

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES
Departamento de Comercialización e Investigación de Mercados



**UNA METODOLOGÍA DE MINERÍA DE DATOS
PARA LA AGRUPACIÓN DE SERIES
TEMPORALES: APLICACIÓN AL SECTOR DE LA
CONSTRUCCIÓN RESIDENCIAL.**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

Magdalena Ferrán Aranz

Bajo la dirección de los doctores

José María Santiago Merino
Lorenzo Escot Mangas

Madrid, 2011

ISBN: 978-84-694-3761-2

© Magdalena Ferrán Aranz, 2011

Universidad Complutense de Madrid

Facultad de Ciencias Económicas y Empresariales

Departamento de Comercialización e Investigación de Mercados



UNA METODOLOGÍA DE MINERÍA DE DATOS PARA

LA AGRUPACIÓN DE SERIES TEMPORALES:

APLICACIÓN AL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL

Tesis Doctoral

Autora: Magdalena Ferrán Aranz

Directores:

José María Santiago Merino
Lorenzo Escot Mangas

Madrid, 2010

Agradecimientos

Tras varios años de esfuerzo, con diversas interrupciones y cambios de rumbo, ha llegado el momento de poner punto y aparte a este trabajo. Antes de nada me gustaría dejar constancia de mi agradecimiento a todos aquellos que me ha acompañado en esta etapa de mi vida.

Gracias a mis directores de tesis, José María Santiago Merino y Lorenzo Escot Mangas, por haber confiado en mi capacidad para llevar a cabo este proyecto. Gracias a mis compañeros de la Escuela Universitaria de Estadística de la UCM y, en particular, del departamento de Estadística e Investigación Operativa III por crear un ambiente de trabajo difícilmente mejorable. Gracias a mis amigos de TINSA por enseñarme tanta cosas.

Gracias especialmente a mis padres y hermanos porque siempre os siento a mi lado. Y gracias a Ernesto, Gabriel y Santiago, a quienes dedico este trabajo, por hacer que cada día sea un gran día.

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO 1. EL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL.....	5
1.1 La relevancia del sector de la construcción residencial	6
1.2 La sensibilidad del sector de la construcción residencial al tipo de interés.....	8
1.3 La construcción residencial en España entre 1983 y 2008.....	10
1.4 Indicadores de actividad en la construcción residencial.....	13
1.5 Tipos de referencia del mercado hipotecario.....	14
CAPÍTULO 2. ANÁLISIS CLUSTER DE SERIES TEMPORALES.....	17
2.1 El Análisis Cluster.....	18
2.1.1 Tipos de agrupación	18
2.1.2 El algoritmo de las k -medias	19
2.1.3 Inicialización del algoritmo de las k -medias.....	21
2.2 Análisis Cluster de Series Temporales	26
2.2.1 Minería de datos temporales.....	26
2.2.2 Similitud entre series temporales numéricas.....	27
2.3 Agrupación de series temporales en el sector de la construcción residencial	31
CAPÍTULO 3. ELEMENTOS TEÓRICOS DE LA METODOLOGÍA DEL HAZ DE RECTAS	34

CAPÍTULO 4. METODOLOGÍA DEL HAZ DE RECTAS: APLICACIÓN AL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL	48
4.1 Condiciones de aplicación	49
4.2 Homogeneización de la escala	52
4.3 Cálculo de los centroides iniciales	54
4.4 Generación de los grupos	57
4.5 Interpretación de la solución	60
Apéndices	65
CAPÍTULO 5. EXTRACCIÓN DEL CONJUNTO DE SERIES RESUMEN A PARTIR DE UNA SERIE COVARIABLE	74
5.1 Relación entre visados y tipo de interés	74
5.2 Extracción del conjunto de series resumen	79
5.2.1 Datos de entrada	79
5.2.2 Condiciones de aplicación	79
5.2.3 Homogeneización de la escala	82
5.2.4 Generación del conjunto de series resumen	82
5.3 Interpretación de la solución	85
Apéndices	1
CAPÍTULO 6. DESCRIPCIÓN DE LA AGRUPACIÓN OBTENIDA	94
6.1 Comparación de los centroides iniciales	94
6.2 Comparación de los centroides finales	96
6.3 Representación factorial de las distancias	99
6.4 Descripción de la serie nacional	102
6.5 Descripción de las series provinciales	105
6.5.1 Descripción de las provincias clasificadas en los grupos quinto y sexto	105
6.5.2 Descripción de las provincias clasificadas en el cuarto grupo	107
6.5.3 Descripción de las provincias clasificadas en el tercer grupo	109
6.5.4 Descripción de las provincias clasificadas en los grupos primero y segundo ..	112
6.5.5 Conclusiones generales	114
CAPÍTULO 7. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	116
BIBLIOGRAFÍA	120
ANEXOS	
I. Un esquema de la metodología	131
II. Aplicación del <i>Algoritmo de las k-medias</i>	136

III.Gráficos provinciales.....	145
Málaga	149
Alicante/Alacant	148
Huesca	149
Tarragona	150
Girona	151
Santa Cruz de Tenerife	152
Balears/Balears	153
Las Palmas	154
Almería	155
La Rioja	156
Lugo	157
Castellón/Castelló	158
León	159
Zaragoza	160
Zamora	161
Murcia	162
Asturias	163
Orense/Ourense	164
Cantabria	165
Álava/Araba	166
Córdoba	167
Valencia	168
Huelva	169
Madrid	170
Guipúzcoa/Gipuzkoa	171
Navarra	172
Segovia	173
Badajoz	174
Valladolid	175
Granada	176
Jaén	177
Sevilla	178
Palencia	179
Barcelona	180
La Coruña/A Coruña	181
Vizcaya/Bizkaia	182
Cáceres	183
Cádiz	184
Salamanca	185
Ávila	186
Cuenca	187
Albacete	188
Teruel	189
Burgos	190
Lleida	191
Pontevedra	192
Toledo	193
Guadalajara	194
Soria	195
Ciudad Real	196

INTRODUCCIÓN

El análisis de series temporales es fundamental no sólo en la economía, sino en campos tan diversos como, por ejemplo, la ingeniería, la biología, la medicina, la psicología, la sociología e incluso las artes, siendo uno de los problemas más comunes el de comparar y relacionar un conjunto grande de series temporales. Las técnicas que facilitan la extracción de información de bases de datos voluminosas forman parte de la Minería de Datos. En este trabajo se presenta una metodología de minería de datos para la comparación de series temporales económicas que midan un mismo fenómeno o variable procedentes de diferentes ámbitos, localidades, agentes, condiciones, etc. Para desarrollar la metodología, realizaremos una aplicación al Mercado de la Vivienda, más específicamente al Sector de la Construcción Residencial, y compararemos la evolución de la oferta de vivienda nueva en las distintas provincias españolas.

La elección del Mercado de la Vivienda se justifica por ser uno de los mercados con mayor peso en la economía española y con mayor protagonismo en la actual situación de crisis económica. En el año 1982 se publicaba el análisis posiblemente más exhaustivo del sector de la construcción residencial en España (Alcaide et al, 1982). El prólogo, de Evelio Verdera y Tuelles, decía lo siguiente:

En periodos de aguda crisis económica como la que estamos viviendo, que exigen de los Gobiernos la más correcta asignación de los recursos, resulta más que evidente la importancia de los análisis sectoriales en general, tanto en los procesos de desagregación de los planes macroeconómicos como en la coordinación de los grandes proyectos.

De modo particular, el sector de la construcción alcanza en la economía española un relieve singular, por sus propias singularidades y su incidencia en la actividad económica en su conjunto, que

aconsejan un tratamiento separado y autónomo, tanto en las contabilidades nacionales como en los estudios regionales.

Es observación generalizada que el sector de la construcción ha sido uno de los motores del desarrollo económico español durante la fase de expansión y constituye el principal generador de desempleo a lo largo de esta prolongada crisis económica. Y, por lo menos, no parece discutible que el espectacular aumento del desempleo, que en diciembre de 1981 alcanzó el 15,4% de la población activa, coincide con la crisis del sector de la construcción, cuyo paro alcanzaba durante ese mismo periodo un 27,06%, es decir, proporcionaba el 20,5 de paro de la economía española. El sector de la construcción aporta a la economía española entre un 6 y un 9% del P.I.B. según datos y estimaciones alternativas, y aunque su importancia relativa, dentro del conjunto nacional, ha disminuido a partir de 1976, no puede desconocerse la importancia de cuanto afecte al mismo, para la evolución de nuestra realidad económica. (Alcaide et al, 1982: vii)

Esta descripción del panorama económico de principios de los ochenta es integralmente aplicable a la reciente situación. Concretamente, más de un cuarto de siglo después Julio Rodríguez López hace el siguiente análisis:

La economía española creció en 2007 en un 3,7%, en línea con el perfil expansivo de la década que terminó en dicho ejercicio. Sin embargo, dicho crecimiento comprendió evoluciones muy dispares entre los dos semestres del año. En el primer semestre de 2007 el ritmo de variación intersemestral anualizado fue del 2%, mientras que en el segundo el ritmo citado se frenó hasta el 1,2%. Esta tendencia hacia un menor crecimiento continuó en el primer semestre de 2008, en el que la variación intersemestral anualizada del PIB fue sólo del 0,4%.

El decreciente crecimiento del PIB estuvo marcado por una evolución paralela a la baja del consumo de los hogares, cuya variación intersemestral fue sólo del 0,7% en el segundo semestre de 2007 (2% en el primero). Pero el cambio más abrupto vino de la mano de la inversión en vivienda, cuyo peso en el PIB se había aproximado al 9,5% en el primer semestre de 2008. La variación intersemestral anualizada de dicho componente de la demanda pasó desde el 5% del primer semestre de 2007 al 0,8% en el segundo semestre, experimentado un ritmo negativo de variación, el -8,1% en el primer semestre de 2008. (Rodríguez, 2008: 755-756)

y añade:

La desaceleración del crecimiento de la economía española ha estado acompañada de un comportamiento del empleo notablemente menos expansivo que el registrado en los años precedentes. De acuerdo con la Contabilidad Nacional Trimestral, el número de empleos a tiempo

completo estimado en dicha fuente estadística desaceleró su expansión desde el 2,9% interanual de 2007 hasta el 0,9% del primer semestre de este ejercicio. El empleo en la construcción, tras aumentar en un 4,8% en 2007, disminuyó en un -4,5% en el primer semestre de 2008 sobre el mismo periodo del año precedente.

La afiliación a la Seguridad Social a fines de septiembre de 2008 había retrocedido en un -2,5% sobre el mismo mes de 2007, correspondiendo en dicho periodo un retroceso de -15,6% a la afiliación en el sector de la construcción. El paro registrado de septiembre de 2008 ascendió a 2.625.000 personas, un 30,1% más que en el mismo mes del año anterior, correspondiente a un aumento absoluto anual de 608.000 en el número de parados. El crecimiento relativo más acusado del desempleo tuvo lugar en la construcción, sector en el que los parados aumentaron en un año en un 64,1%. (Rodríguez, 2008: 757)

Llama poderosamente la atención el impresionante paralelismo entre los dos periodos. En la economía española, como en toda economía de mercado, el fenómeno cíclico se presenta de forma recurrente. Para determinar qué factores intervienen en la trayectoria de la serie nacional agregada es importante analizar en qué consisten las similitudes entre las distintas series provinciales. En esta tesis nos centraremos en el aspecto cíclico de la oferta de vivienda nueva, desarrollando una metodología para la agrupación de las series temporales provinciales. La metodología que se propone consiste, básicamente, en construir los K centroides iniciales del *Algoritmo de las k-medias* para un *Análisis Cluster de Series Temporales*. La solución obtenida permitirá interpretar las similitudes entre las distintas series temporales provinciales. Consideramos que la presente metodología, a la que denominaremos Metodología del haz de rectas, es muy útil como estudio exploratorio previo a la aplicación de cualquier modelo estadístico o económico que persiga objetivos tanto de tipo explicativo como de tipo predictivo.

La tesis consta de siete capítulos. Los dos primeros describen los contextos en los que se aplica y se desarrolla la metodología. En el Capítulo 1 se describe la relevancia del Sector de la construcción residencial en la Economía y se justifican determinadas decisiones tales como la elección del tipo de interés como principal proceso generador de las oscilaciones de la serie de vivienda nueva. El Capítulo 2 está dedicado al marco teórico en el que se desarrolla la metodología: *El Análisis Cluster de Series Temporales*. El capítulo comienza con una introducción al *Análisis Cluster* general, para pasar a exponer más detalladamente el *Algoritmo de las k-medias* y los diversos *Métodos de inicialización* de este algoritmo.

El núcleo de esta tesis corresponde al desarrollo de la metodología propiamente dicho y comprende los capítulos tercero y cuarto. En el Capítulo 3 se exponen los elementos teóricos sobre los que se sustenta la metodología. En el Capítulo 4 se ilustran las sucesivas etapas del proceso de agrupación de las series temporales mediante un caso práctico: se construye un conjunto de series resumen que se utilizarán como

centroides iniciales del *Algoritmo de las k-medias* para la agrupación de las series de visados de vivienda nueva.

La metodología propuesta en este trabajo se fundamenta en la existencia de una serie temporal como proceso generador de las oscilaciones del conjunto de series objeto de análisis aunque, para su aplicación, no sea necesario el conocimiento de dicha serie. En el Capítulo 5 supondremos que las fluctuaciones del tipo de interés son las principales responsables de las fluctuaciones de las series de visados, lo que nos permitirá repetir el proceso de construcción del conjunto de series resumen desarrollado en el Capítulo 4 pero, en este caso, bajo el supuesto de que el proceso generador de las oscilaciones del conjunto de series temporales objeto de análisis es conocido.

En el Capítulo 6 se utilizan los resultados obtenidos en los capítulos 4 y 5 para interpretar la evolución del sector de la construcción residencial en términos de las diferencias y similitudes entre las distintas provincias. Finalmente, en el Capítulo 7 se exponen las conclusiones y se describen las futuras líneas de investigación.

CAPÍTULO 1

EL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL

Recientemente, en el 31º Seminario de Política Económica del Banco de la Reserva Federal de la Ciudad de Kansas dedicado a “Housing, Housing Finance and Monetary Police”, Edgard E. Leamer presentó un artículo de gran repercusión internacional, titulado “Housing is the business cycle”, en el que describe el importante papel que juega la vivienda en el ciclo económico estadounidense. Leamer reflexiona:

The bad news is that I am not a macro-economist. Wicksell and Hayek and Keynes and Friedman and Tobin and Lucas and Prescott speak foreign languages with which I have familiarity but not mastery (Leamer, 2007: 1).

En mi caso la “mala noticia” es que soy matemático. Este *accidente* en mi formación académica ha supuesto un cierto inconveniente a la hora de afrontar la información relativa al sector de la construcción residencial. En el complejo proceso de revisión de la literatura en relación a la estructuración de la información observé que, en definitiva, lo que estaba buscando era argumentos que me permitieran justificar determinadas decisiones, tales como:

- La elección del **Sector de la construcción residencial** para ilustrar la metodología y, en particular, la elección del **volumen de viviendas iniciadas**.
- La elección del **Tipo de interés** como factor responsable de las fluctuaciones de la serie de vivienda nueva.
- La elección del **Periodo objeto de análisis**: los veinticinco años comprendidos entre 1983 y 2008.

- La elección de la **Serie de visados de los Colegios de Arquitectos** como indicador de vivienda iniciada.
- La elección concreta de la serie del **Mibor/Euribor**.

Ante la dificultad de manejar un lenguaje que no es el de mi propia disciplina académica, con el riesgo de incurrir en errores conceptuales, he optado por transcribir literalmente aquellos párrafos procedentes de la literatura especializada que, creo, mejor describen mis decisiones.

1.1. LA RELEVANCIA DEL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL

La elección del sector de la construcción residencial radica en su relevancia dentro de la Economía. En el caso español, al revisar la literatura que justifique dicha relevancia encontramos frases como éstas encabezando la introducción a los artículos: En “La demanda de construcción en España. Razones y tendencias recientes” dice P. Taltavull:

Como es sabido, el sector de la construcción se ha convertido en protagonista del crecimiento económico español desde el último lustro (Taltavull, 2007: 37).

Por otra parte y abundando en la misma idea, en “Los booms inmobiliarios en España. Un análisis de tres periodos” señala Rodríguez:

En la evolución de la economía española durante la segunda mitad del siglo XX y primeros del XXI resulta evidente la trascendencia que ejercen los procesos de boom o de auge del subsector inmobiliario (Rodríguez, 2006: 76).

Pero no sólo ellos destacan la importancia del sector en nuestra economía. Así, Ferraz, en “Financiación del boom de la vivienda” sostiene que:

Uno de los rasgos característicos del intenso y prolongado ciclo expansivo que ha venido disfrutando la economía española a lo largo de la última década ha sido el notable protagonismo del sector de la construcción y, en especial, de la vivienda (Ferraz, 2006: 214)

En este mismo artículo, el autor valora en términos cuantitativos el efecto del sector de la construcción en la economía española en el año 2006:

No se puede negar al sector de la construcción el importante efecto arrastre que ha ejercido sobre el resto de sectores, su contribución a la notable generación de empleo, al vigor de la inversión, al avance de la riqueza de las familias o a la absorción e integración del elevado número de inmigrantes que han entrado en nuestro país en los últimos años y que han contribuido a dinamizar y flexibilizar la economía y a aumentar la capacidad productiva.

La relevancia del sector y su efecto dinamizador resultan evidentes a la vista de su peso en el valor añadido bruto de nuestra economía, el 10,4 por 100, y en el empleo, el 14 por 100, porcentajes que en 1996 eran del 6,6 por 100 y 9,1 por 100 respectivamente. Además, ha sido el sector más dinámico en los últimos años, con un crecimiento medio anual, en la última década, del 5,3 por 100 en términos reales, frente al 3,6 por 100 del conjunto de la economía, y ha generado en el mismo periodo más de 1,3 millones de empleos, el 27,3 por 100 del total. Asimismo, la construcción ha absorbido en los últimos años gran parte de la mano de obra inmigrante que ha llegado a nuestro país, dado que el 19 por 100 de los ocupados en este sector son extranjeros, frente al 11 por 100 de media en el global de la economía (Ferraz, 2006: 214)

Aunque es relativamente sencillo hacer estadísticas que valoren puntualmente el efecto del sector de la construcción en la economía en general, valorar dicho efecto en términos dinámicos es bastante complejo. Valga como ejemplo cómo en la revista *Perspectivas de la economía mundial*, el capítulo dedicado a la “Evolución del ciclo de la vivienda e implicaciones para la política monetaria” nos indica:

La explosión reciente de los precios de la vivienda y de la inversión residencial en muchas economías avanzadas, y las fuertes correcciones que posteriormente experimentaron algunas, han reavivado el debate sobre el vínculo entre la vivienda y el ciclo económico, y sobre la respuesta monetaria correcta frente a las vicisitudes del sector.

Pese al consenso general de que la situación del sector de la vivienda acarrea implicaciones importantes para el nivel de la actividad económica, las causas son aún objeto de polémica. Concretamente, no hay acuerdo sobre la dinámica de la inversión residencial, sus consecuencias para el ciclo económico y el impacto que ejercen las fluctuaciones de los precios de la vivienda en el gasto de los consumidores (Cardarelli et al., 2008: 113).

Parece evidente entonces que el sector de la construcción residencial es clave en la economía no sólo española, sino en la de cualquier país. Sin embargo, la justificación de esta evidencia es una cuestión bastante más compleja. Frases como la anterior ponen de manifiesto la dificultad que supone describir la relación entre el ciclo de la

construcción y el ciclo económico, máxime cuando al referirnos al ciclo de la construcción no está claro si estamos hablando de precios o de volumen de construcción. Volviendo a Leamer podemos obtener alguna aclaración al respecto:

“Housing is the most important sector in our economic recession and any attempt to control the business cycle needs to focus especially on residential investment”. (Leamer, 2007: 2)

“Eight of the ten recessions were preceded by sustained and substantial problems in housing”. (Leamer, 2007: 13)

“The reason housing is so important in recessions is that homes have a volume cycle, not a price cycle”. (Leamer, 2007: 25)

“Though builders are more motivated sellers than existing home owners, for new homes too is a volume cycle, not a price cycle”. (Leamer, 2007: 26)

Estas cuatro afirmaciones encadenadas resumen la importancia del análisis de los ciclos en la construcción de vivienda en términos del volumen y no de los precios. Nos quedamos con esta conclusión. Así, centraremos nuestra atención en el estudio de la trayectoria del volumen de viviendas iniciadas.

1.2. LA SENSIBILIDAD DEL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL AL TIPO DE INTERÉS

Desde las primeras publicaciones sobre el mercado inmobiliario, uno de los aspectos más estudiado ha sido el efecto de las fluctuaciones del tipo de interés sobre el sector de la construcción residencial.

R.F. Muth, en su artículo “Housing market dynamics”, afirma que, durante el periodo de la posguerra mundial, en los años cincuenta, cada episodio de amplias oscilaciones en la construcción de nuevas viviendas estuvo asociado con fluctuaciones de los tipos de interés:

Every episode of wide swings in the construction of new housing in the postwar period has been associated with fluctuations in rates of interest. A variety of explanations have been offered for this association. In the fifties it was ceilings on contract interest rates of VA and FHA mortgages; this explanation lost credence when it was realized that FHA mortgages, at least, were widely discounted [Alberts (1962) and Guttentag (1961)]. During the seventies, the outflow of funds from depository mortgage lenders due to rate ceilings on passbook accounts was often identified as the culprit. Yet, after savings

rates were effectively deregulated in mid-1978 by the introduction of market-rate money market accounts by depository lenders, housing starts have continued to respond to fluctuations in interest rates as before (Muth, 1988: 345).

Por su parte R.J. Pozdena, en “Do interest rates still affect housing?” (Pozdena, 1990), resalta la fuerte relación inversa que ha existido históricamente entre las fluctuaciones de los tipos de interés y de las viviendas iniciadas:

The housing market historically has displayed pronounced cycles in investment activity. In the last 30 years, for example, the variations in the volume of new housing starts has been 11/2 times greater than the variations in GNP over the same period. Economists and central bankers long have been interested in understanding these fluctuations for several reasons. First, the cycles in housing activity empirically have been useful leading indicators of the general business cycle. In most recessions, residential fixed investment appears to have led both declines in business investment and GNP. Its perceived value as a leading indicator has made understanding housing cycles important to economic forecasters.

Second, understanding investment behaviour in the economy is important for understanding aggregate economic volatility. Investment spending is the most volatile component of aggregate demand, and spending on new home and apartment construction (residential fixed investment) is the most volatile component of total investment spending. Thus, although residential investment is a smaller proportion of GNP than business fixed investment, its volatility has a particularly important influence on volatility in national income over time.

Finally, because housing activity apparently has been sensitive to changes in interest rates, the housing sector historically has been an important channel through which monetary policy has influenced economic activity. In fact, there is some evidence that the economy may react more quickly to shocks that make their way through the housing component of aggregate demand than through the business fixed investment component. Thus, when monetary authorities have decided to slow the national economy to bring inflation down, they typically have used monetary restraint to raise interest rates, which tended to contract the housing sector, and in this way, they were able to effect the desired cooling of the overall economy.

A key linkage in this process, of course, has been the relationship between housing activity and the level of interest rates. Historically there does appear, indeed, to have been a strong, inverse relationship between disturbances to interest rates and changes in housing starts (Pozdena, 1990: 3).

Así quedaría justificada la elección del tipo de interés como uno de los principales factores responsables de las fluctuaciones de la serie de vivienda nueva, aunque la cantidad de viviendas ofertadas en un momento de tiempo dependa también de aspectos tales como, por ejemplo, el precio, los costes de construcción o el stock.

1.3. LA CONSTRUCCIÓN RESIDENCIAL EN ESPAÑA ENTRE 1983 Y 2008

Volviendo a “Do interest rates still affect housing?”, Pozdena plantea que la liberalización financiera pudo haber alterado la relación entre vivienda y tipos de interés, distinguiendo un antes y un después de 1983 a la hora de detectar diferencias significativas en dicha relación. Esta fecha resulta crucial en su estudio ya que las diferencias son apreciables a partir de la fecha reseñada:

Activity in the housing market traditionally has been very sensitive to changes in interest rates. This sensitivity has had important implications both for participants in the housing industry and monetary policy. Theory suggests, however, that financial innovations and deregulation in recent years may have altered the link between housing and interest rates. In this paper, the theoretical linkages are discussed and studied empirically for the periods before and after 1983. A significant difference in the strength and nature of the linkages is revealed (Pozdena, 1990: 3).

En lo que se refiere a la liberalización financiera en España y su relación con el sector de la construcción, Taltavull resalta el efecto dinamizador que sobre el mercado residencial tuvo la publicación de la Ley 2/81 de Regulación del Mercado Hipotecario:

El mercado hipotecario, a la par que el financiero, ha experimentado un proceso de liberalización que comenzó a principios de los años ochenta con la publicación de la Ley 2/81 de regulación del mercado hipotecario (Ley 2/1981, de 25 de marzo, y su reglamento, el Real Decreto 685/1982, de 22 de junio). Desde entonces, el panorama de financiación a la vivienda ha experimentado una gran transformación, siendo reconocido que uno de los factores que han dinamizado el mercado residencial ha sido la flexibilidad existente en el mercado hipotecario español (Taltavull, 2004: 95).

Más concretamente, en un trabajo posterior, esta misma autora destaca la existencia de dos grandes periodos expansivos en la evolución de la oferta residencial en España entre 1959 y 2004. Mientras que el primero de ellos, que iría desde inicios de los sesenta hasta mitad de los setenta, lo relaciona con la fuerte demanda asociada al crecimiento de la población y al crecimiento económico, el segundo, que comenzó a

mediados de los noventa¹, lo relaciona con los cambios en las variables financieras y económicas:

Básicamente hay dos grandes periodos expansivos que han configurado la mayor parte del parque residencial: el primero conocido iría desde inicios de los sesenta hasta mitad de los setenta, consecuencia de la fuerte demanda ligada al crecimiento de la población y económico (Taltavull, 2001), y el segundo, inconcluso, que comenzó a mediados de la década de los noventa, tras un proceso corto de boom (1986-1991) y recesión (1992-1994) y ha superado el máximo histórico de intensidad edificadora. Este comportamiento reciente ha generado sorpresa por cuanto que se ha considerado que la oferta existente de viviendas y el ritmo de edificación previo eran suficientes para cubrir las necesidades residenciales existentes en la España de final del siglo XX (Taltavull, 2006: 162).

Si se analiza el caso español por regiones y por periodos en las dos últimas décadas, puede observarse cómo la edificación experimenta intensos crecimientos en algunas comunidades autónomas (Baleares, Canarias, Cantabria, Castilla-la Mancha, Comunidad Valenciana, Murcia y La Rioja) ajenos en parte al aumento de la población registrados durante las distintas etapas de este período. Esto implica que el número de viviendas disponibles por mil habitantes ha aumentado, haciendo referencia clara a un uso distinto al básico del parque de viviendas. El hecho de que este fenómeno se concentre en regiones costeras e islas (salvo Castilla-La Mancha) indica que una parte importante de esta oferta ha sido destinada a un nuevo mercado (residencial-turístico o segundas residencias) más relacionado con los cambios en las variables financieras y económicas que en las demográficas, como sucede en los mercados de primeras viviendas (Taltavull, 2006: 163).

Buscando más argumentos volvemos a “Los booms inmobiliarios en España. Un análisis de tres periodos” (Rodríguez, 2006), donde el autor relaciona el pequeño periodo de auge comprendido entre 1986 y 1991 con los cambios introducidos en el funcionamiento de los sistemas financieros y destaca, además, el papel decisivo que han desempeñado los tipos de interés en el más reciente episodio de auge del sector residencial:

En este trabajo se analizan algunas de las circunstancias que rodearon a los episodios acaecidos en los periodos 1969-1974, 1986-1991 y 1997-2006, considerándose como no terminado el episodio iniciado a partir de los últimos meses de 1997 y que se prolongaba todavía en el verano de 2006 (...)

¹ Actualmente concluido, pero no en las fechas de la publicación del artículo de P. Taltavull.

El segundo periodo de auge analizado (1986-1991) se derivó sobre todo de los cambios introducidos en el funcionamiento de los sistemas financieros en un amplio número de países occidentales (...) (Rodríguez, 2006: 76).

En resumen, la mayor integración de los mercados hipotecarios con el resto del sistema financiero, la posibilidad abierta de generar préstamos a largo plazo a partir de recursos captados a la vista, el fin de las restricciones al crecimiento de la financiación hipotecaria y de los controles sobre los niveles de los tipos de interés, y la desregulación general de los sistemas financieros, facilitaron una mejora sustancial en las condiciones de los préstamos inmobiliarios, en especial de los destinados a la compra de vivienda, tanto en España como en un amplio número de países occidentales. Puede decirse que la gran expansión del crédito hipotecario, que se manifestó en el fuerte aumento de la proporción de dicho crédito respecto del PIB, se inició en el periodo de auge de 1986-1991 (Rodríguez, 2006: 82).

El tercer episodio de auge, no terminado hasta ahora, ha sido sobre todo consecuencia del shock de demanda que han supuesto las excepcionales condiciones de financiación a largo plazo al subsector inmobiliario (...) En el episodio en cuestión ha destacado el intenso volumen de construcción residencial desarrollado y el que alcanzasen una proporción significativa las compras de viviendas por razones de inversión. El auge, en el que los tipos de interés están desempeñando un papel decisivo, se ha extendido a un número amplio de países occidentales, apareciendo España entre los que este episodio ha alcanzado una mayor intensidad y duración (Rodríguez, 2006: 77)

En el momento de la publicación de dicho artículo todavía se prolongaba el episodio de auge iniciado a partir de los últimos meses de 1997. Dos años después, este mismo autor, en “De la crisis inmobiliaria a la crisis financiera. El duro otoño de 2008” (Rodríguez, 2008), presentaba el siguiente panorama:

(...) Pero el cambio más abrupto vino de la mano de la inversión en vivienda, cuyo peso en el PIB se había aproximado al 9,5% en el primer semestre de 2008. La variación intersemestral anualizada de dicho componente de la demanda pasó desde el 5% del primer semestre de 2007 al 0,8% en el segundo semestre, experimentado un ritmo negativo de variación, el -8,1% en el primer semestre de 2008 (Rodríguez, 2008: 755-756).

La justificación de considerar el periodo comprendido entre 1983 y 2008, más concretamente, el delimitado por el tercer trimestre de 1983 y el cuarto de 2008, para la aplicación de la metodología responde a tratar de abarcar la información relativa al

sector de la construcción residencial de los veinticinco últimos años². El periodo comienza en los años precedentes a la primera etapa de auge posterior a la publicación de la Ley 2/81 de Regulación del Mercado Hipotecario en España y finaliza con la depresión más reciente.

1.4. INDICADORES DE ACTIVIDAD EN LA CONSTRUCCIÓN RESIDENCIAL

Existen diversos indicadores para el análisis dinámico del sector de la construcción residencial³. Para desarrollar la metodología se ha elegido la serie de visados de los Colegios de Arquitectos. Frente al resto, esta serie presenta dos grandes ventajas: su representatividad y su longitud, tal como ponen de manifiesto R. Vergés y D. Martín en los dos siguientes párrafos incluidos en el artículo “Series históricas de edificación residencial. Estadística de visados de los colegios de arquitectos 1960-2006”:

La peculiaridad de las series españolas de edificación residencial es que encuentran su fuente en los Colegios de Arquitectos desde 1960, es decir antes de la invitación a la estadística cursada por el art. 5b de la Ley Estatal 2/1974 de Colegios Profesionales (...).

La estadística de edificación nació a finales de los años 50 por necesidades de previsión actuarial de la hermandad de Previsión de Arquitectos, antiguo brazo social del Consejo Superior de Colegios de Arquitectos de España (CSCAE). Desde entonces, la extracción de datos a partir de los expedientes de visado ha sido siempre tarea de los Colegios, recopilándose inicialmente al nivel nacional por la Dirección General para Vivienda y Arquitectura (DGVA) del ministerio de tutela, ya sea de Vivienda, MOPU o Fomento.

Desde 1960 y hasta 1990, la única difusión de dichos datos la realizaba la propia DGVA a través de un boletín de carácter interno que aún subsiste y en el que figuran asimismo unas series de “Viviendas Iniciadas y Terminadas” no observadas sino estimadas a partir de los visados. Conviene recordar que desde 1991, es el CSCAE quien recopila y publica los datos residenciales de sus Colegios, provisionalmente en su Boletín Trimestral de Coyuntura y definitivamente en las series históricas difundidas vía web (...).

En principio y una vez obtenida la licencia, el cliente contrata al aparejador para que asuma la dirección de la ejecución de obra, haciéndole entrega del proyecto de ejecución del arquitecto, último responsable de la obra. El aparejador visa el contrato en su colegio y recoge los datos más relevantes del proyecto. Desde 1992 y por

² El dato más reciente relativo al inicio de vivienda en el momento de preparar este trabajo correspondía al segundo trimestre de 2008, razón por la que se partió del tercer trimestre de 1983, para abarcar así un periodo de veinticinco años exacto. Posteriormente, se incorporó la información relativa al segundo semestre de 2008, ampliando así el periodo de observación en medio año.

³ Para una descripción detallada de los mismos, véase Taltavull, 2001: 173-174.

convenio con el Consejo General de Colegios de Aparejadores, éstos remiten dicha información a la FGPEP que la difunde bajo el título “Obras en Edificación” (Sánchez de Rivera, 1992 y S.).

La estadística de los Aparejadores se solapa pues con la de los Arquitectos y ofrece hoy análoga exhaustividad, pero se distingue de ella al especificar tipología y promotor y al no desglosar, en cambio, el régimen de vivienda. Su interés reside sobre todo en documentar un trámite en principio ulterior a la licencia y por tanto más cercano al inicio efectivo de la obra. Así, el número de viviendas en los proyectos de arquitecto que antes publicaba sólo la DGVA, se encuentra ahora en estas dos otras estadísticas de Fomento (Vergés y Martín, 2007: 566-567).

El material disponible muestra que el proyecto de ejecución visado es un excelente predictor del visado de dirección de obra, el cual a su vez es un evidente predictor del inicio de obra. El análisis previo sugiere también que, salvo incidencia de variables coyunturales como el afloramiento de construcción ilegal o del aplazamiento de proyectos en tiempos difíciles, el visado de proyectos como único indicador avanzado de la edificación real, puede ser además un indicador eficiente (Vergés y Martín, 2007: 589).

Se fija de esta forma el número de visados de los Colegios de Arquitectos como indicador de actividad en la construcción residencial.

1.5. TIPOS DE REFERENCIA DEL MERCADO HIPOTECARIO

La elaboración de los primeros índices de referencia por parte del Banco de España se remonta al año 1986, tal como apuntan J. García-Montalvo y M. Mas, en su artículo “La vivienda y el sector de la construcción en España”:

La generalización de hipotecas a tipos de interés variables recomendó la elaboración de índices de referencia por parte de Banco de España (...). Los primeros índices de referencia son los tres contenidos en la Resolución 20.6.1986 de la DGTPF (...) que continúan publicándose en la actualidad. También continúan publicándose los índices de referencia contenidos en la Resolución DGTPF 5.12.1989 (...). Los índices de referencia oficiales existentes en la actualidad son los siguientes: 1) el tipo interbancario (Mibor) a un año, al que se ha unido recientemente el Euribor; 2) el tipo activo de referencia de cajas de ahorros (indicadores CECA), los tipos medios de los préstamos hipotecarios a más de tres años para adquisición de vivienda libre; 3) de los bancos; 4) de las cajas de ahorros y 5) del conjunto de entidades; y 6) el rendimiento interno de la Deuda Pública en

mercado secundario entre dos y seis años (García-Montalvo y Mas, 2000: 122)

En la década de los noventa los tipos de interés hipotecarios experimentaron una importante caída con importantes repercusiones sobre el sector de la construcción residencial:

En líneas generales, la evolución de los índices de referencia del mercado hipotecario ha seguido una trayectoria paralela a la de los tipos de interés generales de la economía, (...). El aspecto más llamativo (...) es la importante caída experimentada por los tipos de interés hipotecarios en los últimos años. Por ejemplo, el Mibor cayó desde el 13,08% en el año 1991, al 7,35% en el año 1996 hasta el 3,03% en marzo de 1999. Esta caída está teniendo repercusiones muy importantes sobre el sector de la construcción residencial y también sobre la actividad de las entidades financieras que están experimentando elevadas tasas de crecimiento en los créditos hipotecarios, (...) acompañadas de reducciones en la morosidad bancaria por el abaratamiento de los préstamos (García-Montalvo y Mas, 2000: 122).

Dichos índices surgieron como respuesta a la generalización de hipotecas a tipos de interés variables y tuvieron una amplia aceptación:

Según la encuesta realizada por esta institución, a 31 de diciembre de 1997 del total de entidades integrantes de la muestra, el 86,4% había adoptado como referencia en los préstamos a interés variable los tipos oficiales, lo que suponía un 91,3% del importe de los mismos (García-Montalvo y Mas, 2000: 124).

De entre los distintos índices, el Mibor ha sido históricamente el de mayor aceptación hasta que, con la entrada en vigor del euro, es sustituido por otro índice equivalente, el Euribor, que se convierte en el tipo de interés de referencia clave:

Dado que las entidades cuentan con un abanico de seis índices de referencia, una cuestión importante es cuál de ellos recibe mayor atención. Los datos de la encuesta del Banco de España permiten comprobar que el índice Mibor a un año se destaca entre los cinco restantes, puesto que representa el 40,8% del número de préstamos, y el 50,5% del importe. Además, el escenario futuro que contempla la encuesta indica el mantenimiento de la preponderancia del tipo Mibor en el horizonte del año 2010, o el índice equivalente que lo sustituya tras la implantación del euro.

En el nuevo contexto financiero que supone la entrada en vigor del euro, el Euribor se presenta como el tipo de interés de referencia clave. En principio, las diferencias entre el Mibor y el Euribor deberían ser nulas si se cumplen ciertos requisitos: igual calidad crediticia de las entidades encuestadas para la elaboración de ambos índices, idénticas condiciones fiscales o de regulación, y también la posibilidad de que se produzcan relaciones de arbitraje. Así, siendo común el marco regulador para ambos mercados, las diferencias que puedan surgir deben achacarse a diferencias en el rating de los prestatarios y/o a la eficiencia del mercado (...). a comienzos del año 1999, tras la implantación del euro, ambos tipos han convergido. Por otra parte, existe coincidencia en recomendar por parte de los expertos la sustitución paulatina, pero decidida, del Mibor por el Euribor (García-Montalvo y Mas, 2000: 124-125).

Esta es la razón de elegir estos dos índices de referencia para la elaboración de la serie temporal del tipo de interés, el Mibor al principio del periodo de análisis y el Euribor desde 1999.

CAPÍTULO 2

ANÁLISIS CLUSTER DE SERIES TEMPORALES

En la presente Tesis Doctoral se desarrolla una metodología para la agrupación de series temporales que consiste, básicamente, en proporcionar los K centroides iniciales del *Algoritmo de las k -medias*. En este capítulo se describe el marco teórico en el que se desarrolla dicha metodología: El *Análisis Cluster de Series Temporales*. Comenzaremos el capítulo con una introducción al *Análisis Cluster*, ofreciendo una breve descripción de los distintos tipos de agrupación, para pasar a exponer más detalladamente el algoritmo concreto que nos ocupa: el *Algoritmo de las k -medias*, y los diversos *Métodos de Inicialización* de este algoritmo.

El objetivo del *Análisis Cluster* aplicado sobre series temporales es encontrar una agrupación tal que las series en un mismo grupo sean parecidas entre sí y distintas de las series en otro grupo. Para ello se considerarán los algoritmos generales de *Análisis Cluster* y, en particular, el *Algoritmo de las k -medias*. Una forma de proceder consistiría en tratar cada serie temporal como un punto en un espacio de tantas dimensiones como observaciones, definir una distancia para cada par de puntos y aplicar el algoritmo de agrupación sobre la nube de puntos. Sin embargo, aunque la más directa, esta no es la única forma de abordar el problema. Liao (2005a) clasifica los *Métodos Cluster* de series temporales en tres categorías diferentes, dependiendo de cuál sea la unidad sobre la que se aplique el algoritmo de agrupación. El segundo apartado de este capítulo proporciona una breve revisión de estos métodos.

Finalizaremos este capítulo con la exposición de la única aplicación que hemos encontrado en la literatura relativa al problema concreto que nos ocupa: la agrupación de las series temporales correspondientes a visados de la construcción en distintas localidades. Concretamente revisaremos el trabajo de Stock y Watson (2008) en el que, en función de las similitudes entre las series de visados de la construcción para vivienda

nueva en Estados Unidos en el periodo 1969-2007, realizan una agrupación de los estados americanos en cinco regiones.

2.1 EL ANÁLISIS CLUSTER

Los avances tecnológicos de los últimos años han permitido la recogida y almacenado de enormes conjuntos de datos de todo tipo para la posterior interpretación del fenómeno al que representan. La Minería de Datos (DM, Data Mining) puede definirse como un proceso que reúne un conjunto de herramientas de diversas ciencias (Estadística, Informática, Matemáticas, Ingeniería, entre otras) que persigue extraer conocimiento oculto o información no trivial de grandes volúmenes de datos: “*Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner*” (Hand et al., 2001:1). Las bases de la minería de datos se encuentran en la inteligencia artificial y en el análisis estadístico. En un proceso estándar de minería de datos, desarrollado en varias etapas, la agrupación o *clustering* suele ser la primera de ellas (Fayyad, 1996), tanto si se trata de datos estáticos como de datos dinámicos.

2.1.1 Tipos de agrupación

Dada una base de datos formada por un número grande de elementos el *Análisis Cluster* tiene como objetivo crear una partición en grupos basada únicamente en la información proporcionada por los propios datos que describen los elementos y sus relaciones. El objetivo es que la similitud entre los elementos pertenecientes a un mismo grupo sea grande frente a la de elementos en distintos grupos. Cuanto mayor sea la similitud dentro de un grupo y menor la similitud entre grupos más diferenciados estarán y mejor será la solución de agrupación. Cada uno de estos grupos se denomina *cluster*, región en la que la densidad de elementos es localmente más alta que en otras regiones (Peña et al., 1999).

Existen diferentes criterios para establecer la agrupación: completa frente a parcial, estricta frente a difusa y jerárquica frente a particiva:

Completa frente a parcial: Una agrupación es completa cuando todos los elementos están asignados a un grupo, mientras que en caso contrario será parcial.

Estricta frente a difusa: Una agrupación es estricta si cada elemento pertenece sólo a un grupo y difusa si se permite que pertenezca a más de un grupo con distintos grado de pertenencia.

Jerárquica frente a particiva: Es la distinción más frecuente. Dada una base de datos con n elementos, una agrupación es jerárquica si agrupa los n elementos de la base de datos de forma secuencial, siguiendo una estructura de árbol, y particiva si los divide en k

grupos, con k no superior a n , tales que cada uno de ellos contiene al menos un elemento. La mayor parte de los métodos de agrupación puede ser clasificado en una de estas dos grandes categorías, aunque existen otros métodos, menos utilizados, que no pueden serlo, tales como los métodos basados en densidades, los métodos basados en rejillas y los métodos basados en modelos (Han y Lamber, 2001).

Conseguir la solución óptima en una agrupación particiva requeriría la enumeración exhaustiva de todas y cada una de las posibles divisiones de los n elementos en k grupos. La alternativa es utilizar un método heurístico. Si, en particular, se trata de hacer una partición estricta, el algoritmo heurístico de mayor aceptación es el de las k -medias.

2.1.2 El algoritmo de las k -medias

El *Algoritmo de las k -medias* (MacQueen, 1967; Hartigan, 1975), es uno los algoritmos para el *Análisis Cluster* más populares¹. Se trata de un algoritmo particivo para datos numéricos, en el que se supone un número fijo K de grupos².

Dado un valor K y un conjunto formado por N vectores P -dimensionales, con N grande, de la forma:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})' \quad i = 1, \dots, N$$

el objetivo es clasificar los N vectores en K grupos (C_1, C_2, \dots, C_K) de tal forma que la similitud entre los vectores pertenecientes a un mismo grupo sea grande frente a la de vectores en distintos grupos. Al número de vectores en el grupo k -ésimo, C_k , se le denominará n_k .

Si

$$X_{N \times P} = \{x_{ij}\}_{N \times P}$$

es la matriz de datos $N \times P$, el *Algoritmo de las k -medias* lleva a cabo la partición de tal forma que el cuadrado de la distancia euclídea entre un vector cualquiera y el vector centroide de su grupo correspondiente sea menor que el cuadrado de la distancia euclídea a los centroides de los restantes grupos. El centroide del grupo C_k se obtiene como el promedio de los valores en la columna de la matriz de datos para los elementos del grupo correspondiente. Concretamente, la j -ésima componente del centroide del grupo C_k viene dada por:

$$\bar{x}_j^{(k)} = \frac{1}{n_k} \sum_{i \in C_k} x_{ij} \quad [2.1]$$

¹ Aunque la denominación de Algoritmo de las k -Medias es debida a MacQueen (1967), el primero en proponer el algoritmo para la agrupación de datos en el sentido de minimizar la variabilidad dentro de los grupos fue Forgy (1965) (Bock, 2008).

² Son diversos los autores que han propuesto métodos para estimar el número de cluster de una clasificación. Un desarrollo detallado al respecto puede verse en Gordon (1999) o en Steinley (2003).

y el centroide completo por:

$$\bar{x}^{(k)} = (\bar{x}_1^{(k)}, \bar{x}_2^{(k)}, \dots, \bar{x}_p^{(k)})'$$

La versión estándar del *Algoritmo de las k-medias* consiste en un proceso iterativo formado por las siguientes etapas (Steinly y Brusco, 2007):

1. Se definen K vectores P -dimensionales como centroides iniciales,

$$s^{(k)} = (s_1^{(k)}, s_2^{(k)}, \dots, s_p^{(k)})' \quad \text{para } 1 \leq k \leq K$$

2. Se calcula el cuadrado de la distancia euclídea entre el i -ésimo vector y el k -ésimo centroide inicial,

$$d^2(x_i, s^{(k)}) = \sum_{j=1}^P (x_{ij} - s_j^{(k)})^2 \quad i = 1, \dots, N \quad 1 \leq k \leq K$$

Los N vectores se clasifican en K grupos en términos de la mínima distancia. El grupo k -ésimo estará formado por todos aquellos vectores tales que la mínima distancia se obtenga para el k -ésimo centroide inicial.

3. Una vez hecha la clasificación inicial de los N vectores, se calculan sus centroides como en la expresión [2.1].
4. Se calcula la distancia de cada vector a los centroides:

$$d^2(x_i, \bar{x}^{(k)}) = \sum_{j=1}^P (x_{ij} - \bar{x}_j^{(k)})^2 \quad i = 1, \dots, N \quad 1 \leq k \leq K$$

y se clasifica cada vector en el grupo correspondiente al centroide más próximo.

5. Se recalculan los centroides para actualizar la pertenencia a los grupos.
6. Los pasos 4 y 5 se repiten hasta que ningún vector cambie de un grupo³.

Obsérvese que, al tratar de encontrar una “buena” partición del conjunto de datos mediante el algoritmo iterativo de las k -medias, lo que se intenta es minimizar la suma de los cuadrados de los errores o, dicho de otra forma, la variabilidad dentro de los grupos:

³ O se verifique algún criterio de convergencia, por ejemplo, que de una solución a otra sólo cambie de grupo un determinado porcentaje de elementos, que las diferencias entre los centroides de dos soluciones consecutivas sea menor que un determinado valor o que se alcance un número prefijado de iteraciones.

$$\sum_{j=1}^P \sum_{k=1}^K \sum_{i \in C_k} (x_{ij} - \bar{x}_j^{(k)})^2$$

Se puede probar que la variabilidad dentro de los grupos disminuye en cada iteración del algoritmo (Cuadras, 2007), pero no hay garantías de que la solución obtenida sea la óptima en términos globales, dado que depende de cuál sea la elección de los centroides iniciales para la clasificación.

Aunque la distancia euclídea es la elección más frecuente, existen otras distancias que pueden ser utilizadas con este algoritmo. Concretamente, si los elementos objeto de análisis son series temporales, dado que la distancia euclídea ignora la estructura de autocorrelación, puede ser conveniente, como veremos más adelante, considerar otra medida.

2.1.3 Inicialización del algoritmo de las k -medias

En ocasiones, los centroides iniciales son conocidos o se dispone de información *a priori* sobre los grupos. El problema surge cuando no hay ningún tipo de información en este sentido y la elección de los centroides iniciales se convierte en un aspecto crítico para la calidad de la solución de agrupación (Lin et al., 2004). Una alternativa consistiría en aplicar el algoritmo de forma repetida sobre distintos conjuntos de centroides iniciales hasta encontrar la mejor solución. Pero esta forma de proceder, además de poco elegante, sería poco práctica, incluso intratable, en el caso de un conjunto de datos muy voluminoso. Por ello, algunos autores han desarrollado métodos para dar valores a los centroides iniciales, comúnmente denominados *semillas* del algoritmo⁴. Los métodos clásicos son (Anderberg, 1973):

- Elegir los k primeros elementos del conjunto de datos (MacQueen, 1967). Supuesto que el orden de los datos no responde a ningún criterio, esta es la opción más económica y sencilla.
- Numerados los elementos de 1 a m , elegir los situados en las posiciones $m/k, 2m/k, \dots, m(k-1)/k$ y m . Este método es casi tan sencillo como el anterior y trata de eliminar el posible efecto de una secuencia no aleatoria en los datos.
- Elegir de forma subjetiva k elementos del conjunto de datos.
- Numerados los elementos de 1 a m , elegir los elementos correspondientes a k números aleatorios en el rango de 1 a m (McRae, 1971)

⁴ Tal como indican Jain y Dubes (1988), los mejores resultados se obtienen cuando la partición inicial es muy parecida a la solución final. Posiblemente por ello los algoritmos más complejos (en general, los más recientes) buscan obtener una solución inicial parecida a la que será la final, convirtiéndose ellos mismos en algoritmos de agrupación más que de inicialización.

- Generar k puntos-vectores tales que cada coordenada consista en un valor aleatorio para el rango de la variable asociada. Este método presenta el inconveniente de que, salvo que el conjunto de datos “rellene” el espacio, algunos de estos puntos-vectores puedan encontrarse a mucha distancia respecto de todos los elementos del conjunto de datos.
- Dado un valor de distancia d_1 (por ejemplo, el promedio de las distancias euclideas entre cada par de elementos), calcular el número de elementos con distancia inferior a d_1 (la *densidad*) para cada elemento del conjunto de datos. El elemento con mayor densidad será elegido como primera *semilla*. Las restantes se obtienen según orden decreciente de la densidad, con la condición de que cada nueva semilla se encuentre a una distancia mínima d_2 de las previamente seleccionadas. El proceso finaliza cuando todos los elementos restantes tienen densidad cero. Si el número de semillas generadas es superior al deseado, se agrupan mediante un método jerárquico (por ejemplo, el Método del centroide) hasta obtener una solución de k puntos (Astrahan, 1970). El inconveniente de este método radica en la dependencia de la solución respecto de los parámetros d_1 y d_2 .
- Considerar como primera semilla el primer elemento del conjunto de datos. Se calcula la distancia al segundo elemento que, si dicha distancia es mayor que una determinada cota, se convertirá en la segunda semilla; en caso contrario, se pasa al siguiente elemento de la base de datos, hasta encontrar una segunda semilla. Elegida la segunda semilla, se pasa al siguiente elemento y se calcula su distancia a las dos semillas. Si ambas distancias son mayores que una determinada cota, dicho elemento se convertirá en la tercera semilla; en caso contrario, se pasa al siguiente elemento de la base de datos, hasta encontrar una tercera semilla. El proceso continua hasta elegir las k semillas. Este método, denominado SCS (Simple Cluster-Seeking), lo propusieron Tou y Gonzales (1974) y presenta el inconveniente de que, además de depender de la ordenación de los elementos en la base de datos, presenta el inconveniente de que el usuario debe definir *a priori* la cota para las distancias.
- Considerar como primera semilla el centroide del grupo formado por todos los elementos del conjunto de datos; seleccionar las sucesivas semillas comparando cada uno de los elementos, en orden de entrada en el conjunto de datos, con las semillas previamente seleccionadas; el elemento será seleccionado como semilla si su distancia a todas las demás es superior a una cantidad d prefijada. El proceso finaliza cuando se obtengan las k semillas o se hayan evaluado todos los elementos del conjunto de datos (Ball y Hall, 1967). Aunque depende del parámetro d , este método es lo suficientemente sencillo como para probar con distintos valores supuesto que en los primeros intentos se obtengan pocas semillas o se evalúen pocos elementos (cuando las semillas procedan de los elementos situados en las primeras posiciones dentro del conjunto de datos).
- Si c es el centroide correspondiente al grupo formado por todos los elementos en la base, dado un vector aleatorio e con valores pequeños para todas sus componentes, el método propuesto por Linde et al. (1980), denominado (BS), considera una división binaria del grupo (Binary Splitting) en términos de $c+e$ y de $c-e$. Los

elementos se clasifican en dos grupos aplicando el *Algoritmo de las k-medias* sobre dichos centroides iniciales. Para cada uno de los dos centroides de la solución obtenida, se procede como en el caso de un único grupo, haciendo una división binaria de cada uno de ellos en términos del vector aleatorio e , y ejecutando el *Algoritmo de las k-medias* para la obtención de cuatro grupos. El proceso de división continúa repitiéndose hasta que se llega a la solución de k centroides. Este método presenta diversos inconvenientes tales como la ejecución del algoritmo tras cada división y la dependencia de la elección del vector e . Además el número de grupos posible es un valor de la forma 2^r , con $r=0,1,\dots$

Una alternativa a elegir semillas dentro del conjunto de datos es realizar una partición del mismo en k grupos y considerar los centros de los grupos como *semillas*. Algunos métodos para la generación de esta partición son:

- Dado un conjunto de puntos semilla, asignar cada elemento al grupo construido en torno al punto semilla más próximo (Forgy, 1965).
- Dado un conjunto de puntos semilla, supongamos que cada uno de ellos consiste inicialmente en un grupo formado por un único miembro; a continuación se asigna, de uno en uno, cada uno de los elementos del conjunto de datos al grupo más próximo, actualizándose el centroide del grupo con cada asignación en términos del vector de medias para los miembros del grupo (MacQueen, 1967). Si se consideran como puntos semilla iniciales los k primeros elementos del conjunto de datos (tal como propone el propio MacQueen), el proceso comenzará con la asignación del elemento situado en la posición $k+1$. En consecuencia, una forma de refinar el método es dar una segunda pasada sobre el conjunto de datos en términos de los centroides obtenidos con la primera pasada.
- Considerar como partición inicial una solución proporcionada por la aplicación de un método jerárquico. Dependiendo del volumen del conjunto de datos puede ser necesaria la extracción de uno o más subconjuntos de un tamaño adecuado para la aplicación del método jerárquico (Lance y Williams, 1967, o Milligan, 1980)
- Se pueden considerar diversos métodos de asignación aleatoria. Por ejemplo, asignar cada elemento a uno de los k grupos mediante la generación de un número aleatorio comprendido entre 1 y k .
- Crear una partición en términos de algún aspecto presente en los datos relativo al problema concreto objeto de análisis

Obsérvese que algunos de estos métodos clásicos necesitan, a su vez, de un conjunto de *semillas*. Una de las grandes dificultades a la hora de enumerar tanto los métodos clásicos como los modernos para la inicialización del *Algoritmo de las k-medias* es que la gran mayoría de ellos no consiste únicamente en una forma de inicialización, sino que es en sí mismo un método de agrupación que, a su vez, presenta el mismo problema que

el *Algoritmo de las k-medias*: necesita unos puntos o centroides iniciales. Posiblemente ésta es la razón por la que, al revisar la literatura al respecto⁵, los métodos enumerados por los distintos autores no siempre coinciden, especialmente en lo que se refiere a los más recientes. En lo que sigue haremos una breve descripción de aquellos que con mayor frecuencia aparecen referenciados en la literatura.

Kauffman y Rousseeuw (1990): Considerando que el *Algoritmo de las k-medias* intenta encontrar k centroides tales que minimicen la *distorsión* (la suma de los cuadrados de las distancias de cada punto al centroide más próximo), Kauffman y Rousseeuw (1990) eligen como primera semilla el elemento más próximo al centroide del grupo formado por toda la base de datos; la siguiente semilla coincidirá con aquel elemento que proporcione la mayor reducción de la *distorsión*. La tercera y sucesivas semillas se seleccionan con el mismo criterio hasta obtener K .

Babu y Murty (1993): Utilizan algoritmos genéticos (Genetic Algorithms) para la elección de las semillas. La población consiste en varias selecciones para las semillas. La idoneidad de cada selección de semillas se valora ejecutando el *Algoritmo de las k-medias* y calculando el valor de la *distorsión*. La solución ajustada se reproducirá para crear una segunda generación de soluciones. Este proceso se repite hasta que se generen un número predeterminado de generaciones. Aunque con este método es posible alcanzar la solución óptima en numerosas ocasiones, presenta el inconveniente de que el *Algoritmo de las k-medias* debe ser ejecutado por cada solución de cada generación. Un problema adicional que presentan los algoritmos genéticos es que los resultados pueden variar significativamente dependiendo de la elección del tamaño de la población y de las probabilidades de cruce y de mutación (Jain et al., 1999).

Huang y Harris (1993): Proponen una variación del método Binary Splitting (Linde et al., 1980) en la que la división se realiza mediante el Análisis de componentes principales. El método, denominado Direct Search Binary Splitting (DSBS), permite elegir el vector e de forma determinística con la finalidad de mejorar la calidad de la solución de agrupación.

Katsavounidis, Kuo, y Zhen (1994): Proponen un método (al que se denomina KKZ) que proporciona mejores resultados que la técnica BS. El algoritmo comienza con la selección de un elemento de la base de datos situado en una posición extrema como primera semilla. La segunda semilla será el punto más lejano de la primera. A continuación, se calculan las mínimas distancias entre cada uno de los restantes elementos y cada semilla; la siguiente semilla será el elemento que presente el máximo valor de los mínimos anteriores. La elección de este elemento se repite hasta alcanzar las K semillas. Este algoritmo presenta el inconveniente de ser muy sensible a las observaciones aisladas.

Daoud y Roberts (1996): Proponen un primer método en el que dividen todo el dominio de entrada en M celdas del mismo tamaño y de cada una de ellas seleccionan

⁵ Daoud y Roberts (1996), He et al. (2004), Peña et al. (1999), Redmon y Henegan (2007) y Steinley y Brusco (2007).

aleatoriamente un número de semillas proporcional al número de elementos en la celda. El inconveniente radica en que si K es pequeño, M también debe serlo para que las celdas proporcionen al menos una semilla. En el mismo artículo proponen un segundo método como mejora del primero. La mejora consiste en dividir, de forma iterativa, el dominio de entrada en dos volúmenes distintos y calcular para cada uno ellos la suma de los cuadrados de los errores entre los elementos y los centroides más próximos en cada volumen. El número de grupos a considerar en cada volumen que minimice el error viene dado por una solución numérica de una ecuación polinómica, que corresponde a una ecuación cuadrática en el caso bidimensional. En cualquier caso, no demuestran la efectividad del método (He et al., 2004).

Bradley y Fayyad (1998): Proponen un procedimiento tipo *bootstrap* para determinar las semillas iniciales. En primer lugar se eligen 10 conjuntos aleatorios de la base de datos de entrada (aunque puede ser un número diferente), cada uno de ellos formado por la décima parte de los elementos. A continuación se ejecuta el *Algoritmo de las k-medias* sobre cada conjunto, considerando el mismo conjunto de semillas iniciales (elegidas mediante la metodología propuesta por Forgy, 1965), lo que da como resultado 10 conjuntos de K de puntos centrales. El *Algoritmo de las k-medias* se ejecuta 10 veces sobre toda la base de datos, considerando en cada ocasión como semillas iniciales cada uno de los 10 conjuntos de K puntos centrales. Los K puntos centrales resultantes de la décima ejecución se utilizarán como semillas iniciales del *Algoritmo de las k-medias* para toda la base de datos. Bradley y Fayyad (1998) argumentan que con este refinado de las semillas iniciales se elimina el efecto de valores extremos en los datos.

Likas, Vlassis, y Verbeek (2003): Proponen el denominado *Algoritmo de las k-medias* global, que construye los centroides iniciales haciendo una partición recursiva del espacio de representación de la base de datos en subespacios disjuntos utilizando una metodología de árbol- kd . El hiperplano de corte utilizado se define como el plano perpendicular al eje de máxima varianza extraído por componentes principales. La partición recursiva finaliza cuando cada nodo-hoja contenga menos elementos que un número predefinido o se hayan creado tantos nodos hojas como un número predefinido. Los centroides iniciales para el *Algoritmo de las k-medias* serán los centroides de los elementos en los nodos-hoja finales.

Khan y Ahmad (2004): Proponen un algoritmo, al que denominan CCIA (**C**luster **C**enter **I**nitialization **A**lgorithm), que se basa en considerar valores concretos de cada atributo por separado. Aunque el algoritmo puede ser de gran utilidad en determinadas ocasiones, es bastante complejo; además para su aplicación necesita de la ejecución del *Algoritmo de las k-medias* que, a su vez, necesita de algún método de aleatorización para su inicialización (Deelers y Auwatanamongkol, 2007)

2.2 ANÁLISIS CLUSTER DE SERIES TEMPORALES

2.2.1 Minería de datos temporales

La *Minería de Datos Temporales* se refiere al análisis de grandes conjuntos de datos secuenciales, entendiéndose éstos como datos ordenados en relación a algún índice⁶. Las series temporales constituyen una clase concreta de dato secuencial, ya que los registros dentro de la base de datos están indexados u ordenados según el parámetro tiempo.

Una serie temporal es una secuencia de números reales que representan medidas de una variable real en intervalos iguales de tiempo. Si la variable está definida sobre un conjunto numerable de puntos temporales la serie temporal será discreta. La mayor parte de las aplicaciones del *Análisis Cluster de Series Temporales* se refiere a variables discretas.

El *Análisis Clásico de Series Temporales* tiene una larga historia. Las técnicas para la construcción de modelos estadísticos sobre series temporales reales o complejas se utilizan desde hace más de cincuenta años (Box et al., 1994; Chatfield, 1996). El origen de la *Minería de Datos Temporales* es mucho más reciente y se distingue del *Análisis Clásico* en dos aspectos fundamentales.

Una primera diferencia entre ambos enfoques radica en el tamaño y la naturaleza de la base de datos y en la forma en que se recogen los datos. Por un lado, los métodos de minería de datos temporales suelen ser capaces de analizar conjuntos de datos de tamaños prohibitivos para técnicas convencionales para el ajuste de modelos a series temporales; por otro, la secuencias de valores pueden ser de tipo nominal o simbólico, haciendo inaplicables técnicas de ajuste de modelos tales como los ARMA o los ARIMA; y además, a diferencia de la mayoría de las aplicaciones de los métodos estadísticos, en minería de datos se tiene poco o ningún control sobre el proceso de recopilación de los datos, que frecuentemente responde a otros propósitos totalmente diferentes.

El segundo aspecto en el que difieren la *Minería de Datos Temporales* y el *Análisis Clásico de Series Temporales* radica en la clase de información que se desea extraer de los datos. El ámbito de la minería de datos va más allá de las aplicaciones estándar de predicción o control del análisis de series temporales: el conocimiento exacto de los parámetros de un modelo (por ejemplo, los coeficientes de un modelo ARMA o los pesos de una red neuronal) suele ser de poca relevancia, siendo de mayor interés el descubrimiento de tendencias o patrones en los datos, que faciliten la interpretación de las relaciones entre ellos.

La *Minería de Datos Temporales* reúne diversos tipos de tareas aplicables a las series temporales. Laxman y Sastry (2006) establecen la siguiente distinción entre tipos

⁶ Una revisión de los avances existentes en el análisis de datos con interdependencias temporales puede verse en los trabajos de Zhao y Bhowmick (2003) y Laxman y Sastry (2006).

de tarea: predicción, clasificación, agrupación (o *clustering*), búsqueda y recuperación, y descubrimiento de patrones.

Predicción: La predicción en series temporales se refiere al pronóstico de valores futuros de la serie basados en las observaciones pasadas. Para ello es necesario construir un modelo de predicción sobre los datos.

Clasificación: En la clasificación de secuencias, se supone que cada una de ellas pertenece a una de un conjunto finito de varias clases, en un número predefinido, siendo el objetivo el de la determinación automática de la categoría o clase correspondiente para la secuencia dada.

Agrupación (*clustering*): La agrupación de secuencias o series temporales se realiza a partir de alguna medida de similitud. La agrupación de series temporales es especialmente interesante en *Minería de Datos Temporales* ya que ofrece un mecanismo para encontrar de forma automática algún tipo de estructura en conjuntos grandes de datos que, por otra vía, sería un problema difícilmente abordable.

Búsqueda y recuperación: Otra de las tareas relevantes en *Minería de Datos Temporales* es la búsqueda de secuencias en base de datos grandes. Dada una secuencia de datos se trata de, dentro de una base de datos grande formada por datos secuenciales, encontrar y recuperar a partir de ella, secuencias o subsecuencias *muy similares* a la secuencia dada.

Descubrimiento de patrones: A diferencia de las aplicaciones de búsqueda y recuperación de secuencias, en el descubrimiento de patrones no existe una secuencia de referencia sino que se trata de detectar pautas frecuentes de comportamiento en las secuencias de datos. Estas pautas serán de utilidad a la hora de obtener reglas de asociación referidas a las series temporales objeto de estudio.

2.2.2 Similitud entre series temporales numéricas

Cuando la base de datos es una tabla de datos de dimensión $N \times P$ correspondiente a N observaciones sobre un conjunto de P variables, el objetivo del *Análisis Cluster* suele ser el de agrupar las observaciones (las filas de la tabla), pero también puede aplicarse para agrupar variables (las columnas). Si, en particular, las variables son series temporales observadas en los mismos instantes temporales (en las filas), se tratará de un *Análisis Cluster de Series Temporales*. Aunque la más sencilla, ésta no es la única situación de partida para un *Análisis Cluster de Series Temporales*, pero sí es la única en la que, a efectos de la metodología que se desarrolla en este trabajo, se centra nuestro interés.

Dados un conjunto de J series temporales $\{X_j\}$, $j=1, \dots, J$, observadas en los mismos T instantes temporales todas ellas:

$$X_j = (x_{1,j}, \dots, x_{t,j}, \dots, x_{T,j})' \quad j = 1, \dots, J$$

el problema consiste en formar K grupos o subconjuntos tales que las series más similares entre sí pertenezcan a un mismo grupo, donde K puede ser o no un valor conocido.

Para buscar series temporales parecidas es necesario establecer una medida de similitud entre ellas. Concretamente, la similitud entre dos series cualesquiera se mide cuantitativamente mediante una función apropiada a las condiciones específicas del problema objeto de análisis, que debe satisfacer las condiciones propias de una distancia.

Una función de distancia definida sobre un conjunto de series temporales es una función definida sobre cada par de series simétrica, no negativa y tal que la distancia entre una serie y ella misma es nula. Una función de distancia es métrica si verifica las siguientes cuatro condiciones:

$$\begin{aligned} d(X_r, X_s) &> 0 \quad \text{si} \quad X_r \neq X_s \\ d(X_j, X_j) &= 0 \quad \forall X_j \\ d(X_r, X_s) &= d(X_s, X_r) \quad \forall X_r, X_s \\ d(X_r, X_s) &\leq d(X_r, X_q) + d(X_q, X_s) \quad \forall X_q, X_r, X_s, \quad (\text{desigualdad triangular}) \end{aligned}$$

Dada una distancia métrica, el objetivo del *Análisis Cluster de Series Temporales* es encontrar una agrupación que minimice las distancias entre las series en un mismo grupo y maximice las distancias entre series en distintos grupos. Para ello se considerarán los algoritmos generales de *Análisis Cluster* y, en particular, el *Algoritmo de las k-medias* que, recordemos, consistía en un método iterativo de búsqueda local que trataba de minimizar las distancias dentro de los grupos. Una forma de proceder consistiría en tratar cada serie temporal como un punto en un espacio T dimensional, definir una distancia para cada par de puntos y aplicar el algoritmo de agrupación sobre la nube de los J puntos. Sin embargo, aunque la más directa, esta no es la única forma de abordar el problema. Alternativamente, cada serie puede ser resumida en un conjunto de parámetros, correspondiente a características de la serie o al ajuste de algún modelo, de tal forma que se reduzca la dimensión del espacio de representación de la nube de puntos. Así, el algoritmo de agrupación se ejecutaría sobre información extraída de la serie en lugar de sobre la propia serie observada.

Liao (2005a) clasifica los *Métodos Cluster de Series Temporales* en tres categorías diferentes, dependiendo de cuál sea la unidad sobre la que se aplique el algoritmo de agrupación: métodos basados en la forma de las propias series temporales (“raw-data-based approach”), métodos basados en características de las series (“feature-based approach”) y métodos basados en la estimación de parámetros de modelos ajustados sobre las series temporales (“model-based approach”). La medida de similitud

adecuada para comparar las series temporales dependerá del método de agrupación elegido.

Métodos basados en la forma

Se aplican directamente sobre las series temporales observadas (o sobre una transformación de las mismas, si fuera necesaria). Dadas dos series temporales X_r y X_s , con T observaciones cada una de ellas en los mismos instantes temporales, la alternativa más sencilla para medir la similitud entre ellas es considerar las denominadas distancias de Minkowski:

$$L_q(X_r, X_s) = \left[\sum_{t=1}^T |x_{tr} - x_{ts}|^q \right]^{1/q}$$

y, en particular, la distancia euclídea:

$$L_2(X_r, X_s) = \left[\sum_{t=1}^T |x_{tr} - x_{ts}|^2 \right]^{1/2}$$

Las distancias L_q miden la cercanía de cada par de valores observados, siendo necesario que ambas series estén definidas sobre los mismos instantes temporales y en la misma escala de medida. Una forma de solventar este inconveniente es transformar los valores de las distintas series temporales mediante funciones lineales o normalización, aunque no siempre proporcione buenos resultados⁷.

A diferencia de la amplitud, el eje del tiempo no suele ser un problema, dado que sus valores no suelen ser contemplados en el cálculo de las distancias más allá de lo que se refiere al orden, salvo que las observaciones de las series no puedan ser representadas sobre un eje temporal común. El alineamiento dinámico temporal (Dynamic Time Warping, Berndt y Clifford (1996)) tiene como objetivo encontrar un eje temporal común que proporcione la mejor alineación de las series en términos de su forma.

Otra alternativa a la distancia euclídea es considerar una métrica que capture la similitud en forma, por ejemplo, basada en la correlación entre las series. Si denominamos $cc(X_r, X_s)$ al coeficiente de correlación de Pearson entre la series X_r y X_s , es posible definir una métrica sin más que realizar alguna manipulación algebraica sobre su expresión. Concretamente, Goley et al. (1998) consideran dos distancias basadas en el coeficiente de correlación:

$$d(X_r, X_s) = \sqrt{2(1 - cc(X_r, X_s))} \quad \text{y} \quad d(X_r, X_s) = \left(\frac{1 - cc(X_r, X_s)}{1 + cc(X_r, X_s)} \right)$$

⁷ Goldin y Kanellakis (1995), Agrawal et al. (1995), Keogh y Kasetty (2002).

Métodos basados en características

Las distancias enumeradas en el apartado anterior se calculan directamente sobre el espacio de representación de la propia serie temporal. Si la dimensión del espacio es grande, el cálculo de estas distancias puede resultar computacionalmente muy costoso. Una solución consiste en resumir las series en vectores de características y realizar la agrupación en el espacio de proyección de estos vectores.

La primera propuesta de proyección de las series en un nuevo espacio de menor dimensión se debe a Agrawal et al. (1993), que considera el espacio de la transformada rápida de Fourier. Esta proyección se basa en el Teorema de Parseval que afirma que la transformada de Fourier conserva las distancias. A partir de este primer trabajo se han propuesto otras transformaciones entre las que se incluyen las transformaciones *Wavelet* (Huhtala et al., 1999) y *Piecewise Constant Approximation* (Keogh y Pesan, 2000).

Otra alternativa para cuantificar la similitud entre series temporales en un espacio de menor dimensión es, supuesto que las series son estacionarias, proyectarlas en el espacio de sus funciones de probabilidad. Este tipo de proyección permite comparar series con diferentes longitudes. Kakizawa et al. (1998) utilizan la divergencia de Kullback-Leibler y la medida de información de Chernoff.

Aunque la mayoría de estos métodos se suponen de naturaleza genérica, lo cierto es que en la práctica las características consideradas suelen depender de la aplicación concreta. Es decir, un conjunto de características que funciona bien para una determinada aplicación, puede no ser relevante para otra. Si sólo algunas de las propiedades de las series son importantes, la elección de las características adecuadas proporcionará un buen resultado.

Métodos basados en modelos

Con este tipo de aproximación se supone conocido el proceso que ha generado cada serie temporal y la agrupación se realiza bajo este supuesto. Así, dos series temporales se consideran similares cuando los parámetros de los modelos que caracterizan cada serie o los residuos resultantes del ajuste del modelo son parecidos. Los métodos basados en modelos se clasifican en tres grandes categorías: Modelos autorregresivos de medias móviles (ARMA), Modelos ocultos de Markov (HMM) y Modelos de mixturas de polinomios.

Modelos autorregresivos de medias móviles (ARMA): La principal aproximación al problema de la agrupación de series temporales es suponer que los modelos subyacentes han sido generados por un proceso ARMA (Box et al., 1994). El procedimiento de agrupación suele constar de tres etapas: ajustar el modelo a cada una de las series, medir la distancia entre los modelos ajustados y agrupar las series en función de las distancias obtenidas. Este tipo de aproximación es la adoptada por Piccolo (1990), Maharaj (1996, 2000), Baragona (2001) y Tong y Dabas (1990).

Piccolo (1990) utiliza la distancia euclídea entre los vectores de m parámetros de modelos ARIMA de orden m ajustados sobre las distintas series, mientras que Maharaj (1996, 2000) utiliza los parámetros para contrastar las hipótesis de si dos series estacionarias son o no realizaciones de un mismo proceso estocástico. La agrupación se realiza sobre los p -valores asociados a los estadísticos de contraste.

Baragona (2001) y Tong y Dabas (1990) comparan las series temporales en términos de las distancias entre los residuos de los modelos ajustados. Tong y Dabas (1990) consideran el coeficiente de correlación entre los residuos sin desfase, mientras que Baragona (2001) utiliza una función de distancia que reescala la correlación sin desfase mediante la suma de las correlaciones desfasadas.

Modelos ocultos de Markov (HMM): Como alternativa a los modelos ARIMA, algunos autores asumen que, dentro de cada grupo, las series tiene un mismo proceso generador que puede ser descrito mediante una Cadena oculta de Markow (HMM). La diferencia fundamental entre los métodos ARMA y los HMM radica en que con los primeros la agrupación se realiza después de ajustar los modelos, mientras que la mayoría de las aplicaciones con métodos HMM implica la formación de grupos de modelos en cada iteración del algoritmo de agrupación⁸.

Modelos de mixturas de polinomios: Para una buena estimación de los parámetros, en los modelos anteriores es necesario que las series temporales tengan mucha longitud. Para el caso de series temporales cortas y con longitudes variables Gaffney y Smith (1999, 2003) asumen que el modelo subyacente es una mixtura de funciones polinómicas. Otra aplicación de este tipo de metodología puede verse en Bar-Joseph et al. (2002).

2.3 AGRUPACIÓN DE SERIES TEMPORALES EN EL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL

El problema de agrupar series temporales ha surgido en campos tan diversos como la economía⁹, la ingeniería¹⁰, las ciencias¹¹, la medicina¹², el medioambiente¹³, o las artes¹⁴. En el caso del sector de la construcción residencial, el único trabajo que

⁸ Ejemplos de aplicaciones con este tipo de modelos pueden verse en Ramoni et al. (2002), Cadez et al. (2000), Ridgeway (1997), Oates (1999), Oates, Firoui y Smyth (2000), Oates, Schmill y Cohen (2000), Smyth (1997), Zhong (2002), Zhong y Ghosh (2002a, 2002b, 2002c, 2003), Li (2000), Li y Biswas (1999a, 1999b, 1999c, 2000) y Alon et al. (2003).

⁹ Fu et al. (2001), Kalpakis et al. (2001), Kosmelj y Batagelj (1990), Maharaj (2000), Möller-Levet et al. (2003) y van Wijk y van Selow (1999).

¹⁰ Goutte et al. (1999), Owsley et al. (1997), Ramoni et al. (2000), Tran y Wagner (2002), Vlachos et al. (2003), Wang et al. (2002) y Wilpon y Rabiner (1985).

¹¹ Kalpakis et al. (2001), Li et al. (2001), Möller-Levet et al. (2003) y Shumway (2003).

¹² Golay et al. (1998), Goutte et al. (1999), Goutte et al. (2001) y Kalpakis et al. (2001.)

¹³ Bohte et al. (1980), Cowpertwait y Cox (1992), Gantert (1994), Macchiato et al. (1995) y Walden (1994).

¹⁴ Beran y Mazzola (1999).

conocemos se debe a James H. Stock y Mark W. Watson y se titula “The Evolution of National and Regional Factors in U.S. Housing Construction”¹⁵ (Stock y Watson, 2008).

Stock y Watson analizan datos trimestrales de visados de la construcción para vivienda nueva en Estados Unidos en el periodo 1969-2007, con el objetivo de determinar regiones geográficas tales que los distintos estados agrupados dentro de una misma región presenten una dinámica parecida en lo que a evolución de la construcción residencial se refiere. Su interés se centra en comparar las trayectorias correspondientes a los distintos estados americanos. Para ello, consideran un modelo factorial dinámico (Geweke, 1977) con volatilidad estocástica (DFM-SV) en el que suponen que la evolución de la serie de visados en cada estado es función de una componente nacional, de varias componentes regionales y de una componente específica. En consecuencia, para determinar las componentes regionales, la especificación del modelo implica la determinación previa de las regiones.

Formalmente, ciñéndose al periodo 1970-2007, postulan un modelo factorial dinámico de expresión:

$$y_{it} = \alpha_i + \lambda_i F_t + \sum_{j=1}^{N_R} \gamma_{ij} R_{jt} + e_{it}$$

donde y_{it} es el logaritmo de los visados de vivienda en el estado i en el instante t . Se supone que el factor nacional F_t y los N_R factores regionales R_{jt} son paseos aleatorios y que el error idiosincrásico e_{ij} sigue un proceso AR(1):

$$\begin{aligned} F_t &= F_{t-1} + \eta_t \\ R_{jt} &= R_{jt-1} + \nu_{jt} \\ e_{it} &= \rho_i e_{it-1} + \varepsilon_{it} \end{aligned}$$

Además γ_{ij} únicamente puede tomar valor distinto de cero cuando el estado i pertenezca a la región j . Para proceder al ajuste del modelo estiman previamente las regiones. En decir, en un primer paso, postulan un modelo de un único factor:

$$\begin{aligned} y_{it} &= \alpha_i + \lambda_i F_t + u_{it} \\ F_t &= F_{t-1} + \eta_t \\ u_{it} &= \rho_{i1} u_{it-1} + \rho_{i2} u_{it-2} + \varepsilon_{it} \end{aligned}$$

y consideran los residuos:

$$\hat{u}_{it} = y_{it} - \hat{\alpha}_i + \hat{\lambda}_i \hat{F}_t$$

¹⁵ Se trata de un trabajo preparado para el volumen “Essays in Volatility in Finance and Economics, Time Series and Regional Economics: A Festschrift in Honour of Robert F. Engle”.

A continuación, para agrupar los estados en regiones, aplican el *Algoritmo de las k-medias* sobre las series residuales en tasa de incremento anual y considerando la distancia euclídea como medida de similitud. Según indican literalmente:

This algorithm was repeated for multiple random starting values.

We undertook an initial cluster analysis to estimate the number of regions, in which the foregoing algorithm was used with 20,000 random starting values. Moving from 2 to 3 clusters reduced the value of the minimized objective function by approximately 10%, as did moving from 3 to 4 clusters. The improvements from 4 to 5, and from 5 to 6, were less, and for 6 clusters the number of states was as few as five in one of the clusters. Absent a statistical theory for estimating the number of clusters, and lacking a persuasive reason for choosing 6 clusters, we therefore chose $k = 5$.

We then estimated the composition of these 5 regions using 400,000 random starting values. We found that even after 200,000 starting values there were some improvements in the objective function, however those improvements were very small and the switches of states in regions involved were few. We then reestimated the regions for the 1970-1987 and 1988-2007 subsamples, using 200,000 additional random starting values and using the full-sample regional estimates as an additional starting value. (Stock y Watson, 2008: 13)

En definitiva, las semillas iniciales se obtienen a partir de la asignación aleatoria de cada estado en uno de los K grupos. El algoritmo se repite con múltiples semillas iniciales aleatorias comparando distintas soluciones para K , hasta quedarse con la solución de 5 grupos. De nuevo el algoritmo se repite con múltiples semillas iniciales comparando distintas soluciones todas ellas con 5 grupos.

Con el objetivo de establecer la relación entre la volatilidad de la construcción de vivienda en cada estado, medida a través de los visados de edificación, y la “Gran Moderación” de mediados de los ochenta para el conjunto de la actividad económica en Estados, este proceso de regionalización se repite en cada uno de los subperiodos: 1970-1987 y 1988-2007.

CAPÍTULO 3

ELEMENTOS TEÓRICOS DE LA METODOLOGÍA DEL HAZ DE RECTAS

Dado un conjunto grande de series temporales (**Fig. 3.1, Izqda**), se trata de obtener un conjunto pequeño de series resumen (**Fig. 3.1, Dcha**), que nos permita, además de agrupar las series temporales en grupos homogéneos, interpretar y comparar sus trayectorias. Para ello se aplicará un *Análisis Cluster* utilizando como criterio de agrupación el *Algoritmo de las k-medias*.

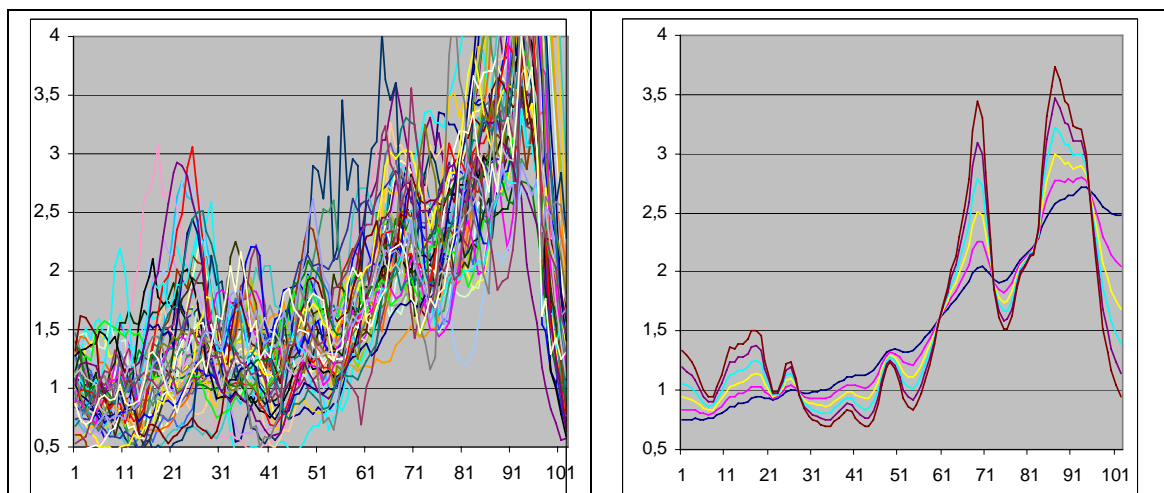


Figura 3.1 *Izqda.*: Series temporales objeto de agrupación; *Dcha.*: Series resumen

La metodología consistirá básicamente en construir un conjunto de K series resumen, que se utilizarán como centroides iniciales del algoritmo. Dicho conjunto se caracteriza porque el correspondiente conjunto de series diferenciadas define un haz de

rectas secantes, razón por la que hemos denominado Metodología del haz de rectas al proceso de agrupación de series temporales propuesto en esta tesis. En este capítulo se exponen los elementos teóricos sobre los que se sustenta la metodología, siendo el punto de partida el siguiente resultado:

Lema 1:

Sea $\{C_{t,k}\}$, $k=1,\dots,K$, un conjunto de K series temporales distintas, todas ellas observadas en los mismos instantes temporales, $t=1,\dots,T$, y con la misma media:

$$\frac{\sum_{t=1}^T C_{t,k}}{T} = \bar{C}_k = \alpha \quad \forall k \quad [3.1]$$

Sea X_t otra serie temporal y supongamos que para cada serie $C_{t,k}$ existen tres coeficientes b_k^x , m_k^x y μ_k^x , con al menos m_k^x distinto de cero, tales que:

$$C_{t,k} = b_k^x \cdot t + m_k^x \cdot X_t + \mu_k^x \quad \forall t, k \quad [3.2]$$

siendo:

$$b_k^x = B_0^x + B_1^x \cdot m_k^x \quad \forall k \quad [3.3]$$

Si $C_{t,q}$, $C_{t,r}$ y $C_{t,s}$, son tres series cualesquiera del conjunto tales que:

$$m_q^x < m_r^x < m_s^x$$

entonces se verifica que:

$$d(C_{t,q}, C_{t,r}) < d(C_{t,q}, C_{t,s})$$

siendo d la distancia euclídea.

Demostración:

Si denominamos:

$$\bar{t} = \frac{\sum_{t=1}^T t}{T} \quad \text{y} \quad \bar{X} = \frac{\sum_{t=1}^T X_t}{T}$$

las expresiones [3.2] y [3.3] implican que:

$$\begin{aligned} \frac{\sum_{t=1}^T C_{t,k}}{T} &= \frac{1}{T} \sum_{t=1}^T (b_k^x \cdot t + m_k^x \cdot X_t + \mu_k^x) = b_k^x \cdot \bar{t} + m_k^x \cdot \bar{X} + \mu_k^x = \\ &= (B_0^x + B_1^x \cdot m_k^x) \cdot \bar{t} + m_k^x \cdot \bar{X} + \mu_k^x = B_0^x \cdot \bar{t} + (B_1^x \cdot \bar{t} + \bar{X}) \cdot m_k^x + \mu_k^x \end{aligned}$$

y, por [3.1]:

$$\alpha = B_0^x \cdot \bar{t} + (B_1^x \cdot \bar{t} + \bar{X}) \cdot m_k^x + \mu_k^x \quad \forall k$$

o, lo que es equivalente:

$$\mu_k^x = (\alpha - B_0^x \cdot \bar{t}) + (-B_1^x \cdot \bar{t} - \bar{X}) \cdot m_k^x \quad \forall k$$

En otras palabras, si denominamos:

$$D_0^x = \alpha - B_0^x \cdot \bar{t} \quad \text{y} \quad D_1^x = -B_1^x \cdot \bar{t} - \bar{X}$$

entonces existen dos coeficientes tales que:

$$\mu_k^x = D_0^x + D_1^x \cdot m_k^x \quad [3.4]$$

y la ecuación [3.2] se puede expresar alternativamente como:

$$\begin{aligned} C_{t,k} &= (B_0^x + B_1^x \cdot m_k^x) \cdot t + m_k^x \cdot X_t + D_0^x + D_1^x \cdot m_k^x \\ &= D_0^x + B_0^x \cdot t + (B_1^x \cdot t + X_t + D_1^x) \cdot m_k^x \end{aligned} \quad [3.5]$$

En consecuencia, si $m_k^x < m_{k'}^x$:

$$\begin{aligned} d(C_{t,k}, C_{t,k'}) &= \left(\sum_{t=1}^T (D_0^x + B_0^x \cdot t + (B_1^x \cdot t + X_t + D_1^x) \cdot m_{k'}^x - (D_0^x + B_0^x \cdot t + (B_1^x \cdot t + X_t + D_1^x) \cdot m_k^x))^2 \right)^{1/2} = \\ &= \left(\sum_{t=1}^T (B_1^x \cdot t + X_t + D_1^x)^2 \cdot (m_{k'}^x - m_k^x)^2 \right)^{1/2} = (m_{k'}^x - m_k^x) \cdot \left(\sum_{t=1}^T (B_1^x \cdot t + X_t + D_1^x)^2 \right)^{1/2} = (m_{k'}^x - m_k^x) \cdot \varepsilon \end{aligned}$$

donde:

$$\varepsilon = \left(\sum_{t=1}^T (B_1^x \cdot t + X_t + D_1^x)^2 \right)^{1/2}$$

es un valor constante no negativo e independiente de t y de k . Luego, si:

$$m_q^x < m_r^x < m_s^x$$

entonces:

$$0 < m_r^x - m_q^x < m_s^x - m_q^x$$

y, en consecuencia:

$$d(C_{t,q}, C_{t,r}) = (m_r^x - m_q^x) \cdot \varepsilon < (m_s^x - m_q^x) \cdot \varepsilon = d(C_{t,q}, C_{t,s})$$

En otras palabras, bajo las hipótesis [3.1], [3.2] y [3.3], las series temporales $C_{t,k}$ pueden ser ordenadas en términos de su similitud. Es decir, dada la secuencia de coeficientes m_k^x ordenados de menor a mayor valor¹:

$$m_{(1)}^x = \min_k m_k^x < m_{(2)}^x < \dots < m_{(K)}^x = \max_k m_k^x$$

si $C_{t,(k)}$ es la serie asociada a $m_{(k)}^x$, $k=1, \dots, K$, entonces las K series temporales $C_{t,k}$ pueden ser ordenadas por su similitud según la secuencia:

$$C_{t,(1)}, C_{t,(2)}, \dots, C_{t,(K)}$$

en el sentido de que, dada cualquier terna de series $C_{t,(q)}$, $C_{t,(r)}$ y $C_{t,(s)}$, si $q < r < s$ entonces las dos series $C_{t,(q)}$ y $C_{t,(s)}$ son las más distantes de las tres.

Proposición 1:

Bajo las hipótesis [3.1], [3.2] y [3.3], si C_t es la serie temporal promedio:

$$C_t = \frac{\sum_{k=1}^K C_{t,k}}{K} \tag{3.6}$$

entonces existen tres coeficientes b^x , m^x y μ^x tales que:

$$C_t = b^x \cdot t + m^x \cdot X_t + \mu^x \tag{3.7}$$

con:

$$b^x = B_0^x + B_1^x \cdot m^x \tag{3.8}$$

¹ Obsérvese que, dadas dos series k y k' , tiene que verificarse $m_k^x \neq m_{k'}^x$. Si fuera $m_k^x = m_{k'}^x$, entonces por [3.3] $b_k^x = b_{k'}^x$, y por [3.4] $\mu_k^x = \mu_{k'}^x$, y en consecuencia, $C_{t,k}$ y $C_{t,k'}$ serían la misma serie, en contradicción con las hipótesis de partida.

Demostración:

Por [3.6] y [3.2]:

$$C_t = \frac{1}{K} \sum_{k=1}^K (b_k^x \cdot t + m_k^x \cdot X_t + \mu_k^x) = \frac{\sum_{k=1}^K b_k^x}{K} \cdot t + \frac{\sum_{k=1}^K m_k^x}{K} \cdot X_t + \frac{\sum_{k=1}^K \mu_k^x}{K}$$

luego, para que se verifique [3.7], basta con considerar:

$$b^x = \bar{b}_k^x = \frac{\sum_{k=1}^K b_k^x}{K}, \quad m^x = \bar{m}_k^x = \frac{\sum_{k=1}^K m_k^x}{K} \quad \text{y} \quad \mu^x = \bar{\mu}_k^x = \frac{\sum_{k=1}^K \mu_k^x}{K}$$

Además, por [3.3]:

$$b^x = \frac{\sum_{k=1}^K (B_0^x + B_1^x \cdot m_k^x)}{K} = B_0^x + B_1^x \cdot \frac{\sum_{k=1}^K m_k^x}{K} = B_0^x + B_1^x \cdot m^x$$

y por [3.4]:

$$\mu^x = \frac{\sum_{k=1}^K (D_0^x + D_1^x \cdot m_k^x)}{K} = D_0^x + D_1^x \cdot \frac{\sum_{k=1}^K m_k^x}{K} = D_0^x + D_1^x \cdot m^x \quad [3.9]$$

Proposición 2:

Bajo las hipótesis [3.1], [3.2] y [3.3] existen tres coeficientes b_k^c , m_k^c y μ_k^c :

$$C_{t,k} = b_k^c \cdot t + m_k^c \cdot C_t + \mu_k^c \quad [3.10]$$

donde los coeficientes b_k^c y m_k^c son tales que:

$$b_k^c = B_0^c + B_1^c \cdot m_k^c \quad [3.11]$$

Además:

$$B_0^c = -B_1^c = B_1^x$$

Demostración:

Dado que, por la *Proposición 1*:

$$C_t = b^x \cdot t + m^x \cdot X_t + \mu^x$$

o, lo que es equivalente:

$$X_t = \frac{-b^x}{m^x} \cdot t + \frac{1}{m^x} \cdot C_t + \frac{-\mu^x}{m^x}$$

entonces:

$$\begin{aligned} C_{t,k} &= b_k^x \cdot t + m_k^x \cdot X_t + \mu_k^x = b_k^x \cdot t + m_k^x \cdot \left(\frac{-b^x}{m^x} + \frac{1}{m^x} \cdot C_t + \frac{-\mu^x}{m^x} \right) + \mu_k^x = \\ &= \left(b_k^x - b^x \cdot \frac{m_k^x}{m^x} \right) + \frac{m_k^x}{m^x} \cdot H_t + \left(\mu_k^x - \mu^x \cdot \frac{m_k^x}{m^x} \right) = b_k^c \cdot t + m_k^c \cdot C_t + \mu_k^c \end{aligned}$$

donde:

$$m_k^c = \frac{m_k^x}{m^x} \quad \text{y} \quad b_k^c = b_k^x - b^x \cdot m_k^c$$

Además, por [3.3]:

$$b_k^c = B_0^x + B_1^x \cdot m_k^x - b^x \cdot m_k^c = B_0^x + B_1^x \cdot m^x \cdot \frac{m_k^x}{m^x} - b^x \cdot m_k^c = B_0^x + (B_1^x \cdot m^x - b^x) \cdot m_k^c$$

y, por la *Proposición 1*:

$$b^x = B_0^x + B_1^x \cdot m^x \quad \text{o, lo que es equivalente} \quad B_0^x = B_1^x \cdot m^x - b^x$$

entonces:

$$b_k^c = B_0^x - B_0^x \cdot m_k^c = B_0^c + B_1^c \cdot m_k^c$$

donde:

$$B_0^c = B_0^x \quad \text{y} \quad B_1^c = -B_0^x \quad \text{[3.12]}$$

Proposición 3:

Bajo las hipótesis [3.1], [3.2] y [3.3], si $C_{t,q}$, $C_{t,r}$ y $C_{t,s}$, son tres series cualesquiera del conjunto tales que:

$$m_q^c < m_r^c < m_s^c$$

entonces se verifica que:

$$d(C_{t,q}, C_{t,r}) < d(C_{t,q}, C_{t,s})$$

donde d es la distancia euclídea.

Demostración:

Por la **Proposición 2**:

$$m_k^c = \frac{m_k^x}{m^x} \quad \text{o, lo que es equivalente,} \quad m_{(k)}^c = \frac{m_{(k)}^x}{m^x}$$

En consecuencia, la secuencia de coeficientes m_k^c ordenados de menor a mayor valor:

$$m_{(1)}^c, m_{(2)}^c, \dots, m_{(K)}^c$$

da lugar a la misma ordenación de las series temporales $C_{t,k}$:

$$C_{t,(1)}, C_{t,(2)}, \dots, C_{t,(K)}$$

que la proporcionada por la secuencia de coeficientes m_k^x ordenados. En otras palabras, el orden obtenido en términos de la serie temporal promedio es el mismo que el obtenido en términos de la serie temporal X_t .

Proposición 4:

Bajo las hipótesis [3.1] y [3.2], si denominamos:

$$c_{t,k} = \nabla C_{t,k} = C_{t,k} - C_{t-1,k} \quad \text{y} \quad x_t = \nabla X_t = X_t - X_{t-1}$$

que los coeficientes b_k^x verifiquen la hipótesis [3.3] implica que el conjunto de K series temporales $\{c_{t,k}\}$, $k=1,\dots,K$, define respecto de x_t un haz de K rectas de vértice:

$$(x_t, c_{t,k}) = (-B_1^x, B_0^x).$$

Si en particular b_k^x es igual a cero $\forall k$, entonces el vértice del haz coincidirá con el origen de coordenadas:

$$(x_t, c_{t,k}) = (0,0)$$

Demostración:

Por [3.2] y por las propias definiciones de $c_{t,k}$ y de x_t :

$$c_{t,k} = b_k^x \cdot t + m_k^x \cdot X_t + \mu_k^x - (b_k^x \cdot (t-1) + m_k^x \cdot X_{t-1} + \mu_k^x) = b_k^x + m_k^x \cdot (X_t - X_{t-1}) = b_k^x + m_k^x \cdot x_t$$

además:

$$c_{t,k} = c_{t,k'} \quad \text{sii}^2 \quad b_k^x + m_k^x \cdot x_t = b_{k'}^x + m_{k'}^x \cdot x_t \quad \text{sii} \quad (m_k^x - m_{k'}^x) \cdot x_t = -(b_k^x - b_{k'}^x) \quad \text{sii}$$

$$(m_k^x - m_{k'}^x) \cdot x_t = -(B_0^x + B_1^x \cdot m_k^x - (B_0^x + B_1^x \cdot m_{k'}^x)) \quad \text{sii} \quad (m_k^x - m_{k'}^x) \cdot x_t = -B_1^x \cdot (m_k^x - m_{k'}^x)$$

En consecuencia:

$$c_{t,k} = c_{t,k'} \quad \text{sii} \quad x_t = -B_1^x$$

y, en dicho caso, por [3.3]:

$$c_{t,k} = b_k^x + m_k^x \cdot x_t = b_k^x - m_k^x \cdot B_1^x = B_0^x$$

Además:

$$b_k^x = 0 \quad \forall k \quad \text{sii} \quad 0 = B_0^x + B_1^x \cdot m_k^x \quad \forall k \quad \text{sii} \quad B_0^x = B_1^x = 0$$

² En lo sucesivo utilizaremos la nomenclatura estándar *sii* para simplificar la expresión *si y sólo si*.

Proposición 5:

Bajo las hipótesis [3.1] y [3.2], que los coeficientes b_k^x verifiquen la hipótesis [3.3] implica que los valores de la serie temporal c_t se encuentran en una recta del haz definido por el conjunto de las K series temporales $\{c_{t,k}\}$, $k=1,\dots,K$, respecto de x_t .

Demostración:

Por la **Proposición 1** y por la propia definición de c_t :

$$c_t = b^x \cdot t + m^x \cdot X_t + \mu^x - (b^x \cdot (t-1) + m^x \cdot X_{t-1} + \mu^x) = b^x + m^x \cdot x_t$$

donde, por [3.8]:

$$b^x = B_0^x + B_1^x \cdot m^x$$

Luego, si $x_t = -B_1^x$:

$$c_t = b^x + m^x \cdot (-B_1^x) = b^x - m^x \cdot B_1^x = B_0^x$$

En otras palabras, los puntos de la forma (x_t, c_t) , $t=1,\dots,T$, se encuentran en una recta que pasa por el punto:

$$(x_t, c_t) = (-B_1^x, B_0^x)$$

que, por la **Proposición 4**, es el vértice del haz de rectas que definen las K series temporales $c_{t,k}$ respecto de x_t .

Proposición 6:

Bajo las hipótesis [3.1] y [3.2], que los coeficientes b_k^x verifiquen [3.3] implica que el conjunto de K series temporales $\{c_{t,k}\}$, $k=1,\dots,K$, define respecto de c_t un haz de K rectas de vértice:

$$(c_t, c_{t,k}) = (B_0^x, B_0^x)$$

Si en particular b_k^x es igual a cero $\forall k$, entonces el vértice del haz coincidirá con el origen de coordenadas:

$$(c_t, c_{t,k}) = (0,0)$$

Demostración:

Por [3.10] y por las propias definiciones de $c_{t,k}$ y de c_t :

$$c_{t,k} = b_k^c \cdot t + m_k^c \cdot C_t + \mu_k^c - (b_k^c \cdot (t-1) + m_k^c \cdot C_{t-1} + \mu_k^c) = b_k^c + m_k^c \cdot c_t$$

Además:

$$c_{t,k} = c_{t,k'} \quad \text{sii} \quad b_k^c + m_k^c \cdot c_t = b_{k'}^c + m_{k'}^c \cdot c_t \quad \text{sii} \quad (m_k^c - m_{k'}^c) \cdot c_t = -(b_k^c - b_{k'}^c) \quad \text{sii}$$

$$(m_k^c - m_{k'}^c) \cdot x_t = -(B_0^c + B_1^c \cdot m_k^c - (B_0^c + B_1^c \cdot m_{k'}^c)) \quad \text{sii} \quad (m_k^c - m_{k'}^c) \cdot c_t = -B_1^c \cdot (m_k^c - m_{k'}^c)$$

En consecuencia:

$$c_{t,k} = c_{t,k'} \quad \text{sii} \quad c_t = -B_1^c$$

y, en dicho caso, por [3.11]:

$$c_{t,k} = b_k^c + m_k^c \cdot c_t = b_k^c - m_k^c \cdot B_1^c = B_0^c$$

Además, por [3.12]:

$$B_0^c = B_0^x \quad \text{y} \quad B_1^c = -B_0^x$$

En definitiva, el haz de K rectas de la forma:

$$c_{t,k} = b_k^c + m_k^c \cdot c_t$$

tiene por vértice el punto:

$$(c_t, c_{t,k}) = (B_0^c, -B_1^c) = (B_0^x, B_0^x)$$

Además, si b_k^x es igual a cero $\forall k$, entonces, por la Proposición 4, $B_0^x = 0$.

Proposición 7:

Bajo las hipótesis [3.1], [3.2] y [3.3], dado un par cualquiera de series del conjunto $\{C_{t,k}\}$, $k=1,\dots,K$, existe al menos un punto en su trayectoria en el que las series se cruzan. Además, en aquellos puntos tales que dos de las series se cruzan, también lo hacen las restantes, así como la serie temporal promedio.

Demostración:

Si dos series r y s no se cruzaran en ningún punto de su trayectoria entonces una de las dos trayectorias estaría siempre por encima de la otra, en contradicción con la hipótesis de medias iguales. Es decir, si:

$$C_{t,r} > C_{t,s} \quad \forall t$$

entonces:

$$\bar{C}_r = \sum_{t=1}^T \frac{C_{t,r}}{T} > \sum_{t=1}^T \frac{C_{t,s}}{T} = \bar{C}_s$$

En consecuencia, existe al menos un punto en la trayectoria de las dos series en el que se cruzan. Además, si t es el instante correspondiente al cruce, entonces:

$$C_{t,r} - C_{t,s} = 0$$

y, por [3.5]:

$$(B_1^x \cdot t + X_t + D_1^x) \cdot (m_r^x - m_s^x) = 0$$

Si fuera:

$$m_r^x = m_s^x$$

entonces, por [3.3] y por [3.4]:

$$b_r^x = b_s^x \quad \text{y} \quad \mu_r^x = \mu_s^x$$

lo que implicaría que $C_{t,r}$ y $C_{t,s}$ fueran la misma serie, en contradicción con la hipótesis [3.1]. Entonces:

$$B_1^x \cdot t + X_t + D_1^x = 0$$

y, en consecuencia, en dicho instante t :

$$(B_1^x \cdot t + X_t + D_1^x) \cdot (m_k^x - m_{k'}^x) = 0 \quad \forall k, k'$$

sii, por [3.5]:

$$C_{t,k} = C_{t,k'} \quad \forall k, k'$$

Es decir, en aquellos puntos tales que las trayectorias de dos de las series se cruzan, también lo hacen las restantes. Por otro lado, en dicho instante t también se verifica:

$$(B_1^x \cdot t + X_t + D_1^x) \cdot (m_k^x - m^x) = 0$$

En consecuencia:

$$(B_1^x \cdot t + X_t + D_1^x) \cdot m_k^x = (B_1^x \cdot t + X_t + D_1^x) \cdot m^x$$

sii:

$$D_0^x + B_0^x \cdot t + (B_1^x \cdot t + X_t + D_1^x) \cdot m_k^x = D_0^x + B_0^x \cdot t + (B_1^x \cdot t + X_t + D_1^x) \cdot m^x$$

sii:

$$(B_0^x + B_1^x \cdot m_k^x) \cdot t + m_k^x \cdot X_t + D_0^x + D_1^x \cdot m_k^x = (B_0^x + B_1^x \cdot m^x) \cdot t + m^x \cdot X_t + D_0^x + D_1^x \cdot m^x$$

Por [3.5]:

$$C_{t,k} = (B_0^x + B_1^x \cdot m_k^x) \cdot t + m_k^x \cdot X_t + D_0^x + D_1^x \cdot m_k^x$$

y, por [3.8] y [3.9]:

$$(B_0^x + B_1^x \cdot m^x) \cdot t + m^x \cdot X_t + D_0^x + D_1^x \cdot m^x = b^x \cdot t + m^x \cdot X_t + \mu^x = C_t$$

Entonces:

$$C_{t,k} = C_t \quad \forall k$$

Es decir, en aquellos puntos tales que las trayectorias de las K series se cruzan, también lo hace la de la serie promedio.

Proposición 8:

Bajo las hipótesis [3.1], [3.2] y [3.3], si existe más de un punto en el que las trayectorias de todas las series del conjunto $\{C_{t,k}\}$, $K=1,\dots,k$, se cruzan, entonces la diferencia entre dos puntos de cruce cualesquiera es independiente de la media de las series.

Demostración:

Por la Proposición 7, si t es el instante correspondiente a un punto de cruce de todas las trayectorias, entonces:

$$B_1^x \cdot t + X_t + D_1^x = 0$$

y, por [3.5] el valor del punto de cruce para las K series del conjunto es:

$$C_{t,k} = D_0^x + B_0^x \cdot t + (B_1^x \cdot t + X_t + D_1^x) \cdot m_k^x = D_0^x + B_0^x \cdot t$$

En consecuencia, la diferencia entre dos puntos de cruce en dos instantes t y t' no depende de cuál sea la media de las series temporales:

$$C_{t,k} - C_{t',k} = D_0^x + B_0^x \cdot t - (D_0^x + B_0^x \cdot t') = B_0^x \cdot (t - t')$$

Resumiendo:

Si tenemos un conjunto de K series temporales $\{C_{t,k}\}$, $k=1,\dots,K$, todas ellas con la misma media, generadas por otra serie temporal X_t mediante una expresión del tipo:

$$C_{t,k} = b_k^x \cdot t + m_k^x \cdot X_t + \mu_k^x$$

y tales que las series diferenciadas correspondientes:

$$c_{t,k} = \nabla C_{t,k} = C_{t,k} - C_{t-1,k} \quad \text{y} \quad x_t = \nabla X_t = X_t - X_{t-1}$$

definen un haz de rectas secantes (Fig. 3.2, Sup. Izqda):

$$c_{t,k} = b_k^x + m_k^x \cdot x_t \quad \text{con} \quad b_k^x = B_0^x + B_1^x \cdot m_k^x$$

entonces la secuencia de pendientes ordenadas de menor a mayor valor:

$$m_{(1)}^x, m_{(2)}^x, \dots, m_{(K)}^x$$

nos permite ordenar las K series temporales $C_{t,k}$, $k=1,\dots,K$, por su similitud (Fig. 3.2, Sup. Dcha).

Además, considerando que la serie promedio C_t (Fig. 3.2, Inf. Izqda) es tal que respecto de la correspondiente serie diferenciada:

$$c_t = \nabla C_t = C_t - C_{t-1} \quad \text{donde} \quad C_t = \frac{\sum_{k=1}^K C_{t,k}}{K}$$

el conjunto de K series temporales $\{c_{t,k}\}$, $k=1,\dots,K$, también define un haz de rectas secantes (Fig. 3.2, Inf. Dcha):

$$c_{t,k} = b_k^c + m_k^c \cdot c_t \quad \text{y} \quad b_k^c = B_0^c + B_1^c \cdot m_k^c$$

entonces para establecer la ordenación no es necesario disponer de la serie X_t , aunque se suponga que existe.

Por otro lado, existe al menos un punto en el que todas las trayectorias de las series temporales del conjunto $\{C_{t,k}\}$ se cruzan; y en aquellos puntos tales que dos de las series se cruzan también lo hacen las restantes, así como la serie promedio (Fig. 3.2, Inf. Izqda). Además, si se cruzan en más de un punto, la diferencia entre dos puntos de cruce cualesquiera es independiente de la media de las series.

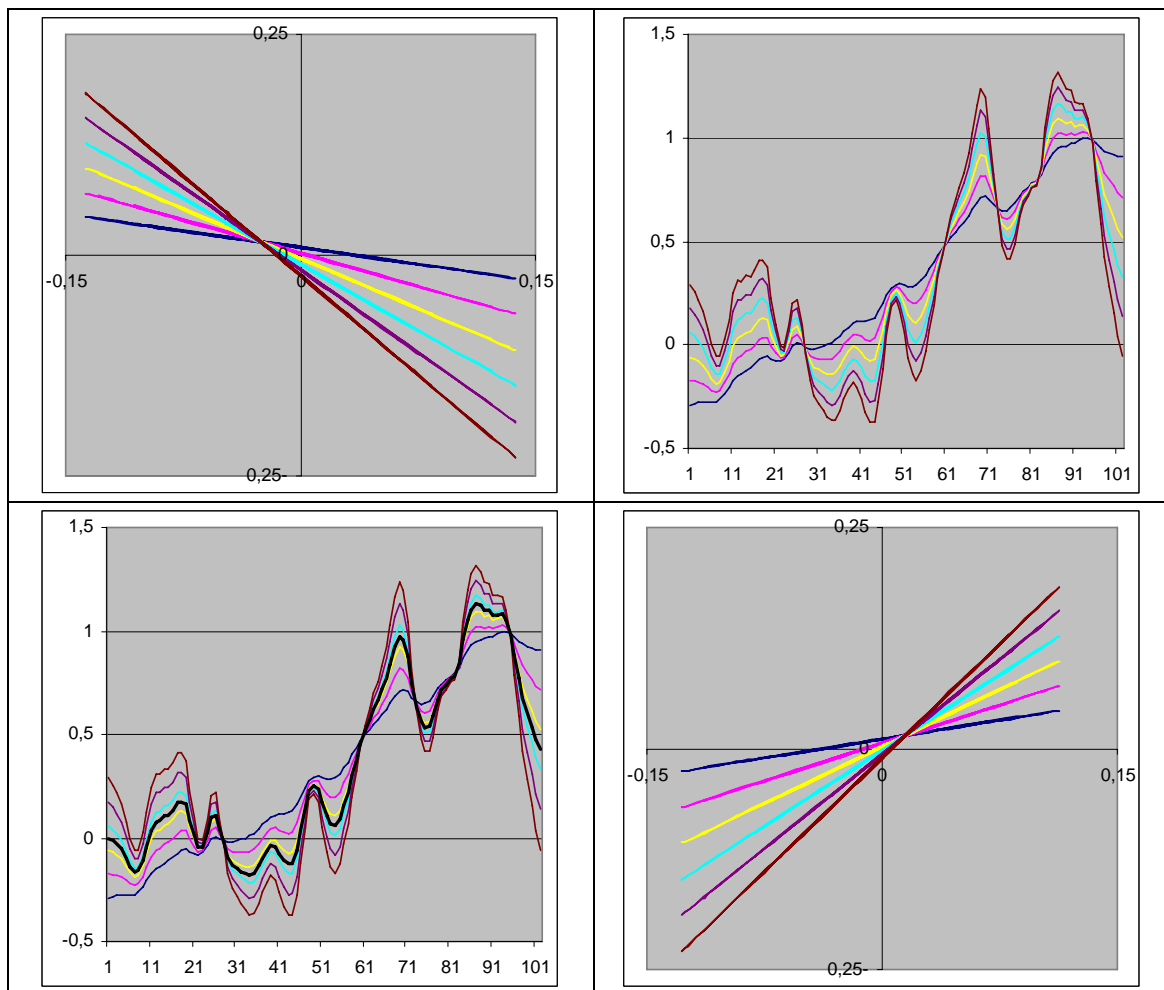


Figura 3.2 *Sup. izqda:* Valores de $c_{t,k}$ frente a x_t ; *Sup. dcha:* Series $C_{t,k}$; *Inf. izqda:* Series $C_{t,k}$ y C_t ; *Inf. dcha:* Valores de $c_{t,k}$ frente a c_t

CAPÍTULO 4

METODOLOGÍA DEL HAZ DE RECTAS: APLICACIÓN AL SECTOR DE LA CONSTRUCCIÓN RESIDENCIAL

Dado el conjunto de las cincuenta series temporales relativas a la oferta de vivienda nueva en cada una de las cincuenta provincias españolas, se trata de obtener un conjunto de K series resumen (con K pequeño) que nos permita, además de agrupar las cincuenta series temporales en K grupos homogéneos, interpretar y comparar sus trayectorias. Para obtener K grupos homogéneos, en el sentido de que las series pertenecientes a un mismo grupo presenten trayectorias parecidas entre sí y distintas a las de series pertenecientes a otro grupo, se aplicará un *Análisis Cluster* utilizando como criterio de agrupación el *Algoritmo de las k -medias*. La metodología que se propone consiste en construir un conjunto de K series resumen del conjunto de series observado que se utilizarán como centroides iniciales del *Algoritmo de las k -medias*. La partición obtenida, así como los centroides finales correspondientes nos permitirán interpretar y comparar las trayectorias de las cincuenta series temporales¹.

Los valores publicados por el Ministerio de Vivienda de la serie de viviendas iniciadas en España no corresponden a observaciones contrastadas², por lo que en la elaboración del modelo se utilizará como indicador de oferta de vivienda nueva la información relativa a los visados de la construcción. De este tipo de información la más adecuada sería la relativa a los visados de obra nueva otorgados por los Colegios de Aparejadores. La serie de visados de los proyectos de ejecución otorgados por los Colegios de Arquitectos, aunque no tan adecuada como la anterior ya que, junto con la

¹ En el Anexo I se proporciona un esquema de la metodología.

² La serie que publica el Ministerio de Vivienda relativa a las viviendas iniciadas se calcula mediante una metodología de desfase aplicada sobre los visados previos al inicio de la construcción, (Vergés, 2001).

nueva edificación, incluye las ampliaciones y reformas, dispone de datos provinciales sobre un periodo histórico de mayor longitud, razón por la que la optaremos por ella a efectos de ilustrar la metodología³. Concretamente, se considerarán las series trimestrales de acumulados anuales⁴ en las cincuenta provincias españolas en el periodo comprendido entre los años 1983 y 2008⁵.

4.1 CONDICIONES DE APLICACIÓN

Sea el conjunto de las J series temporales:

$$\{V_{t,j}\} \quad j=1,\dots,J$$

relativas a los proyectos de construcción de nueva vivienda en el trimestre t , $t=1,\dots,102$ (desde el tercer trimestre de 1983 hasta el cuarto de 2008, ambos inclusive) en cada una de las $J=50$ provincias españolas (**Fig. 4.1**). Para agruparlas en K grupos homogéneos, construiremos un conjunto de K series temporales, $\{C_{t,k}\} k=1,\dots,K$, tales que:

$$C_{t,k} = b_k \cdot t + m_k \cdot \ln X_t + \mu_k \quad \forall k$$

donde $\ln X_t$ será una función resumen del conjunto de series $\ln V_{t,j}$; los coeficientes b_k y m_k verificarán:

$$b_k = B_0 + B_1 \cdot m_k \quad \forall k$$

con, al menos, los m_k no nulos; y, además:

$$\frac{\sum_{t=1}^T C_{t,k}}{T} = \alpha \quad \forall k$$

En otras palabras, construiremos un conjunto de K series temporales, todas ellas en la misma escala, tales que las series diferenciadas correspondientes definan, respecto de la serie diferenciada de $\ln X_t$, un haz de rectas de vértice⁶:

$$(-B_1, B_0)$$

³ Véase el Apartado 1.4 del Capítulo 1 para una justificación más detallada de esta decisión.

⁴ En el Apéndice 1 se muestra la relación entre los datos transformados y los originales.

⁵ Véase el Apartado 1.3 del Capítulo 1 para una justificación de esta decisión.

⁶ Véase la Proposición 4 del Capítulo 3.

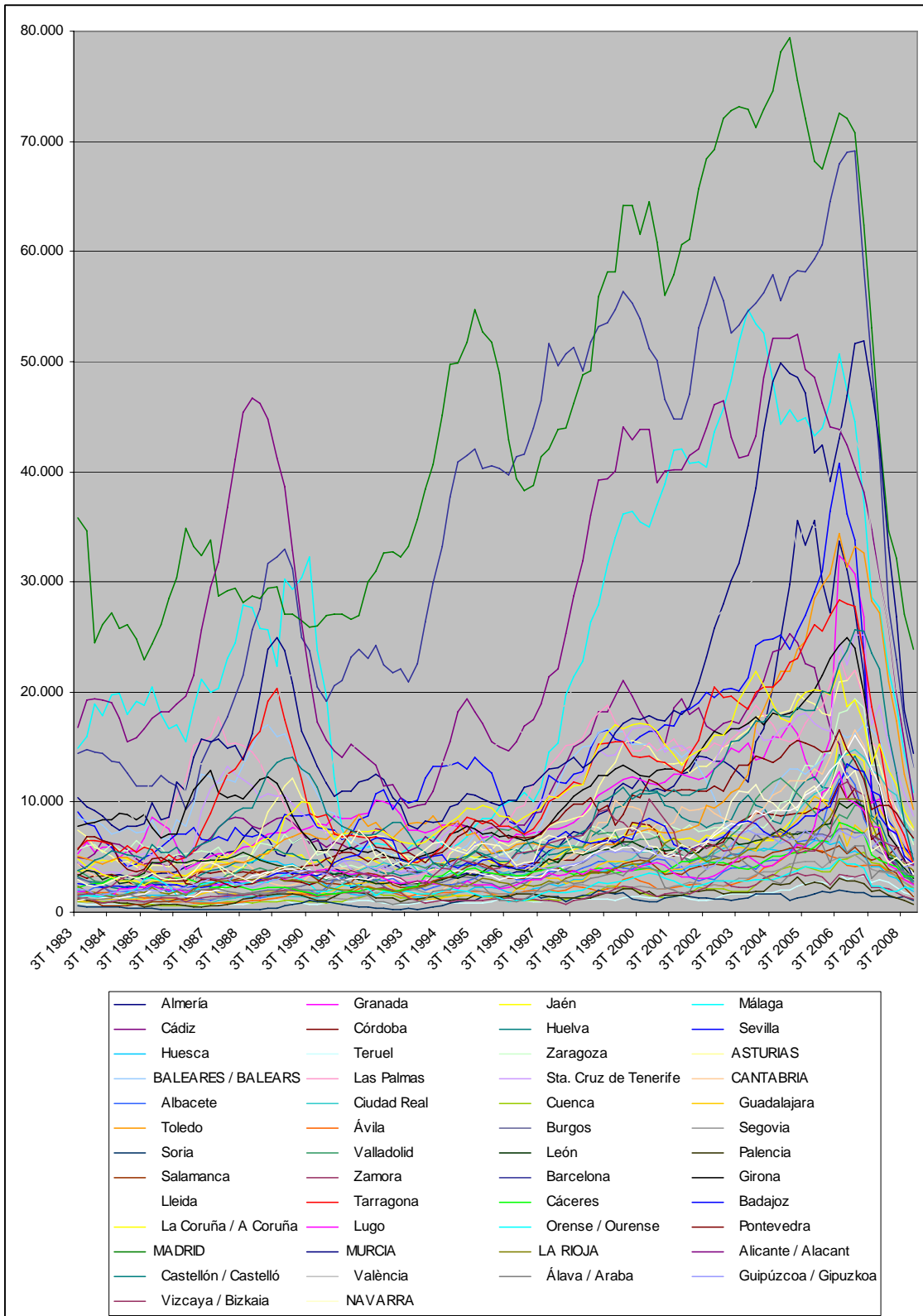


Figura 4.1 Series $V_{t,j}$ Fuente: CSCAE

La construcción de este conjunto de K series temporales para resumir el comportamiento de las J series temporales objeto de análisis estará justificada si la estructura que subyace es la de un haz de rectas. En otras palabras, la metodología será aplicable si, al considerar la serie promedio en términos del logaritmo⁷:

$$\ln V_t = \frac{\sum_{j=1}^J \ln V_{t,j}}{J}$$

y ajustar las J ecuaciones de regresión lineal por mínimos cuadrados ordinarios (MCO):

$$\ln \hat{V}_{t,j} = A_{0,j} \cdot t + A_{1,j} \cdot \ln V_t + A_{2,j} \quad [4.1]$$

se detecta, por un lado, que los coeficientes $A_{1,j}, j=1, \dots, J$, son no nulos y, por otro, que los J puntos de la forma $(A_{1,j}, A_{0,j})$ se concentran en torno a una línea recta.

Por ejemplo, obsérvese (**Fig. 4.2, Sup. Izqda**) que la serie $V_{t,12}$ -en la escala izquierda-, correspondiente a Asturias, presenta una estructura muy similar a la de la serie V_t -en la escala derecha-, correspondiente al promedio. El ajuste de sus valores en términos del promedio según la ecuación [4.1] vendrá dado por (**Fig. 4.2, Sup. Dcha**):

$$\ln \hat{V}_{t,12} = A_{0,12} \cdot t + A_{1,12} \cdot \ln V_t + A_{2,12}$$

donde $A_{0,12}$, $A_{1,12}$ y $A_{2,12}$ son los coeficientes estimados por MCO. Concretamente (**Fig. 4.2, Inferior**):

$$A_{0,12} = -0,003, \quad A_{1,12} = 1,324 \quad \text{y} \quad A_{2,12} = -2,163$$

Al nivel de significación α igual a 0,05, podemos concluir que el coeficiente que nos ocupa, $A_{1,12}$, es significativamente distinto de cero: el p -valor asociado al estadístico de contraste es menor que el nivel de significación α .

El hecho de que excepcionalmente, para alguna de las J series temporales, el coeficiente asociado al promedio, $A_{1,j}$, pueda ser considerado nulo carece de importancia, pero la aplicación de esta metodología tendrá sentido cuando en términos generales, para la mayoría de las series del conjunto de análisis, sea significativamente distinto de cero⁸.

⁷ La justificación de considerar las series transformadas por la función logaritmo se proporciona en el apartado 4.4.

⁸ El Apéndice 2 proporciona información relativa a los valores de estos coeficientes para las 50 provincias.

Una vez calculados los coeficientes $A_{0,j}$, $A_{1,j}$ y $A_{2,j}$, el siguiente paso consistirá en comprobar que los J puntos de la forma $(A_{1,j}, A_{0,j})$ se concentran en torno a una línea recta. Efectivamente, obsérvese (**Fig. 4.3**) que dichos puntos se concentran en torno a la recta de regresión lineal ajustada por MCO de la forma:

$$\hat{A}_{0,j} = B_0 + B_1 \cdot A_{1,j} = 0,015 - 0,015 \cdot A_{1,j} \quad j=1, \dots, J$$

Al nivel de significación α igual a 0,05, podemos concluir que ambos coeficientes son significativamente distintos de cero: el p -valor asociado al estadístico de contraste es menor que el nivel de significación α en ambos casos.

Si B_1 fuera nulo significaría que la recta es paralela al eje horizontal. Este hecho carecería de importancia siempre y cuando dicha recta se cruzara con el eje de ordenadas a la altura del cero y los puntos se concentraran en torno a dicha recta; en otras palabras, siempre y cuando, los coeficientes $A_{0,j}$ fueran todos próximos a cero. En este caso, la componente determinística en la expresión [4.1] sobraría. La única connotación, a efectos de la metodología, sería que la expresión de las series $C_{t,k}$ tampoco incluiría la componente determinística, lo que implicaría que el haz de rectas correspondiente a las series diferenciadas tendría por vértice el origen de coordenadas⁹.

4.2 HOMOGENEIZACIÓN DE LA ESCALA

Para aplicar la metodología será necesario homogeneizar la escala de medida de las J series temporales $V_{t,j}$. Para ello se considerarán J transformaciones de la forma:

$$\ln Y_{t,j} = \ln V_{t,j} + \varepsilon_j \quad j=1, \dots, J$$

donde ε_j será aquel valor constante tal que la media de las J series transformadas sea la misma. Concretamente, consideraremos:

$$\varepsilon_j = \alpha - \alpha_j$$

donde:

$$\alpha_j = \frac{\sum_{t=1}^T \ln V_{t,j}}{T} \quad \text{y} \quad \alpha = \frac{\sum_{t=1}^T \ln Y_t}{T}$$

siendo:

$$Y_t = \frac{V_t}{V_1} \quad t=1, \dots, T$$

⁹ Véase la Proposición 6 del Capítulo 3.

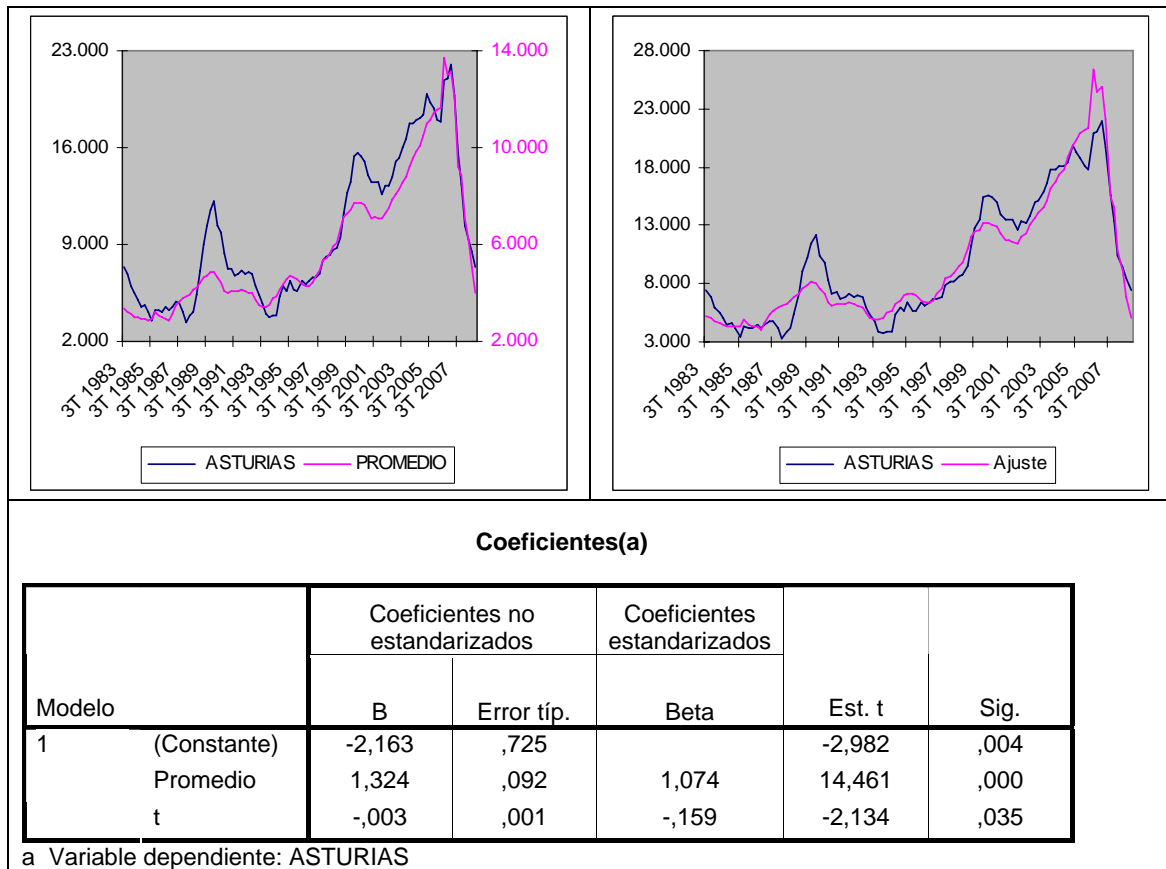


Figura 4.2 Sup. izqda: Series $V_{t,12}$ y v_{12} ; Sup. dcha: Series $V_{t,12}$ y $\hat{V}_{t,12}$; Inferior: Valores de los coeficientes $A_{0,12}$, $A_{1,12}$ y $A_{2,12}$

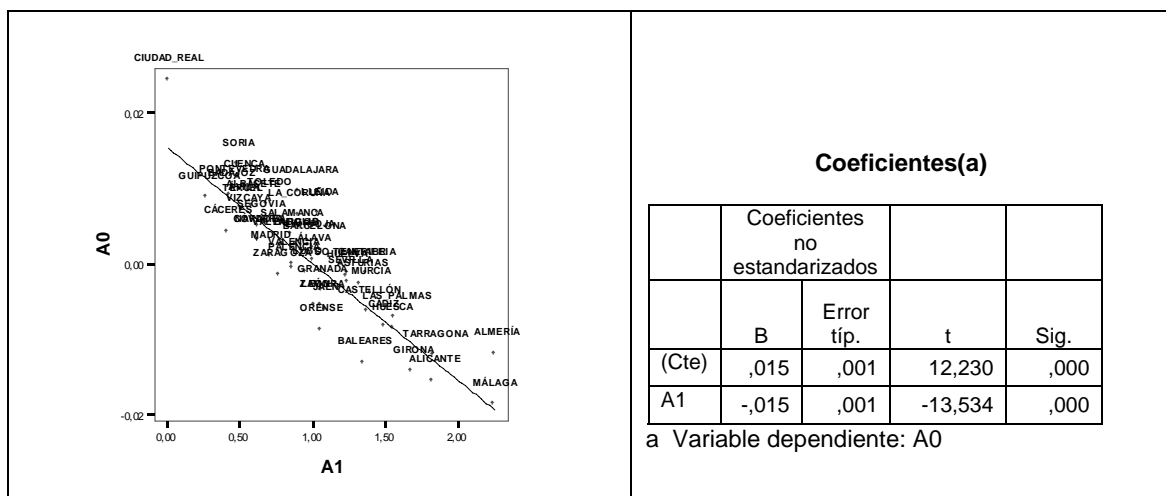


Figura 4.3 Izqda: Valores de $A_{0,j}$ frente a valores de $A_{1,j}$; Dcha: Valores de los coeficientes B_0 y B_1

Así¹⁰,

$$\ln Y_t = \frac{\sum_{j=1}^J \ln Y_{t,j}}{J}$$

y las series transformadas tienen todas la misma media¹¹:

$$\frac{\sum_{t=1}^T \ln Y_{t,j}}{T} = \frac{\sum_{t=1}^T \ln Y_t}{T} = \alpha \quad \forall j$$

La **Fig. 4.4** muestra la representación gráfica de las series promedio y Asturias antes (V_t y $V_{t,12}$, en el panel izquierdo) y después de la transformación (Y_t e $Y_{t,12}$, en el panel derecho). Obsérvese que las series Y_t e $Y_{t,12}$ son tales que mantienen la forma original de V_t y $V_{t,12}$, respectivamente, pero ahora se encuentran en la misma escala, lo que facilita la comparación de sus trayectorias, que se establece en términos de su crecimiento desde el instante inicial¹².

4.3 CÁLCULO DE LOS CENTROIDES INICIALES

Si K es el número de grupos que se desea formar, fijaremos K valores dispersos, m_k , dentro del rango de variación de $A_{1,j}$ y consideraremos:

$$b_k = B_0 + B_1 \cdot m_k \quad k=1,\dots,K$$

Por ejemplo, en términos del rango de variación de $A_{1,j}$ (**Fig. 4.3**), fijaremos $K=6$ coeficientes de la forma:

$$m_1 = \min_j A_{1,j} \quad \text{y} \quad m_k = m_{k-1} + \theta \quad \text{para} \quad k=2,\dots,6$$

siendo:

$$\theta = (\max_j A_{1,j} - \min_j A_{1,j}) / (K - 1)$$

Y, a partir de los valores de B_0 y B_1 , calcularemos los coeficientes b_k :

¹⁰ Véase el Apéndice 3.

¹¹ Véase el Apéndice 4.

¹² Por definición, el valor de Y_t en el primer instante es igual a 1. El valor de la curva en cualquier otro instante será el incremento en dicho instante respecto del inicial. El Apéndice 5 proporciona alguna aclaración adicional relativa a la comparación de la curva promedio con la de las distintas provincias.

$$b_k = 0,015 + (-0,015) \cdot m_k \quad k=1, \dots, 6$$

Estos coeficientes m_k y b_k (**Fig. 4.5, Sup. Izqda.**) se utilizarán para construir el conjunto de K series $C_{t,k}^I$, $k=1, \dots, K$, centroides iniciales del *Algoritmo de las k-medias* para la agrupación de las J series temporales transformadas¹³.

El proceso de construcción de los centroides iniciales se realizará en tres etapas. En primer lugar, si denominamos:

$$\ln y_t = \nabla \ln Y_t = \ln \left(\frac{Y_t}{Y_{t-1}} \right) \quad t=2, \dots, T$$

se calculará el haz de K rectas de la forma:

$$f_{t,k} = b_k + m_k \cdot \ln y_t, \quad t=2, \dots, T \quad \text{y} \quad f_{1,k} = b_k, \quad k=1, \dots, K$$

Concretamente, en nuestro análisis, la primera serie:

$$f_{t,1} = 0,015 + 0,006 \cdot \ln y_t$$

al corresponder al valor m_1 , el mínimo entre los seis posibles de m_k , es la menos sensible a las fluctuaciones de la serie promedio, $\ln y_t$, y, en consecuencia, la menos volátil de las K funciones; por el contrario, la sexta:

$$f_{t,6} = -0,019 + 2,25 \cdot \ln y_t$$

al corresponder al valor m_6 , el máximo entre los seis posibles de m_k , es la más sensible a las fluctuaciones de la serie promedio, $\ln y_t$, y, en consecuencia, la más volátil.

A partir de las series $f_{t,k}$, y como segunda etapa del proceso de construcción de los centroides iniciales, se calcularán otras K series de la forma:

$$g_{t,k} = g_{t-1,k} + f_{t,k}, \quad t=2, \dots, T \quad \text{y} \quad g_{1,k} = f_{1,k}, \quad k=1, \dots, K$$

¹³ Si, tal como es nuestro caso, el objetivo es más el de comparar las series que el de construir un número concreto de grupos, la elección de K puede hacerse a modo de tanteo y, en función de la solución obtenida, corregir su valor si parece necesario y aplicar de nuevo la metodología.

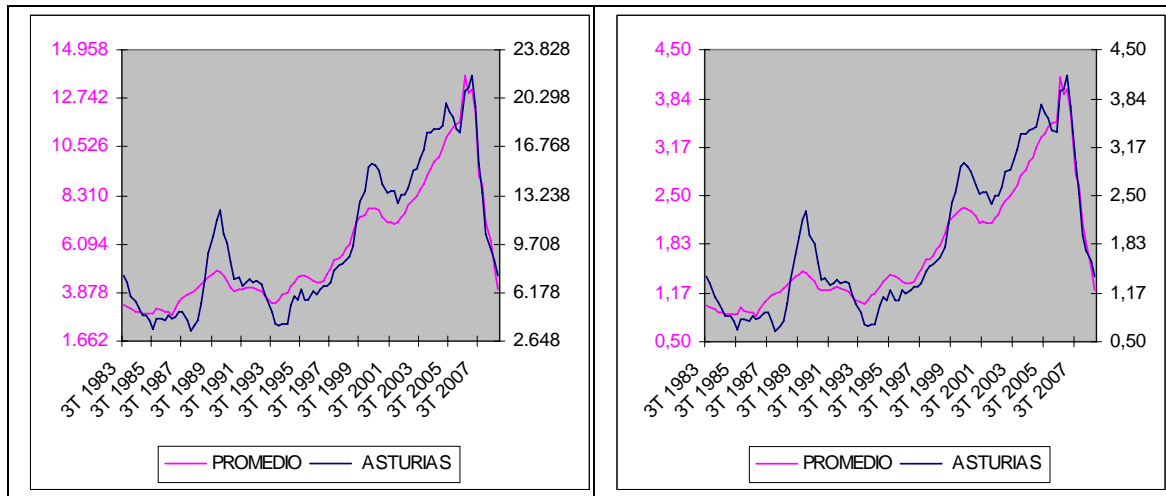


Figura 4.4 Izqda: Series V_t y $V_{t,12}$; Dcha: Series Y_t e $Y_{t,12}$

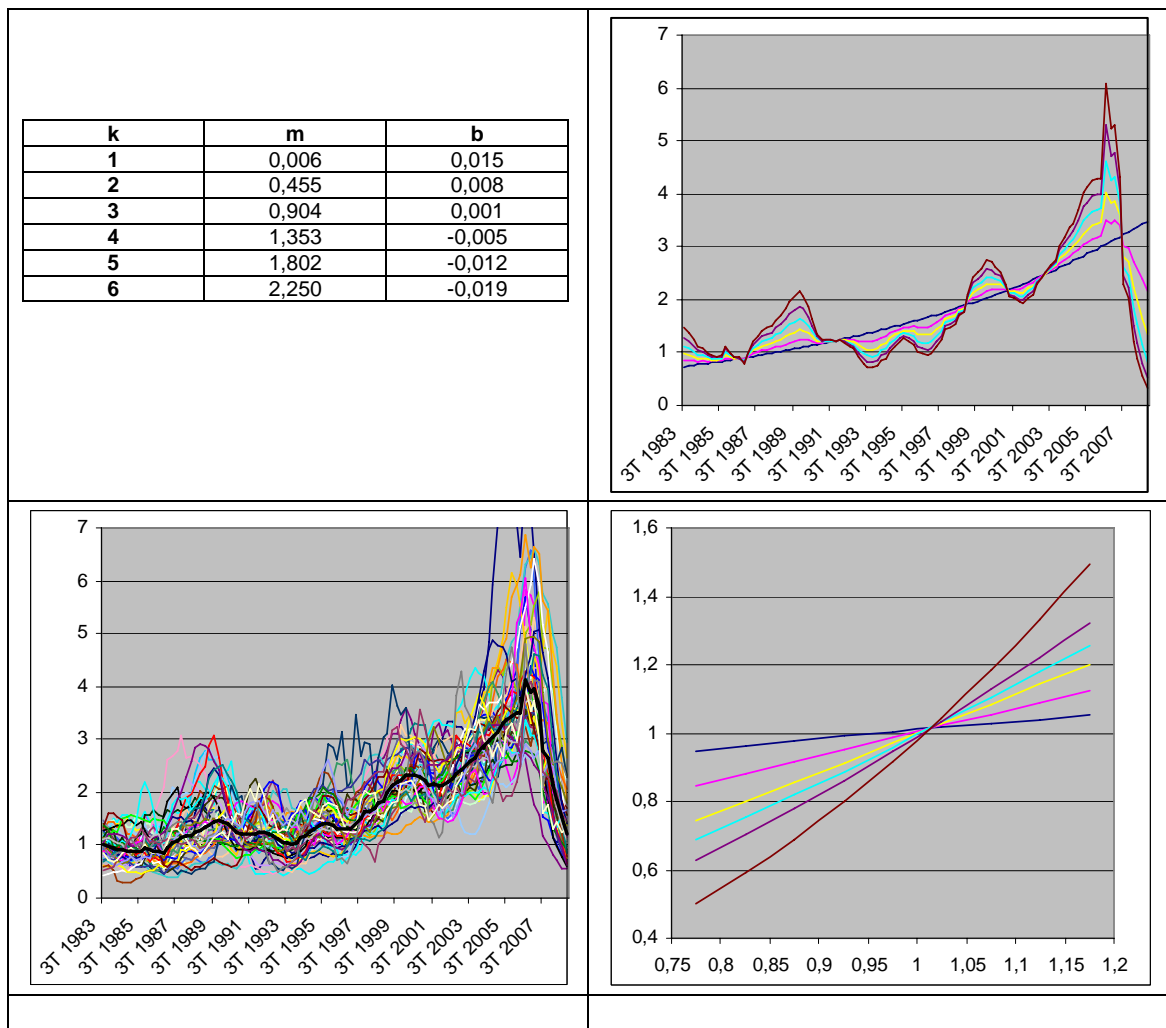


Figura 4.5 Sup. izqda: Tabla de los valores de m_k y de b_k ; Sup. dcha: Series $H_{t,k}^I$; Inf. izqda: Series $Y_{t,j}$; Inf. dcha: Valores de $h_{t,k}^I$ frente a los de y_t

Finalmente, los valores de las series $g_{t,k}$ se homogeneizan en términos del nivel de las series transformadas, $\ln Y_{t,j}$ $j=1, \dots, J$. Sea:

$$\beta_k = \frac{\sum_{t=1}^T g_{t,k}}{T} \quad k=1, \dots, K$$

si consideramos:

$$C_{t,k}^I = g_{t,k} - \beta_k + \alpha$$

Entonces su media coincide con la de las series transformadas¹⁴:

$$\frac{\sum_{t=1}^T C_{t,k}^I}{T} = \alpha \quad \forall k$$

Es decir, los conjuntos de series:

$$\ln Y_{t,j} \quad j=1, \dots, J \quad \text{y} \quad C_{t,k}^I \quad k=1, \dots, K$$

se encuentran al mismo nivel. Alternativamente, las K series de la forma:

$$H_{t,k}^I = \exp C_{t,k}^I$$

se encuentran en la misma escala que las J series temporales $Y_{t,j}$ (**Fig. 4.5**, en los paneles *Sup. Dcha.* e *Inf. Izqda.*, respectivamente).

A efectos de la partición del conjunto de las J series originales $V_{t,j}$ en K grupos homogéneos, aplicaremos el *Análisis Cluster* sobre las series $\ln Y_{t,j}$ utilizando como criterio de agrupación el *Algoritmo de las k-medias* y considerando como centroides iniciales los valores de las K curvas $C_{t,k}^I$.

4.4 GENERACIÓN DE LOS GRUPOS

Dado el conjunto de las J series temporales $\{Y_{t,j}\}$, $j=1, \dots, J$, para agruparlas en K grupos mediante el *Algoritmo de las k-medias* consideraremos en cada iteración la distancia $d_{k,j}^i$:

¹⁴ Véase el Apéndice 6 para la justificación de esta afirmación.

$$d_{k,j}^i = d^2(\ln Y_{t,j}, C_{t,k}^{i-1}) \quad k=1,\dots,K \quad j=1,\dots,J$$

siendo d la distancia euclídea y $C_{t,k}^i$ los K centroides correspondientes a la i -ésima iteración¹⁵. La primera agrupación se hará en términos de los centroides iniciales:

$$d_{k,j}^1 = d^2(\ln Y_{t,j}, C_{t,k}^1) \quad k=1,\dots,K \quad j=1,\dots,J$$

En las sucesivas iteraciones:

$$C_{t,k}^i = \frac{\sum_{j^{(k)}=1}^{n(k)} \ln Y_{t,j^{(k)}}}{n(k)} \quad k=1,\dots,K$$

donde $n(k)$ es el total de series que contiene el grupo k -ésimo en la i -ésima iteración y $j^{(k)}$ es el subíndice correspondiente a las series asignadas a dicho grupo. A los centroides correspondientes a la solución final los denominaremos:

$$C_{t,k}^F \quad k=1,\dots,K \quad t=1,\dots,T$$

La solución obtenida proporciona la siguiente partición¹⁶ (**Fig. 4.6**):

Grupo 1 (panel superior izquierdo):

Provincias = { CIUDAD REAL }

Grupo 2 (panel superior derecho):

Provincias = { ÁVILA, CUENCA, ALBACETE, TERUEL, BURGOS, LLEIDA, PONTEVEDRA, TOLEDO, GUADALAJARA, SORIA }

Grupo 3 (panel central izquierdo):

Provincias = { CÓRDOBA, VALENCIA, HUELVA, MADRID, GUIPÚZCOA, NAVARRA, SEGOVIA, BADAJOZ, VALLADOLID, GRANADA, JAÉN, SEVILLA, PALENCIA, BARCELONA, LA CORUÑA, VIZCAYA, CÁCERES, CÁDIZ, SALAMANCA }

Grupo 4 (panel central derecho):

Provincias = { LA RIOJA, LUGO, CASTELLÓN, LEÓN, ZARAGOZA, ZAMORA, MURCIA, ASTURIAS, ORENSE, CANTABRIA, ALAVA }

¹⁵ En el Apéndice 7 se proporciona una justificación de considerar la distancia sobre los valores transformados por la función logaritmo.

¹⁶ Véase el Anexo II para el proceso de formación de los grupos.

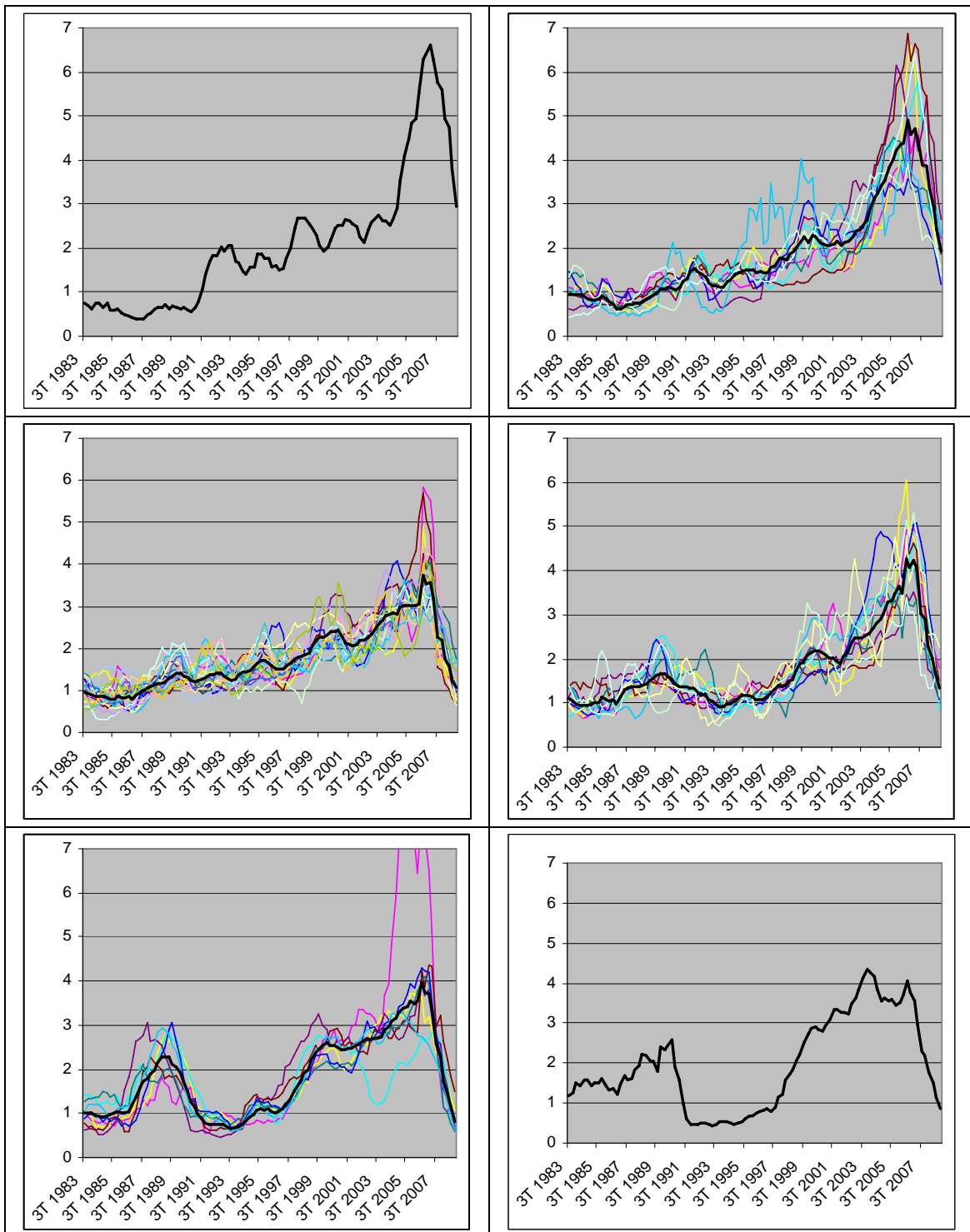


Figura 4.6 De izquierda a derecha y de arriba abajo, serie $\exp(C_{t,k}^F)$ junto con las series $Y_{t,j}$ que forman el grupo final correspondiente, para $k=1,\dots,6$ respectivamente

Grupo 5 (panel inferior izquierdo):

Provincias = {ALICANTE, HUESCA, TARRAGONA, GIRONA, SC TENERIFE, BALEARES, LAS PALMAS, ALMERÍA}

Grupo 6 (panel inferior derecho):

Provincias = {MÁLAGA}

Obsérvese, al comparar los seis grupos obtenidos, que dentro de un mismo grupo las distintas curvas siguen trayectorias próximas a las del centroide correspondiente. Por otro lado, el centroide del primer grupo se parece al del segundo que, a su vez, se parece al del tercero y así sucesivamente hasta el centroide del quinto grupo que es muy parecido al del sexto. En otras palabras, en términos generales, la curva correspondiente a una provincia será parecida a la de las provincias clasificadas en el mismo grupo o en los adyacentes.

Si repitiéramos el análisis sobre un segmento del periodo de análisis, esta solución sufriría variaciones, de menor o mayor magnitud dependiendo del segmento considerado: si eligiéramos un segmento parecido al periodo completo, la solución sería muy parecida a la actual, mientras que si eligiéramos el segmento al inicio o al final del periodo, la solución podría ser bastante distinta¹⁷. En definitiva, la partición obtenida depende del periodo de análisis considerado, por lo que, bajo este punto de vista, no tendríamos una buena solución. Sin embargo, el interés de la metodología que se propone para la agrupación de series temporales no radica tanto en la solución en sí, como en las herramientas que ofrece a la hora de interpretar las similitudes entre las series.

4.5 INTERPRETACIÓN DE LA SOLUCIÓN

Obsérvese (**Fig. 4.5**, *Sup. Dcha*) que la sexta serie (línea marrón) es la que presenta mayores fluctuaciones y que este efecto se va suavizando de la sexta hacia la primera serie. Es más, las series $H_{t,k}^I$ están ordenadas, en el sentido de que, si:

$$m_q < m_r < m_s$$

entonces¹⁸:

$$d(H_{t,q}^I, H_{t,r}^I) < d(H_{t,q}^I, H_{t,s}^I)$$

Este orden responde en definitiva a su sensibilidad frente a la serie promedio. Concretamente, el conjunto de K series:

¹⁷ Esta apreciación ha sido contrastada aplicando la metodología sobre distintos segmentos dentro del periodo de análisis.

¹⁸ Véase el Apéndice 8 para la justificación de esta afirmación.

$$h_{t,k}^I = \Delta H_{t,k}^I = \frac{H_{t,k}^I}{H_{t-1,k}^I} = \exp(b_k + m_k \cdot \ln y_t) \quad k=1,\dots,K$$

es tal que define un haz de curvas (**Fig. 4.5, Inf. Dcha**) concurrentes en el punto¹⁹:

$$(\exp(-B_1), \exp B_0) = (e^{0,015}, e^{0,015}) = (1.015, 1.015)$$

En dicho haz, en un extremo está la sexta curva (línea marrón), la de mayor sensibilidad frente a la curva promedio y en el extremo opuesto (línea azul oscuro), la primera, que se desvía muy poco de la línea horizontal a la altura del valor 1.

Obsérvese además (**Fig. 4.5, Sup. Dcha**) que en aquellos puntos en los que dos de las series de la forma $H_{t,k}^I$ se cruzan, también lo hacen las restantes²⁰. Además, si en dos instantes t y t' :

$$H_{t,k}^I = H_{t,k'}^I \quad \text{y} \quad H_{t',k}^I = H_{t',k'}^I \quad \text{con} \quad k \neq k'$$

entonces:

$$\frac{H_{t',k}^I}{H_{t,k}^I} = \frac{H_{t',k'}^I}{H_{t,k'}^I}$$

En otras palabras, a lo largo de las trayectorias de las K series $H_{t,k}^I$ existen diversos instantes (los correspondientes a los sucesivos puntos de cruce) en los que el incremento entre cualquier par de ellos es el mismo para todas las series. Así, la relación entre los distintos puntos de cruce facilita la interpretación de la evolución conjunta de las J series temporales observadas.

Por ejemplo, obsérvese que el primer punto de cruce de las K trayectorias se produce al finalizar el año 1986 y, aproximadamente, a la altura del valor 0,9. El segundo punto de cruce se produce al finalizar el año 1991, a la altura del valor 1,25 (aproximadamente). En otras palabras, en el periodo de cinco años transcurrido entre estas dos fechas, el valor de todas las curvas se ha incrementado en una proporción de $1,4=1,25/0,9$, aunque con diferente comportamiento: mientras que los valores de la curva correspondiente al valor de k igual a 1 (en azul oscuro) han crecido de forma paulatina, los de la correspondiente a $k=6$ (en marrón), tras presentar una tendencia creciente en los primeros años del periodo, sufre un decrecimiento acentuado en los últimos. Esta situación se invierte en el periodo comprendido entre el segundo punto de corte y el tercero, que se produce al finalizar el año 1998 a la altura del valor 1,9. El incremento en los siete años comprendidos entre el segundo y el tercer punto de corte es de $1,9/1,25=1,5$. También ahora los valores de la curva correspondiente a $k=1$ han

¹⁹ Véase el Apéndice 9 para la justificación de esta afirmación.

²⁰ En el Apéndice 10 se proporciona una justificación de este resultado.

crecido de forma paulatina; sin embargo, los de la correspondiente a $k=6$ lo hacen a la inversa que en el periodo anterior: al principio decrecen y al final crecen. En resumen, si comparamos las trayectorias de las seis curvas en el periodo de veintiún años comprendido entre finales de 1986 y de 2007, dado que los valores inicial y final coinciden (puntos 0,9 y 3,2, respectivamente), podemos concluir que al final del periodo el valor de las seis curvas es más de tres veces y media superior al valor inicial ($3,2/0,9=3,55$); sin embargo la forma de alcanzar el valor final ha sido muy distinta. Mientras que, en un extremo, la curva correspondiente a $k=1$ lo ha hecho con un crecimiento prácticamente constante, en el otro, la correspondiente a $k=6$ lo ha hecho con mucha volatilidad.

A efectos de simplificar la interpretación anterior podría haber sido de utilidad que el primer punto de cruce estuviera a la altura del valor 1. Para ello, sería suficiente con cambiar de escala, corrigiendo las K series por aquel factor tal que el primer punto de cruce de las series corregidas tomara valor igual a 1. Si se llevara a cabo este cambio de escala, sería necesario aplicar la misma corrección sobre las J series de la forma $Y_{t,j}$ ²¹. Esta misma operación podría realizarse sobre cualquier otro punto de cruce que pudiera ser de interés como referencia para los incrementos.

En definitiva, considerando además que la relación entre los puntos de corte no depende de la escala en que representemos el conjunto de series resumen²², para interpretar la trayectoria de una cualquiera de las J series temporales del conjunto de análisis en términos de las restantes, bastará con poner el conjunto de funciones resumen a su misma escala y compararlas. Por ejemplo, si consideramos la serie de visados de la construcción correspondiente a Asturias, $V_{t,12}$, para interpretar su trayectoria en comparación con la de las restantes provincias, transformaremos el conjunto de series $H_{t,k}^I$ de tal forma que conserven la misma forma pero estén a la misma escala que Asturias. Concretamente, para situar las K curvas $H_{t,k}^I$ al nivel de una curva $V_{t,j}$ cualquiera haremos:

$$H_{t,k}^j = \exp(\ln H_{t,k}^I - \gamma_k + \alpha_j) \quad k=1, \dots, K \quad [4.2]$$

donde:

$$\gamma_k = \frac{\sum_{t=1}^T \ln H_{t,k}^I}{T}$$

y, en particular, para situarlas al nivel de Asturias:

²¹ Recordemos que el criterio para homogeneizar la escala era que la serie promedio tuviera su origen en el valor 1; al hacer el cambio de escala, el efecto sería el de un cambio de origen sobre la serie promedio y, dado que todas las proporciones se mantendrían inalteradas, no sería necesario aplicar de nuevo el *Algoritmo de las K medias* ya que proporcionaría la misma solución.

²² Véase el Apéndice 11 para la justificación de esta afirmación.

$$H_{t,k}^{12} = \exp(\ln H_{t,k}^I - \gamma_k + \alpha_{12}) \quad k=1, \dots, K$$

Obsérvese (**Fig. 4.7, Izqda.**) que la trayectoria de Asturias se cruza con la de las distintas series resumen, si no exactamente en los puntos de cruce de todas ellas, en posiciones muy próximas. En este sentido, podemos afirmar que los puntos de cruce previamente descritos resumen con bastante precisión la tendencia del sector de la construcción en Asturias a lo largo del periodo de observación. Por otro lado, recordemos que Asturias había sido clasificada en el cuarto grupo²³; en consecuencia es esperable que, en términos generales, la trayectoria de Asturias se mantenga próxima a la cuarta serie resumen. Obsérvese que, aunque bien es cierto que en la etapa final -desde el año 2005 aproximadamente- dicha trayectoria (línea resaltada en negrita) transcurre próxima a la de las tercera y cuarta series resumen (líneas amarilla y azul), durante la mayor parte del periodo -hasta el año 2001 aproximadamente- se parece más a la sexta (línea marrón). En otras palabras, bajo el supuesto de que existe un factor responsable de las oscilaciones del conjunto de las series de visados en las distintas provincias, en el sentido de que la menor o mayor volatilidad de las oscilaciones depende del mayor o menor grado de sensibilidad frente a las fluctuaciones de dicho factor, podemos afirmar que, en lo que se refiere a Asturias, mientras que la primera etapa del periodo de observación (desde 1983 hasta 2001) es de fuerte sensibilidad, en comparación con las restantes, en la etapa final esta sensibilidad se reduce a un nivel intermedio.

Con el mismo criterio que hemos modificado la escala de las series $H_{t,k}^I$, podemos modificar la de las series $H_{t,k}^F$, así como la de las series relativas a las provincias clasificadas en su mismo grupo. La representación gráfica correspondiente (**Fig. 4.7, Dcha**) permite interpretar la evolución del sector de la construcción residencial en Asturias respecto de las distintas provincias clasificadas en su propio grupo. La interpretación de la evolución de cualquier otra provincia sería equivalente, sin más que modificar las escalas de los conjuntos de series objeto de representación.

En el Capítulo 6 se describe la trayectoria de cada provincia en términos de los centroides iniciales y de la agrupación obtenida. Considerando que el desarrollo de la metodología se fundamenta en la existencia de un determinado factor responsable de las oscilaciones del conjunto de series objeto de análisis, sería de gran utilidad incorporar su información caso de que fuera conocido. En el Capítulo 5 supondremos que dicho factor es el tipo de interés y construiremos, partir de la serie temporal correspondiente, un segundo conjunto de series resumen. Así, en el Capítulo 6, junto con los dos tipos de representaciones mencionadas, utilizaremos un tercer tipo relativo a este segundo conjunto de curvas resumen.

²³ Véase la solución del *Análisis Cluster* en el Apartado 4.4.

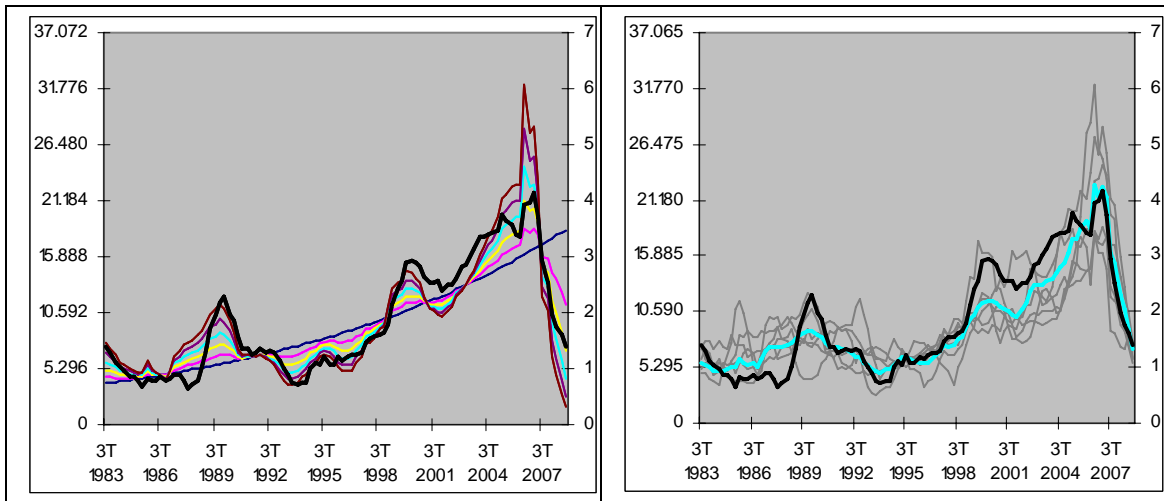


Figura 4.7 Izqda: Series $H_{t,k}^j$ y $V_{t,j}$ (eje izquierdo) y series $H_{t,k}^l$ e $Y_{t,j}$ (eje derecho), para $j=12$ (Asturias); Dcha: Serie $Y_{t,12}$ (Asturias, en negrita), serie $H_{t,4}^F$ (en azul) y series $Y_{t,j}$ correspondientes a las provincias clasificadas en el grupo 4 (en gris), en la escala de la serie promedio (eje derecho) y en la de Asturias (eje izquierdo)

APÉNDICES

Apéndice 1:

El CSCAE publica, para cada una de las provincias españolas, el número de visados otorgados mensualmente. Sin más que sumar los tres datos mensuales correspondientes se obtiene automáticamente la serie trimestral (línea azul en la **Fig. 4.8**). Si, para la j -ésima provincia -por ejemplo para Madrid- denominamos $Z_{t,j}$ a la serie de visados otorgados en el trimestre t (línea azul en la **Fig. 4.8**), para suavizarla, consideraremos el acumulado anual (línea roja en la **Fig. 4.8**), es decir:

$$V_{t,j} = \sum_{s=t-3}^t Z_{s,j}$$

En otras palabras, en cada instante-trimestre el valor de la serie objeto de análisis será el total de visados otorgados en el plazo de los 12 últimos meses.

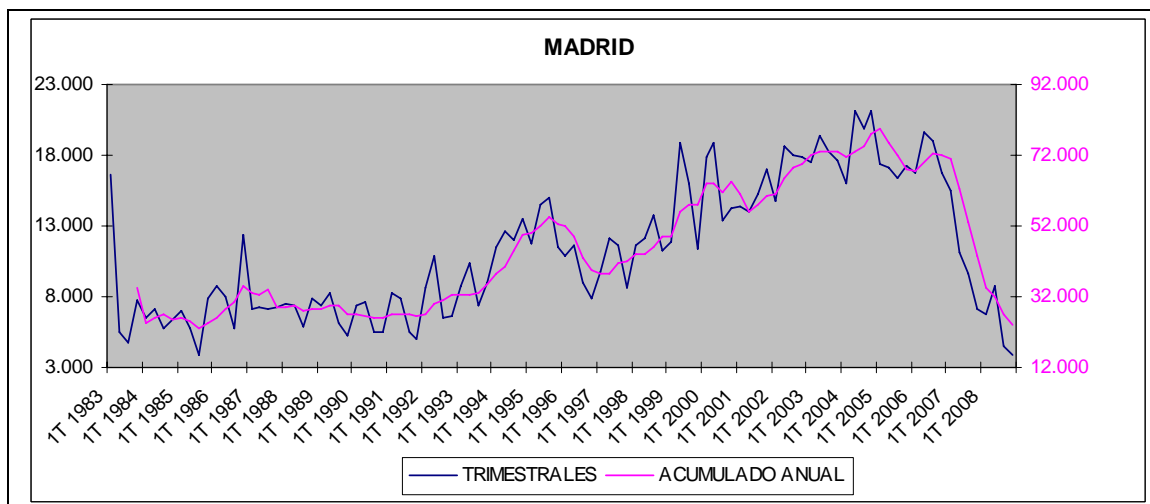


Figura 4.8 Serie trimestral de Visados de Proyectos de Ejecución publicada por el CSCAE y transformación en los valores $V_{t,j}$

Obsérvese (**Fig. 4.9**, *Izqda*) que la serie de visados correspondiente a Sevilla, $V_{t,8}$, está incompleta¹, siendo el último dato válido el correspondiente al segundo trimestre de 2007. Para completar la información de los seis últimos trimestres del periodo de análisis, se ha utilizado la información relativa a su Comunidad Autónoma, Andalucía.

Si $V_{t,j}$ $j=1,\dots,7$, son las series de visados correspondientes a las siete restantes provincias andaluzas, denominemos:

¹ Al menos lo estaba en el momento de la elaboración de este trabajo.

$$V_t^A = \sum_{j=1}^7 V_{t,j} \quad t=1, \dots, T$$

a la serie de visados de Andalucía (salvo Sevilla). Para completar los seis últimos valores de Sevilla se considerará una transformación de V_t^A de la forma:

$$W_t^A = \exp(\ln V_t^A - \alpha^A + \alpha_8) \quad t=1, \dots, T$$

donde:

$$\alpha_8 = \frac{\sum_{t=1}^{T-6} \ln V_{t,8}}{T-6} \quad \text{y} \quad \alpha^A = \frac{\sum_{t=1}^{T-6} \ln V_t^A}{T-6}$$

Así la serie transformada de Andalucía (salvo Sevilla) estará al mismo nivel que la de Sevilla (**Fig. 4.9, Dcha**):

$$\frac{\sum_{t=1}^{T-6} \ln W_t^A}{T-6} = \frac{\sum_{t=1}^{T-6} (\ln V_t^A - \alpha^A + \alpha_8)}{T-6} = \frac{(T-6) \cdot (\alpha_8 - \alpha^A)}{T-6} + \frac{\sum_{t=1}^{T-6} \ln V_t^A}{T-6} = (\alpha_8 - \alpha^A) + \alpha^A = \alpha_8$$

Para completar los seis últimos valores de Sevilla haremos:

$$V_{t,8} = W_t^A \quad t=T-5, \dots, T$$

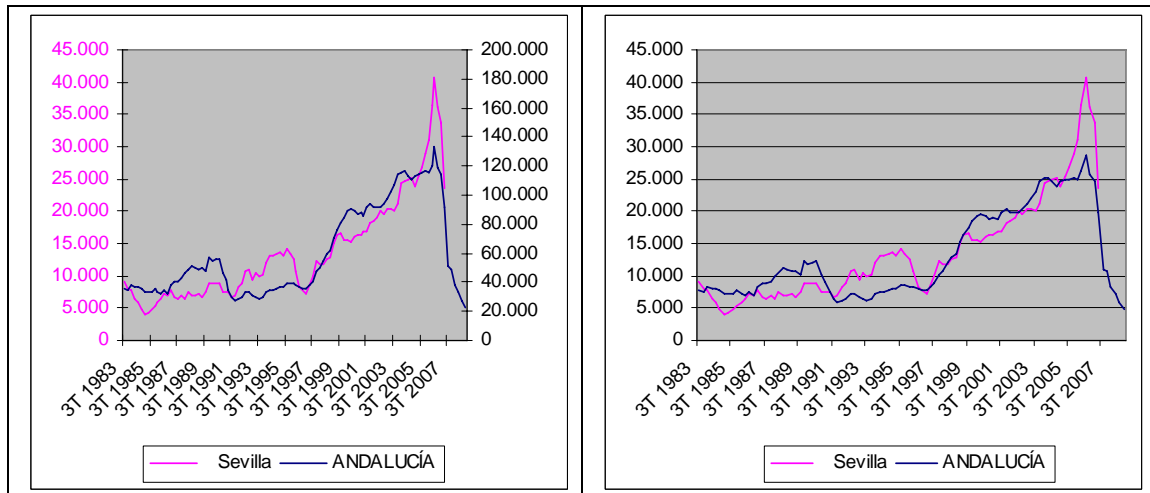


Figura 4.9 Interpolación de los valores de Sevilla en términos de los de Andalucía

Apéndice 2:

La **Fig. 4.10** proporciona una tabla con los valores de $A_{0,j}$ y $A_{1,j}$ para las 50 provincias. Obsérvese que, en general, el coeficiente $A_{0,j}$ toma valores pequeños, pero en la

mayoría de los casos es significativamente distinto de cero. Salvo para Ciudad Real, el coeficiente $A_{1,j}$ es positivo y significativamente distinto de cero, tomando el máximo valor en Almería.

PROVINCIA	$A_{0,j}$	$A_{1,j}$
Almería	-0,012	2,250
Granada	-0,004	1,060
Jaén	-0,006	1,088
Málaga	-0,019	2,242
Cádiz	-0,008	1,489
Córdoba	0,003	0,627
Huelva	-0,002*	1,233
Sevilla	-0,002*	1,236
Huesca	-0,009	1,547
Teruel	0,007	0,518
Zaragoza	-0,001*	0,769
ASTURIAS	-0,003	1,324
BALEARES / BAL	-0,013	1,352
Las Palmas	-0,007	1,555
Sta. Cruz de	-0,001*	1,239
CANTABRIA	-0,001*	1,369
Albacete	0,008	0,582
Ciudad Real	0,024	0,006*
Cuenca	0,010	0,536
Guadalajara	0,010	0,899
Toledo	0,008	0,686
Ávila	0,007	0,515
Burgos	0,002*	0,882
Segovia	0,005	0,624
Soria	0,013	0,485
Valladolid	0,003*	0,791
León	-0,006	1,013
Palencia	-0,001*	0,860
Salamanca	0,004*	0,849
Zamora	-0,005	1,053
Barcelona	0,002*	1,004
Girona	-0,014	1,680
Lleida	0,007	1,037
Tarragona	-0,012	1,829
Cáceres	0,004	0,418
Badajoz	0,009	0,431
La Coruña /	0,006	0,899
Lugo	-0,001*	0,954
Orense / Our	-0,009	1,051
Pontevedra	0,010	0,446
MADRID	0,001*	0,707
MURCIA	-0,004	1,395
LA RIOJA	0,002	0,969
Alicante / A	-0,016	1,820
Castellón /	-0,006	1,373
València	0,000*	0,859
Álava / Arab	0,000*	1,003
Guipúzcoa /	0,009	0,275
Vizcaya / Bi	0,006	0,544
NAVARRA	0,003*	0,629

* No significativo al nivel .05

Figura 4.10 Valores de $A_{0,j}$ y de $A_{1,j}$

Apéndice 3:

Dado que:

$$\begin{aligned} \alpha &= \frac{1}{T} \cdot \sum_{t=1}^T \ln Y_t = \frac{1}{T} \cdot \sum_{t=1}^T \ln V_t - \ln V_1 = \frac{1}{T} \cdot \sum_{t=1}^T \frac{1}{J} \cdot \sum_{j=1}^J \ln V_{t,j} - \ln V_1 = \\ &= \frac{1}{J} \cdot \sum_{j=1}^J \frac{1}{T} \sum_{t=1}^T \ln V_{t,j} - \ln V_1 = \frac{1}{J} \cdot \sum_{j=1}^J \alpha_j - \ln V_1 = \bar{\alpha}_j - \ln V_1 \end{aligned}$$

entonces:

$$\frac{1}{J} \cdot \sum_{j=1}^J \ln Y_{t,j} = \frac{1}{J} \cdot \sum_{j=1}^J (\ln V_{t,j} + \alpha - \alpha_j) = \ln V_t + \alpha - \bar{\alpha}_j = \ln Y_t + \ln V_t + \alpha - \bar{\alpha}_j = \ln Y_t$$

Apéndice 4:

$$\frac{\sum_{t=1}^T \ln Y_{t,j}}{T} = \frac{\sum_{t=1}^T (\ln V_{t,j} + \alpha - \alpha_j)}{T} = \frac{T \cdot (\alpha - \alpha_j)}{T} + \frac{\sum_{t=1}^T \ln V_{t,j}}{T} = (\alpha - \alpha_j) + \alpha_j = \alpha$$

Apéndice 5:

Al superponer en un único gráfico (**Fig. 4.4**, *Izqda*) las series promedio, V_t , y de Asturias, $V_{t,12}$, obsérvese que ambas presentan una evolución muy parecida en términos de los incrementos. Por ejemplo, desde el primer trimestre de 1993 hasta el primero de 2007 el incremento es aproximadamente de:

$$\frac{12.742}{3.878} = \frac{20.298}{6.178} = 3,29$$

mientras que hasta el segundo de 2008 es aproximadamente de:

$$\frac{6.094}{3.878} = \frac{9.708}{6.178} = 1,57$$

lo que implica que, en ambas curvas, entre el primer trimestre de 2007 y el segundo de 2008 se ha producido un decremento de:

$$\frac{6.094}{12.742} = \frac{9.708}{20.298} = 0,48$$

Obsérvese que si, en lugar de considerar las curvas V_t y $V_{t,12}$, consideramos Y_t e $Y_{t,12}$ se simplifica la comparación de sus trayectorias (**Fig. 4.4, Dcha**) al encontrarse en la misma escala de medida.

Por otro lado, al superponer las curvas V_t y $V_{t,12}$ (**Fig. 4.4, Izqda**) la proporción entre los valores de los ejes de escala es de:

$$\frac{23.828}{14.958} = \frac{20.298}{12.742} = \dots = \frac{2.648}{1.662} = 1.59$$

Es decir, a lo largo de su trayectoria los valores de la curva de Asturias son aproximadamente 1.59 veces superiores a los de la curva promedio. Este valor no es más que el resultado de (**Fig. 4.11**):

$$\exp(\alpha_j - \alpha) \cdot \frac{1}{V_1} = \exp(\alpha_{12} - \alpha) \cdot \frac{1}{V_1} = \exp(9,05 - 0,471) \cdot \frac{1}{3,323} = 1,59$$

En otras palabras, los valores de α y α_j nos permiten interpretar la relación entre la curva promedio y la de una provincia cualquiera. En consecuencia, nos permiten interpretar la relación entre las curvas de dos provincias. Por ejemplo, si a lo largo de su trayectoria los valores de la curva de Asturias son aproximadamente 1.59 veces superiores a los de la curva promedio y los de Cantabria 0,94, entonces los de Asturias son aproximadamente 1,69 veces superiores a los de Cantabria ($1,59/0,94=1,69$).

Apéndice 6:

$$\frac{\sum_{t=1}^T C_{t,k}^I}{T} = \frac{\sum_{t=1}^T (g_{t,k} - \beta_k + \alpha)}{T} = \frac{T \cdot (\alpha - \beta_k)}{T} + \frac{\sum_{t=1}^T g_{t,k}}{T} = (\alpha - \beta_k) + \beta_k = \alpha$$

Apéndice 7:

Dado que:

$$d_{k,j}^i = d^2(\ln Y_{t,j}, C_{t,k}^{i-1}) = \sum_{t=1}^T (\ln Y_{t,j} - C_{t,k}^{i-1})^2 = \sum_{t=1}^T (\ln Y_{t,j} - \ln H_{t,k}^{i-1})^2$$

la ventaja de considerar la distancia sobre los valores transformados por la función logaritmo es que cada término de la forma:

$$\ln Y_{t,j} - \ln H_{t,k}^{i-1}$$

es invariante frente a incrementos iguales de las series originales. Es decir, supongamos que:

$$Y_{t+s,j} = \delta \cdot Y_{t,j} \quad \text{y} \quad H_{t+s,k}^{i-1} = \delta \cdot H_{t,k}^{i-1} \quad \text{para } s > 0$$

entonces:

$$\ln Y_{t+s,j} - \ln H_{t+s,k}^{i-1} = \ln Y_{t,j} + \delta - \ln H_{t,k}^{i-1} - \delta = \ln Y_{t,j} - \ln H_{t,k}^{i-1}$$

	α	V_1
PROMEDIO	0,471	3323
PROVINCIA	α_j	$\exp(\alpha_j - \alpha) \cdot \frac{1}{V_1}$
Almería	8,82	1,27
Granada	9,09	1,67
Jaén	8,59	1,01
Málaga	9,91	3,77
Cádiz	9,14	1,76
Córdoba	8,53	0,95
Huelva	8,59	1,01
Sevilla	9,35	2,15
Huesca	7,94	0,53
Teruel	7,03	0,21
Zaragoza	8,81	1,25
ASTURIAS	9,05	1,59
BALEARES / BALEARS	9,17	1,81
Las Palmas	9,13	1,73
Sta. Cruz de Tenerife	9,14	1,74
CANTABRIA	8,52	0,94
Albacete	8,03	0,58
Ciudad Real	8,18	0,67
Cuenca	7,31	0,28
Guadalajara	7,92	0,52
Toledo	8,99	1,51
Ávila	7,62	0,38
Burgos	8,13	0,64
Segovia	7,62	0,38
Soria	6,66	0,15
Valladolid	8,47	0,90
León	8,43	0,86
Palencia	7,23	0,26
Salamanca	8,00	0,56
Zamora	7,39	0,31
Barcelona	10,36	5,93
Girona	9,19	1,83
Lleida	8,30	0,75
Tarragona	9,27	1,99
Cáceres	8,11	0,62
Badajoz	8,50	0,93
La Coruña / A Coruña	9,11	1,70
Lugo	8,12	0,63
Orense / Ourense	7,74	0,43
Pontevedra	8,83	1,28
MADRID	10,65	7,89
MURCIA	9,70	3,08
LA RIOJA	8,11	0,63
Alicante / Alacant	10,15	4,80
Castellón / Castelló	9,10	1,68
València	9,85	3,57
Álava / Araba	7,76	0,44
Guipúzcoa / Gipuzkoa	8,31	0,77
Vizcaya / Bizkaia	8,44	0,87
NAVARRA	8,62	1,04

Figura 4.11 Interpretación de los valores α y α_j

Apéndice 8:

Tenemos K series de la forma:

$$C_{t,k}^I = g_{t,k} - \beta_k + \alpha \quad k=1,\dots,K$$

donde:

$$g_{t,k} = g_{t-1,k} + f_{t,k}, \quad t=2,\dots,T \quad \text{y} \quad g_{1,k} = f_{1,k}, \quad k=1,\dots,K$$

siendo:

$$f_{t,k} = b_k + m_k \cdot \ln y_t, \quad t=2,\dots,T, \quad f_{1,k} = b_k \quad \text{con} \quad b_k = B_0 + B_1 \cdot m_k, \quad k=1,\dots,K$$

Además:

$$\frac{\sum_{t=1}^T C_{t,k}^I}{T} = \alpha$$

En otras palabras:

$$C_{1,k}^I = g_{1,k} - \beta_k + \alpha = f_{1,k} - \beta_k + \alpha$$

$$C_{2,k}^I = g_{2,k} - \beta_k + \alpha = g_{1,k} + f_{2,k} - \beta_k + \alpha = f_{1,k} + f_{2,k} - \beta_k + \alpha$$

$$C_{3,k}^I = g_{3,k} - \beta_k + \alpha = g_{2,k} + f_{3,k} - \beta_k + \alpha = f_{1,k} + f_{2,k} + f_{3,k} - \beta_k + \alpha$$

.....

$$C_{t,k}^I = g_{t,k} - \beta_k + \alpha = g_{t-1,k} + f_{t,k} - \beta_k + \alpha = f_{1,k} + \dots + f_{t,k} - \beta_k + \alpha = \sum_{s=1}^t f_{s,k} - \beta_k + \alpha$$

En consecuencia, en un instante t cualquiera:

$$\begin{aligned} C_{t,k}^I &= \sum_{s=1}^t (b_k + m_k \cdot \ln y_s) - \beta_k + \alpha = b_k \cdot t + m_k \cdot \sum_{s=2}^t \ln y_s - \beta_k + \alpha = \\ &= b_k \cdot t + m_k \cdot (\ln Y_t - \ln Y_1) - \beta_k + \alpha = b_k \cdot t + m_k \cdot \ln Y_t - \beta_k + \alpha \quad ^2 \end{aligned}$$

Si denominamos:

$$\mu_k = -\beta_k + \alpha$$

entonces disponemos de un conjunto de K series temporales $\{C_{t,k}^I\}$, $k=1,\dots,K$, tales que:

² Recuérdese que, por definición, el valor de Y_t en el primer instante es igual a 1.

$$\frac{\sum_{t=1}^T C_{t,k}^I}{T} = \alpha \quad \forall k$$

y para las que existe otra serie temporal, $\ln Y_t$, tal que:

$$C_{t,k}^I = b_k \cdot t + m_k \cdot \ln Y_t + \mu_k$$

donde los coeficientes b_k y m_k son tales que:

$$b_k = B_0 + B_1 \cdot m_k$$

En otras palabras, se verifican las hipótesis [3.1], [3.2] y [3.3] del **Lema 1³** para la metodología propuesta. En consecuencia, si:

$$m_q < m_r < m_s$$

entonces:

$$d(\ln H_{t,q}, \ln H_{t,r}) < d(\ln H_{t,q}, \ln H_{t,s})$$

o, lo que es equivalente:

$$d(H_{t,q}, H_{t,r}) < d(H_{t,q}, H_{t,s})$$

Apéndice 9:

En el **Apéndice 8** hemos comprobado que el conjunto de K series temporales $\{ C_{t,k}^I \} = \{ \ln H_{t,k}^I \}$ verifica las hipótesis [3.1], [3.2] y [3.3] del **Lema 1⁴** para la metodología propuesta. En consecuencia, por la **Proposición 4**, si denominamos:

$$\ln h_{t,k} = \nabla \ln H_{t,k} = \ln H_{t,k} - \ln H_{t-1,k} \quad \text{y} \quad \ln y_t = \nabla \ln Y_t = \ln Y_t - \ln Y_{t-1}$$

entonces las series temporales $\ln h_{t,k}$ $k=1, \dots, K$ definen, respecto de $\ln y_t$, un haz de K rectas de vértice:

$$(\ln y_t, \ln h_{t,k}) = (-B_1, B_0).$$

o, lo que es equivalente, las series temporales $h_{t,k}$ $k=1, \dots, K$ definen, respecto de y_t , un haz de K curvas de vértice:

$$(y_t, h_{t,k}) = (\exp(-B_1^F), \exp(B_0^F))$$

³ Véase el Capítulo 3.

⁴ Véase el Capítulo 3.

Apéndice 10:

En el **Apéndice 8** hemos comprobado que el conjunto de K series temporales $\{C_{t,k}^I\} = \{\ln H_{t,k}^I\}$ verifica las hipótesis [3.1], [3.2] y [3.3] del **Lema 1**⁵ para la metodología propuesta. En consecuencia, por la **Proposición 7**, dado un par cualquiera de series de la forma $C_{t,k}^I$ existe al menos un punto en su trayectoria en el que las series se cruzan. Además, en aquellos puntos tales que dos de las series se cruzan, también lo hacen las restantes. Considerando que:

$$H_{t,k}^I = \exp C_{t,k}^I$$

en aquellos instantes en los que $C_{t,k}^I = C_{t,k'}^I$, entonces también $H_{t,k}^I = H_{t,k'}^I$.

Apéndice 11:

En el **Apéndice 8** hemos comprobado que el conjunto de K series temporales $\{C_{t,k}^I\} = \{\ln H_{t,k}^I\}$ verifica las hipótesis [3.1], [3.2] y [3.3] del **Lema 1**⁶ para la metodología propuesta. En consecuencia, por la **Proposición 8**, la diferencia entre dos puntos de cruce cualesquiera es independiente de la media de las series.

⁵ Véase el Capítulo 3.

⁶ Véase el Capítulo 3.

CAPÍTULO 5

EXTRACCIÓN DEL CONJUNTO DE SERIES RESUMEN A PARTIR DE UNA SERIE COVARIABLE

El desarrollo de la metodología se fundamenta en la existencia de una serie temporal responsable de las fluctuaciones del conjunto de series objeto de análisis. Tal como hemos visto en el capítulo anterior, para aplicar la metodología dicha serie no tiene porqué ser conocida, sino que hemos considerado que su efecto se manifiesta a través de la serie promedio. Así, la construcción de los centroides iniciales y, en consecuencia, la agrupación de las series objeto de comparación se ha realizado a partir de la serie promedio. Para interpretar la agrupación obtenida hemos comprobado que sería de gran utilidad, además de la representación de cada serie junto con las restantes series clasificadas en su mismo grupo, la representación en la propia solución de centroides iniciales. Si la serie temporal responsable de las oscilaciones del conjunto de series objeto de análisis fuera conocida, el proceso de extracción del conjunto de centroides iniciales con base en dicha serie sería el mismo que el seguido con la serie promedio. El conjunto de series resumen así obtenido aportaría información adicional a efectos de la interpretación de la evolución de cada serie y, en consecuencia, de la relación entre las distintas series. Con este objetivo, en la extracción del conjunto de series resumen a partir de una serie covariable supondremos que las oscilaciones de las series de visados responden, con distintos grados de sensibilidad, a las fluctuaciones del tipo de interés.

5.1 RELACIÓN ENTRE VISADOS Y TIPO DE INTERÉS

Como paso previo a la extracción del conjunto de series resumen con base en la serie del tipo de interés analicemos, en términos de los visados, la relación entre la evolución del sector de la construcción residencial y el tipo de interés.

Sea el conjunto de las J series temporales:

$$\{V_{t,j}\} \quad j=1,\dots, 50$$

relativas a los visados de proyectos de ejecución en el trimestre $t=1,\dots,102$ (desde el tercer trimestre de 1983 hasta el cuarto de 2008, ambos inclusive) en cada una de las cincuenta provincias españolas (**Fig. 4.1**). Sea la serie promedio en términos del logaritmo:

$$\ln V_t = \frac{\sum_{j=1}^J \ln V_{t,j}}{J}$$

Si denominamos:

$$Y_t = \frac{V_t}{V_1}$$

la relación que se propone tiene su origen en la información proporcionada por la representación gráfica de las series Y_t e I_t (**Fig. 5.1, Sup. Izqda.**), donde I_t es la serie del *Tipo de interés*¹. En términos generales, a medida que disminuye el tipo de interés aumenta el promedio de visados. La relación entre estas series se hace más evidente al analizar sus tasas de crecimiento. Si denominamos:

$$y_t = \frac{Y_t}{Y_{t-1}} \quad \text{e} \quad i_t = \frac{I_t}{I_{t-1}}$$

la representación gráfica de ambas tasas (**Fig. 5.1, Sup. Dcha.**) muestra cierta simetría en sus trayectorias que, parece, sería más evidente si se desfasara la curva del tipo de interés. Concretamente, si consideramos la función de correlación cruzada entre las series y_t e i_t (**Fig. 5.1, Inf. Izqda**), la máxima correlación entre la serie y_t y una desfasada de i_t se obtiene para $i_{t-s} = i_{t-4}$. Efectivamente, al considerar la representación gráfica de y_t e i_{t-4} (**Fig. 5.1, Inf. Dcha**) ahora sí es evidente la correlación inversa entre las fluctuaciones en visados y tipos de interés.

Concretamente, si denominamos:

$$X_t = I_{t-4} \quad \text{y} \quad x_t = \frac{X_t}{X_{t-1}}$$

¹ Véase el Apéndice 1 para la descripción de la serie I_t .

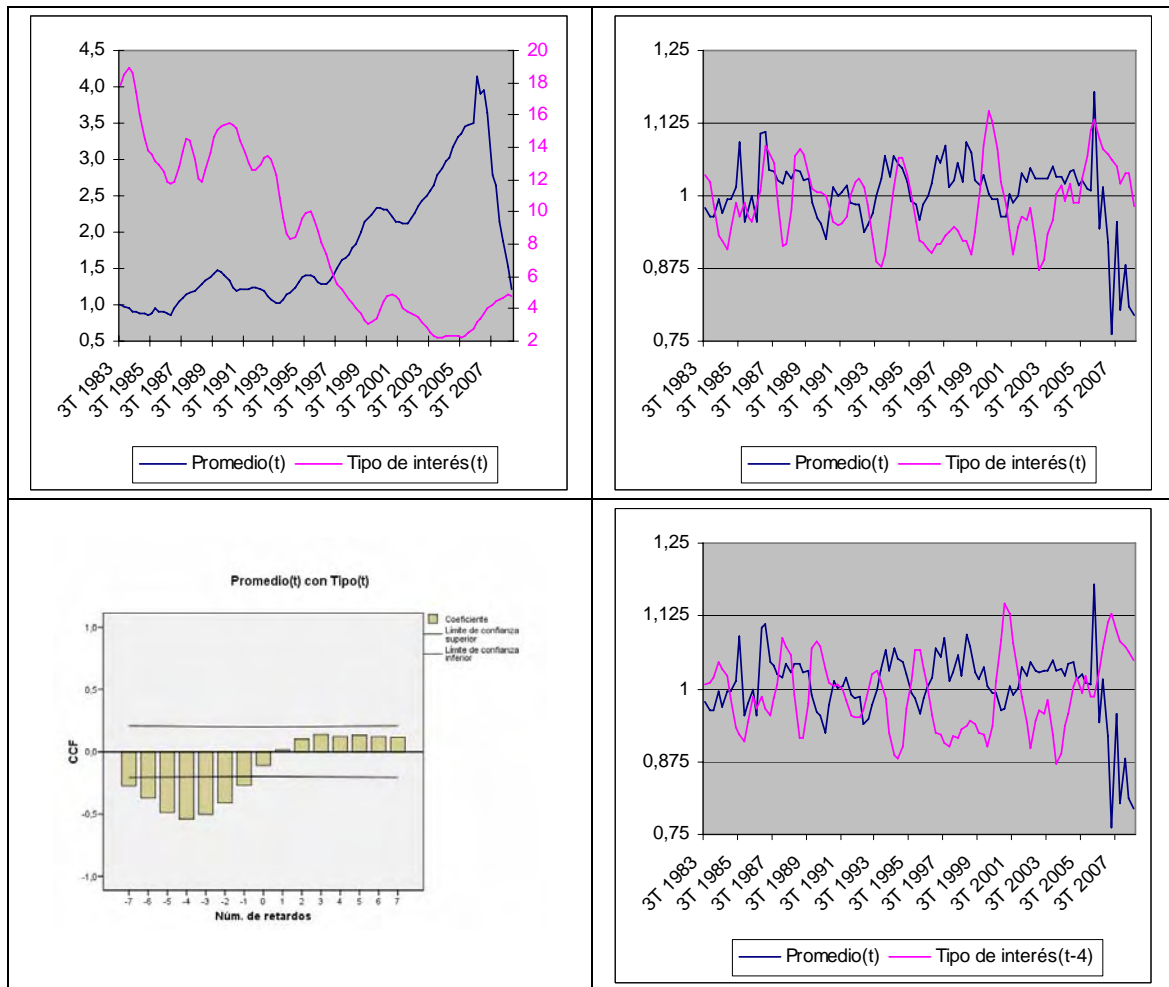


Figura 5.1 *Sup. izqda:* Series Y_t e I_t ; *Sup. dcha:* Series y_t e i_t ; *Inf. izqda:* Función de correlación cruzada entre y_t e i_t ; *Inf. dcha:* Series y_t e i_{t-4}

para interpretar la sensibilidad de sector de la construcción residencial frente al tipo de interés se propone el siguiente modelo de regresión lineal:

$$\ln y_t = \alpha_1 \cdot \ln x_t + \alpha_2 + a_t \tag{5.1}$$

aunque a efectos de la estimación de los parámetros consideraremos el modelo:

$$\ln Y_t = \alpha_0 \cdot t + \alpha_1 \cdot \ln X_t + \alpha_2 + a_t \tag{5.2}$$

donde α_0 , α_1 y α_2 son parámetros desconocidos a estimar, y a_t es la componente aleatoria. Concretamente, el ajuste de los valores de Y_t en términos del tipo de interés según la ecuación [5.2] vendrá dado por:

$$\hat{Y}_t = \exp(A_0 \cdot t + A_1 \cdot \ln X_t + A_2)$$

donde A_0 , A_1 y A_2 son las estimaciones de α_0 , α_1 y α_2 por MCO. Así, el ajuste de los valores de y_t en términos del tipo de interés según la ecuación [5.1] vendrá dado por:

$$\hat{y}_t = \frac{\hat{Y}_t}{\hat{Y}_{t-1}} = \exp(\ln \hat{Y}_t - \ln \hat{Y}_{t-1}) = \exp(A_0 + A_1 \cdot (\ln X_t - \ln X_{t-1})) = \exp(A_0 + A_1 \cdot \ln x_t)$$

Concretamente (**Fig. 5.2**):

$$\hat{y}_t = e^{A_0} \cdot x_t^{A_1} = e^{-0,0000542} \cdot x_t^{-0,595} \tag{5.3}$$

Al nivel de significación α de 0,05, podemos concluir que, aunque el coeficiente A_0 no es significativamente distinto de cero (el p -valor asociado al estadístico de contraste es mayor que el nivel de significación α), el coeficiente A_1 , que al fin y al cabo es el que nos interesa, sí lo es. Interpretemos el ajuste.

Si suponemos una disminución del tipo de interés del 20 por ciento o, lo que es equivalente, un valor de $x_t = i_{t-4}$ de 0,8, entonces, según la Ecuación [5.3], el valor estimado del incremento en visados vendría dado por:

$$\hat{y}_t = \frac{1}{e^{0,0000542} \cdot x_t^{0,595}} = \frac{1}{e^{0,0000542} \cdot 0,8^{0,595}} = 1,14$$

Dicho valor podría haberse obtenido directamente desde la representación gráfica de la **Fig. 5.3**. Consistiría en el punto definido por el corte de la vertical a la altura del 0,8 con la curva de \hat{y}_t .

En definitiva, a pesar de su parsimonia, la estimación del modelo [5.2] reproduce con bastante precisión la trayectoria de la serie Y_t (**Fig. 5.4, Izqda**), medida resumen de las fluctuaciones del sector de la construcción residencial en las cincuenta provincias españolas, por lo que la serie X_t relativa al tipo de interés puede ser una buena alternativa a la serie promedio para la extracción del haz de rectas. Equivalentemente, la estimación del modelo [5.1] reproduce con bastante precisión la trayectoria de la serie y_t (**Fig. 5.4, Dcha**).

Coeficientes(a)						
Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	1,660	,163		10,207	,000
	Tipo	-,595	,051	-,955	-11,572	,000
	T	-5,42E-005	,001	-,004	-,045	,964

a Variable dependiente: Promedio

Figura 5.2 Valores de los coeficientes A_0 , A_1 y A_2

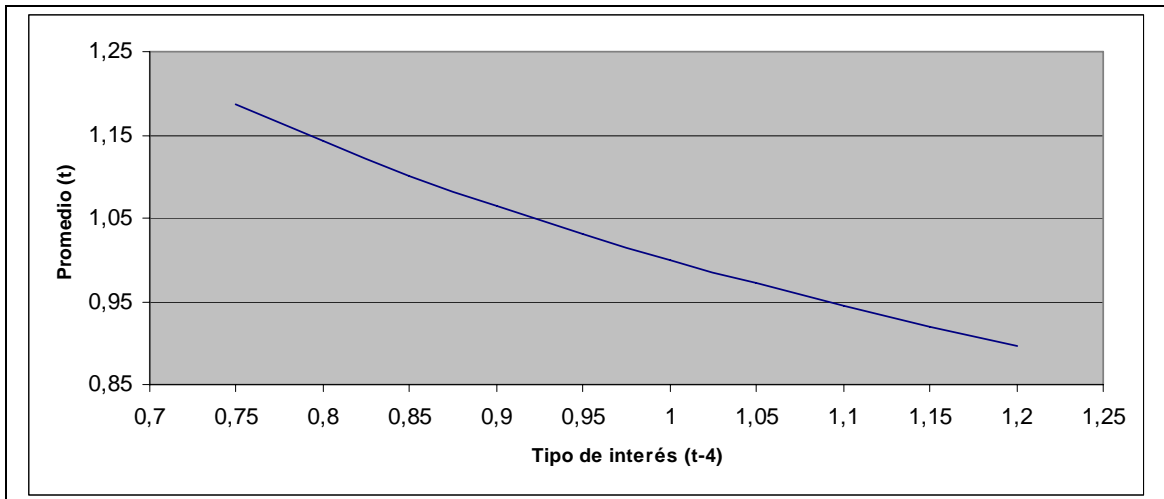


Figura 5.3 Valores de \hat{y}_t frente a los de $x_t = i_{t-4}$

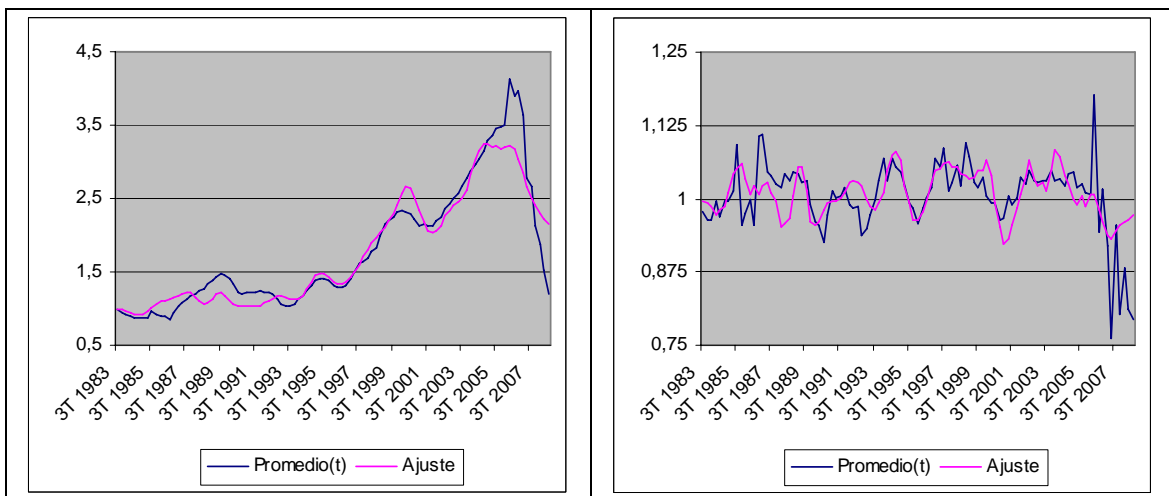


Figura 5.4 Izqda: Series Y_t e \hat{Y}_t ; Dcha: Series y_t e \hat{y}_t

5.2 EXTRACCIÓN DEL CONJUNTO DE SERIES RESUMEN

Para la extracción del conjunto de series resumen a partir de la serie del tipo de interés seguiremos el esquema propuesto en el Anexo I.

5.2.1 Datos de entrada

- Las 50 series de visados observadas en 102 instantes de tiempo:

$$\{V_{t,j}, t=1,\dots,102\} \quad j=1,\dots,50$$

- La serie temporal promedio con referencia en el instante inicial:

$$Y_t = \frac{V_t}{V_1} \quad t=1,\dots,T$$

donde:

$$\ln V_t = \frac{\sum_{j=1}^J \ln V_{t,j}}{J} \quad t=1,\dots,T$$

- La serie relativa al tipo de interés, X_t , que se supone genera las fluctuaciones de las distintas series de visados, $V_{t,j}$.

5.2.2 Condiciones de aplicación

Aplicaremos la metodología, cuando podamos suponer que la estructura que subyace en el conjunto de series temporales $\{V_{t,j}\}$ es la de haz de un rectas.

- Para $j=1,\dots,J$, calcularemos $A_{0,j}$, $A_{1,j}$ y $A_{2,j}$, coeficientes de la ecuación de regresión:

$$\ln \hat{V}_{t,j} = A_{0,j} \cdot t + A_{1,j} \cdot \ln X_t + A_{2,j} \quad [5.4]$$

estimada por MCO.

- Para $j=1,\dots,J$, calcularemos B_0 y B_1 , coeficientes de la ecuación de regresión:

$$\hat{A}_{0,j} = B_0 + B_1 \cdot A_{1,j} \quad [5.5]$$

estimada por MCO.

Será adecuado aplicar la metodología si:

- los coeficientes $A_{1,j}$ son significativamente distintos de cero aunque, excepcionalmente, alguno de ellos sea nulo; y
- los J puntos de la forma $(A_{1,j}, A_{0,j})$ se concentran en torno a una recta.

Por ejemplo, obsérvese (**Fig. 5.5**) que la serie $V_{t,12}$, correspondiente a Asturias, presenta una estructura muy simétrica respecto de la relativa al tipo de interés, X_t . El ajuste de los valores de $V_{t,12}$ en términos del tipo de interés vendrá dado por la expresión [5.4] para el caso particular de j igual a 12:

$$\ln \hat{V}_{t,12} = A_{0,12} \cdot t + A_{1,12} \cdot \ln X_t + A_{2,12}$$

donde:

$$A_{0,12} = -0,003, \quad A_{1,12} = -0,767 \quad \text{y} \quad A_{2,12} = 10,705$$

Al nivel de significación α de 0,05, podemos concluir que el coeficiente $A_{1,12}$, que al fin y al cabo es el que nos interesa, es significativamente distinto de cero, aunque el coeficiente $A_{0,12}$ no lo sea²: el p -valor asociado al estadístico de contraste es menor que el nivel de significación α en el primer caso y mayor en el segundo.

Una vez calculados los J valores de $A_{0,j}$ y de $A_{1,j}$, el siguiente paso consistirá en comprobar que están asociados linealmente. Efectivamente (**Fig. 5.6**), parece que cuanto mayor es el valor de $A_{1,j}$ mayor es el de $A_{0,j}$. Esta percepción se confirma al ajustar la ecuación de regresión [5.5]:

$$\hat{A}_{0,j} = B_0 + B_1 \cdot A_{1,j} = 0,015 + 0,025 \cdot A_{1,j}$$

Al nivel de significación α de 0,05, podemos concluir que, no sólo el coeficiente B_1 , sino también B_0 es significativamente distinto de cero³: el p -valor asociado al estadístico de contraste es menor que el nivel de significación α en ambos casos.

² El Apéndice 2 proporciona información relativa a los valores de estos dos coeficientes para las 50 provincias.

³ La pendiente de la recta puede ser nula; si así fuera, y los puntos se concentraran en torno a la recta, significaría que la componente determinística podría ser ignorada.

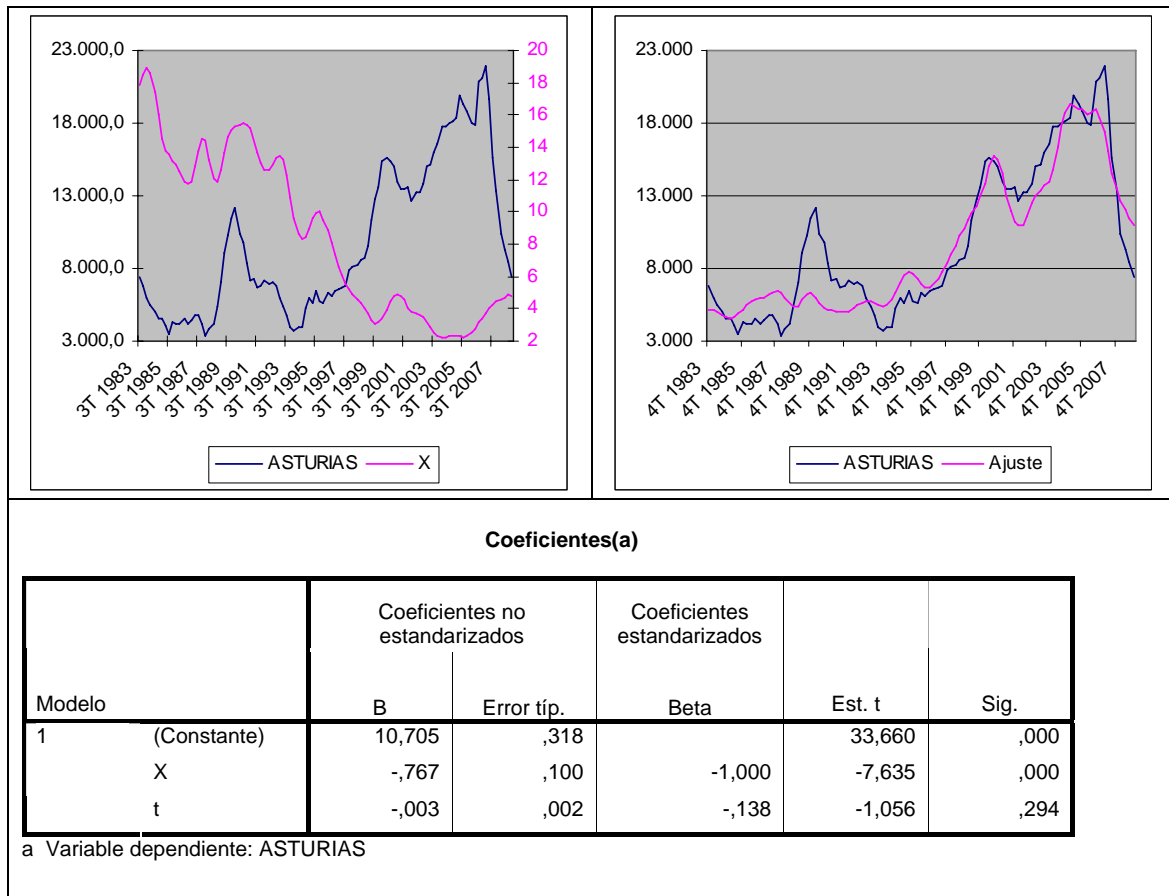


Figura 5.5 Sup. izqda: Series $V_{t,12}$ y X_t ; Sup. dcha: Series $V_{t,12}$ y $\hat{V}_{t,12}$; Centro: Valores de los coeficientes $A_{0,12}$, $A_{1,12}$ y $A_{2,12}$

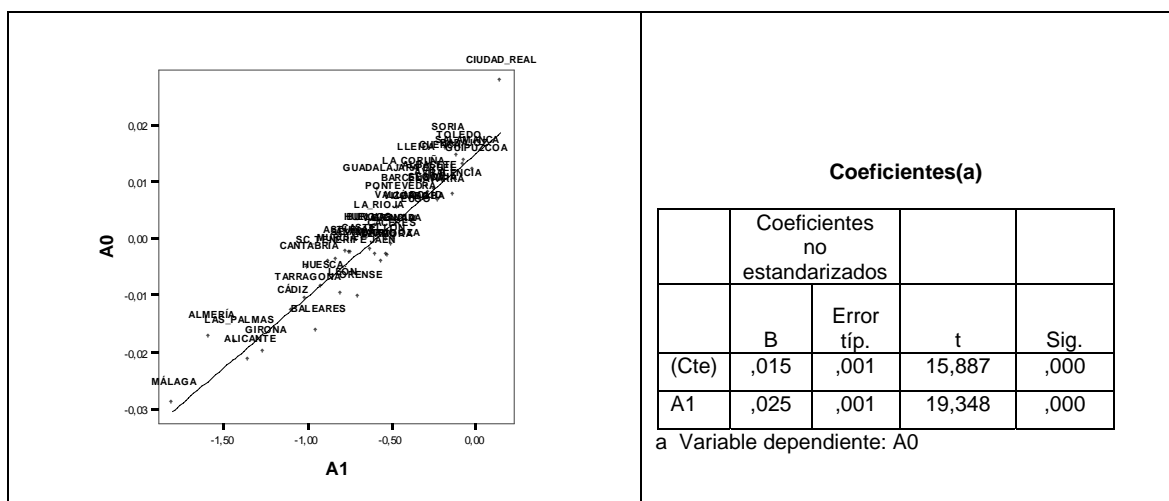


Figura 5.6 Izqda: Valores de $A_{0,j}$ frente a valores de $A_{1,j}$; Dcha: Valores de los coeficientes B_0 y B_1

5.2.3 Homogeneización de la escala

Antes de proceder a la extracción del conjunto de series resumen, es necesario situar las 50 series al mismo nivel; concretamente, el nivel considerado es el de la serie temporal promedio con referencia en el instante inicial y transformado por la función logaritmo, $\ln Y_t$. Aunque la homogeneización es independiente de la serie X_t , recordemos que consistía en:

- Calcular el promedio de la serie $\ln Y_t$:

$$\alpha = \frac{\sum_{t=1}^T \ln Y_t}{T}$$

- Calcular los promedios de cada una de las series $\ln V_{t,j}$:

$$\alpha_j = \frac{\sum_{t=1}^T \ln V_{t,j}}{T} \quad j=1,\dots,J$$

- Transformar las J series temporales $V_{t,j}$ en J nuevas series, $\ln Y_{t,j}$, todas ellas al mismo nivel que $\ln Y_t$:

$$\ln Y_{t,j} = \ln V_{t,j} - \alpha_j + \alpha \quad j=1,\dots,J$$

5.2.4 Generación del conjunto de series resumen

El proceso de construcción de las series resumen $C_{t,k}$, $k=1,\dots,K$ a partir de la serie del tipo de interés se realizará en las siguientes etapas (**Fig. 5.7**).

- En primer lugar fijaremos K valores dispersos, m_k , dentro del rango de variación de $A_{1,j}$ y consideraremos:

$$b_k = B_0 + B_1 \cdot m_k = 0,015 + 0,025 \cdot m_k \quad k=1,\dots,K$$

Por ejemplo, en términos del rango de variación de $A_{1,j}$ (**Fig. 5.6**), fijaremos $K=6$ coeficientes de la forma:

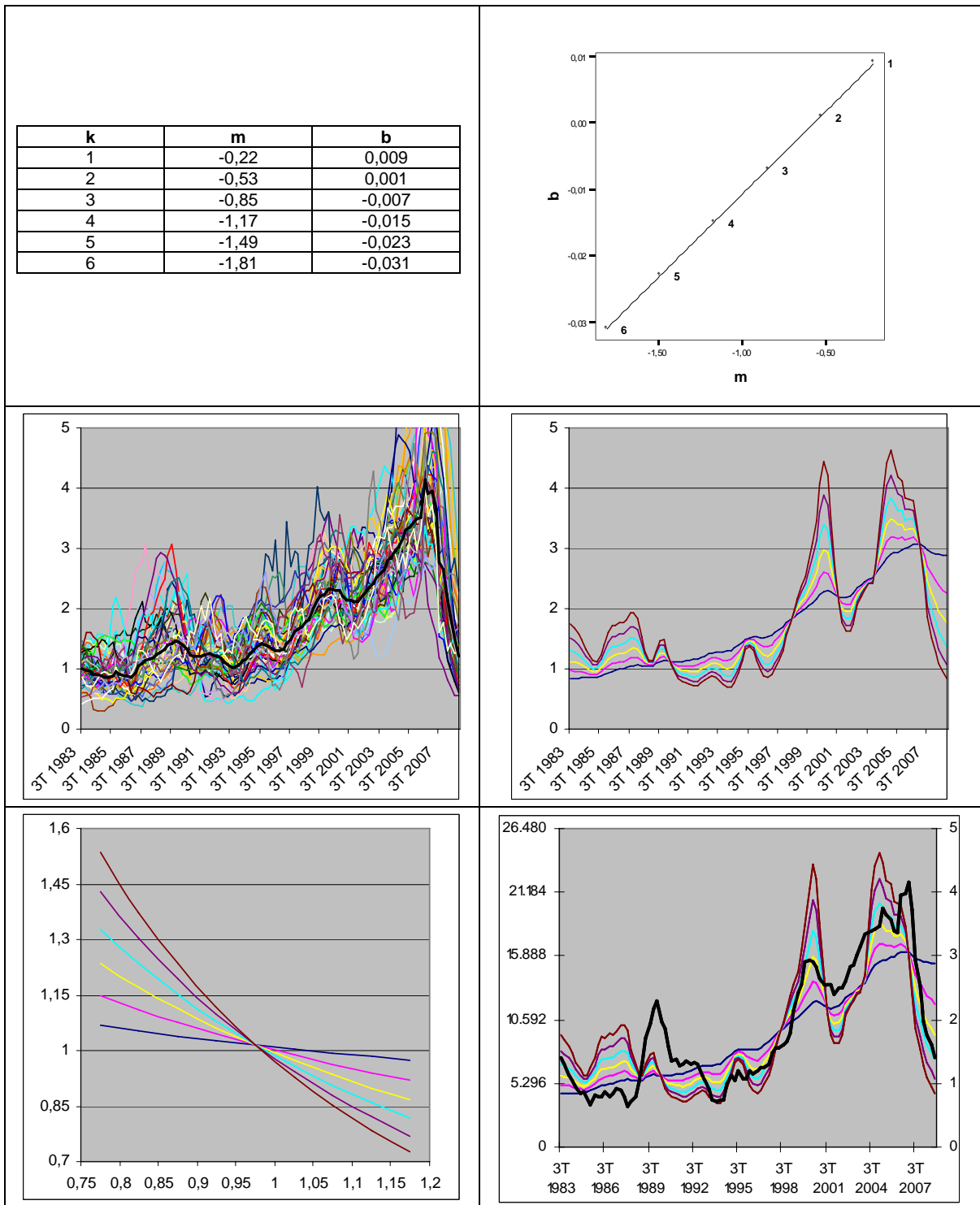


Figura 5.7 *Sup. izqda:* Tabla de los valores de m_k y de b_k ; *Sup. dcha:* Valores de b_k frente a los de m_k ; *Ctro. izqda:* Series $Y_{t,j}$; *Ctro. dcha:* Series $H_{t,k}$; *Inf. Izqda:* Valores de $h_{t,k}$ frente a los de y_t ; *Inf. Dcha:* Series $H_{t,k}^j$ y $V_{t,j}$ (eje izquierdo) y series $H_{t,k}$ e $Y_{t,j}$ (eje derecho), para $j=12$ (Asturias)

$$m_1 = \max_j A_{1,j} = -0,216^4 \quad \text{y} \quad m_k = m_{k-1} - \theta \quad \text{para} \quad k=2, \dots, 6$$

siendo:

$$\theta = (\max_j A_{1,j} - \min_j A_{1,j}) / (K - 1) = 0,318$$

- A continuación calcularemos los valores de $f_{t,k}$, $k=1, \dots, K$:

$$f_{t,k} = b_k + m_k \cdot \ln x_t, \quad t=2, \dots, T \quad \text{y} \quad f_{1,k} = b_k$$

donde:

$$x_t = \frac{X_t}{X_{t-1}} \quad t=2, \dots, T$$

- Seguidamente, calcularemos los valores de $g_{t,k}$, $k=1, \dots, K$:

$$g_{t,k} = g_{t-1,k} + f_{t,k}, \quad t=2, \dots, T \quad \text{y} \quad g_{1,k} = f_{1,k}$$

- Para finalizar el proceso de construcción del conjunto de K series resumen, $\{C_{t,k}\}$, bastará con homogeneizar los valores de la funciones $g_{t,k}$ en términos de la escala de las series transformadas $\ln Y_{t,j}$:

$$C_{t,k} = g_{t,k} - \beta_k + \alpha \quad t=1, \dots, T \quad k=1, \dots, K:$$

donde:

$$\beta_k = \frac{\sum_{t=1}^T g_{t,k}}{T} \quad \forall k$$

Así⁵:

$$\frac{\sum_{t=1}^T C_{t,k}}{T} = \alpha \quad \forall k$$

⁴ Este es el máximo valor significativamente distinto de cero (véase el Apéndice 2). En este caso, a diferencia de la metodología aplicada sobre la serie promedio, comenzaremos por el máximo, para así, dada la correlación inversa entre el promedio y el tipo de interés, mantener la concordancia entre las series resumen de ambas soluciones.

⁵ Véase el Apéndice 6 del Capítulo 4 para la comprobación de este resultado.

En definitiva, a partir de la serie X_t hemos construido un conjunto de K series $\{C_{t,k}\}$ al mismo nivel que el conjunto de J series $\{\ln Y_{t,j}\}$. Alternativamente, las K series de la forma:

$$H_{t,k} = \exp C_{t,k}$$

se encuentran en la misma escala que las J series $Y_{t,j}$ (**Fig. 5.7**, en los paneles centrales).

5.3 INTERPRETACIÓN DE LA SOLUCIÓN

Obsérvese (**Fig. 5.7, Ctro. Dcha**) que la sexta serie (línea marrón) es la que presenta mayores fluctuaciones y que este efecto se va suavizando hacia la primera. Es más, las series $H_{t,k}$ están ordenadas, en el sentido de que si:

$$m_q < m_r < m_s$$

entonces⁶:

$$d(H_{t,q}, H_{t,r}) < d(H_{t,q}, H_{t,s})$$

Este orden responde en definitiva a su sensibilidad frente a la serie del tipo de interés. Concretamente, el conjunto de K series:

$$h_{t,k} = \Delta H_{t,k} = \frac{H_{t,k}}{H_{t-1,k}} = \exp(b_k + m_k \cdot \ln x_t) \quad k=1, \dots, K$$

es tal que define un haz de curvas (**Fig. 5.7, Inf. Izqda**) concurrentes en el punto⁷:

$$(\exp(-B_1), \exp B_0) = (e^{-0.025}, e^{0.015}) = (0.975, 1.015)$$

En dicho haz, en un extremo está la sexta curva (línea marrón), la de mayor sensibilidad frente a la serie del tipo de interés y en el extremo opuesto (línea azul oscuro), la primera, que se desvía poco de la línea horizontal a la altura del valor 1.

Obsérvese además (**Fig. 5.7, Ctro. Dcha**) que en aquellos puntos en los que dos de las series de la forma $H_{t,k}$ se cruzan, también lo hacen las restantes⁸. Además, si en dos instantes t y t' :

⁶ Véase el Apéndice 3 para la justificación de esta afirmación.

⁷ Véase el Apéndice 4 para la justificación de esta afirmación.

⁸ En el Apéndice 5 se proporciona una justificación de este resultado.

$$H_{t,k} = H_{t,k'} \quad \text{y} \quad H_{t',k} = H_{t',k'} \quad \forall k \neq k'$$

entonces:

$$\frac{H_{t',k}}{H_{t,k}} = \frac{H_{t',k'}}{H_{t,k'}}$$

En otras palabras, tal como sucedía en el caso de la extracción del conjunto de series resumen a partir de la serie promedio, a lo largo de las trayectorias de las K series $H_{t,k}$ existen diversos instantes (los correspondientes a los sucesivos puntos de cruce) en los que el incremento entre cualquier par de ellos es el mismo para todas las series. Considerando además que la relación entre los puntos de corte no depende de la escala en que representemos el conjunto de series resumen⁹, también ahora podríamos utilizar la relación entre los distintos puntos de cruce para interpretar la evolución conjunta de las J series temporales observadas.

En definitiva, si suponemos que la menor o mayor volatilidad de las oscilaciones de las J series es consecuencia de la menor o mayor sensibilidad frente a las fluctuaciones de un determinado factor conocido, en nuestro caso, el tipo de interés, para resumir el conjunto de J series temporales podemos buscar una nueva solución de K series basada, en lugar de en los valores de la serie promedio, en la trayectoria de dicho factor. Para comparar las trayectorias de las J series temporales del conjunto de análisis en términos del tipo de interés, bastará con poner el conjunto de series resumen a su misma escala y compararlas.

Por ejemplo, si consideramos de nuevo la serie de visados de la construcción correspondiente a Asturias, $V_{t,12}$, para interpretar su trayectoria transformaremos el conjunto de series resumen de tal forma que conserven la misma forma pero estén en la misma escala que Asturias; concretamente, para situar las K curvas $H_{t,k}$ al nivel de la curva $V_{t,j}$ haremos:

$$H_{t,k}^j = \exp(\ln H_{t,k} - \gamma_k + \alpha_j) \quad k=1,\dots,K \quad [5.5]$$

donde:

$$\gamma_k = \frac{\sum_{t=1}^T \ln H_{t,k}}{T}$$

y, en particular, para situarlas al nivel de Asturias:

$$H_{t,k}^{12} = \exp(\ln H_{t,k} - \gamma_k + \alpha_{12}) \quad k=1,\dots,K$$

⁹ Véase el Apéndice 6 para la justificación de esta afirmación.

Obsérvese (**Fig. 5.7**, *Inf. Dcha*) que, aunque bien es cierto que en la etapa inicial -hasta el año 1994 aproximadamente- la curva de Asturias (línea resaltada en negrita) no fluctúa en consonancia con el conjunto de series resumen, durante el resto del periodo se aproxima bastante a la series de mayor sensibilidad frente al tipo de interés. En otras palabras podemos suponer que, salvo en lo que respecta a la primera etapa del periodo de observación (desde 1983 hasta 1993), las oscilaciones observadas en la evolución del sector de la construcción residencial en Asturias se producen como consecuencia de su sensibilidad frente al tipo de interés.

De la misma forma que hemos representado la trayectoria de Asturias en términos del conjunto de series resumen, podríamos representar la de cualquier otra provincia sin más que modificar la escala de dicho conjunto según la expresión [5.5]¹⁰.

¹⁰ Véase el Capítulo 6 para la interpretación de la trayectoria de cada provincia.

APÉNDICES

Apéndice 1:

Para la creación de la serie *Tipo de interés*, aunque en la etapa más reciente (desde el primer trimestre de 1999) consideraremos el principal indicador utilizado en España para fijar el tipo de interés de las hipotecas, el Euribor a un año, las observaciones iniciales corresponden al Mibor a un año¹. A pesar de que partían de valores muy distantes, a lo largo de la década de los noventa, y como consecuencia del proceso de convergencia al euro de la economía española, ambas series fueron aproximándose hasta prácticamente converger: desde principios del año 1999 las diferencias entre los valores de ambas series son insignificantes (**Fig. 5.8, Izqda**).

Recordemos que, en cada instante-trimestre, el valor de las series de visados provinciales, $V_{t,j}$, consistía en el total de visados otorgados en el plazo de los 12 últimos meses. Obsérvese que el efecto de este suavizado de las series originales de visados, $Z_{t,j}$, implica cierto rezago: en la **Fig. 4.8** la curva roja, correspondiente a la suavizada, se rezaga respecto de la curva azul, correspondiente a la original. Con el objetivo de contemplar este efecto a la hora de calcular la correlación cruzada con las series de visados, la serie del tipo de interés que utilizaremos a efectos del análisis será la serie trimestral de promedio anuales: el dato relativo al trimestre t -ésimo será igual a la media de los valores publicados a lo largo de los doce meses precedentes, incluidos los del propio trimestre. Así, además de suavizar la curva del tipo de interés, el rezago de la curva suavizada (línea roja en **Fig. 5.8, Dcha**) respecto de la original (línea azul) será el mismo que en el caso de las curvas de visados.

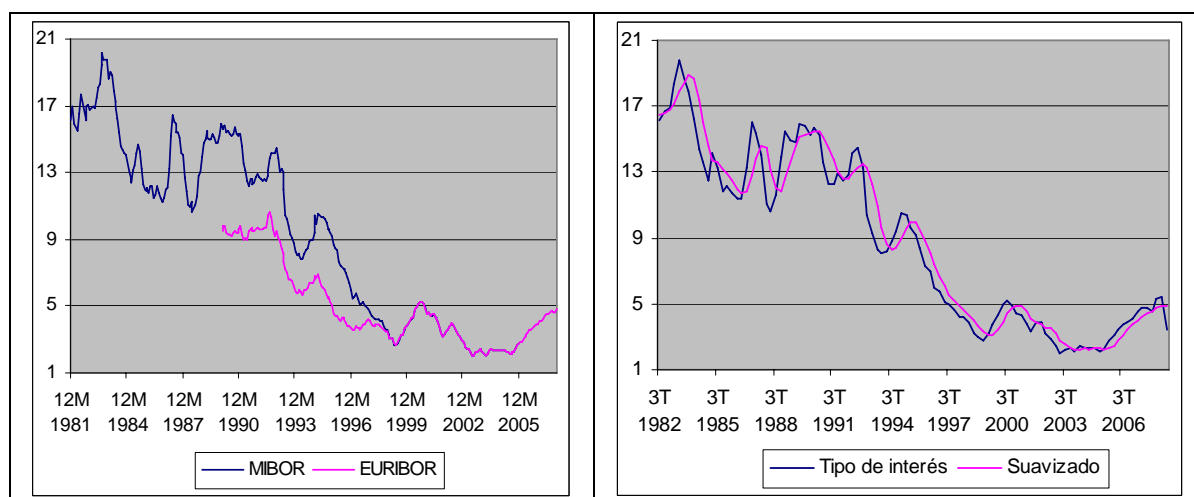


Figura 5.8 Creación de la serie Tipo de interés, X_t

¹ Para una descripción más detallada de los tipos de referencia del mercado hipotecario véase el Apartado 1.5 del Capítulo 1.

Apéndice 2:

La **Fig. 5.9** proporciona una tabla con los valores de $A_{0,j}$ y $A_{1,j}$ para las 50 provincias. Obsérvese que, aunque en todos los casos el valor del coeficiente $A_{0,j}$ es pequeño, en la mayoría es significativamente distinto de cero. Salvo para Ciudad Real, el coeficiente $A_{1,j}$ es negativo, tomando el mínimo valor en Málaga, y sólo en siete ocasiones no es significativamente distinto de cero: Ciudad Real, Toledo, Soria, Palencia, Salamanca, Badajoz y Guipúzcoa.

Apéndice 3:

Tenemos K series de la forma:

$$C_{t,k} = g_{t,k} - \beta_k + \alpha \quad k=1,\dots,K$$

donde:

$$g_{t,k} = g_{t-1,k} + f_{t,k}, \quad t=2,\dots,T \quad \text{y} \quad g_{1,k} = f_{1,k}, \quad k=1,\dots,K$$

siendo:

$$f_{t,k} = b_k + m_k \cdot \ln x_t, \quad t=2,\dots,T, \quad f_{1,k} = b_k \quad \text{con} \quad b_k = B_0 + B_1 \cdot m_k, \quad k=1,\dots,K$$

Además:

$$\frac{\sum_{t=1}^T C_{t,k}}{T} = \alpha$$

En otras palabras:

$$C_{1,k} = g_{1,k} - \beta_k + \alpha = f_{1,k} - \beta_k + \alpha$$

$$C_{2,k} = g_{2,k} - \beta_k + \alpha = g_{1,k} + f_{2,k} - \beta_k + \alpha = f_{1,k} + f_{2,k} - \beta_k + \alpha$$

$$C_{3,k} = g_{3,k} - \beta_k + \alpha = g_{2,k} + f_{3,k} - \beta_k + \alpha = f_{1,k} + f_{2,k} + f_{3,k} - \beta_k + \alpha$$

.....

$$C_{t,k} = g_{t,k} - \beta_k + \alpha = g_{t-1,k} + f_{t,k} - \beta_k + \alpha = f_{1,k} + \dots + f_{t,k} - \beta_k + \alpha = \sum_{s=1}^t f_{s,k} - \beta_k + \alpha$$

PROVINCIA	$A_{0,j}$	$A_{1,j}$
Almería	-0,017	-1,580
Granada	0,000*	-0,482
Jaén	-0,004*	-0,559
Málaga	-0,029	-1,807
Cádiz	-0,013	-1,091
Córdoba	0,004	-0,350
Huelva	0,000*	-0,664
Sevilla	-0,003*	-0,751
Huesca	-0,009	-0,914
Teruel	0,008	-0,243
Zaragoza	-0,003*	-0,527
ASTURIAS	-0,003*	-0,767
BALEARES / BAL	-0,016	-0,944
Las Palmas	-0,018	-1,429
Sta. Cruz de	-0,004*	-0,870
CANTABRIA	-0,005	-0,995
Albacete	0,009	-0,284
Ciudad Real	0,028	0,148*
Cuenca	0,013	-0,216
Guadalajara	0,009	-0,578
Toledo	0,014	-0,113*
Ávila	0,008	-0,282
Burgos	0,000*	-0,642
Segovia	0,007	-0,275
Soria	0,016	-0,175*
Valladolid	0,004*	-0,421
León	-0,010	-0,799
Palencia	0,008	-0,135*
Salamanca	0,013	-0,064*
Zamora	-0,003*	-0,518
Barcelona	0,007	-0,384
Girona	-0,020	-1,261
Lleida	0,012	-0,366
Tarragona	-0,011	-1,012
Cáceres	-0,001*	-0,499
Badajoz	0,013	-0,076*
La Coruña /	0,010	-0,379
Lugo	0,003*	-0,364
Orense / Our	-0,010	-0,693
Pontevedra	0,005*	-0,464
MADRID	-0,003*	-0,596
MURCIA	-0,004*	-0,827
LA RIOJA	0,002*	-0,590
Alicante / A	-0,022	-1,352
Castellón /	-0,002*	-0,620
València	0,000*	-0,528
Álava / Arab	-0,003*	-0,737
Guipúzcoa /	0,012	-0,014*
Vizcaya / Bi	0,003*	-0,430
NAVARRA	0,006	-0,218

* No significativo al nivel .05

Figura 5.9 Valores de $A_{0,j}$ y de $A_{1,j}$;

En consecuencia, en un instante t cualquiera:

$$C_{t,k} = \sum_{s=1}^t (b_k + m_k \cdot \ln x_s) - \beta_k + \alpha = b_k \cdot t + m_k \cdot \sum_{s=2}^t \ln x_s - \beta_k + \alpha =$$

$$= b_k \cdot t + m_k \cdot (\ln X_t - \ln X_1) - \beta_k + \alpha = b_k \cdot t + m_k \cdot \ln X_t - m_k \cdot \ln X_1 - \beta_k + \alpha$$

Si denominamos:

$$\mu_k = \ln X_1 - \beta_k + \alpha$$

entonces disponemos de un conjunto de K series temporales $\{C_{t,k}\}$, $k=1, \dots, K$, tales que:

$$\frac{\sum_{t=1}^T C_{t,k}}{T} = \alpha \quad \forall k$$

y para las que existe otra serie temporal, $\ln X_t$, tal que:

$$C_{t,k} = b_k \cdot t + m_k \cdot \ln X_t + \mu_k$$

donde los coeficientes b_k y m_k son tales que:

$$b_k = B_0 + B_1 \cdot m_k$$

En otras palabras, se verifican las hipótesis [3.1], [3.2] y [3.3] del **Lema 1**² para la metodología propuesta. En consecuencia, si:

$$m_q < m_r < m_s$$

entonces:

$$d(\ln H_{t,q}, \ln H_{t,r}) < d(\ln H_{t,q}, \ln H_{t,s})$$

o, lo que es equivalente:

$$d(H_{t,q}, H_{t,r}) < d(H_{t,q}, H_{t,s})$$

Apéndice 4:

En el **Apéndice 3** hemos comprobado que el conjunto de K series temporales $\{C_{t,k}\} = \{\ln H_{t,k}\}$ verifica las hipótesis [3.1], [3.2] y [3.3] del **Lema 1**³ para la metodología propuesta. En consecuencia, por la **Proposición 4**, si denominamos:

² Véase el Capítulo 3.

³ Véase el Capítulo 3.

$$\ln h_{t,k} = \nabla \ln H_{t,k} = \ln H_{t,k} - \ln H_{t-1,k} \quad \text{y} \quad \ln x_t = \nabla \ln X_t = \ln X_t - \ln X_{t-1}$$

entonces las series temporales $\ln h_{t,k}$ $k=1, \dots, K$ definen, respecto de $\ln x_t$, un haz de K rectas de vértice:

$$(\ln x_t, \ln h_{t,k}) = (-B_1, B_0).$$

o, lo que es equivalente, las series temporales $h_{t,k}$ $k=1, \dots, K$ definen, respecto de x_t , un haz de K curvas de vértice:

$$(x_t, h_{t,k}) = (\exp(-B_1), \exp(B_0))$$

Apéndice 5:

En el **Apéndice 3** hemos comprobado que el conjunto de K series temporales $\{ C_{t,k} \} = \{ \ln H_{t,k} \}$ verifica las hipótesis [3.1], [3.2] y [3.3] del **Lema 1⁴** para la metodología propuesta. En consecuencia, por la **Proposición 7**, dado un par cualquiera de series de la forma $C_{t,k}$ existe al menos un punto en su trayectoria en el que las series se cruzan. Además, en aquellos puntos tales que dos de las series se cruzan, también lo hacen las restantes. Considerando que:

$$H_{t,k} = \exp C_{t,k}$$

en aquellos instantes en los que $C_{t,k} = C_{t,k'}$ entonces también $H_{t,k} = H_{t,k'}$.

Apéndice 6:

En el **Apéndice 3** hemos comprobado que el conjunto de K series temporales $\{ C_{t,k} \} = \{ \ln H_{t,k} \}$ verifica las hipótesis [3.1], [3.2] y [3.3] de la **Lema 1⁵** para la metodología propuesta. En consecuencia, por la **Proposición 8**, la diferencia entre dos puntos de cruce cualesquiera es independiente de la media de las series.

⁴ Véase el Capítulo 3.

⁵ Véase el Capítulo 3.

CAPÍTULO 6

DESCRIPCIÓN DE LA AGRUPACIÓN OBTENIDA

La solución ideal de un *Análisis Cluster* es que los elementos clasificados en un mismo grupo sean muy parecidos entre sí y, en consecuencia, muy parecidos al centroide correspondiente, y distintos de los elementos clasificados en otros grupos. En nuestro caso¹, en términos generales, la distancia de una provincia al centroide final del propio grupo es muy parecida a las distancias a los centroides finales de los grupos adyacentes; incluso hay provincias para las que la serie correspondiente es más parecida al centroide final de otro grupo que algunas del propio grupo. En otras palabras, parece que las provincias clasificadas en un determinado grupo se parecen, además de a las restantes provincias del propio grupo, a las de los grupos adyacentes. Con el objetivo confirmar esta apreciación, comparemos en primer lugar tanto los centroides iniciales como los finales de los seis grupos.

6.1 COMPARACIÓN DE LOS CENTROIDES INICIALES

Recordemos² que en aquellos puntos en los que dos de los centroides iniciales se cruzan también lo hacen los restantes. Estos puntos de corte son los puntos de inflexión o cambio de ciclo entorno a la tendencia media de crecimiento. Permiten describir la trayectoria de los centroides iniciales así como la de la serie promedio (**Fig. 6.1, Sup. Izqda.**): en el periodo de cinco años transcurrido entre las dos fechas correspondientes a los dos primeros puntos de corte (entre finales de 1986 y finales de 1991), el promedio de visados anuales se incrementó en una proporción de 1,39 (**Fig. 6.1, Sup. Dcha.**); en el

¹ Véase la tabla de distancias a los centroides finales en la Fig. II.6 del Anexo II.

² Véase el apartado 4.5 del Capítulo 4.

periodo de siete años transcurrido entre las dos fechas correspondientes al segundo y al tercer puntos de corte (entre finales de 1991 y finales de 1998) se incrementó en una proporción de 1,52 o, lo que es equivalente, en el periodo de doce años transcurrido entre las dos fechas correspondientes al primer y al tercer puntos de corte (entre finales de 1986 y finales de 1998) se incrementó en una proporción de 2,11; ... finalmente, en el periodo de veintiún años comprendido entre finales de 1986 y el tercer trimestre de 2007 el promedio de visados anuales se incrementó en una proporción de 3,56.

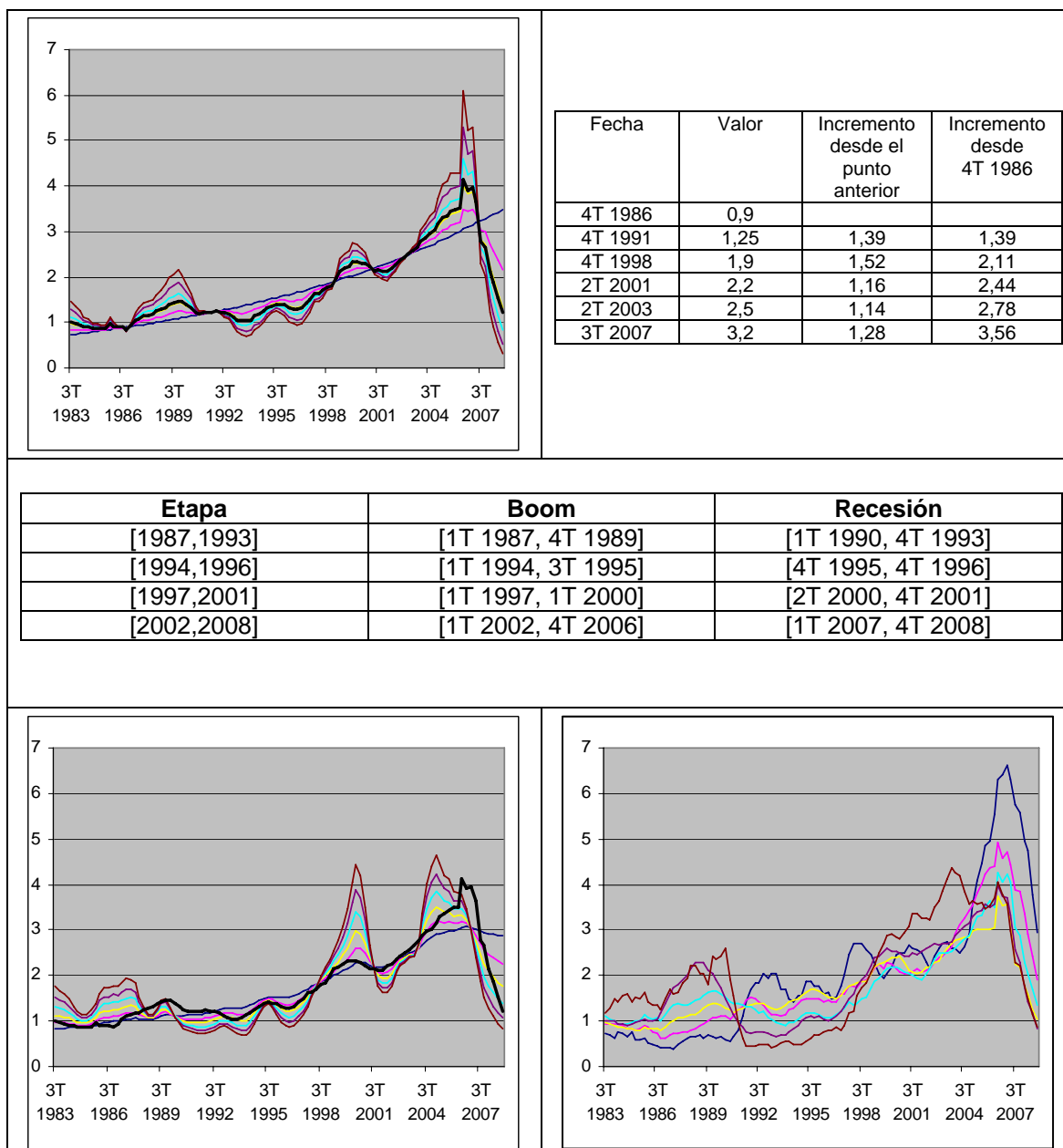


Figura 6.1 *Sup. izqda*: Centroides iniciales junto con la serie promedio; *Sup. dcha*: Puntos de corte en la trayectoria de los centroides iniciales; *Centro*: Etapas en la trayectoria de la serie promedio; *Inf. izqda*: Curvas resumen construidas a partir de la serie del tipo de interés junto con la serie promedio; *Inf. dcha*: Centroides finales

Lo que diferencia a los seis centroides iniciales es la trayectoria seguida entre los puntos de corte. Para compararlas, distinguiremos cuatro etapas (**Fig. 6.1, Centro**). La primera abarca el periodo de tiempo comprendido entre finales de 1986 y finales de 1993 y las tres siguientes el comprendido entre principios de 1994 y finales de 2008. En la primera etapa, la trayectoria del sexto centroide final (la curva marrón) es tal que a un corto episodio de *boom* (1987-1989), sigue otro corto episodio de recesión (1990-1993), mientras que en el periodo comprendido entre 1994 y 2008 se caracteriza por una marcada y prolongada tendencia creciente que se manifiesta fundamentalmente en tres episodios de *boom* y recesión: el primero, corto y débil, entre el primer trimestre de 1994 y el tercero de 1995, seguido de una pequeña recesión entre los cuartos trimestres de 1995 y 1996; el segundo, más largo e intenso, entre el primer trimestre de 1997 y el primero de 2000, seguido de una pequeña recesión entre el segundo trimestre de 2000 y el cuarto 2001; y el tercero, también intenso y prolongado, entre el primer trimestre de 2002 y finales de 2006. Este tercer episodio finaliza con un fuerte repunte en el segundo semestre de 2006 precediendo al fuerte derrumbe final (2007-2008).

El quinto centroide inicial sigue una trayectoria muy similar a la del sexto; la diferencia entre ellos radica en que tanto los episodios de *boom* como los de recesión son menos pronunciados. Así, cada centroide se parece al anterior, siendo los episodios de *boom* y de recesión cada vez menos pronunciados, hasta el primer centroide inicial, que presenta una trayectoria de tendencia creciente constante.

Considerando que la serie promedio es prácticamente igual que el tercer centroide inicial³ (**Fig. 6.1, Sup. Izqda.**), su trayectoria es tal que presenta un corto y débil episodio de *boom* y recesión entre 1987 y 1993, y entre 1994 y 2006 se caracteriza por una marcada y prolongada tendencia creciente de sus valores que se manifiesta fundamentalmente en tres episodios de *boom*: [1T 1994, 3T 1995], [1T 1997, 1T 2000] y [1T 2002, 4T 2006]; en el primero de ellos débilmente y en los dos siguientes con más intensidad. A cada una de estos *booms* sigue un periodo corto de recesión, débil en el caso de los dos primeros, y muy fuerte en el caso del tercero, el que se produce en los dos últimos años del periodo de observación (2007-2008).

Obsérvese que, desde la segunda etapa, las oscilaciones de la trayectoria de la serie promedio están muy en consonancia con las fluctuaciones del tipo de interés (**Fig. 6.1, Inf. Izqda.**). Aunque transcurre muy pegada a la segunda curva resumen (en rojo), en la recesión final la trayectoria tiene una pendiente muy parecida a la de la sexta (en marrón), lo que indica una fuerte sensibilidad a la subida de los tipos de interés en la fase final del periodo de observación. Considerando que las curvas resumen han sido extraídas aplicando un retardo de cuatro trimestres sobre la serie del tipo de interés, podríamos afirmar que aunque globalmente la respuesta frente a las fluctuaciones del tipo de interés se produce con un año de retraso, en la fase final se ha producido no sólo con mayor sensibilidad, sino también con mayor retraso.

³ Véase los coeficientes para el valor de k igual a 3 en el panel superior izquierdo de la Fig. 4.5 del Capítulo 4.

6.2 COMPARACIÓN DE LOS CENTROIDES FINALES

Con el objetivo de simplificar la interpretación de las trayectorias de los seis centroides finales, utilizaremos la representación de cada centroide final en términos de los seis centroides iniciales. Según lo anteriormente expuesto, en la comparación de estas seis formas de evolucionar (**Fig. 6.1**, *Inf. Dcha.*) distinguiremos cuatro etapas: la primera corresponde al periodo de tiempo comprendido entre 1987 y 1993; la segunda, al comprendido entre 1994 y 1996; la tercera, al comprendido entre 1997 y 2001; y la cuarta y última, al comprendido entre 2002 y 2008.

La trayectoria del quinto centroide final (**Fig. 6.2**, *Inf. Izqda.*) es tal que en la primera etapa (1987-1993) presenta un fuerte *boom* seguido de una fuerte recesión; en la segunda (1994-1996), a un débil *boom* sigue una también débil recesión; en la tercera (1997-2001), a un fuerte *boom* sigue una inapreciable recesión; y en la cuarta (2002-2008), a un fuerte *boom* sigue una recesión de tal magnitud que el valor al final de la misma es incluso menor que el de principios de 1994. El sexto centroide final o, lo que es equivalente, Málaga, sigue una trayectoria muy parecida (**Fig. 6.2**, *Inf. Dcha.*). Resaltar únicamente que el fuerte *boom* de finales de los noventa no finaliza en recesión sino que se confunde con el siguiente *boom*, el de la cuarta y última etapa.

El cuarto centroide final tiene una trayectoria muy parecida a la quinto (**Fig. 6.2**, *Centro Dcha.*). La mayor diferencia entre ambas se encuentra en la primera etapa (1987-1993), en la que tanto el *boom* como la recesión son más suaves en el caso del cuarto. También el *boom* de la tercera etapa (1997-2001) es más suave pero, en contrapartida, el *boom* de la cuarta (2002-2008) es más intenso.

El tercer centroide final sigue, con pequeñas fluctuaciones acordes con los episodios de *boom* y recesión detectados en la serie promedio, una trayectoria de tendencia creciente a lo largo de prácticamente todo el periodo de análisis, llegando a triplicar el valor de partida a mediados del año 2006 (**Fig. 6.2**, *Centro Izqda.*). En los dos últimos años, tras un pequeño repunte en el segundo semestre de 2006, decrece con fuerza hasta recuperar el nivel del año 1983.

La trayectoria del segundo centroide final (**Fig. 6.2**, *Sup. Dcha.*), aunque muy parecida a la del tercero en las dos etapas centrales, en la cuarta se caracteriza por un fuerte *boom*, en el que, en el periodo de aproximadamente cinco años comprendido entre principios de 2002 y mediados de 2006, supera su valor de partida en dos veces y media. Este *boom* que precede a la recesión final es aún más acusado en el caso del primer centroide (**Fig. 6.2**, *Sup. Izqda.*), que supera su valor de partida en dos veces y media tan solo en los dos años transcurridos entre 2005 y 2006. Ésta no es la única gran diferencia frente a los restantes centroides finales: hasta 1991 la trayectoria del primer centroide final toma valores muy bajos. En esta etapa inicial los valores del segundo centroide final, aunque no tanto, también son bajos en comparación con los restantes centroides finales.

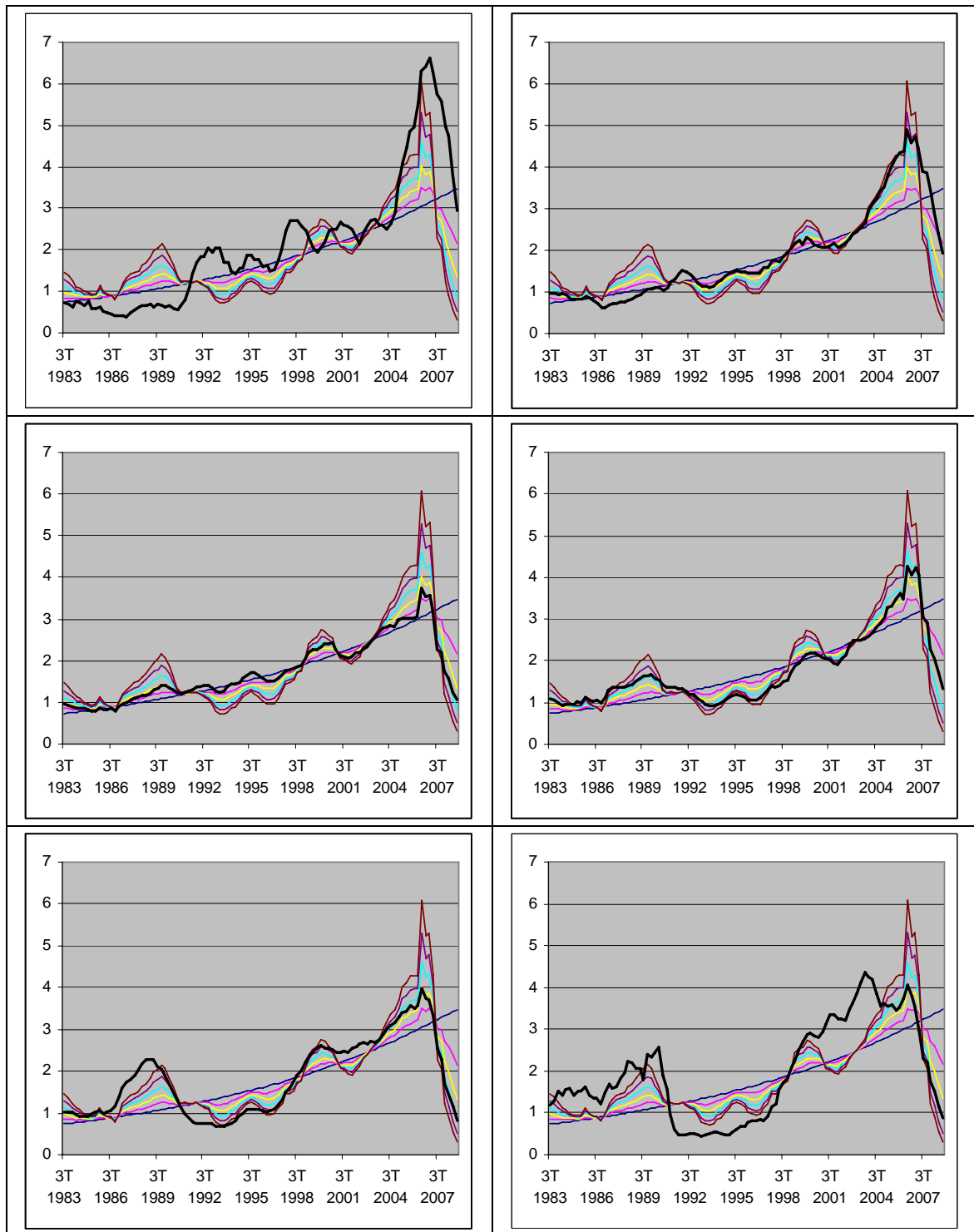


Figura 6.2 Centroides finales en términos de los centroides iniciales (de izqda. a dcha. y de arriba a abajo, grupos 1, 2, 3, 4, 5 y 6, respectivamente)

Parece entonces que, de la misma forma que los centroides iniciales están ordenados⁴, también lo están los finales e, implícitamente, los grupos. Formalicemos esta apreciación.

Dado el conjunto de los K centroides finales $\{H_{t,k}^F\}$, $k=1,\dots,K$, consideremos la matriz $K \times K$ de distancias $d_{k,k'}$:

$$d_{k,k'} = d^2(\ln H_{t,k}^F, \ln H_{t,k'}^F) \quad k, k' = 1, \dots, K$$

siendo d la distancia euclídea. Si representamos las seis columnas de la matriz de distancias, podemos observar (**Fig. 6.3, Izqda.**) que el centroide más próximo al del primer grupo (línea azul) es el del segundo, seguido del tercero y así sucesivamente, lo que implica una secuencia de distancias creciente; por otro lado, el centroide más próximo al del sexto grupo (curva marrón) es el del quinto, seguido del cuarto y así sucesivamente, lo que implica una secuencia de distancias decreciente. En lo que se refiere a cualquier otro centroide, la secuencia será decreciente hasta cero (distancia consigo mismo) y creciente hasta el sexto centroide. En otras palabras, la secuencia de distancias al centroide del primer grupo está positivamente correlada con la secuencia de distancias al centroide del segundo que, a su vez, está positivamente correlada con la secuencia de distancias al centroide del tercero, y así sucesivamente hasta la secuencia de distancias al centroide del quinto grupo que está positivamente correlada con la secuencia de distancias al centroide del sexto grupo. Además, a medida que se separan los grupos, la correlación positiva se va debilitando pasando a ser en negativa, alcanzando su mínimo entre las secuencia correspondientes a los centroides de los grupos extremos, el primero y el sexto. Esta relación entre los centroides finales implica que, si aplicamos un Análisis de Componentes Principales sobre la matriz de distancias entre los centroides finales y representamos la solución en el espacio de los dos primeros componentes (**Fig. 6.3, Dcha.**), observamos el denominado efecto Guttman⁵.

Así, si todas las series clasificadas en un mismo grupo fueran, por un lado, muy parecidas entre sí y, en consecuencia, muy parecidas al centroide correspondiente y, por otro, distintas de las series clasificadas en otro grupo, dadas todas las series de un grupo cualquiera sus correspondientes secuencias de distancias a todas las series estarían muy correladas positivamente. En este caso, al aplicar un Análisis de Componentes Principales sobre la matriz de distancias entre series, y representarlas como puntos en el espacio de los dos primeros componentes, observaríamos no sólo una parábola sino que, además, los puntos correspondientes a series clasificadas en un mismo grupo estarían muy próximos entre sí. Analicemos en estos términos la solución obtenida.

⁴ Véase el Apéndice 8 del Capítulo 4.

⁵ El efecto Guttman se obtiene cuando, al representar las filas o las columnas de una matriz en el espacio de las dos primeras componentes de la solución factorial, la nube de puntos correspondiente tiene forma de arco de parábola.

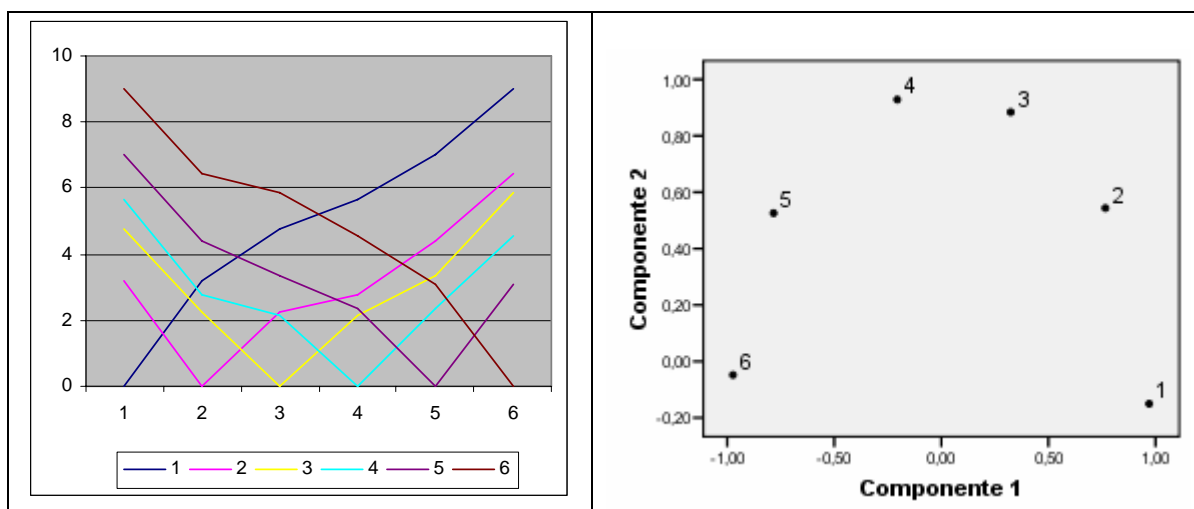


Figura 6.3 Distancias entre las series $H_{t,k}^F, k=1, \dots, 6$ y representación de las correlaciones entre ellas en el espacio factorial

6.3 REPRESENTACIÓN FACTORIAL DE LAS DISTANCIAS

Dado los conjuntos de las J series $\{Y_{t,j}\}, j=1, \dots, J$, y de los K centroides finales $\{H_{t,k}^F\}, k=1, \dots, K$, consideremos la matriz $(J+K) \times (J+K)$ de distancias $d_{i,i'}$, donde, si d es la distancia euclídea.:

$d_{i,i'} = d^2(\ln Y_{t,i}, \ln Y_{t,i'})$	si i e i' están comprendidos ambos entre 1 y J :
$d_{i,i'} = d^2(\ln H_{t,i-J}^F, \ln H_{t,i'-J}^F)$	si i e i' están comprendidos ambos entre $J+1$ y $J+K$:
$d_{i,i'} = d^2(\ln Y_{t,i}, \ln H_{t,i'-J}^F)$	si i está comprendido entre 1 y J e i' entre $J+1$ y $J+K$:
$d_{i,i'} = d^2(\ln H_{t,i-J}^F, \ln Y_{t,i'})$	si i está comprendido entre $J+1$ y $J+K$ e i' entre 1 y J :

Si aplicamos un Análisis de Componentes Principales sobre dicha matriz de distancias y representamos la solución en el espacio de los dos primeros componentes observamos (**Fig. 6.4**) que, si trazáramos una línea uniendo los centroides de los seis grupos obtendríamos una curva muy aproximada a un arco de parábola de vértice en la posición del tercer centroide o en la de Córdoba. Estos dos puntos son los mejor representados por la primera componente principal (**Fig. 6.5**), que separa los dos últimos grupos de los cuatro primeros: el valor de la saturación de las provincias en los grupos quinto y sexto, así como la de sus respectivos centroides, es inferior a 0,25, mientras que para los cuatro primeros grupos es superior. Podríamos interpretar entonces que la primera componente principal, que explica el 41,86 por ciento de la variabilidad total, surge para separar el conjunto de provincias clasificadas en los grupos quinto y sexto del resto. Por otro lado,

la segunda componente principal, que explica el 25,80 por ciento de la variabilidad total, surge para separar los cuatro primeros grupos; además, los puntos correspondientes al quinto centroide y a Alicante son los mejor representados por dicha componente. Simplificando, podría interpretarse que Córdoba representa al grueso de las provincias españolas con una evolución parecida, mientras que Alicante representa al conjunto de provincias de mayor sensibilidad frente a la serie promedio.

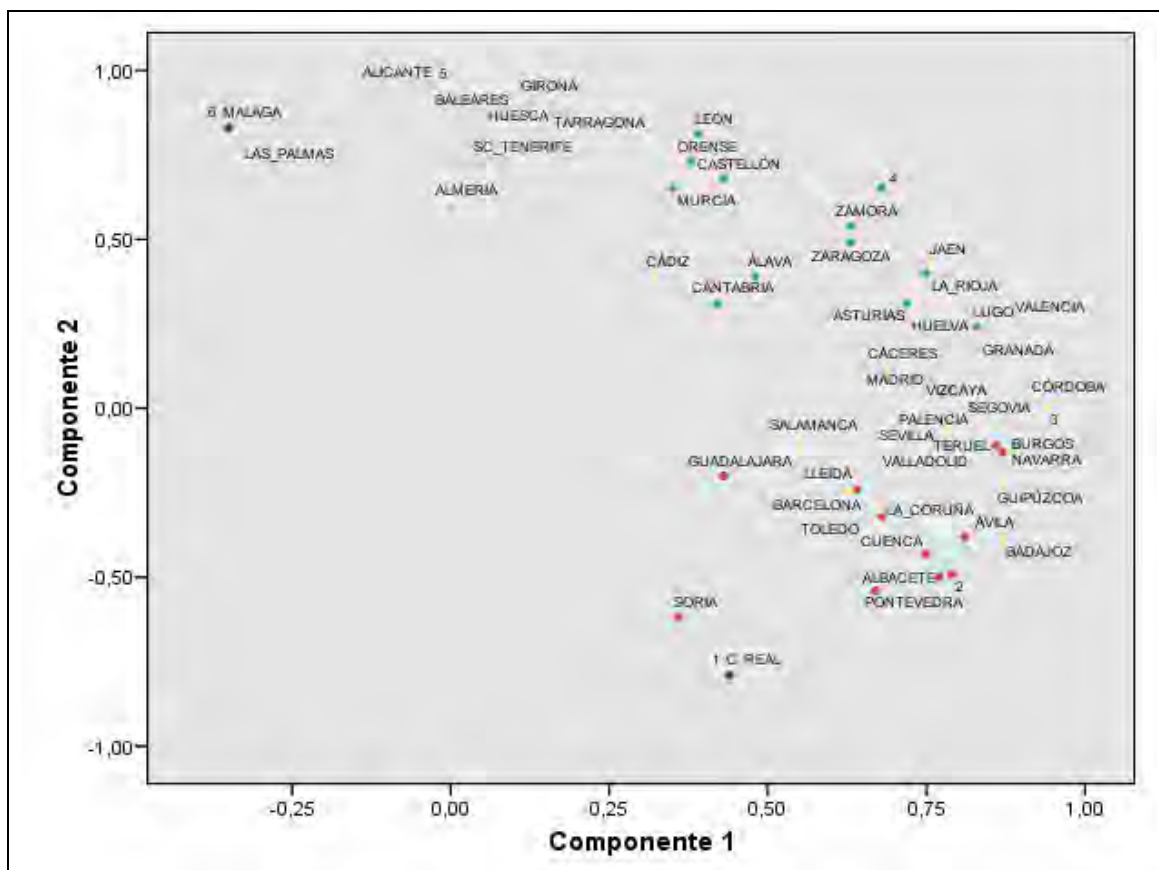


Figura 6.4 Representación factorial de las provincias

En términos generales, podemos afirmar que los distintos grupos de provincias se encuentran en tramos consecutivos del arco de parábola, lo que implica que, aunque algunas series están muy próximas al centroide de su propio grupo, otras se encuentran entre dicho centroide y el de grupo inmediatamente anterior o posterior.

La correcta interpretación de la posición de los puntos en el espacio factorial dependerá de la calidad de su representación. Si consideramos las dos primeras componentes, la calidad de representación de las provincias en conjunto es del 67,66 por ciento, lo que supone que para algunas de ellas en particular es muy alta: tales como Badajoz, Alicante o Valencia, para los que el valor de la comunalidad es mayor que 0,9, mientras que para otras es muy baja: tales como Cádiz, Cantabria o Guadalajara, para los que el valor de la comunalidad es menor que 0,3.

Componente	Autovalor	% de la varianza	% acumulado
1	22,60	41,86	41,86
2	13,93	25,80	67,66

	Componente		Comunalidades
	1	2	
ALMERÍA	0,00	0,60	0,36
GRANADA	0,87	0,13	0,77
JAÉN	0,76	0,43	0,77
CÁDIZ	0,32	0,39	0,26
CÓRDOBA	0,95	0,02	0,89
HUELVA	0,73	0,28	0,62
SEVILLA	0,78	-0,07	0,61
HUESCA	0,16	0,90	0,83
TERUEL	0,86	-0,11	0,75
ZARAGOZA	0,63	0,49	0,63
ASTURIAS	0,72	0,31	0,62
BALEARES	0,01	0,87	0,75
LAS_PALMAS	-0,28	0,71	0,59
SC_TENERIFE	0,09	0,73	0,54
CANTABRIA	0,42	0,31	0,27
ALBACETE	0,77	-0,50	0,85
CUENCA	0,75	-0,43	0,75
GUADALAJARA	0,43	-0,20	0,23
TOLEDO	0,68	-0,32	0,57
ÁVILA	0,81	-0,38	0,80
BURGOS	0,87	-0,13	0,77
SEGOVIA	0,92	-0,03	0,84
SORIA	0,36	-0,62	0,52
VALLADOLID	0,82	-0,15	0,69
LEÓN	0,39	0,81	0,81
PALENCIA	0,86	-0,06	0,75
SALAMANCA	0,55	-0,09	0,31
ZAMORA	0,63	0,54	0,68
BARCELONA	0,63	-0,25	0,45
GIRONA	0,13	0,91	0,84
LLEIDA	0,64	-0,24	0,47
TARRAGONA	0,18	0,88	0,81
CÁCERES	0,69	0,12	0,49
BADAJOS	0,87	-0,39	0,91
LA_CORUÑA	0,68	-0,30	0,55
LUGO	0,83	0,24	0,74
ORENSE	0,38	0,73	0,68
PONTEVEDRA	0,67	-0,54	0,74
MADRID	0,76	0,12	0,59
MURCIA	0,35	0,65	0,54
LA_RIOJA	0,75	0,40	0,72
ALICANTE	-0,11	0,95	0,91
CASTELLÓN	0,43	0,68	0,66
VALENCIA	0,92	0,26	0,91
ÁLAVA	0,48	0,39	0,38
GUIPÚZCOA	0,87	-0,23	0,81
VIZCAYA	0,75	0,01	0,56
NAVARRA	0,88	-0,12	0,79
1 C. Real	0,44	-0,79	0,82
2	0,79	-0,49	0,87
3	0,94	0,00	0,88
4	0,68	0,65	0,88
5	-0,03	0,96	0,92
6 Málaga	-0,35	0,83	0,81

Figura 6.5 Resultados del Análisis de Componentes Principales

En cualquier caso, y siempre bajo el supuesto de que su calidad de representación es alta, si el ángulo que forman desde el origen dos puntos-provincia es muy pequeño, entonces las dos correspondientes secuencias de distancias a todas las provincias están muy correladas positivamente y, en consecuencia, podemos afirmar que las dos provincias son parecidas entre sí; si el ángulo es próximo a los 180 grados, entonces las dos correspondientes secuencias de distancias están muy correladas negativamente y, en consecuencia, podemos afirmar que son distintas entre sí; y finalmente, si el ángulo es próximo a los 90 grados, entonces las dos correspondientes secuencias de distancias están muy incorreladas y, en consecuencia, podemos afirmar que las dos provincias no son ni muy parecidas ni muy distintas.

Con todas estas consideraciones, a la hora de interpretar la trayectoria de una provincia concreta en comparación con las restantes, además de las tres ya mencionadas en ocasiones anteriores, incluiremos otras tres representaciones más: una junto con cada una de las dos provincias mejor representadas por cada una de las componentes principales: Córdoba y Alicante, y la tercera junto con España, por lo que, antes de proceder a la descripción de cada una de las distintas series provinciales, dediquemos un apartado a la descripción de la serie agregada.

6.4 DESCRIPCIÓN DE LA SERIE NACIONAL

Recordemos⁶ que para aplicar la metodología era necesario homogeneizar la escala de medida de las J series temporales $V_{t,j}$ considerando transformaciones de la forma:

$$\ln Y_{t,j} = \ln V_{t,j} + \varepsilon_j \quad j=1,\dots,J$$

donde:

$$\varepsilon_j = \alpha - \alpha_j$$

con:

$$\alpha_j = \frac{\sum_{t=1}^T \ln V_{t,j}}{T} \quad \text{y} \quad \alpha = \frac{\sum_{t=1}^T \ln Y_t}{T}$$

siendo:

$$Y_t = \frac{V_t}{V_1} \quad \text{y} \quad \ln V_t = \frac{\sum_{j=1}^J \ln V_{t,j}}{J} \quad t=1,\dots,T$$

Así,

$$\ln Y_t = \frac{\sum_{j=1}^J \ln Y_{t,j}}{J}$$

⁶ Véase el apartado 4.2 del Capítulo 4.

y las series transformadas tienen todas la misma media:

$$\frac{\sum_{t=1}^T \ln Y_{t,j}}{T} = \frac{\sum_{t=1}^T \ln Y_t}{T} = \alpha \quad \forall j$$

Si denominamos $V_{t,0}$ a la serie temporal de visados en España, para situar sus valores en la misma escala que los de las distintas series provinciales consideraremos la siguiente transformación:

$$\ln Y_{t,0} = \ln V_{t,0} + \varepsilon_0 \quad j=1,\dots,J$$

donde:

$$\varepsilon_0 = \alpha - \alpha_0$$

con:

$$\alpha_0 = \frac{\sum_{t=1}^T \ln V_{t,0}}{T}$$

Así:

$$\frac{\sum_{t=1}^T \ln Y_{t,0}}{T} = \frac{\sum_{t=1}^T \ln Y_t}{T} = \alpha \quad \forall j$$

Obsérvese (**Fig. 6.6, Sup. Izqda.**) que las series $Y_{t,0}$ e Y_t son muy parecidas; en consecuencia, considerando que el tercer centroide inicial y la serie promedio son prácticamente iguales, la serie de España en valores homogeneizados también es muy parecida al tercer centroide, aunque pueda serlo aún más al cuarto (**Fig. 6.6, Sup. Dcha.**). En definitiva, la trayectoria de España, bien resumida por los puntos de corte de los centroides iniciales, es tal que presenta un corto y moderadamente fuerte episodio de *boom* y recesión entre 1987 y 1993; entre 1994 y 2006 se caracteriza por una marcada y prolongada tendencia creciente de sus valores que se manifiesta fundamentalmente en tres episodios de *boom*: [1T 1994, 3T 1995], [1T 1997, 1T 2000] y [1T 2002, 4T 2006]; en el primero de ellos débilmente y en los dos siguientes con más intensidad. A cada una de estos *booms* sigue un periodo corto de recesión, débil en el caso de los dos primeros, y muy fuerte en el caso del tercero, el que se produce en los dos últimos años del periodo de observación (2007-2008). El periodo delimitado por 1994 y 2008 se caracteriza, además, porque las oscilaciones de la trayectoria están muy en consonancia con las fluctuaciones del tipo de interés (**Fig. 6.6, Inf. Izqda.**).

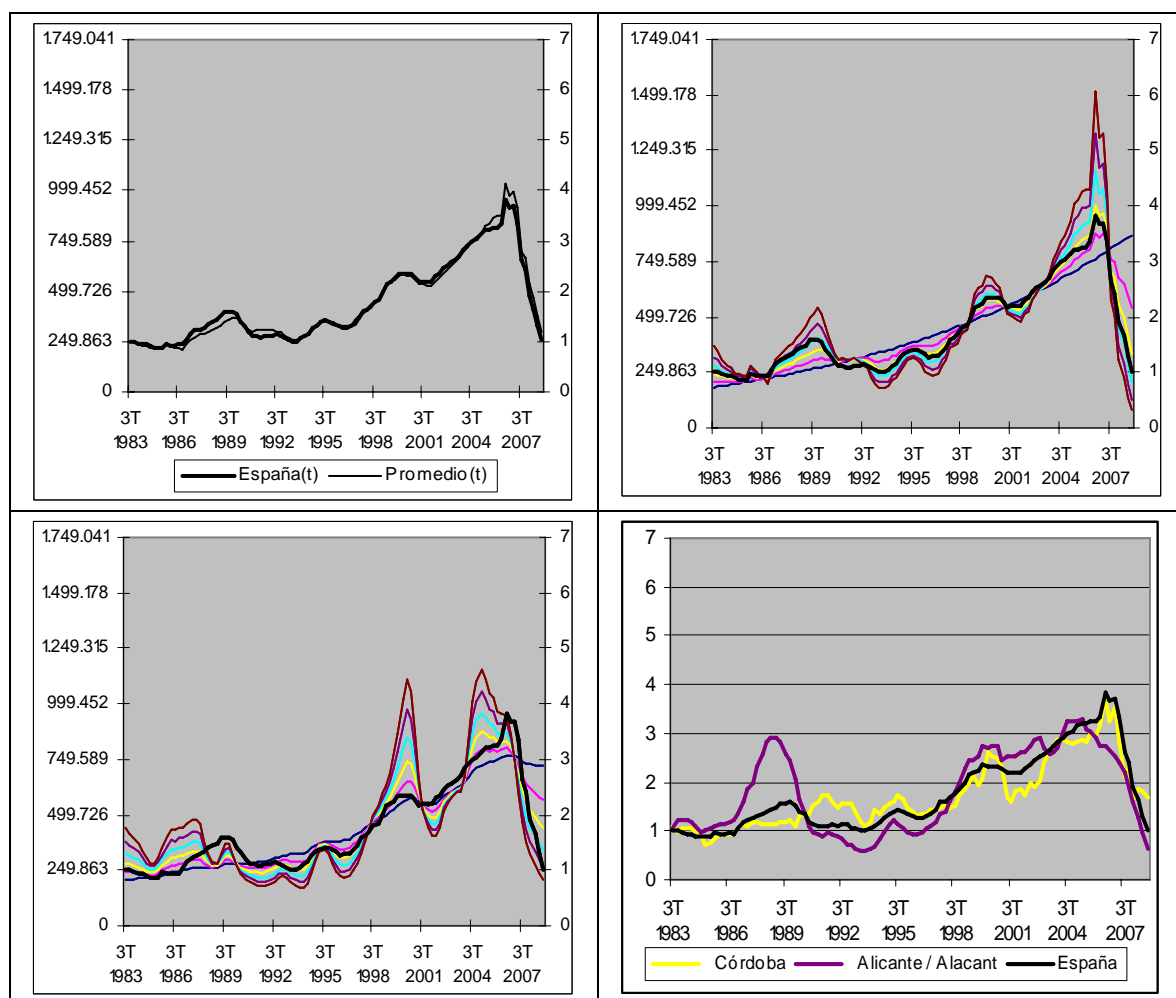


Figura 6.6 Evolución de la construcción residencial en España (acumulado anual y valores homogeneizados). *Sup. izqda:* junto con la serie promedio; *Sup. dcha:* junto con los centroides iniciales; *Inf. izqda:* junto con las curvas resumen construidas a partir de la serie del tipo de interés; *Inf. dcha:* junto con Alicante y Córdoba.

Comparando España con Córdoba y Alicante (**Fig. 6.6, Inf. Dcha.**) que, recordemos, representan al grueso de las provincias españolas con una evolución parecida y al conjunto de provincias de mayor sensibilidad frente a la serie promedio, respectivamente, mientras que las trayectorias de España y Córdoba son bastante parecidas, la de Alicante se caracteriza por la fuerza del primer *boom* y del tercero, así como por la debilidad del cuarto. En el caso de Córdoba parece que el primer episodio de *boom* se retrasa, tal como observábamos en el segundo centroide inicial, coincidiendo con las fechas de la primera recesión en Alicante y España; además, tanto la tercera recesión (2000-2001), como el cuarto *boom* (2002-2006), son más pronunciados que en España; sin embargo, la cuarta recesión (2007-2008) parece haberse frenado en niveles por encima de los de Alicante y España que, en ambos casos, están a la altura del nivel alcanzado tras la primera recesión (finales de 1993).

6.5 DESCRIPCIÓN DE LAS SERIES PROVINCIALES

Los seis gráficos relativos a la interpretación de la trayectoria de cada provincia se proporcionan en el Anexo III. El primero de ellos (panel superior izquierdo) permite interpretarla en términos de las restantes provincias clasificadas en su mismo grupo; el segundo (panel central izquierdo), en términos de los centroides iniciales; el tercero (panel inferior izquierdo), en términos de las fluctuaciones del tipo de interés; y los tres últimos (paneles superior, central e inferior derechos), en términos de Córdoba, de Alicante y de España, respectivamente. En la descripción de las provincias comenzaremos por el sexto grupo y terminaremos con el primero. Dentro de cada grupo el criterio seguido será el de mínima distancia al centroide del grupo correspondiente (**Fig. 6.7**).

6		5		4		3		2		1	
PROV.	DIST.	PROV.	DIST.	PROV.	DIST.	PROV.	DIST.	PROV.	DIST.	PROV.	DIST.
MÁLAGA	0,00	ALICANTE	1,41	LA_RIOJA	1,75	CÓRDOBA	1,39	ÁVILA	1,74	C.REAL	0,00
		HUESCA	1,63	LUGO	1,92	VALENCIA	1,48	CUENCA	1,77		
		TARRAGONA	1,66	CASTELLÓN	1,93	HUELVA	1,69	ALBACETE	1,83		
		GIRONA	1,96	LEÓN	1,94	MADRID	1,76	TERUEL	1,88		
		SC_TENERIFE	2,21	ZARAGOZA	1,97	GUIPUZCOA	1,83	BURGOS	2,08		
		BALEARES	2,97	ZAMORA	2,12	NAVARRA	1,88	LLEIDA	2,44		
		LAS_PALMAS	3,14	MURCIA	2,14	SEGOVIA	1,91	PONTEVEDRA	2,56		
		ALMERIA	3,23	ASTURIAS	2,21	BADAJOS	1,91	TOLEDO	2,65		
				ORENSE	2,64	VALLADOLID	1,99	GUADALAJARA	2,84		
				CANTABRIA	2,82	GRANADA	2,09	SORIA	3,84		
				ÁLAVA	2,90	JAÉN	2,11				
						SEVILLA	2,13				
						PALENCIA	2,16				
						BARCELONA	2,23				
						LA_CORUNA	2,24				
						VIZCAYA	2,36				
						CÁCERES	2,82				
						CÁDIZ	3,03				
						SALAMANCA	3,21				

Figura 6.7 Distancia de cada provincia al centroide final de su grupo

6.5.1 Descripción de las provincias clasificadas en los grupos quinto y sexto

Dado que el sexto grupo está formado por sólo una provincia, Málaga, y que, en consecuencia, la trayectoria del centroide y la de Málaga son una misma cosa, para completar su descripción la compararemos con las provincias del grupo más próximo, el quinto.

Recordemos (**Fig. 6.2, Inf. Izqda.**) que la trayectoria del quinto centroide final es tal que en la primera etapa (1987-1993), presenta un fuerte *boom* seguido de una fuerte recesión; en la segunda etapa (1994 y 1996) a un débil *boom* sigue una también débil recesión; en la tercera (1997-2001), a un fuerte *boom* sigue una inapreciable recesión; y en la cuarta (2002-2008), a un fuerte *boom* sigue una recesión de tal magnitud que el valor al final de la misma es incluso menor que el de principios de 1994. El sexto centroide final o, lo que es equivalente, Málaga, sigue una trayectoria parecida (**Fig. 6.2, Inf. Dcha.**). Las provincias del quinto grupo son parecidas entre sí y, en consecuencia, parecidas al centroide final del grupo. Veamos si siguen también esta misma pauta de comportamiento

Málaga (Fig. III.1): Su trayectoria, parecida a la del quinto centroide final, se caracteriza, en comparación con las provincias clasificadas en el quinto grupo, por valores relativamente bajos entre 1992 y 1997 y relativamente altos entre 2000 y 2005. Además, en los tres últimos años del periodo de observación su trayectoria es prácticamente igual que la del quinto centroide final y que la de España. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales; además, a partir del año 1991, sus oscilaciones están bastante en consonancia con las fluctuaciones del tipo de interés.

Alicante/Alacant (Fig. III.2): Su trayectoria, muy parecida a la del centroide de su grupo, se caracteriza por valores relativamente altos en el periodo comprendido entre 1988 y 1990 y relativamente bajos en el comprendido entre 2005 y 2007. Aunque el punto de corte previo a la recesión final no haya sido alcanzado (la recesión final se produce con algo de adelanto en comparación con la mayor parte de provincias españolas), en términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales; además, a partir del año 1991 sus oscilaciones están bastante en consonancia con las fluctuaciones del tipo de interés.

Huesca (Fig. III.3): Su trayectoria, muy parecida tanto a la del centroide de su grupo como a la de Alicante, se caracteriza porque en general, y fundamentalmente en la primera mitad del periodo de observación, transcurre con algo de retraso frente a ellas. Por otro lado, en el periodo comprendido entre 2005 y 2007 alcanza valores algo más altos que Alicante, lo que la hace más parecida a España. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés.

Tarragona (Fig. III.4): Su trayectoria, muy parecida tanto a la del centroide de su grupo como a la de Alicante, se caracteriza, al igual que la de Huesca, porque en general, y fundamentalmente en la primera mitad del periodo de observación, transcurre con algo de retraso frente a la de Alicante. Por otro lado, y de forma incluso más acusada que Huesca, en el periodo comprendido entre 2005 y 2007 alcanza valores más altos que Alicante, lo que la hace más parecida a España. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés.

Girona (Fig. III.5): Su trayectoria, muy parecida a la del centroide de su grupo, se caracteriza por ser prácticamente igual que la de España, salvo en lo que se refiere al *boom* de finales de los ochenta, etapa en la que toma valores, en términos relativos, más altos. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1990 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés.

Sta. Cruz de Tenerife (Fig. III.6): Su trayectoria, además de ser muy parecida a la del centroide de su grupo a lo largo de todo el periodo de observación, también presenta bastante similitud con la de España e incluso con la de Córdoba en la segunda mitad. En

términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Baleares/Balears (Fig. III.7): Su trayectoria, muy parecida a la del centroide de su grupo y prácticamente igual que la de Alicante hasta el año 2002, se caracteriza por una fuerte recesión entre 2002 y 2003 que no se aprecia en el resto de provincias de su grupo. Salvo en la cuarta y última etapa, la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Las Palmas (Fig. III.8): Su trayectoria, aunque parecida tanto a la del centroide de su propio grupo como a la de Alicante, se caracteriza porque, hasta el año 2000, tanto los episodios de *boom* como los de recesión se producen por adelantado. En los tres últimos años del periodo de observación su trayectoria es prácticamente igual que la de España, lo que implica que toma valores, en términos relativos, más altos que la de Alicante. Salvo en lo que se refiere a la etapa final, sería necesario retardar la trayectoria para que quedara bien resumida por los puntos de corte de los centroides iniciales.

Almería (Fig. III.9): Su trayectoria, aunque bastante parecida tanto a la del centroide de su grupo como a la de Alicante, se caracteriza porque a principios de 2004 sufre un *boom* absolutamente desorbitado, no sólo en comparación con las provincias de su propio grupo, sino en comparación con todas las provincias españolas; por otro lado, entre 1987 y 1990 toma, en comparación con las restantes provincias de su grupo, valores relativamente bajos, lo que implica que en la fase inicial del periodo de observación sea más parecida a España. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

6.5.2 Descripción de las provincias clasificadas en el cuarto grupo

Recordemos (Fig. 6.2, *Centro Dcha.*) que el cuarto centroide final tiene una trayectoria muy parecida a la quinto. La mayor diferencia entre ambas se encuentra en la primera etapa (1987-1993), en la que tanto el *boom* como la recesión son más suaves en el caso del cuarto. También el *boom* de la tercera etapa (1997-2001) es más suave pero, en contrapartida, el *boom* de la cuarta (2002-2008) es más intenso. Las provincias del cuarto grupo son parecidas entre sí y, en consecuencia, parecidas al centroide final del grupo. Veamos si siguen esta misma pauta de comportamiento.

La Rioja (Fig. III.10): Su trayectoria, muy parecida tanto a la del centroide de su propio grupo como a la de España, se caracteriza por tomar valores relativamente altos en 2001 y entre finales de 2006 y principios de 2007; además, la recesión final se produce con algo de retraso en comparación con la mayor parte de provincias españolas. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Lugo (Fig. III.11): Su trayectoria, parecida tanto a la del centroide de su propio grupo como a la de Córdoba, se caracteriza por valores relativamente bajos entre 2002 y 2004 y, fundamentalmente, por un fuerte *boom* entre 2005 y 2006 que comporta una recesión

en los dos últimos años del periodo de observación aún mayor que la observada en la mayor parte de las provincias españolas.

Castellón/Castelló (Fig. III.12): Su trayectoria, aunque bastante parecida a la del centroide de su propio grupo, se caracteriza por valores relativamente altos entre 1989 y 1990, más en consonancia con la trayectoria de Alicante y de otras provincias clasificadas en el quinto grupo; además, tanto el *boom* de finales de los ochenta como la recesión del final del periodo de observación se produce con algo de retraso en comparación con el grueso de las provincias españolas. En términos generales la trayectoria (o una función adelantada de la misma) queda bastante bien resumida por los puntos de corte de los centroides iniciales.

León (Fig. III.13): Su trayectoria, parecida tanto a la del centroide de su propio grupo como a la de España y, desde 1998, a la de Córdoba, se caracteriza por valores relativamente altos entre 1985 y 1988 y relativamente bajos entre 2001 y 2006. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Zaragoza (Fig. III.14): Su trayectoria, aunque parecida a la del centroide de su propio grupo, se caracteriza por un nivel estable sin apenas oscilaciones cíclicas salvo en los cuatro últimos años del periodo de observación, en los que partiendo de niveles inferiores a su grupo se produce un fuerte *boom* (de la segunda mitad de 2004 a finales de 2006) seguido de una fuerte recesión (2007-2008).

Zamora (Fig. III.15): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias; a pesar de ello su trayectoria esta bien representada tanto por la del centroide de su grupo como por la de España o la de Córdoba. En comparación con la de Córdoba, se caracteriza por valores relativamente altos hasta 1992 y relativamente bajos entre 1993 y 2001.

Murcia (Fig. III.16): Su trayectoria, aunque parecida a la del centroide final de su grupo, se caracteriza por su fuerte similitud con el sexto centroide inicial en la primera mitad del periodo de observación (hasta 1996), por valores relativamente bajos entre 1997 y 2001 y, fundamentalmente, por un fuerte *boom* entre 2002 y 2004.

Asturias (Fig. III.17): Su trayectoria, no demasiado parecida a la del centroide de su grupo, se caracteriza por su fuerte similitud con el sexto centroide inicial hasta el año 2000 (salvo entre 1987 y 1989). Obsérvese que el fuerte *boom* que comienza en 1987 en Alicante, en el caso de Asturias se retrasa a 1989. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés. Podría decirse de ella que resume las diferencias y similitudes entre Córdoba, Alicante y España.

Orense/Ourense (Fig. III.18): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias; a pesar de ello su trayectoria esta bien representada tanto por la del centroide de su

grupo como por la de España o la de Córdoba. Se caracteriza por valores relativamente altos hasta 1995 y relativamente bajos desde 1996.

Cantabria (Fig. III.19): Su trayectoria, no demasiado parecida a la del centroide de su grupo, se caracteriza por su fuerte similitud con la del sexto centroide inicial hasta mediados del 2007 (la recesión final se produce con algo de retraso lo que hace que al final del periodo de observación se separe). En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Álava/Araba (Fig. III.20): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias; a pesar de ello su trayectoria está bien representada tanto por la del centroide de su grupo como por la de España o la de Córdoba y, en términos generales, queda bastante bien resumida por los puntos de corte de los centroides iniciales.

6.5.3 Descripción de las provincias clasificadas en el tercer grupo

Recordemos (Fig. 6.2, *Centro Izqda.*) que el tercer centroide final sigue, con pequeñas fluctuaciones acordes con los episodios de *boom* y recesión detectados en la serie promedio, una trayectoria de tendencia creciente a lo largo de prácticamente todo el periodo de análisis, llegando a triplicar el valor de partida a mediados del año 2006. En los dos últimos años, tras un pequeño repunte en el segundo semestre de 2006, decrece con fuerza hasta recuperar el nivel del año 1983. Las provincias del tercer grupo son parecidas entre sí y, en consecuencia, parecidas al centroide final del grupo. Veamos si siguen esta misma pauta de comportamiento.

Córdoba (Fig. III.21): Recordemos que la trayectoria de Córdoba, muy parecida a la del centroide de su propio grupo, se caracteriza por la debilidad del primer episodio de *boom* que, por otro lado, se produce con un cierto retraso respecto del de Alicante y España; además, tanto la tercera recesión (2000-2001), como el cuarto *boom* (2002-2006), son más pronunciados que en España; sin embargo, la cuarta recesión (2007-2008) parece haberse frenado en niveles por encima de los de Alicante y España. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés.

Valencia (Fig. III.22): Su trayectoria, aunque parecida tanto a la del centroide de su grupo como a la de Córdoba, se caracteriza por ser muy parecida a la de España; en consecuencia, en términos generales queda muy bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés.

Huelva (Fig. III.23): Su trayectoria, aunque parecida a la del centroide de su grupo, se caracteriza por ser prácticamente igual que la de España, salvo en la etapa delimitada por los años 1999 y 2002, en la que toma valores más altos, más en consonancia con una fuerte sensibilidad frente al tipo de interés en la segunda mitad del periodo de

observación. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales.

Madrid (Fig. III.24): Su trayectoria, bastante parecida a la del centroide de su grupo y a la de Córdoba, se caracteriza por valores relativamente altos en los trimestres próximos al tercero de 1995 y relativamente bajos en el comprendido entre 2005 y 2007; a partir del año 1999 es muy parecida a la de Alicante y, como en ésta, la recesión final se produce con algo de adelanto en comparación con la mayor parte de provincias españolas.

Guipúzcoa/Gipuzkoa (Fig. III.25): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias; a pesar de ello su trayectoria es prácticamente igual que la de Córdoba; la diferencia más notoria se detecta en los trimestres próximos al tercero de 1995, en los que, como Madrid, toma valores relativamente altos. Aunque el punto de corte previo a la recesión final no haya sido alcanzado, en términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están bastante en consonancia con las fluctuaciones del tipo de interés.

Navarra (Fig. III.26): Su trayectoria, prácticamente igual que la de Córdoba, se caracteriza frente a ésta por la mayor fuerza del primer y del cuarto *booms*. En términos generales queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Segovia (Fig. III.27): Su trayectoria es bastante parecida tanto a la del centroide del grupo, como a la de España y a la de Córdoba. En términos generales queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Badajoz (Fig. III.28): Su trayectoria, bastante parecida tanto a la del centroide del grupo como a la de Córdoba, se caracteriza, en comparación con ellas, por valores relativamente altos entre 1992 y 1994 y entre el tercer cuatrimestre de 2006 y el segundo de 2007. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Valladolid (Fig. III.29): Su trayectoria, aunque parecida a la de Córdoba, en los trimestres próximos al tercero de 1995 toma, al igual que Madrid, valores relativamente altos; pero, a diferencia de ésta, también en el año 2004 toma valores claramente por encima de los de Córdoba. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Granada (Fig. III.30): Su trayectoria, muy parecida tanto a la de Córdoba como a la de España, se caracteriza por tomar valores extremadamente altos en los cuatro trimestres comprendidos entre el tercero de 2006 y el segundo de 2007. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Jaén (Fig. III.31): Su trayectoria, parecida tanto a la de Córdoba como a la de España, se caracteriza por valores relativamente bajos entre 2003 y 2005 y relativamente altos entre el tercer cuatrimestre de 2006 y el segundo de 2007. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Sevilla (Fig. III.32): Su trayectoria, muy parecida a la de Córdoba hasta 1997, se caracteriza por su fuerte similitud con el sexto centroide inicial desde 1998⁷. En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Palencia (Fig. III.33): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias; a pesar de ello su trayectoria está bien representada por Córdoba.

Barcelona (Fig. III.34): Su trayectoria, parecida a la de España, se caracteriza por valores relativamente altos en la parte central del periodo de observación (entre 1995 y 2000). En términos generales la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

La Coruña/A Coruña (Fig. III.35): Su trayectoria, parecida a la de España, se caracteriza por valores relativamente altos en 1999 y en 2002. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés.

Vizcaya/Bizkaia (Fig. III.36): Su trayectoria, parecida a la de Córdoba hasta el año 1995, se caracteriza por valores relativamente altos entre 1996 y 2000, y relativamente bajos en 2004 y 2005, lo que, en comparación con la mayor parte de provincias clasificadas en su mismo grupo, hace que diste bastante del centroide.

Cáceres (Fig. III.37): Su trayectoria, aunque parecida a la de Córdoba desde el año 1995, hasta 1994 son muy distintas, lo que, en comparación con la mayor parte de provincias clasificadas en su mismo grupo, hace que difiera del centroide final del grupo.

Cádiz (Fig. III.38): Su trayectoria, prácticamente igual que la de España hasta 1996, año a partir del que se caracteriza por una fuerte sensibilidad frente al tipo de interés y por un crecimiento más acusado que el de España; además, parece que frente al resto de provincias su respuesta frente a las fluctuaciones del tipo de interés se produce por adelantado⁸. La consecuencia de esta sensibilidad es que entre 1997 y 2005 sus valores están por encima, no sólo de los de España, sino también de los del centroide, y en los

⁷ En cualquier caso la interpretación de esta provincia no es fiable en la etapa final, dado que sus valores no son los reales, sino que proceden de la estimación en términos de Andalucía (véase el Apéndice 1 del Capítulo 4)

⁸ Recuérdese que el conjunto de series resumen ha sido creado a partir de la aplicación de un desfase de cuatro trimestres sobre la serie del tipo de interés.

tres últimos años por debajo. Así, en comparación con la mayor parte de provincias clasificadas en su mismo grupo, es muy distinta del centroide. Salvo en lo que se refiere a la etapa inicial, sería necesario retardar la trayectoria para que quedara bien resumida por los puntos de corte de los centroides iniciales.

Salamanca (Fig. III.39): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias; a pesar de ello su trayectoria está bien representada por Córdoba, caracterizándose por un fuerte *boom* entre 1984 y 1989 y por valores relativamente altos entre 1991 y 1998; estos dos aspectos hacen que, en comparación con la mayor parte de provincias clasificadas en su mismo grupo, difiera del centroide.

6.5.4 Descripción de las provincias clasificadas en los grupos primero y segundo

Dado que el primer grupo está formado por sólo una provincia, Ciudad Real, y que, en consecuencia, la trayectoria de su centroide y la de Ciudad Real son una misma cosa, para completar su descripción la compararemos con las provincias del grupo más próximo, el segundo.

Recordemos que la trayectoria del segundo centroide final (**Fig. 6.2, Sup. Dcha.**), aunque muy parecida a la del tercero en las dos etapas centrales, en la cuarta se caracteriza por un fuerte *boom* en el periodo de aproximadamente cinco años comprendido entre principios de 2002 y mediados de 2006; en el caso del primer centroide (**Fig. 6.2, Sup. Izqda.**), este *boom* que precede a la recesión final se limita a los dos años comprendidos entre 2005 y 2006, y es todavía más acusado. Ésta no es la única gran diferencia frente a los restantes centroides finales: hasta 1991 la trayectoria del primer centroide final toma valores, en términos relativos, muy bajos. Los valores del segundo centroide, aunque no tanto como los del primero, también son bajos en esta fase inicial. Las provincias del segundo grupo son muy parecidas entre sí y, en consecuencia, parecidas al centroide final del grupo. Veamos si siguen esta misma pauta de comportamiento.

Ávila (Fig. III.40): Su trayectoria, muy parecida a la de Córdoba en el periodo comprendido entre 1990 y 2002, se caracteriza, al igual que la mayor parte de las provincias de su grupo, por un fuerte *boom* entre 2003 y 2005 seguido de una fuerte recesión.

Cuenca (Fig. III.41): Su trayectoria, muy parecida a la de Córdoba hasta el año 2002, se caracteriza, al igual que la mayor parte de las provincias de su grupo, por un fuerte *boom* entre 2003 y 2006 seguido de una fuerte recesión.

Albacete (Fig. III.42): Su trayectoria, muy parecida a la de Córdoba hasta el año 2002, se caracteriza, al igual que la mayor parte de las provincias de su grupo, por un fuerte *boom* entre 2003 y 2006 seguido de una fuerte recesión.

Teruel (Fig. III.43): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias, aún así su trayectoria queda bien resumida por la de Córdoba hasta el año 2002. Entre 2004 y 2006, al igual que la mayor parte de las provincias de su grupo aunque más débilmente, se caracteriza por un fuerte *boom* seguido de una fuerte recesión.

Burgos (Fig. III.44): Su trayectoria, aunque muy parecida a la del centroide de su grupo o a la de Córdoba en la primera mitad del periodo de observación, en la segunda se parece más a la de España. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales; además, desde el año 1993 sus oscilaciones están muy en consonancia con las fluctuaciones del tipo de interés.

Lleida (Fig. III.45): Su trayectoria, aunque bastante parecida a la del centroide de su grupo o a la de Córdoba hasta 2002, se caracteriza por su gran similitud con la de España en el periodo comprendido entre 1990 y 2001 y por un fuerte *boom* entre 2002 y 2006 seguido de una fuerte recesión. En términos generales la trayectoria queda muy bien resumida por los puntos de corte de los centroides iniciales.

Pontevedra (Fig. III.46): Su trayectoria, aunque parecida a la del centroide de su grupo, se caracteriza, en comparación con las provincias de su propio grupo, por valores relativamente bajos entre 2005 y 2008. Este hecho implica que desde 1993 sea más parecida tanto a la de Córdoba como a la de España.

Toledo (Fig. III.47): Aunque en la primera mitad del periodo de observación su trayectoria es muy parecida tanto a la del centroide del grupo como a la de Córdoba, se caracteriza por valores relativamente bajos entre 1996 y 2002 y por un fuerte *boom* entre 2002 y 2006 seguido de una fuerte recesión.

Guadalajara (Fig. III.48): Hasta 2001 su trayectoria es bastante parecida tanto a la del centroide de su grupo como a la de Córdoba; entre 2002 y 2005, al igual que la mayor parte de las provincias de su grupo, se caracteriza por un fuerte *boom* seguido de una fuerte recesión. Salvo en lo que se refiere al último tercio del periodo de observación, la trayectoria queda bastante bien resumida por los puntos de corte de los centroides iniciales.

Soria (Fig. III.49): La pequeña magnitud de sus valores hace que las fluctuaciones sean más bruscas que las observadas en la gran mayoría de las provincias. Distinta tanto de las provincias de su grupo como del resto, se caracteriza por un fuerte *boom* en 1990, seguido de una fuerte recesión entre 1991 y 1993, seguido a su vez por otro fuerte *boom* entre 1994 y 1995; desde 1996 mantiene un nivel constante, aunque con bruscas fluctuaciones en torno a él.

Ciudad Real (Fig. III.50): Su trayectoria, distinta tanto de las provincias del segundo grupo como del resto, se caracteriza por valores relativamente bajos entre 1984 y 1991, relativamente altos entre 1993 y 1998 y por un fuerte *boom* entre 2005 y 2006 seguido de una fuerte recesión.

6.5.5 Conclusiones generales

Las trayectorias de Málaga y de la mayor parte de las provincias clasificadas en el quinto grupo: Alicante, Huesca, Tarragona, Girona, Santa Cruz de Tenerife, Baleares, y Las Palmas (**Fig. 6.8, Sup. Izqda.**), se caracterizan por un fuerte *boom* en la segunda mitad de los ochenta, seguido de una fuerte recesión. Almería, aunque clasificada en el quinto grupo, tiene una trayectoria que, al igual que varias de las provincias clasificadas en el primer grupo, se caracteriza más por el fuerte *boom* de la última etapa del periodo de observación, que por el primero.

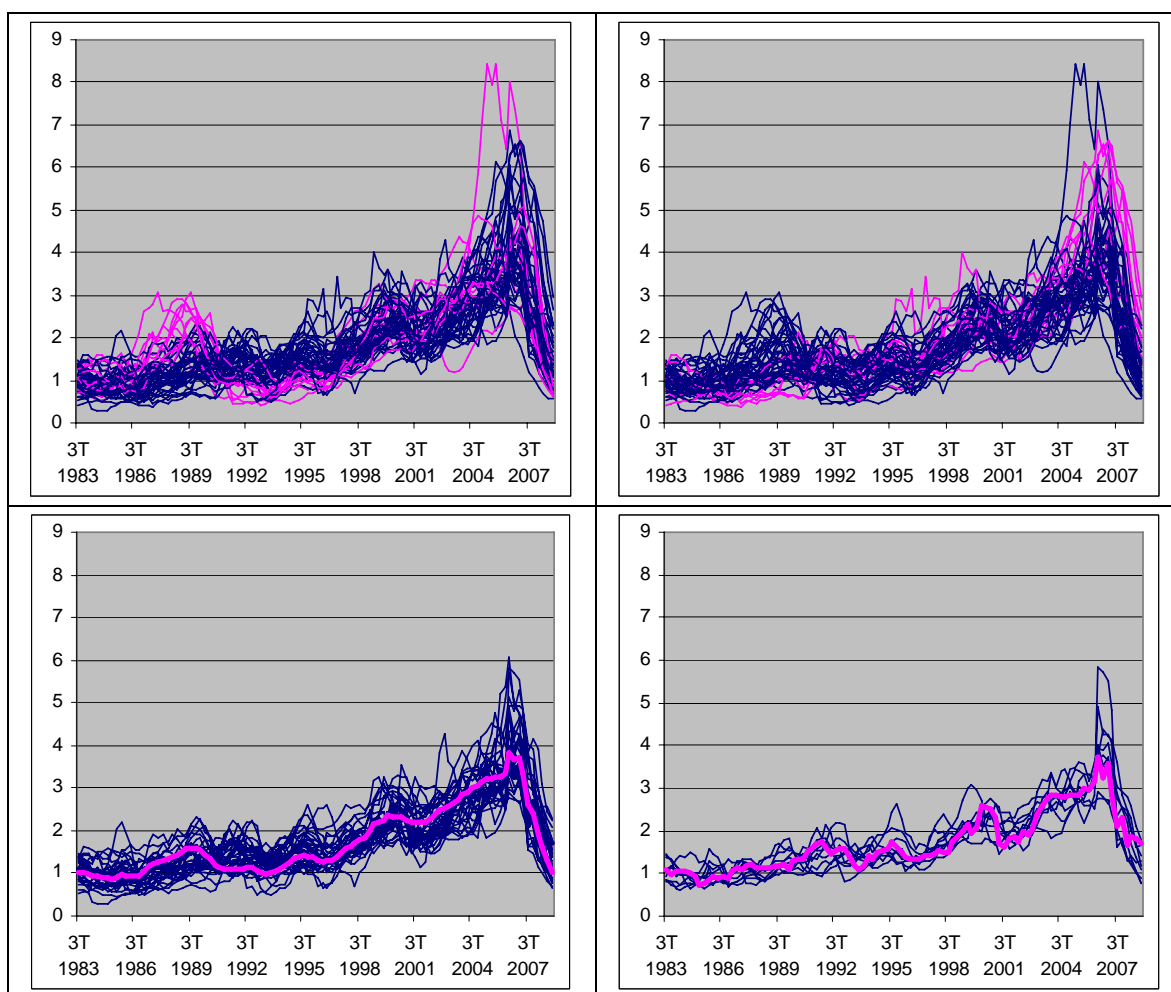


Figura 6.8 Series provinciales en valores homogeneizados. *Sup. izqda*: Resaltadas en rojo Málaga y las provincias clasificadas en el quinto grupo; *Sup. dcha*: Resaltadas en rojo Ciudad Real y las provincias clasificadas en el primer grupo; *Inf. izqda*: Descartadas las provincias que forman los dos primeros grupos, salvo Burgos, las que forman los dos últimos, y Murcia y Castellón, con España resaltada en rojo; *Inf. dcha*: Provincias muy parecidas a Córdoba (resaltada en rojo)

Las trayectorias de Ciudad Real y de la mayor parte de las provincias clasificadas en el primer grupo se caracterizan por presentar valores relativamente bajos hasta 1991: Pontevedra y Soria; o por valores relativamente altos entre 2002 y 2006: Cuenca, Teruel, Lleida, Toledo y Guadalajara; o por ambos aspectos: Ávila, Albacete y Ciudad Real (**Fig. 6.8, Sup. Dcha.**). Burgos es la única provincia del segundo grupo cuya trayectoria no se caracteriza por ninguno de estos dos aspectos, siendo muy parecida a la de Córdoba⁹.

Murcia y Castellón, aunque clasificadas en el cuarto grupo, tienen una trayectoria que, al igual que varias de las provincias clasificadas en el quinto grupo, se caracteriza por presentar valores relativamente altos entre 1987 y 1990 seguidos de valores relativamente bajos entre 1991 y 1997. Por otro lado, al igual que varias de las provincias clasificadas en el segundo grupo, presentan valores relativamente altos entre 2002 y 2006.

Descartadas las provincias que forman los dos primeros grupos (salvo Burgos), las que forman los dos últimos y Murcia y Castellón, queda un conjunto formado por veintinueve provincias: Córdoba, Valencia, Huelva, Madrid, Guipúzcoa, Navarra, Segovia, Badajoz, Valladolid, Granada, Jaén, Sevilla, Palencia, Barcelona, La Coruña, Vizcaya, Cáceres, Cádiz, Salamanca, La Rioja, Lugo, León, Zaragoza, Zamora, Asturias, Orense, Cantabria, Álava y Burgos, que no se caracterizan por ninguno de los aspectos previamente resaltados. Aunque globalmente este conjunto está bien representado por la trayectoria de España (**Fig. 6.8, Inf. Izqda.**), es de destacar la existencia de un subconjunto formado por diez provincias muy parecidas entre sí y muy bien representadas por Córdoba: Valencia, Madrid, Guipúzcoa, Navarra, Segovia, Badajoz, Granada, Jaén, Burgos y la propia Córdoba (**Fig. 6.8, Inf. Dcha.**).

⁹ Aunque, de entre las distancias de Burgos a cada uno de los seis centroides finales, la mínima es la correspondiente al segundo, obsérvese (Fig. II.6) que el valor de la distancia al tercero es prácticamente el mismo.

CAPÍTULO 7

CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

A partir de la información en un conjunto de J series temporales, con J grande, hemos construido un conjunto de K series resumen, con K pequeño, tal que no sólo nos ha permitido crear una partición de las J series temporales en K grupos homogéneos en términos de sus trayectorias, sino que, lo que posiblemente es mucho más interesante, nos ha simplificado la interpretación de cada trayectoria en términos de las restantes. Además, la relación entre las oscilaciones de las J series y las fluctuaciones de un determinado factor conocido, o serie covariable, nos ha permitido construir un segundo conjunto de series resumen para completar la interpretación de la agrupación obtenida y de las trayectorias de las J series temporales.

La aplicación de la metodología propuesta en esta tesis será adecuada siempre y cuando la estructura del conjunto de datos objeto de análisis responda a la de un haz de rectas. Éste será el primer aspecto al que dedicaremos nuestros esfuerzos en el futuro: buscar mecanismos que nos permitan decidir sobre la calidad del ajuste de un haz de rectas a los datos objeto de análisis.

La solución obtenida al considerar las series provinciales de inicio de vivienda es claramente dependiente tanto del número de grupos elegido como de la partición establecida en el rango de parámetros del haz de rectas. Aunque nuestro objetivo haya sido más el de ofrecer un mecanismo de ayuda a la hora de interpretar las similitudes y las diferencias entre las distintas provincias que el de agruparlas en un número concreto de grupos, también es cierto que sería deseable establecer algún tipo de criterio para fijar dicho número. La propia solución podría ayudar en este sentido: según la información proporcionada por la solución para una elección arbitraria del número de grupos, elegir un nuevo valor para el número de grupos y repetir el proceso de agrupación. También la propia solución podría llevar a la reconsideración de un nuevo

conjunto de valores dentro del rango de parámetros del haz de rectas para la generación del conjunto de curvas resumen. En definitiva, la búsqueda de mecanismos tanto para la elección del número de grupos como para fijar los valores que, dentro del haz de rectas, determinan la expresión de los centroides iniciales serán dos aspectos a los que nos gustaría prestar especial atención en trabajos futuros.

A la vista de los resultados obtenidos hemos interpretado los puntos de corte del conjunto de series resumen como cambios de ciclo de las series de visados. ¿Podría considerarse, en términos generales, que la metodología del haz de rectas ofrece un mecanismo de detección de cambios de ciclo para series temporales? Dar respuesta a esta pregunta es otro aspecto que nos parece relevante a la hora de concentrar nuestros esfuerzos futuros.

El desarrollo de la metodología se fundamenta en la existencia de una serie temporal responsable de las fluctuaciones del conjunto de series objeto de análisis, ¿podría generalizarse la metodología al caso de que en lugar de una fueran dos o más las series temporales responsables? Éste es otro gran interrogante que queda pendiente de respuesta.

En definitiva, a la hora de construir un modelo estadístico o económico que persiga objetivos de tipo explicativo o predictivo para un conjunto grande de series temporales, la aplicación de esta metodología puede ser una herramienta de gran utilidad como parte de un estudio exploratorio previo. La correcta interpretación de la agrupación obtenida o de las trayectorias de las series temporales dependerá de la calidad de la solución de K series resumen que, a su vez, dependerá de la adecuación del conjunto de datos a la estructura de haz de rectas para la extracción del conjunto de centroides iniciales. Desde el punto de vista teórico nuestros esfuerzos deberían concentrarse en refinar la metodología pero, a efectos de medir la utilidad de esta herramienta, la pregunta que surge de forma inmediata es ¿además del conjunto de datos objeto de análisis, existen otros conjuntos sobre los que funcionaría esta metodología? o, dicho de otra forma, ¿es generalizable esta metodología a otros problemas del mundo real?

E.J. Keogh y S. Kasetty (2002) realizan una revisión exhaustiva de las aportaciones realizadas en el campo de minería de datos para series temporales con el objetivo de advertir a la comunidad científica en este campo de investigación sobre el hecho de que gran parte de las propuestas tienen muy poca generalización a problemas del mundo real. Específicamente, en lo que se refiere a los problemas de clasificación o agrupación de series temporales, indican que en la mayor parte de los trabajos el objetivo perseguido es el de construir nuevas medidas de similitud entre series temporales para luego ser aplicadas en algoritmos de agrupación conocidos. Afirman que las propuestas realizadas no ofrecen mejores resultados que la sencilla y bien conocida distancia euclídea.

Para justificar esta afirmación comparan los resultados de la agrupación de series temporales con doce medidas de similitud distintas sobre dos conjuntos de datos

ampliamente referenciados en la literatura¹ y para los que el grupo de pertenencia de las distintas series es conocido. Concluyen que ninguna de las medidas mejora la solución proporcionada por la distancia euclídea; más aún, afirman que la mayoría de las distancias proporcionan el mismo resultado que si la asignación hubiera sido realizada aleatoriamente. En definitiva consideran que, aunque el resultado de este experimento no implica necesariamente que las medidas de similitud no tengan su interés, uno debe preguntarse sobre la contribución de una nueva medida que no pueda demostrar su utilidad.

Como colofón a su trabajo incluyen las siguientes sugerencias concretas para todos aquellos investigadores que trabajan en el campo de la minería de datos con series temporales (Keogh y Kasetty, 2002: 367-368):

- *Algorithms should be tested on a wide range of datasets, unless the utility of the approach is only been claimed for a particular type of data. If possible, one subset of the datasets should be used to fine tune the approach, then a different subset of the datasets should be used to do that the actual testing. This methodology is widely used in the machine learning community to help prevent implementation and data bias (Cohen, 1993).*
- *Where possible, experiments should be designed to be free of the possibility of implementation bias. Note that this does not preclude the addition of extensive implementation testing.*
- *Novel similarity measures should be compared to simple strawmen, such as Euclidean distance or Dynamic Time Warping. Some subjective visualization, or objective experiments should justify their introduction.*
- *Where possible, all data and code used in the experiments should be made freely available to allow independent duplication of finding (Bay, 1999).*

En definitiva, a la hora de proponer una metodología, es fundamental demostrar su utilidad. Recogiendo tanto las advertencias como las sugerencias de E.J. Keogh y S. Kasetty, y con el objetivo de comprobar si la herramienta que se propone es generalizable a otros problemas del mundo real, nuestros futuros esfuerzos se concentrarán, en primer lugar, en demostrar que existen diversos conjuntos de datos cuya estructura puede ser resumida mediante un haz de rectas en la línea planteada en nuestra metodología, no sólo en el Sector de la Construcción, sino también en otros sectores económicos en general o en otros campos de investigación.

Para demostrar la utilidad de una nueva herramienta no basta con que solucione el problema para el que ha sido diseñada sino que, además, la solución debe ser, en términos de su calidad y/o de su sencillez, al menos tan buena como la proporcionada por otras herramientas. Por ello, en segundo lugar, nuestros futuros esfuerzos irán dirigidos a comparar nuestra metodología con otras propuestas existentes. En esta línea, el primer trabajo podría consistir en el análisis de los datos de visados de vivienda para

¹ Cylinder-Bell-Funnel y Control-Chart.

el mercado estadounidense. Recordemos la complejidad de la metodología aplicada por J.H. Stock y M.W. Watson a la hora de construir los centroides iniciales²:

This algorithm was repeated for multiple random starting values.

We undertook an initial cluster analysis to estimate the number of regions, in which the foregoing algorithm was used with 20,000 random starting values. Moving from 2 to 3 clusters reduced the value of the minimized objective function by approximately 10%, as did moving from 3 to 4 clusters. The improvements from 4 to 5, and from 5 to 6, were less, and for 6 clusters the number of states was as few as five in one of the clusters. Absent a statistical theory for estimating the number of clusters, and lacking a persuasive reason for choosing 6 clusters, we therefore chose $k = 5$.

We then estimated the composition of these 5 regions using 400,000 random starting values. We found that even after 200,000 starting values there were some improvements in the objective function, however those improvements were very small and the switches of states in regions involved were few. We then reestimated the regions for the 1970-1987 and 1988-2007 subsamples, using 200,000 additional random starting values and using the full-sample regional estimates as an additional starting value. (Stock y Watson, 2008: 13)

Si la estructura del conjunto de datos del mercado estadounidense respondiera a la de un haz de rectas, quizá la metodología propuesta en este trabajo simplificara la solución dada por Stock y Watson para la generación de los centroides iniciales.

² Véase el Apartado 2.3 del Capítulo 2.

BIBLIOGRAFÍA

- Agrawal, R.C., Faloutsos, C. y Swami, A.N. (1993): Efficient Similarity Search in Sequence Databases. *FODO*
- Agrawal, R.C., Lin, I., Sawhney, H.S. y Shim, K. (1995): Fast similarity search in the presence of noise, scaling and translation in time series databases. *Proceedings of the 21st International Conference on very large data bases (VLDB'95)*, 490-500
- Alcaide, A., Fernández Díaz, A. y Rodríguez Saiz, L. (1982): Análisis económico del sector de la construcción. *Colegio Universitario de Estudios Financieros (CUNEF), Madrid*
- Alon, J. Sclaroff, S., Kollios, G. y Pavlovic, V. (2003): Discovering clusters in motion time-series data. *In IEEE Computer Vision and Pattern Recognition Conference (CVPR)*
- Anderberg, M.R.(1973): Cluster analysis for applications. *Academic Press, New York*
- Astrahan, M.M. (1970): Speech Analysis by Clustering, Or the Hyperphone Method, *Stanford Artificial Intelligence Project Memorandum aim-124, Stanford, ca: Stanford University*
- Babu, G.P. y Murty, M.N. (1993): A near-optima initial seed value selection in k -means clustering using a genetic algorithm. *Pattern Recognition Lett.* **14**, 763-769
- Bagnall, A. J., Janacek, G. Iglesia, B. de y Zhang, M. (2003): Clustering time series from mixture polynomial models with discretised data. *In Proceedings of the second Australasian Data Mining Workshop*, 105-120

- Ball, G.H. y Hall, D.J. (1965): ISODATA, A novel Method of Data Analysis and Pattern Classification. *AD 699616. Stanford Res. Inst. Menlo Park, California*
- Bar-Joseph, Z., Gerber, G., Gifford, D. Jaakkola, T. y 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)*, 39-48
- Baragona, R. (2001): A simulation study on clustering time series with meta-heuristic methods. *Quad. Stat.* **3** 1–26
- Bay, S. (1999): UCI Repository of KD databases [<http://kdd.ics.uci.edu/>]. *Irvine, CA: University of California, Department of Information and Computer Science*
- Beran, J. y Mazzola, G. (1999): Visualizing the relationship between time series by hierarchical smoothing models. *J. Comput. Graph. Stat.* **8** (2) 213–238
- Berndt, D.J. y Clifford, J. (1996): Finding patterns in time series: A dynamic programming approach. *Advances in Knowledge discovery and data mining*, 229-248
- Bezdek, J.C. (1987) *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York and London
- Biernacki, C. Celeux, G. y Govaert, G. (2000): Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 719–725
- Bock, H.H. (2008): Origins and extension of the k -means algorithm in cluster analysis. *Electronic Journal for History of Probability and Statistics*, **4**(2)
- Bohte, Z., Cepar, D. y Kosmelj, K. (1980): Clustering of time series. *COMSTAT 1980, Proceeding in Computational Statistics*, 587-593
- Box, G.E.P., Jenkins, G.M. Reinsel, G.C. (1994): *Time Series Analysis: Forecasting and Control*. Singapore: Pearson Education Inc.
- Bradley, P.S. y Fayyad, U.M. (1998): Refining initial points for k -means clustering algorithm. *In: Sharlik, J. (Ed.) Proc. of the 15th International Conference on Machine learning (ICML'98)*, Morgan Kaufmann. San Francisco, CA, 91-99
- Cadez, I.V., Gaffney, S. y Smyth, P. (2000): A general probabilistic framework for clustering individuals and objects. *In Knowledge Discovery and Data Mining*, 140-149
- Cadez, I.V., Heckerman, D., Meek, C., Smyth, P. y White, S. (2000): Visualization of navigation patterns on a web site using model-based clustering. *In Knowledge Discovery and Data Mining*, 280-284

- Cardarelli, R., Igan, D. y Rebucci, A. (2008): La evolución del ciclo de la vivienda y las implicaciones para la política monetaria. *Perspectivas de la economía mundial* Cap. 3 FMI
- Chatfield, C. (1996): The analysis of time series. *5th edn* New York, NY.: Chapman and Hall
- Cohen, W. (1993): Efficient pruning methods for separate-and-conquer rule learning systems. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, Morgan Kaufmann: Chambéry, France, 988-994
- Cowpertwait, P.S.P. y Cox, T.F. (1992): Clustering population means under heterogeneity of variance with an application to a rainfall time series problem. *The Statistician*, **41**, 113-121
- Cuadras, C.M. (2007): Nuevos métodos de análisis multivariante. *CMC Editions, Barcelona*
- Daoud, M.B.A. y Roberts, S.A. (1996): New methods for the initialisation of cluster. *Pattern Recognition Lett.* 17(5), 451-455
- Deelers, S. y Auwatanamongkol, S.(2007): Enhancing K-Means algorithm with Initial Cluster Centers Derived from Data Partitioning along the Data Axis with the Highest Variance. *Proceeding of WASET*, **26**, 323-328
- Everitt, B.S. (1993): Cluster Analysis. *Edward Arnold, Ltd., London, UK*
- Fayyad, U.M., (1996): Data mining and knowledge discovery: Making sense out of data. *IEEE Expert*, **11**(5), 20-25
- Fayyad, U.M., Reina, C.A. y Bradley, P.S. (1998): Initialization of iterative refinement clustering algorithm. In: *Proceeding of the International Conference on Knowledge Discovery and Data Mining (KDD'98)*, New York, 194-198
- Ferraz, R. (2006): La Financiación del boom de la vivienda. *Papeles de Economía Española* **109**, 214-224
- Focardi, S.M. (2001): Clustering economic and financial time series: exploring the existence of stable correlation conditions. *Technical Report 2001-04, The Intertek Group*
- Fu, T.-C., Chung, F.-L., Ng, V. y Luk, R. (2001): Pattern discovery from stock time series using self-organizing maps. *KDD 2001 Workshop on Temporal Data Mining, August 26–29, San Francisco, 27–37*

- Gaffney, S. y 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. y Smyth, P. (2003): Curve clustering with random effects regression mixtures. In C. M. Bishop and B. J. Frey, editors, *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*
- Galeano, P. and Peña, D. (2000): Multivariate analysis in vector time series. *Resenhas*, **4**, 383-404
- Gan, G., Ma, C. y Wu, J. (2007): Data clustering: Theory, Algorithms and Applications. *ASA-SIAM Series on Statistics and Applied Probability*
- Gantert, C. (1994): Classification of trees via the linear state space model. *Biometrical Journal. Journal of Mathematica Methods in Biosciences*, **36**, 825-839
- García-Montalvo, J. y Mas, M. (2000): La vivienda y el sector de la construcción en España. *Editorial CAM*
- Geweke, J. (1977): The Dynamic Factor Analysis of Economic Time Series. In: D.J. Aigner and A.S. Goldberger, eds. *Latent Variables in Socio-Economic Models, North-Holland, Amsterdam*
- Golay, X., Kollias, S. Stoll, G., Meier, D., Valavanis, A. y Boesiger, P. (1998): A new correlation-based fuzzy logic clustering algorithm for fMRI. *Mag. Resonance Med.* **40** 249-260
- Golding, D.Q. y Kanellakis, P.C. (1995): On similarity queries for time series data: Constraint specification and implementation. *Proceedings of the 1st International Conference on Principles and Practice of Constraint Programming (CP'95)*, 137-153
- Gordon, A.D. (1999): Classification. *Chapman and hall, London*
- Goutte, C. Hansen, L.K., Liptrot M.G. y Rostrup, E. (2001): Feature-space clustering for fMRI meta-analysis. *Hum. Brain Mapping* **13**, 165–183.
- Goutte, C., Toft, P. y Rostrup, E. (1999): On clustering fMRI time series. *Neuroimage* **9** (3), 298–310
- Han, J., Kamber, M. (2001): Data Mining: Concepts and Techniques. *Morgan Kauffmann, San Francisco*, 346-389
- Hand, D. Mannila, H. y Smyth, P. (2001): Principles of data mining. *MIT Press*
- Hartigan, J.A. (1975): Clustering algorithms. *New York: John Wiley and Sons*

- He, J., Lan, M., Tan, C., Sung, S. y Low, H. (2004): Initialization of cluster refinement algorithms: A review an comparative study. *Proceedings of International Joint Conference on Neural Networks*
- Huang, C. y Harris, R. (1993): A comparison of several codebook generation approaches. *IEEETrans. Image Process.* **2**(1), 108-112
- Huhtala, Y., Karkkainen, J. y Toivonen, H. (1999): Mining for similarities in aligned time series using wavelets. *Data Mining and Knowledge Discovery: Theory, Tools, and Technology, Belur V. Dasarathy; Ed. Proc. SPIE Vol. 3695*, 150-160,
- Jain, A.K. y Dubes, R.C.(1988): Algorithms for Clustering data. *Prentice Hall, Englewood Cliffs*
- Jain, A.K., Murty, M.N. y Flynn, P.J. (1999): Data clustering: a Review. *ACM Computing Surveys*, **31**(3), 263-323
- Kakizawa, Y., Shumway, R.H. y Taniguchi, N. (1998): Discrimination and clustering for multivariate time series. *J. Amer. Stat. Assoc.* **93** (441), 328–340
- Kalpakis, K., Gada, D. y Puttagunta, V. (2001): Distance measures for effective clustering of ARIMA time-series. *Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA*, 273–280
- Katsavounidis, I., Kuo, C.C.J. y Zhen, Z. (1994): A new initialization technique for generalized Lloyd iteration. *Signal Process. Lett. IEEE*, **1**(10), 144-146
- Kaufman, I. y Rousseeuw, P.J.(1990): Finding groups in data: An introduction to Cluster Analysis, *Wiley, New York*
- Keogh, E.J. y Kasetty, S. (2002): On the need for time series data mining benchmarks: A survey an empirical demonstration. *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD'02)*, 102-111
- Keogh, E.J. y Pazzani, M.J. (2000): A Simple Dimensionality Reduction Technique for Fast Similarity Search in Large Time Series Databases. *PAKDD*, 122-133
- Khan, S.S. y Ahmad, A. (2004): Cluster center initialization algorithm for k -means clustering. *Pattern Recognition Lett.* **25**(11), 1293-1302
- Kosmelj, K. y Batagelj, V. (1990): Cross-sectional approach for clustering time varying data. *J. Classification*, **7** 99-109
- Krishnapuram, R., Joshi, A., Nasraoui, O. y Yi, L. (2001): Low-Complexity fuzzy relational clustering algorithms for web mining, *IEEE Trans. Fuzzy Systems* **9**(4) 595-607

- Kumar, M., Patel, N.R. y Woo, J. (2002): Clustering Seasonality Patterns in the Presence of Errors. *Proceedings of KDD '02, Edmonton, Alberta, Canada*
- Laxman, S. y Sastry, P.S.: (2006): A survey of temporal data mining. *Indian Institute of Science. Bangalore, India Sadhana*, Vol. 31, Part. 2 pp. 173-198
- Lance, G.N. y Williams, W.T. (1967): A General Theory of Classificatory Sorting Strategy. 1. Hierarchical Systems. *Comp. Journal*, **9**(1), 60-64
- Leamer, E. E. (2007): Housing is the Business Cycle. *NBER Working Papers 13428, National Bureau of Economic Research, Inc.*
- Li, C. (2000): A Bayesian approach to temporal data clustering using the hidden Markov model methodology. *PhD thesis, Vanderbilt University, Nashville*
- Li, C. y Biswas, G. (1999a): Clustering sequence data using hidden markov model representation. *In SPIE'99 Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology*, 14-21
- Li, C. y 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. y Biswas, G. (1999c): Temporal pattern generation using hidden Markov model based unsupervised classification. *in: D.J. Hand, J.N. Kok, M.R. Berthold (Eds.), Lecture Notes in Computer Science, vol. 164, IDA '99, Springer, Berlin, 245–256*
- Li, C. y Biswas, G. (2000): Bayesian clustering for temporal data using hidden markov model representation. *In proceedings of the Seventeenth International Conference on Machine Learning*, 543-550
- Li, C., Biswas, G. Dale, M. y Dale, P. (2001): Building models of ecological dynamics using HMM based temporal data clustering—a preliminary study. *in: F. Hoffmann et al. (Eds.), IDA 2001, Lecture Notes in Computer Science, vol. 2189, 53–62*
- Liao T.W. (2005a): Clustering of time series data – A survey. *Pattern Recogn.* **38** 1857-1874
- Liao, T.W. (2005b): Mining of vector time series by clustering *Working paper.*
- Liao, T.W., Bolt, B. Forester, J. Hailman, E. Hansen, C. Kaste, R.C. y O'May, J. (2002): Understanding and projecting the battle state. *23rd Army Science Conference, Orlando, FL, December 2-5*

- Likas, A., Vlassis, N. y Verbeek, J.J. (2003): The global k -means clustering algorithm. *Pattern Recognition Lett.* **36**(5), 451-461
- Lin, J. VLachos, M. Keogh, E. Gunopulos, D. (2004): Iterative incremental clustering of time series. *In EDBT*
- Linde, Y., Buzo, A. y Gray, R.M.(1980): An algorithm for vector quantizer design. *IEEE Trans. Commun.* **28**, 84-95
- Lu, J.F., Tang, J.B., Tang, Z.M. y Yang, J.Y. (2008): Hierarchical initialization approach for K -Means clustering. *Pattern Recognition Lett.* **29**, 787-795
- Macchiato, M.F. La Rotonda, L., Lapenna, L. y Ragosta, M. (1995): Time modelling and spatial clustering of daily ambient temperature: An application in Southern Italy. *EnvironMetrics*, **6**, 31-53
- MacQueen, J. (1967): Some methods for classification and analysis of multivariate observations, en: L.M. Le Cam, J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 281-297
- Maharaj, E.A. (1996) A significance test for classifying ARMA models. *Journal of Statistical Computation and Simulation*, **54**, 305-331
- Maharaj, E.A. (2000): Clusters of time series. *J. Classification* **17** 297–314
- McRae, D.J. (1971): MIKCA: A FORTRAN IV Iterative K -Means Cluster Analysis Program. *Behavioral Science*, **16**(4), 423-424
- Möller-Levet, C.S., Klawonn, F., Cho, K.-H. y Wolkenhauer, O. (2003): Fuzzy clustering of short time series and unevenly distributed sampling points. *Proceedings of the 5th International Symposium on Intelligent Data Analysis, Berlin, Germany, August 28–30*
- Milligan, G.W. (1980): An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, **45**, 325-342
- Milligan, G.W. y Cooper, M. (1988): A study of standardization of variables in cluster analysis. *Multivariate Behavioral Research*, **24**, 163-176
- Muth, R.F. (1988), Housing Market Dynamics. *Regional Science and Urban Economics* **18**, 345-356
- Oates, T. (1999): Identifying distinctive subsequences in multivariate time series by clustering. *In S. Chaudhuri and D. Madigan, editors, Fifth International Conference on Knowledge Discovery and Data Mining*, 322-326, San Diego, CA, USA

- Oates, T., Firoiu, L. y Cohen, P.R. (1999): Clustering time series with hidden Markov models and dynamic time warping. *Proceedings of the IJCAI-99 Workshop on Neural, Symbolic, and Reinforcement Learning Methods for Sequence Learning*
- Oates, T., Firoiu, L. y Cohen, P.R. (2001): Using dynamic time warping to bootstrap HMM-based clustering of time series. *Lecture Notes in Computer Science*, **1828**, 35-52
- Oates, T., Schmill, M.D. y Cohen, P.R. (2000): Identifying qualitatively different outcomes of actions: Gaining autonomy through learning. In *Carles Sierra, Maria Gini, and Jeffrey S. Rosenschein, editors, Proceedings of the Fourth International Conference on Autonomous Agents*, 110-111, Barcelona, Catalonia, Spain
- Ormerod, P. y Mounfield, C. (2000): Localised structures in the temporal evolution of asset prices. *New Approaches to Financial Economics, Santa Fe Conference*
- Owsley, L.M.D., Atlas, L.E. y Bernard, G.D. (1997): Self-organizing feature maps and hidden Markov models for machine-tool monitoring. *IEEE Trans. Signal Process.* **45** (11), 2787–2798
- Peña, D. (2002): Análisis de datos multivariantes. *McGraw Hill, Madrid*
- Peña, J.M. Lozano, J.A. y Larrañaga, P. (1999): An Empirical comparison of four initialization methods for the *K*-Means algorithm. *Pattern Recognition Lett.* **20**, 1027-1040
- Piccolo, D. (1990): A distance measure for classifying *ARMA* models. *J. Time Ser. Anal.* **11** (2), 153–163
- Policker, S. y Geva, A.B. (2000): Nonstationary time series analysis by temporal clustering. *IEEE Trans. Syst. Man Cybernet.-B: Cybernet.* **30** (2), 339–343
- Pozdena, R. J. (1990): Do interest rates still affect housing? *Federal Reserve Bank of San Francisco. Economic Review*, 3-14
- Ramoni, M., Sebastiani P. y Cohen, P. (2000): Multivariate clustering by dynamics. *Proceedings of the 2000 National Conference on Artificial Intelligence (AAAI-2000)*, San Francisco, CA, 633–638
- Ramoni, M., Sebastiani P. y Cohen, P. (2002): Bayesian clustering by dynamics. *Mach. Learning* **47** (1), 91–121
- Redmon, S.J. y Heneghan, C. (2007): A method for initialising the *K*-means clustering algorithm using *kd*-trees. *Pattern Recognition Lett.*, **28**, 965-973

- Ridgeway, G. (1997): Finite discrete Markov processes. *Technical Report MSR-TR-97-24, Microsoft Research*
- Rodríguez López, J. (2006) Los booms inmobiliarios en España. Un análisis de tres períodos. *Papeles de Economía Española* **109**, 76-90
- Rodríguez López, J. (2008) De la crisis inmobiliaria a la crisis financiera. El duro otoño de 2008 *Ciudad y Territorio: Estudios Territoriales* **158**, 751-774
- Shaw, C.T. y King, G.P. (1992): Using cluster analysis to classify time series. *Physica D* **58**, 288–298
- Shumway, R.H. (2003): Time–frequency clustering and discriminant analysis. *Stat. Probab. Lett.* **63**, 307–314
- Smyth, P. (1997) Clustering sequences with hidden markov models. In *Michael C. Mozer, Michael I. Jordan, and Thomas Petsche, editors, Advances in Neural Information Processing Systems*, **9**, 648
- Steinly, D. (2003): *K*-Means clustering: A half-century síntesis. *British journal of Mathematical and Statistical Psychology*, **59**, 1-34
- Steinly, D. y Brusco, M.J.(2007): Initializing *K*-means Batch Clustering: A critical Evaluation of Several Techniques. *Journal of Classification*, **24**, 99-121
- Stock, J.H. y Watson, M.W. (2008): The Evolution of National and Regional Factors in U.S. Housing Construction. *Robert F. Engle Festschrift Conference, San Diego, CA*
- Su, T. y Dy, J.G. (2007): In Search of Deterministic Methods for Initializing *K*-Means and Gaussian Mixture, *Intelligent Data Analysis*, **11**(4), 319-338
- Taltavull de la Paz, P. (2001): Economía de la construcción, *Civitas, Madrid*
- Taltavull de la Paz, P. (2004): Habitatge i finançament hipotecari: l'adaptació permanent. *Revista Económica de Catalunya* **42**, 94-111
- Taltavull de La Paz, P. (2006): La oferta de viviendas y el Mercado inmobiliario en España. *Papeles de Economía Española*. **109**, 156–181
- Taltavull de La Paz, P. (2007) La demanda de construcción en España. Razones y tendencias recientes. *Cuadernos de Información Económica* **197** 37–44
- Tong, P. y Dabas, H. (1990): Cluster of time series models: An example. *Journal of Applied Statistics*, **17**, 187-198

- Tou, J. y Gonzales, R. (1974): Pattern recognition Principles. *Addison Wesley, Reading, MA*
- Tran, D. y Wagner, M. (2002): Fuzzy c-means *clustering*-based speaker verification. in: *N.R. Pal, M. Sugeno (Eds.), AFSS 2002, Lecture Notes in Artificial Intelligence, 2275, 318–324*
- Van Wijk, J.J. y van Selow, E.R. (1999): Cluster and calendar based visualization of time series data. *Proceedings of IEEE Symposium on Information Visualization, San Francisco, CA, October 25-26*
- Vergés, R. y Martín, D. (2007) Series históricas de edificación residencial. Estadística de visados de los colegios de arquitectos 1960-2006. *Estadística Española Vol. 49* **166**, 563-595
- Vlachos, M. Lin, J., Keogh, E. y Gunopulos, D. (2003): A wavelet-based anytime algorithm for *k*-means clustering of time series. *Proceedings of the Third SIAM International Conference on Data Mining, San Francisco, CA*
- Walden, A.T. (1994): Spatial clustering: Using simple summaries of seismic data to find the edge of an oil-field. *Applied Statistics*, **43**, 385-398
- Wang, L., Mehrabi, M.G. y Kannatey-Asibu Jr., E. (2002): Hidden Markov model-based wear monitoring in turning. *J. Manufacturing Sci. Eng.* **124**, 651–658
- Wilpon, J.G. y Rabiner, L.R. (1985): Modified *k*-means clustering algorithm for use in isolated word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **33** (3), 587–594
- Wismüller, A., Lange, O., Dersch, D.R., Leinsinger, G.L., Hahn, K., Pütz B. y Auer, D. (2002): Cluster analysis of biomedical image time series, *Int. J. Comput. Vision* **46** (2), 103–128
- Xiong, Y. y Yeung, D.-Y. (2002): Mixtures of ARMA models for model-based time series clustering. *Proceedings of the IEEE International Conference on Data Mining, Maebaghi City, Japan*
- Zhao, Q. y Bhowmick, S. (2003): Sequential Pattern Mining *Technical Report. Nanyang Technological University. Singapore*
- Zhong, S. (2002): Probabilistic model-based clustering of complex data. *PhD thesis, University of Texas at Austin*
- Zhong, S. y Ghosh, J. (2002a): HMMs and coupled HMMs for multi-channel EEG classification. *In Proc. IEEE Int. Joint Conf. on Neural Networks*

Zhong, S. y Ghosh, J. (2002b): A unified framework for model-based clustering. *In Intelligent Engineering Systems Through Artificial Neural Networks (ANNIE)*

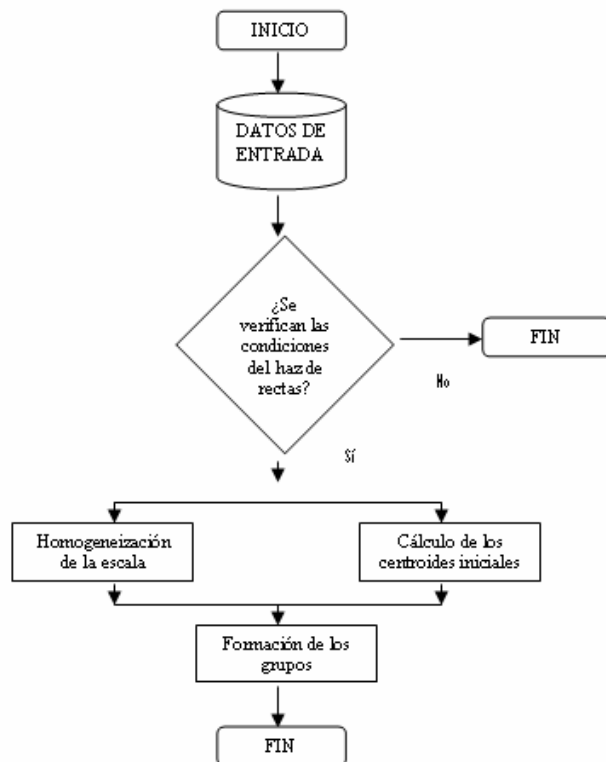
Zhong, S. y Ghosh, J. (2002c): A unified framework for model-based clustering and its application to clustering time sequences. *Technical report, Department of Electrical and Computer Engineering, University of Texas*

Zhong, S. y Ghosh, J. (2003): Scalable, balanced model-based clustering. *In Proceedings of SIAM Int. Conf. on Data Mining*

ANEXO I

UN ESQUEMA DE LA METODOLOGÍA

I.1 DIAGRAMA DE FLUJO



I.2 DATOS DE ENTRADA

- J series temporales observadas en T instantes de tiempo:

$$\{V_{t,j}, t=1,\dots,T\} \quad j=1,\dots,J$$

- La serie temporal promedio con referencia en el instante inicial Y_t :

$$Y_t = \frac{V_t}{V_1} \quad t=1,\dots,T$$

donde:

$$\ln V_t = \frac{\sum_{j=1}^J \ln V_{t,j}}{J} \quad t=1,\dots,T$$

- Una serie temporal X_t generadora de las fluctuaciones de las J series temporales $V_{t,j}$. Si dicha serie es desconocida o no se dispone de sus observaciones, aplicar la metodología sobre la serie Y_t

I.3 CONDICIONES DE APLICACIÓN

- Para $j=1,\dots,J$, cálculo de $A_{0,j}$, $A_{1,j}$ y $A_{2,j}$, coeficientes de la ecuación de regresión:

$$\ln \hat{V}_{t,j} = A_{0,j} \cdot t + A_{1,j} \cdot \ln X_t + A_{2,j}$$

estimada por mínimos cuadrados ordinarios.

Aplicaremos la metodología, cuando podamos suponer que la estructura que subyace en el conjunto de series temporales $\{V_{t,j}\}$ es la de haz de rectas. Si denominamos:

$$y_{t,j} = \frac{V_{t,j}}{V_{t-1,j}} \quad j=1,\dots,J \quad \text{y} \quad x_t = \frac{X_t}{X_{t-1}} \quad t=1,\dots,T$$

el conjunto de series $\ln y_{t,j}$ define un haz de rectas respecto de la serie $\ln x_t$ si existen dos coeficientes $\alpha_{0,j}$ y $\alpha_{1,j}$ tales que:

$$\ln y_{t,j} = \alpha_{0,j} + \alpha_{1,j} \cdot \ln x_t \quad \text{con} \quad \alpha_{0,j} = \beta_0 + \beta_1 \cdot \alpha_{1,j} \quad j=1,\dots,J$$

En consecuencia, será adecuado aplicar la metodología si:

- los coeficientes $A_{1,j}$ son significativamente distintos de cero aunque, excepcionalmente, alguno de ellos sea nulo; y
 - los J puntos de la forma $(A_{1,j}, A_{0,j})$ se concentran en torno a una recta¹.
- Para $j=1, \dots, J$, cálculo de B_0 y B_1 , coeficientes de la ecuación de regresión:

$$\hat{A}_{0,j} = B_0 + B_1 \cdot A_{1,j}$$

estimados por mínimos cuadrados ordinarios.

I.4 HOMOGENEIZACIÓN DE LA ESCALA

- Cálculo de α :

$$\alpha = \frac{\sum_{t=1}^T \ln Y_t}{T}$$

- Cálculo de α_j :

$$\alpha_j = \frac{\sum_{t=1}^T \ln V_{t,j}}{T} \quad j=1, \dots, J$$

- Transformación de las J series temporales originales:

$$Y_{t,j} = \exp(\ln V_{t,j} - \alpha_j + \alpha) \quad j=1, \dots, J \quad t=1, \dots, T,$$

I.5 CÁLCULO DE LOS CENTROIDES INICIALES

- Elección de K , número de grupos
- Fijar los coeficientes m_k , $k=1, \dots, K$:

¹ La pendiente de la recta puede ser nula; si así fuera, y los puntos se concentran en torno a la recta significaría que la componente determinística puede ser ignorada.

Por ejemplo, elegir dos valores dispersos, a y b , dentro del rango de variación de los valores $A_{1,j}$, $j=1,\dots,J$; hacer $\theta = (b-a)/(K-1)$ y considerar:

$$m_1 = a \quad \text{y} \quad m_k = m_{k-1} + \theta \quad k=2,\dots,K$$

- Cálculo de los coeficientes b_k , $k=1,\dots,K$:

$$b_k = B_0 + B_1 \cdot m_k \quad k=1,\dots,K$$

- Cálculo de $f_{t,k}$, $k=1,\dots,K$:

$$f_{t,k} = b_k + m_k \cdot \ln x_t, \quad t=2,\dots,T \quad \text{y} \quad f_{1,k} = b_k$$

donde:

$$x_t = \frac{X_t}{X_{t-1}} \quad t=2,\dots,T$$

- Cálculo de $g_{t,k}$, $k=1,\dots,K$:

$$g_{t,k} = g_{t-1,k} + f_{t,k}, \quad t=2,\dots,T \quad \text{y} \quad g_{1,k} = f_{1,k}$$

- Cálculo de β_k , $k=1,\dots,K$:

$$\beta_k = \frac{\sum_{t=1}^T g_{t,k}}{T}$$

- Cálculo de los centroides iniciales $C_{t,k}^I$, $k=1,\dots,K$:

$$C_{t,k}^I = g_{t,k} - \beta_k + \alpha \quad t=1,\dots,T$$

I.6 FORMACIÓN DE LOS GRUPOS

Aplicar el *Algoritmo de las k-medias* para agrupar las J series temporales $Y_{t,j}$ en K grupos. Utilizar en cada iteración la distancia $d_{k,j}^i$:

$$d_{k,j}^i = d(\ln Y_{t,j}, C_{t,k}^i) \quad k=1,\dots,K \quad j=1,\dots,J$$

siendo d la distancia euclídea y $C_{t,k}^i$ los K centroides correspondientes a la i -ésima iteración. La primera agrupación se hará en términos de los centroides iniciales:

$$d_{k,j}^1 = d(\ln Y_{t,j}, C_{t,k}^1) \quad k=1,\dots,K \quad j=1,\dots,J$$

En las sucesivas iteraciones:

$$C_{t,k}^i = \frac{\sum_{j(k)=1}^{n(k)} \ln Y_{t,j(k)}}{n(k)} \quad k=1,\dots,K$$

donde $n(k)$ es el total de series que contiene el conglomerado k -ésimo en la i -ésima iteración y $j(k)$ es el subíndice correspondiente a las series agrupadas en dicho conglomerado. A los centroides correspondientes a la solución final los denominaremos:

$$C_{t,k}^F \quad k=1,\dots,K \quad t=1,\dots,T$$

ANEXO II

APLICACIÓN DEL ALGORITMO DE LAS K - MEDIAS

Dado el conjunto de las J series temporales $V_{t,j}$, para agruparlas en K grupos mediante el Algoritmo de las K -medias consideraremos en cada iteración la distancia $d_{k,j}^i$:

$$d_{k,j}^i = d(\ln Y_{t,j}, C_{t,k}^{i-1}) \quad k=1,\dots,K \quad j=1,\dots,J$$

siendo d la distancia euclídea y $C_{t,k}^i$ los K centroides obtenidos en la i -ésima iteración. La primera agrupación se hará en términos de los centroides iniciales:

$$d_{k,j}^1 = d(\ln Y_{t,j}, C_{t,k}^I) \quad k=1,\dots,K \quad j=1,\dots,J$$

En las sucesivas iteraciones:

$$C_{t,k}^i = \frac{\sum_{j^{(k)}=1}^{n(k)} \ln Y_{t,j^{(k)}}}{n(k)} \quad k=1,\dots,K$$

donde $n(k)$ es el total de series que contiene el grupo k -ésimo en la i -ésima iteración y $j^{(k)}$ es el subíndice correspondiente a las series agrupadas en dicho grupo. A los centroides correspondientes a la solución final los denominaremos:

$$C_{t,k}^F \quad k=1,\dots,K \quad t=1,\dots,T$$

Concretamente, dados los centroides iniciales de los seis grupos (**Fig. II.1**):

$$\text{Grupo 1: } (C_{1,1}^1, \dots, C_{102,1}^1) = (\ln H_{1,1}^I, \dots, \ln H_{102,1}^I) = (-0.31, \dots, 1.24)$$

$$\text{Grupo 6: } (C_{1,6}^1, \dots, C_{102,6}^1) = (\ln H_{1,6}^I, \dots, \ln H_{102,6}^I) = (0.39, \dots, -1.14)$$

en la primera iteración del algoritmo de las k -medias se calcula la distancia de cada serie provincial a cada uno de los centroides iniciales de los seis grupos (**Fig. II.2**). Las provincias están clasificadas en seis bloques, uno por cada grupo, en términos de la mínima distancia, y dentro de cada bloque, están ordenadas de menor a mayor distancia al centroide del grupo correspondiente. Por ejemplo, considerando que Ciudad Real es la provincia número 18, su distancia a cada uno de los centroides de los seis grupos es:

$$\begin{aligned} d_{1,18}^1 &= d(\ln Y_{t,18}, C_{t,1}^I) = 3,8937 & d_{2,18}^1 &= d(\ln Y_{t,18}, C_{t,2}^I) = 4,2275 \\ d_{3,18}^1 &= d(\ln Y_{t,18}, C_{t,3}^I) = 4,7447 & d_{4,18}^1 &= d(\ln Y_{t,18}, C_{t,4}^I) = 5.3926 \\ d_{5,18}^1 &= d(\ln Y_{t,18}, C_{t,5}^I) = 6,1301 & d_{6,18}^1 &= d(\ln Y_{t,18}, C_{t,6}^I) = 6,9285 \end{aligned}$$

y, dado que:

$$\min_k d_{k,18}^1 = \min_k d(\ln Y_{t,18}, C_{t,k}^I) = d(\ln Y_{t,35}, C_{t,1}^I) = 3,8937$$

es clasificada en el primer grupo. Según la distancia de las restantes provincias a cada uno de los seis centroides iniciales la solución obtenida en la primera iteración es:

Grupo	Provincias
1	CIUDAD REAL
2	SEGOVIA, TERUEL, BADAJOZ, GUIPÚZCOA, VIZCAYA, ÁVILA, CUENCA, LA_CORUÑA, ALBACETE, CÁCERES, PONTEVEDRA, GUADALAJARA, TOLEDO, SORIA
3	VALENCIA, CÓRDOBA, LA RIOJA, BURGOS, LUGO, MADRID, NAVARRA, VALLADOLID, ZARAGOZA, PALENCIA, LLEIDA, BARCELONA, ÁLAVA, SALAMANCA
4	HUELVA, JAÉN, ASTURIAS, LEÓN, GRANADA, HUESCA, CASTELLÓN, SEVILLA, MURCIA, ZAMORA, CANTABRIA, SC_TENERIFE, CÁDIZ, ORENSE
5	LAS_PALMAS, TARRAGONA, GIRONA, ALICANTE, ALMERÍA, BALEARES
6	MÁLAGA

Con esta agrupación, se calculan los centroides correspondientes a la primera iteración. Por ejemplo, si consideramos las $n(k)=n(5)=6$ provincias del quinto grupo:

Provincia	Subíndice j correspondiente al conjunto de las $J=50$ provincias	Subíndice $j(k)$ correspondiente al grupo
LAS_PALMAS	14	1(5)
TARRAGONA	34	2(5)
GIRONA	33	3(5)
ALICANTE	44	4(5)
ALMERÍA	1	5(5)
BALEARES	13	6(5)

Centroides iniciales						
	Grupo					
	1	2	3	4	5	6
3T 1983	-,31	-,17	-,03	,11	,25	,39
4T 1983	-,29	-,17	-,05	,07	,19	,32
1T 1984	-,28	-,18	-,08	,02	,12	,21
2T 1984	-,26	-,19	-,11	-,04	,04	,11
3T 1984	-,25	-,18	-,12	-,05	,02	,08
4T 1984	-,23	-,19	-,14	-,10	-,05	-,01
1T 1985	-,22	-,18	-,14	-,11	-,07	-,04
2T 1985	-,20	-,17	-,15	-,12	-,10	-,07
3T 1985	-,18	-,16	-,13	-,11	-,08	-,06
4T 1985	-,17	-,11	-,05	,00	,06	,12
1T 1986	-,15	-,12	-,09	-,06	-,03	,00
2T 1986	-,14	-,13	-,12	-,10	-,09	-,08
3T 1986	-,12	-,12	-,12	-,11	-,11	-,10
4T 1986	-,11	-,13	-,16	-,18	-,20	-,23
1T 1987	-,09	-,08	-,06	-,05	-,03	-,02
2T 1987	-,08	-,02	,03	,09	,14	,20
3T 1987	-,06	,01	,07	,14	,21	,28
4T 1987	-,04	,03	,11	,19	,27	,35
1T 1988	-,03	,05	,14	,22	,30	,38
2T 1988	-,01	,07	,15	,24	,32	,41
....
1T 2005	1,02	1,08	1,14	1,19	1,25	1,31
2T 2005	1,04	1,11	1,18	1,25	1,32	1,39
3T 2005	1,05	1,12	1,20	1,27	1,34	1,41
4T 2005	1,07	1,14	1,22	1,30	1,37	1,45
1T 2006	1,08	1,16	1,23	1,31	1,38	1,46
2T 2006	1,10	1,17	1,24	1,31	1,38	1,46
3T 2006	1,11	1,25	1,39	1,53	1,67	1,80
4T 2006	1,13	1,23	1,34	1,44	1,55	1,65
1T 2007	1,14	1,25	1,35	1,46	1,56	1,67
2T 2007	1,16	1,22	1,28	1,34	1,40	1,46
3T 2007	1,17	1,10	1,04	,97	,90	,83
4T 2007	1,19	1,09	1,00	,90	,81	,71
1T 2008	1,20	1,00	,80	,60	,40	,20
2T 2008	1,22	,95	,69	,42	,16	-,11
3T 2008	1,23	,86	,50	,13	-,23	-,60
4T 2008	1,24	,77	,29	-,18	-,66	-,14

Figura II.1 Valores de $C_{t,k}^I$

	Distancias					
	1	2	3	4	5	6
CIUDAD_REAL	3,8937	4,2275	4,7447	5,3926	6,1301	6,9285
SEGOVIA	2,1529	1,6682	1,6910	2,2057	2,9664	3,8292
TERUEL	2,1224	1,7904	1,9583	2,5285	3,2985	4,1588
BADAJOS	2,1616	1,9627	2,2272	2,8281	3,6006	4,4564
GUIPUZCOA	2,2202	2,0571	2,3373	2,9365	3,7030	4,5530
VIZCAYA	2,6486	2,3291	2,4007	2,8339	3,4968	4,2840
ÁVILA	2,6295	2,3753	2,5096	2,9805	3,6603	4,4542
CUENCA	2,5566	2,4194	2,6644	3,2053	3,9214	4,7339
LA_CORUNA	2,7918	2,4251	2,4284	2,8003	3,4227	4,1854
ALBACETE	2,8614	2,6289	2,7502	3,1851	3,8282	4,5928
CÁCERES	2,9754	2,6656	2,6994	3,0655	3,6655	4,4050
PONTEVEDRA	3,2045	3,0917	3,2829	3,7317	4,3591	5,0997
GUADALAJARA	3,4004	3,2044	3,3020	3,6697	4,2378	4,9376
TOLEDO	3,4895	3,2878	3,3721	3,7232	4,2757	4,9628
SORIA	4,2681	4,2594	4,4719	4,8766	5,4308	6,0939
VALENCIA	2,4942	1,7552	1,3672	1,6083	2,2875	3,1317
CÓRDOBA	2,2096	1,6263	1,5285	1,9891	2,7396	3,6033
LA RIOJA	2,4997	1,8561	1,6032	1,9038	2,5705	3,3940
BURGOS	2,6416	2,0762	1,8900	2,1824	2,8075	3,5958
LUGO	3,1071	2,4830	2,1461	2,2304	2,6968	3,3911
MADRID	2,9397	2,4253	2,2474	2,4796	3,0290	3,7590
NAVARRA	2,8525	2,4292	2,3652	2,6863	3,2813	4,0306
VALLADOLID	2,9686	2,5059	2,3822	2,6457	3,2020	3,9288
ZARAGOZA	3,2477	2,6910	2,4215	2,5328	2,9825	3,6475
PALENCIA	3,3212	2,7870	2,5363	2,6509	3,0904	3,7420
LLEIDA	3,0021	2,6411	2,6205	2,9475	3,5267	4,2563
BARCELONA	3,4743	3,0300	2,8669	3,0310	3,4761	4,1120
ÁLAVA	3,7266	3,2648	3,0593	3,1605	3,5422	4,1272
SALAMANCA	3,9835	3,6756	3,6167	3,8184	4,2437	4,8338
HUELVA	2,9786	2,2227	1,7140	1,6924	2,1724	2,9160
JAÉN	3,4353	2,6793	2,1179	1,9289	2,2102	2,8245
ASTURIAS	3,3156	2,5916	2,0901	1,9874	2,3365	2,9829
LEÓN	3,5093	2,8116	2,3287	2,2068	2,4992	3,0904
GRANADA	3,3715	2,7000	2,2675	2,2186	2,5751	3,2046
HUESCA	4,0246	3,2431	2,6013	2,2236	2,2475	2,6621
CASTELLÓN	3,9520	3,2534	2,7348	2,5105	2,6561	3,1202
SEVILLA	3,5878	2,9654	2,5776	2,5343	2,8514	3,4303
MURCIA	3,8648	3,2185	2,7766	2,6436	2,8630	3,3666
ZAMORA	3,8218	3,1859	2,7609	2,6502	2,8902	3,4077
CANTABRIA	3,7128	3,1231	2,7662	2,7348	3,0389	3,5943
SC_TENERIFE	3,7512	3,1948	2,8759	2,8747	3,1915	3,7465
CÁDIZ	4,5234	3,8602	3,3570	3,0931	3,1294	3,4567
ORENSE	4,4247	3,8037	3,3596	3,1677	3,2727	3,6491
LAS PALMAS	5,5389	5,0259	4,6658	4,4953	4,5359	4,7823
TARRAGONA	4,2941	3,4171	2,6159	1,9845	1,7211	1,9786
GIRONA	4,6033	3,7702	3,0276	2,4591	2,2039	2,3655
ALICANTE	5,0525	4,2553	3,5520	3,0091	2,7242	2,7778
ALMERÍA	5,2561	4,5136	3,8791	3,4133	3,1911	3,2626
BALEARES	5,1846	4,5315	4,0146	3,6918	3,6152	3,7999
MÁLAGA	6,4608	5,7608	5,1537	4,6758	4,3698	4,2726

Figura II.2 Distancias a los centroides iniciales

el centroide vendrá dado por:

$$C_{t,5}^1 = \frac{\sum_{j=1}^6 \ln Y_{t,j(5)}}{6} = \frac{1}{6} (\ln Y_{t,1(5)} + \ln Y_{t,2(5)} + \dots + \ln Y_{t,6(5)}) =$$

$$= \frac{1}{6} (\ln Y_{t,14} + \ln Y_{t,34} + \ln Y_{t,33} + \ln Y_{t,44} + \ln Y_{t,1} + \ln Y_{t,13})$$

Con este mismo criterio se calculan los cinco centroides restantes:

$$C_{t,k}^1 \quad k=1, 2, 3, 4 \text{ y } 6$$

y se procede a una segunda agrupación en términos de las distancias:

$$d_{k,j}^2 = d(\ln Y_{t,j}, C_{t,k}^1) \quad k=1, \dots, K \quad j=1, \dots, J$$

calculándose la nueva solución de centroides:

$$C_{t,k}^2 = \frac{\sum_{j^{(k)}=1}^{n(k)} \ln Y_{t,j^{(k)}}}{n(k)} \quad k=1, \dots, K$$

Este proceso de agrupación y actualización de los centroides finaliza cuando dos soluciones consecutivas de centroides sean iguales o muy parecidas. Concretamente, en nuestro ejemplo, el proceso converge en la quinta iteración (**Fig. II.3**), en la que los centros de los grupos no presentan ningún cambio respecto de la solución en la iteración anterior¹. En consecuencia, los centroides finales serán los correspondientes a la última iteración:

$$C_{t,k}^F = C_{t,k}^5 \quad k=1, \dots, K \quad t=1, \dots, T$$

Concretamente (**Fig. II.4**):

$$\text{Grupo 1: } (C_{1,1}^F, \dots, C_{102,1}^F) = (\ln H_{1,1}^F, \dots, \ln H_{102,1}^F) = (-0.30, \dots, 1.08)$$

.....

$$\text{Grupo 6: } (C_{1,6}^F, \dots, C_{102,6}^F) = (\ln H_{1,6}^F, \dots, \ln H_{102,6}^F) = (0.17, \dots, -1.14)$$

¹ El criterio de convergencia consiste en una proporción de la distancia mínima entre los centros iniciales de los grupos, por lo que deben estar comprendidos entre 0 y 1. Por ejemplo, si el criterio es igual a 0,01, el proceso iterativo finalizará cuando de una solución a la siguiente ninguno de los centros de los grupos se mueva en una distancia superior al uno por ciento de la distancia menor entre cualquiera de los centros iniciales.

Historial de iteraciones(a)						
Iteración	Cambio en los centroides					
	1	2	3	4	5	6
1	3,894	1,027	,664	,729	1,662	4,273
2	,000	,450	,429	,562	,597	,000
3	,000	,340	,266	,544	,297	,000
4	,000	,000	,123	,192	,000	,000
5	,000	,000	,000	,000	,000	,000

a Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de ,000. La iteración actual es 5. La distancia mínima entre los centros iniciales es de ,982.

Figura II.3 Cambio en los centroides en cada iteración

Centroides finales						
	Grupo					
	1	2	3	4	5	6
3T 1983	-,30	-,03	-,03	,09	,00	,17
4T 1983	-,39	-,02	-,07	,05	-,02	,24
1T 1984	-,49	-,02	-,12	-,01	-,03	,41
2T 1984	-,29	-,03	-,16	-,07	-,10	,35
3T 1984	-,30	-,07	-,14	-,06	-,13	,45
4T 1984	-,40	-,16	-,17	-,05	-,16	,46
1T 1985	-,29	-,19	-,19	-,04	-,13	,36
2T 1985	-,51	-,18	-,24	,01	-,09	,42
3T 1985	-,53	-,15	-,24	,00	-,04	,41
4T 1985	-,49	-,06	-,15	,13	-,01	,49
1T 1986	-,64	-,12	-,18	,07	-,05	,36
....
2T 2006	1,71	1,46	1,12	1,25	1,26	1,31
3T 2006	1,84	1,56	1,32	1,45	1,38	1,40
4T 2006	1,86	1,49	1,26	1,40	1,32	1,33
1T 2007	1,89	1,52	1,27	1,44	1,32	1,27
2T 2007	1,84	1,46	1,16	1,40	1,22	1,09
3T 2007	1,75	1,32	,81	1,11	,95	,83
4T 2007	1,72	1,32	,78	1,06	,82	,79
1T 2008	1,60	1,17	,56	,83	,54	,56
2T 2008	1,56	1,02	,46	,71	,35	,41
3T 2008	1,33	,83	,22	,53	,15	,14
4T 2008	1,08	,61	,05	,29	-,20	-,14

Figura II.4 Valores de $C_{t,k}^F$

A lo largo del proceso de formación de los grupos ha habido cinco soluciones distintas asociadas a las cinco iteraciones del proceso. La **Fig. II.5** proporciona las soluciones final e inicial para cada una de las 50 provincias. Obsérvese que, de las 50 provincias, 30 han sido clasificadas en el mismo grupo según las dos soluciones y que, de las 20 clasificadas en distintos grupos, todas ellas lo están en grupos consecutivos.

Provincia	Grupo Inicial	Grupo Final
CIUDAD_REAL	1	1
TERUEL	2	2
ALBACETE	2	2
CUENCA	2	2
GUADALAJARA	2	2
TOLEDO	2	2
ÁVILA	2	2
SORIA	2	2
PONTEVEDRA	2	2
SEGOVIA	2	3
CÁCERES	2	3
BADAJOS	2	3
LA_CORUÑA	2	3
GUIPÚZCOA	2	3
VIZCAYA	2	3
BURGOS	3	2
LLEIDA	3	2
CÓRDOBA	3	3
VALLADOLID	3	3
PALENCIA	3	3
SALAMANCA	3	3
BARCELONA	3	3
MADRID	3	3
VALENCIA	3	3
NAVARRA	3	3
ZARAGOZA	3	4
LUGO	3	4
LA RIOJA	3	4
ÁLAVA	3	4
GRANADA	4	3
JAÉN	4	3
CÁDIZ	4	3
HUELVA	4	3
SEVILLA	4	3
ASTURIAS	4	4
CANTABRIA	4	4
LEÓN	4	4
ZAMORA	4	4
ORENSE	4	4
MURCIA	4	4
CASTELLÓN	4	4
HUESCA	4	5
LAS PALMAS	4	5
SC_TENERIFE	4	5
ALMERÍA	5	5
BALEARES	5	5
GIRONA	5	5
TARRAGONA	5	5
ALICANTE	5	5
MÁLAGA	6	6

Figura II.5 Soluciones inicial y final

Concretamente:

SEGOVIA, CÁCERES, BADAJOZ, LA CORUÑA, GUIPÚZCOA y VIZCAYA: Clasificadas en el tercer grupo en la solución final mientras que en la inicial estaban en el segundo.

BURGOS y LLEIDA: Clasificadas en el segundo grupo en la solución final mientras que en la inicial estaban en el tercero.

ZARAGOZA, LUGO, LA RIOJA y ÁLAVA: Clasificadas en el cuarto grupo en la solución final mientras que en la inicial estaban en el tercero.

GRANADA, JAÉN, CÁDIZ, HUELVA y SEVILLA: Clasificadas en el tercer grupo en la solución final mientras que en la inicial estaban en el cuarto.

HUESCA, LAS PALMAS y TENERIFE: Clasificadas en el quinto grupo en la solución final mientras que en la inicial estaban en el cuarto.

En definitiva, según la distancia de las distintas provincias a cada uno de los seis centroides finales (**Fig. II.6**) la solución obtenida proporciona la siguiente partición:

Grupo	Provincias
1	CIUDAD REAL
2	ÁVILA, CUENCA, ALBACETE, TERUEL, BURGOS, LLEIDA, PONTEVEDRA, TOLEDO, GUADALAJARA, SORIA
3	CÓRDOBA, VALENCIA, HUELVA, MADRID, GUIPÚZCOA, NAVARRA, SEGOVIA, BADAJOZ, VALLADOLID, GRANADA, JAÉN, SEVILLA, PALENCIA, BARCELONA, LA CORUÑA, VIZCAYA, CÁCERES, CÁDIZ, SALAMANCA
4	LA RIOJA, LUGO, CASTELLÓN, LEÓN, ZARAGOZA, ZAMORA, MURCIA, ASTURIAS, ORENSE, CANTABRIA, ALAVA
5	ALICANTE, HUESCA, TARRAGONA, GIRONA, SC TENERIFE, BALEARES, LAS PALMAS, ALMERÍA
6	MÁLAGA

	Distancias					
	1	2	3	4	5	6
CIUDAD_REAL	,0000	3,1829	4,7409	5,6520	7,0342	8,9911
ÁVILA	4,1280	1,7356	2,7044	3,0327	4,8151	6,5736
CUENCA	3,5519	1,7693	2,9496	3,2139	4,8867	6,9638
ALBACETE	3,5891	1,8263	2,7712	3,4396	5,0818	7,0211
TERUEL	4,3812	1,8790	2,5004	2,2491	4,0167	6,0194
BURGOS	4,7523	2,0774	2,1131	2,5207	3,7790	5,6568
LLEIDA	4,1413	2,4447	2,9738	3,1462	4,2106	6,3070
PONTEVEDRA	3,9155	2,5630	3,1530	4,1583	5,3806	7,0847
TOLEDO	4,3725	2,6481	3,7378	3,4777	5,3639	7,1751
GUADALAJARA	4,3541	2,8372	4,0039	3,4917	4,5302	6,0954
SORIA	5,1362	3,8409	4,4693	5,3308	6,0122	7,5045
CÓRDOBA	5,0059	2,4159	1,3887	2,0500	3,7485	6,0286
VALENCIA	5,1477	2,5786	1,4842	1,7298	3,2078	5,4694
HUELVA	5,1617	2,9143	1,6893	2,4272	2,8488	5,2949
MADRID	5,5792	3,3424	1,7595	3,0025	3,6938	5,9369
GUIPÚZCOA	4,8885	2,8174	1,8260	3,0909	4,6014	7,0397
NAVARRA	5,2520	2,9805	1,8771	2,9222	4,4452	6,6798
SEGOVIA	4,4937	2,1939	1,9082	2,2680	3,7282	5,9684
BADAJOS	3,6655	2,0785	1,9118	3,0635	4,7103	6,9925
VALLADOLID	4,9655	2,8948	1,9918	3,2982	4,0836	6,3335
GRANADA	5,4586	3,0583	2,0924	2,4664	3,9600	5,8888
JAÉN	5,8728	3,3619	2,1066	2,1090	3,2983	5,4642
SEVILLA	5,1270	3,1849	2,1317	3,3282	4,0752	6,2309
PALENCIA	5,3510	3,0951	2,1600	3,0285	4,3776	6,6243
BARCELONA	4,8857	3,3665	2,2286	3,9668	4,3279	6,7916
LA_CORUÑA	4,3715	2,6380	2,2383	3,4044	4,1606	6,3369
VIZCAYA	5,1702	3,1266	2,3579	3,0751	3,8675	5,9557
CÁCERES	5,5037	3,3404	2,8205	3,1093	4,0284	5,9642
CÁDIZ	6,6475	4,5645	3,0320	3,9401	3,4969	5,3547
SALAMANCA	5,8093	4,3253	3,2066	4,2353	4,8418	7,5286
LA RIOJA	4,9785	2,5493	2,3373	1,7456	2,7964	4,9294
LUGO	5,3353	2,9160	2,4665	1,9174	3,7126	5,8948
CASTELLÓN	6,4410	3,8764	3,3479	1,9293	3,0103	4,6699
LEÓN	6,6665	3,9402	2,7671	1,9405	2,2841	4,4417
ZARAGOZA	5,9581	3,3703	2,9857	1,9701	3,3575	5,1244
ZAMORA	6,5897	3,8932	3,0517	2,1201	3,6664	5,4858
MURCIA	6,2090	3,6769	3,5746	2,1416	2,8907	4,6604
ASTURIAS	5,4283	2,7779	2,5569	2,2121	3,1695	4,7962
ORENSE	7,4096	4,7957	3,6294	2,6438	3,4992	5,2360
CANTABRIA	5,3485	3,1587	3,4446	2,8217	3,2892	4,9041
ÁLAVA	6,0967	3,7302	3,5829	2,8957	3,7250	5,4300
ALICANTE	7,8724	5,2570	4,0012	3,1045	1,4126	3,2894
HUESCA	6,8074	4,1997	3,1282	2,2257	1,6267	3,8858
TARRAGONA	6,6712	4,1331	3,0517	2,3519	1,6601	4,1433
GIRONA	7,2439	4,6101	3,4084	2,6154	1,9610	3,8429
SC_TENERIFE	6,3416	4,1518	3,4169	2,9786	2,2068	4,4703
BALEARES	8,0975	5,5501	4,4122	3,4442	2,9710	4,5830
LAS_PALMAS	7,9686	6,0014	5,0511	4,6837	3,1432	4,9128
ALMERÍA	6,6820	4,6568	4,4094	3,7402	3,2252	4,1873
MÁLAGA	8,9911	6,4327	5,8448	4,5448	3,0869	,0000

Figura II.6 Distancias a los centroides finales

ANEXO III

GRÁFICOS PROVINCIALES

Los seis gráficos relativos a la interpretación de cada provincia incluyen la siguiente información:

Panel superior izquierdo: Serie provincial (resaltada en negrita) junto con las restantes provincias clasificadas en su mismo grupo, así como el centroide correspondiente (resaltado en color). Los valores de las series se proporcionan en dos escalas: en la de la serie objeto de interpretación (eje izquierdo) y en valores homogeneizados (eje derecho).

Panel central izquierdo: Serie provincial (resaltada en negrita) junto con los centroides iniciales. Los valores de las series se proporcionan en dos escalas: en la de la serie objeto de interpretación (eje izquierdo) y en valores homogeneizados (eje derecho).

Panel inferior izquierdo: Serie provincial (resaltada en negrita) junto con las series resumen calculadas en términos del tipo de interés. Los valores de las series se proporcionan en dos escalas: en la de la serie objeto de interpretación (eje izquierdo) y en valores homogeneizados (eje derecho).

Panel superior derecho: Serie provincial (resaltada en rojo) junto con la serie de Córdoba. Los valores de las series se proporcionan en dos escalas: en la de la serie objeto de interpretación (eje derecho) y en la de Córdoba (eje izquierdo).

Panel central derecho: Serie provincial (resaltada en rojo) junto con la serie de Alicante. Los valores de las series se proporcionan en dos escalas: en la de la serie objeto de interpretación (eje derecho) y en la de Alicante (eje izquierdo).

Panel inferior derecho: Serie provincial (resaltada en rojo) junto con la serie de España. Los valores de las series se proporcionan en dos escalas: en la de la serie objeto de interpretación (eje derecho) y en la de España (eje izquierdo).

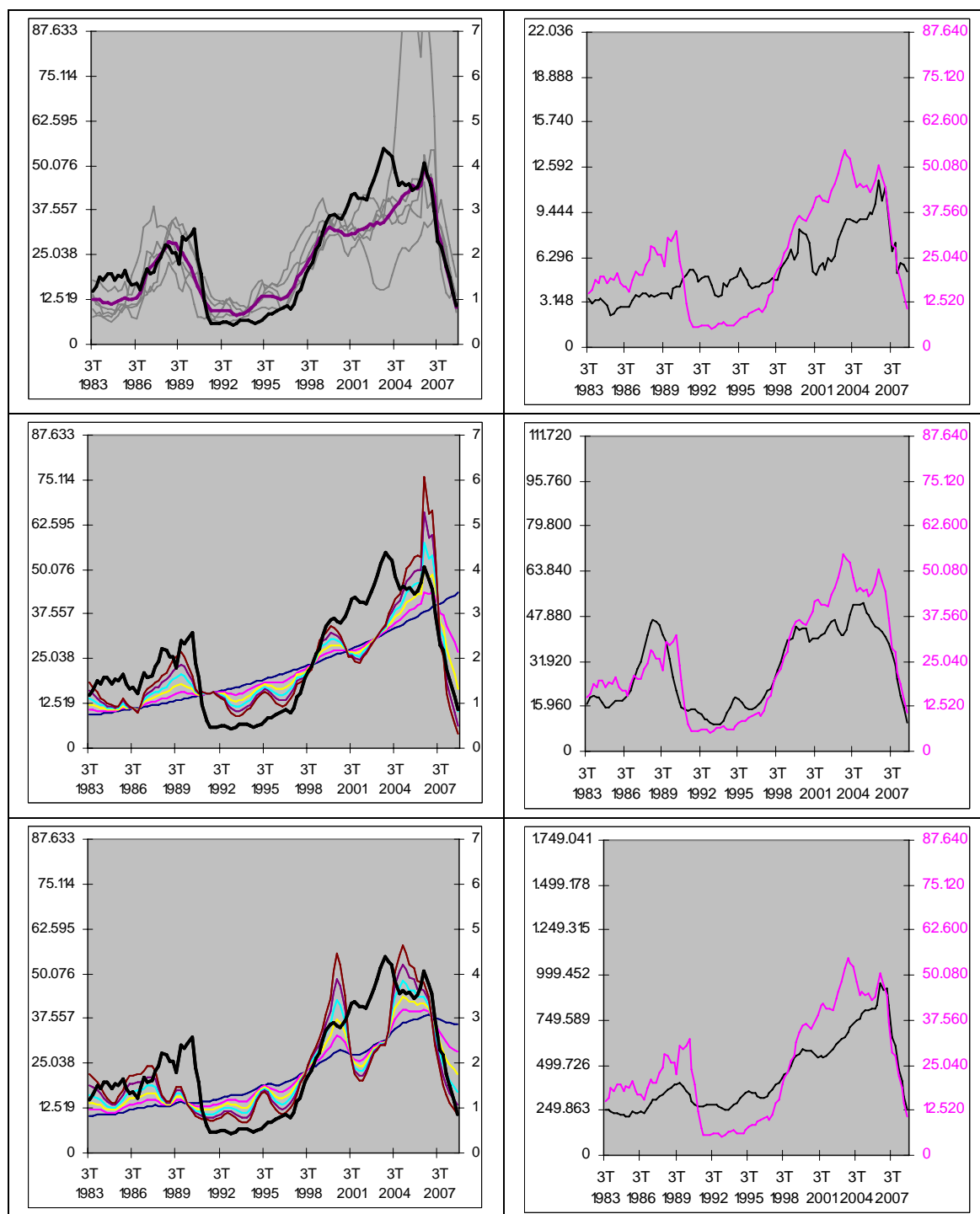


Figura III.1 Málaga

Paneles izquierdos (Málaga resaltada en negrita): *Panel Superior*: Junto con las provincias del quinto grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Málaga: Acumulado anual/valores homogeneizados;

Paneles derechos (Málaga resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Málaga (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

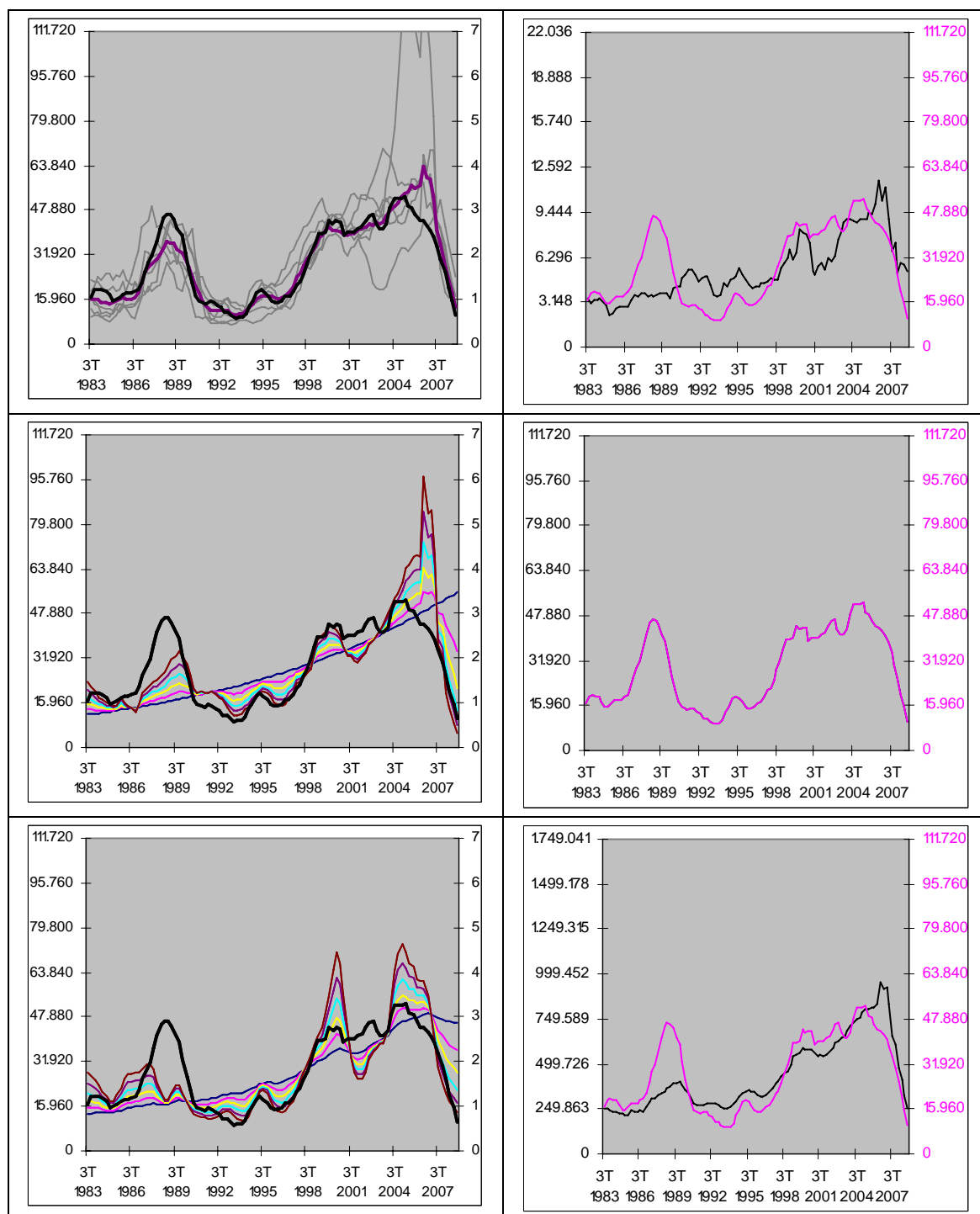


Figura III.2 Alicante/Alacant

Paneles izquierdos (Alicante resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Alicante: Acumulado anual/valores homogeneizados; **Paneles derechos** (Alicante resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Alicante (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

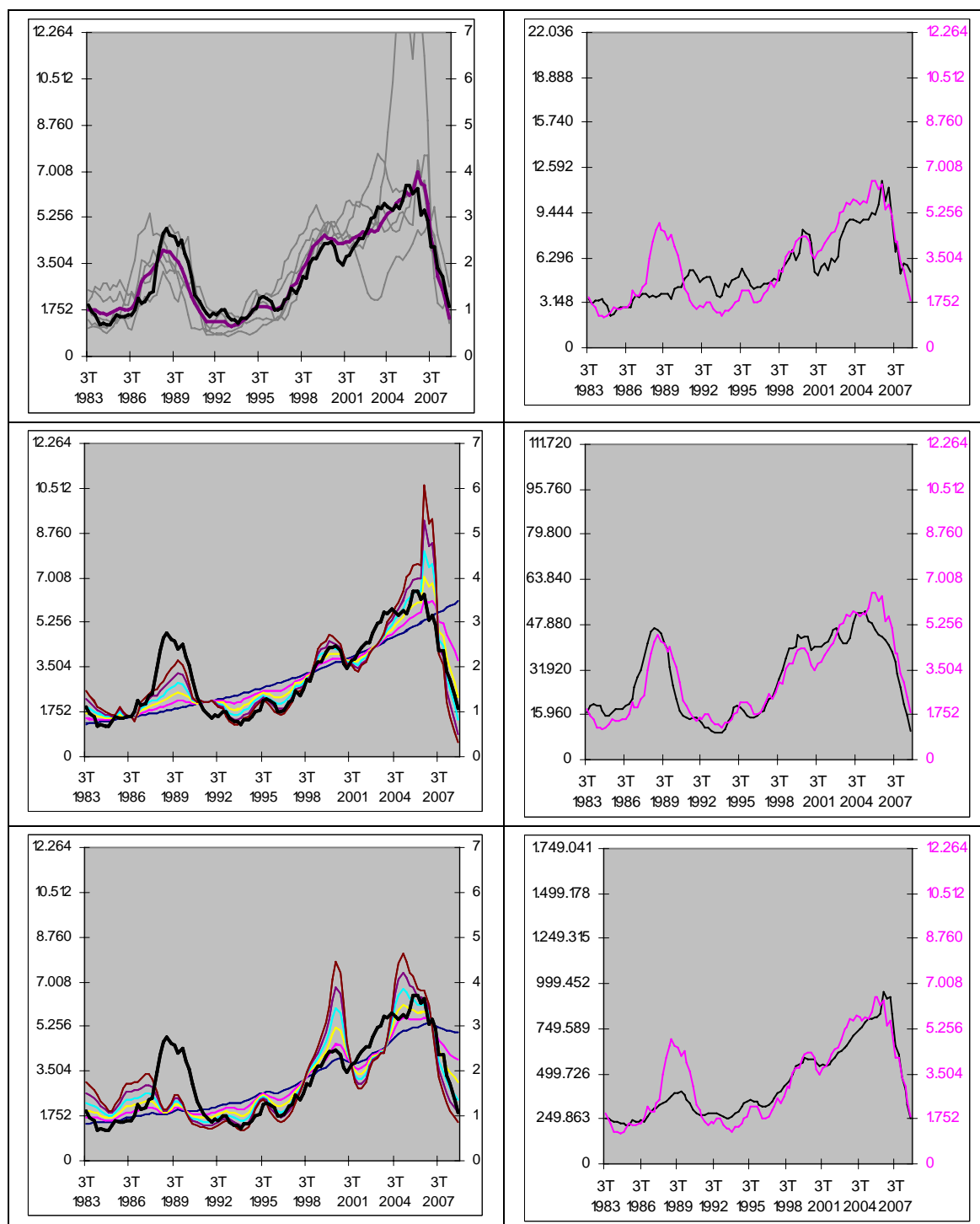


Figura III.3 Huesca

Paneles izquierdos (Huesca resaltada en negrita): *Panel Superior:* Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central:* Junto con los centroides iniciales; *Panel Inferior:* Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho:* Unidades visadas en Huesca: Acumulado anual/valores homogeneizados;

Paneles derechos (Huesca resaltada en rojo): *Paneles Superior/Central/Inferior:* Junto con Córdoba/Alicante/España; *Eje de escala derecho:* Unidades visadas en Huesca (en rojo); *Eje de escala izquierdo:* Unidades visadas en Córdoba/Alicante/España

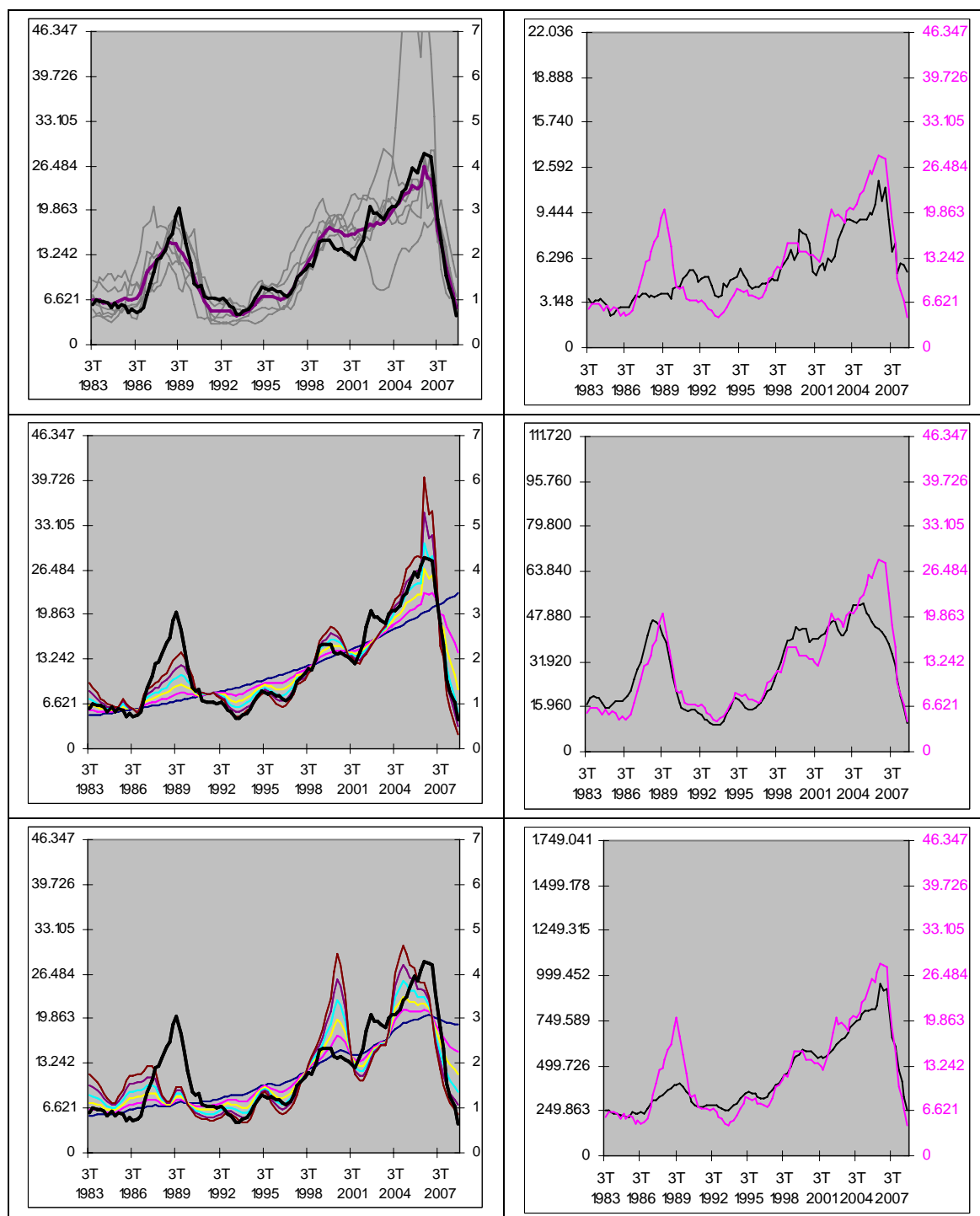


Figura III.4 Tarragona

Paneles izquierdos (Tarragona resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Tarragona: Acumulado anual/valores homogeneizados; **Paneles derechos** (Tarragona resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Tarragona (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

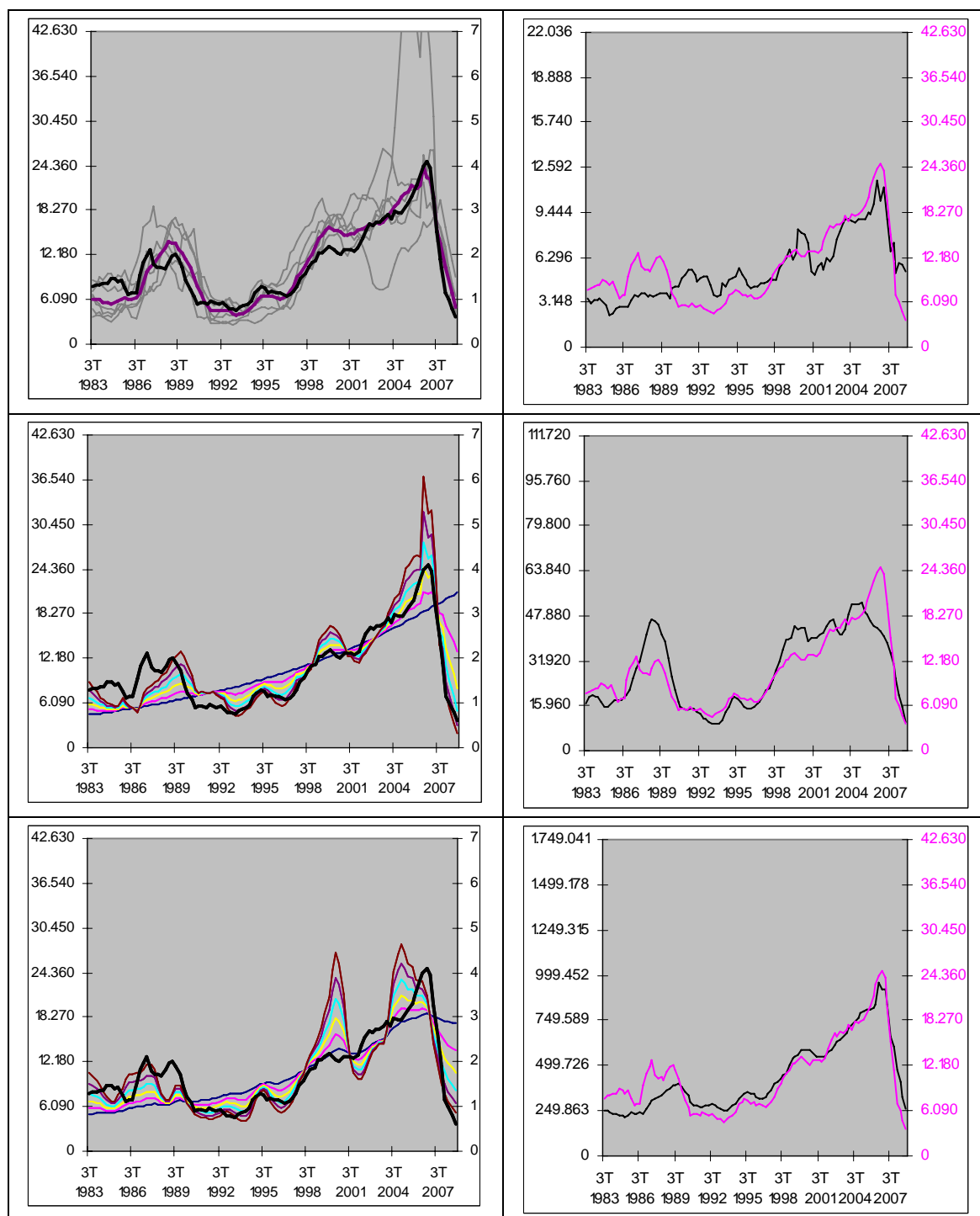


Figura III.5 Girona

Paneles izquierdos (Girona resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierda/derecho*: Unidades visadas en Girona: Acumulado anual/valores homogeneizados; **Paneles derechos** (Girona resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Girona (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

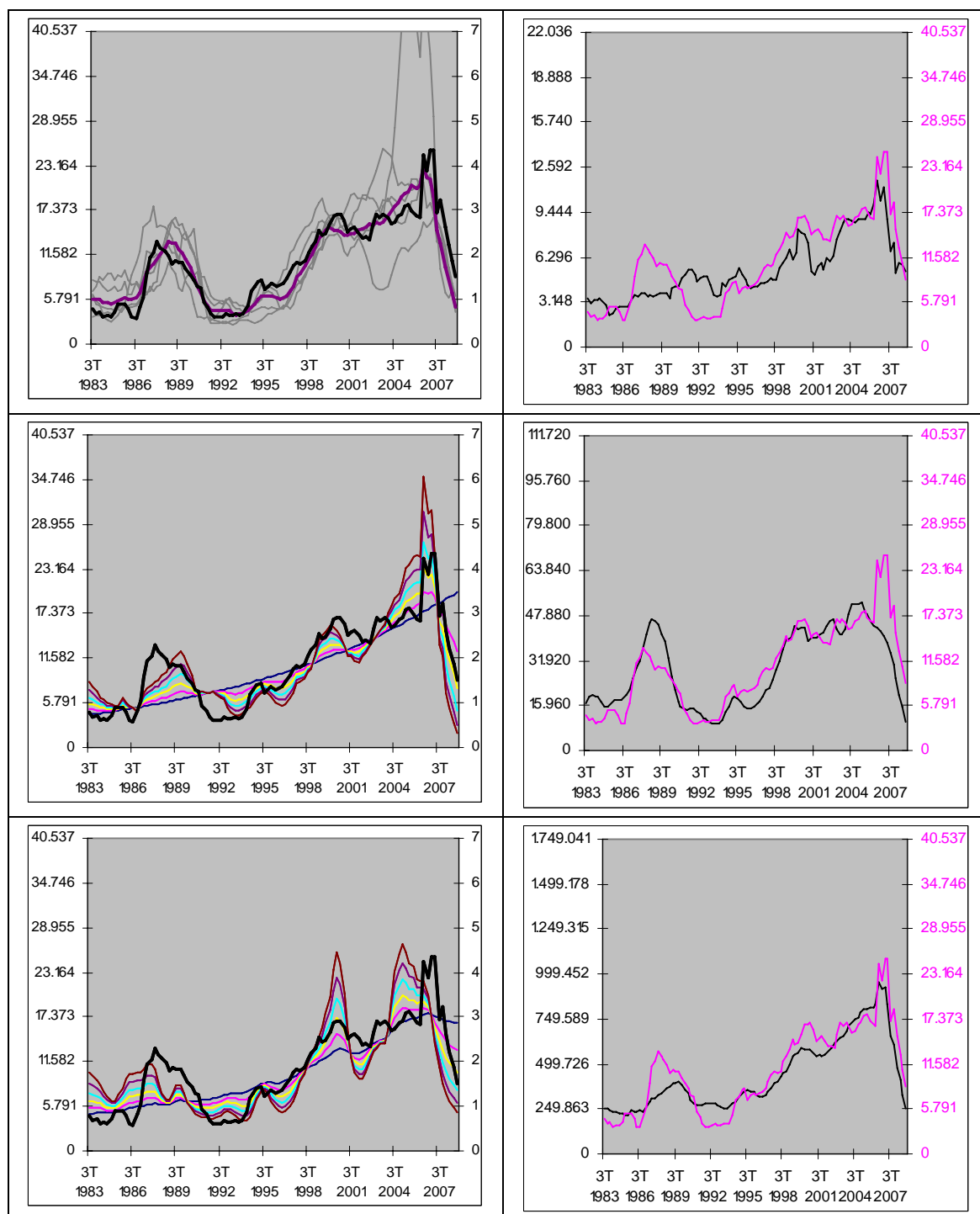


Figura III.6 Sta. Cruz de Tenerife

Paneles izquierdos (Tenerife resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Tenerife: Acumulado anual/valores homogeneizados; **Paneles derechos** (Tenerife resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Tenerife (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

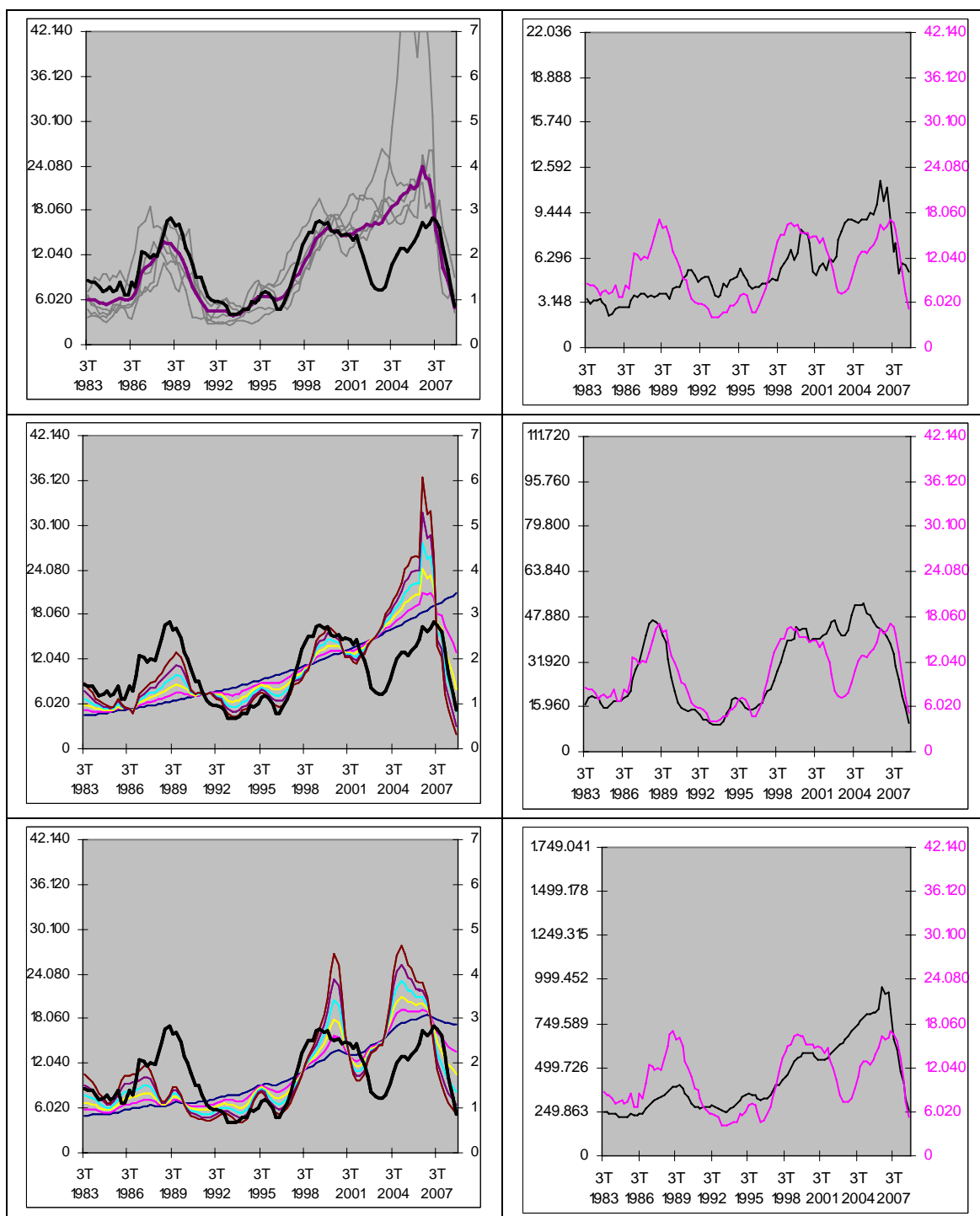


Figura III.7 Baleares/Balears

Paneles izquierdos (Baleares resaltada en negrita): *Panel Superior:* Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central:* Junto con los centroides iniciales; *Panel Inferior:* Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho:* Unidades visadas en Baleares: Acumulado anual/valores homogeneizados; **Paneles derechos** (Baleares resaltada en rojo): *Paneles Superior/Central/Inferior:* Junto con Córdoba/Alicante/España; *Eje de escala derecho:* Unidades visadas en Baleares (en rojo); *Eje de escala izquierdo:* Unidades visadas en Córdoba/Alicante/España

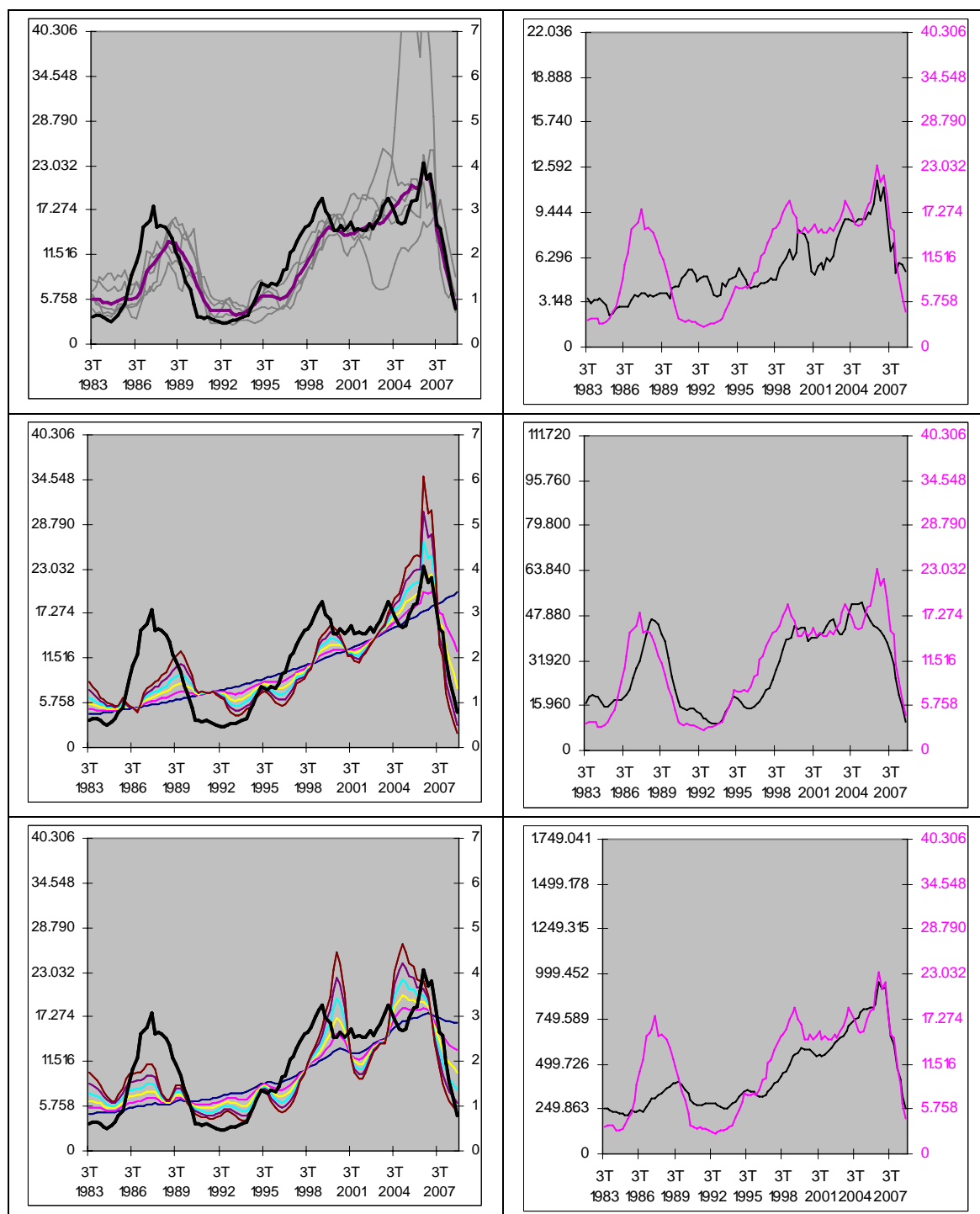


Figura III.8 Las Palmas

Paneles izquierdos (Las Palmas resaltada en negrita): *Panel Superior:* Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central:* Junto con los centroides iniciales; *Panel Inferior:* Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho:* Unidades visadas en Las Palmas: Acumulado anual/valores homogeneizados;

Paneles derechos (Las Palmas resaltada en rojo): *Paneles Superior/Central/Inferior:* Junto con Córdoba/Alicante/España; *Eje de escala derecho:* Unidades visadas en Las Palmas (en rojo); *Eje de escala izquierdo:* Unidades visadas en Córdoba/Alicante/España

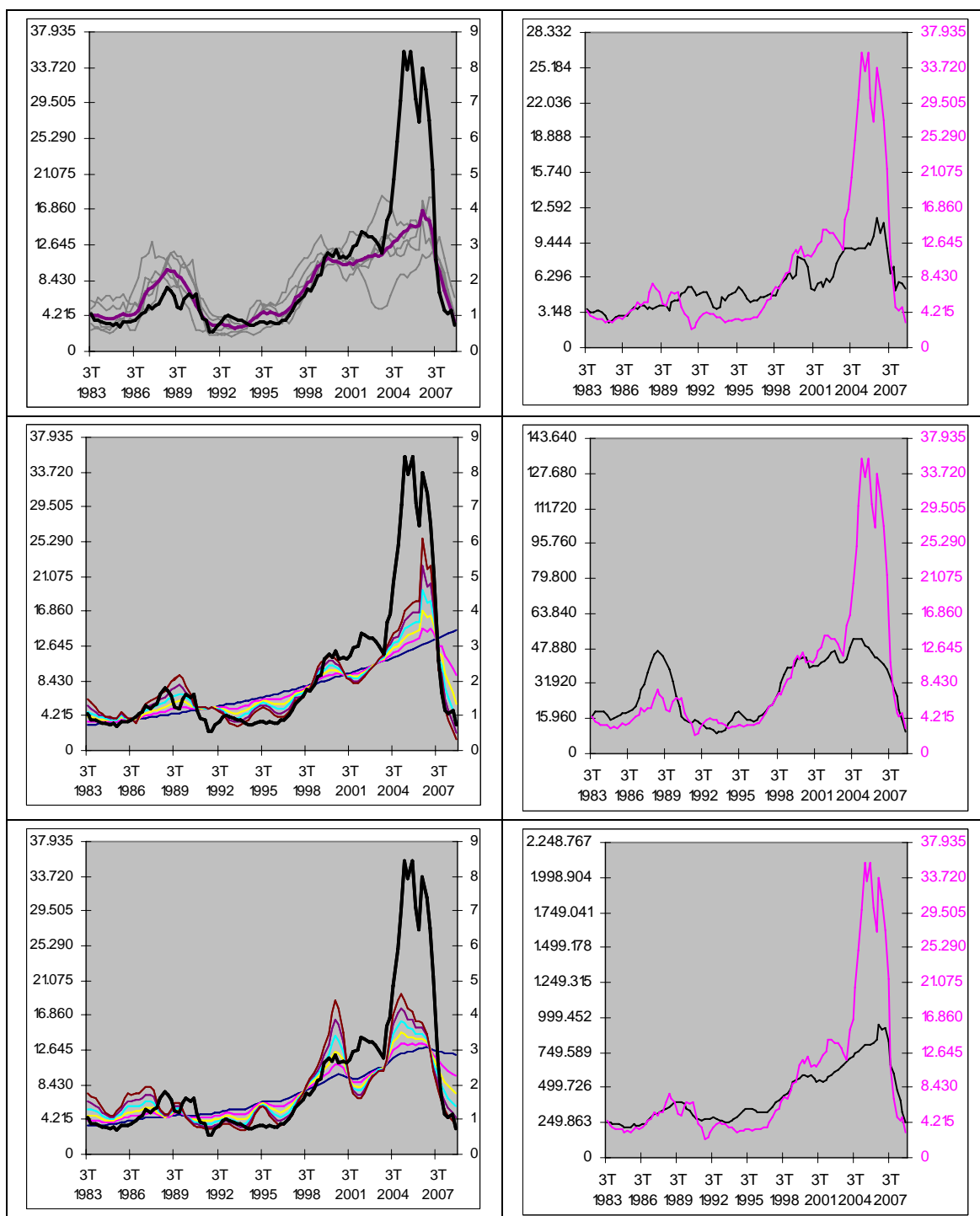


Figura III.9 Almería

Paneles izquierdos (Almería resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Almería: Acumulado anual/valores homogeneizados; **Paneles derechos** (Almería resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Almería (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

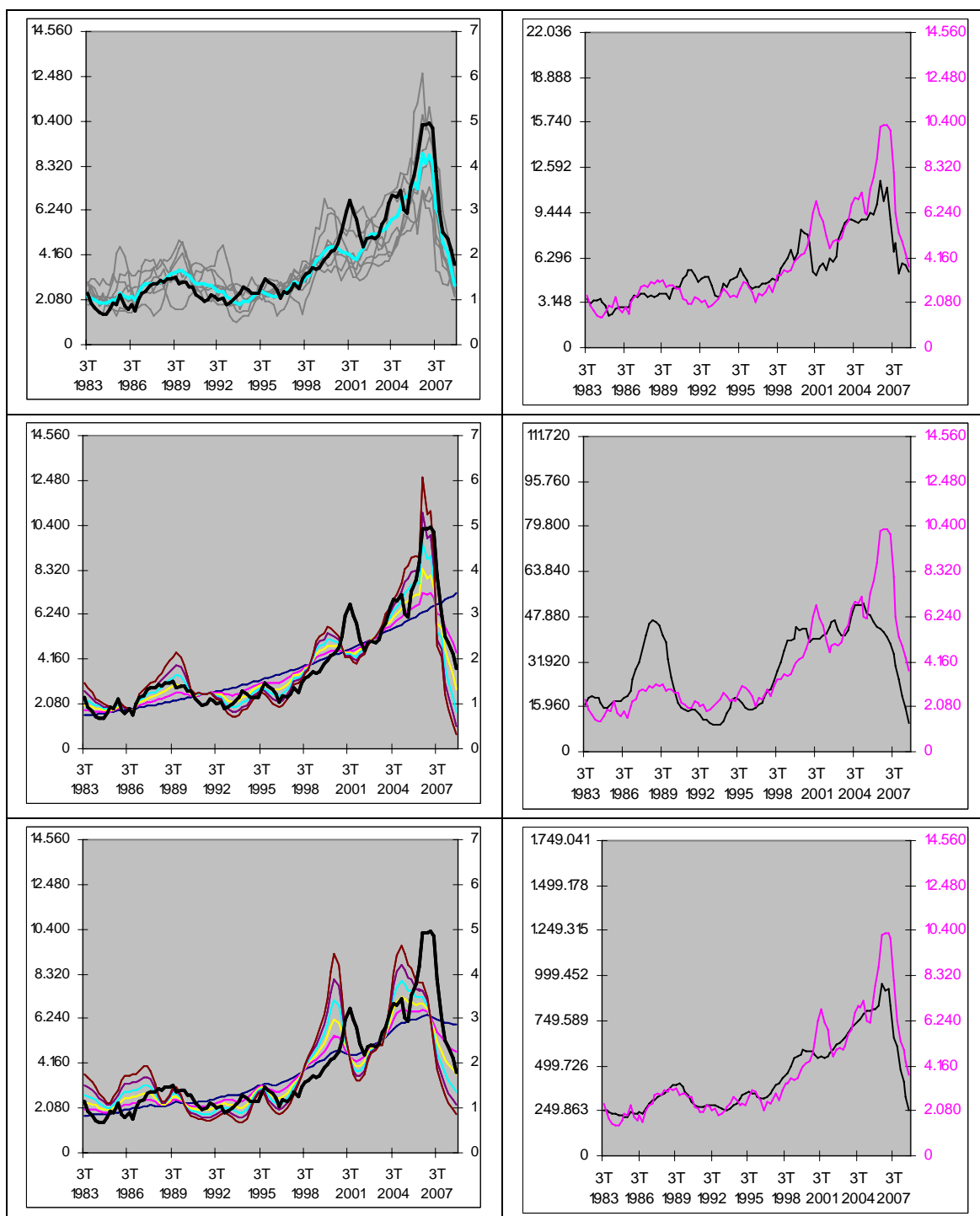


Figura III.10 La rioja

Paneles izquierdos (La Rioja resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en La Rioja: Acumulado anual/valores homogeneizados; **Paneles derechos** (La Rioja resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en La Rioja (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

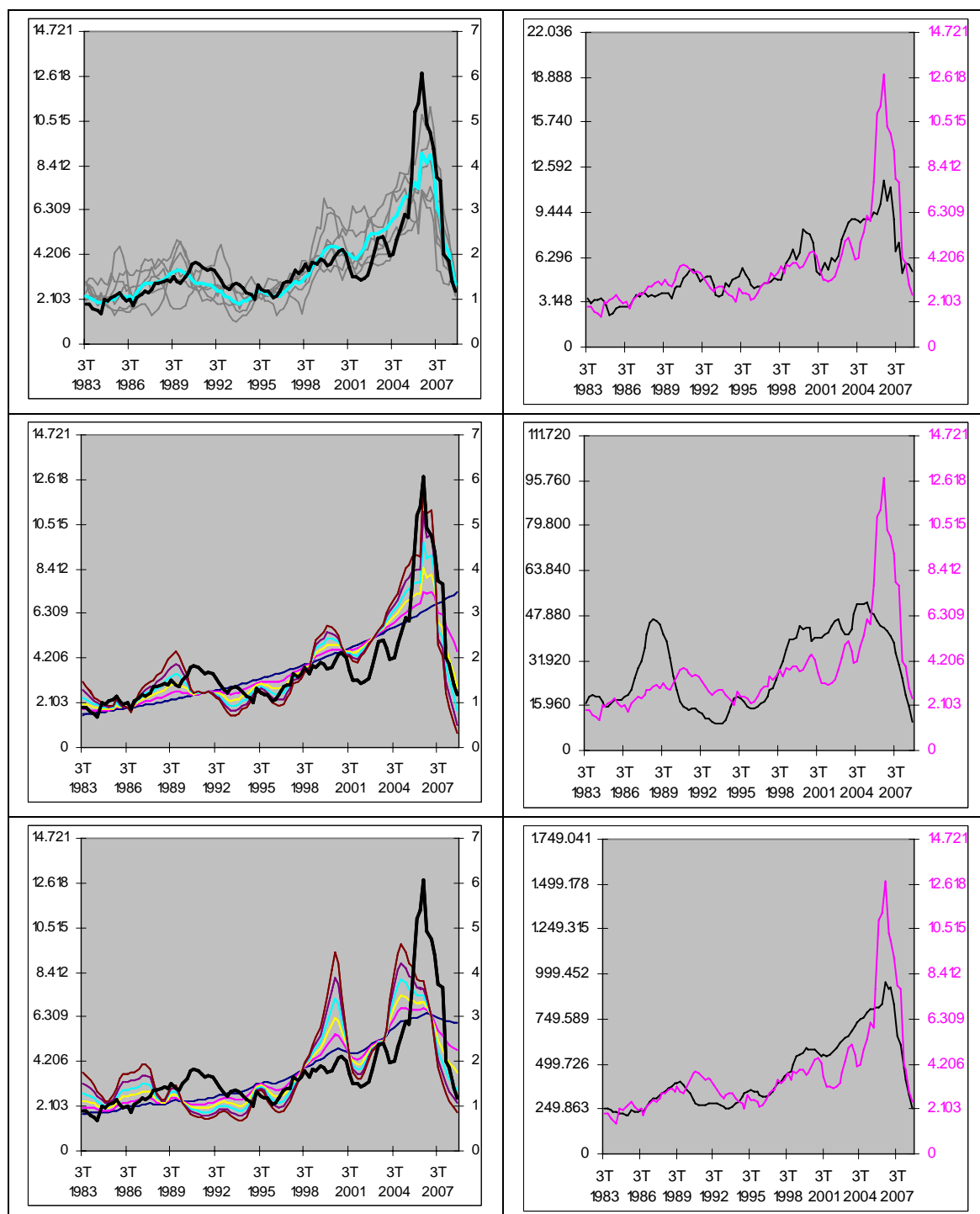


Figura III.11 Lugo

Paneles izquierdos (Lugo resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Lugo: Acumulado anual/valores homogeneizados; **Paneles derechos** (Lugo resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Lugo (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

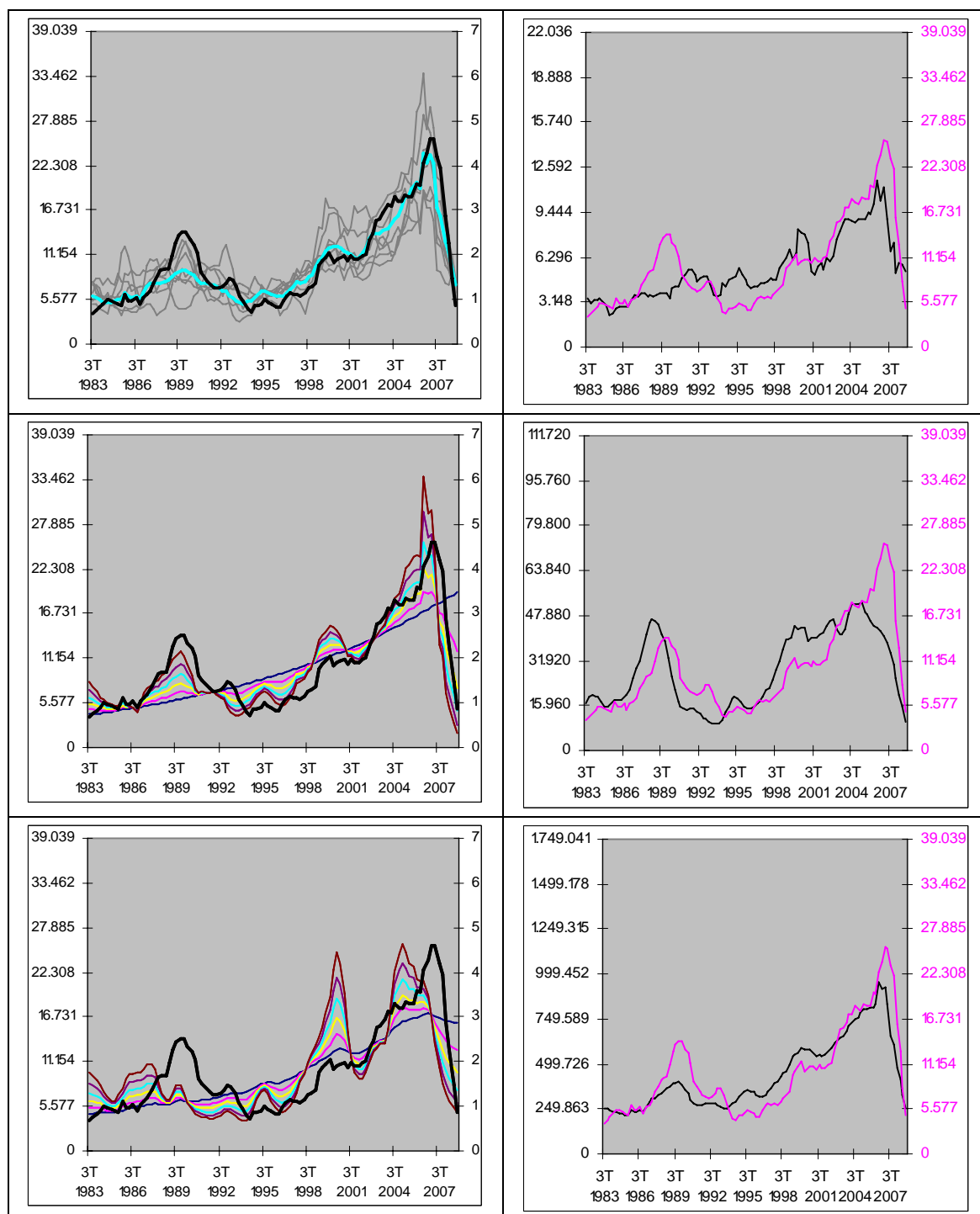


Figura III.12 Castellón/Castelló

Paneles izquierdos (Castellón resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Castellón: Acumulado anual/valores homogeneizados;

Paneles derechos (Castellón resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Castellón (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

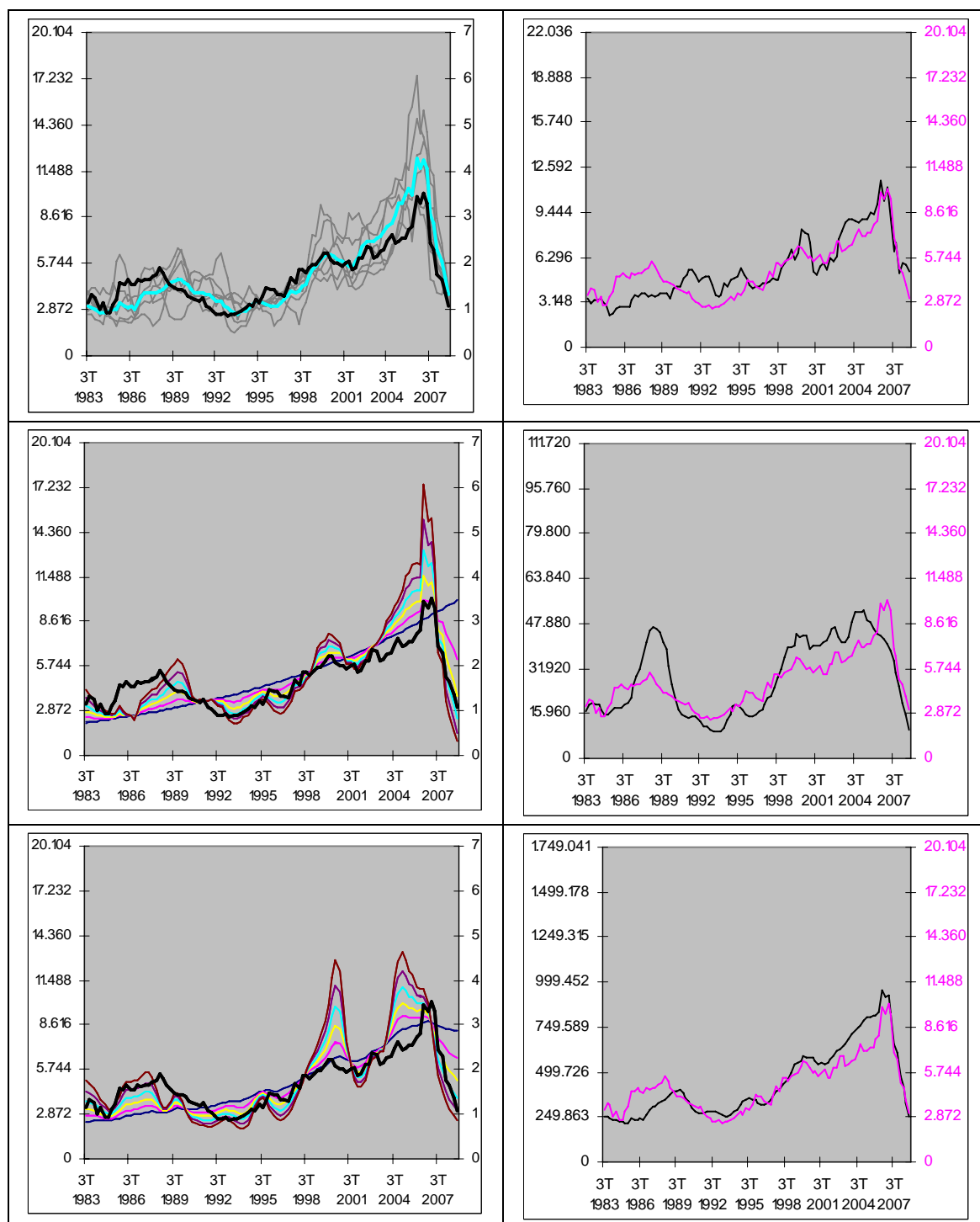


Figura III.13 León

Paneles izquierdos (León resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en León: Acumulado anual/valores homogeneizados;

Paneles derechos (León resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en León (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

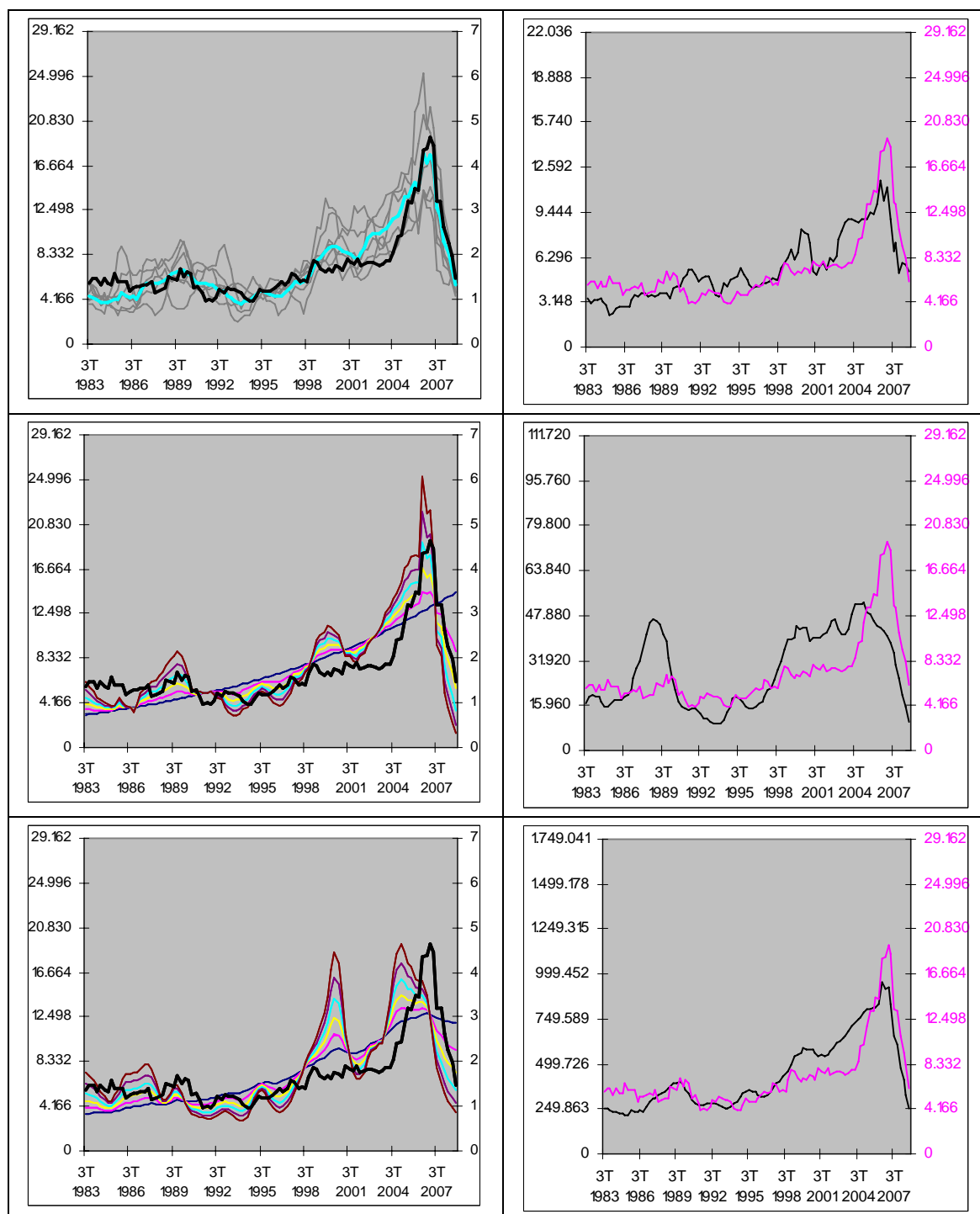


Figura III.14 Zaragoza

Panes izquierdo (Zaragoza resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Zaragoza: Acumulado anual/valores homogeneizados; **Panes derecho** (Zaragoza resaltada en rojo): *Panes Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Zaragoza (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

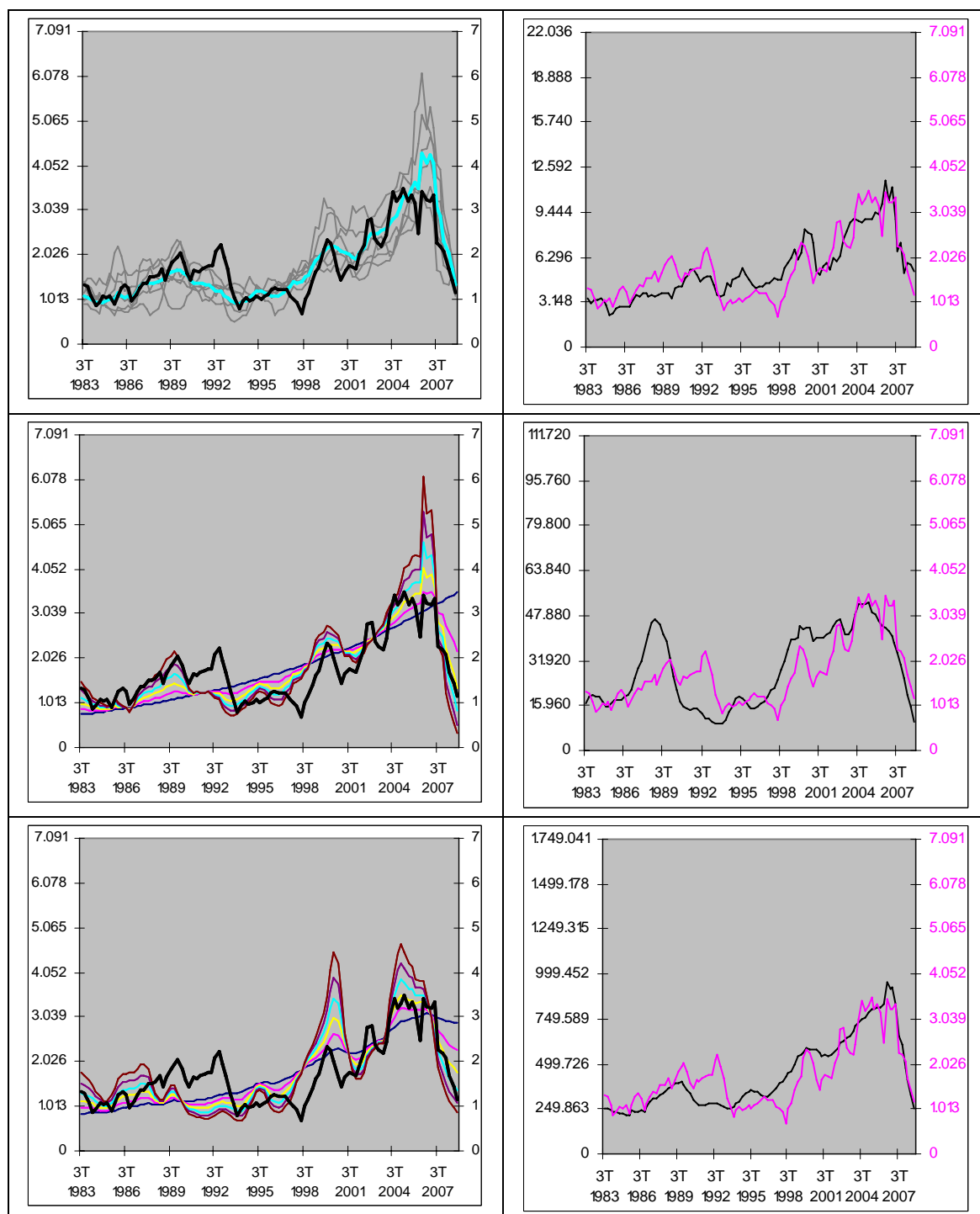


Figura III.15 Zamora

Paneles izquierdos (Zamora resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Zamora: Acumulado anual/valores homogeneizados; **Paneles derechos** (Zamora resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Zamora (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

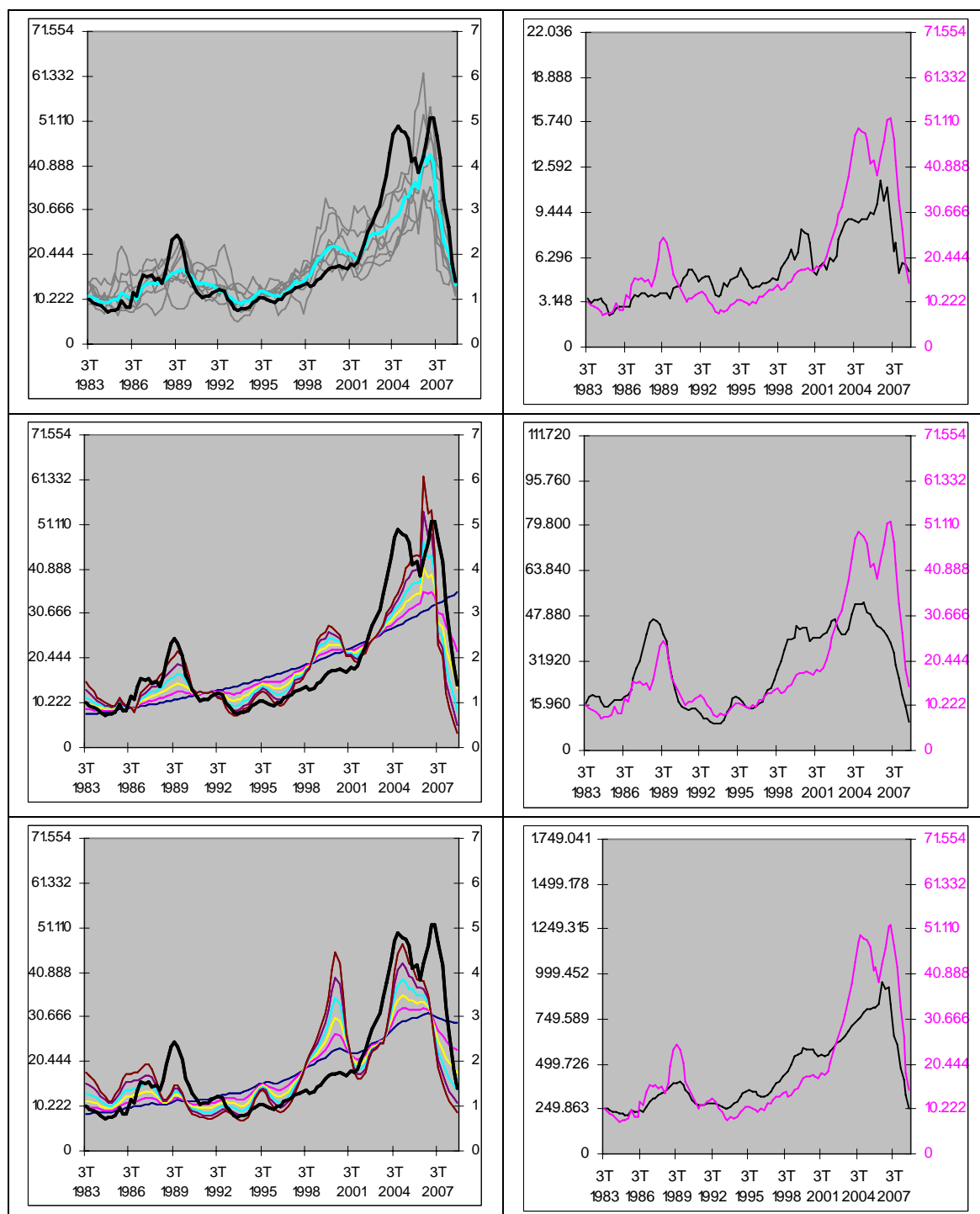


Figura III.16 Murcia

Paneles izquierdos (Murcia resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Murcia: Acumulado anual/valores homogeneizados; **Paneles derechos** (Murcia resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Murcia (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

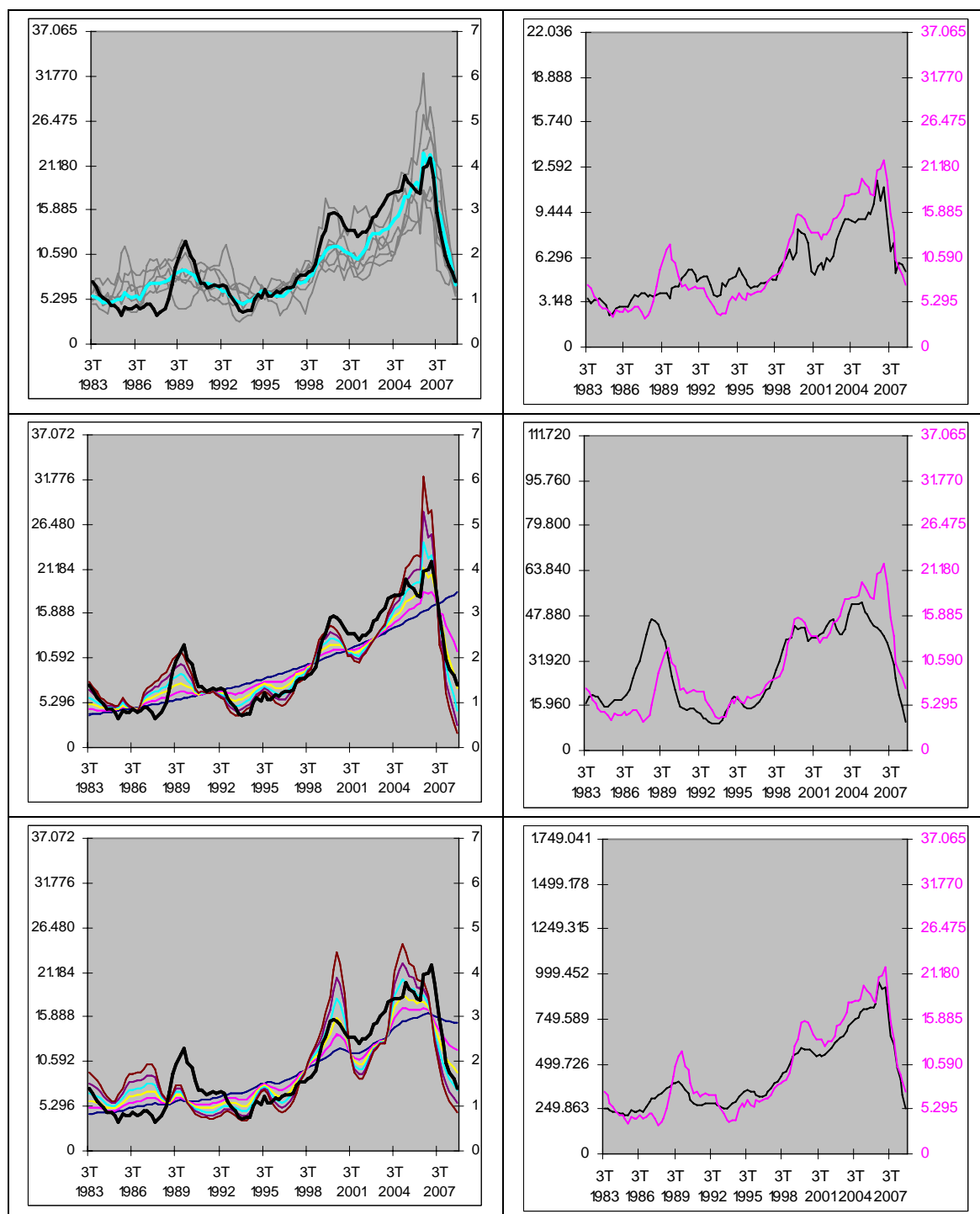


Figura III.17 Asturias

Paneles izquierdos (Asturias resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Asturias: Acumulado anual/valores homogeneizados; **Paneles derechos** (Asturias resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Asturias (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

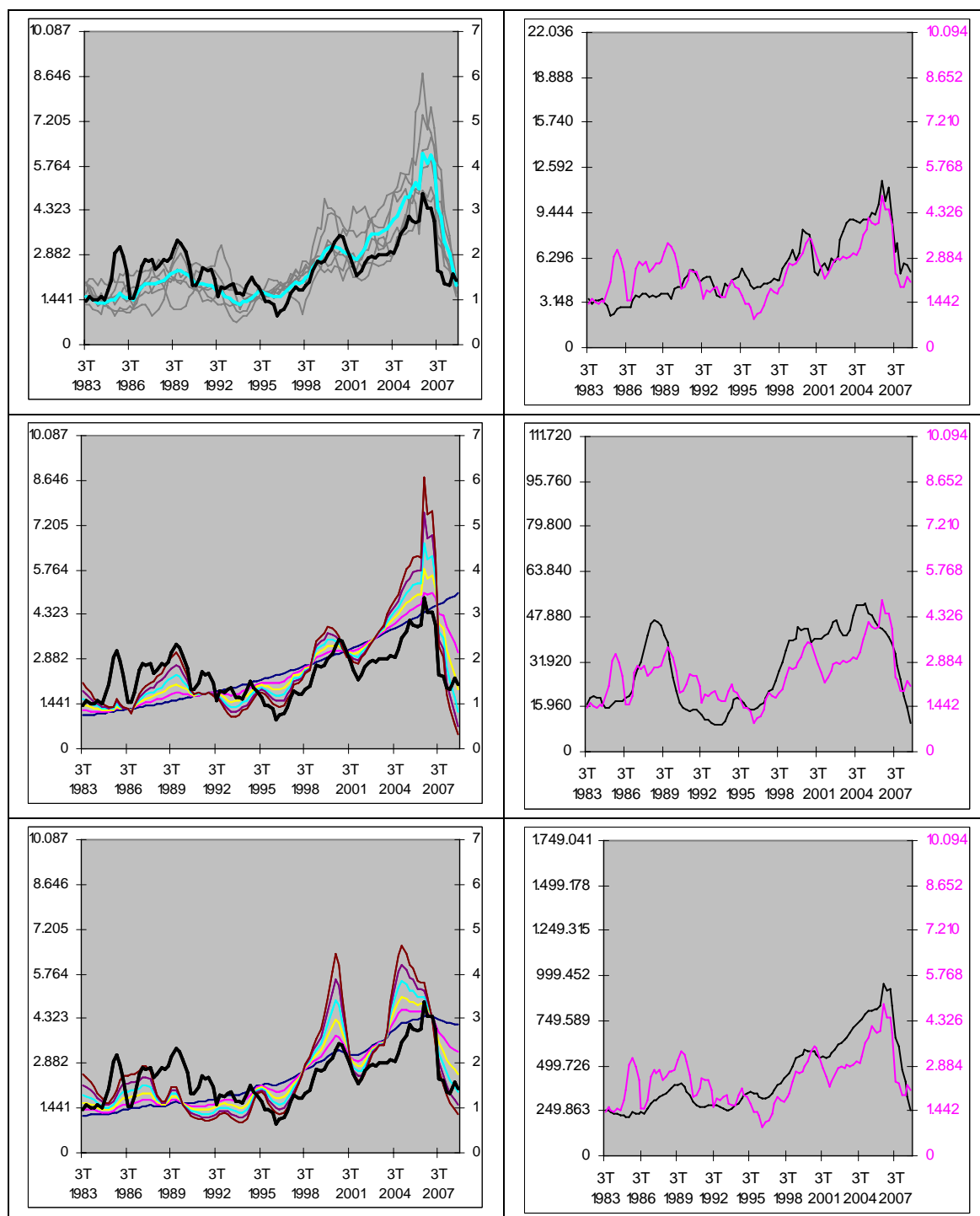


Figura III.18 Ourense/Ourense

Paneles izquierdos (Ourense resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Ourense: Acumulado anual/valores homogeneizados; **Paneles derechos** (Ourense resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Ourense (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

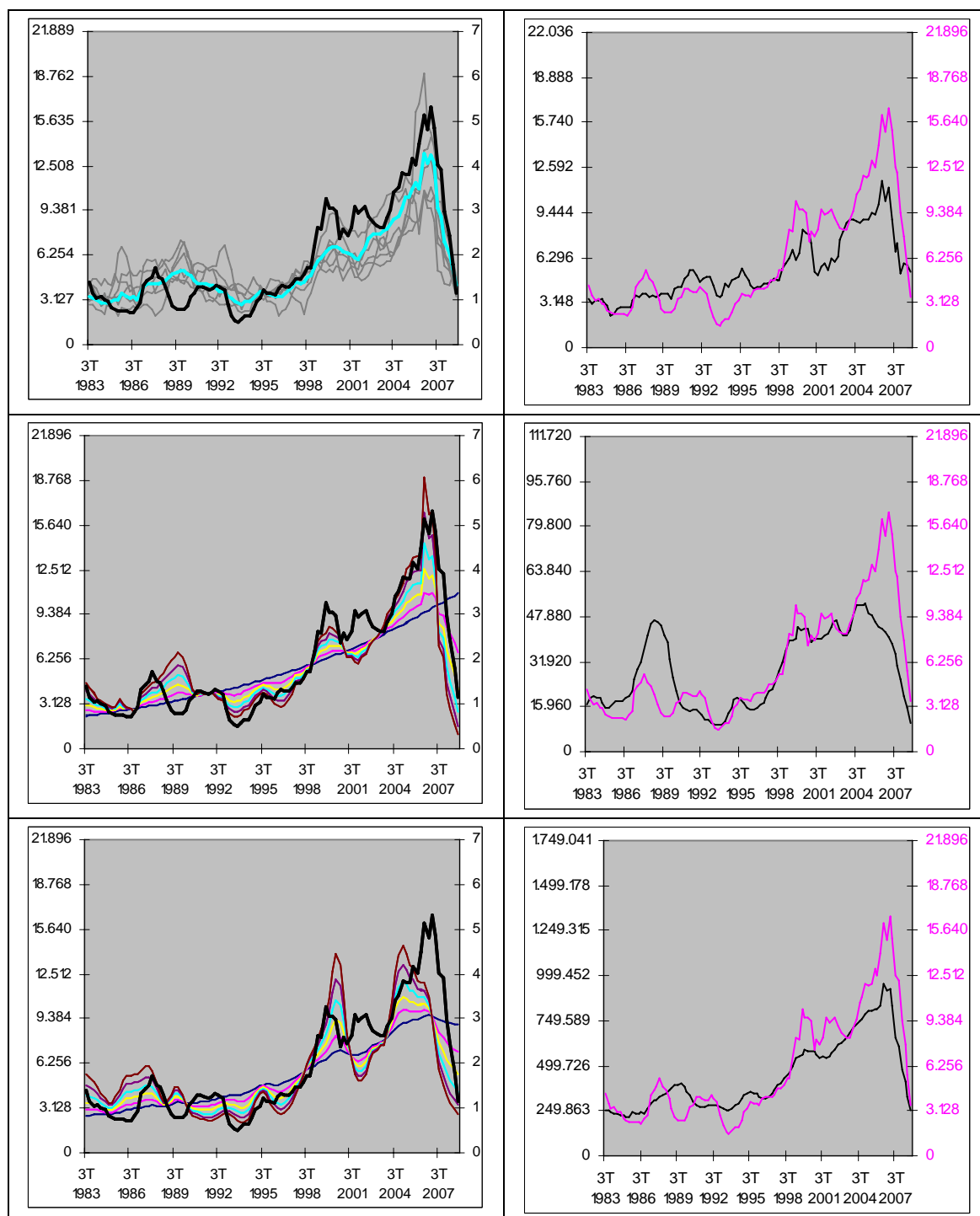


Figura III.19 Cantabria

Paneles izquierdos (Cantabria resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Cantabria: Acumulado anual/valores homogeneizados;

Paneles derechos (Cantabria resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Cantabria (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

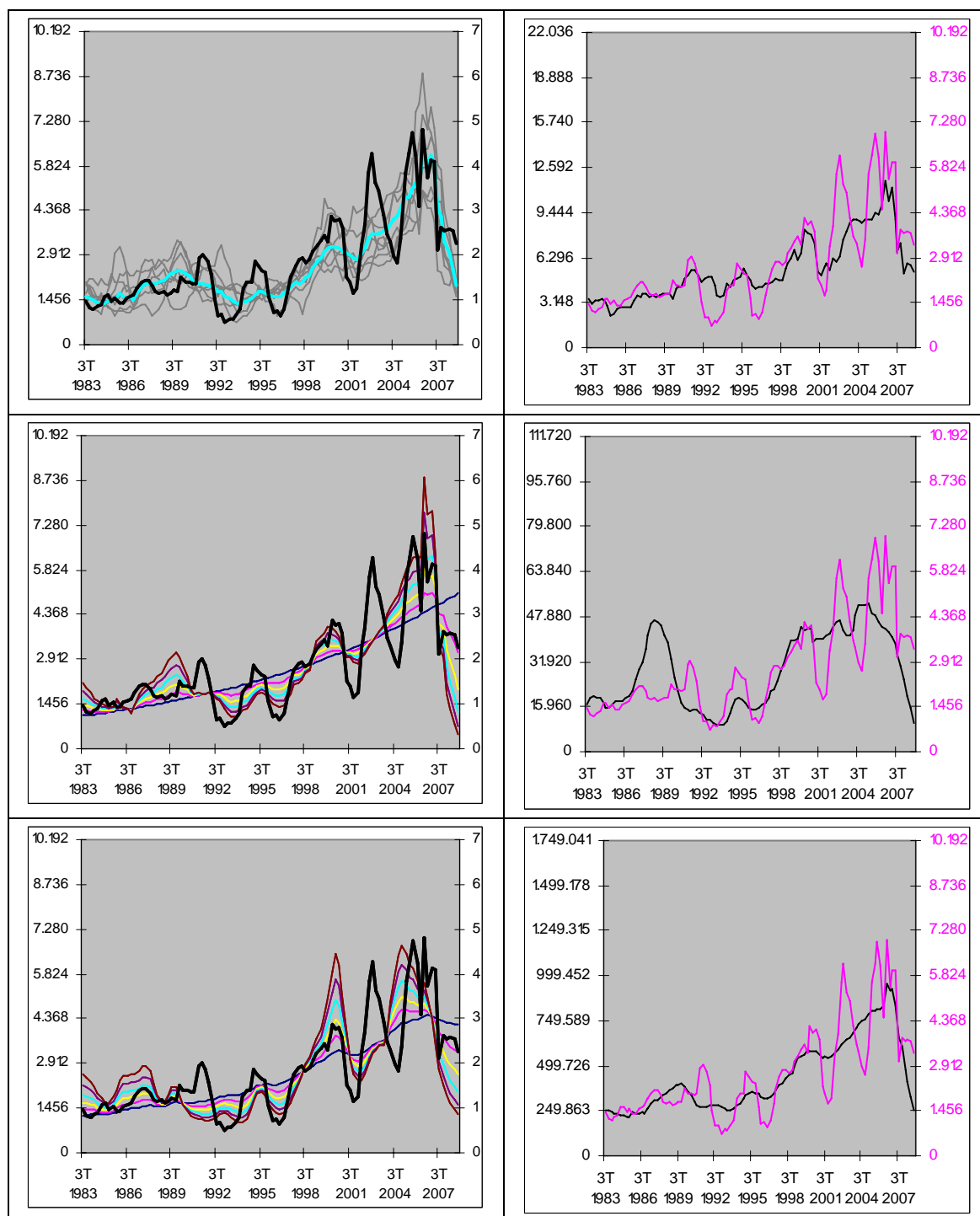


Figura III.20 Álava/Arava

Paneles izquierdos (Álava resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Álava: Acumulado anual/valores homogeneizados; **Paneles derechos** (Álava resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Álava (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

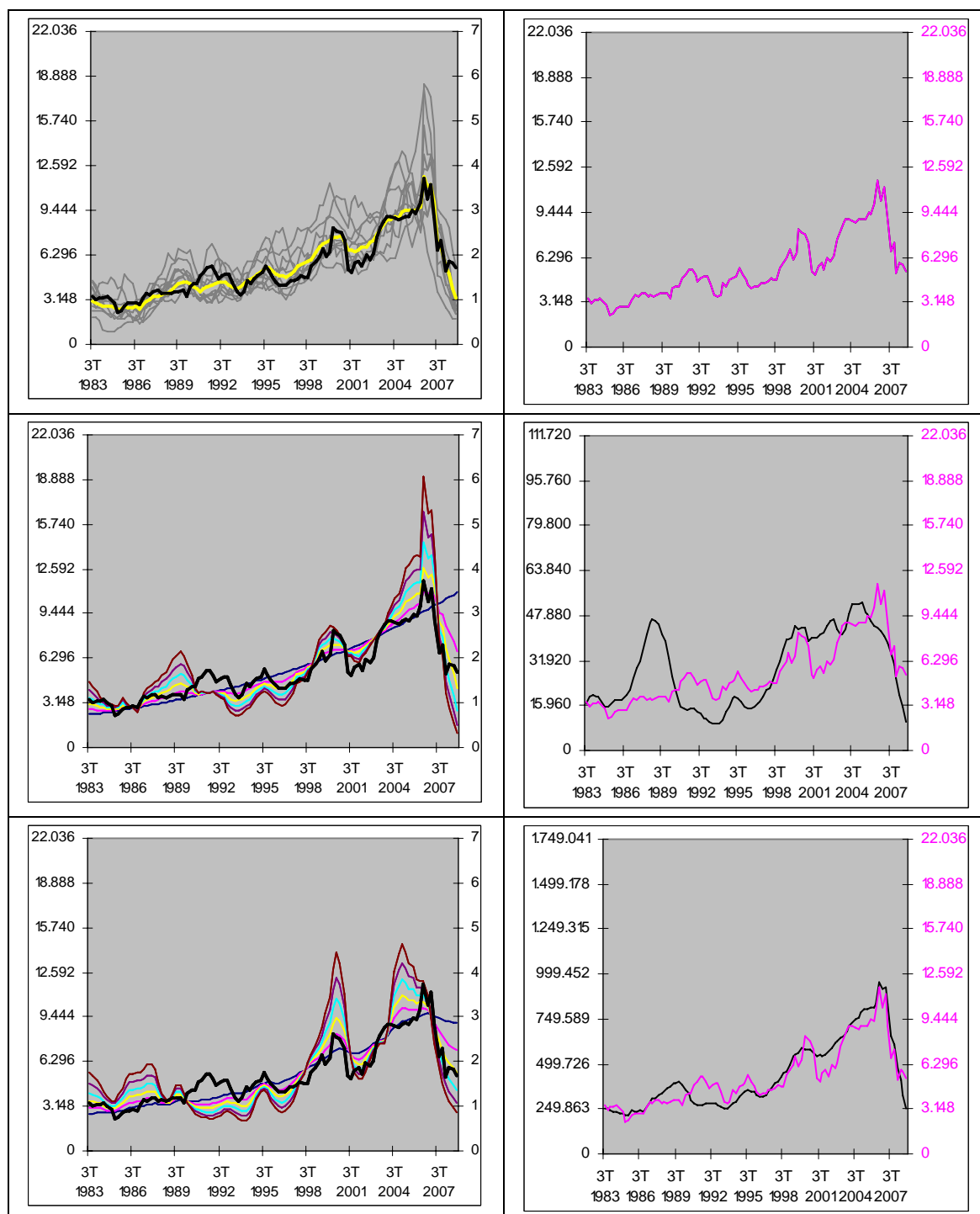


Figura III.21 Córdoba

Paneles izquierdos (Córdoba resaltada en negrita): *Panel Superior:* Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central:* Junto con los centroides iniciales; *Panel Inferior:* Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho:* Unidades visadas en Córdoba: Acumulado anual/valores homogeneizados; **Paneles derechos** (Córdoba resaltada en rojo): *Paneles Superior/Central/Inferior:* Junto con Córdoba/Alicante/España; *Eje de escala derecho:* Unidades visadas en Córdoba (en rojo); *Eje de escala izquierdo:* Unidades visadas en Córdoba/Alicante/España

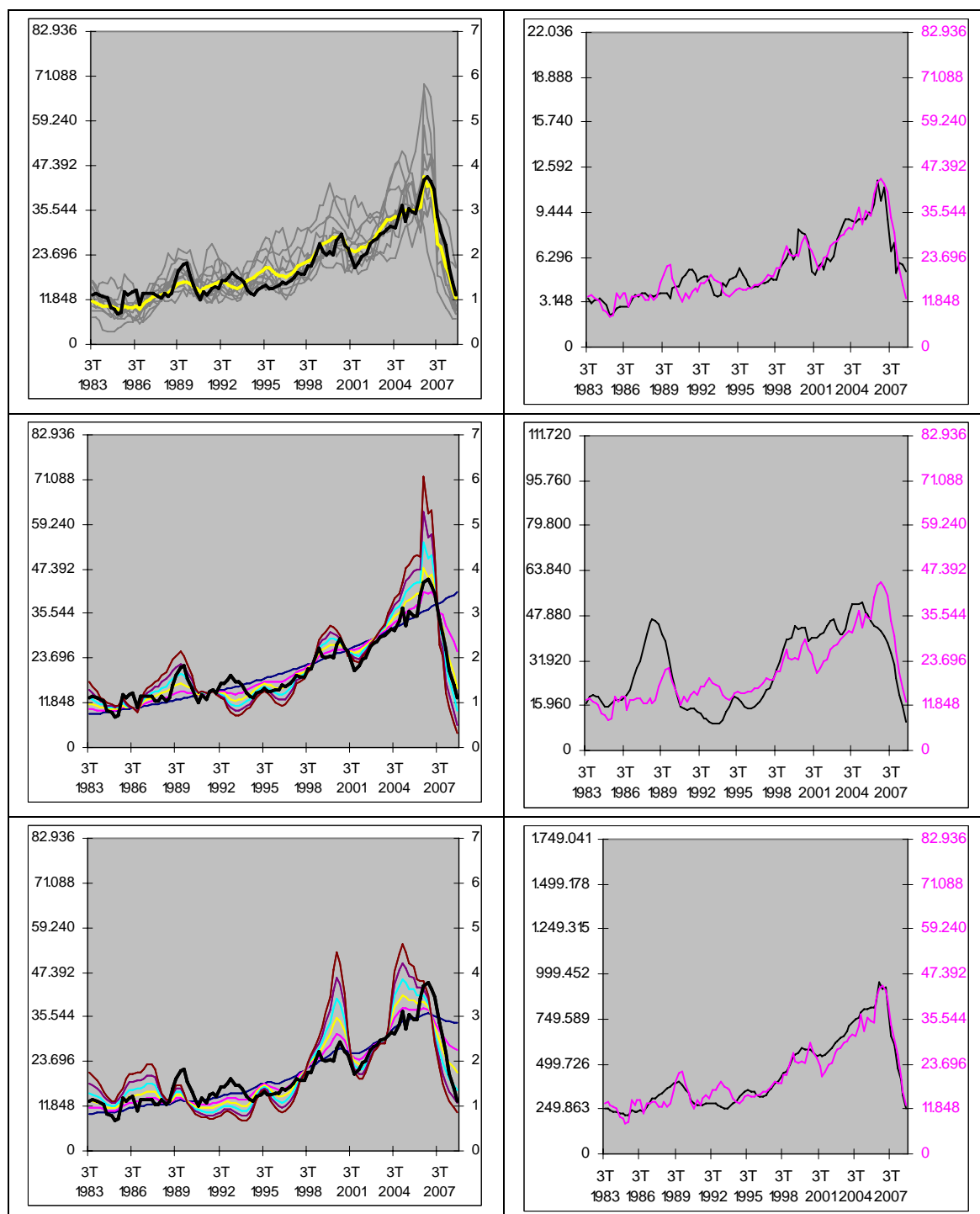


Figura III.22 Valencia

Paneles izquierdos (Valencia resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Valencia: Acumulado anual/valores homogeneizados; **Paneles derechos** (Valencia resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Valencia (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

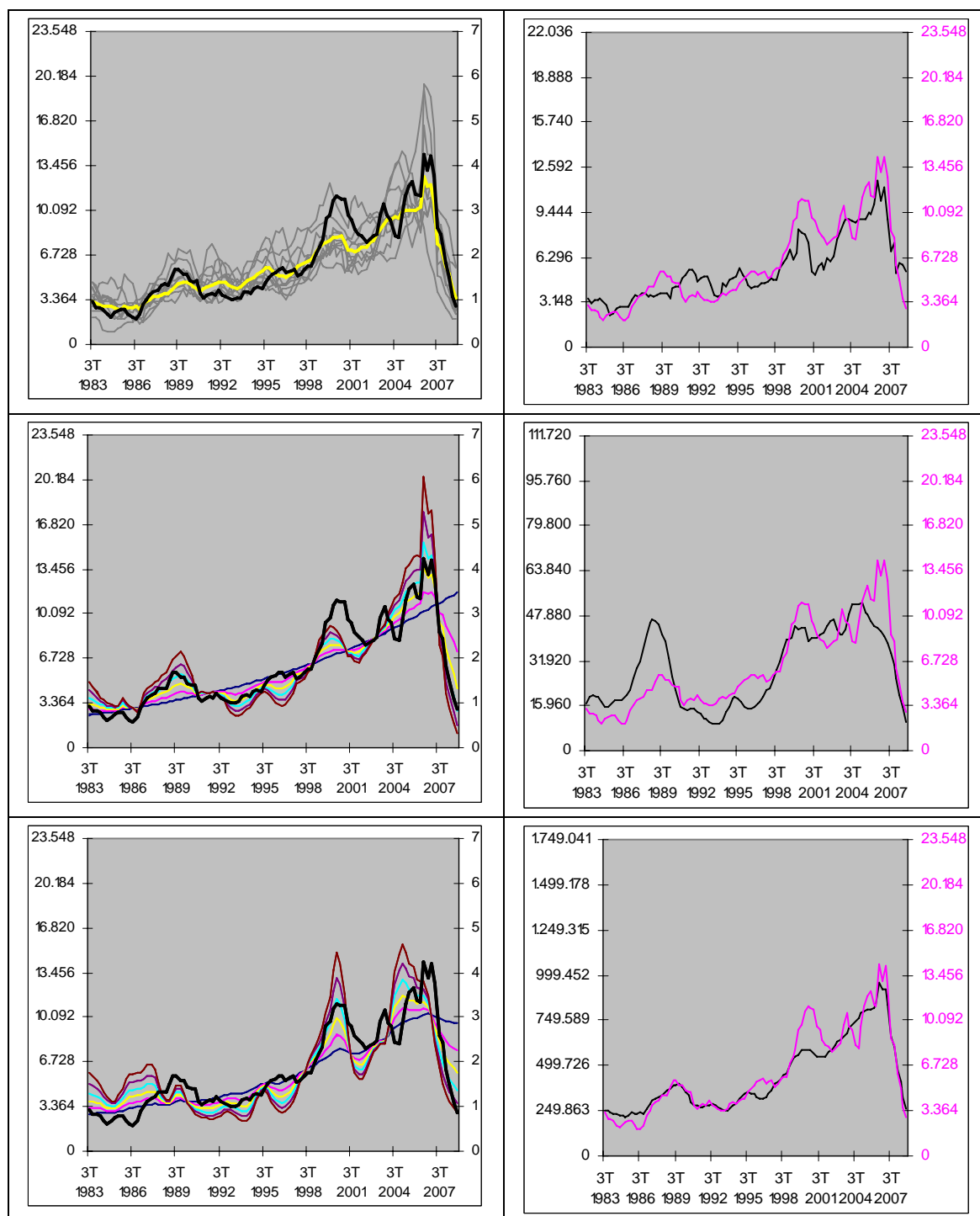


Figura III.23 Huelva

Paneles izquierdos (Huelva resaltada en negra): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Huelva: Acumulado anual/valores homogeneizados; **Paneles derechos** (Huelva resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Huelva (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

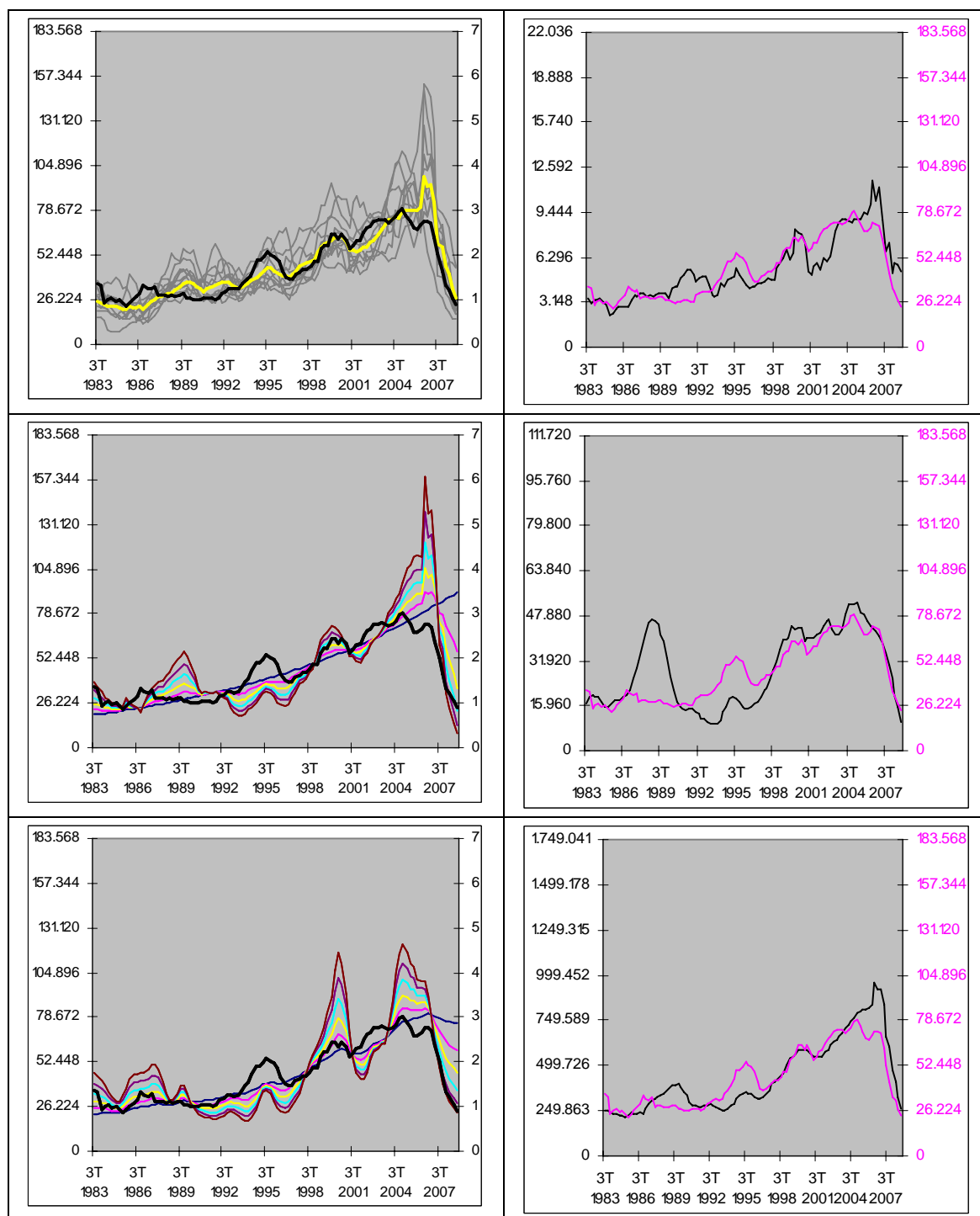


Figura III.24 Madrid

Paneles izquierdos (Madrid resaltada en negrita): *Panel Superior:* Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central:* Junto con los centroides iniciales; *Panel Inferior:* Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho:* Unidades visadas en Madrid: Acumulado anual/valores homogeneizados; **Paneles derechos** (Madrid resaltada en rojo): *Paneles Superior/Central/Inferior:* Junto con Córdoba/Alicante/España; *Eje de escala derecho:* Unidades visadas en Madrid (en rojo); *Eje de escala izquierdo:* Unidades visadas en Córdoba/Alicante/España

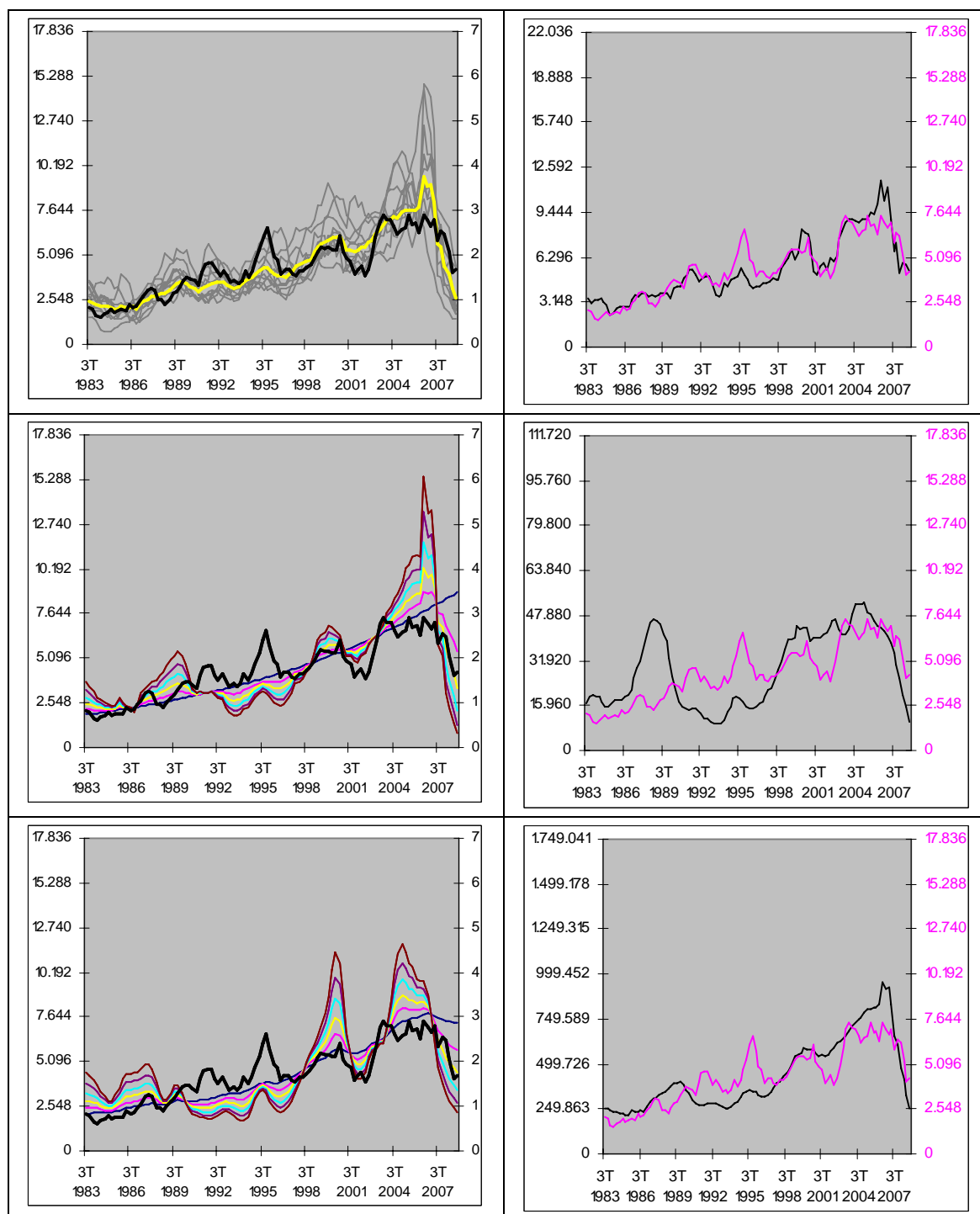


Figura III.25 Guipúzcoa/Gipuzkoa

Paneles izquierdos (Guipúzcoa resaltada en negra): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Guipúzcoa: Acumulado anual/valores homogeneizados; **Paneles derechos** (Guipúzcoa resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Guipúzcoa (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

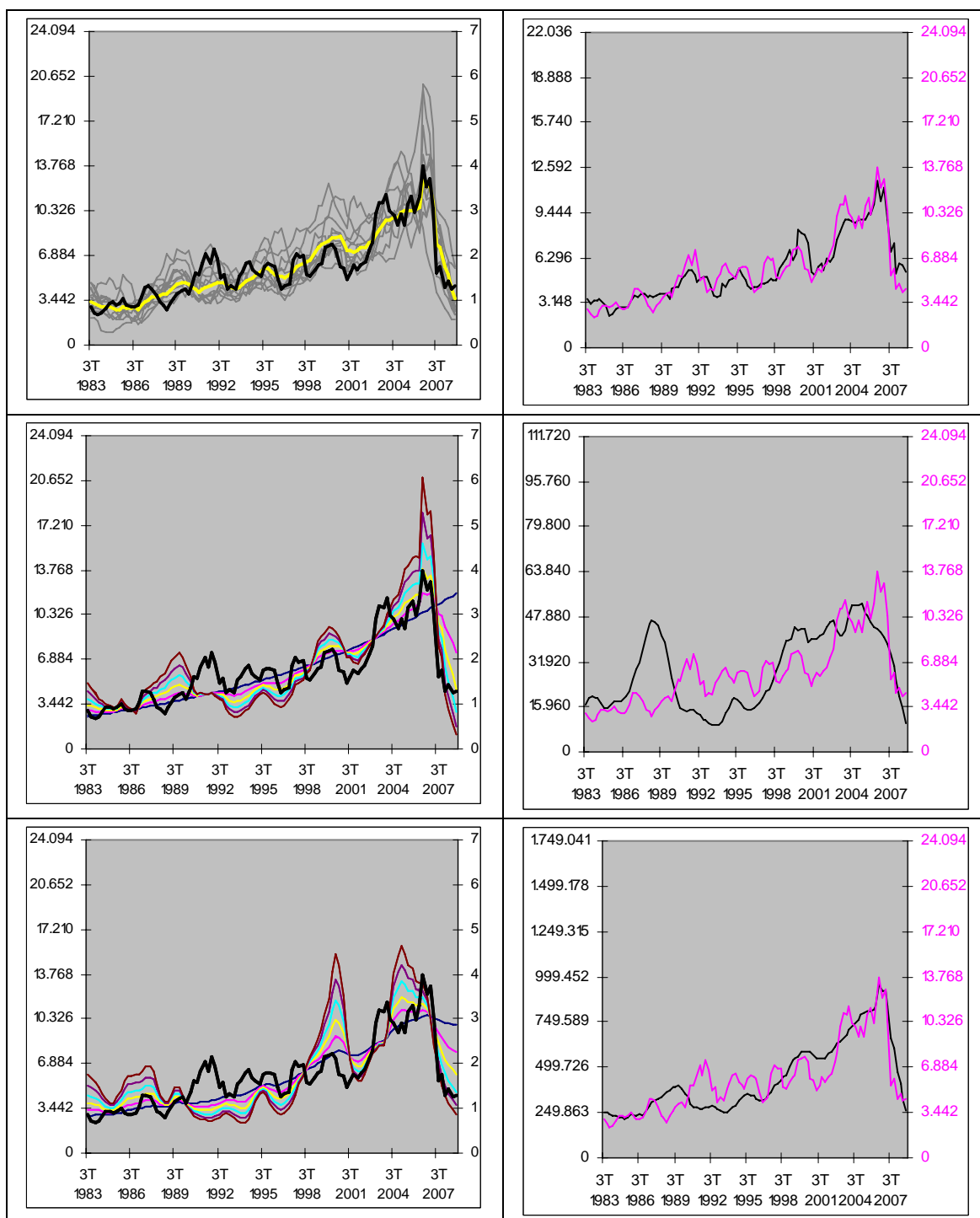


Figura III.26 Navarra

Paneles izquierdos (Navarra resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Navarra: Acumulado anual/valores homogeneizados; **Paneles derechos** (Navarra resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Navarra (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

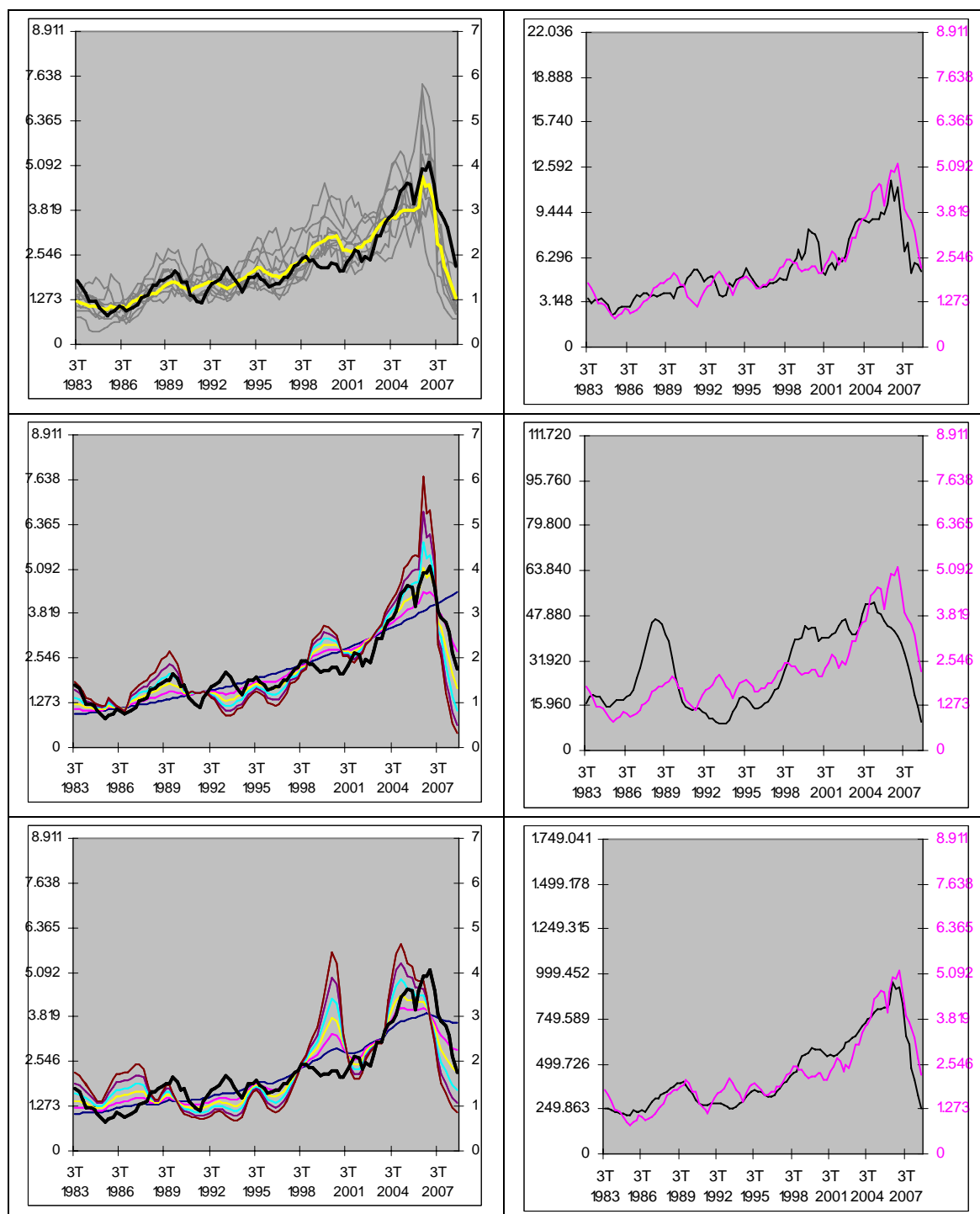


Figura III.27 Segovia

Panes izquierdos (Segovia resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Segovia: Acumulado anual/valores homogeneizados; **Panes derechos** (Segovia resaltada en rojo): *Panes Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Segovia (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

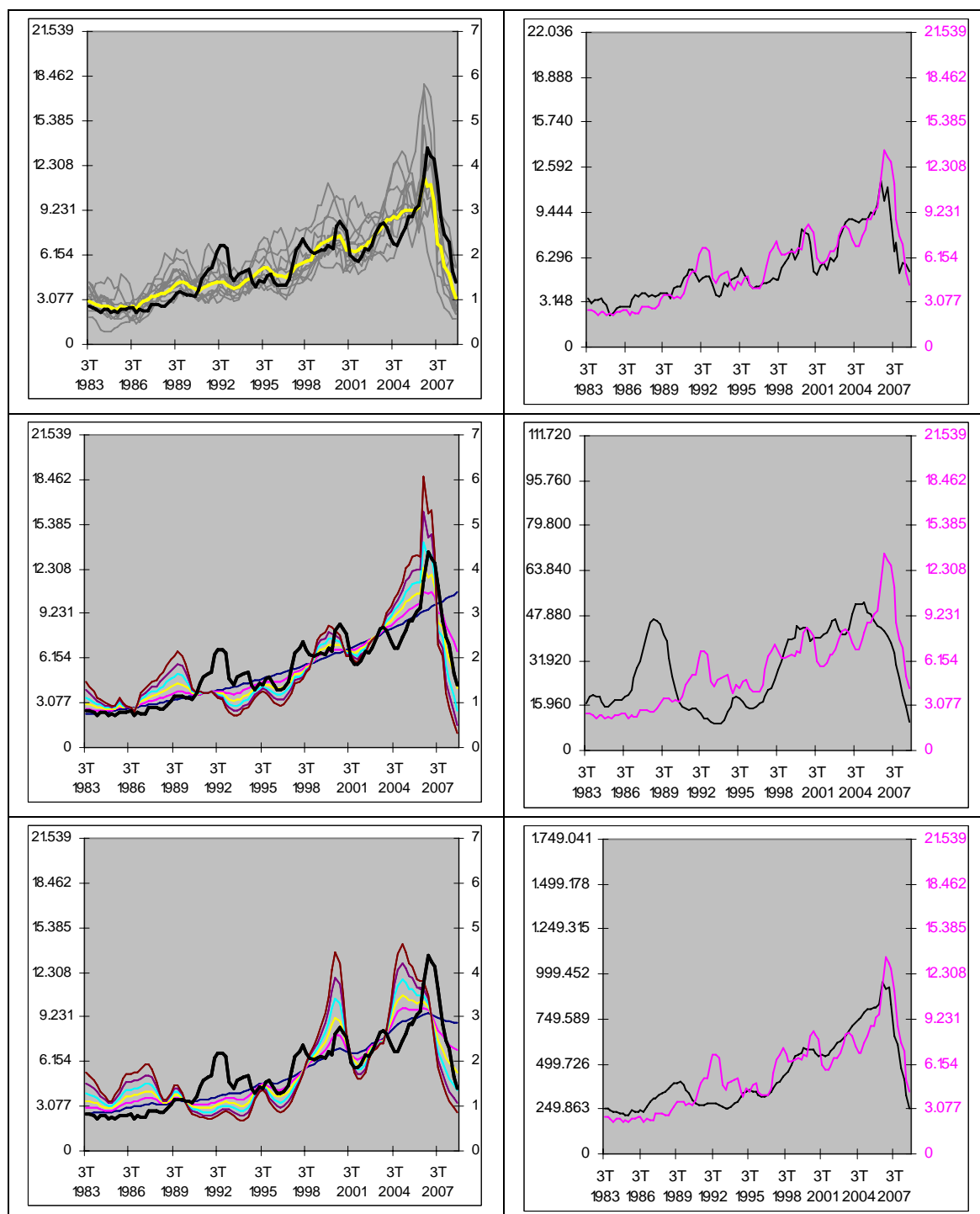


Figura III.28 Badajoz

Paneles izquierdos (Badajoz resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Badajoz: Acumulado anual/valores homogeneizados; **Paneles derechos** (Badajoz resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Badajoz (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

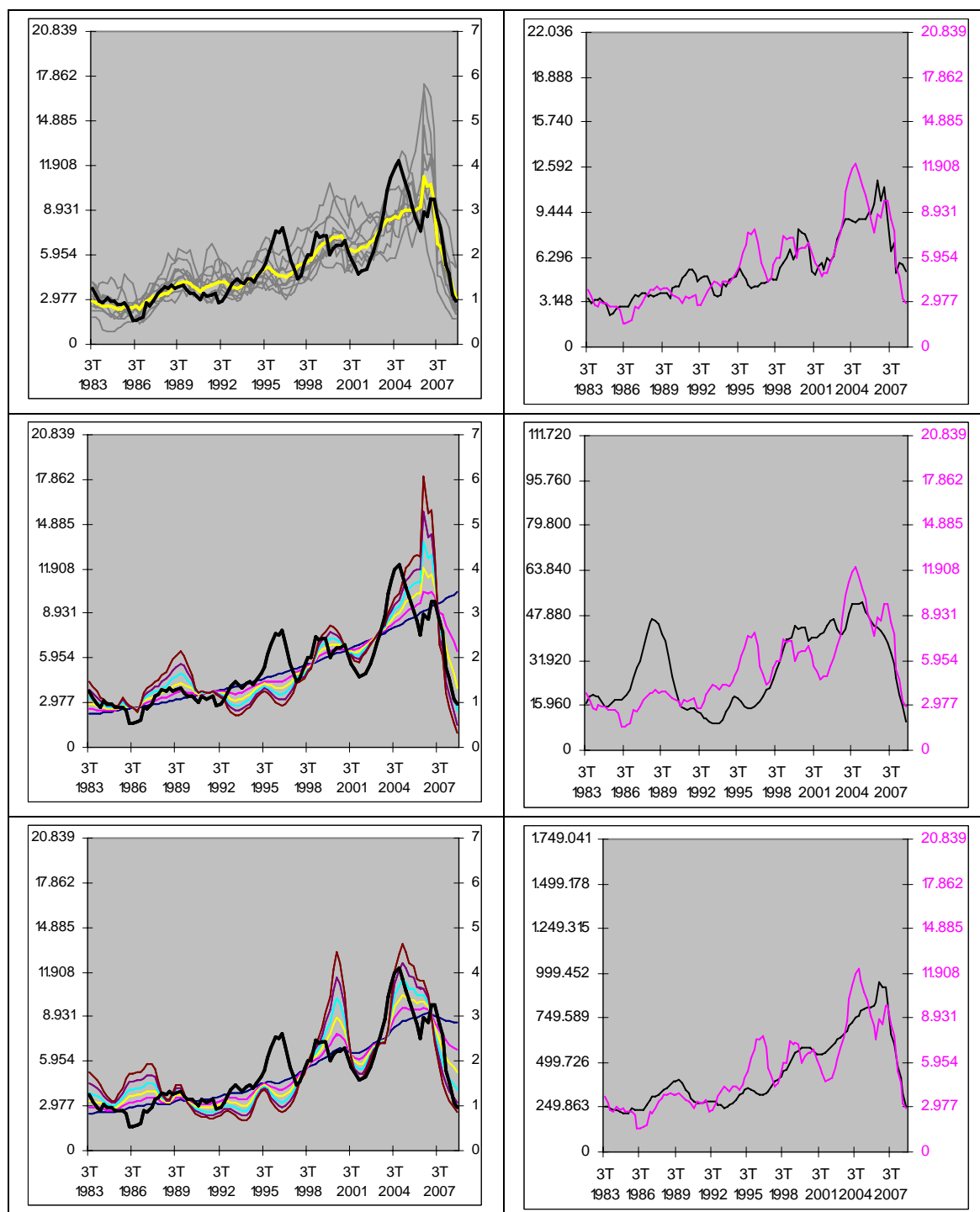


Figura III.29 Valladolid

Paneles izquierdos (Valladolid resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Valladolid: Acumulado anual/valores homogeneizados;

Paneles derechos (Valladolid resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Valladolid (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

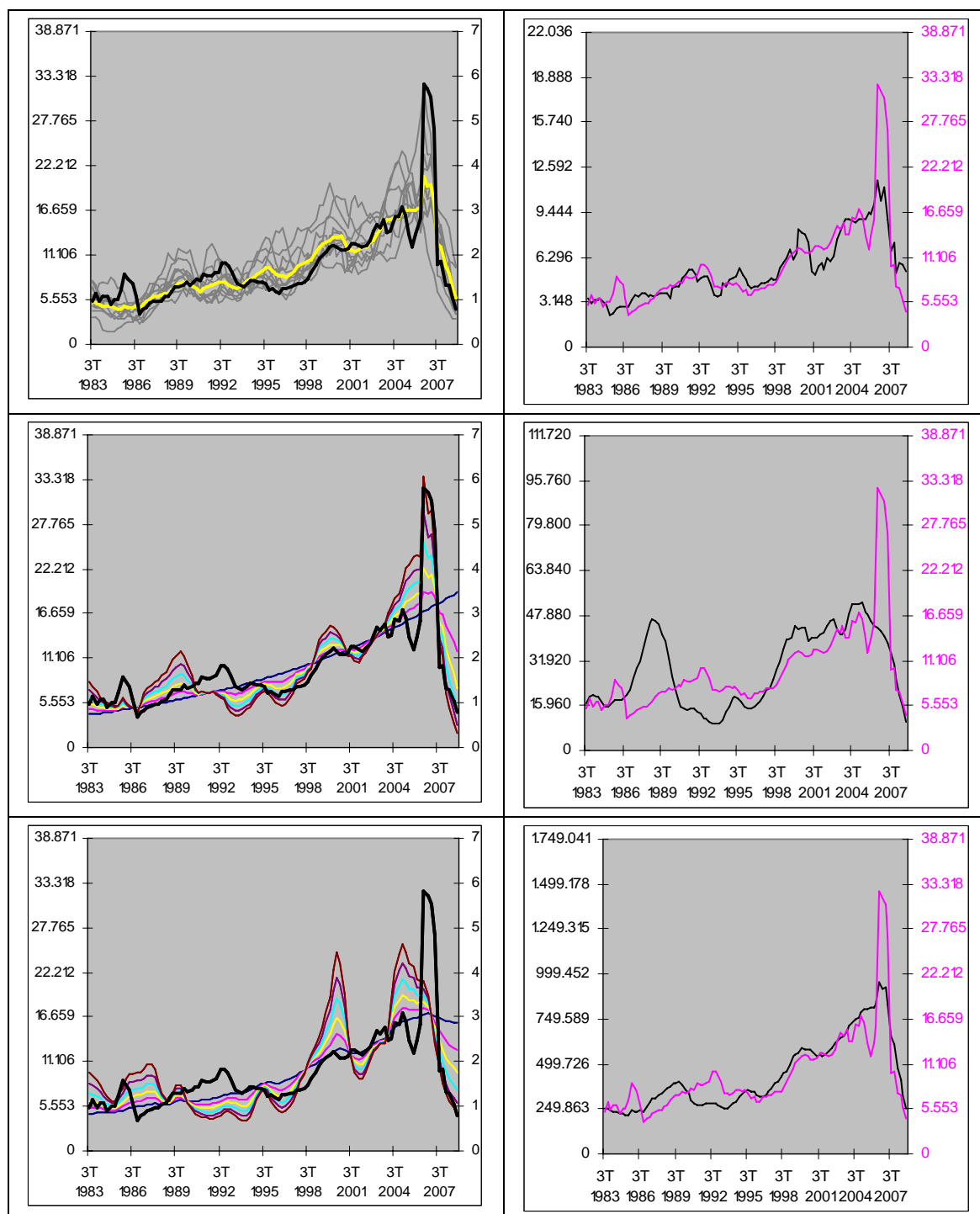


Figura III.30 Granada

Paneles izquierdos (Granada resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Granada: Acumulado anual/valores homogeneizados; **Paneles derechos** (Granada resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Granada (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

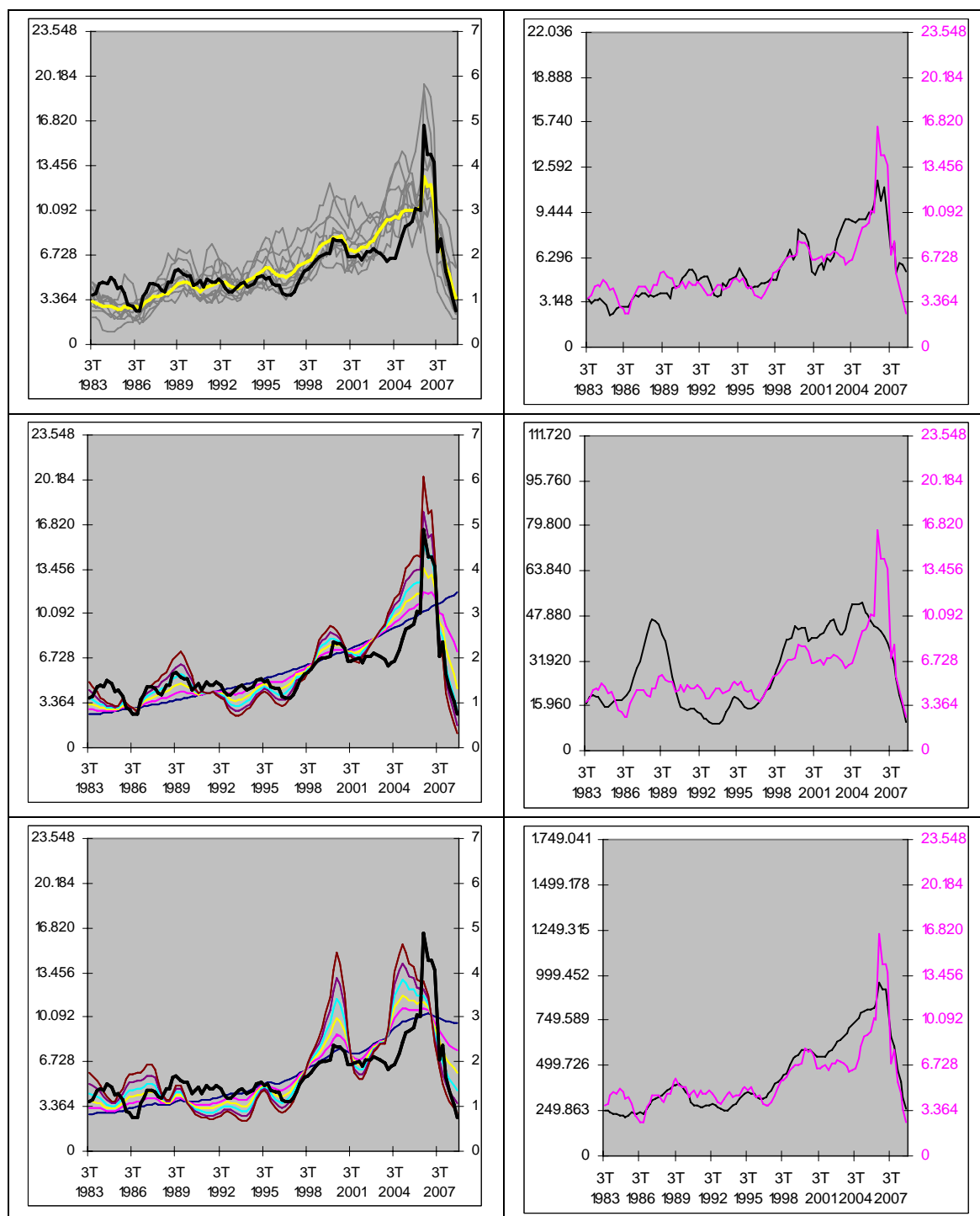


Figura III.31 Jaén

Paneles izquierdos (Jaén resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Jaén; *Eje de escala derecho*: Unidades visadas en Córdoba/Alicante/España; **Paneles derechos** (Jaén resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Jaén (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

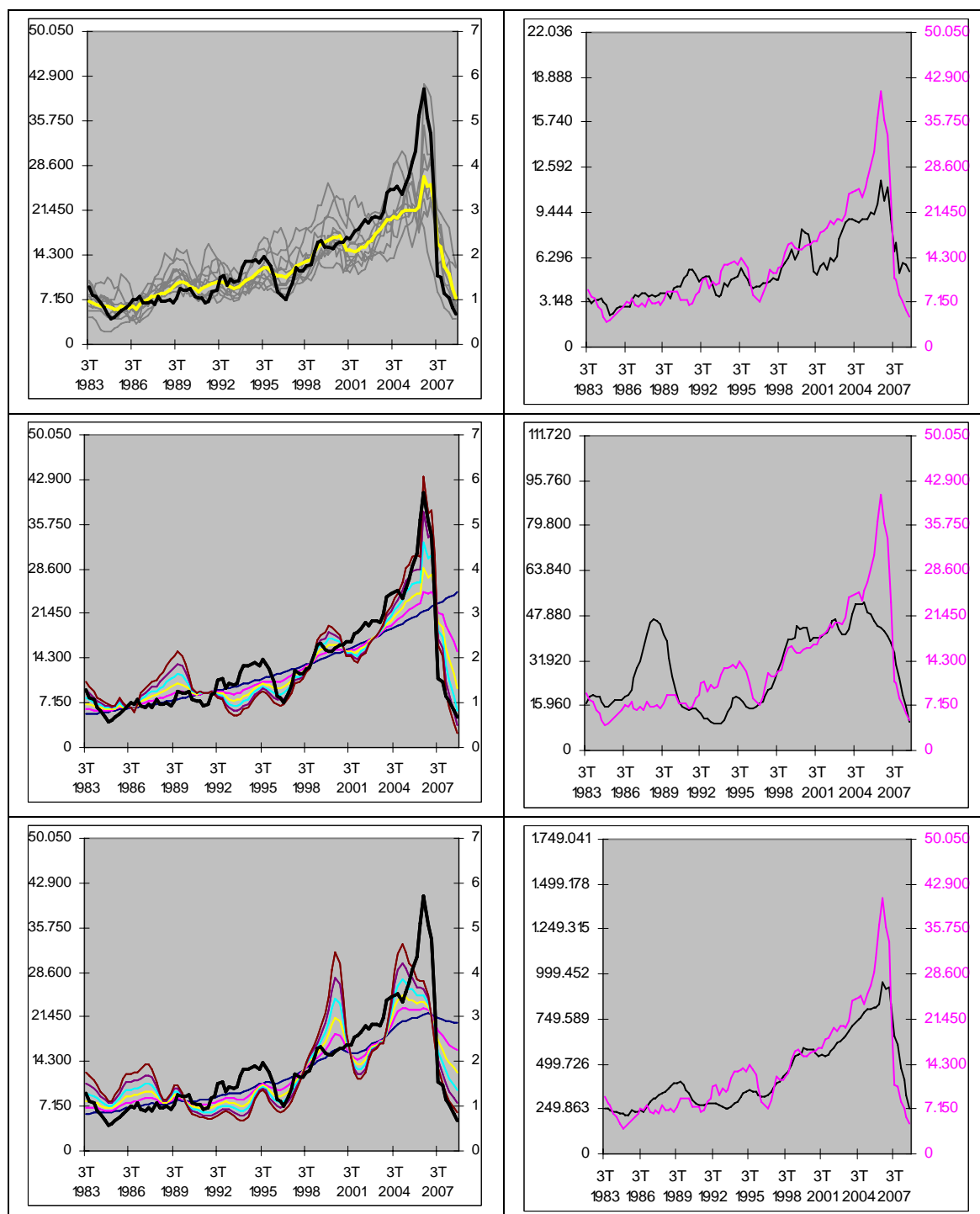


Figura III.32 Sevilla

Paneles izquierdos (Sevilla resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Sevilla: Acumulado anual/valores homogeneizados; **Paneles derechos** (Sevilla resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Sevilla (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

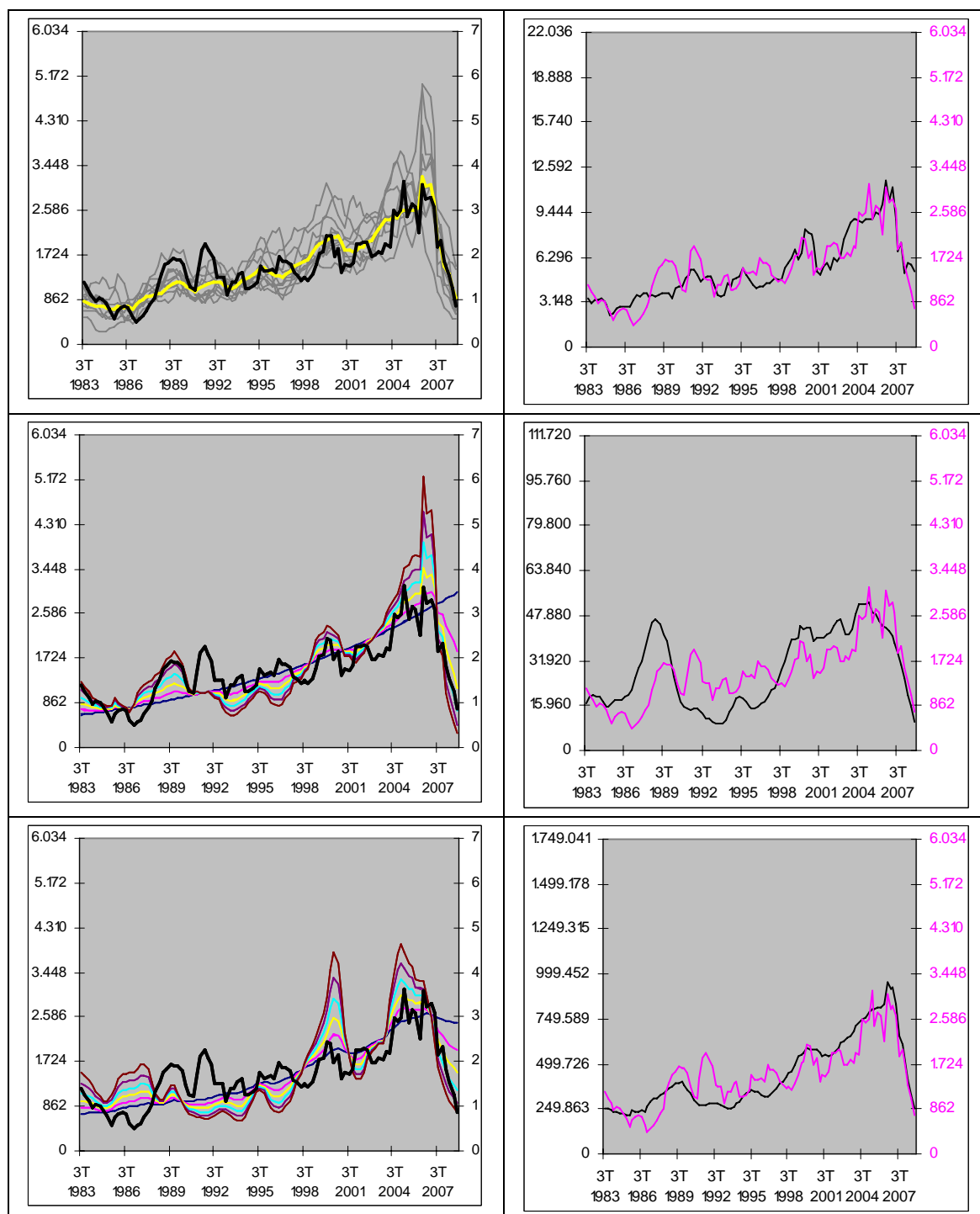


Figura III.33 Palencia

Paneles izquierdos (Palencia resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Palencia: Acumulado anual/valores homogeneizados; **Paneles derechos** (Palencia resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Palencia (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

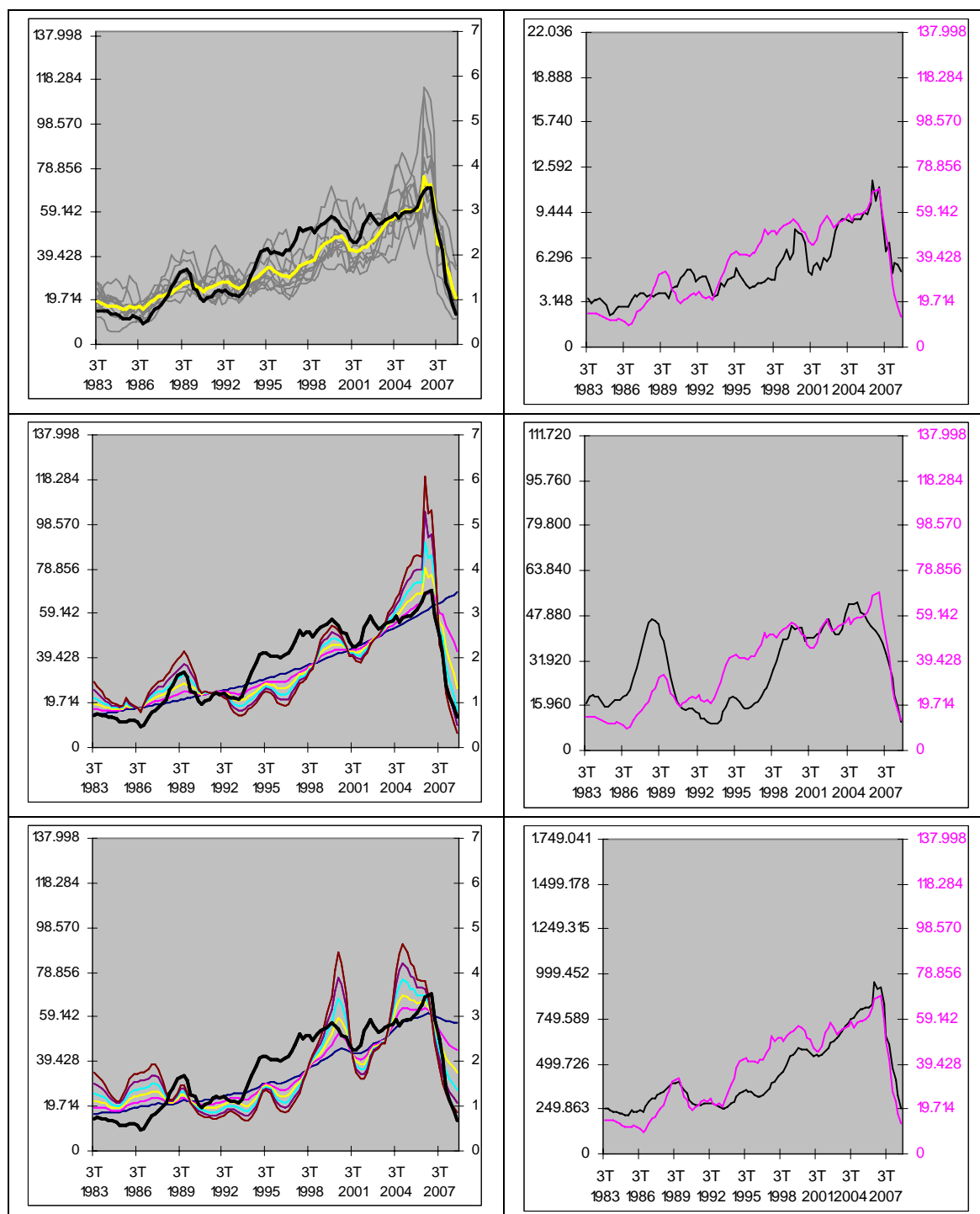


Figura III.34 Barcelona

Paneles izquierdos (Barcelona resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Barcelona: Acumulado anual/valores homogeneizados; **Paneles derechos** (Barcelona resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Barcelona (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

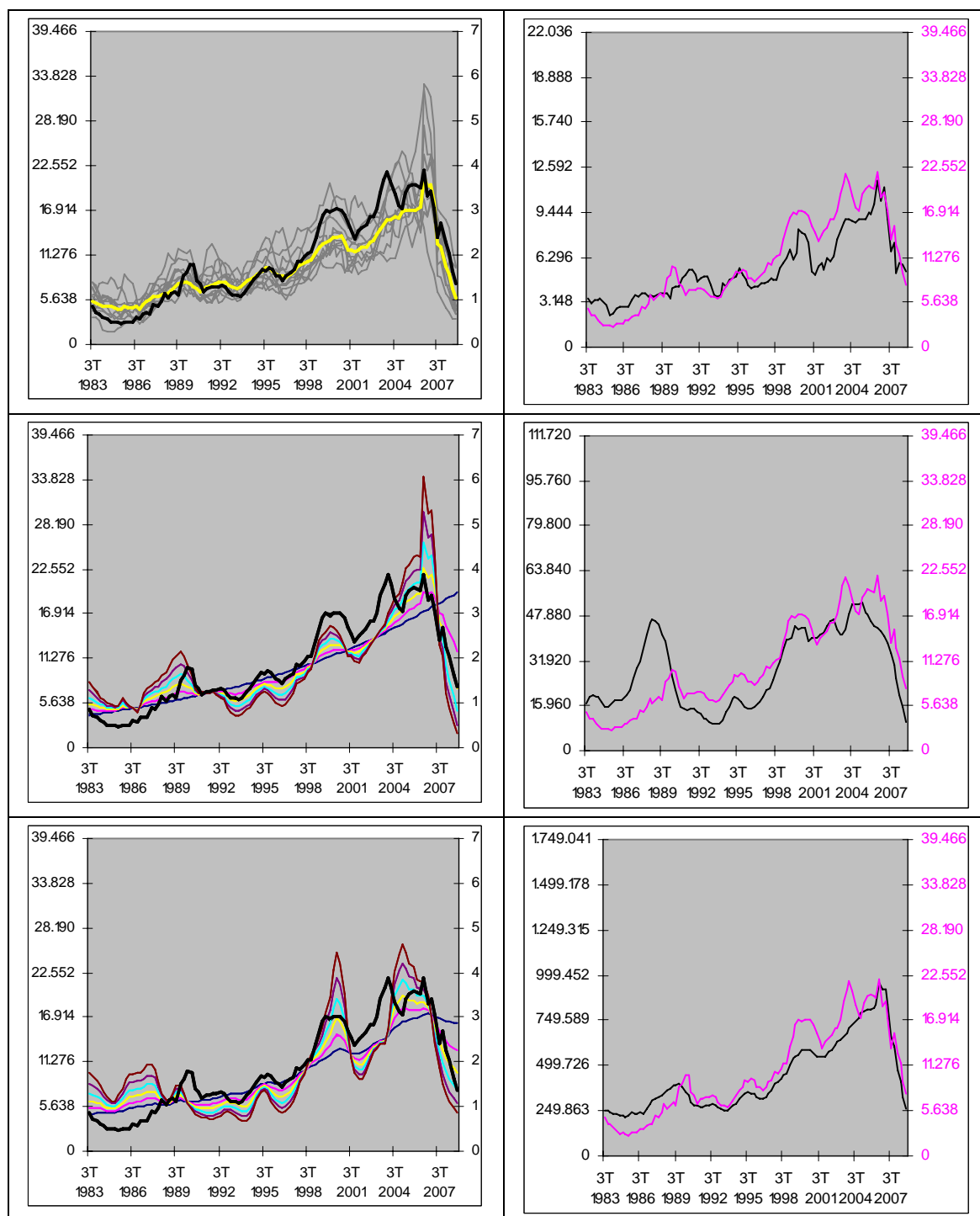


Figura III.35 La Coruña/A Coruña

Paneles izquierdos (La Coruña resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en La Coruña: Acumulado anual/valores homogeneizados; **Paneles derechos** (La Coruña resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en La Coruña (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

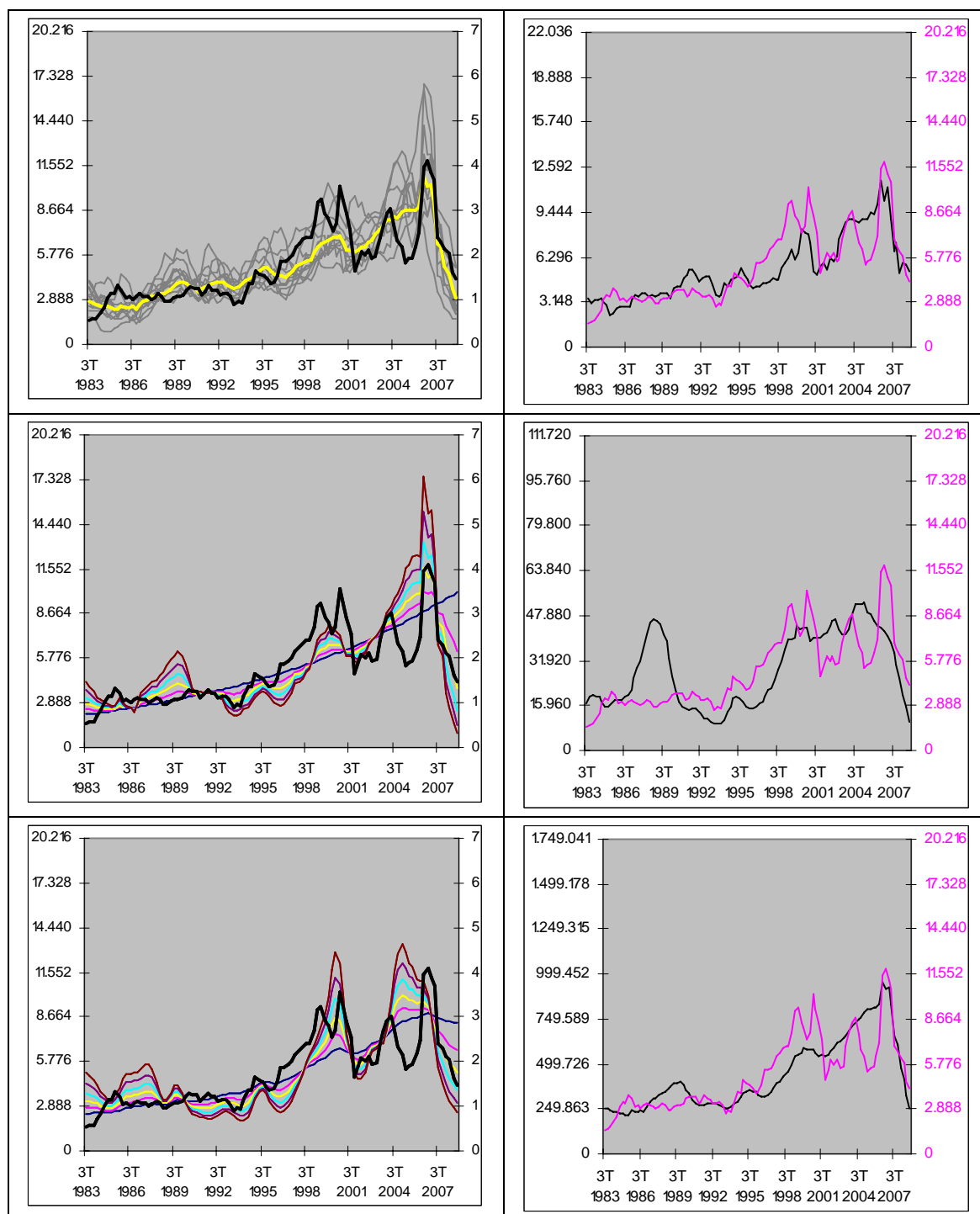


Figura III.36 Vizcaya/Bizkaia

Paneles izquierdos (Vizcaya resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Vizcaya: Acumulado anual/valores homogeneizados; **Paneles derechos** (Vizcaya resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Vizcaya (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

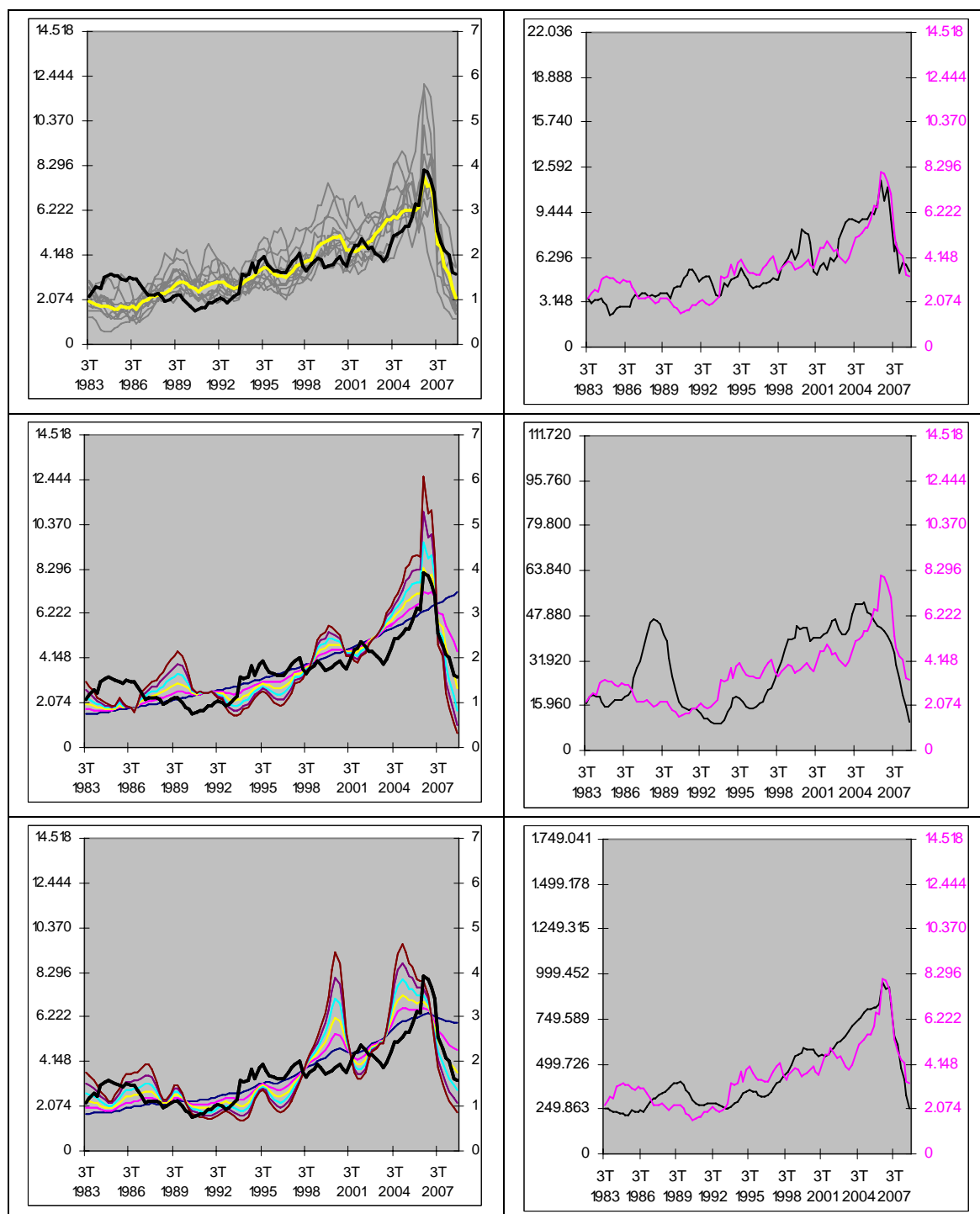


Figura III.37 Cáceres

Paneles izquierdos (Cáceres resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Cáceres: Acumulado anual/valores homogeneizados; **Paneles derechos** (Cáceres resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Cáceres (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

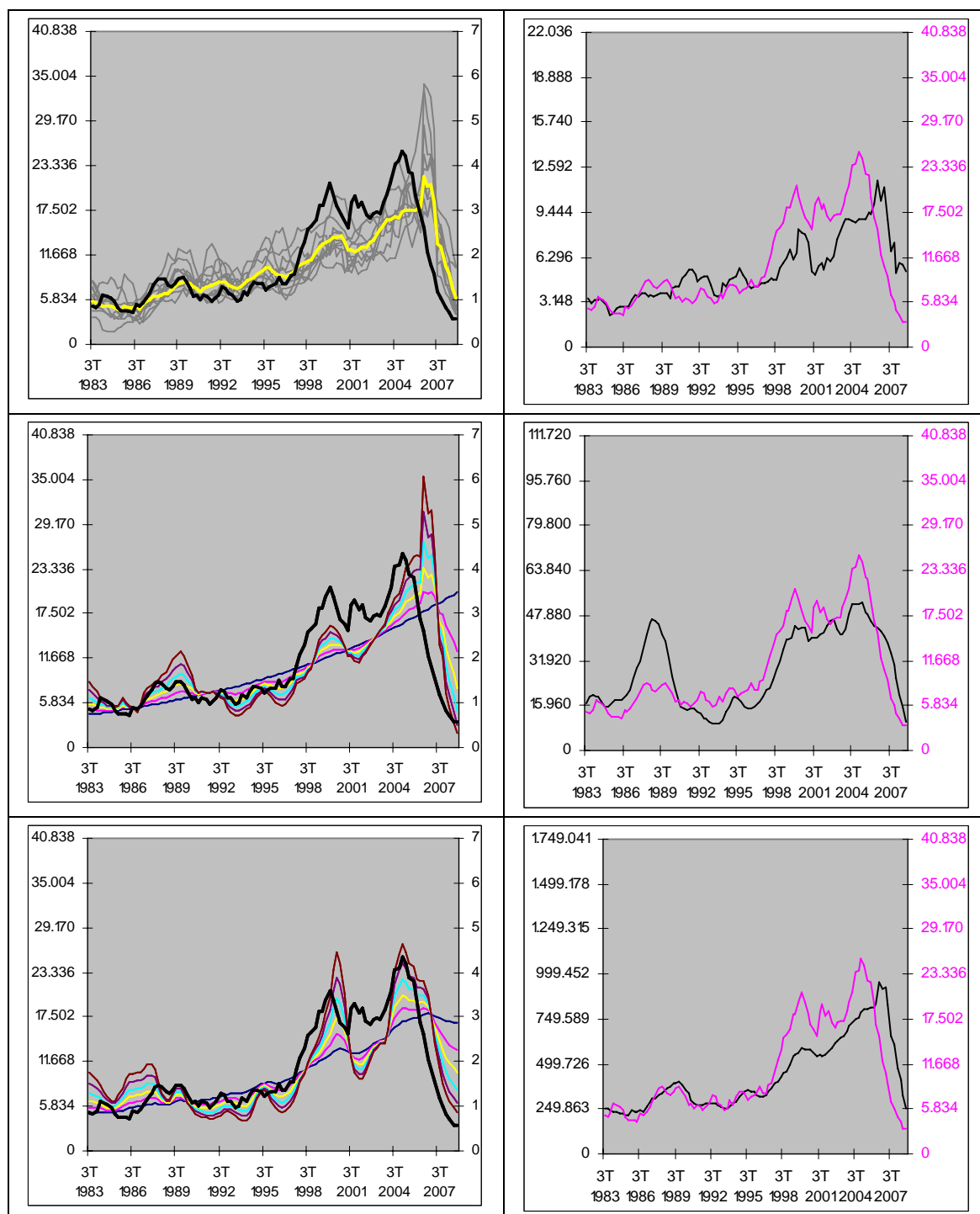


Figura III.38 Cádiz

Paneles izquierdos (Cádiz resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Cádiz: Acumulado anual/valores homogeneizados;

Paneles derechos (Cádiz resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Cádiz (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

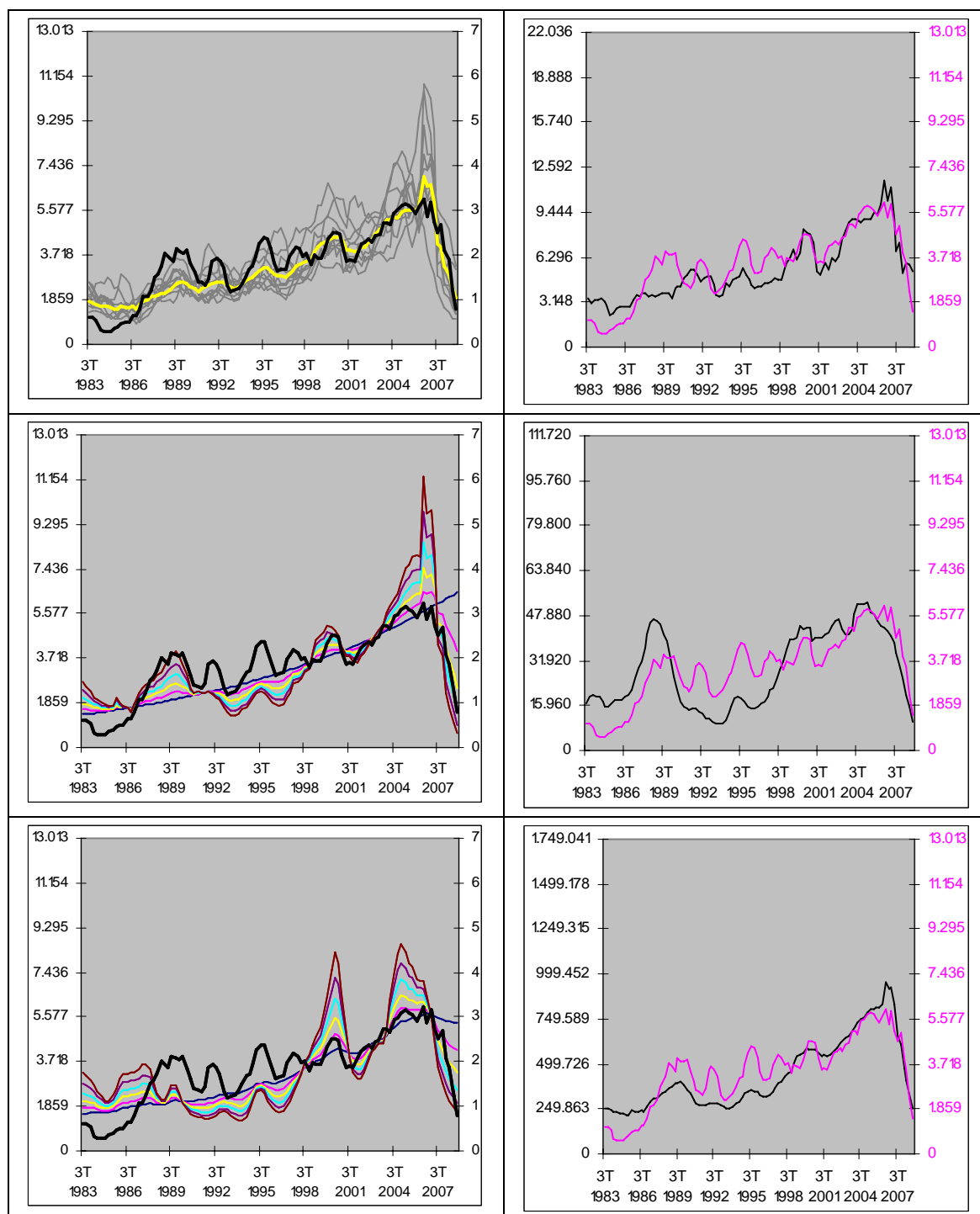


Figura III.39 Salamanca

Paneles izquierdos (Salamanca resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Salamanca: Acumulado anual/valores homogeneizados;

Paneles derechos (Salamanca resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Salamanca (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

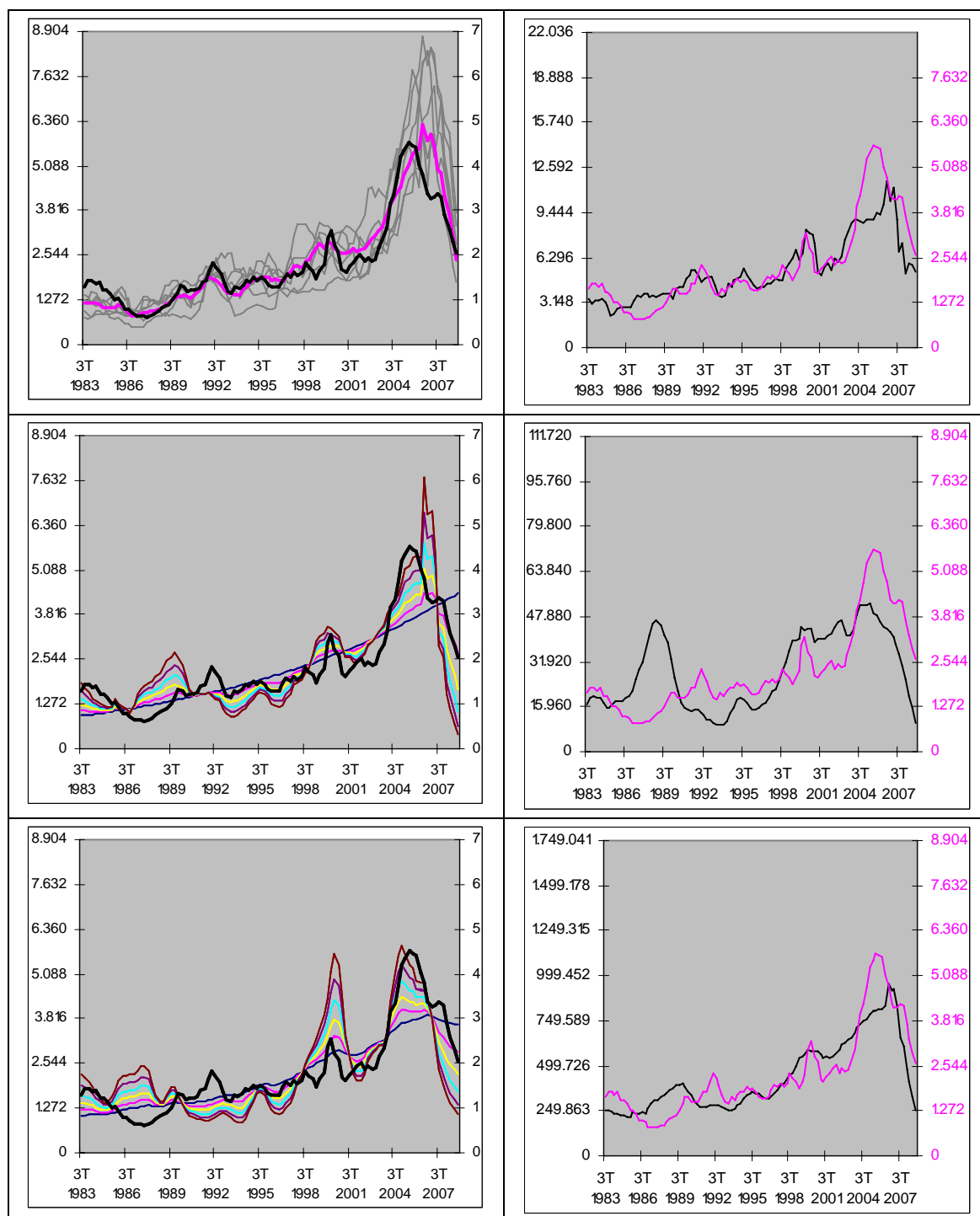


Figura III.40 Avila

Paneles izquierdos (Ávila resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Ávila: Acumulado anual/valores homogeneizados; **Paneles derechos** (Ávila resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Ávila (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

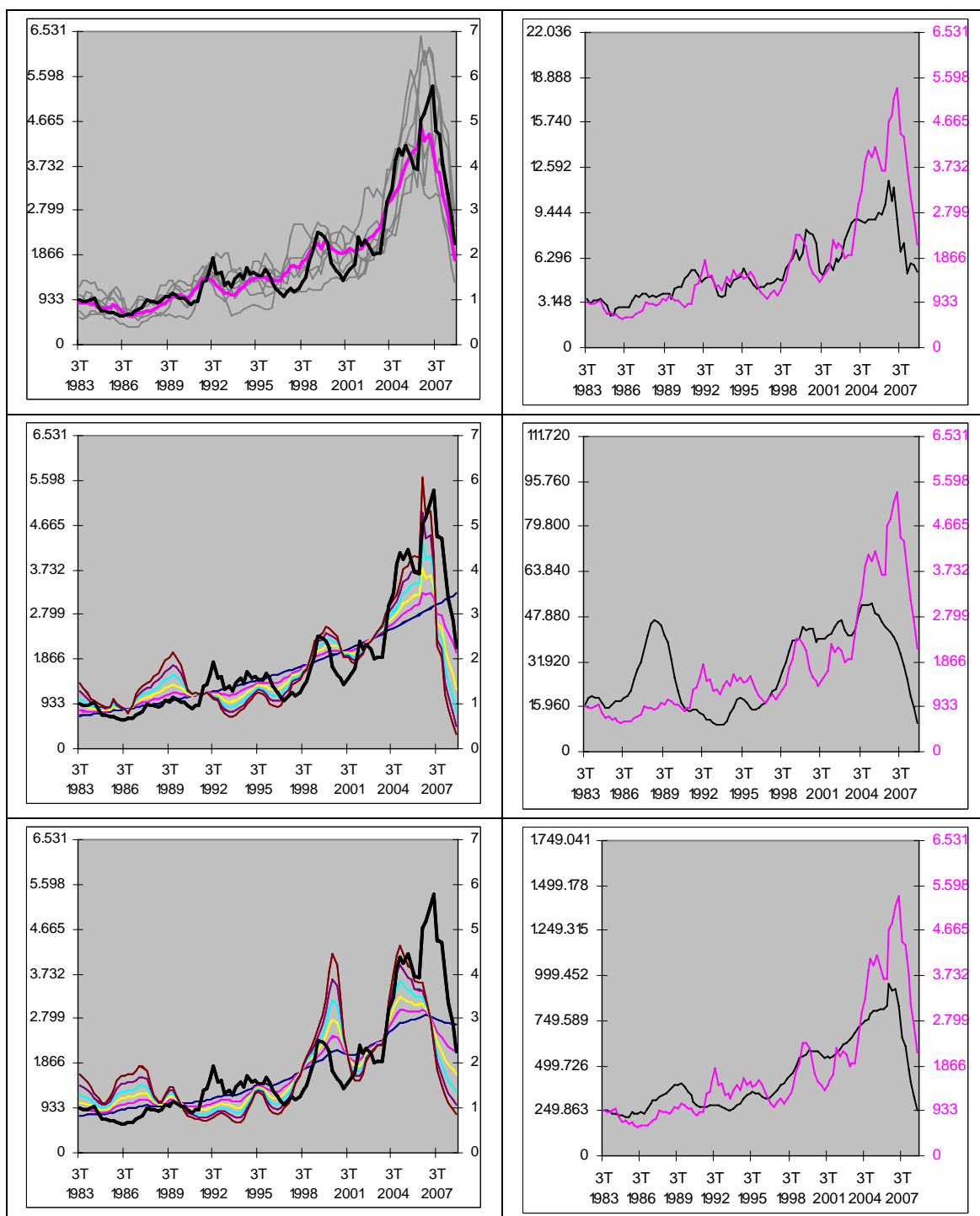


Figura III.41 Cuenca

Paneles izquierdos (Cuenca resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Cuenca: Acumulado anual/valores homogeneizados; **Paneles derechos** (Cuenca resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Cuenca (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

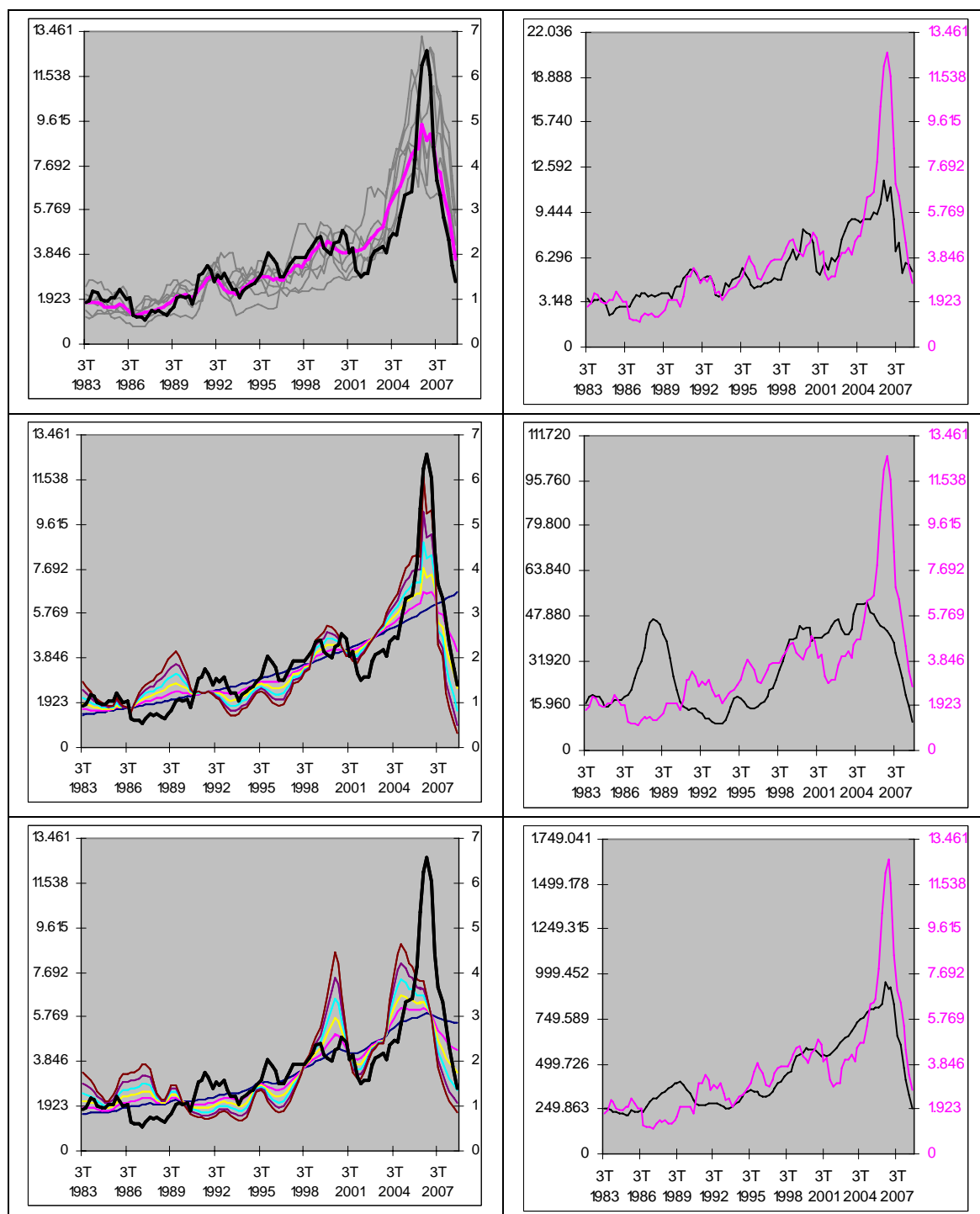


Figura III.42 Albacete

Paneles izquierdos (Albacete resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Albacete: Acumulado anual/valores homogeneizados;

Paneles derechos (Albacete resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Albacete (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

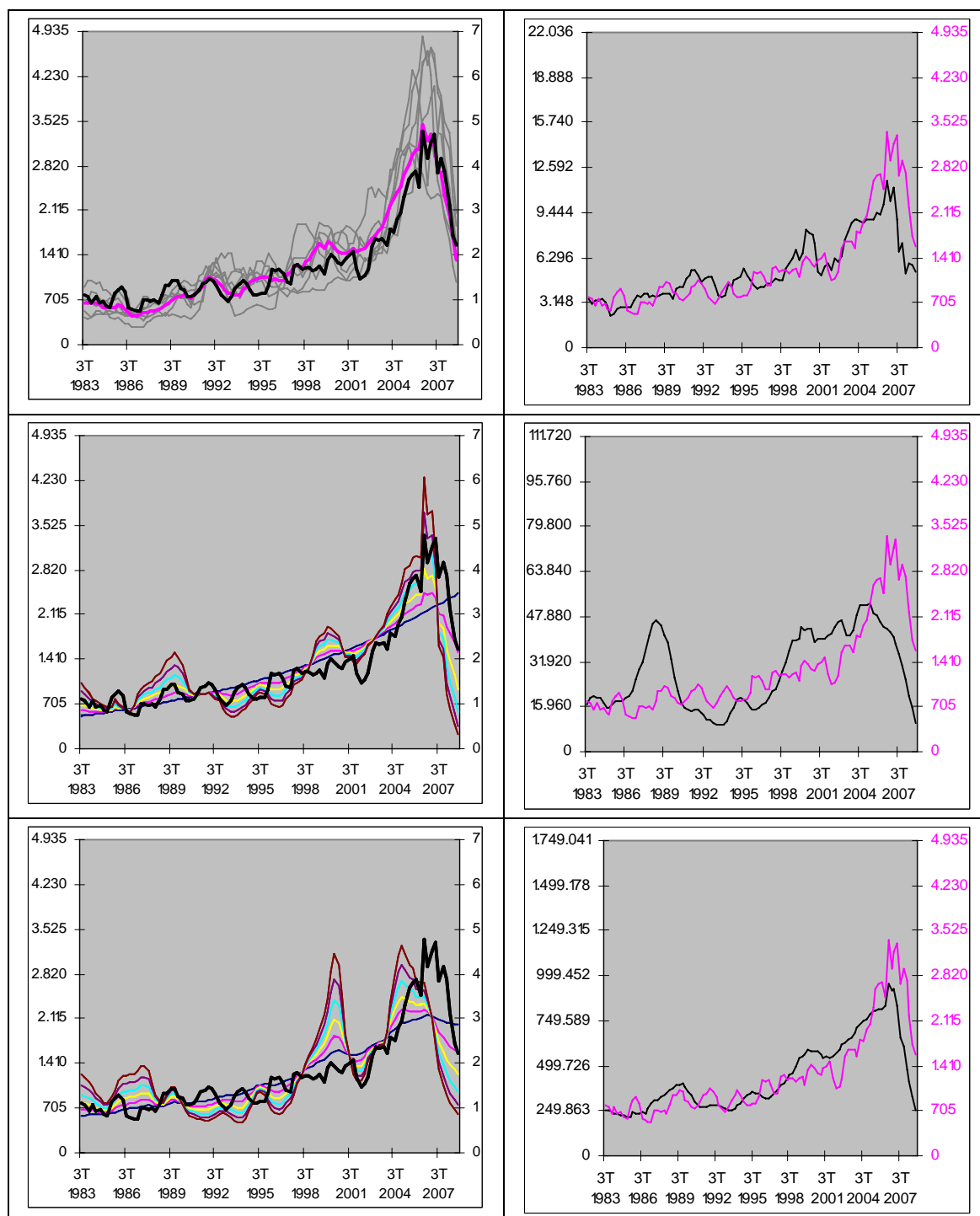


Figura III.43 Teruel

Paneles izquierdos (Teruel resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Teruel: Acumulado anual/valores homogeneizados; **Paneles derechos** (Teruel resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Teruel (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

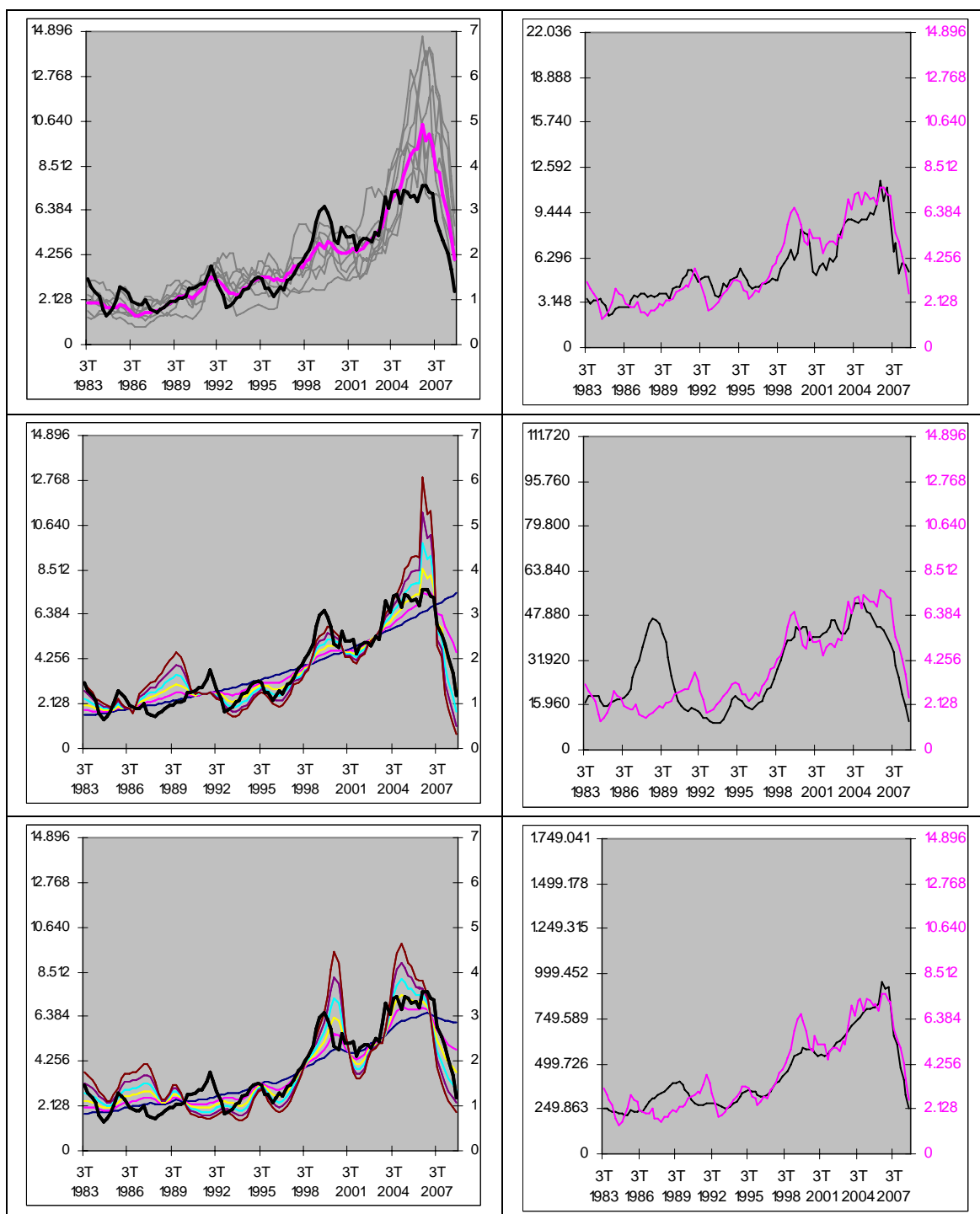


Figura III.44 Burgos

Paneles izquierdos (Burgos resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Burgos: Acumulado anual/valores homogeneizados; **Paneles derechos** (Burgos resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Burgos (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

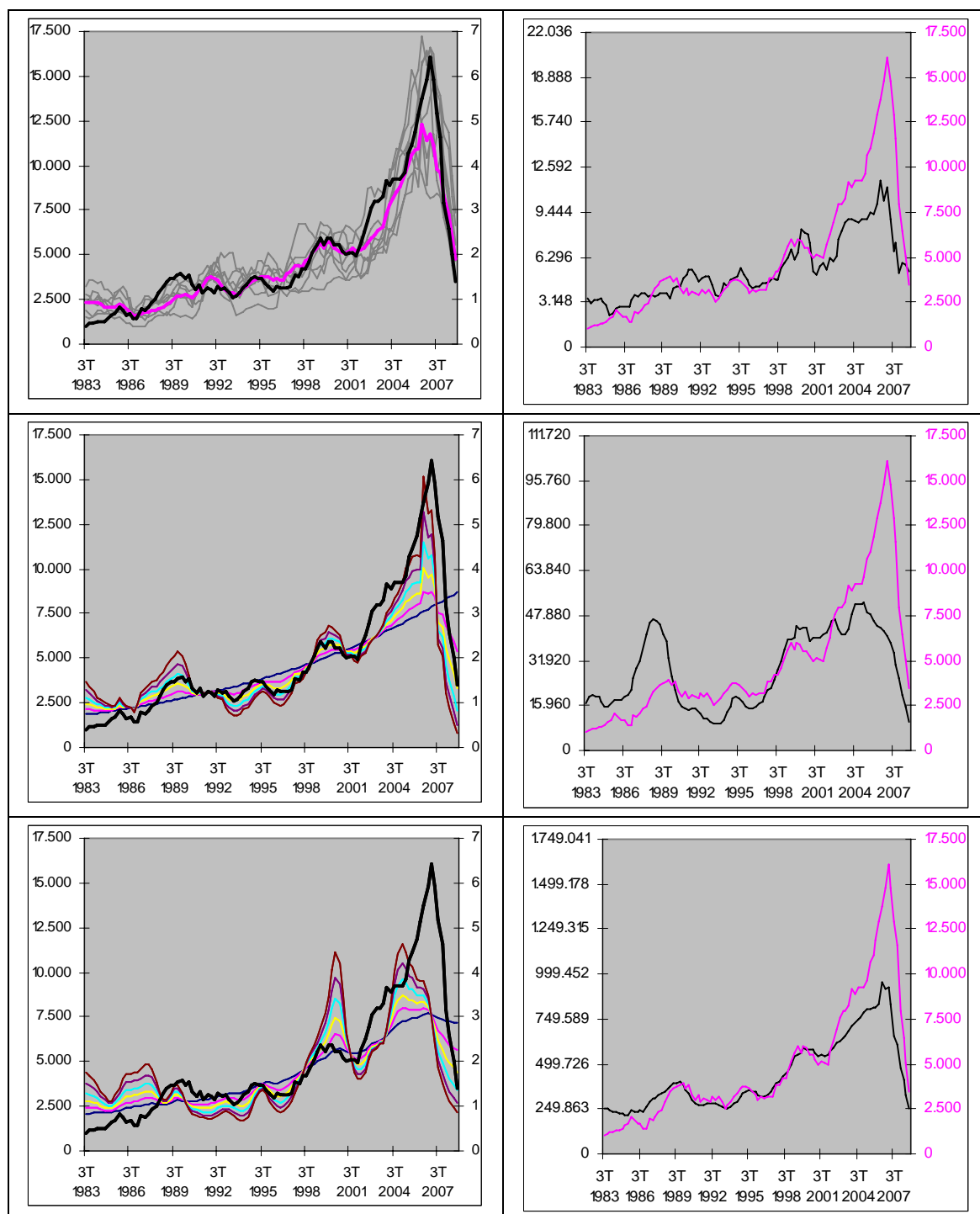


Figura III.45 Lleida

Paneles izquierdos (Lleida resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Lleida: Acumulado anual/valores homogeneizados; **Paneles derechos** (Lleida resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Lleida (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

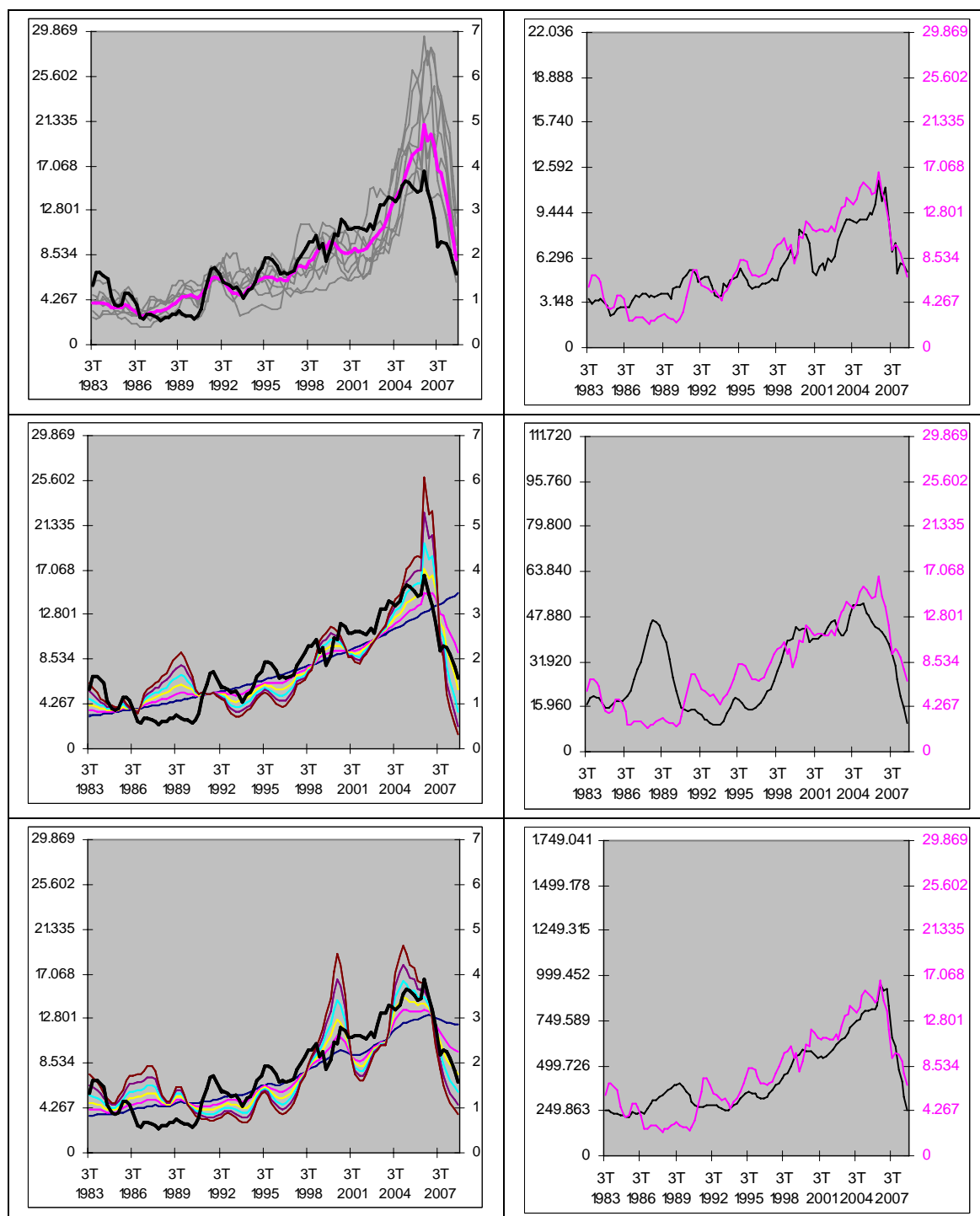


Figura III.46 Pontevedra

Paneles izquierdos (Pontevedra resaltada en negrita): *Panel Superior:* Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central:* Junto con los centroides iniciales; *Panel Inferior:* Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho:* Unidades visadas en Pontevedra: Acumulado anual/valores homogeneizados; **Paneles derechos** (Pontevedra resaltada en rojo): *Paneles Superior/Central/Inferior:* Junto con Córdoba/Alicante/España; *Eje de escala derecho:* Unidades visadas en Pontevedra (en rojo); *Eje de escala izquierdo:* Unidades visadas en Córdoba/Alicante/España

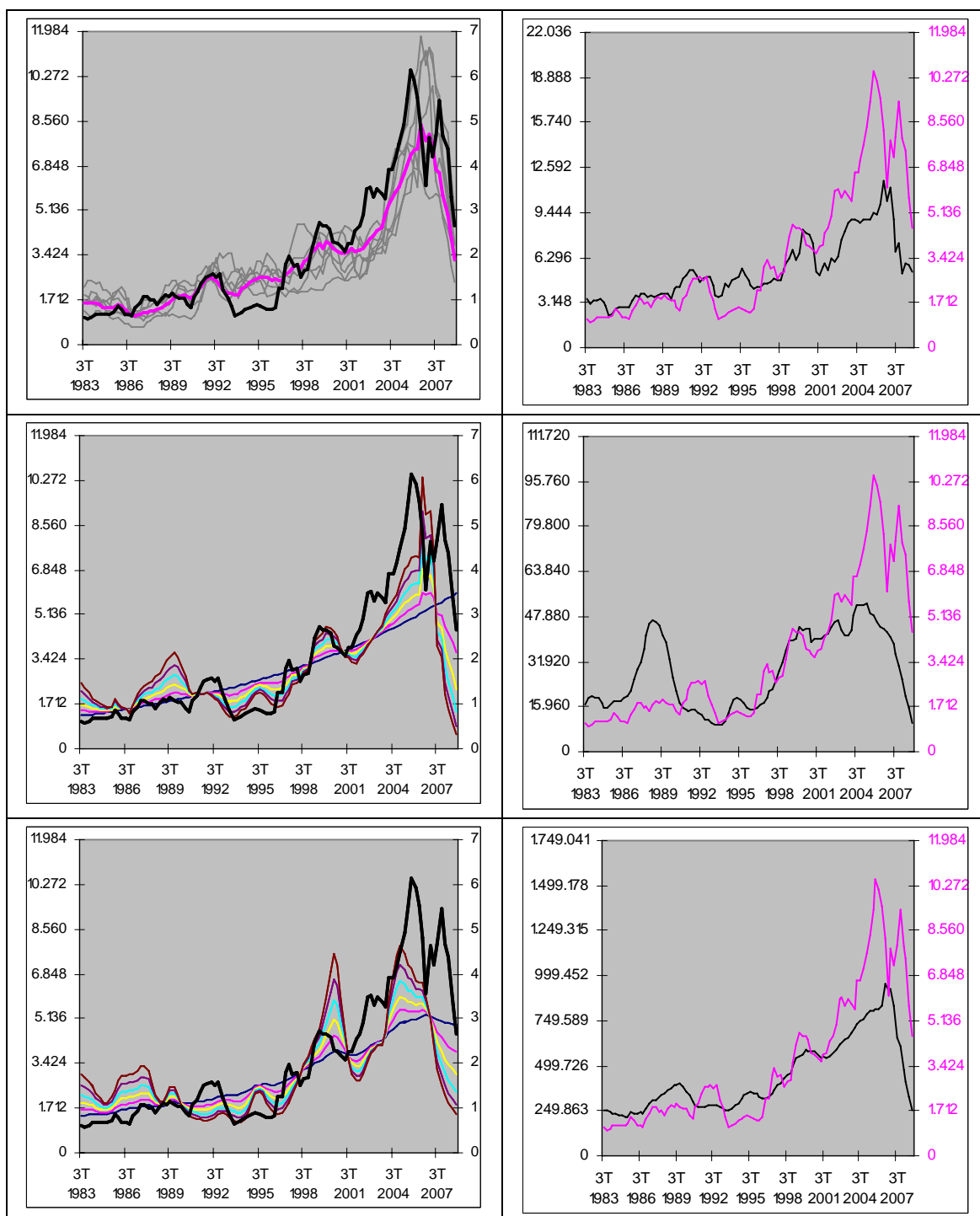


Figura III.48 Guadalajara

Paneles izquierdos (Guadalajara resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Guadalajara: Acumulado anual/valores homogeneizados; **Paneles derechos** (Guadalajara resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Guadalajara (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

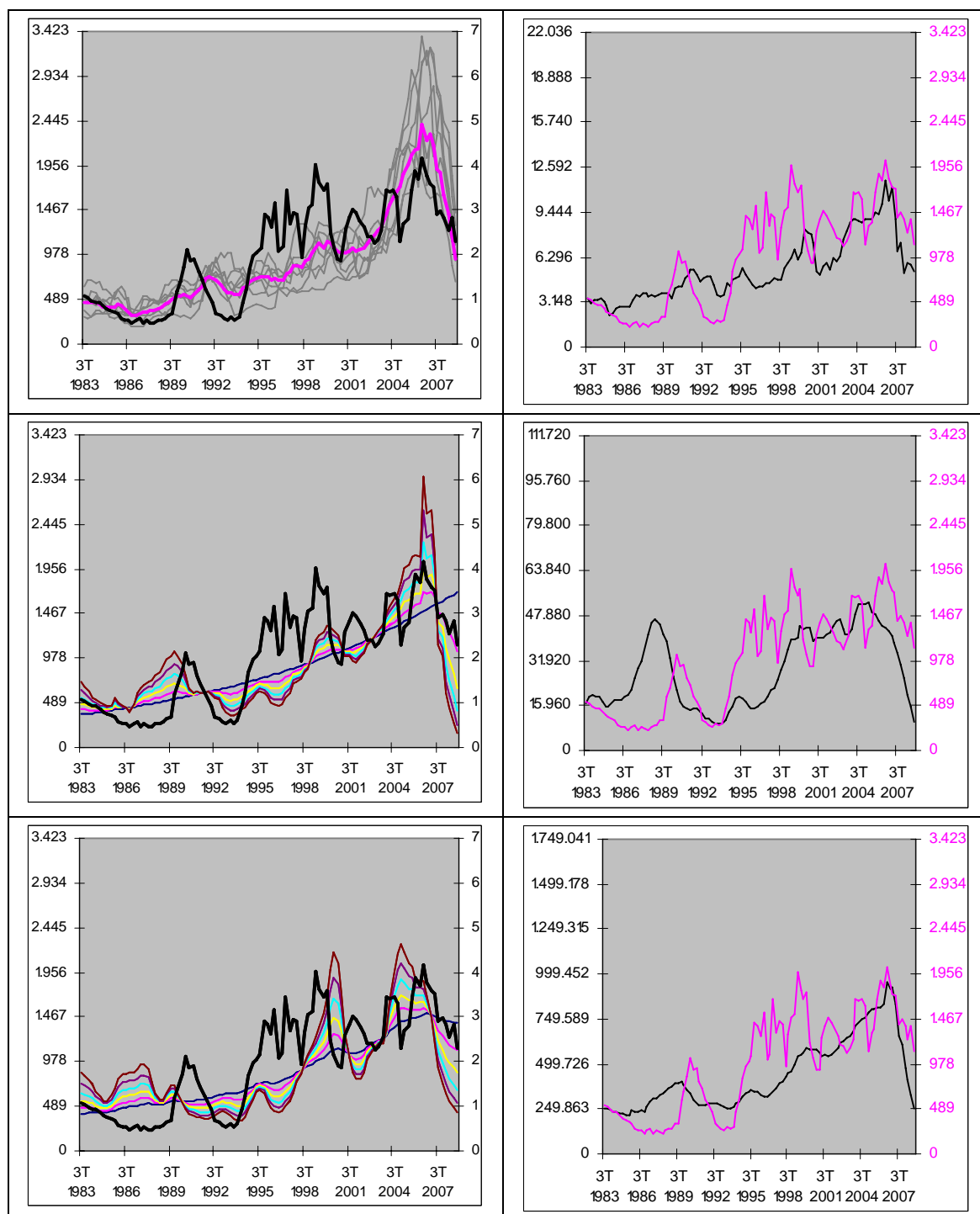


Figura III.49 Soria

Paneles izquierdos (Soria resaltada en negrita): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Soria: Acumulado anual/valores homogeneizados; **Paneles derechos** (Soria resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Soria (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

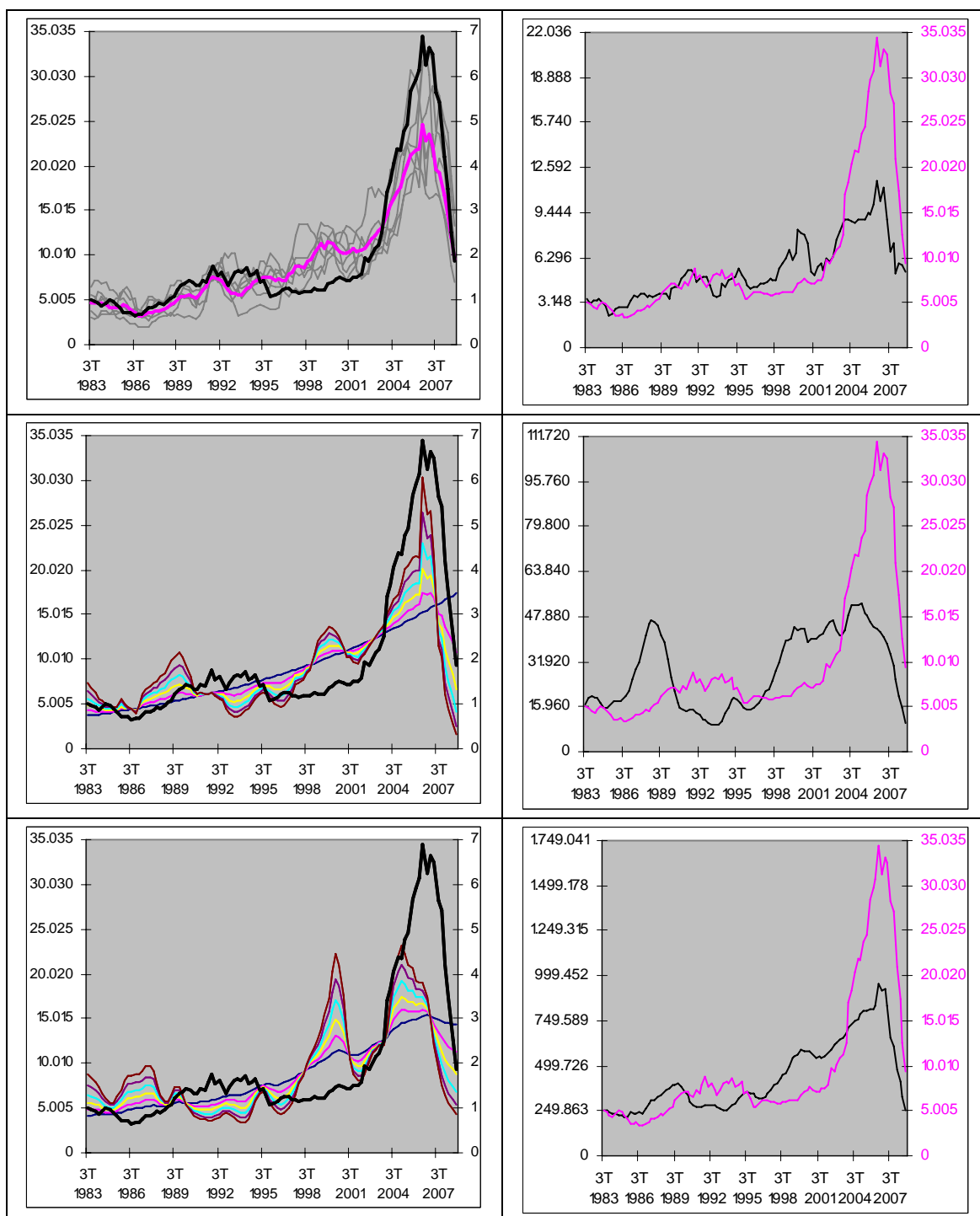


Figura III.47 Toledo

Paneles izquierdos (Toledo resaltada en **negrita**): *Panel Superior*: Junto con las restantes provincias de su grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Toledo: Acumulado anual/valores homogeneizados; **Paneles derechos** (Toledo resaltada en **rojo**): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Toledo (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España

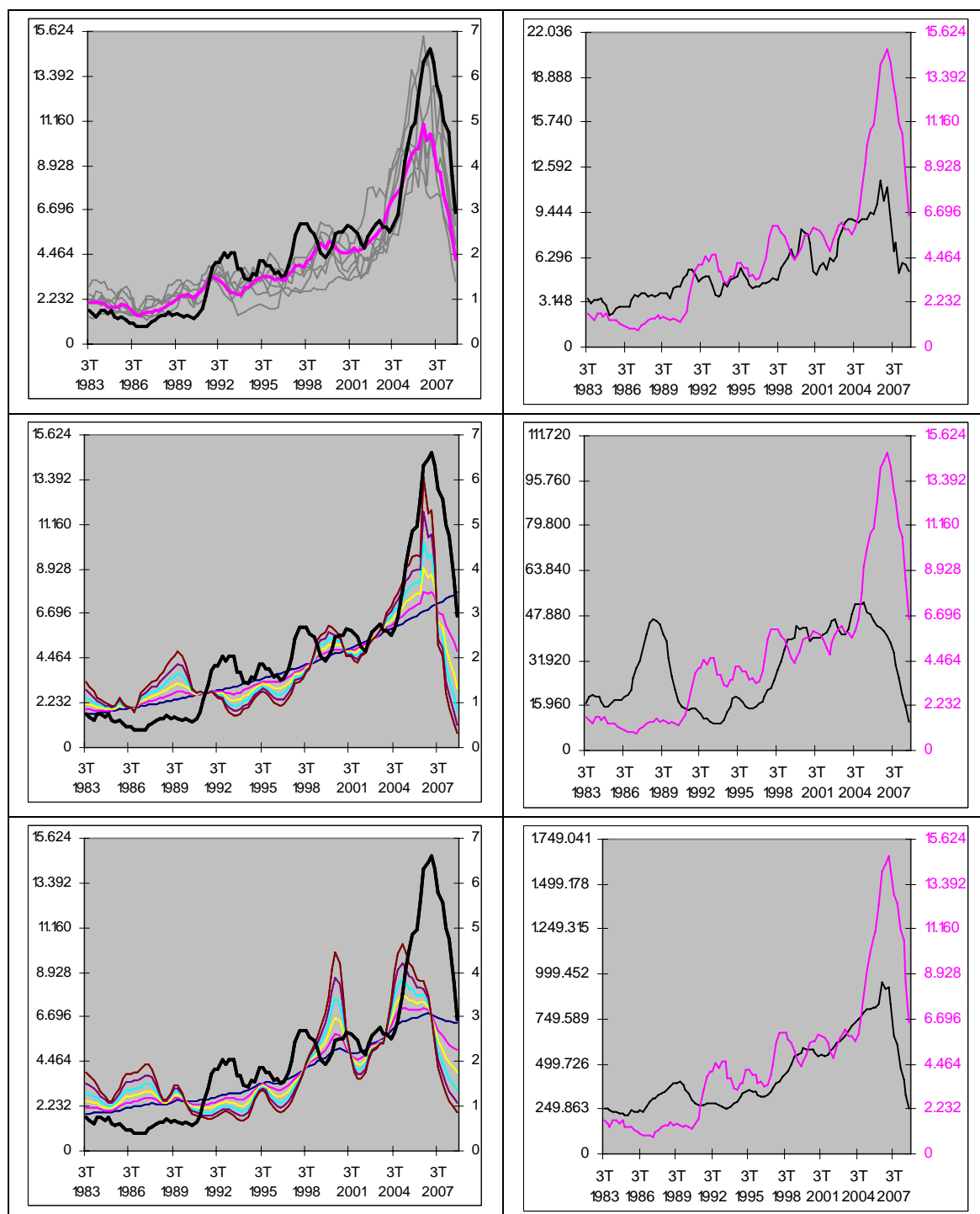


Figura III.50 Ciudad Real

Paneles izquierdos (Ciudad Real resaltada en negrita): *Panel Superior*: Junto con las provincias del segundo grupo y el centroide final correspondiente; *Panel Central*: Junto con los centroides iniciales; *Panel Inferior*: Junto con las curvas resumen extraídas en términos de la evolución del tipo de interés; *Eje de escala izquierdo/derecho*: Unidades visadas en Ciudad Real: Acumulado anual/valores homogeneizados;

Paneles derechos (Ciudad Real resaltada en rojo): *Paneles Superior/Central/Inferior*: Junto con Córdoba/Alicante/España; *Eje de escala derecho*: Unidades visadas en Ciudad Real (en rojo); *Eje de escala izquierdo*: Unidades visadas en Córdoba/Alicante/España