



Munich Personal RePEc Archive

Which spatial weighting matrix? An approach for model selection

Herrera Gómez, Marcos and Mur Lacambra, Jesús and Ruiz Marín, Manuel

ASOCIACION ARGENTINA DE ECONOMIA POLITICA

2011

Online at <https://mpra.ub.uni-muenchen.de/37585/>
MPRA Paper No. 37585, posted 23 Mar 2012 06:36 UTC

¿Cuál Matriz de Pesos Espaciales? Un Enfoque sobre Selección de Modelos

Marcos Herrera^{(1)*}, Jesús Mur⁽¹⁾, Manuel Ruiz⁽²⁾

⁽¹⁾Universidad de Zaragoza; ⁽²⁾Universidad Politécnica de Cartagena

Resumen

En econometría espacial, es habitual especificar una matriz de pesos espaciales, denominada comúnmente W . Esta decisión es importante debido a que la elección de la matriz W condiciona el resto del análisis. Sin embargo, la elección de la matriz W , usualmente, refleja la subjetividad del investigador. Este trabajo revisa la literatura en busca de criterios de selección que ayuden a resolver este problema. Además, un nuevo procedimiento no-paramétrico es introducido. La propuesta se basa en una medida de información, entropía condicional, que utiliza información proveniente de los datos. Se comparan las alternativas mediante un experimento Monte Carlo.

Abstract

In spatial econometrics, it is customary to specify a weighting matrix, the so-called W matrix. The decision is important because the choice of W matrix determines the rest of the analysis. However, the procedure is not well defined and, usually, reflects the priors of the user. In the paper, we revise the literature looking for criteria to help with this problem. Also, a new nonparametric procedure is introduced. Our proposal is based on a measure of the information, conditional entropy, that uses information present in the data. We compare these alternatives by means of a Monte Carlo experiment.

Palabras Claves: *Econometría Espacial, Selección de Modelos, Entropía Simbólica.*

Código JEL: *C21, C52.*

*Autor para correspondencia: Departamento de Análisis Económico, Universidad de Zaragoza. Gran Vía 2-4 (50005). Zaragoza (España). Mail: mherreragomez@gmail.com

1. Introducción

La matriz de pesos espaciales, comúnmente identificada por W , es un elemento muy característico en modelos econométricos que utilizan datos geo-referenciados. Esta matriz permite introducir dependencia entre las unidades y, frecuentemente, es causa de disputas en relación a qué y cómo debería ser especificada. Desde hace medio siglo atrás, después de los pioneros trabajos de Moran (1948) y Whittle (1954), los términos “unidos” o “vinculados” fueron preferidos para iniciar la discusión. El trabajo de Ord (1975) realiza la conversión de estos términos dentro de un arreglo matricial y posiciona este elemento como clave para especificar modelos espaciales. Desde entonces, la matriz de pesos espaciales ha recibido considerable atención (Anselin, 2002), pero a pesar de ello no existen respuestas totalmente convincentes sobre los anteriores interrogantes.

De manera formal, la matriz de pesos espaciales es representada como:

$$W = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,j} & \cdots & w_{1,N} \\ w_{2,1} & 0 & \cdots & w_{2,j} & \cdots & w_{2,N} \\ \vdots & \vdots & \ddots & \cdots & \cdots & \cdots \\ w_{i,1} & w_{i,2} & \vdots & 0 & \cdots & w_{i,N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots \\ w_{N,1} & w_{N,2} & \vdots & w_{N,j} & \vdots & 0 \end{bmatrix}, \quad (1)$$

donde las filas y columnas identifican a las observaciones de corte transversal, siendo N el tamaño del conjunto de datos y $w_{i,j}$ ($i, j = 1, 2, \dots, N$) las ponderaciones o pesos que aproximan la relación entre dos unidades i (filas) y j (columnas). La diagonal principal esta formada por ceros, estableciendo que ninguna observación puede estar relacionada consigo misma (no podemos ser vecinos de nosotros mismos). En determinadas situaciones, suele añadirse un superíndice a la matriz para indicar el orden de contigüidad al que se refiere (en general, $W^{(j)}$ representa al j -ésimo orden de contigüidad, siendo $j \geq 1$). Es decir, por ejemplo, $W^{(2)}$ identificará al conjunto de vecinos de los vecinos de cada observación o vecinos de segundo orden.

Como criterio más ampliamente utilizado para elegir los pesos $w_{i,j}$ se encuentra el de cercanía o distancia geográfica, pero esto no es necesariamente válido para todas las aplicaciones como se pone de manifiesto mediante el conocido principio de *alotopía* (Ancot et al., 1982): “*A menudo lo que sucede en una región se encuentra relacionado con otro fenómeno localizado en otras partes diferentes y remotas del espacio*”. Este principio destaca que la utilización de este tipo de criterio puede ser una posición muy optimista ya que existe una gran incertidumbre que caracteriza a su elección.

El problema de identificar cuáles son las regiones u observaciones vinculadas y cómo introducir las al análisis no es un problema particular de la econometría espacial. En series temporales, un problema similar sucede cuando un analista debe enfrentar la introducción de dependencia aunque tiene en claro ciertas cosas: debido a la naturaleza dinámica de la economía, se deberá tomar en cuenta al pasado e introducirlo y, adicionalmente, se deberá considerar la frecuencia de los datos con los que se trabaja. Ninguna de estas dos cuestiones está libre de controversias cuando los datos provienen de un corte transversal geo-referenciado. El espacio es irregular y heterogéneo (nada comparado a la sucesión monótona en el tiempo) y las influencias pueden ser de cualquier tipo en el espacio. La consideración de la cercanía geográfica como criterio de orden, tal como es mencionada por Tobler (1970), es solo una de las posibilidades.

De acuerdo con Haining (2003, p. 74), “*el primer paso para cuantificar la estructura de dependencia espacial en un conjunto de datos es definir, para el conjunto de puntos o áreas, la relación espacial existen entre ellos*”. Esto es lo que Anselin (1988, p. 16) designa como “*la necesidad de determinar cuál de las otras unidades en el sistema espacial tiene influencia particular sobre la unidad bajo consideración (...) expresada en nociones topológicas de vecindario o vecinos más cercanos*”. Este primer paso es crucial, absolutamente, pero dicho paso puede no ser tan simple y directo como

escribir una matriz binaria o una matriz de pesos estandarizada por filas. En algunos casos, se puede tener información completa sobre la especificación de la matriz de ponderación. En otros casos, esta matriz será una mera hipótesis. Sospechamos que la segunda situación es, por lejos, la más habitual entre los trabajos aplicados.

Desde una perspectiva diferente, la matriz de pesos proviene de un problema de subidentificación que afecta, en general, a la mayoría de los modelos espaciales. Paelinck (1979, p. 20) reconoce que existe un problema de identificación en las especificaciones interdependientes utilizadas para modelizar la conducta espacial. En términos de LeSage y Pace (2009, p. 8), un proceso espacial autoregresivo¹ no restringido puede escribirse como:

$$\left. \begin{aligned} y_i &= \alpha_{ij}y_j + \alpha_{ik}y_k + x_i\beta + \varepsilon_i, \\ y_j &= \alpha_{ji}y_i + \alpha_{jk}y_k + x_j\beta + \varepsilon_j, \\ y_k &= \alpha_{ki}y_i + \alpha_{kj}y_j + x_k\beta + \varepsilon_k, \\ \varepsilon_i; \varepsilon_j; \varepsilon_k &\sim i.i.d.\mathcal{N}(0; \sigma^2), \end{aligned} \right\} \quad (2)$$

“donde es fácil ver su escasa utilidad práctica dado que esto resultará en un sistema con muchos más parámetros que observaciones (...). La solución al problema de la sobreparametrización, que surge cuando permitimos que cada relación de dependencia posea su propio parámetro, es imponer estructura sobre las relaciones de dependencia espacial”. Esta es la razón de por qué necesitamos una matriz de pesos espaciales. Esto es un punto de consenso en la literatura, a pesar del esfuerzo de Folmer y Oud (2008), intentando avanzar hacia lo que ellos denominan enfoque estructural para la matriz de pesos espaciales, o los argumentos de Paci y Usai (2009) en favor del uso de proxies para capturar efectos de desbordamiento.

La problemática de la especificación de la matriz parece más compleja, aunque la práctica usual ha favorecido al tipo de soluciones simples. El enfoque dominante implica un tratamiento exógeno del problema. Las unidades espaciales cercanas o vecinas son identificadas mediante una variable binaria usando como, por ejemplo, el tradicional criterio de adyacencia física o los k vecinos más cercanos, por mencionar los más comunes. Después de esto, la matriz binaria puede ser estandarizada en alguna forma habitual (por filas, columnas o de acuerdo a la suma total). Otras matrices son construidas usando alguna función de distancia entre los centroides de las unidades espaciales y con la posterior estandarización de la matriz. En algunas aplicaciones, la geografía puede ser sustituida por otro dominio como la similitud en las estructuras soio-económicas en orden de obtener las medidas de distancia.

Como procedimientos alternativos al exógeno, recientemente han aparecido los denominados procedimientos endógenos. Entre ellos, podemos mencionar el algoritmo AMOEBA (Getis y Aldsdad, 2004; Aldsdad y Getis, 2006), el método *CCC* de Mur y Paelinck (2010) o el enfoque basado en entropía de Esteban *et al.* (2009). Aunque son muy diferentes entre ellos, la idea básica de los enfoques endógenos apareció en los trabajos de Kooijman (1976) y Openshaw (1977): utilizar la información contenida en los datos, o en los residuos del modelo, en orden de estimar la matriz de pesos. Esto es posible si se tiene un panel de datos geo-referenciados como en Conley y Molinari (2007), Bhattacharjee y Jensen-Butler (2006) y Beenstock *et al.* (2010), pero es riesgoso para el caso de cortes transversales únicos. Finalmente, hay diferentes enfoques bien conocidos que combinan fuertes aprioris sobre los canales de conectividad con algoritmos inferenciales endógenos (Bodson y Peters, 1975; Dacey, 1965).

Bavaud (1998, p. 153), dado el estado del tópico, es claramente escéptico: *“no hay ninguna cosa tal como “verdadera”, “universal” matriz de pesos espaciales óptima en todas las situaciones”* y continúa afirmando que la matriz de pesos *“debe reflejar las propiedades del fenómeno en particular, propiedades que poseen límites diferentes de un campo bajo estudio a otro”*. Compartimos este escepticismo en el sentido que, al final, el problema de elegir una matriz de pesos espaciales entre

¹Es importante destacar que la dependencia espacial puede estar presente en la parte sistemática de las ecuaciones, en el término de error o en ambos términos simultáneamente. Por simplicidad, nos enfocaremos en la especificación espacial autoregresiva como modelo básico, siendo extensibles los comentarios para las demás alternativas.

diferentes posibilidades es un problema de selección de modelos. De hecho, diferentes matrices de pesos, resultarán en diferentes rezagos espaciales de las variable endógenas o exógenas incluidas en el modelo. Diferentes ecuaciones con diferentes regresores significan un problema de selección de modelos, aún cuando la matriz de pesos espaciales pueda aparecer en la parte no sistemática de la ecuación (en el término de error). Ésta es la dirección que se pretende explorar en el presente trabajo como un camino alternativo para tratar con la incertidumbre de especificación inherente a la matriz de pesos espaciales.

La sección 2 sigue con la revisión de las técnicas de selección de modelos que parecen tratar mejor este problema. Posteriormente, se presenta un procedimiento no paramétrico en la sección 3. La sección 4 desarrolla un experimento Monte Carlo en el que se comparan el comportamiento de los diferentes criterios analizados. La sección 5 presenta un resumen de los resultados comparados del experimento. Por último, se presentan las conclusiones de la investigación.

2. Escogiendo una Matriz de Pesos Espaciales

El modelo (2) puede ser escrito en forma matricial:

$$\mathbf{y} = \Gamma\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3)$$

donde \mathbf{y} y $\boldsymbol{\varepsilon}$ son vectores de orden $(n \times 1)$, X es una matriz de orden $(n \times k)$, $\boldsymbol{\beta}$ es un vector de $(k \times 1)$ parámetros y Γ es una matriz de coeficientes de interacción de orden $(n \times n)$. Tal como comentábamos, el modelo se encuentra subidentificado. Una solución, tal vez la más popular, consiste en la introducción de alguna estructura en la matriz Γ , parametrizando la interacción espacial como, por ejemplo: $\Gamma = \rho W$, donde ρ es un parámetro y W una matriz de pesos espaciales. El término $\mathbf{y}_W = W\mathbf{y}$ que, consecuentemente, aparecerá sobre el lado derecho de la ecuación es llamado rezago espacial de la variable endógena. En este punto es bueno recordar un par de cuestiones:

1. La matriz de pesos puede ser construida en diferentes formas siguiendo, por ejemplo, alguna hipótesis de interacción. Cada hipótesis resultará en una matriz de ponderaciones diferentes llevando a un rezago espacial distinto. En resumen, la matriz de ponderaciones contabiliza diferentes modelos.
2. Existen algunos lineamientos generales sobre como especificar la matriz de ponderaciones usando conceptos tales como cercanía, accesibilidad, influencia, etc. Diferentes modelos pueden requerir diferentes canales de interacción que no necesariamente son conocidos. Esto implica incertidumbre.

Corrado y Fingleton (2011) discuten la construcción de la matriz de pesos sobre una base teórica (esto es, ellos consideran, entre otras cosas, que la información sobre las ponderaciones debe provenir de la teoría económica). De manera alternativa, en este trabajo nos focalizaremos en el análisis del tema bajo un tratamiento estadístico de tal incertidumbre.

Asumamos que tenemos un conjunto de Q matrices de pesos espaciales linealmente independientes, $\Upsilon = \{W_1; W_2; \dots; W_Q\}$, siendo $Q \geq 2$. Usualmente Q se corresponde con un número pequeño de matrices diferentes que compiten entre si pero en otros casos este número puede ser bastante grande, reflejando una situación de mayor incertidumbre. Como se comentaba, cada matriz genera un rezago espacial diferente y un diferente modelo espacial. Estas matrices pueden ser relacionadas por medio de distintas restricciones, resultando en una serie de modelos anidados; si las matrices no se encuentran relacionadas, la secuencia de modelos espaciales será del tipo no-anidado.

Dos matrices de contactos pueden estar anidadas, por ejemplo, en el caso de movimientos tipo torre y tipo reina: todas las conexiones de la primera matriz se encuentran incluidas en la de la segunda, tal como se muestra en la Figura 1.

Figura 1: Criterios de Contigüidad

Tipo Torre					Tipo Reina				
		<i>j</i>				<i>j</i>	<i>j</i>	<i>j</i>	
	<i>j</i>	<i>i</i>	<i>j</i>			<i>j</i>	<i>i</i>	<i>j</i>	
		<i>j</i>				<i>j</i>	<i>j</i>	<i>j</i>	

Discriminar entre dos matrices de este tipo no es difícil utilizando técnicas de selección para modelos anidados. Por ejemplo, bajo un enfoque de máxima verosimilitud (se necesitaría el supuesto de normalidad), puede ser suficiente con un contraste de razón de verosimilitudes o multiplicador de Lagrange.

Para el caso de matrices no anidadas, pueden proponerse diferentes métodos. Anselin (1984) provee el estadístico de Cox para este caso:

$$\left. \begin{aligned} H_0 : \mathbf{y} &= \rho_1 W_1 \mathbf{y} + X_1 \beta_1 + \varepsilon_1, \\ H_A : \mathbf{y} &= \rho_2 W_2 \mathbf{y} + X_2 \beta_2 + \varepsilon_2, \end{aligned} \right\} \quad (4)$$

que Leenders (2002) lo convierte en una versión del contraste J usando una regresión aumentada como la siguiente:

$$\mathbf{y} = (1 - \alpha) [\rho_1 W_1 \mathbf{y} + X_1 \beta_1] + \alpha [\hat{\rho}_2 W_2 \mathbf{y} + X_2 \hat{\beta}_2] + \nu, \quad (5)$$

siendo $\hat{\rho}_2$ y $\hat{\beta}_2$ las correspondientes estimaciones máximo-verosímiles de los respectivos parámetros de una estimación separada de H_A . Leenders generaliza además la comparación de la hipótesis nula contra Q diferentes modelos. Kelejian (2008) sostiene el enfoque bajo un esquema $SARAR^2$, que requiere de estimadores de momentos generalizados, GMM :

$$\begin{aligned} \mathbf{y} &= \rho_i W_i \mathbf{y} + X_i \beta_i + u_i = Z_i \gamma_i + \mathbf{u}_i, \\ \mathbf{u}_i &= \lambda_i M_i \mathbf{u}_i + v_i, \end{aligned} \quad (6)$$

con $i = 1, 2, \dots, Q$, $Z_i = (W_i \mathbf{y}, X_i)$ y $\gamma_i' = (\rho_i, \beta_i')$. El contraste J para el problema de selección de la matriz de pesos se corresponde al caso donde $X_i = X$; $W_i = M_i$ pero $W_i \neq W_j, \forall W_i, W_j \subseteq \Upsilon$. En orden de obtener el contraste necesitamos la estimación de una regresión aumentada, similar a la ecuación (5):

$$\mathbf{y}(\hat{\lambda}) = S(\hat{\lambda})\eta + \varepsilon, \quad (7)$$

donde $S(\hat{\lambda}) = [Z(\hat{\lambda}), F]$, $Z(\hat{\lambda}) = (I - \hat{\lambda}W)$ (de igual forma que para \mathbf{y}), siendo $\hat{\lambda}$ la estimación de λ para el modelo de la nula. Además, $F = [Z_1 \hat{\gamma}_1, Z_2 \hat{\gamma}_2, \dots, Z_Q \hat{\gamma}_Q, W_1 Z_1 \hat{\gamma}_1, W_2 Z_2 \hat{\gamma}_2, \dots, W_Q Z_Q \hat{\gamma}_Q]$.

²Según Kelejian, la elección de un modelo inicial como el SARAR es debido a que es uno de los modelos más complejos (incorpora endogeneidad y estructura espacial en el error, simultáneamente). Las demás especificaciones pueden verse como simplificaciones de esta estructura.

La ecuación de (7) puede ser estimada por MC2E usando una como instrumentos: $\hat{S} = [\hat{Z}(\hat{\lambda}), \hat{F}]$, donde $\hat{F} = PF$ (igual que para $Z(\hat{\lambda})$) con $P = H(H'H)^{-1}H$ y $H = [X, WX, W^2X]$. Bajo la hipótesis nula que el modelo es correcto, la estimación por MC2E de η es asintóticamente normal:

$$\hat{\eta} \sim \mathcal{N} \left[\eta_0; \sigma_\epsilon^2 \left(\hat{S}'\hat{S} \right)^{-1} \right], \quad (8)$$

donde $\eta_0 = [\gamma'; 0]$. El contraste J asume, bajo H_0 , que los últimos $2Q$ parámetros del vector η son iguales a cero.

Definamos $\hat{\delta} = A\hat{\eta}$ donde A es una matriz de $2Q \times (k + 1 + 2Q)$ correspondiente a la hipótesis nula: $H_0 : A\eta = 0$, entonces el contraste J puede ser formulado como un estadístico de Wald:

$$\hat{\delta}'\hat{V}^{-1}\hat{\delta} \sim \chi^2(2Q), \quad (9)$$

siendo \hat{V} la covarianza muestral estimada de $\hat{\delta}$.

Burridge y Fingleton (2010) muestran que la distribución asintótica Chi-Cuadrado para el contraste J, bajo la nula, puede ser una pobre aproximación. Como alternativa, proponen un procedimiento de bootstrap que mejora el tamaño y la potencia. Burridge (2011) propone construir el contraste usando estimadores mixtos entre GMM y máxima verosimilitud que controlan más efectivamente el tamaño del contraste. Piras y Lozano (2010) presentan nueva evidencia sobre el uso del contraste J que relaciona la potencia en función de la selección de instrumentos.

El problema de la selección de modelos ha sido tratada muy a menudo, y de forma exitosa, desde una perspectiva bayesiana (Leamer, 1978); esto incluye el caso de selección de una matriz de pesos espaciales en un modelo espacial (Hepple, 1995a,b). El enfoque bayesiano, aunque altamente demandante en términos de información, es atractivo y poderoso. Sobre el método, puede verse una buena introducción en LeSage y Pace (2009). Lo mismo que en el caso del contraste J, el punto de inicio es un conjunto finito de modelos alternativos, $M = \{M_1; M_2; \dots; M_Q\}$. La especificación de cada modelo coincide (regresores, estructura de dependencia, etc.) pero la matriz de pesos espaciales es diferente. Denotando por θ al vector de parámetros, entonces la probabilidad conjunta de los Q modelos con n observaciones corresponde a:

$$p(M, \theta, \mathbf{y}) = \pi(M) \pi(\theta|M) L(\mathbf{y}|\theta, M), \quad (10)$$

donde $\pi(M)$ se refiere a las probabilidades *a priori* o iniciales de los modelos, usualmente $\pi(M) = 1/Q$; $\pi(\theta|M)$ refleja las probabilidades *a priori* del vector de parámetros condicional al modelo y $L(\mathbf{y}|\theta, M)$ es la verosimilitud de los datos condicionados por los parámetros y los modelos. Usando la regla de Bayes:

$$p(M, \theta|\mathbf{y}) = \frac{p(M, \theta, \mathbf{y})}{p(\mathbf{y})} = \frac{\pi(M) \pi(\theta|M) L(\mathbf{y}|\theta, M)}{p(\mathbf{y})}. \quad (11)$$

La probabilidad *a posteriori* de los modelos, condicionada por los datos, resulta de la integración de (10) sobre el vector de parámetros θ :

$$p(M|\mathbf{y}) = \int p(M, \theta|\mathbf{y}) d\theta. \quad (12)$$

Esta última es la medida de probabilidad necesaria en orden de comparar diferentes matrices de pesos. LeSage y Pace (2009) discuten el caso de un modelo gaussiano denominado *SAR*:

$$\left. \begin{aligned} \mathbf{y} &= \rho_i W_i \mathbf{y} + X_i \beta_i + \varepsilon_i, \\ \varepsilon_i &\sim i.i.d. \mathcal{N}(0; \sigma_\epsilon^2). \end{aligned} \right\} \quad (13)$$

La log-verosimilitud marginal de (13) es:

$$p(M|\mathbf{y}) = \int \pi_\beta(\beta|\sigma^2) \pi_\sigma(\sigma^2) \pi_\rho(\rho) L(\mathbf{y}|\theta, M) d\beta d\sigma^2 d\rho. \quad (14)$$

Asumiendo: (1) independencia entre las probabilidades *a priori* asignadas a β y σ^2 , (2) distribución conjugada *a priori* Gamma-Inversa-Normal, y (3) para ρ una distribución *Beta*(d, d) *a priori*. Los cálculos no son simples y, finalmente, “debemos confiar en una integración numérica univariada sobre el parámetro ρ para convertir ésta (expresión 14) a una expresión escalar necesaria para calcular $p(M|y)$ que es útil para la comparación de modelos” (LeSage y Pace, 2009, p 172). El caso del modelo de error espacial, *SEM*, es resuelto en LeSage y Parent (2007); para nuestro conocimiento, el modelo *SARAR* de (6) permanece aún sin resolución.

Recientemente, Hansen (2007) introdujo otra perspectiva al problema de selección de modelos que se relaciona a la confianza de los investigadores sobre los modelos en la hipótesis alternativa. En general, el criterio propuesto minimiza el error cuadrático medio estimado buscando un balance entre sesgo (debido a la incorrecta especificación) y varianza (debido a la estimación de los parámetros). El criterio óptimo debería seleccionar el estimador con menor riesgo. Interpretando de una manera alternativa, dada una colección de matrices espaciales $W = \{W_1; W_2; \dots; W_Q\}$, todas de las cuales pertenecen a un modelo espacial de referencia, la propuesta es elegir la matriz W_q que, combinada con otros términos del modelo, produce un vector de estimaciones, $\hat{\theta}_q(W_q)$, que minimiza el riesgo. Hansen (2007) muestra que el mejor ajuste respecto al error cuadrático medio puede ser obtenido mediante la ponderación de los estimadores. El estimador promedio para θ es:

$$\hat{\theta}(W) = \sum_{q=1}^Q \varpi^q \hat{\theta}_q(W_q). \quad (15)$$

Como establecen Hansen y Racine (2010), la colección de ponderaciones, $\{\varpi^q; q = 1, 2, \dots, Q\}$ debe ser no-negativa y cumplir con la restricción del simplex \mathbb{R}^Q : $\sum_{q=1}^Q \varpi^q = 1$.

Posteriormente, estas ponderaciones pueden ser utilizadas para comparar el ajuste de cada modelo (matriz de pesos) respecto a los datos.

3. Una Propuesta No-Paramétrica para Seleccionar W

La propuesta de esta sección es presentar un nuevo procedimiento no-paramétrico para la selección de la matriz de pesos espaciales. El criterio de selección se basa en la información contenida en la distribución espacial de la relación investigada. La medida de información que utilizamos es una reformulación del índice tradicional de entropía en términos de lo que denominamos *entropía simbólica*.

Como es explicado por Matilla y Ruiz (2008), la idea es, primero, transformar las series en una secuencia de símbolos que capturan información relevante. Posteriormente, se traslada la inferencia al espacio de símbolos mediante la aplicación de técnicas adecuadas.

Comenzando con el proceso de simbolización, asumamos que $\{x_s\}_{s \in S}$ e $\{y_s\}_{s \in S}$ son dos procesos espaciales, donde S es un conjunto de localizaciones en el espacio. Denotamos con $\Gamma_l = \{\sigma_1, \sigma_2, \dots, \sigma_l\}$ al conjunto de símbolos definido por el usuario; siendo σ_i , para $i = 1, 2, \dots, l$, un símbolo. Simbolizar un proceso es definir una función

$$f : \{x_s\}_{s \in S} \rightarrow \Gamma_l, \quad (16)$$

tal que cada elemento x_s es asociado a un único símbolo $f(x_s) = \sigma_{i_s}$, con $i_s \in \{1, 2, \dots, l\}$. Diremos que la localización $s \in S$ es σ_i -*tipo*, relativo a la serie $\{x_s\}_{s \in S}$, si y solo si $f(x_s) = \sigma_{i_s}$. Llamaremos a f *función simbolizadora*. El mismo procedimiento puede ser realizado para el proceso y_s .

Denotemos con $\{Z_s\}_{s \in S}$ al proceso bivalente:

$$Z_s = \{x_s, y_s\}. \quad (17)$$

Para este caso, definimos un conjunto de símbolos Ω_l como el producto directo de los dos conjuntos de símbolos Γ_l , tal que $\Omega_l^2 = \Gamma_l \times \Gamma_l$, cuyos elementos son de la forma $\eta_{ij} = (\sigma_i^x, \sigma_j^y)$. La función simbolizadora del proceso bivalente será

$$g : \{Z_s\}_{s \in S} \rightarrow \Omega_l^2 = \Gamma_l \times \Gamma_l, \quad (18)$$

definida como

$$g(Z_s = (x_s, y_s)) = (f(x_s), f(y_s)) = \eta_{ij} = (\sigma_i^x, \sigma_j^y). \quad (19)$$

Diremos que la localización s es η_{ij} - tipo para $Z = (x, y)$ si y solo si s es σ_i^x - tipo para x y σ_j^y - tipo para y .

Definidos los principales conceptos, haremos uso de la siguiente función simbolizadora f . Sea M_e^x la mediana del proceso espacial univariante $\{x_s\}_{s \in S}$ y definamos una función indicadora

$$\tau_s = \begin{cases} 1 & \text{si } x_s \geq M_e^x, \\ 0 & \text{en otro caso.} \end{cases} \quad (20)$$

Sea $m \geq 2$ la dimensión de encaje, definida por el investigador. Para cada $s \in S$, sea N_s el conjunto de vecinos formado por los $(m-1)$ vecinos de s . Usaremos en término m - entorno para designar al conjunto formado por cada s y su N_s , tal que el m - entorno $x_m(s) = (x_s, x_{s_1}, \dots, x_{s_{m-1}})$. Ahora, definamos una función indicadora para cada s_i con $i = 1, 2, \dots, m-1$:

$$\iota_{ss_i} = \begin{cases} 0 & \text{si } \tau_s \neq \tau_{s_i}, \\ 1 & \text{en otro caso.} \end{cases} \quad (21)$$

Finalmente, presentamos una función simbolizadora para el proceso espacial $\{x_s\}_{s \in S}$ como $f : \{x_s\}_{s \in S} \rightarrow \Gamma_m$, donde:

$$f(x_s) = \sum_{i=1}^{m-1} \iota_{ss_i}, \quad (22)$$

con $\Gamma_m = \{0, 1, \dots, m-1\}$. La cardinalidad de Γ_m es igual a m .

Adicionalmente, necesitamos introducir algunas definiciones fundamentales:

Definición 1: La entropía de Shannon, $h(x)$, de una variable aleatoria discreta x es:

$$h(x) = -\sum_{i=1}^n p(x_i) \ln(p(x_i)).$$

Definición 2: La entropía $h(x, y)$ de un par de variables aleatorias discretas (x, y) con distribución conjunta $p(x, y)$ es:

$$h(x, y) = -\sum_x \sum_y p(x, y) \ln(p(x, y)).$$

Definición 3: La entropía condicional $h(x|y)$ con distribución $p(x, y)$ es definida como:

$$h(x|y) = -\sum_x \sum_y p(x, y) \ln(p(x|y)).$$

La última definición, $h(x|y)$, es la entropía x que permanece cuando y ha sido observada.

Estas medidas de entropía pueden ser adaptadas a la distribución empírica de los símbolos. Una vez que las series han sido simbolizadas, para una dimensión de encaje $m \geq 2$, podemos estimar la frecuencia absoluta y relativa de la colección de símbolos $\sigma_{i_s}^x \in \Gamma_l$ y $\sigma_{j_s}^y \in \Gamma_l$.

La frecuencia absoluta del símbolo σ_i^x es:

$$n_{\sigma_i^x} = \# \{s \in S | s \text{ es } \sigma_i^x \text{ - tipo para } x\}. \quad (23)$$

Similarmente, para la serie $\{y_s\}_{s \in S}$, la frecuencia absoluta del símbolo σ_j^y es:

$$n_{\sigma_j^y} = \# \{s \in S | s \text{ es } \sigma_j^y - \text{tipo para } y\}. \quad (24)$$

Luego, la frecuencia relativa puede ser estimada como:

$$p(\sigma_i^x) \equiv p_{\sigma_i^x} = \frac{\# \{s \in S | s \text{ es } \sigma_i^x - \text{tipo para } x\}}{|S|} = \frac{n_{\sigma_i^x}}{|S|}, \quad (25)$$

$$p(\sigma_j^y) \equiv p_{\sigma_j^y} = \frac{\# \{s \in S | s \text{ es } \sigma_j^y - \text{tipo para } y\}}{|S|} = \frac{n_{\sigma_j^y}}{|S|}, \quad (26)$$

donde $|S|$ denota la cardinalidad del conjunto S ; en general $|S| = R$.

Similarmente, calculamos la frecuencia relativa para $\eta_{ij} \in \Omega_i^2$:

$$p(\eta_{ij}) \equiv p_{\eta_{ij}} = \frac{\# \{s \in S | s \text{ es } \eta_{ij} - \text{tipo}\}}{|S|} = \frac{n_{\eta_{ij}}}{|S|}. \quad (27)$$

Finalmente, la *entropía simbólica* para la serie espacial *bidimensional* $\{Z_s\}_{s \in S}$ es:

$$h_Z(m) = - \sum_{\eta \in \Omega_m^2} p(\eta) \ln(p(\eta)). \quad (28)$$

Además podemos obtener las entropías simbólicas marginales como

$$h_x(m) = - \sum_{\sigma^x \in \Gamma_m} p(\sigma^x) \ln(p(\sigma^x)), \quad (29)$$

$$h_y(m) = - \sum_{\sigma^y \in \Gamma_m} p(\sigma^y) \ln(p(\sigma^y)). \quad (30)$$

A su vez, podemos obtener la entropía simbólica de y , condicionada por la ocurrencia del símbolo σ^x en x como:

$$h_{y|\sigma^x}(m) = - \sum_{\sigma^y \in \Gamma_m} p(\sigma^y|\sigma^x) \ln(p(\sigma^y|\sigma^x)). \quad (31)$$

Puede estimarse adicionalmente la entropía condicional simbólica de y_s dado x_s :

$$h_{y|x}(m) = \sum_{\sigma^x \in \Gamma_m} p(\sigma^x) h_{y|\sigma^x}(m). \quad (32)$$

Ahora, estamos en condiciones de presentar el problema de la selección de la matriz de pesos espaciales para la relación entre las variables x e y . Esta selección será realizada entre un conjunto finito de matrices, relevantes para la relación analizada. Denotemos por $\mathcal{W}(x, y) = \{W_q | q \in Q\}$ al conjunto de matrices, donde Q es un conjunto de índices. Nos referiremos a $\mathcal{W}(x, y)$ como el conjunto de estructuras espacio-dependientes entre x e y .

Denotamos con \mathcal{K} al subconjunto de Γ_m y sea $W \in \mathcal{W}(x, y)$ un miembro del conjunto de matrices. Así podemos definir

$$\mathcal{K}_W^x = \{\sigma^x \in \mathcal{K} | \sigma^x \text{ es admisible para } Wx\}, \quad (33)$$

donde *admisible* indica que la probabilidad de ocurrencia del símbolo es positiva.

Por Γ_m^x denotaremos al conjunto de símbolos que son admisibles para $\{x_s\}_{s \in S}$. Sea $W_0 \in \mathcal{W}(x, y)$ la matriz de pesos espaciales más informativa para la relación entre x e y . Dado un proceso espacial $\{y_s\}_{s \in S}$, existe un subconjunto $\mathcal{K} \subseteq \Gamma_m$ tal que $p(\mathcal{K}_{W_0}^x | \sigma^y) > p(\mathcal{K}_W^{*x} | \sigma^y)$ para todo $\mathcal{K}^* \subseteq \Gamma_m$, $W \in \mathcal{W}(x, y) \setminus \{W_0\}$ y $\sigma^y \in \Gamma_m^y$. Entonces

$$\begin{aligned}
h_{W_0x|y}(m) &= - \sum_{\sigma^y \in \Gamma^y} p(\sigma^y) \left[\sum_{\sigma^x \in \mathcal{K}_{W_0}^x} p(\sigma^x|\sigma^y) \ln(p(\sigma^x|\sigma^y)) \right] \\
&\leq - \sum_{\sigma^y \in \Gamma^y} p_{\sigma^y} \left[\sum_{\sigma^x \in \mathcal{K}_W^{*x}} p(\sigma^x|\sigma^y) \ln(p(\sigma^x|\sigma^y)) \right] = h_{Wx|y}(m).
\end{aligned} \tag{34}$$

Así hemos demostrado el siguiente teorema:

Teorema 1: Sean $\{x_s\}_{s \in S}$ e $\{y_s\}_{s \in S}$ dos procesos espaciales. Para una dimensión de encaje fija $m \geq 2$, con $m \in \mathbb{N}$, si la más informativa matriz de pesos espaciales que revela la estructura espacio-dependiente entre x e y es $W_0 \in \mathcal{W}(x, y)$ entonces

$$h_{W_0x|y}(m) = \min_{W \in \mathcal{W}(x, y)} \{h_{Wx|y}(m)\}. \tag{35}$$

Dado el Teorema 1 y usando la siguiente propiedad: $h_{Wx|y} \leq h_{Wx}$, proponemos el siguiente criterio para la selección entre matrices de igual número de vecinos:

$$pseudo - R^2 = 1 - h_{Wx|y}(m)/h_{Wx}(m).$$

La selección de la matriz se realiza por medio del valor más elevado del $pseudo - R^2$.

4. Experimento Monte Carlo

En esta sección, generamos un gran número de muestras desde diferentes procesos generadores de datos (*P.G.D.*) para estudiar la conducta de las diferentes propuestas: criterio Hansen-Racine, criterio bayesiano, contraste J y entropía condicional simbólica.

Nuestro principal interés es detectar la matriz de pesos espaciales más informativa entre diferentes alternativas. Para esto, hemos supuesto una variable explicativa, x , que es la misma para todos los modelos, pero estructuras espaciales diferentes, tal que $W = W_q$, donde q es la matriz del q -ésimo modelo alternativo.

Una gran variedad de criterios para la especificación de matrices espaciales son posibles, sin embargo en este estudio restringiremos nuestra atención al criterio de k -vecinos. Además, pueden especificarse una amplia variedad de modelos: en este trabajo consideramos, por simplicidad, modelos con variables explicativas espacialmente dependientes.

Cada experimento se inicia mediante la obtención de un mapa aleatorio en una hipotética superficie bidimensional. Este mapa irregular es reflejado por su correspondiente matriz de contactos W estandarizada. En un primer paso, W es una matriz binaria de 1 y 0 denotando si las regiones son vecinas o no, respectivamente, y subsecuentemente se estandariza por la suma de los elementos de cada fila tal que la suma de los pesos sea igual a 1.

Los siguientes parámetros globales son considerados en el *P.G.D.* para modelos anidados:

$$N \in \{100, 400, 700, 1000\}, k \in \{4, 5, 7\}, \tag{36}$$

donde N es el tamaño muestral y el parámetro k es el número de vecinos de cada observación utilizado para construir la matriz W . Es decir:

- $W_4 = 4 - \text{vecinos} - \text{más} - \text{ceranos}$.
- $W_5 = 5 - \text{vecinos} - \text{más} - \text{ceranos}$.
- $W_7 = 7 - \text{vecinos} - \text{más} - \text{ceranos}$.

donde W_7 anida a la matriz W_5 y W_5 anida a la matriz W_4 , previa estandarización.

En el caso de los modelos no-anidados, todas las matrices de pesos espaciales contienen 4 vecinos pero se modifica el criterio de vecindad. En estos casos, hemos supuesto las siguientes matrices no-anidadas:

- $W^{(1)} = 4 - \text{vecinos} - \text{más} - \text{cercanos}$.
- $W^{(2)} = 5^\circ - \text{al} - 8^\circ - \text{vecinos} - \text{más} - \text{cercanos}$.
- $W^{(1-2)} = 1^\circ - 2^\circ - 5^\circ - 6^\circ - \text{vecinos} - \text{más} - \text{cercanos}$.

En el experimento, hemos simulado relaciones lineales y no-lineales entre las variables x e y .

En el primer caso, linealidad, controlamos la relación mediante el coeficiente de determinación esperado basado en la siguiente especificación:

$$\mathbf{y} = \beta\mathbf{x} + \theta W\mathbf{x} + \varepsilon, \quad (37)$$

y la fortaleza de la relación puede ser deducida por el coeficiente R^2 esperado.

Bajo la ecuación (37), el coeficiente de determinación esperado entre las variables, asumiendo varianza unitaria de x y ε así como incorrelación entre ambas variables, es igual a:

$$R^2 = \frac{\beta^2 + (\theta^2/k)}{\beta^2 + (\theta^2/k) + 1}.$$

Para este coeficiente hemos considerado diferentes valores:

$$R^2 \in \{0,4; 0,6; 0,8\} \quad (38)$$

Por simplicidad, en todos los casos hemos mantenido $\beta = 0,5$. El parámetro de rezago espacial de x , θ , es obtenido por deducción: $\theta = \sqrt{\frac{(-k)(\beta^2(1-R^2)-R^2)}{1-R^2}}$.

Habiendo definido los valores de los parámetros involucrados en la simulación, podemos presentar los diferentes procesos analizados.

PGD0: Lineal

$$\mathbf{y} = \beta\mathbf{x} + \theta W\mathbf{x} + \varepsilon \quad (39)$$

PGD1: No-lineal 1

$$\mathbf{y} = \exp \left[(\beta\mathbf{x} + \theta W\mathbf{x} + \varepsilon)^{1,25} \right] \quad (40)$$

PGD2: No-lineal 2

$$\mathbf{y} = 1/(\beta\mathbf{x} + \theta W\mathbf{x} + \varepsilon)^2 \quad (41)$$

En todos los casos: $x_i \sim \mathcal{N}(0, 1)$, $\varepsilon_i \sim \mathcal{N}(0, 1)$ y $Cov(x_i, \varepsilon_i) = 0$.

5. Resultados del Experimento

Los resultados para los modelos anidados se presentan en los Cuadros 1-6. Cuando el proceso es lineal, *PGD0*, la selección realizada por los criterios de Hansen-Racine y Bayesiano es cercana al 100 % en casi todas las situaciones. El comportamiento de los contrastes J y LM es parecido, con resultados que superan el 85 % de selección correcta en casi todos los casos. El LM es levemente superior para el caso de W_7 , Cuadro 2, aproximándose al comportamiento de Hansen-Racine y Bayesiano.

Cuadro 1: Proceso Lineal. Modelos Anidados

Criterios		Hansen-Racine			Bayesiano		
Matrices		W_4	W_5	W_7	W_4	W_5	W_7
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	92.7	82.1	86.0	91.5	83.9	85.6
	0,6	99.7	97.3	98.7	99.6	98.0	98.4
	0,8	100.0	100.0	99.9	100.0	100.0	99.9
$N = 400$	0,4	100.0	98.9	99.6	100.0	99.3	99.4
	0,6	100.0	100.0	100.0	100.0	100.0	100.0
	0,8	100.0	100.0	100.0	100.0	100.0	100.0
$N = 700$	0,4	100.0	100.0	100.0	100.0	100.0	100.0
	0,6	100.0	100.0	100.0	100.0	100.0	100.0
	0,8	100.0	100.0	100.0	100.0	100.0	100.0
$N = 1000$	0,4	100.0	100.0	100.0	100.0	100.0	100.0
	0,6	100.0	100.0	100.0	100.0	100.0	100.0
	0,8	100.0	100.0	100.0	100.0	100.0	100.0

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuadro 2: Proceso Lineal. Modelos Anidados

Criterios		Contraste J			LM		
Matrices		W_4	W_5	W_7	W_4	W_5	W_7
N	R^2	% Selección					
$N = 100$	0,4	71.2	53.1	55.8	89.7	73.3	60.9
	0,6	89.5	86.5	87.1	90.3	91.0	97.3
	0,8	87.4	85.6	88.1	88.0	90.5	99.9
$N = 400$	0,4	89.5	88.7	88.5	90.4	92.3	99.9
	0,6	87.8	85.3	87.7	88.7	91.5	100.0
	0,8	89.9	86.2	88.0	91.0	90.4	100.0
$N = 700$	0,4	87.4	87.4	89.2	88.8	93.7	100.0
	0,6	89.1	85.2	87.1	89.4	91.7	100.0
	0,8	91.0	86.8	87.4	91.8	92.7	100.0
$N = 1000$	0,4	88.4	86.6	89.5	89.1	92.2	100.0
	0,6	90.6	87.8	90.5	91.7	93.1	100.0
	0,8	87.8	86.2	89.6	89.1	91.9	100.0

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuando el proceso generador es no-lineal, $PGD1$, los resultados se modifican sensiblemente (Cuadros 3-4). El criterio Bayesiano es el de mejor comportamiento con un valor máximo del 89% de selección cuando el tamaño muestral es igual a 1000. El contraste LM tiende a seleccionar matrices subidentificadas y por ello se observan altos porcentajes de selección para W_4 . En este caso, los R^2 solo son presentados para identificar el valor que interviene en la generación de θ .

Cuadro 3: Proceso No-lineal 1. Modelos Anidados

Criterios		Hansen-Racine			Bayesiano		
Matrices		W_4	W_5	W_7	W_4	W_5	W_7
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	27.4	24.4	37.3	64.5	49.4	60.7
	0,6	26.7	24.2	34.2	75.6	61.8	68.7
	0,8	28.2	20.2	29.9	76.2	59.7	71.2
$N = 400$	0,4	30.0	25.0	41.8	80.5	69.8	75.3
	0,6	28.7	22.0	37.1	85.4	73.9	80.5
	0,8	38.0	17.6	37.3	84.1	68.9	76.9
$N = 700$	0,4	28.5	23.3	42.2	86.9	72.9	80.4
	0,6	28.8	22.0	41.0	88.7	80.0	82.8
	0,8	41.7	18.9	40.8	87.2	75.0	80.7
$N = 1000$	0,4	29.3	23.8	46.9	89.0	79.1	83.2
	0,6	33.0	22.9	40.0	88.8	82.0	85.0
	0,8	39.5	20.7	37.5	88.0	76.3	81.5

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuadro 4: Proceso No-lineal 1. Modelos Anidados

Criterios		Contraste J			LM		
Matrices		W_4	W_5	W_7	W_4	W_5	W_7
N	R^2	% Selección					
$N = 100$	0,4	14.9	4.5	10.3	88.6	21.8	12.7
	0,6	28.4	13.2	18.3	88.1	35.2	23.1
	0,8	28.7	12.7	20.0	91.1	32.1	24.8
$N = 400$	0,4	42.3	26.8	34.4	89.5	50.4	38.7
	0,6	53.8	33.7	42.9	89.1	53.7	46.5
	0,8	48.1	28.0	35.3	91.0	50.6	39.3
$N = 700$	0,4	58.2	35.3	43.4	90.3	55.7	47.7
	0,6	61.7	45.9	53.1	90.5	63.9	58.4
	0,8	54.1	33.8	42.3	88.7	57.7	47.1
$N = 1000$	0,4	66.4	50.1	51.4	91.1	65.4	58.0
	0,6	67.6	52.2	58.8	89.2	68.5	62.2
	0,8	60.5	39.0	47.0	88.6	56.1	52.2

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuando la no-linealidad se acentúa, PGD_2 , ningún criterio de los presentados es capaz de brindarnos información adecuada sobre el verdadero proceso generador. En este caso, se puede observar como el LM tiende a seleccionar la matriz con menor cantidad de vecinos en todos los casos, muestra de esto son los bajos valores para W_5 y W_7 .

Cuadro 5: Proceso No-lineal 2. Modelos Anidados

Criterios		Hansen-Racine			Bayesiano		
Matrices		W_4	W_5	W_7	W_4	W_5	W_7
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	38.8	25.2	32.7	30.9	23.5	27.2
	0,6	42.7	27.3	36.5	31.1	23.8	31.0
	0,8	42.7	31.7	44.3	34.3	22.2	29.2
$N = 400$	0,4	40.6	23.3	36.2	34.9	24.2	29.3
	0,6	40.7	26.8	38.7	30.1	23.1	28.2
	0,8	44.9	32.5	44.3	29.2	23.1	28.2
$N = 700$	0,4	39.3	25.1	33.1	35.2	22.5	32.3
	0,6	41.4	26.9	38.8	33.4	21.6	29.7
	0,8	46.7	33.1	43.2	29.9	21.1	27.6
$N = 1000$	0,4	37.7	22.6	33.9	31.5	24.4	31.9
	0,6	42.6	26.5	36.4	33.0	21.5	28.3
	0,8	43.2	30.9	42.0	33.0	22.4	29.5

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuadro 6: Proceso No-lineal 2. Modelos Anidados

Criterios		Contraste J			LM		
Matrices		W_4	W_5	W_7	W_4	W_5	W_7
N	R^2	% Selección					
$N = 100$	0,4	0.3	0.4	0.2	89.4	5.5	0.6
	0,6	0.5	0.0	0.6	90.6	5.3	1.6
	0,8	0.4	0.0	0.5	90.0	3.3	0.9
$N = 400$	0,4	0.4	0.0	0.2	89.9	5.1	0.3
	0,6	0.3	0.1	0.3	89.1	4.3	0.9
	0,8	0.2	0.0	0.1	90.8	3.5	0.3
$N = 700$	0,4	0.4	0.2	0.4	90.7	5.1	1.0
	0,6	0.1	0.0	0.0	90.4	4.7	1.0
	0,8	0.1	0.0	0.3	90.3	4.7	0.9
$N = 1000$	0,4	0.7	0.0	0.4	88.6	4.0	1.2
	0,6	0.4	0.0	0.0	89.0	5.0	0.3
	0,8	0.4	0.1	0.3	89.7	3.2	0.7

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

A continuación, en los Cuadros 7-12, se presentan los resultados para modelos no anidados. De igual manera como sucedía en el caso de los modelos anidados, cuando el proceso es lineal, $PGD0$, la selección realizada por los criterios de Hansen-Racine y Bayesiano es prácticamente del 100%. El comportamiento del contraste J se mantiene en torno al 88% de selección correcta. Respecto al criterio de entropía condicional, su comportamiento mejora a medida que el R^2 y el tamaño muestral se incrementa, superando en la mayoría de las situaciones al contraste J.

Cuadro 7: Proceso Lineal. Modelos No Anidados

Criterios		Hansen-Racine			Bayesiano		
Matrices		$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$	$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	99.6	99.4	99.8	99.6	99.5	99.8
	0,6	100.0	100.0	100.0	100.0	100.0	100.0
	0,8	100.0	100.0	100.0	100.0	100.0	100.0
$N = 400$	0,4	100.0	100.0	100.0	100.0	100.0	100.0
	0,6	100.0	100.0	100.0	100.0	100.0	100.0
	0,8	100.0	100.0	100.0	100.0	100.0	100.0
$N = 700$	0,4	100.0	100.0	100.0	100.0	100.0	100.0
	0,6	100.0	100.0	100.0	100.0	100.0	100.0
	0,8	100.0	100.0	100.0	100.0	100.0	100.0
$N = 1000$	0,4	100.0	100.0	100.0	100.0	100.0	100.0
	0,6	100.0	100.0	100.0	100.0	100.0	100.0
	0,8	100.0	100.0	100.0	100.0	100.0	100.0

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuadro 8: Proceso Lineal. Modelos No Anidados

Criterios		Contraste J			Entropía Condicional		
Matrices		$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$	$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	87.1	87.4	86.7	56.2	47.6	43.2
	0,6	88.7	88.1	86.4	81.0	66.6	66.3
	0,8	88.0	89.1	86.8	95.2	91.0	91.8
$N = 400$	0,4	88.7	86.9	88.0	83.7	59.5	67.4
	0,6	86.1	85.7	85.3	99.2	96.3	95.5
	0,8	87.9	88.6	87.8	100.0	100.0	100.0
$N = 700$	0,4	86.5	87.6	87.6	94.8	77.7	80.5
	0,6	86.9	88.7	87.9	100.0	99.5	99.3
	0,8	85.8	89.1	86.6	100.0	100.0	100.0
$N = 1000$	0,4	88.1	87.9	86.8	98.5	85.8	86.0
	0,6	89.0	88.1	87.5	100.0	99.8	99.7
	0,8	87.9	87.7	87.7	100.0	100.0	100.0

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

En el primer proceso no-lineal, $PGD1$, el criterio Bayesiano mantiene su buen comportamiento con valores superiores al 90% en casi todos los casos. El contraste J, mantiene su comportamiento con valores superiores al 80%, excepto para N igual a 100. La entropía condicional se comporta adecuadamente, alcanzando valores del 100% en numerosas situaciones.

Cuadro 9: Proceso No-lineal 1. Modelos No Anidados

Criterios		Hansen-Racine			Bayesiano		
Matrices		$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$	$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	14.6	13.4	26.9	81.8	84.8	79.8
	0,6	13.6	12.8	26.3	93.1	91.6	91.0
	0,8	21.3	22.8	27.8	95.0	94.7	95.0
$N = 400$	0,4	14.6	14.7	26.1	95.0	96.1	95.6
	0,6	19.3	17.4	27.4	97.9	98.5	97.2
	0,8	28.7	27.8	31.2	98.0	97.4	97.5
$N = 700$	0,4	16.8	13.9	26.2	97.5	97.1	97.0
	0,6	20.7	22.7	24.1	98.5	98.3	98.8
	0,8	32.8	31.7	28.8	98.3	98.3	99.9
$N = 1000$	0,4	15.5	15.5	21.9	98.8	98.5	97.9
	0,6	21.2	23.4	28.8	99.1	99.3	99.0
	0,8	35.4	34.3	33.3	99.3	99.0	99.0

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuadro 10: Proceso No-lineal 1. Modelos No Anidados

Criterios		Contraste J			Entropía Condicional		
Matrices		$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$	$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	51.1	50.2	47.3	56.2	38.6	43.2
	0,6	70.9	68.0	67.5	81.0	66.7	66.3
	0,8	71.6	73.6	73.4	95.2	91.2	91.8
$N = 400$	0,4	80.0	79.9	77.7	85.4	59.5	68.1
	0,6	83.3	83.3	81.6	99.3	96.3	95.5
	0,8	84.0	85.2	83.2	100.0	100.0	100.0
$N = 700$	0,4	80.9	85.0	80.9	96.0	75.8	79.0
	0,6	84.1	86.0	83.3	100.0	99.2	98.7
	0,8	84.6	86.2	86.0	100.0	100.0	100.0
$N = 1000$	0,4	86.3	83.5	83.6	98.5	86.8	86.7
	0,6	85.2	85.9	83.9	100.0	99.8	99.7
	0,8	85.3	86.0	84.7	100.0	100.0	100.0

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Para el proceso no-lineal 2, igual como sucedía en los casos anidados, los criterios Hansen-Racine, Bayesiano y el contraste J no brindan información útil sobre el verdadero proceso generador. El criterio de entropía es claramente superior, excepto para tamaños muestrales reducidos, alcanzando el 100 % de selección cuando el proceso utiliza $W^{(1)}$ y $N = 1000$.

Cuadro 11: Proceso No-lineal 2. Modelos No Anidados

Criterios		Hansen-Racine			Bayesiano		
Matrices		$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$	$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	35.7	34.3	37.0	30.3	30.0	29.9
	0,6	40.0	39.9	41.8	27.9	31.6	27.2
	0,8	46.1	44.9	43.9	21.8	24.2	22.7
$N = 400$	0,4	34.8	32.5	36.1	26.1	27.1	28.2
	0,6	42.2	42.1	41.3	27.5	28.4	29.3
	0,8	48.7	49.0	47.6	24.1	22.8	23.7
$N = 700$	0,4	34.6	34.4	34.8	27.7	28.2	30.8
	0,6	41.5	39.6	40.9	27.4	28.0	30.7
	0,8	48.0	47.1	46.3	22.5	23.1	20.9
$N = 1000$	0,4	35.9	36.2	35.8	28.6	30.8	26.6
	0,6	43.5	41.1	41.6	27.9	27.3	27.9
	0,8	50.3	48.0	47.3	21.3	22.7	23.2

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

Cuadro 12: Proceso No-lineal 2. Modelos No Anidados

Criterios		Contraste J			Entropía Condicional		
Matrices		$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$	$W^{(1)}$	$W^{(2)}$	$W^{(1-2)}$
N	R^2	% Selección	% Selección	% Selección	% Selección	% Selección	% Selección
$N = 100$	0,4	2.9	2.6	2.1	44.2	31.7	31.0
	0,6	2.2	1.7	1.3	52.7	40.3	42.9
	0,8	1.0	1.0	0.5	62.2	51.7	52.4
$N = 400$	0,4	2.2	2.1	2.4	68.0	42.6	44.9
	0,6	1.5	1.4	1.6	84.1	70.4	72.6
	0,8	0.3	0.4	0.3	92.2	84.6	87.2
$N = 700$	0,4	1.8	2.0	2.1	79.4	57.6	63.8
	0,6	0.9	1.3	1.2	95.0	86.2	85.9
	0,8	0.4	0.2	0.3	98.2	96.3	96.2
$N = 1000$	0,4	2.1	2.2	1.5	85.6	61.3	67.2
	0,6	0.9	1.2	1.0	97.7	93.1	93.0
	0,8	0.3	1.0	0.1	100.0	99.2	98.9

Nota: % Selección es el porcentaje de veces que cada matriz W es elegida "correctamente". Repeticiones: 1000.

6. Conclusiones

El trabajo intenta brindar criterios adecuados para seleccionar la matriz de pesos espaciales. El punto de partida es que el problema de elegir una matriz de pesos espaciales es un problema de selección de modelos. Basados en esta dirección hemos explorado enfoques alternativos que tratan con la incertidumbre de especificación.

Entre los criterios conocidos, el criterio bayesiano es el de mejor comportamiento. El contraste J, que ha sido propuesto en numerosas oportunidades para este tipo de selección, no es adecuado en la mayoría de las situaciones contempladas.

Respecto al criterio de entropía condicional, puede mencionarse como ventaja su simplicidad y buen comportamiento. Obsérvese que no es necesaria una determinada especificación, únicamente se asume que existe una estructura espacial que relaciona las variables bajo análisis. Esto contrasta fuertemente con los métodos hasta ahora conocidos, donde en general se requiere linealidad, especificación correcta (rezago espacial en la variable endógena, en el error o en los regresores según corresponda), generalmente normalidad, y adicionalmente se debe estimar de manera adecuada los parámetros que intervienen. Como una limitación, podemos mencionar que el criterio de entropía condicional solo es válido para estructuras espaciales con igual número de vecinos. Esto genera que no pueda usarse para la selección de estructuras anidadas.

En la futura agenda de investigación queda por presentar el comportamiento de los mismos criterios ante modelos espacialmente dinámicos y ampliar el criterio de entropía condicional para comparar modelos anidados.

Referencias

- [1] Aldstadt, J. y A. Getis (2006). "Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters", *Geographical Analysis*, 38, pp. 327-343.
- [2] Ancot, L, Paelinck, J., Klaassen, L. y W. Molle (1982). "Topics in Regional Development Modelling", en M. Albegov, A. Andersson y F. Snickars (eds), *Regional Development Modelling in Theory and Practice*, Amsterdam, North Holland.
- [3] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- [4] Anselin, L. (2002). "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models", *Agricultural Economics*, 17, pp. 247-267.
- [5] Bavaud, F. (1998). "Models for Spatial Weights: a Systematic Look", *Geographical Analysis*, 30, pp. 153-171.
- [6] Beenstock, M., Ben Zeev, N. y D. Felsenstein (2010). "Nonparametric Estimation of the Spatial Connectivity Matrix using Spatial Panel Data", *Working Paper*, Department of Geography, Hebrew University of Jerusalem.
- [7] Bhattacharjee, A. y C. Jensen-Butler (2006). "Estimation of Spatial Weights Matrix, with an Application to Diffusion in Housing Demand", *Working Paper*, School of Economics and Finance, University of St. Andrews, UK.
- [8] Bodson, P. y D. Peters (1975). "Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation: An Application to a Belgium Labor Demand Function", *Environment and Planning A*, 7, pp. 455-472.
- [9] Burrige, P. (2011). "Improving the J Test in the SARAR Model by Likelihood-based Estimation", *Working Paper*, Department of Economics and Related Studies, University of York.

- [10] Burridge, P. y B. Fingleton (2010). "Bootstrap Inference in Spatial Econometrics: The J-Test", *Spatial Economic Analysis*, 5, pp. 93-119.
- [11] Conley, T. y F. Molinari (2007). "Spatial Correlation Robust Inference with Errors in Location or Distance", *Journal of Econometrics*, 140, pp. 76-96.
- [12] Corrado, L. y B. Fingleton (2011). "Where is Economics in Spatial Econometrics?", Working Paper, Department of Economics, University of Strathclyde.
- [13] Dacey, M. (1965). "A Review on Measures of Contiguity for Two and k-Color Maps", en J. Berry y D. Marble (eds.), *A Reader in Statistical Geography*. Englewood Cliffs, Prentice-Hall.
- [14] Fernández, E., Mayor, M. y J. Rodríguez (2009). "Estimating Spatial Autoregressive Models by GME-GCE Techniques", *International Regional Science Review*, 32, pp. 148-172.
- [15] Folmer, H. y J. Oud (2008). "How to get rid of W? A Latent Variable Approach to Modeling Spatially Lagged Variables", *Environment and Planning A*, 40, pp. 2526-2538.
- [16] Getis, A. y J. Aldstadt (2004). "Constructing the Spatial Weights Matrix Using a Local Statistic Spatial", *Geographical Analysis*, 36, pp. 90-104.
- [17] Haining, R. (2003). *Spatial Data Analysis*. Cambridge, Cambridge University Press.
- [18] Hansen, B. (2007). "Least Squares Model Averaging", *Econometrica*, 75, pp. 1175-1189.
- [19] Hansen, B. y J. Racine (2010). "Jackknife Model Averaging", *Working Paper*, Department of Economics, McMaster University.
- [20] Hepple, L. (1995a). "Bayesian Techniques in Spatial and Network Econometrics: 1 Model Comparison and Posterior Odds. *Environment and Planning A*, 27, pp. 447-469.
- [21] Hepple, L. (1995b). "Bayesian Techniques in Spatial and Network Econometrics: 2 Computational Methods and Algorithms", *Environment and Planning A*, 27, pp. 615-644.
- [22] Kelejian, H. (2008). "A spatial J-test for Model Specification Against a Single or a Set of Non-Nested Alternatives", *Letters in Spatial and Resource Sciences*, 1, pp. 3-11.
- [23] Kooijman, S. (1976). "Some Remarks on the Statistical Analysis of Grids Especially with Respect to Ecology", *Annals of Systems Research*, 5.
- [24] Leamer, E. (1978). *Specification Searches: Ad Hoc Inference with Non Experimental Data*. New York: John Wiley and Sons, Inc.
- [25] Leenders, R. (2002). "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix", *Social Networks*, 24, pp. 21-47.
- [26] Lesage, J. y K. Pace (2009). *Introduction to Spatial Econometrics*. Boca Raton, CRC Press.
- [27] Lesage, J. y O. Parent (2007). "Bayesian Model Averaging for Spatial Econometric Models", *Geographical Analysis*, 39, pp. 241-267.
- [28] Matilla, M. y M. Ruiz (2008). "A Non-parametric Independence Test using Permutation Entropy", *Journal of Econometrics*, 144, pp. 139-155.
- [29] Moran, P. (1948). "The Interpretation of Statistical Maps", *Journal of the Royal Statistical Society B*, 10, pp. 243-251.
- [30] Mur, J. y J. Paelinck (2010). "Deriving the W-matrix via P-median Complete Correlation Analysis of Residuals", *The Annals of Regional Science*, DOI: 10.1007/s00168-010-0379-3.

- [31] Openshaw, S. (1977). "Optimal Zoning Systems for Spatial Interaction Models", *Environment and Planning A*, 9, pp. 169-184.
- [32] Ord, K. (1975). "Estimation Methods for Models of Spatial Interaction", *Journal of the American Statistical Association*, 70, pp. 120-126.
- [33] Paci, R. y S. Usai (2009). "Knowledge Flows across European Regions", *The Annals of Regional Science*, 43, pp. 669-690.
- [34] Paelinck, J y L. Klaassen (1979). *Spatial Econometrics*. Farnborough: Saxon House.
- [35] Piras, G y N. Lozano (2010). "Spatial J-test: Some Monte Carlo Evidence", *Statistics and Computing*, DOI: 10.1007/s11222-010-9215-y.
- [36] Tobler, W. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region", *Economic Geography*, 46, pp. 234-240.
- [37] Whittle, P. (1954). "On Stationary Processes in the Plane", *Biometrika*, 41, pp. 434-449.