

Aplicación de un modelo de econometría espacial a datos agregados de asistencia escolar en la Argentina

Juan Carlos Cid

1 - INTRODUCCIÓN

Los trabajos sobre Econometría Espacial experimentaron un notable incremento en los últimos años. Entre los factores que influyeron puede mencionarse el desarrollo de los sistemas de información geográfica (GIS); la mayor velocidad de procesamiento y capacidad de almacenamiento de las computadoras y la posibilidad de disponer de la ubicación en la superficie terrestre de las observaciones de los fenómenos (llamada geo-referenciación). A todo esto se agregó la creciente atención prestada por los investigadores a la evidencia de que los fenómenos socioeconómicos ocurren no solamente en determinado tiempo sino, además, en cierto espacio.

Sin embargo, en nuestro país han sido escasos los avances en esta materia, fundamentalmente porque la información estadística generada por el sistema oficial no suele estar geo-referenciada.

Desde el punto de vista de un modelo de regresión lineal que utiliza datos de sección cruzada, la cuestión espacial sería irrelevante si las observaciones fueran estadísticamente independientes¹. Pero existen muchas situaciones en las que el valor observado en un determinado punto o región geográfica depende de los valores que tienen las observaciones vecinas o próximas.

¹ En este documento no se considera un segundo caso de violación de los supuestos del modelo de regresión lineal que se presenta cuando existe heterogeneidad espacial, es decir que la relación funcional varía en el espacio.

Específicamente en el campo de la economía de la educación, los análisis de la decisión de un individuo de asistir a un establecimiento educativo incluyen normalmente variables que corresponden a determinadas características personales y del hogar: edad y sexo de la persona, ingreso per cápita familiar, nivel educativo alcanzado por los padres, cantidad de menores, presencia de libros, etc. A estos factores se agregan a veces otros ambientales, por ejemplo el área de residencia, suponiendo que la población urbana tiene mayor probabilidad que la rural de asistir a una institución educativa. Nótese que esa hipótesis introduce una dimensión espacial en la estimación del modelo.

Cabría preguntarse sin embargo si, a igualdad de todas las restantes variables explicativas, un joven residente en una zona rural del departamento La Caldera en Salta tiene la misma probabilidad de asistir a una universidad que otro que vive en la zona rural del departamento Rivadavia de la misma provincia. La respuesta probablemente será negativa, pues el primero se halla a escasos kilómetros, y pocos minutos de viaje, de la sede de la Universidad Nacional de Salta, mientras que para el segundo resulta imposible el traslado diario. La incorporación de variables ficticias por departamento o municipio, en la medida en que lo permita el número de observaciones, aparentemente resolvería el problema. Sin embargo, puede ocurrir que en un departamento de Formosa o de Chaco colindante con Rivadavia funcione una institución de enseñanza superior y vuelva a aparecer así la cuestión espacial.

En segundo lugar, además de la influencia que en la decisión de asistir ejerce el ambiente educativo hogareño –medido por medio de los años de instrucción formal del jefe de hogar o la cantidad de libros que hay en la vivienda–, es probable que el ambiente social en que se desenvuelve el joven –el vecindario, los amigos, las relaciones, etc. – también afecte a la decisión. La dificultad en tal caso radicaría en identificar una variable adecuada para cuantificarlo.

En este documento nos propusimos examinar la aplicación de un modelo sencillo de econometría espacial a los datos censales de asistencia a un establecimiento educativo de los jóvenes entre 20 y 24 años. En la siguiente sección se hace una breve explicación del marco teórico y de los datos utilizados. La tercera está destinada a presentar los resultados de la estimación del modelo y del uso de otras técnicas espaciales de análisis de los datos. Las conclusiones y las posibles líneas de investigación futuras se consignan en la cuarta sección. Todos los Cuadros y Figuras se incluyen en un Anexo final. Los mapas temáticos de la Argentina fueron confeccionados por Olver Figueroa, de la división cartográfica de la Dirección de Estadísticas de Salta, que también es responsable de los diversos procesos desarrollados con el programa ArcView.

2 - DATOS Y METODOLOGÍA

En nuestro análisis se usaron los datos del Censo Nacional de Población, Hogares y Viviendas de 2001 (CEN 2001), a los que se puede acceder en el sitio del Instituto Nacional de Estadística y Censos (INDEC) a través de un sistema de consulta con el software Redatam.

Consideramos la tasa de escolarización de los jóvenes comprendidos en el tramo entre 20 y 24 años, etapa en la que se encontrarían cursando estudios superiores. La variable está referida al indicador departamental, es decir la proporción de personas en esas edades que asistía a un establecimiento educativo respecto al total de la población, en cada uno de los 511 departamentos (llamados partidos en la provincia de Buenos Aires) en que está dividida la Argentina². Mencionemos que en el CEN 2001 había aproximadamente 3,2 millones de habitantes en ese grupo etario.

Los antecedentes sobre el tema en nuestro país habitualmente consideraron al individuo como unidad de observación, por ejemplo recurriendo a las bases de la Encuesta Permanente de Hogares. Estimaron entonces modelos donde la variable dependiente es dicotómica. Optamos aquí por emplear el departamento como unidad de observación porque en el sistema estadístico oficial existe una geo-referenciación del mapa del territorio argentino con su división departamental, pero no de cada uno de los hogares empadronados en un censo.

Respecto a las variables explicativas, utilizamos el nivel educativo del jefe de hogar, el grado de urbanización, el nivel de bienestar material, la tasa de desocupación y el tamaño de la población.

El nivel educativo del jefe de hogar tiene un importante poder explicativo en la decisión del joven de asistir a un establecimiento escolar. Al igual que para el resto de las variables, por medio del sistema de consulta remota de INDEC se estimó una medida departamental agregada: La proporción de jefes que como mínimo habían cursado algo del nivel medio. Es decir que se suman las personas con secundario incompleto y completo, terciario o universitario completo o incompleto. La hipótesis es que el coeficiente de esta variable es mayor que cero.

La *proxi* para el grado de urbanización en el departamento fue la proporción de hogares con provisión de agua corriente. En cuanto al bienestar material

² A los fines de nuestro estudio la Ciudad Autónoma de Buenos Aires se consideró como una sola jurisdicción. En la provincia de Tierra del Fuego, Antártida e Islas del Atlántico Sur se analizaron exclusivamente los departamentos de Río Grande y Ushuaia. No se incluyen las Islas Malvinas por carecerse de datos ni tampoco la población del continente antártico, por sus características especiales y escaso volumen.

promedio, se midió a través de la proporción de hogares no pobres según la metodología del Indicador de Privación Material de los Hogares (IPMH) diseñada por INDEC para su aplicación en el Censo 2001³. Los coeficientes estimados de estas dos variables debieran ser positivos.

La tasa de desocupación fue calculada específicamente en relación a la Población Económicamente Activa (PEA) en el tramo de 20 a 29 años de cada departamento. Aclaremos que el censo poblacional no es una fuente adecuada para estimar este indicador por varios motivos. El cuestionario censal no tiene la complejidad necesaria para determinar con precisión la condición de actividad de una persona y el censista carece del suficiente entrenamiento para indagar en esta temática. La situación del país hacia fines de 2001 tampoco contribuyó para obtener una buena medición. INDEC señaló que los resultados del relevamiento en general subestimaron la PEA y sobreestimaron la desocupación. A pesar de estas serias limitaciones, tuvimos que recurrir a los datos censales porque la Encuesta Permanente de Hogares, una fuente alternativa mucho más confiable, cubre solamente a la población urbana de la Argentina y no alcanza la desagregación utilizada en nuestro estudio. Respecto al efecto de esta variable explicativa sobre la matriculación, no es claro: En principio, al reducir el costo de oportunidad de destinar el tiempo a la educación, el signo debería ser positivo. Sin embargo la tasa de desocupación del tramo de 20 a 29 años tiene alta correlación con la tasa general para todas las edades. Un valor elevado reduciría las posibilidades de los hogares de financiar la educación de los jóvenes por lo que, de prevalecer esta fuerza, el coeficiente estimado para la variable resultaría negativo.

Por último, también se consideró la cantidad de población (en miles) comprendida en el tramo de edades de 20 a 24 como una variable explicativa del modelo. Nuestra hipótesis es que cuanto mayor sea aquel número, más justificación tiene el funcionamiento de instituciones de nivel superior en la jurisdicción, por lo que su coeficiente sería positivo. Aun en presencia de otras condiciones favorables (elevado nivel educativo de los jefes, alto grado de urbanización, etc.), la población de jóvenes de un departamento debe alcanzar cierto nivel crítico para que se plantee la apertura de un establecimiento de esas características.

Sin embargo, existiría también una causalidad en el sentido inverso. En departamentos como Capital, en la provincia de Córdoba, o La Plata, en la provincia de Buenos Aires, no solo hay universidad porque viven muchos jóvenes sino que además hay muchos jóvenes (una parte de ellos, provenientes de otros lugares) porque funciona allí una universidad. En cambio en los dos departamentos de Tierra del Fuego –Ushuaia y Río Grande– se observó una baja tasa de matriculación. La explicación residiría en que los

³ La metodología del Indicador de Privación Material de los Hogares puede ser consultada en <http://www.indec.gov.ar/nuevaweb/cuadros/7/sesd-metod-ipmh.pdf>

jóvenes que decidieron continuar una carrera universitaria emigraron a otras provincias. Entonces, los empadronados en Tierra del Fuego fueron los que se habían quedado porque no estudiaban.

El Modelo Espacial General ⁴ se puede escribir de la siguiente forma:

$$y = \rho W_1 y + X\beta + u \quad (1)$$

$$u = \lambda W_2 u + \varepsilon \quad (2)$$

$$\varepsilon \sim N(0, \sigma_\varepsilon^2 I_n) \quad (3)$$

Donde y es un vector columna de n observaciones de la variable dependiente; X es una matriz $n \times k$ con los valores observados de las k variables explicativas; u y ε son los vectores de términos de error de las respectivas ecuaciones; el vector columna β y los escalares ρ y λ son parámetros a ser estimados.

Las matrices W_1 y W_2 , ambas de dimensión $n \times n$, se denominan de ponderaciones espaciales. La matriz W_1 refleja la autocorrelación espacial que afecta a y . Es decir que sobre el valor de una observación de la variable explicada influyen, además de los valores que tienen las variables explicativas, los asumidos por y en las observaciones vecinas. El término de error de la ecuación principal también está autocorrelacionado espacialmente y esa influencia se muestra por medio de W_2 .

LeSage (1999) menciona dos razones por las que puede ocurrir que el dato observado en un punto del espacio dependa de los valores encontrados en otras ubicaciones. La primera es que se refleje allí un error de medición originado en que las divisiones geográficas adoptadas en la recolección de la información (por ejemplo, los departamentos provinciales) no se corresponden con la naturaleza del proceso subyacente. En segundo lugar, una cuestión más relevante sería que existe una dimensión espacial en la esencia del problema que se analiza. Por ejemplo, la presencia de interacción espacial, de efectos de difusión o de externalidades espaciales.

Como se explicó previamente, y en nuestra estimación es la proporción, en tanto por ciento, de las personas de 20 a 24 años que asistían a un establecimiento educacional a la fecha del censo y n es 511, el número de jurisdicciones en que dividimos el territorio nacional.

Señalemos que uno de los criterios más difundidos de vecindad, y el elegido en este trabajo para determinar la matriz de ponderaciones espaciales, es la contigüidad. Se definen como contiguos los territorios (departamentos en

⁴ En esta parte del documento se sigue el desarrollo y la simbología de LeSage (1999).

nuestro caso) que tienen una porción de límite en común: En la celdilla de la fila i y de la columna j de la matriz inicial $n \times n$ hay un 1 cuando el departamento i es vecino al departamento j ⁵. El resto de las celdas se completa con ceros. Posteriormente, esa matriz inicial se normaliza haciendo que la suma de los elementos sea igual a uno en todas sus filas.

En nuestro caso, partimos de un mapa geo-referenciado de la Argentina con división política. Con el programa ArcView se obtuvo un archivo digital con la latitud y la longitud de los 511 puntos que constituyen el centro geográfico de cada uno de los departamentos. Las coordenadas a su vez se utilizaron como insumos para construir con el programa MatLab una matriz de contigüidad W . El procedimiento en MatLab se denomina **xy2cont** e integra el paquete de programas de Econometría Espacial, desarrollado por LeSage, que está disponible en el sitio www.spatial-econometrics.com.

La contigüidad también puede establecerse trabajando con un mapa del territorio y cargando los contactos en una hoja de cálculo. Este procedimiento manual es laborioso y no ofrecería ventajas significativas en comparación con la rutina establecida en el *software*. En la siguiente sección se presenta una comparación entre ambos sistemas para una región, el noroeste argentino.

No existen razones a priori para sostener que en nuestro modelo haya dos matrices W_1 y W_2 distintas. Sin embargo se discuten, también en el apartado de la comprobación empírica, distintas alternativas.

Volviendo a las ecuaciones 1 a 3 del modelo espacial, obsérvese que si ρ y λ no son significativamente distintos de cero el planteo se reduce al modelo lineal de mínimos cuadrados ordinarios:

$$y = X\beta + u \quad (4)$$

$$u \sim N(0, \sigma^2_u) \quad (5)$$

Si ρ es igual a cero, en un mapa temático con los valores de y para las 511 observaciones se observará una distribución geográfica aleatoria de los mismos. En cambio, si existe correlación espacial en la variable dependiente se hallarán *clusters* de jurisdicciones próximas entre sí que presentaron valores similares.

Para λ valen las mismas consideraciones, en este caso referidas a los residuos de la regresión lineal.

⁵ La relación de vecindad es recíproca. Se supone que ninguna observación es vecina de sí misma, de manera que la diagonal principal está ocupada por ceros.

Uno de los estadísticos más usados para detectar la autocorrelación espacial de los residuos de un modelo de mínimos cuadrados es el I Moran, que se calcula con la siguiente fórmula:

$$I = u' W u / u' u \quad (6)$$

Donde u es el vector columna de n residuos de la regresión del modelo lineal y W una matriz de ponderaciones espaciales.

El índice de Moran, cuando se lo estandariza restándole la media y dividiendo por el desvío estándar, se distribuye asintóticamente según la normal.

3 - RESULTADOS OBTENIDOS

En primer lugar, la Figura 1 del Anexo permite comparar dos matrices de contigüidad construidas para los 117 departamentos de las seis provincias de la región del noroeste. La que denominamos generada es la que se obtuvo al aplicar la función **xy2cont** al archivo de latitudes y longitudes de los puntos que son el centro geográfico de cada jurisdicción. La denominada real es la diseñada luego de examinar los mapas provinciales para establecer cuáles eran los departamentos (polígonos) que tenían un tramo de límites en común y grabar en un archivo los vecinos de cada uno. Obvia decir que la tarea es tediosa y demanda bastante tiempo completar en este caso las 13.689 celdas de una planilla de 117 filas y columnas.

Como los departamentos están ordenados dentro de cada provincia, claramente se forman 6 nubes de puntos sobre la diagonal principal porque las vecindades se dan predominantemente entre jurisdicciones de una misma provincia. Además las provincias están ordenadas alfabéticamente, por eso las nubes a medida que nos desplazamos de la celda (1;1) hasta la (117;117) identifican sucesivamente a Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán. Las contigüidades que se establecen lejos de la diagonal corresponden a departamentos de los bordes. Por ejemplo, Santa Rosa, provincia de Catamarca, comparte límites con Choya y Guasayán, de Santiago del Estero, y con Graneros y La Cocha, de Tucumán.

Las contigüidades reales que determinamos luego de examinar los mapas fueron 592 (4,3% de los casos posibles de las 13.689 celdas). En promedio, cada departamento tiene 5 vecinos. En cambio, el procedimiento estimó un número mayor, 672 (4,9% de casos). Los errores por exceso aparecen en la Figura 1 como un círculo vacío, los errores por defecto con el símbolo de suma y las coincidencias se denotan por el "+" inscripto en la circunferencia. Se

produjeron 518 coincidencias (es decir que el procedimiento algorítmico pronosticó una vecindad que efectivamente existía en el terreno), 74 contigüidades detectadas por nosotros en base a los mapas no fueron reconocidas por la función y, en sentido inverso, ésta generó 154 contactos “falsos”⁶.

En base a esta experiencia, y a que la determinación manual de la matriz cuadrada de 117 filas demandó toda una jornada de trabajo mientras que el procedimiento **xy2cont** solo una fracción de segundo, se decidió utilizar este último para el cálculo de la matriz de ponderaciones de todo el país, que posee 511 filas y columnas y 261.121 celdas.

El Cuadro 1 del Anexo muestra estadísticos descriptivos de las variables departamentales⁷. Mencionemos que once departamentos tenían una tasa de asistencia inferior a 5%. Tres pertenecían a La Pampa (Lihuel Calel, Chical Có y Limay Mahuida), otros tres a Jujuy (Rinconada, Susques y Valle Grande), y el resto a Catamarca (Ancasti), Chubut (Mártires), Santiago del Estero (Mitre), Formosa (Bermejo) y Neuquén (Catán Lil). Los porcentajes más elevados se hallaron en 7 jurisdicciones que superaron 45%: los departamentos Capital de las provincias de Córdoba, Mendoza y San Juan, la Ciudad Autónoma de Buenos Aires tomada en bloque y los partidos de La Plata, San Isidro y Vicente López, en la provincia de Buenos Aires. Además tenían en general valores relativamente altos todos los departamentos donde se asienta una capital provincial (o próximos a ella, como Palpalá en Jujuy y Yerbabuena en Tucumán) o una sede universitaria tradicional (como Tandil y Bahía Blanca en Buenos Aires y Río Cuarto en Córdoba).

El modelo que estimamos inicialmente fue el de las ecuaciones 4 y 5 de la sección anterior. Se reitera que este sería el método válido si ρ y λ no fueran significativamente distintos de cero. Los resultados de la estimación lineal por mínimos cuadrados ordinarios se muestran en el Cuadro 2.

Todos los coeficientes tienen los signos esperados y son estadísticamente significativos. El de la tasa de desempleo señalaría que, de las dos alternativas discutidas en la sección anterior, el sistema educativo estaría actuando como

⁶ La función **xy2cont** aplica un algoritmo basado en la triangulación de los puntos centrales. Las provincias con formas muy irregulares de división política (Catamarca, Salta) originan más errores que aquéllas que se subdividen en polígonos más regulares (La Pampa, Chubut). Por otra parte, existe cierto tipo de errores del procedimiento que cabría catalogar como aciertos. Por ejemplo, puede ocurrir que se califique como contiguos a dos departamentos próximos entre sí pero separados por una estrecha franja de tierra perteneciente a un tercero.

⁷ Debe recordarse que se está considerando una muestra compuesta por 511 observaciones departamentales. Las medias del Cuadro 1 son los promedios de esas observaciones, sin ponderar por su tamaño. Así por ejemplo la media de la tasa de asistencia es 20,6% (el promedio simple de las tasas de los 511 departamentos) pero la global calculada sobre los 3,2 millones de jóvenes es 30,8%.

sustituto de una incorporación más temprana de los jóvenes a la fuerza de trabajo.

Como ya se ha mencionado antes, en ausencia de correlación espacial de la variable dependiente se observará una distribución geográfica aleatoria de los valores de las observaciones.

En el Anexo la Figura 2 es el mapa de la Argentina con los valores de la tasa de asistencia departamental. Vemos que los departamentos con valores similares tienden a estar agrupados geográficamente. En términos generales, si y_i e y_j son departamentos contiguos, es probable que la tasa de escolaridad de sus jóvenes sea del mismo orden de magnitud.

En la Figura 3 se graficaron, siempre sobre la división política del territorio argentino, los residuos de la regresión lineal según que sean valores positivos o negativos. Lejos de distribuirse aleatoriamente, los del mismo signo conforman *clusters* geográficos, lo que estaría abonando la hipótesis de la existencia de correlación espacial de los errores.

El test de Moran permite verificar esta posibilidad. En el Cuadro 3 del mismo Anexo se presentan los resultados obtenidos al aplicarlo sobre los residuos del modelo lineal. Como el valor estandarizado es 2,44 se puede rechazar la hipótesis nula de ausencia de autocorrelación de los disturbios al nivel de significación de 5%.

La estimación del Modelo Espacial General arrojó los resultados que muestra el Cuadro 4. Todos los coeficientes de las variables explicativas mantuvieron su significación estadística, con alteraciones relativamente pequeñas en su valor. Los parámetros que más se modificaron en comparación el modelo de mínimos cuadrados ordinarios fueron el de la proporción de población no pobre y el término constante. Los coeficientes ρ y λ son significativos y evidencian una correlación positiva tanto para la variable explicada como para el término de error.

En la estimación utilizamos como matriz W_1 de retardos espaciales de y a la de contigüidad de primer orden generada por el procedimiento *xy2cont*. Como W_2 , la matriz de correlación espacial de los errores, se adoptó la de segundo orden. Un retardo espacial de segundo orden significa que la definición de contigüidad se extiende a los vecinos de los vecinos. Aplicar ese criterio al modelo implica suponer que la estructura de los disturbios involucra una dependencia espacial de orden superior, por ejemplo que exista una segunda onda de efectos en el fenómeno que se está analizando⁸.

Por supuesto, carecemos de una fundamentación teórica para adoptar ese criterio. Por eso se probó aplicar en forma inversa las matrices de retardos

⁸ Véase LeSage (1999), páginas 87-88.

espaciales (es decir la de segundo orden en la ecuación 1 y la de primer orden en la ecuación de los errores) y también la misma matriz de primer orden para ambas ecuaciones. Los resultados de esas estimaciones se presentan en los Cuadros 5 y 6.

No surgen mayores diferencias en los coeficientes de las variables explicativas que, por otra parte, conservan su significación. Los efectos del cambio entre una y otra alternativa se observan, como era lógico esperar, en los parámetros ρ y λ .

4 - CONCLUSIONES

Como se aclaró en la Introducción del documento, el principal propósito de nuestra investigación era probar la utilización de técnicas de econometría espacial en datos censales del tema educativo. En el sistema estadístico oficial argentino aún no se cuenta con información geo-referenciada con el grado de detalle que sería deseable. Probablemente en un futuro próximo se podrá conocer la ubicación precisa de cada una de las observaciones sobre el territorio, tanto en los censos de población como en los económicos y en otros operativos. El tema suscitará seguramente controversia porque plantea interrogantes acerca del resguardo del secreto de los datos individuales que es tradición en todas las oficinas estadísticas.

Con la limitación entonces de haber restringido el análisis a las variables agregadas por departamento, la conclusión más interesante es que en la determinación de la tasa de asistencia a un establecimiento educativo para el tramo de edades considerado se detecta la interdependencia espacial. La estimación de un modelo espacial incluyendo interdependencia en la variable y así como autocorrelación de los errores arrojó coeficientes significativos para los parámetros ρ y λ , que ponderan los respectivos retardos. Además, el test de Moran para los residuos del modelo lineal ordinario permite rechazar, con probabilidad de error inferior a 1,5%, la hipótesis nula de ausencia de correlación espacial. Sin embargo, debe reconocerse que ese modelo no mejora radicalmente la bondad de ajuste del lineal simple ni tampoco altera en gran proporción las estimaciones de cada uno de los parámetros ni su significación estadística.

Los resultados de las estimaciones con distintas alternativas de matrices de ponderaciones espaciales de primero y segundo orden -contenidos en los Cuadros 4, 5 y 6- destacan la necesidad de ahondar en el criterio de definición de la vecindad entre observaciones, el que estará vinculado a la naturaleza del problema estudiado. En el tema educativo es posible que la simple contigüidad

entre jurisdicciones no alcance y sea necesario considerar distancias y medios de transporte y comunicación. Creemos que en el modelo existe una variable omitida, porque habría que considerar de alguna manera las ubicaciones de los centros de estudios superiores existentes. Incluir la distancia y dificultad de acceso desde cada departamento a la entidad educativa más próxima posiblemente aumentaría el poder explicativo del modelo. Quizás en ese caso los parámetros ρ y λ dejarían de ser significativos.

Referencias bibliográficas

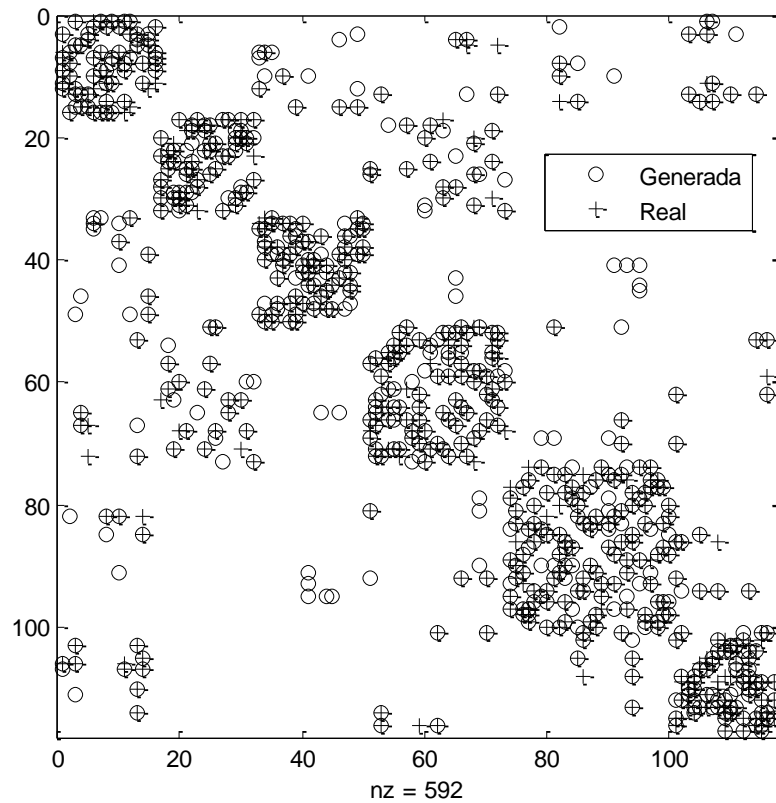
INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS. Censo Nacional de Población, Hogares y Viviendas 2001. En el sitio www.indec.gov.ar

LESAGE, J. P. (1999). The Theory and Practice of Spatial Econometrics. En el sitio www.spatial-econometrics.com

LESAGE, J. P. y PACE, R. K. (2009). Introduction to Spatial Econometrics. CRC Press. Boca Raton.

ANEXO DE CUADROS Y FIGURAS

Figura 1: Comparación de la matriz de contigüidad real con la generada por el procedimiento del programa MatLab. Departamentos de la Región NOA



**Cuadro 1: Descriptivos de las variables departamentales
República Argentina, 2001**

Variable	Mínimo	Máximo	Media	Desvío típico
Asistencia a est. educativo	0,0	54,5	20,6	8,8
Nivel educativo de jefes	5,5	73,8	30,6	12,5
Hogares con agua de red	0,0	100,0	78,6	20,0
Hogares no pobres	0,6	72,0	37,3	17,0
Tasa de desocupación	3,2	65,1	29,9	9,0
Jóvenes de 20 a 24 años (miles)	0,04	234,3	6,3	16,0

Cuadro 2: Estimación del modelo lineal convencional

Variable Dependiente: Tasa de asistencia por departamento

Variable explicativa	Coefficiente	estadístico t
Constante	-5,466	-5,023 (***)
Nivel educativo de Jefes	0,384	14,534 (***)
Hogares con agua de red	0,041	3,398 (***)
Hogares no pobres	0,106	6,039 (***)
Tasa de desocupación	0,230	9,399 (***)
Jóvenes de 20 a 24 años	0,057	3,891 (***)
R ²	0,722	
R ² corregido	0,720	
σ ²	21,784	
F	262,931	
Durbin-Watson	1,553	

*** significación al 1%

Cuadro 3: Test de I Moran para la correlación espacial de los residuos

I Moran	0,05939
Media de I	-0,00320
Desvío estándar de I	0,02566
Estadístico I Moran (estandarizado)	2,43989
Probabilidad marginal	0,01469 (**)

** significación al 5%

Cuadro 4: Estimación de un modelo espacial general

Variable Dependiente: Tasa de asistencia por departamento

W₁ es la matriz de contigüidad de primer orden

W₂ es la matriz de contigüidad de segundo orden

Variable explicativa	Coefficiente	estadístico t (asintótico)
Constante	-6,791	-6,260 (***)
Nivel educativo de Jefes	0,382	14,637 (***)
Hogares con agua de red	0,041	3,422 (***)
Hogares no pobres	0,097	5,582 (***)
Tasa de desocupación	0,231	9,469 (***)
Jóvenes de 20 a 24 años	0,058	3,939 (***)
ρ	0,078	39,148 (***)
λ	0,107	26,191 (***)
R ²	0,726	
R ² corregido	0,724	
σ ²	21,237	
Log-likelihood	-1329,161	

Número de iteraciones 23

*** significación al 1%

Cuadro 5: Estimación de un modelo espacial general (primera variante)

Variable Dependiente: Tasa de asistencia por departamento

W_1 es la matriz de contigüidad de segundo orden

W_2 es la matriz de contigüidad de primer orden

Variable explicativa	Coefficiente	estadístico t (asintótico)
Constante	-8,999	-8,382 (***)
Nivel educativo de Jefes	0,379	14,602 (***)
Hogares con agua de red	0,040	3,377 (***)
Hogares no pobres	0,097	5,562 (***)
Tasa de desocupación	0,236	9,719 (***)
Jóvenes de 20 a 24 años	0,055	3,813 (***)
ρ	0,190	35,954 (***)
λ	0,064	35,389 (***)
R^2	0,729	
R^2 corregido	0,726	
σ^2	21,025	
Log-likelihood	-1326,890	
Número de iteraciones	43	

*** significación al 1%

Cuadro 6: Estimación de un modelo espacial general (segunda variante)

Variable Dependiente: Tasa de asistencia por departamento

W_1 y W_2 son la matriz de contigüidad de primer orden

Variable explicativa	Coefficiente	estadístico t (asintótico)
Constante	-6,781	-6,255 (***)
Nivel educativo de Jefes	0,383	14,710 (***)
Hogares con agua de red	0,041	3,470 (***)
Hogares no pobres	0,099	5,654 (***)
Tasa de desocupación	0,230	9,416 (***)
Jóvenes de 20 a 24 años	0,058	3,947 (***)
ρ	0,074	38,087 (***)
λ	0,072	37,599 (***)
R^2	0,726	
R^2 corregido	0,724	
σ^2	21,233	
Log-likelihood	-1329,150	
Número de iteraciones	33	

*** significación al 1%