MODELOS DE REGRESIÓN ESPACIO TEMPORALES EN LA ESTIMACIÓN MUNICIPAL DE LA RENTA. ESTIMACIÓN DE LA RFBDpc MUNICIPAL EN LA REGION DE MURCIA.

López Hernández, Fernando A.
Departamento de Métodos Cuantitativo e Informáticos
Universidad Politécnica de Cartagena.
e-mail: Fernando.Lopez@upct.es

García Córdoba, José Antonio.
Departamento de Métodos Cuantitativo e Informáticos
Universidad Politécnica de Cartagena.
e-mail: Josea.Garcia@upct.es

Ruiz Marín, Manuel.
Departamento de Métodos Cuantitativo e Informáticos
Universidad Politécnica de Cartagena.
e-mail: manuel.ruiz@upct.es

Resumen

.
Muy recientemente han aparecido estudios en los que haciendo uso de técnicas de econometría espacial, se han realizado nuevas estimaciones de la renta municipal eliminando los problemas de autocorrelación espacial que presentaban los modelos anteriores. Estos modelos poseen una elevada capacidad explicativa, pero tienen ciertas carencias al utilizarlos con intención predictiva.
El objetivo de este estudio es plantear una alternativa a la estimación de la renta municipal mediante modelos espacio-temporales de regresión que simultáneamente solucione dos problemas: el problema de autocorrelación espacial y a la falta de capacidad predictiva de ciertos modelos espaciales de regresión. Se utiliza el modelo teórico desarrollado con el fin de realizar una estimación de la Renta Familiar Bruta Disponible per cápita de los municipios de la Región de Murcia.

Palabras Clave: Econometría Espacial, Renta Familiar Disponible, Región de Murcia.

1. Introducción.

La renta per cápita es el indicador sintético más utilizado para medir el grado de desarrollo de la población que habita una determinada área geográfica. Sin embargo, su estimación presenta importantes dificultades cuando la escala territorial del análisis es inferior a la provincial. El método que con más frecuencia se ha utilizado para la estimación de este indicador de desarrollo municipal ha sido el método indirecto basado en una estimación econométrica. Básicamente esta metodología consiste en la realización de una estimación econométrica a un nivel geográfico superior, el provincial, que posteriormente se extrapola a un nivel geográfico inferior, el municipal.

Para la realización de estas estimaciones en ningún caso se atiende a los problemas de autocorrelación espacial que previsiblemente presente la estimación mínimo cuadrática del modelo econométrico. Con la intención de solventar este problema, en los últimos dos años han aparecido diversos estudios en los que haciendo uso de técnicas de econometría espacial, se han realizado nuevas estimaciones de la renta municipal (Alañón A. 2002, Chasco C. 2003). Mediante estos nuevos modelos de regresión se eliminan los problemas de autocorrelación espacial que efectivamente presentan los modelos anteriores.

Los modelos de regresión espacial utilizados para llevar a cabo estas estimaciones, no sólo han resuelto el problema de la dependencia espacial en los residuos, sino que han aportado también un elevado poder explicativo en cuanto al comportamiento de la distribución espacial de la renta. A pesar de las ventajas que aporta el uso de esta técnica, tanto en cuanto a que dota a los estimadores de óptimas propiedades como de la nueva perspectiva que introduce, este tipo de modelos presenta ciertas carencias cuando se utilizan con intención predictiva.

El objetivo de este artículo es plantear una alternativa a la estimación de la renta que solucione simultáneamente dos problemas: el problema de la autocorrelación espacial en los modelos estimados por mínimos cuadrados y el problema de la incapacidad predictiva del modelo espacial de regresión con estructura autorregresiva residual.

La alternativa que se propone en este trabajo viene de la mano de los modelos espacio-temporales de regresión, conocidos en econometría espacial como SUR (Seemingly Unrelated Regression) espaciales. Estas estructuras utilizan la doble dimensión espacio temporal que permite a la vez, eliminar los problemas de autocorrelación espacial y dotar de una mayor capacidad predictiva a los modelos espaciales de regresión.

A pesar de que el modelo SUR espacial tiene una especificación muy genérica que hace de ellos modelos con un alto grado de complejidad a la hora de realizar la estimación de los parámetros, en el caso que se propone se han introducido una serie de restricciones sobre este modelo general que le confieren una mayor operatividad a la vez que se adaptan al problema concreto que se desea abordar. Por último, se utiliza el modelo teórico desarrollado con el fin de realizar una estimación de la Renta Familiar Bruta Disponible per cápita de los municipios de la Región de Murcia (RFBDpc).

El trabajo se desarrolla siguiendo el siguiente guión: En el segundo apartado se presentan una serie de reflexiones sobre la capacidad explicativa y predictiva de los modelos espaciales de regresión. En la tercera sección se especifica el modelo general de regresión espacio temporal y se establecen las restricciones que determinan el modelo concreto que será desarrollado. En el cuarto apartado se realiza la estimación de la RFBDpc a nivel provincial. En el quinto apartado se realiza la extrapolación para el caso de los municipios de la Región de Murcia. El artículo se cierra presentando las conclusiones y desarrollos futuros.

2. Modelos espaciales de regresión: Capacidad explicativa versus capacidad predictiva.

En el análisis de los procesos económicos son muchas las situaciones en las que las unidades en que se suministra la información coincide con las divisiones geopolíticas de las distintas regiones. En este caso, el investigador únicamente puede seleccionar el nivel de desagregación espacial en el que se toma la muestra transversal. Es habitual que en este tipo de observaciones esté presente el fenómeno de pérdida de independencia de las observaciones fruto de su localización, bien por falta de concordancia entre la unidad de observación y la extensión del fenómeno, bien por los naturales flujos y reflujos entre economías vecinas.

Es en esta situación, en la que un determinado fenómeno se extiende sobre las regiones vecinas, cuando se presenta la llamada autocorrelación espacial positiva (Cliff y Ord 1981) de tal forma que, el hecho de encontrar una determinada observación en cierta localización, hace más verosímil de lo que sería normal, encontrar observaciones semejantes en su entorno.

La modelización de este tipo de estructuras en el espacio discreto viene de la mano de la llamada matriz de pesos o ponderaciones W. Esta es una matriz cuadrada que se introduce como un factor exógeno del modelo, y cuyos elementos w_{ij} reflejan a criterio del investigador, la intensidad de la relación entre las regiones i y j. Mediante estos coeficientes w_{ij} pueden plantearse relaciones bidireccionales y asimétricas entre las distintas regiones en estudio.

La matriz W se introduce en la especificación del modelo de regresión clásico con el fin de recoger la estructura de dependencia espacial del proceso económico. Estos son los modelos de regresión espaciales. Hay una amplia tipología de modelos de regresión espaciales que recogen la interdependencia que entre distintas unidades (regiones, distritos, etc.) una buena parte de ellos pueden consultarse en Anselin 1988 o Moreno y Vayá 2000.

Dentro de esta amplia variedad de modelos deben destacarse dos estructuras que con más frecuencia se utilizan en econometría espacial: los llamados modelos de retardo espacial que también llamaremos de *comunicación o contagio* y modelos del error espacial o de *comparación*.

Los primeros, los modelos de retardo espacial, recogen la estructura de dependencia espacial del proceso mediante la inclusión de un retardo espacial como factor explicativo de la variable endógena, con la siguiente especificación:

$$\begin{gathered} \mathbf{Y} = \rho\mathbf{WY} + \mathbf{Xb} + \mathbf{e} \\ \mathbf{e} \sim N(0,\sigma^2\mathbf{I}) \end{gathered} \qquad (1)$$

Donde como es habitual, Y es un vector columna nx1, X es una matriz kxn que recoge una serie de variables exógenas (donde opcionalmente se puede incluir una columna de unos para el término constante), W es una matriz de conexiones nxn exógena que define la estructura de vecindades, WY es el retardo espacial de la variable Y.

La magnitud y el signo del parámetro de dependencia espacial viene recogida en el modelo por un parámetro autorregresivo ρ que determina tanto la intensidad como el carácter positivo o negativo de esta dependencia. Finalmente, el vector **e** se corresponde con el término de perturbación que se supone un ruido blanco.

La segunda de especificación más utilizada, a la que nos referiremos como modelos de *comparación,* introduce la estructura de dependencia espacial en el término de perturbación del modelo atendiendo a la siguiente especificación:

$$\begin{aligned} \mathbf{Y} &= \mathbf{Xb} + \mathbf{u} \\ \mathbf{u} &= \rho \mathbf{W} \mathbf{u} + \mathbf{e} \\ \mathbf{e} &\sim N(0,\sigma^2\mathbf{I}) \end{aligned} \tag{2}$$

como en el caso anterior Y es un vector columna nx1, X es una matriz kxn que recoge una serie de variables exógenas, W es una matriz de conexiones nxn exógena que define la estructura de vecindades y ρ es el parámetro de dependencia espacial.

Capacidad Explicativa

Las dos especificaciones anteriores pueden expresarse de manera que muestren de una forma más transparente la estructura autorregresiva que se especifica. Así el primero de los modelos descritos (1) puede expresarse como:

$$y_i = \mu_i + \rho \sum_{j \in N_i} w_{ij} y_j + e_i \qquad i=1,...,n \tag{3}$$

Donde por μ_i se denotará la parte de y_i que es explicada a través de los factores exógenos. En nuestro caso:

$$\mu_i = \beta_1 x_{1i} + \ldots + \beta_k x_{ki} \qquad (4)$$

por N_i se denota el conjunto de índices de las zonas que se consideran vecinas o que interactúan con la zona i.

$$N_i = \{k \,;\, k \text{ es vecino de } i\}.$$

De esta forma, el valor de la variable endógena localizado en la zona i-ésima puede interpretarse como el resultado de dos contribuciones:

En primer lugar, y por esto se entiende como modelo de *comunicación* o *contagio*, la influencia que sobre la propia variable endógena ejercen las zonas vecinas con las que interactúa. Esta influencia se recoge mediante una media ponderada - a través de los valores w_{ij} con la que se cuantifica la influencia que ejerce la zona j sobre la i - de los valores de la propia variable en las zonas vecinas. Los valores asignados a los elementos de la matriz w_{ij} como anteriormente se ha comentado, dependen del criterio del investigador y en cierta forma recogen su teoría sobre la estructura de interrelaciones. Este factor de *comunicación,* se pondera por un coeficiente ρ[1] estimado en el modelo.

En segundo lugar y como en un modelo de regresión clásico, se recoge el efecto de una serie de variables exógenas localizadas en la zona i-ésima (μ_i), al que finalmente se acompaña de un término de error.

El segundo de los modelos (2), el modelo del error espacial, también puede expresarse de una forma más transparente como:

$$y_i = \mu_i + \rho \sum_{j \in N_i} w_{ij} \left(y_j - \mu_j \right) + e_i \qquad i=1, \ldots, n. \qquad (5)$$

En esta segunda especificación, como en el caso anterior, el valor de la variable endógena localizado en la i-ésima zona es el resultado de dos contribuciones. La primera de estas contribuciones es la tendencia del proceso[2] (μ_i) que como es natural recoge la influencia de los factores exógenos en la zona i. La segunda de las

[1] Si la matriz W está estandarizada por filas como suele ser habitual, este coeficiente oscilará entre -1 y 1.
[2] Esta es efectivamente la media del procesos estocástico espacial $\{Y_i : i=1,\ldots,n\}$ implícito.

contribuciones se trata de una modificación de la tendencia del proceso fruto de la comparación con su entrono. Este segundo término *corregirá* el valor de la tendencia en función de los valores obtenidos en localizaciones que sean vecinas dependiendo de sí son superiores o inferiores a su valor esperado. Así, aquellas zonas vecinas infravaloradas por la tendencia del proceso contribuirán de forma positiva a incrementar el valor de la variable en la zona i-ésima, mientras que aquellas zonas vecinas $\{y_j;\ j\chi N_i\}$ sobrevaloradas por la tendencia del proceso (μ_j) contribuirán de forma negativa y disminuirán el valor de la variable en la localización "i". Al igual que en el caso anterior, vendrá ponderado por un coeficiente ρ de dependencia espacial estimado en el modelo, y que podrá ser de signo positivo o negativo dependiendo del tipo de autocorrelación presente.

Estos dos modelos, al igual que otras especificaciones que introducen un mayor grado de complejidad, poseen una elevada capacidad explicativa del proceso que se analiza. Por un lado determinan las contribuciones que cada zona aporta a aquellas con las que se relaciona, y por otro, la aportación que recibe cada zona de las zonas vecinas con las que interacciona. Esta interdependencia se cuantifica mediante la estimación del parámetro ρ de dependencia espacial. Se recoge así por tanto, el efecto mimetismo producto de las interrelaciones y flujos de intercambio entre regiones, natural en cualquier proceso estocástico de carácter socioeconómico que se analice en el espacio.

Dependiendo del tipo de autocorrelación presente en la variable, autocorrelación espacial sustantiva o residual será adecuada una u otra especificación. En este sentido hay una amplia batería de contrastes (Moreno y Vayá 2000) con capacidad para detectar cuál de las dos estructuras es mas adecuada al problema que se analice y que no son objeto de este trabajo.

Capacidad Predictiva.

Demostrada la capacidad de explicar la forma en que se expande sobre el espacio el proceso estocástico, cabría preguntarse cuál es la capacidad predictiva de estos modelos.

En primer lugar, debemos notar que la palabra predicción está asociada al tiempo: con la información presente es posible determinar la información futura. Habituados a los modelos temporales quizás no tenga sentido *predecir* con modelos con un solo corte temporal. Debe por tanto, en una primera reflexión, analizarse cuál es el significado de la predicción en el contexto espacial.

¿Qué sentido tiene la predicción en el contexto espacial?. En primer lugar, el modelo estimado podría utilizarse para predecir valores de localizaciones no observadas de forma equivalente al proceso de Krikeage de la Estadística Espacial. Esta estimación se aplicaría bien sobre áreas o regiones de la superficie analizada para la que no se dispusiera de información o bien sobre zonas o regiones frontera. Este primer uso de los modelos de regresión espaciales, se entiende como interpolación.

En segundo lugar, el modelo puede utilizarse, desplazándolo sobre otra superficie y suponiendo que el comportamiento de los factores es semejante. Esta es la principal utilidad predictiva de los modelos de regresión espaciales y será en este sentido en el que se desarrollará este trabajo.

En cualquiera de los casos, los dos modelos descritos en el apartado anterior presentan diferente comportamiento en cuanto a su capacidad predictiva. Desde un punto de vista teórico ambos utilizan la propia variable endógena como factor explicativo del proceso, pero en el caso del primero de los modelos señalados, el modelo de *comunicación*, podemos hacer uso de algún artificio aritmético para conferirle poder predictivo.

Así, para especificaciones con retardo de la endógena, es posible re-especificar el modelo con el fin de evitar la presencia de la endógena espacialmente retardada como variable explicativa del modelo, así:

$$\hat{Y} = (I - \hat{\rho}W)^{-1}X\hat{\beta} \tag{6}$$

En este caso se ha eludido la presencia de la propia variable dependiente y es posible utilizar esta formulación con fines de predicción y extrapolación.

Por el contrario en el caso del modelo con estructura autorregresiva en los errores, o modelo de *comparación*, las estimaciones que obtendríamos deben involucrar obligatoriamente a la propia variable endógena, y no es posible utilizar el recurso del anterior modelo:

$$\hat{Y} = \hat{\rho}WY + (I - \hat{\rho}W)X\hat{\beta} = X\hat{\beta} + \hat{\rho}W(Y - X\hat{\beta}) \qquad (7)$$

Si una vez estimado este modelo, se desea trasladar los resultados obtenidos de un nivel geográfico a otro de nivel inferior como es el caso de la metodología utilizada en la estimación de la RFBDpc resulta imposible ya que es necesario el propio valor que se desea estimar como puede apreciarse en (7). Cabría entonces preguntarse por otras alternativas distintas que confirieran poder predictivo a este modelo.

Este es el objetivo de este trabajo, sugerir alternativas planteando posibles vías para solucionar el problema y evaluar su viabilidad. Podían sugerirse distintas alternativas:

Una primera idea sería asignarle a Y algún valor anterior en el tiempo. Esta sería la solución más simple y también la menos formal.

Una segunda idea sería realizar un proceso iterativo de estimación de la variable. Asignándole a Y en primer lugar la tendencia del proceso y progresivamente rectificando su valor, puesto que como hemos dicho se tarta de un modelo de comparación.

En tercer lugar, y esta es la línea que proponemos, para conferirle un verdadero poder predictivo a los modelos espaciales de regresión, es necesario involucrar la dimensión temporal. Por tanto es necesario introducir los modelos de regresión espacio temporales.

Por otra parte, el modelo de regresión con estructura autorregresiva en los errores (2) presupone que si la autocorrelación espacial es debida a flujos de intercambio entre zonas colindantes, estos se producen en el mismo instante o intervalo de tiempo en que se analiza el proceso. Esta hipótesis de transferencia inmediata no tiene porqué ser adecuada para todos los procesos económicos. Por el contrario, lo natural es que exista cierto desfase temporal, más o menos corto, en el que se produzca la transmisión e intercambio de la variable endógena entre las distintas zonas que interactúan.

3. Los modelos SUR.

Los modelos SUR, originalmente propuestos por Séller (1962) están diseñados para recoger interacción en una doble dimensión la espacial y la temporal. Una completa descripción de estos modelos la podemos encontrar en Anselín 1988.

Como punto de partida consideremos la siguiente especificación del modelo de regresión que expresa una amplia variedad de estructuras espacio temporales.

$$y_{it} = X_{it}\,\beta_{it} + e_{it} \quad ; \quad E[e_{it}] = 0 \quad ; E[e_{it}\ e_{js}] \neq 0 \ ; \quad i = 1,\ldots,n \ ; t = 1,\ldots,T. \qquad (8)$$

Siendo y_{it} una observación de la variable endógena en la unidad espacial i (i=1,…,n) y en el periodo temporal t (t=1,…,T), X_{it} es un vector fila de observaciones para la unidad espacial i en el instante t, β_{it} es un vector de parámetros espacio temporal, y e_{it} es el término de error.

Una expresión particular de este modelo general, es el conocido como SUR espacial:

$$y_{it} = X_{it}\,\beta_t + e_{it} \quad ; \quad E[e_{it}] = 0 \ ; E[e_{it}\ e_{is}] = \sigma_{ts}; \quad i = 1,\ldots,n \ ; t = 1,\ldots,T. \qquad (9)$$

En forma matricial, la ecuación para cada periodo de tiempo t será:

$$Y_t = X_t\,\beta_t + e_t \qquad (10)$$

Donde Y_t y e_t son vectores nx1, X_t es una matriz n por k_t de variables explicativas. En este caso, el número de variables exógenas, k_t puede ser diferente para cada ecuación (cada periodo de tiempo).

En este caso β_t es un vector de coeficientes que toma diferentes valores para cada unidad temporal, pero que se considera constante en el espacio. Este tipo de modelos será apropiado en aquellas situaciones en las que el número de periodos temporales T sea inferior al número de observaciones espaciales (n>T) como suele ocurrir en Ciencia Regional.

Modelo SUR espacial con autocorrelación espacial residual.

Como este modelo SUR espacial consta de una ecuación para cada instante de tiempo, es posible introducir una estructura autorregresiva de forma semejante a lo se hace en el modelo (2) mediante un esquema autorregresivo de primer orden:

$$\begin{aligned} Y_t &= X_t \beta_t + u_t \\ u_t &= \rho_t W u_t + e_t \end{aligned} \tag{11}$$

Donde

$$E[e_t\, e_s'] = \sigma_{ts}\, I \tag{12}$$

Este modelo es el llamado Modelo SUR espacial con esquema autorregresivo residual.

El modelo que se propone en este trabajo es una simplificación de este modelo con el objetivo de poder salvar los complejos problemas de inferencia que plantea la estimación de este tipo de estructuras para las que no hay software estándar capaz de realizarla. Las simplificaciones que se van a introducir, tienen con objetivo poder realizar el proceso de estimación mediante el software específico de econometría espacial SpaceStat.

Son dos las restricciones que se introducen en este modelo:

En primer lugar, se supondrá que el coeficiente de dependencia espacial ρ_t es constante en el tiempo, y por tanto, que el grado y la intensidad de la dependencia espacial no varía en la dimensión. Este supuesto de estabilidad temporal de la dependencia espacial que puede ser restrictivo cuando se analizan periodos de tiempo amplios, presumiblemente no lo es cuando los periodos de tiempo son muy próximos, como será el caso que posteriormente se analizará.

En segundo lugar, se impondrá la restricción de ausencia de correlación temporal en los residuos del modelo, esto es:

$$\begin{aligned} E[e_t\, e_s'] &= 0 \text{ si } s \neq t \\ E[e_t\, e_t'] &= \sigma^2 I_n \end{aligned} \tag{13}$$

Esta restricción no resta al modelo de cierto grado de interrelación espacio temporal como veremos en el siguiente desarrollo.

Con el fin de simplificar la notación y puesto que la aplicación empírica que se desarrolla en el siguiente apartado se realiza para el caso de dos periodos de tiempo, se puede escribir:

$$\begin{bmatrix} Y_t \\ Y_{t+1} \end{bmatrix} = \begin{bmatrix} X_t & 0 \\ 0 & X_{t+1} \end{bmatrix} \begin{bmatrix} \beta_t \\ \beta_{t+1} \end{bmatrix} + \begin{bmatrix} u_t \\ u_{t+1} \end{bmatrix} ; \begin{bmatrix} u_t \\ u_{t+1} \end{bmatrix} = \rho \begin{bmatrix} W_{tt} & 0 \\ W_{tt+1} & W_{t+1t+1} \end{bmatrix} \begin{bmatrix} u_t \\ u_{t+1} \end{bmatrix} + \begin{bmatrix} e_t \\ e_{t+1} \end{bmatrix} \quad (14)$$

Siendo Y_i un vector nx1 observado en dos instantes de tiempo i = t, t+1, X_i la matriz de orden $k_i$xn que recoge los k_i factores exógenos, β_i es un vector columna de orden k_i que incluye los coeficientes que acompañan al vector X_i (en este caso no supondremos la presencia de término independiente) y ρ es el coeficiente de dependencia espacial.

La matriz que determina la estructura de vecindades se ha definido por bloques con la siguiente notación:

$$W^* = \begin{bmatrix} W_{tt} & 0 \\ W_{tt+1} & W_{t+1t+1} \end{bmatrix} \quad (15)$$

Esta matriz W* es una matriz definida por bloques de orden 2x2 de tal forma que sea triangular inferior por bloques, con el fin de no dar lugar a relaciones futuras. Cada uno de estos bloques está formado por una matriz W_{ij} de orden nxn mediante la que el investigador cuantifica la estructura de vecindades en cada periodo de tiempo para los elementos de la diagonal principal, y para los elementos debajo de la diagonal principal pueden también establecerse de vecindades que recojan la interacción espacio temporal.

Finalmente, supondremos que los residuos del modelo se ajustan a una distribución Normal

$$e = [e_t \; e_{t+1}]' \sim N(0, \sigma^2 I_{2n}) \quad (16)$$

La especificación de la matriz W* es determinante en la construcción del modelo de regresión espacio temporal. Se analizarán en este trabajo tres alternativas diferentes. En primer lugar supondremos que la estructura de vecindades y las redes de interacción existentes entre las distintas unidades espaciales no varían en el tiempo, de tal forma que W_{ij} = W. En este caso, es posible nombrar tres matrices de interacción espacio temporal:

$$W_A = \begin{bmatrix} W & 0 \\ 0 & W \end{bmatrix} \quad ; \quad W_B = \begin{bmatrix} W & 0 \\ W & W \end{bmatrix} \quad ; \quad W_C = \begin{bmatrix} W & 0 \\ W & 0 \end{bmatrix} \tag{17}$$

La primera de estas matrices W_A daría lugar a un modelo que sólo contemple la dependencia espacial contemporánea, de manera que la dependencia espacial entre distintos instantes de tiempo se considera nula.

$$\begin{gathered} Y_t = X_t\beta_t + u_t \ ; \ \ u_t = \rho W u_t + e_t \\ Y_{t+1} = X_{t+1}\beta_{t+1} + u_{t+1} \ ; \ \ u_{t+1} = \rho W u_{t+1} + e_{t+1} \end{gathered} \tag{18}$$

La segunda matriz W_B da lugar a una dependencia espacial en el instante t y a una doble dependencia espacial y temporal en el instante t+1.

$$\begin{gathered} Y_t = X_t\beta_t + u_t \ ; \ \ u_t = \rho W u_t + e_t \\ Y_{t+1} = X_{t+1}\beta_{t+1} + u_{t+1} \ ; \ \ u_{t+1} = \rho W u_t + \rho W u_{t+1} + e_{t+1} \end{gathered} \tag{19}$$

Finalmente, la matriz W_C daría lugar a una dependencia espacial en el instante t y a una dependencia espacio-temporal en t+1 con un retardo temporal

$$\begin{gathered} Y_t = X_t\beta_t + u_t \ ; \ \ u_t = \rho W u_t + e_t \\ Y_{t+1} = X_{t+1}\beta_{t+1} + u_{t+1} \ ; \ \ u_{t+1} = \rho W u_t + e_{t+1} \end{gathered} \tag{20}$$

Este último modelo permite realizar las siguientes operaciones al sustituir valores de la primera ecuación en la segunda:

$$Y_{t+1} = X_{t+1}\beta_{t+1} + \rho W(Y_t - X_t\beta_t) + e_{t+1} \tag{21}$$

Si escribimos para cada una de las localizaciones el valor estimado según la ecuación anterior:

$$y_{i\,t+1} = \mu_{i\,t+1} + \rho \sum_{j \in N_i} w_{ij} \left(y_{jt} - \mu_{jt} \right) + e_{it+1} \qquad i=1,...,n \tag{22}$$

Donde por μ_{it} hemos denotado la tendencia del modelo, en base a los correspondientes factores exógenos.

$$\mu_{i\,t} = \sum_{j=1}^{k_t} \beta_{jt} x_{jti} \tag{23}$$

Estas dos últimas ecuaciones permiten conferirle capacidad predictiva al modelo de regresión espacio temporal, de tal forma que el valor de la variable en un instante t+1 y en una localización i vendrá determinado por la siguiente expresión:

$$\hat{y}_{i\,t+1} = \hat{\mu}_{i\,t+1} + \hat{\rho} \sum_{j \in N(i)} w_{ij} \left(y_{jt} - \hat{\mu}_{jt} \right) \qquad i=1,...,n \tag{24}$$

Con

$$\hat{\mu}_{i\,t} = \sum_{j=1}^{k_t} \hat{\beta}_{jt} x_{jti} \tag{25}$$

La estimación de los parámetros que determinan este modelo puede realizarse mediante el software Spacestat. Las dos restricciones introducidas en el modelo inducen una distribución Normal Multivariante en el proceso estocástico subyacente $Y \equiv N(X\beta;\sigma^2[(I-\rho W^*)^T(I-\rho W^*)]^{-1})$ que permite realizar la estimación Máximo verosímil de los parámetros del modelo con este software.

4. Estimación de la RFBDpc a nivel provincial.

En este apartado se realizará una estimación de la RFBDpc a nivel provincial mediante un modelo SUR espacial con estructura autorregresiva en los errores introduciendo las restricciones indicadas en el apartado anterior. El modelo estimado se utilizará posteriormente para realizar una extrapolación a los municipios de la Región de Murcia, obteniendo la estimación de la RFBDpc de cada uno de ellos.

La metodología que se utiliza para realizar esta estimación de la RFBDpc a nivel provincial, es el método basado en las técnicas de regresión espaciales, utilizando como factores explicativos de la Renta, las variables *clásicas* en este tipo de estimaciones. De los indicadores seleccionados se debe disponer de información al doble nivel de desagregación espacial, el provincial y el municipal, para el periodo de tiempo señalado. En este trabajo se considera el periodo 96-98 puesto que la última estimación de la RFBDpc en la Región de Murcia se refiere al año 1996.

De forma esquemática, son dos las etapas en las que se va a realizar el proceso de estimación: En la primera etapa se realizará la selección de indicadores para el periodo en estudio (años 1996 y 1998) que de forma significativa expliquen el comportamiento de la Renta atendiendo al modelo especificado en el apartado anterior. La estimación del modelo (14) se realiza a nivel provincial con las tres matrices de contacto seleccionadas (17), estableciendo comparaciones en cuanto a la verosimilitud de la estimación de cada una de ellas. Se realizará la selección del modelo que se considere mas adecuado.

En la segunda etapa, los resultados obtenidos en el paso anterior, se extrapolarán a los municipios de la Región de Murcia, obteniendo una estimación de la RFBDpc municipal para esta Región en el año 1998.

Modelo SUR espacial: Selección de indicadores.

La variable Renta, exógena en nuestro modelo, se ha obtenido de la publicación de la Fundación BBVA. Las variables que se introducen como factores explicativos de la RFBDpc a tanto nivel provincial como municipal se han extraído de distintas fuentes: El Anuario Comercial de España elaborado por el Instituto Lawrence R. Klein. (varios años). Con esta importante fuente de datos es posible disponer de las variables a un doble nivel geográfico que permitirá realizar la extrapolación del nivel provincial al nivel municipal. También se ha recogido información sobre el IVA del Instituto de Estudios Fiscales (http://www.minhac.es/ief/principal.htm) y de Econet, el servidor económico-estadístico de la Región de Murcia http://www.carm.es/chac/dgep/econet/

Los criterios utilizados para seleccionar los indicadores explicativos del nivel de Renta han sido los siguientes:

1. Los indicadores seleccionados como factores explicativos del modelo deben ser indicadores de Renta al doble nivel geográfico. No sólo deben ser significativos estadísticamente en el modelo que se plantee a nivel provincial, sino que deben ser variables que expliquen el nivel de renta también a nivel municipal ya que el objetivo es extrapolar los resultados a la RFBDpc de los municipios y no para las provincias.

2. Con el fin de dotar de estabilidad temporal a la estimación y aunque el modelo permite utilizar distintas variables exógenas para distintos años, se ha considerado adecuado introducir los mismos factores para los dos periodos de tiempo considerados. Otras alternativas estudiadas, no aumentaban la verosimilitud del modelo de forma considerable.

Tras introducir un amplio grupo de variables en el modelo de regresión, las variables que finalmente se han seleccionado han sido cuatro:

- **IVA**: Valor del IVA repercutido (miles de euros) por mil habitantes.
- **AU:** Automóviles por mil habitantes.
- **TF:** Líneas Telefónicas por mil habitantes.
- **CA:** Número de Camiones por mil habitantes.

Modelo SUR espacial: Matriz de vecindades.

Tal y como se planteaba en el apartado anterior se han considerado tres matrices de contactos que definidas por boques nombramos como,

$$W_A = \begin{bmatrix} W & 0 \\ 0 & W \end{bmatrix} ; W_B = \begin{bmatrix} W & 0 \\ W & W \end{bmatrix} ; W_C = \begin{bmatrix} W & 0 \\ W & 0 \end{bmatrix} \tag{26}$$

Donde cada uno de estos bloque es una matriz nxn W={w_{ij}} definida como:

$$w_{ij} = \begin{cases} \dfrac{P_j}{P_i} \text{ si las provincias i,j tienen frontera comun.} \\ 0 \text{ si las provincias no tienen frontera comun.} \end{cases} \tag{27}$$

Donde por P_i se denota la población de la provincia i-ésima en el año 1996.

Esta matriz se estandariza por filas como suele ser habitual, de tal forma que finalmente la matriz W con la que se desarrolla el modelo de regresión espacial pondera mediante la siguiente regla:

$$w_{ij} = \begin{cases} \dfrac{P_j}{\sum_{k \in N_i} P_k} & \text{si las provincias i,j tienen frontera comun.} \\ 0 & \text{si las provincias no tienen frontera comun.} \end{cases} \quad (28)$$

Esta definición de matriz W que introduce las poblaciones de cada provincia, frente a la alternativa más habitual de asignarle valores de 0 ó 1, tiene un doble objetivo. En primer lugar, con la matriz W así definida el retardo espacial WY es una definición correcta de retardo espacial. En este caso, el elemento i-ésimo del vector WY es efectivamente la RFBDpc media de los vecinos de la zona i y no la media aritmética de las RFBDpc.

Por otra parte, la matriz así definida da lugar a relaciones asimétricas de tal forma que la influencia que ejerzan dos provincias (o municipios) será función de las poblaciones que posean cada uno de ellos. Las regiones vecinas de una determinada zona i ejercerán su influencia en función de la población que posean. Se ha comprobado que estas matrices de relaciones asimétricas aumentan significativamente la verosimilitud del modelo de regresión.

En cuanto a las matrices de dependencia espacial utilizadas, se han realizado las regresiones con las tres matrices W_A, W_B y W_C propuestas en el apartado anterior. De las tres alternativas la verosimilitud del modelo es máxima para W_C con un valor de -781,236, obteniendo para W_A una verosimilitud de -787,123 y para W_B un valor de -783,488. Será por tanto con la matriz W_C con la que se realice finalmente la estimación del modelo.

Modelo SUR espacial: Estimación Provincial.

Los resultados correspondientes a la estimación provincial se detallan en la Tabla 1 con el conocido formato del Software SpaceStat.

Este modelo no está sustentado en un modelo MCO con problemas de autocorrelación en los residuos como suele ser habitual en econometría espacial, sino que se parte de la base de que este fenómeno está presente y se intenta modelar de forma adecuada. La hipótesis de suponer la presencia de la estructura espacial está confirmada con los resultados obtenidos[3].

Tabla 1: Estimación RFBDpc Nivel Provincial. SUR espacial. Estimación MV.

```
SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
DATA SET     SUR5               SPATIAL WEIGHTS MATRIX      WCP
DEPENDENT VARIABLE        R6       OBS 100      VARS   8      DF  92
R2          0.8917     Sq. Corr.   0.7256     R2(Buse)    0.9027
LIK        -781.236    AIC         1578.47    SC          1599.31
SIG-SQ      319056.  (     564.850 )

VARIABLE    COEFF       S.D.      z-value        Prob
    IVA6     458.382     76.9191    5.959274    0.000000
     AU6      9.9023      2.1941    4.513139    0.000006
     CA6     25.7434     5.17686    4.972780    0.000001
     TF6      6.2715      2.0826    3.011375    0.002601
    IVA8     237.927     73.9907    3.215633    0.001302
     AU8     6.98065     2.02499    3.447250    0.000566
     CA8     14.8156     4.32973    3.421835    0.000622
     TF8     12.8628      2.0861    6.165935    0.000000
  LAMBDA    0.754245   0.0573705   13.146906    0.000000

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF      VALUE        PROB
Breusch-Pagan test     8     4.754119     0.783504
Spatial B-P test       8     4.766703     0.782195
DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHTS MATRIX      WCP (row-standardized
weights)
WARNING: weights matrix contains zero rows
TEST                       DF      VALUE        PROB
Likelihood Ratio Test       1    39.263441     0.000000
LAGRANGE MULTIPLIER TEST ON SPATIAL LAG DEPENDENCE
WEIGHT     STAND  ZERO    DF      VALUE        PROB
  WCP    yes    yes     1     0.506750     0.476549
```

Elaboración propia. Resultados obtenidos con SpaceStat 1.91

[3] Se han probado otras especificaciones del modelo de la forma habitual. Esto es, partir de un modelo estimado por MCO y detectar la presencia de dependencia espacial en los residuos para finalmente presentar el modelo de regresión espacial que posee una mayor verosimilitud. Esta línea ha funcionado correctamente y son tres las variables que aparecen IVA, AU y AI (Actividades Industriales por mil habitantes) junto con una constante. Este modelo presenta la misma verosimilitud que el finalmente se propone pero a nuestro criterio posee una peor capacidad predictiva a nivel municipal.

De los resultados obtenidos para este modelo se debe destacar la presencia significativa de los cuatro factores, con una verosimilitud (LIK) de -781,23 y un coeficiente R2(Buse) superior a 0,90. No se detectan problemas de heterocedasticidad puesto que el test B-P con un valor de 4,75 no rechaza la hipótesis nula con una probabilidad superior al 78%.

En cuento al parámetro de dependencia espacial (que en los resultados de la tabla aparece como LAMBDA), toma un valor de 0,7542 sobre un máximo de 1 (W está estandarizada por filas). Este parámetro es indudablemente significativo con un z-valor muy elevado. También la significatividad de estos parámetros está confirmada por test de cociente de verosimilitudes. Por último, en este modelo no cabe esperar encontrar dependencia espacial de tipo Lag como se observa con los resultados correspondientes al test de multiplicadores de Lagrange.

5. Extrapolación de la RFBDpc a nivel Municipal. Estimación de la RFBDpc en los Municipios de la Región de Murcia.

Hasta la fecha se han realizado tres estimaciones de la RFBDpc a nivel municipal en la Región de Murcia, Esteban y Pedreño (1988); Beyaert, Buendía y Esteban (1993); Buendía y Clavo-Flores (1.999). En estos tres trabajos se ha utilizado la metodología basada en los modelos de regresión clásicos sin tener en cuenta los posibles problemas de autocorrelación espacial en los residuos.

La presencia de una fuerte estructura de dependencia espacial en esa variable está demostrada, tanto a nivel provincial como a nivel municipal en el caso de la Región de Murcia en un análisis descriptivo de la distribución espacial de la RFBDpc en los municipios de la Región de Murcia (López F. et al. 2002).

En este sentido, en los dos últimos años han surgido dos estudios en los que se contempla el problema de la autocorrelación espacial. En cada uno de estos trabajos se utiliza un modelo de regresión con distinta estructura espacial autorregresiva.

El estudio de Alañón A. (2002) es un trabajo el que se realiza la estimación de la Renta para todos los municipios del territorio Español. En este trabajo la estructura espacial que plantea es la de un modelo de regresión espacial de tipo *comparativo*, con un único corte temporal correspondiente al año 1991. En la extrapolación que realiza a los municipios después de obtener los coeficientes de la estimación a nivel provincial, no hace uso del parámetro de dependencia espacial. En este sentido, los coeficientes que obtiene gozan de las buenas propiedades de los estimadores MV, sin los problemas que plantean los estimadores obtenidos por OLS en presencia de dependencia espacial[4], pero del modelo como predicción sólo hace uso de la tendencia del proceso mediante los factores exógenos y no rectifica esta tendencia utilizando la parte *comparativa* del modelo. Pierde por tanto toda la potencialidad del modelo al no poder hacer uso del parámetro de dependencia espacial a nivel municipal, puesto que obviamente desconoce el segundo sumando de la ecuación (7).

El segundo de los modelos de estimación que utiliza técnicas de la econometría espacial es el presentado por Chasco C. (2003). En este caso la estructura autorregresiva que aparece como mas adecuada es la de tipo sustantivo, y por tanto si que permite realizar la extrapolación a nivel municipal si los problemas que plantea en anterior modelo. De esta manera, si que hace uso de la estructura espacial del proceso mediante el parámetro de dependencia espacial, asignando parte de la renta de un municipio en función de sus vecinos.

Puesto que el tipo de dependencia espacial que mayor verosimilitud introduce en el modelo propuesto en este artículo es de tipo residual, se ha hecho uso de los resultados obtenidos en el apartado anterior, estimando la RFBDpc a nivel municipal según la siguiente ecuación:

$$\begin{aligned}
\hat{\mu}_{i98} &= 237{,}9IVA8_{i,98} + 6{,}9AU8_{i,98} + 14{,}8CA8_{i,98} + 12{,}8TF8_{i,98} \\
\hat{\mu}_{i96} &= 458{,}3IVA8_{i,96} + 9{,}9AU8_{i,96} + 25{,}7CA8_{i,96} + 6{,}2TF8_{i,96} \\
\hat{y}_{i98} &= \hat{\mu}_{i98} + 0{,}75\sum_{j\in N_i} w_{ij}\left(y_{jt} - \hat{\mu}_{jt}\right) \qquad i=1,...,n
\end{aligned} \tag{29}$$

Algunas precisiones se deben de realizar cuando se aplica esta formulación con el fin de estimar la Renta para el caso específico de los municipios de la Región de Murcia:

[4] Los estadísticos obtenidos por MCO con presencia de dependencia especial son ineficientes debido a la no esfericidad de la matriz de varianzas y covarianzas. Esta no esfericidad induce una sobreestimación de los coeficientes β y del coeficiente de determinación R^2.

1. La estimación de la RFBDpc publicada en Econet para el total de la Región de Murcia correspondiente al año 1996 no coincide con la suministrada en la última estimación realizada por el BBVA. Con el fin de establecer conexión lógica entre ambas estimaciones se ha procedido a una rectificación de la estimación a nivel municipal de Econet que mantenga la distribución proporcional entre los distintos municipios.
2. La variable TF presenta tres acusados valores extremos para los municipios de Los Alcázares, San Javier y San Pedro del Pinatar en los dos años estudiados. Este problema que inicialmente era previsible al introducir esta variable en el modelo, es producto del elevado número de segundas residencias en estos municipios costeros que hacen desproporcionado el número de líneas telefónicas frente al número de habitantes. Se ha optado por realizar una corrección de estos valores siguiendo la línea dada en el último informe del CES (Consejo Económico y Social de la Región de Murcia) restando a estos municipios un 25% de sus teléfonos y redistribuyéndolos entre el resto de los municipios en proporción al número de líneas que poseen. Esto es una simplificación del ajuste realizado en el informe antes mencionado.
3. Otra observación extrema que desvirtúa la estimación la encontramos en la variable IVA para los dos años considerados en el Municipio de Santomera. Con el fin de reducir este problema se ha optado por sustituir el valor observado por la media de los valores.
4. Finalmente, para el municipio de Ulea no hay información en el anuario de la Caixa para el año 1998 ya que en este periodo tenía menos de 1000 habitantes (no así en 1996 que si tenía población superior a los mil habitantes). Se ha realizado una estimación para los valores de las variables en aquellos casos en los que no hay información en Econet.

Los resultados así obtenidos se presentan en la Tabla 2. Con el fin de contrastar los resultados obtenidos se asocia a cada estimación obtenida por este método el nivel de Renta obtenido para este año por el Anuario Comercial.

Tabla 2: Estimación de la RFBDpc de los Municipios de la CARM

Municipio	RFBDpc	Nivel SUR*	Nivel CX*
Abanilla	7:745 €	4	4
Abarán	6:579 €	2	4
Águilas	6:892 €	3	4
Albudeite	6:002 €	1	2
Alcantarilla	8:111 €	4	4
Alcázares (Los)	10:317 €	6	5
Alguazas	7:186 €	3	4
Alhama de Murcia	8:152 €	4	3
Archena	7:207 €	3	4
Beniel	7:708 €	4	4
Blanca	6:103 €	2	3
Bullas	7:832 €	4	3
Calasparra	7:269 €	3	3
Campos del Río	6:361 €	2	2
Caravaca de la Cruz	7:342 €	3	3
Cartagena	8:680 €	4	5
Cehegín	6:593 €	2	3
Ceutí	6:905 €	3	4
Cieza	6:745 €	2	4
Fortuna	7:622 €	4	3
Fuente-Alamo	10:265 €	6	4
Jumilla	7:623 €	4	3
Librilla	7:508 €	3	3
Lorca	8:393 €	4	4
Lorquí	8:146 €	4	4
Mazarrón	9:090 €	5	4
Molina de Segura	9:205 €	5	4
Moratalla	6:253 €	2	2
Mula	6:454 €	2	3
Murcia	8:848 €	5	5
Pliego	7:614 €	4	3
Puerto Lumbreras	8:190 €	4	4
Ricote	5:528 €	1	2
San Javier	10:736 €	6	5
San Pedro del Pinatar	9:249 €	5	5
Santomera	8:695 €	4	4
Torre-Pacheco	9:554 €	5	4
Torres de Cotillas	8:187 €	4	4
Totana	7:676 €	4	4
Ulea	6:124 €	2	
Unión (La)	6:673 €	2	4
Villanueva del Río S.	5:719 €	1	2
Yecla	9:120 €	5	4

*Nivel CX: Nivel de Renta asignado por el Anuario Comercial de España.
Nivel SUR: Nivel de Renta asignado por el modelo SUR atendiendo a loa inérvalos del Anuario Comercial de España.

6. Conclusiones:

La dependencia espacial está presente en esta variable y no tiene porque estar asociada a una inadecuada especificación del modelo. Aún en el hipotético caso de que el modelo no presentase problemas de dependencia espacial a nivel provincial con toda seguridad que este debe estar presente a nivel municipal. Esto es presumible puesto que la dependencia espacial aumenta al disminuir la unidad espacial de observación. Además debe existir una falta de concordancia natural entre la extensión del fenómeno económico y la unidad de observación.

Estas estructuras de dependencia espacial no están presenten únicamente en la dimensión espacial sino que es de esperar que las interrelaciones se presenten en la doble dimensión espacio temporal. Al plantear un modelo que capte esta doble dimensión deben incluirse no sólo retardos espaciales sino también un adecuado retardo espacio-temporal.

El modelo propuesto en este artículo recoge esta doble dimensionalidad. Por un lado es capaz de dar solución a este problema de una forma simple, y además tiene la ventaja añadida de que sus coeficientes pueden ser estimados haciendo uso de un software disponible y no es necesario realizar una programación específica lo que supondría un importante coste.

En cuanto a la estimación de la RFBDpc de los municipios de la CARM haremos aquí sólo unos breves comentarios. En primer lugar los resultados obtenidos son consistentes con la estructura económica de los municipios de la Región existiendo un fuerte paralelismo entre la estimación aquí realizada y la obtenida por el Anuario Comercial. En este sentido destacamos en primer lugar que nuestra estimación introduce una mayor variabilidad que la estimación dada por el Anuario Comercial de la Caixa con niveles de renta máximo de 6 y mínimos de 1 frente a una franja de 2 a 5 que se obtiene en el Anuario.

Finalmente a nuestro criterio consideramos que son dos las líneas en las que se debe seguir investigando.

Por una parte es esencial prestar más atención a la selección de variables exógenas que sea capaz de explicar los niveles de renta al doble nivel geográfico. En este sentido y tal y como se realiza en Chasco C. (2003) la técnica del Análisis de Componentes Principales sea adecuada para resolver este problema.

Por otra parte se debe realizar un esfuerzo para programar mediante Matlab el proceso de inferencia del modelo SUR general.

7. Bibliografía.

Alañón A. (2002): Estimación de la Renta de los Municipios Españoles en 1991 Mediante Técnicas de Econometría Espacial. *XXVIII Reunión de Estudios Regionales. Murcia Noviembre 2002*.

Anselin L. (1.988): *Spatial Econometrics: Methods and Models.* Edit. Kluwer Academic Publishers.

Banco Bilbao Vizcaya (BBVA): *Renta Nacional de España y su distribución Provincial Año 1.995 y Avances 1996-1999*. Bilbao.

Beyaert, A, Buendía J.D. Esteban, M. (1.993): *Distribución intrarregional de la renta en Estructura económica de la Región de Murcia*. Ed. Civitas, Madrid. Cap,24, pp 715-738.

Buendía y Calvo-Flores A. (1999): *Informe sobre la distribución intermunicipal de la renta. Disparidades intermunicipales de la Región de Murcia durante el periodo 1986-1996*. Consejo Económico y Social de la Región de Murcia.

Caja de Ahorro y pensiones de Barcelona (La Caixa) (varios años): *Anuario Comercial de España.*

Chasco C. (2003): *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Consejería de Economía e Innovación Tecnológica, Comunidad de Madrid. Madrid, abril de 2003.

Cliff, A. D. y Ord, J. K. (1.981): Spatial *processes: Models and Applications.* Pion Limited, London

Esteban J. y Pedreño, A. (1988): *Renta municipal de la Comunidad Autónoma de la Región de Murcia* Cuadernos de Investigación. Caja de Ahorros de Alicante y Murcia. Alicante.

López F.A., Palacios M.A. y García J.A. (2002): Distribución Espacial de la Renta Familiar Disponible en los Municipios de la Región de Murcia. Un Análisis Descriptivo. *XVI Reunión Asepelt Madrid Junio 2002*.

Moreno, R. y Vayá, E (2000): *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*. Ediciones Universidad de Barcelona.

Zellner A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regressions and Test of Aggregation Bias *Journal of the American Statistical Association*, 57, 348-68