

**ASOCIACION ESPAÑOLA DE CIENCIA REGIONAL.
XXIX REUNION DE ESTUDIOS REGIONALES. SANTANDER
27-28 de noviembre de 2003**

El test de Vuong y otros indicadores de dinámica espacial (I).

Jesús Mur (*)

Ana Angulo (**)

Departamento de Análisis Económico
Universidad de Zaragoza
Gran Vía, 2-4. (50005). Zaragoza
Tel: 976-761815

(*) jmur@posta.unizar.es

(**) aangulo@posta.unizar.es

RESUMEN:

En este trabajo se analiza, en términos generales, el problema de la especificación de modelos econométricos de corte transversal sobre la dimensión espacio.

La selección de la estructura dinámica del modelo es un aspecto fundamental que determina la utilidad final de la especificación. Sin embargo, en la práctica habitual esta cuestión se resuelve utilizando un enfoque poco elaborado, consistente en proponer inicialmente una ecuación lo más sencilla posible, la cual se va corrigiendo en función de la evidencia acumulada. Esta estrategia no parece la más adecuada para alcanzar una especificación óptima.

En el papel se examinan algunas estrategias alternativas a la descrita. La primera es de tipo clásico y se basa en la sistematización de una serie de resultados obtenidos en un contexto de estimación máximo-verosímil. Este enfoque se completa con la propuesta de Vuong (1989), donde se desarrolla una estrategia de selección de modelos que puede ser adecuada para el análisis de la dinámica espacial.

1.- Introducción

En un papel reciente, Florax, Folmer y Rey (1998) discuten de forma bastante ácida el método de trabajo consolidado en el área que se ha dado en denominar Econometría espacial. Según ellos, en la actualidad predomina enfoque de tipo *clásico*, estructurado en tres etapas:

- (i)- En primer lugar, se estima un modelo simple, habitualmente estático y especificado bajo condiciones ideales.
- (ii)- A continuación, se aplican una serie de contrastes de autocorrelación, residual y sustantiva, sobre la ecuación estimada.
- (iii)- Caso de rechazar alguna hipótesis, se aplicarán los ajustes correspondientes (reformulando la ecuación, filtrando las variables, incorporando elementos de dinámica espacial, etc.).

Es decir, inicialmente se tiende a especificar un modelo lo más simple posible, el cual se va corrigiendo a medida que la evidencia disponible lo demanda. En términos econométricos (Mizon, 1984), este planteamiento podría denominarse como '*de lo particular a lo general*'. A pesar de su popularidad, y de los nuevos contrastes de subespecificación propuestos en los últimos años, más potentes y versátiles (ver Florax y de Graaff, 2003, para una panorámica), no parece la mejor estrategia para alcanzar una especificación óptima.

Con este trabajo queremos retomar alguna de las cuestiones planteadas en la referencia de Florax, Folmer y Rey (1998), singularmente las asociadas a la especificación de la estructura dinámica de la ecuación¹. En la sección segunda sistematizamos una serie de resultados, en gran parte ya conocidos, relativos al método máximo-verosímil con los cuales se puede articular una estrategia coherente de modelización. La tercera sección la dedicamos a presentar la propuesta de Vuong (1989), como marco de análisis potencialmente útil y complementario con el enfoque anterior. En la cuarta sección discutimos algunas cuestiones relativas a la aplicación de este método en un contexto espacial. La última sección se reserva para las conclusiones y comentario finales.

¹ Fuera de la discusión quedan, por cuestiones de espacio y de agilidad en la exposición, otro tipo de efectos (heterogeneidad por ejemplo), a los cuales se debería prestar la misma atención.

2.- Modelo amplio y estructuras de anidación.

En la literatura se encuentran plenamente consolidados los conceptos de *autocorrelación residual* y *autocorrelación sustantiva*, en referencia a los mecanismos de dependencia espacial (potencialmente) existentes en una especificación de corte transversal. Sin embargo, en la práctica parece que se emplean de forma excluyente: no es habitual analizar incorrelación en el término de perturbación de una ecuación en la que se ha incluido una estructura dinámica de la variable endógena, como tampoco suele ser normal cuestionar la dinámica de la ecuación cuando se ha decidido introducir dependencia espacial en el término de error. Esta práctica limita innecesariamente el horizonte de la aplicación por cuanto la estructura de interrelaciones puede ser más general. En el Cuadro 1 presentamos el marco de referencia al cual limitaremos nuestra discusión.

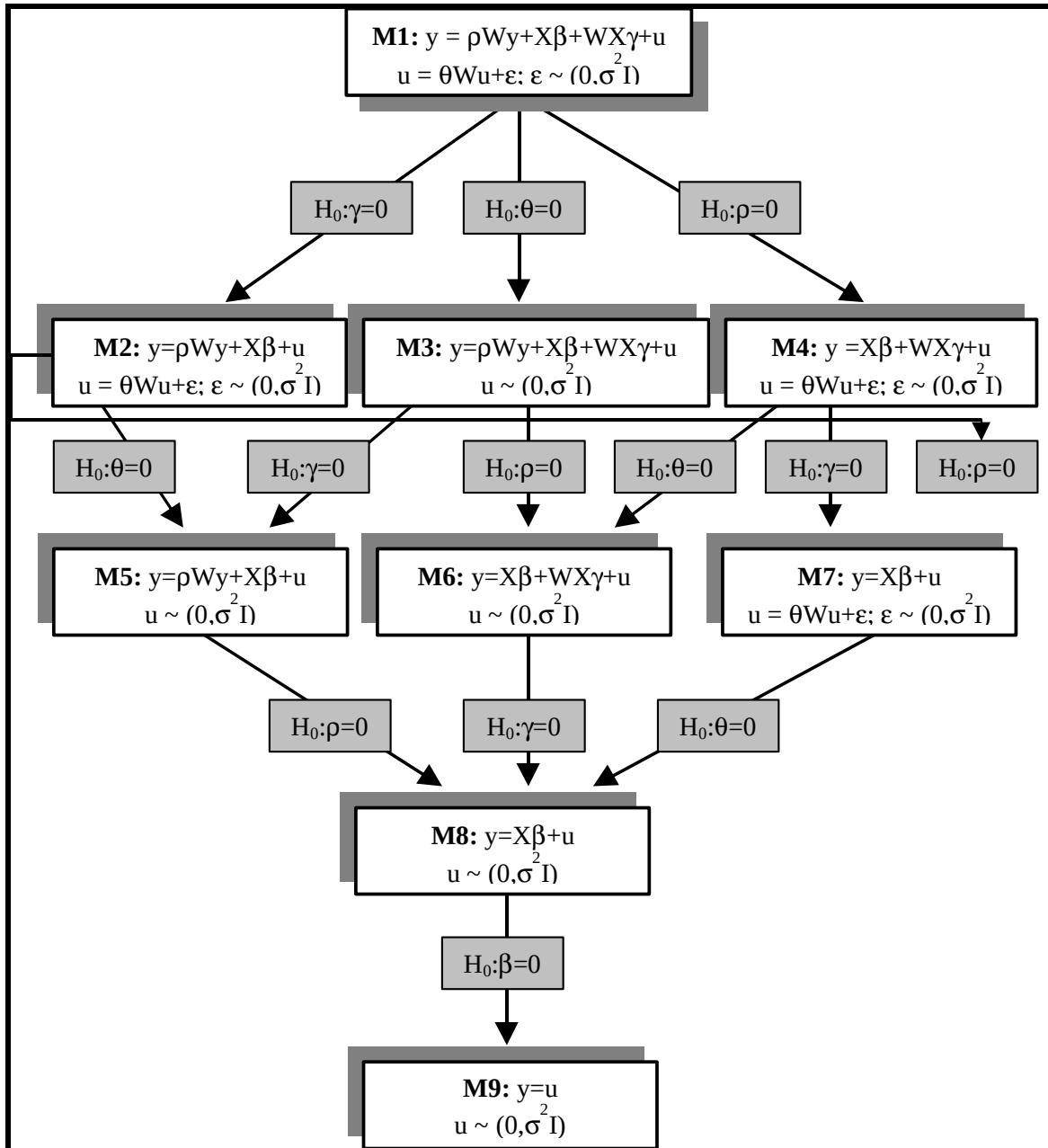
En esa secuencia de anidaciones, y es el vector ($R \times 1$) de observaciones de la variable endógena, X es la matriz ($R \times k$) de observaciones de las variables exógenas en las R unidades de observación, u es el vector ($R \times 1$) de términos de error y ε es un vector de perturbaciones del mismo orden. La matriz de pesos que se emplea es la misma en todos los casos, W , en principio de tipo binario y con ceros en su diagonal principal. Por último, ρ y θ son parámetros de dependencia, sustantiva y residual respectivamente, y β y γ son vectores (ambos $k \times 1$) de parámetros asociados a las variables exógenas (el último vector lleva implícito un mecanismo de dependencia transversal).

Con respecto a la estructura de anidación del Cuadro 1, útil sobre todo para estructurar la exposición, es conveniente hacer algunas precisiones:

- (a)- No es necesario emplear matrices de contactos diferentes. Salvo circunstancias excepcionales, lo más natural es mantener la misma hipótesis de interacción para todas las variables del modelo. La única restricción (como indica Anselin, 1988a) es que, si se combina una estructura autoregresiva en la variable endógena con una estructura del mismo tipo en la perturbación, deben intervenir también otro tipo de elementos sistemáticos en la ecuación (basta con un factor de escala) para asegurar la identificación de los parámetros.
- (b)- La inclusión de una estructura de retardos en las variables exógenas viene motivada por la existencia de *efectos de desbordamiento*, pero no introduce peculiaridad alguna en términos econométricos. Sus consecuencias se limitan a la aparición de nuevos regresores, probablemente muy relacionados con los

originales. Para simplificar la notación, incluiremos ambos tipos de términos en la matriz $Z=[X, WX]$, de orden $(R \times 2k)$, la cual se asocia al vector de parámetros $\delta=[\beta, \gamma]'$, de orden $(2k \times 1)$.

Cuadro 1: Estructura de anidación para un análisis de dinámica espacial



(c)- En el Cuadro 1 se ha especificado una estructura autoregresiva en el término de error de la ecuación, aunque se podrían haber introducido otro tipo de procesos, una media móvil ($u = \varepsilon + \theta W\varepsilon$) por ejemplo. Las estructuras autoregresivas, SAR, y de media móvil, SMA, son las utilizadas más habitualmente. Sin embargo, no son intercambiables y su especificación requiere un análisis adicional.

(d)- En general, no se dispondrá de información sobre la distribución de probabilidad de la perturbación. El procedimiento habitual consiste en asumir inicialmente normalidad, lo cual simplifica el ejercicio de inferencia. Posteriormente esa hipótesis es sometida a contraste.

El modelo M8 es el punto de partida de lo que se ha denominado método clásico. Esta especificación se corrige hacia M5 ó M7 si los contrastes de subespecificación habituales lo aconsejan, aunque raramente se alcanzará M1 ya que no se suele contemplar como alternativa explícita de modelización.

Como se ha dicho, nuestra propuesta invierte la secuencia de análisis de forma que sea precisamente M1, como modelo más amplio, el punto de partida. El principal argumento que puede esgrimirse contra ese modelo es la complejidad de las relaciones, propias de unas estructuras fuertemente no lineales. Esa complejidad existe, y para superarla será necesario aplicar potencia de cálculo a unos algoritmos de estimación eficientes. En cualquier caso, creemos que los beneficios derivados de invertir la secuencia de análisis compensan los costes incurridos en la mayor necesidad de cálculo. Los resultados que presentamos se encuentran en esta misma línea.

Asumiendo normalidad en el término de perturbación, la especificación M1 se concreta en:

$$\left. \begin{aligned} y &= \rho Wy + Z\delta + u \\ u &= \theta Wu + \varepsilon \\ \varepsilon &\sim N[0, \sigma^2 I] \end{aligned} \right\} \quad (2.1)$$

La función de log-verosimilitud tiene la estructura habitual:

$$\begin{aligned} l(y/\varphi) &= -\frac{R}{2} \ln 2\pi - \frac{R}{2} \ln \sigma^2 - \frac{[By - Z\delta]' D' D [By - Z\delta]}{2\sigma^2} + \ln|B| + \ln|D| \\ \ln|B| &= |I - \rho W| = \sum_{r=1}^R (1 - \rho \lambda_r) \\ \ln|D| &= |I - \theta W| = \sum_{r=1}^R (1 - \theta \lambda_r) \end{aligned} \quad (2.2)$$

donde λ_r es la raíz característica r-ésima de la matriz W y φ es el vector de parámetros de la especificación $\varphi = [\delta, \rho, \theta, \sigma^2]'$. El sistema de condiciones necesarias es no lineal, haciendo imposible la obtención de una solución analítica:

$$\frac{\partial l}{\partial \delta} = \frac{1}{\sigma^2} [Z' D' D (By - Z\delta)] = \frac{1}{\sigma^2} Z' D' \varepsilon \quad (2.3a)$$

$$\frac{\partial l}{\partial \rho} = \frac{1}{\sigma^2} [y' W D' D (By - Z\delta)] - \sum_{r=1}^R \frac{\lambda_r}{1 - \rho \lambda_r} = \frac{1}{\sigma^2} [y' W D' \varepsilon] - \sum_{r=1}^R \frac{\lambda_r}{1 - \rho \lambda_r} \quad (2.3b)$$

$$\frac{\partial l}{\partial \theta} = \frac{1}{\sigma^2} [(By - Z\delta)' W D (By - Z\delta)] - \sum_{r=1}^R \frac{\lambda_r}{1 - \theta \lambda_r} = \frac{1}{\sigma^2} [u' W \varepsilon] - \sum_{r=1}^R \frac{\lambda_r}{1 - \theta \lambda_r} \quad (2.3c)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{R}{2\sigma^2} + \frac{[(By - Z\delta)' D' D (By - Z\delta)]}{2\sigma^4} = -\frac{R}{2\sigma^2} + \frac{\varepsilon' \varepsilon}{2\sigma^4} \quad (2.3.d)$$

La estimación MV del vector de parámetros de posición, δ , y del parámetro de dispersión, σ^2 , puede obtenerse de forma condicionada a la de los parámetros de dependencia transversal:

$$\tilde{\delta} = [Z' \tilde{D}' \tilde{D} Z]^{-1} Z' \tilde{D}' \tilde{D} \tilde{B} y \quad (2.4a)$$

$$\tilde{\sigma}^2 = \frac{[\tilde{B} y - Z \tilde{\delta}]' \tilde{D}' \tilde{D} [\tilde{B} y - Z \tilde{\delta}]}{R} = \frac{\tilde{\varepsilon}' \tilde{\varepsilon}}{R} \quad (2.4b)$$

con $\tilde{D} = I - \tilde{\theta} W$ y $\tilde{B} = I - \tilde{\rho} W$. Es decir, el ejercicio de optimización, aunque costoso, puede resolverse actuando sobre la función de log-verosimilitud concentrada, tal como indica Ord (1975).

Por otra parte, la matriz de información correspondiente al modelo general de (2.1) es:

$$I(\varphi) = -E \left[\frac{\partial^2 l}{\partial \varphi \partial \varphi'} \right] = \begin{bmatrix} \frac{Z' D' D Z}{\sigma^2} & \frac{Z' D' D W B^{-1} Z \delta}{\sigma^2} & 0 & 0 \\ \frac{\delta' Z' B^{-1} W D' D Z}{\sigma^2} & t(\rho) & 2tr W^2 B^{-1} D^{-1} & \frac{tr W B^{-1}}{\sigma^2} \\ 0 & 2tr W^2 B^{-1} D^{-1} & t(\theta) & \frac{tr W D^{-1}}{\sigma^2} \\ 0 & \frac{tr W B^{-1}}{\sigma^2} & \frac{tr W D^{-1}}{\sigma^2} & \frac{R}{2\sigma^4} \end{bmatrix} \quad (2.5)$$

siendo
$$t(\rho) = \frac{\delta' Z' B^{-1} W D' D W B^{-1} Z \delta}{\sigma^2} + tr D^{-2} B^{-1} W D' D W B^{-1} + \sum_{r=1}^R \left(\frac{\lambda_r}{1 - \rho \lambda_r} \right)^2,$$

$$t(\theta) = tr D^{-1} W^2 D^{-1} + \sum_{r=1}^R \left(\frac{\lambda_r}{1 - \theta \lambda_r} \right)^2$$
 y $W^2 = W W$.

La inversa de la matriz de información desempeña un papel central en cualquier ejercicio de inferencia MV, por lo que se hace imprescindible su obtención. La notación que vamos a emplear es la siguiente:

$$I(\varphi)^{-1} = \begin{bmatrix} V(\delta) & Cov(\delta, \alpha) \\ Cov(\alpha, \delta) & V(\alpha) \end{bmatrix} \quad (2.6)$$

siendo $\alpha = [\rho, \theta, \sigma^2]'$. Los resultados que se obtienen, de forma sintética, son:

$$V(\delta) = \sigma^2 [Z' D' D Z]^{-1} [I_{2k} + V(\rho) Z' D' D W B^{-1} Z \delta \delta' Z' B^{-1} W D' D Z (Z' D' D Z)^{-1}] \quad (2.7)$$

En esta expresión I_{2k} es la matriz identidad de orden $(2k \times 2k)$.

$$Cov(\delta, \alpha) = \begin{bmatrix} -V(\rho)(Z'D'DZ)^{-1}Z'D'DW B^{-1}Z\delta \\ -Cov(\rho, \theta)(Z'D'DZ)^{-1}Z'D'DW B^{-1}Z\delta \\ -Cov(\rho, \sigma^2)(Z'D'DZ)^{-1}Z'D'DW B^{-1}Z\delta \end{bmatrix} \quad (2.8)$$

$$V(\alpha) = \begin{bmatrix} V(\rho) & Cov(\rho, \theta) & Cov(\rho, \sigma^2) \\ Cov(\theta, \rho) & V(\theta) & Cov(\theta, \sigma^2) \\ Cov(\sigma^2, \rho) & Cov(\sigma^2, \theta) & V(\sigma^2) \end{bmatrix} \quad (2.9)$$

con:

$$V(\rho) = \left[\frac{e}{\sigma^2} + 2R \text{var}(m_\rho)(1 - \text{corr}^2(m_\rho, m_\theta)) \right]^{-1} \quad (2.9a)$$

$$V(\theta) = \frac{1}{\text{var}(m_\theta)} \left[\frac{1}{2R} + V(\rho) \text{var}(m_\rho) \text{corr}^2(m_\rho, m_\theta) \right] \quad (2.9b)$$

$$V(\sigma^2) = \frac{2\sigma^4}{R} \left[1 + \mu^2(m_\theta) \right] + 4\sigma^4 V(\rho) \text{var}(m_\rho) \left[\mu(m_\rho) - \mu(m_\theta) \text{corr}^2(m_\rho, m_\theta) \right]^2 \quad (2.9c)$$

$$Cov(\rho, \theta) = -V(\rho) \left(\frac{\text{var}(m_\rho)}{\text{var}(m_\theta)} \right)^2 \text{corr}(m_\rho, m_\theta) \quad (2.9d)$$

$$Cov(\rho, \sigma^2) = -2\sigma^2 V(\rho) \text{var}(m_\rho) \left[\mu(m_\rho) - \mu(m_\theta) \text{corr}(m_\rho, m_\theta) \right] \quad (2.9e)$$

$$Cov(\theta, \sigma^2) = \frac{\sigma^2}{\sqrt{\text{var}(m_\theta)}} \left\{ 2V(\rho) \text{var}(m_\rho) \text{corr}(m_\rho, m_\theta) \left[\mu(m_\rho) - \mu(m_\theta) \text{corr}(m_\rho, m_\theta) \right] - \frac{\mu(m_\theta)}{R} \right\} \quad (2.9f)$$

siendo $e = \delta' Z' B^{-1} W D' M_{ZD} D W B^{-1} Z \delta$ y $M_{ZD} = I - D Z (Z' D' D Z)^{-1} Z' D'$. En general, las distintas varianzas y covarianzas de (2.9) se construyen a partir de unas variables artificiales, $m_r(\rho)$ y $m_r(\theta)$, las cuales dependen, a su vez, de las raíces características de W , $\{\lambda_r, r=1, 2, \dots, R\}$, y de los parámetros de dependencia espacial, ρ y θ . Las definiciones precisas son:

$$m_r(\rho) = \frac{\lambda_r}{1 - \rho \lambda_r} \quad \bar{m}(\rho) = \sum_{r=1}^R \frac{m_r(\rho)}{R} \quad \text{var}(m_\rho) = \frac{\sum_{r=1}^R [m_r(\rho) - \bar{m}(\rho)]^2}{R} \quad (2.10a)$$

$$m_r(\theta) = \frac{\lambda_r}{1 - \theta \lambda_r} \quad \bar{m}(\theta) = \sum_{r=1}^R \frac{m_r(\theta)}{R} \quad \text{var}(m_\theta) = \frac{\sum_{r=1}^R [m_r(\theta) - \bar{m}(\theta)]^2}{R} \quad (2.10b)$$

$$\mu(m_\rho) = \frac{\bar{m}(\rho)}{\sqrt{\text{var}(m_\rho)}} \quad \mu(m_\theta) = \frac{\bar{m}(\theta)}{\sqrt{\text{var}(m_\theta)}} \quad (2.10c)$$

$$corr(m_\rho, m_\theta) = \frac{\sum_{r=1}^R [m_r(\rho) - \bar{m}(\rho)] [m_r(\theta) - \bar{m}(\theta)]}{\sqrt{\sum_{r=1}^R [m_r(\rho) - \bar{m}(\rho)]^2 \sum_{r=1}^R [m_r(\theta) - \bar{m}(\theta)]^2}} \quad (2.10d)$$

Una vez que se disponga de todos estos elementos, es relativamente sencillo resolver la estructura de anidación del Cuadro 1². Así, por ejemplo, el modelo M4 se encuentra anidado en M1 al imponer la restricción de que el parámetro de autocorrelación espacial sustantiva del modelo amplio es cero:

$$\left. \begin{array}{l} H_0: \rho = 0 \\ H_A: \rho \neq 0 \end{array} \right\} \quad (2.11)$$

permitiendo la existencia de una estructura de dependencia transversal en la perturbación. El contraste de (2.11) puede resolverse utilizando el multiplicador de Lagrange, para lo cual necesitamos la estimación máximo-verosímil restringida de M4:

$$\left. \begin{array}{l} \tilde{\delta}_R = [Z' \tilde{D}' \tilde{D} Z]^{-1} Z' \tilde{D}' \tilde{D} y \\ \tilde{\sigma}_R^2 = \frac{[y - Z \tilde{\delta}_R]' \tilde{D}' \tilde{D} [y - Z \tilde{\delta}_R]}{R} = \frac{\tilde{u}_R' \tilde{D}' \tilde{D} \tilde{u}_R}{R} = \frac{\tilde{\varepsilon}_R' \tilde{\varepsilon}_R}{R} \\ \tilde{u}_R = y - Z \tilde{\delta}_R \\ \tilde{\varepsilon}_R = \tilde{D} \tilde{u}_R \\ \tilde{D} = I - \tilde{\theta}_R W \end{array} \right\} \quad (2.12)$$

Utilizando estos resultados, el estadístico de contraste es (ver también Anselin y Bera, 1998):

$$LM(\rho | \tilde{\theta}_R) = V(\tilde{\rho}_R) \left[\frac{y' W \tilde{D} \tilde{\varepsilon}_R}{\tilde{\sigma}_R^2} \right]^2 \underset{as}{\sim} \chi^2(1) \quad (2.13)$$

En esta expresión interviene la varianza asintótica del estimador MVR de ρ , $V(\tilde{\rho}_R)$, evaluada en el punto de estimación MVR de (2.12). La expresión de referencia es la de (2.9a) en la que $\tilde{\theta}_R$ se sustituirá por su correspondiente estimación MVR y $\tilde{\rho}_R$ se hará igual a cero.

En el contraste de Wald utilizamos los resultados de la estimación máximo-verosímil no restringida correspondiente al modelo amplio de M1:

² La discusión que se reproduce a continuación coincide, parcialmente, con la de Anselin y Bera (1998). El aspecto diferencial de nuestro trabajo es que introducimos, de forma explícita, el concepto de modelo amplio lo que nos permite obtener resultados más completos en este aspecto.

$$W(\tilde{\rho}) = \frac{\tilde{\rho}^2}{V(\tilde{\rho})} \underset{as}{\sim} \chi^2(1) \quad (2.14)$$

A diferencia de lo que ocurre con el LM de (2.13), ahora la varianza asintótica del estimador de ρ , $V(\tilde{\rho})$, se evalúa en el punto correspondiente a la estimación MV no restringida del sistema de ecuaciones de (2.3).

De la misma forma, también podemos contrastar independencia en el término de error, asumiendo autocorrelación de tipo sustantiva en la ecuación (nuevamente, ver Anselin y Bera, 1998). El modelo amplio vuelve a ser M1 mientras que el restringido es M3. La hipótesis del contraste se concreta en:

$$\left. \begin{array}{l} H_0: \theta = 0 \\ H_A: \theta \neq 0 \end{array} \right\} \quad (2.15)$$

Una vez más, el Multiplicador de Lagrange se perfila como la opción preferida, dado que únicamente requiere los resultados de la estimación MVR. La expresión del estadístico de prueba en este caso es:

$$LM(\theta|\tilde{\rho}_R) = V(\tilde{\theta}_R) \left[\frac{\tilde{u}_R' W \tilde{u}_R}{\tilde{\sigma}_R^2} \right] \underset{as}{\sim} \chi^2(1) \quad (2.16)$$

en la que introducimos los resultados de la estimación MVR de M3:

$$\left. \begin{array}{l} \tilde{\delta}_R = [Z'Z]^{-1} Z' \tilde{B}y \\ \tilde{\sigma}_R^2 = \frac{[\tilde{B}y - Z \tilde{\delta}_R]' [\tilde{B}y - Z \tilde{\delta}_R]}{R} = \frac{\tilde{u}_R' \tilde{u}_R}{R} \\ \tilde{u}_R = \tilde{B}y - Z \tilde{\delta}_R \\ \tilde{B} = I - \tilde{\rho}_R W \end{array} \right\} \quad (2.17)$$

La estimación de la varianza asintótica del estimador MVR de θ , $V(\tilde{\theta}_R)$, se obtiene al introducir los resultados de (2.17), junto con $\tilde{\theta}_R = 0$, en la varianza de (2.9b).

La expresión del estadístico de Wald correspondiente a este caso es simple:

$$W(\tilde{\theta}) = \frac{\tilde{\theta}^2}{V(\tilde{\theta})} \underset{as}{\sim} \chi^2(1) \quad (2.18)$$

La varianza asintótica del estimador MV de θ es, otra vez, la que aparece en (2.9b), evaluada en el punto de estimadores máximo-verosímiles.

Por último, lo que podría entenderse como contraste de dinámica espacial global relaciona los modelos M1 y M8 mediante la hipótesis conjunta:

$$\left. \begin{array}{l} H_0: \theta = \rho = 0 \\ H_A: \theta \neq 0 \vee \rho \neq 0 \end{array} \right\} \quad (2.19)$$

Replicando, una vez más, el planteamiento basado en los Multiplicadores de Lagrange, el estadístico resultante es el conocido LM-SARMA (Anselin, 1988b). El atractivo de este contraste es que se basa, únicamente, en la estimación MCO del modelo restringido de M8:

$$LM - SARMA = V(\tilde{\rho}_R) \left[\frac{y'Wy}{\tilde{\sigma}_R^2} - \frac{\tilde{u}_R'W\tilde{u}_R}{\tilde{\sigma}_R^2} \right]^2 - \frac{1}{2S_0} \left[\frac{\tilde{u}_R'W\tilde{u}_R}{\tilde{\sigma}_R^2} \right]^2 \underset{as}{\sim} \chi^2(2) \quad (2.20)$$

donde:

$$\left. \begin{array}{l} \tilde{\delta}_R = [Z'Z]^{-1}Z'y \\ \tilde{\sigma}_R^2 = \frac{[y - Z\tilde{\delta}_R]'[y - Z\tilde{\delta}_R]}{R} = \frac{\tilde{u}_R'\tilde{u}_R}{R} \\ \tilde{u}_R = y - Z\tilde{\delta}_R \end{array} \right\} \quad (2.21)$$

La varianza asintótica del estimador MVR de ρ , $V(\tilde{\rho}_R)$, debe evaluarse en el punto de estimación máximo-verosímil restringida, $\tilde{\rho}_R = \tilde{\theta}_R = 0$, resultando:

$$\tilde{V}_R(\rho) = \frac{\tilde{\sigma}_R^2}{\tilde{\delta}_R'Z'WMWZ\tilde{\delta}_R} \quad (2.22)$$

con $M = I - Z(Z'Z)^{-1}Z'$. El factor de escala que afecta al segundo término de la parte derecha de (2.20), $1/(2S_0)$, es la estimación de la varianza asintótica del estimador MVR de θ . En definitiva, en ambos casos se está actuando sobre las expresiones (2.9a) y (2.9b), en las que se introduce la hipótesis nula del contraste: $\tilde{\rho}_R = \tilde{\theta}_R = 0$.

Por otro lado, el test de Wald para la hipótesis de (2.19) es de tipo bidireccional:

$$W(\tilde{\rho}; \tilde{\theta}) = \frac{1}{1 - r_{(\tilde{\rho}; \tilde{\theta})}^2} \left[W(\tilde{\rho}) + W(\tilde{\theta}) - 2r_{(\tilde{\rho}; \tilde{\theta})}^2 \sqrt{W(\tilde{\rho})W(\tilde{\theta})} \right] \underset{as}{\sim} \chi^2(2) \quad (2.23)$$

siendo $r_{(\tilde{\rho}; \tilde{\theta})}$ el coeficiente de correlación de los estimadores MV de ρ y de θ ³. El estadístico de (2.23) se convierte en la suma de los estadísticos unidireccionales de (2.14) y (2.18) cuando ese coeficiente de correlación sea cero. Esto es, cuando ambos

³ Es decir $r_{(\tilde{\rho}; \tilde{\theta})} = \frac{Cov(\tilde{\rho}; \tilde{\theta})}{\sqrt{V(\tilde{\rho})V(\tilde{\theta})}}$, donde se utilizarán los resultados de (2.9a), (2.9b), (2.9d).

estimadores se encuentren asintóticamente incorrelados, circunstancia que, por lo general, no se producirá salvo en circunstancias muy excepcionales. Por otra parte, puede demostrarse que cuando los parámetros de dependencia transversal, ρ y θ , toman valores similares el coeficiente de correlación asociado se aproxima a la unidad ($r_{(\tilde{\rho};\tilde{\theta})} \rightarrow 1$). En este caso, el estadístico conjunto de (2.23) tiende hacia el cuadrado de la diferencia de los estadísticos unidireccionales, expresados estos últimos en versión t-ratio, y normalizada la expresión para descontar la fuerte relación de colinealidad existente entre los dos estimadores. Esto es:

$$W_{\rho \rightarrow \theta}(\tilde{\rho};\tilde{\theta}) \rightarrow \frac{1}{1-r_{(\tilde{\rho};\tilde{\theta})}^2} \left(t(\tilde{\rho}) - t(\tilde{\theta}) \right)^2 \quad (2.24)$$

con $t(\tilde{\rho}) = \sqrt{W(\tilde{\rho})}$ y $t(\tilde{\theta}) = \sqrt{W(\tilde{\theta})}$.

3.- Medidas de información y otras técnicas de selección de modelos.

Como se ha dicho en la sección de introducción, no se han publicado muchos trabajos en los que se contemple el problema de la selección de modelos en un contexto puramente espacial. El documento ya comentado de Florax, Folmer y Rey (1998) es una excepción, además del Capítulo 14 del manual de Anselin (1988a). La relación se completa con una colección de trabajos de orden menor y ámbito más estricto. (ver Chasco, 2003, para una revisión reciente).

La naturaleza de los datos (ya sea espacial o temporal) no es un aspecto con la suficiente relevancia como para alterar el modo habitual de enfocar el problema. Sin embargo, sí que existen peculiaridades que es preciso reconocer. Por ejemplo, la finalidad de un modelo de corte transversal en el espacio raramente será el de la predicción, tal como esta última se entiende en un contexto temporal. Además, la naturaleza de las relaciones que predominan sobre la dimensión espacio es multidireccional y vienen acompañadas por múltiples efectos externos (de desbordamiento, de contagio, etc.). Estas circunstancias hacen de muy difícil asimilación el planteamiento básico enunciado por Judge et al. (1980) en torno al método econométrico: ‘... many of the choice procedures discussed (...) recommend splitting the data into two sets; one set for model choice and estimation and the other set for assessing the empirical construct’.

En cualquier caso, queremos insistir en que la reflexión sobre el problema de la selección de modelos, en aplicaciones de econometría espacial, ha sido escasa hasta el momento. Los trabajos que se publican en el área muestran una acusada preferencia por

el uso de estadísticos simples, que condensan de forma muy efectiva la evidencia disponible. Entre estos estadísticos destaca sobremanera el AIC de Akaike (ver Aznar, 1989, para una revisión de la literatura sobre selección de modelos), el cual se halla implementado en los paquetes informáticos más populares, como el Spatial Econometrics⁴ o el SpaceStat⁵. El estadístico de Akaike se basa en la medida de información de Kullback-Leibler y propone una aproximación preferencialista al problema. El enfoque confirmativo⁶ parece tener menos adeptos, aunque el trabajo de Anselin y Griffith (1988), donde se examina el comportamiento del Cp de Mallows bajo distintas estructuras de heterogeneidad espacial, es una remarcable excepción.

En definitiva, y a pesar del tiempo transcurrido, parece que continúa teniendo vigencia la afirmación de Anselin (1988a, p. 297) cuando dice que: *‘these techniques can be related to important bodies of theory and methodology in regional science and geography. (...) and have been suggested at various points in time as the new and improved research directions to follow’*. El presente trabajo pretende seguir esta recomendación, aunque los objetivos son más modestos. En concreto, nuestra motivación es explorar el potencial de la metodología de selección de modelos propuesta por Vuong (1989).

El enfoque de este autor se estructura en torno a la medida de información de Kullback-Leibler, y su aspecto diferencial es que utiliza el criterio de información para determinar qué modelo, de entre los dos considerados⁷, se encuentra más próximo al verdadero proceso generador de datos (PGD), mientras que en el AIC se usa para medir la proximidad del modelo bajo consideración al PGD.

La estrategia que propone Vuong entraña cierta complejidad. Por un lado se dispone de un vector aleatorio X_t de orden $(m \times 1)$, particionado en $X_t = [Y_t, Z_t]$ de órdenes

⁴ Disponible en: <http://www.spatial-econometrics.com>

⁵ Disponible en: <http://www.terraseer.com/spacestat.html>

⁶ Tal como lo plantea Aznar (1989, p. 126): *‘.. in the first case (confirmation strategy), there was information about one of the models being the true model and that only if the sample evidence was quite clearly against it, could we reject it. For the hypothesis testing strategy, all the models were equally acceptable from the beginning’*.

⁷ Esta es otra peculiaridad de la propuesta, dado que solo se consideran dos alternativas rivales.

n y k respectivamente ($n+k=m$). La función de distribución, cierta, del vector X_t es H_X^0 , si bien nos interesa específicamente la condicional de Y_t (endógenas) respecto a Z_t (explicativas), $H_{Y|Z}^0$. Los vectores X_t satisfacen la cláusula iid y las funciones de distribución y de densidad ($h_{Y|Z}^0$ y $h_{Y|Z}^0$), condicionales e incondicionales, tienen buen comportamiento.

Se desconocen los detalles relativos a la función de distribución verdadera por lo que se contemplan dos posibles familias rivales: $F_\theta = \{F_{Y|Z}(\theta); \theta \in \Theta \subset \mathfrak{R}^p\}$ y $G_\gamma = \{G_{Y|Z}(\gamma); \gamma \in \Gamma \subset \mathfrak{R}^q\}$, las cuales pueden hallarse anidadas, solaparse o ser estrictamente independientes. Nuevamente, las funciones de distribución y de densidad asociadas tienen buen comportamiento, de modo que existen las matrices de segundas derivadas:

$$\left. \begin{aligned} A_f(\theta) &= E^0 \left[\frac{\partial^2 \lg f(-; \theta)}{\partial \theta \partial \theta'} \right] & A_g(\gamma) &= E^0 \left[\frac{\partial^2 \lg g(-; \gamma)}{\partial \gamma \partial \gamma'} \right] \\ B_f(\theta) &= E^0 \left[\frac{\partial \lg f(-; \theta)}{\partial \theta} \frac{\partial \lg f(-; \theta)}{\partial \theta'} \right] & B_g(\gamma) &= E^0 \left[\frac{\partial \lg g(-; \gamma)}{\partial \gamma} \frac{\partial \lg g(-; \gamma)}{\partial \gamma'} \right] \\ B_{fg}(\theta) &= E^0 \left[\frac{\partial \lg f(-; \theta)}{\partial \theta} \frac{\partial \lg g(-; \gamma)}{\partial \gamma'} \right] \end{aligned} \right\} \quad (3.2)$$

donde E^0 significa valor esperado de acuerdo a la función de distribución cierta (y desconocida) de X_t .

Cualquiera que sea el caso, los estimadores máximo-verosímiles de θ y γ serán consistentes en el sentido de White (1982), convergiendo, si el modelo se encuentra mal especificado, hacia los *valores pseudo-verdaderos* de Sawa (1978). Conservarán igualmente la normalidad asintótica:

$$\left. \begin{aligned} \text{plim } \tilde{\theta}_n &= \theta^* & \sqrt{n}[\tilde{\theta}_n - \theta^*] &\xrightarrow{D} N[0; A_f^{-1}(\theta^*) B_f(\theta^*) A_f^{-1}(\theta^*)] \\ \text{plim } \tilde{\gamma}_n &= \gamma^* & \sqrt{n}[\tilde{\gamma}_n - \gamma^*] &\xrightarrow{D} N[0; A_g^{-1}(\gamma^*) B_g(\gamma^*) A_g^{-1}(\gamma^*)] \end{aligned} \right\} \quad (3.3)$$

siendo θ^* y γ^* los puntos de convergencia, y n el tamaño de la muestra. Este resultado permite, a su vez, estimar consistentemente las matrices de segundas derivadas de (3.2) y las matrices de covarianzas que intervienen en (3.3).

Por otro lado, la diferencia de las log-verosimilitudes:

$$LR_n(\tilde{\theta}_n; \tilde{\gamma}_n) = L_f(\tilde{\theta}_n) - L_g(\tilde{\gamma}_n) = \sum_{t=1}^n \lg \frac{f(Y_t | Z_t; \tilde{\theta}_n)}{g(Y_t | Z_t; \tilde{\gamma}_n)} \quad (3.4)$$

es una suma de n términos independientes, por lo que, tras normalizar la relación, se alcanzan resultados de convergencia casi segura:

$$\frac{LR_n(\tilde{\theta}_n; \tilde{\gamma}_n)}{n} \xrightarrow{as} E^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] \quad (3.5)$$

y en distribución:

$$\sqrt{n} \left[\frac{LR_n(\tilde{\theta}_n; \tilde{\gamma}_n)}{n} - E^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] \right] \xrightarrow{D} N(0; \omega^{*2}) \quad (3.6a)$$

$$2LR_n(\tilde{\theta}_n; \tilde{\gamma}_n) \xrightarrow{D} M(p+q; \lambda^*) \quad (3.6b)$$

El resultado de (3.6a) se obtiene para el caso en el que los modelos no son equivalentes, $f(Y_t | Z_t; \theta^*) \neq g(Y_t | Z_t; \gamma^*)$. En esa expresión, ω^{*2} se corresponde con la varianza poblacional del cociente entre las log-verosimilitudes, evaluado en los parámetros de convergencia:

$$\omega^{*2} = V^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] = E^0 \left[\left(\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right)^2 \right] - \left(E^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] \right)^2 \quad (3.7)$$

El resultado de (3.6b) se mantiene bajo el supuesto de que ambos modelos son observacionalmente equivalentes, $f(Y_t | Z_t; \theta^*) = g(Y_t | Z_t; \gamma^*)$, y la principal dificultad es que la distribución de convergencia no es estándar. En concreto, $M(m, \lambda)$ se corresponde con la suma de los cuadrados de m variables normales, ponderada cada una de ellas por el correspondiente elemento del vector λ : $M(m, \lambda) = \sum_{j=1}^m \lambda_j u_j^2$;

$u_j \sim \text{iid}N(0,1)$. En la distribución de convergencia de (3.6b), λ^* es el vector de (p+q) raíces características de la matriz W:

$$W = \begin{bmatrix} B_f(\theta^*) A_f^{-1}(\theta^*) & -B_{fg}(\theta^*; \gamma^*) A_g^{-1}(\gamma^*) \\ -B_{gf}(\theta^*; \gamma^*) A_f^{-1}(\theta^*) & B_g(\gamma^*) A_g^{-1}(\gamma^*) \end{bmatrix} \quad (3.8)$$

Por otro lado, el estadístico de varianza de (3.7) también puede utilizarse para medir la distancia existente entre ambos modelos. Tal como pone de manifiesto el propio Vuong, si los modelos son equivalentes la varianza será cero lo que conduce al contraste:

$$\left. \begin{aligned} H_0^{\bar{\omega}} : \omega^{*2} &= 0 \\ H_A^{\bar{\omega}} : \omega^{*2} &\neq 0 \end{aligned} \right\} \quad (3.9)$$

El estadístico de prueba es la varianza muestral de la sucesión de (3.4):

$$\tilde{\omega}_n^2 = \frac{1}{n} \sum_{t=1}^n \left(\lg \frac{f(Y_t | Z_t; \tilde{\theta}_n)}{g(Y_t | Z_t; \tilde{\gamma}_n)} \right)^2 - \left(\frac{1}{n} \sum_{t=1}^n \lg \frac{f(Y_t | Z_t; \tilde{\theta}_n)}{g(Y_t | Z_t; \tilde{\gamma}_n)} \right)^2 \quad (3.10)$$

o su versión corregida:

$$\hat{\omega}_n^2 = \tilde{\omega}_n^2 + \left(\frac{1}{n} LR_n(\tilde{\theta}_n; \tilde{\gamma}_n) \right)^2 \quad (3.11)$$

Bajo la hipótesis nula, ambos estadísticos cuentan con límites probabilísticos bien definidos:

$$plim \tilde{\omega}_n^2 = \omega^{*2} \quad (3.12a)$$

$$plim \hat{\omega}_n^2 = \omega^{*2} + E^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] \quad (3.12b)$$

Su distribución asintótica resulta ser la de (3.6b):

$$\left. \begin{aligned} n \tilde{\omega}_n^2 \\ n \hat{\omega}_n^2 \end{aligned} \right\} \xrightarrow{D} M(p+q; \lambda^*) \quad (3.13)$$

Estos resultados permiten desarrollar una estrategia de selección de modelos en la que se distinguen tres casos.

Modelos no anidados

Esta circunstancia se dará cuando $F_\theta \cap G_\gamma = \emptyset$, lo que implica que las funciones de densidad tendrán que ser, necesariamente, diferentes: $f(-; \theta^*) \neq g(-; \gamma^*)$. En este caso, el ratio de verosimilitudes conduce a un contraste de hipótesis nula simple frente a alternativa bilateral:

$$\left. \begin{aligned} H_0 : E^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] &= 0 \\ H_A : \left\{ \begin{aligned} H_f : E^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] &> 0 \\ H_g : E^0 \left[\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right] &< 0 \end{aligned} \right. \end{aligned} \right\} \quad (3.14)$$

El Teorema 5.1 de Vuong (1989, p. 318) establece que: ‘... if F_θ y G_γ are strictly non-nested, then:

- (i) under $H_0 : n^{-1/2} LR_n[\tilde{\theta}_n; \tilde{\gamma}_n] / \tilde{\omega}_n \xrightarrow{D} N(0;1)$
- (ii) under $H_f : n^{-1/2} LR_n[\tilde{\theta}_n; \tilde{\gamma}_n] / \tilde{\omega}_n \xrightarrow{as} +\infty$
- (iii) under $H_g : n^{-1/2} LR_n[\tilde{\theta}_n; \tilde{\gamma}_n] / \tilde{\omega}_n \xrightarrow{as} -\infty$
- (iv) properties (i) to (iii) hold ifi $\tilde{\omega}_n$ is replaced by $\hat{\omega}_n$ ‘.

Dado que los dos modelos son diferentes, la aceptación de la hipótesis nula debe interpretarse en el sentido de que la información muestral disponible no permite discriminar entre ambas alternativas.

Modelos solapados

Esta situación se producirá cuando existan funciones de distribución comunes $(F_\theta \cap G_\gamma \neq \emptyset)$ entre las dos especificaciones, pero no se encuentren anidadas ellas mismas $(F_\theta \not\subset G_\gamma \text{ y } G_\gamma \not\subset F_\theta)$. En estas circunstancias, es posible que ambas especificaciones sean iguales por lo que, en primer lugar, deberá contemplarse esa posibilidad. Una vez descartada la equivalencia, se procederá a discriminar entre los modelos. El contraste de la primera etapa es:

$$\left. \begin{array}{l} H_0 : f(-; \theta^*) = g(-; \gamma^*) \\ H_A : f(-; \theta^*) \neq g(-; \gamma^*) \end{array} \right\} \Leftrightarrow \left. \begin{array}{l} H_0^\omega : \omega^{*2} = 0 \\ H_A^\omega : \omega^{*2} \neq 0 \end{array} \right\} \quad (3.15)$$

para el que podrá utilizarse el estadístico varianza, definido en (3.10), junto con los resultados (3.11) a (3.13). En este sentido, el Teorema 6.1 de Vuong (1989, p. 321) indica que:

- ‘(i) under H_0^ω , for any $x \geq 0$, $\Pr[n \tilde{\omega}_n^2 \leq x] - M_{p+q}(x; \tilde{\lambda}_n^2) \xrightarrow{as} 0$
- (ii) under $H_A^\omega; n \tilde{\omega}_n^2 \xrightarrow{as} +\infty$
- (iii) properties (i) and (ii) hold for $n \hat{\omega}_n^2$ ‘.

La dificultad de esta propuesta reside en el cálculo de los valores críticos del contraste, dado que la distribución de convergencia no es estándar. Si se rechaza la hipótesis nula, la conclusión es que ambos modelos no son observacionalmente equivalentes y se podrá utilizar el estadístico LR, como en (3.14), a efectos de discriminar entre ellos.

Modelos anidados

En este caso, una de las alternativas contempladas es un caso particular de la otra ($G_\gamma \subset F_\theta$). Cuando el modelo amplio se encuentra bien especificado, la estrategia basada en el estadístico LR de (3.4) desemboca en el caso habitual, cuya distribución de referencia es la $\chi^2(p - q)$.

4.- El factor espacio en el planteamiento de Vuong.

El problema que pretendemos abordar con la estrategia de selección descrita en la sección anterior es muy particular. En concreto, se trata de elegir entre un modelo con *autocorrelación sustantiva* o con *autocorrelación residual* (M2 frente a M4 en la situación descrita en el Cuadro 1), una vez que se han obtenido indicios claros de subespecificación en la relación estática inicial (M8 ó M6). En la sección segunda se ha defendido una estrategia de actuación diferente basada en formular, inicialmente, un modelo lo más general posible que se tratará de simplificar a continuación. Sin embargo, en la práctica habitual domina la preferencia por la sencillez de la ecuación simple de M8 (en lugar de la matriz X se usará Z y el vector δ reemplaza a β , englobando tanto a M6 como a M8) para empezar la discusión. En este caso, si se aprecian carencias en la estructura dinámica de la ecuación, las alternativas pueden ser tanto M2 como M4. El problema que formulamos es el de determinar, sobre bases sólidas, qué modelo resulta preferible de entre ambos.

Tal como se acaba de definir, parece claro que se trata de un ejercicio de discriminación entre modelos no anidados, dado que previamente se ha descartado la validez del modelo estático de M8, ó M6. Los modelos M2 y M4 se encontrarían solapados solo en el caso de que se mantuviese la ecuación estática como tercera alternativa, pero esta circunstancia ha sido descartada por los contrastes de especificación habituales.

Otra cuestión a tener en cuenta es el alcance de la cláusula *iid* sobre la cual descansan, de forma explícita, los resultados de Vuong. Sus implicaciones parecen determinantes en el escenario que hemos descrito, caracterizado por la heterocedasticidad y la autocorrelación transversal. Sin embargo, vamos a comprobar que ninguna de estas circunstancias presenta mayores inconvenientes. Para ello basta con filtrar las variables del modelo, utilizando los vectores característicos de la matriz de pesos W empleada en la aplicación. Tal como se indica en Mur (1999), si esta matriz se ha especificado de tipo binario, admitirá la descomposición espectral básica:

$$W = Q\Lambda Q' \Leftrightarrow \begin{cases} \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_R \end{bmatrix} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_R) \\ Q = \begin{bmatrix} q_{11} & q_{21} & \dots & q_{R1} \\ q_{12} & q_{22} & \dots & q_{R2} \\ \dots & \dots & \dots & \dots \\ q_{1R} & q_{2R} & \dots & q_{RR} \end{bmatrix} = [q_1 \quad q_2 \quad \dots \quad q_R] \end{cases} \quad (4.1)$$

Los elementos de la diagonal principal de Λ son las R raíces características de W , mientras que cada columna de la matriz Q (q_j , $j=1, 2, \dots, R$) se corresponde con un vector característico de esa matriz.

Suponiendo que en el modelo existe autocorrelación sustantiva y se filtran las series por la matriz Q , el resultado es:

$$\left. \begin{aligned} y &= \rho W y + Z \delta_1 + u \\ u &\sim N(0, \sigma_1^2 I) \end{aligned} \right\} \Rightarrow \left. \begin{aligned} y^* &= \rho \Lambda y^* + Z^* \delta_1 + u^* \\ u^* &\sim N(0, \sigma_1^2 I) \end{aligned} \right\} \quad (4.2)$$

siendo $y^* = Q'y$, $Z^* = Q'Z$ y $u^* = Q'u$ las series resultantes del filtrado previo. La consecuencia es que el modelo inicial de (4.2), caracterizado por una estructura SAR en la variable endógena, se convierte en un modelo estático en las variables filtradas, no lineal en parámetros y cuya perturbación es heterocedástica:

$$\left. \begin{aligned} y^* &= \Delta^{-1} Z^* \delta_1 + v^* \\ v^* &\sim N(0, \sigma_1^2 \Delta^{-2}) \end{aligned} \right\} \quad (4.3)$$

con $\Delta = [I - \rho \Lambda]$ y $v^* = \Delta^{-1} u^*$. El aspecto importante de (4.3) es que se han eliminado las relaciones de dependencia transversal existentes en la ecuación de partida. Por otro lado, caso de existir una estructura de autocorrelación residual, el procedimiento de filtrado conduce a:

$$\left. \begin{aligned} y &= Z \delta_2 + u \\ u &= \rho W u + \varepsilon \\ \varepsilon &\sim N(0, \sigma_2^2 I) \end{aligned} \right\} \Rightarrow \left. \begin{aligned} y^* &= Z^* \delta_2 + u^* \\ u^* &= \Delta^{-1} \varepsilon^* \\ \varepsilon^* &\sim N(0, \sigma_2^2 I) \end{aligned} \right\} \quad (4.4)$$

donde $\varepsilon^* = Q'\varepsilon$. Las consecuencias se limitan a la aparición de un término de error heterocedástico ($u^* \sim N(0, \sigma_2^2 \Delta^{-2})$), libre de relaciones de dependencia transversal.

La función de verosimilitud del modelo de (4.2) es:

$$l(y^* | \phi_1) = -\frac{R}{2} \ln 2\pi - \frac{R}{2} \ln \sigma_1^2 - \sum_{r=1}^R \frac{[(1 - \rho \lambda_r) y_r^* - z_r^* \delta_1]^2}{2 \sigma_1^2} + \sum_{r=1}^R \ln(1 - \rho \lambda_r) \quad (4.5)$$

con $\phi_1 = [\delta_1; \rho; \sigma_1^2]$ y $z_r^* = [z_{1r}^*; z_{2r}^*; \dots; z_{2kr}^*]$ el vector de observaciones filtradas de las variables exógenas, en el punto r ; además se cumple que: $(1 - \rho\lambda_r)y_r^* - z_r^{*'}\delta_1 = \varepsilon_r^* \sim N(0, \sigma_1^2)$. La log-verosimilitud de (4.5) es una suma de R términos aleatorios independientes ($l(y^* | \phi_1) = \sum_{r=1}^R l_r(y_r^* | \phi_1)$), y admite una Ley de los Grandes Números (LGN; Davidson, 2000). Si el PGD es el de (4.2), los estimadores serán consistentes con respecto a ϕ_1 y se cumplirá que:

$$E^0[l_r(y_r^* | \phi_1)] = -\frac{1}{2}(\ln 2\pi + \ln \sigma_1^2 + 1) + \ln(1 - \rho\lambda_r) \quad (4.6a)$$

$$plim \frac{l(y^* | \phi_1)}{R} = -\frac{1}{2}(\ln 2\pi + \ln \sigma_1^2 + 1) + \ln(1 - \rho\lambda) = E^0[l(y^* | \phi_1)] \quad (4.6b)$$

siendo $\ln(1 - \rho\lambda) = \lim \frac{\sum_{r=1}^R \ln(1 - \rho\lambda_r)}{R}$. Si el PGD no se corresponde con las ecuaciones de (4.2), la convergencia de los estimadores MV se producirá hacia los *valores pseudo-verdaderos*, ϕ_1^* , y el resultado de (4.6b) se leerá en términos generales:

$$plim \frac{l(y^* | \phi_1^*)}{R} = E^0[l(y^* | \phi_1^*)] \quad (4.7)$$

Por otro lado, la función de log-verosimilitud asociada al modelo de (4.4) es:

$$l(y^* | \phi_2) = -\frac{R}{2} \ln 2\pi - \frac{R}{2} \ln \sigma_2^2 - \sum_{r=1}^R \frac{[(y_r^* - z_r^{*'}\delta_2)(1 - \theta\lambda_r)]^2}{2\sigma_2^2} + \sum_{r=1}^R \ln(1 - \theta\lambda_r) \quad (4.8)$$

con $\phi_2 = [\delta_2; \rho; \sigma_2^2]$, cumpliéndose que $(y_r^* - z_r^{*'}\delta_2)(1 - \theta\lambda_r) = \varepsilon_r^* \sim N(0, \sigma_2^2)$. Al igual que antes, si el PGD se corresponde con las ecuaciones de (4.4), los estimadores MV serán consistentes en relación a ϕ_2 y se verificará que:

$$E^0[l_r(y_r^* | \phi_2)] = -\frac{1}{2}(\ln 2\pi + \ln \sigma_2^2 + 1) + \ln(1 - \theta\lambda_r) \quad (4.9a)$$

$$plim \frac{l(y^* | \phi_2)}{R} = -\frac{1}{2}(\ln 2\pi + \ln \sigma_2^2 + 1) + \ln(1 - \theta\lambda) = E^0[l(y^* | \phi_2)] \quad (4.9b)$$

con $\ln(1 - \theta\lambda) = \lim \frac{\sum_{r=1}^R \ln(1 - \theta\lambda_r)}{R}$. Si las ecuaciones de (4.4) no se corresponden con el PGD, los estimadores serán consistentes con respecto a los *valores pseudo-verdaderos*, ϕ_2^* , y el resultado de convergencia se mantendrá aunque solo en términos generales:

$$plim \frac{l(y^* | \varphi_2^*)}{R} = E^0[l(y^* | \varphi_2^*)] \quad (4.10)$$

Combinando los resultados anteriores, el ratio de verosimilitudes de (3.4) se expresa como:

$$LR_R(\tilde{\varphi}_1; \tilde{\varphi}_2) = l(y^* | \tilde{\varphi}_1) - l(y^* | \tilde{\varphi}_2) = -\frac{R}{2} \ln \frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_2^2} + \sum_{r=1}^R \ln \frac{1 - \tilde{\rho} \lambda_r}{1 - \theta \lambda_r} \quad (4.11)$$

cuyo límite de convergencia puede establecerse, al igual que en (3.5), en términos generales⁸:

$$\frac{LR_R(\tilde{\varphi}_1; \tilde{\varphi}_2)}{R} \xrightarrow{p} E^0[l(y^* | \varphi_1^*)] - E^0[l(y^* | \varphi_2^*)] = E^0 \left[\lg \frac{L(y^* | \varphi_1^*)}{L(y^* | \varphi_2^*)} \right] \quad (4.12)$$

donde $L(-)$ se refiere a la función de verosimilitud. En definitiva, retomamos el resultado tradicional de que la media muestral (el ratio de verosimilitudes normalizado) de los términos de una sucesión de variables aleatorias independientes, es un estimador consistente del momento poblacional respectivo. Además, las funciones de log-verosimilitud de (4.5) y de (4.8), evaluadas en los parámetros del PGD (φ_1 y φ_2) o en los *seudo-verdaderos* de convergencia (φ_1^* y φ_2^*), admiten un Teorema Central del Límite (TCL; Davidson, 2000), el cual se extiende al propio ratio de verosimilitudes:

$$\sqrt{R} \left[\frac{LR_R(\varphi_1^*; \varphi_2^*)}{R} - E^0 \left(\lg \frac{L(y^* | \varphi_1^*)}{L(y^* | \varphi_2^*)} \right) \right] \xrightarrow{D} N(0; \varpi^{*2}) \quad (4.13)$$

con $\varpi^{*2} = V^0 \left(\lg \frac{L(y^* | \varphi_1^*)}{L(y^* | \varphi_2^*)} \right)$, lo mismo que ocurría en (3.7). Este último resultado

garantiza la aplicabilidad, en el contexto descrito, de la segunda parte del Teorema 3.3

de Voung (1989, p. 313): ‘(ii) if $f(-; \theta^*) \neq g(-; \gamma^*)$ (...) then

$$n^{-1/2} LR_n(\hat{\theta}_n; \hat{\gamma}_n) - n^{-1/2} E^0 \left(\lg \frac{f(Y_t | Z_t; \theta^*)}{g(Y_t | Z_t; \gamma^*)} \right) \xrightarrow{D} N(0; \varpi^{*2}), \quad (4.14)$$

lo cual completa la discusión.

⁸ Al ser modelos no anidados, el PGD no puede corresponderse, simultáneamente, con los dos. Al menos uno de ellos, o ambos, deberá ser erróneo.

En definitiva, se han identificado dos modelos rivales, el de (4.2) y el (4.4), los cuales se encuentran no anidados. La hipótesis nula y alternativa se concreta en:

$$H_0: E^0 \left[\lg \frac{L(y^* | \phi_1^*)}{L(y^* | \phi_2^*)} \right] = 0$$

$$H_A: \left\{ \begin{array}{l} H_1: E^0 \left[\lg \frac{L(y^* | \phi_1^*)}{L(y^* | \phi_2^*)} \right] > 0 \\ H_2: E^0 \left[\lg \frac{L(y^* | \phi_1^*)}{L(y^* | \phi_2^*)} \right] < 0 \end{array} \right\} \quad (4.15)$$

El estadístico de prueba es el de (4.11), el cual usaremos sabiendo que:

$$(i) \text{ bajo } H_0: R^{-1/2} LR_R[\tilde{\phi}_1; \tilde{\phi}_2] / \tilde{\omega}_R \xrightarrow{D} N(0;1)$$

$$(ii) \text{ bajo } H_1: R^{-1/2} LR_R[\tilde{\phi}_1; \tilde{\phi}_2] / \tilde{\omega}_R \xrightarrow{as} +\infty$$

$$(iii) \text{ bajo } H_2: R^{-1/2} LR_R[\tilde{\phi}_1; \tilde{\phi}_2] / \tilde{\omega}_R \xrightarrow{as} -\infty$$

siendo:

$$\tilde{\omega}_R^2 = \frac{1}{R} \sum_{r=1}^R \left(\lg \frac{L_r(y_r^* | \tilde{\phi}_1)}{L_r(y_r^* | \tilde{\phi}_2)} \right)^2 - \left(\frac{1}{R} \sum_{r=1}^R \lg \frac{L_r(y_r^* | \tilde{\phi}_1)}{L_r(y_r^* | \tilde{\phi}_2)} \right)^2 \Rightarrow$$

$$\Rightarrow \tilde{\omega}_R^2 = \frac{1}{R} \sum_{r=1}^R \left(\lg \frac{L_r(y_r^* | \tilde{\phi}_1)}{L_r(y_r^* | \tilde{\phi}_2)} \right)^2 - (LR_R(\tilde{\phi}_1, \tilde{\phi}_2))^2 \quad (4.16)$$

5.- Comentarios finales.

El programa de investigación en Econometría espacial se ha robustecido considerablemente durante los últimos años. Se han propuesto nuevos enfoques para tratar problemas relevantes, al mismo tiempo que se han revisado y actualizado diferentes instrumentos que forman parte de la tradición en esta disciplina. Sin embargo, existen ciertas facetas sobre las que tan apenas se registran aportaciones específicas como, por ejemplo, la de la selección de modelos en un contexto espacial. El papel que concluimos sección se encuentra en esta línea. No se trata, como es evidente, de un trabajo acabado y cerrado en sí mismo, sino que es un simple esbozo de un proyecto de investigación que todavía debe madurar a medio plazo.

Referencias

- Anselin, L. 1988a. *Spatial Econometrics: Methods and Models*. Dordrech: Kluwer.
- Anselin, L. 1988b. Lagrange Multiplier Tests Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20, 1-17.
- Anselin, L. y D. Griffith 1988. Do Spatial Effects Really Matter in Regression Analysis?. *Papers of the Regional Science Association*, 65, 4-34.
- Anselin, L. y A. Bera 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. En *Handbook of Applied Economics Statistics* editado por A. Ullah y D. Giles, 237-289. Nueva York: Marcial Dekker.
- Aznar, A. 1989. *Econometric Model Selection: A New Approach*. Dordrecht: Kluwer.
- Chasco, C. 2003. *Econometría espacial aplicada a la predicción-extrapolación de los datos microterritoriales*. Madrid: Consejería de Economía e innovación tecnológica, Comunidad de Madrid.
- Davidson, D. 2000. *Econometric Theory*. Oxford: Blackwell Publishers.
- Florax, R., H. Folmer y S. Rey 1998. *The Relevance of Hendry's Econometric Methodology*. Discussion Paper TI 98-125/4. Tinbergen Institute: Rotterdam.
- Florax, R. y T. de Graaff 2003. The Performance of Diagnostics Tests for Spatial Dependence in Linear Regression Models: A Meta-Analysis of Simulation Studies. En (eds.) por L. Anselin, R. Florax y S. Rey: *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin: Springer.
- Judge, G, W. Griffiths, R. Hill y T. Lee 1980. *The Theory and Practice of Econometrics*. Nueva York: John Wiley & Sons.
- Mizon, G. 1984. The Encompassing Approach in Econometrics. En (eds.) D. Hendry y K. Wallis: *Econometrics and Quantitative Economics*. Londres: Basil Blackwell.
- Mur, J. 1999. Heterocedasticidad y autocorrelación espacial: una propuesta de identificación de series espaciales, *Estadística Española*, 144, 79-104.
- Ord, J. 1975. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70, 120-126.
- Sawa, T. 1978. Information Criteria for Discriminating Among Alternative Regression Models, *Econometrica*, 46, 1273-1291.
- Vuong, Q. 1989. Likelihood Ratio-Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307-333.
- White, H. 1982. Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25.