

Mediciones de Pobreza: Pobreza Energética y Necesidades Básicas Insatisfechas. Un Caso de Estudio desde la Econometría Espacial. Parte 2: Autocorrelación Espacial Global

Poverty Measurements: Energy Poverty and Unsatisfied Basic Needs. A Case Study from Spatial Econometrics. Part 2: Global Spatial Autocorrelation

Presentación: 13 y 14 de septiembre de 2023

Gustavo Schweickardt

CONICET-Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay
gaocerin@gmail.com

Resumen

En este trabajo se presentan los pasos para identificar la existencia de Autocorrelación Espacial Global en las dos variables que son consideradas en la Primera Parte: El Indicador de Dificultad de Acceso a la Energía (IDAENE), vinculado a la Pobreza Energética, y el Indicador de Necesidades Básicas Insatisfechas (NBI), para cada Partido de la Provincia de Buenos Aires. Este procedimiento puede ubicarse en la frontera que separa el Análisis Exploratorio del Análisis Confirmatorio de los Datos Espaciales (AEDE y ACDE, respectivamente), puesto que se intenta confirmar, si los patrones espaciales observados por inspección visual en los Mapas Coropléticos de ambas variables, son estadísticamente significativos. Desde la teoría, son considerados, para tal fin, estadísticos basados en los datos registrados para sendas variables, y en una Matriz que define la Estructura de Vecindad de cada Unidad Espacial (Partido), los cuales son sometidos a una prueba de hipótesis.

Palabras clave: Autocorrelación Espacial Global, Análisis Exploratorio de Datos Espaciales, Índice y Test de Moran

Abstract

This paper presents the steps to identify the existence of Global Spatial Autocorrelation in the two variables considered in the First Part: The Energy Access Difficulty Index (EADI), associated with Energy Poverty, and Unsatisfied Basic Needs (UBN), for each County of the Province of Buenos Aires. This procedure can be set on the 'border' between the Exploratory from the Confirmatory Analysis of Spatial Data (ESDA and CSDA, respectively) since it tries to confirm whether the spatial patterns observed by visual inspection in the Choropleth Maps of both variables are statistically significant. For this purpose, considering the theory, indices based on the data recorded for each variable and on a Matrix that defines the Neighborhood Structure of each Spatial Unit (County) are contrasted through a hypothesis test.

Keywords: Global Spatial Autocorrelation, Exploratory Spatial Data Analysis, Moran's Index and Test

Introducción

Las técnicas de Análisis Exploratorio de Datos Espaciales incluyen dos Efectos Espaciales: la Dependencia Espacial o Autocorrelación Espacial y la Heterogeneidad Espacial (Anselin, 1988). La Dependencia o Autocorrelación Espacial, efecto de interés para la investigación que aquí se reporta, surge de la existencia de una relación entre lo que sucede en un lugar determinado y lo que sucede en otro punto considerado 'vecino' (Cliff y Ord, 1973).

Las Medidas Globales, como es el caso de la Autocorrelación Global, utilizan la información completa del conjunto de datos, con el fin de obtener un valor promedio para todo el espacio geográfico. En cambio, las Medidas Locales, examinan la Dependencia Espacial en un subconjunto de datos definidos, con respecto al valor asumido en cada localización y su entorno local, generándose un valor para cada localización (Paelinck y Klaassen, 1979).

El efecto de Autocorrelación Espacial, se soporta, en gran medida, sobre lo que Tobler (1979) formuló bajo la denominación de "Primera Ley de la Geografía", la cual establece que "Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas entre sí, que con las cosas distantes".

En los conceptos vertidos a continuación y en el desarrollo del trabajo reportado, a excepción de que se indiquen otros autores, se sigue a Herrera Gómez (2015).

En primer término, debe ser introducido el concepto de Matriz de Pesos Espaciales, la cual permite operacionalizar el término 'vecino' y 'Estructura de Vecindad', para el análisis general de los Efectos Espaciales.

Esta Matriz, que se indicará como W , describe la conectividad entre n unidades que se encuentran localizadas en un espacio bidimensional.

La construcción de W está basada en, al menos, dos supuestos fundamentales relativos a la Estructura Espacial:

1) Un Criterio de Conectividad, que define qué Unidades Espaciales pueden ser consideradas 'vecinas' entre sí, y

2) Un supuesto de Pesos Espaciales, que operacionaliza cómo los 'vecinos' se afectan entre sí.

La Matriz W , resultará ser Cuadrada de orden n , siendo n el número de Unidades Espaciales (149 x 149, para el caso aquí analizado). Cada elemento de W es denominado Peso Espacial, w_{ij} . Por convención, ninguna región o unidad espacial puede interactuar consigo misma, de forma que $w_{ii} = 0$ (diagonal principal de W).

Se tienen diferentes Criterios para definir W . Realizando una breve síntesis introductoria para el objetivo aquí pretendido, se consideran solamente Criterios Geográficos, y en particular dos:

a) Criterio de Contigüidad: Dados los Polígonos que representan cada Unidad Espacial (Partidos de la Provincia de Buenos Aires - incluida la Ciudad Autónoma de Buenos Aires-), estos serán 'vecinos' si comparten un límite. En cuanto a la estructura para definir la 'vecindad' de cada Unidad Espacial, existen diferentes Tipos. El más utilizado es el Tipo 'Reina'; el tipo 'Alfil' y el tipo 'Torre' se utilizan menos (la Estructura de Vecindad, en estos casos, se conforma según los movimientos posibles, para una de esas 'piezas', en una posición, conforme el juego del Ajedrez). Para el objetivo del presente trabajo, se considerará una Matriz de Pesos Espaciales, bajo el Criterio de Contigüidad, Tipo 'Reina'. Los Pesos Espaciales utilizados son binarios ($w_{ij} = 1 \rightarrow$ las unidades i y j son vecinas; $w_{ij} = 0$, no lo son).

b) Criterio de Distancia - k Vecinos más Cercanos: Utilizando la información de los Centroides de cada Polígono que representa las Unidades Espaciales, se va eligiendo como 'vecino' al Centroide más cercano hasta obtener el número de 'vecinos' establecidos, k . Bajo este criterio todas las regiones poseerán la misma cantidad de 'vecinos', evitando el problema de unidades aisladas o unidades con excesiva cantidad de 'vecinos'. Para el objetivo de este reporte, se considerará, también, una Matriz de Pesos Espaciales bajo el Criterio k vecinos más cercanos, donde k resulta del valor promedio (o próximo a él) de 'vecinos' por Unidad Espacial (Partido) que arroje la construcción de la W por el Criterio de Contigüidad Tipo 'Reina'. Los Pesos Espaciales utilizados, también son binarios.

Definidos los Pesos Espaciales, lo habitual es trabajar con alguna transformación de la Matriz W debido a que mejora las propiedades estadísticas de los estimadores. La transformación más utilizada es la Normalización por Fila.

La ecuación (1) expresa genéricamente una Matriz de Pesos Espaciales, y la ecuación (2) la Normalización por Fila de sus elementos ($W^* \rightarrow$ Matriz de Pesos Espaciales Normalizada por Filas).

$$W = \begin{bmatrix} 0 & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & 0 \end{bmatrix} \quad (1)$$

$$w_{ij}^* = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}} ; \sum_{j=1}^n w_{ij}^* = 1 \quad (2)$$

El Índice de Moran:

El contraste más ampliamente utilizado para la medida de Autocorrelación Espacial Global, es el estadístico denominado I de Moran (Índice de Moran), introducido por Moran (1950). Se define mediante la siguiente expresión:

$$I = \frac{n}{S_0} \times \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} ; S_0 = \sum_i \sum_j w_{ij} \quad (3)$$

Donde: \bar{y} es la media de los valores y_i que adopta la variable en estudio, y , y n es el número de observaciones. Las sumatorias se expanden desde 1 a n .

En el restudio aquí reportado, se emplea otro contraste utilizado con frecuencia: es el estadístico c propuesto por Geary (1954). Se define mediante la siguiente expresión:

$$c = \frac{n-1}{S_0} \times \frac{\sum_i \sum_j w_{ij} (y_i - y_j)^2}{\sum_i (y_i - \bar{y})^2} \quad (4)$$

Se tiene un tercer contraste de uso frecuente, pero que aquí no se considera: el estadístico G propuesto por Getis y Ord (1992), y referido como G de Getis, en alusión al primer autor. La aplicación de este contraste, requiere que la Matriz de Pesos Espaciales tenga como entradas w_{ij} valores binarios. Se define mediante la expresión:

$$G = \frac{\sum_i \sum_{j \neq i} w_{ij} y_i y_j}{\sum_i y_i y_j} \quad (5)$$

Estos contrastes intentan captar la Autocorrelación Espacial, en este caso Global, combinando en su construcción dos tipos de similitudes:

1.- Similitud de Valores o Atributos: se refiere a la posibilidad de expresar el estadístico correspondiente como una función, $f(y_i, y_j)$, del valor que asume la variable en estudio, y , en dos localizaciones diferentes, (i y j), considerando si tales valores son '(muy) parecidos' o son '(muy) diferentes';

2.- Similitud de Localización: referida a los Pesos Espaciales de la matriz W considerada, contemplando si las Unidades Espaciales, i y j , están cerca o están alejadas, mediante el peso espacial consignado (w_{ij}).

Se describe a continuación, el Test e I de Moran, sin pérdida de generalidad y por límites de espacio en el manuscrito, ya que es el más empleado.

El estadístico I , asume valores entre -1 y 1. La Autocorrelación, entonces, puede ser positiva o negativa. Se demuestra que el valor esperado del Índice de Moran es una variable aleatoria cuya esperanza es $(-1/(n-1))$. Entonces, en rigor, si $I \text{ Moran} > (-1/(n-1)) \rightarrow$ Autocorrelación Positiva y si $I \text{ Moran} < (-1/(n-1)) \rightarrow$ Autocorrelación Negativa.

El Índice de Moran presenta el inconveniente, para saber si es o no significativo, que su Distribución de Probabilidad es desconocida, pues, valga la redundancia, para poder hacer inferencia y saber si la Autocorrelación eventualmente detectada es significativa, se necesitaría conocer la Distribución de Probabilidad del estadístico I . A los efectos de salvar esta situación y realizar el Test, se consideran dos alternativas: 1ra) La generación de una 'Distribución de Referencia' y 2da) el Empleo de Resultados Asintóticos.

1ra) Es posible generar una 'Distribución de Referencia', a partir de una aproximación computacional que se denomina "Test de Permutaciones" (aleatorización), permutando los valores R veces en el espacio, y computando el valor del I de Moran (en las aplicaciones, $R = 500$ permutaciones suele ser un valor suficiente). Con R permutaciones son calculados R (por ejemplo 500) I de Moran, uno para cada permutación. A partir de esos R valores de I , se construye la 'Distribución de Referencia', la cual resultaría la distribución empírica del Índice de Moran bajo aleatorización.

Esa sería, entonces, la distribución que debería tener el estadístico I si no existiera un Patrón Espacial, Autocorrelación Espacial Global, para el caso. Dicho de otro modo: en ausencia de Autocorrelación Espacial Global, los valores se distribuirían aleatoriamente en todo el espacio geográfico.

Entonces, el Test de Permutaciones considera tal 'Distribución de Referencia', suponiendo que la Hipótesis Nula, H_0 , es cierta, siendo H_0 : 'Existe Aleatoriedad, es decir que 'No Existe Autocorrelación Espacial Global en la variable considerada', caso contrario, se rechaza H_0 aceptando la Hipótesis Alternativa, H_1 : 'No Existe Aleatoriedad, es decir que Existe Autocorrelación Espacial Global en la variable considerada'.

El Test arroja un valor llamado 'Pseudo p ', por la forma de construcción de la 'Distribución de Referencia'. Operacionalmente se trata como un p -valor, y si el mismo resulta menor que un cierto nivel de significancia (se adopta 0,05), se rechaza H_0 , y se acepta que existe Autocorrelación Espacial Global en la variable considerada.

2da) El Empleo de Resultados Asintóticos, supone que se está considerando un número importante de datos (para el caso en análisis son suficientes). Bajo Hipótesis Nula de Aleatoriedad, H_0 , la Distribución Asintótica del I de Moran es Normal (o converge a una Distribución Normal).

Sobre los Diagramas de Dispersión de Moran: Las letras en los Cuadrantes, son las iniciales en inglés de 'High' → 'Alto', y 'Low' → 'Bajo': 1er Cuadrante: 'H-H' → 'Alto-Alto'; 2do Cuadrante: 'L-H' → 'Bajo-Alto'; 3er Cuadrante: 'L-L' → 'Bajo-Bajo'; y 4to Cuadrante: 'L-H' → 'Alto-Bajo', refiriéndose a los valores de la variable en estudio, y , y a su rezago espacial, $W y$.

Estos cuadrantes se definieron en base al número cero. Vale observar este punto, pues aparecen números negativos y positivos tanto en el eje horizontal (variable y) como en el vertical (rezago espacial $W y$), y esto no significa que se tengan tales valores negativos en la variable o en su rezago espacial, puesto que no es lógico.

Como en la expresión (1) se opera con desvíos respecto de la media de y , el cero no quiere decir que la variable y asuma el valor cero, sino que la referencia se ubica tal media.

Lo mismo puede decirse respecto del eje vertical: se representan los valores de los rezagos espaciales, considerando la normalización por fila de la Matriz W , y que, multiplicando la misma por el vector de observaciones, y , se obtiene, en cada componente i -ésimo del vector de rezagos espaciales ($n \times 1$), $W y$, y cada componente de este vector es el valor promedio de y en el vecindario de la Unidad Espacial que representa, entonces, también en el eje vertical se opera en relación a la media (valor promedio).

Se menciona, por último, respecto del estadístico c Geary, que resultado de c asume valores entre 0 y 2; no asume valores negativos. El valor esperado del Índice de Geary es 1. Siguiendo la misma lógica que para el I de Moran, hay que interpretar el c de Geary (pues el hecho de que no asuma valores negativos, no implica que no exista Autocorrelación Espacial Negativa). La siguiente conclusión tiene que ver con la expresión (4) para el cálculo de c : si c asume valores superiores a 1, es decir entre 1 y 2 → Autocorrelación Espacial Negativa; y si c asume valores inferiores a 1, entre 0 y 1 → Autocorrelación Espacial Positiva. Bajo la Hipótesis Nula, H_0 , la Distribución Asintótica es Normal.

Desarrollo

Son presentados a continuación, los Diagramas de Dispersión de Moran y el Resultado de las pruebas de Hipótesis para las dos variables en estudio: 'Índice de Dificultad de Acceso a la Energía' (IDAENE) y 'Necesidades Básicas Insatisfechas' (NBI). Es decir que, para ambas variables de interés, se intenta confirmar el Patrón Espacial (Autocorrelación Espacial Global) detectado por inspección visual en el AEDE descriptivo realizado en la Parte 1 de la investigación que se reporta. La Figura 1, Presenta el resultado que arroja el Software STATA 16, para la variable IDAENE, empleando tanto la Matriz W tipo Reina, como la $W_k(=5)$ Vecinos más Cercanos, normalizadas por filas. La Figura 2, presenta lo mismo para la variable NBI. En ambas Figuras, se ha utilizado la Distribución de Referencia, con $R = 500$ permutaciones. La Tabla 1, presenta las pruebas de hipótesis, considerando una Distribución Asintótica, incluyendo el estadístico c de Geary, para la Matriz W Tipo Reina normalizada por filas.

En ambas Figuras, 1 y 2, con los valores de I que se presentan, se observa que el Pseudo p -valor = 0,002, de manera que se rechaza H_0 , y se acepta la Existencia de Autocorrelación Espacial Global en las variables IDAENE

y NBI. Existe una pequeña diferencia ente los valores de I según se considere la Matriz de Pesos Espaciales W Tipo Reina o $W_k(=5)$, debido a que son diferentes las Matrices y, con ello, los rezagos espaciales W_{IDAENE} y W_{NBI} , conforme sea la W considerada.

Análogamente, en la Tabla 1, al utilizar resultados asintóticos que convergen a una Distribución Normal, aparece el valor z para comparar con los estadísticos I y c . En todos los casos, el valor $p = 0,000$, rechazándose la Hipótesis H_0 , y confirmando la existencia de un Patrón Espacial (Autocorrelación Espacial Global) en ambas variables, IDAENE y NBI. El valor del I de Moran es el mismo, correspondientemente, que el obtenido en los reportes anteriores.

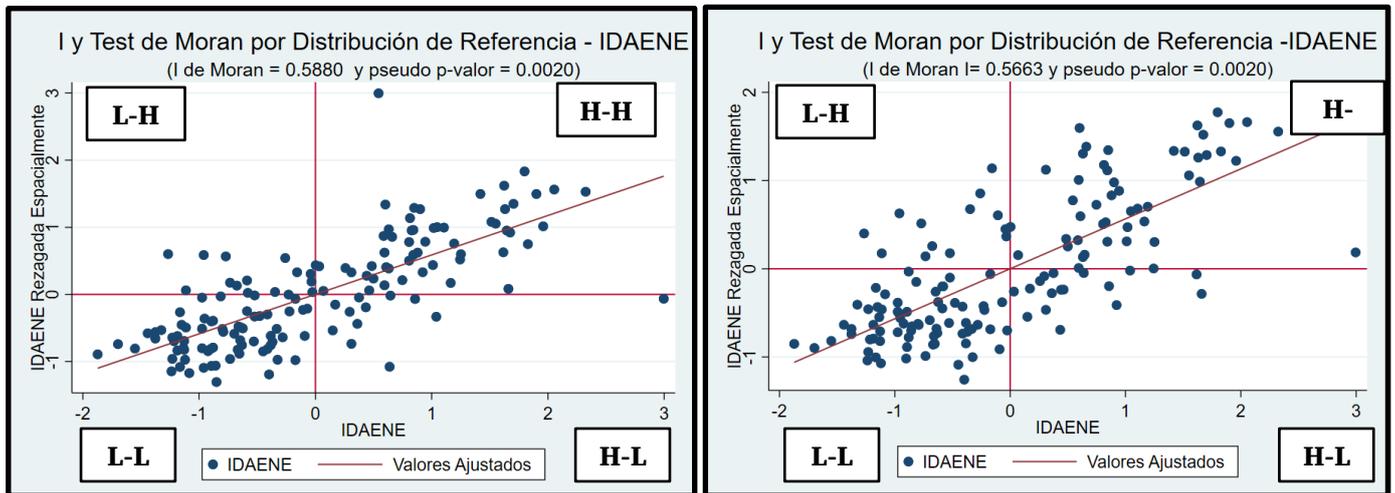


Figura 1. Diagrama de Dispersión e Índice de Moran reportados por STATA 16 para la Variable ‘Índice de Dificultad de Acceso a la Energía’ (IDAENE). Test utilizando la Distribución de Referencia. Izquierda: WReina – Derecha: WK5

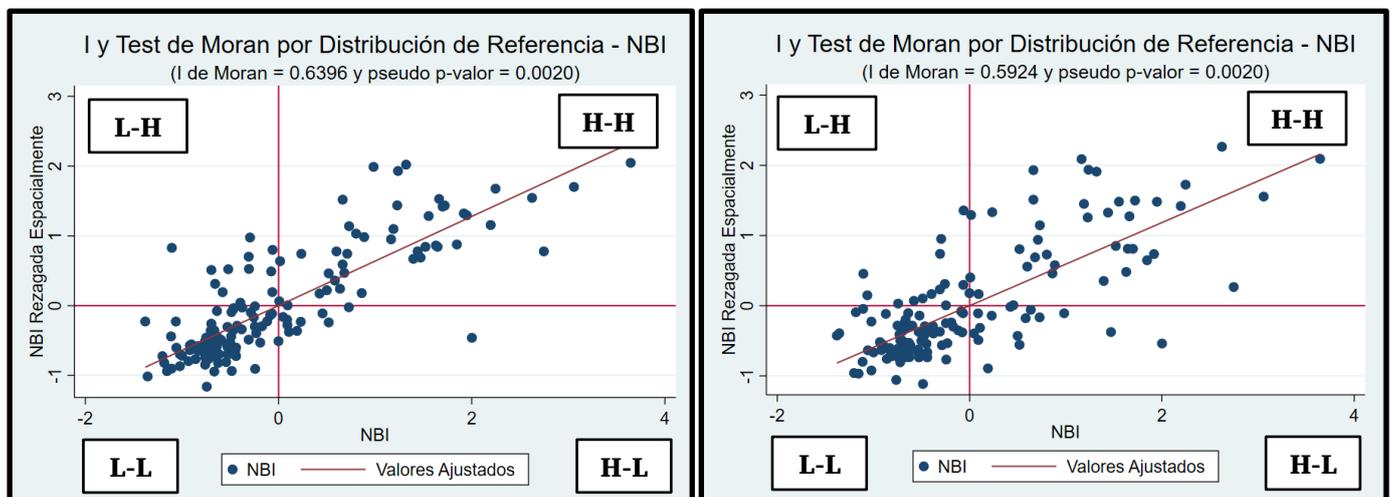


Figura 2. Diagrama de Dispersión e Índice de Moran reportados por STATA 16 para la Variable ‘Necesidades Básicas Insatisfechas’ (NBI). Test utilizando la Distribución de Referencia. Izquierda: WReina – Derecha: WK5

Name: WReina Type: Distance-based (inverse distance) Distance band: c1.c2 < d <= c3.c4 Row-standardized: Yes						Name: WReina Type: Distance-based (inverse distance) Distance band: c1.c2 < d <= c3.c4 Row-standardized: Yes					
Moran's I						Moran's I					
Variables	I	E(I)	sd(I)	z	p-value*	Variables	I	E(I)	sd(I)	z	p-value*
IDAENE	0.588	-0.007	0.051	11.595	0.000	NBI	0.640	-0.007	0.051	12.671	0.000
Geary's c						Geary's c					
Variables	c	E(c)	sd(c)	z	p-value*	Variables	c	E(c)	sd(c)	z	p-value*
IDAENE	0.403	1.000	0.053	-11.187	0.000	NBI	0.351	1.000	0.056	-11.662	0.000
*2-tail test						*2-tail test					

Tabla 1. Test I de Moran y c de Geary reportados por STATA 16 empleando Aproximación Asintótica. Izquierda: Variable 'Índice de Dificultad de Acceso a la Energía' (IDAENE) – Derecha: Variable 'Necesidades Básicas Insatisfechas' (NBI). Matriz de Pesos Espaciales utilizada: WReina

Conclusiones

Más allá de los resultados reportados, se presentan los siguientes puntos a modo de conclusiones finales:

1ero) Ha sido descripta de la forma más simple posible, la metodología que debe seguirse en para la confirmación estadística de la Autocorrelación Espacial Global en las dos variables de interés (IDAENE y NBI), detectada en el Análisis Exploratorio de Datos Espaciales, en forma descriptiva, en la Parte 1 de la investigación que se reporta.

2da) Siempre resulta conveniente trabajar con más de una Matriz de Pesos Espaciales (basadas en Criterios de Contigüidad y Criterios de Distancia). Asimismo, adoptar varios índices (al menos el estadístico de Moran y el de Geary, pues existen otros que pueden ser consultados en la literatura especializada) para las pruebas de hipótesis correspondientes, puesto que en caso que los índices arrojen valores similares para cada Matriz W considerada, y se rechace H_0 , la confirmación de la Autocorrelación Espacial Global resultará más sólida.

3era) Cuando en la teoría se habla de 'rezago espacial' o 'variable rezagada espacialmente', se está haciendo referencia específica a la Autocorrelación Espacial. Cabe destacar que el análisis que se realiza es de Corte Transversal, en un momento determinado. Cuando se trabaja en econometría no espacial con series de tiempo, se habla de 'rezago' para referirse a la Autocorrelación que una variable bajo estudio tiene en términos de los valores que la misma puede asumir en diferentes periodos. Desde allí que el término prevalezca para este efecto espacial.

Referencias

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic. 294 pages.
- Cliff, A.D., Ord, J.K. (1973). *Spatial Autocorrelation*. London: Pion. 178 pages.
- Geary, R. C. (1954). "The contiguity ratio and statistical mapping". *The Incorporated Statistician*, Nov., 1954, Vol. 5, No. 3, pp. 115-127+129-146.
- Getis, A., Ord, J. K. (1992). "The analysis of spatial association by use of distance statistics". *Geographical analysis*, 24(3), pp. 189–206.
- Herrera Gómez, M. (2015). "Econometría espacial usando Stata. Breve guía aplicada para datos de corte transversal". Instituto de Estudios Laborales y del Desarrollo Económico (ielde), Facultad de Ciencias Económicas, Jurídicas y Sociales, Universidad Nacional de Salta (UNSa). Disponible en: <http://hdl.handle.net/11336/7116>
- Moran, P. A. (1950). "Notes on continuous stochastic phenomena". *Biometrika*, Vol. 37, No. ½, pp. 17–23.
- Paelinck, J.H.P., Klaassen, L.H. (1979). *Spatial Econometrics*. Farnborough: Saxon House. 211 pages.
- Tobler, W.R. (1979). "Cellular Geography". In: Gale, S., Olsson, G. (eds) *Philosophy in Geography*. Theory and Decision Library, Vol 20. Springer, pp. 379-386.