

## **Econometría espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas**

Rosina Moreno Serrano y Esther Vayá Valcarce\*

**RESUMEN:** El objetivo del presente artículo es contribuir a la difusión de las técnicas de econometría espacial en nuestro país, ofreciendo una somera visión de los aspectos teóricos y aplicados de la misma. En concreto, se introducen los dos tipos de análisis de datos espaciales existentes: el análisis exploratorio, ideado para el estudio univariante, y el estudio de la autocorrelación espacial en un modelo de regresión. Asimismo, se realiza una aplicación de parte de las técnicas abordadas al caso de las regiones europeas.

**Clasificación JEL:** C49, O47, R12.

**Palabras Clave:** Autocorrelación espacial; Análisis Exploratorio de Datos Espaciales; Economía Regional.

**ABSTRACT:** The goal of this paper is helping in the diffusion in our country of the techniques given by Spatial Econometrics, offering an overview of its theoretical and applied aspects. Specifically, we introduce the two types of spatial data analysis: the exploratory analysis, focused on a univariate study, as well as the study of spatial autocorrelation in a regression model. Additionally, we apply some of the techniques to the case of the European Union regions.

**JEL classification:** C49, O47, R12.

**Key Words:** Spatial Autocorrelation; Exploratory Spatial Data Analysis; Regional Economics.

### **1. Introducción**

En los últimos cuarenta años, la economía regional y urbana ha experimentado un fuerte desarrollo metodológico basado en la necesidad de trabajar con la especial natu-

---

\* Miembros del grupo de investigación «AQR» (Anàlisi Quantitativa Regional). Profesoras de la Facultad de Ciencias Económicas y Empresariales, Universidad de Barcelona. E-mail: [rmore@eco.ub.es](mailto:rmore@eco.ub.es) y [evaya@eco.ub.es](mailto:evaya@eco.ub.es)

raleza de los datos de corte transversal. Cuando se utiliza este tipo de datos suelen aparecer los denominados *efectos espaciales*: la heterogeneidad y la dependencia espacial.

El primer efecto aparece cuando se utilizan datos de unidades espaciales muy distintas para explicar un mismo fenómeno. En ese caso aparecen problemas como la heteroscedasticidad o la inestabilidad estructural, los cuales pueden ser resueltos mediante las técnicas econométricas existentes para series temporales. La dependencia o autocorrelación espacial surge siempre que el valor de una variable en un lugar del espacio está relacionado con su valor en otro u otros lugares del espacio. No obstante, y a diferencia de lo ocurrido con la heterogeneidad espacial, la dependencia espacial no puede ser tratada por la econometría estándar. Ello es debido a la multidireccionalidad que domina las relaciones de interdependencia entre unidades espaciales.

Con el objetivo de resolver los problemas que supone la presencia de efectos espaciales, especialmente el segundo de ellos, surgió la econometría espacial como subdisciplina de la econometría general que proporciona las técnicas de contrastación y de estimación necesarias para trabajar con datos que presentan problemas de heterogeneidad y/o dependencia espacial. Respecto al origen de la econometría espacial cabe destacar que, tras una fase inicial de reconocimiento del problema (que se remonta a Student en 1914), es en la década de los setenta cuando surge el término de econometría espacial como tal, acuñado por Paelinck y Klaassen (1979) y originariamente referido a los esfuerzos realizados para abordar la autocorrelación espacial en el término de perturbación de una regresión. En un sentido más amplio, Anselin (1988a) la define como la colección de técnicas que tratan las peculiaridades causadas por el espacio en el análisis estadístico de los modelos tradicionales de la ciencia regional. Según Anselin y Bera (1998), en esta última definición se estarían incorporando específicamente las regiones, la localización y la interacción espacial, formando la base de los trabajos empíricos en economía regional y urbana.

El gran desarrollo de la econometría espacial se ha producido, sin embargo, en los años ochenta y noventa. Entre los primeros estudios que analizan de manera general los aspectos metodológicos más importantes se encuentran los trabajos de Cliff y Ord (1981), Blommestein (1983) y Anselin (1980, 1988a, b). Posteriormente han ido apareciendo en las revistas de economía regional contribuciones concretas en el campo de la econometría espacial, así como la aparición de varias colecciones de artículos como en Anselin y Florax (1995a) y Anselin y Rey (1997). De esta forma, la importancia y relevancia de los métodos que analizan los efectos espaciales en los modelos econométricos ha ido incrementando de forma notable. Ello es debido, entre otros aspectos, al renovado interés por el papel del espacio y la interacción espacial en la teoría económica (sirvan de ejemplo los aspectos relacionados con la nueva geografía económica, Krugman 1991), a la creciente disponibilidad de extensas bases de datos socio-económicos geo-referenciados y, por último, al desarrollo de una tecnología eficiente y poco costosa para tratar con dichos datos, tanto en forma de Sistemas de Información Geográfica como de *software* útil para el análisis de datos espaciales.

No obstante, a pesar de esta aparente mayor difusión de la econometría espacial, la distancia que la separa, en términos de su conocimiento y difusión, de la econometría clásica todavía es notable (Anselin y Griffith, 1988; Anselin y Hudak, 1992). Este

hecho es incuestionable a nivel teórico dado que, en la actualidad, la mayoría de los manuales de econometría y estadística no hacen referencia al problema de los efectos espaciales y, en caso de hacerlo, se limitan a un breve apunte. Por otra parte, la todavía escasa difusión de la econometría espacial también es evidente en el terreno aplicado, especialmente en el ámbito español, mostrando la necesidad de acercar las técnicas econométricas espaciales a los investigadores en el campo económico<sup>1</sup>.

Teniendo todo ello en cuenta, el propósito del presente artículo es contribuir a la difusión de las técnicas de econometría espacial en nuestro país, ofreciendo una somera visión de los aspectos teóricos y aplicados de la misma. El trabajo se estructura de la siguiente manera. En el segundo apartado se presentan los efectos espaciales, la heterogeneidad y la autocorrelación espacial, analizando a su vez la matriz de pesos espaciales como instrumento básico para reflejar las interdependencias en el espacio. El objetivo del análisis exploratorio de datos espaciales así como las principales herramientas del mismo se ofrecen en el tercer apartado, mientras que el cuarto se dedica al estudio de la dependencia espacial en un modelo de regresión. En el apartado quinto se realiza un breve comentario de algunas debilidades de la econometría espacial. En el apartado sexto se presenta una aplicación de las técnicas analizadas a lo largo del trabajo para el caso de la Unión Europea en el estudio de la convergencia regional. Por último, en el apartado séptimo se presentan las principales conclusiones del mismo.

## 2. Efectos espaciales

### 2.1. Heterogeneidad espacial

La heterogeneidad espacial consiste en la variación de las relaciones en el espacio. De forma general, se puede decir que existen dos aspectos distintos de heterogeneidad espacial: la inestabilidad estructural y la heteroscedasticidad. En el primer caso, la heterogeneidad espacial se refiere a la falta de estabilidad en el espacio del comportamiento de la variable bajo estudio que lleva a que la forma funcional y los parámetros de una regresión puedan variar según la localización siendo, por tanto, no homogéneos en toda la muestra. El segundo aspecto, la heteroscedasticidad, proviene de la omisión de variables u otras formas de errores de especificación que llevan a la aparición de errores de medida.

El tratamiento de la heterogeneidad espacial puede realizarse por medio de las técnicas econométricas estándar, entre otras, la consideración explícita de parámetros variantes, coeficientes aleatorios (Hildreth y Houck, 1968) y *switching regressions* (Quandt, 1958) o las técnicas de filtraje adaptativo espacial (Foster y Gorr, 1983, 1984, 1986). Sin embargo, en el caso específico de la heterogeneidad espacial, en donde ésta viene causada por cuestiones inherentes a la localización de las observaciones, el conocimiento teórico de la estructura espacial de los datos puede conducir-

<sup>1</sup> Para un detalle de los trabajos de economía regional en los que se han aplicado las técnicas de econometría espacial, véase Moreno y Vayá (2000).

nos a procedimientos más complejos pero más eficientes como la expansión espacial de parámetros (Casetti, 1972) o las regresiones ponderadas geográficamente (Fotheringham *et al.*, 1998). Sin embargo, dado que este efecto puede tratarse en la mayoría de los casos mediante las técnicas econométricas tradicionales, el presente trabajo se centra en la problemática en torno a la autocorrelación espacial.

## 2.2. Autocorrelación espacial

La dependencia o autocorrelación espacial aparece como consecuencia de la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que ocurre en otro lugar (Cliff y Ord, 1973; Paelink y Klaassen, 1979; Anselin, 1988a).

La autocorrelación espacial puede ser positiva o negativa. Si la presencia de un fenómeno determinado en una región lleva a que se extienda ese mismo fenómeno hacia el resto de regiones que la rodean, favoreciendo así la concentración del mismo, nos hallaremos ante un caso de autocorrelación positiva. Por el contrario, existirá autocorrelación negativa cuando la presencia de un fenómeno en una región impida o dificulte su aparición en las regiones vecinas a ella, es decir, cuando unidades geográficas cercanas sean netamente más disímiles entre ellas que entre regiones alejadas en el espacio. Por último, cuando la variable analizada se distribuya de forma aleatoria, no existirá autocorrelación espacial. Dos son las principales causas que pueden inducir a la aparición de dependencia espacial: la existencia de errores de medida y de fenómenos de interacción espacial, de efectos desbordamiento y de jerarquías espaciales.

Tras la definición anterior, es posible detectar una cierta similitud entre los conceptos de autocorrelación espacial y temporal en la medida en que, en ambos casos, se produce un incumplimiento de la hipótesis de independencia entre las observaciones muestrales, se hallen éstas referidas a unidades de corte transversal o a series temporales. Sin embargo, una importante diferencia aparece entre ellas: la dependencia temporal es únicamente unidireccional (el pasado explica el presente), mientras que la dependencia espacial es multidireccional (una región puede no sólo estar afectada por otra región contigua a ella sino por otras muchas que la rodean, al igual que ella puede influir sobre aquéllas). Este hecho imposibilita la utilización del operador de retardos  $L$  presente en el contexto temporal, el cual recoge únicamente una relación unidireccional.

La solución al problema de la multidireccionalidad en el contexto espacial pasa por la definición de la denominada matriz de pesos espaciales, de retardos o de contactos,  $W$ , una matriz cuadrada no estocástica cuyos elementos  $w_{ij}$  reflejan la intensidad de la interdependencia existente entre cada par de regiones  $i$  y  $j$ . Con respecto a cómo definir los mencionados pesos, cabe destacar que no existe una definición unánimemente aceptada, si bien se ha de cumplir que dichos pesos sean no negativos y finitos (Anselin, 1980). A pesar de ello, de forma habitual se recurre al concepto de contigüidad física de primer orden, utilizado inicialmente por Moran (1948) y Geary (1954), donde  $w_{ij}$  es igual a 1 si las regiones  $i$  y  $j$  son físicamente adyacentes o a 0 en caso contrario (se asume por definición que  $w_{ii} = 0$ )<sup>2</sup>. Siguiendo el criterio de la pro-

ximidad física, diversos autores han propuesto definiciones de  $W$  basadas en la utilización de la distancia entre regiones (Cliff y Ord, 1973, 1981; Dacey, 1968). En este sentido, Anselin (1980) propone la utilización de una matriz inversa de distancias al cuadrado, de manera que claramente la intensidad de la interdependencia entre dos regiones disminuye con la distancia que separa sus respectivos centros<sup>3</sup>. Por último, cabe destacar que, de forma habitual, se recurre a la estandarización de la matriz  $W$ , dividiendo cada elemento  $w_{ij}$  por la suma total de la fila a la que pertenece, de forma que la suma de cada fila de la matriz estandarizada sea igual a la unidad. Finalmente, cabe comentar que, a partir de la matriz de pesos espaciales, es posible obtener el denominado *retardo espacial* de una variable  $x$  sin más que premultiplicar la matriz  $W$  por la citada variable ( $Wx$ ). Así, cada elemento de dicha variable retardada espacialmente se muestra como un promedio ponderado de los valores de la variable en el subgrupo de observaciones vecinas  $S_i$ , dado que  $w_{ij} = 0$  para  $j \notin S_i$ .

### 3. Análisis exploratorio de datos espaciales (AEDE)

En los estudios aplicados de economía regional y urbana y, muy especialmente en aquellos centrados en temas relativos a la localización de la actividad y la desigualdad regional, es habitual proceder inicialmente a un análisis descriptivo de la distribución en el espacio de la variable (o variables) objeto de estudio a través de la representación en un mapa de la citada variable, complementándose dicha etapa descriptiva con el cómputo de diversos índices de concentración/desigualdad como el índice de Gini, de Theil, de Atkinson o la  $\sigma$ -convergencia. Sin embargo, dichas técnicas presentan importantes limitaciones. Así, si bien la observación de un mapa podría ofrecernos una idea intuitiva acerca de, por ejemplo, la posible correlación entre regiones vecinas o la presencia de algunas regiones próximas con valores elevados o bajos de la citada variable, la información así obtenida, aunque útil, es subjetiva y altamente dependiente, entre otras cosas, del número de intervalos seleccionados para representar a la variable. Asimismo, muy diferentes distribuciones espaciales, con distintos patrones de dependencia espacial, podrían suministrar valores similares de los citados índices de desigualdad (Vayá, 1998) dado que no consideran de forma explícita el espacio en el cual están localizadas las regiones. Por ello queda justificada la necesidad de un análisis complementario, como el análisis exploratorio de datos espaciales (AEDE), centrado de forma explícita en los efectos espaciales y consistente en el conjunto de técnicas que permiten describir distribuciones espaciales, identificar localizaciones atípicas (*outliers* espaciales), descubrir esquemas de asociación espacial (*cluster* espacial) y sugerir diferentes regímenes espaciales u otras for-

<sup>2</sup> Asimismo, es posible obtener matrices de contigüidad de órdenes superiores. Así, diremos que dos regiones  $i$  y  $j$  serán contiguas de segundo orden si ambas están separadas por una tercera región  $h$  que es contigua de primer orden a ambas. La misma idea es extensible para órdenes superiores.

<sup>3</sup> Dado que estas definiciones están excesivamente relacionadas con aspectos físicos de las unidades espaciales, algunos autores han propuesto definiciones alternativas (Bodson y Peeters, 1975; Case *et al.*, 1993; López-Bazo *et al.*, 1999; Moreno *et al.*, 2002; Vayá *et al.*, 1998, 2002).

mas de inestabilidad espacial. El centro de estos conceptos lo ocupa la noción de autocorrelación espacial.

### 3.1. Estadísticos globales y locales de autocorrelación espacial

Los estadísticos globales de autocorrelación constituyen la aproximación más tradicional al efecto de dependencia espacial, permitiendo contrastar la presencia o ausencia de un esquema de dependencia espacial a nivel univariante, es decir, contrastar si se cumple la hipótesis de que una variable se encuentra distribuida de forma totalmente aleatoria en el espacio o si, por el contrario, existe una asociación significativa de valores similares o disímiles entre regiones *vecinas*. Entre dichos estadísticos globales, los cuales resumen en un indicador único el esquema general de dependencia, se encuentran la I de Moran (Moran, 1948) y la G(d) (Getis y Ord, 1992), presentándose sus expresiones en el cuadro 1. De forma general, todos ellos permiten contrastar la hipótesis nula de no autocorrelación espacial, es decir, la hipótesis nula de la existencia de una distribución aleatoria de la variable a lo largo del territorio.

Dado que los tests de autocorrelación espacial global analizan todas las observaciones de la muestra de forma conjunta, no resultan sensibles a situaciones donde predomine una importante inestabilidad en la distribución espacial de la variable objeto de estudio (procesos no estacionarios espacialmente), es decir, no contemplan la posibilidad de que el esquema de dependencia detectado a nivel global (por ejemplo, ausencia de autocorrelación espacial) pueda no mantenerse en todas las unidades del espacio analizado. Dicha limitación es fácilmente superable por medio del cálculo de los denominados contrastes locales de asociación espacial entre los que se encuentran el estadístico local de Moran,  $I_i$ , (Anselin, 1995a) y los test New- $G_i(d)$  y New- $G_i^*(d)$  (Ord y Getis, 1995). En el cuadro 2 se ofrecen sus expresiones. En este caso, se obtendrá un valor de dichos estadísticos para cada observación de la muestra, pudiendo así analizar la situación de cada unidad espacial por separado. Finalmente, es necesario resaltar la complementariedad de la información suministrada por estos índices, tanto globales como locales (Moreno y Vayá, 2000).

**Cuadro 1.** Algunos estadísticos de asociación espacial global

| <i>I de Moran</i>   | <i>G(d) de Getis y Ord</i>   |
|---|--|
| $I = \frac{N}{S_0} \cdot \frac{\sum_{ij}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_{i1}^N (x_i - \bar{x})^2} \quad i \neq j$   | $G(d) = \frac{\sum_{i1}^N \sum_{j1}^N w_{ij}(d) x_i x_j}{\sum_{i1}^N \sum_{j1}^N x_i x_j} \quad i \neq j$  |
| <p><math>x_i</math>: valor de la variable <math>x</math> en la región <math>i</math>; <math>\bar{x}</math>: media muestral de la variable <math>x</math>; <math>w_{ij}</math>: pesos de la matriz <math>W</math>; <math>N</math>: tamaño muestral; <math>S_0 = \sum_i \sum_j w_{ij}</math>.</p> | <p>donde dos pares de regiones <math>i</math> y <math>j</math> son vecinas si se encuentran dentro de una distancia <math>d</math> determinada.</p>  |
| <p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> <li>• <math>Z(I) &gt; 0</math> y signif.: autocorrelación positiva</li> <li>• <math>Z(I) &lt; 0</math> y signif.: autocorrelación negativa</li> </ul>   | <p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> <li>• <math>Z(G) &gt; 0</math> y signif.: concentración de valores elevados</li> <li>• <math>Z(G) &lt; 0</math> y signif.: concentración de valores bajos</li> </ul> |

**Cuadro 2.** Algunos estadísticos de asociación espacial local

| <i>Local de Moran <math>I_i</math></i>  | <i>G(d) de Getis y Ord</i>   |
|---|--|
| $I_i = \frac{z_i}{\sum_i z_i^2 / N} \sum_{j \in J_i} w_{ij} z_j$  | $New - G_i^* = \frac{\sum_{j=1}^N w_{ij} x_j - W_i^* \bar{x}}{s \{ [(NS_{ii}^* - W_i^{*2}) / (N-1)]^{1/2} \}}$   |
| <p><math>z_i</math>: valor correspondiente a la región <math>i</math> de la variable normalizada; <math>J_i</math>: conjunto de regiones vecinas a <math>i</math>.</p>  | $W_i^* = W_i + w_{ii}, S_{ii}^* = \sum_j w_{ij}^2, s^2 = \frac{1}{N-1} \sum_j (x_j - \bar{x})^2$   |
| <p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> <li>• <math>Z(I_i) &gt; 0</math> y signif.: cluster de valores similares alrededor de <math>i</math></li> <li>• <math>Z(I_i) &lt; 0</math> y signif.: cluster de valores disímiles alrededor de <math>i</math></li> </ul> | <p>Interpretación:</p> <ul style="list-style-type: none"> <li>• <math>New - G_i^* &gt; 0</math> y signif.: cluster de valores similares elevados alrededor de <math>i</math></li> <li>• <math>New - G_i^* &lt; 0</math> y signif.: cluster de valores similares bajos alrededor de <math>i</math></li> </ul> |

### 3.2. Técnicas de visualización de la presencia de efectos espaciales

Los estadísticos de dependencia espacial analizados en el apartado anterior se muestran como instrumentos útiles para detectar la existencia de patrones de asociación espacial específicos en la distribución de una variable en el territorio. No obstante, el AEDE incluye otras herramientas complementarias que permiten, por medio de un entorno gráfico, enriquecer los resultados obtenidos tras el cómputo de los citados contrastes. Entre ellas, podemos destacar el denominado *box map*, una extensión del mapa cuartil junto con los *outliers* superiores e inferiores (definidos estos últimos como las observaciones fuera de los límites en un *box plot*), el cual se muestra como una forma sencilla de identificar *outliers* espaciales.

Una aproximación más formal a la visualización de la dependencia espacial se basa en el concepto del *scatterplot* de Moran y su asociado *scatter map*. El *scatterplot* de Moran es un tipo de gráfico que muestra en el eje de abscisas las observaciones de la variable objeto de estudio normalizada y en el de ordenadas el retardo espacial de la misma también normalizado. De este modo, los cuatro cuadrantes reproducen diferentes tipos de dependencia espacial. Si la nube de puntos está dispersa en los cuatro cuadrantes es indicio de ausencia de correlación espacial. Si por el contrario los valores se encuentran concentrados sobre la diagonal que cruza los cuadrantes I (derecha superior) y III (izquierda inferior), existe una elevada correlación espacial positiva de la variable, coincidiendo su pendiente con el valor de la  $I$  de Moran. La dependencia será negativa si los valores se concentran en los dos cuadrantes restantes. Finalmente, el *scattermap* es la traslación de la información del *scatterplot* a un mapa del territorio, donde cada uno de los cuatro cuadrantes es resaltado con un color diferente.

Como se puede derivar de las herramientas comentadas, la introducción del mapa como una posible vista adicional de los datos es básica. De esta forma, muchas de las técnicas de análisis exploratorio de datos espaciales aprovecha la capacidad de visua-

lización y manipulación de los datos que proporcionan los Sistemas de Información Geográfica (SIG)<sup>4</sup>. Por ejemplo, el SIG *ArcView* (ESRI, 1995) se organiza en torno a varias vistas vinculadas de los datos (un mapa, una tabla y varios tipos de gráficos), lo cual permite un cierto grado de interacción dinámica en el sentido de que una selección determinada en una de las vistas se refleja de forma inmediata en el resto. Asimismo, se desarrollaron en los años noventa varias vinculaciones de *ArcView* con otros módulos de *software*, convirtiéndolo en una herramienta adecuada para el análisis exploratorio de datos espaciales. Así, por ejemplo, se ha vinculado el *software SpaceStat* para el análisis de datos espaciales creado por Anselin (1992, 1995b) con el programa *ArcView* en un entorno de Microsoft Windows (Anselin y Bao, 1996, 1997; <http://www.spacestat.com>). El principio tras dicho vínculo es la transferencia de resultados específicos desde *SpaceStat* a *ArcView* de manera que los resultados numéricos obtenidos pueden ser visualizados rápidamente sobre un mapa.

#### 4. Autocorrelación espacial en el modelo de regresión

En el apartado anterior se han analizado los estadísticos espaciales habitualmente utilizados para contrastar la presencia de dependencia espacial a nivel univariante. No obstante, de igual modo, es posible que el citado efecto espacial esté presente en el contexto de un modelo de regresión, ya sea como consecuencia de la existencia de variables sistemáticas (endógena y/o exógenas) correlacionadas espacialmente o como consecuencia de la existencia de un esquema de dependencia espacial en el término de perturbación.

En caso de que la variable endógena de un modelo de regresión lineal esté correlacionada espacialmente, la solución pasa por especificar el siguiente modelo:

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &\sim N(0, \sigma^2 I) \end{aligned} \quad [1]$$

donde  $y$  es un vector ( $N \times 1$ ),  $W y$  el retardo espacial de la variable  $y$ ,  $X$  una matriz de  $K$  variables exógenas,  $u$  un término de perturbación ruido blanco,  $N$  el número de observaciones  $y$ , por último,  $\rho$  el parámetro autorregresivo que recoge la intensidad de las interdependencias entre las observaciones muestrales.

De igual forma, la autocorrelación espacial podría estar únicamente presente en el término de perturbación:

$$\begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= \lambda W \varepsilon + u \\ u &\sim N(0, \sigma^2 I) \end{aligned} \quad [2]$$

<sup>4</sup> Se puede decir que el epicentro del SIG es una base de datos que combina de forma eficiente información de los valores (atributos) de las variables de interés con información locacional o topológica (disposición espacial). Es decir, un SIG es un poderoso conjunto de herramientas para la recogida, almacenamiento, transformación y disposición de datos espaciales del mundo real con unos objetivos determinados.

donde  $u$  es un término ruido blanco y  $\lambda$  el parámetro autorregresivo que refleja la intensidad de las interdependencias.

Si bien la estructura autorregresiva es la más utilizada a nivel tanto teórico como empírico, es posible modelizar el término de perturbación de forma diferente. Así, Cliff y Ord (1981) sugieren utilizar un proceso de medias móviles<sup>5</sup> espacial de orden 1, SMA(1):

$$\begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= \theta W_1 u + u \\ u &\sim N(0, \sigma^2 I) \end{aligned} \quad [3]$$

donde  $\theta$  es el coeficiente de medias móviles espacial y  $u$  es el término de perturbación incorrelacionado (obviamente, el esquema definido es generalizable a órdenes superiores)<sup>6</sup>.

A partir de la expresión [1] es fácil observar como, en caso de omitir de forma errónea un retardo espacial de la variable endógena y/o exógenas, la dependencia espacial se trasladaría directamente al término de perturbación, el cual pasaría a estar correlacionado espacialmente. Este tipo de autocorrelación espacial es conocido con el nombre de autocorrelación espacial sustantiva y su solución pasa por la inclusión en el modelo de un retardo espacial de la variable correlacionada espacialmente. Por el contrario, cuando la dependencia espacial residual no está causada por la omisión errónea de un retardo de alguna de las variables sistemáticas, nos hallamos ante un caso de autocorrelación espacial residual como *nuisance* o perturbación. Dicha situación podría verse explicada por la omisión de variables no cruciales que se hallen correlacionadas espacialmente o por la existencia de errores de medida. En este caso, la solución implica la consideración explícita de un esquema de dependencia espacial en el término de error.

Existe una amplia batería de estadísticos espaciales para contrastar las anteriores estructuras: los tests *ad-hoc*, los cuales son válidos para contrastar la hipótesis nula de no autocorrelación espacial residual (si bien no presentan una hipótesis alternativa definida que explícitamente describa el proceso generador del término de perturbación); y aquellos basados en el principio de máxima verosimilitud (los cuales están rigurosamente estructurados en términos de una hipótesis nula y alternativa específicas). Entre los primeros destacan los tests de la I de Moran y el test K-R, mientras que en el segundo grupo se incluyen el contraste de Wald, el de razón de verosimilitud y el de los multiplicadores de Lagrange, si bien los más utilizados son estos últimos (véase el cuadro 3, para un resumen de los contrastes más comúnmente utilizados).

<sup>5</sup> Ver Anselin y Bera (1998) para un análisis de las implicaciones económicas que se derivan de la distinta matriz de varianzas y covarianzas del término de error que se obtienen en una estructura autorregresiva o una media móvil.

<sup>6</sup> Los modelos presentados no agotan todas las posibilidades de modelos de autocorrelación espacial. Así, existen formas más generales, como el modelo mixto regresivo espacial regresivo con perturbaciones autorregresivas y heteroscedásticas (Anselin, 1988a), así como distintos esquemas para la autocorrelación residual como el modelo de Kelejian y Robinson (1993) analizado en detalle en Anselin y Moreno (2003).

**Cuadro 3.** Algunos estadísticos de autocorrelación espacial en el modelo de regresión

|            |   |   |
|------------|---|---|
| I de Moran | $I = \frac{N}{S} \frac{e'We}{e'e}$  | e: residuos MCO; N: tamaño muestral; S: suma de todos los elementos $w_{ij}$ de la matriz W. Debe estandarizarse.                     |
| LM-ERR     | $LM - ERR = \frac{\left[ \frac{e'We}{s^2} \right]^2}{T_1} \sim \chi_1$  | e: residuos MCO; $s^2$ : estimación de la varianza residual; $T_1 = \text{tr}(W'W + W^2)$ .   |
| LM-EL      | $LM - EL = \frac{\left[ \frac{e'We}{s^2} - T_1 (RJ_{\rho-\beta})^{-1} \frac{e'Wy}{s^2} \right]^2}{\left[ T_1 - T_1^2 (RJ_{\rho-\beta})^{-1} \right]} \sim \chi_1$ | $RJ_{\rho-\beta} = [T_1 + (WX\beta)'M(WX\beta)/s^2]$ ; $M = I - X(X'X)^{-1}X'$ , y: var. endógena, y el resto de notación como antes. |
| LM-LAG     | $LM - LAG = \frac{\left[ \frac{e'Wy}{s^2} \right]^2}{RJ_{\rho-\beta}} \sim \chi_1$  | Notación como antes.  |
| LM-LE      | $LM - LE = \frac{\left[ \frac{e'Wy}{s^2} - \frac{e'We}{s^2} \right]^2}{RJ_{\rho-\beta} - T_1} \sim \chi_1$  | Notación como antes.  |

En relación a las consecuencias derivadas de la autocorrelación espacial en la estimación habitual por MCO, es preciso distinguir entre los dos tipos de autocorrelación ya comentados: la dependencia espacial residual y sustantiva. Por lo que hace referencia a las consecuencias de la presencia de residuos correlacionados espacialmente en la estimación MCO, éstas son similares a las conocidas en el contexto temporal. Así, si bien las estimaciones de los parámetros seguirán siendo insesgadas, sin embargo serán ineficientes dado que la matriz de varianzas y covarianzas del término de perturbación será no esférica, a la vez que la varianza residual será sesgada, siendo las predicciones MCO ineficientes. Todo ello lleva a sesgar la inferencia basada en los tests de significación de la t-student y en el coeficiente de determinación  $R^2$  (el cual estará sobrevalorado). Asimismo, también se verá afectada la validación de un número importante de contrastes utilizados para detectar especificaciones erróneas como, por ejemplo, los tests de inestabilidad estructural como el test de Chow o los tests de heteroscedasticidad. Por contra, las estimaciones MCO en presencia de un retardo espacial de la variable endógena serán sesgadas e inconsistentes, aun cuando el término de perturbación no esté correlacionado espacialmente.

De todo ello se desprende que, con independencia del esquema de dependencia espacial detectado, la estimación MCO no es la adecuada. En este sentido, la aproximación máximo verosímil (MV) se ha mostrado como una de las alternativas más ampliamente utilizadas. Con respecto a dicho método de estimación, cabe decir que, de idéntica forma que en el contexto temporal, los estimadores MV deberán ser obtenidos a partir de la maximización del logaritmo de la función de verosimilitud aso-

ciada al modelo espacial especificado. En este sentido, partiendo de una función de densidad normal conjunta para los errores del modelo más general con autocorrelación tanto residual como substantiva (retardo espacial de la endógena), se obtiene una función de verosimilitud cuya expresión es (en logaritmos):

$$\text{Ln } L(\theta) = -\frac{N}{2} \ln 2\pi - \frac{1}{2} \ln |\Omega| + \ln |B| + \ln |A| - \frac{1}{2} v'v \quad [4]$$

donde  $\theta = [\rho, \beta', \lambda, \sigma^2]$ ,  $A = (I - \rho W)$ ,  $B = (I - \lambda W)$ ,  $\Omega$  es la matriz de varianzas y covarianzas del término de perturbación y  $v'v = (Ay - X\beta)'B'\Omega^{-1}B(Ay - X\beta)$ . A partir de la expresión [4], los estimadores MV podrán ser obtenidos igualando a 0 el vector de derivadas parciales de [4] respecto a  $\theta^7$ . En este trabajo, si bien no se desarrollará el proceso de derivación de dichos estimadores (ver Anselin, 1988a, para un análisis exhaustivo), es necesario destacar que el sistema resultante de primeras derivadas es altamente no lineal. Para solucionarlo será preciso recurrir a métodos numéricos dada la imposibilidad de disponer de una solución analítica válida. No obstante, esta dificultad se ve reducida cuando se imponen determinadas restricciones en el modelo general, pasando a trabajar con modelos más sencillos como los definidos en [1] y [2]. En este caso, la solución al problema de la no linealidad viene de la mano de una función de verosimilitud concentrada, la cual será lineal para todos los parámetros menos uno.

Además de la estimación máximo-verosímil, han sido propuestos en la literatura métodos de estimación alternativos, entre los que destacan la estimación por variables instrumentales (Haining, 1978; Bivand, 1984; Anselin, 1988a) o el método generalizado de los momentos (Kelejian y Prucha, 1997). Dichos métodos de estimación se revelan como notablemente más simples que la estimación MV, incluso cuando el tamaño muestral es elevado. Asimismo, bajo determinadas condiciones, dichos estimadores son consistentes aun cuando el término de error no siga una distribución normal.

## 5. Algunas limitaciones de la Econometría Espacial

A pesar de los avances experimentados por la Econometría Espacial en las últimas décadas, y la mayor utilización que se hace de la misma, antes de concluir nos gustaría destacar algunas debilidades sobre las que debería trabajarse de forma más intensa en el futuro.

En primer lugar, en relación a las estimaciones MV, es preciso destacar que, si bien dicho método de estimación es el más ampliamente utilizado, esta técnica pre-

<sup>7</sup> Existe una limitación en el rango de valores posibles que pueden alcanzar los parámetros espaciales, a fin de que los estimadores máximo-verosímiles alcancen sus propiedades asintóticas. Específicamente, se deberá de cumplir que  $1/\omega_{\min} < \rho < 1/\omega_{\max}$  y que  $1/\omega_{\min} < \lambda < 1/\omega_{\max}$ , donde  $\omega_{\max}$  y  $\omega_{\min}$  son, respectivamente, el mayor y menor valor propio en términos reales de  $W$ . En caso de que  $W$  esté estandarizada,  $\omega_{\max}$  alcanza un valor igual a 1, mientras que  $1/\omega_{\min} \leq -1$ , de manera que los valores positivos de los parámetros espaciales deberán ser menores que 1.

senta algunas complicaciones derivadas tanto de la necesidad de recurrir a métodos numéricos para resolver los procesos de optimización no lineales como de sus propiedades asintóticas y de las restricciones impuestas sobre el valor de los parámetros autorregresivos. Asimismo, es necesario mencionar otro problema importante: el coste computacional que supone la implementación de la estimación máximo verosímil en caso de trabajar con tamaños muestrales medios o elevados. Sin embargo, tal y como comentan Pace (1996) y Pace y Barry (1996, 1997), la mayor parte de las matrices utilizadas en dichos cálculos se caracterizan por incluir un gran número de elementos iguales a 0, incrementando la proporción de estos últimos con relación al número total de observaciones a medida que aumenta el tamaño muestral. Este hecho posibilita la aplicación de técnicas de matrices *sparse* basadas en la aplicación de un conjunto de algoritmos, lo cual reduce de forma notable el tiempo invertido en la estimación MV así como la cantidad de información que es preciso almacenar de forma temporal y permanente.

Por otra parte, en relación a la modelización espacio-temporal, no contemplada en este trabajo por razones de espacio, la simple adaptación directa de los métodos de contrastación y estimación desde el contexto transversal presenta la desventaja de que la naturaleza del panel de datos no se acaba considerando. En este caso, la solución viene de la mano, por una parte, de los denominados modelos SUR (Modelos de ecuaciones aparentemente no relacionadas) y SUR espacial y, por la otra, de los modelos de componentes del error. Sin embargo, la utilización de dichos modelos ha resultado, hasta el momento, escasa.

## 6. Aplicación empírica de las técnicas de Econometría Espacial

El estudio del crecimiento económico y la convergencia, especialmente en el ámbito regional, ha sido uno de los tópicos que ha despertado un mayor interés en los últimos años, dando lugar a la aparición de un número muy notable de contribuciones tanto de índole teórica (centradas en su mayoría en la discusión entre modelos neoclásicos y modelos endógenos de crecimiento) como aplicada. En este último caso, es habitual encontrar trabajos en los que, para diferentes ámbitos geográficos, se analice la existencia o no de convergencia regional en base a la estimación de la ecuación tradicional de  $\beta$ -convergencia (ver Barro y Sala-i-Martin, 1995), donde la tasa de crecimiento del producto per capita (o de la productividad laboral) se regresa contra el valor inicial de dicha variable,

$$g_y = a - (1 - e^{-\beta T}) \ln y + e \quad [5]$$

siendo  $g_y$  la tasa de crecimiento de la variable entre el período 0 y  $T$ ,  $\ln y$  el logaritmo de la variable en el año inicial,  $a$  es una constante que refleja el estado estacionario o nivel de equilibrio a largo plazo de la misma y  $e$  el término de perturbación. Muy brevemente, cabe decir que en este caso concluiríamos en la presencia de convergencia

absoluta si se contrasta la existencia de una relación significativa y negativa entre ambas variables.

Una de las principales críticas que cabría hacer a este tipo de análisis, extrapolable por otra parte a muchos otros trabajos aplicados en el ámbito regional y local, es que prescinden por completo del espacio, como si las regiones fueran unidades aisladas dentro de un territorio mayor ajenas al comportamiento del resto. Así, son muy pocos los trabajos que, en mayor o menor medida, suplen esta deficiencia, analizando la existencia de efectos espaciales en la muestra analizada (ver Armstrong, 1995; López-Bazo *et al.*, 1999; Rey y Montouri, 1999; Ying, 2000; Fingleton y McCombie, 1998; Moreno *et al.*, 2002; Vayá *et al.*, 2002). Este hecho podría ser relativamente aceptable si el omitir de forma errónea dichos efectos tuviera escasas consecuencias sobre el modelo de regresión o si, por ejemplo, la implementación de las técnicas espaciales analizadas previamente fuera muy costosa. Sin embargo, tal y como se ha mostrado en los apartados anteriores, ninguna de estas razones es sostenible.

Teniendo ello en cuenta, y a modo de ejemplo, seguidamente se procederá a aplicar las técnicas espaciales estudiadas en los apartados 3 y 4 al estudio de la existencia de  $\beta$ -convergencia entre las regiones europeas. Para ello, hemos seleccionado una muestra formada por 108 regiones de la Unión Europea (UE) para las cuales se ha obtenido información de la variable Producto Interior Bruto medido en millones de ECUS y el empleo total para el período 1975-1992, utilizando para ello la base de datos REGIO del EUROSTAT. A partir de esta información se ha construido la variable productividad laboral, la cual ha sido relativizada al promedio europeo.

### 6.1. Análisis exploratorio de la productividad laboral en las regiones europeas

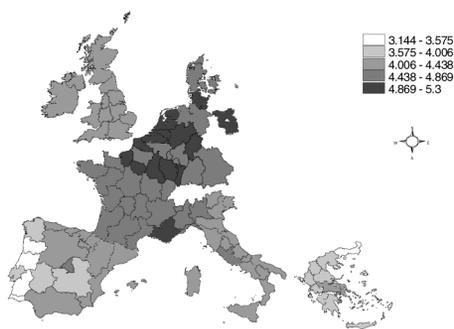
Como paso previo a la utilización de la variable productividad laboral en el modelo de regresión expresado en [5] se hace necesario proceder a un estudio exploratorio de la distribución espacial de la misma.

Teniendo ello en cuenta, iniciamos el análisis con la figura 1, la cual muestra el mapa en el que aparece representada la variable LnPIB por ocupado para 1975. Dos son las principales conclusiones que cabe extraer de la misma. En primer lugar, parece existir una clara no estacionariedad espacial en la distribución de la productividad como lo muestra la existencia de una cierta heterogeneidad espacial. Así, mientras que de forma generalizada las regiones de los estados tradicionalmente pobres de Portugal, Grecia, Irlanda y España (además de la mayoría de regiones de Italia y de Gran Bretaña) presentan valores bajos de la variable estudiada al principio del período, no ocurre lo mismo con las regiones centrales, las cuales concentran los valores más elevados de dicha variable. En segundo lugar, esta distribución parece encontrarse lejos de poder ser considerada como aleatoria. Por el contrario, parece existir, en términos globales, una asociación espacial positiva entre regiones próximas en la medida en que regiones vecinas muestran valores similares de la citada variable.

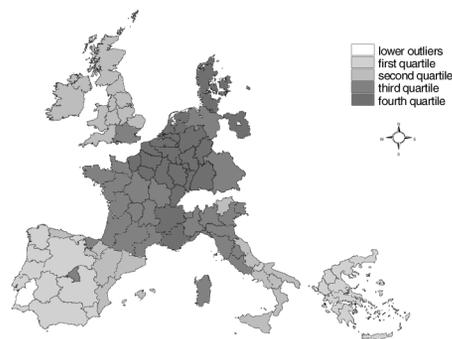
La figura 2 representa el *box map* asociado, a partir del cual parece reforzarse lo que ya se derivaba de la figura 1, esto es, la aparente existencia de dependencia espacial en la distribución de la productividad en 1975, observándose además un comportamiento notablemente diferenciado de las regiones centrales respecto al resto de regiones europeas. Así, se observa como claramente las regiones que forman parte del mismo cuartil se encuentran a su vez agrupadas en el espacio. Únicamente, la región de Madrid rompe este esquema ya que, perteneciendo al tercer cuartil, se encuentra completamente rodeada de regiones que se sitúan en el primer cuartil de la distribución. Por último, cabe destacar la existencia de un *outlier* de bajos valores observado en la región portuguesa de Alentejo.

No obstante, con relación a la posible asociación espacial que parece derivarse de las figuras anteriores, es preciso remarcar que los resultados que puedan deducirse especialmente de la figura 1 son altamente sensibles, entre otros aspectos, al número de intervalos definidos para representar a la variable bajo estudio.

**Figura 1.** Distribución espacial del Ln PIB por ocupado, 1975



**Figura 2.** *Box Map* del Ln PIB por ocupado, 1975



Por ello, es preciso llevar a cabo un estudio objetivo y exhaustivo de la posible presencia de dependencia espacial por medio del cómputo de los contrastes de autocorrelación espacial ya definidos en el apartado 3.1. En este sentido, en el cuadro 4 se muestran los resultados de los contrastes de autocorrelación espacial global I de Moran y G de Getis y Ord, utilizando una matriz *W* definida en términos de contigüidad física de primer orden estandarizada.

**Cuadro 4.** Contrastes de autocorrelación espacial globales

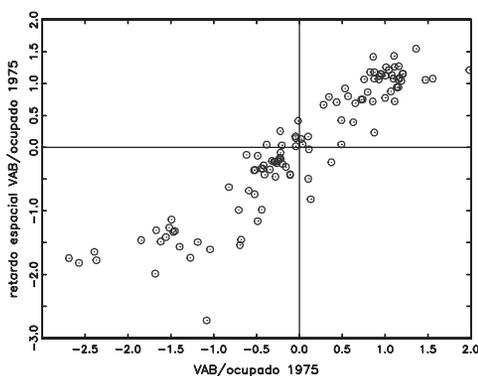
| Variable                 | <i>I</i> de Moran | <i>G</i> de Getis y Ord |
|--------------------------|-------------------|-------------------------|
| Ln PIB por ocupado, 1975 | 11.605*           | 2.543*                  |

Nota: \* denota significativo al 1%.

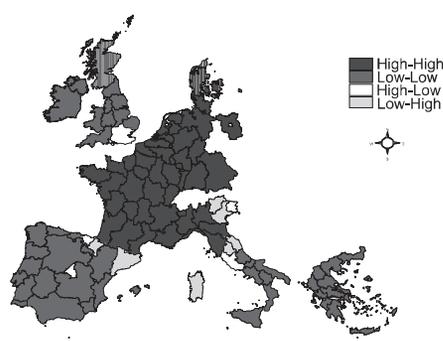
Como se puede observar, el logaritmo del PIB por ocupado en 1975 presenta una muy significativa dependencia espacial positiva, hecho que corrobora las primeras intuiciones derivadas de las figuras 1 y 2. De esta forma, el contraste de autocorrelación espacial I de Moran es claramente significativo a un nivel del 1%, rechazando la existencia de una distribución aleatoria de dicha variable y reflejando por tanto como, en términos globales, regiones próximas en el espacio muestran valores similares de la misma. Asimismo, el test G de Getis y Ord es significativo y positivo, hecho que lleva a concluir a favor de la presencia de una clara asociación en el espacio, mayoritariamente de valores elevados.

Otros instrumentos útiles en el análisis del grado de dependencia espacial de una variable son el *scatterplot* de Moran (figura 3) y su *scatter map* asociado (figura 4). Como se puede observar, existe una notable concentración de la masa de puntos en los cuadrantes I y III, corroborándose por tanto el predominio de una concentración en el espacio de valores similares de productividad laboral. Por el contrario, son muy pocos los casos en los que se observa una discrepancia entre el valor alcanzado por una región y el promedio observado en sus vecinas. Únicamente, como se observa en la figura 4, cabe resaltar que las regiones de País Vasco, Madrid, South East, Friuli-Venezia Giulia y Lazio concentran valores elevados de la productividad laboral (es decir, superiores a la media) mientras que sus regiones colindantes muestran la situación opuesta. La situación contraria aparece en las regiones de Navarra, Cataluña, Trentino-Alto Adige, Veneto, Umbria, Marche y Sardeña.

**Figura 3.** Scatterplot de Moran del Ln PIB por ocupado, 1975



**Figura 4.** Scatter Map del Ln PIB por ocupado, 1975

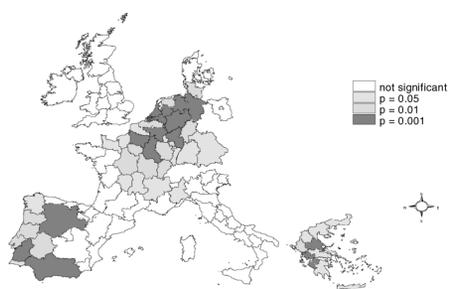


En la figura 5 se muestran los resultados del contraste  $I_i$ , resaltando únicamente aquellas regiones que presentan valores significativos de dicho contraste para tres diferentes niveles de significación, mientras que la figura 6 informa acerca de si las agrupaciones significativas detectadas por la  $I_i$  concentran valores elevados o bajos de

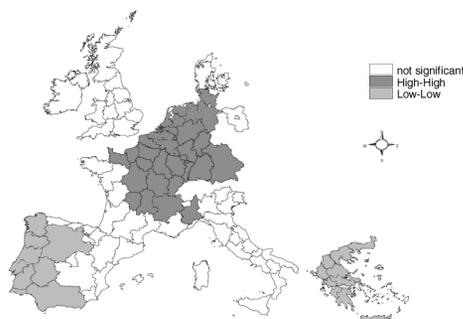
la variable (información que puede ser extraída tanto de la posición de cada región en los cuadrantes del *scatterplot* de Moran como del signo obtenido tras el cálculo del contraste  $New-G_i^*$ ). Por último, la figura 7 muestra las regiones con un valor significativo del test  $New-G_i^*$ <sup>8</sup>.

Las principales conclusiones que cabe extraer de las tres figuras comentadas se pueden resumir en tres puntos. En primer lugar, los *clusters* significativos detectados no se encuentran dispersos en el territorio sino que se sitúan próximos entre si y concentrados en determinadas zonas del sur y del centro de Europa. En segundo lugar, todos los *clusters* señalados concentran valores similares de la variable en cuestión, no detectándose ninguna región con un comportamiento significativamente disímil al mostrado por sus vecinas. Por último, destacar las similitudes entre las agrupaciones significativas detectadas por el contraste local de Moran y por el test  $New-G_i^*$ .

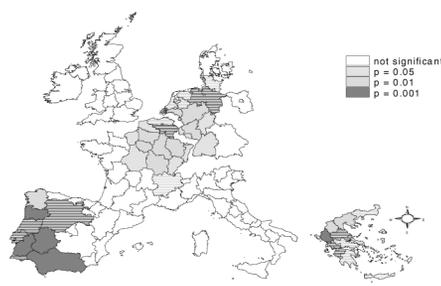
**Figura 5.** Significación del contraste local de Moran,  $I_i$ , Ln VAB por ocupado 1975



**Figura 6.** Distribución espacial del contraste local de Moran,  $I_i$ , Ln VAB por ocupado 1975



**Figura 7.** Significación del contraste  $New-G_i$ , Ln VAB por ocupado 1975



<sup>8</sup> En el caso de la  $I_i$ , se han calculado los pseudo-niveles de significación obtenidos de una distribución empírica derivada siguiendo un criterio de permutación.

Asimismo, cabe decir que únicamente el test  $I_i$  permite conocer la contribución de cada región al valor del contraste global  $I$  de Moran (cosa que no sucede con el  $New-G_i^*$  respecto al test global  $G(d)$ ). Este hecho hace posible detectar aquellas regiones con un  $I_i$  con una contribución excesiva al contraste global. En concreto, asumiremos el criterio de que únicamente tendrán una contribución elevada al test de la  $I$  de Moran aquellas regiones que presenten un valor del test  $I_i$  que sobrepase las bandas definidas por 1,5 veces el rango intercuartílico de la distribución de  $I_i$ , criterio utilizado por defecto en el programa *SpaceStat*. Ello nos lleva a detectar a las regiones portuguesas de Norte, Centro, Alentejo, Algarve y a la región griega de Borio Aigaio como concentraciones excesivas de valores bajos del logaritmo de la productividad laboral en 1975.

Llegado este punto, cabe decir que si bien se ha replicado el estudio para el año 1992, por razones de espacio no se presentarán los resultados obtenidos. En este sentido, únicamente cabe destacar la similitud de los resultados obtenidos, en términos de la persistencia de un claro esquema de dependencia espacial positiva en la distribución de la variable al final del período y de un mantenimiento de los dos grandes grupos de *clusters*, uno de valores bajos localizado básicamente en las regiones griegas y portuguesas, así como en las regiones españolas colindantes a aquellas y en algunas regiones del sur de Italia, y otro de valores elevados de productividad, ubicado en las regiones centrales de la UE. No obstante, en relación a este último grupo, cabe destacar que se ha detectado una cierta relocalización del centro europeo hacia las regiones del sur de Alemania y del norte de Italia, alejándose así de las regiones del norte de Francia. Finalmente, es preciso apuntar que se ha llevado a cabo también un análisis exploratorio de la tasa de crecimiento de la productividad laboral para el período 1975-1992, a partir del cual se ha rechazado la hipótesis de distribución aleatoria en el espacio a favor de la presencia de un esquema altamente significativo de dependencia espacial positiva.

## 6.2. Estudio de la dependencia espacial en la ecuación de $\beta$ -convergencia

Tras el análisis exploratorio de la variable  $\ln$  PIB por ocupado se ha comprobado como dicha variable muestra un esquema claramente significativo de dependencia espacial positiva, de manera que regiones próximas en el espacio tienden a mostrar valores similares de productividad laboral. Llegado este punto, cabría plantearnos cuáles son las razones que podrían explicar este resultado. En este sentido, cabría identificar dos tipos de factores explicativos de la asociación espacial detectada. En el primer grupo podríamos incluir a todos aquellos factores, ajenos o exógenos al comportamiento de las propias regiones, que afectarían por igual a regiones vecinas como, por ejemplo, factores de tipo institucional, climático, geográfico (ubicación geográfica respecto a los grandes mercados) o factores relacionados con la dotación de recursos. En cambio, el origen del segundo tipo de factores explicativos cabría encontrarlo en la presencia de externalidades que excediesen de las propias fronteras regionales.

En cualquier caso, es importante destacar que los resultados obtenidos en el análisis exploratorio nos alertan de la posible presencia de problemas de dependencia espacial en una ecuación, como la de convergencia antes mostrada, donde las variables principales son la productividad laboral y su tasa de crecimiento. Teniendo ello en cuenta, y bajo el supuesto de que no disponemos de ningún modelo teórico previo que incluya de partida la existencia de un esquema de dependencia espacial<sup>9</sup>, seguiremos la siguiente estrategia propuesta en Florax y Folmer (1992).

En primer lugar, estimaremos la ecuación de convergencia donde se regresa la tasa de crecimiento de la productividad para el período 1975-1992 contra la productividad en el año inicial, estimándola por MCO. Posteriormente, se calcularán los contrastes de autocorrelación espacial. En caso de que a partir de dichos contrastes se rechace la hipótesis nula, aceptaremos la ecuación de crecimiento tradicional como la correcta, concluyendo que no existe ningún tipo de interdependencia entre las regiones. Si tal fuera el caso, se concluiría que la tecnología y la productividad de una región dependen exclusivamente de factores propios de la región.

Si, por el contrario, el contraste I de Moran y el LM-ERR y su asociado robusto LM-EL fueran significativos, o al menos estos dos últimos mostrasen una probabilidad menor que la del test LM-LAG y su robusto LM-LE, estimaríamos el modelo del error espacial por máxima verosimilitud, donde el parámetro autorregresivo  $\lambda$  mediría la intensidad de la dependencia espacial entre los residuos. Una vez realizada esta estimación se debería contrastar la consistencia de este tipo de modelo mediante el cálculo del test de factores comunes (test de COMFAC)<sup>10</sup>. Asimismo, se podría incorporar un retardo espacial de la variable explicativa, es decir, del nivel inicial de productividad laboral ( $Wlny$ ), contrastando su significación mediante un contraste de la razón de verosimilitud. En cualquier caso, cabe decir que la autocorrelación espacial residual detectada podría ser consecuencia de la omisión de variables exógenas correlacionadas espacialmente o de la presencia de errores de medida derivados de una escasa correspondencia entre las fronteras administrativas de las unidades espaciales consideradas y las fronteras naturales en las que los fenómenos de crecimiento tienen lugar. Asimismo, este modelo implicaría, en términos económicos, que el crecimiento de una región  $i$  no se vería afectado de forma sistemática por el crecimiento en sus regiones vecinas sino únicamente cuando éstas últimas mostrasen un crecimiento anómalo (relacionado con shocks aleatorios o anticipados en sus vecinos).

Finalmente, si el contraste I de Moran y el LM-LAG y su asociado robusto LM-LE fueran significativos, o al menos estos dos últimos tuvieran un valor de la probabilidad menor que el LM-ERR y su robusto LM-EL, el modelo adecuado a estimar sería aquél que incluyera un retardo espacial de la variable endógena, habiendo de ser estimado también por máxima verosimilitud y donde el parámetro  $\gamma$  mediría la intensidad de las dependencias entre regiones vecinas. Llegada esta situación se podría considerar también la posibilidad de incorporar un retardo espacial de la varia-

<sup>9</sup> Ver Vayá *et al.* (2002) para un modelo de crecimiento económico en el que, de partida, se introduce la presencia de dependencia espacial fundamentada en la existencia de externalidades interregionales.

<sup>10</sup> Ver Anselin (1988a).

ble explicativa. Finalmente, cabe decir que si este fuera el modelo seleccionado, estaríamos concluyendo a favor de la existencia de interdependencias en el crecimiento entre regiones europeas vecinas derivadas, entre otros posibles factores, de la presencia de economías externas (estáticas y/o tecnológicas) que exceden de las fronteras regionales.

Con todo ello en cuenta, en primer lugar, se estima la ecuación de convergencia sin efectos espaciales, obteniendo los resultados presentados en el cuadro 5.

**Cuadro 5.** Estimación MCO de la ecuación de crecimiento para el caso regional de la UE

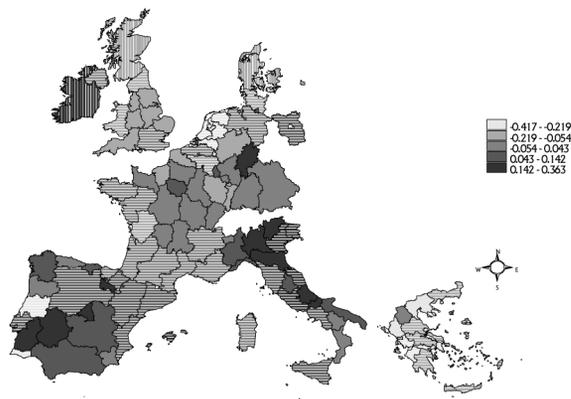
|                |                    |
|----------------|--------------------|
| Lny            | -0,3652<br>(0,029) |
| B implícita    | 0,0267             |
| R <sup>2</sup> | 0,6002             |
| lnL            | 61,62              |
| I-Moran        | 7,46*              |
| LM-ERR         | 49,50*             |
| LM-EL          | 9,94*              |
| LM-LAG         | 39,59*             |
| LM-LE          | 0,03               |

Notas: La variable dependiente es la diferencia logarítmica del PIB por ocupado entre 1975 y 1992 y lny es el logaritmo del PIB por ocupado en 1975. lnL es el valor del máximo del logaritmo de la función de verosimilitud. Desviación estándar entre paréntesis. \*Denota significativo al 1%.

Como se puede observar, la tasa de convergencia estimada se sitúa ligeramente por encima del 2% anual, en concreto el 2,7%. Dicho resultado se encuentra en el rango de valores obtenidos en otros trabajos utilizando una muestra de corte transversal. El ajuste del modelo (R<sup>2</sup> del 0,60) es razonable, si bien es preciso recordar la poca fiabilidad de dicha medida en caso de estar presente un esquema de autocorrelación espacial en los residuos del modelo. Sin embargo, a pesar del buen ajuste del modelo, existe algún tipo de problema de especificación. Así, tanto la I de Moran de los residuos como los contrastes habituales de multiplicadores de Lagrange rechazan las respectivas hipótesis nulas de ausencia de dependencia espacial, siendo el valor del contraste LM-ERR (49,5) mayor que el de LM-LAG (39,6), lo que sugeriría que debería reespecificarse el modelo incluyendo un esquema autorregresivo en el término de error. Esta conclusión parece confirmarse por los valores de las formas robustas de estos contrastes, ya que el test LM-EL sigue siendo significativo mientras que el LM-LE deja de serlo.

El mapa de los residuos obtenidos de la estimación MCO (figura 8) sugiere claramente la presencia de un esquema de dependencia espacial positiva, en la medida en que valores similares de dichos residuos se encuentran próximos en el espacio.

**Figura 8.** Mapa de la distribución espacial de los residuos MCO de la ecuación de convergencia



Sin embargo, tan solo la especificación formal de los contrastes nos indica el modelo espacial alternativo. Por tanto, reestimamos la ecuación de convergencia incluyendo una estructura autorregresiva espacial en el término de error. El resultado de esta estimación se muestra en la primera columna del cuadro 6. El parámetro autorregresivo espacial del término de error ( $\lambda$ ) es altamente significativo y positivo, confirmando lo que los contrastes de dependencia nos sugerían tras la estimación cuadrática. El contraste de la razón de verosimilitud de la dependencia espacial en el término de error nos lleva a la misma conclusión que el contraste de la *t* de la estimación máximo-verosímil. Asimismo, el contraste COMFAC de coherencia interna de la especificación del modelo del error espacial no es capaz de rechazar dicha coherencia, por lo que parece ser que el modelo del error espacial resulta más adecuado a la hora de estimar la ecuación de crecimiento para la muestra de las regiones europeas escogida. A pesar de ello, hemos procedido también a estimar el modelo incluyendo únicamente el retardo espacial, para confirmar que no es tan adecuado como el del error. Los resultados de dicha estimación se presentan en la segunda columna del cuadro 6. La velocidad de convergencia, en este caso, disminuye sustancialmente, pasando de un 2,6% en la estimación MCO en el cuadro 5 a un valor del 1,2% anual. El parámetro autorregresivo espacial del retardo de la endógena es positivo y significativo, resultado corroborado por el valor del estadístico de la razón de verosimilitud para dicho parámetro. De esta forma, parece rechazarse la hipótesis de que la inclusión del retardo de la endógena no es necesaria. Sin embargo, en términos de ajuste, la verosimilitud obtenida para el modelo del error espacial (85,79) es mayor que la del retardo espacial (82,37) lo que, junto con los resultados de los estadísticos, nos llevan a mantener que el modelo con una estructura autorregresiva en el término de error es más adecuado.

**Cuadro 6.** Estimación espacial de la ecuación de crecimiento para el caso regional de la UE

|                   | <i>Modelo del error espacial</i> | <i>Modelo del retardo espacial</i> |
|-------------------|----------------------------------|------------------------------------|
| Lny               | -0,3479<br>(0,043)               | -0,1846<br>(0,035)                 |
| $\beta$ implícita | 0,0252                           | 0,0120                             |
| $\gamma$          |                                  | 0,5692<br>(0,073)                  |
| $\lambda$         | 0,6453<br>(0,071)                |                                    |
| R <sup>2</sup>    | 0,5447                           | 0,6937                             |
| lnL               | 85,79                            | 82,37                              |
| LR-ERR            | 48,34*                           |                                    |
| LR-COMFAC         | 0,54                             |                                    |
| LR-LAG            |                                  | 41,49*                             |

Notas: La variable dependiente es la diferencia logarítmica del PIB por ocupado entre 1975 y 1992 y lny es el logaritmo del PIB por ocupado en 1975. lnL es el valor del logaritmo de la función de verosimilitud. Desviación estándar entre paréntesis. \*Denota significativo al 1%.

## 7. Conclusiones

Con el presente trabajo se ha perseguido el objetivo principal de contribuir a la difusión de las técnicas de econometría espacial en nuestro país, ofreciendo una somera visión de los aspectos teóricos y aplicados de la misma. Para ello, en primer lugar, se han revisado las principales técnicas econométricas existentes para el estudio de la dependencia espacial. En concreto, se han introducido los dos tipos de análisis de datos espaciales existentes: el análisis exploratorio, ideado para el estudio univariante, y el análisis de la dependencia espacial en el seno de un modelo de regresión. Asimismo, se ha incluido una breve reflexión acerca de algunas de las principales limitaciones de la Econometría Espacial en la actualidad.

En segundo lugar, y tomando como base la ecuación tradicional de  $\beta$ -convergencia aplicada al estudio de la productividad laboral en el seno de las regiones europeas para el período 1975-1992, se han presentado los resultados tanto del análisis exploratorio de la variable productividad laboral como del estudio de la presencia de dependencia espacial en dicha ecuación de crecimiento. Dos son las principales conclusiones que pueden extraerse del estudio realizado. Primero, la variable analizada presenta un claro esquema de dependencia espacial positiva entre regiones vecinas, de manera que regiones próximas en el espacio muestran niveles de productividad laboral similares. Y segundo, la ecuación clásica de convergencia adolece de importantes problemas de dependencia espacial que deben ser considerados de forma explícita y que afectarían tanto a su especificación como a los métodos de estimación que deberían utilizarse. Este último resultado se revela como la excusa perfecta para alertar a todos aquellos investigadores aplicados,

tanto del ámbito regional como urbano, acerca de la necesaria revisión y aplicación de las técnicas de econometría espacial que han sido recogidas o referenciadas en este trabajo.

## Bibliografía

- Anselin, L. (1980): «Estimation methods for spatial autoregressive structures». Ithaca NY: Cornell University, *Regional Science Dissertation and Monograph Series #8*.
- Anselin, L. (1988a): *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, The Netherlands.
- Anselin, L. (1988b): «Lagrange multiplier test diagnostic for spatial dependence and spatial heterogeneity», *Geographical Analysis*, 20(1):1-17.
- Anselin, L. (1992): «SpaceStat: A program for the analysis of spatial data», *National Center for Geographic Information and Analysis*. University of California, Santa Barbara, CA.
- Anselin, L. (1995a): «Local indicators of spatial association-LISA», *Geographical Analysis*, 27:93-115.
- Anselin, L. (1995b): *SpaceStat version 1.80 user's guide*. Morgantown, WV, Regional Research Institute, West Virginia University.
- Anselin, L. y Griffith, D.A. (1988): «Do spatial effects really matter in regression analysis?», *Papers Regional Science Association*, 65:11-34.
- Anselin, L. y Hudak, S. (1992): «Spatial econometrics in practice. A review of software options», *Regional Science and Urban Economics*, 22(3):509-536.
- Anselin, L. y Florax, R. (1995a): *New Directions in Spatial Econometrics*, Ed: Springer. Berlin.
- Anselin, L. y Florax, R. (1995b): «Small sample properties of tests for spatial dependence in regression models: some further results», *New Directions in Spatial Econometrics*, 21-74. Ed: Springer. Berlin.
- Anselin, L. y Bao, S. (1996): «SpaceStat.apr User's guide», *Regional Research Institute*, West Virginia University, Morgantown, WV.
- Anselin, L. y Bao, S. (1997): «Exploratory spatial data analysis linking SpaceStat and ArcView», en Fischer, M. y A. Getis (eds.): *Recent developments in spatial analysis-spatial statistics, behavioural modelling and neurocomputing*. Berlin, Springer-Verlag.
- Anselin, L. y Rey, S.J. (1997): «Introduction to the special issue on spatial econometrics», *International Regional Science Review*, 20(1, 2):1-8.
- Anselin, L. y Bera, A.K. (1998): «Spatial dependence in linear regression models with an introduction to spatial econometrics», *Handbook of Applied Economic Statistics* (eds.) Aman Ullah and D. E. A. Giles. New York: Marcel Dekker, Inc.
- Anselin, L. y Moreno, R. (2003): «Properties of test of spatial error components», *Regional Science and Urban Economics*. En prensa.
- Armstrong, Harvey W. (1995): «Convergence among regions of the European Union, 1950-1990», *Papers in Regional Science*, 74:143-152.
- Barro, R. y Sala-i-Martin, X. (1995): *Economic Growth*. McGraw-Hill, New York.
- Bivand, R.S. (1984): «Regression modeling with spatial dependence: an application of some class selection and estimation methods», *Geographical Analysis*, 16(1):25-37.
- Blommestein, H. (1983): «Specification and estimation of spatial econometric models: A discussion of alternative strategies for spatial economic modelling», *Regional Science and Urban Economics*, 13:251-270.
- Bodson, P. y Peeters, D. (1975): «Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation. An application to a Belgian labour-demand function», *Environment and Planning A*, 7:455-472.
- Case, A., Rosen, H. y Hines, J. (1993): «Budget spillovers and fiscal policy interdependence: evidence from the states», *Journal of Public Economics*, 52:285-307.

- Casetti, E. (1972): «Generating models by the expansion method: applications to geographical research», *Geographical Analysis*, 4:81-91.
- Cliff, A. y Ord, J. (1973): *Spatial Autocorrelation*. London, Pion.
- Cliff A. y Ord, J. (1981): *Spatial Process. Models and Applications*. London, Pion.
- Dacey, M. (1968): «A review of measures of contiguity for two and K-Color Maps», *In spatial analysis: a reader in statistical geography*, 479-495. Editado por B. Berry.
- Fingleton, Bernard y McCombie, John S.L. (1998): «Increasing returns and economic growth: Some evidence for manufacturing from the European Union regions», *Oxford Economic Papers*, 50:89-105.
- Florax, R. y Folmer, H. (1992): «Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators», *Regional Science and Urban Economics*, 22:404-432.
- Foster, S. y Gorr, W. (1983): «Adaptive filtering approaches to spatial modeling», *Modeling and Simulation*, 14:745-750.
- Foster, S. y Gorr, W. (1984): «Spatial adaptive filtering», *Modeling and Simulation*, 15:29-34.
- Foster, S. y Gorr, W. (1986): «An adaptive filter for estimating spatially-varying parameters: Application to modeling police hours spent in response to calls for service», *Management Science*, 32:878-889.
- Fotheringham, A.S., Charlton, M. y Brunson, C. (1998): «Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis», *Environment and Planning A*; 30(11):1905-27.
- Geary, R. (1954): «The contiguity ratio and statistical mapping», *The Incorporated Statistician*, 5:115-145.
- Getis, A. y Ord, J.(1992): «The analysis of spatial association by use of distance statistics», *Geographical Analysis*, 24:189-206.
- Haining, R. (1978): «Estimating spatial interaction models», *Environment and Planning A*, 10:305-320.
- Hildreth, C. y Houck, J. (1968): «Some estimators for a linear model with random coefficients», *Journal of the American Statistical Association*, 63:584-595.
- Kelejian, H.H. y Prucha, I.R. (1997): «A generalized moments estimator for the autoregressive parameter in a spatial model», *International Economic Review*, 40(2):509-536.
- Kelejian, H.H. y Robinson, D. (1993): «A suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a county expenditure Model», *Papers in Regional Science*, 72:297-312.
- Krugman, P. (1991): *Geography and Trade*. MIT Press, Cambridge MA.
- López-Bazo, E., Vayá, E., Mora, A.J. y Suriñach, J. (1999): «Regional economic dynamics and convergence in the European Union», *The Annals of Regional Science*, 33(3):343-370.
- Moran, P. (1948): «The interpretation of statistical maps», *Journal of the Royal Statistical Society B*, 10:243-251.
- Moreno, R. (1998): «Infraestructuras, externalidades y crecimiento regional. Algunas aportaciones para el caso español». Tesis Doctoral. Universidad de Barcelona.
- Moreno, R. y Vayá, E. (2000): *Técnicas Económicas para el tratamiento de datos espaciales: La econometría espacial*. UB 44 manuals, Edicions Universitat de Barcelona.
- Moreno, R., López-Bazo, E., Vayá, E. y Artís, M. (2002): «Externalities, public capital and costs of production». Próxima publicación como capítulo del libro *First Volume in a Series on Spatial Econometrics and Statistics*, Springer-Verlag, Eds.
- Ord, J.K. y Getis, A. (1995): «Local spatial autocorrelation statistics: distributional issues and an application», *Geographical Analysis*, 27(4):286-306.
- Pace, R.K. (1996): «Performing large spatial regressions and autoregressions», *Economic Letters*, 54:283-291.
- Pace, R.K. y Barry, R. (1996): «Sparse spatial autoregressions», *Statistics and Probability Letters*, 2158, 1-7.
- Pace, R. K. y Barry, R. (1997): «Quick computation of spatial autoregressive estimators», *Sgeographical Analysis*, 29:232-247.
- Paelinck, J.H.P y Klaassen, L.H. (1979): *Spatial Econometrics*. Farnborough, Saxon House.
- Quandt, R. (1958): «The estimation of the parameters of a linear regression system obeying two separate regimes», *Journal of the American Statistical Association*, 53:873-880.

## 106 Rosina Moreno y Esther Vay 

- Rey, Sergio y Montouri, Brett D. (1999): «U.S. regional income convergence: A spatial econometric perspective», *Regional Studies*, 33:143-156.
- Student (1914): «The elimination of spurious correlation due to position in time or space», *Biometrika*, 10:179-180.
- Vay , E. (1998): «Localizaci n, crecimiento y externalidades regionales. Una propuesta basada en la econometr a espacial», Tesis Doctoral, Universidad de Barcelona.
- Vay , E., L pez-Bazo, E. y Art s, M. (1998): «Growth, convergence and (why not?) regional externalities», *Working Paper* n m. E98/31, Divisi  II, Universidad de Barcelona.
- Vay , E., L pez-Bazo, E., Moreno, R. y Suri ach, J. (2002): «Growth and externalities across economies. An empirical analysis using spatial econometrics». Pr xima publicaci n como cap tulo del libro *First Volume in a Series on Spatial Econometrics and Statistics*, Springer-Verlag Eds.
- Ying, Long G. (2000): «Measuring the spillover effects: some Chinese evidence», *Papers in Regional Science*, 79:75-89.