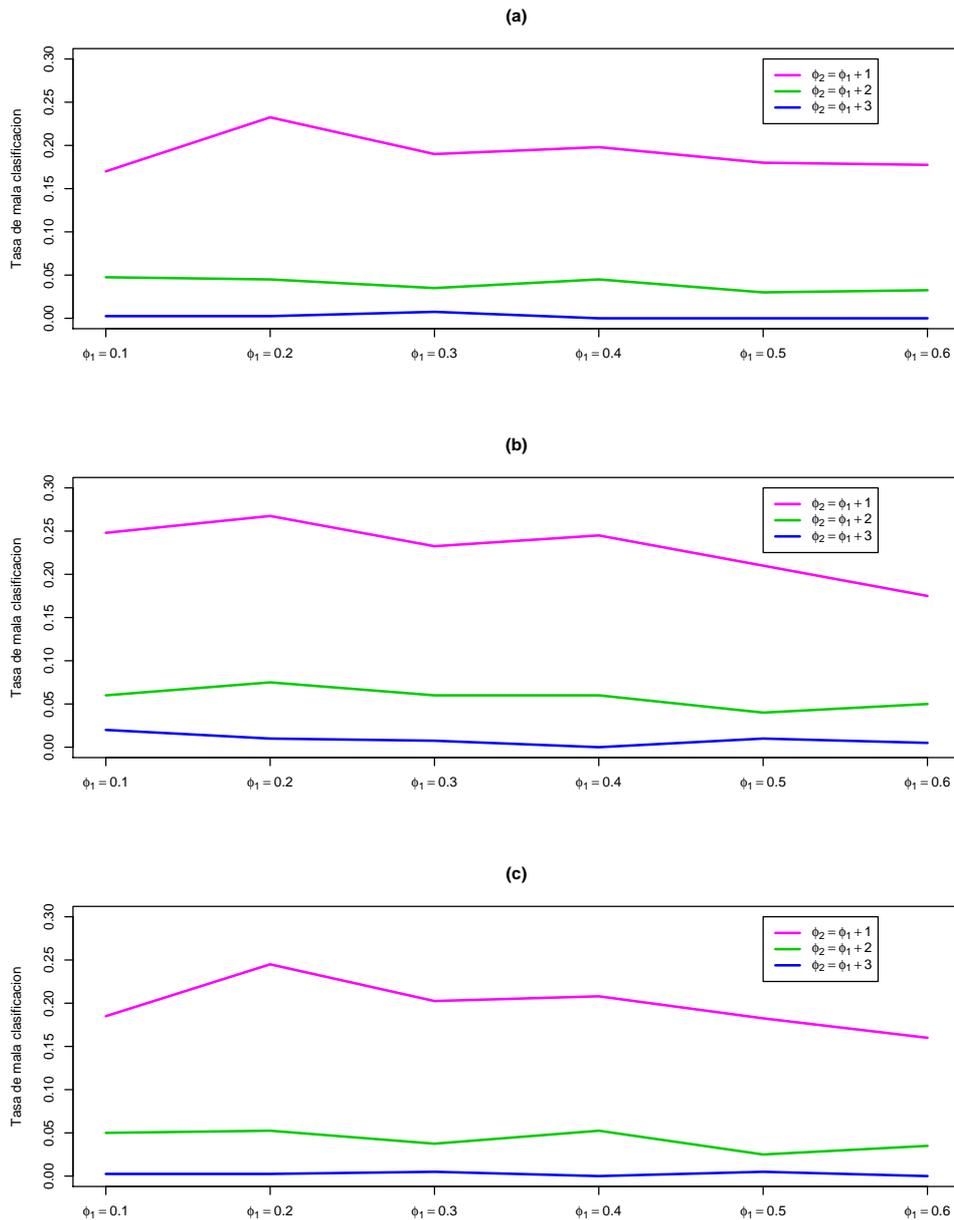
Según se puede deducir, el procedimiento propuesto presenta, a pesar de la ausencia de resultados teóricos en este sentido, un buen comportamiento a la hora de discriminar entre procesos de distribución marginal exponencial, con tasas de mala clasificación similares a

Figura 3.9: Densidad espectral correspondiente a procesos autorregresivos AR(1) con distribución marginal exponencial estándar según los valores del parámetro ϕ en (3.7).



las ya obtenidas en el caso de discriminar entre procesos autorregresivos con errores de distribución normal. De nuevo se aprecia, como era de esperar, una importante mejoría en los resultados a medida que la distancia entre los parámetros ϕ_1 y ϕ_2 que definen los grupos de estudio aumenta, ya que también en este caso esa distancia determina la separabilidad entre las clases Π_1 y Π_2 . Según se muestra en la Tabla 3.13 la separabilidad, en términos de la tasa de J-divergencia, es prácticamente nula entre clases de procesos definidas por parámetros autorregresivos muy similares. Así, las mayores tasas de mala clasificación se sitúan en torno al 25 %, detectándose siempre en aquellas situaciones en las que ϕ_1 y ϕ_2 toman valores muy próximos ($k=1$). Estas probabilidades de error se reducen sustancialmente a medida que aumenta la separabilidad entre ambas poblaciones, situándose por debajo del 5 %, de modo que para alguna de las combinaciones todas las series llegan a clasificarse con éxito en la categoría de procedencia. Si comparamos la separabilidad entre las clases de procesos consideradas con la que se obtenía para procesos AR(1) con errores gaussianos, expresada en términos de la tasa de J-divergencia, observamos valores

Figura 3.10: Tasas de error para discriminar entre procesos AR(1) con distribución marginal exponencial estándar con diferentes parámetros ϕ_1 y ϕ_2 . Se clasificaron 200 series de cada clase con longitud $n=400$. Resultados con: (a) Periodograma suavizado \hat{f}_{DLS} , (b) Log-periodograma suavizado por mínimos cuadrados \hat{f}_{LS} , (c) Log-periodograma suavizado por máxima verosimilitud \hat{f}_{LK} .



semejantes que justifican la similitud en las tasas de error obtenidas.

Tabla 3.13: Separabilidad entre procesos AR(1) con distribución marginal exponencial estándar y diferentes parámetros autorregresivos ϕ_1 , ϕ_2 , evaluada mediante la tasa de J-divergencia.

	ϕ_2							
$\phi_1 \downarrow$	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,1	0,011	0,046	0,113	0,226	0,418	0,763	1,485	3,709
0,2		0,012	0,054	0,137	0,292	0,582	1,208	3,170
0,3			0,015	0,067	0,182	0,417	0,946	2,644
0,4				0,019	0,092	0,269	0,698	2,130
0,5					0,027	0,141	0,467	1,628
0,6						0,044	0,257	1,140
0,7							0,085	0,673
0,8								0,251

Por otro lado, al comparar los resultados según el estimador espectral utilizado, y aunque los resultados no muestran diferencias importantes, de nuevo se observa que el estimador \hat{f}_{LS} proporciona tasas de error ligeramente más elevadas que cualquiera de los otros dos estimadores, \hat{f}_{DLS} o \hat{f}_{LK} . Es con este procedimiento con el que se obtienen las tasas de error más altas, por encima del 25 % en el caso de poca separabilidad entre los grupos, pero que en todo caso caen en torno al 5 % a medida que las categorías son diferentes.

Siguiendo en esta misma línea, se procedió también a estudiar el comportamiento del estadístico D_W para discriminar entre procesos autorregresivos AR(1) con diferente distribución marginal. Así, se realizó un ejercicio de simulación con la finalidad de discriminar entre la clase de procesos AR(1) con distribución marginal exponencial estándar de parámetro $\phi_1=0.5$ y la clase de procesos AR(1) con errores gaussianos $N(0,1)$ y parámetro autorregresivo $\phi_2=0.5$. De nuevo, se generaron $s = 100$ series de longitud $n = 400$ de cada una de las clases de procesos consideradas, que fueron clasificadas de acuerdo con el estadístico D_W según el procedimiento experimental descrito en el Apartado 3.1.1. Las tasas de error obtenidas fueron de un 9.5 % para el caso de trabajar con el estimador periodograma suavizado, \hat{f}_{DLS} , de un 13.5 % para el estimador log-periodograma suavizado por mínimos cuadrados, \hat{f}_{LS} , y de un 9.0 % para el estimador espectral basado en criterios de máxima verosimilitud, \hat{f}_{LK} . Dichos resultados reflejan tasas de mala clasificación bajas, si se tiene en cuenta además que la separabilidad de ambas poblaciones, en términos de la tasa de J-divergencia, es sólo de 0.042. Este tipo de procesos resultan tanto más similares cuanto menor sea el valor del parámetro autorregresivo que los define. Así, se

repitió el experimento para procesos autorregresivos de distribución marginal normal y exponencial de parámetro 0.3, lo que conduce a una separabilidad entre las clases de sólo 0.004. Las tasas de mala clasificación obtenidas fueron en ese caso de un 29,5 % para el estimador periodograma suavizado y de un 32 % para el estimador log-periodograma suavizado por mínimos cuadrados y para el estimador basado en máxima verosimilitud. Se procedió igualmente a discriminar entre procesos AR(1) de distribución marginal normal y exponencial con parámetro autorregresivo $\Phi=0.7$, con una separabilidad ahora de 0.235. Las tasas de mala clasificación obtenidas fueron de un 3 % para los estimadores espectrales \hat{f}_{DLS} y \hat{f}_{LS} y de un 2 % para el estimador \hat{f}_{LK} .

b) Discriminación de procesos AR(1) con distribución marginal doble exponencial de parámetro λ .

Siguiendo el esquema anterior, se procedió a utilizar el estadístico propuesto para la clasificación de procesos autorregresivos AR(1) con distribución marginal doble exponencial. Se consideró ahora bajo Π_j , el siguiente modelo:

$$X_t = \phi_j X_{t-1} + \varepsilon_t, \quad |\phi_j| < 1, \tag{3.8}$$

donde $\{\varepsilon_t\}$ es una secuencia de variables aleatorias i.i.d. según

$$\varepsilon_t = \begin{cases} 0 & \text{con probabilidad } \phi_j^2 \\ Y_t & \text{con probabilidad } 1 - \phi_j^2 \end{cases}$$

siendo $\{Y_t\}$ variables aleatorias i.i.d. con distribución doble exponencial de parámetro λ_j , $j = 1, 2$. Esto es:

$$Y_t = \begin{cases} -\frac{1}{\lambda_j} \ln(2U) & \text{si } U < 0,5 \\ \frac{1}{\lambda_j} \ln(2(1-U)) & \text{si } U \geq 0,5 \end{cases}$$

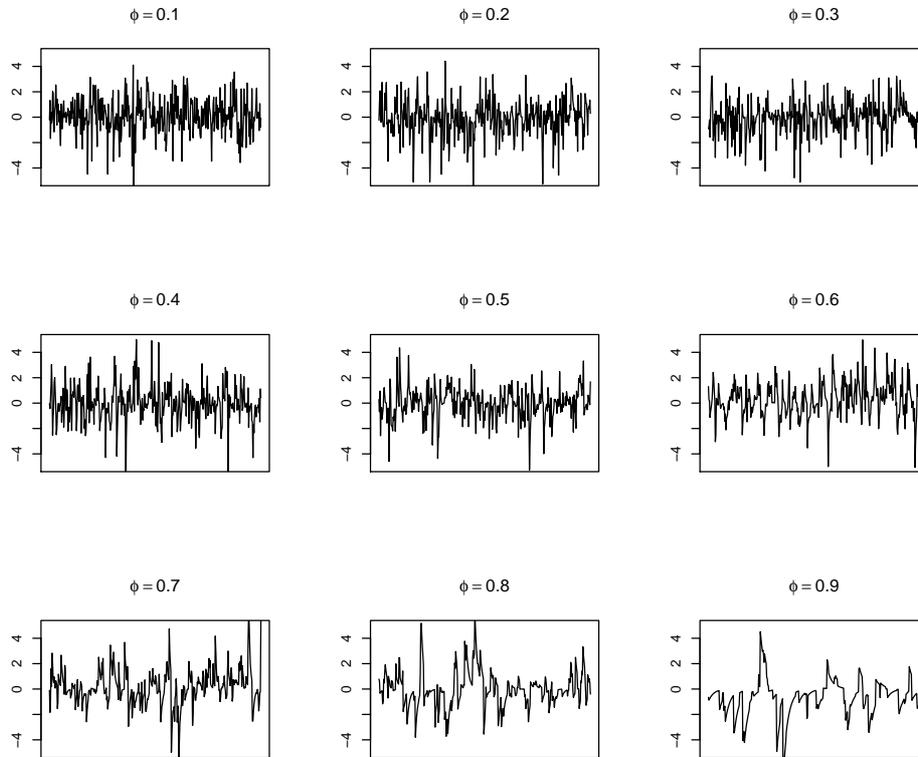
con U una variable aleatoria con distribución uniforme $U(0, 1)$.

Igual que en el apartado anterior, se generaron $s = 200$ series de longitud $n = 400$ a partir de Π_j , ($j = 1, 2$), de acuerdo con la ecuación (3.8), con valores del parámetro $\phi_1 = 0,1, 0,2, 0,3, \dots, 0,6$, $\phi_2 = \phi_1 + 0,1 * k$, para $k = 1, 2, 3$, y tomando siempre $\lambda_1 = \lambda_2 = 1$. En la Figura 3.11 se muestran realizaciones aleatorias de cada uno de los procesos utilizados en la discriminación.

La densidad espectral de cada una de estos procesos, al igual que en el caso anterior, vendrá dada por la ecuación:

$$f(\lambda) = \frac{1}{2\pi\lambda^2} \frac{1 - \phi_j^2}{1 - 2\phi_j \cos(\lambda) + \phi_j^2}, \quad \lambda \in (-\pi, \pi), \tag{3.9}$$

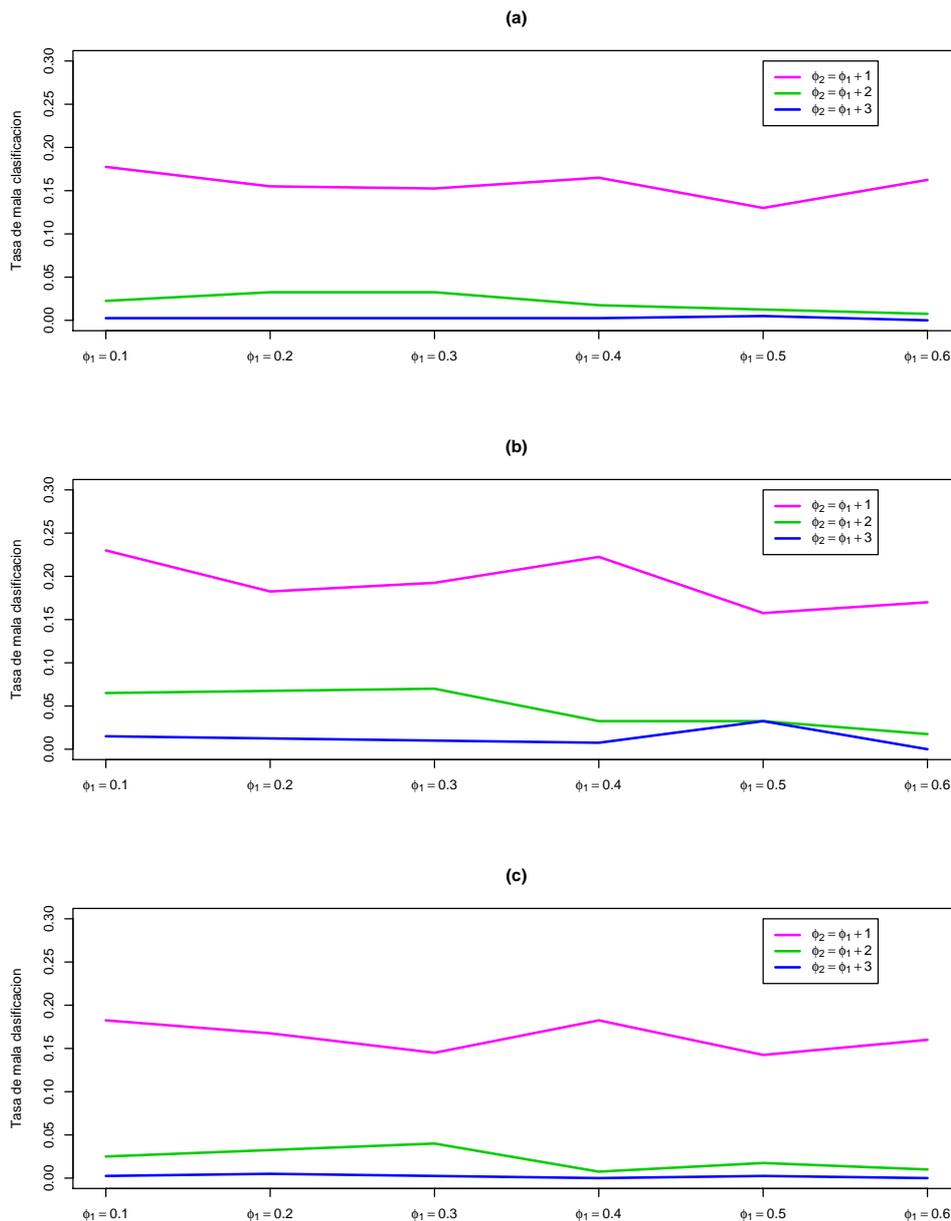
Figura 3.11: Realizaciones de procesos autorregresivos AR(1) con distribución marginal doble exponencial de parámetro $\lambda = 1$, según diferentes valores del parámetro ϕ en (3.8).



observándose la estructura de dependencia típica de este tipo de procesos.

Cada una de las series generadas se clasificó como perteneciente a una de las dos poblaciones de acuerdo con el estadístico discriminante D_W , utilizando cada uno de los tres estimadores espectrales, con los mismos parámetros que en las simulaciones anteriores. Se clasificaron $s = 200$ series de cada una de las poblaciones a discriminar y se calculó la tasa global de mala clasificación. Los resultados obtenidos se muestran en la Figura 3.12, obteniéndose conclusiones análogas que en el caso anterior: tasas de error en torno al 20% que disminuyen a medida que aumenta la diferencia entre los parámetros autorregresivos de las series que se comparan, y probabilidades de mala clasificación ligeramente peores en caso de que la estimación de la densidad espectral se realice mediante el estimador \hat{f}_{LS} .

Figura 3.12: Tasas de error para discriminar entre procesos AR(1) de parámetros ϕ_1 y ϕ_2 con distribución marginal doble exponencial estándar. Se clasificaron 200 series de cada clase con longitud $n=400$. Resultados con: (a) Periodograma suavizado \hat{f}_{DLS} , (b) Log-periodograma suavizado por mínimos cuadrados \hat{f}_{LS} , (c) Log-periodograma suavizado por máxima verosimilitud \hat{f}_{LK}



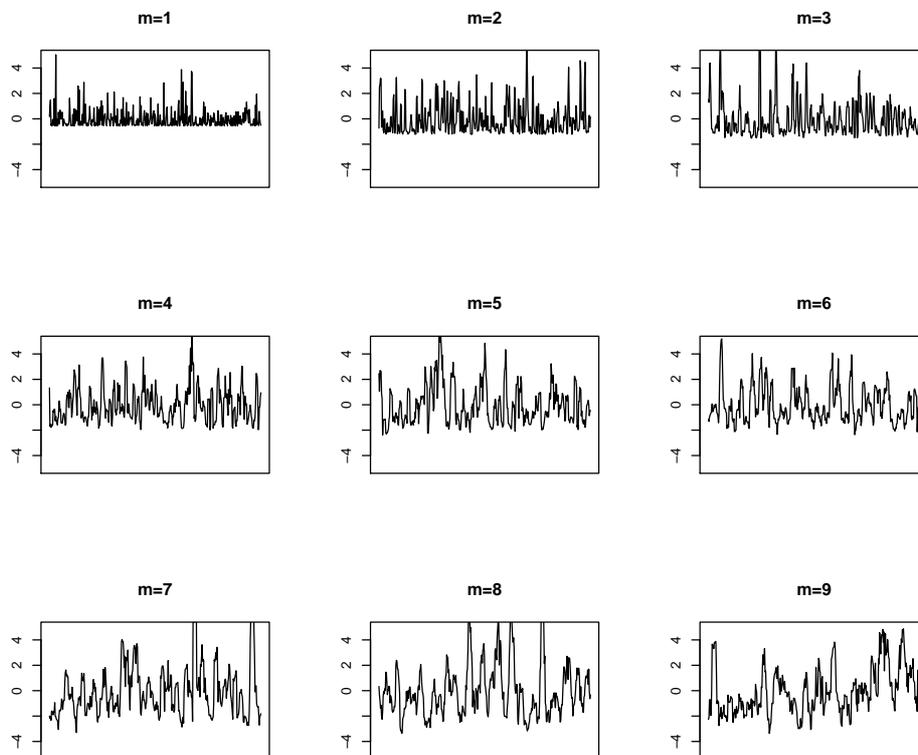
c) Discriminación de series m -dependientes con distribución marginal gamma

Se consideró ahora bajo Π_j , el siguiente modelo:

$$X_t = \frac{1}{2} \sum_{i=t}^{t+m_j-1} Y_i^2 \quad (3.10)$$

donde $\{Y_1, Y_2, \dots, Y_{m_j+n-1}\}$ es una secuencia de variables aleatorias i.i.d. con distribución normal estándar $N(0, 1)$. De esta forma, en cada instante, X_t es igual a la suma de m_j variables aleatorias con distribución gaussiana. Las series así generadas son m -dependientes con distribución marginal gamma $\Gamma(1, \frac{m}{2})$.

Figura 3.13: Realizaciones de procesos m -dependientes con distribución marginal gamma. Series desestacionalizadas en media.



Del mismo modo que en los casos anteriores, se generaron series de longitud $n = 400$ a partir de Π_j , ($j = 1, 2$), de acuerdo con la ecuación (3.10), con valores del parámetro $m_1 = 1, 2, 3, \dots, 7$, $m_2 = m_1 + k$, para $k = 1, 2, 3$. En la Figura 3.13 se muestra una realización aleatoria de cada uno de los procesos utilizados en la discriminación, una vez desestacionalizadas en media.

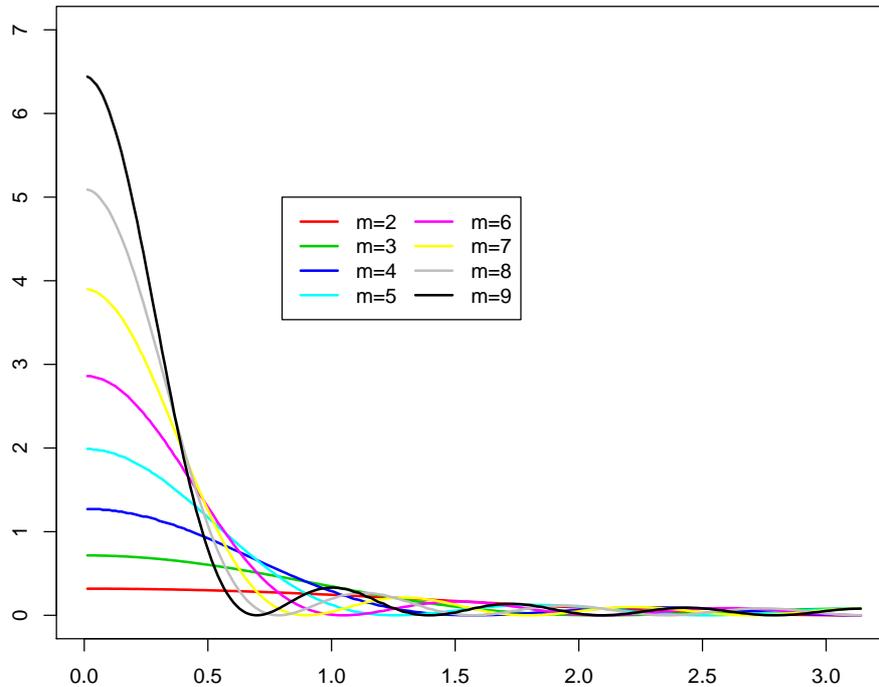
En este caso, la función de densidad espectral correspondiente al proceso (3.10) vendrá dada por la ecuación:

$$f(\lambda) = \frac{1}{2\pi} \left\{ \frac{1}{2}m + \sum_{h=1}^{m-1} (m-h)\cos(\lambda h) \right\} \quad (3.11)$$

según se muestra en la Figura 3.14. Dicho espectro se corresponde a una función de autocovarianzas determinada por la estructura:

$$\gamma(h) = \begin{cases} \frac{1}{2}(m-h) & \text{si } 0 \leq h < m \\ 0 & \text{si } h \geq m \end{cases} \quad (3.12)$$

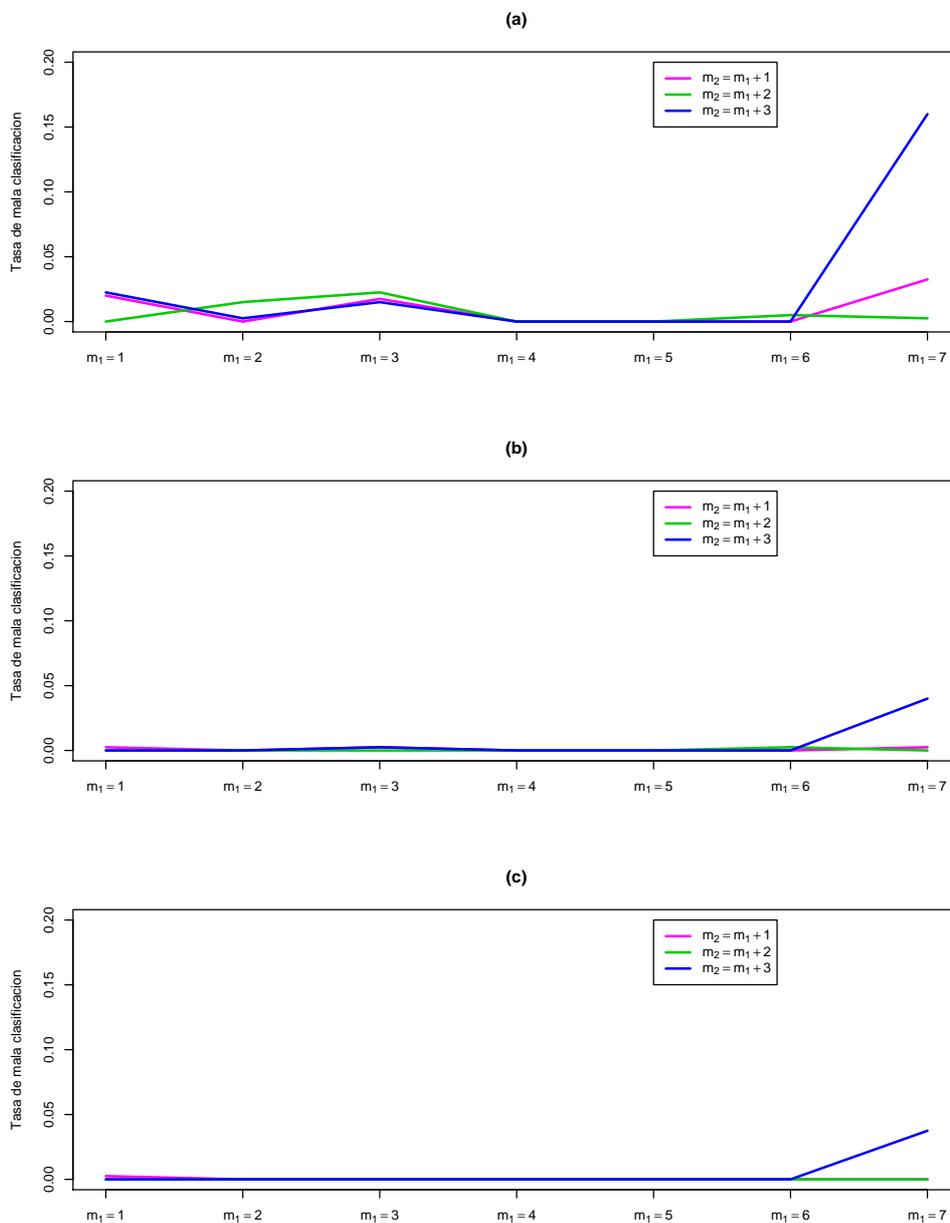
Figura 3.14: Densidad espectral correspondiente a procesos m-dependientes con distribución marginal gamma.



Se generaron $s = 200$ series generadas a partir de cada una de las poblaciones a discriminar y se clasificaron de acuerdo con el estadístico D_W utilizando los tres estimadores espectrales que han sido estudiados con idénticos parámetros que en las simulaciones anteriores. Los resultados alcanzados se muestran en la Figura 3.15.

En este caso, se obtuvieron tasas de error muy bajas, por debajo del 2% en la mayoría de situaciones analizadas, y que se incrementan a medida que los parámetros m

Figura 3.15: Tasas de error para discriminar entre series m -dependientes con distribución marginal gamma. (a) Periodograma suavizado \hat{f}_{DLS} , (b) Log-periodograma suavizado por mínimos cuadrados \hat{f}_{LS} , (c) Log-periodograma suavizado por máxima verosimilitud \hat{f}_{LK}



que definen la estructura de dependencia de las series aumenta. Para valores mayores de m la estructura de autocorrelación de las series entre las que se discrimina, tal y como se deduce de (3.12), es muy similar, lo que justifica que empeoren los resultados obtenidos en la clasificación. En cuanto a los resultados obtenidos con cada uno de los tres

estimadores del espectro, en este caso no se observan peores resultados para el estimador log-periodograma suavizado, obteniéndose tasas de mala clasificación similares con los tres estimadores utilizados, resultando ligeramente más altas en el caso de trabajar con el periodograma suavizado, \widehat{f}_{DLS} .

d) Discriminación de series m -dependientes con distribución marginal log-normal

Se consideró ahora bajo Π_j , el siguiente modelo:

$$X_t = \prod_{i=t}^{t+m_j-1} \exp(Y_i) = \exp\left\{ \sum_{i=t}^{t+m_j-1} Y_i \right\} \quad (3.13)$$

donde $\{Y_1, Y_2, \dots, Y_{m_j+n-1}\}$ es una secuencia de variables aleatorias i.i.d. con distribución normal estándar $N(0, \frac{\sigma}{\sqrt{m_j}})$. Las series así generadas son m -dependientes con distribución marginal *Lognormal*(0, σ).

Igual que en las simulaciones anteriores, se generaron series de longitud $n = 400$ a partir de Π_j , ($j = 1, 2$), de acuerdo con la ecuación (3.13), con valores del parámetro $m_1 = 1, 2, 3, \dots, 7$, $m_2 = m_1 + k$, para $k = 1, 2, 3$. Las funciones de autocorrelación muestral correspondientes vendrán dadas según la estructura de autocovarianzas:

$$\gamma(h) = \begin{cases} \exp\left(\sigma^2\left(2 - \frac{h}{m}\right)\right) - \exp(\sigma^2) & \text{si } 0 \leq h < m \\ 0 & \text{si } h \geq m \end{cases} \quad (3.14)$$

que se corresponde con la función de densidad espectral dada por (Figura 3.16):

$$f(\lambda) = \frac{\cos(\lambda h)}{\pi} \left\{ \exp(2\sigma^2) - \exp(\sigma^2) + \sum_{h=1}^{m-1} \exp\left(\sigma^2\left(2 - \frac{h}{m}\right)\right) - \exp(\sigma^2) \right\}. \quad (3.15)$$

Se clasificaron $s = 200$ series generadas a partir de cada una de las poblaciones a discriminar y se clasificaron de acuerdo con el estadístico D_W utilizando los tres estimadores espectrales que han sido estudiados con idénticos parámetros que en las simulaciones anteriores. Los resultados alcanzados se muestran en la Figura 3.17.

En este caso se obtuvieron tasas de mala clasificación más elevadas que en los casos anteriores, que resultaron ser mayores para valores grandes de los parámetros m_1 y m_2 . De nuevo esto puede venir justificado por la estructura de autocorrelaciones de las series entre las que se discrimina, que resultan tanto más similares cuanto mayores sean estos valores. Aunque las diferencias son menos relevantes, de nuevo se aprecian tasas de error ligeramente más elevadas cuando se trabaja con el estimador log-periodograma suavizado, \widehat{f}_{LS} que con los otros dos estimadores del espectro, \widehat{f}_{DLS} y \widehat{f}_{LK} .

Figura 3.16: Densidad espectral correspondiente a procesos m -dependientes con distribución marginal lognormal.

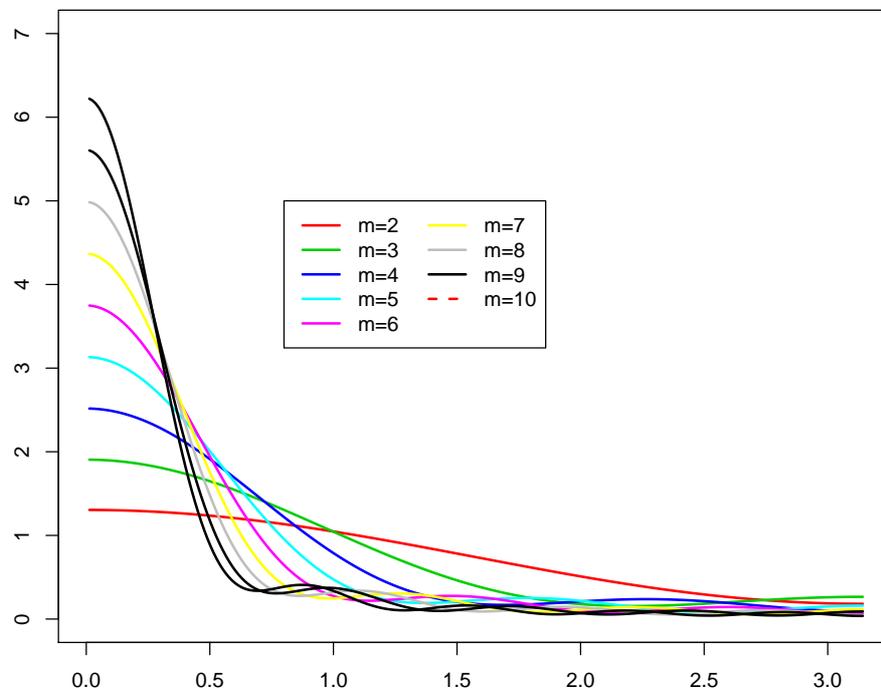
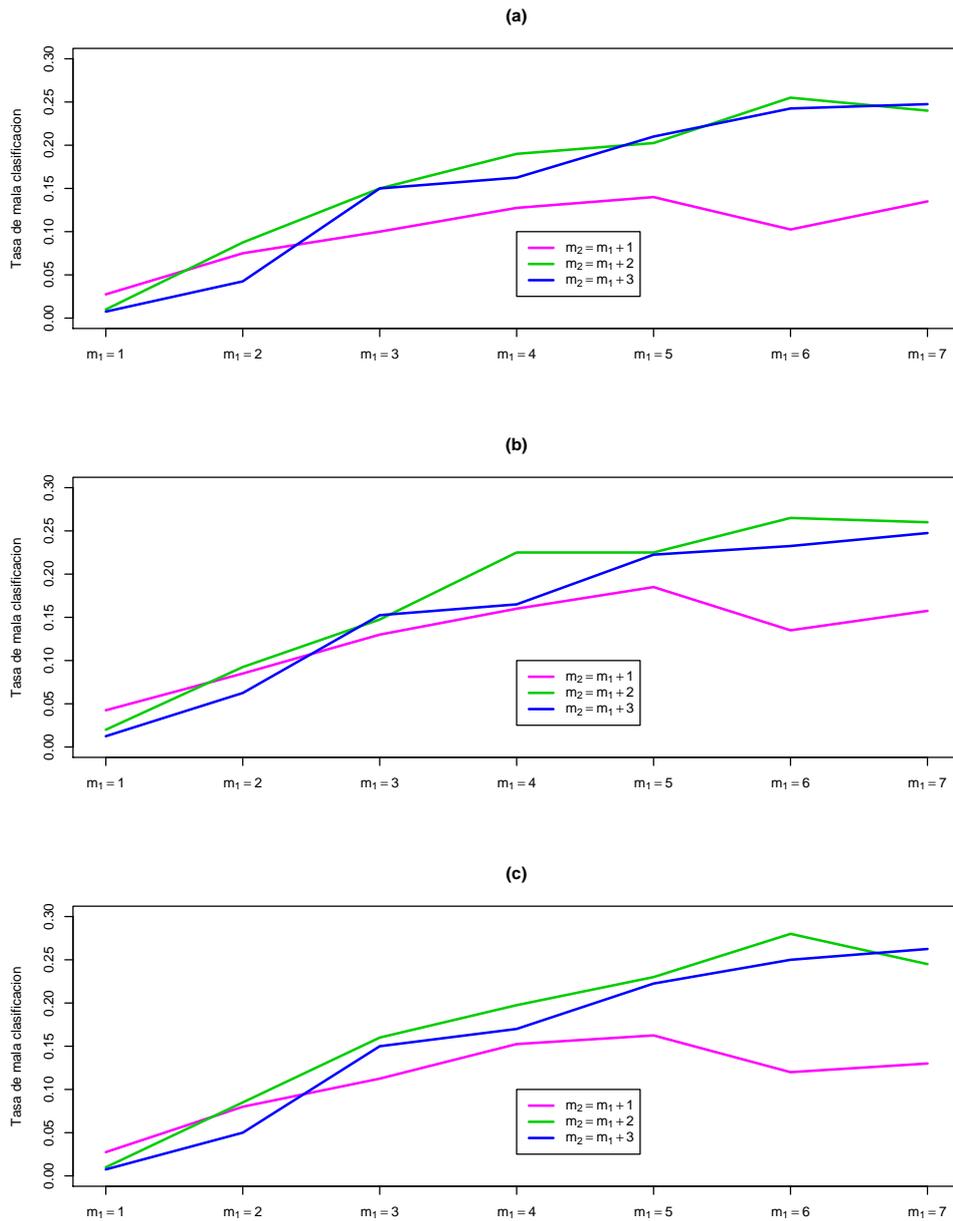


Figura 3.17: Tasas de error para discriminar entre series m-dependientes con distribución marginal lognormal. (a) Periodograma suavizado \hat{f}_{DLS} , (b) Log-periodograma suavizado por mínimos cuadrados \hat{f}_{LS} , (c) Log-periodograma suavizado por máxima verosimilitud \hat{f}_{LK} .



3.1.4. Conclusiones

Tras el estudio de simulación llevado a cabo en la presente sección, y después de una observación detallada de los resultados obtenidos, se pueden extraer las siguientes conclusiones:

- El procedimiento discriminante propuesto presenta un buen comportamiento para la clasificación de realizaciones de procesos estacionarios Gaussianos.
- Las tasas de mala clasificación alcanzadas con el estadístico discriminante D_W resultan competitivas si se comparan con las obtenidas por otros procedimientos utilizados en la literatura para la clasificación de procesos estacionarios Gaussianos.
- Las tasas de error disminuyen a medida que aumenta la separabilidad entre las clases de procesos entre los que se discrimina, definida en términos de su estructura espectral o, equivalentemente, de su estructura de autocorrelaciones.
- Los resultados del análisis discriminante mejoran notablemente cuando se clasifican series de mayor longitud.
- Las propiedades asintóticas de cada uno de los tres estimadores del espectro considerados (\hat{f}_{DLS} , \hat{f}_{LS} , \hat{f}_{LK}) influyen parcialmente en el comportamiento del estadístico discriminante D_W .
- Los peores resultados, en términos de la tasa de mala clasificación, se alcanzan cuando para el cómputo del estadístico discriminante D_W se utiliza el suavizador log-periodograma suavizado, \hat{f}_{LS} , al ser este menos eficiente que los otros dos estimadores del espectro.
- El comportamiento del estadístico discriminante D_W resulta muy similar cuando se utilizan como estimadores del espectro \hat{f}_{DLS} o \hat{f}_{LK} , a pesar de que cabría esperar un mejor comportamiento con el estimador \hat{f}_{LK} , al presentar un menor sesgo que \hat{f}_{DLS} en las regiones con picos del espectro.
- De modo general, la distribución marginal de las series no afecta a las tasas de mala clasificación obtenidas con el procedimiento discriminante propuesto, pese a la ausencia de resultados teóricos acerca de la distribución asintótica del estadístico discriminante en este tipo de situaciones.

3.2. Aplicación a datos reales: clasificación de fonemas.

Utilizaremos los datos descritos por Hastie, Buja y Tibshirani (1995) para ilustrar la aplicación del método discriminante propuesto en el Capítulo 2 a un conjunto de datos reales. Las series utilizadas proceden a su vez de una base de datos más amplia (*TIMIT Acoustic-Phonetic Continuous Speech Corpus, NTIS, US Department of Commerce*), que constituye una fuente de datos repetidamente utilizada en estudios de reconocimiento de patrones de voz.

Se ha utilizado únicamente un subconjunto del total de datos disponibles, de forma que se seleccionaron cinco fonemas para su discriminación a partir de registros de voz digitalizados. Los cinco fonemas utilizados pueden transcribirse como: “sh” (tal y como se pronunciaría en el término inglés “she”), “dcl” (como en “dark”), “iy” (como en “she”), “aa” (como la vocal en “dark”) y “ao” (tal y como se pronuncia la primera vocal en “water”). Partiendo de los registros de voz de un total de 12 varones, se seleccionaron 4509 series de datos de 32 ms. de duración cada una, con una tasa de muestreo de 16 kHz (es decir, 1 observación cada $1/16000=0,0000625$ segundos= $0,0625$ msec). Cada serie corresponde a la pronunciación de uno de los cinco fonemas anteriores, y consta así de un total de 512 observaciones equiespaciadas. Los datos disponibles no corresponden a las series originales, sino al logaritmo de los periodogramas correspondientes. La distribución de los 4509 registros disponibles, según el fonema al que corresponden es como sigue:

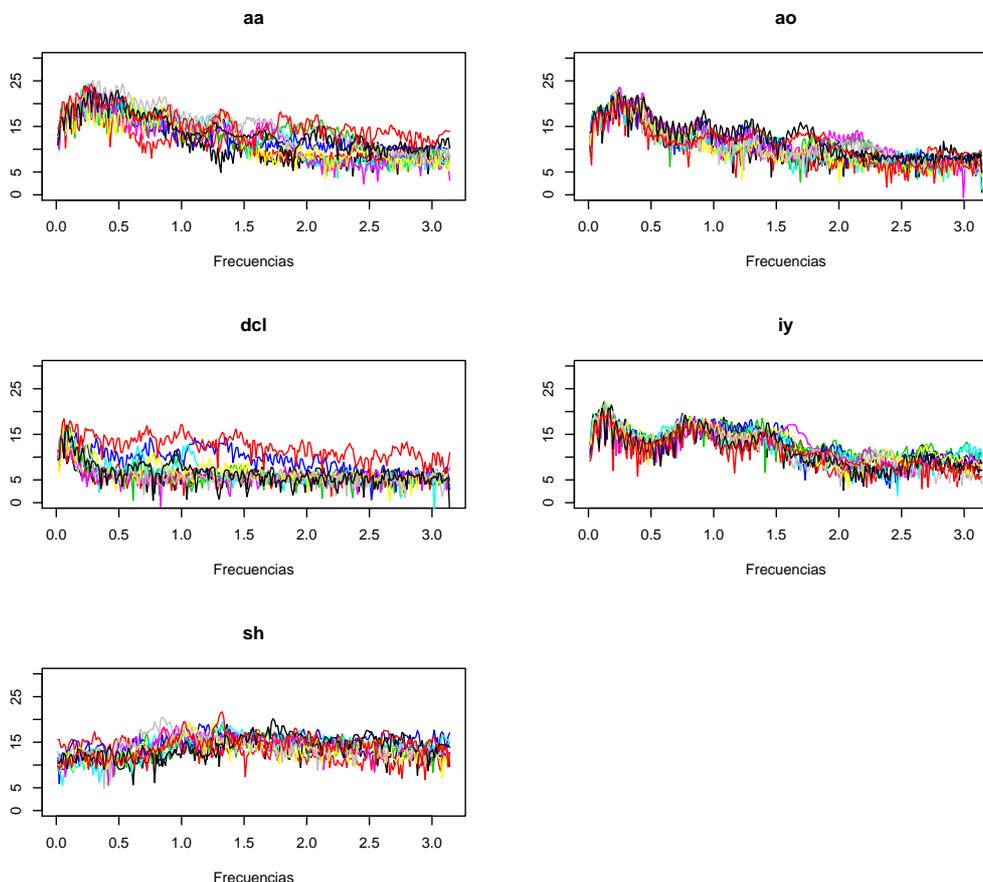
	Fonema				
	aa	ao	dcl	iy	sh
Número de series	695	1022	757	1163	872

Para cada una de las series el valor del periodograma (equivalentemente, de su logaritmo) se calcula sobre las frecuencias de Fourier en $(0, \pi)$:

$$I(\lambda_k) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{-it\lambda_k} \right|^2, \quad \lambda_k = \frac{2\pi k}{n}, \quad k = 1, 2, \dots, [n/2],$$

siendo $n = 512$ la longitud de las series consideradas. Los datos disponibles consisten así en 4509 log-periodogramas de longitud $[n/2]=256$, correspondientes a un rango de frecuencias de 0-8 kHz, de los cuales se conoce además la clase (fonema) al que corresponden. En la Figura 3.18 se representa una muestra aleatoria de 10 log-periodogramas correspondientes a cada una de las clases de fonemas consideradas. Como se puede observar, existe una enorme variabilidad entre las series correspondientes a un mismo fonema, que se traslada a los correspondientes periodogramas (Figura 3.19). Obviamente este hecho debería dificultar el problema de discriminación entre los grupos de estudio.

Figura 3.18: Logaritmo del periodograma correspondiente a una muestra aleatoria de 10 series de cada una de las clases de fonemas sujetas a discriminación.

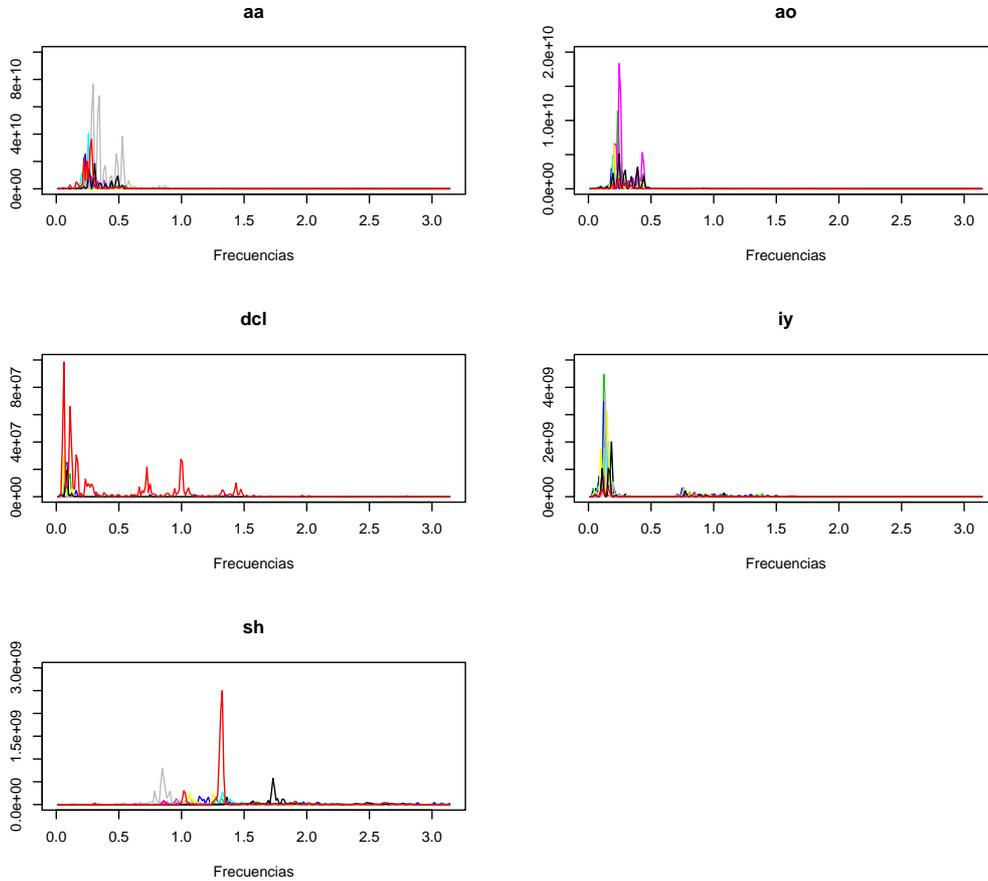


Inicialmente, se exploró el comportamiento del estadístico discriminante D_W utilizando una muestra de entrenamiento de 250 log-periodogramas seleccionados de forma aleatoria, 50 por cada una de las clases de fonemas disponibles. Siguiendo el procedimiento descrito en la Sección 2.4 de la presente memoria, se estimó la densidad espectral correspondiente a cada una de las series de entrenamiento utilizando los estimadores espectrales propuestos por Fan y Kreutzberger (1998): \hat{f}_{DLS} , \hat{f}_{LS} y \hat{f}_{LK} . Para cada una de las clases de fonemas consideradas, la densidad espectral correspondiente a cada clase se estimó promediando las estimaciones del espectro obtenidas para las series de entrenamiento en dicho grupo.

A su vez, se utilizaron para la clasificación 50 series por grupo, constituyendo una muestra test de un total de 250 log-periodogramas. Cada una de las series utilizadas como test fueron clasificadas en una de las cinco clases de fonemas de acuerdo con la medida de disparidad espectral

$$D_W(\hat{f}_n, \hat{f}_i) = \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\hat{f}_n(\lambda) \hat{f}_i^{-1}(\lambda)) d\lambda, \quad i = 1, 2, \dots, 5, \quad (3.16)$$

Figura 3.19: Periodograma correspondiente a una muestra aleatoria de 10 series de cada una de las clases de fonemas sujetas a discriminación.



siendo \hat{f}_n un estimador de la densidad espectral correspondiente a la serie considerada y $\hat{f}_i, i = 1, \dots, 5$ el estimador no paramétrico de la densidad espectral correspondiente a la clase de fonemas i . Se utilizó la función

$$W(x) = \log(\alpha x + (1 - \alpha)) - \alpha \log(x) \quad (3.17)$$

como función de divergencia, tomando $\alpha = 0,5$ en todos los casos. Para el cómputo de los estimadores no paramétricos de la densidad espectral se utilizó el núcleo de Epanechnikov, y el parámetro ventana se determinó en cada caso minimizando la función de validación cruzada sobre un rango de valores razonables de h . Para el cómputo del estimador log-periodograma suavizado basado en el criterio de máxima verosimilitud local, se empleó el algoritmo iterativo de Newton-Raphson.

Cada serie se clasificó así en el grupo para el que se minimizaba la distancia (3.16) entre su densidad espectral y la densidad espectral de cada grupo, estimada ésta a partir

del promedio de las densidades espectrales correspondientes a las series de entrenamiento en ese grupo. Con este criterio, se calculó el número de series mal clasificadas en cada una de las clases de fonemas, y se promediaron las tasas de mala clasificación para obtener una tasa global de error. La Tabla 3.14 muestra los resultados obtenidos utilizando el periodograma suavizado por mínimos cuadrados locales en la estimación de la densidad espectral.

Tabla 3.14: Tasas de mala clasificación sobre una muestra test de 250 series de fonemas. La densidad espectral en cada clase se estimó promediando las estimaciones del espectro de las series de entrenamiento, utilizando el estimador periodograma suavizado por mínimos cuadrados locales, \hat{f}_{DLS} .

Fonema	Fonema					Tasa de error
	“aa”	“ao”	“dcl”	“iy”	“sh”	
“aa”	30 %	68 %	2 %	0	0	70 %
“ao”	2	90 %	8 %	0	0	10 %
“dcl”	0	4	96 %	0	0	4 %
“iy”	0	56 %	14 %	30 %	0	70 %
“sh”	0	4	0	8 %	88 %	12 %

Tasa global de mala clasificación=33.2 %

Se obtuvo una tasa global de mala clasificación del 33.2 %, que resulta muy elevada si se compara con la alcanzada con otros procedimientos utilizados con anterioridad para la clasificación de este mismo conjunto de datos. Así, Hastie, Buja y Tibshirani (1995) refieren tasas de error en torno a un 9 % empleando una versión regularizada del análisis discriminante lineal (LDA), que denominan análisis discriminante penalizado (PDA). A su vez, Ferraty y Vieu (2003), utilizando un enfoque funcional, obtienen tasas de mala clasificación que no superan el 13 %.

Como alternativa, se exploró la posibilidad de realizar el análisis discriminante promediando en primer lugar los periodogramas de las 50 series de entrenamiento disponibles en cada grupo. En este caso, la densidad espectral correspondiente a cada una de las clases de fonemas consideradas se estimó suavizando el promedio de los periodogramas en cada clase.

Se utilizaron de nuevo para la clasificación 50 series por grupo, constituyendo una muestra test de un total de 250 log-periodogramas. Cada una de las series test fueron clasificadas en una de las cinco clases de fonemas de acuerdo con la medida de disparidad espectral (3.16), donde ahora $\hat{f}_i, i = 1, \dots, 5$ denota al estimador no paramétrico de la

densidad espectral correspondiente a la clase i obtenido a partir del promedio de los periodogramas de las series de entrenamiento. En este caso se utilizó siempre el estimador periodograma suavizado, \hat{f}_{DLS} , estudiado por Fan y Kreutzberger (1998) (nótese que al suavizar el promedio de los periodogramas el estimador log-periodograma suavizado por mínimos cuadrados, \hat{f}_{DLS} , resulta no ser un estimador insesgado de la densidad espectral correspondiente. De igual forma, si en lugar de promediar los periodogramas de las series de entrenamiento, se calculara el promedio de los log-periodogramas correspondientes y se utilizara éste para obtener estimaciones del espectro de cada clase, el estimador basado en máxima verosimilitud, \hat{f}_{LK} , resultaría ser un estimador sesgado).

Manteniendo el resto de parámetros como en el caso anterior, se calculó el número de series asignadas a cada uno de los grupos de fonemas y la tasa global de mala clasificación. La Tabla 3.15 muestra los resultados obtenidos utilizando el periodograma suavizado por mínimos cuadrados locales en la estimación de la densidad espectral.

Tabla 3.15: Tasas de mal clasificación sobre una muestra test de 250 series de fonemas. La densidad espectral en cada clase se estimó suavizando el promedio de los periodogramas de las series de entrenamiento, utilizando el estimador periodograma suavizado.

Fonema	Fonema					Tasa de error
	“aa”	“ao”	“dcl”	“iy”	“sh”	
“aa”	20 %	64 %	16 %	0	0	80 %
“ao”	0	62 %	38 %	0	0	38 %
“dcl”	0	0	100 %	0	0	0 %
“iy”	0	10 %	24 %	66 %	0	34 %
“sh”	0	0	0	8 %	92 %	8 %

Tasa global de mala clasificación=32 %

La tasa global de error fue así de un 32 %, resultando de nuevo muy elevada en relación a la alcanzada con los procedimientos descritos por Hastie, Buja y Tibshirani (1995) y Ferraty y Vieu (2003). La tasa de mala clasificación alcanzó el 80 % para las series correspondientes al fonema “aa”, manteniéndose por debajo del 10 % únicamente para los grupos “sh” y “dcl”, siendo éste último el único para el cual todas las series se ubicaron correctamente.

Al examinar con detenimiento los resultados del análisis discriminante y los periodogramas de las series a clasificar, se sospechó que la elevada tasa de error obtenida podría deberse a que, al promediar los periodogramas (o, equivalentemente, las densidades espectrales estimadas) de las series de entrenamiento, se está eliminando gran parte de la

variabilidad inherente a las series dentro de cada grupo propia del ejemplo que nos ocupa, tal y como habíamos constatado en las Figuras 3.18 y 3.19. Como ejemplo, en la Figura 3.20 se muestra la densidad espectral estimada de una muestra aleatoria de series correspondientes al fonema “aa”, junto con la densidad espectral estimada para las clases de fonemas “aa” y “ao”. Tal y como se puede observar, y debido a la enorme variabilidad que existe entre los periodogramas correspondientes a un mismo fonema, existen series cuyas densidades se encuentran notablemente más próximas a la densidad espectral del grupo “ao” que a la de su propia clase. Este hecho podría justificar los malos resultados del procedimiento discriminante. Nótese que otros criterios de clasificación utilizados con anterioridad para discriminar este mismo conjunto de datos, como el de Ferraty y Vieu (2003) obviaban este problema “eliminando” aquellas series más lejanas a la hora de calcular la distancia de la serie a cada uno de los grupos. De hecho, su procedimiento se basa en estimar no paramétricamente las probabilidades a posteriori de que una serie pertenezca a cada una de las clases. Para ello, se introduce un parámetro de suavizado h de modo que para la estimación de cada una de las probabilidades a posteriori se utilizan únicamente aquellas curvas en cada grupo que distan de la serie a clasificar a lo sumo una cantidad h .

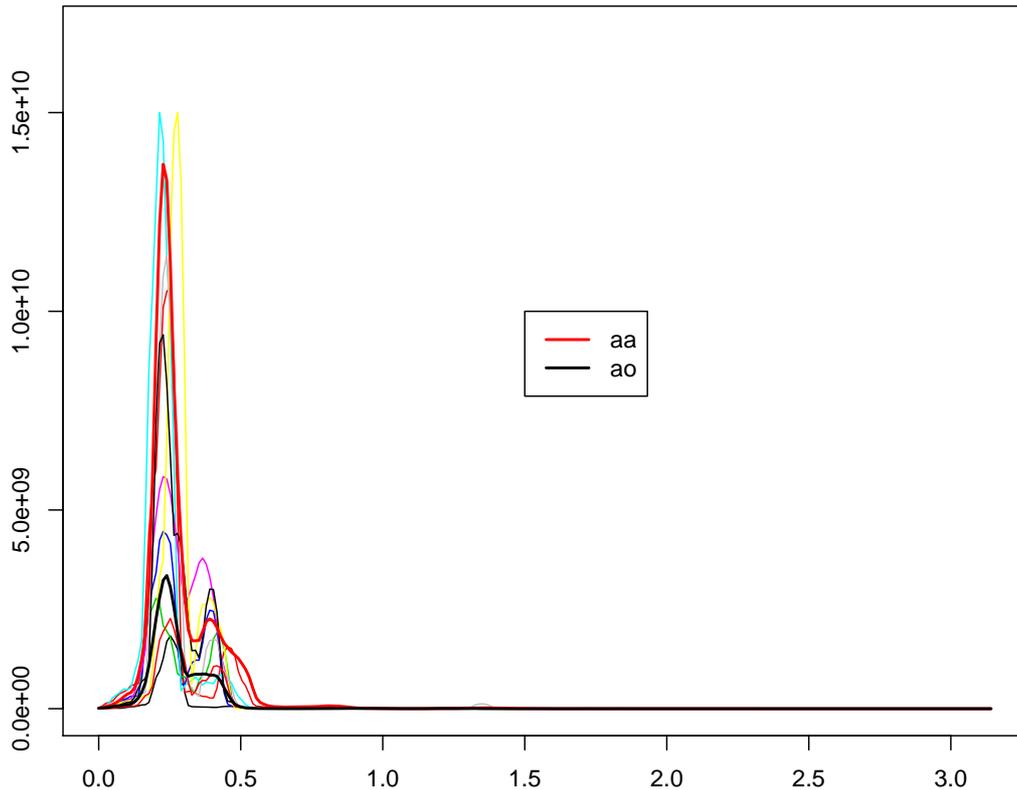
Una forma de solventar este problema consistiría en adoptar otro criterio como medida de proximidad entre cada serie y cada uno de los grupos de fonemas. Para ello se puede recurrir a cualquiera de los criterios habitualmente utilizados en el análisis cluster, como el del vecino más próximo o enlace simple o el método de enlace promedio. Se optó por calcular la proximidad entre una serie y cada uno de los grupos de fonemas promediando las distancias espectrales entre dicha serie y las series de entrenamiento en cada grupo. Así, la distancia entre cada serie X_n y la clase i , $i = 1, 2, \dots, 5$ se calculó como:

$$D_W(\hat{f}_n, G_i) = \frac{1}{r_i} \sum_{j=1}^{r_i} D_W(\hat{f}_n, \hat{f}_{ij})$$

siendo \hat{f}_n un estimador de la densidad espectral correspondiente a cada una de las series a clasificar y \hat{f}_{ij} , $j = 1, \dots, r_i$ la densidad espectral estimada de cada una de las r_i series en la clase de fonemas i -ésima, $i = 1, 2, \dots, 5$. Cada serie se clasificó en el grupo para el que se minimizaba la distancia anterior, utilizando cada uno de los tres estimadores del espectro con idénticos parámetros que en los casos anteriores.

Con este criterio, se abordó finalmente el problema de discriminación. Se utilizó una muestra de entrenamiento de 750 log-periodogramas seleccionados de forma aleatoria, 150 por cada una de las clases de fonemas. A su vez, entre las series restantes, se seleccionó para su clasificación una muestra test de 1250 log-periodogramas, 250 series por grupo. Este procedimiento se repitió 50 veces, de modo que se dispuso de un total de 50 muestras test

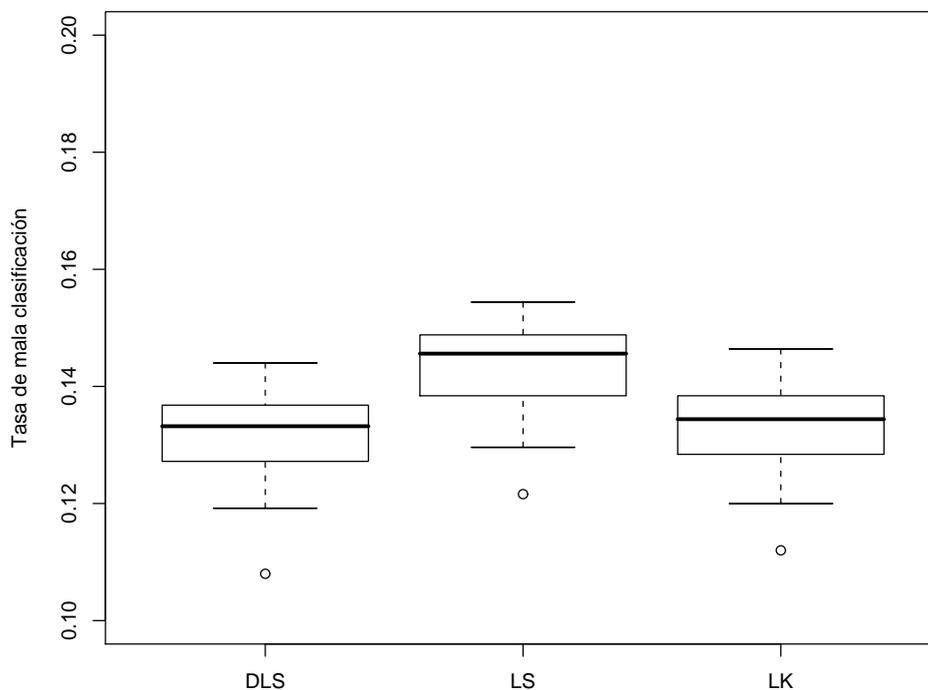
Figura 3.20: Densidad espectral estimada de una muestra aleatoria de series correspondientes a la clase de fonemas “aa”. Se representan conjuntamente (en grueso) las densidades espectrales de las clases “aa” y “ao”, estimadas a partir del promedio de los espectros suavizados de las muestras de entrenamiento.



diferentes para su clasificación. En cada iteración, cada una de las series utilizadas como test fue clasificada en uno de los cinco grupos de fonemas promediando las distancias entre su densidad espectral y las densidades espectrales de las series de entrenamiento en cada grupo. Para ello se utilizaron cualquiera de los tres estimadores espectrales estudiados. Para el cómputo del estadístico D_W se utilizó la función de divergencia (3.17) con diferentes valores de α en todos los casos. En la estimación de la densidad espectral, y con cualquiera de los tres suavizadores, se utilizó el núcleo de Epanechnikov y el ancho de banda h se determinó minimizando la función de validación cruzada sobre un rango razonable de valores de h . Para cada iteración, se calculó el número de series mal clasificadas en cada

grupo y la tasa global de error. En la Figura 3.21 se muestran los diagramas de cajas para las tasas globales de mala clasificación obtenidas en las 50 iteraciones con cada uno de los tres estimadores espectrales considerados, tomando $\alpha = 0,5$ para el cómputo de la función de divergencia en (3.17).

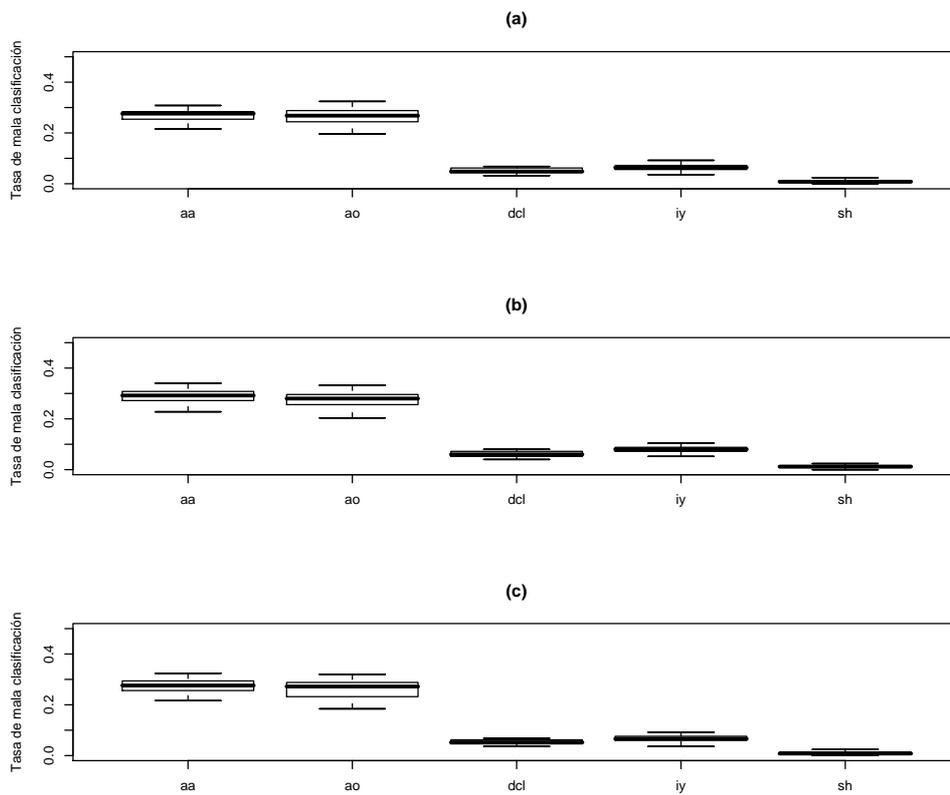
Figura 3.21: Tasas de error sobre 50 muestras test para la clasificación de las series de fonemas. La distancia a cada grupo se calcula como el promedio de las distancias a las series de entrenamiento en ese grupo, según $D_W(\hat{f}_n, \hat{f}_{ij})$, utilizando la función (3.17) con $\alpha = 0.5$. Resultados con los estimadores \hat{f}_{DLS} , \hat{f}_{LS} y \hat{f}_{LK} .



Tal y como se puede observar, la tasa global de mala clasificación se situó en torno al 13 %, independientemente del estimador del espectro considerado, si bien de nuevo parece objetivarse la influencia de las propiedades asintóticas de los suavizadores locales sobre los resultados en la clasificación. Así, los resultados con el estimador log-periodograma suavizado tienden a ser ligeramente peores que con los otros dos estimadores del espectro (Mediana de la tasa de error: 14,6 % para \hat{f}_{LS} versus 13,3 % para \hat{f}_{DLS} y 13,4 % para \hat{f}_{LK}). Si se analizan las tasas de mala clasificación según la clase de origen de las series, los peores resultados se alcanzaron para los grupos de fonemas “aa” y “ao”, con tasas de error

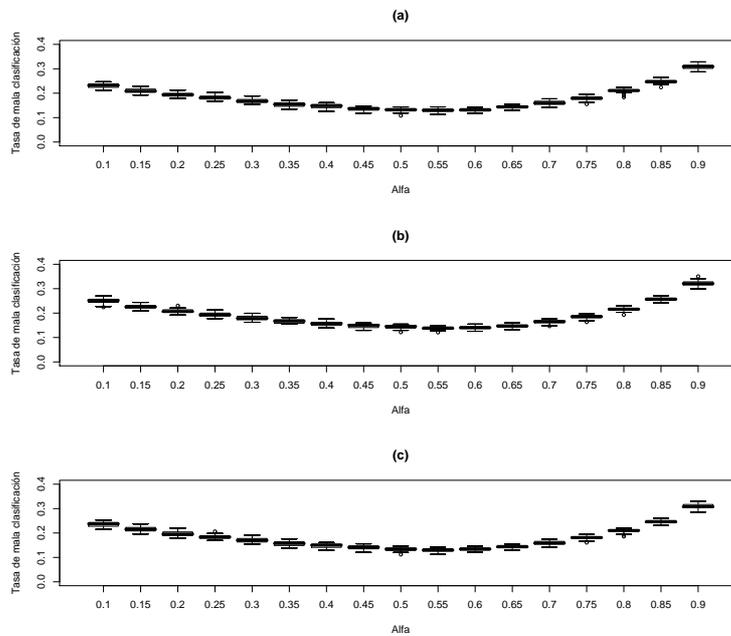
promedio en torno al 27 %, independientemente del estimador del espectro considerado. Para el resto de las clases, se obtienen tasas de error alrededor del 5 % para la clase “dcl”, en torno al 6 % para la clase “iy” y del 1 % para el fonema “sh” (Figura 3.22).

Figura 3.22: Tasas de error sobre 50 muestras test para la clasificación de cada una de las clases de fonemas. La distancia a cada grupo se calcula como el promedio de las distancias a las series de entrenamiento en ese grupo, según $D_W(\hat{f}_n, \hat{f}_j)$, utilizando la función (3.17) con $\alpha = 0.5$. Resultados con los estimadores: (a) \hat{f}_{DLS} , (b) \hat{f}_{LS} y (c) \hat{f}_{LK} .



Finalmente, se exploró la influencia del parámetro α en (3.17) sobre los resultados alcanzados (Figura 3.23). Las mayores tasas de error se alcanzaron para valores extremos de α , mientras que la mínima tasa de error se obtuvo con $\alpha = 0.55$. Con este valor, la tasa media de error fue de un 12,96 % para el estimador periodograma suavizado, de un 13,82 % para el estimador log-periodograma suavizado y de un 13,03 % para el estimador basado en máxima verosimilitud.

Figura 3.23: Tasas de error sobre 50 muestras test para la clasificación de las series de fonemas. La distancia a cada grupo se calcula como el promedio de las distancias a las series de entrenamiento en ese grupo, según $D_W(\hat{f}_n, \hat{f}_{ij})$, utilizando la función (3.17) con distintos valores de $\alpha =$. Resultados con los estimadores: (a) \hat{f}_{DLS} , (b) \hat{f}_{LS} y (c) \hat{f}_{LK} .



Capítulo 4

Análisis cluster de series temporales

4.1. Introducción

El análisis cluster es un proceso de clasificación no supervisada cuyo objetivo se centra en encontrar similitudes y diferencias entre un conjunto más o menos grande de observaciones a fin de poder clasificarlas en un número reducido de clases, cada una de ellas constituida por unidades de características similares. A diferencia del análisis discriminante, las características de cada grupo y, en muchos casos, el número de clusters que deben identificarse, se desconocen a priori y deben determinarse a partir de los propios datos.

La mayoría de los trabajos sobre análisis cluster se han centrado en la clasificación de datos estáticos, distinguiéndose generalmente entre dos tipos de algoritmos cluster: clustering jerárquico o clustering partitivo (ver, por ejemplo, Jain y Dubes (1988)). Los métodos jerárquicos se basan en agrupar los datos de forma secuencial. Generalmente se distinguen dos tipos de soluciones jerárquicas, y se habla a su vez de métodos aglomerativos o métodos divisivos. En los métodos aglomerativos, se parte de la solución inicial donde cada individuo es un grupo, de modo que en cada paso dos objetos o grupos de objetos se unen dando lugar a un nuevo cluster, hasta llegar a la solución final con un sólo grupo que contiene todas las observaciones. Para los métodos divisivos, se procede de la forma contraria. Dentro de este tipo de técnicas, existen distintos algoritmos según cuál sea el criterio elegido para la sucesiva transición entre grupos pequeños y otros de mayor tamaño: vecino más próximo o enlace simple (que maximiza la mínima distancia entre los objetos dentro del mismo grupo), vecino más lejano o enlace completo (que minimiza la máxima distancia entre los objetos dentro del mismo grupo), agrupación de medias o

enlace promedio (basado en el promedio de las distancias entre los objetos en diferentes grupos), método de Ward (basado en minimizar la variación intragrupal de la estructura formada), etc.

Los métodos partitivos requieren a su vez conocer a priori el número k de clusters que se quiere identificar. La forma intuitiva de proceder consiste en dividir los m datos observados en k clusters de forma que se optimice algún tipo de criterio o función de coste asociada. Entre los métodos cluster partitivos, el más utilizado es el algoritmo de k -medias, donde cada cluster se representa por el valor medio de los elementos en ese grupo. Se trata de un método iterativo que persigue minimizar la distancia entre los clusters. Una generalización de este método es el algoritmo EM (Expectation Maximizing), que tiene en cuenta además la matriz de covarianzas entre los elementos de cada cluster. Aunque existen otros métodos, estos algoritmos se encuentran entre los más utilizados de entre los métodos partitivos. Además de los métodos jerárquicos y partitivos, otros algoritmos cluster, menos utilizados, son los denominados “overlapping clusters”, métodos basados en densidades, “grid-based methods” y “model-based methods” (Han y Kamber (2001)).

Uno de los puntos clave en el análisis cluster es la elección de una distancia o similitud adecuada para medir la discrepancia entre cada par de observaciones. La mayoría de los algoritmos cluster antes mencionados requieren disponer de una matriz de proximidad que mida el nivel de similitud o disimilitud entre todos los posibles pares de observaciones disponibles. El cálculo de esta matriz es el primer paso del análisis y la elección de una medida de proximidad dependerá, además de la finalidad del proceso de clasificación, del tipo de observaciones a clasificar. Entre las distancias más habitualmente utilizadas están la distancia euclídea, la distancia de Mahalanobis, de Minkowski, el coeficiente de congruencia, el coeficiente de correlación, etc.

El trabajo con series de tiempo resulta de gran interés debido a la preponderancia de este tipo de datos en ámbitos tan diferentes como la ingeniería, física, finanzas y economía, medicina, sismología, meteorología, reconocimiento de patrones, robótica o zoología. Así, el cluster de series de tiempo puede resultar de utilidad en estas áreas para permitir detectar unos pocos patrones representativos de una muestra más grande, predecir comportamientos futuros, cuantificar la afinidad entre distintas entidades, etc. Algunos ejemplos en la literatura son: la clasificación de modelos de series de producción industrial (Piccolo (1990)), la comparación de datos sismológicos, como el problema de distinguir entre las ondas provocadas por terremotos y explosiones nucleares (Kakizawa, Shumway y Taniguchi (1998)), la organización de datos industriales (Galbraith y Jiaqing (1999)), la clasificación de niños de acuerdo con las similitudes en su comportamiento a lo largo del tiempo (Hirsch y DuBois (1991)), la clasificación de países de acuerdo con los registros de emisiones de CO_2 (Alonso *et al.* (2006)), el cluster de modelos de procesos ecológicos (Li

et al. (2001)) o la clasificación de series de cotización bursátil (*Coronnello et al.* (2005)).

Sin embargo, los trabajos relacionados con el análisis cluster de procesos estocásticos resultan escasos si se comparan con la multitud de estudios que se centran en el cluster de observaciones estáticas. Cabe destacar que las características propias de las series temporales hacen que las soluciones clásicas no siempre resulten adecuadas para abordar el proceso de clasificación de este tipo de datos. Tal y como se señaló anteriormente, uno de los puntos clave para llevar a cabo un análisis cluster consiste en determinar una medida de similitud o de distancia entre dos observaciones, de modo que una vez calculada la matriz de distancias entre todos los posibles pares de datos, se pueda utilizar cualquiera de los algoritmos cluster tradicionales para agrupar las observaciones. Sin embargo, el concepto de similitud entre dos series de tiempo no resulta elemental y puede establecerse de formas muy diferentes. De hecho, una medida de similitud concreta puede resultar o no más apropiada que otra dependiendo del objetivo con el que se lleve a cabo el análisis cluster o incluso del propio contexto de aplicación. Así, algunas medidas de disimilitud convencionales (por ejemplo, la distancia Euclídea entre los datos observados) pueden no comportarse adecuadamente al trabajar con series de tiempo, por estar basadas en la cercanía de los valores observados y no en la estructura de autocorrelaciones de los datos. En definitiva, algunas medidas de proximidad convencionales pueden considerar información que es inherentemente estática, y que ignora por lo tanto la evolución de las series de tiempo.

Otra característica peculiar de las series temporales es que con frecuencia la longitud de las series es grande, de modo que la alta dimensionalidad de las observaciones puede conllevar problemas computacionales que hagan más complicada la labor de clasificación. En particular, este hecho puede dificultar la aplicación de los métodos multivariantes clásicos en el dominio de tiempo, motivando la necesidad de desarrollar métodos cluster específicos para la clasificación de series temporales, recurriendo además al dominio de frecuencias.

El creciente interés en el cluster de series temporales, debido a sus importantes repercusiones prácticas, junto con las peculiaridades anteriormente mencionadas, han hecho que en los últimos años se produzca un incremento de la investigación en este sentido, dando lugar a diferentes técnicas diseñadas para detectar similitudes en la estructura de datos de tipo temporal. Una completa revisión de los trabajos realizados en este sentido se puede encontrar en Liao (2005), quien clasifica los métodos cluster de series temporales en tres grupos diferentes, según se basen directamente en los datos observados (“raw-data-based approach”), en alguna característica obtenida a partir de las series sujetas a clasificación (“feature-based approach”) o en un modelo paramétrico ajustado a los datos (“model-based approach”). Los primeros algoritmos trabajan directamente con las series

observadas (bien en el dominio de tiempo o en el dominio de frecuencias), reemplazando las medidas de distancia o similitud habitualmente utilizadas con datos estáticos por una que resulte apropiada en el contexto de series de tiempo. Los otros enfoques se basan en obtener a partir de las series observadas, o de un modelo paramétrico ajustado a cada una de ellas, ciertos parámetros a partir de los cuales afrontar la clasificación. En cualquier caso, la principal aportación consiste en proponer nuevas medidas de distancia o similitud (ya sea entre los datos observados o entre algún parámetro que caracterice a las series) que tengan en cuenta la naturaleza dinámica de este tipo de datos y, en consecuencia, mejoren los resultados de la clasificación. A partir de esta medida de distancia, los diferentes autores han optado por utilizar indistintamente cualquiera de los algoritmos de clasificación cluster tradicionales que se han mencionado con anterioridad.

Muchos de los trabajos dedicados al análisis cluster de series temporales asumen que los procesos generadores de las series observadas siguen un modelo paramétrico determinado y realizan el análisis cluster bajo esta hipótesis. En este sentido, se suelen asumir para los procesos generadores estructuras ARIMA (Maharaj (1996, 2000), Piccolo (1990), Tong y Dabas (1990), Kalpakis, Gada y Puttagunta (2001), Baragona (2001), Xiong y Yeung (2002)), mixturas de polinomios (Gaffney y Smyth (1999,2003), Bar-Joseph *et al.* (2002), Bagnall, Janakec y Zhang (2003)), o modelos de cadena de Markov (MC) o cadenas ocultas de markov (HMM) (Ramoni, Sebastiani y Cohen (2002), Cadez, Gaffney y Smyth (2000), Oates (1999), Oates, Firoui y Cohen (1999)(2001), Oates, Schmill y Cohen (2000), Smyth (1997), Zhong (2002), Zhong y Ghosh (2002,2002b, 2002c,2003)). La mayoría de los trabajos que optan por este enfoque paramétrico asumen que las series observadas proceden de un proceso ARIMA, estiman los parámetros del modelo a partir de los datos, y proponen realizar el análisis cluster basándose en la similitud entre los parámetros de los modelos ajustados. Así, Piccolo (1990) propone ajustar un modelo ARIMA a cada una de las series observadas y luego clasificarlas utilizando como medida de distancia la distancia euclídea entre los coeficientes de las expansiones autorregresivas asociadas. Maharaj (1996) utiliza la distancia propuesta por Piccolo (1990) para desarrollar un test de hipótesis con el fin de contrastar si dos series estacionarias son o no realizaciones de un mismo proceso estocástico. El estadístico asociado, así como el p-valor correspondiente son propuestos como medidas que pueden ser utilizadas a su vez en el análisis cluster de series temporales. Posteriormente, la misma autora ha extendido este procedimiento al caso de trabajar con series no necesariamente independientes entre sí (Maharaj (2000)).

Una forma alternativa de definir la distancia entre las modelizaciones ARIMA de las series de datos se basa en los residuos del modelo ajustado (Tong y Dabas (1990), Baragona (2001)). Así, Tong y Dabas (1990) clasifican diferentes modelos ARIMA ajustados a una misma serie de tiempo en función de la distancia entre los residuos obtenidos tras cada

ajuste, definida de diferentes formas. Es obvio que un enfoque similar puede utilizarse para el cluster de modelos ajustados a diferentes series temporales. De hecho, Baragona (2001) utiliza también una distancia basada en los coeficientes de autocorrelación de las series de residuos para abordar este problema.

Siguiendo en esta misma línea, más recientemente, Kalpakis, Gada y Puttagunta (2001) han propuesto diferentes algoritmos para el cluster de series de tiempo a partir de modelizaciones ARIMA, transformando los datos observados en los coeficientes cepstral asociados. Así mismo, Xiong y Yeung (2002) han desarrollado un nuevo método de clasificación basado también en la modelización de los datos mediante mixturas de modelos ARMA.

Además de la modelización ARIMA, muchos autores abordan el problema del análisis cluster de series de tiempo asumiendo que las series dentro de cada cluster tienen un mismo proceso generador que puede ser descrito mediante una cadena de Markov (CM) o una cadena oculta de Markov (HMM). La principal diferencia con el enfoque basado en modelos ARIMA es que bajo la hipótesis de un modelo ARIMA el ajuste se realiza sobre cada una de las series de datos antes de realizar el análisis cluster, mientras que aquí se suele obtener un modelo diferente a partir de los clusters identificados en cada una de las iteraciones del proceso de clasificación. Algunas referencias en este sentido son Ramoni, Sebastiani y Cohen (2002), Cadez, Gaffney y Smyth (2000), Ridgeway (1997), Oates (1999), Oates, Firoui y Cohen (1999, 2001), Oates, Schmill y Cohen (2000), Smyth (1997), Zhong (2002), Zhong y Ghosh (2002, 2002b, 2002c, 2003), Li (2000), Li y Biswas (1999, 1999b, 1999c, 2000), Li *et al.* (2001) y Alon *et al.* (2003).

Otra alternativa que puede encontrarse en la literatura consiste en asumir que el modelo subyacente es una mixtura de funciones polinómicas o mixtura de modelos Gaussianos. Gaffney y Smyth (1999, 2003) asumen un modelo de regresión de mixturas, utilizando un algoritmo EM para estimar las probabilidades de pertenencia a cada cluster. En esta misma línea, Bar-Joseph *et al.* (2002) adoptan un modelo de mixturas de splines para datos genéticos, utilizando también el método de clasificación EM.

Parece lógico pensar que el buen funcionamiento de estos procedimientos se basa en que realmente se verifiquen las hipótesis sobre las que se asientan, es decir, que las series de tiempo sujetas a clasificación realmente se ajusten a los modelos propuestos, pudiendo conducir a resultados equívocos en caso contrario. Un enfoque alternativo consiste en no asumir ningún modelo teórico para los procesos generadores de las series observadas, y medir la similitud o discrepancia entre dos series a partir de los propios datos o de alguna característica que pueda obtenerse a partir de ellos (“model free approaches”). La solución más sencilla consiste en tratar las series de tiempo como si fuesen vectores multidimensionales de datos independientes y utilizar como medida de distancia la distancia Euclídea

entre los datos observados. Tal y como ya señalaron Galeano y Peña (2000), esta métrica resulta invariante a cualquier transformación que modifique el orden de las observaciones a lo largo del tiempo, y por lo tanto no toma en cuenta la estructura de autocorrelación que caracteriza a las series temporales. Así, por ejemplo, sería lógico considerar similares los beneficios de dos empresas si siguen una evolución similar, independientemente de que el volumen de ventas de una de ellas duplique las ventas de la otra compañía. Una alternativa consiste en utilizar una métrica que tenga en cuenta dicha estructura, bien a partir de la correlación entre cada par de series (Bohte *et al.* (1980)) o bien a partir de las funciones de autocorrelación, autocorrelación parcial o autocorrelación inversa de cada una de las series observadas (Galeano y Peña (2000), Caiado, Crato y Peña (2006), Caiado y Crato (2005)). Aunque este tipo de distancias permitiría clasificar de forma eficiente series con diferentes estructuras de autocorrelación, los resultados obtenidos podrían depender del número de coeficientes considerados. Otros autores proponen realizar primero una transformación de los datos y utilizar después una métrica asociada. Así, por ejemplo, Bagnall, Janacek y Zhang (2003) sugieren realizar el cluster una vez discretizados los datos. A su vez, Agrawal, Faloutsos y Swami (1993) proponen como medida de distancia la distancia euclídea entre los primeros coeficientes de la transformada discreta de Fourier de las series de tiempo, mientras que Gavrilov *et al.* (2000) utilizan el análisis de componentes principales.

Como puede observarse, prácticamente todas las soluciones propuestas para el cluster de series temporales persiguen de un modo u otro reducir la alta dimensionalidad que caracteriza a este tipo de datos. En este sentido, el enfoque a través del dominio de frecuencias puede resultar especialmente interesante, y ha sido defendido por diversos autores. Es evidente que en muchas ocasiones las similitudes y diferencias entre series de tiempo pueden caracterizarse en términos de sus estructuras de covarianzas o, equivalentemente, de sus densidades espectrales, por lo que es lógico que el cluster de series de tiempo pueda realizarse a partir de distancias definidas en el dominio espectral. Así, por ejemplo, Shaw y King (1992) han abordado el problema del análisis cluster de series de tiempo utilizando como medida de distancia la distancia euclídea entre los espectros normalizados de las series observadas. En esta misma línea, Caiado, Crato y Peña (2006) han propuesto utilizar como medida de disparidad la distancia euclídea entre los periodogramas normalizados (o su logaritmo) asociados a las series de partida. Estos mismos autores proponen también utilizar la aproximación espectral a la medida de información de Kullback-Leibler, reemplazando las matrices de covarianzas por los periodogramas normalizados correspondientes. En este sentido, Kakizawa, Shumway y Taniguchi (1998) propusieron una medida de disparidad espectral que extiende algunas medidas de disparidad clásicas definidas en el dominio de tiempo, como la información discriminante de Kullback-Leibler o la medida de información de Chernoff.

Un enfoque distinto para abordar el problema del análisis cluster de series temporales es el propuesto recientemente por Alonso *et al.* (2006), donde se pone de manifiesto cómo el objetivo con el que se lleva a cabo la clasificación es esencial a la hora de determinar una medida de disparidad adecuada. Así, dichos autores abordan el problema del análisis cluster de series de tiempo cuando el interés se centra en clasificar las series de acuerdo con el valor que tomarán en un horizonte de tiempo determinado. Es obvio que en este caso la proximidad entre series no puede definirse únicamente a partir de su estructura de autocovarianzas, ni con un enfoque basado en modelos. Los autores proponen así un nuevo método cluster basado en las densidades de predicción de cada una de las series observadas en un momento dado de tiempo, utilizando estimadores tipo núcleo de la densidad y un procedimiento “sieve bootstrap”. Las diferencias entre cada par de densidades estimadas mediante técnicas bootstrap proporcionan una matriz de distancias que se utiliza como punto de partida para el análisis cluster.

Finalmente, trabajos recientes han abordado específicamente el problema de la clasificación de procesos no estacionarios. En particular, los procesos localmente no estacionarios, en los que el espectro asociado varía en el tiempo, han sido objeto de diversas investigaciones que han aportado soluciones diversas al análisis cluster de este tipo de procesos (Hirukawa (2006), Sakiyama (2002), Shumway (2003))

A pesar de la preponderancia, en los últimos años, de trabajos dedicados al análisis cluster de series de tiempo, se ha estudiado con poco detalle el comportamiento de cada una de estas distancias en diferentes contextos de clasificación y por lo tanto se desconoce, en situaciones concretas, cuál de los diferentes enfoques proporciona mejores resultados. Los trabajos de Kaplakis, Gada y Puttagunta (2001), Boets, De Cock y De Moor (2005) y Caiado, Crato y Peña (2006) son los únicos que conocemos en este sentido. Mientras que los dos primeros exploran el comportamiento de diferentes medidas de disparidad en el análisis cluster de procesos ARMA, el trabajo de Caiado, Crato y Peña (2006) se centra en la clasificación de procesos estacionarios y no estacionarios. En un trabajo anterior, Caiado y Crato (2005) habían investigado el comportamiento de diferentes métricas para discriminar entre un proceso lineal determinista y un “paseo aleatorio”. El mismo autor ha estudiado posteriormente el comportamiento de las mismas distancias para la clasificación de procesos estacionarios y no estacionarios con diferentes características de persistencia (Caiado (2006)).

En este capítulo se propondrán nuevas medidas, de corte no paramétrico, para abordar el análisis cluster de series temporales. Algunas de estas distancias se basan en la medida de disparidad espectral propuesta por Kakizawa, Shumway y Taniguchi (1998), y extienden al contexto del análisis cluster el procedimiento discriminante descrito en el Capítulo 2. Así, la distancia entre dos series de tiempo se definirá en términos de la discre-

pancia entre las densidades espectrales correspondientes, estimadas no paramétricamente mediante técnicas de regresión polinómica local. Se propondrán además otras medidas de discrepancia basadas en estadísticos de corte no paramétrico originalmente desarrollados para contrastar la igualdad del logaritmo del espectro de dos procesos estocásticos. Dada la flexibilidad de las técnicas de corte no paramétrico, cabe esperar que con estas distancias se mejoren los resultados de las medidas basadas en modelos, a la vez que se contrastará su comportamiento, en diferentes contextos de clasificación, con el de otras distancias definidas tanto en el dominio de tiempo como en el dominio espectral.

En la Sección 4.2 se realizará una revisión más pormenorizada de algunas de las distancias entre series de tiempo que han sido propuestas por otros autores para abordar el análisis cluster de procesos estocásticos. En la Sección 4.3 se introducirán nuevas medidas de corte no paramétrico para evaluar la disparidad entre este tipo de datos, y se indicará su utilización en el análisis cluster de procesos estocásticos. A continuación, en la Sección 4.4 se describirán los resultados de un estudio de simulación llevado a cabo para comparar el comportamiento de diversas distancias en la clasificación de series temporales. Esta Sección se estructurará en cinco apartados: en primer lugar, se describirá el procedimiento experimental y a continuación se estudiará la clasificación de series de tiempo en tres contextos diferentes: (a) clasificación de procesos como estacionarios y no estacionarios, (b) clasificación de diferentes procesos ARMA y (c) clasificación de procesos no lineales. Finalmente, se resumirán algunas conclusiones del estudio de simulación. La última sección del capítulo se dedicará a analizar el comportamiento de las medidas propuestas para la clasificación de un conjunto de datos reales referidos a los resultados del electrocardiograma de una serie de pacientes aquejados de diferente patología.

Los resultados presentados en este capítulo pueden verse en Pértega y Vilar (2007).

4.2. Algunas medidas de disparidad entre series de tiempo

Tal y como se ha señalado en la sección anterior, uno de los aspectos clave del análisis cluster es la función utilizada para medir la similitud entre cada par de datos que se van a clasificar. En esta sección se describirán algunas de las medidas de disimilitud entre series de tiempo que han sido utilizadas en otros trabajos para el análisis cluster de procesos estocásticos.

Sean $\mathbf{X}_n = (X_1, \dots, X_n)$ e $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ realizaciones parciales de dos procesos escalares $X = \{X_t, t \in \mathbb{Z}\}$ e $Y = \{Y_t, t \in \mathbb{Z}\}$, respectivamente.

Una forma sencilla de valorar la discrepancia o distancia entre dos series de tiempo

consiste en considerar la distancia Euclídea entre los datos observados:

$$d_E(X, Y) = \left\{ \sum_{t=1}^n (X_t - Y_t)^2 \right\}^{1/2}. \quad (4.1)$$

La distancia Euclídea se basa en medir directamente la cercanía de los valores observados, independientemente de la estructura de autocorrelación inherente a las series de tiempo. Así, los datos se tratan como si fuesen independientes de modo que d_E resulta ser una medida invariante a permutaciones en el tiempo. Por lo tanto, es obvio que no puede ser considerada una buena medida de disparidad entre series temporales, tal y como señalan Galeano y Peña (2000).

Una alternativa para tener en cuenta la estructura temporal de las series de tiempo consiste en medir la discrepancia entre dos series a partir de las funciones de autocorrelación estimadas (ACFs) (Galeano y Peña (2000)). Así, sean $\hat{\rho}_X = (\hat{\rho}_{1,X}, \dots, \hat{\rho}_{L,X})^t$ y $\hat{\rho}_Y = (\hat{\rho}_{1,Y}, \dots, \hat{\rho}_{L,Y})^t$ los vectores de coeficientes de autocorrelación estimados a partir de las series de tiempo \mathbf{X}_n e \mathbf{Y}_n , respectivamente, para algún L tal que $\hat{\rho}_{i,X} \approx 0$ y $\hat{\rho}_{i,Y} \approx 0$, para $i > L$. De esta forma, se puede definir la distancia entre las series como:

$$d_{ACF}(X, Y) = \{(\hat{\rho}_X - \hat{\rho}_Y)^t \mathbf{\Omega} (\hat{\rho}_X - \hat{\rho}_Y)\}^{1/2}, \quad (4.2)$$

donde $\mathbf{\Omega}$ es alguna matriz de pesos. Algunas posibles elecciones para los pesos en $\mathbf{\Omega}$ son (Caiado, Crato y Peña (2006)):

(i) Pesos uniformes, tomando $\mathbf{\Omega} = \mathbf{I}$. En este caso, la distancia d_{ACF} consiste simplemente en la distancia Euclídea entre las funciones de autocorrelación estimadas:

$$d_{ACFU}(X, Y) = \left\{ \sum_{i=1}^L (\hat{\rho}_{i,X} - \hat{\rho}_{i,Y})^2 \right\}^{1/2}. \quad (4.3)$$

(ii) Pesos geométricos, disminuyendo con el retardo de las autocorrelaciones, de modo que d_{ACF} vendrá dada por:

$$d_{ACFG}(X, Y) = \left\{ \sum_{i=1}^L p(1-p)^i (\hat{\rho}_{i,X} - \hat{\rho}_{i,Y})^2 \right\}^{1/2}, \quad \text{con } 0 < p < 1. \quad (4.4)$$

(ii) La distancia de Mahalanobis entre los coeficientes de autocorrelación (d_{ACFM}), que resulta de tomar $\mathbf{\Omega} = (\text{cov}(\hat{\rho}^{-1}))$, la inversa de la matriz de covarianzas de los coeficientes de autocorrelación, calculada según la fórmula de Bartlett truncada (Brockwell y Davis (1991)).

Igualmente, se han introducido distancias similares a partir de los coeficientes de autocorrelación parcial (PACF) de las series de tiempo consideradas, en lugar de los coeficientes

de autocorrelación simple (Caiado, Crato y Peña (2006)). Así, se define la distancia entre dos series de tiempo como:

$$d_{PACF}(X, Y) = \left\{ \left(\widehat{\phi}_X - \widehat{\phi}_Y \right)^t \mathbf{\Omega} \left(\widehat{\phi}_X - \widehat{\phi}_Y \right) \right\}^{1/2}, \quad (4.5)$$

donde $\widehat{\phi}_{ii}$ son los coeficientes de autocorrelación parcial estimados de cada una de las series de tiempo consideradas y $\mathbf{\Omega}$ es una matriz de pesos que puede ser definida como en el caso anterior. Así, en particular, podrá considerarse la distancia Euclídea entre los vectores de coeficientes de autocorrelación parcial estimados con pesos uniformes (d_{PACFU}), o con pesos geométricos (d_{PACFG}). De igual forma, se han definido distancias similares entre las funciones de autocorrelación inversa propuestas por Cleveland (1972) y desarrolladas por Chatfield (1979) (Caiado, Crato y Peña (2005)).

Todas las distancias anteriores se definen directamente a partir de los datos observados, sin exigir hipótesis adicionales sobre los procesos generadores. Tal y como se vio en la sección anterior, un enfoque alternativo se basa en asumir que cada una de las series de tiempo es generada a partir de un modelo paramétrico determinado, por ejemplo, a partir de un modelo ARIMA. Un ejemplo es la métrica propuesta por Piccolo (1990) para la clase de procesos ARIMA invertibles, definida como la distancia Euclídea entre los coeficientes de los operadores $AR(\infty)$ de la estructura ARIMA asociada a cada una de las series de tiempo consideradas:

$$d_{PIC}(X, Y) = \sqrt{\sum_{j=1}^{\infty} (\pi_{jx} - \pi_{jy})^2},$$

donde π_{jx} y π_{jy} son los coeficientes de los operadores $AR(\infty)$ de X e Y , respectivamente. En la práctica, en lugar de ajustar el modelo ARIMA correspondiente a cada serie de tiempo, se puede modelizar la estructura autorregresiva utilizando un criterio como el criterio de información de Akaike (AIC). Así, la distancia anterior puede calcularse como:

$$d_{PIC}(X, Y) = \left\{ \sum_{j=1}^k (\widehat{\pi}_{j,X} - \widehat{\pi}_{j,Y})^2 \right\}^{1/2}, \quad (4.6)$$

con $\widehat{\mathbf{\Pi}}_X = (\widehat{\pi}_{1,X}, \dots, \widehat{\pi}_{k_1,X})^t$ y $\widehat{\mathbf{\Pi}}_Y = (\widehat{\pi}_{1,Y}, \dots, \widehat{\pi}_{k_2,Y})^t$ los vectores de los parámetros $AR(k_1)$ y $AR(k_2)$ estimados de las series observadas \mathbf{X}_T e \mathbf{Y}_T , respectivamente, y $k = \max(k_1, k_2)$, donde k_1 y k_2 son el orden de los modelos AR ajustados a cada una de dichas series.

Basándose en esta distancia, Maharaj (1996) introdujo posteriormente otras dos medidas de disparidad en la clase de procesos ARMA, basándose en un test de hipótesis para determinar si dos series de tiempo procedían o no del mismo proceso generador. Más específicamente, el test propuesto viene dado como:

$$d_M(X, Y) = \sqrt{T} \left(\widehat{\mathbf{\Pi}}_X - \widehat{\mathbf{\Pi}}_Y \right)^t \widehat{\mathbf{V}}^{-1} \left(\widehat{\mathbf{\Pi}}_X - \widehat{\mathbf{\Pi}}_Y \right), \quad (4.7)$$

con $\widehat{\mathbf{V}}$ un estimador de $\mathbf{V} = \sigma_X^2 \mathbf{R}_X^{-1}(k) + \sigma_Y^2 \mathbf{R}_Y^{-1}(k)$, siendo σ_X^2 y σ_Y^2 las varianzas de los procesos ruido blanco asociados con \mathbf{X}_n e \mathbf{Y}_n , respectivamente, y \mathbf{R}_X y \mathbf{R}_Y las matrices de covarianza muestrales de ambas series.

Maharaj (1996) demostró que d_M , bajo la hipótesis nula de igualdad de los procesos generadores, sigue asintóticamente una distribución chi-cuadrado con k grados de libertad. Puesto que d_M , y también su p-valor asociado, satisfacen las condiciones de no negatividad y simetría, Maharaj (1996) propone utilizarlas como medidas de disparidad entre los procesos X e Y en el análisis cluster de series de tiempo.

Otras medidas de distancia definidas en el contexto del análisis cluster de procesos estocásticos han sido propuestas en el dominio espectral. Sean I_X e I_Y los periodogramas de las series de tiempo \mathbf{X}_n e \mathbf{Y}_n , respectivamente, evaluadas en las frecuencias de Fourier $\lambda_k = 2\pi k/n$, $k = 1, \dots, N$ con $N = [(n-1)/2]$. Caiado, Crato y Peña (2006) han considerado hasta tres distancias diferentes basándose en los datos de los periodogramas. En primer lugar, proponen utilizar como distancia entre dos series la distancia Euclídea entre las ordenadas de sus correspondientes periodogramas:

$$d_P(X, Y) = \frac{1}{N} \sqrt{\sum_{k=1}^N (I_X(\lambda_k) - I_Y(\lambda_k))^2}. \quad (4.8)$$

Los citados autores mencionan que si no se está interesado en la escala de los procesos que se desea clasificar, sino únicamente en su estructura de autocorrelación, resulta más adecuado trabajar con los periodogramas normalizados, reemplazando el periodograma $I(\lambda_k)$ por $NI(\lambda_k) = I(\lambda_k)/\widehat{\gamma}_0$, donde $\widehat{\gamma}_0$ es la varianza muestral de la serie de tiempo correspondiente. Así, proponen tomar como distancia entre cada par de series la distancia euclídea entre los periodogramas normalizados:

$$d_{NP}(X, Y) = \frac{1}{N} \sqrt{\sum_{k=1}^N (NI_X(\lambda_k) - NI_Y(\lambda_k))^2}. \quad (4.9)$$

Finalmente, y puesto que la varianza de las ordenadas del periodograma resulta proporcional al valor del espectro en las frecuencias correspondientes, Caiado, Crato y Peña (2006) precisan que tiene sentido trabajar con el logaritmo del periodograma normalizado:

$$d_{LNP}(X, Y) = \frac{1}{N} \sqrt{\sum_{k=1}^N (\log NI_X(\lambda_k) - \log NI_Y(\lambda_k))^2}, \quad (4.10)$$

y, obviamente, también con el logaritmo de los correspondientes periodogramas

$$d_{LP}(X, Y) = \frac{1}{N} \sqrt{\sum_{k=1}^N (\log I_X(\lambda_k) - \log I_Y(\lambda_k))^2}. \quad (4.11)$$

Dichos autores demostraron que estas métricas están relacionadas con las distancias basadas en los coeficientes de autocorrelación descritas anteriormente, de forma que se verifica que:

$$d_{NP}(X, Y) = (2\sqrt{n})d_{ACFU}(X, Y).$$

En ocasiones alguna de las distancias entre series de tiempo definidas en el dominio espectral resultan de una simple aproximación de medidas que han sido definidas previamente en el dominio de tiempo, y que se utilizan con el fin de evitar los graves problemas computacionales que pueden surgir al trabajar con series de gran longitud. Así, Caiaido, Crato y Peña (2006) proponen utilizar la aproximación espectral de la información discriminante de Kullback-Leibler:

$$d_{KLT D}(X, Y) = \text{tr}(\Sigma_X \Sigma_Y^{-1}) - \log(|\Sigma_X|/|\Sigma_Y|) - n,$$

donde Σ_X , Σ_Y denotan las matrices de covarianzas asociadas a las series \mathbf{X}_n e \mathbf{Y}_n , respectivamente. Dicha aproximación espectral se calcula a partir de las coordenadas del periodograma normalizado de cada una de las series:

$$d_{KLF D}(X, Y) = \sum_{k=1}^N \left[\frac{NI_X(\lambda_k)}{NI_Y(\lambda_k)} - \log \frac{NI_X(\lambda_k)}{NI_Y(\lambda_k)} - 1 \right].$$

En esta misma línea, la medida de disparidad espectral (2.10) estudiada en la Sección 2 en el contexto del análisis discriminante y propuesta por Kakizawa, Shumway y Taniguchi (1998)

$$D_W(f_X, f_Y) = \frac{1}{4\pi} \int_{-\pi}^{\pi} W(f_X(\lambda)f_Y^{-1}(\lambda))d\lambda$$

incluye, como casos particulares, a las aproximaciones espectrales de medidas clásicas de disparidad entre dos series de tiempo como la información discriminante de Kullback-Leibler o la información de Chernoff. Dichos autores proponen utilizar la distancia desarrollada para el análisis cluster de series temporales, utilizando una simple transformación para garantizar la simetría de la medida propuesta. Así, la distancia entre dos series de tiempo \mathbf{X}_n e \mathbf{Y}_n puede computarse como:

$$d_W(X, Y) = D_W(\hat{f}_X, \hat{f}_Y) + D_W(\hat{f}_Y, \hat{f}_X)$$

donde $\hat{f}_X(\cdot)$, $\hat{f}_Y(\cdot)$ denotan las densidades espectrales estimadas de los procesos X e Y , respectivamente, y $W(\cdot)$ es una función que satisface ciertas condiciones de regularidad necesarias para garantizar que $D_W(\cdot, \cdot)$ sea una cuasidistancia. En particular:, se exigirá que la función de divergencia $W(\cdot)$ sea tres veces continuamente diferenciable en $(0, \infty)$ con un único mínimo en 1 tal que $W(1) = 0$.

Cabe señalar que mientras alguna de las medidas anteriores han sido específicamente diseñadas para procesos estacionarios, como las distancias propuestas por Piccolo (1990)

y Maharaj (1996), aquellas distancias basadas en los coeficientes de autocorrelación o en el espectro de los series pueden calcularse para cualquier tipo de procesos. Es cierto que tanto los coeficientes de autocorrelación como el espectro se definen habitualmente para procesos estacionarios, por lo que las métricas basadas en las autocorrelaciones muestrales y en los periodogramas correspondientes asumen implícitamente que las series sobre las que se trabaja son estacionarias. No obstante, su definición puede ser extendida a procesos integrados (Peña y Poncela (2006)) y en este sentido, las distancias descritas podrán utilizarse también en la clasificación de otro tipo de procesos.

4.3. Una vía no paramétrica para el análisis cluster de series de tiempo

En esta sección se introducirán nuevas medidas de corte no paramétrico que pueden utilizarse en el análisis cluster de series temporales y que no han sido consideradas hasta el momento en este contexto.

En primer lugar, se propondrá utilizar la medida de disparidad espectral (2.10) propuesta en el Capítulo 2 en el contexto del análisis discriminante, y extender su utilización al análisis cluster de series temporales. Según esta medida, la discrepancia entre dos series de tiempo puede evaluarse según:

$$D_W(X, Y) = \frac{1}{4\pi} \int_{-\pi}^{\pi} W(f_X(\lambda)f_Y^{-1}(\lambda))d\lambda,$$

donde $\hat{f}_X(\cdot)$, $\hat{f}_Y(\cdot)$ denotan las densidades espectrales estimadas de los procesos X e Y , respectivamente, y $W(\cdot)$ es una función que satisface las condiciones de regularidad necesarias para que $D_W(X, Y)$ sea una cuasi-distancia.

Puesto que la medida $D_W(\cdot, \cdot)$ no resultar ser, de modo general, una medida simétrica, se requiere realizar una alguna modificación para poder ser utilizada como punto de partida en el análisis cluster. En particular, si la función de divergencia $W(\cdot)$ se modifica según

$$\tilde{W}(x) = W(x) + W(x^{-1}), \tag{4.12}$$

la nueva medida de disparidad espectral basada en $\tilde{W}(\cdot)$ vendrá dada por la expresión:

$$\begin{aligned} d_W(X, Y) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \tilde{W}(\hat{f}_X(\lambda)\hat{f}_Y^{-1}(\lambda)) = \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\hat{f}_X(\lambda)\hat{f}_Y^{-1}(\lambda))d\lambda + \frac{1}{4\pi} \int_{-\pi}^{\pi} W(\hat{f}_Y(\lambda)\hat{f}_X^{-1}(\lambda))d\lambda = \\ &= D_W(\hat{f}_X, \hat{f}_Y) + D_W(\hat{f}_Y, \hat{f}_X), \end{aligned}$$

resultando así una medida simétrica que puede utilizarse en el análisis cluster de series de tiempo. En particular, se propone utilizar los estimadores espectrales desarrollados en Fan y Kreutzberger (1998) para aproximar la densidad espectral de cada proceso, denotando la medida de disparidad resultante de la siguiente forma:

- (i) $d_{W(DLS)}$, cuando el espectro se estime utilizando el suavizador lineal local del periodograma, obtenido mediante mínimos cuadrados locales (\hat{f}_{DLS})
- (ii) $d_{W(LS)}$, cuando el espectro se estime mediante la exponencial del suavizador lineal local del logaritmo del periodograma, utilizando también mínimos cuadrados locales (\hat{f}_{LS})
- (iii) $d_{W(LK)}$, cuando la estimación de la densidad espectral se realice mediante la exponencial del estimador lineal local del log-periodograma, obtenido utilizando el criterio de máxima verosimilitud local en lugar de mínimos cuadrados locales (\hat{f}_{LK})

Puesto que la aplicación de los algoritmos cluster tradicionales únicamente requiere de la definición de una medida que pueda resultar apropiada para medir la disparidad entre dos series de tiempo, de acuerdo con el objetivo de la clasificación, de modo general cualquier estadístico originalmente diseñado para contrastar la igualdad de los procesos generadores de dos series de tiempo podría utilizarse como una medida a partir de la cual abordar el análisis cluster de series temporales, tal y como propone Maharaj (1996, 2000). Sin embargo, se trata de una alternativa poco explorada en la literatura. En este trabajo, proponemos utilizar como medida de discrepancia en el análisis cluster de procesos estocásticos dos nuevas distancias definidas a partir de estadísticos originalmente diseñados para contrastar la igualdad de la densidad espectral (más concretamente, del logaritmo de la densidad espectral) de dos procesos. En particular, se plantea el siguiente contraste de hipótesis:

$$\begin{aligned} H_0 : m_X(\cdot) &= m_Y(\cdot) \\ H_1 : m_X(\cdot) &\neq m_Y(\cdot) \end{aligned} \tag{4.13}$$

con $m_X(\lambda) = \log(f_X(\lambda))$ y $m_Y(\lambda) = \log(f_Y(\lambda))$ el logaritmo de la densidad espectral de los procesos X e Y , respectivamente.

En primer lugar, nos centraremos en el test de razón de verosimilitud generalizada propuesto originalmente por Fan y Zhang (2004) para contrastar si la densidad espectral $f_X(\cdot)$ asociada a una serie de tiempo \mathbf{X}_n pertenece o no a una determinada familia paramétrica $\{f_\theta(\cdot), \theta \in \Theta\}$. Dichos autores abordan por lo tanto el problema del contraste de hipótesis:

$$\begin{aligned} H_0 : f_X(\cdot) &= f_\theta(\cdot) \\ H_1 : f_X(\cdot) &\neq f_\theta(\cdot), \end{aligned}$$

que puede ser expresado de forma equivalente a través del logaritmo de la densidad espectral correspondiente:

$$\begin{aligned} H_0 : m_X(\cdot) &= m_\theta(\cdot) \\ H_1 : m_X(\cdot) &\neq m_\theta(\cdot). \end{aligned}$$

Si X es un proceso lineal Gaussiano, las ordenadas del periodograma evaluadas sobre las frecuencias de Fourier satisfacen, como vimos en (1.23), el siguiente modelo de regresión heterocedástico:

$$I_X(\lambda_k) = f_X(\lambda_k)V_k + R_{n,k},$$

donde $R_{n,k}$ denota un término asintóticamente despreciable y las V_k son variables aleatorias independientes e idénticamente distribuidas según una distribución exponencial estándar para todo $k \neq 0$. Por lo tanto, tomando logaritmos:

$$Y_k = \log I_X(\lambda_k) = m_X(\lambda_k) + \varepsilon_k + r_k,$$

donde ahora $m_X(\lambda_k) = \log(f_X(\lambda_k))$, $\varepsilon_k = \log(V_k)$ son variables aleatorias i.i.d. con función de densidad $f_\varepsilon(x) = \exp(-\exp(x) + x)$ y $r_k = \log(1 + R_k/f(\lambda_k))$ denota un término asintóticamente despreciable.

La función de log-verosimilitud asociada resulta ser, por lo tanto:

$$\sum_{k=1}^N \{-\exp(Y_k - m(\lambda_k)) + Y_k - m(\lambda_k)\} K_{h_n}(\lambda_k - \lambda),$$

obtenida utilizando una función núcleo K y ventana h_n .

Así, el estimador de máxima verosimilitud local para el logaritmo de la densidad espectral, $\hat{m}_{LK}(\lambda)$, vendrá dado por el valor \hat{a} para el que se maximiza la expresión:

$$\sum_{k=1}^N \{-\exp(Y_k - a - b(\lambda_k - \lambda)) + Y_k - a - b(\lambda_k - \lambda)\} K_{h_n}(\lambda_k - \lambda).$$

Bajo la hipótesis nula, la función de log-verosimilitud asociada es igual a:

$$\sum_{k=1}^N \{-\exp(Y_k - m_\theta(\lambda_k)) + Y_k - m_\theta(\lambda_k)\},$$

de modo que el estimador de máxima verosimilitud de θ será el valor $\hat{\theta}$ que maximiza la expresión anterior.

El test de razón de verosimilitud generalizada propuesto por Fan y Zhang (2004) se define entonces como el cociente de la verosimilitud bajo cada una de las hipótesis:

$$T_{LK} = \sum_{k=1}^N \{\exp(Y_k - \hat{m}_\theta(\lambda_k)) + \hat{m}_\theta - \exp(Y_k - \hat{m}_{LK}(\lambda_k)) - \hat{m}_{LK}(\lambda_k)\},$$

de modo que la hipótesis nula se rechazará para valores grandes del estadístico T_{LK} .

Una ligera modificación de dicho estadístico permite ajustar el procedimiento al caso del contraste de la igualdad del logaritmo de los espectros dado en (4.13). En este caso, se verifica que:

$$Z_k = \log(I_X(\lambda_k)) - \log(I_Y(\lambda_k)) = \mu(\lambda_k) + R_k^* + r_k^*,$$

donde ahora $\mu(\lambda) = m_X(\lambda) - m_Y(\lambda)$, R_k^* denota una secuencia de variables aleatorias i.i.d. con función de densidad $h(x) = \exp(x)/(\exp(x) + 1)^2$ y r_k^* es un término asintóticamente despreciable.

La función de log-verosimilitud generalizada asociada al modelo de regresión anterior viene dada por:

$$\sum_{k=1}^N \{Z_k - \mu(\lambda_k) - 2 \log(\exp(Z_k - \mu(\lambda_k)) + 1)\},$$

que, bajo la hipótesis nula resulta igual a:

$$\sum_{k=1}^N \{Z_k - 2 \log(\exp(Z_k) + 1)\}.$$

De forma que el contraste (4.13) puede resolverse mediante el estadístico basado en la razón de verosimilitudes:

$$\begin{aligned} d_{GLK}(X, Y) &= \sum_{k=1}^N \left[Z_k - \hat{\mu}(\lambda_k) - 2 \log \left(1 + e^{\{Z_k - \hat{\mu}(\lambda_k)\}} \right) \right] - \\ &\quad - \sum_{k=1}^N \left[Z_k - 2 \log (1 + e^{Z_k}) \right], \end{aligned} \quad (4.14)$$

donde $Z_k = \log(I_X(\lambda_k)) - \log(I_Y(\lambda_k))$, $\mu(\lambda_k) = m_X(\lambda_k) - m_Y(\lambda_k)$ y $\hat{\mu}(\lambda_k)$ es el estimador de máxima log-verosimilitud local de $\mu(\lambda_k)$ obtenido mediante un ajuste lineal local.

Finalmente, el contraste de hipótesis (4.13) puede abordarse también mediante un procedimiento clásico, a partir del estadístico basado en la distancia funcional de tipo Cramér-von-Mises entre los estimadores no paramétricos del logaritmo de los espectros, $m_X(\lambda)$ y $m_Y(\lambda)$. En particular, se considerará la distancia dada por

$$d_{CM}(X, Y) = \int (\hat{m}_X(\lambda) - \hat{m}_Y(\lambda))^2 d\lambda, \quad (4.15)$$

donde $\hat{m}_X(\lambda)$ y $\hat{m}_Y(\lambda)$ denotan un estimador del logaritmo del espectro de cada uno de los procesos. En particular, se podrán utilizar cualquiera de los estimadores espectrales propuestos en Fan y Kreutzberger (1998).

Los estadísticos anteriores (4.14) y 4.15) satisfacen las condiciones generales de una distancia y, por lo tanto, pueden ser consideradas como punto de partida para el análisis cluster de un conjunto de series temporales.

4.4. Estudio de simulación

En este capítulo se presentan los resultados de un estudio de simulación llevado a cabo para analizar y comparar el comportamiento de las medidas de similitud descritas en la Sección 4.3 al trabajar sobre muestras finitas, cuando éstas son utilizadas en el análisis cluster de diferentes clases de procesos estocásticos. Asimismo, se compararán los resultados obtenidos con los alcanzados con otras medidas utilizadas en la literatura para el análisis cluster de series de tiempo, según lo descrito en la Sección 4.2.

Es lógico que si una medida de disparidad guarda relación con alguna de las propiedades de los procesos que se quieren clasificar se espere que los resultados obtenidos tras realizar el análisis cluster a partir de esta medida sean mejores que los obtenidos al utilizar cualquier otra distancia. Es importante tener en cuenta además que el análisis cluster es un proceso de clasificación no supervisada, de forma que en la práctica el análisis cluster se lleva a cabo sin conocer el grupo de pertenencia de las series con las que se trabaja. Por lo tanto, es esencial analizar el comportamiento de las distintas medidas de disimilitud en diferentes contextos de clasificación, a fin de poder valorar la idoneidad de los resultados que podamos obtener en un futuro. En particular, se han considerado tres contextos de clasificación diferentes: (i) en primer lugar, se planteó el problema de distinguir entre procesos estacionarios frente a procesos no estacionarios, (ii) en segundo lugar, se estudió la clasificación de diferentes tipos de procesos ARMA estacionarios, (iii) finalmente, se abordó el problema de clasificar diferentes modelos de series de tiempo no lineales. Consideramos que estos tres contextos de clasificación son adecuados para valorar de una forma completa el comportamiento de las medidas de disparidad propuestas, por su generalidad y capacidad para representar un amplio espectro de las situaciones que se pueden dar en la práctica.

Esta sección se estructura a lo largo de cinco apartados. En primer lugar, se ofrece una descripción general del estudio de simulación que se ha llevado a cabo. A continuación, se procede a describir los resultados de las simulaciones para cada uno de los tres contextos analizados. Finalmente, se realizará una discusión de los resultados alcanzados y se expondrán las conclusiones que se han derivado de este estudio de simulación.

4.4.1. Procedimiento experimental

En cada uno de los tres contextos de clasificación estudiados se siguió un mismo esquema para llevar a cabo las simulaciones, tal y como se detalla a continuación:

1. Especificación de los procesos a clasificar.

Para cada uno de los tres contextos considerados, se determinaron una serie de procesos que se sometieron a clasificación. En general, se han seleccionado modelos que dificulten las labores de clasificación, es decir, modelos con índices de separabilidad bajos. Así, se ha decidido utilizar modelos previamente considerados en otros trabajos siguiendo los argumentos que se indican en cada sección.

2. Generación de las series de tiempo.

Una vez determinados los modelos, se obtuvieron $s \geq 1$ realizaciones de longitud n de cada uno de ellos, tomando n diferentes valores. Este paso se repitió un número T de veces, de forma que de cada proceso se disponía de $T \times s$ realizaciones convenientemente etiquetadas según el proceso de origen.

3. Cálculo de la matriz de distancias.

A partir de cada conjunto de datos, en cada iteración se obtuvieron una serie de matrices de disimilitud entre las series observadas (una matriz por cada una de las distancias consideradas). Para ello, la disparidad entre cada par de series de tiempo se calculó utilizando diferentes métricas que se resumen a continuación:

- (a) Distancia Euclídea entre los datos observados, d_E , dada en (4.1).
- (b) Distancias basadas en los coeficientes de autocorrelación: la distancia Euclídea entre los vectores de coeficientes de autocorrelación, d_{ACFU} , dada en (4.3), y la distancia Euclídea entre los coeficientes de autocorrelación con pesos geométricos disminuyendo con cada retardo, d_{ACFG} , dada en (4.4). En ambos casos el número de autocorrelaciones consideradas fue $L = 10$.
- (c) Distancias basadas en los coeficientes de autocorrelación parcial, d_{PACFU} y d_{PACFG} , calculadas como d_{ACFU} y d_{ACFG} a partir de los coeficientes de autocorrelación parcial en lugar de los coeficientes de autocorrelación. También en este caso se tomó $L = 10$. Los pesos utilizados para el cómputo tanto de d_{ACFG} como de d_{PACFG} se tomaron con $p = 0,05$.
- (d) Distancias propuestas por Piccolo (1990) y Maharaaj (1996), d_{PIC} y d_M , descritas en (4.6) y (4.7). En ambos casos, cada una de las series de tiempo generadas se ajustó mediante un modelo autorregresivo $AR(k)$, seleccionando el orden k (hasta un máximo $k = 10$) de acuerdo con el Criterio de Información de Akaike. Los estimadores de los parámetros autorregresivos se utilizaron para el cómputo de ambas distancias.
- (e) Distancias basadas en los periodogramas. En particular, se consideró la distancia Euclídea entre: las ordenadas de los periodogramas (d_P), las ordenadas de los

periodogramas normalizados (d_{NP}), el logaritmo de los periodogramas (d_{LP}) y el logaritmo de los periodogramas normalizados (d_{LNP}).

- (f) Medidas de disparidad espectral definidas como d_W en (4.13), donde las densidades espectrales se estimaron mediante el periodograma suavizado por mínimos cuadrados locales ($d_{W(DLS)}$), el log-periodograma suavizado por mínimos cuadrados locales ($d_{W(LS)}$) y el log-periodograma suavizado mediante máxima verosimilitud local ($d_{W(LK)}$). La función de divergencia W se tomó siempre igual a $W(x) = \log(\alpha + (1 - \alpha)) - \alpha \log(x)$, con $\alpha = 0,5$ y modificada de acuerdo con (4.12) para obtener una versión simetrizada de la medida de disparidad. En todos los casos, se utilizó el núcleo de Epanechnikov y, en la estimación del espectro de cada una de las series de tiempo, la ventana se determinó mediante validación cruzada sobre un rango adecuado de valores.
- (g) Distancia basada en la razón de verosimilitud generalizada, d_{GLK} y distancia funcional tipo Cramer-von-Mises, d_{CM} , definidas en (4.14) y (4.15), respectivamente. Igual que con las distancias basadas en d_W , se utilizaron los estimadores espectrales de máxima verosimilitud local para el cómputo tanto de d_{GLK} como d_{CM} , con el núcleo de Epanechnikov y el selector de ventana basado en la función de validación cruzada.

4. Aplicación de un algoritmo cluster.

A continuación, la matriz de distancias obtenida con cada una de las métricas utilizadas se procesó mediante un algoritmo cluster jerarquizado para obtener el correspondiente dendograma. Este paso se realizó con cada uno de los conjuntos de datos utilizando cuatro métodos diferentes: enlace completo, enlace simple, enlace promedio y método de Ward.

5. Evaluación de los resultados de clasificación.

Los resultados del análisis cluster se evaluaron comparando las soluciones cluster obtenidas experimentalmente con las verdaderas clases de pertenencia. Esta comparación se realizó de diferentes modos dependiendo del contexto de clasificación considerado y será detallado por lo tanto a lo largo de las siguientes secciones.

4.4.2. Clasificación de series como estacionarias o no estacionarias.

En un estudio reciente, Caiado, Crato y Peña (2006) han descrito los resultados de un estudio de simulación diseñado para comparar el comportamiento de diferentes medidas de similitud a la hora de clasificar series de tiempo como estacionarias o no estacionarias. El primer objetivo de nuestro estudio de simulación consistió en extender el trabajo de

Caiado, Crato y Peña (2006) incluyendo, además de las medidas allí consideradas, las medidas de similitud no paramétricas propuestas en la Sección 4.3 de esta memoria.

Considérese un proceso $ARIMA(p, d, q)$ definido según

$$\phi(B)(1 - B)^d M_t = \theta(B)\omega_t, \quad t = 0, \pm 1, \dots,$$

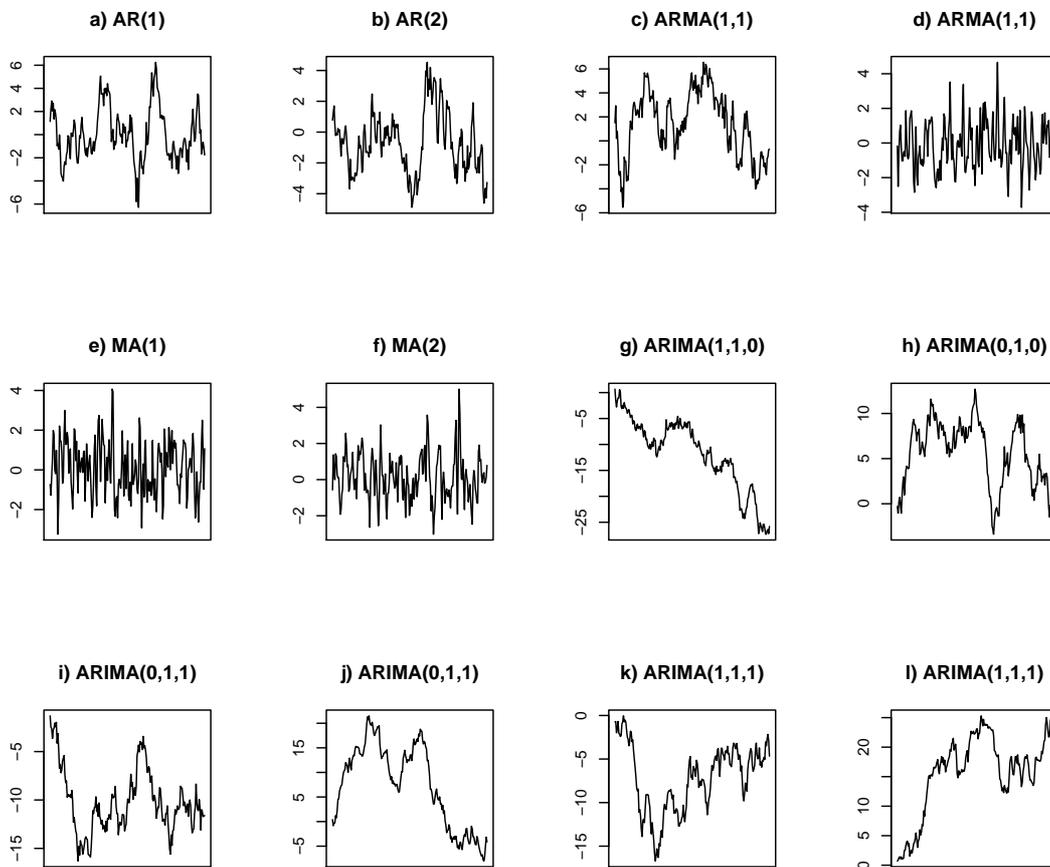
donde B es el operador retardo tal que $B^r M_t = M_{t-r}$, $\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$ es el operador autorregresivo de orden p , $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ es el operador de medias móviles de orden q , d es el orden de diferenciación (de modo que $d = 0$ si el proceso es estacionario y $d \geq 1$ si el proceso es no estacionario) y w_t es una secuencia de variables aleatorias incorreladas con media y varianza constante. Se asume que cada proceso $N(t) = (1 - B)^d M_t$ es causal e invertible. Al igual que en Caiado, Crato y Peña (2006), se obtuvo una realización de cada uno de los siguientes doce procesos ARIMA, seis de ellos estacionarios ((a)-(f)) y seis no estacionarios ((g)-(l)):

(a) AR(1)	$\phi_1 = 0,9$	(g) ARIMA(1, 1, 0)	$\phi_1 = -0,1$
(b) AR(2)	$\phi_1 = 0,95, \phi_2 = -0,1$	(h) ARIMA(0, 1, 0)	
(c) ARMA(1, 1)	$\phi_1 = 0,95, \theta_1 = 0,1$	(i) ARIMA(0, 1, 1)	$\theta_1 = 0,1$
(d) ARMA(1, 1)	$\phi_1 = -0,1, \theta_1 = -0,95$	(j) ARIMA(0, 1, 1)	$\theta_1 = -0,1$
(e) MA(1)	$\theta_1 = -0,9$	(k) ARIMA(1, 1, 1)	$\phi_1 = 0,1, \theta_1 = -0,1$
(f) MA(2)	$\theta_1 = -0,95, \theta_2 = -0,1$	(l) ARIMA(1, 1, 1)	$\phi_1 = 0,05, \theta_1 = -0,05$

En todos los casos el proceso de error se consideró ruido blanco con media cero y varianza unidad. Tal y como indican Caiado, Crato y Peña (2006), la elección de tales procesos se realizó con la finalidad de dificultar la clasificación de las series de tiempo en estacionarias o no estacionarias, al tener todos ellos valores de los parámetros próximos a los de un proceso ruido blanco integrado. En la Figura 4.1 se observa la forma típica de una realización aleatoria de dichos modelos.

Una vez generadas las doce series de tiempo, se procedió al análisis cluster, según lo descrito en la Sección 4.4.1, considerándose en cada dendograma la solución final de dos grupos. Uno de ellos fue definido como el “cluster estacionario” (aquel que englobaba a un mayor número de realizaciones de procesos estacionarios) y el otro como “no estacionario”. Como criterio de evaluación del cluster se tomó el porcentaje de éxitos en la clasificación, es decir, el porcentaje de series de tiempo clasificadas en el cluster correcto de acuerdo con su procedencia. Este proceso fue replicado $T = 300$ veces, de modo que el porcentaje de éxitos se promedió a lo largo de todas las iteraciones. Los resultados obtenidos utilizando

Figura 4.1: Series de tiempo generadas de los procesos estacionarios y no estacionarios (a)-(l).



el método del enlace completo con series de longitud n , para diferentes valores de n (50, 200 y 500) se muestran en la Tabla 4.1.

Centraremos la atención, en primer lugar, en los resultados obtenidos al clasificar series de longitud $n = 200$. Como puede observarse, existe una serie de métricas con un comportamiento claramente superior al del resto de distancias. Este grupo estaría formado por las distancias basadas en los coeficientes de autocorrelación, d_{ACFU} y d_{ACFG} , y aquellas basadas en el periodograma normalizado, d_{NPF} y d_{LNPF} , todas ellas con porcentajes de éxito por encima del 80 % (entre el 82 % y el 84 %). En este mismo grupo podrían incluirse las distancias no paramétricas evaluadas sobre el rango de frecuencias más bajas, con porcentajes de éxito en torno al 80 %. Tal y como era de esperar, las distancias euclídeas entre los datos observados, d_E y entre las coordenadas de los periodogramas, d_P , mostraron los peores resultados, con porcentajes de éxito inferiores al 67 % en ambos casos.

Tabla 4.1: Porcentaje de éxito en la clasificación de las series (a)-(l) como estacionarias y no estacionarias en $T = 300$ iteraciones. n es la longitud de las series. Las frecuencias bajas corresponden a las ordenadas 1 a \sqrt{n} . Las frecuencias altas a las ordenadas $\sqrt{n} + 1$ a $[n/2]$.

Medida	n			Medida	n		
	50	200	500		50	200	500
<i>Distancia Euclídea</i>				<i>No paramétricas</i>			
d_E	65.90	66.90	68.20	$d_{W(DLS)}$	73.50	71.00	72.40
<i>Autocorrelaciones</i>				Freq. bajas	78.06	80.92	82.65
d_{ACFU}	75.40	84.10	83.50	Freq. altas	62.28	65.92	71.43
d_{ACFG}	76.00	83.20	82.10	$d_{W(LS)}$	67.10	70.70	72.60
d_{PACFU}	74.40	75.00	75.00	Freq. bajas	71.83	79.81	80.10
d_{PACFG}	74.40	75.00	75.00	Freq. altas	64.42	67.44	71.94
<i>Periodogramas</i>				$d_{W(LK)}$	69.30	71.50	71.90
d_P	66.60	65.80	65.50	Freq. bajas	75.80	79.50	84.00
d_{LP}	66.00	73.10	74.80	Freq. altas	63.50	68.10	72.60
d_{NLP}	72.00	81.80	82.80	d_{GLK}	63.80	70.72	73.60
d_{LNP}	70.00	84.20	94.40	Freq. bajas	63.83	79.08	80.8
Freq. bajas	64.70	73.80	78.60	Freq. altas	62.50	68.00	72.60
Freq. altas	69.20	83.80	95.10	d_{CM}	69.50	72.00	74.50
<i>Basadas en modelos</i>				Freq. bajas	75.89	79.89	85.90
d_{PIC}	69.60	74.90	75.00	Freq. altas	63.11	68.89	72.10
d_M	71.80	75.00	75.00				

Esta baja tasa de éxito viene a corroborar la importancia de la elección de una medida de disparidad adecuada para la clasificación de series de tiempo. Finalmente, todas las otras distancias consideradas proporcionaron resultados similares entre ellos, con porcentajes de éxito entre el 70 % y el 75 %.

En general, los porcentajes de clasificación correcta aumentaron cuando se consideraron series de mayor longitud ($n = 500$). Los resultados obtenidos con la distancia euclídea entre el logaritmo de los periodogramas normalizados, d_{LNP} , fueron especialmente buenos, alcanzando una tasa de éxito del 95 %. También las distancias no paramétricas $d_{W(DLS)}$, $d_{W(LK)}$ y d_{CM} evaluadas sobre el rango de frecuencias más bajas mejoraron sustancialmente, obteniendo porcentajes de éxito de un 82,65 %, 84 % y un 85,9 %, respectivamente (los mejores tras d_{LNP}). Por el contrario, las distancias basadas en los coeficientes de autocorrelación (d_{ACFU} y d_{ACFG}) obtuvieron tasas de éxito ligeramente inferiores que las obtenidas con $n = 200$. Esto puede ser debido a que el número de autocorrelaciones consideradas se mantuvo en $L = 10$, y los coeficientes autorregresivos de los modelos considerados (tanto de los estacionarios como de los no estacionarios) son similares para los

primeros retardos.

A su vez, los resultados alcanzados con $n = 50$ fueron bastante peores para todas las distancias, con porcentajes de éxito por debajo del 79% en todos los casos. Destacar que las distancias no paramétricas $d_{W(DLS)}$, $d_{W(LK)}$ y d_{CM} (evaluadas sobre el rango de frecuencias más bajas) se situaron de nuevo entre las métricas con mejores resultados, con porcentaje de éxito por encima de d_{ACFU} y d_{LNP} .

Los resultados obtenidos tras el estudio de simulación son consistentes con lo reflejado por Caiado, Crato y Peña (2006). En su estudio, la distancia entre el logaritmo de los periodogramas normalizados, d_{LNP} , obtuvo muy buenos resultados, al igual que las métricas basadas en los coeficientes de autocorrelación tomando $L = n/10$.

Resulta especialmente interesante analizar lo que ocurre con las distancias de corte no paramétrico, así como con la distancia propuesta por Maharaj (1996), ya que ninguna de ellas fueron consideradas en el trabajo de Caiado, Crato y Peña (2006). Tal y como cabía esperar, la distancia de Maharaj, d_M , no proporciona buenos resultados en este contexto de clasificación, obteniendo porcentajes de éxito que no superaron el 75% independientemente de la longitud de las series consideradas. Este mal comportamiento puede deberse a las características particulares de los procesos que se han considerado. Los modelos ARMA etiquetados como (a), (b) y (c) y los modelos ARIMA (g)-(l) tienen secuencias autorregresivas π_j muy similares para los primeros retardos. De este modo, es lógico que la distancia de Maharaj, que se basa en estimaciones de estas secuencias, sea incapaz de discriminar entre estos dos grupos de modelos. Tras inspeccionar los dendogramas obtenidos utilizando la distancia d_M en cada una de las iteraciones esta explicación parece plausible, ya que los resultados muestran típicamente una solución de tres clusters: $\{(d),(e),(f)\}$, $\{(a),(b),(c)\}$ y el cluster no estacionario $\{(g)-(l)\}$. Esta peculiaridad fue también reflejada por Caiado, Crato y Peña (2006) en relación con la distancia propuesta por Piccolo (1990).

Ninguna de las medidas de corte no paramétrico se comportó de forma plenamente satisfactoria. Las densidades espectrales de los procesos considerados resultaron ser altamente asimétricas, tomando valores muy altos en un pequeño rango de frecuencias menores, y valores más bajos para frecuencias altas. Así, los estimadores no paramétricos del espectro parecen tender a sobresuavizar la densidad espectral justo en el rango de frecuencias bajas, donde se concentran las principales diferencias. Este argumento justifica además el buen comportamiento de estas distancias cuando se evaluaron únicamente sobre las frecuencias menores.

4.4.3. Clasificación de procesos ARMA

A continuación se realizó un nuevo estudio experimental diseñado para explorar el comportamiento de las diferentes medidas de disparidad consideradas en el análisis cluster de procesos ARMA estacionarios. Para ello se consideraron los siguientes procesos (todos ellos con error ruido blanco de media cero y varianza unidad):

- (i) AR(1) $\phi_1 = 0,5$
- (ii) MA(1) $\theta_1 = 0,7$
- (iii) AR(2) $\phi_1 = 0,6, \phi_2 = 0,2$
- (iv) MA(2) $\theta_1 = 0,8, \theta_2 = -0,6$
- (v) ARMA(1,1) $\phi_1 = 0,8, \theta_1 = 0,2$

Este conjunto de modelos han sido utilizados en otro trabajo por Maharaj (1996) para estudiar el comportamiento de la distancia d_M definida en (4.7) en el análisis cluster de procesos ARMA. Es por ello que se consideró un ejemplo adecuado para explorar el comportamiento de otro tipo de distancias en este contexto. En la Figura 4.2 se muestra el aspecto típico de las realizaciones de cada uno de estos modelos.

Se generaron $s = 4$ series de longitud $n = 200$ de cada uno de los procesos anteriores. Se realizó un análisis cluster a partir de las veinte así generadas, utilizando cada una de las medidas de disparidad especificadas en la Sección 4.4.2. Este esquema se repitió $T = 100$ veces y, en cada iteración, se examinaron las soluciones obtenidas con cuatro y cinco clusters. En esta ocasión, la evaluación de los resultados del análisis cluster se basó en el índice propuesto por Gavrilov *et al.* (2000), definido como el promedio de la máxima similitud de cada clase de procesos con cada uno de los clusters identificados en el análisis:

$$Sim(G, C) = \frac{1}{5} \sum_{i=1}^5 \max_{1 \leq j \leq k} Sim(G_j, C_i), \quad (4.16)$$

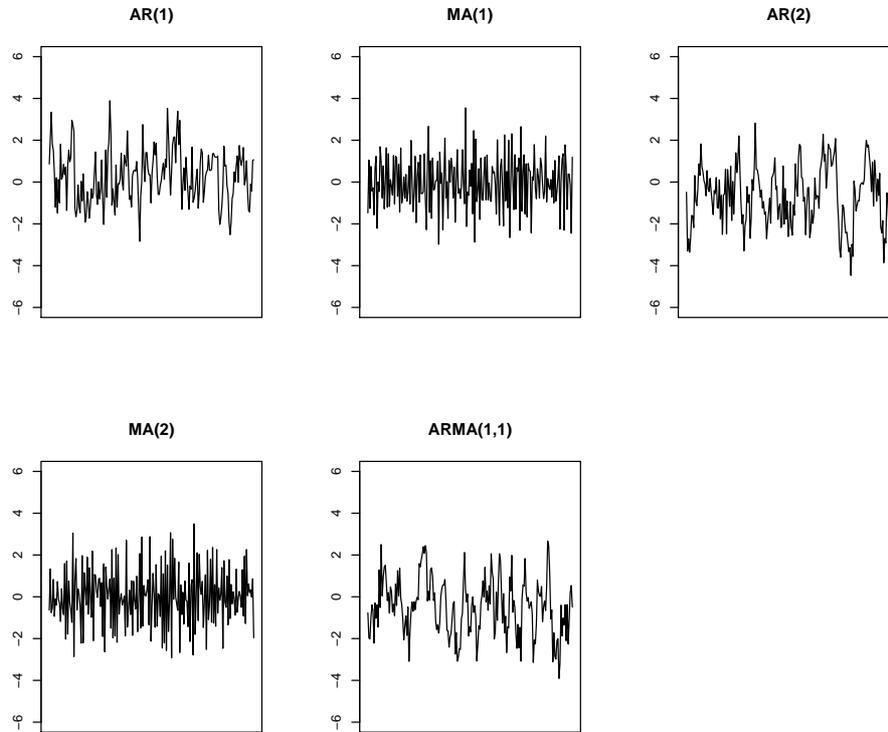
donde $C = \{C_1, \dots, C_5\}$ denota al conjunto de las 5 clases de procesos que existen en realidad, $G = \{G_1, \dots, G_k\}$, $k = 4, 5$ a la solución de k clusters obtenida tras el análisis, y

$$Sim(G_j, C_i) = \frac{2|G_j \cap C_i|}{|G_j| + |C_i|} \quad (4.17)$$

con $|\cdot|$ indicando el número de elementos en cada conjunto, de modo que la similitud de una determinada clase de procesos con cada cluster sea tanto mayor cuanto mayor sea el número de elementos que tienen en común.

El índice de evaluación definido en (4.16) puede tomar así valores entre 0 y 1, de modo que $Sim(G, C) = 1$ para una solución cluster completamente correcta, empeorando la calidad de los resultados conforme disminuye el valor de $Sim(G, C)$.

Figura 4.2: Ejemplo de series de tiempo generadas de los procesos ARMA (i)-(v).



En la Tabla 4.2 se muestra el índice de evaluación promedio (a lo largo de las $T = 100$ iteraciones) para las soluciones de $k = 4, 5$ clusters obtenidas con el método del enlace completo y utilizando cada una de las medidas de disparidad consideradas.

Tal y como se puede observar, los mejores resultados corresponden a las distancias basadas en los coeficientes de autocorrelación parcial y a la métrica propuesta por Maharaj (1996), con índices por encima de 0.8 en todos los casos. El buen comportamiento de estas medidas no resulta sorprendente puesto que, por un lado, las funciones de autocorrelación parcial de los procesos considerados son muy diferentes entre sí y, por otro, la distancia propuesta por Maharaj (1996), d_M , ha sido específicamente definida para clasificar procesos ARMA. Este también es el caso de la distancia propuesta por Piccolo (1990), d_{PIC} , que ha mostrado sin embargo un peor comportamiento.

Las distancias no paramétricas $d_{W(DLS)}$, $d_{W(LS)}$ y $d_{W(LK)}$ se sitúan en una posición intermedia en cuanto a los índices de evaluación obtenidos, alrededor de 0.8. Mientras que $d_{W(DLS)}$ y $d_{W(LK)}$ han alcanzado resultados muy parecidos, el peor comportamiento se ha registrado para la distancia construida a partir del logaritmo del periodograma suavizado mediante mínimos cuadrados locales, $d_{W(LS)}$. Este hecho viene a corroborar, en el caso

Tabla 4.2: Cluster de los procesos ARMA (i)-(v): Índice de evaluación definido en (4.16) para la solución de k clusters, con $k = 4, 5$. Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.

Medida	k		Medida	k	
	4	5		4	5
<i>Distancia Euclídea</i>			<i>Basadas en modelos</i>		
d_E	0.457	0.475	d_{PIC}	0.706	0.703
<i>Autocorrelaciones</i>			d_M	0.825	0.807
d_{ACFU}	0.716	0.751	<i>No paramétricas</i>		
d_{ACFG}	0.732	0.765	$d_{W(DLS)}$	0.783	0.793
d_{PACFU}	0.820	0.816	$d_{W(LS)}$	0.762	0.774
d_{PACFG}	0.828	0.820	$d_{W(LK)}$	0.778	0.790
<i>Periodogramas</i>			d_{GLK}	0.730	0.732
d_P	0.552	0.583	d_{CM}	0.769	0.786
d_{LP}	0.684	0.704			
d_{NP}	0.612	0.648			
d_{LNP}	0.703	0.740			

de trabajar con muestras finitas, la ineficiencia asintótica que se había probado para el estimador \hat{f}_{LS} con respecto tanto a \hat{f}_{DLS} como a \hat{f}_{LK} . A su vez, la distancia de tipo Cramer-von-Mises, d_{CM} , también ha alcanzado unos resultados razonablemente buenos, por encima de la distancia $d_{W(LS)}$ y muy cercana a $d_{W(DLS)}$ y $d_{W(LK)}$. Por el contrario, entre las distancias de corte no paramétrico, la métrica basada en el test de la razón de verosimilitudes generalizadas fue la que peores resultados obtuvo, con índices que apenas superaron el valor de 0.73.

El resto de las medidas consideradas proporcionaron peores resultados, incluyendo aquellas que habían mostrado un mejor comportamiento a la hora de clasificar series como estacionarias o no estacionarias en el apartado anterior: d_{ACFU} , d_{ACFG} , d_{NP} y d_{LNP} . En particular, la distancia Euclídea entre los log-periodogramas normalizados, d_{LNP} , (que había resultado ser la mejor en ese contexto) obtuvo, en el mejor de los casos, un índice de calidad de 0.74.

Aunque la solución adecuada al problema que se ha planteado es la formada por 5 clusters, los índices de evaluación obtenidos con la solución de 5 clusters son tan sólo ligeramente mejores que los calculados para la solución con 4 clusters, independientemente de la medida considerada. De hecho, los scores de aquellas métricas para las cuales se han observado los mejores resultados, d_M , d_{PACFU} y d_{PACFG} , resultan incluso más bajos para la solución con 5 clusters que para la solución con 4 clusters. Tras examinar los

dendogramas obtenidos para algunas de las iteraciones, se observó que lo que ocurría habitualmente era que uno de los grupos correctamente identificados en la solución de 4 clusters se dividía en dos nuevos grupos en la solución de 5 clusters correspondiente. Es decir, alguno de los clusters que contenían series procedentes de diferentes procesos resultaba más compacto que otro cluster correctamente identificado.

Para entender mejor los resultados obtenidos, se analizó además el número medio de veces que cada proceso fue correctamente identificado tras utilizar cada una de las distancias. La Tabla 4.3 muestra esta información para la solución de 4 clusters.

Tabla 4.3: Porcentaje de veces que se identificó correctamente cada uno de los procesos ARMA (i)-(v) en la solución de 4 clusters. Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.

Medida	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)
<i>Distancia Euclídea</i>					
d_E	0.0	0.0	0.0	0.0	0.0
<i>Autocorrelaciones</i>					
d_{ACFU}	36.0	2.0	12.0	12.0	1.0
d_{ACFG}	42.0	2.0	17.0	17.0	1.0
d_{PACFU}	33.0	1.0	99.0	97.0	0.0
d_{PACFG}	39.0	0.0	99.0	97.0	0.0
<i>Periodogramas</i>					
d_P	0.0	0.0	0.0	2.0	0.0
d_{LP}	0.0	0.0	69.0	66.0	0.0
d_{NP}	2.0	0.0	0.0	0.0	1.0
d_{LNP}	2.0	0.0	66.0	63.0	0.0
<i>Basadas en modelos</i>					
d_{PIC}	0.0	0.0	76.0	78.0	0.0
d_M	34.0	0.0	100.0	100.0	0.0
<i>No paramétricas</i>					
$d_{W(DLS)}$	24.0	1.0	96.0	95.0	0.0
$d_{W(LS)}$	12.0	1.0	98.0	93.0	0.0
$d_{W(LK)}$	24.0	0.0	98.0	95.0	0.0
d_{GLK}	3.0	0.0	90.0	89.0	0.0
d_{CM}	20.0	0.0	97.0	94.0	0.0

Como puede observarse, aquellas medidas que proporcionaron los mejores resultados (las basadas en los coeficientes de autocorrelación parcial, la métrica propuesta por Maharaj (1996) y las distancia no paramétricas) identificaron correctamente los procesos MA(1) y MA(2) en casi todas las iteraciones. Por el contrario, las series generadas a partir de

procesos AR(2) y ARMA(1,1) no fueron correctamente clasificadas en ninguna de las iteraciones, pero esto ocurrió con todas las medidas consideradas. Estas mismas medidas fueron capaces de clasificar correctamente a las series procedentes de procesos AR(1) entre un 20% y un 40% de las ocasiones, salvo las distancias $d_{W(LS)}$ y d_{GLK} . En particular, esta última medida casi nunca identificó las series AR(1), de ahí el peor índice de calidad que reflejaba con respecto a las otras medidas no paramétricas consideradas.

Con respecto a aquellas distancias que mostraron un peor comportamiento, resulta interesante observar su escasa capacidad para identificar las realizaciones de procesos MA(1). Así, por ejemplo, la distancia Euclídea entre el logaritmo de los periodogramas normalizados, d_{LNP} , únicamente agrupó correctamente las series MA alrededor de un 65% de las ocasiones y casi nunca identificó correctamente los procesos AR(1). De la misma forma, resulta sorprendente que la distancia de Piccolo (1990) no lograra identificar los procesos AR(1) en ninguna de las iteraciones.

Como un criterio de evaluación alternativo, se ha considerado también el número medio de clusters correctamente identificados en cada iteración. Más concretamente, para cada iteración se ha calculado: a) El número de clusters correctamente identificados (CC), b) El número de clusters incompletos (CI), en el sentido de que están formados por series pertenecientes a un mismo proceso generador pero no incluye al total de cuatro series generadas de ese modelo, y c) El número de clusters con series mezcladas (CM), es decir, clusters que engloban al menos dos series de distintos procesos generadores. Estos valores, promediados a lo largo de 100 iteraciones, y para la solución de 4 clusters, se muestran en la Tabla 4.4.

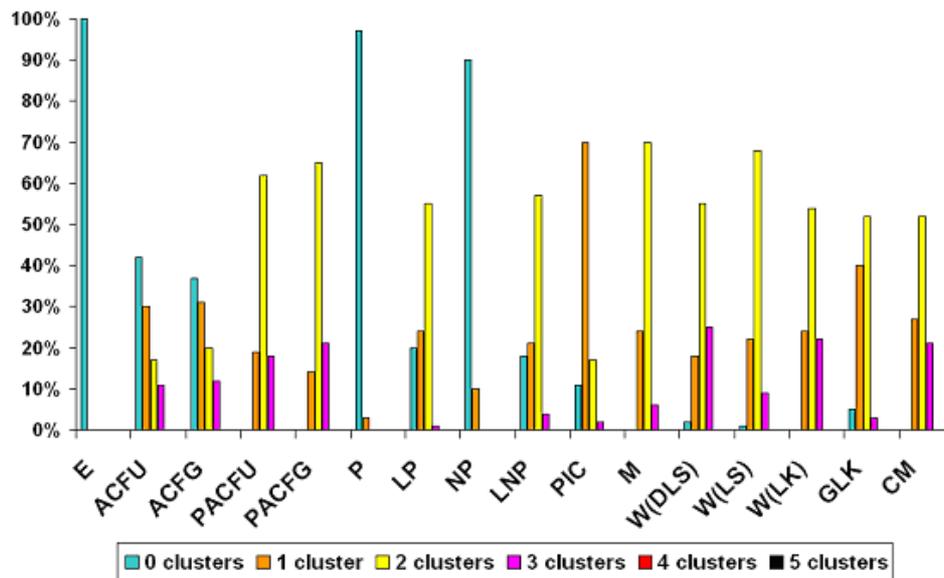
Los mejores resultados, en términos del número medio de clusters correctamente identificados (CC), corresponden a las medidas con mejor comportamiento según el índice de calidad en la Tabla 4.2, con valores entre 2 y 2.35 clusters correctos en cada iteración. Ninguna de las otras medidas utilizadas en la clasificación es capaz de detectar una media de dos clusters correctos. Además, sólo las métricas basadas en los coeficientes de autocorrelación, la medida propuesta por Maharaj (1996) y las distancias no paramétricas resultaron capaces de identificar correctamente 3 clusters en un porcentaje significativo de las iteraciones. Esto puede verse en la Figura 4.3, donde se muestra la distribución del número de clusters correctamente identificados en cada iteración. De dicha figura se deduce que la distancia entre el logaritmo de los periodogramas normalizados no identificó ningún cluster en aproximadamente un 32% de las ocasiones, confirmando su mal comportamiento en este contexto de clasificación.

De forma análoga, se puede analizar con más detalle lo que ocurre cuando se considera la solución con 5 clusters. En la Tabla 4.5 se muestra el número medio de veces que cada

Tabla 4.4: Resultados del cluster de los procesos ARMA (i)-(v) para la solución con 4 clusters. Los resultados muestran el número de: clusters correctamente identificados (CC), clusters incompletos que no incluyen series de diferentes procesos (CI) y clusters con series de distintos procesos generadores (CM). Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.

Medida	CC	IC	MC	Medida	CC	IC	MC
<i>Distancia Euclídea</i>			<i>Basadas en modelos</i>				
d_E	0.00	1.63	2.37	d_{PIC}	1.54	1.10	1.36
<i>Autocorrelaciones</i>			<i>No paramétricas</i>				
d_{ACFU}	0.63	0.69	2.68	$d_{W(DLS)}$	2.16	0.42	1.42
d_{ACFG}	0.79	0.62	2.59	$d_{W(LS)}$	2.04	0.49	1.47
d_{PACFU}	2.30	0.26	1.44	$d_{W(LK)}$	2.17	0.53	1.30
d_{PACFG}	2.35	0.22	1.43	d_{GLK}	1.82	0.43	1.75
<i>Periodogramas</i>			d_{CM}				
d_P	0.02	1.61	2.37		2.11	0.60	1.29
d_{LP}	1.35	0.89	1.76				
d_{NP}	0.03	1.07	2.90				
d_{LNP}	1.31	0.73	1.96				

Figura 4.3: Cluster de los procesos ARMA (i)-(v): Distribución del número de clusters correctamente identificados en cada iteración para la solución de 4 clusters. Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.



uno de los procesos fue correctamente identificado, para la solución de 5 clusters.

Tabla 4.5: Porcentaje de veces que se identificó correctamente cada uno de los procesos ARMA (i)-(v) en la solución de 5 clusters. Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.

Medida	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)
<i>Distancia Euclídea</i>					
d_E	0.0	0.0	0.0	0.0	0.0
<i>Autocorrelaciones</i>					
d_{ACFU}	39.0	2.0	29.0	26.0	1.0
d_{ACFG}	42.0	2.0	33.0	29.0	1.0
d_{PACFU}	32.0	1.0	93.0	75.0	1.0
d_{PACFG}	35.0	0.0	97.0	75.0	0.0
<i>Periodogramas</i>					
d_P	0.0	0.0	0.0	2.0	1.0
d_{LP}	2.0	0.0	70.0	65.0	0.0
d_{NP}	9.0	0.0	0.0	0.0	1.0
d_{LNP}	6.0	1.0	73.0	67.0	0.0
<i>Basadas en modelos</i>					
d_{PIC}	2.0	0.0	52.0	54.0	0.0
d_M	14.0	0.0	75.0	83.0	0.0
<i>No paramétricas</i>					
$d_{W(DLS)}$	30.0	3.0	90.0	80.0	0.0
$d_{W(LS)}$	17.0	0.0	89.0	79.0	0.0
$d_{W(LK)}$	35.0	0.0	87.0	76.0	0.0
d_{GLK}	6.0	0.0	73.0	74.0	0.0
d_{CM}	33.0	0.0	86.0	75.0	0.0

Tal y como puede observarse, las medidas con un mejor comportamiento (las basadas en los coeficientes de autocorrelación parcial, la distancia propuesta por Maharaj (1996) y las distancias no paramétricas) identifican correctamente los procesos MA(1) y MA(2) en un porcentaje de ocasiones menor que cuando se consideraba la solución de 4 clusters. En particular, la distancia de Maharaj ha pasado de identificar correctamente los procesos MA en todas las ocasiones a hacerlo en, a lo sumo, el 83% de las iteraciones. El resto de estas distancias identifican correctamente los procesos MA(2) en un porcentaje que no supera el 80% de las iteraciones. Estas mismas medidas siguen sin identificar en la solución de 5 clusters las series procedentes de los modelos AR(2) y ARMA(1,1) y clasifican correctamente las realizaciones de procesos AR(1) sólo entre un 6% y un 42% de las ocasiones. Con respecto a aquellas distancias que mostraron un peor comportamiento, pese a alcanzar resultados ligeramente mejores, siguen sin ser capaces de clasificar correctamente

las realizaciones de procesos AR y han logrado identificar los procesos MA en a lo sumo un 73% de las iteraciones.

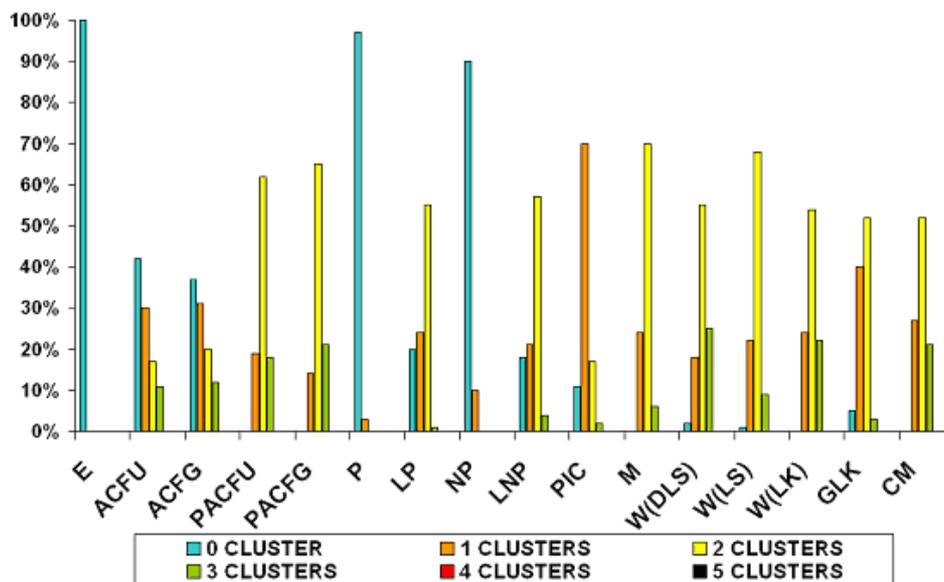
También para la solución de 5 clusters se ha considerado el número medio de clusters correctamente identificados en cada iteración (CC), el número de clusters incompletos (CI) y el número de clusters que incluían series de distintos procesos generadores (CM) (Tabla 4.6). De nuevo, los resultados son ligeramente peores que los obtenidos para la solución de 4 clusters. Los mejores resultados, en términos del número medio de clusters correctamente identificados (CC) vuelven a corresponder a las medidas con mejor comportamiento según el índice de calidad (4.16), si bien sólo las distancias basadas en los coeficientes de autocorrelación parcial, d_{PACFU} y d_{PACFG} , y una de las distancias no paramétricas, $d_{W(DLS)}$, alcanzan un promedio de 2 clusters correctos en cada iteración. De nuevo, las distancias basadas en los coeficientes de autocorrelación y las de corte no paramétrico son las que proporcionaron en un mayor número de ocasiones una solución con tres clusters correctos (Figura 4.4).

Tabla 4.6: Resultados del cluster de los procesos ARMA (i)-(v) para la solución con 5 clusters. Los resultados muestran el número de: clusters correctamente identificados (CC), clusters incompletos que no incluyen series de diferentes procesos (CI) y clusters con series de distintos procesos generadores (CM). Longitud de las series $n = 200$. Número de iteraciones $T = 100$. Método de enlace completo.

Medida	CC	IC	MC	Medida	CC	IC	MC
<i>Distancia Euclídea</i>				<i>Basadas en modelos</i>			
d_E	0.00	0.74	4.26	d_{PIC}	1.10	0.85	3.05
<i>Autocorrelación simple o parcial</i>				<i>No paramétricas</i>			
d_{ACFU}	0.97	1.17	2.86	d_M	1.82	0.94	2.24
d_{ACFG}	1.07	1.25	2.68	$d_{W(DLS)}$	2.03	1.04	1.93
d_{PACFU}	2.02	0.85	2.13	$d_{W(LS)}$	1.85	1.00	2.15
d_{PACFG}	2.07	0.84	2.09	$d_{W(LK)}$	1.98	0.99	2.03
<i>Periodogramas</i>				d_{GLK}	1.53	0.58	2.89
d_P	0.03	1.59	3.38	d_{CM}	1.94	0.99	2.07
d_{LP}	1.37	0.78	2.85				
d_{NP}	0.10	1.68	3.22				
d_{LNP}	1.47	0.94	2.59				

Tal y como se mencionó previamente, los dendogramas obtenidos en cualquiera de las iteraciones pueden resultar de ayuda en el análisis de los resultados alcanzados en esta sección. Así, por ejemplo, en la Figura 4.5 se muestran los dendogramas obtenidos mediante el método de enlace completo a partir de las distancias d_M , $d_{W(DLS)}$ y d_{LNP} .

Figura 4.4: Cluster de los procesos ARMA (i)-(v): Distribución del número de clusters correctamente identificados en cada iteración para la solución de 5 clusters. Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.



Como puede observarse, tanto la distancia propuesta por Maharaj (1996) como $d_{W(DLS)}$ llegan a identificar correctamente tres clusters (aquellos que corresponden a los procesos AR(1), MA(1) y MA(2)). Por el contrario, con la distancia d_{LNP} no se consigue identificar ningún cluster correcto tanto en la solución de 4 como de 5 clusters.

4.4.4. Clasificación de procesos no lineales

Por último, se desarrolló un experimento de simulación para evaluar el comportamiento de las diferentes medidas consideradas en la Sección 4.4.1 para la clasificación de procesos no lineales. En este caso, se consideraron cuatro modelos diferentes, de cada uno de los cuales se generaron $s = 4$ series de longitud $n = 200$. Más concretamente, los procesos considerados fueron:

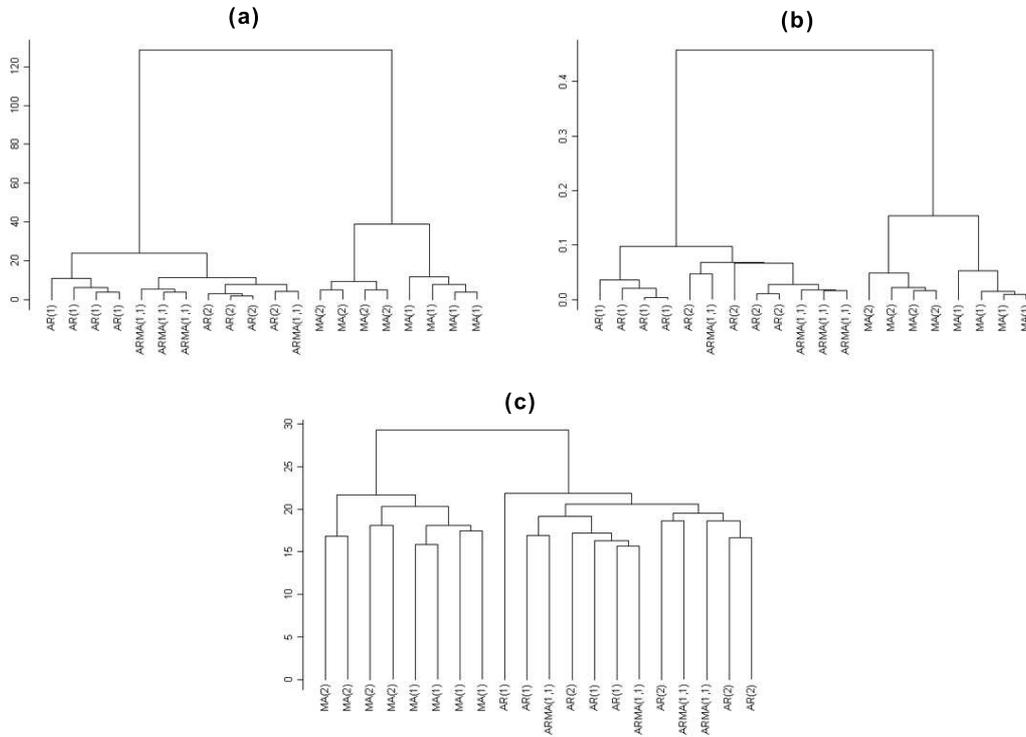
(1) Modelo TAR (Threshold Autoregressive):

$$X_t = 0,5X_{t-1}I(X_{t-1} \leq 0) - 2X_{t-1}I(X_{t-1} > 0) + \varepsilon_t$$

(2) Modelo EXPAR (Exponential Autoregressive):

$$X_t = (0,3 - 10 \exp\{-X_{t-1}^2\}) X_{t-1} + \varepsilon_t$$

Figura 4.5: Cluster de los procesos ARMA (i)-(v): Dendogramas obtenidos mediante el método de enlace completo a partir de (a) d_M , (b) $d_{W(DLS)}$ y (c) d_{LNP} para una iteración particular. Longitud de las series: $T = 200$.



(3) Modelo MA(1) lineal (MA):

$$X_t = \varepsilon_t + 0,4\varepsilon_{t-1}$$

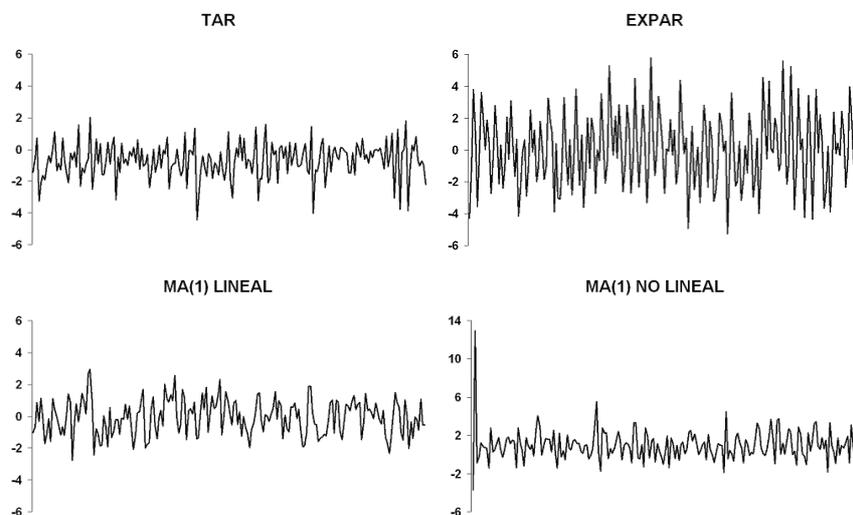
(4) Modelo MA(1) no lineal (NLMA):

$$X_t = \varepsilon_t - 0,5\varepsilon_{t-1} + 0,8\varepsilon_{t-1}^2$$

En todos los casos el proceso de error ε_t consistió en variables gaussianas de media cero y varianza unidad. Estos modelos han sido utilizados previamente para estudiar las propiedades de diversos estadísticos diseñados para testar la no linealidad de series de tiempo (ver Tong y Yeung (1991)). Nuestra intención consistió básicamente en discriminar entre diferentes formas lineales y examinar el comportamiento de las diferentes distancias en este contexto. Una realización arbitraria de cada uno de estos procesos se muestra en la Figura 4.6.

En este caso, la evaluación de los resultados del análisis cluster se realizó con el mismo

Figura 4.6: Ejemplo de series de tiempo generadas de los procesos (1)-(4).



criterio que en la sección anterior, según se definió en (4.16), donde ahora

$$Sim(G, C) = \frac{1}{4} \sum_{i=1}^4 \max_{1 \leq j \leq k} Sim(G_j, C_i), \quad (4.18)$$

siendo $C = \{C_1, \dots, C_4\}$ el conjunto de 4 clases de procesos que existen en realidad, $G = \{G_1, \dots, G_k\}$ la solución de k clusters obtenida tras el análisis, y donde la medida de similitud se define como en (4.17).

Al igual que en la sección anterior, los valores de este índice se promediaron a lo largo de 100 iteraciones, dando lugar a los resultados que se muestran en la Tabla 4.7.

Definitivamente, y tal y como cabía esperar, los mejores resultados en este contexto se obtuvieron con las medidas de disimilitud de corte no paramétrico. Todas ellas proporcionaron índices de calidad alrededor de 0.9, excepto la métrica basada en el cociente de verosimilitudes generalizadas, d_{GLK} . Sin embargo, incluso esta distancia, con un índice de calidad de 0.818 ha mostrado un comportamiento claramente superior al de resto de medidas consideradas. A su vez, las distancias paramétricas propuestas por Piccolo (1990) y Maharaj (1996) han visto afectado su rendimiento por la mala especificación de los procesos generadores y se han situado en una posición intermedia (con índices de calidad en torno a 0.78), junto con las medidas basadas en los coeficientes de autocorrelación y d_{LP} . El resto de distancias se han situado bastante por debajo, con índices de calidad en torno sólo a 0.57. En particular, la distancia Euclídea entre los logaritmos del periodograma normalizado, que había resultado ser la mejor métrica en el caso de clasificar procesos como estacionarios o no estacionarios, ha obtenido un índice de calidad muy bajo, de 0.574.

Tabla 4.7: Cluster de los procesos (1)-(4): Índice de evaluación definido en (4.18) para la solución de 4 clusters. Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.

Medida	Índice	Medida	Índice
<i>Distancia Euclídea</i>		<i>Basadas en modelos</i>	
d_E	0.537	d_{PIC}	0.769
<i>Autocorrelaciones</i>		d_M	0.781
d_{ACFU}	0.752	<i>No paramétricas</i>	
d_{ACFG}	0.777	$d_{W(DLS)}$	0.920
d_{PACFU}	0.784	$d_{W(LS)}$	0.895
d_{PACFG}	0.795	$d_{W(LK)}$	0.912
<i>Periodogramas</i>		d_{GLK}	0.818
d_P	0.485	d_{CM}	0.913
d_{LP}	0.786		
d_{NP}	0.576		
d_{LNP}	0.574		

Al igual que en la sección anterior, y para valorar de una forma más detallada los resultados alcanzados, se ha computado también el número medio de veces que se ha identificado correctamente cada una de las clases de procesos y el número medio de clusters correctamente identificados en cada iteración. Estos datos se muestran en las Tablas 4.8 y 4.9, respectivamente.

Los resultados en la Tabla 4.8 permite concluir que las series generadas de los procesos EXPAR y MA forman los clusters más homogéneos. Cuando se utilizan las mejores medidas no paramétricas ($d_{W(DLS)}$, $d_{W(LK)}$ y d_{CM}), en términos de los índices de calidad obtenidos, las series de este par de procesos se agrupan correctamente en su correspondiente cluster alrededor del 91 % de las ocasiones. Con las métricas basadas en los coeficientes de autocorrelación, o con aquellas propuestas por Piccolo y Maharaj se obtuvieron peores resultados, especialmente para las series EXPAR, con tasas de éxito alrededor del 25 %. Las distancias basadas en las coordenadas de los periodogramas, excepto d_{LP} , son incluso incapaces de separar correctamente la clase MA de procesos lineales. Ha resultado mucho más difícil agrupar correctamente las series generadas a partir de los modelos TAR y NLMA con cualquiera de las distancias consideradas. Este hecho se corrobora tras observar los datos de la Tabla 4.9, donde se observa que las distancias no paramétricas permite identificar correctamente entre dos y tres clusters en cada iteración (la media más alta de número de procesos correctamente identificados fue de 2.56, correspondiente a $d_{W(DLS)}$). Ninguna de las otras métricas fue capaz de identificar un promedio de dos

Tabla 4.8: Porcentaje de veces que se identificó correctamente cada una de las clases de procesos (1)-(4) en la solución de 4 clusters. Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.

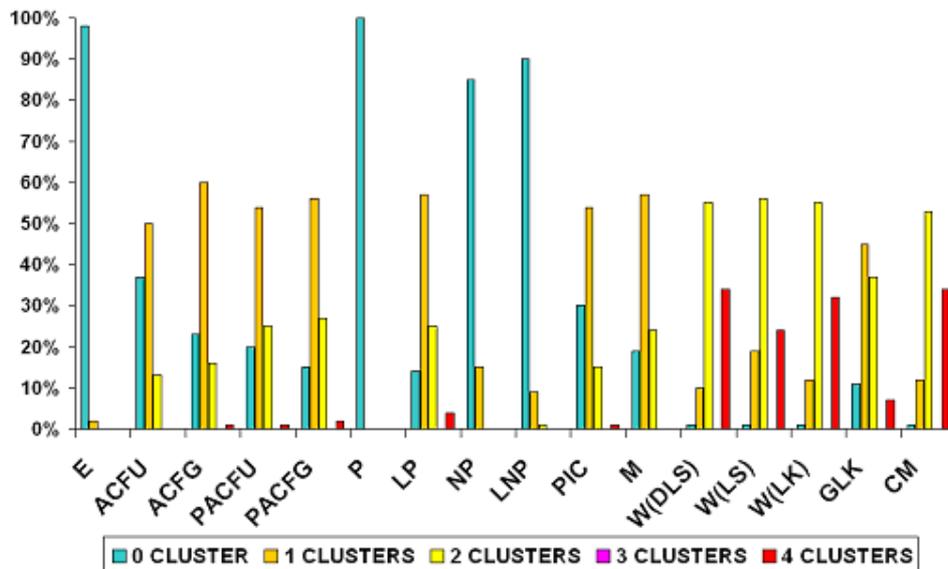
Medida	TAR	EXPAR	MA	NLMA
<i>Distancia Euclídea</i>				
d_E	0.0	0.0	0.0	2.0
<i>Autocorrelaciones</i>				
d_{ACFU}	0.0	17.0	50.0	9.0
d_{ACFG}	1.0	22.0	59.0	14.0
d_{PACFU}	2.0	25.0	69.0	12.0
d_{PACFG}	3.0	27.0	75.0	13.0
<i>Periodogramas</i>				
d_P	0.0	0.0	0.0	0.0
d_{LP}	4.0	72.0	36.0	11.0
d_{NP}	0.0	2.0	9.0	4.0
d_{LNP}	0.0	1.0	8.0	2.0
<i>Basadas en modelos</i>				
d_{PIC}	1.0	18.0	59.0	10.0
d_M	0.0	21.0	74.0	10.0
<i>No paramétricas</i>				
$d_{W(DLS)}$	34.0	93.0	95.0	34.0
$d_{W(LS)}$	24.0	95.0	81.0	27.0
$d_{W(LK)}$	32.0	94.0	91.0	33.0
d_{GLK}	7.0	73.0	54.0	13.0
d_{CM}	34.0	94.0	91.0	35.0

clusters correctos. Por lo tanto, excepto para las medidas no paramétricas, ninguna de las métricas estudiadas parecen resultar apropiadas para realizar el cluster de procesos no lineales. Esta apreciación se observa gráficamente en la Figura 4.7, donde se muestra la distribución de probabilidad del número de clusters correctamente identificados en cada iteración para cada una de las medidas consideradas. Se observa que las métricas no paramétricas proporcionan una solución completamente correcta cerca del 40% de las veces. Cabe destacar, además, el alto porcentaje de ocasiones en las que las medidas basadas en los periodogramas, y en particular d_{LNP} , resultan incapaces de identificar correctamente ni uno sólo de las clases consideradas.

Tabla 4.9: Resultados del cluster de los procesos (1)-(4) para la solución con 4 clusters. Los resultados muestran el número de: clusters correctamente identificados (CC), clusters incompletos que no incluyen series de otros procesos (CI) y clusters con series de distintos procesos generadores (CM). Longitud de las series: $n = 200$. Número de iteraciones: $T = 100$. Método de enlace completo.

Medida	CC	IC	MC	Medida	CC	IC	MC
<i>Distancia Euclídea</i>			<i>Basadas en modelos</i>				
d_E	0.02	2.54	1.44	d_{PIC}	0.88	1.37	1.75
<i>Autocorrelaciones</i>			<i>No paramétricas</i>				
d_{ACFU}	0.76	1.46	1.78	$d_{W(DLS)}$	2.56	0.68	0.76
d_{ACFG}	0.96	1.36	1.68	$d_{W(LS)}$	2.27	0.80	0.93
d_{PACFU}	1.08	1.13	1.79	$d_{W(LK)}$	2.50	0.67	0.83
d_{PACFG}	1.18	1.15	1.67	d_{GLK}	1.47	0.87	1.66
<i>Periodogramas</i>			d_{CM}				
d_P	0.00	3.00	1.00		2.54	0.65	0.81
d_{LP}	1.23	1.08	1.69				
d_{NP}	0.15	1.95	1.90				
d_{LNP}	0.11	1.46	2.43				

Figura 4.7: Cluster de procesos (1)-(4): Distribución del número de clusters correctamente identificados en cada iteración para la solución de 4 clusters.



4.4.5. Conclusiones

Mediante este experimento de simulación, se ha examinado el comportamiento de diversas medidas de disparidad paramétricas y no paramétricas en diferentes contextos de

clasificación. Más concretamente, se consideraron tres problemas diferentes: la clasificación de procesos como estacionarios y no estacionarios, el cluster de procesos ARMA y la clasificación de procesos no lineales. En cada uno de estos escenarios, se diseñó un estudio de simulación para evaluar los resultados obtenidos tras utilizar las diferentes distancias.

Los resultados obtenidos muestran que el comportamiento de una determinada métrica depende en gran medida del tipo de procesos que se quieren clasificar. Así, una determinada medida puede proporcionar resultados muy buenos en un contexto de clasificación y no resultar apropiada en otro. Por ejemplo, a partir de los resultados que se muestran en esta sección se concluye que tanto las distancias basadas en los periodogramas normalizados como las métricas basadas en los coeficientes de autocorrelación proporcionan tasas de éxito muy elevadas para la clasificación de procesos como estacionarios y no estacionarios (confirmando así los resultados ya obtenidos por Caiado, Crato y Peña (2006)). Sin embargo, estas medidas proporcionan unos de los peores resultados cuando se utilizan para la clasificación de diferentes procesos ARMA o procesos no lineales. De igual forma, aquellas medidas basadas en asumir un modelo generador determinado pueden llevar a resultados poco adecuados si el modelo que se asume no se adapta al contexto considerado. Este es el caso de las distancias propuestas por Piccolo (1990) o Maharaj (1996), diseñadas específicamente bajo la hipótesis de modelos ARMA o ARIMA. Así, mientras que proporcionan buenos resultados en el experimento diseñado para la clasificación de procesos ARMA (especialmente la distancia propuesta por Maharaj (1996), que resultó la mejor en este contexto), resultaron entre las peores cuando se utilizaron para la clasificación de procesos no lineales.

Entre todas las medidas de disparidad consideradas, se incluyeron cinco distancias de corte no paramétrico. Básicamente, consistieron en medidas de disparidad entre los espectros (o el logaritmo de los espectros) suavizados previamente. Tal y como cabía esperar, estas medidas proporcionaron resultados razonablemente buenos en los tres contextos de clasificación considerados, confirmando su robustez frente al resto de distancias consideradas. Más concretamente, todas las medidas no paramétricas proporcionaron resultados sustancialmente mejores a los obtenidos con el resto de distancias para la clasificación de procesos no lineales, alcanzaron resultados muy próximos a los mejores para la clasificación de procesos ARMA y las tasas de éxito obtenidas para la clasificación de procesos como estacionarios o no estacionarios resultaron competitivas cuando se evaluaron sobre el rango de frecuencias más bajas. Entre las distancias de corte no paramétrico consideradas, $d_{W(DLS)}$, $d_{W(LK)}$ y d_{CM} obtuvieron el mejor comportamiento y por lo tanto pueden considerarse las “ganadoras” del estudio de simulación que se ha llevado a cabo.

Finalmente, debe señalarse que los peores resultados se alcanzaron con la distancia Euclídea entre las series de tiempo observadas (d_E), las ordenadas de los periodogramas

(d_P) y entre el logaritmo de las ordenadas del periodograma (d_{LP}). Esto permite concluir que estas medidas no resultan adecuadas para determinar la afinidad entre series de tiempo, y recalcar la importancia de la elección de una medida de disparidad adecuada a partir de la que realizar el análisis cluster de procesos estocásticos.

4.5. Aplicación a datos reales: clasificación de registros electrocardiográficos.

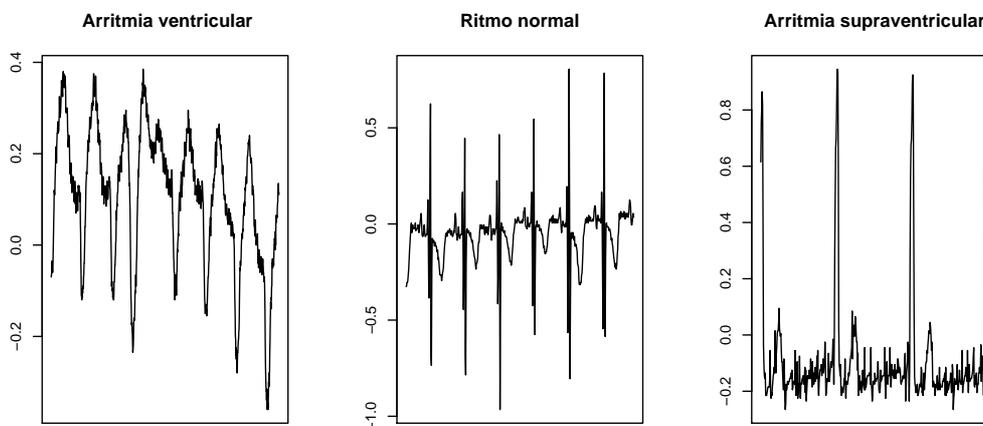
Para ilustrar la aplicación práctica de los procedimientos cluster desarrollados en la sección anterior, se ha recurrido a un conjunto de datos reales obtenidos de la base de datos de electrocardiogramas mantenida por PsysioNet¹, servicio público del Instituto Nacional de Imagen Biomédica y Bioingeniería (NIBIB) y del Instituto Nacional de Ciencias Médicas Generales (NIGMS). Estos datos han sido previamente utilizados en la literatura para valorar los resultados de otros procedimientos de cluster de series temporales. En particular, Kalpakis *et al.* (2001) los han empleado para evaluar el comportamiento de un procedimiento de clasificación que ellos proponen basándose en la distancia euclídea entre los coeficientes cepstrales de dos series de tiempo.

El conjunto de datos utilizados incluye los registros electrocardiográficos de tres grupos de pacientes diferentes. El primer grupo incluye 22 series de tiempo de longitud $n = 1000$ correspondientes a los registros electrocardiográficos, durante un periodo de 2 segundos, de otros tantos pacientes a los que se les ha diagnosticado una arritmia ventricular. La arritmia se define como cualquier trastorno o irregularidad en el ritmo o frecuencia cardíaca natural del corazón. La arritmia ventricular es aquella que se origina en los ventrículos y es el tipo de arritmia más severo que se conoce, resultando en algunos casos de peligro vital para el paciente. El segundo grupo incluye los registros, de 2 segundos de duración, de 18 pacientes sanos. La tasa de muestreo en este caso es menor, dando lugar a series de longitud $n=512$. Finalmente, el tercer grupo incluye los registros electrocardiográficos de 24 pacientes con arritmia supraventricular ($n=512$). Este tipo de arritmia se origina en las cámaras superiores del corazón o aurículas y no resulta habitualmente peligrosa para la vida de los enfermos. En la Figura 4.8 se muestra un ejemplo de los registros electrocardiográficos de pacientes en cada uno de estos grupos.

En primer lugar, se procedió a la clasificación de los procesos correspondientes a los grupos 1 y 2 antes descritos. Es decir, se procedió a examinar el comportamiento de los procedimientos cluster propuestos en la Sección 4.3 para la clasificación de los registros electrocardiográficos de pacientes con diagnóstico de arritmia ventricular y pacientes

¹Physiobank Archive. www.physionet.org/physiobank/database.

Figura 4.8: Ejemplo de los registros electrocardiográficos de los pacientes en cada uno de los tres grupos de estudio.

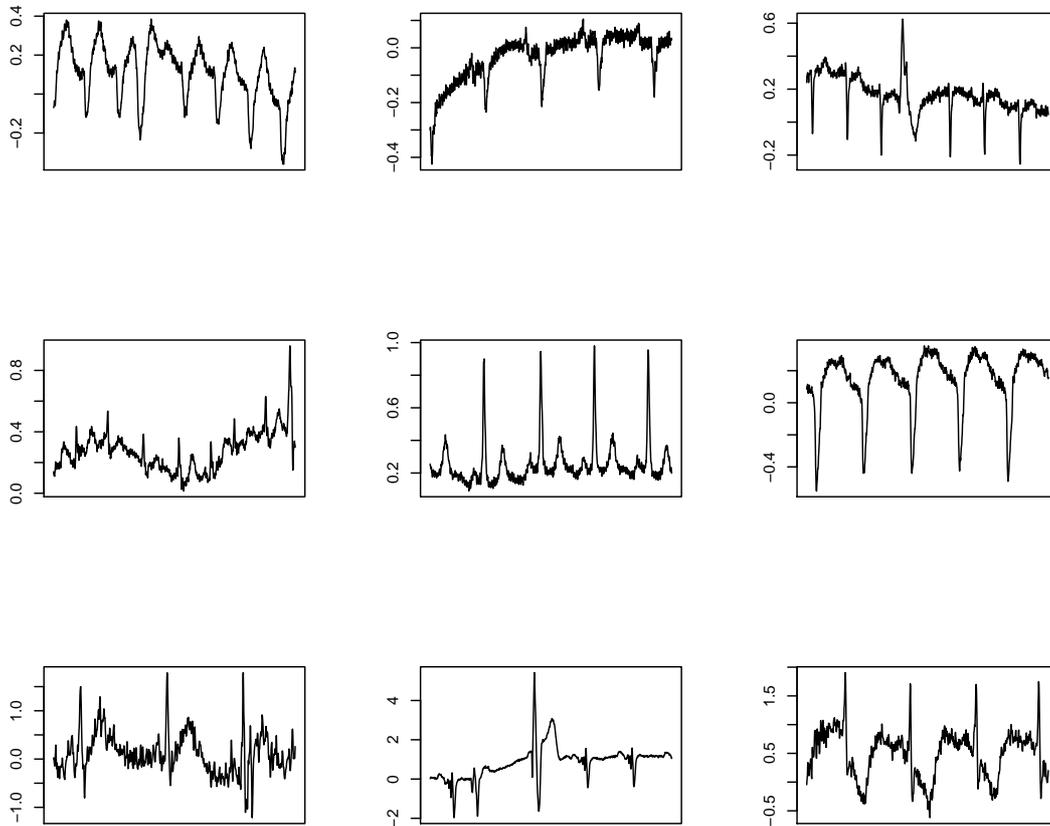


sanos. Tras examinar con detenimiento los datos, se constató que existía una alta variabilidad entre los registros de los pacientes dentro de un mismo grupo. Así, en la Figura 4.9 se muestran 9 de los 22 registros disponibles de pacientes con arritmia ventricular. Tal y como se puede apreciar, los patrones observados pueden resultar sustancialmente diferentes incluso en pacientes aquejados de una misma patología. De la misma forma, se observan diferencias importantes entre los registros de pacientes sin ninguna alteración electrocardiográfica (Figura 4.10).

Puesto que los datos correspondientes a los registros de pacientes con arritmia ventricular eran claramente no estacionarios, se procedió a diferenciar las series para conseguir la estacionariedad en media. Tanto a las series correspondientes a pacientes sanos como aquellas series correspondientes a pacientes con arritmia ventricular (éstas una vez diferenciadas) se les restó la media muestral para conseguir series estacionarias de media cero.

A continuación, se emplearon las medidas propuestas en la Sección 4.3 para calcular la matriz de distancias entre las series observadas, a partir de la cual realizar el análisis cluster de los datos. En particular, se utilizaron las medidas de disparidad espectral definidas como d_W en (4.13), utilizando para la estimación del espectro cada uno de los tres suavizadores propuestos por Fan y Kreutzberger (1998) ($d_{W(DLS)}$, $d_{W(DLS)}$ y $d_{W(DLS)}$), y la distancia funcional de tipo Cramer-von-Mises, d_{CM} , dada por (4.15). Para el cómputo de las distancias se utilizaron los mismos parámetros que en el estudio de simulación anterior. Nótese que en este caso concreto no se pudo emplear la distancia basada en el

Figura 4.9: Ejemplo de registros electrocardiográficos de pacientes diagnosticados de arritmia ventricular.

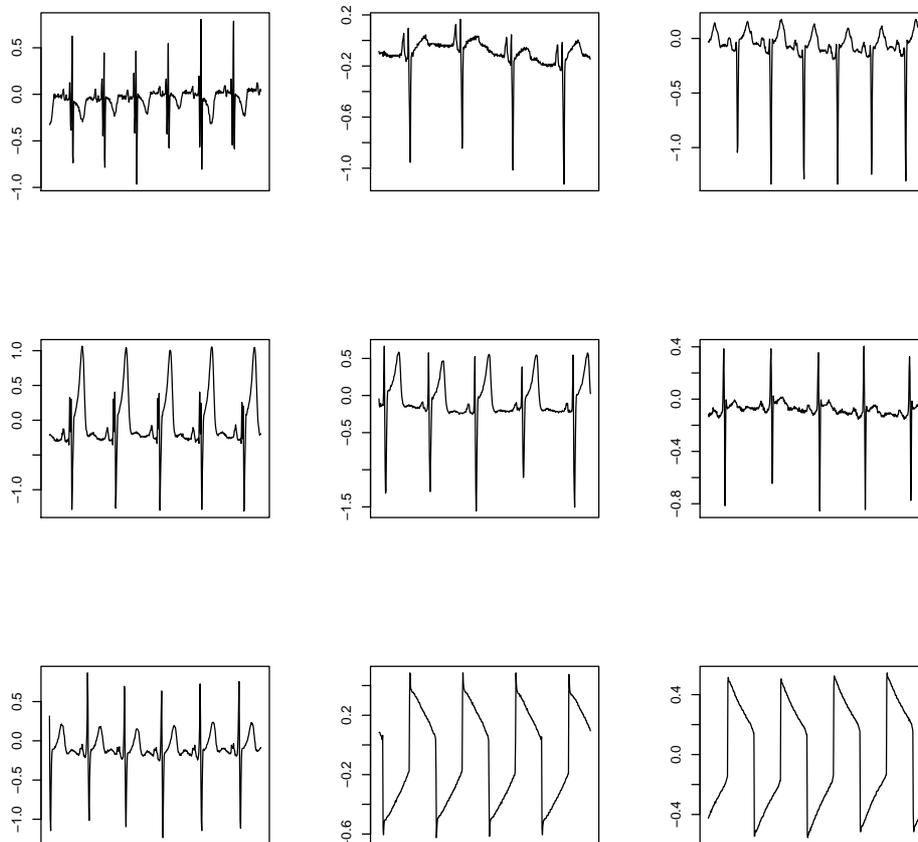


test de razón de verosimilitud generalizada, d_{GLK} , definida en (4.14), por tener las series en ambos grupos diferente longitud, y no poder por lo tanto evaluar los periodogramas correspondientes sobre las mismas frecuencias.

La matriz de distancias obtenida con cada una de estas métricas se procesó mediante un algoritmo cluster jerarquizado (método de enlace completo) para obtener el correspondiente dendograma, considerándose la solución final de dos grupos. Para cada distancia, los resultados del análisis cluster se evaluaron utilizando los siguientes indicadores:

1. El porcentaje de series mal clasificadas en la solución de 2 clusters.
2. El índice propuesto por Gavrilov et al. (2000), según se definió en (4.16), donde

Figura 4.10: Ejemplo de registros electrocardiográficos de pacientes sin ninguna patología.



ahora

$$Sim(G, C) = \frac{1}{2} \sum_{i=1}^2 \max_{1 \leq j \leq k} Sim(G_j, C_i),$$

siendo $C = \{C_1, C_2\}$ los dos grupos de procesos que existen en realidad, $G = \{G_1, G_2\}$, la solución de 2 clusters obtenida tras el análisis, y donde la medida de similitud se define como en (4.17).

3. Los valores silueta, $s(i)$, definidos tal y como se explica a continuación.

Para cada una de las series i consideradas, denotaremos por A el cluster al que pertenece dicha serie i y por $a(i)$ al promedio de las distancias entre la serie i y cada una de las otras series clasificadas en el mismo cluster A . Para cada uno del resto de clusters identificados en el análisis C , sea $D(i, C)$ el promedio de las distancias entre la serie i y todas las series clasificadas en C . Sea $b(i) = \min_{C \neq A} D(i, C)$, es decir,

$b(i)$ puede verse como la distancia entre la serie i y su cluster “vecino” (el segundo mejor cluster para i). El valor silueta $s(i)$ puede definirse entonces como

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \quad (4.19)$$

El valor $s(i)$ se encuentra siempre, por lo tanto, entre -1 y 1 de forma que:

- Las series con un valor grande de $s(i)$ (más cercano a 1) serán series bien clasificadas,
- Las series con valores pequeños de $s(i)$ (próximos a 0) serán series que se encuentran entre dos clusters
- Las series con valores negativos de $s(i)$ (próximos a -1) estarán probablemente mal clasificadas

Una forma útil de representar gráficamente la solución de k grupos obtenida tras un análisis cluster es mediante el gráfico de silueta, en el que se representan los valores silueta $s(i)$ para cada una de las series. El gráfico de silueta proporciona así una herramienta para valorar la estabilidad de la solución cluster obtenida y la calidad del análisis que resulta de gran utilidad en situaciones en las que, como en esta, se desconoce la solución correcta.

En la Tabla 4.10 se muestran los resultados obtenidos, en términos de los indicadores anteriores, para la clasificación de los registros electrocardiográficos del grupo de pacientes sanos y pacientes con arritmia ventricular, con cada una de las distancias empleadas. En la misma tabla se muestran además los resultados obtenidos en el estudio de Kalpakis *et al.* (2000) sobre el mismo conjunto de datos, utilizando diferentes medidas de proximidad para la clasificación: la distancia euclídea entre los coeficientes cepstrales de las series de tiempo (LPCCEP), la distancia euclídea entre los coeficientes de la transformada finita de Fourier de las series (FFT), la distancia euclídea entre los coeficientes de la transformada finita de Fourier de las funciones de autocorrelación (FFT(ACF)), la transformada wavelet discreta (DWT) o el análisis de componentes principales (PCA).

Tal y como se puede observar, la distancia FFT(ACF) es la que proporciona mejores resultados en términos del índice (4.16), si bien el ancho del gráfico de silueta es el menor de los obtenidos con todos los procedimientos testados, indicando la escasa estabilidad de la solución obtenida. El porcentaje de series mal clasificadas, asimismo, fue menor con una de las distancias no paramétricas testadas, d_{CM} , siendo sólo de un 7.5 %. De modo general, las distancias no paramétricas proporcionaron resultados similares a los alcanzados en el estudio de Kalpakis *et al.* (2000) con la distancia basada en los coeficientes cepstrales, que había resultado ser la mejor entre todas las distancias que habían considerado en su trabajo. De nuevo, se observa la influencia de las propiedades teóricas de los estimadores

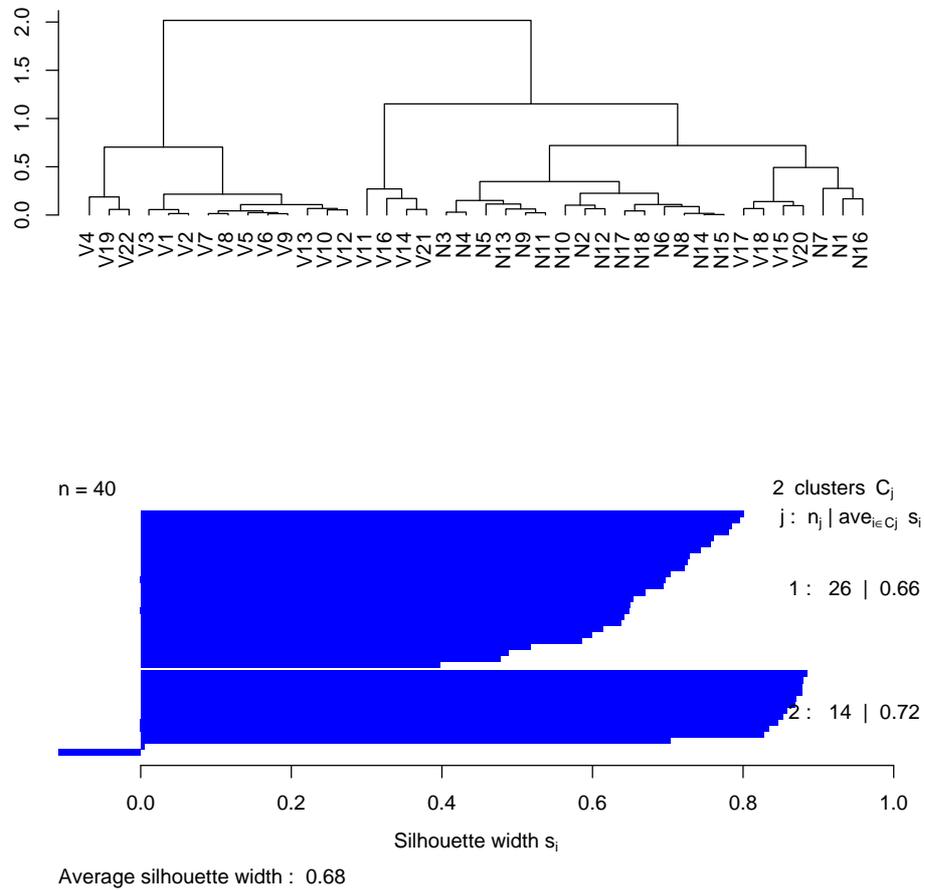
Tabla 4.10: Resultados del cluster de los registros electrocardiográficos correspondientes a pacientes sanos y pacientes con arritmia ventricular.

Método	Tasa de error	Sim(G,C)	Ancho silueta
$d_{W(DLS)}$	20 %	0.798	0.670
$d_{W(LS)}$	37.5 %	0.625	0.690
$d_{W(LK)}$	20 %	0.798	0.680
d_{CM}	7.5 %	0.744	0.770
LPCCEP	20 %	0.771	0.502
FFT	—	0.629	0.539
FFT (ACF)	10 %	0.881	0.299
DWT	—	0.587	0.523
PCA	—	0.587	0.377

del espectro sobre los resultados de la clasificación, obteniéndose con d_{LS} una mayor tasa de error y peores índices de fiabilidad que con el resto de las distancias de corte no paramétrico. Si bien con la distancia de tipo Cramer-von-Mises se obtiene una tasa de mala clasificación más baja y un ancho de silueta mayor, con las distancias d_{DLS} y d_{LK} se alcanzan los mayores índices de fiabilidad según la medida $Sim(G, C)$, en torno a 0.8. Los resultados de la clasificación con las distancias d_{LK} y d_{CM} se muestran en las Figuras 4.11 y 4.12.

A continuación se procedió a utilizar las mismas distancias para clasificar las series correspondientes a los registros electrocardiográficos de pacientes sanos y pacientes con arritmia supraventricular. Las series de sujetos con arritmia supraventricular fueron también diferenciadas con el objetivo de trabajar con series estacionarias de media cero. Los resultados obtenidos, con el método de enlace completo, se muestran en la Tabla 4.11, junto con los reportados por Kalpakis *et al.* (2000) para las distancias mencionadas con anterioridad. En esta situación, los resultados alcanzados por las distancias no paramétricas propuestas en la presente memoria son ligeramente peores que los obtenidos con la distancia basada en los coeficientes cepstrales, LPCCEP, si los comparamos según el índice $Sim(G, C)$, con índices de estabilidad similares en términos del valor silueta promedio. En cualquier caso, su comportamiento resulta claramente superior al del resto de medidas analizadas. Las elevadas tasas de error, por encima del 25 % en todos los casos, no reflejan adecuadamente las soluciones cluster alcanzadas. A la vista de los dendogramas obtenidos se observa cómo con estas medidas se logra separar adecuadamente las series correspondientes a sujetos sanos y a enfermos con arritmia supraventricular, si bien la solución con dos clusters no parece ser la solución natural que surge de estos resultados, observándose en general una solución con tres clusters en las que las series correspondientes a pacientes

Figura 4.11: Cluster de los registros electrocardiográficos correspondientes a pacientes sanos y pacientes con arritmia ventricular. Resultados con la distancia d_{LK} .



con arritmia supraventricular quedan divididas en dos grupos (Figuras 4.13, 4.14).

Figura 4.12: Cluster de los registros electrocardiográficos correspondientes a pacientes sanos y pacientes con arritmia ventricular. Resultados con la distancia d_{CM} .

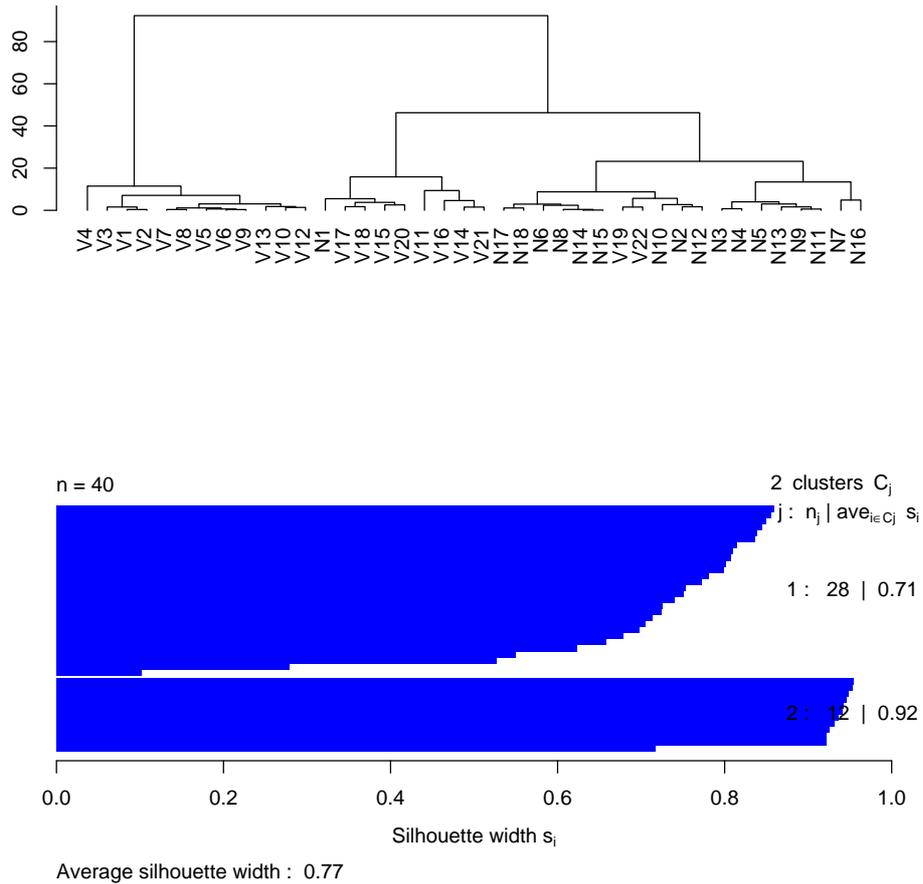


Tabla 4.11: Resultados del cluster de los registros electrocardiográficos correspondientes a pacientes sanos y pacientes con arritmia supraventricular.

Método	Tasa de error	Sim(G,C)	Ancho silueta
$d_{W(DLS)}$	33.3 %	0.654	0.470
$d_{W(LS)}$	59.5 %	0.606	0.510
$d_{W(LK)}$	28.6 %	0.708	0.480
d_{CM}	28.6 %	0.708	0.540
LPCCEP	—	0.779	0.519
FFT	—	0.579	0.649
FFT (ACF)	—	0.593	0.425
DWT	—	0.561	0.635
PCA	—	0.601	0.521

Figura 4.13: Cluster de los registros electrocardiográficos correspondientes a pacientes sanos y pacientes con arritmia supraventricular. Resultados con la distancia d_{LK} .

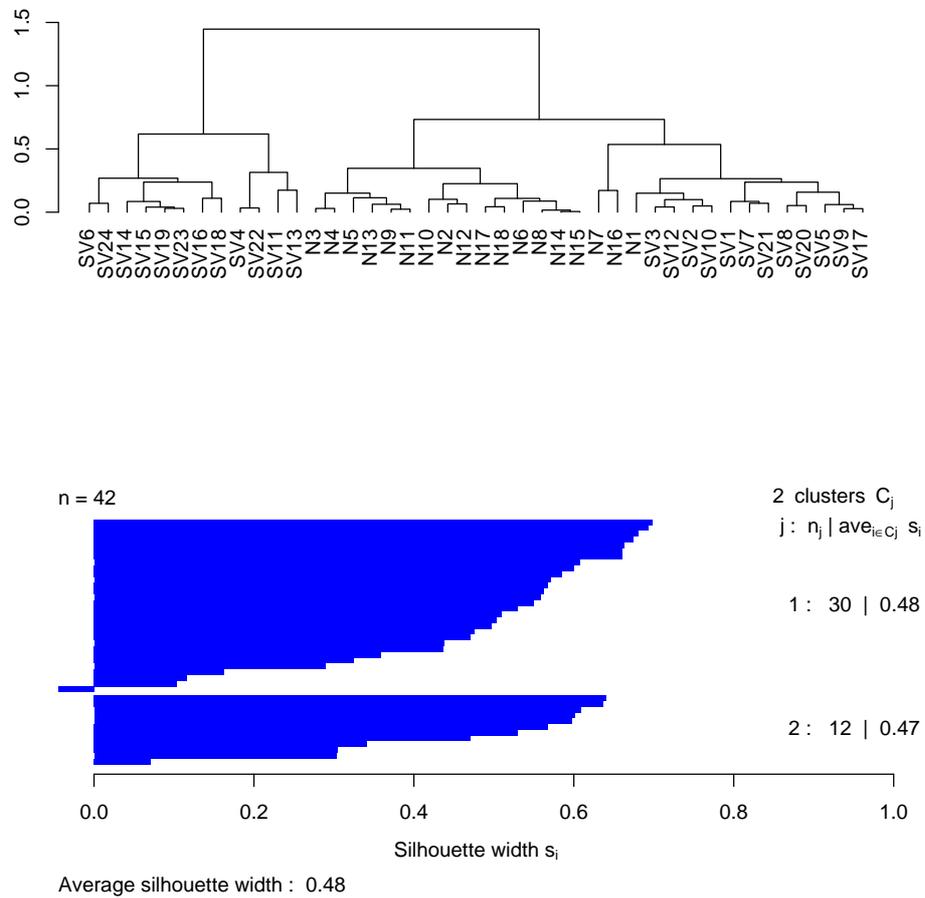
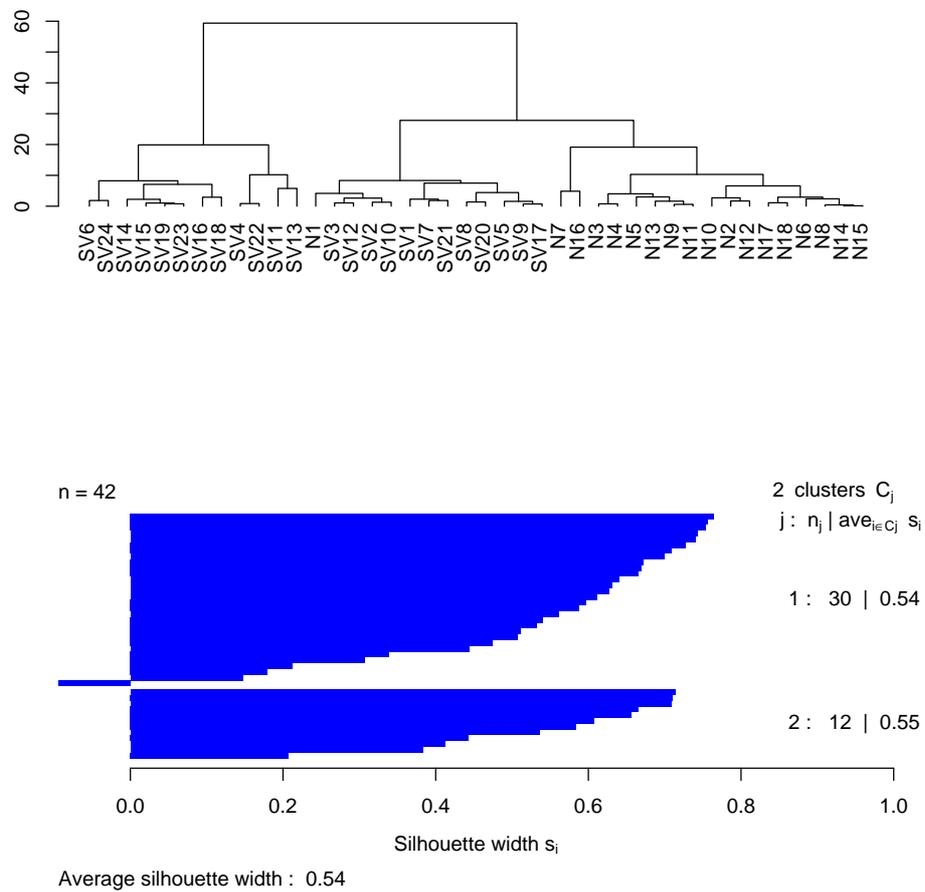


Figura 4.14: Cluster de los registros electrocardiográficos correspondientes a pacientes sanos y pacientes con arritmia supraventricular. Resultados con la distancia d_{CM} .



Capítulo 5

Análisis de las series de cotización bursátil del sector bancario en España

5.1. Introducción

Es conocido que los métodos de estimación no paramétrica de curvas proporcionan una herramienta estadística muy potente para explorar la estructura subyacente a un conjunto de datos. Estos métodos se caracterizan, frente a otros procedimientos más tradicionales, por no requerir de ninguna hipótesis paramétrica acerca de los datos, lo que las convierte en técnicas extremadamente versátiles que pueden ser aplicadas en multitud de situaciones prácticas y en diferentes contextos.

A lo largo de la presente memoria se han desarrollado algunos procedimientos novedosos para el análisis discriminante y el análisis cluster de series temporales, desde una perspectiva no paramétrica. Estos procedimientos se fundamentan en general en la definición de una distancia entre series de tiempo definida a partir de sus correspondientes densidades espectrales, que son estimadas a su vez mediante técnicas de regresión polinómica local. Ya se ha recalado en diversas ocasiones a lo largo de este trabajo la multitud de situaciones prácticas en las que pueden tener aplicación este tipo de técnicas. Nótese que, cuando varias series temporales se analizan conjuntamente, además de obtener un modelo que las ajuste correctamente para comprender cómo se comportan dichas series, puede resultar también interesante investigar el grado de similitud entre ellas. Las técnicas aquí propuestas resultarán de utilidad, por lo tanto, en este tipo de análisis y pueden ser utilizadas junto con otras técnicas paramétricas o no paramétricas para realizar un análisis

exhaustivo de un conjunto de series temporales.

En este capítulo se propone una estrategia de análisis general, basada en técnicas de corte no paramétrico, para investigar el grado de afinidad entre un conjunto de series de tiempo. Dicha estrategia consiste en estudiar separadamente las tendencias y las componentes estocásticas de cada una de las series observadas, utilizando para ello técnicas tipo núcleo. Más específicamente, el procedimiento de análisis que se propone consta de las siguientes etapas:

(i) En primer lugar, la tendencia de cada una de las series temporales es estimada mediante técnicas de regresión polinómica local asumiendo que se trata de funciones suaves.

(ii) En segundo lugar, las tendencias así ajustadas se clasifican en grupos similares utilizando para ello técnicas cluster habituales.

(iii) Para cada uno de los grupos de tendencias identificados en el análisis cluster, se contrasta la hipótesis de igualdad de tendencias mediante tests de corte no paramétrico que tomen en cuenta la estructura de dependencia de los datos.

(iv) Finalmente, y una vez completado el análisis de las tendencias, se desarrolla un análisis cluster de las series de residuos no paramétricos correspondientes utilizando, para ello, las técnicas cluster introducidas en el Capítulo 4 de la presente memoria.

Para ilustrar la aplicación de la metodología propuesta se utilizará un conjunto de datos reales correspondiente a las series de cotización bursátil del sector de la banca en España, durante los años 2001 y 2002. El análisis del sector bancario resulta especialmente interesante desde un punto de vista económico, dada su gran importancia en el mercado bursátil (su valor bursátil representa un porcentaje importante del valor de la bolsa en España) y la gran heterogeneidad de las compañías de este sector, en cuanto a tamaño, expectativas de crecimiento, riesgos y estrategias de negocio. En cualquier caso, el objetivo de este capítulo no consiste en estudiar las implicaciones económicas o financieras que puedan derivarse del análisis, sino el demostrar la versatilidad y la potencia de las técnicas no paramétricas para estimar, comparar y clasificar las series observadas. De hecho, esta misma estrategia de análisis podrá aplicarse a cualquier otro conjunto de datos temporales, siguiendo los mismos pasos.

Este capítulo se estructurará de la siguiente forma. En la Sección 5.2 se presentarán la series de cotización bursátil que se analizaron. En la Sección 5.3 se expondrán los resultados del análisis de las componentes deterministas de las series. La estimación de las tendencias se abordará mediante técnicas de regresión polinómica local, para realizar posteriormente un análisis cluster de las mismas. Se contrastará asimismo la igualdad de las tendencias de las series consideradas, y de aquellas series agrupadas en un mismo cluster.

Puesto que los tests estadísticos empleados están basados en técnicas no paramétricas, su distribución es en general desconocida y se considerarán tanto aproximaciones asintóticas como bootstrap de la misma. En la Sección 5.4 se abordará el análisis cluster de las series de residuos, utilizando para ello las técnicas descritas en el Capítulo 4 de este trabajo. Se realizará un análisis exhaustivo de los resultados alcanzados según la medida de disparidad utilizada para evaluar la discrepancia o distancia entre las series. En la Sección 5.5 se demostrará la utilidad general de la estrategia de análisis propuesta a través de su aplicación a un conjunto de datos simulados, comparando el comportamiento de los procedimientos cluster propuestos en esta memoria con los alcanzados con otros criterios desarrollados con anterioridad en la literatura. Finalmente, se expondrán algunas conclusiones en la Sección 5.6.

Los resultados expuestos en este capítulo han sido condensados en Vilar, Vilar y Pértega (2004, 2007).

5.2. Presentación de las series de cotización bursátil

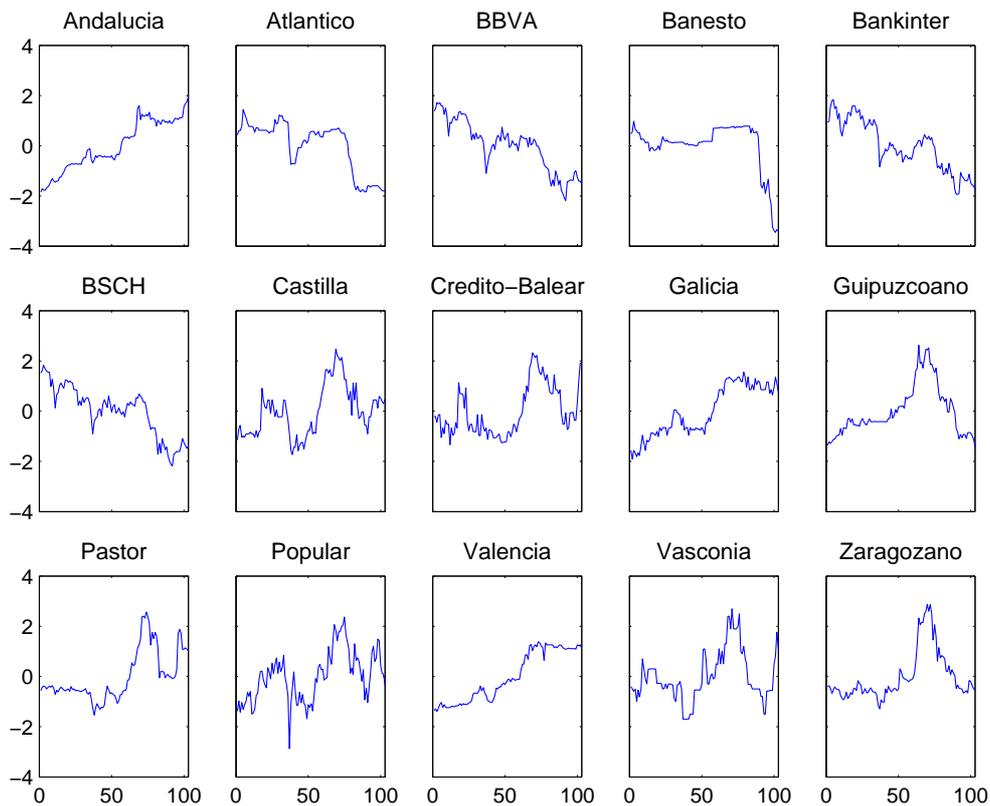
Se analizan quince series de datos financieros que recogen el precio semanal (expresado en euros por acción) de otros tantos bancos que cotizan en el mercado bursátil español durante un periodo de dos años (2001 y 2002). Las quince series corresponden a las siguientes entidades bancarias: Andalucía, Atlántico, BBVA, Banesto, Bankinter, BSCH, Castilla, Crédito-Balear, Galicia, Guipuzcoano, Pastor, Popular, Valencia, Vasconia y Zaragoza. Dichos bancos representan la totalidad del sector de la banca en España que cotiza en bolsa y cinco de ellos (Banesto, Bankinter, BBVA, BSCH and Popular) pertenecen al IBEX-35, que agrupa a las 35 compañías de mayor peso en el mercado bursátil español. Los datos han sido obtenidos de la página web www.joramon.com.

Cada una de las series de tiempo $\mathbf{X}_l = (X_{l,1}, \dots, X_{l,n})$, $l = 1, \dots, 15$ fue evaluada en los mismos momentos de tiempo. Más específicamente, cada serie consta de $n=103$ observaciones semanales correspondientes a los precios por acción, en euros, registrados todos los jueves durante los años 2001 y 2002. En aquellas semanas en las que el jueves fuese día festivo la observación correspondiente fue reemplazada por el precio por acción del miércoles anterior, dado que en otros estudios donde el precio se registró los viernes han obtenido conclusiones erróneas debido al llamado efecto fin de semana.

A efectos de comparación, todas las series han sido estandarizadas de forma que presentasen media cero y varianza unidad. Las series resultantes se muestran en la Figura 5.1. Como se puede apreciar, existe un gran parecido entre alguna de las series observadas (por ejemplo, las correspondientes a los bancos BBVA y BSCH muestran ambas

patrones claramente descendentes a lo largo de todo el periodo de estudio), mientras que otras muestran una evolución claramente diferente (por ejemplo, las correspondientes al banco Andalucía y el BBVA). Utilizando un paquete estadístico estándar se ha podido comprobar que cada una de las series observadas puede ser modelizada razonablemente bien mediante un modelo autorregresivo o un modelo de medias móviles. Sin embargo, en el presente trabajo, el interés no se centra en ajustar un determinado modelo a cada una de las series sino en clasificarlas en grupos similares. Por lo tanto, el objetivo final consiste en encontrar similitudes y diferencias entre las series observadas con el fin de clasificarlas en un número pequeño de grupos. Dicha clasificación puede ser de utilidad para detectar unos pocos patrones representativos, predecir comportamientos futuros, cuantificar el grado de afinidad, etc.

Figura 5.1: Precios semanales estandarizados (en euros por acción) del sector bancario español en bolsa durante los años 2001-2002.



Las técnicas cluster habituales no resultan adecuadas para afrontar este proceso de clasificación, ya que no toman en cuenta la caracterización probabilística de las series de tiempo. Dado que de la Figura 5.1 se deduce claramente la existencia de un patrón o tendencia subyacente a cada una de las series observadas, para un análisis adecuado, y tal y como ya se comentó, se dividirá cada serie en su componente determinista y su componente aleatoria, que serán analizadas de modo independiente.

5.3. Análisis de las tendencias

El primer paso del análisis consistió en estimar la tendencia individual correspondiente a cada serie, para después clasificar y comparar dichas tendencias entre sí.

Puesto que todas las series han sido evaluadas en los mismos puntos, los datos pueden ajustarse a un modelo de regresión de diseño fijo, esto es

$$X_{l,t} = \varphi_l(z_t) + \varepsilon_{l,t}, \quad t=1,2, \dots, n, \quad l=1,2, \dots, k, \quad (5.1)$$

donde $n=103$ es la longitud de las series, $k=15$ el número de series consideradas, $\varphi_l(\cdot)$ es la tendencia de la l -ésima serie, $z_t = t/n$ y $\varepsilon_{l,t}$ denota a la componente aleatoria correspondiente.

Para estimar cada una de las tendencias individuales φ_l a partir de los datos observados se han utilizado técnicas de regresión polinómica local. Este tipo de suavizadores presentan ventajas importantes frente a otros métodos de estimación tipo núcleo, como: mejor comportamiento en las fronteras, permiten estimar además las derivadas de la función de regresión, son más sencillos de obtener en términos computacionales y presentan buenas propiedades minimax (ver Fan y Gijbels (1996)).

Tal y como se expuso en la Sección 1.1, el estimador polinómico local se obtiene ajustando localmente un polinomio de grado p a los datos mediante el método de mínimos cuadrados locales ponderados. Dada cada una de las series $\{(z_t, X_{l,t})\}_{t=1}^n$, que se consideran generadas a partir del modelo (5.1) para $l \in \{1, \dots, k\}$, y asumiendo que existen las primeras $(p + 1)$ derivadas de la función $\varphi_l(x)$ y son continuas, las técnicas de regresión polinómica local permiten estimar el vector $\beta_l(z) = (\beta_{l,0}(z), \beta_{l,1}(z), \dots, \beta_{l,p}(z))^t$, donde $j! \beta_{l,j}(z) = \varphi_l^{(j)}(z)$, $j = 0, 1, \dots, p$, minimizando la función

$$\Psi(\beta_l(z)) = \sum_{t=1}^n \left(X_{l,t} - \sum_{j=0}^p \beta_{l,j}(z)(z_t - z)^j \right)^2 w_{l,t}, \quad (5.2)$$

donde los pesos $w_{l,t}$ vienen dados como $w_{l,t} = (nh_l)^{-1} K((z_t - z)/h_l)$, con $K(\cdot)$ una función núcleo que corresponde a una función de densidad unimodal y simétrica en torno a cero, y $h_l = h_{l,n}$ el parámetro de suavizado o ventana. El estimador de $\beta_l(z)$, obtenido como solución al problema de mínimos cuadrados locales ponderados dado en (5.2), suele denominarse estimador núcleo polinómico local o suavizador polinómico local.

El ancho de banda h_l se determinó en todos los casos mediante un algoritmo plug-in específicamente diseñado para solventar el problema de dependencia de los datos (Francisco, Opsomer y Vilar (2004)). La clave de este algoritmo consiste en estimar un modelo de

correlación paramétrico que se asume para el proceso de errores y “colocar” ese estimador en la expresión asintótica del error cuadrático medio integrado del suavizador. De forma más detallada, se trata de buscar la ventana h_l para la cual se minimiza el error cuadrático medio integrado asintótico, dado por

$$h_{l,n}^{opt} = C_p(K) \left(\frac{\Gamma_l}{n \int (\varphi_l^{(p+1)}(z))^2 dz} \right)^{\frac{1}{2p+3}}, \quad (5.3)$$

donde $C_p(K)$ es un número real que depende del núcleo $K(\cdot)$ elegido, y $\Gamma_l = \sum_{\tau=-\infty}^{\infty} v_l(\tau)$, con $v_l(\tau) = E(\varepsilon_{l,i} \varepsilon_{l,i+\tau})$. Por lo tanto, la ventana plug-in $\hat{h}_{l,n}$ puede obtenerse directamente reemplazando las cantidades desconocidas Γ_l y $\varphi_l^{(p+1)}$ en (5.3) por estimaciones apropiadas.

Existen diversos enfoques para estimar Γ_l y $\varphi_l^{(p+1)}$. En lo que atañe a Γ_l , una primera alternativa consiste en asumir una estructura paramétrica de dependencia para el proceso de errores que involucre a un número pequeño de parámetros (por ejemplo, un modelo ARMA). En ese caso, los parámetros del modelo podrán ser estimados a partir de los residuos no paramétricos $\hat{\varepsilon}_{l,t} = X_{l,t} - \hat{\varphi}_{l,h_{l,pilot}}(z_t)$, generados tras utilizar una ventana piloto $h_{l,pilot}$. Un enfoque alternativo consiste en estimar Γ_l directamente a partir de los datos utilizando métodos basados en diferencias. Así, Müller y Stadtmüller (1988) propusieron un estimador para Γ_l basado en las diferencias de primer orden de $X_{l,t}$ para errores m -dependientes. En la misma línea, Hermann, Gasser y Kneip (1992) sugirieron un estimador basado en las diferencias de segundo orden de $X_{l,t}$ asumiendo que los errores satisfacen algunas condiciones mixing.

En lo que respecta a la estimación de $\varphi_l^{(p+1)}$, básicamente puede utilizarse un enfoque paramétrico o no paramétrico. La primera alternativa consiste fundamentalmente en asumir que la función de regresión puede ajustarse mediante un polinomio de grado $(p+3)$, $\tilde{\varphi}_l$, y estimar posteriormente $\varphi_l^{(p+1)}$ calculando la derivada de orden $(p+1)$ de la curva ajustada, $\tilde{\varphi}_l^{(p+1)}(x)$. El enfoque no paramétrico se basa en estimar $\varphi_l^{(p+1)}$ utilizando regresión polinómica local tal y como se explicó anteriormente.

Una explicación más detallada del problema de selección de la ventana en el contexto de datos dependientes, incluyendo el algoritmo plug-in anterior, puede verse en Francisco-Fernández y Vilar-Fernández (2001, 2004).

De acuerdo con todo lo expuesto, para la estimación de cada una de las tendencias φ_l correspondientes a las series estudiadas se utilizó un suavizador polinómico local con $p = 1$ (suavizador lineal local). En todos los casos, la ventana utilizada se estimó utilizando el algoritmo plug-in descrito anteriormente. En particular, Γ_l se estimó asumiendo una estructura ARMA para los procesos de errores, cuyos parámetros se estimaron a partir de

los residuos no paramétricos correspondientes. Las ventanas así obtenidas se muestran en la Tabla 5.1.

Tabla 5.1: Ventanas utilizadas para estimar la tendencia de las series, calculadas mediante un algoritmo plug-in.

Banco	Andalucía	Atlántico	BBVA	Banesto	Bankinter
h	0,134229	0,087754	0,167225	0,175560	0,175827
Banco	BSCH	Castilla	Crédito-Balear	Galicia	Guipuzcoano
h	0,178553	0,111946	0,080087	0,092995	0,114639
Banco	Pastor	Popular	Valencia	Vasconia	Zaragozano
h	0,115431	0,136151	0,158361	0,112908	0,167420

Además de cada una de las tendencias individuales φ_l , se estimó la tendencia media del conjunto de 15 series, que se denotará por φ_c . La curva φ_c representa así la evolución de los precios semanales promedio (en euros por acción) del sector bancario español durante los años 2001 y 2002, y se estimó también no paramétricamente a partir de la nube de puntos $\{(z_t, \bar{X}_{\cdot,t}), t = 1, \dots, 103\}$ mediante un suavizador lineal local, donde $\bar{X}_{\cdot,t} = (1/15) \sum_{l=1}^{15} X_{l,t}$. La Figura 5.2 muestra un gráfico con cada una de las series observadas, junto con el estimador lineal local de las tendencias individuales y de la curva promedio φ_c .

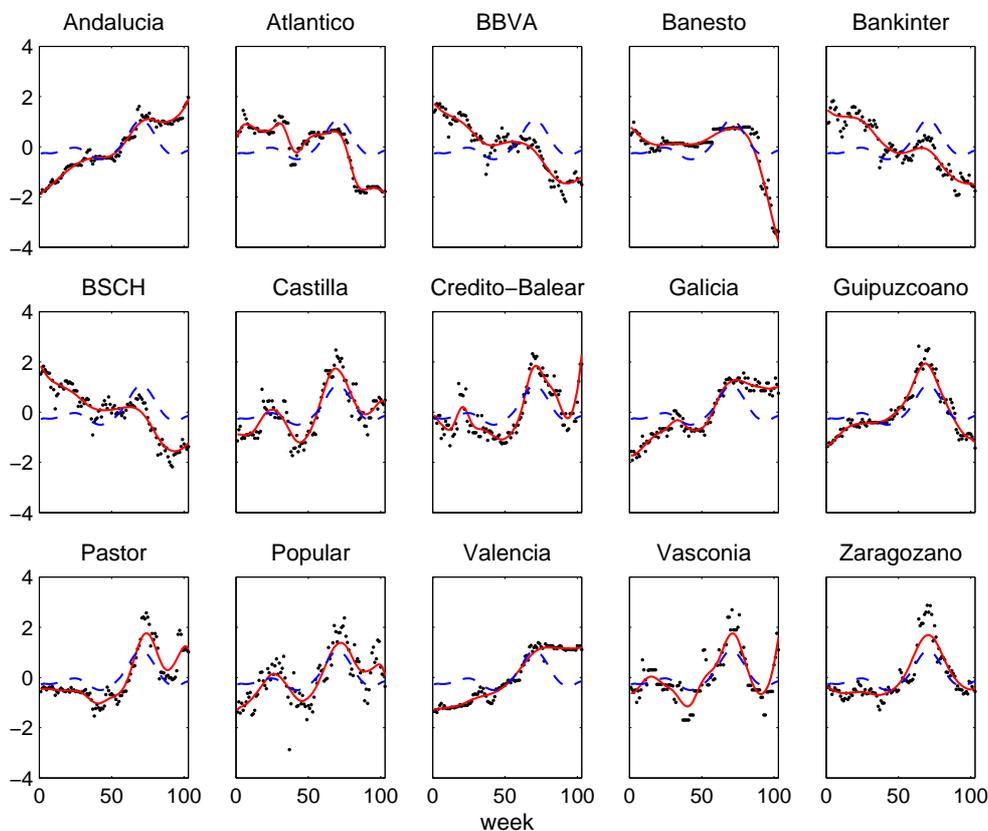
La Figura 5.2 sugiere que no puede asumirse que exista una tendencia común a todas las series, sino que existen diferencias importantes entre las tendencias individuales de todas ellas. Desde un punto de vista formal, cabe plantearse entonces el siguiente contraste de hipótesis:

$$\begin{aligned}
 H_0 : \varphi_1 = \dots = \varphi_{15} = \varphi_c \text{ versus} \\
 H_1 : \text{Existen } (l, j) \text{ tales que } \varphi_l \neq \varphi_j, l, j \in \{1, \dots, 15\}.
 \end{aligned}
 \tag{5.4}$$

El problema de contraste de igualdad de k funciones de regresión mediante técnicas no paramétricas ha sido ampliamente abordado en la literatura estadística durante los últimos años. Algunas referencias importantes a este respecto son Härdle y Marron (1990), Hall y Hart (1990), King, Hart y Werhly (1991), Kulasekera (1995), Koul y Schick (1997), Kulasekera y Wang (1997, 1998), Dette y Neumeyer (2001) y Neumeyer y Dette (2003), entre otras. Sin embargo, todos estos trabajos se centran en el caso de que las observaciones disponibles sean independientes, hipótesis que no se verifica en el contexto aquí planteado.

En un trabajo reciente, Vilar-Fernández y González Manteiga (2004) han abordado el problema de contraste de la igualdad de k funciones de regresión con errores dependientes

Figura 5.2: Gráficos de las series observadas junto con el el suavizador lineal local correspondiente a cada tendencia individual φ_l , $l = 1, \dots, 15$, (línea rosa) y a la tendencia media φ_c (línea azul).



en un contexto muy general. Para ello, proponen utilizar como test estadístico una distancia funcional entre los estimadores no paramétricos de las diferentes funciones de regresión φ_l . Hasta el momento, no se dispone de otros trabajos que hayan abordado este problema en contextos de dependencia. En la presente memoria, utilizaremos el estadístico propuesto por dichos autores para abordar el contraste de hipótesis en (5.4), proporcionando a continuación una breve descripción del procedimiento utilizado.

De acuerdo con Vilar-Fernández y González Manteiga (2004), se propone realizar el contraste de igualdad de $k \geq 2$ curvas de regresión mediante un test estadístico basado en una distancia funcional del tipo Crámer-von-Mises entre los estimadores núcleo de cada una de las funciones de regresión desconocidas φ_l . De modo más detallado, dicho estadístico vendrá dado por:

$$\hat{Q}_n^{(1)} = \sum_{l=2}^k \sum_{s=1}^{l-1} \int (\hat{\varphi}_{l,h}(z) - \hat{\varphi}_{s,h}(z))^2 \omega(z) dz, \quad (5.5)$$

donde $w(z)$ es una función peso definida en el soporte de las variables del diseño, $C = [0, 1]$, y, para cada $l = 1, \dots, k$, $\hat{\varphi}_{l,h}$ es un estimador tipo núcleo de φ_l calculado con una ventana h y función kernel K .

Puede demostrarse que $\hat{Q}_n^{(1)}$ es un estimador consistente de

$$Q = \sum_{l=2}^k \sum_{s=1}^{l-1} \int (\varphi_l - \varphi_s)^2 \omega,$$

de modo que se rechazará la hipótesis nula de igualdad de las funciones de regresión sólo si $Q > 0$.

Diversos autores han estudiado el comportamiento asintótico del estadístico $\hat{Q}_n^{(1)}$ en un contexto de independencia. Algunas referencias interesantes son King, Hart y Wehrly (1991), Kulasekera (1995) o Kulasekera y Wang (1997, 1998) para el caso en el que $k = 2$. Más recientemente, Dette y Neumeyer (2001) han estudiado el caso más general en el que $k > 2$ bajo condiciones de heterocedasticidad y diferentes tipos de diseño en cada grupo, además de demostrar la normalidad asintótica del estadístico en (5.5). La consistencia de una versión bootstrap del mismo test ha sido también establecida en dicho trabajo.

Para el contexto de datos dependientes, Vilar-Fernández y González-Manteiga (2004) han estudiado también el comportamiento asintótico del estimador (5.5) para $k \geq 2$ funciones de regresión y condiciones generales sobre el diseño en cada uno de los grupos. En particular, si la función de autocovarianzas de cada uno de los procesos de error en (5.1), $v_l(\tau) = E(\varepsilon_{l,t}\varepsilon_{l,t+\tau})$ satisface

$$\sum_{\tau=-\infty}^{\infty} |\tau| |v_l(\tau)| < \infty, \text{ para } l = 1, \dots, k,$$

el Teorema 3 en Vilar-Fernández y González-Manteiga (2004) establece que, bajo la hipótesis nula de igualdad de las curvas de regresión, el estadístico (5.5) tiene una distribución asintótica normal según:

$$\sqrt{n^2 h} \left(\hat{Q}_n^{(1)} - \frac{k-1}{nh} I_\omega C_K \sum_{l=1}^k \Gamma_l \right) \xrightarrow{D} N(0, \sigma_{Q^{(1)}}^2), \quad (5.6)$$

con $\Gamma_l = \sum_{\tau=-\infty}^{\infty} v_l(\tau)$, para $l = 1, \dots, k$, $I_\omega = \int \omega(z) dz$, $C_K = \int K^2(z) dz$ y

$$\sigma_{Q^{(1)}}^2 = 2 \left(\int (K * K)^2 \right) \left(\int \omega^2 \right) \left[(k-1)^2 \sum_{l=1}^k \Gamma_l + \sum_{j=1}^k \sum_{l=1, l \neq j}^k \Gamma_j \Gamma_l \right]$$

donde $*$ denota al operador convolución.

Los parámetros Γ_l en la expresión asintótica (5.6) son desconocidos por lo que, en la práctica, debe obtenerse alguna estimación de dichos parámetros, $\hat{\Gamma}_l$, a partir de las muestras observadas. Si se consideran estimaciones adecuadas $\hat{\Gamma}_l$, entonces la convergencia en (5.6) se mantiene cuando Γ_l es reemplazado por $\hat{\Gamma}_l$, $l = 1, \dots, k$, (ver Teorema 2 en Vilar-Fernández y González-Manteiga (2004)). Por lo tanto, la hipótesis nula de igualdad de las k curvas de regresión se rechazará con un nivel de significación α cuando

$$\tilde{Q}_n^{(1)} = \sqrt{n^2 h} \hat{\sigma}_{Q^{(1)}}^{-1} \left(\hat{Q}_n^{(1)} - \frac{k-1}{nh} I_\omega C_K \sum_{l=1}^k \hat{\Gamma}_l \right) > z_\alpha, \quad (5.7)$$

donde z_α es tal que $\Phi(z_\alpha) = 1 - \alpha$, denotando $\Phi(\cdot)$ a la función de distribución de una variable normal estándar.

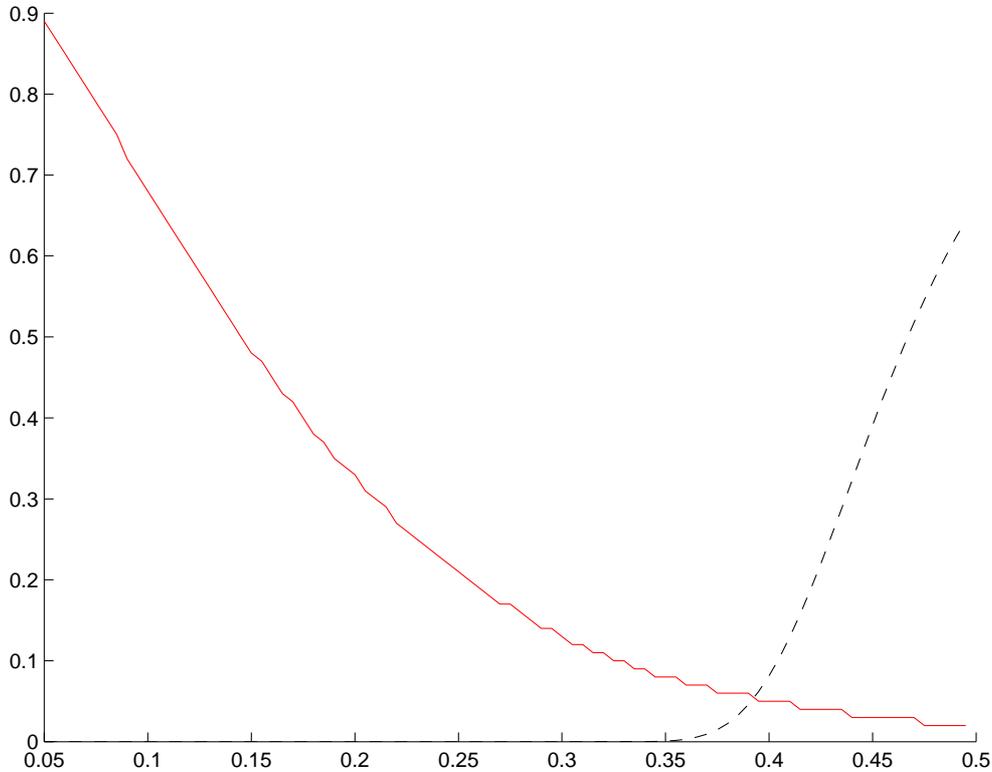
Se empleó el criterio (5.7) para contrastar la igualdad de las tendencias de las quince series estudiadas. En particular, para estimar cada una de las tendencias individuales φ_l se empleó el estimador de Nadaraya-Watson (correspondiente a un suavizador polinómico local con $p = 0$) y núcleo Gaussiano. Se tomó como función peso $\omega(z) = 1/0,8$ en $[0,1,0,9]$ y 0 en otro caso. Los residuos no paramétricos obtenidos de la forma $\hat{\varepsilon}_{l,t} = X_{l,t} - \hat{\varphi}_{l,h}(z_t)$ se utilizaron para estimar Γ_l , $l = 1, \dots, 15$, y el estadístico $\tilde{Q}_n^{(1)}$ se calculó para distintos valores de la ventana h . Los resultados obtenidos se muestran en la Figura 5.3 y la Tabla 5.2. La Figura 5.3 muestra los valores de $\tilde{Q}_n^{(1)} \times 10^{-2}$ y los p-valores asociados a cada $\tilde{Q}_n^{(1)}$ como función de la ventana h . A su vez, en la Tabla 5.2 se muestran los resultados para algunos valores concretos de h .

Tabla 5.2: Test de igualdad de las tendencias de las quince series observadas según el estadístico $\tilde{Q}_n^{(1)}$ para algunos valores específicos de h

h	0,100	0,200	0,300	0,350	0,400	0,500
$\tilde{Q}_n^{(1)}$	68,126	32,574	13,091	8,085	5,005	3,135
p -value	0,000	0,000	0,000	0,001	0,081	0,390

Tal y como cabía esperar, en la Figura 5.3 se aprecia cómo la igualdad en las tendencias se rechaza para un amplio rango de valores razonables de la ventana h . A medida que el ancho de banda h aumenta, se tiende a aceptar la hipótesis nula de igualdad. Este comportamiento resulta lógico, pues cuanto mayor sea la ventana, se introducirá un mayor grado de suavización y las tendencias individuales estimadas serán más parecidas, aproximándose al promedio de las observaciones. Es más, por otro lado, y puesto que las series han sido estandarizadas previamente, la curva promedio coincide con la línea $Y = 0$, por lo que el test $\tilde{Q}_n^{(1)}$ en (5.7) tiende a cero a medida que la ventana h se incrementa. En consecuencia, para valores grandes de h la hipótesis nula se acepta con mayor facilidad y

Figura 5.3: Gráfico de $\tilde{Q}_n^{(1)} \times 10^{-2}$ (línea rosa) y los p-valores asociados con $\tilde{Q}_n^{(1)}$ (línea azul) en función del ancho de banda h .



el test pierde potencia. En definitiva, el ancho de banda h ejerce una gran influencia en las decisiones que se puedan derivar de dicho test, de modo que una elección adecuada del parámetro ventana es esencial para obtener un test con un poder estadístico aceptable.

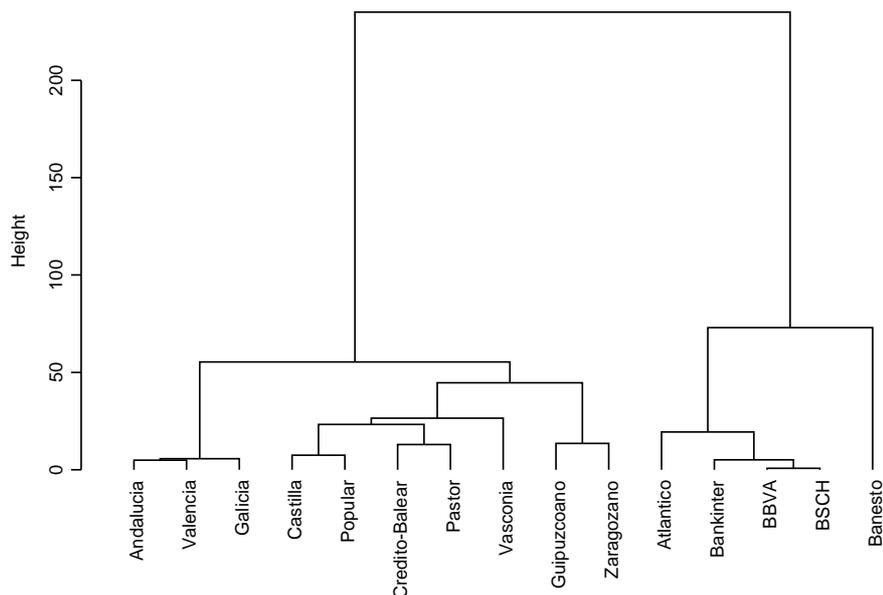
Tras rechazar la hipótesis nula de igualdad de las tendencias de las quince series se procedió a realizar un análisis cluster sobre las tendencias individuales correspondientes, estimadas no paramétricamente, con el fin de analizar el grado de afinidad entre ellas y agruparlas en un número reducido de clases. Para ello, se realizaron diferentes análisis de cluster jerárquico utilizando como medida de proximidad entre dos patrones de comportamiento el cuadrado de la distancia euclídea. De un modo más detallado, la distancia entre dos tendencias estimadas, $d(\hat{\varphi}_{l,h_l}, \hat{\varphi}_{s,h_s})$, para $l, s \in \{1, 2, \dots, 15\}$, vendrá dada por

$$d(\hat{\varphi}_{l,h_l}, \hat{\varphi}_{s,h_s}) = \sum_{t=1}^{103} [\hat{\varphi}_{l,h_l}(z_t) - \hat{\varphi}_{s,h_s}(z_t)]^2.$$

Se emplearon diferentes criterios para la sucesiva agrupación de las tendencias en cada paso del proceso jerárquico, incluyendo el método de Ward y los métodos de enlace simple, enlace completo y enlace promedio. En todos los casos se obtuvieron resultados simila-

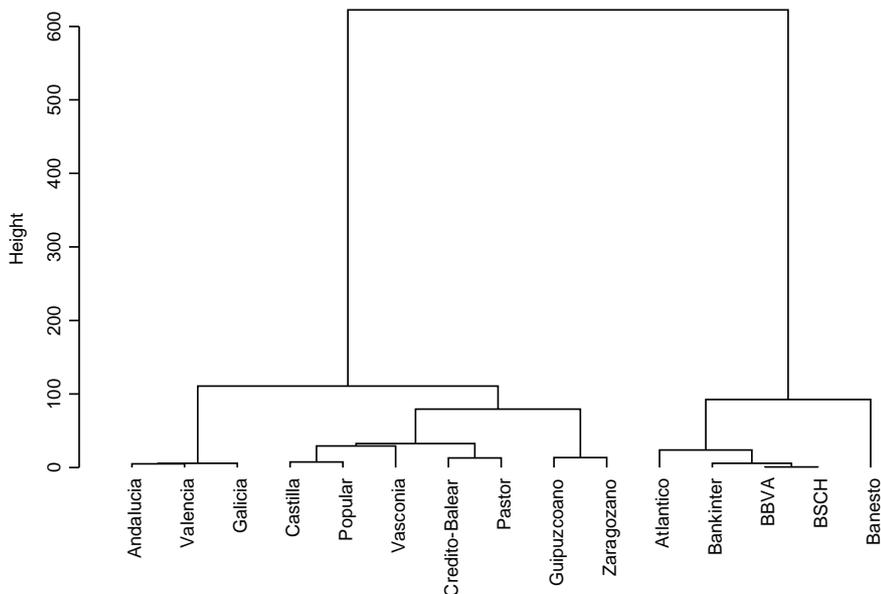
res, que pueden ser resumidos gráficamente mediante los correspondientes dendogramas. Las Figuras 5.4 y 5.5 muestran los dendogramas correspondientes al proceso de cluster empleando el método de enlace promedio y el método de Ward, respectivamente.

Figura 5.4: Dendograma del análisis cluster sobre los suavizadores lineales locales de las tendencias de cada una de las quince series financieras. Resultados con la distancia euclídea al cuadrado y el método de enlace promedio.



Los dendogramas que se muestran en las Figuras 5.4 y 5.5 identifican un total de cinco clusters. Existe un primer grupo, formado por los bancos BBVA, BSCH, Bankinter, Atlántico y Banesto, claramente diferenciado del resto. Dentro de este primer grupo, $C1 = \{BBVA, BSCH, Bankinter, Atlántico\}$ forman un cluster compacto, con una tendencia a la baja a lo largo de todo el periodo de estudio y sin muchas oscilaciones, mientras que $C2 = \{Banesto\}$ se mantiene aislado hasta las últimas etapas del proceso cluster, debido probablemente a que su cotización se mantiene más o menos constante hasta aproximadamente las últimas veinticinco semanas de estudio, momento en el que se registra una notable caída en los precios por acción. Con respecto al resto de las series, se observan tres clusters diferenciados: $C3 = \{\text{Andalucía, Valencia, Galicia}\}$, $C4 = \{\text{Guipuzcoano, Zaragoza}\}$ y $C5 = \{\text{Castilla, Popular, Crédito-Balear, Pastor, Vasconia}\}$. En oposición al primer cluster $C1$, la cotización de los bancos incluidos en $C3$ muestran una tendencia creciente en su cotización bursátil durante todo el periodo de estudio. Para el grupo $C5$, el patrón es una curva oscilante que presenta dos modas en torno a las semanas 25 y 70. Finalmente, el patrón en el cluster $C4$ es similar al del grupo $C5$ pero aquí desaparece la

Figura 5.5: Dendrograma del análisis cluster sobre los suavizadores lineales locales de las tendencias de cada una de las quince series financieras. Resultados con la distancia euclídea al cuadrado y el método de Ward.

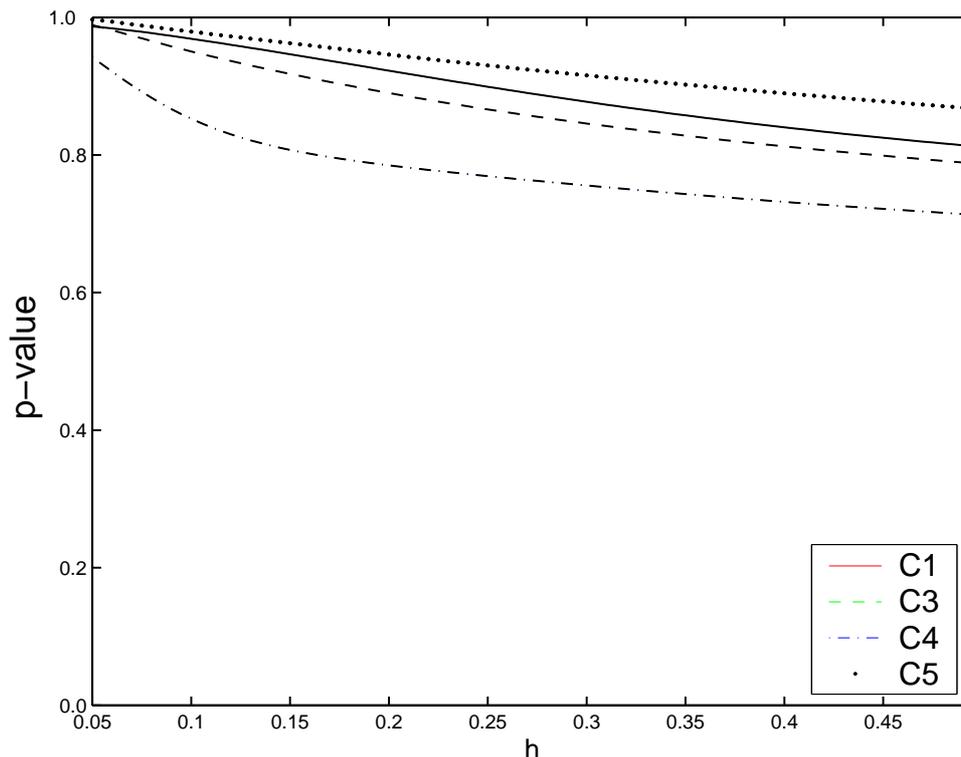


primera moda, de modo que la cotización es más o menos constante en el periodo inicial de estudio, y la segunda moda es todavía más pronunciada. Cabe destacar que, a pesar de que el banco Popular pertenece al IBEX-35, no aparece incluido en el primer cluster, C_1 , en el que sí se incluyen el resto de los bancos del IBEX-35 (BBVA, BSCH, Bankinter and Banesto), excepto en el caso del Banesto que, en cualquier caso, se encuentra próximo a este grupo.

Llegados a este punto, es importante recordar que los resultados del análisis cluster no son necesariamente únicos, dado que dependen de la medida de distancia utilizada, del método cluster y del propio investigador. Así, puede ser necesario estudiar con mayor detenimiento hasta qué punto son similares las series clasificadas en un mismo grupo, antes de concluir que cualquiera de ellas puede ser representativa de todas las series en ese cluster. Con esa finalidad, una vez analizados los resultados del análisis cluster, se procedió a contrastar la igualdad de las tendencias de las series que han quedado clasificadas en cada uno de los grupos. Para ello se utilizó de nuevo el test estadístico $\tilde{Q}_n^{(1)}$ en (5.7) para contrastar la igualdad de las tendencias incluidas en cada uno de los clusters: C_1 , C_3 , C_4 y C_5 . De nuevo, la Figura 5.6 muestra un gráfico con los p-valores obtenidos para cada grupo como función de la ventana h utilizada para suavizar las tendencias. De dicha

figura se concluye que la hipótesis de igualdad puede aceptarse en cualquiera de los cuatro clusters para cualquiera de los anchos de banda considerados. En todos los casos, además, el p-valor asociado al contraste ha resultado mayor de 0.7.

Figura 5.6: Gráfico de los p-valores asociados a $\tilde{Q}_n^{(1)}$ como función de h para el contraste de igualdad de las tendencias incluidas en cada uno de los grupos identificados en el análisis cluster.



Puede resultar también interesante contrastar la igualdad dos a dos de las tendencias agrupadas o no dentro de un mismo cluster. Nuevamente, desde un punto de vista más formal, tomando $k = 2$ en (5.4), se trata de abordar el contraste de hipótesis:

$$H_0 : \varphi_l - \varphi_s = 0 \quad \text{versus} \quad H_1 : \varphi_l - \varphi_s \neq 0. \quad (5.8)$$

para $l, s \in \{1, \dots, k\}, l \neq s$.

Existen diversos métodos basados en estimaciones no paramétricas para contrastar la igualdad de $k = 2$ curvas de regresión aunque, tal y como ya se ha indicado, han sido estudiados generalmente bajo condiciones de independencia. Además, incluso bajo condiciones de independencia, la distribución muestral de estos estadísticos resulta difícil de obtener y solamente se dispone de aproximaciones asintóticas a las mismas. Puesto que dichas aproximaciones con frecuencia involucran a parámetros poblacionales desconocidos

(como ocurría, por ejemplo, en el caso del estadístico $\tilde{Q}_n^{(1)}$), y presentan tasas de convergencia bastante lentas, algunos autores han propuesto la utilización de procedimientos de remuestreo para poder aproximar la distribución muestral de dichos estadísticos. En este sentido, se han desarrollado algunos algoritmos bootstrap también bajo condiciones de dependencia.

En este trabajo se han empleado hasta cuatro tests estadísticos diferentes de tipo no paramétrico para abordar el contraste de hipótesis (5.8) en el que se investiga la igualdad de cada uno de los pares de tendencias de los cinco bancos incluidos en el IBEX-35: BBVA, Banesto, Bankinter, BSCH y Popular. Además del estadístico $\tilde{Q}_n^{(1)}$ dado en (5.7), se han empleado otros tres estadísticos que se denotarán por $\tilde{Q}_n^{(2)}$, $\tilde{Q}_n^{(3)}$ y $\tilde{Q}_n^{(4)}$ y que se describen a continuación.

a) El primero de dichos estadísticos, que se denotará por $\tilde{Q}_n^{(2)}$, se calcula como la diferencia entre una estimación no paramétrica de la varianza de la muestra conjunta, $\hat{\sigma}_P^2$, y una combinación convexa de estimaciones no paramétricas de la varianza de cada una de las muestras individuales. De un modo más detallado, $\tilde{Q}_n^{(2)}$ vendrá dado por

$$\tilde{Q}_n^{(2)} = \hat{\sigma}_P^2 - \frac{1}{2n} (\hat{\sigma}_l^2 + \hat{\sigma}_s^2), \quad (5.9)$$

con

$$\hat{\sigma}_P^2 = \frac{1}{2n} \left\{ \sum_{t=1}^n [X_{l,t} - \hat{\varphi}_{p,g}(z_t)]^2 + \sum_{t=1}^n [X_{s,t} - \hat{\varphi}_{p,g}(z_t)]^2 \right\}$$

y, para $j = l, s$,

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_{t=1}^n [X_{j,t} - \hat{\varphi}_{i,h_j}(z_t)]^2,$$

donde $\hat{\varphi}_{p,g}$ denota al suavizador núcleo con ancho de banda g de las dos muestras combinadas y $\hat{\varphi}_{j,h_j}$ denota al suavizador núcleo con ventana h_j de cada una de las muestras individuales para $j = l, s$. Bajo condiciones de independencia, el estadístico $\tilde{Q}_n^{(2)}$ ha sido estudiado por Dette y Neumeyer (2001). Previamente, el estimador de la varianza había sido propuesto por Hall y Marron (1990).

b) El segundo test estadístico que se utilizará, y que denotaremos por $\tilde{Q}_n^{(3)}$, es de tipo ANOVA y vendrá dado por

$$\tilde{Q}_n^{(3)} = \frac{1}{2n} \left\{ \sum_{t=1}^n [\hat{\varphi}_{p,g}(z_t) - \hat{\varphi}_{l,h_l}(z_t)]^2 + \sum_{t=1}^n [\hat{\varphi}_{p,g}(z_t) - \hat{\varphi}_{s,h_s}(z_t)]^2 \right\}, \quad (5.10)$$

siendo $\hat{\varphi}_{p,g}$, $\hat{\varphi}_{l,h_l}$ y $\hat{\varphi}_{s,h_s}$ los suavizadores tipo núcleo utilizados para calcular $\tilde{Q}_n^{(2)}$. Este test había sido propuesto por Young y Bowman (1995) a partir del clásico análisis de la varianza de un factor. Nótese que existe una clara relación entre $\tilde{Q}_n^{(2)}$ y $\tilde{Q}_n^{(3)}$.

c) Finalmente, se empleará para el contraste el estadístico $\tilde{Q}_n^{(4)}$, propuesto por Neumeyer y Dette (2003). Dicho test se basa en la diferencia de dos procesos "marked empirical", contruidos a partir de los residuos obtenidos bajo la hipótesis nula de la igualdad de ambas tendencias. De este modo, bajo la hipótesis nula, los residuos se obtendrán como $\hat{\varepsilon}_{j,t}^0 = X_{j,t} - \hat{\varphi}_{p,g}(z_t)$, $i = l, s$, y la diferencia entre los correspondientes "marked empirical processes" viene dada por

$$\hat{R}(r) = \frac{1}{2n} \left[\sum_{t=1}^n \hat{\varepsilon}_{l,t}^0 I(z_t \leq r) - \sum_{t=1}^n \hat{\varepsilon}_{s,t}^0 I(z_t \leq r) \right],$$

donde $r \in [0, 1]$ e $I(\cdot)$ denota a la función "indicador". Entonces $\tilde{Q}_n^{(4)}$ puede definirse como

$$\tilde{Q}_n^{(4)} = \int_0^1 \hat{R}^2(r) dr. \quad (5.11)$$

El comportamiento asintótico bajo condiciones de independencia del estadístico $\tilde{Q}_n^{(4)}$ ha sido estudiado en Neumeyer y Dette (2003). En particular, en dicho trabajo se demuestra la capacidad de dicho test para detectar hipótesis alternativas que tienden a la hipótesis nula a una tasa $n^{-1/2}$ y se desarrolla una versión "wild bootstrap" de dicho método.

Con cualquiera de los cuatro tests propuestos, el criterio a seguir consistirá en rechazar la hipótesis nula de igualdad de cada par de tendencias para valores grandes del estadístico correspondiente. En la práctica, se necesitará conocer la distribución muestral de $\tilde{Q}_n^{(i)}$, para $i = 1, 2, 3, 4$, a fin de obtener los valores críticos del test. Desafortunadamente, tal y como hemos comentado, resulta extremadamente complicado determinar la distribución de dichos estadísticos bajo condiciones de dependencia. En algunos casos, como para el test $\tilde{Q}_n^{(1)}$, ha sido posible derivar su distribución asintótica en el caso de trabajar con observaciones dependientes (Vilar-Fernández y González-Manteiga (2004)). En la práctica, sin embargo, un procedimiento alternativo para solventar este problema consiste en estimar los parámetros desconocidos que aparecen en esa distribución estimándolos de algún modo a partir de la muestra, y luego considerar una versión plug-in del test. De cualquier modo, es bien sabido que la tasa de convergencia de la distribución del test resulta normalmente muy lenta y se necesita un tamaño muestral considerable para obtener valores críticos del test razonables.

Una forma alternativa y sencilla de aproximar la distribución de cualquiera de los estadísticos $\tilde{Q}_n^{(i)}$, para $i = 1, 2, 3, 4$, es mediante técnicas bootstrap. En un trabajo reciente, Vilar, Vilar y González (2007) han comparado en un estudio de simulación el comportamiento de tres métodos bootstrap diferentes para aproximar la distribución de los tests estadísticos que aquí consideramos, teniendo en cuenta además las condiciones de dependencia. La idea fundamental de los métodos bootstrap en el contexto de series temporales

radica en asumir que existe alguna estructura paramétrica subyacente a la dependencia de los datos (por ejemplo, un modelo ARMA), hecho que debe tomarse en cuenta en el algoritmo de remuestreo. Así, un primer método bootstrap, que será el utilizado en el presente trabajo, consiste en asumir una estructura paramétrica determinada para el proceso de errores (en este caso, un modelo autorregresivo) y replicar así la estructura de dependencia de los datos. Existen otros algoritmos de remuestreo, más generales, en los que no se asume un modelo de dependencia paramétrico. En dicho caso, el método bootstrap se centra en replicar la dependencia remuestreando un bloque entero de observaciones. Una revisión más detallada de las diferentes técnicas puede verse en Vilar, Vilar y González (2007).

En particular, en la presente memoria, la distribución muestral de cada uno de los tests estadísticos $\tilde{Q}_n^{(i)}$, para $i = 1, 2, 3, 4$, se aproximará mediante un algoritmo bootstrap siguiendo los siguientes pasos:

Paso 1. Se calcula el test estadístico $\tilde{Q}_n^{(i)}$, para $i = 1, 2, 3, 4$, a partir de la muestra inicial dada por $\{(z_t, X_{l,t}, X_{s,t})\}_{t=1}^n$.

Paso 2. Bajo la hipótesis nula de igualdad de las dos funciones de regresión, se estimarán los residuos no paramétricos según

$$\hat{\varepsilon}_{j,t}^0 = X_{j,t} - \hat{\varphi}_{p,g}(z_t), \text{ para } t = 1, \dots, n \text{ y } j = l, s,$$

donde $\hat{\varphi}_{p,g}(\cdot)$ denota el suavizador núcleo del total de la muestra conjunta con ventana g .

Paso 3. Se obtendrá una muestra bootstrap de los residuos estimados en el Paso 2, asumiendo que el proceso de errores sigue una estructura autorregresiva. De un modo más detallado, se procederá como sigue:

Paso 3.1. En base a los residuos estimados, $\hat{\varepsilon}_{j,t}^0$, $j = l, s$, se estimarán los vectores de parámetros, ϕ_j , $j = l, s$, asociados a la estructura autorregresiva del proceso de errores.

Paso 3.2. Puesto que la representación autorregresiva de los procesos de error se asume invertible, pueden obtenerse estimaciones $\{\hat{\varepsilon}_{j,t}, t > p_j\}$ del ruido de los modelos autorregresivos a partir de $\{\hat{\varepsilon}_{j,t}^0\}$ y ϕ_j , $j = l, s$. Las series de ruido estimadas serán centradas para obtener las secuencias $\tilde{\varepsilon}_{j,t} = \hat{\varepsilon}_{j,t} - \hat{\varepsilon}_{j,\cdot}$, para $t > p_j$, donde $\hat{\varepsilon}_{j,\cdot} = \frac{1}{n-p_j} \sum_{t=p_j+1}^n \hat{\varepsilon}_{j,t}$, para $j = l, s$.

Paso 3.3. Se deriva la distribución empírica de $\tilde{\varepsilon}_{j,t}$ para $j = l, s$ como

$$\hat{F}_j(z) = \frac{1}{n-p_j} \sum_{t=p_j+1}^n 1_{\tilde{\varepsilon}_{j,t} \leq z}$$

Paso 3.4. Se obtiene una muestra de variables aleatorias independientes e idénticamente distribuidas $\{e_{j,-M}^*, \dots, e_{j,-1}^*, e_{j,0}^*, e_{j,1}^*, \dots, e_{j,p_j}^*\}$, con $M > 0$ a partir de \widehat{F}_j , $j = l, s$.

La secuencia $\{e_{j,t}^*\}_{t=-M}^{p_j}$ se utiliza junto con las estimaciones de los parámetros autorregresivos ϕ_j para generar una muestra bootstrap del proceso de error, $\{\varepsilon_{j,t}^*\}$, para $j = l, s$.

Paso 4. Se obtiene una muestra bootstrap $\left\{ \left(z_t, X_{l,t}^*, X_{s,t}^* \right) \right\}_{t=1}^n$ según

$$X_{j,t}^* = \widehat{\varphi}_{p,g}(z_t) + \widehat{\varepsilon}_{j,t}^*, t = 1, \dots, n, j = l, s$$

El test estadístico $\tilde{Q}_n^{(v)}$ se computa entonces sobre la muestra bootstrap.

Paso 5. El Paso 3.4 y el Paso 4 se repiten un número B grande de veces, de forma que se obtenga una secuencia de valores del estadístico $\tilde{Q}_n^{(v)}$, $\tilde{Q}_{n,1}^{(v)*}, \dots, \tilde{Q}_{n,B}^{(v)*}$. Se tomará entonces como región crítica del test a un nivel de significación α la definida por

$$\tilde{Q}_n^{(v)} > \tilde{Q}_{n,((1-\alpha)B)}^{(v)*}$$

donde $[\cdot]$ denota la parte entera de un número real y $\{\tilde{Q}_{n,(i)}^{(v)*}\}_{i=1}^B$ es la muestra $\{\tilde{Q}_{n,i}^{(v)*}\}_{i=1}^B$ ordenada de menor a mayor valor.

En la Tabla 5.3 se incluyen los p-valores asociados a las comparaciones dos a dos de cada uno de los cinco bancos incluidos en el IBEX-35 (BBVA, Banesto, Bankinter, BSCH y Popular), utilizando cada uno de los estadísticos $\tilde{Q}_n^{(1)}$, $\tilde{Q}_n^{(2)}$, $\tilde{Q}_n^{(3)}$ y $\tilde{Q}_n^{(4)}$, definidos en (5.7), (5.9), (5.10) y (5.11), respectivamente. Los p-valores fueron obtenidos, en todos los casos, mediante la aproximación bootstrap descrita anteriormente (columnas 3 a 6), aunque para el estadístico $\tilde{Q}_n^{(1)}$ se obtuvieron además los p-valores asociados utilizando su distribución asintótica (columna 2).

A partir de la Tabla 5.3 puede concluirse que no existen diferencias significativas entre las tendencias de los precios semanales por acción de los bancos BBVA, Bankinter y BSCH, puesto que con todos los procedimientos se obtienen valores grandes de significación al comparar dichos bancos dos a dos. Estos resultados apoyan la homogeneidad del grupo C_1 que se había identificado previamente en el análisis cluster. Dado que los tres bancos pertenecen al IBEX-35, cualquiera de sus tendencias podría ser elegida como patrón representante de la cotización bancaria del IBEX-35 en el periodo de estudio. Por otro lado, las filas 4, 9 y 11 en la Tabla 5.3 establecen que la tendencia de los precios semanales por acción del banco Popular difieren significativamente de aquellas del BBVA, Bankinter y BSCH. De hecho, este resultado se concluye de todos los procedimientos utilizados. Por lo tanto, resulta claro que aunque el banco Popular pertenece al IBEX-35, su comportamiento es significativamente distinto al del resto de los bancos en este grupo

Tabla 5.3: p -valores de las comparaciones dos a dos de los bancos incluidos en el IBEX-35 utilizando los estadísticos $\tilde{Q}_n^{(k)}$, $k = 1, 2, 3, 4$, y aproximando su distribución muestral mediante técnicas bootstrap. En el caso de $\tilde{Q}_n^{(1)}$, los p -valores se calculan además a partir de su distribución asintótica.

	$\tilde{Q}_n^{(1)}$ asintótica	$\tilde{Q}_n^{(1)}$ bootstrap	$\tilde{Q}_n^{(2)}$ bootstrap	$\tilde{Q}_n^{(3)}$ bootstrap	$\tilde{Q}_n^{(4)}$ bootstrap
BBVA/Banesto	0,103	0,048	0,008	0,024	0,258
BBVA/Bankinter	0,738	0,782	0,332	0,924	0,934
BBVA/Popular	0,000	0,000	0,000	0,000	0,016
BBVA/BSCH	0,854	0,960	0,644	0,994	0,998
Banesto/Bankinter	0,072	0,034	0,004	0,022	0,192
Banesto/Popular	0,436	0,204	0,014	0,132	0,190
Banesto/BSCH	0,208	0,094	0,028	0,078	0,282
Bankinter/Popular	0,000	0,002	0,000	0,000	0,008
Bankinter/BSCH	0,718	0,702	0,434	0,850	0,932
Popular/BSCH	0,000	0,002	0,000	0,000	0,024

y su ubicación en el grupo C_5 , el cluster más alejado del C_1 , está completamente justificada. Por último, con respecto a los datos de cotización del Banesto, los diferentes tests de hipótesis empleados no permiten llegar a resultados concluyentes. No existe evidencia suficiente para rechazar la hipótesis nula o, al menos, dicha evidencia no es sustentada de modo unánime por los distintos procedimientos utilizados. Las comparaciones dos a dos han sido también realizadas sobre el resto de series bancarias, siendo los resultados congruentes con aquellos alcanzados tras el análisis cluster.

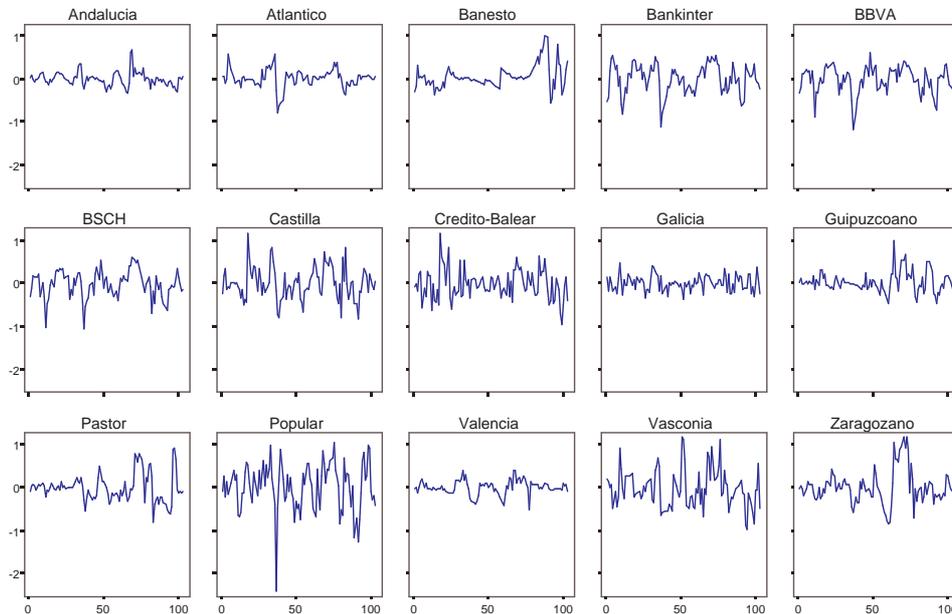
5.4. Análisis cluster de las componentes aleatorias

Como siguiente paso en el análisis, se procedió a eliminar las tendencias estimadas mediante regresión lineal local del modelo (5.1) para obtener y analizar posteriormente las secuencias de ruido, esto es, los residuos no paramétricos $\hat{\varepsilon}_{l,t}$ dados por

$$\hat{\varepsilon}_{l,t} = X_{l,t} - \hat{\varphi}_{l,h_l}(z_t), \quad t = 1, 2, \dots, n, \quad l = 1, 2, \dots, 15. \quad (5.12)$$

En la Figura 5.7 se muestran las quince series correspondientes a los residuos así estimados.

Figura 5.7: Gráfico de las series de residuos no paramétricos $\hat{\varepsilon}_{l,t}$ obtenidos para cada una de las series de acuerdo con (5.12).

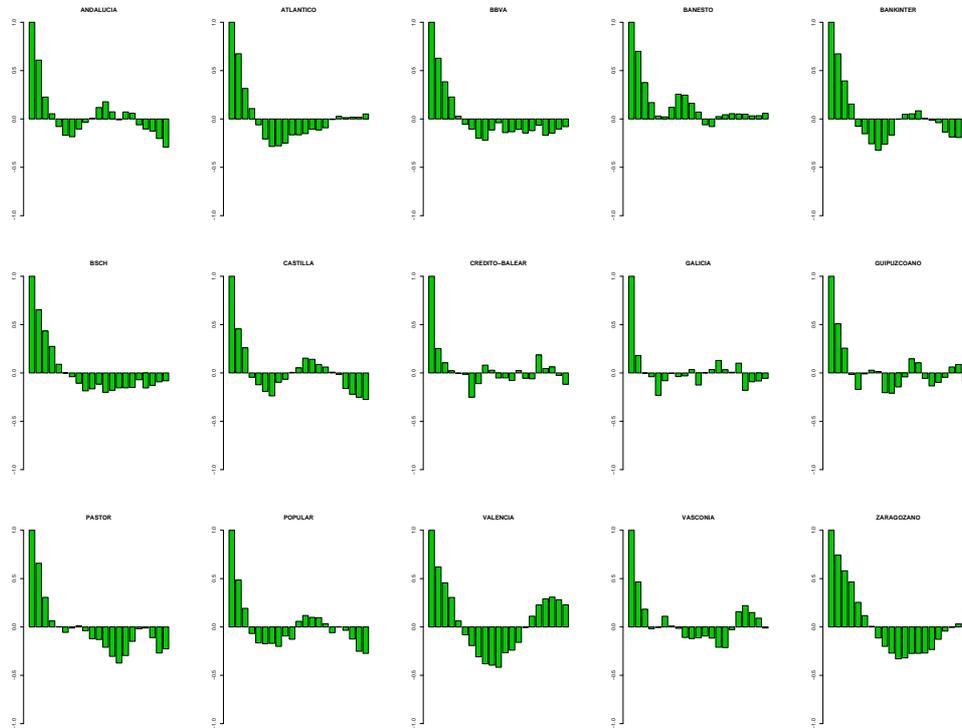


La principal conclusión que se puede extraer de la Figura 5.7 es que todas las series de residuos son, aparentemente, series estacionarias de media cero aunque no parece observarse un patrón común que pueda describir la estructura de dependencia de las quince series en su conjunto. Para corroborar este hecho, se han calculado también las funciones de autocorrelación muestral de las quince series y las funciones de autocorrelación parcial, que se muestran en las Figuras 5.8 y 5.9, respectivamente. De la observación de dichas figuras pueden extraerse conclusiones análogas.

Mediante paquetes estadísticos estándar se ha comprobado que todas las series de residuos pueden ser modelizadas razonablemente bien mediante un modelo autorregresivo o un proceso de medias móviles. No obstante, tal y como ya se ha apuntado, el objetivo en este ejemplo no se centra en ajustar un modelo que describa la estructura de dependencia de las series de residuos, sino más bien en clasificarlas en grupos similares. En este caso, las técnicas de análisis cluster estándar no proporcionarán resultados adecuados al menos que la medida utilizada para valorar la similitud entre las series tome en cuenta la caracterización probabilística de las mismas.

En este capítulo, se utilizarán las técnicas descritas en el Capítulo 4 de la presente memoria para llevar a cabo un análisis cluster de las quince series de residuos. Para ello,

Figura 5.8: Funciones de autocorrelación muestral de las quince series de residuos no paramétricos, $\hat{\varepsilon}_{l,t}$.



se siguieron los siguientes pasos:

Paso 1. Se obtuvieron los valores de los periodogramas asociados a cada una de las series de residuos sobre las frecuencias de Fourier. Esto es, para la l -ésima serie, $l = 1, \dots, 15$, se calculó la función

$$I_{l,n}(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n \hat{\varepsilon}_{l,t} \exp(-i\lambda t) \right|^2, \quad \lambda \in [-\pi, \pi],$$

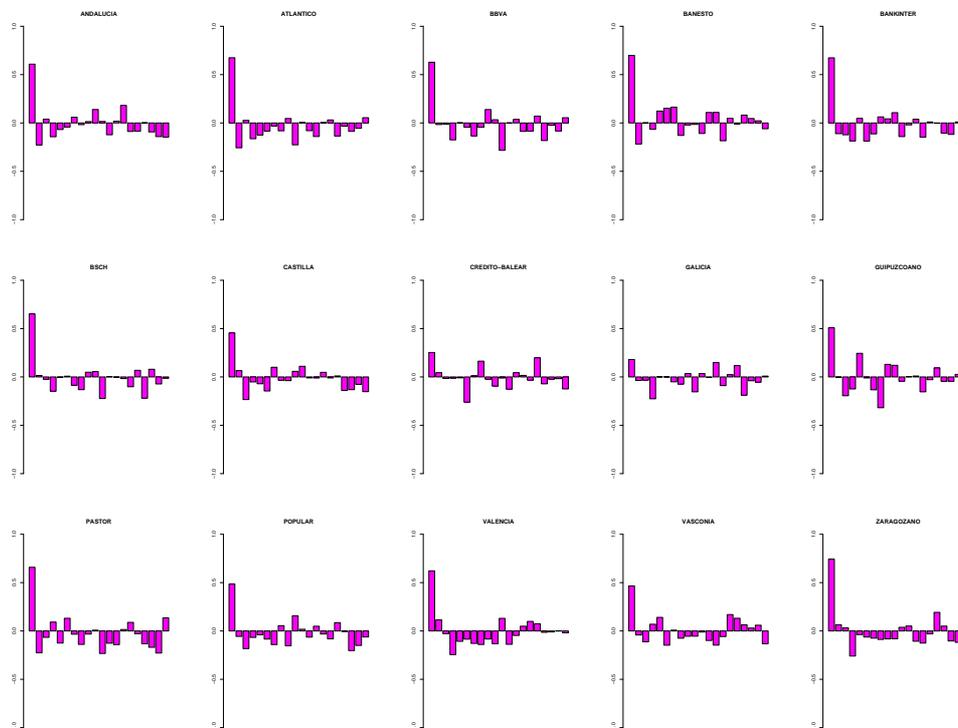
sobre las frecuencias $\lambda_k = 2\pi k/n$, con $k = -N, \dots, N$, $N = [(n-1)/2]$, siendo $n = 103$ la longitud de cada una de las series consideradas.

Paso 2. A continuación se estimó la densidad espectral de cada una de las series de errores, $f_l(\lambda)$, para $l = 1, \dots, 15$ utilizando cualquiera de los tres estimadores propuestos en Fan y Kreutzberger (1998):

i) El suavizador lineal local del periodograma, donde el ajuste lineal local se realiza de la manera habitual mediante mínimos cuadrados locales ponderados ($\hat{f}_{l,DLS}$, $l = 1, \dots, 15$).

(ii) El exponencial del suavizador lineal local mediante mínimos cuadrados ponderados del logaritmo del periodograma de cada una de las series de residuos, $\hat{f}_{l,LS}$, $l = 1, \dots, 15$.

Figura 5.9: Funciones de autocorrelación parcial estimadas de las quince series de residuos no paramétricos, $\hat{\varepsilon}_{l,t}$.



(iii) El estimador $\hat{f}_{l,LK}$, $l = 1, \dots, 15$, obtenido de modo similar a $\hat{f}_{l,LS}$, excepto que el ajuste lineal local del logaritmo del periodograma se hace en este caso utilizando un criterio de máxima verosimilitud local en lugar del de mínimos cuadrados. Puesto que en este caso el suavizador debe obtenerse mediante algún procedimiento numérico, se utilizó el algoritmo de Newton-Raphson para este propósito.

Paso 3. La distancia entre cada par de series residuales, $\{\hat{\varepsilon}_{l,t}\}$ y $\{\hat{\varepsilon}_{s,t}\}$, $l, s \in \{1, \dots, 15\}$, se calculó mediante la medida de disparidad espectral $d_W(\cdot, \cdot)$ introducida en la Sección 4.3, según:

$$\begin{aligned} d_W(\hat{\varepsilon}_{l,t}, \hat{\varepsilon}_{s,t}) &= D_W(\hat{f}_l, \hat{f}_s) + D_W(\hat{f}_s, \hat{f}_l) = \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ W \left(\frac{\hat{f}_l(\lambda)}{\hat{f}_s(\lambda)} \right) + W \left(\frac{\hat{f}_s(\lambda)}{\hat{f}_l(\lambda)} \right) \right\} d\lambda, \end{aligned} \quad (5.13)$$

donde \hat{f}_r es cualquiera de los tres suavizadores lineales locales indicados en el Paso 2 para estimar el r -ésimo espectro, $r = l, s$, y $W(\cdot)$ es una función de divergencia que, en este caso, se tomó igual a

$$W(x) = \log \frac{x+1}{2} - \frac{1}{2} \log x. \quad (5.14)$$

Paso 4. Utilizando la matriz de distancias entre las series de errores obtenida basándose en d_W se han utilizado diferentes métodos (Ward, enlace promedio, enlace completo) para llevar a cabo un procedimiento cluster jerárquico con el fin de clasificar las series.

En la Figuras 5.10, 5.11 y 5.12 se muestran los dendogramas obtenidos tras realizar el análisis cluster con diferentes métodos de clasificación jerárquica y los estimadores espectrales $\hat{f}_{i,DLS}$, $\hat{f}_{i,LS}$ y $\hat{f}_{i,LK}$, respectivamente. Dichos dendogramas muestran soluciones cluster muy similares, que de modo general proporcionan una solución con tres clusters, aunque con ligeras discrepancias a la hora de clasificar las series de residuos correspondientes a los bancos Banesto, Galicia y Guipuzcoano. Estas pequeñas diferencias pueden justificar una solución con un mayor número de grupos. Independientemente del estimador espectral y de la medida de similitud entre grupos utilizados, los bancos $S1 = \{\text{Vasconia, Castilla, Crédito-Balear, Popular}\}$ conforman uno de los grupos identificados. También en todos los casos las series de residuos correspondientes a los bancos $S2 = \{\text{Valencia, Andalucía, Atlántico}\}$ se clasifican conjuntamente en un mismo grupo, al que se añadiría el Galicia o el Banesto dependiendo del suavizador espectral empleado ($\hat{f}_{i,LS}$ o $\hat{f}_{i,LK}$, respectivamente). Finalmente, un tercer cluster estaría formado por las entidades $S3 = \{\text{Zaragozano, Pastor, Bankinter, BBVA, BSCH}\}$, en el que se incluirían también el Banesto, Galicia o banco Guipuzcoano en función del estimador espectral utilizado.

Para evaluar la calidad de las soluciones cluster así obtenidas se utilizaron los valores silueta, $s(i)$, $i = 1, \dots, 15$, definidos tal y como se explicó en (4.19). La Figura 5.13 muestra los gráficos de silueta obtenidos al utilizar el método de clasificación descrito con cada uno de los tres estimadores espectrales y el método de enlace completo. Como se puede observar, en todos los casos aquellas series con un valor silueta más bajo fueron las correspondientes a Banesto, Galicia y Guipuzcoano, señalando la dificultad para ubicarlas en cualquiera de los clusters identificados. También la serie correspondiente al banco Atlántico presenta valores silueta asociados bajos, lo que sugiere que su clasificación junto con los bancos Valencia y Andalucía quizá no sea adecuada. El resto de las series parecen estar correctamente clasificadas en cada uno de los tres clusters identificados, con valores silueta $s(i)$ que se encuentran, en general, por encima de 0.5.

Puesto que algunos de los dendogramas obtenidos sugerían una solución con un número mayor de clusters se optó por obtener los gráficos de silueta bajo la solución de $k = 5$ clusters, que se muestran en la Figura 5.14. Los resultados obtenidos corroboran la dificultad para la clasificación de la serie correspondiente al Banesto y sugieren la identificación de un nuevo cluster formado por los bancos $S4 = \{\text{Galicia, Guipuzcoano}\}$. Los gráficos de silueta para el resto de soluciones de cluster jerárquico (enlace promedio, método de Ward) permiten extraer conclusiones análogas.

Figura 5.10: Dendrograma del análisis cluster de las 15 series de residuos utilizando $\hat{f}_{i,DLs}$ y d_W como estimador espectral y medida de similitud entre series, respectivamente. Resultados con el método de enlace promedio (a), enlace completo (b) y método de Ward (c).

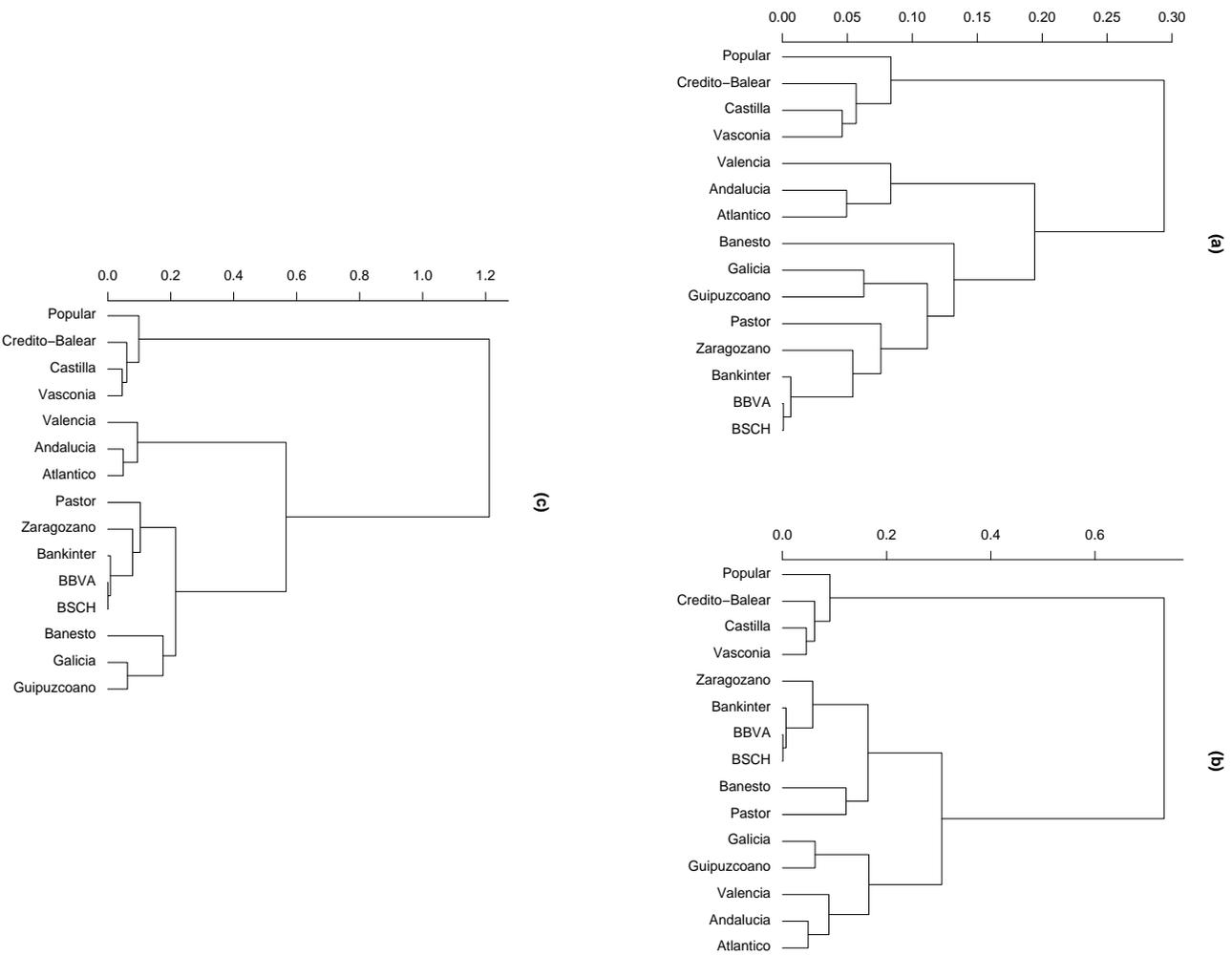


Figura 5.11: Dendograma del análisis cluster de las 15 series de residuos utilizando $\hat{f}_{i,LS}$ y d_W como estimador espectral y medida de similitud entre series, respectivamente. Resultados con el método de enlace promedio (a), enlace completo (b) y método de Ward (c).

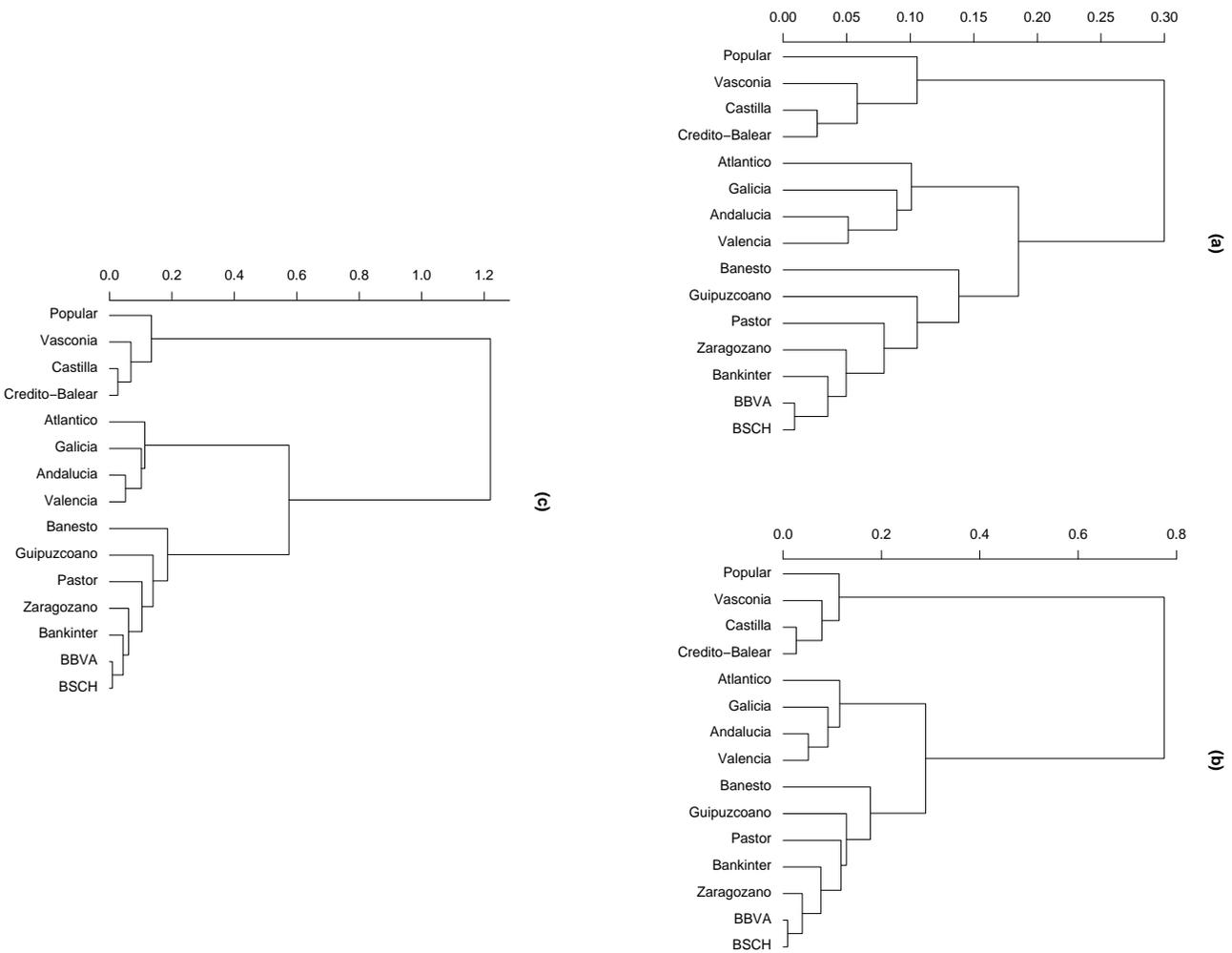


Figura 5.12: Dendograma del análisis cluster de las 15 series de residuos utilizando $\hat{f}_{i,LK}$ y d_W como estimador espectral y medida de similitud entre series, respectivamente. Resultados con el método de enlace promedio (a), enlace completo (b) y método de Ward (c).

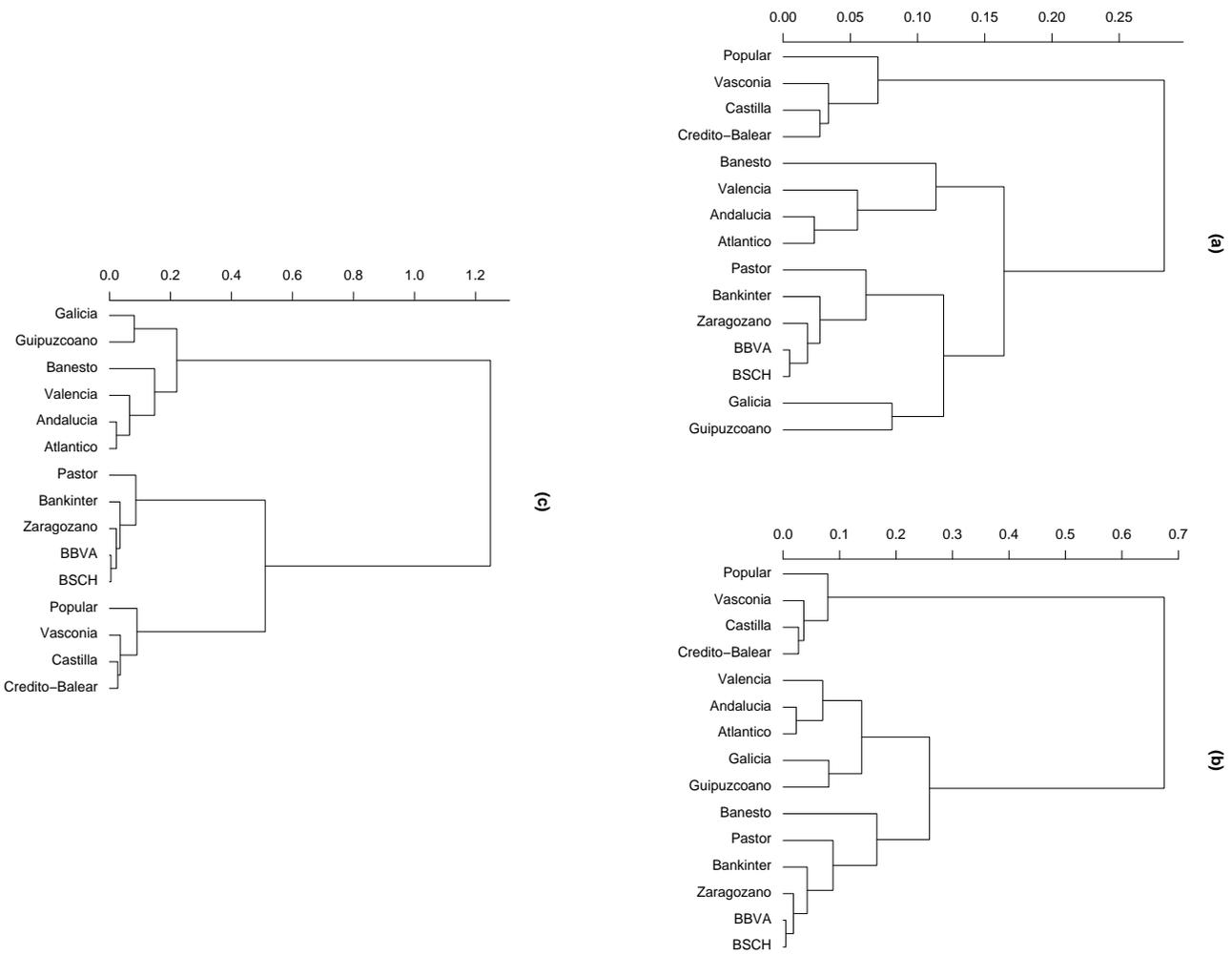


Figura 5.13: Gráficos de silueta para la solución de 3 grupos del análisis cluster de las 15 series de residuos utilizando d_W y el método de enlace completo como distancia entre series y entre grupos, respectivamente, a partir cada uno de los estimadores espectrales: (a) $\hat{f}_{i,DLS}$, (b) $\hat{f}_{i,LK}$ y (c) $\hat{f}_{i,LK}$.

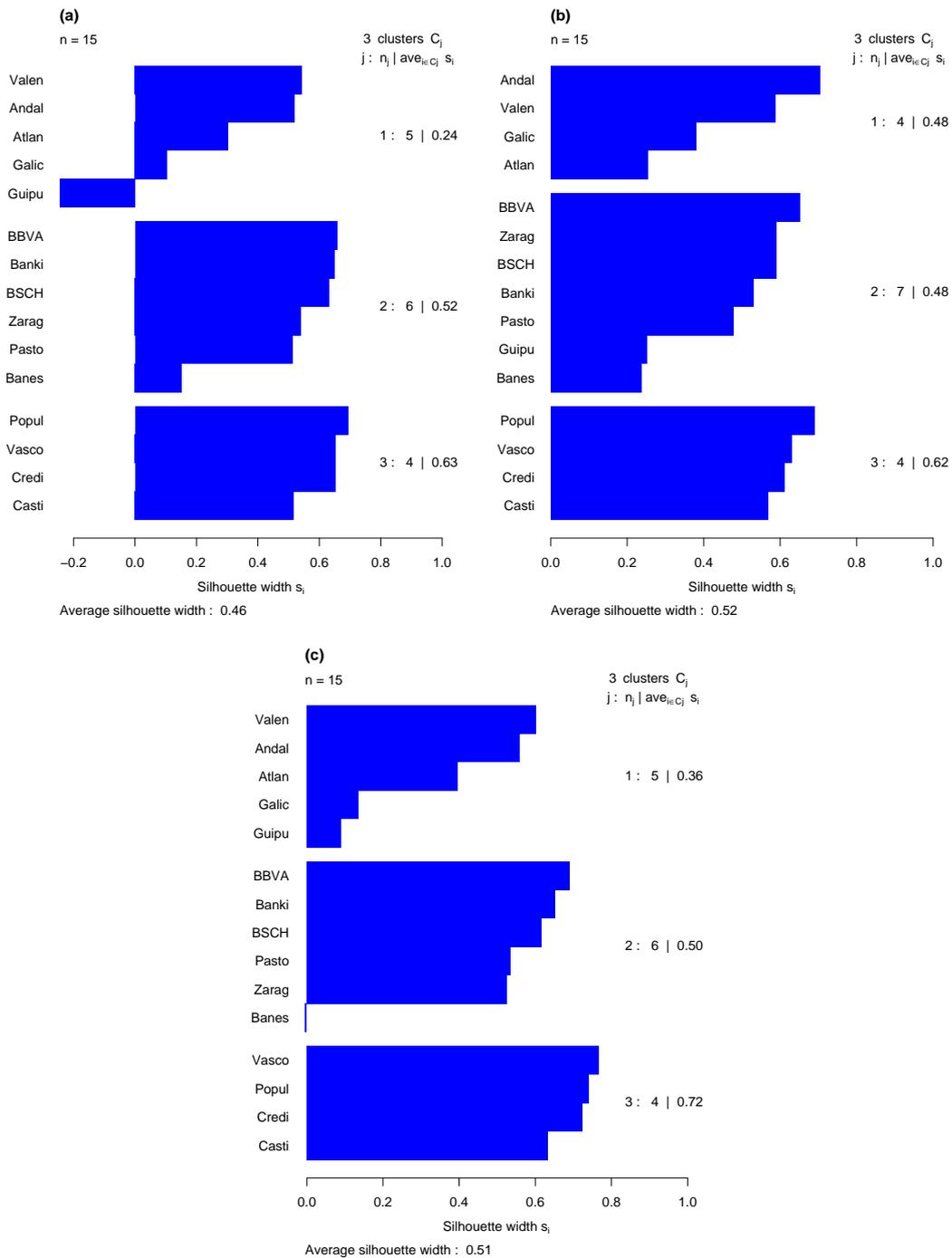
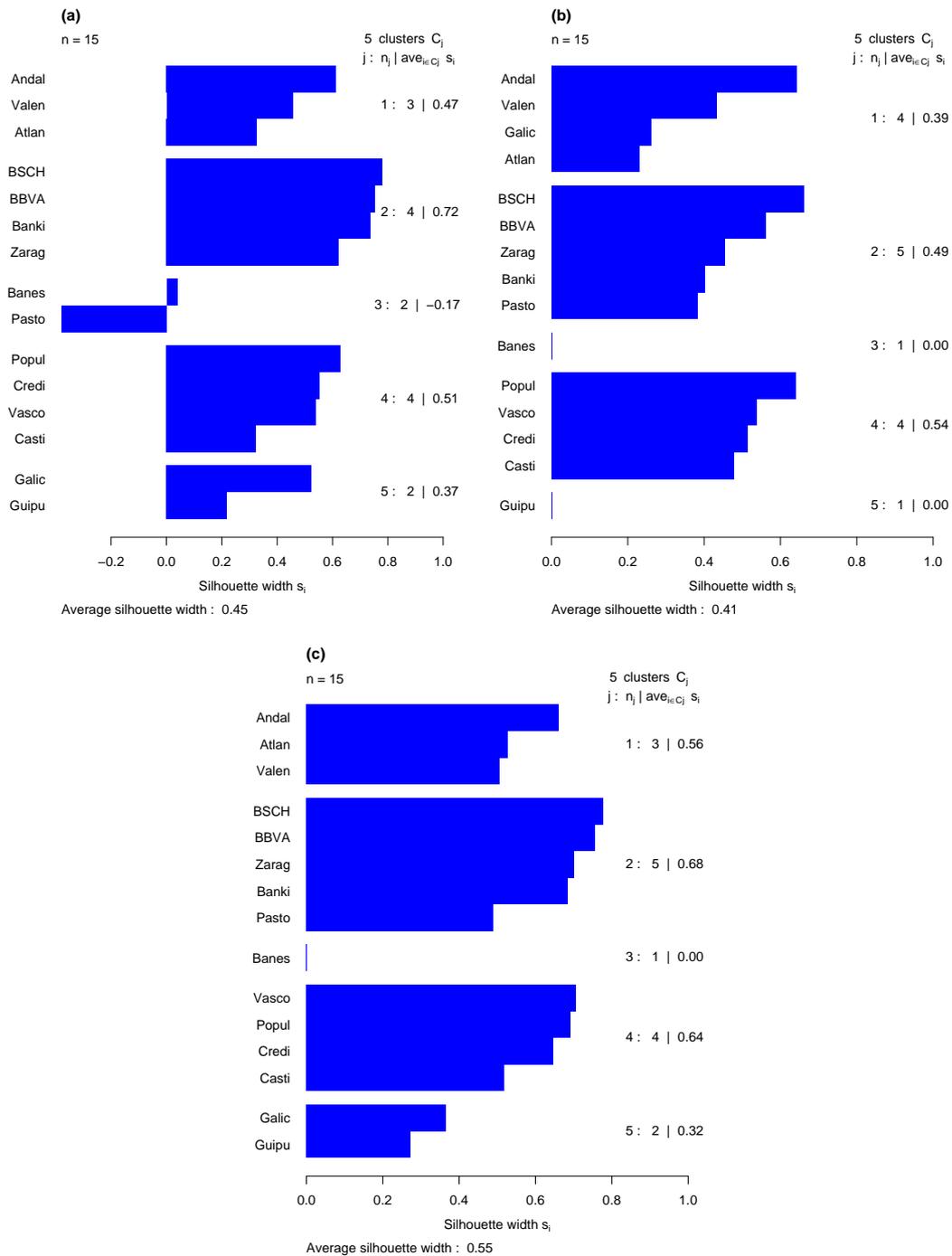


Figura 5.14: Gráficos de silueta para la solución de 5 grupos del análisis cluster de las 15 series de residuos utilizando d_W y el método de enlace completo como distancia entre series y entre grupos, respectivamente, a partir cada uno de los estimadores espectrales: (a) $\hat{f}_{i,DLS}$, (b) $\hat{f}_{i,LK}$ y (c) $\hat{f}_{i,LK}$.



Se comparó además la solución obtenida con el algoritmo cluster propuesto con las soluciones alcanzadas en caso de utilizarse otros criterios para evaluar la discrepancia entre series de tiempo. En particular, se evaluaron las medidas de disparidad entre series de tiempo consideradas en la Sección 4.2. Más concretamente, se realizó el cluster de las series de residuos utilizando como medidas de distancia entre ellas:

- (a) La distancia euclídea entre las series, d_E .
- (b) La distancia euclídea entre los $L = 10$ primeros coeficientes de autocorrelación muestral, d_{ACFU} .
- (c) La distancia euclídea entre los $L = 10$ primeros coeficientes de autocorrelación parcial, d_{PACFU} .
- (d) La distancia propuesta por Piccolo (1990), d_{PIC} , dada por (4.6).
- (e) La distancia propuesta por Maharaj (1996), d_M , definida en (4.7).
- (f) La distancia euclídea entre las ordenadas de los periodogramas, d_P , del logaritmo de los periodogramas, d_{LP} , de los periodogramas normalizados, d_{NP} y del logaritmo de los periodogramas normalizados, d_{LNP} .
- (g) La distancia basada en la razón de verosimilitud generalizada, d_{GLK} , definida en (4.14).
- (h) La distancia funcional tipo Cramer-von-Mises, d_{CM} , definida en (4.15).

En la Tabla 5.4 se muestran los valores silueta promedio correspondientes a las soluciones con $k = 3, 4$ ó 5 clusters obtenidas con cada una de las anteriores distancias, junto con los valores asociados a la distancia d_W . Tal y como se puede observar, los valores silueta más altos correspondieron a la distancia no paramétrica d_W , calculada con cualquiera de los tres estimadores espectrales y a las otras dos distancias de corte no paramétrico consideradas, d_{GLK} y d_{CM} , con valores silueta promedio en torno a 0.50 para la solución con $k = 3$ clusters. De hecho, la distancia funcional de tipo Crámer-von-Mises, d_{CM} , fue la que proporcionó mejores resultados, con un valor silueta promedio de 0.56, mientras que la distancia basada en la razón de verosimilitud generalizada, d_{GLK} , presentó un valor silueta promedio de 0.49 para la solución con 3 clusters. En la Figura 5.15 se muestran los dendogramas obtenidos con cada una de estas dos distancias y el método de enlace completo. La distancia d_{GLK} coincide en identificar los clusters $S1 = \{\text{Vasconia, Castilla, Crédito-Balear, Popular}\}$, $S2 = \{\text{Valencia, Andalucía, Atlántico}\}$ y $S3 = \{\text{Zaragozano, Pastor, Bankinter, BBVA, BSCH}\}$, incluyendo al Banesto en este último y agrupando en $S1$ a las series correspondientes a Galicia y Guipuzcoano para la solución con 3 clusters. Por su parte, la

distancia d_{CM} proporciona una solución ligeramente diferente, disgregando el primero de los clusters en los grupos {Vasconia, Popular} y {Castilla, Crédito-Balear} y asignando los bancos Galicia y Guipuzcoano a $S2$. Si analizamos el gráfico de silueta asociado a la solución de tres clusters (Figura 5.16) observamos que, pese a alcanzar un valor silueta promedio alto, la solución obtenida con la distancia d_{CM} parece clasificar incorrectamente a las series correspondientes a Castilla y Crédito-Balear, confirmando su pertenencia a $S1$.

Tabla 5.4: Cluster de las series de residuos. Silueta promedio para las soluciones con $k = 3$, 4 y 5 clusters. Método de enlace completo.

Medida	Silueta promedio			Medida	Silueta promedio		
	$k = 3$	$k = 4$	$k = 5$		$k = 3$	$k = 4$	$k = 5$
<i>Distancia Euclídea</i>				<i>Basadas en modelos</i>			
d_E	0.21	0.12	0.12	d_{PIC}	0.20	0.19	0.17
<i>Autocorrelaciones</i>				<i>No paramétricas</i>			
d_{ACFU}	0.26	0.20	0.21	d_M	0.36	0.30	0.28
d_{PACFU}	0.16	0.14	0.14	$d_{W(DLS)}$	0.46	0.48	0.45
<i>Periodogramas</i>				$d_{W(LS)}$	0.52	0.47	0.41
d_P	0.35	0.30	0.31	$d_{W(LK)}$	0.51	0.46	0.55
d_{LP}	0.17	0.16	0.14	d_{GLK}	0.49	0.39	0.24
d_{NP}	0.17	0.16	0.21	d_{CM}	0.56	0.44	0.35
d_{LNP}	0.10	0.10	0.09				

Con respecto al resto de distancias, todas ellas obtuvieron valores silueta promedio por debajo a 0.50, lo que sugiere soluciones cluster menos estables que las obtenidas a partir de las distancias de corte no paramétrico. Los dendogramas obtenidos con cada una de estas distancias se muestran en las Figuras 5.17, 5.18, 5.19 y 5.20.

Figura 5.15: Dendrograma del análisis cluster de las 15 series de residuos no paramétricos $\varepsilon_{i,t}$. Resultados con el método de enlace completo y la distancia basada en la razón de verosimilitud generalizada, d_{LK} (a) y la distancia funcional de tipo Cramer-von-Mises, d_{CM} (b).

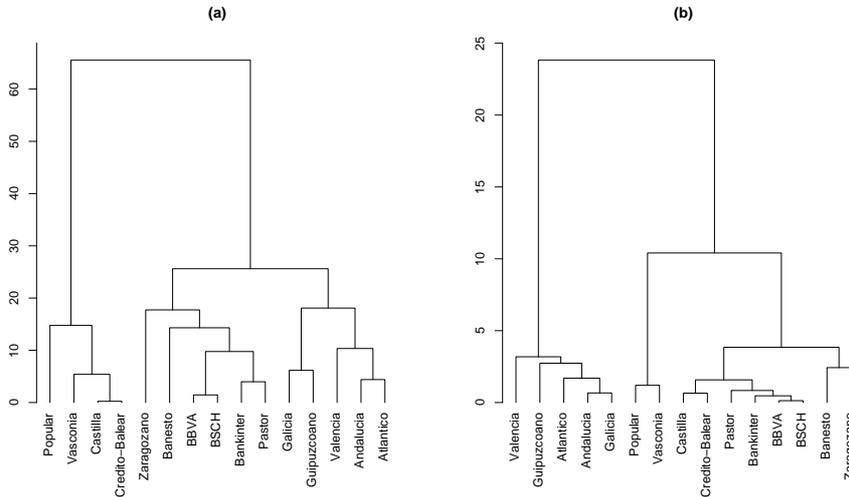


Figura 5.16: Gráficos de silueta para la solución de 3 grupos del análisis cluster de las 15 series de residuos utilizando el método de enlace completo como distancia entre grupos y como distancia entre las series: (a) Distancia basada en la razón de verosimilitud generalizada, (b) Distancia funcional de tipo Crámer-von-Mises.

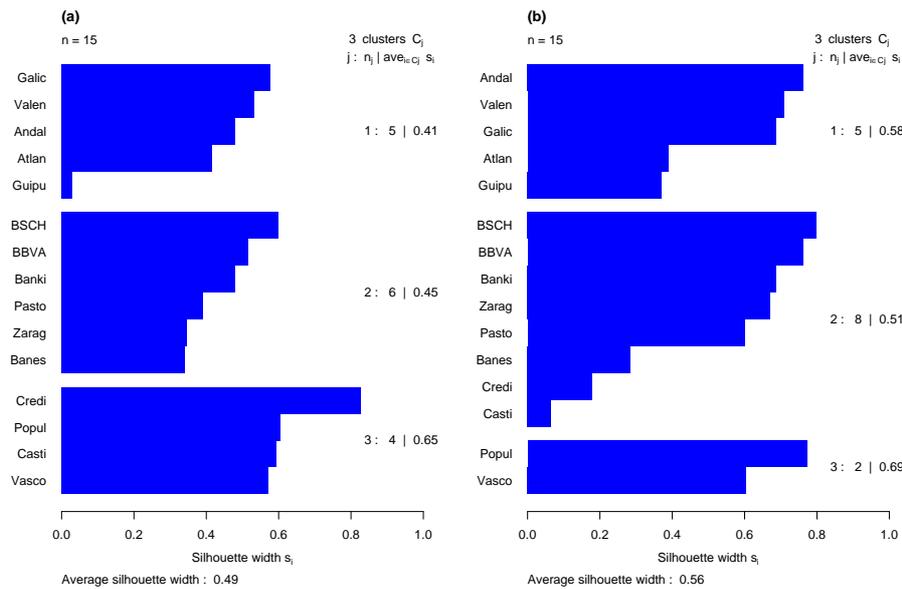


Figura 5.17: Dendrograma del análisis cluster de las 15 series de residuos no paramétricos $\varepsilon_{i,t}$. Resultados con la distancia euclídea entre las series, d_E y el método de enlace completo.

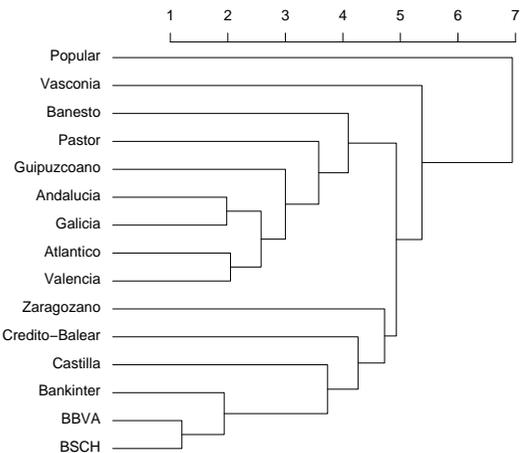


Figura 5.18: Dendrograma del análisis cluster de las 15 series de residuos no paramétricos $\varepsilon_{i,t}$. Resultados con el método de enlace completo y la distancia euclídea entre los coeficientes de autocorrelación, d_{ACFU} (a) y autocorrelación parcial, d_{PACFU} (b).

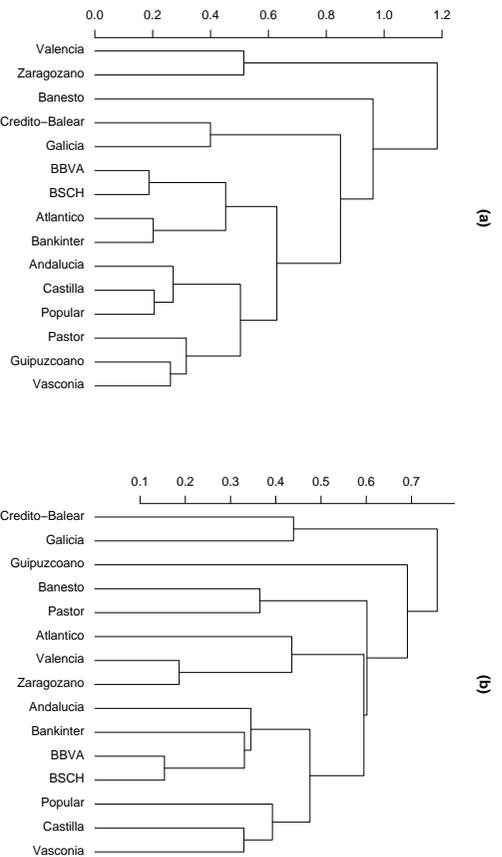


Figura 5.19: Dendrograma del análisis cluster de las 15 series de residuos no paramétricos $\varepsilon_{i,t}$. Resultados con el método de enlace completo y las distancias propuestas por Piccolo, d_{PIC} (a) y Maharaaj, d_M (b).

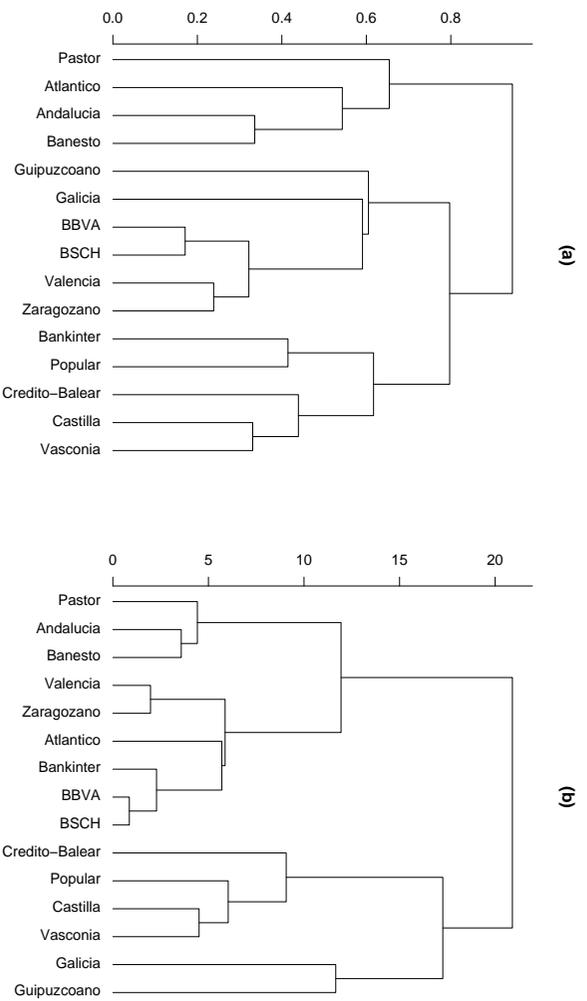
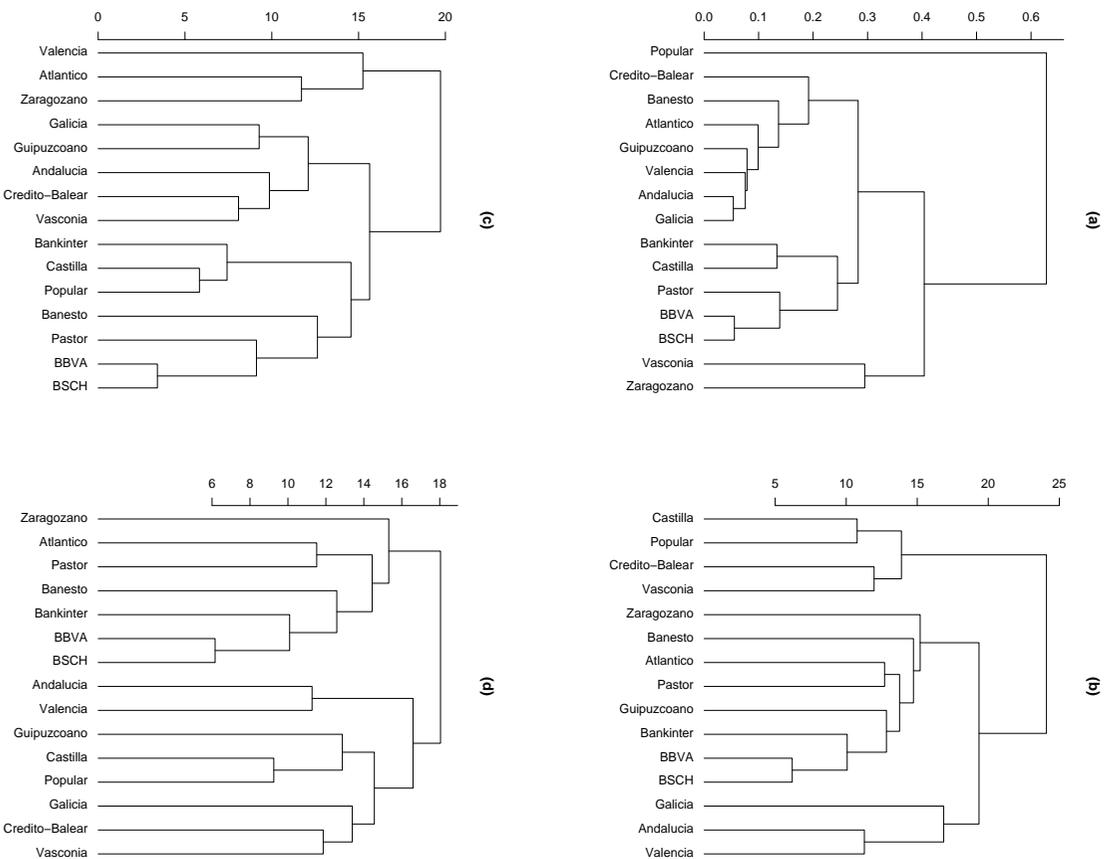


Figura 5.20: Dendrograma del análisis cluster de las 15 series de residuos no paramétricos $\varepsilon_{i,t}$. Resultados con el método de enlace completo y la distancia euclídea entre las ordenadas del periodograma, d_P (a), log-periodograma, d_{LP} (b), periodograma normalizado, d_{NP} (c) y log-periodograma normalizado, d_{LNP} .



5.5. Aplicación a datos simulados

El proceso utilizado para el análisis de las series de cotización bursátil del sector bancario español puede ser utilizado para otro conjunto cualquiera de series siguiendo exactamente los mismo pasos:

- (i) Estimación no paramétrica de las tendencias
- (ii) Análisis cluster de las tendencias estimadas
- (iii) Análisis cluster de las series de residuos

Con el fin de poder evaluar el posible comportamiento del procedimiento de análisis propuesto en otras situaciones diferentes a la aquí planteada, se llevó a cabo un estudio de simulación. Se generó un conjunto de series temporales de acuerdo con el modelo (5.1), de modo que cada serie $X_{l,t}$ consiste en una componente determinista de tendencia ($\varphi_l(z_t)$) y una componente aleatoria ($\varepsilon_{l,t}$). Se escogieron tres tendencias diferentes y tres estructuras de error que se combinaron para obtener nueve clases de series de tiempo distintas, con igual tendencia y componente aleatoria. Las tendencias consideradas fueron:

- (i) $\varphi_1(z) = \sin(1,5\pi z)$,
- (ii) $\varphi_2(z) = 5(z + 0,1)(z - 0,6)(z - 1,6)$,
- (iii) $\varphi_3(z) = 3(z + 0,2)(z - 0,6)(z - 1,5)$,

y los tres modelos de dependencia utilizados para los procesos de error:

- (i) Un proceso de medias móviles MA(1) lineal

$$\varepsilon_{1,t} = \eta_{1,t} - 0,4\eta_{1,t-1},$$

- (ii) Un proceso de medias móviles MA(1) no lineal

$$\varepsilon_{2,t} = \eta_{2,t} - 0,5\eta_{2,t-1} + 0,8\eta_{2,t-1}^2,$$

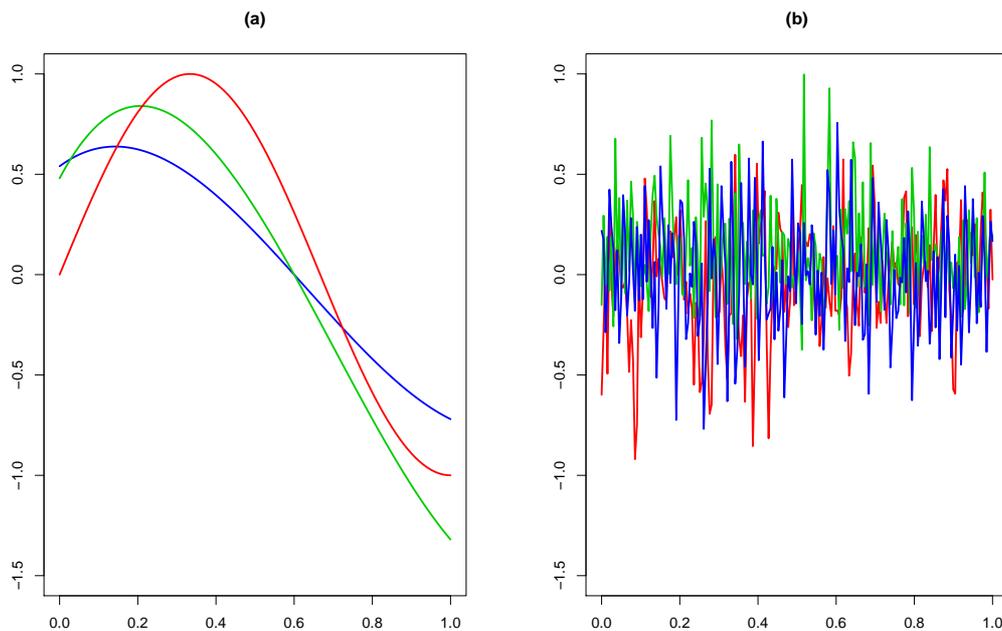
- (iii) Un proceso exponencial autorregresivo (EXPAR)

$$\varepsilon_{3,t} = (0,3 - 10 \exp(-\varepsilon_{3,t-1}^2))\varepsilon_{3,t-1} + \eta_{3,t},$$

donde $\eta_{i,t}$ para $i = 1, 2, 3$ son variables independientes con distribución Normal de media cero y varianza unidad.

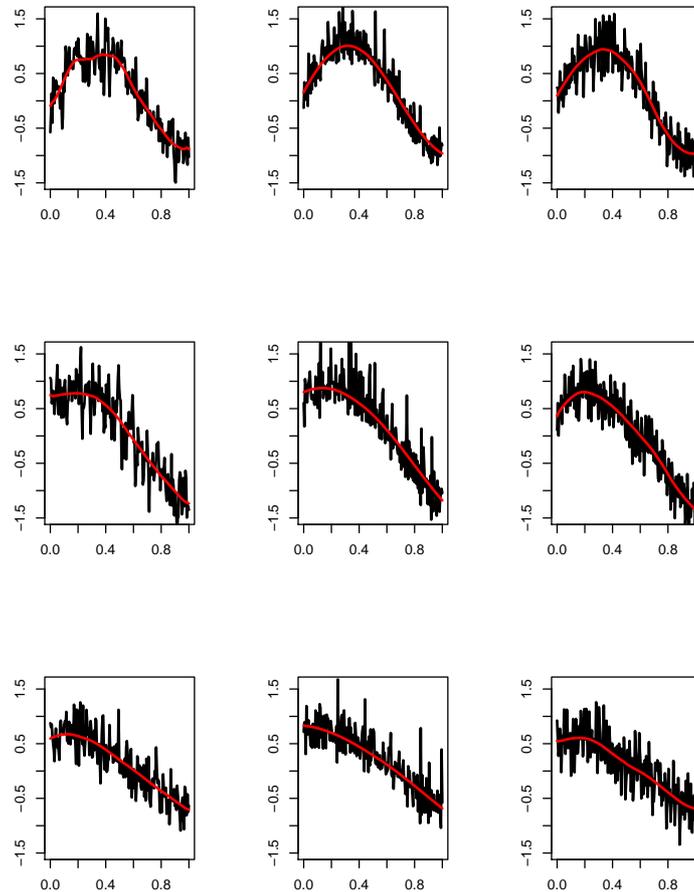
Con el objetivo de dificultar el proceso de clasificación, todas las secuencias de error $\varepsilon_{j,t}, j = 1, 2, 3$ fueron reescaladas apropiadamente para tener desviación estándar 0.3. En las Figuras 5.21a y 5.21b se pueden observar cada una de las tres tendencias consideradas y una realización de cada uno de los procesos de error considerados, respectivamente. En ambos casos se puede apreciar la similitud entre los datos, lo que sugiere un escaso índice de separabilidad entre las series de tiempo sujetas a clasificación.

Figura 5.21: (a) Gráfico de las tendencias consideradas en el estudio de simulación: $\varphi_1(x)$ (rojo), $\varphi_2(z)$ (verde) y $\varphi_3(z)$ (azul). (b) Realizaciones de los procesos de error: MA(1) lineal (rojo), MA(1) no lineal (verde) y EXPAR (azul), adecuadamente reescaladas para tener una desviación estándar igual a 0.3.



Se generaron dos series de longitud $n = 200$ de cada uno de los nueve modelos resultantes. De este modo, el conjunto de datos considerado consistió en 18 series de forma que existían tres grupos de seis series con la misma tendencia y tres grupos de seis series con el mismo proceso de error asociado. Una realización de cada uno de los nueve modelos considerados se muestra en la Figura 5.22. De nuevo, dicha figura sugiere la dificultad del proceso de clasificación de las series consideradas.

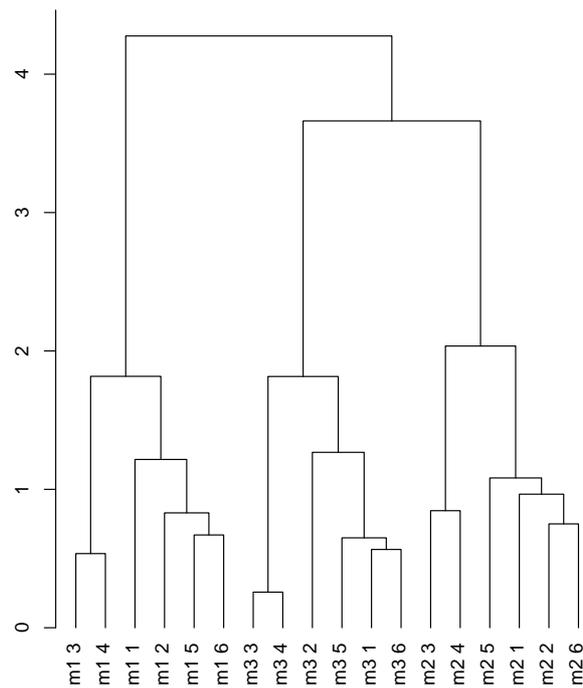
Figura 5.22: Series de longitud $n=200$ generadas a partir del modelo $X_t = \varphi_i(z_t) + \varepsilon_{j,t}$, con $i, j \in \{1, 2, 3\}$ y $\varepsilon_{j,t}$ convenientemente reescalado, junto con las tendencias correspondientes, estimadas mediante regresión lineal local.



De acuerdo con el procedimiento propuesto, el primer paso del análisis consistió en estimar la tendencia de cada una de las series de tiempo consideradas utilizando, igual que con el ejemplo anterior, técnicas de regresión lineal local. Las tendencias así estimadas se muestran conjuntamente con cada una de las realizaciones de cada grupo en la Figura 5.22. Se realizó un análisis cluster de las tendencias así estimadas mediante técnicas jerárquicas, clasificándose correctamente en tres grupos de seis series, a pesar de la gran similitud observada. El dendograma obtenido con el método de enlace promedio y la distancia Euclídea al cuadrado como medida de distancia entre los datos se muestra en la Figura 5.23. Estos resultados vienen a corroborar el buen funcionamiento del suavizador lineal local, a pesar de haberse obtenido sin ninguna hipótesis adicional sobre la forma funcional de las tendencias consideradas. Se obtuvieron resultados análogos con otros métodos de

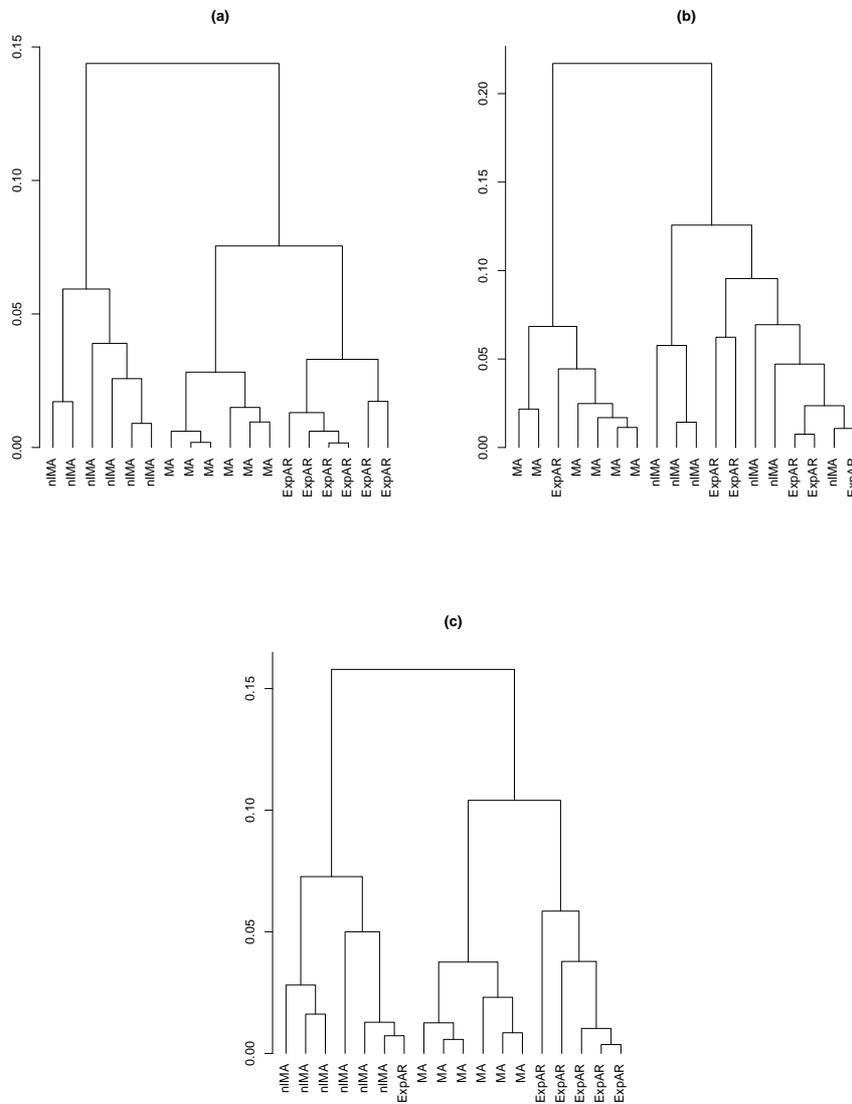
clasificación jerárquica.

Figura 5.23: Dendograma resultado del análisis cluster de las tendencias de las 18 series simuladas, estimadas mediante regresión lineal local. La distancia Euclídea al cuadrado y el método de enlace promedio se utilizaron como matriz de proximidad inicial y medida de similitud entre grupos, respectivamente.



A continuación, las tendencias estimadas mediante regresión lineal local se restaron de las series simuladas para obtener las series de residuos no paramétricos correspondientes. Procediendo igual que en el caso de las series de cotización bancaria, las series de residuos se clasificaron de acuerdo con el algoritmo descrito en el Capítulo 4. Así, la similitud entre dos series de residuos l y s se evaluó mediante la distancia $d_W(\varepsilon_l, \varepsilon_s) = D_W(\hat{f}_l, \hat{f}_s) + D_W(\hat{f}_s, \hat{f}_l)$, siendo \hat{f}_l y \hat{f}_s estimadores de las densidades espectrales f_l y f_s , respectivamente. El análisis cluster se realizó con cada uno de los tres estimadores no paramétricos \hat{f}_{DLS} , \hat{f}_{LS} y \hat{f}_{LK} , igual que en la sección anterior. Los dendogramas obtenidos, utilizando cada uno de estos tres estimadores espectrales, y el método de enlace completo, se muestran en la Figura 5.24.

Figura 5.24: Dendogramas del análisis cluster de las series de residuos de las 18 series simuladas. Resultados con el método de enlace completo utilizando como distancia entre series d_W , a partir de los estimadores espectrales: (a) $\hat{f}_{i,DLS}$, (b) $\hat{f}_{i,LK}$ y (c) $\hat{f}_{i,LK}$.

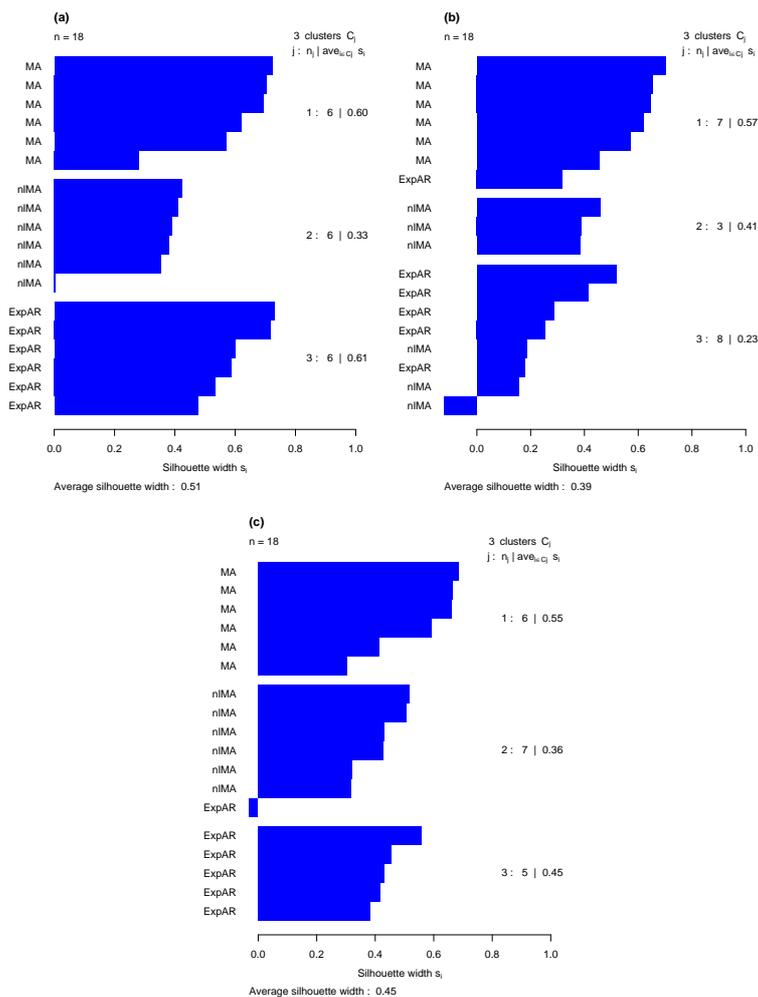


Tal y como se puede observar, los resultados obtenidos con la distancia d_W y cada uno de los tres estimadores espectrales considerados sugiere una solución con tres clusters. Los mejores resultados correspondieron a la distancia calculada a partir de los estimadores \hat{f}_{DLS} y \hat{f}_{LK} , tal y como cabría esperar si tenemos en cuenta sus propiedades asintóticas. Con estos dos estimadores, las 18 series de residuos se agrupan correctamente de acuerdo con su proceso generador (salvo una de ellas en el caso de trabajar con el estadístico

\hat{f}_{LK}). Si se emplea la distancia d_W computada a partir del estadístico log-periodograma suavizado, \hat{f}_{LS} , el porcentaje de series mal clasificadas aumenta.

Si se evalúan estos mismos resultados mediante los correspondientes gráficos de silueta para la solución con $k=3$ clusters (Figura 5.25), observamos que los clusters se encuentran bien delimitados cuando d_W se calcula a partir del estimador periodograma suavizado, excepto para una de las series MA no lineales, muy próxima a clasificarse incorrectamente como EXPAR. Resultados análogos se observan con \hat{f}_{LK} , clasificándose en este caso de forma incorrecta una serie EXPAR en el grupo de MA no lineales. El valor silueta promedio fue, en estos casos, de 0.51 y 0.45, respectivamente. Cuando se utilizó el estimador \hat{f}_{LS} el valor silueta promedio fue de 0.39.

Figura 5.25: Gráficos de silueta para la solución de $k = 3$ grupos en el análisis cluster de las series de residuos de las 18 series simuladas. Resultados con el método de enlace completo utilizando como distancia entre series d_W , a partir de los estimadores espectrales: (a) $\hat{f}_{L, DLS}$, (b) $\hat{f}_{L, LK}$ y (c) $\hat{f}_{L, LK}$.



La clasificación de las series de residuos se llevó a cabo utilizando también otras medidas de distancia entre las series, al igual que se hizo en la sección anterior. El objetivo consistió en comparar los resultados del procedimiento de clasificación no paramétrico basado en la distancia d_W con los obtenidos utilizando otras métricas propuestas en la literatura en el contexto del análisis cluster de series temporales, en particular: (i) la distancia Euclídea entre las series d_E , (ii) la distancia Euclídea entre los coeficientes de correlación (d_{ACFU}) y los coeficientes de autocorrelación parcial (d_{PACFU}), (iii) las distancias entre series de tiempo propuestas por Piccolo (1990) (d_{PIC}) y Maharaj (1996) (d_M), (iv) la distancia euclídea entre las ordenadas del periodograma (d_P), del periodograma normalizado (d_{NP}), del log-periodograma (d_{LP}) y del log-periodograma normalizado (d_{LNP}), (v) la distancia basada en el test de razón de verosimilitud generalizada de Fan, d_{GLK} y (vi) la distancia de tipo Crámer-von-Mises, d_{CM} .

Las Figuras 5.26 a 5.29 muestran los gráficos de silueta obtenidos cuando las 18 series de residuos se clasificaron utilizando las diferentes distancias consideradas, tomando en todos los casos la solución correspondiente a $k = 3$ clusters y el método de enlace completo. Además de los valores silueta, definidos como en (4.19), la calidad de la solución cluster obtenida con cada una de las distancias se evaluó teniendo en cuenta el índice propuesto por Gavrilov, definido de forma similar a como se hizo en (4.16). Los valores silueta promedio para la solución de $k = 3$ clusters, junto con el índice de evaluación (4.16), correspondiente al análisis cluster con cada una de las distancias consideradas y el método de enlace completo se muestra en la Tabla 5.5.

Tabla 5.5: Cluster de las series de residuos. Índice de evaluación definido en (4.16) y silueta promedio para la solución de 3 clusters. Longitud de las series: $T = 200$. Método de enlace completo.

Medida	Índice de evaluación	Silueta promedio	Medida	Índice de evaluación	Silueta promedio
<i>Distancia Euclídea</i>			<i>Basadas en modelos</i>		
d_E	0.463	0.02	d_{PIC}	0.765	0.26
<i>Autocorrelaciones</i>			d_M	1.000	0.55
d_{ACFU}	0.822	0.33	<i>No paramétricas</i>		
d_{PACFU}	1.000	0.30	$d_{W(DLS)}$	1.000	0.51
<i>Periodogramas</i>			$d_{W(LS)}$	0.768	0.39
d_P	0.765	0.12	$d_{W(LK)}$	0.944	0.45
d_{LP}	0.653	0.06	d_{GLK}	0.831	0.34
d_{NP}	0.886	0.10	d_{CM}	0.944	0.46
d_{LNP}	0.659	0.06			

Se obtuvo una solución completamente correcta con tres de las distancias consideradas (d_{PACFU} , d_M y $d_{W(DLS)}$). Sin embargo, la solución cluster proporcionada resultó ser más compacta con las distancias d_M y $d_{W(DLS)}$, que con la distancia d_{PACFU} , con índices silueta promedio de 0.55, 0.51 y 0.30, respectivamente. Las distancias no paramétricas, salvo $d_{W(LS)}$, obtienen índices de evaluación altos, por encima de 0.8, y lo mismo ocurre con la distancia entre los periodogramas normalizados y la distancia entre los coeficientes de autocorrelación muestral. Sin embargo, estas distancias proporcionan una solución cluster mucho más inestable, con valores silueta promedio de, a lo sumo, 0.33. En conclusión, las distancias no paramétricas, junto con la distancia de Maharaj proporcionan en este caso la solución más adecuada, tanto en términos de la clasificación final como teniendo en cuenta la estabilidad de la solución alcanzada.

Figura 5.26: Gráficos de silueta para la solución de $k = 3$ grupos en el análisis cluster de las series de residuos de las 18 series simuladas. Resultados con el método de enlace completo utilizando como distancia entre series la distancia euclídea entre los coeficientes de autocorrelación (a) y entre los coeficientes de autocorrelación parcial (b).

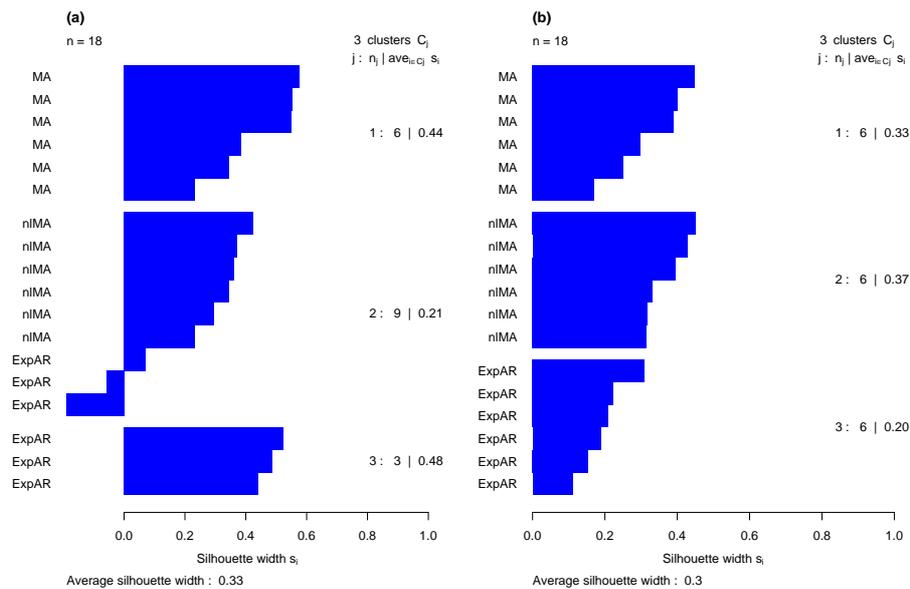


Figura 5.27: Gráficos de silueta para la solución de $k = 3$ grupos en el análisis cluster de las series de residuos de las 18 series simuladas. Resultados con el método de enlace completo utilizando como distancia entre series la propuesta por Piccolo (a) y Maharaj (b).

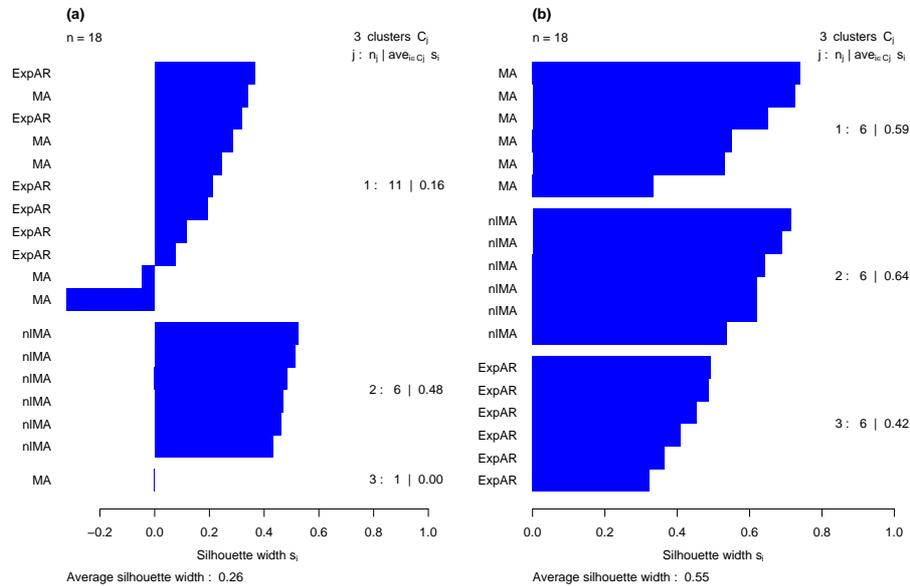


Figura 5.28: Gráficos de silueta para la solución de $k = 3$ grupos en el análisis cluster de las series de residuos de las 18 series simuladas. Resultados con el método de enlace completo utilizando como distancia entre series la distancia basada en el test de razón de verosimilitud generalizada (a) y la distancia funcional de tipo Cramer-von-Mises (b).

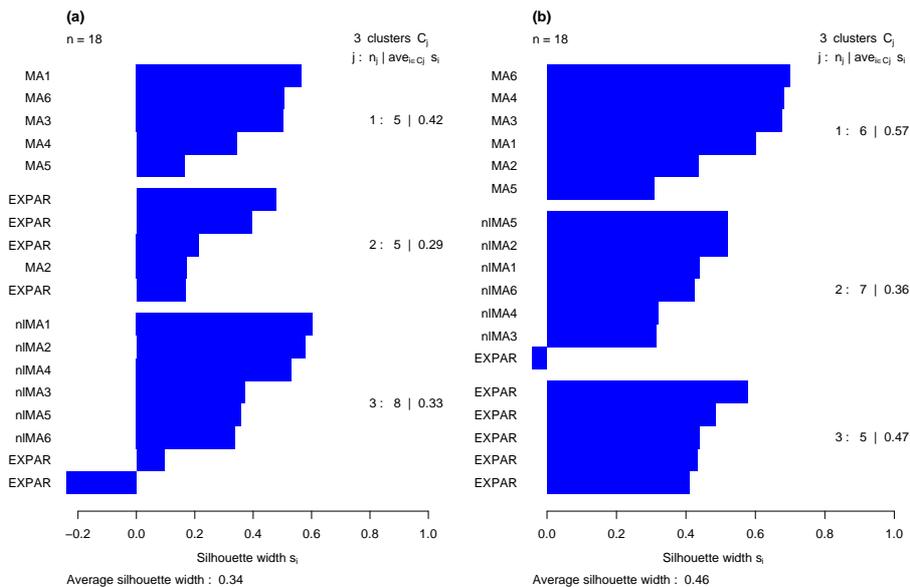
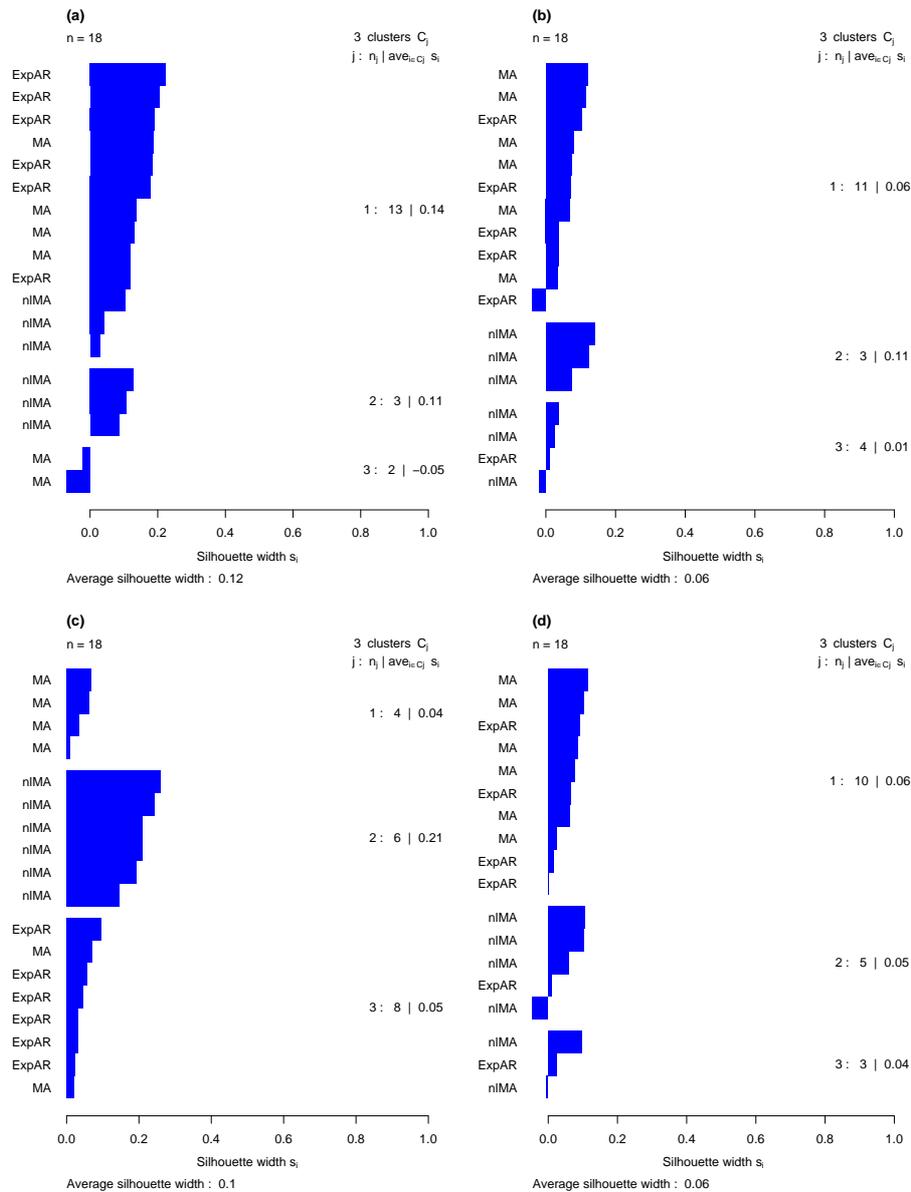


Figura 5.29: Gráficos de silueta para la solución de $k = 3$ grupos en el análisis cluster de las series de residuos de las 18 series simuladas. Resultados con el método de enlace completo utilizando como distancia entre series la distancia euclídea entre periodogramas (a), log-periodogramas (b), periodogramas normalizados (c) y log-periodogramas normalizados (d).



5.6. Conclusiones

El principal objetivo de este capítulo era el de demostrar cómo la metodología no paramétrica proporciona un amplio espectro de procedimientos que permiten realizar un análisis exhaustivo y completo de un conjunto de series de tiempo, incluyendo los procesos de estimación, contraste de hipótesis y análisis cluster. Además de que permiten obviar cualquier tipo de hipótesis paramétrica sobre los datos, las técnicas no paramétricas pueden ajustarse perfectamente al trabajo con datos dependientes, tal y como se ha demostrado a lo largo de este capítulo.

Para ilustrar su utilidad en la práctica, se ha analizado un conjunto de series de tiempo consistente en los datos de cotización semanal de las entidades del sector de la banca en España que cotiza en bolsa, durante los años 2001 y 2002. El principal objetivo fue el de clasificar las entidades bancarias en un número reducido de categorías, de forma que las series en una misma categoría presenten un patrón de comportamiento similar. Por lo tanto, el objetivo no consistía en tratar de ajustar un modelo determinado a cada una de las series, sino en encontrar similitudes y diferencias entre las series observadas. Para alcanzar tal objetivo, se propone analizar independientemente las tendencias y las componentes estocásticas de cada serie.

La estimación mediante técnicas de suavización lineal local de las tendencias de las series nos permite obtener una visión general del conjunto de componentes deterministas, sugiriendo, en el ejemplo propuesto, que probablemente existan patrones de comportamiento distintos entre las tendencias. Un test no paramétrico recientemente propuesto permite rechazar la hipótesis nula de igualdad de las tendencias, confirmando así desde un punto de vista estadístico las diferencias observadas. Un análisis cluster estándar de las tendencias estimadas sugiere una solución con los siguientes cinco clusters: $C1=\{BBVA, BSCH, Bankinter, Atlántico\}$, $C2=\{Banesto\}$, $C3=\{Andalucía, Valencia, Galicia\}$, $C4=\{Guipuzcoano, Zaragoza\}$, $C5=\{Castilla, Popular, Crédito-Balear, Pastor, Vasconia\}$. Los resultados del análisis cluster y las estimaciones no paramétricas de las componentes deterministas permiten describir el patrón de comportamiento que caracteriza a cada uno de los grupos durante el periodo de estudio. Finalmente, se han realizado comparaciones dos a dos de las series agrupadas dentro de un mismo cluster utilizando para ello hasta cuatro tests estadísticos diferentes. En todos los casos los resultados permiten corroborar la solución proporcionada por el análisis cluster y confirmar el esquema de clasificación anterior. Cualquiera de los cuatro tests estadísticos empleados son de corte no paramétrico y su distribución muestral fue aproximada mediante técnicas bootstrap específicamente diseñadas para trabajar con datos dependientes.

Una vez clasificadas y caracterizadas las componentes deterministas de las series, la atención se concentró en el análisis de las series de residuos no paramétricos correspondientes. El problema se centró así en la clasificación de series estacionarias de media cero, de forma que las similitudes y diferencias entre ellas deben ser determinadas mediante algún tipo de distancia que tome en cuenta su estructura de dependencia. En particular, se utilizó la distancia d_W definida en el Capítulo 4 y calculada a partir de cualquiera de los tres estimadores de la densidad espectral propuestos por Fan y Kreutzberger (1998). Los resultados del análisis cluster sugieren la identificación de cinco grupos: $S1=\{\text{Vasconia, Castilla, Crédito-Balear, Popular}\}$, $S2=\{\text{Valencia, Andalucía, Atlántico}\}$, $S3=\{\text{Zaragozano, Guipuzcoano}\}$, $S4=\{\text{Banesto}\}$ y $S5=\{\text{Galicia, Guipuzcoano}\}$. Tal clasificación resulta consistente con los correspondientes procesos ARMA ajustados, de forma que el procedimiento cluster propuesto resulta capaz de identificar la series de acuerdo con su estructura de dependencia. Este hecho resulta especialmente importante, ya que no se ha exigido ninguna hipótesis previa sobre la estructura de dependencia de los datos, en contraposición con otros métodos descritos en la literatura específicamente diseñados para trabajar con series de tiempo con una estructura en particular.

Finalmente, el procedimiento cluster fue comparado con otros procedimientos alternativos específicamente diseñados para el trabajo con series de tiempo. La comparación se realizó mediante un conjunto de datos simulados y los resultados alcanzado muestran como los procedimientos de corte no paramétrico propuestos en este trabajo proporcionan resultados más adecuados. En definitiva, puede concluirse que, además de presentar adecuadas propiedades teóricas, el enfoque no paramétrico propuesto se comporta razonablemente bien y puede aplicarse con resultados satisfactorios a otras situaciones prácticas diferentes a la aquí planteada.

Apéndice: Resultados auxiliares

Teorema (Cramer-Wold)

Sean X, X_1, X_2, \dots vectores aleatorios de dimensión k . Entonces, $X_n \xrightarrow{d} X$ si y sólo si para todas las elecciones de $a = (a(1), \dots, a(k)) \in \mathfrak{R}^k$:

$$\sum_{i=1}^k a(i)X_n(i) \xrightarrow{d} \sum_{i=1}^k a(i)X(i).$$

Teorema (Convergencia dominada)

Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y $\{f_n\}$ una sucesión de funciones medibles, tales que $|f_n| \leq g \forall n$, (siendo g medible e integrable), y $f_n \xrightarrow{cs} f$. Entonces f es integrable y además

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu,$$

siendo $\int_{\Omega} f d\mu$ la integral de f respecto de la medida μ .

Teorema (Teorema central del límite de Lindeberg)

Sean $\{X_k\}_{k \in \mathbb{N}}$ variables aleatorias independientes de media m_k y varianza σ_k^2 finitas. Sea $c_n^2 = \sum_{k=1}^n \sigma_k^2$ y sea $F_k(\cdot)$ la función de distribución de X_k . Si $\varepsilon > 0$ se cumple

$$\frac{1}{c_n^2} \sum_{k=1}^n \int_{\{|x-m_k| \geq \varepsilon c_n\}} (x - m_k)^2 dF_k(x) \xrightarrow{n \rightarrow \infty} 0,$$

entonces

$$\frac{S_n - E(S_n)}{c_n} \xrightarrow{L} N(0, 1).$$

Teorema (Fuller (1976))

Sea $\{\mathbf{X}_n\}$ una secuencia de variables aleatorias k -dimensionales de elementos $\{X_{jn}, j = 1, 2, \dots, k\}$, y sea $\{\mathbf{r}_n\}$ una secuencia de vectores k -dimensionales con elementos reales y positivos $\{r_{jn}, j = 1, 2, \dots, k\}$ tales que

$$\begin{aligned} X_{jn} &= O_P(r_{jn}), \quad j = 1, 2, \dots, t \\ X_{jn} &= o_P(r_{jn}), \quad j = t + 1, \dots, k \end{aligned}$$

Sea $g_n(\mathbf{x})$ una secuencia de funciones reales (Borel-medibles) definidas en un espacio Euclídeo k -dimensional, y sea $\{s_n\}$ una secuencia de números reales positivos. Sea $\{\mathbf{a}_n\}$ una secuencia de vectores k -dimensionales no aleatorios. Si

$$g_n(\mathbf{a}_n) = o(s_n)$$

para todas las secuencias $\{\mathbf{a}_n\}$ tales que

$$\begin{aligned} a_{jn} &= O(r_{jn}), \quad j = 1, 2, \dots, t \\ a_{jn} &= o(r_{jn}), \quad j = t + 1, \dots, k \end{aligned}$$

entonces

$$g_n(\mathbf{X}_n) = o_P(s_n)$$

Corolario (Fuller (1976)) Sea $\{X_n\}$ una secuencia de variables aleatorias escalares tales que

$$X_n = a + O_P(r_n), \tag{5.15}$$

donde $r_n \rightarrow 0$ cuando $n \rightarrow \infty$. Si $g(x)$ es una función con s derivadas continuas en $x = a$, entonces

$$g(X_n) = g(a) + g^{(1)}(a)(X_n - a) + \dots + \frac{1}{(s-1)!} g^{(s-1)}(a)(X_n - a)^{s-1} + O_P(r_n^s)$$

Bibliografía

- Agrawal, R., Faloutsos, C., Swami, A.N. (1993). Efficient similarity search in sequence databases. In: Lomet, D. (editor), *Proceedings of the 4th International Conference of Foundations of Data Organization and Algorithms (FODO)*, pp. 69-84. Springer Verlag, Chicago, Illinois.
- Alagón, J. (1989). Spectral discrimination for two groups of time series. *J. Time Ser. Anal.*, **10**, 203-214.
- Alon, J., Sclaroff, S., Kollios, G., Pavlovic, V. (2003). Discovering clusters in motion time-series data. In: *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*.
- Alonso, A., Berrendero, J., Hernandez, A., Justel, A. (2004). Time series clustering based on forecast densities. *Comput. Stat. Data An.*, **51**, 762-776.
- Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.
- Bagnall, A.J., Janakec, G., Zhang, M. (2003). *Clustering time series from mixture polynomial models with discretised data*. Technical Report CMP-C03-17, School of Computing Sciences, University of East Anglia.
- Baragona, R. (2001). A simulation study on clustering time series with metaheuristic methods. *Quaderni di Statistica*, **3**, 1-26.
- Bar-Joseph, Z., Gerber, G., Glifford, D., Jaakkola, T., Simon, I. (2002). A new approach to analyzing gene expression time series data. In: *Proceedings of the Sixth Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, pp. 39-48.
- Boets, J., De Cock, K., Espinoza, M., De Moor, B. (2005). Clustering time series, subspace identification and cepstral distances. *Commun. Inf. Syst.*, **5**, 69-96.

- Bohte, Z.D., Cepar, D., Kosmelj, K. (1980). Clustering of time series. *Proceedings in Computational Statistics, COMPSTAT 80*, pp. 587-593.
- Brillinger, D.R. (1981). *Time Series Analysis: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.
- Brockwell P.J., Davis R.A. (1996). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York.
- Cadez, I.V., Gaffney, S., Smyth, P. (2000). A general probabilistic framework for clustering individuals and objects. In: *Proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 140-149.
- Caiado, J. (2006). *Distance-based methods for classification and clustering of time series*. Tesis Doctoral. Universidad Técnica de Lisboa.
- Caiado, J., Crato, N. (2005). Discrimination between deterministic trend and stochastic trend processes. In: Janssen, J., Lenca, P. (editors), *Proceedings of the XIth International Conference on Applied Stochastic Models and Data Analysis*, pp. 1419-1424.
- Caiado, J., Crato, N, Peña, D. (2006). A periodogram-based metric for time series classification. *Comput. Stat. Data An.*, **50**, 2668-2684.
- Capon, J. (1965). An asymptotic simultaneous diagonalization procedure for pattern recognition. *J. Inform. Cont.*, **8**, 264-281.
- Chan, H.T., Chinipardaz, R., Cox, T.F. (1996). Discrimination of AR, MA and ARMA time series models. *Commun. Stat-Theor. M.*, **25**, 1247-1260.
- Chatfield, C. (1979). Inverse autocorrelations. *J. R. Statist. Soc. A*, **142**, 363-377.
- Claeskens, G., Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Stat.*, **31**, 1852-1884.
- Cleveland, W.S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics*, **14**, 277-298.
- Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, **74**, 829-836.
- Cleveland W.S., Devlin S.J. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.*, **83**, 597-610.
- Coronnello, C., Tumminello, M., Lillo, F., Miccichè, S., Mantegna, R.N. (2005). Sector identification in a set of stock return time series traded at the London stock exchange. *Acta Phys. Pol. B*, **36**: 2653-2679.

- Dargahi-Noubary, G.R. (1992). Discrimination between Gaussian time series based on their spectral differences. *Commun. Stat-Theor. M.*, **21**, 2439-2458.
- Dargahi-Noubary, G.R. (1999). A linear discriminant for Gaussian time series. *J. Time Ser. Anal.*, **20**, 145-153.
- Dargahi-Noubary, G.R., Laycock, P.J. (1981). Spectral ratio discriminants and information theory. *J. Time Ser. Anal.*, **2**, 71-86.
- Davenport, W.B., Root, W.L. (1958). *An Introduction to the Theory of Random Signals and Noise*. McGraw-Hill, New York.
- Dette, H., Neumeier, N. (2001). Nonparametric analysis of covariance. *Ann. Stat.*, **29**, 1361-1400.
- Eubank R.L. (1988). *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.
- Fan, J. (1992). Design-adaptive nonparametric regression. *J. Am. Stat. Assoc.*, **87**, 998-1004.
- Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall, London.
- Fan, J., Heckman, N.E., Wand, M.P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Am. Stat. Assoc.*, **90**, 141-150.
- Fan, J., Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scand. J. Statist.*, **25**, 359-369.
- Fan, J., Zhang, W. (2004). Generalised likelihood ratio tests for spectral density. *Biometrika*, **91**, 195-209.
- Ferraty, F., Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Comput. Stat. Data An.*, **44**, 161-173.
- Fraley, C., Raftery, A.E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.*, **97**, 611-631.
- Francisco, M., Opsomer, J., Vilar, J.M. (2004). A plug-in bandwidth selector for local polynomial regression with correlated errors. *J. Nonparametr. Stat.*, **16**, 127-151.
- Francisco-Fernández, M., Vilar, J.M., Vilar, J.A. (2003). On the uniform strong consistency of local polynomial regression under dependence conditions. *Commun. Stat-Theor. M.*, **32**, 2415-2440.

- Francisco Fernández, M., Vilar Fernández J.M. (2001). Local polynomial regression estimation with correlated errors. *Commun. Stat-Theor. M.*, **30**, 1271-1293.
- Francisco Fernández, M., Vilar Fernández J.M. (2004). Weighted local nonparametric regression with dependent errors: study of real private residential fixed investment in the USA. *Stat. Inference Stoch. Process.*, **7**, 69-93.
- Fuller, W.A. (1976). *Introduction to statistical time series*. John Willey & Sons, New York.
- Gaffney, S., Smyth, P. (1999). *Trajectory clustering with mixtures of regression models*. Technical Report 99-15, Department of Information and Computer Science, University of California.
- Gaffney, S., Smyth, P. (2003). Curve clustering with random effects regression mixtures. In: Bishop, C.M., Frey, B.J. (editors). *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*.
- Galbraith, J., Jiaqing, L. (1999). Cluster and discriminant analysis on time series as a research tool. In: *UTIP Working Paper Number 6*. Austin: Lyndon B. Johnson School of Public Affairs. The University of Texas at Austin.
- Galeano, P., Peña, D. (2000). Multivariate analysis in vector time series. *Resenhas*, **4**, 383-403.
- Gasser, T., Müller, H.G. (1979). Kernel estimation of regression functions. In: *Smoothing Techniques for Curve Estimation*. Lecture Notes in Mathematics, **757**, pp. 23-68. Springer-Verlag, New York.
- Gavrilov, M., Anguelov, D., Indyk, P., Motwani, R. (2000). Mining the stock market: which measure is best?. In: *Proceedings of the KDD*, pp. 487-496.
- Gersch, W., Martinelli, F., Yonemoto, J., Lew, M.D., McEwan, J.A. (1979). Automatic classification of electroencephalograms: Kullback-Leibler nearest neighbor rules. *Science*, **205**, 193-195.
- Gersch, W., Yonemoto, J. (1977). Automatic classification of multivariate EEG, using an amount of information measure and the eigenvalues of parametric time series model features. *Comput. Biomed. Res.*, **10**, 297-316.
- Gevins, A.S., Veager, C.L., Diamond, S.L., Spire, J., Zeitlin, G., Gevins, A. (1975). Automated analysis of the electrical activity of the human brain (EEG): A progress report. *Proc. IEEE*, **63**, 1382-1399.

- Györfi, L., Härdle, W., Sarda, P., Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*. Lecture Notes in Statistics, **60**. Springer-Verlag, Berlin
- Hall, P., Hart, J.D. (1990). Bootstrap test for differences between means in nonparametric regression. *J. Am. Stat. Assoc.*, **85**, 1039-1049.
- Hall, P., Marron, J.S. (1990). On variance estimation in nonparametric regression. *Biometrika*, **77**, 415-419.
- Han, J., Kamber, M. (2001). *Data mining: concepts and techniques*. Morgan Kaufmann, San Francisco, pp. 346-389.
- Hand, D.J. (1982). *Kernel discriminant analysis*. Electronic & Electrical Engineering Research Studies: Pattern Recognition & Image Processing Series 2. Research Studies Press. John Wiley & Sons, Chichester-New York.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Boston.
- Härdle, W., Marron, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.*, **18**, 63-89.
- Hastie, T., Buja, A., Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Stat.*, **23**, 73-102.
- Hastie, T., Loader, C. (1993). Local regression: automatic kernel carpentry. *Stat. Sci.*, **8**, 120-143.
- Hastie T., Tibshirani R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Hastie, T., Tibshirani, R. (1996). Discriminant analysis by gaussian mixtures. *J. Roy. Stat. Soc. B*, **58**, 155-176.
- Helstrom, C.W. (1968). *Statistical Theory of Signal Detection*. Pergammon Press, Oxford.
- Hermann, E., Gasser, T., Kneip, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, **79**, 783-795.
- Hirsch, B., DuBois, D. (1991). Self-esteem in early adolescence: the identification and prediction of contrasting longitudinal trajectories. *J. Youth Adoles.*, **20**, 53-72.
- Hirukawa, J. (2006). Cluster analysis for non-Gaussian locally stationary processes. *Int. J. Theor. Appl. Finance*, **9**, 113-132.

- Huang, T.S., Schreiber, W.F., Tretiak, O.J. (1971). Image processing. *Proc. IEEE*, **59**, 1586-1609.
- Jain, A. K., Dubes, R.C. (1988). *Algorithms for clustering data*. Prentice Hall Inc, New York.
- Johnson, R.A., Wichern, D.W. (1982). *Applied Multivariate Statistical Analysis* (3rd ed.). Prentice-Hall, Englewood Cliffs, New York.
- Jones, M.C. (1997). A variation on local linear regression. *Stat. Sinica*, **7**, 1171-1180.
- Kakazos, D., Papantoni-Kakazos, P. (1980). Spectral distance measuring between Gaussian processes. In: *IEEE Transactions on Automatic Control*, *AC-25*, pp. 950-959.
- Kakizawa, Y. (1996). Discriminant analysis for non-gaussian vector stationary processes. *J. Nonparametr. Stat.*, **7**, 187-203.
- Kakizawa, Y., Shumway R.H., Taniguchi M. (1998). Discrimination and clustering for multivariate time series. *J. Am. Stat. Assoc.*, **93**, 328-340.
- Kalpakis, G.D., Gada, K., Puttagunta, V. (2001). Distance measure for effective clustering of ARIMA time-series. In: *Proceedings of the IEEE International Conference on Data Mining*, San Jose, CA, pp. 273-280.
- King, E.C., Hart, J.D., Wehrly, T.E. (1991). Testing the equality of two regression curves using linear smoothers. *Stat. Probab. Lett.*, **12**, 239-247.
- Kooperberg, C., Stone, C.J., Truong, Y.K. (1995). Rate of convergence for logspline spectral density estimation. *J. Time Ser. Anal.*, **16**, 389-401.
- Kosmelj, K., Batagelj, V. (1990). Cross-sectional approach for clustering time-varying data. *J. Classif.*, **7**, 99-109.
- Koul, H.L., Schick, A. (1997). Testing for the equality of two nonparametric regression curves. *J. Stat. Plan. Infer.*, **65**, 293-314.
- Krzysko, M., Wolynski, W. (1997). Linear discriminant functions for stationary time series. *Biometrical J.*, **39**, 955-973.
- Kulasekera, K.B. (1995). Comparison of regression curves using quasi-residuals. *J. Am. Stat. Assoc.*, **90**, 1085-1093.
- Kulasekera, K.B., Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Am. Stat. Assoc.*, **92**, 500-511.

- Kulasekera, K.B., Wang, J. (1998). Bandwidth selection for power optimality in a test of equality of regression curves. *Statist. Probab. Lett.*, **37**, 287-293.
- Kullback, S. (1978). *Information theory and statistics*. Peter Smith, Gloucester, MA.
- Kullback, S., Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Stat.*, **22**, 79-86.
- Lejeune, M. (1985). Estimation non-paramétrique ar noyaux: régression polynomiale mobile. *Revue de Statist. Appliq.*, **33**, 43-68.
- Li, C. (2000). *A Bayesian approach to temporal data clustering using the hidden Markov model methodology*. PhD thesis, Vanderbilt University, Nashville.
- Li, C., Biswas, G. (1999). Clustering sequence data using hidden markov model representation. In: *SPIE'99 conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology*, pp. 14-21.
- Li, C., Biswas, G. (1999b). Profiling of dynamic system behaviors using hidden markov model representation. In: *Proceedings of the ICSC'99 Advances in Intelligent Data Analysis (AIDA'99)*.
- Li, C., Biswas, G. (1999c). Temporal pattern generation using hidden markov model based unsupervised classification. In: Hand, D., Kok, K, Berthold, M. (editors). *Advances in Intelligent Data Analysis*, Lecture Notes in Computer Science, pp. 1642. Springer.
- Li, C., Biswas, G. (2000). Bayesian clustering for temporal data using hidden markov model representation. In: *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 543-550.
- Li, C., Biswas, G., Dale, M., Dale, P. (2001). Building models of ecological dynamics using HMM based temporal data clustering-a preliminary study. In: Hoffmann, F. et al. (Eds.), IDA 2001, *Lecture Notes in Computer Science*, 2189, pp. 53-62.
- Liao, T.W. (2005). Clustering of time series data-a survey. *Pattern Recogn.*, **38**, 1857-1874.
- Liggett, W.S. Jr. (1971). On the asymptotic optimality of spectral analysis for testing hypotheses about time series. *Ann. Math. Stat.*, **42**, 1348-1358.
- Maharaj, E.A. (1996). A significance test for classifying ARMA models. *J. Stat. Comput. Sim.*, **54**, 305-331.
- Maharaj, E.A. (2000). Clusters of time series. *J. Classif.*, **17**, 297-314.

- Markel, J.D., Gray, A.H. Jr. (1976). *Linear Prediction of Speech*. Springer, Berlin.
- Masry E. (1996). Multivariate regression estimation: Local polynomial fitting for time series. *Stoch. Proc. Appl.*, **65**, 81-101.
- Masry E. (1996b). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Ser. Anal.*, **17**, 571-599.
- Masry E., Fan J. (1997). Local polynomial estimation of regression function for mixing processes. *Scand. J. Stat.*, **24**, 165-179.
- McCullagh P., Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- McLachlan, G.J. (1992). *Discriminant analysis and statistical pattern recognition*. Wiley, New York.
- Meisel, W.S. (1972). *Computer Oriented Approaches to Pattern Recognition*. Academic Press, New York.
- Müller H.G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Lecture Notes in Statistics, **46**. Springer-Verlag, Berlin.
- Müller, H.G., Stadtmüller, U. (1988). Detecting dependencies in smooth regression models. *Biometrika*, **75**, 639-650.
- Nadaraya, E.A. (1964). Remarks on nonparametric estimates for density functions and regression curves. *Theor. Probab. Appl.*, **15**, 134-137.
- Neumeyer, N., Dette, H. (2003). Nonparametric analysis of covariance. *Ann. Stat.*, **31**, 880-920.
- Nielsen, S.F. (2005). Local linear estimating equations: uniform consistency and rate of convergence. *J. Nonparametr. Stat.*, **17**, 493-511.
- Oates, T. (1999). Identifying distinctive subsequences in multivariate time series by clustering. In: Chaudhuri, S., Madigan, D. (editors). *Fifth International Conference on Knowledge Discovery and Data Mining*, pp. 322-326, San Diego, CA, USA.
- Oates, T., Firoiu, L., Cohen, P.R. (1999). Clustering time series with hidden markov models and dynamic time warping. In: *Proceedings of the IJCAI-99 Workshop on Neural, Symbolic and Reinforcement Learning Methods for Sequence Learning*, pp. 17-21.
- Oates, T., Firoiu, L., Cohen, P.R. (2001). Using dynamic time warping to bootstrap HMM-based clustering time series. *Lect. Notes Comput. Sc.*, **1828**, 35- 52.

- Oates, T., Schmill, M.D., Cohen, P.R. (2000). Identifying qualitatively different outcomes of actions: Gaining autonomy through learning. In: Sierra, C., Gini, M., Rosenschein, J.S. (editors). *Proceedings of the Fourth International Conference on Autonomous Agents*, pp. 110-111. Barcelona, Catalonia, Spain.
- Otnes, R.K., Enochson, L. (1978). *Applied Time Series Analysis*. Wiley, New York.
- Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scand. J. Stat.*, **27**, 143-176.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065-1076.
- Parzen, E. (1990). *Time series, statistics and information*. IMA preprint series 663, Institute of mathematics and its applications, University of Minnesota.
- Peña, D., Poncela, P. (2006). Nonstationary dynamic factor analysis. *J. Stat. Plan. Infer.*, **136**, 1237-1257.
- Pértega, S., Vilar, J.A. (2007). A comparative study of several parametric and nonparametric approaches for time series classification. Enviado a *Int. Stat. Rev.*
- Piccolo, D. (1990). A distance measure for classifying ARIMA models. *J. Time Ser. Anal.*, **11**, 153-164.
- Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York.
- Priestley, M.B. (1989). *Spectral Analysis of Time Series*. Probability and Newline Mathematical Statistics. A Series of Monograph and Text books. Academic Press.
- Priestley, M.B., Chao, M.T. (1972). Nonparametric function fitting. *J. Roy. Stat. Soc. B*, **34**, 384-392.
- Ramoni, M., Sebastiani, P., Cohen, P. (2002). Bayesian clustering by dynamics. *Mach. Learn.*, **47**, 91-121.
- Ridgeway, G. (1997). *Finite discrete Markov processes*. Technical Report MSR-TR-97-21, Microsoft Research.
- Ripley, B.D. (1996). *Pattern recognition and neural networks*. Cambridge University Press, Cambridge, UK.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832-837

- Rosenfeld, A., Weszka, J.S. (1976). Picture recognition. In: Fu, K.S. (ed.), *Digital Pattern Recognition*, pp. 135-166. Springer, Berlin.
- Ruppert, D., Wand, M.P. (1994). Multivariate weighted least squares regression. *Ann. Stat.*, **22**, 1346-1370.
- Sakiyama, K. (2002). Some statistical applications for locally stationary processes. *Sci. Math. Jpn.*, **56**, 231-250.
- Selin, I. (1965). *Detection Theory*. Princeton Univ. Press, Princeton.
- Shaw, C.T., King, G.P. (1992). Using cluster analysis to classify time series. *Physica D, Nonlin. Phenom.*, **58**, 288-298.
- Shumway, R.H. (1982). Discriminant Analysis for Time Series. *Handbook of Statistics*, **2**, 1-46.
- Shumway, R.H. (1996). Statistical approaches to seismic discrimination. In: Husebye, E.S., Dainty, A.M. (eds.) *Monitoring a comprehensive test ban treaty*, pp. 791-803. Kluwer, Dainty, Dordrecht.
- Shumway, R.H. (2003). Time-frequency clustering and discriminant analysis. *Statist. Probab. Lett.*, **63**, 307-314.
- Shumway, R.H., Blandford R. (1993). An examination of some new and classical short period discriminants. *Tech. Rept. No. TR-74-10*, Seismic Data Analysis Center, Alexandria, U.S.A.
- Shumway, R.H., Stoffer, D.S. (2000). *Time series analysis and its applications*. Springer-Verlag, New York.
- Shumway, R.H., Unger, A.N. (1974). Linear discriminant functions for stationary time series. *J. Am. Stat. Assoc.*, **69**, 948-956.
- Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
- Smyth, P. (1997). Clustering sequences with hidden markov models. In: Mozer, M.C., Jordan, M.I., Petsche, T. (editors). *Advances in Neural Information Processing Systems*, 9, pp. 648. The MIT Press.
- Stone C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595-645.
- Taniguchi, M., Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series*. Springer-Verlag, New York.

- Tjøstheim D. (1975). Autorregresive representation of seismic P-wave signals with application to the vowel recognition problem. *Geophys. J. Roy. Astron. Soc.*, **43**, 269-291.
- Tong, H., Dabas, P. (1990). Cluster of time series models: an example. *J. Appl. Stat.*, **17**, 187-198.
- Tong, H., Yeung, I. (2000). On tests for self-exciting threshold autoregressive type non-linearity in partially observed time series. *Appl. Stat.*, **40**, 43-62.
- Van Trees, H.L. (1968). *Detection Estimation and Modulation Theory, Parts I, II*. Wiley, New York.
- Vilar Fernández, J.M. (1987). *Estimación no paramétrica de la función de densidad y función de predicción en series de tiempo*. Tesis Doctoral. Universidad de Santiago de Compostela.
- Vilar, J.A., Pértega, S. (2004). Discriminant and cluster analysis for Gaussian stationary processes: local linear fitting approach. *Nonparametr. Stat.*, **16**, 443-462.
- Vilar, J.M., González, W. (2004). Nonparametric comparison of curves with dependent errors. *Statistics*, **38**, 81-99.
- Vilar, J.M., Vilar, J.A., González, W. (2007). Bootstrap tests for nonparametric comparison of regression curves with dependent errors. *Test*, **16**, 123-144.
- Vilar, J.M., Vilar, J.A., Pértega, S. (2004). *Nonparametric classification of time series: application to the bank share prices in spanish stock market*. Reports in Statistics and Operational Research, No. 04-03, Ed. Nino, Santiago de Compostela.
- Vilar, J.M., Vilar, J.A., Pértega, S. (2007). Classifying Time Series Data: A Nonparametric Approach. Enviado a *J. Classif.*
- Wahba G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- Watson, G.S. (1964). Smooth regression analysis. *Sankhya, A*, **26**, 359-372.
- Wedderburn R.W.M. (1974). Quasilikelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, **61**, 439-447.
- Welch, P.D., Wimpres, R.S. (1961). Two multivariate statistical computer programs and their application to the vowel recognition problem. *J. Acoust. Soc. Amer*, **33**, 426-434.
- Wolf, J.J. (1976). Speech recognition and understanding. In: Fu, K.S. (ed.), *Digital Pattern Recognition*, pp. 167-203. Springer, Berlín.

- Xiong, Y., Yeung, D.Y. (2002). Mixtures of ARMA models for model-based time series clustering. In: *Proceedings of the IEEE International conference on Data Mining*, Maebaghi City, Japan.
- Young, S.G., Bowman, A.W. (1995). Nonparametric analysis of covariance. *Biometrics*, **51**, 920-931.
- Zhang, G., Taniguchi, M. (1994). Discriminant analysis for stationary vector time series. *J. Time Ser. Anal.*, **15**, 117-126.
- Zhang, G., Taniguchi, M. (1995). Nonparametric approach for discriminant analysis in time series. *J. Nonparametr. Stat.*, **5**, 91-101.
- Zhong, S. (2002). Probabilistic model-based clustering of complex data. PhD thesis, University of Texas at Austin.
- Zhong, S., Ghosh, J. (2002). HMMs and coupled HMMs for multi-channel EEG classification. In: *Proceedings IEE Int. Joint Conf. on Neural Networks*.
- Zhong, S., Ghosh, J. (2002b). A unified framework for model-based clustering. In *Intelligent Engineering Systems Through Artificial Neural Networks (ANNIE)*.
- Zhong, S., Ghosh, J. (2002c). A unified framework for model-based clustering and its application to clustering time sequences. Technical report, Department of Electrical and Computer Engineering, University of Texas.
- Zhong, S., Ghosh, J. (2003). Scalable, balanced model-based clustering. In: *Proceedings of SIAM Int. Conf. on Data Mining*.