

Convergencia de las tasas de paro regionales en España: un análisis de cointegración¹

Alfonso Moral de Blas

Universidad de Valladolid

Dpto. de Fundamentos del Análisis Económico

Avda. Valle Esqueva, 6

47011 Valladolid

RESUMEN

Las diferencias entre el nivel de vida de los distintos países junto con los recientes procesos de integración supranacional han aumentado el interés por el tema de la convergencia. Con este marco de referencia nos hemos planteado la convergencia real entre las distintas Comunidades Autónomas españolas, y para ello nos hemos centrado en el estudio del mercado de trabajo.

En este trabajo trataremos de ver si tiene algún sentido hablar de convergencia entre regiones en términos de tasas de paro. Para ello, utilizando técnicas de cointegración, buscaremos relaciones de largo plazo entre las tasas de paro de cada región y el agregado nacional. Estas relaciones nos permitirán analizar la existencia de convergencia real entre las Comunidades Autónomas españolas, y la influencia que tiene el grado de homogeneidad del grupo estudiado sobre los resultados obtenidos.

¹ El autor quiere agradecer los comentarios realizados por los dos evaluadores anónimos, en cualquier caso las posibles deficiencias existentes en el trabajo son responsabilidad exclusiva de su autor.

1. INTRODUCCIÓN

Las últimas décadas del siglo XX han venido marcadas por una auténtica proliferación de estudios centrados en la convergencia económica. La razón de este hecho habría que buscarla en los recientes procesos de globalización de la economía y en la formación de entidades supranacionales.

No obstante, y aunque la literatura sobre el tema parece apuntar lo contrario, la cuestión de la convergencia real no tiene por qué circunscribirse exclusivamente al ámbito internacional. También tiene sentido preguntarse si este progresivo acercamiento entre las economías de los distintos países se ha visto correspondido por un mayor grado de integración a escala nacional.

El propósito de este trabajo va más en esa segunda línea. Nos preguntamos si el proceso de integración europea en que se ha visto inmersa la economía española ha dado lugar a un mayor grado de homogeneidad (si ha habido o no ha habido convergencia real) entre las regiones españolas. Más concretamente, se va a realizar un análisis de la convergencia real entre las distintas Comunidades Autónomas españolas en términos de tasas de paro. Y para ello vamos a servirnos de un método basado en técnicas de cointegración.

Como en cualquier análisis regional, el estudio de la convergencia del desempleo debe tener en cuenta dos tipos de factores explicativos, los que afectan a nivel nacional, y los particulares de cada Comunidad Autónoma. En este sentido, las condiciones económicas e institucionales de cada país o las políticas económicas aplicadas a nivel nacional, pueden indicar una evolución similar del desempleo en todas las regiones. Sin embargo, las diferencias regionales en cuanto a la composición sectorial, la dotación de factores o la estructura productiva, junto con los problemas de movilidad que se presentan en el caso español, pueden afectar en sentido contrario.

En términos generales son dos los objetivos de este trabajo; por un lado trataremos de ver si las tasas de desempleo regionales siguen, o no, un proceso de acercamiento a las tasas medias de la nación; por otro, intentaremos ver que influencia ejerce en ese proceso de convergencia el grado de homogeneidad del grupo al cual se refieren las tasas de desempleo. Grado de homogeneidad que será medido a través de variables tales como la edad, el sexo o el nivel de estudios.

El resto del trabajo se organiza de la siguiente manera. En el epígrafe 2 se intenta aclarar cual es el funcionamiento del método de cointegración que vamos a emplear para el análisis empírico. El epígrafe 3 se dedica a estudiar la convergencia mediante la aplicación de la metodología anterior, y se intenta corroborar la idea de que el

grado de convergencia es mayor cuando utilizamos grupos poblacionales cada vez más homogéneos. Finalmente, en el epígrafe 4, se recogen las principales conclusiones del trabajo.

2. COINTEGRACIÓN APLICADA AL ANÁLISIS DE CONVERGENCIA

Los métodos descritos en este apartado surgen de la necesidad de dar un tratamiento sistemático al componente temporal en los estudios de convergencia, y como respuesta a otros que se basaban en la utilización de datos de corte transversal (básicamente los de β -convergencia, tanto absoluta como condicionada, y σ -convergencia)².

El hecho de trabajar con series temporales nos obliga a realizar un análisis previo para determinar si los procesos que representan dichas series son estacionarios o no. Las variables son no estacionarias cuando su distribución de probabilidad depende del tiempo, y en estos casos los modelos tradicionales de estimación pueden llevar a aceptar como válidas relaciones de tipo espurio. La existencia de estos procesos no estacionarios, o también llamados tendencias estocásticas (como ocurre con las series de tasa de paro) se detecta mediante el cálculo del orden de integración de la serie o determinando el número de raíces unitarias que presenta. Esta ausencia de estacionariedad hace que los contrastes de convergencia más adecuados sean los propuestos por Bernard y Durlauf (1996). Estos tests están basados en la utilización de técnicas de cointegración.

En esta línea se encuentran los trabajos de Fernández-Jardón y Caneda (1999), Pallardo y Esteve (1997), Avilés y cols. (1997) o Bernard and Durlauf (1994). Todos ellos introducen series de tiempo en sus estudios sobre convergencia mediante la utilización de técnicas de cointegración. En este apartado nos limitaremos a las ideas más esenciales sobre la teoría de la cointegración, aunque un desarrollo más amplio de tales cuestiones puede encontrarse en las referencias bibliográficas reseñadas.

La existencia de una relación de cointegración entre dos variables implica que las dos tienen un mismo orden de integración, y que existe una combinación lineal de ellas que presenta un orden de integración inferior. En nuestro caso concreto, las variables serán integradas de orden uno y esperaremos una combinación lineal de las variables que sea estacionaria. Esta combinación lineal se interpreta como una ten-

² Sobre estos métodos pueden consultarse los trabajos de Barro y Sala i Martín (1990, 1991, 1992)

dencia común o como una relación de equilibrio de largo plazo. Posteriormente se estudia esa relación de equilibrio y a partir de dicho estudio se deduce si el proceso que siguen ambas variables es convergente o no.

Estos métodos de convergencia basados en técnicas de cointegración permiten establecer varios estratos o niveles de convergencia (Camarero y cols. 1995, Olloqui y cols. 2002, Pallardó y Esteve 1997).

El primer caso (el tipo de convergencia más fuerte), también denominado convergencia de largo plazo, se da cuando la diferencia entre dos series genera otra cuyo valor medio tiende a cero en el largo plazo. Desde un punto de vista más formal, diremos que existe convergencia a largo plazo entre el valor que toma la variable X en la región i y el que toma en la región j sí:

$$\lim_{k \rightarrow \infty} E(X_{i,t+k} - X_{j,t+k} / \hat{I}_t) = 0$$

donde X es la variable estudiada en las regiones i y j en los momentos t y $t+k$, y \hat{I}_t recoge la información relevante en el momento t .

En este primer caso debemos hacer un supuesto adicional para que tenga lugar la convergencia, y es que la varianza de esa diferencia sea finita; en caso contrario, podría existir divergencia entre las variables aunque se produzca un acercamiento entre sus esperanzas.

De acuerdo con la teoría de la cointegración³, esta definición nos indica la ausencia de una raíz unitaria en la diferencia existente entre X_i y X_j . Si la variable X es integrada de orden 1 en las dos áreas geográficas, la existencia de este tipo de convergencia implica una relación de cointegración entre las dos series que vendrá dada por el vector $(1, -1)$. Además, ésta será la única relación a largo plazo posible entre ellas, porque solo hay dos variables, y de acuerdo con la teoría de la cointegración, solo puede existir un vector de cointegración. Esto quiere decir que, para que tenga lugar la convergencia a largo plazo, la diferencia entre las dos variables no ha de presentar ni tendencia estocástica ni tampoco determinista; es decir, esa diferencia ha de ser estacionaria sin que para ello haya que introducir ninguna tendencia ni término independiente. Un término independiente indicaría una diferencia de nivel entre las series, aunque su evolución temporal es similar, sin embargo, una tendencia determi-

³ Para profundizar en temas de cointegración puede recurrirse a los trabajos de Yule (1926), Phillips (1986), Engle y Granger (1987) y Suriñach y cols. (1995).

nista indicaría una diferencia entre las series, que evoluciona de forma estable con el tiempo y que nos indica ciertos aspectos diferenciadores entre regiones.

Este primer concepto nos estaría indicando que la diferencia entre los niveles que presenta la variable dentro de cada uno de los entes territoriales va disminuyendo a lo largo del tiempo, y por lo tanto, ambos niveles convergen hacia un mismo estado de equilibrio de largo plazo.

El segundo tipo de convergencia que podemos identificar a través del análisis de cointegración está asociado a un menor nivel de exigencia, y se designa a través del término *catching up*. En este caso lo que ocurre es que la brecha entre ambas series va disminuyendo a lo largo del tiempo pero cada una puede tender hacia su propio equilibrio de largo plazo. Este tipo de convergencia se da cuando

$$E(X_{i,T} - X_{j,T} / \hat{\nu}_t) < X_{i,t} - X_{j,t}$$

donde t indica el momento actual y T algún momento futuro para el que se hace la proyección.

En este segundo caso seguimos manteniendo que la diferencia entre $X_{i,t}$ y $X_{j,t}$ debe ser estacionaria, pero se permite la existencia de una tendencia determinista; lo que no debe aparecer nunca es una tendencia estocástica en esa diferencia de variables. Nuevamente $X_{i,t}$ y $X_{j,t}$ han de ser cointegradas, y como en el caso de la convergencia de largo plazo $(1, -1)$, es el único vector de cointegración que puede existir entre ellas.

El tercer caso de convergencia —el menos estricto de todos— se da cuando $X_{i,t}$ y $X_{j,t}$ tienen tendencias diferentes, pero es posible establecer una relación funcional entre ambas tendencias. La representación formal de este caso sería la siguiente:

$$\lim_{k \rightarrow \infty} E(X_{i,t+k} - \hat{b}X_{j,t+k} / \hat{\nu}_t) = 0$$

siendo $\hat{a} > 0$

En este caso las series $X_{i,t}$ y $X_{j,t}$, también deben estar cointegradas pero no es preciso que la relación de cointegración sea la que refleja el vector $(1, -1)$, sino que viene dada por el vector $(1, -b)$, con $b > 0$. Estamos pues admitiendo que se corrigen las tendencias estocásticas de las series, pero que éstas afectan con distinta intensidad a cada ente territorial, y podemos saber cual es la diferencia existente entre esas series a largo plazo.

Este último caso admite interpretaciones muy distintas dependiendo de qué valor tome b (que sea mayor o menor que uno), de cuál sea el valor inicial de cada serie, y

de qué tipo de evolución (ascendente o descendente) siga cada una. Podría ocurrir que las series se alejaran entre sí a lo largo del tiempo; y también puede ocurrir que dos series con tendencias distintas se estén acercando durante un período determinado y al cabo de cierto tiempo acaben convergiendo (aunque, si esas tendencias distintas se mantienen, después de ese momento de convergencia, las dos series empezarían a divergir).

El análisis de convergencia a través de series posee algunas deficiencias; y la más importante de estas deficiencias tiene que ver con la posibilidad de que existan «saltos» o cambios estructurales en las series (Quah 1993). Estos cambios estructurales pueden falsear los resultados de los tests de raíces unitarias y hacer, por ejemplo, que aceptemos la presencia de raíces unitarias cuando éstas no existen. En tales casos puede ocurrir que se acepte, o que se rechace, erróneamente la hipótesis de convergencia⁴.

Desde el punto de vista empírico existen bastantes trabajos, algunos de ellos ya citados, donde se analizan variables como la renta per capita, el nivel de precios o la tasa de desempleo y donde se utiliza este tipo de metodología.

Para países europeos, y utilizando el PIB per cápita, Greasley y Oxley (1997) comprueban que solo existe convergencia si se analizan los países dos a dos, pero nunca de forma conjunta. En este mismo sentido Pallardó y Esteve (1997) se vuelven a centrar en los países europeos y solamente encuentran evidencia de acercamiento entre las rentas per cápita de Suecia y Dinamarca. Camarero y cols. (2001) examinan los precios y la inflación de algunos países europeos y rechazan la convergencia de largo plazo en todos los casos, aunque si encuentran *catching up* de algunos países con Alemania y con la media europea. Usando esta misma metodología para el ámbito regional, McGuinness y Sheehan (1998) no encuentran una clara evidencia de convergencia entre los ingresos per capita de las distintas regiones del Reino Unido.

Para el caso español Olloqui y cols. (2002) analizan la convergencia de los precios de las distintas provincias y solo encuentran evidencia de un acercamiento en seis de las cincuenta. Suriñach y cols. (1995) se centran en las Comunidades Autónomas y analizan la convergencia en las tasas de desempleo encontrando evidencia de acercamiento con el agregado nacional en cinco de ellas. Avilés y cols. (1997)

⁴ En este sentido, Rappoport y Reinlich (1989) y Perrón (1989) mantienen que los test de raíces unitarias están sesgados a favor de la aceptación de dichas raíces cuando aparecen cambios estructurales en las series estudiadas.

también se centran en el mercado de trabajo y aunque encuentran evidencia de acercamiento entre la población activa de Andalucía y España, ese acercamiento desaparece cuando se analizan las tasas de paro y ocupación. Finalmente Fernández-Jardón y Caneda (1999), tampoco encuentran evidencia de convergencia entre Galicia y España, cuando analizan variables como la producción o la población.

3. ANÁLISIS EMPÍRICO

Ya hemos dicho en el apartado anterior que con este método se pretende descubrir la existencia de relaciones de largo plazo entre las variables. Posteriormente serán esas relaciones las que nos indiquen si existe convergencia o divergencia entre las variables estudiadas.

Desde un punto de vista práctico, nuestro estudio se estructurará en varias fases. En primer lugar, se llevará a cabo el análisis de convergencia entre la tasa de paro de cada región y la media nacional, tomando como base la tasa de paro de toda la población. Posteriormente, se repetirá el análisis para grupos de población cada vez más homogéneos en términos de edad, sexo y nivel de estudios, intentando extraer las conclusiones pertinentes.

En lo referente a los datos también hemos de hacer algunas consideraciones previas. En primer lugar, la totalidad de los datos utilizados en este trabajo provienen de la Encuesta de Población Activa (*EPA*) proporcionada por el Instituto Nacional de Estadística (*INE*), y en segundo lugar, el período de análisis es el comprendido entre 1977 y 2001, y se van a emplear datos trimestrales.

El principal problema de los datos utilizados proviene de la representatividad cuando utilizamos muestras de tamaño reducido. Para solucionarlo hemos optado por centrar nuestro análisis en el grupo de varones mayores de 25 años, ya que dicho grupo es el más numeroso y proporciona datos suficientes. No obstante, en las regiones donde existían problemas estadísticos con los datos al concretar en los que poseen estudios superiores, se han «reconstruido» los valores ausentes de acuerdo con el procedimiento que describimos más adelante.

Por su parte, la tasa de paro de los distintos grupos, como variable fundamental de nuestro estudio, se ha elaborado a partir de las series de activos y ocupados de la *EPA*, todas ellas desagregadas por Comunidades Autónomas, sexo, grupo de edad y nivel de estudios.

En cuanto al estudio de la convergencia, esta se realiza entre cada Comunidad Autónoma y el correspondiente agregado nacional, entendiendo por agregado nacio-

nal al resto de regiones que componen España. Por lo tanto, a la hora de elaborar la tasa de paro del agregado nacional, se eliminan los datos de esa región y de este modo se evita la posible relación que pudiera existir entre la Comunidad Autónoma y España por el mero hecho de que la primera esté incluida dentro de la segunda.

Finalmente, conviene señalar que el criterio de elección de grupos cada vez más homogéneos, está relacionado, como se ha señalado previamente, con la disponibilidad de datos. El primer grupo estudiado es el conjunto de la población trabajadora. Después, nos hemos centrado en el grupo formado por los varones con más de 25 años; y la razón de esta selección es que se trata de un grupo suficientemente amplio, que admite desagregaciones posteriores sin plantear problemas serios de disponibilidad de datos. Finalmente, hemos centrado nuestro análisis en un subgrupo del anterior: los varones con más de 25 años que, además de reunir esas características, poseen estudios superiores. Este grupo, aunque no es demasiado amplio, presenta un comportamiento muy regular y esto facilita su tratamiento estadístico.

Nuestro análisis econométrico constará de dos fases: una primera, donde se verá si todas las series presentan una raíz unitaria; y una segunda, donde se procederá a estudiar si la diferencia existente entre la tasa de paro de cada región y el correspondiente agregado nacional sigue un proceso estacionario, y si dicho proceso está o no está condicionado por tendencias deterministas. De esta manera vamos a poder concluir si existe convergencia real, y también si ésta se debe a un proceso de *catching up*, o si, por el contrario, el equilibrio de largo plazo es el mismo para ambas series.

Una vez hechas estas consideraciones, el primer paso de nuestro trabajo debe de ser la elaboración de las series que vamos a utilizar. Las tasas de paro que utilizaremos se definen de la siguiente manera:

$$TP^i = \frac{P^i}{A^i}$$

$$TPE^i = \frac{PE - P^i}{AE - A^i}$$

donde TP^i es la tasa de paro de la región i calculada como cociente entre parados de la región i (P^i) y activos de esa región (A^i). Por su parte, TPE^i es la tasa de paro del agregado nacional correspondiente a esa región, y se define como el cociente entre el

total de parados de España (PE) menos los que aporta esa región (P^i), y el total de activos de España (AE) menos los correspondientes a la región i .

El siguiente paso, una vez definidas nuestras variables, consiste en determinar el grado de integración que tienen las series, y fundamentalmente, demostrar que las tasas de paro de cada región y las de su correspondiente agregado nacional presentan una raíz unitaria. De esta manera se puede establecer si existe o no existe una relación de cointegración entre ellas, relación que además será única.

Para determinar el número de raíces unitarias, vamos a utilizar dos métodos distintos: por un lado, la metodología propuesta por Dickey y Fuller (1979, 1981); y por otro lado, la recogida en los trabajos de Phillips (1987) y Phillips y Perron (1988). El primer método da lugar al test que se denomina *augmented Dickey-Fuller* (ADF), y el segundo método al test *Phillips-Perron* (PP).

Ya hemos comentado previamente que el principal problema que tienen los trabajos de convergencia realizados a través de técnicas de cointegración proviene del hecho de que puedan existir cambios estructurales en las series utilizadas. Estos cambios pueden llevar a conclusiones erróneas sobre el grado de integración de las series. Para intentar solucionar este problema hemos procurado homogeneizar las series, y con este fin se han eliminando tanto los posibles efectos estacionales que causa el estar trabajando con datos trimestrales, como los posibles cambios estructurales que pueda provocar en una serie la presencia de datos atípicos especialmente relevantes.

Existen varios métodos a la hora de homogeneizar las series. Uno de ellos consiste en filtrar la serie una vez que se han determinado los efectos estacionales y los datos atípicos que son especialmente representativos (Fernández-Jardón y Caneda, 1999). Existe una segunda opción, que es la que hemos utilizado en este trabajo, y que consiste en aplicar conjuntamente dos procedimientos distintos: en primer lugar se aplica el método *TRAMO* (*Time series Regression with ARIMA noise, Missing values, and Outliers*) que elimina el efecto de los datos atípicos más influyentes reconstruyendo además los ausentes, y en segundo lugar se aplica el método *SEATS* (*Signal Extraction in ARIMA Time Series*) que elimina la estacionalidad de las series. La aplicación de esos dos métodos se puede conseguir de forma simultánea a través del programa *TSW* (*TRAMO-SEATS para WINDOWS*) (Maravall y Gómez, 1993, 1994, 2001 y Maravall y cols. 1999).

Este segundo procedimiento corrige los efectos de series con periodicidad inferior a la anual y nos proporciona una serie desestacionalizada. Además, como ya hemos mencionado nos permite depurar las series eliminando el efecto de los «outliers» (datos atípicos) especialmente relevantes. Y finalmente, una tercera ventaja que nos proporciona el uso de este método, es que nos permite estimar el valor de obser-

vaciones perdidas o ausentes. De esta manera, en regiones donde el alto grado de desagregación aplicado nos impide tener una muestra fiable, este procedimiento nos permite disponer de un dato verosímil que podemos utilizar en el análisis empírico. Esta última aplicación es importante en las regiones con menor tamaño en términos poblacionales; fundamentalmente, La Rioja, Murcia, Navarra, Baleares y Extremadura.

Una vez que poseemos las series depuradas, ya podemos estudiar cual es su grado de integración. El resultado más relevante de la aplicación de estos tests nos permite comprobar que las series de tasas de paro utilizadas se pueden considerar integradas de orden 1⁵.

Después de haber homogeneizado las series y comprobado que todas ellas poseen una raíz unitaria, el siguiente paso consiste en investigar si existe una relación de cointegración entre ellas, y más concretamente, si esta relación viene dada por el vector $(1, -1)$. Para saber si esto es verdad, analizamos lo que ocurre con la diferencia existente entre la tasa de paro de cada región y la del correspondiente agregado nacional. Dependiendo de cómo sea esa diferencia (estacionaria o no estacionaria), podremos deducir si existe o no existe convergencia entre cada región y su agregado nacional. Además, si esa diferencia resulta ser estacionaria, sin término independiente ni tendencia, podremos afirmar que existe convergencia a largo plazo y que la Comunidad Autónoma y el agregado nacional tienden a un mismo estado estacionario. Sin embargo, si aparecen tendencias deterministas, la convergencia será del tipo «catching-up» y podremos tener estados estacionarios diferentes.

El análisis de cointegración se realiza en tres fases distintas, a lo largo de las cuales intentamos conseguir grupos poblacionales cada vez más homogéneos. En primer lugar, se estudia la existencia de relaciones de largo plazo entre las tasas de paro de cada Comunidad Autónoma con su agregado correspondiente, referidas dichas tasas a la totalidad de la población. En segundo lugar, se repite el mismo proceso pero utilizando las tasas de paro de los varones mayores de 25 años; finalmente, se buscará un grupo aún más homogéneo, y se vuelven a estudiar las relaciones de largo plazo entre cada región y el agregado nacional, utilizando las tasas de paro de los varones mayores de 25 años con estudios superiores.

Para todas las regiones se han aplicado los dos tests a los que ya nos hemos referido (*PP* y *ADF*) con el fin de obtener resultados más concluyentes. En este epígrafe únicamente presentaremos una tabla —Tabla 1— que nos resumirá los resultados

⁵ Para consultar los resultados concretos, tanto de la homogeneización como de los test de raíces unitarias no incluidos en el apéndice, pueden ponerse en contacto con el autor.

obtenidos y que nos proporcionará un soporte visual a la hora de aportar conclusiones sobre la existencia de convergencia real entre las tasas de paro de las regiones españolas. No obstante, en las tablas A1, A2 y A3 del apéndice, aparecen recogidos los resultados de todos los tests de raíces unitarias aplicados a la diferencia de variables, y que son la base del análisis de cointegración.

Tabla 1. Comunidades Autónomas donde se cumple que la diferencia con el agregado nacional es estacionaria

	CONVERGENCIA A LARGO PLAZO	CATCHING-UP
CONJUNTO DE LA POBLACION	MADRID, MURCIA.	ANDALUCIA, ARAGON, VALENCIA, LA RIOJA.
VARONES MAYORES DE 25 AÑOS	MURCIA	ANDALUCIA, ASTURIAS, BALEARES, CANTABRIA, VALENCIA, PAIS VASCO, LA RIOJA.
VARONES MAYORES DE 25 AÑOS CON ESTUDIOS SUPERIORES	ANDALUCIA, EXTREMADURA, MADRID, MURCIA.	ARAGON, ASTURIAS, CANARIAS, CANTABRIA, CASTILLA Y LEON, CASTILLA LA MANCHA, CATALUÑA, VALENCIA, GALICIA, PAIS VASCO, NAVARRA, LA RIOJA.

Fuente: Elaboración propia a partir de datos EPA.

El resultado que cabe esperar de todo este análisis sería el aumento progresivo del número de regiones que convergen con el agregado nacional, según vamos avanzando en el grado de homogeneidad del grupo poblacional estudiado. Este resultado estaría en concordancia con el hecho de que cuanto mejor definido esté el grupo estudiado menos son las fuentes de divergencia existentes a nivel regional.

Los resultados recogidos en la TABLA 1 parecen estar en concordancia con nuestra hipótesis inicial: el grado de convergencia aumenta efectivamente a medida que el grupo poblacional se define mejor y se hace más homogéneo. Al menos esto es lo que parece indicar el aumento experimentado por el número de Comunidades Autónomas que presentan un resultado estacionario al ser comparadas con su correspondiente agregado nacional, aumento que se observa claramente según nos vamos centrando en un mercado laboral cada vez más homogéneo en términos de sexo, edad y nivel de estudios.

Profundizando un poco más en los resultados obtenidos, y haciendo un estudio separado para cada grupo de población, podemos extraer algunas conclusiones adicionales.

En primer lugar, y refiriéndonos a la tasa específica de paro de toda la población, podemos apreciar que solo seis regiones presentan un resultado estacionario al compararlas con el agregado nacional (Andalucía, Aragón, Madrid, Murcia, La Rioja y Valencia). Además, Madrid solo muestra estacionariedad cuando bajamos el nivel de exigencia y Andalucía solo lo cumple para uno de los test aplicados.

En cuanto al tipo de convergencia, podemos apreciar que solamente en dos regiones se produce convergencia a largo plazo (Murcia y Madrid). El resto alcanzan una convergencia del tipo *catching-up*, lo cual indica la existencia de componentes deterministas, ya sean tendencias o términos independientes. Estos componentes, aunque no impiden la existencia de un cierto equilibrio en la evolución seguida por las series, sí impiden que se alcance la convergencia completa de las tasas de desempleo. Esto estaría indicando la existencia de más de un estado estacionario en la evolución seguida por las tasas de paro regionales.

Pasemos ahora al análisis de la convergencia de las tasas específicas de paro de los varones mayores de 25 años. En este caso se produce un ligero aumento del número de regiones que presentan un cierto grado de convergencia (Andalucía, Asturias, Baleares, Cantabria, Murcia, País Vasco, La Rioja y Valencia), aunque algunas de ellas, como ya ocurría en el caso anterior, solo cumplen la condición de estacionariedad para uno de los dos test de raíces unitarias aplicados, y otras vuelven a exigir un menor nivel de confianza.

Resulta un poco sorprendente el hecho de que dos regiones que antes convergían (Aragón y Madrid) no sigan haciéndolo ahora. No obstante, este resultado deja de parecer tan anormal si se tiene en cuenta que el grupo de los mayores de 25 años sigue siendo un grupo muy amplio y por lo tanto muy heterogéneo. Al agrupar a la gente por edades de esta manera podrían aflorar discrepancias que antes permanecían ocultas o diluidas en el conjunto de la población.

En cuanto al grado de convergencia, vuelve a ser Murcia la única Comunidad Autónoma que mantiene el equilibrio a largo plazo sin término independiente ni tendencia (recordemos que Madrid ya no cumple la condición de convergencia). El resto sigue manteniendo componentes deterministas que impiden la convergencia de las tasas de paro de todas las regiones hacia un único estado estacionario.

El paso final que damos con este estudio se centra en un grupo mucho más definido: el de los varones mayores de 25 años con estudios superiores. En este caso, la situación ya es mucho más clara, y el aumento de regiones que cumplen la condición de estacionariedad es más que considerable. Ahora, la única Comunidad Autónoma que no presenta relación de equilibrio con su correspondiente agregado nacional es Baleares. Este resultado, que podría sorprendernos si tenemos en cuenta que Balea-

res aparecía como estacionaria en el caso anterior, no es demasiado extraño después de todo: los datos de Baleares presentaban graves problemas de muestreo y han tenido que ser reconstruidos a través del programa TSW. Aparte de Baleares, la única Comunidad que presenta problemas es el País Vasco. La condición de estacionariedad solo se cumple para uno de los tests, y para esto debemos reducir el nivel de significación del test. No obstante, este resultado es coherente con la evolución seguida por esta región a lo largo de nuestro estudio.

Finalmente, habría que destacar que la convergencia estricta a largo plazo solo se da en cuatro regiones —Andalucía, Extremadura, Madrid y Murcia—. En el resto se produce convergencia del tipo *catching-up*; es decir, el equilibrio a largo plazo de cada región esta condicionado por la existencia de términos independientes o tendencias. Este resultado pone de manifiesto que no existe un único estado estacionario al que converjan las tasas de paro de todas las Comunidades Autónomas españolas, incluso en el caso de grupos que eliminan la distorsión procedente de variables poblacionales como la edad, el sexo o el nivel de estudios.

6. CONCLUSIONES

Con este trabajo se ha intentado corroborar desde el punto de vista econométrico la idea de que, aunque la convergencia real entre distintas áreas geográficas se presume bastante complicada, cuanto más homogéneo es el grupo de población estudiado, tanto más probable resulta la convergencia entre ellas. No obstante, este resultado puede verse condicionado por las características especiales del grupo seleccionado y debería ponerse en cuarentena en tanto no se realicen estudios comparativos.

Con este fin, y tras aplicar las correspondientes técnicas de cointegración sobre las distintas tasas de desempleo, los resultados nos permiten obtener las siguientes conclusiones:

- Por un lado, podemos confirmar la ausencia de convergencia absoluta entre las regiones; cuando usamos toda la población. Son pocas las regiones que presentan cointegración, y además es necesario introducir componentes deterministas (es decir, aspectos diferenciadores) en la mayoría de las regiones para admitir la convergencia.
- Por otro lado, se corrobora que las características de la población son sumamente importantes en este tipo de análisis; según vamos concretando el grupo en términos de edad, sexo y nivel de estudios, se aprecia un incremento considerable en número de regiones que convergen con su agregado nacional correspondiente.

7. REFERENCIAS BIBLIOGRÁFICAS

- AVILÉS, S.; GÁMEZ, C. y TORRES, J. L. (1997), «La convergencia real de Andalucía: Un análisis de cointegración del mercado de trabajo», *Revista de Estudios Regionales*, 47, 15-36.
- BARRO, R. J. y SALA I MARTÍN, X. (1990), «Economic Growth and Convergence across the United States», *NBER Working Paper*, 3419.
- (1991), «Convergence across States and Regions», *Brooking Papers on Economic Activity*, 1, 107-182.
- (1992), «Convergence», *Journal of Political Economy*, 2, 223-251.
- BERNARD, A. B. y DURLAUF, S. N. (1996), «Interpreting test of the convergence hipótesis», *Journal of Econometrics*, vol. 71, 161-173.
- CAMARERO, M.; ESTEVE, V. y TAMARIT, C. (1995), «Medición de la convergencia en tasas de inflación: España versus Alemania y el SME», *Papeles de Economía Española*, 63, 62-76.
- DICKEY, D. A. y FULLER, W. A. (1979), «Distribution of the Estimators for Autoregressive Time Series with a Unit Root», *Journal of American Statistical Association*, 74, 427-431.
- (1981), «Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root», *Econometrica*, 49, 1057-1072.
- ENGLÉ, R. y GRANGER, C. W. J. (1987), «Cointegration and error correction: representation, estimation and testing», *Econometrica*, 55, 251-276.
- FERNÁNDEZ-JARDÓN, C. M. y CANEDA, A. (1999), *El nivel educativo de los trabajadores como factor de crecimiento económico: Un análisis comparativo entre Galicia y España*, Tesis doctoral.
- GREASLEY, D. y OXLEY, L. (1997), «Time-series based test of convergence hypothesis: Some positive results», *Economic letters*, 56, 143-147.
- MARAVALL, A. y GÓMEZ, V. (1993), Initializing the Kalman Filter with incompletely specified initial conditions in approximate Kalman Filter (Series on Approximation and Decomposition), Chen, R. (ed), London: World Scientific Publ. Co.
- (1994), «Estimation, prediction and interpolation for nonstationary series with the Kalman Filter», *Journal of the American Statistical Association*, 89, 611-624.
- (2001), *Automatic modeling methods for univariate series*, en D. Peña, G.C. Tiao y R.S. Tsay (ed.), *A Course in Time Series Analysis* New York, Wiley, J. and Sons Ch. 7, 170-201.
- MARAVALL, A.; GÓMEZ, V. y PEÑA, D. (1999), «Missing observations in ARIMA models: Skipping Approach versus Additive Outlier Approach», *Journal of Econometrics*, 88, 341-363.
- MCGUINNESS, S. y SHEEHAN, M. (1998), «Regional convergence in U. K.», *Applied Economic Letters*, 5, 653-658.
- OLLOQUI, I.; SOSVILLA, S. y ALONSO, J. (2002), «Convergencia en precios en las provincias españolas», *ICE Tribuna de Economía*, 797, febrero, 160-178.
- PALLARDÓ, V. J. y ESTEVE, V. (1997), «Convergencia real en la Unión Europea», *Revista de Eco-*

- nomía Aplicada*, 14 (vol. V), 25-49.
- PERRON, P. (1989), «The Great Crash, the Oil Price Shock, and the Unit Roots Hypothesis», *Econometrica*, 57, 1361- 1401.
- PHILLIPS, P. C. B. (1986), «Understanding Spurious Regressions in Econometrics», *Journal of Econometrics*, 33, 311-340.
- PHILLIPS, P. C. B. y PERRON, P. (1988), «Testing for a Unit Root in Time Series Regression». *Biometrika*, 75, 335-346.
- QUAH, D.(1993), «Galton's Fallacy and Test of the Convergence Hypothesis». *Scandinavian Journal of Economics*, vol. 95, núm. 4, págs. 427-443.
- RAPPOPORT, P. y RICHLIN, L. (1989), «Segmented Trends and Non-Stationary Time Series». *The Economic Journal*, 99, 168-177.
- SURIÑACH, J.; ARTÍS, M.; LÓPEZ, E. y SANSÓ, A. (1995), Análisis económico regional. Nociones básicas de la teoría de la cointegración, Barcelona: Antoni Bosch Editor.
- YULE, G. U. (1926), «Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series», *Journal of the Royal Statistical Society*, 89, 1-64.

APÉNDICES

Apéndice 1. Tasas de paro de toda la población.
Test de raíces unitarias sobre diferencias entre variables

	ADF			P-P		
	NIVELES		DIFERENCIAS	NIVELES		DIFERENCIAS
ANDALUCÍA	-2.2394 (-2.8915)		-12.5132 (-1.9434)	-2.5413* (-2.8912)	C	—
ARAGÓN	-4.3177 (-3.4561)	C T	—	-4.3727 (-3.4561)	C T	—
ASTURIAS	-2.0172 (-3.4561)		-9.5639 (-1.9434)	-2.0836 (-3.4561)		-9.5600 (-1.9434)
BALEARES	-1.7878 (-2.8918)		-4.9274 (-3.4576)	-2.4587 (-2.8912)		-11.1598 (-3.4566)
CANARIAS	-1.2112 (-1.9434)		-4.9184 (-1.9435)	-1.1411 (-1.9434)		-11.0948 (-1.9434)
CANTABRIA	-2.4954 (-2.8912)		-6.5562 (-1.9435)	-2.5287 (-2.8912)		-10.4411 (-1.9434)
CASTILLA-LEÓN	-1.8876 (-3.4561)		-10.1458 (-1.9434)	-1.9330 (-3.4561)		-10.1404 (-1.9434)
CASTILLA-LA MANCHA	-2.0092 (-2.8912)		-10.9515 (-1.9434)	-1.9911 (-2.8912)		-10.9436 (-1.9434)
CATALUÑA	-2.2749 (-3.4571)		-4.4060 (-1.9435)	-1.8711 (-3.4561)		-7.8769 (-1.9434)
EXTREMADURA	-0.9276 (-1.9434)		-13.2742 (-1.9434)	-2.9980 (-3.4561)		-13.1345 (-1.9434)
GALICIA	-1.9327 (-3.4571)		-5.2464 (-3.4571)	-1.7253 (-3.4561)		-9.7226 (-3.4566)
MADRID	-1.6663* (-1.9434)		—	-1.7746* (-1.9434)		—
MURCIA	-3.4355 (-1.9434)		—	-3.3863 (-1.9434)		—
NAVARRA	0.2794 (-1.9434)		-11.5900 (-1.9434)	-2.0577 (-3.4561)		-11.5547 (-1.9434)
PAÍS VASCO	-1.9410 (-3.4566)		-12.1187 (-3.4566)	-1.9451 (-3.4561)		-12.2704 (-3.4566)
RIOJA	-3.4208 (-2.8918)	C	—	-3.1320 (-2.8912)	C	—
VALENCIA	-3.1315 (-2.8912)	C	—	-2.9787 (-2.8912)	C	—

Entre paréntesis aparecen los valores críticos tabulados por McKinnon para un nivel de significación del 5%. C y T, indican la existencia de término independiente o tendencia respectivamente.

*Se admite la existencia de estacionariedad para un nivel de significación del 10%.

Fuente: Elaboración propia a partir de datos del INE.

Apéndice 2. Tasas de paro de los varones mayores de 25 años.
Test de raíces unitarias sobre diferencias entre variables

	ADF			P-P		
	NIVELES		DIFERENCIAS	NIVELES		DIFERENCIAS
ANDALUCÍA	-3.1320* (-2.8912)	C	—	-2.6463* (-2.8912)	C	—
ARAGÓN	-0.1259 (-3.4561)		-10.4811 (-3.4566)	-0.1626 (-3.4561)		-10.4519 (-3.4566)
ASTURIAS	-3.2392* (-3.4561)	C T	—	-3.0087 (-3.4561)		-12.7363 (-1.9434)
BALEARES	-3.3195 (-2.8912)	C	—	-3.2012 (-2.8912)	C	—
CANARIAS	-1.2112 (-1.9434)		-11.0692 (-1.9434)	-1.1411 (-1.9434)		-11.0948 (-1.9434)
CANTABRIA	-2.6721* (-2.8912)	C	—	-2.5866* (-2.8912)	C	—
CASTILLA-LEÓN	-0.4484 (-1.9434)		-8.9328 (-1.9434)	-0.50156 (-1.9434)		-8.9383 (-1.9434)
CASTILLA-LA MANCHA	-2.1989 (-2.8912)		-9.7675 (-3.4566)	-2.2015 (-2.8912)		-9.7809 (-3.4566)
CATALUÑA	-2.0238 (-3.4561)		-8.5703 (-1.9434)	-0.3492 (-1.9434)		-8.5066 (-1.9434)
EXTREMADURA	-0.4586 (-1.9434)		-9.6377 (-1.9434)	-0.5056 (-1.9434)		-9.6428 (-1.9434)
GALICIA	-2.3011 (-3.4566)		-12.8804 (-1.9434)	-2.7905 (-3.4561)		-12.6484 (-1.9434)
MADRID	-1.1288 (-1.9434)		-12.4439 (-1.9434)	-2.0353 (-2.8912)		-12.5548 (-1.9434)
MURCIA	-2.5492 (-1.9434)		—	-2.4789 (-1.9434)		—
NAVARRA	-1.4760 (-2.8912)		-10.6678 (-1.9434)	-1.47008 (-2.8912)		-10.6350 (-1.9434)
PAÍS VASCO	-2.2929 (-2.8918)		-9.0671 (-1.9435)	-3.2091* (-3.4561)	C T	—
RIOJA	-3.3572 (-2.8912)	C	—	-3.6283 (-3.4561)	C T	—
VALENCIA	-2.5730* (-2.8912)	C	—	-2.4785 (-2.8912)		-11.5744 (-1.9434)

Entre paréntesis aparecen los valores críticos tabulados por McKinnon para un nivel de significación del 5%. C y T, indican la existencia de término independiente o tendencia respectivamente.

*Se admite la existencia de estacionariedad para un nivel de significación del 10%.

Fuente: Elaboración propia a partir de datos del INE.

Apéndice 3. Tasas de paro de varones mayores de 25 años con estudios superiores.
Test de raíces unitarias sobre diferencias entre variables

	ADF			P-P		
	NIVELES		DIFERENCIAS	NIVELES		DIFERENCIAS
ANDALUCÍA	-2.0809 (-1.9434)		—	-2.7150 (-1.9434)		—
ARAGÓN	-4.6291 (-3.4561)	C T	—	-4.6657 (-3.4561)	C T	—
ASTURIAS	-5.7702 (-3.4561)	C T	—	-5.7584 (-3.4561)	C T	—
BALEARES	-2.9265 (-3.4561)		-6.7393 (-1.9435)	-2.9024 (-3.4561)		-10.2073 (-1.9434)
CANARIAS	-5.3437 (-2.8912)	C	—	-5.3966 (-2.8912)	C	—
CANTABRIA	-3.2876* (-3.4566)	C T	—	-4.2915 (-3.4561)	C T	—
CASTILLA-LEÓN	-4.5781 (-2.8912)	C	—	-4.5019 (-2.8912)	C	—
CASTILLA-LA MANCHA	-3.6981 (-2.8912)	C	—	-3.9094 (-2.8912)	C	—
CATALUÑA	-3.3595 (-2.8912)	C	—	-3.2727 (-2.8912)	C	—
EXTREMADURA	-3.2573 (-1.9434)		—	-5.5411 (-2.8912)	C	—
GALICIA	-5.1266 (-3.4561)	C T	—	-5.1695 (-3.4561)	C T	—
MADRID	-2.4194 (-1.9435)		—	-3.7476 (-1.9434)		—
MURCIA	-4.7169 (-1.9435)		—	-5.0145 (-1.9434)		—
NAVARRA	-5.2086 (-3.4561)	C T	—	-5.2803 (-3.4561)	C T	—
PAÍS VASCO	-2.3148 (-2.8915)		-12.6276 (-1.9434)	-2.8440* (-2.8912)	C	—
RIOJA	-3.7541 (-2.8912)	C	—	-3.6449 (-2.8912)	C	—
VALENCIA	-4.8255 (-2.8912)	C	—	-4.9074 (-2.8912)	C	—

Entre paréntesis aparecen los valores críticos tabulados por McKinnon para un nivel de significación del 5%. C y T, indican la existencia de término independiente o tendencia respectivamente.

*Se admite la existencia de estacionariedad para un nivel de significación del 10%.

Fuente: Elaboración propia a partir de datos del INE.