

CONVERGENCIA DEL DESEMPLEO REGIONAL: UN ANALISIS DE DESCOMPOSICION DE LA VARIANZA

Alfonso MORAL DE BLAS
Departamento de Fundamentos del Análisis Económico.
Universidad de Valladolid.
Avda. Valle Esgueva, 6, 47011, Valladolid
amoral@esgueva.eco.uva.es

1.- INTRODUCCION:

Con este trabajo hemos pretendido hacernos eco de una realidad muy concreta, la existencia de altas tasas de paro en España y la persistencia de grandes diferencias entre las tasas de paro regionales.

Con el fin de profundizar en este problema, y a través de un proceso de descomposición de la varianza, se ha pretendido dar respuesta a dos preguntas fundamentales. En primer lugar, cual es el peso del componente regional a la hora de analizar las tasas de paro regionales, y como evoluciona cuando variamos el grado de definición del grupo estudiado. Y en segundo lugar, si se ha producido, a lo largo del tiempo, una disminución en el peso que tiene el componente regional como variable explicativa de las diferencias que presentan las tasas de paro regionales. Dicho de otra manera, si se ha producido un proceso de convergencia de las tasas de paro regionales.

El resto del artículo se organiza de la siguiente manera. En el segundo apartado se realizará una revisión fundamentalmente empírica de las principales fuentes que originan las disparidades regionales. En el apartado tres nos vamos a centrar en el análisis de las fuentes de datos y la metodología empleada, e intentaremos explicar el porque del periodo estudiado, de las regiones analizadas y de la periodicidad de nuestra muestra. En el apartado cuatro vamos a profundizar en los resultados obtenidos tras el análisis de descomposición de la varianza, y finalmente, terminaremos nuestro estudio con una serie de conclusiones sobre si se ha tenido lugar un proceso de convergencia entre las tasas de paro de las regiones incluidas en nuestro trabajo.

2.- FUENTES DE DISPARIDAD REGIONAL

En un principio, la principal fuente de disparidad entre regiones está en las distintas tasas de crecimiento económico que estas presentan. Un crecimiento económico que va a depender de las dotaciones de factores, del stock de capital, de la tecnología y de las características que presente la población, y consiguientemente la fuerza laboral. Por otro lado, las diferencias que presentan las regiones en términos de empleo, desempleo y participación también pueden tener su origen en distorsiones y fallos del mercado que impiden su ajuste. En este ultimo apartado estarían la presencia de competencia imperfecta en los mercados de trabajo y de otros factores o los problemas que surgen a la hora de canalizar la información.

De acuerdo con el resultado económico publicado por la OCDE en el año 2000, se demuestra la importancia de características demográficas como la edad, el género o el nivel educacional a la hora de explicar la dispersión de las tasas de paro regionales en la mayoría de los países de la OCDE. En ese mismo estudio también se demuestra la correlación positiva entre la tasa de paro y de inactividad que, por un lado, admite el predominio del efecto trabajador desanimado, y por otro, está concediendo importancia, aunque de una manera indirecta, a la composición de la población y a la menor actividad de grupos sociales como las mujeres, los mayores y los trabajadores menos cualificados. Por lo tanto, la existencia de diferencias en la composición poblacional de cada región aparece como un factor explicativo muy importante a la hora de analizar las disparidades regionales en la tasa de desempleo.

En la literatura aparecen varios trabajos que corroboran la importancia de estos factores. Gabriel et al (1993 y 1995) y Berguer and Blomquist (1992) reflejan el papel desempeñado por los factores sociales y demográficos a la hora de explicar los diferenciales existentes en el mercado de trabajo. Partridge and Rickman (1997a y 1997b) también investigan en este sentido y afirman la existencia de unas fuerzas de “desequilibrio”, entre las que engloba a las características demográficas, que determinan la existencia de unos mercados locales diferenciados con sus propios niveles salariales y sus propias tasas naturales de desempleo.

A la hora de hablar de la dimensión regional del empleo, además de las características demográficas y sociales que a las que ya nos hemos referido, el principal factor que debemos tener en cuenta es la capacidad que tiene cada región para crear empleo, no en vano el crecimiento del empleo es condición necesaria aunque no suficiente a la hora de reducir el desempleo. Esta capacidad de crear empleo va a estar influenciada por la estructura industrial o la composición sectorial de la actividad económica desarrollada en cada región. También existen otros factores específicos como son la posesión de recursos naturales, el nivel de la actividad innovadora, la facilidad del acceso a mercados de factores y productos, la localización geográfica o la calidad de las infraestructuras, todos ellos muy importantes a la hora de analizar la vitalidad del empleo regional.

Existen trabajos empíricos que tratan este apartado y que recogen la importancia de estos factores, en especial resaltan la importancia de la composición sectorial que es la base de la actividad económica regional. Partridge and Rickman (1997c) llegan a esta misma conclusión en un estudio hecho para los Estados Unidos, y Taylor and Bradley (1997) corroboran este resultado en otro trabajo hecho para Alemania, Italia y Reino Unido.

Otro de los aspectos importantes a la hora de estudiar las diferencias regionales, es el peso de determinadas instituciones del mercado laboral como pueden ser los sindicatos. Ya hemos visto en el apartado teórico que los sindicatos obstaculizan el libre funcionamiento del mercado de trabajo, que causan aumentos salariales y provocan mayores niveles de desempleo. En este sentido, sería de esperar una correlación positiva entre el poder sindical que sufra cada región y los niveles de desempleo que presente.

Por otro lado, este factor se encuentra muy relacionado con otra importante fuente de disparidad regional como son las diferencias salariales. Estas diferencias salariales tienen un comportamiento contrario, y su dispersión tiene una correlación negativa con el nivel de desempleo que presente cada región, esto es debido a que esta dispersión estaría recogiendo las diferencias, en términos de productividad, existentes entre los distintos entes territoriales.

El estudio de la OCDE del año 2000, al que ya nos hemos referido, a través de un análisis de descomposición de la varianza, demuestra que el componente regional es más importante a la hora de estudiar el empleo que en el caso del nivel salarial. Este hecho tiene su explicación en el funcionamiento de los sindicatos, porque la negociación colectiva suele realizarse sector a sector, pero los resultados son tenidos en cuenta a un nivel nacional, por lo tanto no se produce una acción diferenciada por regiones que pueda tener en cuenta los distintos niveles de

productividad que tiene cada una. Únicamente en el caso de que existan sectores que sean exclusivos de una región concreta, puede darse una situación en la que la presión sindical ejerza una acción diferenciada entre distintas áreas territoriales. Alonso e Izquierdo (1999), encuentran en la ausencia de flexibilidad de la negociación salarial efectuada por los sindicatos una explicación a las diferencias existentes entre las tasas de paro regionales, y a su persistencia a lo largo del tiempo.

La última fuente de disparidad que vamos a tratar en este estudio viene del grado de especialización existente. Siguiendo nuevamente la teoría neoclásica, la eliminación de restricciones al intercambio y la libre movilidad de factores, junto con otros aspectos naturales y de localización, facilitan la especialización desde un punto de vista regional. Además, el aprovechamiento de las economías de escala junto con otro tipo de externalidades han favorecido la concentración de determinados sectores de producción en algunas regiones concretas.

Kim (1997 y 1999) y Krugman (1993), afirman que en Estados Unidos ha surgido un proceso de concentración, asociado principalmente a sectores como la agricultura y la industria, y sobre la base de una determinada localización de factores, unas determinadas infraestructuras o unas determinadas condiciones climatológicas. Un proceso de concentración que es menos importante en el caso del sector servicios.

Este proceso generalizado de concentración y especialización, puede provocar que cualquier perturbación asociada a un sector productivo concreto tenga un efecto diferenciado a nivel regional, afectando fundamentalmente a la producción y el empleo de la región o regiones donde se haya concentrado ese sector concreto, y en consecuencia provocando una acentuación de las disparidades existentes. Krugman (1993) y Duranton y Puga (1999) afirman que las regiones con una base industrial muy concreta y especializada, corren un mayor riesgo de sufrir un lento crecimiento y un mayor nivel de desempleo.

De acuerdo a todo el análisis realizado, podemos concluir que es la composición poblacional de las distintas regiones junto con su composición sectorial, su capacidad de crecimiento y las rigideces desde el punto de vista salarial, componen las principales fuentes de disparidad a la hora de llevar a cabo un análisis sobre las tasas de paro regionales.

3.- METODOLOGIA Y FUENTES ESTADÍSTICAS:

Vamos a comenzar este epígrafe haciendo un análisis pormenorizado de cuales son los datos empleados en nuestro estudio, y de cuales son los principales problemas que ha planteado su utilización. Y posteriormente intentaremos describir cual es la metodología econométrica que hemos empleado, así como la interpretación de los datos que proporciona.

3.1 Fuentes estadísticas

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, grupos de edad y nivel de estudios. Sin embargo, la serie del PIB, proviene de la contabilidad regional que proporciona la base de datos TEMPUS.

El hecho de usar estas dos fuentes distintas explica porque el periodo no es tan largo como el proporcionado por la EPA. La descomposición de la varianza, utiliza el PIB como variable

explicativa de la tasa de paro, que proviene de la contabilidad regional y de la que se dispone de un menor número de datos, más concretamente desde 1980 hasta 1999. Esa misma diversidad de fuentes, en concreto la periodicidad del PIB regional, también explica porque hemos realizado un estudio anual cuando la EPA proporciona datos trimestrales.

En este sentido, si bien se produce una disminución sensible en el número de datos, el hecho de que para este análisis se utilicen todas las series de manera conjunta hace que, aunque la periodicidad sea anual, el número de datos siga siendo suficientemente alto (180 en el caso de menos desagregación), y en cuanto a la posible pérdida de información, también podemos considerar que el componente temporal aparece igualmente representado con las variables anuales. Por otro lado este cambio de periodicidad también proporciona una serie de ventajas, en primer lugar disminuye sensiblemente el número de variables que debemos introducir, solo necesitamos una por año en lugar de una por trimestre, además se eliminan los posibles problemas estacionariedad que conlleva el trabajar con datos trimestrales. Para llevar a cabo este cambio, en las series desagregadas de actividad y ocupación, se ha considerado que el dato anual era el que proporcionaba el segundo trimestre del año correspondiente. La razón de elegir este trimestre esta en que, dentro de la estacionariedad que sufren la mayoría de las variables referidas al mercado laboral, suele ser este el trimestre más representativo y menos sujeto a las fluctuaciones estacionales del trabajo en España.

Otro problema que nos hemos encontrado, y que nos ha llevado a disminuir el número de Comunidades Autónomas estudiadas, es el causado por el nivel de desagregación que hemos introducido en nuestro análisis. Hemos de tener en cuenta que, conforme vamos obteniendo grupos más homogéneos, el número de personas que lo integran es cada vez menor, y por lo tanto pueden surgir problemas de muestreo a la hora de ser estudiados. En este sentido e INE considera que cuando los datos nos proporcionan valores inferiores a 5000, pueden aparecer estos errores de muestreo a los que nos hemos referido. Para solucionar este problema, en este trabajo se han usado dos métodos, por un lado eliminar las Comunidades Autónomas más afectadas de este problema de muestreo, y por otro intentar reducir un poco el nivel de desagregación (considerar sólo dos grupos de nivel de estudios, o solo dos grupos de edad) para evitar que el número de regiones estudiadas fuera demasiado bajo.

3.2 Metodología

La metodología que vamos a utilizar en este trabajo es la denominada descomposición de la varianza, y consiste en calcular el porcentaje de la varianza que explica cada variable o grupo de variables consideradas dentro de una regresión. A la hora de descomponer la varianza podemos recurrir a dos métodos distintos.

El primer método consiste en analizar, dentro de cada regresión, el porcentaje de reducción en el término residual como consecuencia de añadir la variable correspondiente como explicativa, cuando todas las demás ya han sido tenidas en cuenta. A la hora de realizar esta descomposición también debemos tener en consideración que el orden de introducción de las variables no afecta al cálculo de los porcentajes, y que la suma de todos los porcentajes no tiene que corresponder necesariamente con el R^2 de cada regresión. Un análisis similar puede encontrarse en el estudio sobre el empleo que hace la OCDE en el año 2000.

El segundo método de descomposición de la varianza que vamos a mencionar, y que es el que vamos a utilizar en nuestro estudio, emplea para su cálculo el valor que toma el coeficiente de determinación en cada una de las regresiones (Simón Pérez, 2001). En este método se calculan dos efectos distintos, un *efecto total*, que coincide con el valor que toma el estadístico R^2 (multiplicado por cien), cuando solo se introduce como variables explicativas las que reflejan la importancia del subconjunto considerado. Y un *efecto marginal*, que se calcula como el cambio en magnitud (multiplicado por cien), experimentado por el coeficiente de determinación

debido a la inclusión de un determinado grupo de variables cuando todas las demás ya han sido tenidas en cuenta.

Una vez calculados estos dos efectos distintos, el porcentaje de la varianza que explica cada variable estará comprendido entre el intervalo de valores que presentan estos dos efectos para cada una de ellas. En este método, la suma de los distintos porcentajes, en su variante marginal, si que coincide con el valor que toma el R^2 , y además, podemos considerar que el porcentaje que resta hasta el 100% sería aplicable a una varianza residual que se vería explicada por otros factores no incluidos en nuestra regresión.

El único problema que surge de la descomposición en efecto total y marginal proviene del hecho de que esos resultados dependen de las variables introducidas. Si existe colinealidad entre las variables, dependiendo de si esta es positiva o negativa, puede ocurrir que el efecto marginal se vea aumentado o disminuido, es decir, si las covarianzas no son cero puede que se altere el resultado. No obstante, los resultados si nos sirven como referencia para comprobar la importancia de ciertas variables en la regresión.

El conjunto de regresiones se realizará mediante la estimación por mínimos cuadrados ordinarios de funciones del siguiente tipo:

$$TP = f(C, Y, R_i, T_i, D_i)$$

donde TP es la tasa de paro de cada grupo, región y año, C es el término independiente, Y es el Producto Interior Bruto de cada región en el año correspondiente, R_i es un conjunto de nueve variables ficticias que recogen el efecto de estar en una u otra región, T_i es otro conjunto de variables ficticias, una por cada año estudiado, que reflejan la importancia del componente temporal en nuestro modelo y D es un tercer conjunto de variables cualitativas que van a recoger características como sexo, edad y nivel de estudios. La variable R_i solo refleja la existencia de nueve regiones debido a un problema de datos, hemos de tener en cuenta que determinados grupos de desagregación, debido a su tamaño, provocan resultados que no son fiables a efectos muestrales y que pueden afectar negativamente a los resultados de nuestro análisis. No obstante, podemos considerar que las regiones incluidas (Andalucía, Aragón, Castilla y León, Castilla la Mancha, Cataluña, Comunidad Valenciana, Galicia, Madrid y País Vasco), son lo suficientemente representativas como para dar validez a nuestro estudio.

En lo referente a cualidades como sexo, edad y estudios, están recogidas por tres únicas variables, SEXO que distingue entre que hombres y mujeres, EDAD que desdobra entre mayores de 25 años y menores de esa edad y ESTUDIOS que hace la distinción entre los que tienen estudios medio o superiores y los que no los tienen. También conviene señalar que en todas las regresiones se ha eliminado la ficticia que corresponde al año 1980 y el término independiente, esto es debido a que se han introducido todas las ficticias regionales y, por lo tanto, se podría incurrir en un problema de multicolinealidad perfecta, aunque si que se han tenido en cuenta cuando se eliminan variables para el cálculo de los efectos total y marginal. Por último conviene resaltar que, debido a que la serie de PIB regional que proporciona la base de datos TEMPUS del INE es más corta, hemos tenido que restringir nuestro trabajo al periodo comprendido entre 1980 y 1999.

La diferencia entre unas regresiones y otras radica en el número de datos utilizados, en los valores que toma TP y en el número de variables cualitativas que se van a introducir en el modelo. Cuando realizamos una desagregación, lo que estamos haciendo es desdoblar cada grupo correspondiente en otros, eso provoca que los datos se multipliquen, que haya que introducir nuevas ficticias que recojan ese desdoblamiento, y además, que los valores que toma la tasa de paro sean distintos, porque también lo son los grupos estudiados.

En total se van a realizar dos regresiones distintas para ver el peso del componente regional, una primera que analizará la importancia del sexo, la edad y el nivel de estudios junto al PIB, al componente regional y temporal a la hora de analizar el problema del desempleo. Y una segunda que eliminará todas las variables demográficas del estudio, y analiza el resto de cualidades representadas en nuestro modelo, es decir, el componente regional y temporal de nuestro modelo, además de la variable que recoge el nivel de producción. Posteriormente se realizarán otras cinco regresiones, una por cada división de cuatro años del periodo original de veinte que tenemos, para ver como evoluciona ese componente regional a lo largo del tiempo y poder saber si hay convergencia.

4.- RESULTADOS DE LA DESCOMPOSICIÓN DE LA VARIANZA:

Con este apartado pretendemos demostrar la importancia que tiene el componente regional en la explicación de la tasa de paro. Por un lado veremos su comportamiento según el grado de homogeneidad del grupo estudiado, y por otro veremos su evolución a lo largo del tiempo con el fin de poder concluir si se produce, o no, un proceso de convergencia.

Una vez analizada la metodología empleada, y estudiadas las peculiaridades de los datos empleados, ya podemos dedicarnos a revisar los resultados obtenidos con nuestro estudio econométrico.

Como ya se ha dicho en el epígrafe anterior, la primera regresión que vamos a realizar, y también la que mayor nivel de desagregación representa, estudia la importancia de características como el sexo, la edad y el nivel de estudios sobre la tasa de desempleo, además del PIB, del tiempo y el componente regional que son comunes a todos los modelos. Los datos que vamos a estudiar en este modelo son 1440 y resultan de multiplicar las 9 regiones por los 20 años, por los dos sexos, por los dos grupos de edad y por las dos divisiones hechas en el nivel de estudios. Por lo tanto, tendremos 18 variables categóricas que recogen el efecto temporal, una para cada año estudiado, toman el valor 1 cuando estamos en el año correspondiente y 0 en el resto de los años, estas son A1980, A1981..., A1999. Tenemos otras nueve variables que recogen el componente regional y que toman el valor 1 cuando estamos en la región correspondiente y 0 en el resto de los casos, las variables son ANDA, ARAG, CASL, CASM, CATA, CVAL, GALI, MADR y PVAS, y se corresponden con Andalucía, Aragón, Castilla y León, Castilla la Mancha, Cataluña, Comunidad Valenciana, Galicia, Comunidad de Madrid y País Vasco respectivamente. También aparece la variable PIB que refleja el valor que toma la producción para cada región y año, la variable SEXO, que toma el valor 1 cuando nos referimos a hombres y 0 cuando son mujeres, la variable EDAD que toma el valor 1 para los menores de 25 años y 0 para los mayores, y la variable ESTUDIOS que toma el valor 1 para los que poseen estudios medios y superiores y 0 para los demás.

Los resultados de esta primera regresión aparecen recogidos en la TABLA 1, y de una primera observación podemos extraer ya una serie de conclusiones relevantes. En primer lugar, tanto el R^2 como el R^2 ajustado presentan un valor cercano al 80%, esta nos da muestra de la importancia que tienen los grupos de variables introducidas a la hora de explicar la variabilidad de la tasa de desempleo. En este mismo sentido también podemos reflejar, a la vista de los valores que toma el estadístico t, que todas las variables introducidas en nuestro modelo son significativas a un nivel del 1%.

INSERTAR TABLA 1

Si nos fijamos en los valores que toman los coeficientes estimados también podemos realizar una serie de apreciaciones. Los coeficientes referidos a las variables anuales presentan todos valores positivos, siendo especialmente elevados los referidos a los años centrales tanto de la década de los ochenta como de los noventa y que vienen a corresponder con los años más

recesivos desde el punto de vista del empleo en España. Las variables que estudian el componente regional también presentan coeficientes positivos, siendo especialmente representativo el valor que toma Andalucía, aunque por otro lado no es un resultado sorprendente si tenemos en cuenta los elevados niveles que presenta la tasa de paro en esta comunidad. El PIB presenta un signo negativo, este podría considerarse adecuado en cuanto que viene a reflejar el hecho de que valores elevados de la producción se asocian con reducciones de la tasa de desempleo. El valor tan pequeño que presenta el coeficiente esta asociado a la magnitud en que esta expresada la producción. La variable que expresa la importancia del género también presenta un signo negativo, y viene a presentar el hecho de que las tasas de paro femeninas presentan niveles más elevados. La variable EDAD tiene un signo positivo, que también es el adecuado, y que viene a expresar las mayores tasas de paro que sufren los jóvenes en comparación a los mayores de 25 años. La última variable que estudiamos en este primer modelo es la variable ESTUDIOS, y presenta un signo positivo que refleja mayores tasas de paro para los que tienen estudios medios o superiores, esto que en primera instancia pudiera parecer ilógico no lo es tanto si tenemos en cuenta la clara influencia negativa que tienen los poseedores de estudios medios a la hora de evaluar el nivel de desempleo.

Las dos últimas columnas de esta TABLA 1 se refieren al cálculo del porcentaje que, cada conjunto de variables explicativas, representa a la hora de explicar la variable dependiente. Como ya se ha explicado anteriormente, se han calculado dos efectos distintos, efecto total y efecto marginal, y que vienen a representar los límites entre los que se va a mover el porcentaje que representa cada variable concreta. A la vista de los resultados obtenidos podemos destacar la importancia (mas del 50%) que tiene la variable EDAD a la hora de explicar la evolución que siguen las tasas de paro y que refleja los altos niveles de desempleo que presentan los segmentos más jóvenes de nuestra población activa. A continuación, y con porcentajes entre el siete y el ocho por ciento, aparecen las variables que recogen el componente temporal, las que recogen el componente regional y la que se refiere al género. Todas ellas pueden ser consideradas importantes a la hora de explicar nuestra variable dependiente. Finalmente aparecen otras dos variables, la que refleja la influencia de la producción y la referida al nivel de estudios, ambas con porcentajes inferiores al uno por ciento y por lo tanto con poca importancia a la hora de explicar la tasa de paro de nuestra economía.

El siguiente paso que vamos a dar en nuestro análisis de descomposición de la varianza, y que supone un menor nivel de desagregación, solo introduce como regresores las variables que recogen el componente espacial, las que recogen el componente temporal y la que refleja el nivel de producción. Este último modelo elimina los desdoblamientos que se hacían de acuerdo a las variables demográficas, y por lo tanto supone una reducción del número de datos (en este caso 180), además provoca la desaparición de los regresores que recogían esas fuentes de variación.

Los resultados de esta última estimación aparecen recogidos en la TABLA 2. Las conclusiones iniciales de esta estimación vuelven a coincidir con las que aparecían en el modelo anterior, todos los regresores vuelven a ser significativos al 1%, con la excepción de la variable que recoge el PIB que en este caso solo es significativa al 5%. Por lo que se refiere al coeficiente de determinación, presenta un valor superior a 0.85, y por lo tanto sigue confirmando que la especificación realizada es la adecuada, y que las variables introducidas como regresores recogen la mayor parte de la variabilidad que presenta la tasa de paro de nuestro país.

INSERTAR TABLA 2

Si nos fijamos en el valor que toman los coeficientes, tampoco hay grandes modificaciones con respecto al caso anterior, todos siguen manteniendo el mismo signo, positivo para las variables regionales y temporales y negativo para el PIB, y siguen siendo los mismos años y regiones (fundamentalmente Andalucía) los más influyentes a la hora de determinar el valor que toma la tasa de desempleo. No obstante, si parece apreciarse una disminución global en el valor

que toman los coeficientes, esto podría ser debido a la disminución que sufren las tasas de paro en sus valores máximos cuando consideramos conjuntamente a hombres y a mujeres, jóvenes y viejos, y educados junto a los no educados.

Finalmente, si nos referimos al porcentaje de la varianza que explica cada grupo de regresores, lógicamente se produce un aumento de los porcentajes de todas las variables, tanto por haber aumentado el R^2 como por haber disminuido el número de variables explicativas que aparecen en nuestra regresión. Especialmente relevante en el desarrollo de nuestro análisis, es el aumento sufrido por el porcentaje de la varianza que se explica por el conjunto de variables que recogen la dimensión espacial de nuestra regresión. Esto significa que la convergencia regional es menor según nos vamos centrando en el estudio de grupos de población más desagregados.

La primera conclusión que se deriva de todo el análisis que hemos realizado, es que se puede apreciar un considerable aumento del porcentaje de la varianza que explican las variables regionales según nos vamos centrando en grupos de población cada vez menos desagregados. Pasamos de porcentajes en torno al 7,5% cuando nos centramos en grupo de población muy definidos, donde aparece la población dividida por sexo, edad y nivel de estudios, hasta porcentajes superiores al 50% en el modelo menos desagregado donde únicamente hacemos la división por regiones y años.

Estos resultados nos permiten admitir la existencia de un mayor nivel de convergencia entre las tasas de paro regionales cuando nos centramos en grupos cada vez más homogéneos, y por lo tanto, dentro de esos grupos poblacionales mejor definidos por características de sexo, edad y estudios, tiene menos importancia el estar en una u otra región a la hora de determinar la tasa de paro soportada.

Como ya decíamos en el inicio de este apartado, podríamos intentar avanzar un poco más en nuestro estudio de la convergencia y realizar un análisis similar al anterior con el fin de estudiar lo que ocurre con el porcentaje de la varianza explicado por las variables regionales a lo largo del tiempo. Lo que nos estamos planteando es si el componente regional, o la importancia que tiene el estar en una u otra Comunidad Autónoma, ha sufrido una disminución según vamos acercándonos a nuestros días. Para llevar a cabo este nuevo análisis, vamos a partir de la primera de nuestras regresiones, que era la que recogía el modelo más completo y también la que incluía mayor número de datos, y la vamos a intentar dividir en cinco regresiones distintas, una para cada periodo de cuatro años en que vamos a dividir nuestra muestra. Tendremos por lo tanto cinco estimaciones, cada una con 288 datos, donde se va a regresar la tasa de paro contra las nueve regiones, contra las variables SEXO, EDAD y ESTUDIOS y contra un conjunto de variables anuales, cuatro en cada caso, y que se refieren a los años incluidos en cada submuestra. Se han realizado grupos de cuatro años, en primer lugar porque tenemos 20 años en total, y de esta manera nos quedan cinco grupos iguales, y en segundo lugar, porque de esta manera se suavizan un poco las fluctuaciones que podríamos obtener en el cálculo de los porcentajes si lo hacemos cada año, o en periodos de dos años.

Una vez realizadas las cinco estimaciones que ya hemos dicho, nuestro propósito es calcular el porcentaje de la varianza que representa el componente regional en cada una de ellas, y concluir de ese estudio la eventual convergencia, o divergencia, de las tasas de paro regionales en función de que ese porcentaje vaya aumentando o disminuyendo a lo largo del tiempo. Para el cálculo de ese porcentaje se van a seguir los mismos pasos que en las regresiones iniciales, es decir calcularemos un efecto total y uno marginal, que se obtienen a partir de la evolución seguida por los distintos estadísticos R^2 , según ya se ha explicado en el inicio de este apartado, y que van a determinar el intervalo de variación de este porcentaje de la varianza que explica cada variable o grupo de variables.

INSERTAR TABLA 3

Los resultados de estas nuevas estimaciones aparecen recogidas en la TABLA 3, donde reflejamos el valor que presenta el coeficiente de determinación en cada submuestra estudiada, así como el porcentaje de la varianza que es explicado por las variables que la dimensión espacial de nuestro modelo, más concretamente, aparecen los dos valores entre los que se espera que fluctue este porcentaje y que son los efectos total y marginal de los que ya hemos hablado.

A la vista de los resultados recogidos en esta última tabla podemos apreciar que los coeficientes de terminación siguen presentando valores muy elevados superando holgadamente el 80%, y en tres de los casos incluso el 85%. En lo referente a la evolución temporal que sigue el porcentaje de la varianza explicado por las variables que recogen el componente regional podemos apreciar, que tras empezar con un nivel muy bajo en el primer cuatrienio, aumenta en los dos siguientes para volver a disminuir en los dos últimos. Desde el punto de vista de la convergencia podemos decir que tras una primer periodo de divergencia, que se produce durante los ochenta, durante la década de los noventa si que se produce una cierta convergencia o acercamiento de las tasas de desempleo regionales que viene reflejado por la disminución que sufre el componente regional durante este periodo.

5.- CONCLUSIONES

En el ámbito de la teoría neoclásica, y asumiendo aspectos tan restrictivos como la existencia de información perfecta, homogeneidad de los factores de producción, total flexibilidad de precios o mercados perfectamente competitivos, los participantes del mercado de trabajo asumen la existencia de un equilibrio. Este se consigue a través del libre movimiento entre regiones como respuesta a diferencias salariales o del nivel de desempleo.

En contraposición a la afirmación teórica previa, la realidad económica española nos presenta un escenario completamente diferente. En España conviven altas tasas de desempleo generalizadas, con grandes diferencias entre los niveles de paro que presentan las distintas Comunidades Autónomas. Esta situación descrita nos obliga a poner en cuarentena los restrictivos supuestos de la teoría neoclásica, así como asumir la existencia de imperfecciones en el mercado laboral que nos alejan de una situación perfectamente competitiva.

Desde un punto de vista econométrico, y de acuerdo a los resultados obtenidos, debemos corroborar la importancia que tiene el componente regional, así como la composición demográfica que tenga cada Comunidad Autónoma, a la hora de estudiar el nivel de desempleo y poder explicar las disparidades regionales.

Finalmente, y en lo referente a la posible convergencia, debemos concluir que si que se ha apreciado, en los últimos cuatrienios, una paulatina disminución del peso del componente regional que había aumentado en los primeros subperiodos de nuestra muestra. Aunque no se obtienen resultados demasiado concluyentes No obstante será necesario el estudio de futuros cuatrienios para ver si se sigue avanzando en esta disminución de disparidades.

BIBLIOGRAFIA

ALONSO, J. E IZQUIERDO, M., Disparidades regionales en el empleo y el desempleo, Papeles de Economía Española, 0(80), 1999, pp 79-99.

BERGUER, M. C. y GLENN C. B., Movility and Destination in Migration Decisions: The roles of Earnings, Quality Life and Housing Prices, Journal of Housing Economics, 2(1), 1992, pp 37-59.

DURANTON, G Y PUGA, D., Diversity and Specialisation in Cities: Why, where and when does it matter?, Discussion Paper, 433, Centre for Economic Performance, London School of Economics, 1999.

GABRIEL, S. A., MATHEY, J. P. y WASCHER, W. L., The Demise of California reconsidered: Interstate Migration over the Economic Cicle, Federal Reserve Bank of San Francisco Economic Review, 1995, pp 30-45.

GABRIEL, STUART A., SHACK-MARQUEZ, J. y WASCHER, W. L., Does Migration Arbitrage Regional Labor Market Differentials?, Regional Science and Urban Economics, 32(2), 1993, pp 211-233.

KIM, S., Economic Integration and Convergencr; US Regions, 1840-1987, Working Paper, 6335, National Bureau of Economic Research, 1997.

KIM, S., Urban Development in the United States – 1690 to 1990, Working Paper, 7120, National Bureau of Economic Research, 1999.

KRUGMAN, P., Lessons of Massachusetts for EMU, Torres, F. and Giavazzi, F. (eds), Adjustment and Growth in the European Monetary Union, Cambridge University Press, Cambridge, 1993, pp 241-261.

OECD, Disparities in Regional Labour Markets. OECD Employment outlook. Junio 2000, pp 31-77.

PARTRIDGE, M. D. y RICKMAN, D. S., County Unemployment in Georgia: The Causes and Potencial Role for Economic Development Pollicy, Review of Regional Studies, 1997a, pp 17-39.

PARTRIDGE, M. D. y RICKMAN, D. S., The Dispersion of US State Unemployment Rates: The Role of Market and Non-market Equilibrium Factors, Regional Studies, 31(6), 1997b, pp 593-606.

PARTRIDGE, M. D. y RICKMAN, D. S., State Unemployment Differentials: Equilibrium Factors vs. Differential Employment Growth, Growth and Change, 28(3), Summer 1997c, pp 360-379.

SIMÓN PÉREZ, H., La Estructura de Salarios Pactados en España. Estudios de Economía Aplicada, 19, 2001, pp 171-187.

TAYLOR, J. BRADLEY, S., Unemployment in Europe: A Comparative Analysis of Regional Disparities in Germany, Italy and the UK, Kiklos, 50(2), 1997, pp 221-45.

TABLA 1 .

	COEFICIENTES	t ²	% DE LA VARIANZA	
			Efecto total	Efecto marginal
A1980	-	-	7.88	7.92
A1981	0.0312	2.67		
A1982	0.0585	5.02		
A1983	0.0846	7.25		
A1984	0.1129	9.67		
A1985	0.1392	11.89		
A1986	0.1368	11.64		
A1987	0.1204	10.14		
A1988	0.1188	9.85		
A1989	0.0910	7.42		
A1990	0.0789	6.32		
A1991	0.0727	5.75		
A1992	0.0989	7.79		
A1993	0.1587	12.60		
A1994	0.1883	14.78		
A1995	0.1697	13.06		
A1996	0.1792	13.62		
A1997	0.1557	11.52		
A1998	0.1342	9.63		
A1999	0.0923	6.42		
ANDA	0.2559	11.01	7.52	7.16
ARAG	0.0450	4.24		
CASL	0.1172	9.16		
CASM	0.0566	5.28		
CATA	0.1993	6.10		
CVAL	0.1331	7.47		
GALI	0.0597	4.83		
MADR	0.1668	6.37		
PVAS	0.1631	12.23		
PIB	-2.08E-08	-4.21	1.05	0.25
SEXO	-0.0816	-22.13	6.94	7.00
EDAD	0.2326	63.04	56.83	56.86
ESTUDIOS	0.0183	4.97	0.35	0.35
R ²	0.80			

TABLA 2 .

	COEFICIENTES	t ²	% DE LA VARIANZA		
			Efecto total	Efecto marginal	
A1980	-	-	32.73	31.42	
A1981	0.026089	2.730249			
A1982	0.041210	4.314014			
A1983	0.058100	6.080841			
A1984	0.083034	8.684519			
A1985	0.101255	10.56630			
A1986	0.096578	10.05137			
A1987	0.086548	8.919779			
A1988	0.084015	8.538012			
A1989	0.062434	6.251950			
A1990	0.050552	4.967815			
A1991	0.049155	4.784977			
A1992	0.067978	6.588173			
A1993	0.111359	10.86448			
A1994	0.131485	12.69240			
A1995	0.117898	11.21960			
A1996	0.113907	10.67900			
A1997	0.103688	9.518514			
A1998	0.083009	7.453794			
A1999	0.055216	4.825958			
ANDA	0.238755	12.05975	55.48	47.81	
ARAG	0.093498	4.722673			
CASL	0.093960	11.16901			
CASM	0.097979	11.64684			
CATA	0.106620	12.51059			
CVAL	0.113616	13.33146			
GALI	0.075871	7.217789			
MADR	0.106701	10.15077			
PVAS	0.178572	6.388652			
PIB	-7.80E-09	-1.852729	10.17	0.26	
R²	0.884772				

TABLA 3

PERIODO	R²	% DE LA VARIANZA EXPLICADO POR EL COMPONENTE REGIONAL	
		Efecto Total	Efecto Marginal
1980-1983	0.846569	6.37	5.43
1984-1987	0.861171	8.30	6.22
1988-1991	0.856116	14.20	14.69
1992-1995	0.876223	10.17	11.64
1996-1999	0.835382	9.99	11.83