Divergencias regionales en el mercado de trabajo español

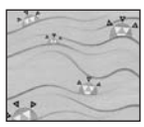
Un analisis de descomposición de la varianza

*Alfonso Moral de Blas**

Son muchos los estudios realizados sobre la convergencia en términos de renta per capita, pero pocos los centrados en otro tipo de variables. En este trabajo hemos optado por analizar la convergencia de las tasas de paro, que es una cuestión muy importante en el caso español, y que ha recibido mucha menos atención dentro de la literatura económica. En concreto, en esta tesis trataremos de ver si tiene algún sentido hablar de convergencia entre regiones en términos de tasas de paro, a pesar de que, a simple vista, la evidencia parece apuntar en sentido contrario. Para ello recurrimos al análisis econométrico basado en la descomposición de la varianza de las tasas de paro regionales.

Palabras clave: convergencia económica, mercado de trabajo, tasa de desempleo, varianza.

Clasificación JEL: R23.



1. Introducción

La cuestión de la convergencia real no tiene por qué circunscribirse únicamente 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. Esto es lo que planteamos en este trabajo, nos preguntamos si los procesos recientes de globalización de la economía han dado lugar a un mayor grado de homogeneidad (si ha habido o no ha habido convergencia real) entre las regiones españolas.

El grueso de la literatura dedicada al análisis de convergencia real se ha centrado en el estudio de las rentas per capita, olvidándose de otras variables como podrían ser los salarios o las tasas de desempleo. Este hecho, junto con la peculiar configuración institucional del mercado de trabajo español, nos ha llevado a centrarnos en un aspecto particular de la convergencia entre regiones: la convergencia de las tasas de paro regionales.

El resto del artículo se organiza de la siguiente manera. En el apartado 2 revisa las distintas alternativas utilizadas en la literatura de convergencia y se comentan los resultados obtenidos para España. El

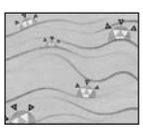
* Departamento de Fundamentos del Análisis Económico, Facultad de Ciencias Sociales Jurídicas y de la Comunicación, Universidad de Valladolid.

apartado 3 se dedica al análisis de la metodología utilizada en el trabajo. En el apartado 4 aparece el análisis empírico realizado y los resultados obtenidos, Finalmente, en el apartado 5 recogemos las principales conclusiones de este trabajo.

2. Análisis de convergencia

Son cuatro los métodos más frecuentemente utilizados en el marco del análisis de convergencia: los dos inspirados en la noción de β-convergencia (β-convergencia absoluta y β-convergencia condicionada); el método conocido como σ-convergencia; y finalmente el método (o métodos) de cointegración.

La β-convergencia se basa en el análisis de la relación existente entre la tasa de crecimiento de una variable y el nivel que esta presenta. En el caso español, hay que señalar la existencia de trabajos que contrastan la hipótesis β-convergencia absoluta y condicionada tanto a nivel provincial como en el caso de las Comunidades Autónomas. Dolado *et al.* (1994), Mas *et al.* (1994) y Villaverde (1999) encuentran evidencia de esta convergencia, y demuestran que esta es más acelerada en el caso de la β-convergencia condicionada.



COLABORACIONES

Por su parte la σ-convergencia analiza la dispersión existente en las variables estudiadas, Raymond y García (1994), García. Raymond y Villaverde (1995) y Villaverde (1999) encuentran resultados muy similares tanto para las Comunidades Autónomas como para las provincias españolas, y concluyen que existe convergencia, aunque ésta se ve frenada a partir de 1980.

La cointegración tiene su fundamento en el análisis de las relaciones de largo plazo entre las variables y en el estudio del grado de integración que presentan las series, Olloqui *et al.* (2002) analizan la convergencia de los precios de las distintas provincias y sólo encuentran evidencia de un acercamiento en seis de las cincuenta, Suriñach *et al.* (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 *et al.* (1997) también se centran en el mercado de trabajo y aunque encuentra 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. Descomposición de la varianza

Aunque este método se ha utilizado con profusión en multitud de estudios econométricos, parece ser que aún no se han explotado sus posibilidades en relación con el tema de la convergencia. Para investigar la existencia o inexistencia de convergencia en términos de renta per capita, tasa de desempleo o cualquier otra variable entre varias áreas geográficas el proceso sería muy simple. Habría que calcular qué porcentaje de la varianza de la variable en cuestión se explica a través del componente territorial, y cómo evoluciona ese porcentaje a lo largo del tiempo.

La descomposición de la varianza se puede realizar a través de métodos distintos, dos de los cuales pasamos a describir a continuación. En el primero de ellos el porcentaje de la varianza explicado por una variable cualquiera (o conjunto de variables) es el que resulta de calcular el porcentaje en que se reduce el

término residual de la regresión cuando se añade una (o varias) variable(s) explicativa(s) a un conjunto dado de regresores (1).

El segundo método utilizado para realizar la descomposición de la varianza —que es el que vamos a utilizar en este trabajo— emplea para su cálculo, en lugar del término residual, el valor del coeficiente de determinación de la regresión (2). Este método permite calcular dos efectos distintos: un *efecto total* y un *efecto marginal,* que pasamos a explicar a continuación, y que son los que determinan el porcentaje de la varianza explicado por cada variable (o variables).

El efecto total coincide con el valor que toma el estadístico R^2 (multiplicado por cien), cuando solamente se tiene en cuenta una variable explicativa (o un conjunto de variables), además del término independiente, y se ignoran todos los demás regresores del modelo.

El *efecto marginal* se mide a través del incremento porcentual del coeficiente de determinación de la regresión cuando se añade una variable explicativa (o un conjunto de variables) al conjunto formado por el resto de regresores del modelo. Ese conjunto de variables añadidas son las que se utilizan para reflejar la cualidad estudiada en ese momento.

Una vez calculados estos dos efectos, el porcentaje de la varianza que explica cada variable (o variables) se situará dentro del intervalo que marcan estos dos efectos.

(1) Como ejemplo de este tipo de procedimiento ponemos el *OCDE* employment outlook, del año 2000. En este informe se analiza la composición de la varianza del desempleo en función de factores regionales, poblacionales o de nivel de estudios.

(2) Este procedimiento se puede encontrar en algunos trabajos como el de Simón Pérez (2001), donde se analizan los factores que influyen en la determinación de los salarios pactados en la negociación colectiva de ámbito sectorial en España,

4. Análisis empírico

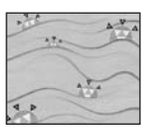
La totalidad de los datos utilizados en este trabajo provienen de la Encuesta de Población Activa *(EPA)* y de la base de datos *TEMPUS,* ambas proporcionadas por el Instituto Nacional de Estadística *(INE).* La tasa de paro de los distintos grupos, como variable fundamental de nuestro estudio, se ha 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, Sin embargo, la serie del Producto Interior Bruto *(PIB),* proviene de la contabilidad regional que proporciona la base de datos *TEMPUS.*

En conjunto, este apartado tiene dos partes. Una primera donde se analiza las variaciones que sufre el porcentaje de la varianza explicado por el componente regional según avanzamos en el grado de desagregación de la tasa de desempleo. Una segunda donde se analiza la evolución temporal de ese porcentaje, y que sería la que realmente nos proporcionaría evidencia sobre la convergencia.

Todas las regresiones se realizarán mediante la estimación por mínimos cuadrados ordinarios de funciones del siguiente tipo:

$$TP = \Phi(C, Y, R, T, D)$$

donde *TP* es la tasa de paro de un conjunto de individuos clasificados según edad, estudios, sexo, región y año; *C* es el término independiente; *Y* es el Producto Interior Bruto de cada región en cada uno de los años considerados; *R* es un conjunto de nueve variables ficticias que recogen el efecto de estar en una u otra región; *T* es otro conjunto de veinte variables ficticias, una por cada año estudiado, que reflejan la influencia del componente temporal; y *D* es un tercer conjunto de variables



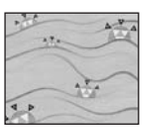
CUADRO 1
...

Efectos		R^2	VARIABLES					
			Años	Regiones	Sexo	Edad	Estudios	PIB
A	Total	0,88	32,73	55,48	—	—	—	10,17
	Marginal		31,42	47,81	—	—	—	0,26
B	Total	0,83	17,23	26,73	37,82	—	—	3,75
	Marginal		17,32	24,92	37,82	—	—	0,70
C	Total	0,86	8,09	8,88	7,51	60,93	—	0,90
	Marginal		8,50	8,72	7,51	60,93	—	0,45
D	Total	0,80	7,86	7,52	6,99	56,83	0,35	1,05
	Marginal		7,92	7,16	7,00	56,86	0,35	0,25

Fuente: Elaboración propia a partir de datos del INE.

cualitativas que recogen características tales como el sexo, la edad o el nivel de estudios.

La variable *R* solo refleja la existencia de nueve regiones debido a un problema de datos, La desagregación de los datos a ciertos niveles no es fiable; eso se debe a que las muestras disponibles son demasiado pequeñas, por lo que no resultan suficientemente significativas. Este problema es particularmente grave en las regiones de escasa población, por lo que dichas regiones han sido excluidas del análisis. A pesar de todo, podemos considerar que las regiones incluidas en este apartado (Andalucía, Aragón, Castilla y León, Castilla la Mancha, Cataluña, Comunidad Valenciana, Galicia, Comunidad de Madrid y País Vasco) son lo suficientemente representativas como para dar validez al estudio realizado.



COLABORACIONES

Por su parte, el vector *D* está formado por las tres variables siguientes: *sexo,* que distingue entre hombres y mujeres; *edad,* que diferencia entre mayores de 25 años y menores de esa edad; y *estudios,* que establece la distinción entre individuos con estudios medios o superiores e individuos por debajo de ese nivel.

La primera regresión que vamos a realizar incluye únicamente, como variables explicativas de la tasa de paro, las que reflejan el componente regional, las referidas al aspecto temporal y el *PIB* de cada región. A partir de aquí vamos introduciendo nuevas variables con el fin de obtener grupos cada vez mejor definidos. Así la segunda regresión incluye, además de las anteriores, la variable *sexo.* La tercera añade la variable *edad,* y la cuarta introduce el nivel de formación a través de la variable *estudios.* Estas regresiones las vamos a denominar A, B, C y D respectivamente.

Hay que observar que cada una de estas regresiones exige una redefinición de la tasa de paro que cada vez se va desagregando más, además de modificar el número de datos empleados. En la primera regresión, la tasa de paro es simplemente la media de cada región, sin embargo. En la segunda regresión ya tenemos tasas de desempleo masculinas y femeninas. En la tercera tenemos tasas masculinas y femeninas diferenciadas por edad, y en la cuarta, la diferenciación se extiende también al nivel de estudios.

Los resultados de estas regresiones aparecen recogidos en el Cuadro 1. En ese cuadro aparecen las cuatro regresiones. El coeficiente de determinación de cada una de ellas, y los efectos total y marginal de cada grupo de variables que se han introducido en las regresiones.

A la vista de los resultados obtenidos podemos destacar, en primer lugar los altos niveles de significación de los modelos, todos con niveles de significación superiores al 80 por 100. Esto esta indicándonos la adecuación de la especificación

aplicada a la hora de explicar los valores que presenta la tasa de desempleo.

En segundo lugar, y en relación con la hipótesis de la convergencia podemos apreciar cómo va disminuyendo la importancia del componente regional según vamos centrándonos en grupos poblacionales cada vez mejor definidos. Estos resultados vienen a corroborar nuestra idea inicial de que la convergencia regional es mayor cuando nos centramos en el estudio de grupos de población más agregados. No obstante, también podemos apreciar que sigue existiendo un porcentaje de la varianza explicado por las variables regionales (un 7 por 100 en el último modelo) que impide la completa igualación de las tasas de paro regionales para grupos de población homogéneos. Este residuo puede deberse, simplemente, al hecho de que las variables que hemos utilizado como controles de las características de la fuerza de trabajo *(edad, sexo, estudios)* son medidas inevitablemente toscas y no captan demasiado bien toda la heterogeneidad de la mano de obra.

De todos modos, se puede apreciar una considerable disminución del porcentaje de la varianza que explican las variables regionales según nos vamos centrando en grupos de población cada vez menos desagregados. Así, pasamos de un porcentaje algo superior al 50 por 100 en el caso de menor desagregación a otro en torno al 25 por 100 cuando se desagrega en función de la edad, porcentaje que luego se reduce hasta un 8-9 por 100 cuando se desagrega en función de la edad y el sexo, y hasta un 7-8 por 100 cuando la desagregación se hace máxima. Es decir, cuando se introducen los estudios como fuente de variación.

El análisis precedente tiene un inconveniente en la medida en que da la impresión de que los efectos regionales son de magnitud fija a lo largo del tiempo. Con el fin de ver si dichos efectos varían con los años, vamos a proceder de la siguiente manera. En primer lugar, tomaremos como punto de partida el último modelo, que es el más desagregado y también el que incluye un mayor número de datos, y a partir de él, fraccionaremos la regresión en veinte regresiones distintas, una para cada año de la muestra.

Vamos a disponer pues de veinte estimaciones, cada una de ellas con 72 observaciones. En las que se regresará la tasa de paro contra todas las variables regionales, el *sexo,* la *edad* y el nivel de *estudios,* Nuestro propósito no es otro que calcular el porcentaje de la varianza que representa el componente regional en cada una de ellas, y ver si dicho porcentaje aumenta o disminuye a lo largo del tiempo.

Para calcular el porcentaje en cuestión seguiremos el mismo camino de antes; es decir, calcularemos los efectos total y marginal, cuyos significados ya conocemos, Los resultados de estas nuevas estimaciones aparecen recogidos en el Gráfico 1.

A la vista de los resultados expuestos en este gráfico, podemos apreciar que, tras empezar con un nivel muy bajo en los primeros años de la muestra. El porcentaje explicado por el componente regional sufre un aumento progresivo y alcanza su máximo a principios de los 90. Posteriormente, sigue una senda descendente, que se ve truncada en los últimos años de la muestra.

Desde el punto de vista de la convergencia podemos decir que tras un primer periodo de divergencia, que se produce durante los ochenta, durante la década de los noventa sí que se produce una cierta convergencia o acercamiento de las tasas de desempleo regionales, Dicho acercamiento se refleja en la disminución que sufre el componente regional durante este periodo, y se ve bruscamente truncado en los últimos tres años de la muestra.

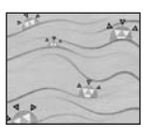
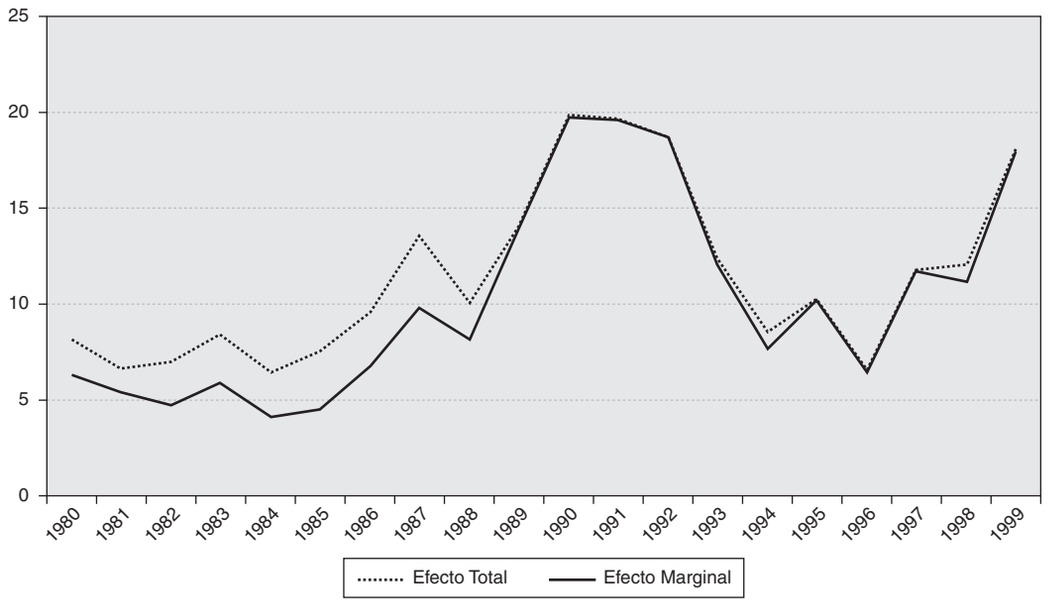


GRÁFICO 1
EVOLUCIÓN DE LOS EFECTOS TOTAL Y MARGINAL ENTRE LOS QUE VARÍA LA IMPORTANCIA DEL COMPONENTE REGIONAL



Fuente: Elaboración propia de datos de INE.

Si podemos apreciar la clara influencia que tiene el ciclo económico en dicha evolución. En este sentido, podemos concluir que la importancia del componente regional aumenta en las fases expansivas del ciclo económico que se producen a finales de los 80 y de los 90, y disminuye en las recesivas de principios de las dos décadas. Este resultado, sin duda, se relaciona con el comportamiento más dinámico de las regiones con menos desempleo. Esto les hace crecer más rápido en las expansiones, pero también decrecer más en las recesiones.

5. Conclusiones

El primer gran resultado es la ausencia de un claro proceso de convergencia, lo cual se concreta en dos resultados: por un lado, el porcentaje de la varianza explicado por el componente regional alcanza valores relativamente altos; y, por otro lado, dicho porcentaje no tiene una clara evolución decreciente, como sería de esperar si la hipótesis de la convergencia fuese cierta.

También se observa la gran importancia del sexo, la edad y el nivel de estudios, y se comprueba que cuando se tienen en cuenta dichas características de la población el componente regional pierde importancia como factor explicativo del desempleo regional.

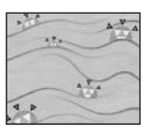
Se puede concluir, por lo tanto que no existe un proceso de convergencia real generalizado, en términos de tasa de desempleo, entre las distintas Comunidades Autónomas españolas. No obstante, sí que se puede admitir que, a medida que nos vamos centrando en grupos de población con características más y más homogéneas, se percibe un cierto acercamiento entre ellas. A pesar de todo, es preciso destacar la persistencia de un componente residual no explicado en las tasas de desempleo regionales.

Este factor regional podría reflejar dos

hechos: 1) los grupos de población caracterizados por edad, sexo y nivel de estudios no son todavía suficientemente homogéneos; y 2) las características regionales que pueden dar lugar a diferencias de desempleo, no han sido adecuadamente medidos, Si pudiésemos afinar más en la definición de los grupos, y si pudiésemos medir mejor todas esas características regionales, sería de esperar que el residuo acabase por desaparecer.

Bibliografía

1. 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,* nº 47, pp. 15-36.
2. DOLADO, J., GONZÁLEZ-PÁRAMO, J.M. y ROLDÁN, J.M. (1993): «Convergencia Económica entre las Provincias Españolas: Evidencia Empírica (1955-1989)», *Moneda y crédito,* nº 198, pp. 81-132.
3. 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.
4. GARCÍA, A., RAYMOND, J.L. y VILLAVERDE, J. (1995): «La convergencia de las provincias españolas», *Papeles de Economía Española,* nº 64, pp. 38-53.
5. MAS, M., MAUDOS, J., PÉREZ, F. y URIEL, E. (1994): «Disparidades regionales y convergencia en las CCAA Españolas», *Revista de Economía Aplicada,* vol. II (4), pp. 37-58.
6. OECD (2000): *Employment Oulook:* junio 2000, (OECD, París).
7. OLLOQUI, I., SOSVILLA, S. y ALONSO, J. (2002): «Convergencia en precios en las provincias españolas», *Información Comercial Española Tribuna de Economía,* nº 797, febrero, pp. 160-178.
8. RAYMOND, J.L. y GARCÍA, A. (1994): «Las disparidades en el PIB per cápita entre las Comunidades Autónomas y la hipótesis de convergencia», *Papeles de Economía Española,* nº 59, pp. 37-58.
9. SIMÓN PÉREZ, H.J. (2001): «La estructura de salarios pactados en España». *Estudios de Economía Aplicada,* nº 19 (diciembre), pp. 171-187.
10. SURIÑACH, J., ARTÍS, M., LÓPEZ. E. y SANSO, A. (1995): *Análisis Económico Regional: Nociones Básicas de la Teoría de la Cointegración,* Antoni Bosch, Editor, y Bosch i Gimpera, Fundació.
11. VILLAVERDE, J. (1999): *Diferencias regionales en España y Unión Monetaria Europea,* Colección Economía XXI. Ediciones Pirámide, S.A., Madrid.



INFORMACIÓN COMERCIAL ESPAÑOLA *en* INTERNET

www.revistas**ICE**.com