



Munich Personal RePEc Archive

**Introduction to Spatial Econometrics:
An application to the study of fertility in
Argentina using R.**

Herrera Gómez, Marcos and Cid, Juan Carlos and Paz, Jorge
Augusto

2012

Online at <https://mpra.ub.uni-muenchen.de/41138/>
MPRA Paper No. 41138, posted 09 Sep 2012 11:32 UTC

Introducción a la Econometría Espacial. Una Aplicación al Estudio de la Fecundidad en la Argentina usando *R*.

Marcos Herrera^{(2)*}, Jorge A. Paz⁽¹⁾, Juan C. Cid⁽²⁾

⁽¹⁾CONICET- IELDE, Universidad Nacional de Salta

⁽²⁾IELDE, Universidad Nacional de Salta

Resumen

La econometría espacial es una rama econométrica relativamente joven pero con un gran crecimiento en las últimas décadas. La complejidad del análisis espacial ha sido uno de los principales obstáculos para la aplicación en estudios empíricos. El objetivo de este trabajo es contribuir a la difusión de las herramientas espaciales desarrolladas. Concretamente, el documento realiza una concisa revisión de los aspectos teóricos que involucran el tratamiento espacial. Asimismo, se presenta una aplicación empírica de las técnicas abordadas. Usando el programa estadístico *R*, se analizan los determinantes de la fecundidad en la Argentina.

Palabras Clave: *Econometría Espacial, Autocorrelación Espacial, Fecundidad, Programa Estadístico R.*

Código JEL: *C21, R12, J13.*

Abstract

Spatial econometrics is a relatively young branch econometric but with a great growth in the last decades. The complexity of spatial analysis and the estimation of spatial models has been the major obstacle for applied studies. The aim of this paper is to contribute to the diffusion of spatial tools developed. Specifically, this paper performs a concise review of the theoretical aspects that involve the spatial treatment. We also present an empirical application of the techniques discussed. Using the statistical program *R*, we analyze the determinants of fertility in Argentina.

Keywords: *Spatial Econometrics, Spatial Autocorrelation, Fertility, R Programming Language.*

JEL Codes: *C21, R12, J13.*

*Autor para correspondencia: IELDE, Universidad Nacional de Salta. Av. Bolivia 5150 (A4408FVY), Salta (Argentina). Mail: mherreragomez@gmail.com

1. Introducción

La econometría espacial es un sub-campo relativamente joven dentro de la econometría que incorpora el tratamiento de los efectos espaciales y los problemas que los mismos provocan en los modelos econométricos que utilizan datos de corte transversal y de panel. Esta disciplina ha tenido un importante crecimiento en las últimas décadas debido a tres principales razones (Anselin y Florax, 1995): (1) La importancia del espacio geográfico dentro de la teoría económica para explicar aspectos como los rendimientos crecientes y las economías de aglomeración. Específicamente, la Nueva Geografía Económica (Krugman, 1991) ha reconocido la importancia de la localización como factor influyente en la determinación de las variables económicas; (2) La existencia de grandes bases de datos socio-económicos georeferenciados como REIS en Estados Unidos, sostenida por el departamento de Análisis Económico (BEA), y REGIO en Europa, sostenida por el departamento estadístico de la Comisión Europea (Eurostat); (3) El avance tecnológico en el manejo de grandes bases de datos con información espacial: Sistemas de Información Geográfica y programas específicos para manipular y modelizar estos datos.

Los mencionados avances han generado que la econometría espacial adquiera especial relevancia dentro de las principales líneas de investigación econométrica. Pero a pesar de su relativa importancia, en la Argentina la presencia de este tipo de investigación económica es notoriamente escasa. Es por ello que nuestro trabajo busca incentivar, desde una revisión teórica y posterior ilustración empírica, la investigación en esta área.

El presente trabajo analiza la incorporación de los efectos espaciales¹ en un contexto de regresión lineal. En particular el efecto espacial que analizaremos es la dependencia espacial, comúnmente denominada autocorrelación espacial². En este caso, la econometría clásica ha confinado el tratamiento de la dependencia dentro de un contexto temporal. La mayoría de libros de texto tales como Goldberger (1991), Greene (1993), Davidson y MacKinnon (1993), Kennedy (1998), Ruud (2000), Hayashi (2000), por citar algunos, no mencionan la posibilidad de autocorrelación espacial en datos de corte transversal (para este tipo de datos el principal problema que los textos destacan es la heterocedasticidad). Sin embargo, otras áreas científicas como la geología, ecología, epidemiología han incorporado el tratamiento de dependencia espacial como sumamente relevante (véase Fischer y Getis, 2010, y sus referencias). En geografía, la primera ley establecida por Tobler (1979) destaca que “todo está relacionado con todo, pero las cosas próximas están más relacionadas que las distantes”, dando a entender que la dependencia espacial es más una norma que una excepción.

La estructura del trabajo es de la siguiente manera. La segunda sección se limita a revisar la incorporación de la dependencia espacial para datos de corte transversal (no se discute la dependencia en espacio-tiempo ni modelos con variable dependiente limitada). La tercera sección presenta una aplicación empírica para los departamentos³ de la Argentina. En esta sección aplicamos los conceptos desarrollados al análisis de la fecundidad mediante el uso del programa estadístico *R*. La cuarta sección presenta las principales conclusiones del trabajo. Adicionalmente, se incorpora un apéndice con la secuencia de comandos en *R* utilizados para la estimación de los diferentes modelos.

¹Los efectos espaciales se pueden dividir en dos tipos: dependencia y heterogeneidad. El tratamiento de la heterogeneidad espacial hace referencia a la inestabilidad estructural en los coeficientes del modelo (coeficientes variables o regímenes espaciales) o la presencia de heterocedasticidad en el término de error. En este trabajo este tipo de efecto espacial no será analizado (véase Anselin, 1988).

²Claramente los dos conceptos no son sinónimos, en términos estadísticos, siendo el concepto de dependencia un término más amplio. Por simplicidad, en este trabajo, utilizaremos ambos términos de forma intercambiable.

³El uso del término *departamentos* es válido para la mayoría de las provincias de la Argentina. Las dos excepciones son la provincia de Buenos Aires que denomina a estas áreas como *partidos* y la Ciudad Autónoma de Buenos Aires que se divide en *distritos*. Por simplicidad en este trabajo utilizaremos el término *departamento* para referirnos indistintamente a este nivel de agregación regional del país.

2. Incorporando la Dependencia Espacial en el Modelo Econométrico

Desde un punto de vista metodológico, la inclusión de la dependencia espacial en los modelos econométricos ha sido motivada por dos vías alternativas. Por un lado, el espacio se ha incorporado mediante fundamentos teóricos, es decir, siguiendo una especificación formal dentro de un modelo económico tal que refleje la interacción de los agentes y/o la interacción social (Brock y Durlauf, 1995). Por otro lado, desde los inicios ha existido una solución pragmática, introduciendo el espacio como una peculiaridad de los datos utilizados tales como la interdependencia espacial y la relevancia de los factores localizados en otros lugares (Paelinck y Klaassen, 1979).

Formalmente, la autocorrelación espacial puede definirse mediante la autocovarianza de la siguiente manera:

$$\text{Cov}(y_i, y_j) = E(y_i y_j) - E(y_i) \cdot E(y_j) \neq 0, \quad \forall i \neq j, \quad (1)$$

donde y_i e y_j son observaciones de una variable aleatoria localizada en i y j sobre el espacio.

Suponiendo que el conjunto de observaciones de un corte transversal es igual a N , en total existen $[N \cdot (N - 1) / 2]$ autocovarianzas espaciales para todos los posibles pares de observaciones (considerando la simetría de la matriz de autocovarianzas). Dado que el número de parámetros a estimar excede al número de observaciones es imposible estimar cada una de las autocovarianzas. Adicionalmente, incrementar el tamaño muestral no soluciona el problema ya que el número de parámetros a estimar también crece, siendo este caso un ejemplo típico del problema de parámetros incidentales (Lancaster, 2000).

Desde una perspectiva univariante, la dependencia espacial puede ser representada mediante un modelo espacial autoregresivo no restringido tal como:

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

siendo A una matriz de coeficientes de interacción con la siguiente estructura:

$$A = \begin{bmatrix} 0 & \alpha_{ij} & \alpha_{ik} \\ \alpha_{ji} & 0 & \alpha_{jk} \\ \alpha_{ki} & \alpha_{kj} & 0 \end{bmatrix}.$$

El modelo (2) es de "(...) *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 una estructura sobre las relaciones de dependencia espacial*" (LeSage y Pace, 2009, p. 8).

En otros términos, el modelo (2) se encuentra subidentificado y la solución clásica radica en imponer alguna estructura en la matriz A , parametrizando la interacción espacial como, por ejemplo: $A = \rho W$, donde ρ es un parámetro a estimar y W una matriz de contactos (también denominada matriz de contigüidades, pesos, ponderaciones, distancias o interacciones espaciales).

2.1. Matriz de Contactos W . Su Importancia Clave

La matriz de contactos permite incorporar el espacio dentro del análisis econométrico y ocupa una posición central dado que, esencialmente, define el conjunto de vecinos para cada localización. Tal como hemos enunciado, la necesidad de una matriz de contactos proviene del problema de subidentificación que afecta a la mayoría de los modelos espaciales. Haining (2003, p. 74) menciona que "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". Este primer

paso es crucial y, en puntuales casos, se puede tener información completa sobre la especificación de dicha matriz. En otros casos, en su gran mayoría, esta matriz es una mera hipótesis de trabajo.

Utilizando la estructura $A = \rho W$, el modelo (2) se transforma en un modelo restringido:

$$\mathbf{y} = \rho W \mathbf{y} + \varepsilon, \quad (3)$$

donde el término $W \mathbf{y}$ es comúnmente denominado rezago espacial de la variable endógena.

Suponiendo que el tamaño muestral es igual a N , la matriz W es de orden $N \times N$ y puede representarse de la siguiente manera:

$$W = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,j} & \cdots & w_{1,N} \\ w_{2,1} & 0 & \cdots & w_{2,j} & \cdots & w_{2,N} \\ \vdots & \vdots & \ddots & \cdots & \cdots & \cdots \\ w_{i,1} & w_{i,2} & \vdots & 0 & \cdots & w_{i,N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots \\ w_{N,1} & w_{N,2} & \vdots & w_{N,j} & \cdots & 0 \end{bmatrix}, \quad (4)$$

donde las columnas y filas corresponden a las observaciones de corte transversal y los pesos $w_{i,j}$ ($i, j = 1, 2, \dots, N$) aproximan la relación entre dos localizaciones i (filas) y j (columnas). La diagonal principal esta formada por ceros, estableciendo que ninguna observación puede estar relacionada consigo misma (no puede ser que la misma observación sea vecina de si misma). Obsérvese que se ha intercambiado α_{ij} (parámetro a estimar dentro del modelo) por w_{ij} (hipótesis externa al modelo).

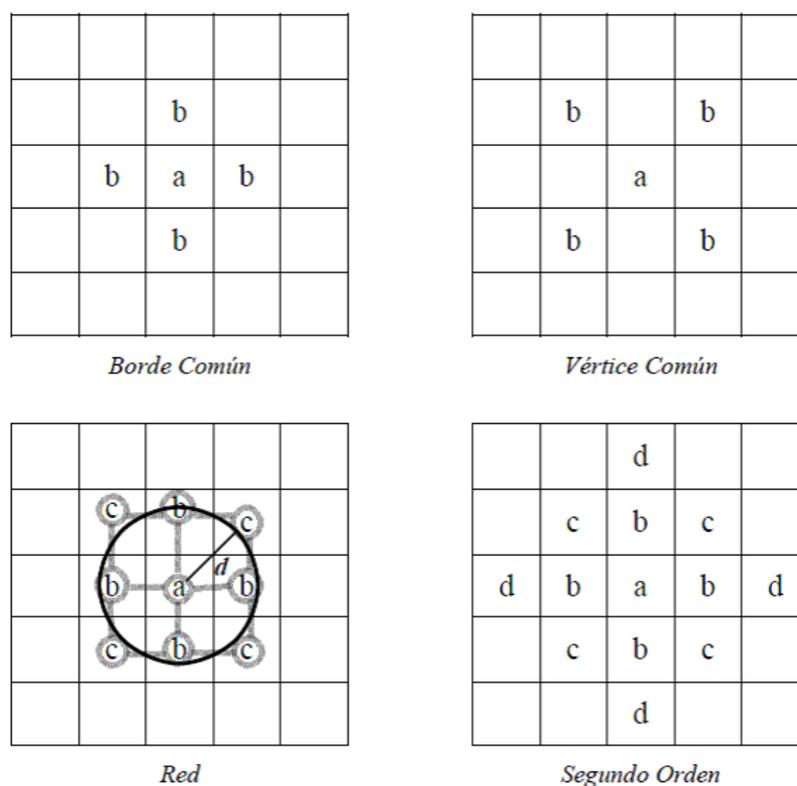
Llegado a este punto, el problema radica en cómo se construyen los pesos de la matriz W . Existen diferentes criterios sobre su construcción siguiendo, por ejemplo, alguna hipótesis de interacción. Cada hipótesis resultará en una matriz de ponderaciones diferentes llevando a un rezago espacial distinto. Es común que la representación del conjunto de vecinos obtenidos por algún criterio de selección se realice mediante una elección binaria, con $w_{i,j} = 1$ cuando i y j son vecinos, y $w_{i,j} = 0$ cuando no lo son.

En la práctica econométrica, la matriz de contactos es a menudo construida desde la geografía o geometría, usando los conceptos de contigüidad y distancia. Por ejemplo, suponiendo que las observaciones se distribuyen sobre un mapa regular como el representado en la Figura 1, hay varias alternativas para establecer el conjunto de vecinos para la celda a . Una posibilidad es considerar vecinos a aquellas celdas que poseen un borde común, siendo para nuestro ejemplo cada celda b del mapa superior izquierdo. Otra posibilidad es considerar como vecinos a aquellas celdas que poseen un vértice común, como en el mapa superior derecho. Las elecciones de estos conjuntos de vecinos son denominadas, respectivamente, *criterio tipo torre* y *criterio tipo alfil*, en analogía al movimiento de las piezas de ajedrez. De igual forma, podríamos elegir vecinos mediante una combinación de ambos criterios dando lugar al *criterio tipo reina*.

Las áreas o polígonos de un mapa pueden transformarse en puntos, y viceversa. La elección de un punto representativo de un polígono es típicamente resuelto por criterios geométricos mediante el punto central o centroide del polígono. Dado el centroide de cada celda, como en la gráfica inferior izquierda (Figura 1), es posible definir una red de puntos o nodos. En este caso, nos basaríamos en un criterio de distancia para definir a los vecinos de a tal que cada centroide que se encuentre dentro de una distancia máxima al centroide de a sería considerado vecino.

Un problema con la elección de vecinos por medio de la distancia (y contigüidad) es la existencia de puntos aislados que pueden no contener vecino alguno para un radio determinado. Esto sucede habitualmente cuando la densidad de los puntos sobre el plano no es regular o cuando se encuentran algunos nodos distribuidos por agrupamientos (*clusters*) y otros aislados. Para salvar este problema se suele determinar un radio d de amplitud tal que asegure que cada observación tienen al menos un vecino.

Figura 1: Contigüidad sobre Mapa Regular



Fuente: Anselin (1988).

Un criterio alternativo es el de k -vecinos más cercanos. En este caso, considerando la distancia geométrica entre las regiones, seleccionaríamos a los k vecinos más cercanos de cada punto. La ventaja de este criterio es que todas las unidades poseerán la misma cantidad de vecinos evitando el problema de unidades aisladas. Bajo mapas regulares, como los representados en la Figura 1, si k es igual a 4, los vecinos más cercanos a la localidad a serán los mismos que los elegidos por el *criterio tipo torre* y, si k es igual a 8, los vecinos más cercanos a la localidad a serán los mismos que los elegidos por el *criterio tipo reina*.

Por otro lado, es posible considerar varios ordenes de contigüidad o vecindad. En la Figura 1, la gráfica inferior derecha muestra el grupo de vecinos de segundo orden para la celda a . Este conjunto de vecinos, identificado por las letras c y d , incluye a los vecinos de los vecinos de a definidos por el *criterio tipo torre*. Para distinguir el orden de vecindad se añade un supra-índice a la matriz: $W^{(j)}$, $\forall j \geq 2$, siendo j el orden de vecindad.

Como alternativa a los pesos binarios, pueden considerarse aquellas funciones de distancia entre unidades espaciales que combinan la distancia, el perímetro y otras características geográficas. A continuación, Cuadro 1, presentamos algunas especificaciones utilizadas en la literatura aplicada.

Determinadas especificaciones plantean posibles problemas de endogeneidad que deberán ser considerados en el momento de la estimación del modelo econométrico. Este tipo de problema puede surgir con propuestas que utilizan variables socio-económicas, como el nivel de empleo o el producto bruto per cápita, para la elección de los pesos. Exceptuando estos casos, los procedimientos presentados en el Cuadro 1 pueden encasillarse como procedimientos exógenos, es decir, aquellos que determinan la estructura de la matriz en función, únicamente, del arreglo espacial de los datos.

Desde una perspectiva diferente, podrían seleccionarse los pesos espaciales mediante una descripción estadística de los datos. Las propuestas de este tipo suelen denominarse procedimientos endógenos. En este caso, los vecinos cercanos o lejanos son establecidos mediante la información

brindada por los propios datos mediante un estadístico que sirve de guía, como en el modelo estadístico local, *SLM*, de Getis y Aldstadt (2004), el procedimiento *AMOEB*A de Aldstadt y Getis (2006), el de modelo de ecuaciones estructurales de Folmer y Oud (2008), la técnica de filtrado espacial de Tiefelsdorf y Griffith (2007), el procedimiento de máxima entropía general de Fernandez *et al.* (2008) o el procedimiento del coeficiente de correlación completo, *CCC*, de Mur y Paelinck (2010).

Cuadro 1: Especificaciones Alternativas de Pesos Espaciales

Referencia	Modelo	Descripción
Dacey (1968)	$w_{ij} = \gamma_{ij} \alpha_i \beta_{i(j)}$	γ_{ij} : factor de contigüidad binario (1,0). α_i : proporción del área de la unidad i respecto al total de área de todas las unidades del sistema. $\beta_{i(j)}$: proporción del perímetro de la unidad i en contacto con la unidad j .
Cliff y Ord (1973)	$w_{ij} = d_{ij}^{-a} [\beta_{i(j)}]^b$	a, b : parámetros positivos d_{ij} : distancia entre los puntos o regiones (i, j) .
Bodson y Peters (1975)	$w_{ij} = \sum_{n=1}^N K_n \left\{ \frac{a}{1+b \cdot e^{-c_j d_{ij}}} \right\}$	K_n : importancia del medio de comunicación n . N : total de medios de comunicación considerados. a, b, c_j : parámetros a estimar. d_{ij} : distancia entre los puntos o regiones (i, j) .
Anselin (1980)	$w_{ij} = d_{ij}^{-2}$	d_{ij} : distancia entre los puntos o regiones (i, j) .
Cliff y Ord (1981)	$w_{ij} = [c + d_{ij}]^{-a}$	c : término constante positivo. d_{ij} : distancia entre los puntos o regiones (i, j) .
Case <i>et al.</i> (1993)	$w_{ij} = \frac{1}{ x_i - x_j }$	x : variable socioeconómica (ej.: PBI per cápita)
Molho (1995)	$w_{ij} = \frac{E_j^{-a d_{ij}}}{\sum_{k \neq i} E_k^{-a d_{ik}}}$, $\forall i \neq j$	E : volumen de empleo. d_{ij} : distancia entre los puntos o regiones (i, j) .
Ma <i>et al.</i> (1997)	$w_{ij} = e^{-d_{ij}^a}$, $w_{ij} = \left(\frac{l_{ij}}{l_i} \right)^a$, $w_{ij} = \frac{\left(\frac{l_{ij}}{l_i} \right)^a}{d_{ij}^{-b}}$	l_{ij} : longitud de frontera entre las regiones (i, j) . l_i : perímetro de la región i . d_{ij} : distancia entre los puntos o regiones (i, j) .
Toral (2000)	$w_{ij} = \delta_{ij} \frac{k_i k_j p_i p_j}{d_{ij}^a}$	$\delta_{ij} = 1$, si las unidades espaciales i, j tienen una frontera en común y cero, si no la tienen. p : población. k : longitud de carreteras (km). d_{ij} : distancia por carretera entre las capitales i, j . a : parámetro positivo, con valores 0, 1 ó 2.

Fuente: Chasco Yrigoyen (2003).

Una vez elegidos los pesos espaciales, es raro que se utilicen en su forma natural. Lo habitual es trabajar con alguna transformación de la matriz en el modelo espacial ya que mejora las propiedades estadísticas de los estimadores y sus estadísticos. La transformación más utilizada es la normalización por fila, en donde los nuevos pesos son obtenidos como $w_{ij}^* = w_{ij} / \sum_j w_{ij}$, de tal forma que la suma de cada fila de la matriz sea igual a la unidad: $\sum_j w_{ij}^* = 1$ (representaremos como W^* a la matriz de contactos normalizada).

2.2. Detección Global de Dependencia Espacial Univariante

Una primera pregunta de análisis en los estudios empíricos suele ser: ¿Presenta la variable de estudio autocorrelación espacial?. Si la respuesta es afirmativa, entonces deberemos formular modelos econométricos más complejos donde se incorpore explícitamente el efecto del espacio.

La cuestión, entonces, pasa por conocer aquellas herramientas que nos permiten detectar la presencia de autocorrelación espacial. Desde una perspectiva descriptiva, una primera aproximación cualitativa puede realizarse mediante el análisis exploratorio espacial (Haining, 2003).

El análisis exploratorio espacial permite la visualización mediante diferentes gráficos del comportamiento de la variable bajo estudio, siendo el mapa un elemento central. Existe una amplia variedad de mapas y formas para describir datos continuos sobre polígonos irregulares⁴ ("lattice data"), una de las más usuales es el mapa de coropletas. En este tipo de mapa se representa la distribución espacial de una variable o atributo mediante diferentes tonalidades. El número de tonalidades representa a los diferentes intervalos y los mismos pueden ser definidos por el usuario. La realización de este análisis puede llevarse a cabo mediante el programa estadístico OpenGeoDa, posiblemente la opción existente más integrada en cuanto a análisis exploratorio espacial.

El análisis exploratorio espacial nos brindará los primeros indicios de dependencia aunque necesitaremos herramientas inferenciales para poder obtener una cuantificación de la misma. La formulación de estas herramientas varía a través de diferentes estadísticos, aunque cada uno de ellos puede ser visto como un caso especial del test de Mantel (Mantel, 1967).

El test de Mantel estima la correlación existente entre dos matrices B y C donde dichas matrices capturan el grado de similitud en el espacio y en la distribución de la variable, respectivamente. Dicho estadístico se define como:

$$\Gamma = \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N b_{ij} c_{ij}, \quad (5)$$

donde b_{ij} es un elemento de la matriz B que captura la similitud espacial de las localizaciones i y j , y c_{ij} es un elemento de la matriz C que captura la similitud de los valores de una determinada variable entre las localizaciones i y j .

Reformulando del estadístico Γ , tal que los elementos b_{ij} sean los pesos espaciales de la matriz de contactos W y los elementos c_{ij} representen el cuadrado de la diferencia de valores de la variable de interés entre localizaciones, puede obtenerse el estadístico C de Geary (1954):

$$C = \frac{N-1}{2S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ji} (y_i - y_j)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (6)$$

donde $S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij} = \mathbf{1}'W\mathbf{1}$, siendo $\mathbf{1}$ un vector $(N \times 1)$ de unos y \bar{y} es la media muestral.

Los momentos del estadístico de Geary, bajo la hipótesis nula de aleatoriedad, son:

$$E[C] = 1, \quad (7)$$

$$V[C] = \frac{(2S_1 + S_2)(N-1) - 4S_0^2}{2(N+1)S_0^2}, \quad (8)$$

donde $S_1 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (w_{ij} + w_{ji})^2$ y $S_2 = \sum_{j=1}^N (w_{j.} + w_{.j})^2$, siendo $w_{j.} = \sum_{i=1}^N w_{ij}$.

⁴Existen diferentes tipos de datos georeferenciados. En este trabajo nos enfocamos en el manejo de variables continuas distribuidas en un espacio discreto o *lattice data*. Véase Cressie (1993) para una exposición de los diferentes tipos de datos.

Su distribución asintótica es normal:

$$\sqrt{N} [C - 1] \underset{as}{\sim} \mathcal{N} [0; \mathbf{V} (C)]. \quad (9)$$

Posiblemente el test más ampliamente utilizado, y que también puede derivarse del test de Mantel, es el estadístico I de Moran (1950):

$$I = \frac{N \sum_i \sum_j (y_i - \bar{y}) w_{ji} (y_j - \bar{y})}{S_0 \sum_{i=1}^N (y_i - \bar{y})^2}. \quad (10)$$

Los momentos del estadístico de Moran, bajo la hipótesis nula de aleatoriedad, son:

$$E [I] = -\frac{1}{N-1}, \quad (11)$$

$$\mathbf{V} [I] = \frac{(3S_0^2 + S_1 R^2 - N S_2)}{S_0 (N+1) (N-1)} - \frac{1}{(N-1)^2}. \quad (12)$$

Su distribución asintótica, nuevamente, es normal:

$$\sqrt{N} [I - E (I)] \underset{as}{\sim} \mathcal{N} [0; \mathbf{V} (I)]. \quad (13)$$

Diferentes experimentos Monte Carlo (Anselin y Florax, 1995) revelan que la aproximación a la distribución normal funciona razonablemente bien con tamaños muestrales medios ($N > 50$). El comportamiento del estadístico empeora sensiblemente cuando reducimos el tamaño muestral o se añaden problemas nuevos como heterocedasticidad, atípicos, etc., lo que significa que los resultados deben interpretarse con cautela.

Como nota final, tanto el test C de Geary como el test I de Moran capturan la dependencia global, es decir, la autocorrelación existente en todo el espacio geográfico. Existen otros estadísticos que permiten capturar la dependencia local utilizando sub-muestras del espacio geográfico. Entre los más utilizados se destacan el análisis *LISA* de Anselin (1995) y los tests $G_i (d)$ y $G_i^* (d)$ de Getis y Ord (1996) y Ord y Getis (2001). El análisis de estos estadísticos se encuentra fuera de los objetivos del presente trabajo (para mayor detalle, véase las citadas referencias).

El rechazo de la hipótesis nula de aleatoriedad de alguno de estos tests de autocorrelación espacial implica un análisis más detallado sobre los determinantes de la variable bajo estudio. Detectada la importancia del espacio sobre la variable, ahora estamos en condiciones de introducir los diferentes modelos econométricos que permiten la incorporación de los efectos espaciales.

2.3. Modelos Econométricos y Tests de Independencia Espacial

En esta sección desarrollaremos la metodología más habitual en el campo aplicado. La misma consiste en el planteo inicial de un modelo estático simple y a partir de allí se incorporan elementos espaciales según las diferentes pruebas estadísticas.

El modelo estático más simple considera que existe una variable dependiente y un conjunto de variables explicativas. La ecuación de trabajo es la siguiente:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\beta + \varepsilon, \\ \varepsilon &\sim \mathcal{N} (0, \sigma^2 \mathbf{I}_N), \end{aligned} \quad (14)$$

siendo la variable dependiente \mathbf{y} un vector de dimensión $(N \times 1)$, \mathbf{X} es una matriz de variables explicativas, incluyendo una constante, de orden $(N \times k)$, β es un vector de parámetros desconocidos de orden $(k \times 1)$ y ε es el término de error de dimensión $(N \times 1)$.

Comenzar con el modelo (14) tiene su lógica en el sentido de que la dependencia espacial univariante, detectada en la sección 2.2, puede ser explicada por el conjunto de variables independientes X . Si este fuese el caso, la sola incorporación de estas variables explicativas evitaría introducir modelos más complejos.

Por otra parte, utilizaremos como modelo de referencia, que incorpora la dependencia espacial residual, el Modelo de Error Espacial (Spatial Error Model):

$$\begin{aligned} \mathbf{y} &= X\beta + \mathbf{u}, \\ \mathbf{u} &= \theta W\mathbf{u} + \varepsilon \Rightarrow \mathbf{u} = B^{-1}\varepsilon, \\ \varepsilon &\sim \mathcal{N}(0, \sigma^2 I_N), \end{aligned} \quad (15)$$

siendo $B = [I - \theta W]$.

Si el modelo (15) se estima por mínimos cuadrados ordinarios, *MCO*, las consecuencias serán $\hat{\beta}$ insesgado y consistente, con una matriz de covarianzas diferente a la habitual:

$$\hat{\beta} = [X'X]^{-1} X'\mathbf{y} = \beta + [X'X]^{-1} X'\mathbf{u}, \quad (16)$$

$$E[\hat{\beta}] = \beta, \quad (17)$$

$$V[\hat{\beta}] = \sigma^2 [X'X]^{-1} X' (B'B)^{-1} X [X'X]^{-1}. \quad (18)$$

Además, el estimador *MCO* del parámetro de dispersión será sesgado pero consistente:

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N-k} = \frac{\mathbf{u}'M\mathbf{u}}{N-k} = \frac{\varepsilon' B^{-1} M B^{-1} \varepsilon}{N-k}, \quad (19)$$

$$E[\hat{\sigma}^2] = \sigma^2 \frac{\text{tr}(B^{-1} M B^{-1})}{N-k}, \quad (20)$$

$$= \frac{\sigma^2}{N-k} \text{tr}\left(B^{-2} - (X'X)^{-1} X' B^{-2} X\right) \neq \sigma^2, \quad (21)$$

$$\text{plim}[\hat{\sigma}^2] = \sigma^2, \quad (22)$$

siendo $M = I - X(X'X)^{-1}X'$.

Estas consecuencias son las esperadas cuando se estima por *MCO* un modelo lineal con matriz de covarianzas no escalar en el término de error.

Por lo tanto, tal como sucede en la práctica econométrica clásica, podemos implementar la secuencia habitual de estimar el modelo estático (14) por *MCO* y, a continuación, analizar los posibles problemas de especificación.

En concreto, analizaremos la presencia de autocorrelación espacial a través del examen residual de la estimación *MCO*. Planteado de esta forma, podemos utilizar los tests de dependencia global adaptados al marco de regresión lineal general.

En este caso, el test I de Moran ha tenido un rol preponderante en esta etapa. Aplicado sobre los residuos *MCO* del modelo (14), la expresión del estadístico es:

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N \hat{u}_i w_{ji} \hat{u}_j}{\sum_{i=1}^N \hat{u}_i^2}. \quad (23)$$

Las expresiones de los momentos del estadístico, bajo la hipótesis nula de aleatoriedad espacial, se modifican de la siguiente forma:

$$E[I] = \frac{N \operatorname{tr}(MW)}{S_0(N-k)} = -\frac{N \operatorname{tr}\left[\left(X'X\right)^{-1}X'WX\right]}{S_0(N-k)}, \quad (24)$$

$$E[I^2] = \frac{\left(\frac{N}{S_0}\right)^2 \operatorname{tr}(MWMW') + \operatorname{tr}(MWMW) + [\operatorname{tr}(MW)]^2}{(N-1)(N-k+2)}. \quad (25)$$

La varianza puede obtenerse como: $V[I] = E[I^2] - E[I]^2$.

La distribución probabilística del estadístico I de Moran es desconocida para muestras finitas, por lo que comúnmente se utiliza una aproximación empírica por permutación. Otra opción es plantear la misma aproximación asintótica del caso univariante:

$$\sqrt{N}[I - E(I)] \underset{as}{\sim} \mathcal{N}[0; V(I)]. \quad (26)$$

El problema con el test I de Moran es que el rechazo de la hipótesis nula no brinda información adicional sobre el modelo bajo la hipótesis alternativa.

Otra alternativa para la detección de dependencia espacial residual son los estadísticos que resultan de la aplicación del principio de máxima verosimilitud. En este caso, podemos utilizar el Mutiplicador de Lagrange que para las hipótesis nula y alternativa:

$$\begin{aligned} H_0 &: \theta = 0, \\ H_1 &: \theta \neq 0, \end{aligned}$$

posee la siguiente estructura:

$$LM_{ERR} = \frac{1}{2S_0} \left(\frac{\hat{\mathbf{u}}'W\hat{\mathbf{u}}}{\hat{\sigma}^2} \right)^2 \underset{as}{\sim} \chi^2_{(1)}, \quad (27)$$

donde $\hat{\mathbf{u}}$ son los residuos MCO del modelo bajo la hipótesis nula. El resultado es simple y puede generalizarse a otros tipos de casos más complejos en la estructura del error.

Otro modelo de referencia que puede plantearse es el Modelo de Rezago Espacial (Spatial Lag Model). Este modelo incorpora dependencia sustantiva en el sentido de que la variable endógena depende de su propio rezago espacial, entre otros elementos. La estructura de dependencias generada por este tipo de procesos es mucho más intensa que la contemplada en el caso de dependencia residual.

El Modelo de Rezago Espacial tiene la siguiente estructura:

$$\begin{aligned} \mathbf{y} &= \rho W\mathbf{y} + X\beta + \mathbf{u}, \\ \mathbf{u} &\sim \mathcal{N}(0, \sigma^2 I_N), \end{aligned} \quad (28)$$

La implicaciones econométricas derivadas de la omisión de elementos dinámicos, como $W\mathbf{y}$, son muy significativas. Suponiendo que el modelo (28) es el correcto pero se ha estimado el modelo (14), puede comprobarse que los estimadores de los parámetros de posición MCO son sesgados:

$$\begin{aligned} \hat{\beta} &= [X'X]^{-1} X'\mathbf{y} = [X'X]^{-1} X'(B^{-1}X\beta + B^{-1}\mathbf{u}), \\ &= [X'X]^{-1} X'B^{-1}X\beta + [X'X]^{-1} X'B^{-1}\mathbf{u}, \\ E[\hat{\beta}] &= [X'X]^{-1} X'B^{-1}X\beta \neq \beta, \end{aligned} \quad (29)$$

Los residuos, la estimación del parámetro de dispersión y su esperanza tienen la siguiente forma:

$$\hat{\mathbf{u}} = \mathbf{y} - X\hat{\beta} = M^{-1}B^{-1}(X\beta + \mathbf{u}), \quad (30)$$

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N - k} = \frac{(X\beta + \mathbf{u})' B^{-1} M B^{-1} (X\beta + \mathbf{u})}{N - k}, \quad (31)$$

$$E[\hat{\sigma}^2] = \frac{\beta' X' B^{-1} M B^{-1} X \beta + \sigma^2 \text{tr}(B^{-1} M B^{-1})}{N - k}. \quad (32)$$

Finalmente, el sesgo no desaparece en un contexto asintótico:

$$\text{plim}[\hat{\beta}] = \Sigma_{X'X}^{-1} \Sigma_{XBX} \beta \neq \beta, \quad (33)$$

$$\text{plim}[\hat{\sigma}^2] \cong \sigma^2 \text{plim} \frac{\text{tr}(B^{-2})}{N}, \quad (34)$$

siendo $\Sigma_{X'X}^{-1} = \lim \left[\frac{X'X}{N} \right]^{-1}$, $\Sigma_{XBX} = \lim \left[\frac{X' B^{-1} X}{N} \right]$.

La severidad de estos resultados aconseja chequear la posible omisión de elementos dinámicos en el modelo. El contexto más detallado para afrontar esta cuestión se corresponde con el planteamiento de máxima-verosimilitud, que desarrollaremos a continuación.

Partiendo de la habitual función de log-verosimilitud, correspondiente en este caso al modelo (28):

$$l(\mathbf{y}|\varphi) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \left[\frac{(\mathbf{B}\mathbf{y} - X\beta)' (\mathbf{B}\mathbf{y} - X\beta)}{2\sigma^2} \right] + \ln |B|, \quad (35)$$

donde $\ln |B| = \ln |I - \rho B| = \sum_{i=1}^N \ln(1 - \rho \lambda_i)$.

El sistema de condiciones necesarias es no lineal, imposibilitando la obtención de una solución analítica:

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2} [X' (\mathbf{B}\mathbf{y} - X\beta)] = \frac{1}{\sigma^2} X' \mathbf{u}, \quad (36)$$

$$\begin{aligned} \frac{\partial l}{\partial \rho} &= \frac{1}{\sigma^2} [\mathbf{y}' W (\mathbf{B}\mathbf{y} - X\beta)] - \sum_{i=1}^N \frac{\lambda_i}{1 - \rho \lambda_i}, \\ &= \frac{1}{\sigma^2} [\mathbf{y}' W \mathbf{u}] - \sum_{i=1}^N \frac{\lambda_i}{1 - \rho \lambda_i}, \end{aligned} \quad (37)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{(\mathbf{B}\mathbf{y} - X\beta)' (\mathbf{y} - X\beta)}{2\sigma^4} = -\frac{N}{2\sigma^2} + \frac{\mathbf{u}' \mathbf{u}}{2\sigma^4}. \quad (38)$$

Para poder solucionar este sistema debe recurrirse a algoritmos numéricos o a procedimientos basados en la log-verosimilitud concentrada. En este último caso resulta relativamente sencillo obtener los estimadores máximo-verosímiles, MV , del vector β y del parámetro de dispersión σ^2 , condicionados ambos al parámetro de dependencia ρ :

$$\begin{aligned} \tilde{\beta} &= [X'X]^{-1} X' B \mathbf{y}, \\ &= [X'X]^{-1} X' \mathbf{y} - \rho [X'X]^{-1} X' \mathbf{y} = \hat{\beta} - \rho \hat{\beta}_{W\mathbf{y}}, \end{aligned} \quad (39)$$

$$\hat{\beta} = [X'X]^{-1} X' \mathbf{y}, \quad \hat{\beta}_{W\mathbf{y}} = [X'X]^{-1} X' W \mathbf{y}, \quad (40)$$

$$\tilde{\sigma}^2 = \frac{\tilde{\mathbf{u}}' \tilde{\mathbf{u}}}{N} = \frac{[\hat{\mathbf{u}}_{\mathbf{y}} - \rho \hat{\mathbf{u}}_{W\mathbf{y}}]' [\hat{\mathbf{u}}_{\mathbf{y}} - \rho \hat{\mathbf{u}}_{W\mathbf{y}}]}{N}, \quad (41)$$

$$\hat{\mathbf{u}} = \mathbf{y} - X\hat{\beta}, \quad \hat{\mathbf{u}}_{W\mathbf{y}} = W\mathbf{y} - X\hat{\beta}_{W\mathbf{y}}. \quad (42)$$

Sustituyendo estos resultados en (35) se obtiene la función de log-verosimilitud concentrada que depende solo del parámetro ρ . A continuación se optimiza esta función mediante iteraciones hasta alcanzar convergencia.

Con el objetivo de contrastar la existencia de elementos dinámicos en la ecuación principal del modelo (28), bajo la hipótesis nula y alternativa:

$$\begin{aligned} H_0 &: \rho = 0, \\ H_1 &: \rho \neq 0, \end{aligned}$$

el Multiplicador de Lagrange es igual a:

$$LM_{LAG} = \frac{\left(\frac{\mathbf{y}'W\hat{\mathbf{u}}}{\hat{\sigma}^2}\right)^2}{V[\hat{\rho}_N]} = \frac{\left(\frac{\mathbf{y}'W\hat{\mathbf{u}}}{\hat{\sigma}^2}\right)^2}{\left(\frac{\hat{\beta}'X'WMWX\hat{\beta}}{\hat{\sigma}^2}\right)} \underset{as}{\sim} \chi^2_{(1)}. \quad (43)$$

Es interesante mencionar que los casos de autocorrelación residual pueden entenderse como casos particulares de dependencia sustantiva, lo cual se hace evidente en el test denominado *COMFAC* (de factores comunes). Supongamos que nuestro modelo de partida es un modelo con autocorrelación residual tal como:

$$\begin{aligned} \mathbf{y} &= X\beta + \mathbf{u}, \\ \mathbf{u} &= \rho W\mathbf{u} + \varepsilon \\ \varepsilon &\sim \mathcal{N}(0, \sigma^2 I_N), \end{aligned} \quad (44)$$

Manipulando el término de error tal que: $\mathbf{u} = \rho W(\mathbf{y} - X\beta) + \varepsilon$, y reintroduciendo este resultado en (44), obtenemos el Modelo Espacial de Durbin:

$$\mathbf{y} = \rho W\mathbf{y} + X\beta + WX\gamma + \varepsilon. \quad (45)$$

La ecuación final (45) se corresponde con un modelo dinámico (aparece un rezago de la endógena como variable explicativa, $W\mathbf{y}$), en el que se ha añadido un rezago espacial de las exógenas WX , siendo γ el vector de parámetros que acompaña a estos rezagos. Dado que la perturbación del modelo, ε , cumple con los supuestos básicos, la ecuación puede estimarse por máxima verosimilitud. El test de factores comunes se concreta en:

$$\begin{aligned} H_0 &: \gamma + \rho\beta = 0, \\ H_1 &: \gamma + \rho\beta \neq 0, \end{aligned}$$

el cual puede resolverse fácilmente aplicando la razón de verosimilitudes, *LR*. En este caso, el modelo amplio es el de la expresión final de (45), el cual se estimará por *MV* sin restricciones. El modelo restringido, bajo hipótesis nula, es el modelo (15)⁵. Este mismo modelo también se estimará por *MV* para poder obtener finalmente el estadístico *LR*:

$$LR_{COMFAC} = 2 \left[l_{|H_1} - l_{|H_0} \right] \underset{as}{\sim} \chi^2_q \quad (46)$$

⁵Bajo H_0 , $\gamma = -\rho\beta$, entonces reemplazando en el modelo (45):

$$\mathbf{y} = \rho W\mathbf{y} + X\beta + WX(-\rho\beta) + \varepsilon \Rightarrow \mathbf{y} = \rho W\mathbf{y} + X\beta - \rho WX\beta + \varepsilon \Rightarrow (I - \rho W)\mathbf{y} = (I - \rho W)X\beta + \varepsilon.$$

La última expresión puede resumirse en el Modelo de Error Espacial:

$$\mathbf{y} = X\beta + (I - \theta W)^{-1} \varepsilon,$$

donde $\rho = \theta$.

siendo $l_{|H_1}$ la log-verosimilitud obtenida en la estimación del modelo amplio y $l_{|H_0}$ la correspondiente al modelo de la hipótesis nula; q es el número de restricciones e igual al número de parámetros incluidos en β , sin considerar la constante.

Otro punto de contacto entre ambos tipos de dependencia se produce en el test denominado *SARMA*. Para obtener este último, debe plantearse explícitamente un modelo amplio que incluya dependencia sustantiva y residual. Esta especificación se denomina Modelo de Rezago y Error Espacial (*SARAR*(1, 1), Anselin y Florax, 1995):

$$\mathbf{y} = \rho W\mathbf{y} + X\beta + \mathbf{u} \quad (47)$$

$$\mathbf{u} = \theta W\mathbf{u} + \varepsilon \quad (48)$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$$

En el test *SARMA* se analiza la no existencia de efectos espaciales en el modelo, combinando información de los estadísticos base LM_{ERR} y LM_{LAG} :

$$H_0 : \rho = 0; \theta = 0,$$

$$H_1 : \rho \neq 0; \theta \neq 0,$$

El estadístico Multiplicador de Lagrange, bajo hipótesis nula, es igual a:

$$LM_{SARMA} = \frac{\left[\left(\frac{\mathbf{y}' W \hat{\mathbf{u}}}{\hat{\sigma}^2} \right) - \left(\frac{\hat{\mathbf{u}}' W \hat{\mathbf{u}}}{\hat{\sigma}^2} \right) \right]^2}{V[\tilde{\rho}_N]} + \frac{1}{2S_0} \left(\frac{\hat{\mathbf{u}}' W \hat{\mathbf{u}}}{\hat{\sigma}^2} \right)^2 \underset{as}{\sim} \chi_{(2)}^2. \quad (49)$$

Los diferentes Multiplicadores de Lagrange presentados hasta ahora cubren los principales aspectos de la especificación de un modelo econométrico de corte transversal. El inconveniente que presentan es que son altamente sensibles a diferentes tipos de errores de especificación. Por ejemplo, el LM_{ERR} no debería reaccionar si el error consiste en que se ha omitido un rezago de la variable endógena en la parte derecha de la ecuación, y el LM_{LAG} no debería reaccionar cuando el error consiste en la omisión de un rezago en la perturbación. Sin embargo, esto no es lo habitual y ambos estadísticos mostrarán valores significativos en respuesta a un error en la especificación, ya sea de tipo sustantiva o residual.

En este contexto, Anselin *et al.* (1996) proponen dos nuevos Multiplicadores de Lagrange diseñados para que su comportamiento sea robusto a diferentes errores de especificación. El LM_{EL} analiza la falta de correlación en los residuos, siendo robusto a errores de especificación en la estructura dinámica de la ecuación principal. El modelo de referencia, la hipótesis nula y el test estadístico son los siguientes:

$$\mathbf{y} = X\beta + \mathbf{u}, \quad (50)$$

$$\mathbf{u} = \theta W\mathbf{u} + \varepsilon,$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N),$$

$$H_0 : \theta = 0,$$

$$H_1 : \theta \neq 0,$$

$$LM_{EL} = \frac{\left[\left(\frac{\hat{\mathbf{u}}' W \hat{\mathbf{u}}}{\hat{\sigma}^2} \right) - \left(\frac{2S_0}{2S_0 + V[\tilde{\rho}_N]} \right) \left(\frac{\mathbf{y}' W \hat{\mathbf{u}}}{\hat{\sigma}^2} \right) \right]^2}{\left(2S_0 - \frac{(2S_0)^2}{2S_0 + V[\tilde{\rho}_N]} \right)} \underset{as}{\sim} \chi_{(1)}^2. \quad (51)$$

Por otra parte, el test LM_{LE} permite detectar estructura dinámica de la ecuación, siendo robusto a estructuras espaciales en el término de error. El modelo de referencia, la hipótesis nula y el test

estadístico son los siguientes:

$$\mathbf{y} = \rho W\mathbf{y} + X\beta + \mathbf{u}, \quad (52)$$

$$\mathbf{u} \sim \mathcal{N}(0; \sigma^2 I_N),$$

$$H_0 : \rho = 0,$$

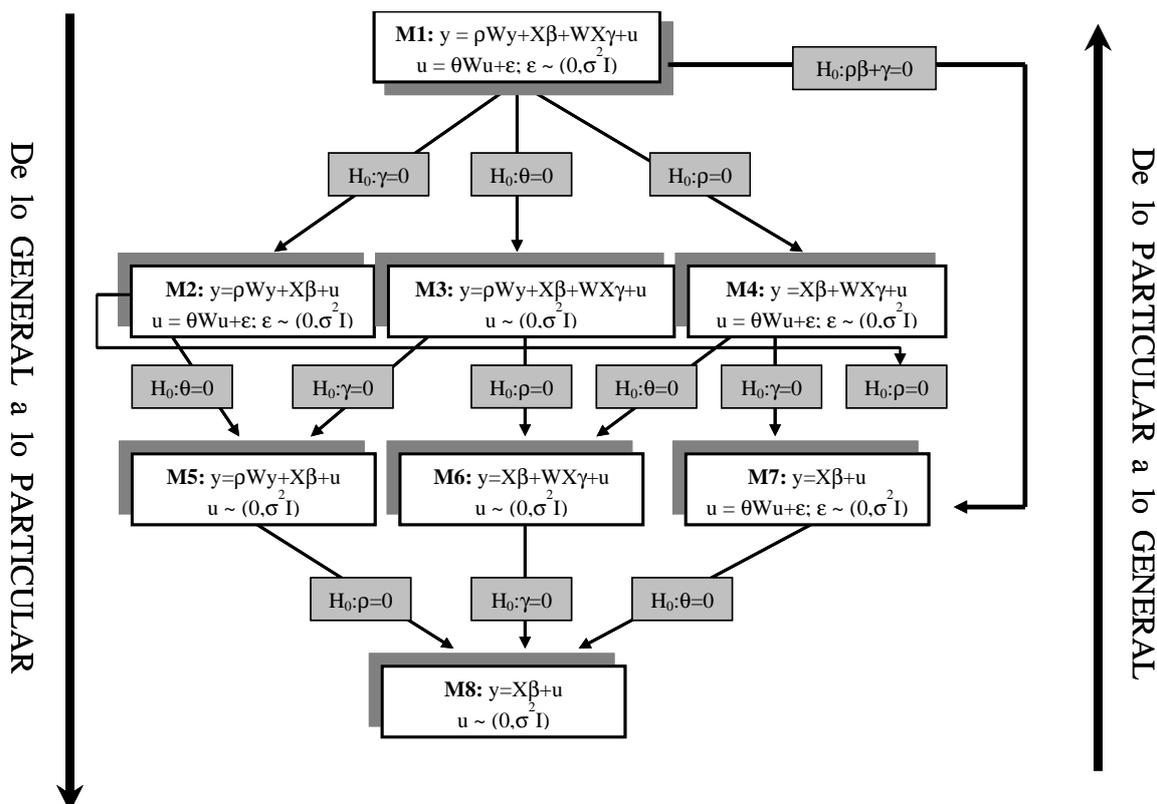
$$H_1 : \rho \neq 0,$$

$$LM_{LE} = \frac{\left[\left(\frac{\tilde{\mathbf{u}}' W \tilde{\mathbf{u}}}{\tilde{\sigma}^2} \right) - \left(\frac{\tilde{\mathbf{u}}' W \tilde{\mathbf{u}}}{\tilde{\sigma}^2} \right) \right]^2}{V[\tilde{\rho}_N]} \underset{as}{\sim} \chi_{(1)}^2. \quad (53)$$

Para un mayor detalle de los diferentes estadísticos de dependencia espacial en modelos lineales véase las excelentes revisiones provistas por Anselin y Bera (1998), Anselin (2001) y Florax y de Graaff (2004).

Hasta el momento, se ha bosquejado la estrategia de especificación usual en econometría espacial. Este tipo de estrategia es denominada “de lo particular a lo general”. Sin embargo, algunos modelos espaciales no han sido considerados y es posible plantearlos utilizando una estrategia de especificación alternativa que es “de lo general a lo particular”. Ambos tipo de estrategias se presentan en la Figura 2.

Figura 2: Estrategias Alternativas de Especificación Espacial



Fuente: Basado en Mur y Angulo (2009).

Con respecto a la estructura de la Figura 2, solo haremos algunos comentarios generales. En los modelos en que aparece más de una matriz de contactos, es posible especificar diferentes en las

estructuras de dependencia en la ecuación principal y en la estructura del error. Lo natural en los trabajos aplicados ha sido mantener la misma hipótesis de interacción para todas las variables del modelo.

La especificación del término de error ha sido planteada como autoregresiva espacial, $SAR(1)$, $\mathbf{u} = \theta W\mathbf{u} + \varepsilon$, pero pueden introducirse más rezagos de orden p , $SAR(p)$: $\mathbf{u} = \theta_1 W\mathbf{u} + \theta_2 W^2\mathbf{u} + \dots + \theta_p W^p\mathbf{u} + \varepsilon$. Otra estructura habitual es la de media móvil, $SMA(1)$: $\mathbf{u} = \varepsilon + \lambda W\varepsilon$ o una generalización de la misma, $SMA(q)$: $\mathbf{u} = \varepsilon + \lambda_1 W\varepsilon + \lambda_2 W^2\varepsilon + \dots + \lambda_q W^q\varepsilon$. Las estructuras autoregresivas, $SAR(1)$, y de media móvil, $SMA(1)$, son las más utilizadas habitualmente.

En general, se supone inicialmente normalidad, lo cual simplifica la inferencia estadística y la estimación de los diferentes modelos. En este trabajo hemos desarrollado toda la taxonomía considerando normalidad. En la actualidad hay vías alternativas como la estimación GMM que no necesita dicho supuesto.

El modelo $M1$ se corresponde en términos econométricos con el modelo amplio, denominado Modelo de Cliff y Ord. Un argumento negativo para comenzar el análisis desde $M1$ es la complejidad de las relaciones que contempla. Desde otra perspectiva, el enfoque “de lo general a lo particular” ha mostrado un mejor comportamiento en la selección del modelo cuando existe heterocedasticidad y valores atípicos (Mur y Angulo, 2009).

3. Ilustración: Análisis de la Fecundidad en la Argentina⁶

Dado el objetivo del documento, el tratamiento de la fecundidad desarrollado en esta sección sirve como ilustración, no pretendiendo ser el tema principal de análisis. Se reconoce que el estudio de la fecundidad es muy complejo y requiere una discusión teórica más detallada así como la incorporación de un número muy amplio de variables. En lo que sigue hemos destacado los principales aspectos teóricos mencionados en la literatura y se han considerado las variables explicativas más pertinentes y que se encuentran disponibles para la Argentina.

El comportamiento moderno de la fecundidad femenina posee un patrón claramente distintivo. La fecundidad ha tendido a declinar, primero a través de las naciones de alto ingreso en las últimas décadas del siglo XIX, y posteriormente a través de la mayoría de las naciones de ingreso medio en el siglo XX (Kuznets, 1966).

A nivel urbano-regional, Sharlin (1986) destaca que, en una primera etapa, se observa una declinación de la fecundidad en las ciudades y posteriormente esa declinación se expande hacia las zonas rurales y aldeñas.

Existen distintas teorías que explican la transición de la fecundidad. Desde el punto de vista económico, diferentes modelos de asignación del tiempo entre las actividades de consumo y producción permiten plantear diversas hipótesis que vinculan cambios en el entorno económico en donde la gente vive y sus demandas reproductivas y de fecundidad (Becker, 1960). Por otra parte, una propuesta alternativa al enfoque económico se relaciona al *cambio de ideas* y la influencia de las interacciones sociales (Cleland y Wilson, 1987). A lo largo del tiempo, los cambios de ideas en las zonas urbanas sobre la demanda de niños sirven como fundamento para la difusión hacia el exterior de las ciudades. Simultáneamente, las mejoras en las comunicaciones y el transporte pueden generar que las personas de zonas centrales y periféricas interactúen rápidamente generando una difusión más acentuada de las ideas sobre fecundidad.

La distinción entre las diferentes teorías y el rol que juegan los grupos externos a la familia y sus interacciones ha postulado estrategias de teorías de juego cooperativas y no cooperativas dentro de las familias. Sin embargo, la complejidad de los factores que probablemente afecten la formación de grupos genera que empíricamente determinados efectos no sean identificables (Manski, 1995).

Siguiendo la teoría de Cleland y Wilson (1987), las ciudades (vistas como regiones centrales dominantes) son el inicio de esta declinación mediante el desarrollo de nuevas ideas de organización

⁶El análisis estadístico de esta sección puede replicarse en su totalidad mediante la secuencia de comandos en R que se encuentra en el apéndice.

social y eventualmente las zonas periféricas las siguen. En términos económicos, la mayor densidad y estructura económica no-agrícola de los lugares urbanos son un entorno propicio para reducir la demanda de niños. La menor cantidad de niños genera, en particular sobre la mujer, mayores oportunidades para desarrollar su nivel educativo y laboral. Este comportamiento produce que la participación femenina sea más acentuada en la fuerza laboral, volviéndose financieramente más independiente y generando nuevos incentivos para limitar los niveles de reproducción (Weeks *et al.*, 2004).

La universalidad del comportamiento de la fecundidad puede ser matizada por diferencias culturales que pueden producir distintos niveles de fecundidad entre zonas y dentro de las mismas. Los estudios a nivel regional, en su vasta mayoría, examinan la fecundidad por regiones basándose en simples dicotomías urbano-rural, como si el espacio fuese uniforme dentro de estos dos grupos. La estrategia habitual es la utilización de variables ficticias que capturan la pertenencia de las observaciones a cada una de las áreas. Sin embargo, es posible que existan importantes variaciones en la fecundidad entre las diferentes áreas. Este punto es el inicio de nuestra investigación empírica, intentando brindar evidencia de la falta de homogeneidad en el espacio de la variable fecundidad para el caso de la Argentina.

Nuestro argumento es que los patrones espaciales son importantes debido a que ofrecen indicios sobre las principales causas y potenciales consecuencias del comportamiento. Como menciona Weeks (2003), existe una escasa literatura que presta atención a las causas y consecuencias sociales de la tendencia de la fecundidad a nivel local.

Puede argumentarse que los patrones espaciales son obvios, pero esta hipótesis ha sido pobremente contrastada para el caso argentino. Desde nuestra opinión, no es que esta hipótesis no sea relevante o que sea un resultado trivial, más bien puede deberse a la limitada difusión del herramental desarrollado en campos científicos como la estadística espacial o la econometría espacial que basan su análisis en los sistemas de información geográfica.

3.1. Hipótesis de Trabajo y Datos

Nuestra hipótesis de partida es que los niveles de fecundidad son determinados sólo parcialmente por la clase social de pertenencia, medida por el capital humano. Si bien los microdeterminantes constituyen el eje de la teoría económica relacionada con este tema (Bryan y Zick, 2006), planteamos aquí que debe considerarse además un componente de fecundidad que puede ser explicado por la localización geográfica. Esto que es lo que en la literatura se denomina *efecto contextual* o *conducta demográfica* (Weeks, 2003). Debido a que el ser humano es un ser inherentemente social, se encuentra influenciado por el contacto con quienes interactúa. En otros términos, la hipótesis puede representarse de la siguiente manera:

$$\text{Fecundidad} = f[\text{clase social, contexto geográfico}], \quad (54)$$

es decir, la fecundidad es función de factores diferentes: uno de pertenencia social y otro de pertenencia geográfica o contextual.

Usando información agregada por departamentos del Censo 2001, nuestras variables son las siguientes:

(1). *HIJOS*: variable que captura la fecundidad como el número promedio de hijos nacidos vivos por mujer. El rango de edad relevante de la mujer ha sido definido entre 15 y 29 años.

(2). *CONVIV*: Porcentaje de mujeres que conviven con hombres (casadas o en pareja), mujeres de 15 a 29 años.

(3). *EDUC*: Promedio de años de educación formal, mujeres de 15 a 29 años.

(4). *URBAN*: Porcentaje de población urbana dentro del total de mujeres de 15 a 29 del respectivo departamento.

(5). *ACTIVA*: Tasa de actividad porcentual definida como la relación entre la población económicamente activa y la población total, mujeres de 15 a 29 años.

(6). *HNOPOB*: Porcentaje de hogares no pobres según criterio *IPMH* (Indicador de Privación Material de los Hogares) considerando el grupo (1) de dicho índice.

En el Cuadro 2 se presentan los estadísticos descriptivos. De las variables utilizadas, es necesario aclarar la construcción de *HNOPOB*. La misma utiliza el índice de Privación Material de los Hogares que identifica a los hogares según su situación respecto a la privación material en cuanto a dos dimensiones: recursos corrientes y patrimonial, elaborado por el *INDEC*. La primera dimensión se mide con un índice de condiciones habitacionales, que considera si en la vivienda existían materiales precarios. La dimensión patrimonial se estima a través de un indicador de capacidad económica, construido a partir de la relación entre la cantidad de ocupados y jubilados del hogar y el número total de miembros. En el cálculo se consideran los años de instrucción, el sexo, la edad y el lugar de residencia. El *IPMH* define cuatro grupos de hogares: (1) Sin privación (no pobres); (2) Con privación sólo de recursos corrientes; (3) Con privación sólo patrimonial y (4) Con privación de ambos tipos simultáneamente (para mayor detalle véase Alvarez *et al.*, 2004).

Cuadro 2: Estadísticos Descriptivos

Variable	Observ.	Media	Desv. Est.	Min.	Max.
<i>HIJOS</i>	531	0.9183	0.2616	0.19	1.83
<i>CONVIV</i>	531	38.2854	6.2231	21.19	68.89
<i>URBAN</i>	531	67.1106	29.9461	0.00	100.00
<i>EDUC</i>	531	9.3681	1.1717	4.83	12.78
<i>ACTIVA</i>	531	41.3712	9.2530	18.65	66.81
<i>HNOPOB</i>	531	48.5865	22.2683	1.23	91.32

La inclusión de las variables explicativas *CONVIV*, *URBAN*, *EDUC*, *ACTIVA* y *HNOPOB* busca capturar el efecto social sobre la fecundidad. Un argumento bien conocido es que las áreas locales pueden tener una amplia variedad de patrones espaciales dependiendo de las diferentes clases sociales que las integran (Knox, 1994). Si existe segregación residencial por clase social y la clase social determina en gran medida la fecundidad, estos patrones residenciales definirán un patrón espacial de reproducción en la región estudiada. También se reconoce desde hace siglos que la tendencia de fecundidad es más baja para las clases sociales altas, como Adam Smith menciona en *La Riqueza de las Naciones* (1776).

Es posible que el efecto de la clase social (capturada por las variables explicativas) determine completamente el lugar donde uno habita así como cuántos hijos decide tener, entonces el componente espacial será prácticamente explicado por el estatus social. En caso contrario, uno esperaría encontrar dependencia espacial asociada con los patrones de fecundidad.

3.2. Matriz de Contactos (*W*) y Análisis Exploratorio

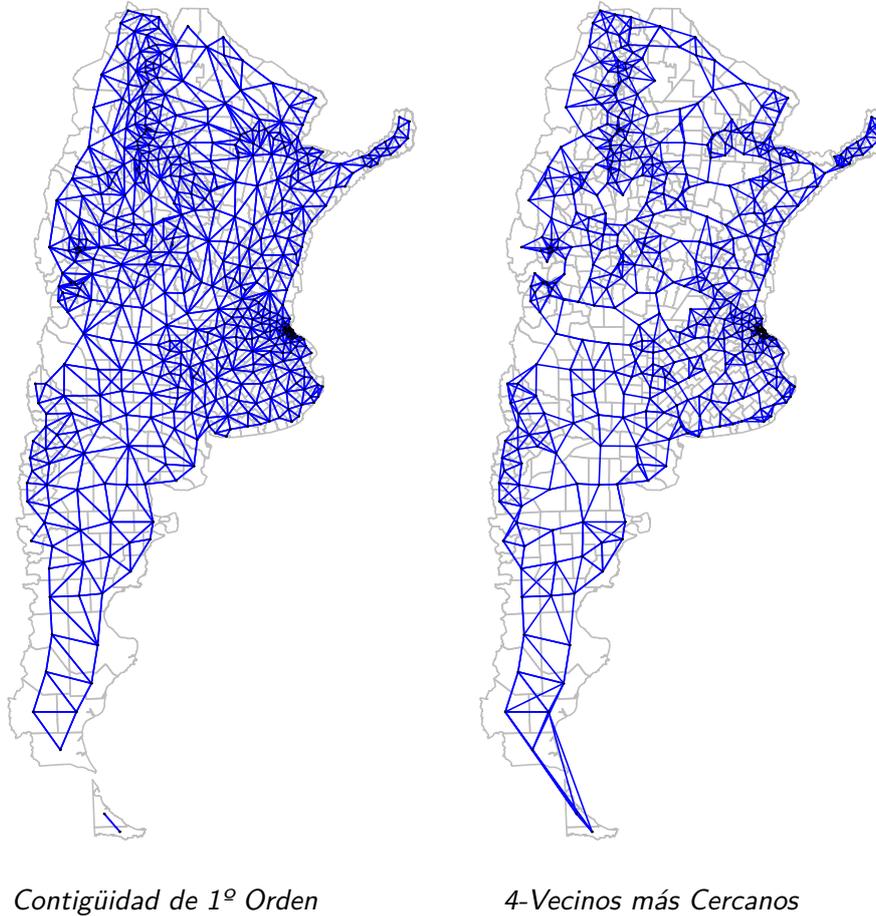
Siguiendo lo bosquejado en la sección 2, la primera cuestión a resolver es la construcción de la matriz de contactos. Dado que no tenemos una teoría que nos guíe en construir la matriz de contactos para los departamentos de la Argentina, consideraremos dos criterios geográficos ampliamente utilizados: tipo reina (contigüidad) y 4-vecinos más cercanos ($4 - nn$) (Figura 3).

En el caso de áreas irregulares, el criterio de contigüidad arroja similares resultados que el criterio tipo reina. Para aplicar los criterios, previamente se ha definido el centroide de cada departamento. Una alternativa al centroide es identificar la ciudad cabecera del departamento, alterando en mayor medida la estructura de vecindad del criterio 4-vecinos más cercanos.

Obsérvese, que en el caso del criterio de contigüidad, los departamentos de Tierra del Fuego quedan desconectados del resto de departamentos. Cuando se utiliza el criterio de los 4-vecinos más cercanos, todas las unidades tienen el mismo número de conexiones. Como se ha mencionado previamente, el uso de criterios geográficos genera un tratamiento exógeno de la matriz de contactos,

evitando problemas de inferencia. La selección entre las matrices de contactos puede ser realizada en una etapa más avanzada, una vez definido el modelo espacial más adecuado.

Figura 3: Mapas de Contactos bajo Criterios Alternativos



A continuación se presenta el mapa de coropletas para la variable *HIJOS*. En la Figura 4, puede observarse que las zonas de baja fecundidad se encuentran distribuidas en la región central del país, esto es, los departamentos de las provincias de Buenos Aires, Córdoba y Santa Fe. La alta fecundidad se concentra en las regiones del norte y sur de la Argentina.

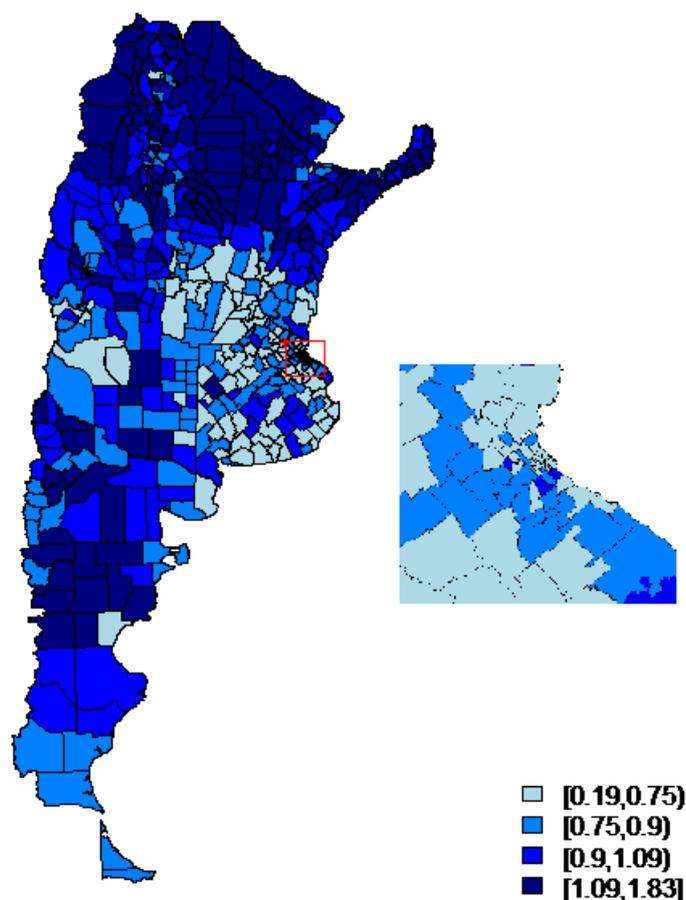
La Figura 4 refleja un patrón espacial que era de esperar, relacionando baja fecundidad con los principales centros urbanos del país. Si bien el mapa brinda información, no puede determinarse si tal información es cuantitativamente relevante. Para obtener una medida cuantitativa de dependencia espacial utilizaremos las matrices espaciales previamente definidas para construir los Diagramas de Dispersión de Moran (Figura 5).

En la Figura 5, bajo ambas matrices de contactos se obtiene una autocorrelación espacial positiva. Es decir, el comportamiento de la tasa de fecundidad en el departamento *i* es similar al de sus vecinos: si los vecinos tienen baja (alta) tasa de fecundidad entonces el departamento analizado tenderá a tener una tasa baja (alta) de fecundidad respecto al promedio nacional.

En los diagramas de dispersión de la Figura 5 si incluye el valor de estadístico *I* de Moran. Las variables graficadas han sido previamente estandarizadas, es decir:

$$hijos_i = (HIJOS_i - \overline{HIJOS}) / \sqrt{\sum (HIJOS_i - \overline{HIJOS})^2 / N - 1}. \quad (55)$$

Figura 4: Distribución Espacial de Fecundidad



Cuando se trabaja simultáneamente con la matriz de contactos y la variable estandarizadas, un camino simple para obtener el valor del test I de Moran es realizar la siguiente regresión:

$$W^* \cdot hijos_i = \beta \cdot hijos_i + \varepsilon_i. \quad (56)$$

El valor del coeficiente estimado es igual al valor I de Moran⁷:

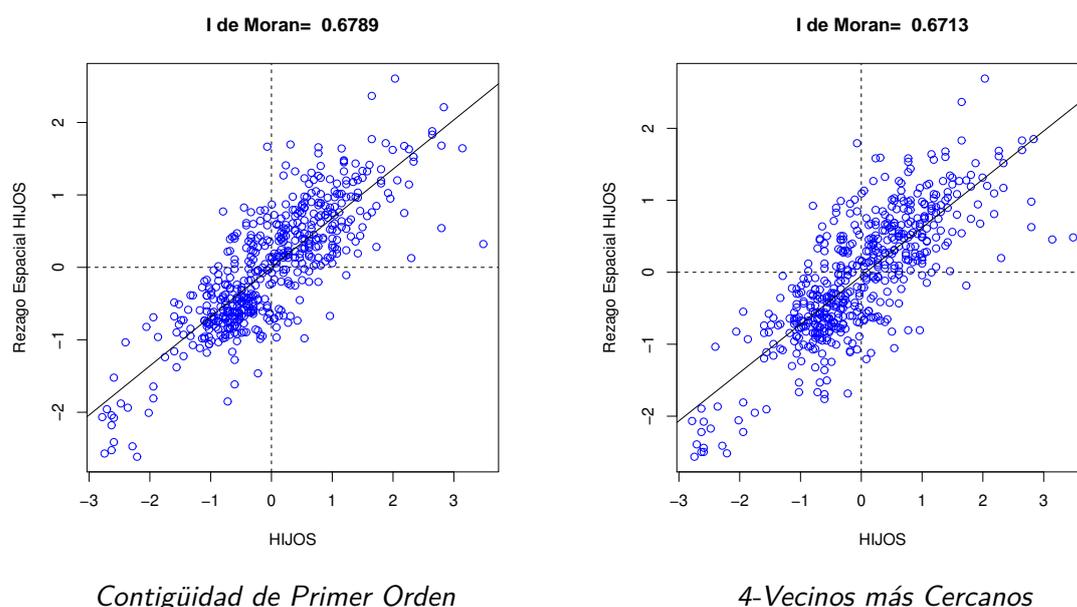
$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y} = \frac{\mathbf{hijos}'W^* \cdot \mathbf{hijos}}{\mathbf{hijos}'\mathbf{hijos}} = I. \quad (57)$$

Necesitaremos calcular la varianza y la media del estadístico para utilizar la aproximación a la distribución Normal y así conocer si los valores obtenidos son significativos estadísticamente. Los valores del test I de Moran normalizado son 25,34, con el criterio de contigüidad, y 23,60, con el criterio de 4-vecinos más cercanos, siendo ambos significativos.

En resumen, la variable fecundidad muestra dependencia espacial significativa tal como inicialmente se había visualizado en el mapa de coropletas.

⁷En el apéndice se presenta el código para realizar este gráfico en *R*.

Figura 5: Diagrama de Dispersión



3.3. Análisis de Regresión Espacial

Detectada la dependencia espacial de la fecundidad, es posible que el conjunto de variables explicativas pueda capturar la variación espacial de la misma. Por lo tanto, nuestro primer modelo es simple:

$$HIJOS_i = \beta_0 + \beta_1 CONVIV_i + \beta_2 URBAN_i + \beta_3 EDUC_i + \beta_4 ACTIVA_i + \beta_5 HNOPOB_i + \varepsilon_i. \quad (58)$$

Los resultados de la estimación del modelo (58) por *MCO* se presentan en el Cuadro 3.

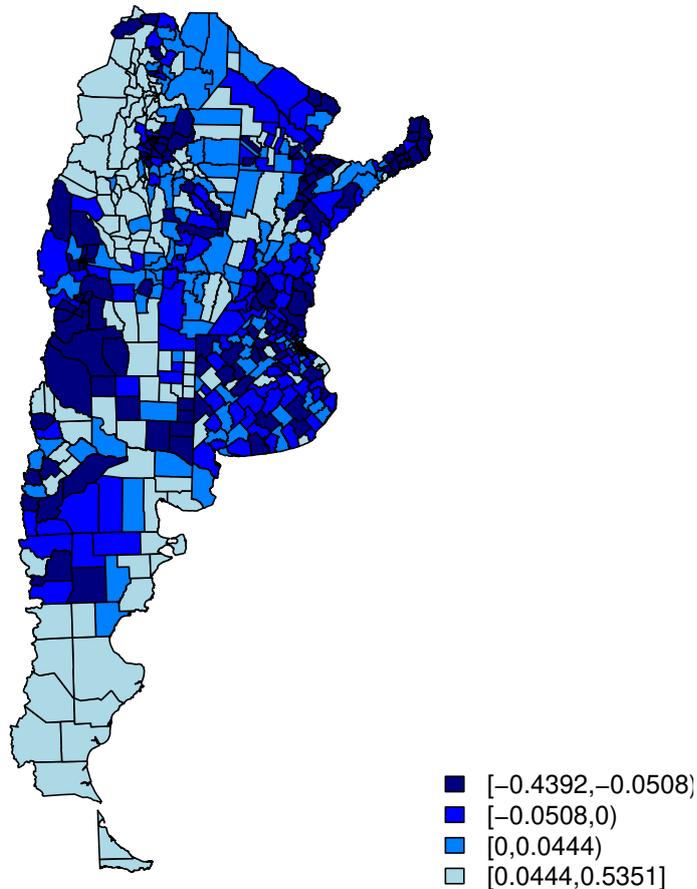
Cuadro 3: Estimación por *MCO*

Variable	Coficiente	t
<i>CONSTANTE</i>	1,2911	12,290
<i>CONVIV</i>	0,0128	12,842
<i>URBAN</i>	-0,0006	3,685
<i>EDUC</i>	-0,0448	4,552
<i>ACTIVA</i>	-0,0039	4,774
<i>HNOPOB</i>	-0,0048	10,568
\bar{R}^2	0,8733	

Los coeficientes tienen el signo esperado. La variable *CONVIV*, que captura el efecto de estar en pareja a la fecha del censo, tiene un efecto positivo sobre la tasa de fecundidad. Las variables *URBAN*, *EDUC*, *ACTIVA* y *HNOPOB* reflejan que la mayor proporción de población urbana, mujeres más educadas en promedio o con mayor participación en el mercado laboral y mayor nivel económico afectan negativamente a la decisión de procreación promedio.

La significancia estadística no puede ser interpretada hasta tanto no se compruebe la hipótesis de independencia en los residuos. Nuevamente, recurrimos al mapa de coropletas para visualizar la distribución de los mismos (Figura 6).

Figura 6: Distribución Espacial de los Residuos MCO



La distribución residual representada por colores no parece tener un patrón aleatorio. La aplicación de la batería de estadísticos espaciales confirma que la distribución residual refleja dependencia espacial (Cuadro 4).

Cuadro 4: Tests de Independencia Espacial

Matrices test	Contigüidad de Primer Orden		4-Vecinos más cercanos	
	Valor	p-valor	Valor	p-valor
<i>I de Moran</i>	0,41	0,0000	0,34	0,0000
<i>LM_{ERROR}</i>	239,93	0,0000	147,67	0,0000
<i>LM_{EL}</i>	143,93	0,0000	85,67	0,0000
<i>LM_{LAG}</i>	102,69	0,0000	72,76	0,0000
<i>LM_{LE}</i>	6,70	0,0096	10,75	0,0010
<i>LM_{SARMA}</i>	246,63	0,0000	158,43	0,0000

El estadístico *I* de Moran muestra un p-valor inferior al nivel de significancia del 1%, detectando dependencia espacial aunque sin informarnos sobre el posible modelo espacial a seguir. Los estadísticos *LM* son más útiles en este sentido.

El test *LM_{ERROR}* y su versión robusta *LM_{EL}* detectan presencia de autocorrelación espacial en la estructura del error, indicándonos como ecuación alternativa un modelo de error espacial. Por su parte, el test *LM_{LAG}* y la versión robusta del mismo *LM_{LE}* detecta indicios de autocorrelación en la estructura de la ecuación y el modelo adecuado será un modelo de rezago espacial.

Dada la información de los tests *LM* podemos optar por estimar los modelos *SLM*, *SEM* y *SARAR*, este último incorpora simultáneamente estructura dinámica en la ecuación y en el término de error. El Cuadro 5 muestra los resultados utilizando las posibles matrices de contactos.

Cuadro 5: Estimación por Máxima Verosimilitud

Matrices	Contigüidad de Primer Orden			4 Vecinos más cercanos		
	<i>SLM</i>	<i>SEM</i>	<i>SARAR</i>	<i>SLM</i>	<i>SEM</i>	<i>SARAR</i>
Variable	Coef. Estim./(t)			Coef. Estim./(t)		
<i>CONSTANTE</i>	0,6837* (6,12)	1,1395* (11,83)	1,1783* (9,99)	0,8311* (7,45)	1,2380 (12,25)	1,1340* (9,68)
<i>CONVIV</i>	0,0130* (14,34)	0,0158* (17,73)	0,0158* (17,24)	0,0128* (13,77)	0,0145* (15,18)	0,0146* (15,09)
<i>URBAN</i>	-0,0011* (6,43)	-0,0004* (2,54)	-0,0003* (2,29)	-0,0010* (5,68)	-0,0003* (1,92)	-0,0004* (2,42)
<i>EDUC</i>	-0,0212* (2,28)	-0,0537* (5,78)	-0,0549* (5,85)	-0,0275* (2,91)	-0,0595* (6,10)	-0,0554* (5,60)
<i>ACTIVA</i>	-0,0018* (2,30)	-0,0034* (4,55)	-0,0034* (4,57)	-0,0018* (2,30)	-0,0026* (3,30)	-0,0024* (3,02)
<i>HNOPOB</i>	-0,0039* (9,39)	-0,0031* (6,55)	-0,0030* (6,38)	-0,0043* (10,21)	-0,0037* (7,57)	-0,0038* (7,89)
ρ	0,3015* (10,01)	-	-0,0330 (0,68)	0,2325* (8,58)	-	0,0694 (1,95)
θ	-	0,7470* (22,46)	0,7700* (21,00)	-	0,6439* (17,09)	0,5824* (12,31)
<i>log MV</i>	556,99	616,40	616,52	544,87	580,19	581,65

Notas: * Significativo al 5 %.

Los resultados de la estimación de los modelos *SLM*, *SEM* y *SARAR* mantienen el signo esperado de las variables explicativas. En el caso de los modelos *SLM* y *SEM*, los coeficientes que capturan la dependencia espacial son positivos y significativos. El problema es que no podemos elegir entre alguno de los dos modelos debido a que no se encuentran anidados.

La estimación del modelo *SARAR* nos permite seleccionar, en este caso aplicado, entre ambos modelos. Como puede observarse, a pesar de que utilizamos dos matrices de contactos diferentes, la evidencia estadística indica que debemos seleccionar el modelo *SEM*: el coeficiente $\hat{\rho}$ no es significativo en ninguno de los modelos *SARAR* y al eliminarlo de la ecuación obtenemos el modelo de error espacial.

Un modelo aún no presentado es el Modelo Espacial de Durbin. Recuérdese que este modelo, si se cumplen las restricciones correspondientes, es posible reducirlo a un modelo de error espacial, *SEM*.

Los resultados de la estimación del modelo de Durbin, bajo las diferentes matrices de contactos, se presentan en el Cuadro 6. Nuevamente, los signos de las variables explicativas son los esperados. En este modelo, adicionalmente se incorporan efectos de vecindad de las variables explicativas y un término de dependencia sustantiva en la ecuación.

El resultado más importante es el estadístico *LR*. Usando la matriz de contactos de contigüidad el resultado es:

$$LR_{COMFAC} = 2 \left[l_{|H_1} - l_{|H_0} \right] = 28,92 \underset{as}{\sim} \chi_5^2, \quad (59)$$

$$p - valor = 0,00002,$$

rechazándose la hipótesis nula.

Bajo la matriz de contactos 4 - nn, el resultado es:

$$LR_{COMFAC} = 2 \left[l_{|H_1} - l_{|H_0} \right] = 47,93 \sim \chi_5^2, \quad (60)$$

$$p - \text{valor} = 0,00000,$$

nuevamente, rechazamos la hipótesis nula.

Cuadro 6: Modelo de Durbin. Estimación por MV

Matrices	Contigüidad de Primer Orden	4 Vecinos más cercanos
Variable	Coef. Estim./(t)	Coef. Estim./(t)
<i>CONSTANTE</i>	0,5370* (3,60)	0,5603* (3,77)
<i>CONVIV</i>	0,0165* (18,34)	0,0162* (16,72)
<i>URBAN</i>	-0,0004* (2,88)	-0,0004* (2,31)
<i>EDUC</i>	-0,0542* (5,73)	-0,0594* (6,13)
<i>ACTIVA</i>	-0,0030* (4,06)	-0,0020* (2,60)
<i>HNOPOB</i>	-0,0023* (4,62)	-0,0027* (5,30)
<i>W · CONVIV</i>	-0,0137* (8,98)	-0,0108* (7,14)
<i>W · URBAN</i>	0,00008 (0,32)	-0,0004 (1,77)
<i>W · EDUC</i>	0,0345* (2,32)	0,0513* (3,54)
<i>W · ACTIVA</i>	0,0014 (1,14)	-0,0006 (0,52)
<i>W · HNOPOB</i>	0,0009 (1,29)	0,0005 (0,72)
ρ	0,6735* (17,44)	0,5547* (12,99)
<i>LR_{COMFAC}</i>	28,92	47,93
<i>log MV</i>	630,86	604,16

Notas: * Significativo al 5%.

En conclusión el modelo de Durbin es el adecuado, según los estadísticos de análisis.

Ante la disyuntiva de seleccionar entre un modelo con dependencia en el error como el *SEM* y otro modelo más complejo como el de Durbin la literatura se inclina por este último. La elección se debe a las consecuencias inferenciales que se derivan de ambos modelos. Si se estima un modelo de error espacial, pero se ha omitido un parámetro de dependencia sustantiva, los coeficientes estimados son inconsistentes. En el caso de estimarse un modelo de Durbin, omitiendo un parámetro de dependencia residual, los estimadores serán consistentes a pesar de no ser eficientes.

Finalmente, tenemos un modelo espacial elegido pero queda la cuestión de qué tipo de matriz de contactos se debe seleccionar. En este paso, podemos utilizar el ajuste del modelo respecto a la función de máxima verosimilitud. Usando la información de la última fila del Cuadro 6, la matriz de contactos más adecuada será la de 4-vecinos más cercanos.

Como último comentario, es posible explorar modelos más complejos tales como los presentados en la Figura 2, pero los mismos escapan a la finalidad del presente trabajo y se revisarán en futuros avances.

4. Conclusiones

Este trabajo ha pretendido introducir al lector dentro de los principales modelos desarrollados en econometría espacial. La revisión ha incluido cuestiones aún abiertas y de intensa investigación como es el tema de la especificación y selección de la matriz de contactos. Esta matriz es un elemento clave en econometría espacial ya que condiciona todo el análisis posterior.

Adicionalmente, el trabajo presenta la metodología más habitual que es “de lo particular a lo general” en cuanto a la introducción de elementos espaciales. Entre los modelos espaciales más simples podemos mencionar el modelo de error espacial, *SEM*, y el modelo de rezago espacial, *SLM*. Dependiendo de la complejidad estudiada se pueden incorporar elementos espaciales en el error y en la ecuación principal simultáneamente, modelo *SARAR*. Otro modelo más avanzado es el modelo espacial de Durbin.

El trabajo presenta una aplicación empírica con la idea de mostrar la estimación de los diferentes modelos espaciales bajo el entorno del paquete estadístico *R*. Para ello utilizamos como ejemplo el análisis de la fecundidad en la Argentina. El tema aplicado es por si mismo valioso debido a la falta de referencia de similares estudios para nuestro país.

La especificación final del modelo de fecundidad ha sido desarrollada mediante la guía estadística. Se reconoce que el análisis de la fecundidad es muy complejo y requiere una discusión teórica más detallada así como la incorporación de un número más amplio de variables. En futuros avances pretendemos unificar teoría económica e información estadística para arribar a una especificación más completa.

Como comentario final queremos destacar la información brindada en el apéndice, en donde se incorpora el código en *R* sobre la estimación de los diferentes modelos. Esperamos con ello incentivar la investigación dentro de esta área econométrica.

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] Alvarez, G., Gómez, A., Lucarini, A., Mario, S. y F. Olmos (2004). "El estudio de la pobreza según el Censo Nacional de Población, Hogares y Viviendas 2001 Índice de Privación Material de los Hogares (IPMH)," *DNESyP/DEP/P5/PID, Serie Pobreza, DT Metodológico*, Instituto Nacional de Estadística y Censos de la República Argentina (INDEC).
- [3] Anselin, L. (1980). "Estimation Methods for Spatial Autoregressive Structures," *Regional Science Dissertation and Monograph Series*, Ithaca, New York.
- [4] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer
- [5] Anselin, L. (1995). "Local Indicators of Spatial association-LISA," *Geographical Analysis*, 27, pp. 93-115.
- [6] Anselin, L. (2001). "Rao's Score Test in Spatial Econometrics," *Journal of Statistical Planning and Inferences*, 97, pp. 113-139.
- [7] Anselin, L. (2002). "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models," *Agricultural Economics*, 17, pp. 247-267.
- [8] Anselin, L. y A. Bera (1998). "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," en A. Ullah y D. Giles (eds.), *Handbook of Applied Economic Statistics*. New York: Marcel Dekker.
- [9] Anselin, L., Bera, A., Florax, R. y M. Yoon (1996). "Simple Diagnostic Tests for Spatial Dependence," *Regional Science and Urban Economics*, 26, pp. 77-104.
- [10] Anselin, L. y R. Florax (eds.) (1995). *New Directions in Spatial Econometrics*. Berlin: Springer-Verlag.
- [11] Becker, G.S. (1960). "An Economic Analysis of Fertility," en *Demographic and Economic Change in Developed Countries*, Princeton, NJ: Princeton University Press.
- [12] Bivand, R. S., E. J. Pebesma y V. Gómez-Rubio (2008). *Applied Spatial Data Analysis with R*. New York: Springer.
- [13] Bivand, R. S., W. Müller y M. Reder. (2009). "Power calculations for global and local Moran's I," *Computational Statistics and Data Analysis*, 53, pp. 2859-2872.
- [14] Bodson, P. y D. Peeters (1975). "Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation. An Application to a Belgian Labour-Demand Function," *Environment and Planning A*, 7, pp. 455-472.
- [15] Bryan, K. y C. Zick (2006). *The Economic Organization of the Household*. 2ª Edición. London: Cambridge University Press.
- [16] Brock, W. y S. Durlauf (1995). "Discrete choice with social interactions I: Theory". Working Paper N° W5291. *National Bureau of Economic Research*.
- [17] Case, A., Rosen, H. y J. Hines (1993). "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States," *Journal of Public Economics*, 52, pp. 285-307.
- [18] Chasco Yrigoyen, C. (2003). *Econometría Espacial Aplicada a la Predicción-Extrapolación de Datos Microterritoriales*. Consejería de Economía e Innovación Tecnológica de la Comunidad de Madrid.

- [19] Cleland, J. y C. Wilson (1987). "Demand Theories of the Fertility Transition: An Iconoclastic View," *Population Studies*, 41(1), pp. 5-30.
- [20] Cliff, A. D. y J. K. Ord (1969). "The problem of Spatial autocorrelation," en A. J. Scott (ed.), *London Papers in Regional Science 1*, Studies in Regional Science. London: Pion.
- [21] Cliff, A. D. and J. K. Ord (1973). *Spatial autocorrelation*. London: Pion.
- [22] Cliff, A. D. and J. K. Ord (1981). *Spatial Processes, Models and Applications*. London: Pion.
- [23] Cressie, N. (1993). *Statistics for Spatial Data*. Revised edition. New York: Wiley.
- [24] Dacey, M. (1969). "Similarities in the Areal Distributions of Houses in Japan and Puerto Rico," *Area*, 3, pp. 35-37.
- [25] Davidson, R. y J. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- [26] Fernández, E., Mayor, M. y J. Rodríguez (2009). "Estimating Spatial Autoregressive Models by GME-GCE Techniques," *International Regional Science Review*, 32, pp. 148-172.
- [27] Fischer, M. y A. Getis (eds) (2010). *Handbook of Applied Spatial Analysis. Software Tools, Methods and Applications*. Berlin: Springer-Verlag.
- [28] Florax, R. y T. de Graaff (2004). "The Performance of Diagnostic Tests for Spatial Dependence in Linear Regression Models: A Meta-analysis of Simulation Studies," en L. Anselin, R. Florax y S. Rey (eds.), *Advances in Spatial Econometrics*. Berlin: Springer-Verlag.
- [29] 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.
- [30] Geary, R. C. (1954). "The contiguity ratio and statistical mapping," *The Incorporated Statistician*, 5, pp. 115-145.
- [31] Getis, A. y J. Aldstadt (2004). "Constructing the Spatial Weights Matrix Using a Local Statistic Spatial," *Geographical Analysis*, 36, pp. 90-104.
- [32] Getis, A. y J. Ord (1992). "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis*, 24, pp. 189-206.
- [33] Getis, A. y J. Ord (1996). "Local Spatial Statistics: An Overview," en Longley, P. y M. Batty (eds), *Spatial Analysis: Modelling in a GIS Environment*. Cambridge: GeoInformation International.
- [34] Goldberger, A. (1991). *A Course in Econometrics*. Harvard University Press,
- [35] Greene, W. (1993). *Econometric Analysis*. 2ª edición. New York: Macmillan.
- [36] Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
- [37] Hayashi, F. (2000). *Econometrics*. Princeton University Press.
- [38] Kennedy, P. (1998). *A Guide to Econometrics*. 4ª edición. Cambridge: MIT Press.
- [39] Knox, P. (1994). *Urbanization: An introduction to Urban Geography*. Englewood Cliffs, NJ: Prentice Hall.
- [40] Krugman, P. (1991). *Geography and trade*. Cambridge, MA: MIT Press.

- [41] Kuznets, S. (1966). *Modern Economic Growth: Rate Structure and Spread*. New Haven: Yale University Press.
- [42] Lancaster, T. (2000). "The incidental parameter problem since 1948," *Journal of Econometrics*, 95, pp. 391–413.
- [43] Lesage, J. y K. Pace (2009). *Introduction to Spatial Econometrics*. Boca Raton: CRC Press.
- [44] Ma, J., Haining, R. y S. Wise (1997). "SAGE User's Guide," *Dpto. de Geografía y Sheffield Center for Geographic Information and Spatial Analysis*, Universidad de Sheffield.
- [45] Manski, C.F. (1995). *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press.
- [46] Mantel, N. (1967). "The Detection of Disease Clustering and a Generalized Regression Approach," *Cancer Research*, 27, pp. 209-220.
- [47] Molho, I. (1995). "Spatial Autocorrelation in British Unemployment," *Journal of Regional Science*, 35, pp. 641-658.
- [48] Moran, P. (1948). "The Interpretation of Statistical Maps," *Journal of the Royal Statistical Society B*, 10, pp. 243-251.
- [49] Moran, P. (1950). "Notes on Continuous Stochastic Phenomena," *Biometrika*, 37 (1), pp. 17–23.
- [50] Mur, J. y A. Angulo (2009). "Model Selection Strategies in a Spatial Setting," *Regional Science and Urban Economics*, 39, pp. 200-213.
- [51] 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.
- [52] Ord, J. y A. Getis (2001). "Testing for local spatial autocorrelation in the presence of global autocorrelation," *Journal of Regional Science*, 41, pp. 411–432.
- [53] Paelinck, J. y L. Klaassen (1979). *Spatial Econometrics*. Farnborough: Saxon House.
- [54] Racine, J. y R. Hyndman (2002). "Using R to Teach Econometrics," *Journal of Applied Econometrics*, 17(2), pp. 149-174.
- [55] Ruud, P. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press.
- [56] Sharlin, A. (1986). "Urban-Rural differences in fertility in Europe during the demographic transition," en A. Coale y S. Watkins (eds.), *The decline of fertility in Europe*. Princeton, NJ: Princeton University Press.
- [57] Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. E. Cannan (ed), Chicago: The University Press [1976].
- [58] Tiefelsdorf M. y D. Griffith (2007). "Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach," *Environment and Planning A*, 39(5), pp. 1193-1221.
- [59] Tobler, W. (1979). "Cellular Geography," en S. Gale y G. Olsson (eds.), *Philosophy in Geography*, pp. 379–386. Dordrecht: Reidel.
- [60] Toral, A. (2001). "El Factor Espacial en la Convergencia de las Regiones de la UE: 1980- 1996," Tesis doctoral, ICADE, Universidad de Comillas. Madrid.

- [61] Weeks, J. (2003). "The Role of Spatial Analysis in Demographic Research," en F. Goodchild y D. Janelle (eds.), *Spatially Integrated Social Science: Examples in Best Practice*. New York: Oxford University Press.
- [62] Weeks, J., Getis, A., Hill, A., Gadalla, M. y T. Rashed (2004). "The Fertility Transition in Egypt: Intraurban Patterns in Cairo," *Annals of the Association of American Geographers*, 94(1), pp. 74–93.

Apéndice: Datos, Archivos, Paquetes y Código en *R*

El análisis empírico ha sido realizado usando datos del Censo 2001 (*INDEC*), disponibles por departamento. Los datos pueden descargarse por medio del programa *REDATAM SP* desde la siguiente página web: www.indec.gov.ar/Redatam+SP.

Los archivos shapefile por departamentos de Argentina pueden descargarse de manera gratuita desde la siguiente página web: www.gadm.org/country.

En la actualidad, los investigadores aplicados disponen de una amplia gama de herramientas para estimar modelos espaciales. Estas herramientas se han implementado en programas comerciales tales como SAS, Matlab o Stata. Lamentablemente, estos programas comerciales pueden ser muy costosos y a veces prohibitivos. Una excelente alternativa para iniciarse en econometría espacial es OpenGeoDa (<https://geodacenter.asu.edu/software/downloads>), aunque presenta las limitaciones de que su código no es modificable y carece de los últimos avances de estimación⁸. Afortunadamente, existe otra alternativa a OpenGeoDa y a los programas comerciales que tiene como virtud su actualización constante incluyendo la incorporación de los últimos avances en econometría espacial. El programa es denominado "*R*" y se encuentra disponible bajo licencia pública GNU. Para obtener este programa simplemente debe ingresarse al sitio <http://www.r-project.org> y a continuación se deben seguir las instrucciones para su descarga.

Básicamente, *R* es un lenguaje de programación de código abierto y presenta varias ventajas: libre, portable, programable y permite el lenguaje matricial (Racine y Hyndman, 2002). Como desventaja puede mencionarse que funciona por línea de comandos y algunas funciones pueden no encontrarse disponibles ya que el funcionamiento de *R* se encuentra fundamentado en el uso de *paquetes*. Cada *paquete* es una colección de funciones que permiten la realización de determinadas tareas que previamente debe instalarse (situación similar a los "*toolboxes*" de Matlab). Debido a las contribuciones provenientes de diferentes áreas científicas, es posible que algunas funcionalidades entre los *paquetes* se solapen ofreciendo estimaciones similares.

Para el área de econometría espacial, *R* dispone de diversos *paquetes* que se complementan. A continuación presentamos el código con los *paquetes* disponibles para replicar el análisis empírico:

```
## Cargando los paquetes espaciales
library(maps)          ## Projections
library(maptools)     ## Data management
library(sp)           ## Data management
library(spdep)        ## Spatial autocorrelation
library(RColorBrewer) ## Visualization
library(classInt)     ## Class intervals
library(lmtest)       ## LM Tests
### Carga de Datos y matrices ###
argentina <- read.csv("arg_centroid_cid.txt",header=TRUE,skip=1)
summary(argentina)
polgal <- read.gal("arg_cidQUEEN.GAL",override.id=TRUE)
attributes(polgal)
print(polgal)
```

⁸El reciente lanzamiento de GeoDaSpace ha cubierto algunas limitaciones en referencia al método de estimación de momentos generalizados o GMM (Generalized Method of Moments).

```

summary(polgal)
attach(argentina)
polgwt <- read.gwt2nb("arg_cid4NN.GWT",region.id=PAIS_ID)
summary(polgwt)
### Mapas de Contactos ###
map <- readShapePoly("fec_ampliado",IDvar="PAIS_ID")
map_crd<-coordinates(map)
par(mfrow=c(1,2),mar=c(0,0,1,0))
plot(map,border="gray")
plot(polgal,coords=map_crd,pch=19, cex=0.1, col="blue", add=T)
plot(map,border="gray")
plot(polgwt,coords=map_crd,pch=19, cex=0.1, col="blue", add=T)
### Mapa HIJOS ###
hijos.palette <- colorRampPalette(c("lightblue", "#007FFF", "blue", "#00007F"),
+ space = "rgb")
palH <- hijos.palette(4)
classes_hijos <- classIntervals(HIJOS, n=4,rtimes = 1)
colsH <- findColours(classes_hijos,palH)
par(mar=rep(0,4))
plot(map,col=colsH, pretty=T, border="black")
legend(x="bottomright",cex=1,fill=attr(colsH,"palette"),bty="n",
+ legend=names(attr(colsH, "table")),ncol=1)
### I de Moran HIJOS ###
polqueen <- nb2listw(polgal)
pol4nn <- nb2listw(polgwt)
moranHijos_queen<-moran.test(HIJOS,polqueen,randomisation=FALSE,
+ alternative="two.sided") print(moranHijos_queen)
moranHijos_4nn<-moran.test(HIJOS,pol4nn,randomisation=FALSE,
+ alternative="two.sided")
print(moranHijos_4nn)
### Gráfico de Moran HIJOS ###
x<-HIJOS
zx<-(x-mean(x))/sd(x)
mean(zx)
var(zx)
# W reina #
w1file<-polqueen
w1zx<-lag.listw(w1file,zx)
morlm<-lm(w1zx~zx)
aa1<-morlm$coefficients[1]
mori1<-morlm$coefficients[2]
par(pty="s")
plot(zx,w1zx,xlab="HIJOS",ylab="Rezago Espacial HIJOS", col="blue")
abline(aa1,mori1)
abline(h=0,lty=2)
abline(v=0,lty=2)
title(paste("I de Moran= ",format(round(mori1,4))))
# W 4-vecinos más cercanos #
w2file<-pol4nn
w2zx<-lag.listw(w2file,zx)
morlm<-lm(w2zx~zx)
aa2<-morlm$coefficients[1]
mori2<-morlm$coefficients[2]
par(pty="s")
plot(zx,w2zx,xlab="HIJOS",ylab="Rezago Espacial HIJOS", col="blue")
abline(aa2,mori2)
abline(h=0,lty=2)
abline(v=0,lty=2)

```

```

title(paste("I de Moran= ",format(round(mori2,4))))
### Regresión Lineal ###
mod.lm <- lm(HIJOS ~ CONVIVEN+EDUCA+URBANA+HOGNOPOB, data=argentina)
summary(mod.lm)
### Chequeo de Dependencia Espacial ###
res <- mod.lm$residuals
res.palette <- colorRampPalette(c("#00007F","blue","#007FFF","lightblue"),
+ space = "rgb")
pal <- res.palette(4)
classes_fx <- classIntervals(res, n=4,style="fixed",
+ fixedBreaks=c(-0.4392,-0.0508,0,0.0444,0.5351), rtimes = 1)
cols <- findColours(classes_fx,pal)
par(mar=rep(0,4))
plot(map,col=cols, pretty=T, border="black")
legend(x="bottomright",cex=1,fill=attr(cols,"palette"),bty="n",
+ legend=names(attr(cols, "table")),ncol=1)
res.moranQ <- lm.morantest(mod.lm,polqueen,alternative="two.sided")
print(res.moranQ)
res.moran4 <- lm.morantest(mod.lm,pol4nn,alternative="two.sided")
print(res.moran4)
Tests.LMQ<-lm.LMtests(mod.lm,listw=polqueen,test="all")
tresQ<-t(sapply(Tests.LMQ, function(x) c(x$Statistic, x$parameter,x$p.value)))
colnames(tresQ) <- c("Statistic", "df", "p-value")
printCoefmat(tresQ)
Tests.LM4<-lm.LMtests(mod.lm,listw=pol4nn,test="all")
tres4<-t(sapply(Tests.LM4, function(x) c(x$Statistic, x$parameter,x$p.value)))
colnames(tres4) <- c("Statistic", "df", "p-value")
printCoefmat(tres4)
### Estimación Máxima Verosimilitud ###
mod.slmQ <- lagsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=polqueen)
summary(mod.slmQ)
mod.slm4 <- lagsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=pol4nn)
summary(mod.slm4)
mod.semQ <- errorsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=polqueen)
summary(mod.semQ)
mod.sem4 <- errorsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=pol4nn)
summary(mod.sem4)
### Modelo SARAR ###
mod.sararQ <- sacsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=polqueen)
summary(mod.sararQ)
mod.sarar4 <- sacsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=pol4nn)
summary(mod.sarar4)
### Modelo Espacial de Durbin ###
mod.sdmQ <- lagsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=polqueen,type="mixed")
summary(mod.sdmQ)
anova(mod.sdmQ, mod.semQ)
mod.sdm4 <- lagsarlm(HIJOS ~ CONVIVEN+URBANA+EDUCA+ACTIVAS+HOGNOPOB,
+data=argentina,listw=pol4nn,type="mixed")
summary(mod.sdm4)
anova(mod.sdm4, mod.sem4)

```