

Una propuesta metodológica para el análisis gráfico de series temporales regionales: una aplicación a las tasas de paro provinciales en España

Magdalena Ferrán Aranz*, Lorenzo Escot**

Recepción: 30 diciembre 2016

Aceptación: 26 febrero 2019

RESUMEN:

En este trabajo proponemos una metodología para el análisis longitudinal comparado de series temporales regionales que combina la metodología del haz de rectas, propuesta por Ferrán (2011), y el análisis de escalamiento multidimensional. El interés de esta metodología radica en que permite visualizar las similitudes y diferencias entre las dinámicas de cada una de las regiones. La presentación de esta metodología se ilustra mediante una aplicación al estudio provincial de las tasas de paro en España a lo largo del periodo 1991-2018. Los resultados del análisis confirman la relevancia de los componentes espaciales en la evolución de la elasticidad del desempleo a lo largo del ciclo económico.

PALABRAS CLAVE: tasa de paro regional; comparación de series temporales geográficas; metodología del haz de rectas; escalamiento multidimensional.

CLASIFICACIÓN JEL: R23; C23; J64.

A methodological approach for regional time series' graphical analysis: an application to unemployment rates in Spanish provinces

ABSTRACT:

This paper presents a methodology for longitudinal comparative analysis of regional time series. This methodology is integrated by the sheaf of straight lines methodology, proposed by Ferrán (2011), and the multidimensional scaling analysis. The interest of this methodology is that it visualizes the similarities and differences between the dynamics of each of the regions. We present this methodology through an application to the provincial study of unemployment rates in Spain over the period 1991-2018. The results of the analysis confirm the relevance of spatial components in the evolution of the unemployment elasticity over the economic cycle.

KEY WORDS: regional unemployment rate; geographic time series comparison; sheaf of straight lines methodology; multidimensional scaling

JEL CLASSIFICATION: R23; C23; J64.

* Dpto. de Estadística y Ciencia de los Datos. Facultad de Estudios Estadísticos. Universidad Complutense de Madrid

** Dpto. de Economía Aplicada, Pública y Política. Facultad de Estudios Estadísticos. Universidad Complutense de Madrid. Av. Puerta de Hierro n.1 - 28040 Madrid.

Autor responsable de la correspondencia: escot@ucm.es.

1. INTRODUCCIÓN

La disponibilidad de series temporales largas de ámbito regional permite el análisis longitudinal comparado del comportamiento diferencial de cada una de las áreas geográficas que integran un territorio. En este trabajo presentamos una metodología para la visualización y el análisis exploratorio de este tipo de series temporales regionales que combina dos técnicas de análisis estadístico. En primer lugar, la denominada metodología del *haz de rectas* (Ferrán 2011 y 2013) que permite, por un lado, comparar gráficamente el comportamiento dinámico de un conjunto de series que presentan diferentes elasticidades respecto a un agregado común mediante la extracción de series auxiliares resumen del conjunto y, por otro lado, visualizar la dinámica de cada serie respecto de su propio pasado. En segundo lugar, utilizaremos el análisis de *escalamiento multidimensional* (Mead, 1992; Borg y Groenen, 2005) aplicado sobre la matriz de distancias del conjunto de series formado por las series temporales objeto de análisis y las generadas mediante la metodología del *haz de rectas*, lo que facilitará la interpretación, mediante los denominados mapas perceptuales, de los gráficos proporcionados por la metodología del *haz de rectas*.

Esta propuesta metodológica es especialmente interesante para el análisis de la economía regional y urbana cuando se aplica para el análisis exploratorio del comportamiento dinámico de cada región respecto a una serie de ámbito suprarregional, que actúa como serie agregada o promedio común. En este sentido la metodología propuesta permite visualizar las diferentes elasticidades de cada región respecto a la serie promedio suprarregional y, lo que es más importante, cómo han ido cambiando dichas elasticidades a lo largo del tiempo, proporcionando a su vez un método gráfico para la representación de regiones más cercanas y más lejanas entre sí en cuanto a este comportamiento dinámico.

La presentación de esta metodología se lleva a cabo mediante una aplicación sobre el estudio provincial de las tasas de paro en España durante el periodo comprendido entre 1991 y finales de 2018. La detección de semejanzas en los patrones dinámicos de la tasa de paro mediante la metodología propuesta en este artículo constituye en sí un objetivo de suficiente entidad e importancia para el análisis regional ya que, por un lado, permite aislar el comportamiento dinámico de cada provincia del efecto de escala que supone analizar regiones no siempre homogéneas en cuanto a su tamaño, y por otro, permite poner de manifiesto la importancia de los componentes espaciales en la evolución del desempleo¹.

Existen en la literatura diferentes enfoques o metodologías para el análisis de las diferencias y similitudes regionales del mercado laboral: desde los trabajos que aplican técnicas de econometría espacial para hacer agrupaciones geográficas no necesariamente coincidentes con las administrativas (Alonso, Del Río y Toharia, 2009; López-Bazo et al. 2002 y 2005; Viñuela, Rubiera y Fernández 2014), pasando por el análisis de cointegración y cambio estructural (Suriñach et al., 1995; Fernández Macho, 1997; y García, Romero y Usabiaga, 2015), o los que aplican diferentes técnicas de clustering para la agrupación regional (López-Bazo et al., 2005; Bande et al., 2008; Bande y Karanassou, 2009, 2013 y 2014; López-Bazo y Motellón, 2013; y Sala y Trivín, 2014). A diferencia de estos enfoques, en este trabajo queremos realizar un análisis exploratorio para interpretar y detectar diferencias y similitudes en el comportamiento dinámico a lo largo de las diferentes fases del ciclo, tomando como referencia la representación gráfica de las correspondientes series temporales de las tasas de paro provinciales a lo largo del periodo 1991-2018.

La forma estándar de representar una serie temporal es a través de un gráfico de líneas². La creciente capacidad de computación actual ofrece magníficas herramientas para el diseño de nuevas técnicas de representación y el desarrollo de software para la visualización de datos espacio-temporales (Andrienko et al. 2003).

¹ La propuesta que realizamos en este trabajo no es tanto la de profundizar en las causas de la disparidad regional de las tasas de paro, objetivo que queda fuera del alcance del presente artículo, sino la de avanzar en el diagnóstico de los diferentes comportamientos regionales del desempleo a lo largo del ciclo económico. Para un análisis de la importancia de la región y localización de los individuos a la hora de explicar las diferencias sistemáticas en las tasas de paro en España véase, entre otros, Viñuela, Rubiera y Cueto (2012) y Cueto, Mayor y Suarez (2017).

² La representación gráfica de series temporales se remonta a finales del siglo XVIII y se considera que el pionero, no sólo de los gráficos de líneas, sino también de los gráficos de barras y de sectores fue William Playfair (Tufte, 2001). Desde entonces los gráficos estadísticos han desempeñado un papel fundamental en el análisis exploratorio de los datos, en la formulación de hipótesis o en el desarrollo de

A pesar de ello sorprende la escasez de trabajos relativos a la representación de conjuntos masivos de series temporales (Lin et al., 2005). Los tres trabajos más referenciados en este campo posiblemente sean los de Hochheiser and Shneiderman (2001), van Wijk y van Selow (1999) y Weber et al. (2001). La duda que surge es la de si en estas tres propuestas y, en términos generales, en las del campo de la visualización, la estructura del gráfico responde más a un objetivo estético que funcional. Para abrir un debate al respecto, la revista *Statistical Computing & Graphics Newsletter* ofrece un apartado monográfico (vol 22 –nº1, 2011) que aborda estos dos grandes enfoques existentes en este ámbito. Por un lado, Kosara (2011) ofrece el punto de vista del experto en informática que construye herramientas y técnicas para la visualización de datos que, además de una atractiva presentación, tengan una gran capacidad interactiva. Por otro lado, Gelman y Unwin (2011) ofrecen el punto de vista del experto en estadística, que utiliza los gráficos como herramienta de ayuda en la toma de decisiones en cualquier punto del proceso de análisis estadístico de los datos.

El interés de este trabajo se centra en la comparación gráfica de series temporales bajo este segundo punto de vista. La simple inspección visual del gráfico del conjunto de series temporales regionales difícilmente puede ayudar en la detección de patrones de comportamiento en la trayectoria temporal de los datos, más allá de los grandes ciclos comunes al conjunto de regiones. Cuando se intenta analizar un conjunto de múltiples series temporales (por ejemplo, una serie temporal para cada región o zona geográfica) surgen una serie de problemas para su adecuada comparación. Tanto si se representa cada serie regional por separado como si se representan todas ellas conjuntamente mediante su superposición en un único gráfico (bien en la escala original de los datos o bien con los datos transformados a una escala común) la comparación se hace muy complicada en cuanto el número de series es mínimamente elevado.

Es por ello que, para el análisis de un conjunto de múltiples series temporales, se hayan propuesto diferentes técnicas de comparación o agrupación en lo que ha venido a denominarse análisis clúster de series temporales (Liao 2005). Estas técnicas tratan de clasificar las series temporales en grupos en función de la similitud o la distancia entre ellas, de modo que las series clasificadas dentro de un mismo grupo sean similares entre sí y diferentes a las del resto de grupos. Estas técnicas requieren en un primer paso elaborar una métrica apropiada de distancia/similitud entre series (la distancia Euclídea, de Manhattan, de Hamming, la norma máxima, etc.) para, posteriormente, utilizar las técnicas de agrupamiento habituales, tales como las *k*-medias, la agrupación jerárquica u otras (Zhao, 2015).

En los dos siguientes apartados se describe la propuesta metodológica que combina la metodología del haz de rectas y el análisis de escalamiento multidimensional para la visualización de las similitudes y diferencias en el comportamiento dinámico de un conjunto formado por un número elevado de series temporales. La metodología se ilustra con el conjunto formado por las series temporales de tasas de paro en las provincias españolas durante el periodo 1991-2018. En el cuarto apartado retomamos las principales conclusiones del análisis elaborado.

2. APLICACIÓN DEL HAZ DE RECTAS A LAS TASAS DE PARO PROVINCIALES

Los datos utilizados para ilustrar nuestra propuesta metodológica son las series trimestrales de los 28 años comprendidos entre 1991 y el tercer trimestre de 2018 de las tasas de paro provinciales de la *Encuesta de Población Activa* publicados por el INE. Utilizamos las series suavizadas mediante el promedio del valor observado en el trimestre correspondiente y los tres inmediatamente anteriores.

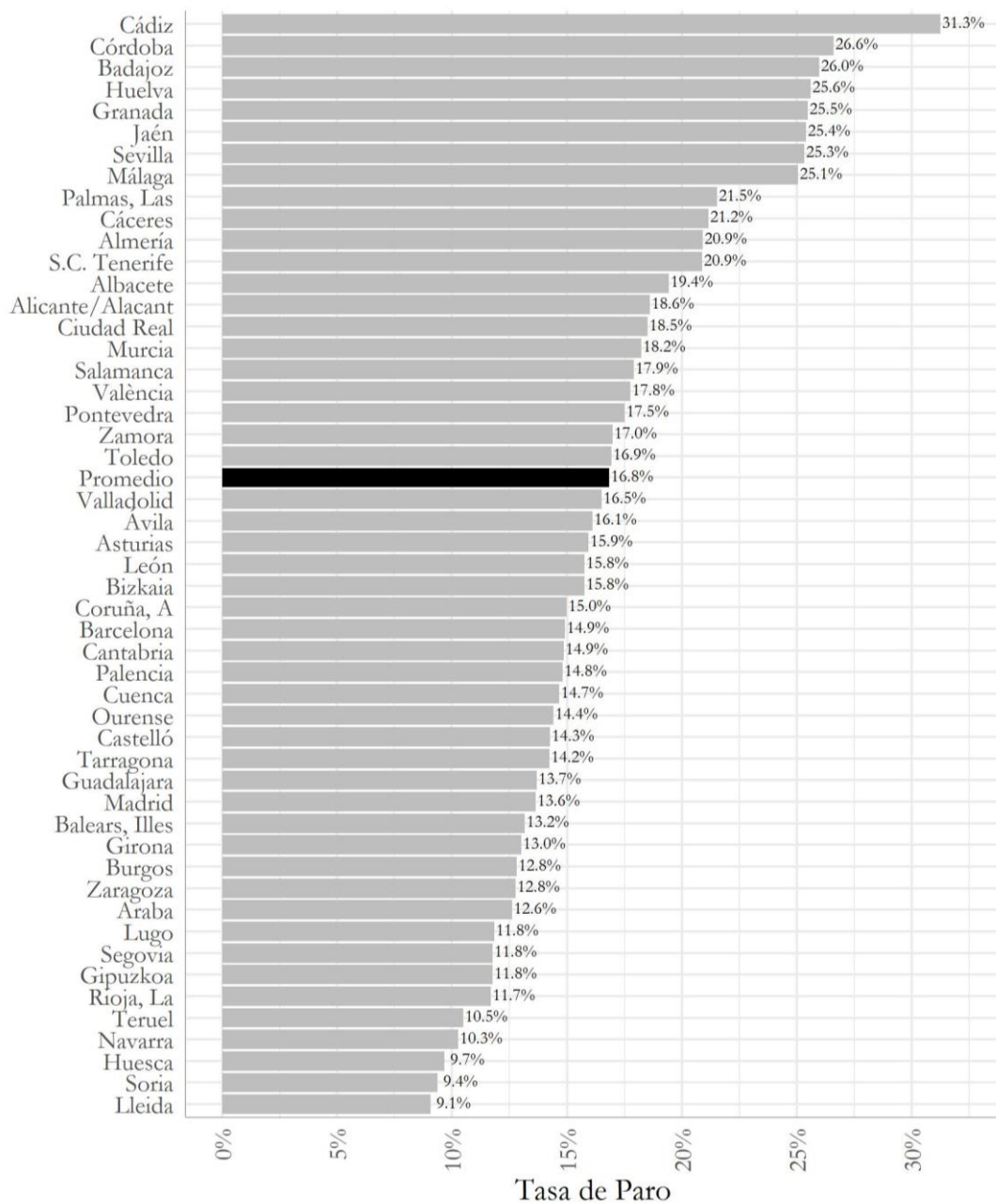
La primera fase de la propuesta metodológica consistirá en aplicar la metodología del *haz de rectas* (Ferrán, 2011 y 2013) a las series de desempleo provinciales. Dado un conjunto de series temporales sobre un mismo fenómeno, pero procedentes de diferentes ámbitos, territorios, agentes, condiciones, etc., la metodología del *haz de rectas* consiste en construir un conjunto de series, procedentes de un haz de rectas, que resuma gráficamente el comportamiento dinámico de las series objeto de análisis. Los fundamentos teóricos de la metodología se describen en el primer apartado del **Anexo I**, y el proceso de construcción de las series *resumen*,

modelos, pero la mayor parte de las propuestas se ha centrado en la mejora de los gráficos convencionales; en el caso particular de la representación de series temporales, en la mejora de los gráficos de líneas o de barras (Zhao et al., 2011).

en el segundo. El resultado final de esta metodología gráfica consiste en la representación de cada serie sobre el gráfico de líneas correspondiente al conjunto de series *resumen*, que facilitará la interpretación de las similitudes y diferencias entre las distintas series a lo largo del periodo analizado.

Con la aplicación de la metodología del *baz de rectas* sobre las series correspondientes a las tasas de paro provinciales se trata no tanto de detectar qué provincias son similares en términos del nivel de tasa de paro, sino de compararlas según su patrón de evolución a lo largo del ciclo económico. Así será posible determinar, por ejemplo, qué provincias son las que han sufrido de una manera relativamente más intensa el problema del paro durante la última crisis económica y cuáles están teniendo una mayor capacidad de recuperación.

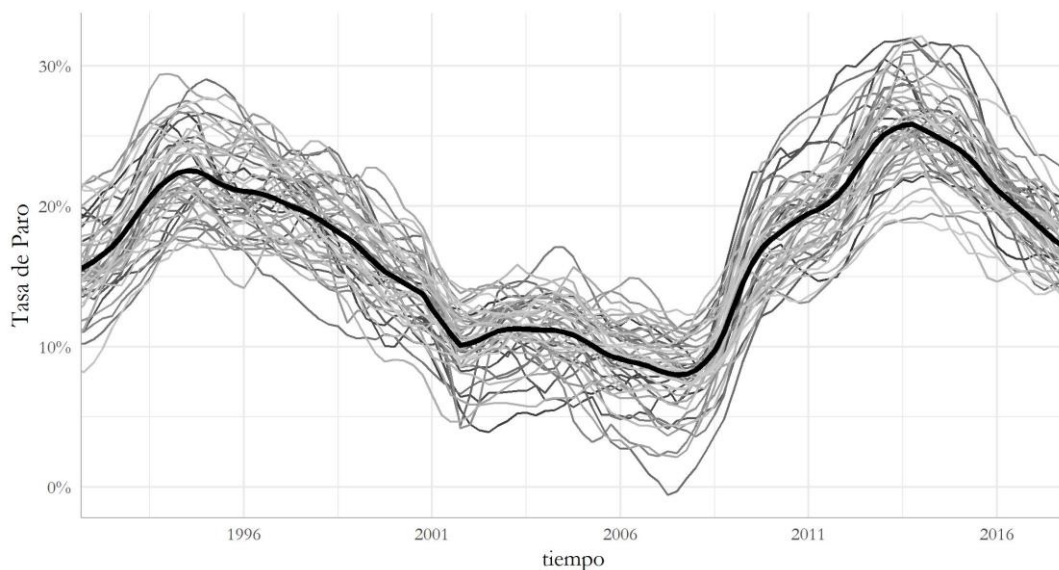
FIGURA 1.
Tasa de paro provincial media en el periodo comprendido entre 1991 y 2018



Fuente: Elaboración propia.

Las tasas de paro medias de las diferentes provincias distan mucho de ser iguales (**Figura 1**). Como el interés de esta metodología no se encuentra en los efectos escala, es decir, en las diferencias en los niveles de desempleo, la aplicación la metodología requiere que las series estén homogenizadas de tal forma que compartan la misma media. Para ello, en primer lugar, hay que construir la serie promedio del conjunto de provincias y reescalar cada una de las series de tal forma que sus medias sean iguales a la de la serie promedio (**Figura 2**). Al homogenizar las series provinciales, se elimina de la comparación entre provincias el efecto de la escala y, por consiguiente, el análisis comparativo regional se circunscribe a la amplitud de las oscilaciones de cada serie respecto a la serie promedio.

FIGURA 2.
Evolución del conjunto de series correspondiente a las tasas de paro provinciales 1991T1-2018T3, suavizadas con una media móvil de orden 4 no centrada y reescaladas en media, junto con la serie promedio (en negrita)



Fuente: Elaboración propia.

La trayectoria de esta serie promedio de desempleo recoge la evolución agregada del conjunto de provincias. En los últimos 28 años el desempleo, en media (serie resaltada en la Figura 2), se ha caracterizado por su evolución anticíclica: la fase inicial, desde 1991, se caracteriza por un crecimiento de las tasas de paro hasta el primer máximo en 1994; a continuación, se produce una fase de decrecimiento, hasta alcanzar los mínimos en 2007; a partir de 2008, con el inicio de la crisis económica, las tasas de paro inician una nueva fase de crecimiento, con un máximo en 2013; finalmente, a partir de 2014, con los primeros indicios de recuperación económica, comienza una nueva fase de reducción media del desempleo. El estudio de en qué medida las diferentes provincias replican con mayor o menor intensidad a lo largo de las diferentes fases del ciclo este comportamiento promedio, permite el análisis comparativo del comportamiento regional del desempleo.

De esta forma, una vez homogenizadas las J series observadas, el siguiente paso de la metodología del *haz de rectas* consiste en la construcción del conjunto de K series *resumen*, $\{C_{t,k}\}$. Con este fin se procede a la estimación de la sensibilidad de cada serie provincial respecto a la serie promedio mediante el ajuste de J ecuaciones de regresión lineal:

$$\hat{Z}_{t,j} = A_{0,j} \cdot t + A_{1,j} \cdot X_t + A_{2,j}$$

donde X_t es la serie promedio del conjunto de series observadas y $Z_{t,j}$ es la serie resultante de reescalar la serie observada (**Tabla 1**). La obtención de las series *resumen* requiere la estimación del vértice $(-B_1, B_0)$ del haz

TABLA 1.

Extracción del *baz de rectas*. Estimación por mínimos cuadrados de la ecuación [4], correspondiente a cada una de las provincias, y de la ecuación [5] (Anexo I) ; tasas medias provinciales de paro, α_j , y tasa promedio, α

PROVINCIA	R_j^2	$A_{0,j}$	$A_{1,j}$	$A_{2,j}$	Tasa de paro Media α_j
Albacete	0,923	-0,0029	1,415	-6,829	19,435
Alicante	0,928	0,0004	1,142	-2,413	18,621
Almería	0,920	0,0824	1,382	-11,058	20,919
Álava	0,881	-0,0589	0,761	7,322	12,621
Asturias	0,924	-0,0327	0,898	3,552	15,938
Ávila	0,948	0,0523	1,073	-4,157	16,116
Badajoz	0,912	-0,0555	1,253	-1,150	25,975
Balears (Illes)	0,829	0,0433	0,859	-0,048	13,158
Barcelona	0,940	-0,0225	0,999	1,283	14,909
Vizcaya	0,945	-0,1142	0,834	9,191	15,765
Burgos	0,933	-0,0152	0,782	4,518	12,819
Cáceres	0,802	0,0021	0,935	0,982	21,578
Cádiz	0,919	-0,0747	1,490	-4,074	31,262
Cantabria	0,932	-0,0640	0,913	5,055	14,858
Castellón	0,880	0,1003	1,139	-7,963	14,267
Ciudad Real	0,968	0,1000	1,161	-8,309	18,519
Córdoba	0,894	-0,0197	1,160	-1,590	26,618
Coruña (A)	0,856	-0,0300	0,651	7,554	14,986
Cuenca	0,943	0,0567	1,037	-3,793	14,650
Guipúzcoa	0,928	-0,1034	0,803	9,101	11,767
Girona	0,905	0,0824	0,787	-1,028	13,007
Granada	0,936	-0,0084	1,390	-6,110	25,499
Guadalajara	0,958	0,0304	0,915	-0,269	13,684
Huelva	0,914	-0,0391	1,267	-2,305	25,614
Huesca	0,932	0,0242	0,691	3,857	9,656
Jaén	0,928	0,0008	1,354	-6,007	25,418
León	0,935	-0,0030	0,895	1,934	15,769
Lleida	0,923	0,0683	0,580	3,253	9,055
Lugo	0,821	0,0032	0,640	5,882	11,844
Madrid	0,937	-0,0206	0,918	2,534	13,636
Málaga	0,968	-0,0668	1,582	-6,063	25,050
Murcia	0,966	-0,0015	1,247	-4,080	18,234
Navarra	0,943	0,0048	0,712	4,586	10,254
Orense	0,696	0,0341	0,754	2,237	14,421
Palencia	0,932	-0,0301	0,984	1,960	14,818
Palmas (Las)	0,819	0,0496	1,353	-8,722	21,531
Pontevedra	0,936	-0,0042	0,889	2,104	17,512
Rioja (La)	0,942	0,0156	0,853	1,607	11,695
Salamanca	0,900	-0,0946	0,844	7,930	17,911
S. C. de Tenerife	0,912	0,0193	1,288	-5,928	20,878
Segovia	0,845	0,0218	0,567	6,068	11,766
Sevilla	0,910	-0,0458	1,217	-1,088	25,333
Soria	0,944	0,0365	0,644	3,957	9,372
Tarragona	0,962	0,0562	0,984	-2,880	14,248
Teruel	0,954	0,0281	0,869	0,633	10,488
Toledo	0,979	0,1104	1,085	-7,618	16,950
Valencia	0,963	0,0021	1,178	-3,109	17,765
Valladolid	0,964	-0,1049	0,828	8,774	16,520
Zamora	0,916	-0,0169	0,982	1,245	16,975
Zaragoza	0,958	0,0041	1,018	-0,527	12,770
			Promedio (α)		16,85
Eje del <i>baz de rectas</i> estimado	$B_0=$,0025	$B_1=$	-,0025	

Fuente: Elaboración propia.

de rectas que subyace en las J series provinciales respecto a la serie promedio. Dicho vértice se obtiene a partir del ajuste de la regresión lineal sobre los J puntos $(A_{1,j}, A_{0,j})$:

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

A partir de los coeficientes estimados B_0 y B_1 y eligiendo K valores representativos dentro del rango de variación de las diferentes sensibilidades o elasticidades respecto a la serie promedio medidas por los valores, $A_{1,j}$, $j = 1, \dots, J$, concretamente eligiendo K valores m_k :

$$m_k = m_{k-1} + \theta, \text{ con } \theta = (\max_j A_{1,j} - \min_j A_{1,j}) / (K - 1) \text{ y } m_1 = \min_j A_{1,j}$$

las expresiones de las K series resumen vendrán dadas por:

$$C_{t,k} = b_k \cdot t + m_k \cdot X_t + \mu_k \quad k = 1, \dots, K$$

donde $\mu_k = \alpha - \bar{g}_k$, siendo $g_{t,k} = b_k \cdot t + m_k \cdot X_t$ y $b_k = B_0 + B_1 \cdot m_k$

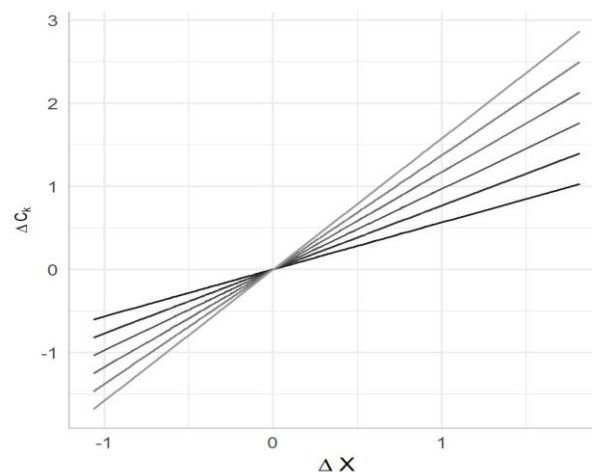
El conjunto de las K series resumen $\{C_{t,k}\}$ es tal que el correspondiente conjunto de series $\{c_{t,k} = \Delta C_{t,k}\}$ verifica la siguiente relación respecto de la serie $x_t = \Delta X_t$:

$$c_{t,k} = b_k + m_k \cdot x_t \quad \forall t, k \quad \text{donde} \quad b_k = B_0 + B_1 \cdot m_k \quad \forall k$$

La relación lineal entre los coeficientes b_k y m_k implica que las K rectas $c_{t,k} = b_k + m_k \cdot x_t$ se cortan en un mismo punto, $(x_t, c_{t,k}) = (-B_1, B_0)$; dicho de otra manera, que las K rectas forman un *haz de rectas* de vértice $(-B_1, B_0)$ tal que la primera de ellas, $c_{t,1}$, la correspondiente al mínimo valor del coeficiente m_k , es la menos elástica respecto de la serie promedio diferenciada, y la última, $C_{t,k}$, la correspondiente al máximo valor, es la más elástica **Figura 3**.

FIGURA 3.

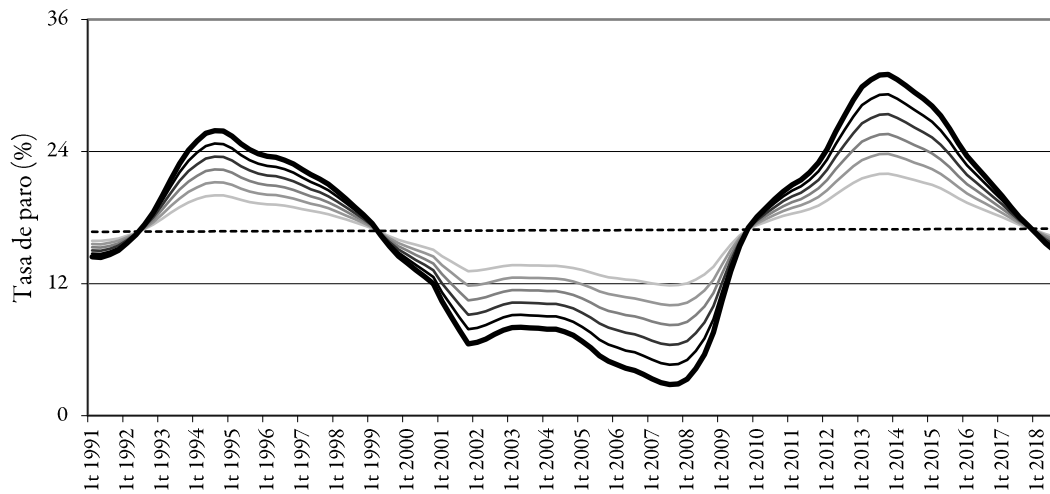
Haz de rectas subyacente en las series resumen de las tasas de paro respecto a la tasa de paro promedio



Fuente: Elaboración propia.

Cuando en el conjunto de las series resumen subyace una estructura de haz de rectas, entonces dichas series replican la misma forma y presentan las mismas fases cíclicas que la serie promedio (Figura 4), pero con una intensidad o amplitud de sus fluctuaciones que se van pronunciando desde la primera (representada mediante la línea más clara), la más rígida, a la última (representada mediante la línea más oscura), la más elástica. Este patrón de curvas es el que permitirá la comparación de las trayectorias de las series observadas. Obsérvese que la recta que pasa por los puntos de corte de las series resumen $\{C_{t,k}\}$ es la tendencia lineal a largo plazo, común a las K series (recta de trazo discontinuo en la Figura 4).

FIGURA 4.
Trayectorias de las series de los conjuntos resumen $\{C_{t,k}\}$



Fuente: Elaboración propia.

En cuanto a la elección de K , el número de series *resumen* a utilizar en el análisis, debe ser tal que permita representar un rango suficientemente amplio de comportamientos más o menos elásticos como para poder hacer comparaciones con las series reales. Sin pérdida de generalidad, en este trabajo hemos extraído seis series *resumen*, $K = 6$, asumiendo diferentes sensibilidades o distancias respecto a la serie promedio, esto es, seis diferentes valores de los coeficientes m_k .

Un resultado de interés de la metodología del *haz de rectas* es el que hace referencia a los puntos de corte de las diferentes series *resumen* $\{C_{t,k}\}$ ³: siempre existirá al menos un punto de corte entre dos series *resumen* cualesquiera, y, además, los puntos de corte de cualquier par de series lo son de todas ellas. Dichos puntos de corte permiten identificar los cambios de ciclo económico, esto es, son los puntos a partir de los que se entra en fases expansivas (se pasa de tasas de paro que superan la media global del periodo, a tasas inferiores) o, en su caso, en fases recesivas (se pasa de tasas inferiores a la media global de todo el periodo a tasas superiores).

En definitiva, el conjunto de series $\{C_{t,k}\}$ permite resumir el comportamiento provincial de las tasas de paro según distintos grados de sensibilidad respecto a la serie promedio y delimita, a partir de los puntos de corte, las distintas fases de evolución a lo largo de todo el periodo considerado.

Así, considerando que el conjunto de series *resumen* presenta cuatro puntos de corte (a mediados de 1992, a principios de 1999, a finales de 2009 y a finales de 2017), la metodología del *haz de rectas* ha identificado tres etapas bien diferenciadas del comportamiento de las tasas de paro en España: la primera, comprendida entre 1991 y 1998, caracterizada por altas tasas de paro; la segunda, comprendida entre 1999 y 2009, caracterizada por bajas tasas de paro; y la tercera, comprendida entre 2010 y 2017, caracterizada por altas tasas de paro.

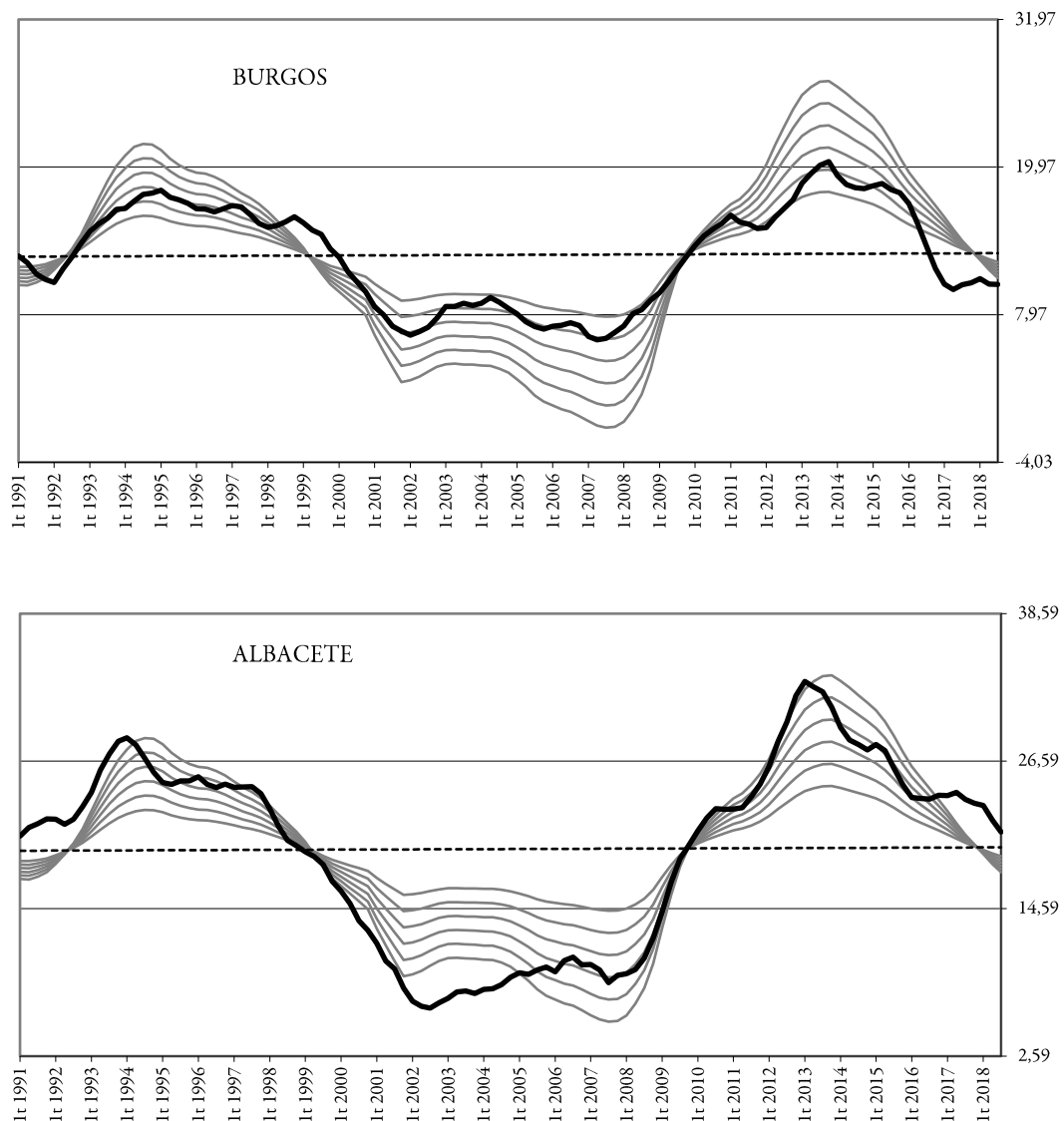
A efectos de interpretar la trayectoria de las tasas de paro en una provincia concreta en comparación con las del resto, la representación gráfica podría hacerse en la escala homogeneizada, esto es, en la escala de la serie promedio; sin embargo, teniendo en cuenta que la diferencia entre los puntos de corte de las series del conjunto $\{C_{t,k}\}$ no depende de la escala de representación⁴, puede reescalararse cada una de las componentes del conjunto de series *resumen* a la escala original de la serie.

³ Apartado B del Resultado fundamental la metodología (Anexo I)

⁴ Apartado C del Resultado fundamental la metodología (Anexo I)

En definitiva, la solución de la metodología del *baz de rectas* consiste en las representaciones de las tasas de paro en cada provincia sobre el correspondiente conjunto de series *resumen* reescaladas. Cada representación permitirá identificar cuándo la evolución de la tasa de paro de la provincia correspondiente se está alejando mucho o poco respecto al comportamiento promedio.

FIGURA 5.
Evolución de las tasas de paro en Albacete y Burgos (series suavizadas) representadas sobre el conjunto de series resumen reescaladas



Fuente: Elaboración propia.

Por ejemplo, al representar (panel superior de la **Figura 5**) la trayectoria de la tasa de paro en Albacete sobre el correspondiente conjunto de series *resumen* (las seis series $C_{t,k}^j$ se representan todas con el mismo formato, siendo la primera, $k = 1$, la más interior y la sexta, $k = 6$, la más exterior), se observa que se corta con ellas prácticamente en los cuatro puntos de corte comunes; además, en los tres periodos delimitados por los puntos de corte, presenta fuertes desviaciones frente a la tendencia estimada (línea discontinua) transcurriendo próxima a la serie *resumen* más exterior, lo que implica que la relación entre las cotas de las tres etapas es muy parecida a la descrita para dicha serie. En otras palabras, cuando la tasa de paro es alta en el conjunto

de las provincias (etapas primera y tercera) la de Albacete es acusadamente alta, y cuando es baja (segunda etapa), acusadamente baja; siempre en términos relativos, es decir, según el nivel de sus propios valores. En el extremo opuesto, la trayectoria de la tasa de paro de Burgos (panel inferior de la **Figura 5**) evoluciona pegada a la *curva resumen* más interior, lo que significa que es moderadamente baja en la etapa expansiva (la segunda) y moderadamente alta en las recesivas (primera y tercera etapas). Comparando a su vez las escalas en las que se representan ambos niveles de desempleo, estas dos figuras permiten evidenciar además que el nivel de desempleo en Albacete es superior al de Burgos.

La construcción del conjunto de series *resumen* tiene sentido sólo si la estructura que subyace en el conjunto de series observadas es la de un *haz de rectas*. Es decir, el correspondiente conjunto de series diferenciadas debe presentar un comportamiento similar al de un haz de rectas: por un lado, todas ellas deben estar relacionadas linealmente con la serie promedio diferenciada y, por otro, los coeficientes de los correspondientes modelos lineales ajustados también deben presentar relación lineal. La estimación de los coeficientes del correspondiente modelo lineal ajustado da lugar a la estimación del vértice del haz de rectas que subyace en las series diferenciadas, y a partir del mismo pueden construirse las series resumen que ayudarán a interpretar la dinámica de cada serie provincial de desempleo.

Más adelante se mostrará la solución relativa a las 50 provincias españolas mediante esta metodología, esto es, la representación de cada serie provincial sobre el conjunto de series *resumen* reescaladas a su nivel (Figuras 11 a 14). Aunque estas series simplifican la interpretación de la evolución de la tasa de paro de cada provincia en función de su diferente elasticidad respecto a la serie promedio, no proporcionan una visión global de la situación de las provincias que permita extraer conclusiones sobre la posible influencia del factor espacial en dichos resultados individuales. Así, para ayudar a la interpretación de la trayectoria de cada serie y establecer y representar espacialmente las similitudes y diferencias entre todas ellas, sería muy conveniente establecer algún criterio de ordenación y de reducción de estas representaciones gráficas. Veamos en lo que sigue cómo aplicar el fundamento teórico que subyace en el proceso de construcción de las series *resumen* para simplificar esta comparación mediante la aplicación del *escalamiento multidimensional*. La solución correspondiente será la que dará lugar a la agrupación y ordenación de las evoluciones provinciales, que como veremos presenta un relevante componente espacial.

3. ANÁLISIS DE LAS TASAS DE PARO PROVINCIALES MEDIANTE EL ESCALAMIENTO MULTIDIMENSIONAL

Una simple comparación de las tasas medias de paro provincial en el periodo comprendido entre 1991 y 2018 permite realizar una clasificación de las provincias en cuanto a la gravedad del problema del paro en España. Este análisis de los niveles de desempleo permite evidenciar que las provincias de Extremadura, Andalucía y Canarias presentan promedios de paro por encima del veinte por ciento de su población activa, mientras que las provincias situadas más al norte de la península presentan promedios inferiores (**Figura 6**). Esta información, aunque interesante, no es suficiente para realizar una comparación de las dinámicas del desempleo entre provincias: el hecho de que dos provincias presenten unas tasas de paro con niveles parecidos en el periodo analizado no implica que su trayectoria haya sido también parecida.

Con el objetivo de, a la hora de presentar los correspondientes cincuenta gráficos, establecer un orden que simplifique la comparación de las trayectorias de las tasas de paro provinciales aplicaremos un análisis de *escalamiento multidimensional* (Mead, 1992; Real y Varela, 2003; o Borg y Groenen, 2005) sobre la matriz de distancias euclídeas entre cada par de series del conjunto de series homogeneizadas, lo que permitirá resumir las similitudes y diferencias entre las dinámicas provinciales y proporcionar así otro tipo de representación espacial de las provincias, en este caso no atendiendo a los niveles de desempleo, sino a las dinámicas o elasticidades de la evolución del desempleo a lo largo del ciclo económico.

FIGURA 6.
Mapa de representación provincial de las tasas medias de paro para el periodo 1991 – 2018



Fuente: Elaboración propia.

Dada una matriz de distancias o similitudes entre objetos, el *escalamiento multidimensional* es una técnica de análisis multivariante cuyo objetivo es proporcionar una representación gráfica de los objetos como puntos en un espacio de pequeña dimensión (en principio bidimensional), de tal forma que las distancias entre ellos reproduzcan con la mayor precisión posible las correspondientes proximidades en la matriz de entrada: si dos objetos son más similares entre sí que respecto de cualquier otro, entonces en la representación gráfica en dichas dimensiones (denominada *mapa perceptual*) la distancia entre ellos debe ser menor que a cualquier otro punto.

En la propuesta metodológica que aquí se presenta se utiliza un resultado novedoso de la metodología del haz de rectas que permite dotar de interpretación a los mapas perceptuales elaborados con el conjunto de series provinciales de desempleo. En efecto, como complemento al resultado fundamental de la metodología del *haz de rectas* (Anexo I), presentamos en este trabajo las siguientes propiedades del conjunto de series *resumen*:

- Si $C_{t,k}$ y $C_{t,k'}$ son dos series temporales cualesquiera del conjunto de series $\{C_{t,k}\}$ tales que $m_k < m_{k'}$, entonces:

$$d(C_{t,k}, C_{t,k'}) = (m_{k'} - m_k) \cdot \varepsilon$$

donde ε es un valor constante positivo independiente de t , de k y de k' .

La demostración de este resultado se ofrece en el Anexo II y, del mismo, se deduce automáticamente el siguiente resultado:

- Si $C_{t,q}$, $C_{t,r}$ y $C_{t,p}$, son tres series temporales cualesquiera del conjunto de series $\{C_{t,k}\}$ tales que $m_q < m_r < m_p$, entonces la distancia euclídea entre las dos más distantes es igual a la suma de las distancias de cada una de ellas a la tercera:

$$d(C_{t,q}, C_{t,r}) + d(C_{t,r}, C_{t,p}) = d(C_{t,q}, C_{t,p})$$

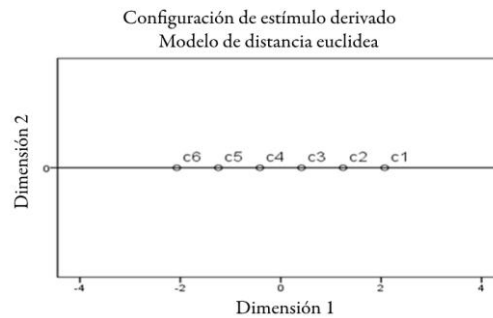
Este resultado implica que las distancias entre las tres series queda resumida en un segmento: el extremo inferior corresponderá a $C_{t,q}$ y el superior a $C_{t,p}$, y entre ellos estará el punto correspondiente a $C_{t,r}$, a la distancia $d(C_{t,q}, C_{t,r})$ del primero y $d(C_{t,r}, C_{t,p})$ del tercero; así la distancia del primero al tercero será igual a la suma de ambas distancias.

Si, en particular, $m_{k+1} = m_k + \theta, \forall k$, entonces $d(C_{t,k+1}, C_{t,k}) = \theta \cdot \varepsilon$ y, en consecuencia, la matriz de distancias entre los K puntos correspondientes a las K series resumen queda resumida en un segmento en el que el extremo inferior corresponderá a la serie $C_{t,1}$ y el superior a $C_{t,K}$; el resto de los puntos correspondientes a las restantes series estarán entre ellos con una distancia igual a $\theta \cdot \varepsilon$ entre cada par.

En consecuencia, si ejecutamos un *escalamiento multidimensional* sobre la matriz de distancias euclídeas entre las seis series del conjunto $\{C_{t,k}\}$ la representación en una única dimensión reproduce a la perfección las distancias entre ellas (Figura 7).

FIGURA 7.
Matriz de distancias entre las series del conjunto $\{C_{t,k}\}$ y solución gráfica del *escalamiento multidimensional*

Matriz de proximidades						
	Distancia euclídea					
	c1	c2	c3	c4	c5	c6
c1	,0	10,9	21,9	32,8	43,7	54,7
c2	10,9	,0	10,9	21,9	32,8	43,7
c3	21,9	10,9	,0	10,9	21,9	32,8
c4	32,8	21,9	10,9	,0	10,9	21,9
c5	43,7	32,8	21,9	10,9	,0	10,9
c6	54,7	43,7	32,8	21,9	10,9	,0



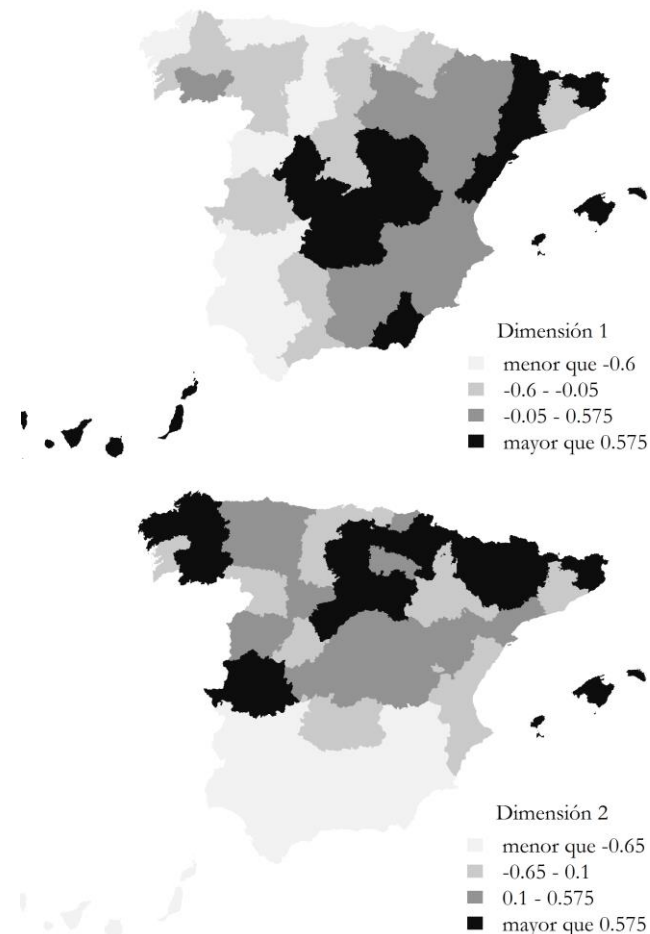
Fuente: Elaboración propia.

A partir de este resultado, utilizaremos el *escalamiento multidimensional* para simplificar la representación conjunta de todas las series homogeneizadas de desempleo (la nacional, las 50 provinciales y las 6 series resumen) Figura 8.

FIGURA 8.
Dimensiones estimadas del análisis de *escalamiento multidimensional* y su representación en el mapa espacial de provincias.

Provincia	Dimensión	
	Dim 1	Dim 2
Albacete	0,528	-1,518
Alicante	0,486	-0,466
Almería	2,384	-0,801
Álava	-1,425	0,563
Asturias	-0,795	0,242
Ávila	1,119	0,050
Badajoz	-1,166	-1,063
Balears	1,157	0,739
Barcelona	-0,375	-0,014
Vizcaya	-2,480	0,023
Burgos	-0,469	0,620
Cáceres	-0,517	0,820
Cádiz	-1,392	-1,884
Cantabria	-1,422	0,048
Castellón	2,607	0,050
C. Real	2,084	0,081
Córdoba	-0,605	-0,691
Coruña (A)	-1,067	0,972
Cuenca	1,174	0,274
Guipúzcoa	-2,293	0,114
Girona	1,532	1,078
Granada	0,099	-1,386
Guadalajara	0,550	0,386
Huelva	-0,632	-1,143
Huesca	0,222	1,010
Jaén	0,179	-1,271
León	-0,285	0,305
Lleida	0,924	1,554
Lugo	-0,577	1,200

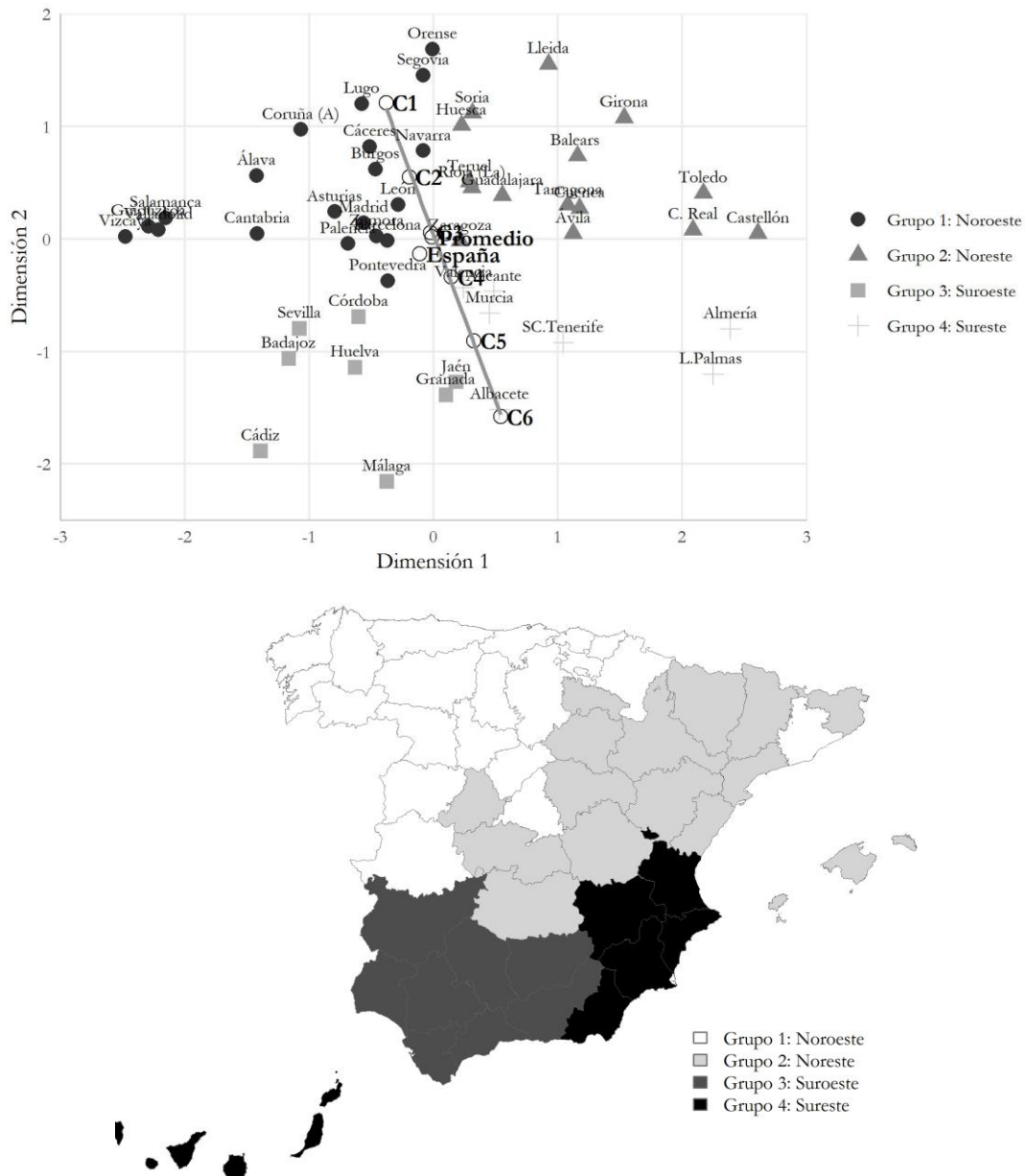
Provincia	Dimensión	
	Dim 1	Dim 2
Madrid	-0,564	0,142
Málaga	-0,377	-2,155
Murcia	0,448	-0,661
Navarra	-0,086	0,784
Orense	-0,010	1,687
Palencia	-0,691	-0,042
L.Palmas	2,248	-1,203
Pontevedra	-0,371	-0,373
Rioja (La)	0,306	0,455
Salamanca	-2,157	0,186
SC.Tenerife	1,043	-0,923
Segovia	-0,085	1,455
Sevilla	-1,081	-0,795
Soria	0,310	1,120
Tarragona	1,079	0,301
Teruel	0,288	0,508
Toledo	2,167	0,407
Valencia	0,241	-0,432
Valladolid	-2,215	0,082
Zamora	-0,459	0,026
Zaragoza	0,211	-0,017
C1	-0,382	1,211
C2	-0,197	0,551
C3	-0,026	0,058
C4	0,142	-0,339
C5	0,323	-0,905
C6	0,538	-1,579
Promedio	-0,013	0,015
España	-0,112	-0,133



Fuente: Elaboración propia.

La interpretación del correspondiente *mapa perceptual* y del espacial (Figura 9) permite extraer información adicional más allá de la mera cercanía o semejanza de la dinámica de las diferentes tasas de paro, proporcionando a su vez información sobre cómo es dicha dinámica (más o menos elástica, o con elasticidad estable, creciente o decreciente) a lo largo de todo el periodo considerado, desde 1991 hasta 2018.

FIGURA 9.
Mapa perceptual y mapa espacial de provincias clasificadas en los grupos noroeste (NO), noreste (NE), suroeste (SO) y sureste (SE).

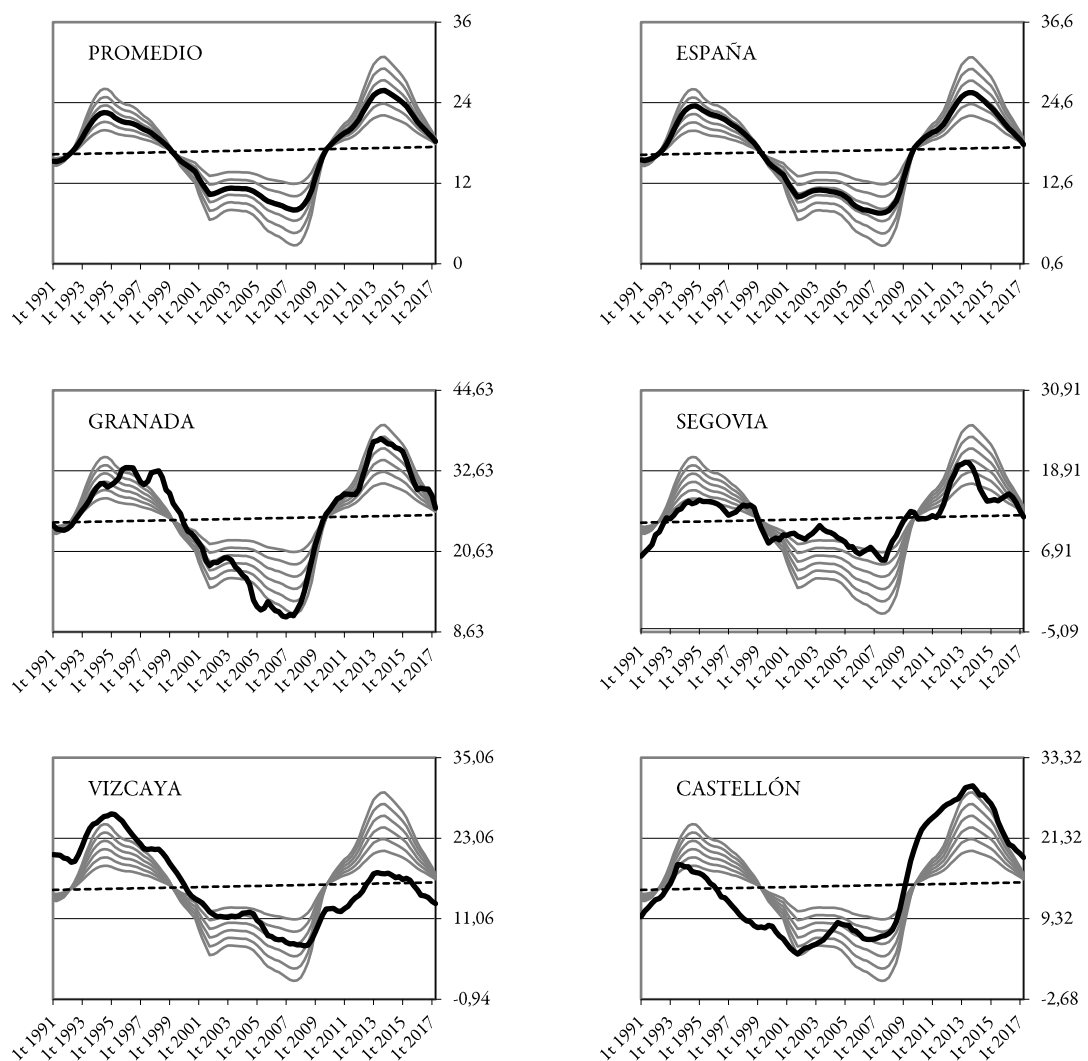


Fuente: Elaboración propia.

Según se deriva del resultado anterior, las seis series *resumen* $\{C_{t,k}\}$ están mapeadas en el segmento de una recta que, cortando el eje de abscisas en un punto próximo al origen de coordenadas, forma un ángulo pequeño con el de ordenadas; en particular, la tercera serie se encuentra muy próxima al origen de coordenadas.

Esta propiedad de las series *resumen* permite interpretar la representación del conjunto de provincias en el *mapa perceptual*. La dimensión asociada al eje de ordenadas (dimensión 2) está relacionada con el grado de elasticidad respecto de la serie promedio (panel superior izquierdo de la **Figura 10**), también próxima al origen de coordenadas: las series mapeadas en la parte negativa del eje de ordenadas son las más elásticas, mientras que las mapeadas en la positiva son las menos elásticas. Así, al comparar las trayectorias de dos provincias mapeadas prácticamente sobre el eje de abscisas, como Granada y Segovia (panel superior de la **Figura 10**), con coordenadas para la dimensión 2 negativa y positiva, respectivamente, se observa (paneles centrales de la **Figura 10**) que la trayectoria de Granada transcurre próxima a la de la sexta serie *resumen* procedente del haz (la más elástica), mientras que la de Segovia transcurre próxima a la de la primera (la menos elástica).

FIGURA 10.
Tasas de paro promedio, nacional, en Granada, Segovia, Vizcaya y Castellón, representadas sobre el conjunto de series generadas por el *haz de rectas* rescaladas.



Fuente: Elaboración propia.

Para interpretar la dimensión correspondiente al eje de abscisas (dimensión 1) del *mapa perceptual* comparemos las trayectorias de las provincias más extremas, Vizcaya y Castellón. Obsérvese (panel inferior de la **Figura 10**) que, mientras la de Vizcaya transcurre próxima a la serie más elástica en la primera etapa y a la

menos elástica en las dos siguientes, la de Castellón transcurre próxima a la menos elástica en las dos primeras etapas y a la más elástica en la tercera. Podemos interpretar, por tanto, esta dimensión 1 del eje de abscisas como un indicador de los cambios o evolución de la elasticidad de una serie a lo largo del tiempo: valores negativos de la dimensión 1 implican una reducción de la elasticidad a medida que transcurre el tiempo; valores positivos, que dicha elasticidad aumenta con el paso del tiempo; y valores próximos a cero, que la elasticidad es estable a lo largo de todo el periodo⁵.

La representación geográfica de las dimensiones obtenidas por cada provincia (**Figura 8**) confirma la relevancia de los componentes espaciales en la evolución del desempleo. Las provincias de la mitad norte de la península están mapeadas, en general, por encima del eje horizontal (lo que indica que son menos elásticas que la media), mientras que las de la mitad sur lo están por debajo (lo que indica que son más elásticas que la media); por otro lado, las provincias de la mitad este están mapeadas, en general, a la derecha del eje vertical (lo que indica que presentan elasticidad creciente a lo largo del tiempo), mientras que las de la mitad oeste lo están a la izquierda (lo que indica que presentan elasticidad decreciente en el periodo considerado).

Estos resultados permiten simplificar la interpretación de las similitudes y diferencias entre las evoluciones o comportamientos de las distintas provincias. Por ejemplo, puede considerarse la solución de cuatro grupos correspondientes a la partición arbitraria proporcionada por el corte en el *mapa perceptual* de un eje paralelo al de ordenadas, que se cruce con el eje de abscisas entre las proyecciones de Jaén y Zaragoza, y un eje paralelo al de abscisas, que se cruce con el eje de ordenadas entre las proyecciones de Palencia y Valencia. Se delimitan así cuatro áreas, en función de una mayor o menor elasticidad que el comportamiento promedio (dimensión 2), y de una elasticidad decreciente o creciente a lo largo del periodo considerado (dimensión 1). Según este resultado, denominaremos a los cuatro grupos de la partición obtenida Noreste, Noroeste, Suroeste y Sureste, cuya composición es la siguiente (**Figura 9**):

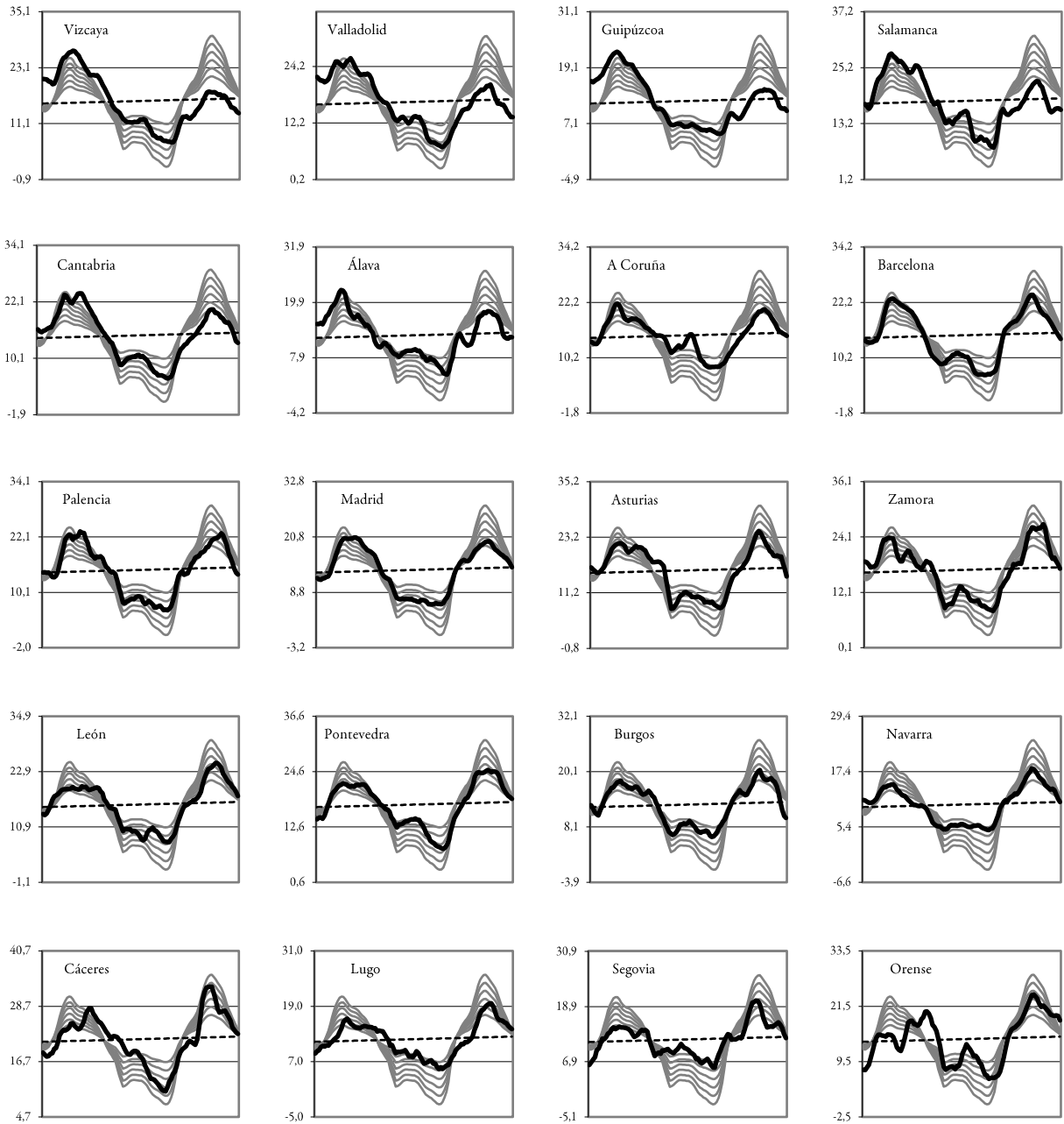
Noroeste: Formado por las provincias de las CCAA de Galicia, Asturias, País Vasco, Navarra, Castilla-León (salvo Ávila y Soria) y Madrid, con las excepciones de Barcelona y Cáceres: Vizcaya, Valladolid, Guipúzcoa, Salamanca, Cantabria, Álava, Barcelona, Palencia, Madrid, Asturias, Zamora, León, Pontevedra, Burgos, Navarra, Cáceres, A Coruña, Lugo, Segovia y Ourense. Este conjunto de provincias responde en general al patrón de elasticidad baja o media en las tres etapas, salvo en determinados casos (aquellas provincias que presentan valores extremadamente bajos en el eje de ordenadas) en los que la elasticidad en la primera etapa es alta (**Figura 11**).

Noreste: Formado por las provincias de las CCAA de Castilla-La Mancha (salvo Albacete), Cataluña (salvo Barcelona), Baleares, Aragón y La Rioja, con las excepciones de Castellón, Ávila y Soria: Castellón, Toledo, Ciudad Real, Ávila, Cuenca, Tarragona, Baleares, Girona, Lleida, Zaragoza, Guadalajara, La Rioja, Teruel, Huesca y Soria. Este conjunto de provincias responde en general al patrón de elasticidad baja o media en las tres etapas, salvo en determinados casos (aquellas provincias que presentan valores extremadamente altos en el eje de ordenadas) en los que la elasticidad en la tercera etapa es acusada (**Figura 12**).

Suroeste: Formado por las provincias de las CCAA de Andalucía (salvo Almería) y Extremadura (salvo Cáceres): Cádiz, Badajoz, Sevilla, Córdoba, Huelva, Jaén, Granada y Málaga. Este conjunto de provincias responde en general al patrón de elasticidad alta en las tres etapas. En determinados casos (aquellas provincias que presentan valores más bajos en el eje de ordenadas) la elasticidad en la primera etapa es relativamente más alta que en la tercera (**Figura 13**).

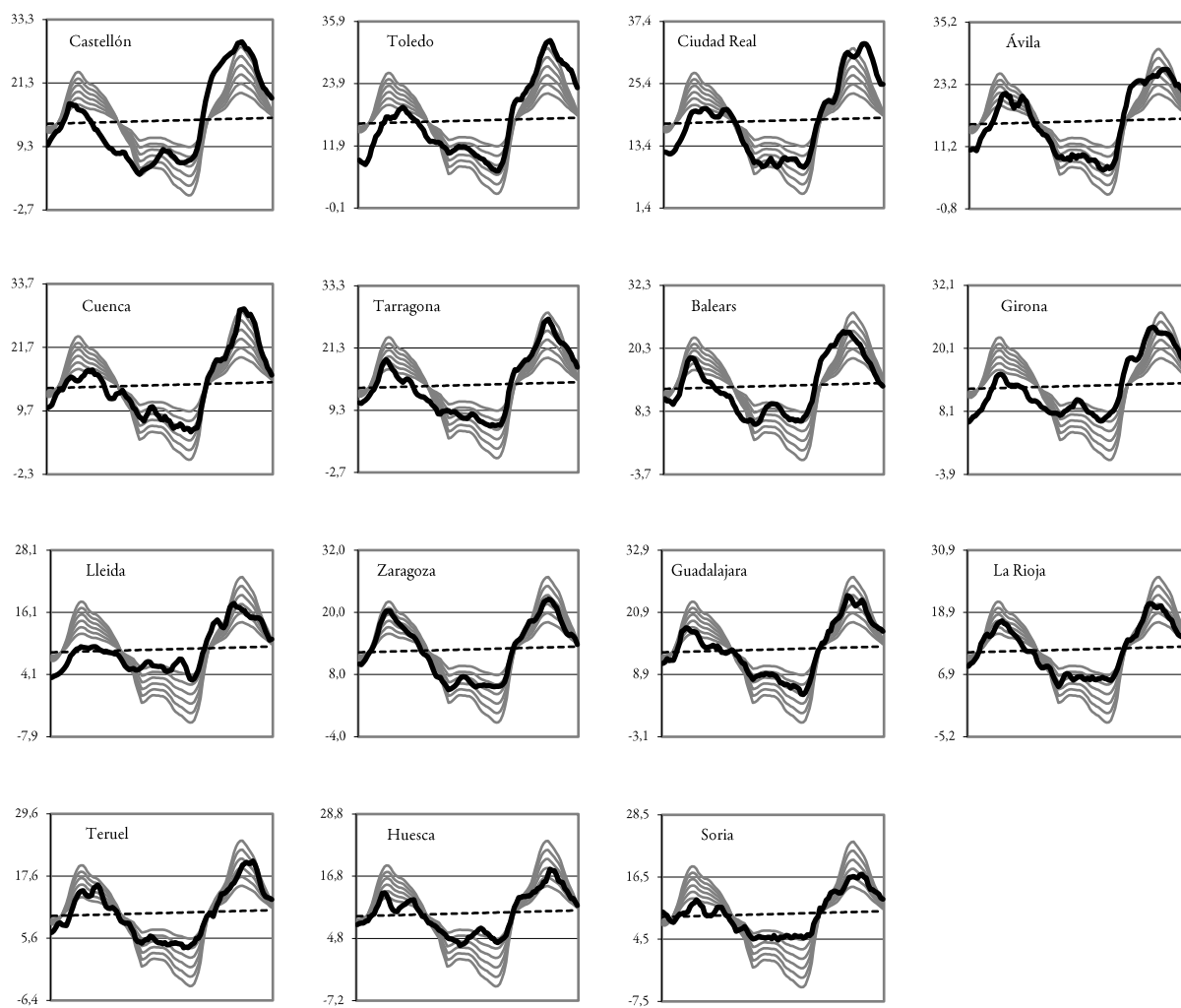
⁵ La trayectoria de la serie agregada es prácticamente igual a la de la serie promedio en las etapas segunda y tercera (paneles superiores de la Figura 10); sin embargo, en la primera etapa se observa una ligera diferencia: mientras que la serie promedio transcurre pegada a la tercera *curva resumen* procedente del haz, la agregada lo hace pegada a la cuarta. Esta ligera desviación se traduce en que, en el *mapa perceptual*, aunque las series promedio y agregada se sitúen muy próximas entre sí y con valores cercanos al origen de coordenadas, exista una diferencia entre la serie promedio y la serie agregada mayor en la coordenada correspondiente al eje de ordenadas (dimensión 2) que entre las coordenadas correspondientes al eje de abscisas (dimensión 1).

FIGURA 11.
Tasa de paro en las provincias del grupo Noroeste.



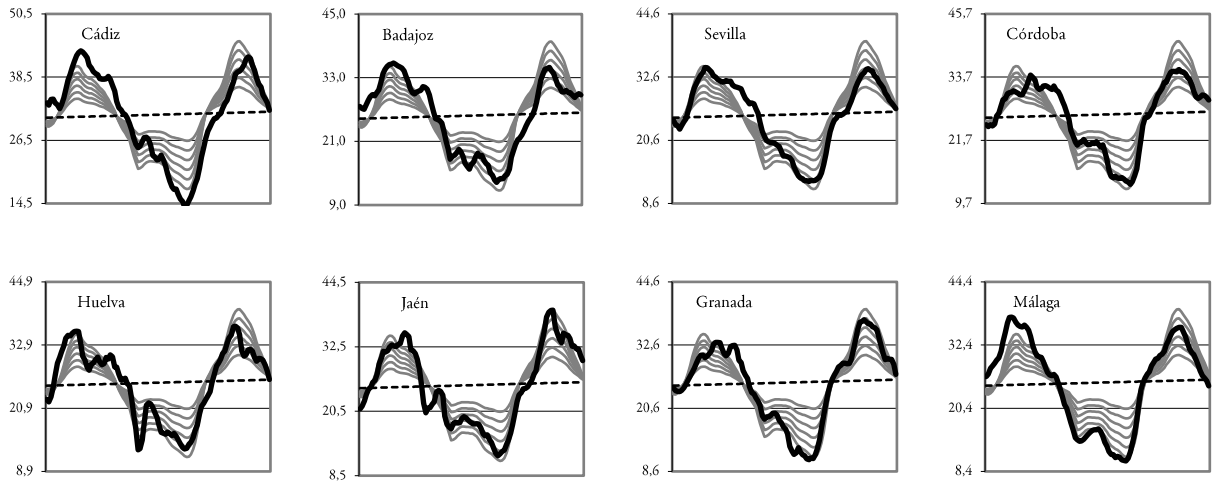
Fuente: Elaboración propia.

FIGURA 12.
Tasa de paro en las provincias del grupo Noreste.



Fuente: Elaboración propia.

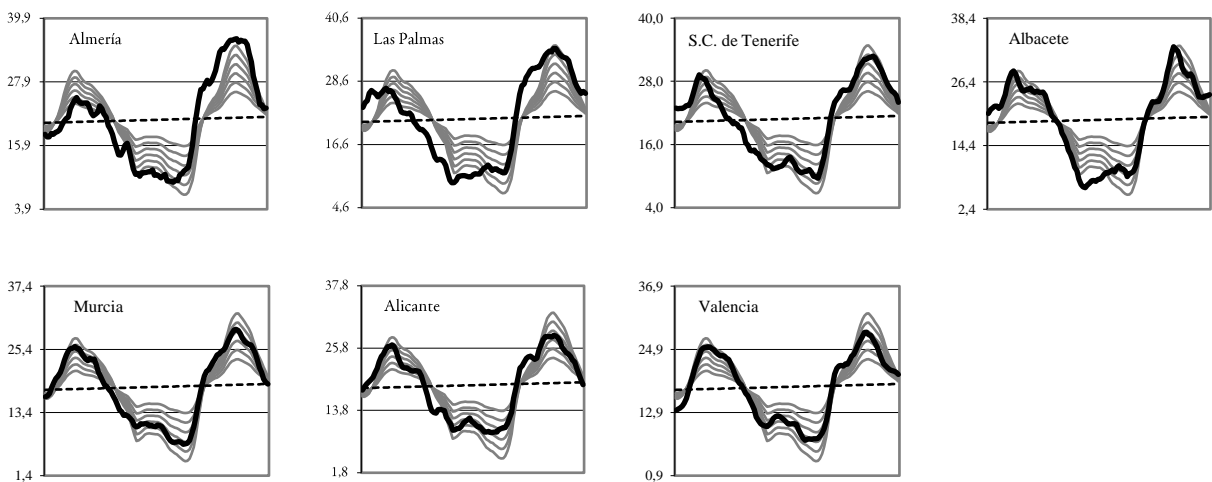
FIGURA 13.
Tasa de paro en las provincias del grupo Suroeste



Fuente: Elaboración propia.

Sureste: Formado por las provincias de las CCAA de Canarias, Murcia y la Comunidad Valenciana (salvo Castellón), con las excepciones de Almería y Albacete: Almería, Las Palmas, Santa Cruz de Tenerife, Albacete, Murcia, Alicante y Valencia. Este conjunto de provincias responde en general al patrón de elasticidad alta en las tres etapas. En determinados casos (aquellas provincias que presentan valores acusadamente altos en el eje de ordenadas) la elasticidad en la tercera etapa es relativamente más alta que en la primera (Figura 14).

FIGURA 14.
Tasa de paro en las provincias del grupo Sureste



Fuente: Elaboración propia

4. CONCLUSIONES

En este trabajo se presenta una nueva propuesta metodológica que permite visualizar y detectar diferencias y semejanzas en los patrones del comportamiento dinámico de series temporales regionales, eliminando el efecto de escala que surge al analizar regiones heterogéneas en cuanto a su tamaño. Esta propuesta metodológica para el análisis descriptivo de series temporales regionales combina la metodología del *haz de rectas* (Ferrán, 2011) con el análisis de *escalamiento multidimensional*.

La metodología del *haz de rectas* establece que, si entre un conjunto de series temporales regionales subyace un *haz de rectas*, es decir, existe una relación lineal entre los coeficientes de las regresiones de cada serie regional y la serie promedio de todas ellas, entonces podemos crear un nuevo conjunto de series que resume gráficamente el patrón dinámico de todas ellas. Con estas series resumen se puede verificar, visualmente, cuándo una determinada provincia o región es más elástica que la media, y para qué periodos ha habido un cambio en dicha elasticidad. Además, cuando subyace un *haz de rectas*, a través del escalamiento multidimensional es posible resumir en un único gráfico bidimensional los comportamientos dinámicos de las diferentes series, proporcionando los resultados del *haz de rectas* la interpretación de esas dos dimensiones. Por ejemplo, en el caso analizado de las tasas de paro, una de las dimensiones representa la elasticidad del desempleo de cada provincia respecto a la serie promedio, y la otra dimensión representa los cambios ocurridos en dicha elasticidad (si ha aumentado o ha disminuido o a permanecido estable) a lo largo de las diferentes fases del ciclo económico analizado. De esta forma se facilita tanto la comparación visual de los diferentes patrones dinámicos como la agrupación de regiones con patrones similares.

Como aplicación práctica de la metodología se han analizado las diferencias y similitudes en el comportamiento de las tasas de desempleo de las 50 provincias españolas a lo largo del periodo comprendido entre 1991 y 2018. Las diferencias entre provincias se encuentran, además de en sus niveles medios de desempleo, en los grados de disminución de las tasas de desempleo en las fases alcistas del ciclo económico, y en los grados de aumento del desempleo en las fases recesivas. Es decir, en sus patrones dinámicos, en la elasticidad de cada una de las provincias respecto a la serie promedio y en los cambios a lo largo de los ciclos en dicha elasticidad.

Una vez eliminado el efecto escala derivado de que no todas las provincias tienen las mismas tasas medias de desempleo, con la aplicación de la propuesta metodológica se ha podido evidenciar que existen provincias que han sufrido relativamente más los efectos de las crisis en términos de desempleo, provincias que han mejorado su capacidad de recuperación tras una crisis y otras que han reducido su vulnerabilidad ante la última gran crisis del 2008.

En este sentido, uno de los resultados del análisis de las series de desempleo es que se confirma la relevancia de los componentes espaciales en la dinámica de su evolución a lo largo de los ciclos. En efecto, la propuesta metodológica ha permitido comparar qué provincias son más parecidas a otras y en qué términos. Cada dimensión del análisis, elasticidad y cambio en la elasticidad, separa a las provincias españolas en dos grupos claramente diferenciados. Así, haciendo referencia a la dimensión que refleja la elasticidad respecto a la serie promedio, las provincias de la mitad sur de la península y de Canarias presentan en general grados de variación del desempleo elevados en comparación con las de la mitad norte, lo que implica un mayor aumento relativo de las tasas durante la crisis, pero, a su vez, también una mayor disminución relativa del desempleo en las fases de recuperación económica.

En relación a la segunda dimensión, los cambios en la elasticidad, existe un grupo de provincias, la mayoría situadas en la mitad oeste de la península, que en general han reducido la elasticidad de sus respectivos mercados de trabajo, esto es, que han reducido el impacto de las crisis y se han visto relativamente menos afectadas por el aumento generalizado de la tasa de paro en la reciente crisis. En el extremo opuesto existe otro grupo de provincias, en general situadas en la mitad este de la península, para las que, en comparación con su propio pasado, la reciente crisis ha afectado de una forma claramente más acusada a sus tasas de desempleo.

La intersección de dicha delimitación en dos grupos de provincias en cada una de las dos dimensiones de la dinámica del desempleo ha dado lugar a cuatro conjuntos de provincias que siguen un claro patrón de agrupación espacial norte-sur-este-oeste, en la misma línea de otros trabajos como los de Cueto Iglesias et al.(2017, p72) que también encuentran que las regiones situadas en el sur y el este son las que resisten mejor las crisis y las que más empleo crean durante la recuperación, mientras que el norte y, particularmente, las regiones del noroeste presentan un patrón opuesto, con baja resistencia y baja recuperación.

A partir de estos patrones detectados, en futuras aportaciones, sería conveniente profundizar en los factores que podrían estar explicando dicho comportamiento diferencial y en los patrones provinciales de desempleo detectados en este trabajo. Dicho análisis queda fuera del alcance de este trabajo, cuyo propósito fundamental es el de ofrecer una propuesta metodológica para la identificación y agrupación de los patrones dinámicos de series temporales regionales.

5. REFERENCIAS BIBLIOGRÁFICAS

- Alonso-Villar, O.; Del Río, C. y Toharia, L. (2009): “Un análisis espacial del desempleo por municipios”, *Revista de Economía Aplicada*, 49 (XXVII), páginas 47-80.
- Andrienko, N.; Andrienko, G. y Gatalsky, P. (2003): “Exploratory spatio-temporal visualization: an analytical review”, *Journal of Visual Languages & Computing*, 14: 503-541.
- Bande, R.; Fernández, M.; y Montuenga V. (2008): “Regional unemployment in Spain: Disparities, business cycle and wage setting”. *Labour Economics*, Volume 15, Issue 5, October 2008, pp. 885-914.
- Bande, R. and Karanassou, M. (2009): “Labour market flexibility and regional unemployment rate dynamics: Spain 1980–1995”. *Papers in Regional Science*, 88: 181–207.
- Bande R, Karanassou, M. (2013) “The NRU and the evolution of regional disparities in Spanish unemployment”. *Urban Studies* 50(10):2044–2062.
- Bande, R, Karanassou, M. (2014): “Spanish regional unemployment revisited: the role of capital accumulation”. *Regional Studies*. Volume 48, Issue 11, 2014.
- Borg, I. and Groenen, P. (2005): *Modern Multidimensional Scaling: Theory and Applications* (2nd ed.). New York: Springer-Verlag.
- Cueto Iglesias, M.B., Mayor Fernández, M y Suarez Cano, P. (2017): *La resiliencia de las regiones españolas, después de la gran recesión*. Consejo Económico y Social del Principado de Asturias. Colección de Estudios, núm. 20.
- Fernández Macho, F.J. (dir.) (1997): *Cointegración y convergencia en la Unión Europea*, Colección Economía, Servicio Editorial de la Universidad del País Vasco.
- Ferrán Aranaz, M. (2011): *Una metodología de minería de datos para la agrupación de series temporales: aplicación al sector de la construcción residencial*, Universidad Complutense de Madrid, Tesis Doctoral.
- Ferrán Aranaz, M. (2013): “El haz de rectas para la comparación gráfica de series temporales geográficas”, *Estadística Española*, 181: 123-148.
- García Cintado, A.; Romero Ávila, D.; y Usabiaga, C. (2015): “Can the hysteresis hypothesis in Spanish regional unemployment be beaten? New evidence from unit root tests with breaks”. *Economic Modelling*, Volume 47, June 2015, Pages 244-252.
- Gelman, A. y Unwin, A. (2011): “Visualization, Graphics and Statistics”, *Statistical Computing & Graphics Newsletter*, 22: 9-12.
- Hochheiser, H. y Shneiderman, B. (2001): “Interactive exploration of time series data”, *In: The 4th International conference on Discovery Science (Washington, DC)*, Springer-Verlag, Berlin, 441-446.

- Kosara, R. (2011): "Visualization: is more than Pictures!", *Statistical Computing & Graphics Newsletter*, 22: 5-8.
- Liao, T.W. (2005): "Clustering of time series data – a survey", *Pattern Recognition*. 38: 1857-1874.
- Lin, J.; Keogh, E. y Lonardi, S. (2005): "Visualizing and discovering non-trivial patterns in large time series databases". *Information Visualization*, 4: 61-82.
- López-Bazo, E.; Del Barrio, T. y Artís, M. (2002): "The Regional Distribution of Spanish Unemployment. A Spatial Analysis", *Papers in Regional Science*, 81, páginas 365- 389.
- López-Bazo, E.; Del Barrio, T. y Artís, M. (2005): "Geographical Distribution of Unemployment in Spain", *Regional Studies*, 39 (3), pp 305-318.
- López-Bazo E, Motellón E (2013): "The regional distribution of unemployment: what do micro-data tell us?", *Papers in Regional Science* 92(2), pp383–405.
- Mead, A (1992): "Review of the Development of Multidimensional Scaling Methods", *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. 41, No. 1, pp. 27-39.
- Real Deus, J.E. y Varela Mallou, J (2003): "Esclamiento multidimensional" en Lévy Mangin, J.P y Varela Mallou, J (dirs.): *Análisis multivariable para las Ciencias Sociales*, Pearson Prentice-Hall, Madrid, pp 451-505.
- Sala, H. y Trivín, P. (2014): "Labour market dynamics in Spanish regions: evaluating asymmetries in troublesome times". *SERIEs*, August 2014, Volume 5, Issue 2, pp 197-221.
- Suriñach, J; Artís, M.; López, E.; y Sansó, A. (1995): *Análisis económico regional, nociones básicas de la Teoría de la Cointegración*. Antoni Bosch Editor.
- Tufte, E.R. (2001): *The visual display of quantitative information*, Cheshire, Conn: Graphic Press.
- Van Wijk, J.J. y van Selow, E.R. (1999): "Cluster and calendar based visualization of time series data", *In: 1999 IEEE Symposium on Information Visualization (San Francisco, CA)*, 4-9.
- Viñuela Jiménez, A; Rubiera Morollón, F.; y Cueto Iglesias, B. (2012): "Espacio y empleabilidad. ¿Importa el concepto de región?", *Información Comercial Española, Revista de Economía*, 865: 155-167.
- Viñuela Jiménez, A.; Rubiera Morollón, F.; y Fernández Vázquez, E.(2014): "Applying economic-based analytical regions: A study of the spatial distribution of employment in Spain", *The Annals of Regional Science*, January 2014, Volume 52, Issue 1, pp .87-102.
- Weber, M.; Alexa, M. y Muller, W. (2001): "Visualizing time series on spirals", *In: 2001 IEEE Symposium on Information Visualization (San Diego, CA)*, 7-14.
- Zhao, J; Chevalier, F.; Pietriga, E. y Balakrishnan, R. (2011): "Exploratory Analysis of Time-Series with ChronoLenses", *IEEE Transaction on visualization and computers graphics*, 17(12): 2422-2431.
- Zhao, Y (2015): *R and Data Mining: Examples and Case Studies*. Academic Press, Elsevier.

ORCID

Lorenzo Escot <https://orcid.org/0000-0002-6734-6327>

ANEXO I: LA METODOLOGÍA DEL HAZ DE RECTAS

1. Conceptos básicos

La metodología del *haz de rectas* es una herramienta estadística que calcula un conjunto reducido de *series resumen* con las que se simplifica la comparación de un conjunto amplio de trayectorias⁶ (Ferrán 2011a, 2011b y 2013). El concepto clave y el resultado fundamental de la metodología son:

Estructura de haz de rectas: *Un conjunto $\{C_{t,k}\}$ de K series temporales distintas, todas ellas definidas en los mismos instantes, $t = 1, \dots, T$, tiene estructura de haz de K rectas si existe otra serie x_t tal que para cada $C_{t,k}$ existen cuatro coeficientes b_k, m_k, B_0 y B_1 , con al menos m_k diferente de cero, tales que:*

$$C_{t,k} = b_k + m_k \cdot x_t \quad \forall t, k \quad \text{donde} \quad b_k = B_0 + B_1 \cdot m_k \quad \forall k$$

El vértice del haz de rectas será igual a $(-B_1, B_0)$.

Resultado fundamental la metodología del haz de rectas: *Sea $\{C_{t,k}\}, k = 1, \dots, K$, un conjunto de K series temporales distintas, todas ellas definidas en los mismos instantes temporales, $t = 1, \dots, T$, y con la misma media:*

$$\bar{C}_k = \frac{1}{T} \cdot \sum_{t=1}^T C_{t,k} = \alpha \quad \forall k \quad (1)$$

Sea X_t otra serie temporal y supongamos que para cada $C_{t,k}$ existen cinco coeficientes b_k, m_k, μ_k, B_0 y B_1 con al menos m_k distinto de cero, tales que:

$$C_{t,k} = b_k \cdot t + m_k \cdot X_t + \mu_k \quad \forall t, k \quad (2)$$

siendo:

$$b_k = B_0 + B_1 \cdot m_k \quad \forall k \quad (3)$$

Entonces:

- A) *El conjunto de series $\{C_{t,k} = \Delta_s C_{t,k}\}$ presenta estructura de haz de rectas respecto de la serie $x_t = \Delta_s X_t$ con vértice en $(\Delta_s X_t, \Delta_s C_{t,k}) = (-sB_1, sB_0)$.*
- B) *Para cualquier par de series temporales del conjunto $\{C_{t,k}\}$ existe al menos un punto en su trayectoria en el que se cortan. Además, los puntos de corte de cualquier par de trayectorias son los puntos de corte de todas ellas.*
- C) *Si las trayectorias se cortan en más de un punto entonces la diferencia entre dos puntos de corte cualesquiera es independiente de la media de las series temporales. Además, todos los puntos de corte están situados sobre una línea recta, la correspondiente al coeficiente $m_k = 0$.*
- D) *Si $C_{t,q}, C_{t,r}$ y $C_{t,p}$ son tres series temporales cualesquiera del conjunto $\{C_{t,k}\}$ tales que $m_q < m_r < m_p$ entonces la distancia euclídea entre la primera y la segunda es menor que la distancia euclídea entre la primera y la tercera:*

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

⁶ Utilizaremos el término trayectoria para referirnos a la línea continua que conecta la secuencia de puntos en la representación gráfica de una serie temporal.

2. Construcción del conjunto de series tal que el correspondiente conjunto de series diferenciadas presenta estructura de haz de rectas

A partir del conjunto de J series temporales $\{Y_{t,j}\}$, $t = 1, \dots, T$ y $j = 1, \dots, J$, la metodología del *haz de rectas* consistirá en la construcción de un conjunto reducido de series $\{C_{t,k}\}$, $k = 1, \dots, K$, tales que el correspondiente conjunto $\{c_{t,k} = \Delta_s C_{t,k}\}$ presente estructura de haz de rectas respecto de la serie $x_t = \Delta_s X_t$, donde X_t es la serie promedio del conjunto $\{Y_{t,j}\}$. Las series del conjunto $\{C_{t,k}\}$ se reescalarán al nivel de cada serie del conjunto $\{Y_{t,j}\}$, creando J nuevos conjuntos de *series resumen* $\{C_{t,k}^j\}$, con el objetivo de que la representación gráfica de la trayectoria de cada serie $Y_{t,j}$ sobre el correspondiente conjunto de *series resumen* reescaladas facilite la interpretación de las similitudes y diferencias respecto de las restantes series.

La metodología consta de los siguientes pasos:

Paso 1: Cálculo de la serie promedio, $X_t = \frac{1}{J} \cdot \sum_{j=1}^J Y_{t,j}$, y creación de las series $Z_{t,j}$:

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

donde $\alpha_j = \bar{Y}_j$ y $\alpha = \bar{X}_j$. Mediante esta transformación, la escala de medida del nuevo conjunto de series es la misma que la de la serie promedio: $\bar{Z}_j = \alpha, \forall j$.

Paso 2: Cálculo de las J ternas de coeficientes $(A_{0,j}, A_{1,j}, A_{2,j})$, $j = 1, \dots, J$, mediante el ajuste de las J ecuaciones de regresión lineal:

$$\hat{Z}_{t,j} = A_{0,j} \cdot t + A_{1,j} \cdot X_t + A_{2,j} \quad (4)$$

Paso 3: Cálculo de los coeficientes B_0 y B_1 mediante el ajuste de la ecuación de regresión lineal:

$$\hat{A}_{0,j} = B_0 + B_1 \cdot A_{1,j} \quad (5)$$

sobre los J puntos $(A_{1,j}, A_{0,j})$.

Paso 4: Dentro del rango de variación de los valores $A_{1,j}$, $j = 1, \dots, J$, elección de K valores representativos. Por ejemplo, $m_k = m_{k-1} + \theta$, donde:

$$\theta = (\max_j A_{1,j} - \min_j A_{1,j}) / (K - 1) \quad \text{y} \quad m_1 = \min_j A_{1,j}$$

Paso 5: Cálculo de los K coeficientes $b_k = B_0 + B_1 \cdot m_k$.

Paso 6: Cálculo de los K coeficientes μ_k :

$$\mu_k = \alpha - \bar{g}_k, \text{ donde } g_{t,k} = b_k \cdot t + m_k \cdot X_t \quad (6)$$

Paso 7: Cálculo del conjunto de K series temporales $\{C_{t,k}\}$:

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

Así, el conjunto de K series $\{c_{t,k} = \Delta_s C_{t,k}\}$, al verificar las condiciones (1), (2) y (3) respecto de $x_t = \Delta_s X_t$, presenta estructura de *haz de rectas* y, además, el conjunto de series $\{C_{t,k}\}$ está en la misma escala que el conjunto $\{Z_{t,j}\}$.

Paso 8: Cálculo de los J conjuntos de series temporales $\{C_{t,k}^j\}$:

$$C_{t,k}^j = C_{t,k} - \alpha + \alpha_j$$

Así, el conjunto de series $\{c_{t,k}^j = \Delta_s C_{t,k}^j\}$, $k = 1, \dots, K$, al verificar las condiciones (1), (2) y (3) respecto de $x_t = \Delta_s X_t$, presenta estructura de *haz de rectas* y, además, cada uno de los conjuntos $\{C_{t,k}^j\}$ está en la misma escala que la serie $Y_{t,j}$.

La representación de cada serie $Y_{t,j}$ sobre el correspondiente conjunto $\{C_{t,k}^j\}$ facilitará la interpretación de las similitudes y diferencias entre todas ellas.

ANEXO II

Demostración:

Según la expresión (2), la distancia euclídea entre dos series del conjunto $\{C_{t,k}\}$ viene dada por:

$$\begin{aligned} d(C_{t,k}, C_{t,k'}) &= \left(\sum_{t=1}^T ((b_{k'} \cdot t + m_{k'} \cdot X_t + \mu_{k'}) - (b_k \cdot t + m_k \cdot X_t + \mu_k))^2 \right)^{1/2} \\ &= \left(\sum_{t=1}^T ((b_{k'} - b_k) \cdot t + (m_{k'} - m_k) \cdot X_t + (\mu_{k'} - \mu_k))^2 \right)^{1/2} \end{aligned}$$

Según la expresión (6): $\mu_k = \alpha - b_k \cdot \bar{t} + m_k \cdot \bar{X}_t$.

En consecuencia, según (3): $\mu_k = (\alpha - B_0 \cdot \bar{t}) + (B_1 \cdot \bar{t} + \bar{X}_t) \cdot m_k$.

Luego, si denominamos $B'_0 = \alpha - B_0 \cdot \bar{t}$ y $B'_1 = B_1 \cdot \bar{t} + \bar{X}_t$, entonces:

$$\mu_k = B'_0 + B'_1 \cdot m_k \quad (7)$$

En consecuencia, según (3) y (7):

$$\begin{aligned} d(C_{t,k}, C_{t,k'}) &= \left(\sum_{t=1}^T ((B_1 \cdot m_{k'} - B_1 \cdot m_k) \cdot t + (m_{k'} - m_k) \cdot X_t + (B'_1 \cdot m_{k'} - B'_1 \cdot m_k))^2 \right)^{1/2} \\ &= \left(\sum_{t=1}^T ((m_{k'} - m_k) \cdot B_1 \cdot t + (m_{k'} - m_k) \cdot X_t + (m_{k'} - m_k) \cdot B'_1)^2 \right)^{1/2} \\ &= (m_{k'} - m_k) \left(\sum_{t=1}^T (B_1 \cdot t + X_t + B'_1)^2 \right)^{1/2} = (m_{k'} - m_k) \cdot \varepsilon \end{aligned}$$

donde $\varepsilon = \left(\sum_{t=1}^T (B_1 \cdot t + X_t + B'_1)^2 \right)^{1/2}$ es un valor constante, positivo e independiente de t , de k y de k' .