

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

DEPARTAMENTO ACADÉMICO DE ESTADÍSTICA E INFORMÁTICA



TRABAJO MONOGRÁFICO

DESCRIPCIÓN METODOLÓGICA DEL MODELO ESPACIAL

AUTORREGRESIVO EN EL ERROR

Presentado para optar el título de Ingeniero Estadístico e Informático

MARCO POLO LUCERO

Modalidad: Examen Profesional

Lima – Perú

2017

INDICE

RESUMEN

I. INTRODUCCIÓN	1
II. MODELO AUTORREGRESIVO ESPACIAL.....	3
2.1 Conceptos Básicos.....	3
2.1.1 Variable de Interés	3
2.1.2 Unidad Espacial.....	3
2.1.3 Autocorrelación Espacial	5
2.1.4 Matriz de Pesos Espaciales.....	6
2.1.5 Estadísticos de Asociación Espacial Univariante	9
2.1.6 Estadísticos de Autocorrelación Espacial en un Modelo de Regresión.....	11
2.2 Modelo Autorregresivo Espacial.....	12
III. MODELO ESPACIAL AUTORREGRESIVO EN EL ERROR	15
3.1 Prueba para la Autocorrelación Espacial Residual Basado en el Estadístico I de Moran	16
3.2 Prueba para la Autocorrelación Espacial Residual Basada en la Estimación de Máxima Verosimilitud	16
IV. METODOLOGÍA DE LA APLICACIÓN DE UN MODELO ESPACIAL AUTORREGRESIVO EN EL ERROR	19
4.1 Análisis de las variables en estudio	19
4.2 Elección de la Unidad Espacial	19
4.3 Elección de la Matriz de Pesos	19
4.4 Detección de Autocorrelación Espacial Global	19
4.5 Análisis de Regresión Espacial.....	20
4.6 Análisis del Modelo Espacial Autorregresivo en el Error (SEM).....	20

V. EJEMPLO DE LA METODOLOGÍA PARA MODELO ESPACIAL AUTORREGRESIVO EN EL ERROR	21
5.1 Datos	21
5.2 Resultados.....	22
5.2.1 Análisis de las Variables en Estudio	22
5.2.2 Elección de la Unidad Espacial para el Estudio de la Fecundidad	23
4.2.3 Elección de la Matriz de Pesos para el Modelo de la Fecundidad	25
5.2.4 Detección de Autocorrelación Espacial Global en la Fecundidad	27
5.2.5 Análisis de Regresión Espacial de la Fecundidad.....	30
5.2.6 Análisis del Modelo Espacial Autorregresivo en el Error (SEM)	32
VI. CONCLUSIONES.....	34
VII. RECOMENDACIONES	35
VIII. REFERENCIAS BIBLIOGRÁFICAS	36
IX. ANEXOS	37

TABLA DE FIGURAS

Figura 1 - Unidades Espaciales	4
Figura 2 - Unidad Espacial y Matriz de Pesos	7
Figura 3 - Matriz de Contigüidad Torre.....	8
Figura 4 - Matriz de Contigüidad Reina.....	9
Figura 5 - Partidos y/o Departamentos en Argentina.....	24
Figura 6 - Mapa de Relaciones de los Centroides de Cada Partido, Según el Tipo de Matriz de Contactos.....	26
Figura 7 - Mapa de la Fecundidad en Argentina.....	28
Figura 8 - Diagramas de Dispersión I de Moran	29
Figura 9 - Distribución Espacial de los Errores por MCO	31

TABLA DE CUADROS

Cuadro 1 - Estadísticos de Asociación Espacial Global.....	10
Cuadro 2 - Estadísticos de Asociación Espacial Local	11
Cuadro 3 - Estadísticos de Autocorrelación Espacial en el Modelo de Regresión.....	12
Cuadro 4 - Estadísticos descriptivos de las variables en estudio.....	22
Cuadro 5 - Estimación de los Parámetros de las Variables Explicativas por MCO	30
Cuadro 6 - Tests de Autocorrelación Espacial.....	32
Cuadro 7 - Estimación de los Parámetros del Modelo SEM	33

RESUMEN

El propósito de este trabajo es presentar y explicar la metodología de un modelo espacial autorregresivo en el error. Para esto se comenzó explicando la teoría necesaria para comprender la estructura que tiene un modelo espacial autorregresivo en el error. En esta parte se estableció que para este modelo es necesario utilizar variables de tipo de corte transversal, que la unidad espacial es un área geográfica delimitada por un polígono y que el comportamiento de la variable de interés debe ser diferente en cada unidad espacial en el área de estudio, esto para no tener problemas de autocorrelación espacial. Finalmente, este modelo utiliza una matriz de pesos que tiene la función de controlar y capturar la autocorrelación espacial, obteniendo un modelo con parámetros estimados insesgados y consistentes.

Para explicar la aplicación del modelo espacial autorregresivo en el error, se utilizaron los resultados de una investigación que se hizo en Argentina acerca de la Fecundidad (promedio de hijos nacidos vivos al nacer por mujer) en mujeres entre 15 y 29 años de edad. Donde primero se obtuvieron estadísticas básicas de la variable dependiente (variable en interés) y de las independientes. Luego se dividió en 531 unidades espaciales (partidos) al territorio Argentino. Después se eligió dos tipos de matrices de pesos (reina y 4-vecinos más cercanos). Posteriormente se probó con la estadística de I Moran, que si existía autocorrelación espacial global utilizando solo la variable fecundidad. Luego se utilizó la variable fecundidad con sus variables explicativas y se concluyó utilizando las pruebas estadísticas LM-ERR y LM-EL que la autocorrelación espacial se encontraba en la estructura del error. Lo anterior sugería que el modelo espacial autorregresivo en el error era el recomendable a utilizar. Finalmente se presentaron y estimaron a los dos modelos espaciales autorregresivos, uno utilizando la matriz de pesos tipo “reina” y el otro modelo utilizando la matriz de pesos tipo “4-vecinos más cercanos”, ambos modelos con variables explicativas significativas y capturando la autocorrelación espacial en la estructura del error, concluyendo que ambos modelos espaciales autorregresivos en el error son igual de óptimos para el estudio de la Fecundidad en Argentina.

I. INTRODUCCIÓN

Una variable en estudio tiene un determinado comportamiento de acuerdo al espacio geográfico donde se encuentre. Los lugares geográficos o unidades espaciales son áreas delimitadas por polígonos donde la variable de interés tiene el mismo comportamiento dentro del área del polígono pero diferente a otras. En la práctica estas unidades espaciales presentan problemas de dependencia o autocorrelación con sus unidades espaciales vecinas. Un modelo autorregresivo espacial tiene la particularidad de aplacar la autocorrelación espacial, ya que considera de antemano, que el comportamiento de la variable de interés sufre un proceso autorregresivo entre sus unidades espaciales vecinas.

Los modelos espaciales autorregresivos son utilizados cuando se detecta autocorrelación espacial. El primero llamado modelo de autorregresivo simple cuando la variable independiente es únicamente un rezago espacial de la variable dependiente (no hay otras variables explicativas exógenas). El segundo es el modelo espacial autorregresivo con rezago espacial, que además de tener un rezago espacial, también tiene otras variables explicativas exógenas. El tercero es el modelo autorregresivo de error espacial, que se caracteriza por tener solo variables exógenas, pero sus errores o perturbaciones siguen un proceso autorregresivo espacial. Por último existe un modelo que une en simultáneo al segundo y tercer modelo anteriormente mencionados.

El modelo autorregresivo de error espacial se utiliza cuando a un modelo de regresión se le especifica un proceso espacial a las perturbaciones o errores. El error de este modelo sigue un proceso autorregresivo mediante un coeficiente espacial influenciando directamente a la covarianza del error adicionándose además un error no correlacionado y homocedástico.

El presente trabajo tiene como objetivos describir los conceptos básicos necesarios para la estructura de un modelo autorregresivo espacial, analizar las referencias de diferentes autores y principalmente describir la metodología del modelo espacial autorregresivo en el error. A continuación se presenta la estructura del trabajo.

En primer lugar, este trabajo de investigación explica qué es un modelo autorregresivo espacial con todos sus argumentos teóricos. En el siguiente capítulo, se describe teóricamente al modelo

autorregresivo de error espacial. Para que luego en el capítulo consiguiente, se detalle paso a paso la forma metodológica del uso del modelo autorregresivo de error espacial. Finalmente el último capítulo explica un ejemplo utilizando la forma metodológica descrita en el capítulo anterior.

En conclusión, esta monografía evidencia que un modelo espacial autorregresivo en el error es una alternativa cuando los errores de un modelo de regresión se encuentran autocorrelacionados espacialmente, teniendo como variables en estudio una dependiente y otras independientes basadas en datos de corte transversal (en un momento específico en el tiempo). La metodología de este modelo presentó seis pasos que se detallan en el capítulo IV de esta monografía.

II. MODELO AUTORREGRESIVO ESPACIAL

2.1 Conceptos Básicos

Para explicar un modelo de autorregresivo espacial, primero se detallarán algunos conceptos básicos como: la variable de interés, la unidad espacial o geográfica, la autocorrelación espacial, la matriz de contactos o pesos y el rezago o retardo espacial.

2.1.1 Variable de Interés

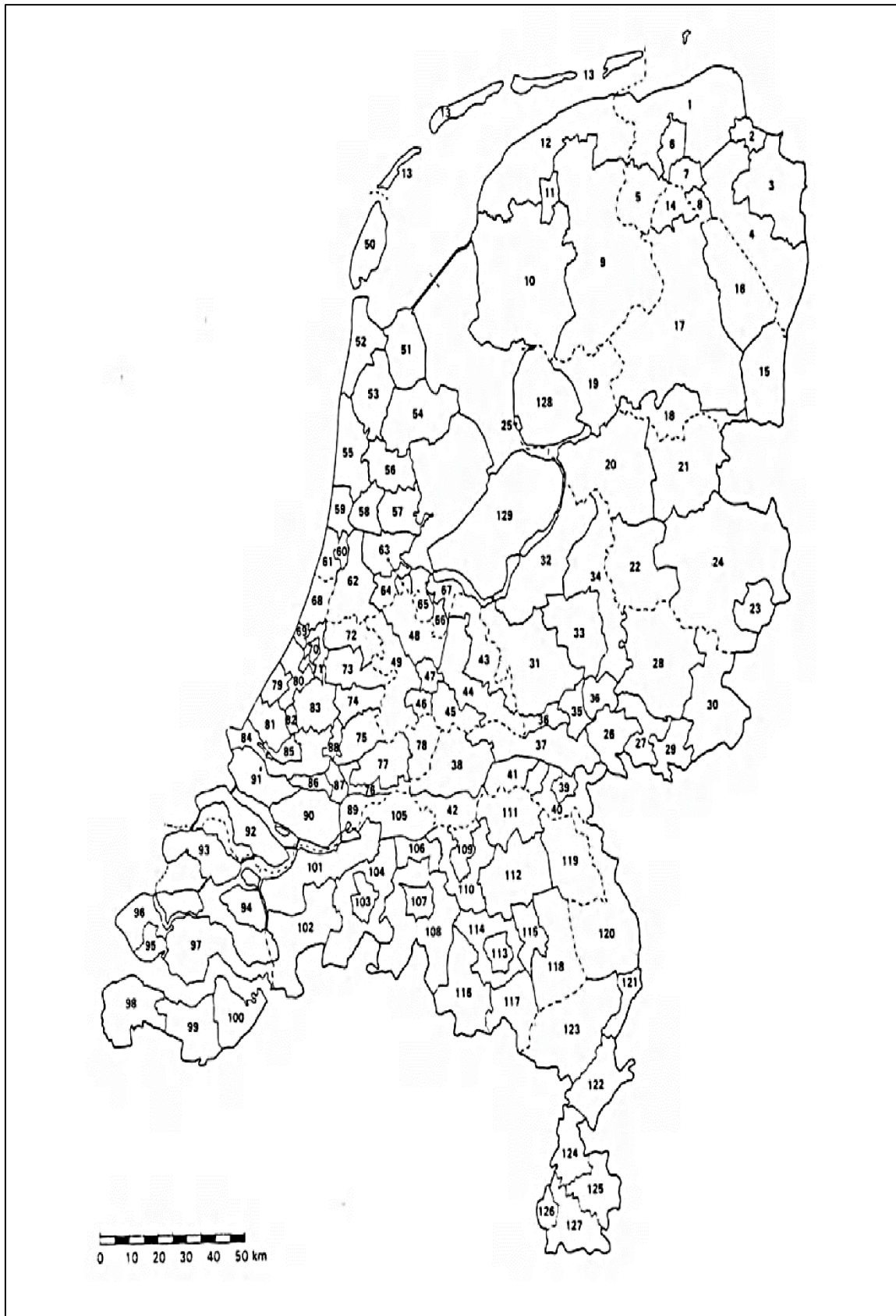
La variable de interés se analiza a partir de datos de corte transversal mediante un modelo de regresión con variables exógenas, que correspondan a un periodo específico en el tiempo. También, se puede usar datos tipo panel, que a diferencia de los mencionados anteriormente, se pueden analizar en tiempos diferentes. El presente estudio utiliza el modelo autorregresivo espacial para datos de corte transversal.

2.1.2 Unidad Espacial

En la mayoría de los ejercicios empíricos en la ciencia regional aplicada, se obtienen datos para las observaciones que se ordenan en el espacio y tiempo o solo en el espacio. En estas situaciones, las observaciones pueden ser caracterizadas por su localización absoluta, usando un sistema de coordenadas, o por su relativa locación, basada en una métrica distante. Es decir, los datos están organizados por unidades espaciales de observación, en el sentido más general (Anselin 1988:11).

Según (Haining 2003:20), la importancia de las relatividades espaciales para explicar la variación de los atributos depende de la escala, es decir, el papel de tales relatividades depende de la escala de la unidad espacial a través de la cual los sucesos se observan y miden en relación con los procesos subyacentes. En palabras sencillas, la unidad espacial es un área geográfica delimitada por un polígono donde la variable en estudio tiene un comportamiento específico, por las diferentes características del lugar geográfico muy aparte de la influencia de las variables exógenas. Un ejemplo real de una delimitación por unidades espaciales de un espacio geográfico se muestra en la Figura 1.

Figura 1 - Unidades Espaciales



Regionalización de los Países Bajos en 127 regiones geográficas económicas (unidades espaciales)

Fuente: Anselin (1995)

2.1.3 Autocorrelación Espacial

La autocorrelación espacial surge siempre que el valor de una variable de interés en un lugar del espacio está relacionado con su valor en otro u otros lugares del espacio (Moreno y Vallá, 2004:84). La autocorrelación espacial puede ser positiva o negativa. Si la presencia de un fenómeno determinado en una unidad espacial lleva a que se extienda ese mismo fenómeno hacia el resto de unidades espaciales que la rodean, favoreciendo así la concentración del mismo, nos hallaremos ante un caso de autocorrelación positiva. Por el contrario, existirá autocorrelación negativa cuando la presencia de un fenómeno en una unidad espacial impida o dificulte su aparición en las unidades espaciales vecinas a ella, es decir, cuando unidades geográficas cercanas sean netamente más disímiles entre ellas que entre unidades geográficas alejadas en el espacio. Por último, cuando la variable analizada se distribuya de forma aleatoria en todo el espacio en estudio, no existirá autocorrelación espacial.

En términos formales, la autocorrelación puede definirse mediante la autocovarianza de la siguiente manera:

$$Cov(y_i, y_j) = E(y_i, y_j) - E(y_i) * E(y_j) \neq 0 \dots (1)$$

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 $\frac{N(N-1)}{2}$ 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, 2000:95).

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

$$y = Ay + \varepsilon \dots (2)$$

Donde A es una matriz de coeficientes de interacción con la siguiente estructura:

$$A = \begin{bmatrix} 0 & \alpha_{i,j} & \alpha_{i,k} \\ \alpha_{j,i} & 0 & \alpha_{j,k} \\ \alpha_{k,i} & \alpha_{k,j} & 0 \end{bmatrix} \dots (3)$$

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, 2009:8). En otros términos, el modelo (2) se encuentra subidentificado. Entiéndase por modelo subidentificado cuando el número de parámetros a ser estimados es mayor que el número de varianzas y covarianzas (Byrne, citado por Gutiérrez 2008:25). 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).

2.1.4 Matriz de Pesos Espaciales

La matriz de pesos espaciales es la expresión formal de la dependencia espacial entre las observaciones (Anselin 1988:21). Según Moreno y Vallá (2004:85), la matriz de pesos espaciales es un instrumento básico para reflejar las interdependencias en el espacio.

Haining (2003:74) 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 la mayoría de casos, esta matriz es una mera hipótesis de trabajo. Utilizando la estructura $A = \rho W$, el modelo (2) se transforma en un modelo restringido:

$$y = \rho W y + \varepsilon \dots (4)$$

Donde el término $W y$ es comúnmente denominado rezago espacial de la variable independiente. Suponiendo que el tamaño muestral es igual a N, la matriz W de orden $N \times N$, puede representarse de la siguiente manera:

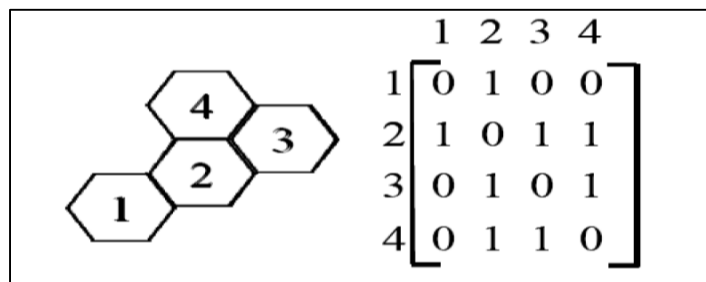
$$W = \begin{bmatrix} 0 & w_{1,2} & \dots & w_{1,j} & \dots & w_{1,N} \\ w_{1,2} & 0 & \dots & w_{2,j} & \dots & w_{2,N} \\ \vdots & \vdots & \ddots & \dots & \dots & \dots \\ w_{i,1} & w_{i,2} & \vdots & 0 & \dots & w_{i,N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots \\ w_{N,1} & w_{N,2} & \vdots & w_{N,j} & \vdots & 0 \end{bmatrix} \dots (5)$$

Donde 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 está formada por ceros, estableciendo que ninguna observación puede estar relacionada consigo misma (no puede ser que la misma observación

sea vecina de sí misma). Obsérvese que se ha intercambiado $\alpha_{i,j}$ (parámetro a estimar dentro del modelo) por $w_{i,j}$ (hipótesis externa al modelo).

Con lo anterior Wy resulta del producto de la matriz de pesos W por el vector de observaciones de una variable aleatoria Y , en que w_{ij} son pesos espaciales y y_j son observaciones de la variable aleatoria $N \times 1$. Para entender lo explicado anteriormente, en la Figura 2 podemos ver 4 unidades espaciales (polígonos), por lo que el vector de la variable aleatoria Y sería de tamaño 4×1 y la matriz de pesos W de tamaño 4×4 , donde se adjudica 1 cuando una unidad espacial es vecina a otra y 0 caso contrario o cuando una misma unidad espacial interacciona consigo misma.

Figura 2 - Unidad Espacial y Matriz de Pesos



Fuente: Aroca (2000)

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, tal como se muestra en el ejemplo de la Figura 2.

- Especificación de la matriz de pesos

(Griffith 1996:80), citado por Getis y Aldstadt (2010:92), concluye que cinco reglas básicas para la especificación de matrices de pesos.

1. Es mejor postular una matriz de pesos geográficos razonables que asumir independencia. Esto implica que se debe buscar o teorizar sobre una W apropiada y que se obtenga mejores resultados cuando se tiene en cuenta la distancia.
2. Se sugiere que para los datos planares, una especificación entre cuatro y seis vecinos es mejor que por encima de seis o por debajo de cuatro.
3. Debería emplearse un número relativamente grande de unidades espaciales, $n > 60$. Siguiendo

la ley de grandes números, la mayoría de las investigaciones espaciales, especialmente debido a unidades espaciales de tamaño desigual, requerirían muestras bastante grandes.

4. Deben darse preferencias a los modelos espaciales de orden inferior a los de orden superior. Siguiendo el principio científico de la parsimonia, siempre es aconsejable elegir modelos menos complicados cuando se presenta la oportunidad.

5. En general, es mejor aplicar una matriz de pesos algo menos especificada (menos vecinos) en lugar de una matriz de pesos superpuestos (extra vecinos)". La sobreespecificación reduce la potencia del análisis.

- Formas experimentales de la matriz de pesos

Hay varias formas experimentales de W , y para este estudio se mencionan cinco, las cuales se denominan “geométricas”. Por geométrica se entenderá que las matrices son principalmente una función de la configuración de las celdas y/o las distancias que las separan. En todos los casos, los W están estandarizados por filas. En otras palabras, se divide cada elemento $w_{i,j}$ entre la suma total de la fila a la que pertenece. De esta forma, la suma de los pesos asociados a cada área es igual a 1. A continuación se presentan cinco formas geométricas de W .

a) Contigüidad – Torre, los cuatro vecinos de cada célula en las direcciones cardinales reciben el valor 1, todos los demás 0. Esta es la formulación más popular de W .

Figura 3 - Matriz de Contigüidad Torre

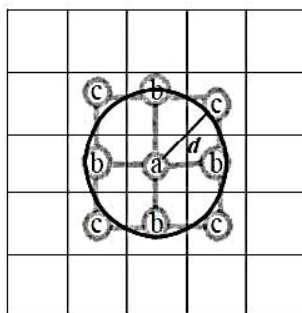
		b		
	b	a	b	
		b		

	b		b	
		a		
	b		b	

Fuente: Anselin (1988)

b) Contigüidad – Reina, los ocho vecinos de cada celda en todas las direcciones reciben el valor 1, todos los demás 0.

Figura 4 - Matriz de Contigüidad Reina



Fuente: Anselin (1988)

c) Distancia Inversa ($1/d$), toma la distancia entre vecinos cercanos como 1, los recíprocos de todos los pares de distancias se calculan y entran en W .

d) Distancia inversa ($1/d^2$), es igual que en 3, excepto que las distancias son cuadradas. Esta formulación de W es probablemente la más popular de toda la distancia basada en W .

e) Distancia inversa ($1/d^5$), es similar a los dos anteriores, excepto en este caso el énfasis está en los vecinos cercanos de cada celda. Cuanto mayor es el exponente, mayor es la influencia de las celdas vecinas en oposición a una mayor distancia entre las celdas. Esta formulación, en muchos aspectos, es comparable a la contigüidad W .

Un criterio alternativo es el de k -vecinos más cercanos. En este caso, considerando la distancia geométrica entre las unidades espaciales, 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.

2.1.5 Estadísticos de Asociación Espacial Univariante

Los modelos autorregresivos utilizan las matrices de pesos o de contactos en su composición para neutralizar la autocorrelación espacial ya sea global o local. Entiéndase de autocorrelación espacial global cuando todas las unidades espaciales de la muestra no se encuentran distribuidas aleatoriamente o si lo hacen conforme a un patrón determinado (Barnio et al. 2012:13) y por autocorrelación espacial local cuando la ubicación de cada observación están correlacionados con sus valores alrededor de este geográficamente hablando.

Según lo comentado por Moreno y Vallá (2004:88) “entre dichos estadísticos globales, los cuales resumen en un indicador único el esquema general de dependencia, se encuentran la I de

Moran (Moran, 1948) y la G(d) (Getis y Ord, 1992)”, se presentan algunas pruebas de autocorrelación global en el cuadro 1.

Cuadro 1 - Estadísticos de Asociación Espacial Global

I de Moran	G(d) de Getis y Ord
$I = \frac{N}{S_0} * \frac{\sum_{i,j}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_{i-1}^N (x_i - \bar{x})^2} \quad i \neq j$ <p> x_i: valor de la variable x en la región i \bar{x}: media muestral de la variable x w_{ij}: pesos de la matriz W N: tamaño muestral $S_0 \sum_i \sum_j w_{ij}$ </p>	$G(d) = \frac{\sum_{i-1}^N \sum_{j-1}^N w_{ij}(d) x_i x_j}{\sum_{i-1}^N \sum_{j-1}^N x_i x_j} \quad i \neq j$ <p> Donde los pares de regiones i y j son vecinas si se encuentran dentro de una distancia determinada </p>
Interpretación tras su estandarización: <ul style="list-style-type: none"> • Z(I)>0 y signif.: autocorrelación (+) • Z(I)<0 y signif.: autocorrelación (-) 	Interpretación tras su estandarización: <ul style="list-style-type: none"> • Z(G)>0 y signif.: concentración de valores elevados • Z(G)<0 y signif.: concentración de valores bajos

Fuente: Moreno y Vallá (2004)

Los estadísticos presentados en el Cuadro 1 permiten contrastar la hipótesis nula de no autocorrelación espacial, es decir, la hipótesis nula de la existencia de una distribución aleatoria de la variable a lo largo del territorio.

Sobre los estadísticos de autocorrelación espacial local (Cuadro 2) Moreno y Vallá (2004:88) mencionan lo siguiente:

Las pruebas de autocorrelación espacial global analizan todas las observaciones de la muestra de forma conjunta, no resultan sensibles a situaciones donde predomine una importante inestabilidad en la distribución espacial de la variable objeto de estudio (procesos no estacionarios espacialmente), es decir, no contemplan la posibilidad de que el esquema de dependencia detectado a nivel global (por ejemplo, ausencia de autocorrelación espacial) pueda no mantenerse en todas las unidades del espacio analizado. Dicha limitación es fácilmente superable por medio del cálculo de los denominados contrastes locales de asociación espacial entre los que se encuentran el estadístico local de Moran, I_i (Anselin, 1995) y los test New-Gi(d) y New-Gi*(d) (Ord y Getis, 1996).

Cuadro 2 - Estadísticos de Asociación Espacial Local

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

Fuente: Moreno y Vallá (2004)

Los estadísticos presentados en el Cuadro 2 obtienen un valor para cada observación de la muestra, con el objetivo de analizar la situación de cada unidad espacial por separado.

2.1.6 Estadísticos de Autocorrelación Espacial en un Modelo de Regresión

Los estadísticos mostrados anteriormente son univariantes, por lo que es necesario tener pruebas estadísticas cuando la variable de interés sea explicada por otras variables exógenas como sucede en los modelos de regresión. Entre los principales destacan el I de Moran, razón de verosimilitud y el basado en los multiplicadores de Lagrange que obtienen de los de razón de verosimilitud (Cuadro 3).

Cuadro 3 - Estadísticos de Autocorrelación Espacial en el Modelo de Regresión

<p>I de Moran</p> $I = \frac{N}{S} * \frac{e'We}{e'e}$	<p>e: residuos MCO N: tamaño muestral S: suma de todos los elementos w_{ij} de la matriz W Debe estandarizarse</p>
<p>LM-ERR</p> $LM - ERR = \frac{[e'We/s^2]^2}{T_1} \sim X_1$	<p>e: residuos MCO s^2: estimación de la varianza residual $T_1: tr(W'W + W^2)$</p>
<p>LM-EL</p> $LM - EL = \frac{\left[\frac{e'We}{s^2} - T_1(RJ_{\rho-\beta})^{-1} e'Wy/s^2 \right]^2}{\left[T_1 - T_1^2(RJ_{\rho-\beta})^{-1} \right]} \sim X_1$	<p>$RJ_{\rho-\beta} = \left[T_1 + \frac{(WX\beta)'M(WX\beta)}{s^2} \right]$ $M = I - X(X'X)^{-1}X'$ y: variable endógena El resto de notación como antes</p>
<p>LM-LAG</p> $LM - LAG = \frac{[e'We/s^2]^2}{RJ_{\rho-\beta}} \sim X_1$	<p>Notación como antes</p>
<p>LM-LE</p> $LM - EL = \frac{\left[\frac{e'Wy}{s^2} - e'We/s^2 \right]^2}{\left[RJ_{\rho-\beta} - T_1 \right]} \sim X_1$	<p>Notación como antes</p>

Fuente: Moreno y Vallá (2004)

2.2 Modelo Autorregresivo Espacial

El modelo autorregresivo espacial captura la relación del valor de una variable en un punto con sus valores en el resto del sistema espacial (Anselin, 1988:33). También, la covarianza entre unidades espaciales dependen de la distancia o rezago entre estas (multidireccional).

El primer modelo autorregresivo espacial, se utiliza cuando el rezago espacial se da por la misma variable de interés sin influencia de variables exógenas. Esto equivale a un modelo espacial autorregresivo de primer orden:

$$y = \rho Wy + \varepsilon \dots (4)$$

Donde ρ es el coeficiente de autocorrelación espacial (su signo determina la relación entre sus efectos vecinos), W es la matriz de contactos o de pesos y ε es el error que asume una distribución normal $\varepsilon \sim N(0, \Omega)$. El término ρWy es la estructura autorregresiva en la variable dependiente que intenta explicar el modelo (el rezago espacial en la variable dependiente). Como el error esta correlacionado con la variable explicativa Wy , los estimadores MCO serán sesgados e ineficientes.

Los siguientes modelos a presentar incluyen la presencia de variables exógenas, pero tienen diferente estructura de acuerdo al tipo de autocorrelación espacial que se quiera controlar.

A diferencia del modelo (4), en el modelo (6) la variable endógena depende de las variables exógenas. A este modelo se lo conoce como modelo de rezago espacial (modelo Lag) o modelo de dependencia espacial sustantiva (rezago más otras variables explicativas) y tiene la siguiente forma:

$$y = \rho Wy + X\beta + \varepsilon \dots \dots \dots (6)$$

Donde y es un vector $N \times 1$, Wy el retardo espacial de la variable dependiente y X una matriz de K variables exógenas, ε un término de perturbación ruido blanco, y ρ el parámetro espacial autorregresivo. Este modelo tiene el mismo problema que el modelo (4), ya que los estimadores MCO serán sesgados e ineficientes.

El tercer modelo (7) llamado modelo espacial autorregresivo en el error, se utiliza cuando la autocorrelación espacial está vinculada al término de error. Plantea un modelo cuya estructura formal tiene un proceso espacial que captura la autocorrelación espacial generada por los errores. Este modelo también es conocido como modelo de error espacial (SEM) y tiene la siguiente forma:

$$\left. \begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= \lambda W\varepsilon + \mu \end{aligned} \right\} \dots (7)$$

Donde y es un vector $N \times 1$, X una matriz de K variables exógenas, $W\varepsilon$ como el rezago espacial en el término de error, λ como el coeficiente autorregresivo y $\mu \sim N(0, \sigma^2 I)$. Los estimadores MCO serán insesgados y consistentes, pero poco eficientes.

Existe un cuarto modelo (8) que se utiliza cuando la autocorrelación es causada por fenómenos de interacción espacial y también por errores de medida, en otras palabras cuando hay dependencia sustantiva (6) y residual (7).

$$\left. \begin{aligned} y &= \rho W y + X \beta + \varepsilon \\ \varepsilon &= \lambda W \varepsilon + \mu \end{aligned} \right\} \dots (8)$$

Donde y es un vector $N \times 1$, X una matriz de K variables exógenas, W_y el retardo espacial de la variable dependiente, $W\varepsilon$ como el rezago espacial en el término de error, ρ el parámetro espacial autorregresivo de W_y , λ como el coeficiente autorregresivo de $W\varepsilon$ y $\mu \sim N(0, \sigma^2 I)$.

III. MODELO ESPACIAL AUTORREGRESIVO EN EL ERROR

Cuando los términos de perturbación en un modelo de regresión muestran dependencia espacial, la suposición estándar de una matriz de covarianza de error esférico no se cumple. En su lugar, esta matriz es de una forma general $\Omega(\theta)$, típicamente parametrizada en función de un pequeño número de coeficientes. En la econometría espacial, estos coeficientes están asociados con un patrón de interacción espacial o estructura espacial que se supone que causa la dependencia (Anselin, 1988:100).

La suposición más utilizada para la forma de dependencia espacial es una especificación autorregresiva espacial. El modelo espacial autorregresivo en el error se denota de la siguiente manera

$$\left. \begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= \lambda W\varepsilon + \mu \end{aligned} \right\} \dots (7)$$

Donde λ es un coeficiente autorregresivo espacial, W es la matriz de peso espacial usual, y μ es un término de error que satisface las suposiciones clásicas de distribución idéntica independiente, con varianza constante σ^2 . Además la varianza de error correspondiente es de la forma:

$$\Omega = E[\varepsilon. \varepsilon'] = E\{(I - \lambda W)^{-1}\mu\}[(I - \lambda W)^{-1}\mu]'\} \dots \dots (9)$$

O también de la siguiente manera:

$$\sigma^2. \Omega(\lambda) = \sigma^2. (I - \lambda W)^{-1}. [(I - \lambda W)^{-1}]' \dots \dots (10)$$

Fisher, citado por Anselin (1988:101), sugiere el uso de $\Omega = A^{1/2}W. A^{1/2}$, donde A es una matriz diagonal de varianzas de error, y W es una matriz de peso espacial simétrica. En el trabajo de Pocock y Cook, citado por Anselin (1988:101), propone que la dependencia espacial en el término de error está parametrizada como una correlación en función de la distancia que separa las observaciones.

Se considera a la matriz de varianza de error como no esférica y asociada sólo con la forma extenderse para aplicar a las otras formas de autocorrelación espacial residual de una manera bastante directa, esto no implica una pérdida de generalidad.

Existen diferentes pruebas para detectar la autocorrelación espacial residual. A continuación se muestra algunas de estas.

3.1 Prueba para la Autocorrelación Espacial Residual Basado en el Estadístico I de Moran

El estadístico I de Moran (Moran, 1947) para la autocorrelación se muestra a continuación y puede aplicarse a los residuos de regresión de manera directa.

$$I = \frac{N}{S} * \frac{\varepsilon'W\varepsilon}{\varepsilon'\varepsilon} \dots \dots \dots (11)$$

Donde ε es un vector de residuos MCO, W es una matriz de contactos o pesos espaciales, N es el número de observaciones, y S es un factor de estandarización, igual a la suma de todos los elementos de la matriz de pesos. Para una matriz de pesos que se normaliza de modo que los elementos de fila se suman a uno, la expresión (11) se simplifica en:

$$I = \frac{\varepsilon'W\varepsilon}{\varepsilon'\varepsilon} \dots \dots \dots (12)$$

La hipótesis nula se basa en la ausencia de dependencia espacial, y no existe una expresión precisa de la hipótesis alternativa. Intuitivamente, se toma la matriz de peso espacial para representar el patrón de interacción espacial potencial que causa dependencia, pero no se especifica la naturaleza del proceso estocástico subyacente. Normalmente se supone que es de una forma autorregresiva espacial. Sin embargo, el coeficiente (12) es matemáticamente equivalente a una regresión MCO de $W\varepsilon$ en ε , en lugar de ε sobre $W\varepsilon$ que correspondería a un proceso autorregresivo. Por otra parte, como ha demostrado Burridge, citado por Anselin (1988:102), la prueba de Moran es proporcional a una prueba de multiplicador de lagrange contra un modelo espacial autorregresivo. En términos generales si se rechaza la hipótesis nula no podemos afirmar que dependencia espacial es por la existencia de errores de medida para las observaciones en unidades espaciales contiguas, ya que también puede ser causado por otros fenómenos de iteración espacial.

3.2 Prueba para la Autocorrelación Espacial Residual Basada en la Estimación de Máxima Verosimilitud

Las pruebas Ratio de Verosimilitud (LR), Wald (W) y Multiplicador de Lagrange (LM-ERR) son aproximaciones asintóticas basadas en la estimación de máxima verosimilitud.

En contraste con el enfoque de la prueba de Moran, las pruebas basadas en Máxima Verosimilitud están rigurosamente estructuradas en términos específicos de hipótesis nulas y alternativas. Formalmente, las hipótesis nula y alternativa son:

$$\left. \begin{array}{l} H_0: \lambda = 0 \\ H_1: \lambda \neq 0 \end{array} \right\} \dots (13)$$

La prueba Ratio de Verosimilitud para la autocorrelación espacial residual es:

$$LR = N \cdot [\ln(\sigma_0^2) - \ln(\sigma_1^2)] + 2 \ln |I - \lambda W| \sim X_1^2 \dots (14)$$

Donde σ_0^2 es la varianza residual estimada para el modelo sin autocorrelación espacial residual (una regresión simple) y σ_1^2 es la varianza residual estimada para el modelo espacial.

Otra prueba basada en la estimación de máxima verosimilitud es la de Wald, que está formulada de la siguiente manera:

$$Wald = \lambda^2 \cdot \left[t_2 + t_3 - \frac{1}{N} (t_1)^2 \right] \sim X_1^2 \dots (15)$$

Donde $t_1 = tr[W \cdot B^{-1}]$, $t_2 = tr[(W \cdot B^{-1})^2]$ y $t_3 = tr[(W \cdot B^{-1})'(W \cdot B^{-1})]$. Además λ se reemplaza por su estimador de máxima verosimilitud.

Utilizando el multiplicador de Lagrange se puede obtener una tercera prueba con la siguiente forma:

$$LM - ERR = \frac{1}{T} * \left[\frac{\varepsilon' W \varepsilon}{\sigma^2} \right]^2 \sim X_1^2 \dots (16)$$

Donde $T = tr\{[W + W'] \cdot W\}$

Finalmente está la prueba LM-EL, que es un Multiplicador de Lagrange ajustado propuesto por Bera y Yo, citado por Anselin y Florax (1995:26), tiene la particularidad de ser robusto cuando hay errores en la especificación de la estructura dinámica de la ecuación principal. La prueba se calcula de la siguiente manera:

$$LM - EL = \frac{\left(\frac{\varepsilon' W \varepsilon}{\sigma^2} - T(R\tilde{J}_{\rho-\beta})^{-1} \left(\frac{\varepsilon' W y}{\sigma^2} \right) \right)^2}{T - T^2(R\tilde{J}_{\rho-\beta})^{-1}} \sim X_1^2 \dots (17)$$

Donde:

$$(R\tilde{J}_{\rho-\beta})^{-1} = \left[T + \frac{(WX\beta)' M(WX\beta)}{\sigma^2} \right]^{-1}$$

Donde $WX\beta$ es un retraso espacial de los valores predichos de una regresión MCO y $M = I - X(X'X)^{-1}X'$, es la matriz de proyección familiar. Al igual que su contraparte LM-ERR, esta estadística también se distribuye como X^2 con un grado de libertad.

Existen más pruebas para detectar la autocorrelación residual. Sin embargo, el presente estudio se enfocará en las más importantes que fueron explicadas anteriormente.

IV. METODOLOGÍA DE LA APLICACIÓN DE UN MODELO ESPACIAL AUTORREGRESIVO EN EL ERROR

Existen varias formas de aplicar un modelo espacial en el error. El presente estudio explica una metodología que se usa cuando a un modelo de regresión lineal se le incluye unidades espaciales y presenta evidencias de algún tipo de autorrelación espacial. Específicamente, autocorrelación causada por los términos de perturbación o errores. Esta metodología se basa en la investigación de Herrera, Paz y Cid (2012).

4.1 Análisis de las variables en estudio

Se realizará un análisis exploratorio de la variable en interés y de cada variable explicativa a través de medidas estadísticas tales como la media, desviación estándar, coeficiente de variabilidad, mínimo y máximo.

4.2 Elección de la Unidad Espacial

Se debe delimitar áreas geográficas (unidades espaciales) más pequeñas, homogéneas tanto en el contexto social como geográfico. Considerando la influencia de las variables explicativas sobre la variable o las variables de interés.

4.3 Elección de la Matriz de Pesos

En la mayoría de casos se utilizan las matrices geométricas como las de contigüidad tipo torre, contigüidad tipo reina y las de distancia inversa. Además, se encuentran también las que están bajo el criterio de los K-vecinos más cercanos. Existen muchas formas, pero para el estudio se utilizarán la de contigüidad tipo reina y la de 4-vecinos más cercanos de las mencionadas anteriormente.

4.4 Detección de Autocorrelación Espacial Global

Para constatar si la variable en estudio se encuentra distribuida de forma totalmente aleatoria en el espacio se utilizarán dos procedimientos. En el primero se construirá una gráfica en donde se pueda observar el comportamiento de la variable de interés en cada unidad espacial y las interrelaciones entre estas. Segundo, se utilizará la prueba estadística I de Moran. Si la prueba resulta significativa, entonces se puede afirmar que existe autocorrelación espacial global.

4.5 Análisis de Regresión Espacial

Se estimarán los parámetros para el modelo de regresión espacial utilizando los mínimos cuadrados ordinarios (MCO). Luego procederá a graficar los residuos MCO en rangos de intensidad para cada unidad espacial con el fin de analizar el comportamiento de estos. También se utilizarán las pruebas estadísticas como la de I de Moran (12) y otras más específicas como las basadas por los estimación de máxima verosimilitud: el multiplicador de Lagrange (16) y el multiplicador de Lagrange ajustado (17) . Si las dos pruebas anteriores resultan significativas, entonces se tiene que utilizar el modelo autorregresivo con dependencia espacial en el error (7) para controlar este tipo de autocorrelación espacial.

4.6 Análisis del Modelo Espacial Autorregresivo en el Error (SEM)

Se estimarán los parámetros del modelo espacial autorregresivo en el error (7). Luego se probará si el modelo es capaz de capturar la autocorrelación residual utilizando las pruebas de Multiplicador de Lagrange (16) y el ajustado (17). Si las pruebas resultan significativas se puede afirmar que el modelo captura el efecto de la autocorrelación espacial residual.

V. EJEMPLO DE LA METODOLOGÍA PARA MODELO ESPACIAL AUTORREGRESIVO EN EL ERROR

En este capítulo se procederá se a explicar, con un ejemplo real, la metodología explicada en el capítulo IV.

5.1 Datos

Los datos utilizados en el presente trabajo provienen del censo de Argentina aplicado en el año 2011. Los mismos se recopilaron y analizaron en el estudio “Introducción a la econometría espacial. Una aplicación al estudio de la fecundidad en la Argentina usando R” en el año 2012, por los investigadores Marcos Herrera, Jorge Paz y Juan Carlos Cid.

Los datos para el ejemplo son de corte transversal, ya que se obtuvieron en un momento específico en el 2011 (Censo del 2011 en Argentina).

La variable en interés (Y) lleva el nombre “Fecundidad” y se mide a partir de la variable “HIJOS”, que es el número promedio de hijos nacidos vivos por mujer que se encuentra entre los 15 y 29 años de edad.

Las variables explicativas (Xs) para la variable Fecundidad son las siguientes:

- CONVIV: Porcentaje de mujeres que conviven con hombres (casadas o pareja), mujeres de 15 a 29 años.
- EDUC: Promedio de años de educación formal, mujeres de 15 a 29 años de edad.
- URBAN: Porcentaje de población urbana dentro del total de mujeres de 15 a 29 años de edad.
- ACTIVA: Tasa de actividad porcentual definida con la relación entre la población económicamente activa y la población total, mujeres de 15 a 29 años
- HNOPOB: Porcentaje de hogares no pobres según criterio IPMH (Indicador de Privación Material de los Hogares)

Los datos y el código en R que utilizaron los investigadores se detallan en el [Anexo 1](#).

5.2 Resultados

A continuación se analizarán los resultados del caso explicado anteriormente.

5.2.1 Análisis de las Variables en Estudio

En el cuadro 4 se muestra el resumen estadístico de las variables en estudio. En este cuadro se observó que la variable explicativa “HNOPOB” resultó ser la más variable (45.83%) y la variable explicativa “EDUC” la menos variable (12.51%). Por otro lado la variable en estudio “HIJOS” resultó con un mínimo de 0.19 hijos nacidos vivos promedio y un máximo de 1.83 hijos nacidos vivos promedio. Además la media estimada de la variable “HIJOS” fue 0.9183 hijos nacidos vivos promedio.

Cuadro 4 - Estadísticos descriptivos de las variables en estudio

Variable	Observ.	Media	Desv. Est.	C.V	Min.	Max.
HIJOS	531	0.9183	0.2616	28.49%	0.19	1.83
CONVIV	531	38.2854	6.2231	16.25%	21.19	68.89
URBAN	531	67.1106	29.9461	44.62%	0.00	100.00
EDUC	531	9.3681	1.1717	12.51%	4.83	12.78
ACTIVA	531	41.3712	9.2530	22.37%	18.65	66.81
HNOPOB	531	48.5865	22.2683	45.83%	1.23	91.32

Fuente: Herrera (2012)

El modelo planteado utilizó variables explicativas como CONVIV, URBAN, EDUC, ACTIVA y HNOPOB para capturar el efecto social sobre la fecundidad y también delimitó la zona en estudio en unidades espaciales para capturar el efecto espacial sobre la fecundidad.

$$\text{Fecundidad} = f(\text{Efecto Social, Contexto Geográfico})$$

5.2.2 Elección de la Unidad Espacial para el Estudio de la Fecundidad

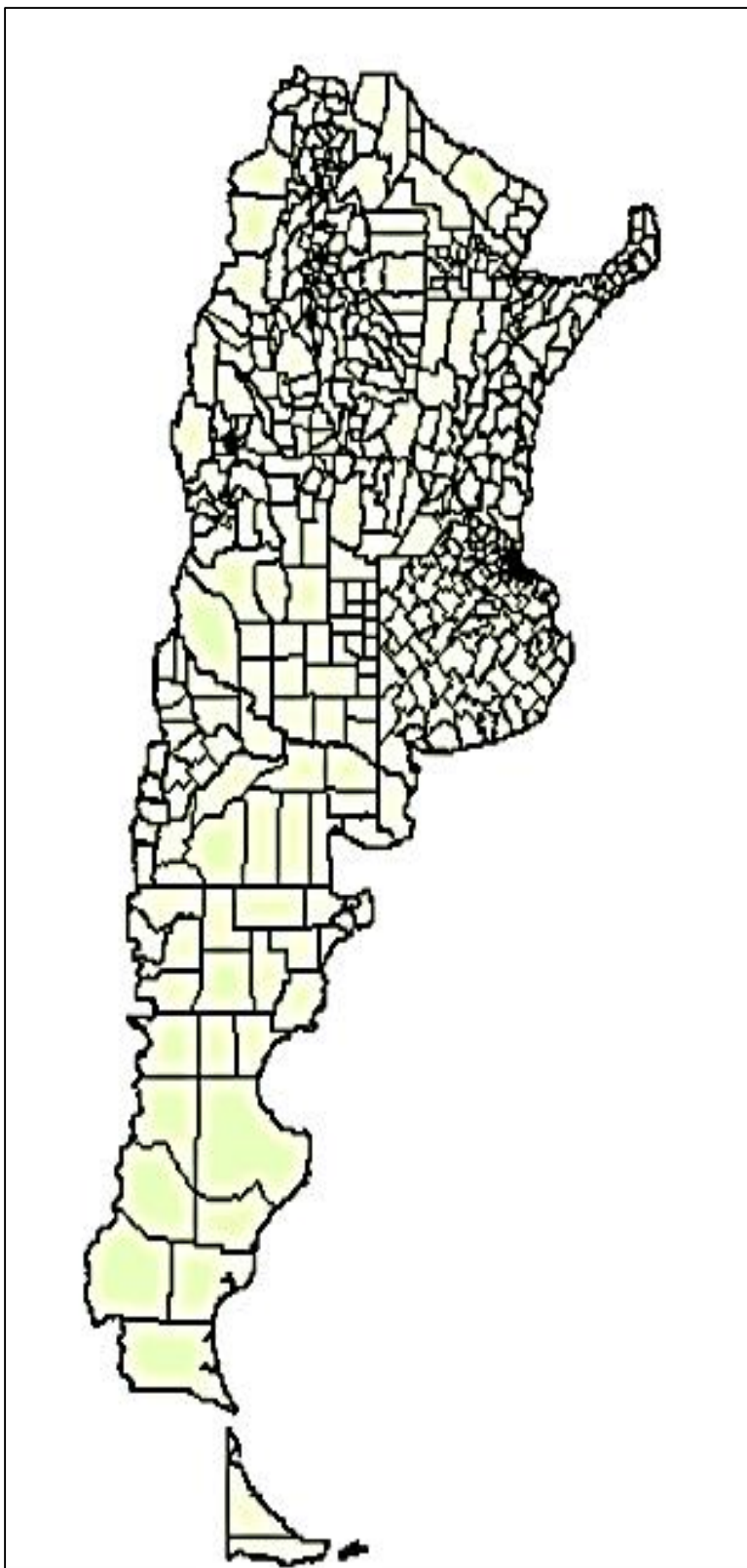
Para la elección de la unidad espacial para el estudio de la fecundidad (número promedio de hijos nacidos vivos por mujer entre 15 y 29 años) en Argentina, se partió de la incertidumbre que los estudios a nivel regional, en su vasta mayoría, examinan la fecundidad por regiones basándose en simples dicotomías urbano-rural, como si el espacio fuese uniforme dentro de estos dos grupos. Por otro lado, hay evidencias de que la universalidad del comportamiento de la fecundidad puede ser matizada por diferencias culturales que pueden producir diferentes niveles de fecundidad entre zonas y dentro de las mismas. En otras palabras, es posible que existan importantes variaciones en la fecundidad entre las diferentes áreas.

El 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.

Finalmente, siguiendo la idea anterior, en vez de tener un dato regional se subdividió a cada provincia en todos sus departamentos o partidos, siendo cada departamento o partido una unidad espacial (figura 5), ya que había indicios que cada partido tenía un comportamiento homogéneo de la variable fecundidad.

Cabe aclarar que en el Perú la provincia está incluida en el departamento, en cambio en Argentina el departamento está incluido en la provincia. Para evitar confusiones se utilizará el término de “partido” en lugar de “departamento”.

Figura 5 - Partidos y/o Departamentos en Argentina



Fuente: www.educ.ar

En la figura 5 se muestran todos los polígonos de los partidos del territorio Argentino, siendo estos 531 partidos o unidades espaciales.

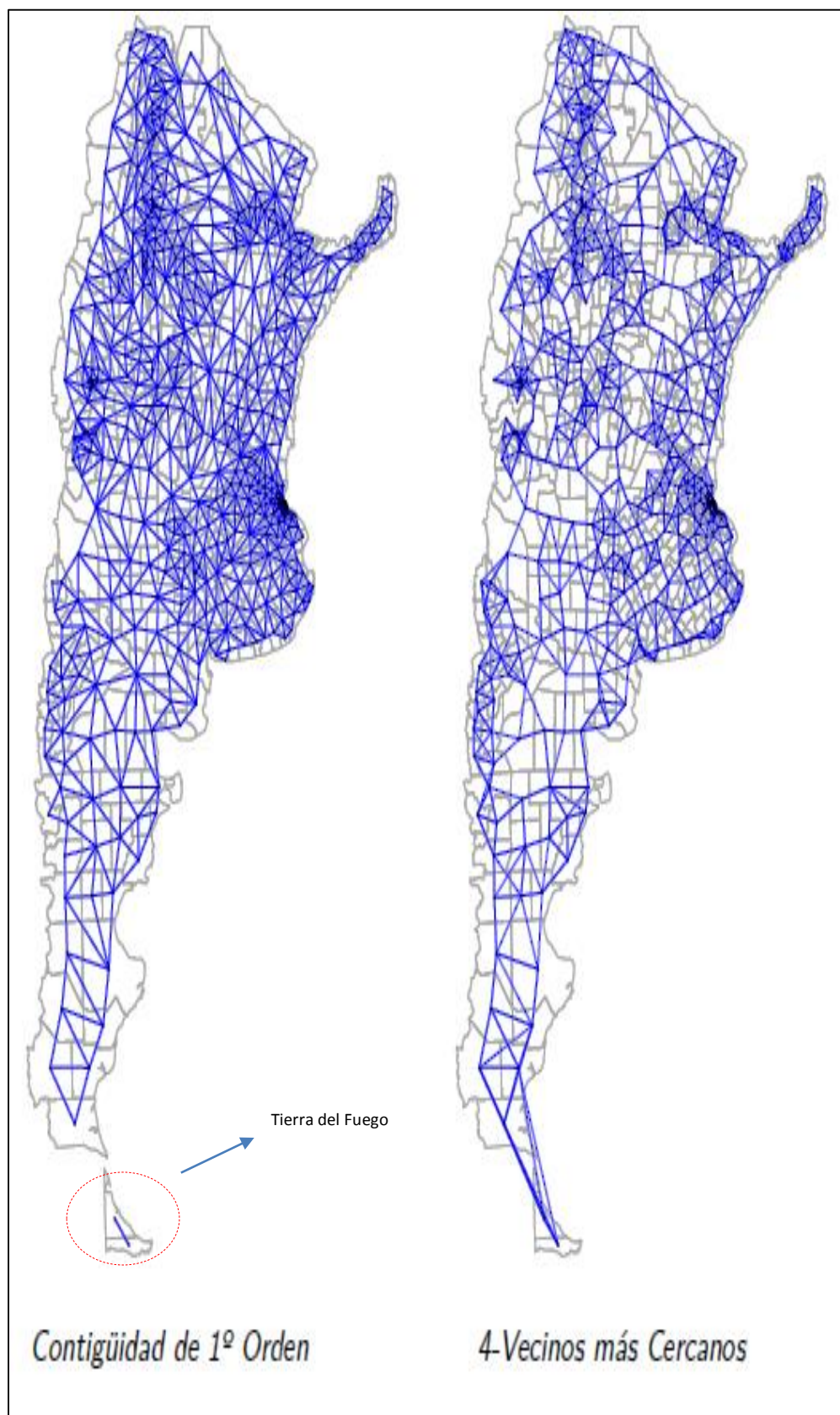
4.2.3 Elección de la Matriz de Pesos para el Modelo de la Fecundidad

La primera cuestión a resolver fue la construcción de la matriz de contactos. Dado que no se encontró una teoría que sirva de guía para construir la matriz de contactos para los departamentos de la Argentina, se consideraron dos criterios geográficos ampliamente utilizados: tipo reina (contigüidad) y 4-vecinos más cercanos ($4 - nn$).

Para aplicar los criterios, previamente se estableció el centroide de cada partido. Una alternativa al centroide se obtuvo al identificar la ciudad cabecera del partido, alterando en mayor medida la estructura de vecindad del criterio 4-vecinos más cercanos (figura 6).

Se detectó, que en el caso del criterio de contigüidad tipo reina, los partidos ubicados en la “Tierra del Fuego” quedaron desconectados del resto de departamentos. Esto es, porque este tipo de criterio no obliga a los polígonos que se encuentran aislados geográficamente a relacionarse o conectarse con los demás. En cambio, cuando se utilizó el criterio de los 4-vecinos más cercanos, todas las unidades espaciales (partidos) tenían 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.

Figura 6 - Mapa de Relaciones de los Centroides de Cada Partido, Según el Tipo de Matriz de Contactos



Fuente: Herrera (2012)

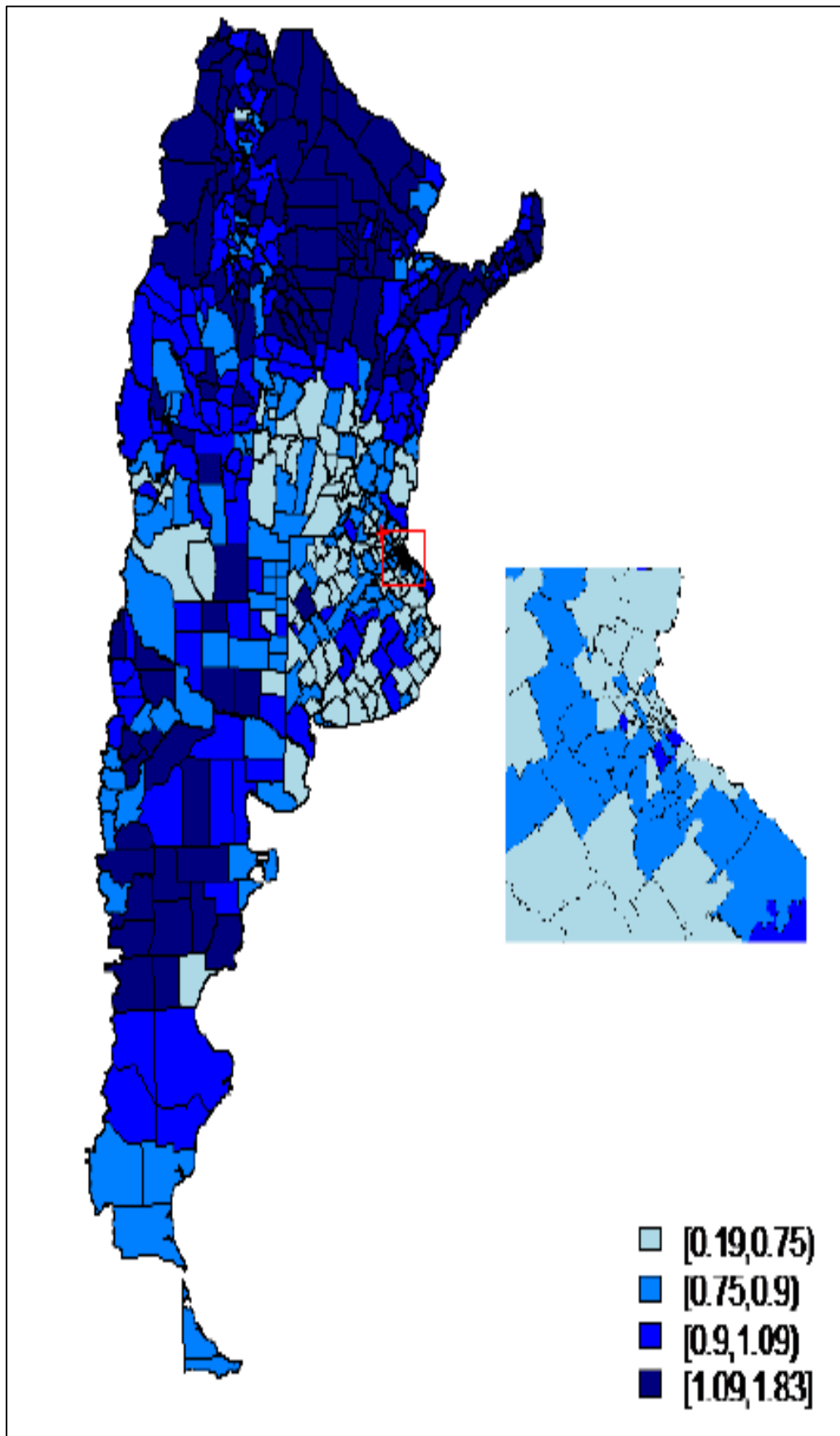
5.2.4 Detección de Autocorrelación Espacial Global en la Fecundidad

En este punto se visualizó el comportamiento espacial de la variable fecundación (HIJOS), observando los niveles de intensidad del número promedio de hijos nacidos de las mujeres entre 15 y 29 en cada unidad espacial (partido).

Para ver este efecto espacial de la fecundidad, se creó rangos de número promedio de hijos nacidos, donde a cada rango se le asignó un color diferente con el objetivo de visualizar fácilmente y empíricamente las interrelaciones del comportamiento de la fecundidad en las diferentes unidades espaciales. Estas unidades se pintaron en distinto tono de color celeste y azul según el rango del número promedio de hijos nacidos. De color celeste claro a las que tuvieron un rango de 0.19 a menos de 0.75, de color celeste oscuro las que tuvieron un rango de 0.75 a menos de 0.9, de color azul claro las que tuvieron un rango de 0.9 a menos de 1.09 y de color azul oscuro las que tuvieron un rango de 1.09 a 1.83 hijos nacidos vivos promedio.

A continuación se presenta el mapa para la variable HIJOS (Figura 7), donde se observó que las zonas de baja fecundidad se encuentran distribuidas en la región central del país (partidos de las provincias de Buenos Aires, Córdoba y Santa Fe). La alta fecundidad se concentró en las regiones del norte y sur de la Argentina. También, se reflejó 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.

Figura 7 - Mapa de la Fecundidad en Argentina



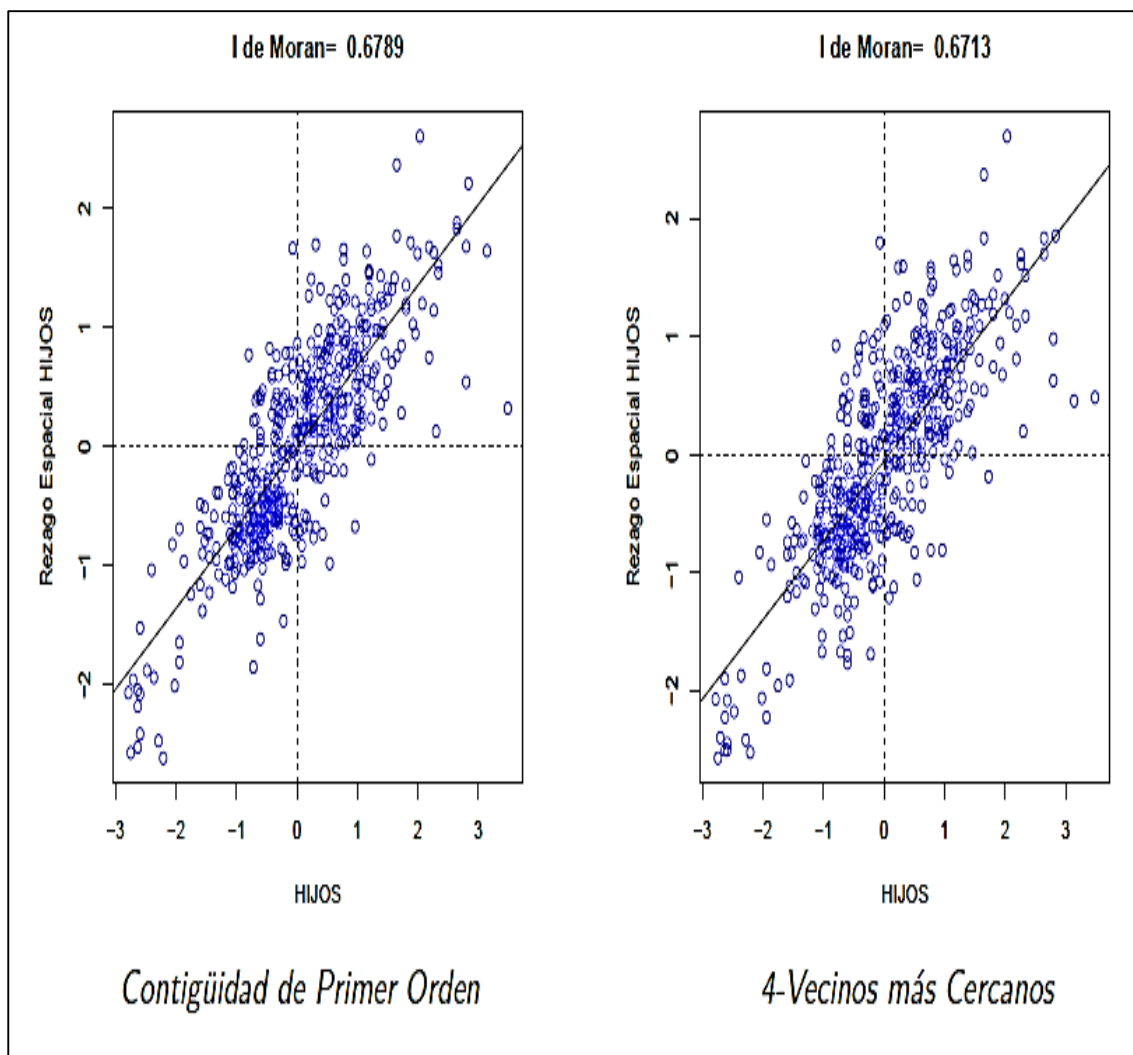
Fuente: Herrera (2012)

Para estimar la dependencia espacial a través de los diagramas de dispersión de Moran, se utilizaron las matrices espaciales contigüidad tipo reina y 4-vecinos más cercanos. En la Figura 8 se muestran estos diagramas así como sus respectivos estadísticos.

En el Figura 8, bajo ambas matrices de contactos se obtuvo una autocorrelación espacial positiva. Es decir, el comportamiento de la tasa de fecundidad en el partido 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.

Los test de I de Moran estandarizados (cuadro1) resultaron ser mayores a cero, por lo que resultó significativa la prueba en ambos casos, por lo que se puede afirmar que la variable fecundidad muestra dependencia o autocorrelación espacial tal como se había visualizado empíricamente en la figura 7. Cabe aclarar que el estudio donde se obtuvo estas estadísticas no mostró los p -valores.

Figura 8 - Diagramas de Dispersión I de Moran



Fuente: Herrera (2012)

5.2.5 Análisis de Regresión Espacial de la Fecundidad

Una vez detectada la autocorrelación espacial de la variable fecundidad, se evaluó la influencia de las variables explicativas en el modelo. Por lo que se propone utilizar el siguiente modelo de regresión lineal:

$$HIJOS_i = \beta_0 + \beta_1 CONVIV_i + \beta_2 URBAN_i + \beta_3 EDUC_i + \beta_4 ACTIVA_i + \beta_5 HNOPOB_i + \varepsilon_i \dots (18)$$

Los resultados de la estimación del modelo (18) por MCO (mínimos cuadrados ordinarios) se presentan en el Cuadro 5.

Cuadro 5 - Estimación de los Parámetros de las Variables Explicativas por MCO

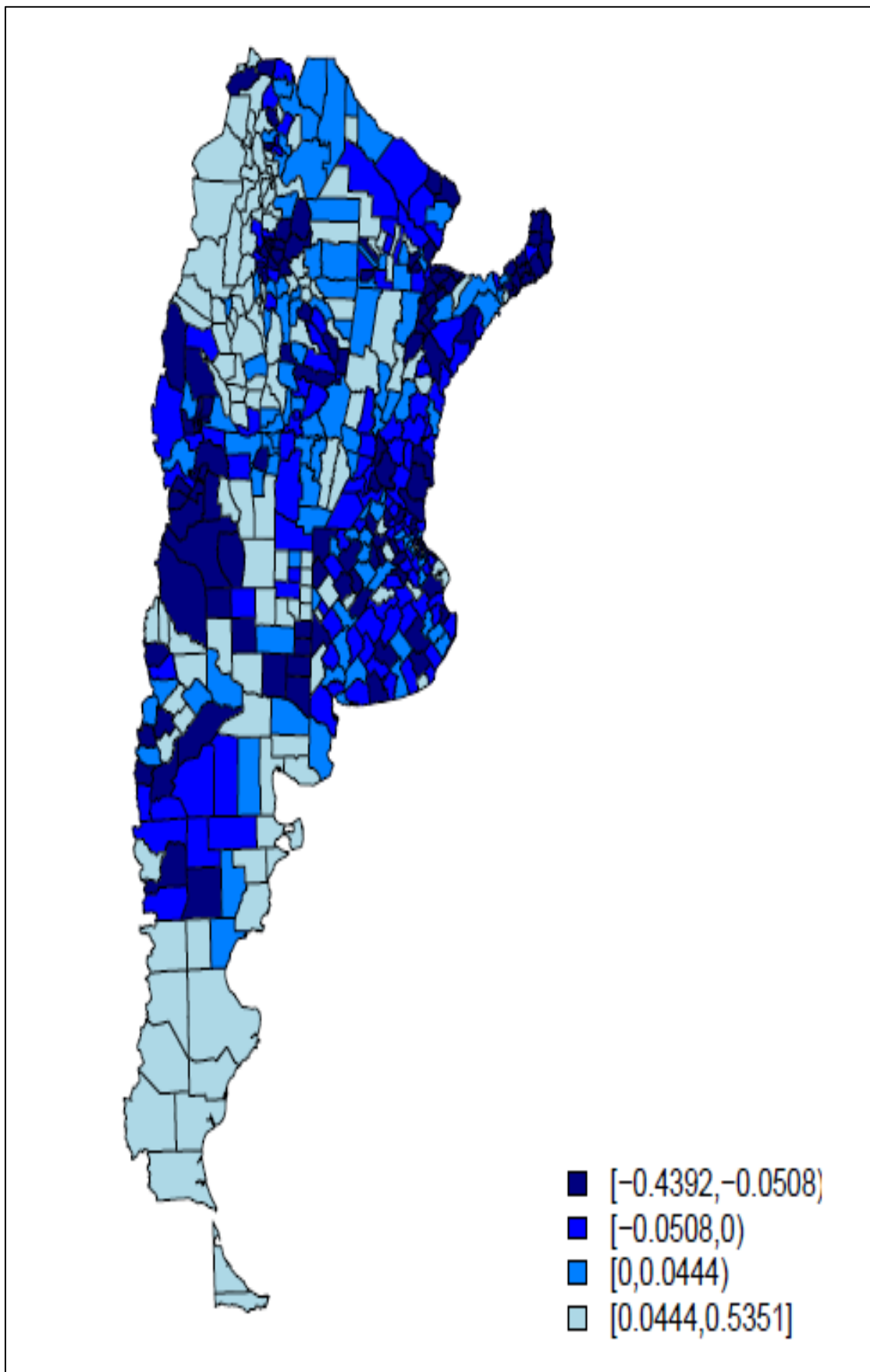
Variable	Coefficiente	t
CONSTANTE	1.2911	12.290
CONVIV	0.0128	12.842
URBAN	-0.0006	3.685
EDUC	-0.0448	4.552
ACTIVA	-0.0039	4.774
HNOPOB	-0.0048	10.568

Fuente: Herrera (2012)

La variable CONVIV, que captura el efecto de que la mujer convive con una pareja a la fecha del estudio, tuvo un efecto positivo sobre la tasa de fecundidad. Por otro lado, las variables URBAN, EDUC, ACTIVA Y HNOPOB reflejaron que la mayor proporción de población urbana, mujeres más educadas en promedio o con mayor participación en el mercado laboral y mayor nivel económico afectaron 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 errores. Nuevamente, se recurre al mapa para visualizar la distribución de los mismos (Figura 9).

Figura 9 - Distribución Espacial de los Errores por MCO



Fuente: Herrera (2012)

La distribución residual representada por colores, aparentemente no tuvo un patrón aleatorio, por lo que se afirmó de manera empírica que la autocorrelación espacial está en la estructura del error

La aplicación de los estadísticos espaciales como el I de Moran y en forma más específica el test LM-ERR y su forma robusta LM-EL detectarán si la autocorrelación espacial está en la estructura del error. Los resultados se muestran en el Cuadro 6.

Cuadro 6 - Tests de Autocorrelación Espacial

Test	Matriz Contigüidad Tipo Reina		Matriz 4-Vecinos más Cercanos	
	Valor	P-value	Valor	P-value
I de Moran	0.41	0.0000	0.34	0.0000
LM-ERR	239.93	0.0000	147.67	0.0000
LM-EL	143.93	0.0000	85.67	0.0000

Fuente: Herrera (2012)

La prueba I de Moran resultó significativa, esto evidenció dependencia o autocorrelación espacial, pero sin información sobre el posible modelo espacial a seguir. Luego para verificar si la autocorrelación espacial se encontraba en la estructura del error, se procedió en utilizar los estadísticos de Multiplicador de Lagrange (LM). La prueba LM-ERR y su versión robusta LM-EL resultaron altamente significativas, que corrobora presencia de autocorrelación espacial en la estructura del error. Finalmente el uso de un modelo de error espacial (SEM) resultó adecuado

5.2.6 Análisis del Modelo Espacial Autorregresivo en el Error (SEM)

Previamente se comprobó que existe una autocorrelación espacial en la estructura del error. Por esta razón se utilizó un proceso autorregresivo en el error del modelo. En terminos sencillos es utilizar un rezago espacial en el error.

El modelo espacial autorregresivo en el error o modelo espacial del error (SEM) se muestra a continuación:

$$\left. \begin{aligned}
 HIJOS_i &= \beta_0 + \beta_1 CONVIV_i + \beta_2 URBAN_i + \beta_3 EDUC_i + \beta_4 ACTIVA_i + \beta_5 HNOPOB_i + \varepsilon_i \\
 \varepsilon_i &= \lambda [W\varepsilon]_i + \mu \\
 \mu &\sim N(0, \sigma^2 I)
 \end{aligned} \right\} \dots (19)$$

Para el ejemplo se obtuvo dos tipos de matriz de pesos espaciales (W), por lo que se estimarán dos modelos distintos.

Para verificar si hay efecto del coeficiente autorregresivo (λ) en el rezago espacial del error ($W\varepsilon$) se plantearon las siguientes hipótesis:

$$\left. \begin{array}{l} H_0: \lambda = 0 \\ H_1: \lambda \neq 0 \end{array} \right\} \dots (13)$$

Por lo que si se rechaza la H_0 , se puede decir que existe suficiente evidencia estadística para afirmar que el modelo (SEM) captura la autocorrelación espacial en la estructura del error.

En el Cuadro 7 se muestran las estimaciones de los parámetros con sus respectivas pruebas de significancia para cada modelo (contigüidad tipo reina y 4-vecinos más cercanos)

Cuadro 7 - Estimación de los Parámetros del Modelo SEM

Variable	Contigüidad tipo Reina		4-Vecinos más cercanos	
	Coefficiente Estimado	(t)	Coefficiente Estimado	(t)
CONSTANTE	1.1395*	11.83	1.2380	12.25
CONVIV	0.0158*	17.73	0.0145*	15.18
URBAN	-0.0004*	2.54	-0.0003*	1.92
EDUC	-0.0537*	5.78	-0.0595*	6.10
ACTIVA	-0.0034*	4.55	-0.0026*	3.30
HNOPOB	-0.0031*	6.55	-0.0037*	7.57
λ	0.7470*	22.46	0.6439*	17.09

“*” significa que el coeficiente estimado es significativo

Fuente: Herrera (2012)

Los resultados de la estimación del modelo SEM, evidenciaron en primera instancia que los signos de los coeficientes estimados de las variables explicativas resultaron iguales que los signos de los coeficientes estimados del modelo de regresión espacial (ecuación 18). Ahora, observando la significancia del coeficiente autorregresivo λ se pudo afirmar que este captura la dependencia espacial y en forma positiva.

En conclusión, el uso de un modelo espacial autorregresivo en el error resultó idóneo para este ejemplo de la fecundación en Argentina.

VI. CONCLUSIONES

1. Un modelo espacial autorregresivo tiene en su estructura la manera de capturar la autocorrelación espacial ya sea residual o sustantiva, lo que un modelo de regresión múltiple o de otro tipo no se puede controlar, ya que sus estructuras no son adecuadas para datos espaciales o geográficos.
2. Los parámetros estimados de un modelo espacial autorregresivo en el error son insesgados y consistentes. En cambio para el modelo espacial autorregresivo de primer orden y el modelo de rezago espacial sus parámetros estimados resultan sesgados.
3. Utilizando el ejemplo de la fecundidad en Argentina se logró describir la aplicación de la metodología para la utilización de un modelo espacial autorregresivo en el error. En primer lugar, se realizó un análisis exploratorio de la variable de interés (fecundidad) y sus variables explicativas. En segundo lugar se explicó la delimitación de la unidad espacial, para luego determinar la matriz de pesos o contactos. Como siguiente paso se explicó la aplicación de una prueba estadística que sugiere la presencia de la autocorrelación espacial global univariante. Luego, se describió el proceder de cómo es un análisis de regresión espacial y finalmente se realizó el análisis y aplicación del modelo espacial autorregresivo en el error.

VII. RECOMENDACIONES

1. Realizar la descripción metodológica del resto de modelos espaciales autorregresivos que se tienen actualmente.
2. Realizar un estudio que involucre la aplicación de los modelos espaciales autorregresivos con énfasis en el modelo espacial autorregresivo en el error en el contexto social y geográfico del Perú.

VIII. REFERENCIAS BIBLIOGRÁFICAS

- [1] Anselin y Florax (1995) “New Directions in Spatial Econometrics”.
Science Dissertation and Monograph Series, Ithaca, New York.
- [2] Daniel A. Griffith (2003). “Spatial Autocorrelation and Spatial Filtering”
- [3] Arthur Getis, Jared Aldstadt (2010). “Constructing the Spatial Weights Matrix Using A Local Statistic”
- [4] Moreno y Vallá (2004). “Econometría espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas”
- [5] Michael Tiefelsdorf (2000). “Modelling Spatial Processes: The identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran’s I”
- [6] Robert Haining (2003). “Spatial Data Analysis: Theory and Practice”
- [7] James P. LeSage (1998). “Spatial Econometrics”
- [8] Lesage, J. y K. Pace (2009). “Introduction to Spatial Econometrics. Boca Raton: CRC Press”
- [9] Luc Anselin (1988). “Spatial Econometrics: Methods and Models”
- [10] Boots, Okabe y Thomas (2002). “Modelling Geographical Systems”
- [11] Barnio, Vianco y Rabanal (2012). “Introducción a la Econometría Espacial”
- [12] Lancaster, (2000). “The incidental parameter problem since 1948”
- [13] Aroca, (200), “Econometría Espacial: Una Herramienta Para el Análisis de la Economía Regional”
- [14] Getis y Ord, (1996). “Local Spatial Statistics: An Overview, en Longley, P. y M.Batty (eds), Spatial Analysis: Modelling in a GIS Environment”
- [15] Getis y Ord, (1992). “The Analysis of Spatial Association by Use of Distance Statistics,”
- [16] Herrera, Paz y Cid (2012) “Introducción a la Econometría Espacial. Una Aplicación al Estudio de la Fecundidad en la Argentina usando R”

IX. ANEXOS

Anexo 1 – Datos y Códigos en R

Datos

El análisis empírico ha sido realizado usando datos del Censo 2001 (INDEC), disponibles por Partido y/o 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.

Código en R

```
## 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)
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.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)
```