

Aplicaciones con OpenGeoda entre la Producción Interna Bruta Total (PIB) y la Población Económicamente Activa (PEA) en Panamá, Años 2007-2015

Applications with OpenGeoda between Total Gross Domestic Production (GDP) and the Economically Active Population (EAP) in Panama, Years 2007-2015

Pérez, Jimmy¹, Ashaw, María², Henríquez, Teresita³, González, Jaime Manuel⁴

- (1) Estudiante de tesis; Centro regional de Coclé (C.R. Coclé); Lic. En mercadeo y comercio Internacional. Universidad Tecnológica de Panamá (UTP). agustinperez27@hotmail.es
- (2) Docente Universitaria (UP); Lic. En matemática, Máster en Matemática Aplicada. jetziba_beula@yahoo.es
- (3) Docente Universitaria (UP); Lic en Estadística, Master en Estadística Aplicada. terehen2@yahoo.com
- (4) Docente Universitario (ULP, UP y UTP); Lic. En Economía, Máster en Economía y Desarrollo. jaimemanuelg@hotmail.com
<https://orcid.org/0000-0002-7718-1896>

Págs.: 38 - 64

Recibido: 05/Jul. /2019

Aprobado: 17/Jul./2019

Artículo

3

Resumen

Dentro de los estudios de mercado y potencial económico de un individuo, empresa y jurisdicciones geográfica se hace importante la utilización de Software apropiado para el análisis espacial, como es el OpenGeoda.

Para el caso de Panamá, se utilizó una base de datos del producto Interno Bruto de Panamá (PIB) provincial, extrayéndose algunos mapas ilustrativos de las variables de producción y del mercado de trabajo en Panamá. De igual forma, se trabajó con una matriz inversa y se pudo ilustrar y calcular el índice de Moran.

El software nos permitió trabajar modelos de econometría espacial que se pueden apreciar desde un mínimo cuadrado ordinario hasta modelos dinámicos en un análisis espacial y regional en Panamá.

Palabras claves: Econometría Espacial, Correlación Espacial, Índice de Moran

Abstract

Within the market studies and economic potential of an individual, company and geographic jurisdictions it is important to use appropriate software for spatial analysis, such as OpenGeoda.

In the case of Panama, a database of the provincial gross domestic product (GDP) of Panama was used, extracting some illustrative maps of the variables of production and the labor market in Panama. In the same way, we worked with an inverse matrix and we could illustrate and calculate the Moran index.

The software allowed us to work spatial econometrics models that can be appreciated from a minimum ordinary square to dynamic models in a spatial and regional analysis in Panama.

Keywords: Spatial Econometrics, spatial correlation, Moran index

Introducción

Como estudioso de los fenómenos económicos, sociales o ambientales seguramente se habrá percatado que cada vez se realiza una mayor difusión de información georreferenciada. Es decir, las variables aparecen vinculadas a su dimensión espacial y pueden ser manejadas en potentes mapas en los que se van superponiendo capas de información.

A la par de la difusión de datos georreferenciados, también se han desarrollado modernos paquetes computacionales, conocidos como *Sistemas de Información Geográfica* (GIS por sus siglas en inglés), ello ha permitido dar impulso a una nueva subdisciplina de la econometría conocida como econometría espacial.

La econometría espacial fue definida a principios de los años setenta por **Jean Paelinck** como el creciente cuerpo de la literatura en ciencia regional que trata primordialmente con la estimación y prueba de problemas encontrados en la implantación de modelos econométricos multirregionales.

Luc Anselin (1988) uno de los pioneros y grandes impulsores de la econometría espacial considera que el campo de esta disciplina está formado por: *“...aquellos métodos y técnicas que, sustentados en una representación formal de la estructura de la dependencia y heterogeneidad espacial, provee el medio para llevar a cabo la adecuada especificación, estimación, prueba de hipótesis y predicción para modelos en la ciencia regional.”*

Materiales y métodos

Los métodos desarrollados por la econometría espacial permiten atender problemas de violación a los supuestos del modelo de regresión ordinario, que no es posible resolverlos en el marco de los modelos econométricos de Mínimos Cuadrados Ordinarios (**MCO**).

Estos problemas son típicos en los datos espaciales y se refieren a:

- 1) Dependencia espacial entre observaciones: **Correlación espacial.**
- 2) Heterogeneidad espacial entre observaciones: **Heteroscedasticidad espacial.**

El caso al que se le ha dedicado mayor atención es al primero, debido a que el segundo ha podido estudiarse en el marco de modelos de panel y otras técnicas similares en donde la heteroscedasticidad y el cambio estructural juegan un papel relevante.

Vecindad y dependencia espacial

Usualmente cuando el economista maneja series económicas, sociales o ambientales lo hace desde una perspectiva en la cual toma como dadas las coordenadas de localización geográfica de las variables. El obviar el contexto espacial significa una pérdida importante de información, sólo piense lo que ocurriría sí, por ejemplo, un estudio de criminalidad en una ciudad no considerará el efecto que tiene la situación que priva en las ciudades vecinas.

En ese sentido, los datos generalmente presentan algún tipo de dependencia o auto correlación espacial, la cual puede definirse como la existencia de una relación funcional entre lo que ocurre en un punto del espacio y lo que sucede en otro lugar, lo cual se explica fundamentalmente por razones de interacción humana con su entorno físico-ambiental. En tal

sentido, las variables socioeconómicas, físicas, ambientales y otras; generalmente presentan algún tipo de dependencia o auto correlación espacial.

La dependencia espacial implicaría que al tomar en consideración una variable, para diferentes localidades, esperaríamos características más similares en localidades vecinas, que en aquéllas separadas por grandes distancias. La dependencia espacial puede ser positiva o negativa, de ser positiva la presencia de un atributo en una localidad se extendería a las regiones vecinas y, en caso de ser negativa, obstaculizaría su presencia en sus vecindades.

Los datos espaciales se pueden clasificar de acuerdo con el objeto espacial al que se refieren y al nivel de medida de las variables. Dicha clasificación puede ilustrarse matricialmente como en la **figura N°1**

$z_1(1)$	$z_2(1)$	$z_k(1)$	S(1)
$z_1(2)$	$z_2(2)$	$z_k(2)$	S(2)
....
....
....
$z_1(n)$	$z_2(n)$		$z_k(n)$	S(n)

Figura N° 1: Matriz de datos Espaciales

Donde tenemos k variables $\{z_1, z_2, \dots, z_k\}$ medidas en la localización $s(i)$ donde $i=1, 2, \dots, n$. Si incorporamos el factor de temporalidad, podríamos tener una matriz de este tipo para cada período del tiempo. Las relaciones entre las variables y localizaciones clasificadas en la matriz de datos pueden establecerse a través de conectividad o vecindad.

Matriz de vecindad por contigüidad

La noción de vecindad se puede establecer de forma binaria; en tal caso, si dos unidades espaciales tienen una frontera común se les asigna un uno, en caso contrario se le asigna un cero. Bajo esta sencilla idea, una variable particular podría referenciarse en un mapa, a partir del cual

es posible establecer sus fronteras y, en consecuencia, identificar sus vecindades. **Luc Anselin (1988)**, plantea diferentes medidas de vecindad, las cuales se asemejan a un tablero de ajedrez y

que podemos apreciar en la **figura N° 2**:

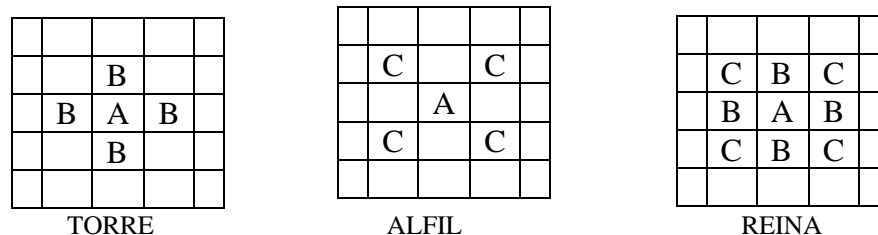


Figura N° 2: Diferentes Vecindades

La vecindad entre puntos también puede ser de orden superior, sí se consideran series de bandas concéntricas alrededor de la localidad bajo consideración.

			D		
	C	B	A	B	C
D	B	A	A	B	D
	C	B	A	B	C
			D		

Figura N° 3: Vecindades de Orden superior

Por ejemplo, en la **figura N°3** y considerando vecindad tipo torre, las celdas C y D son contiguas de segundo orden a la celda A, y son contiguas de primer orden a B.

En un mapa geográfico, como en la **figura N°4**, es posible construir cualquiera de los tipos de matrices de vecindad descritas anteriormente.

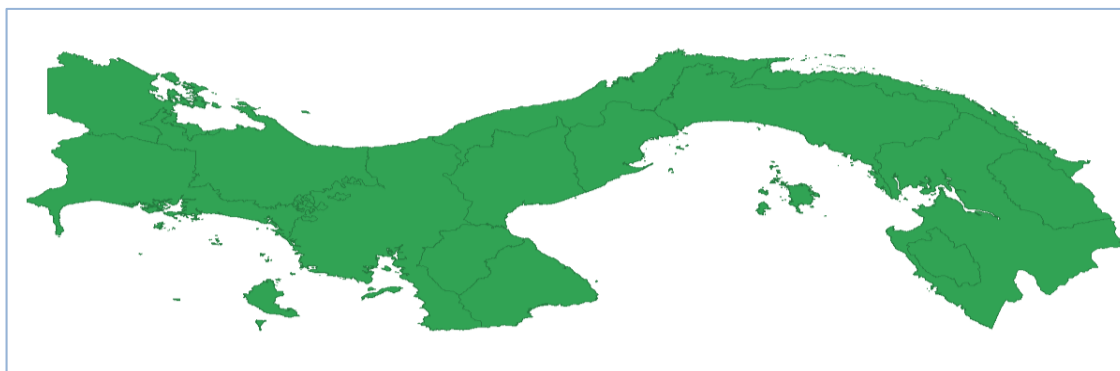


Figura N° 4: Mapa del istmo Panamá

Construcción analítica de una matriz de vecindad

Para ilustrar la forma en la que se construye una matriz binaria de vecindades por **Anselin (1988)**. Suponga que la localización de diferentes variables podría ubicarse en un mapa cuadrado como el siguiente:

1	2B	3
4B	5A	6B
7	8B	9

Figura N° 5: Mapa cuadrado

A cada localidad le asignamos un número y tomando como punto de referencia la localidad 5 calculamos vecindades tipo torre. La localidad 1 y la 3 no tienen vecindad, por ello se les asigna un cero en la matriz de vecindades. La vecindad de una localidad consigo misma es contabilizada también con un cero. La matriz de contactos resultante aparece en la figura siguiente:

Localidad	1	2	3	4	5	6	7	8	9
1	1	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	1

Figura N° 6: Matriz de contactos

Matriz de vecindad por distancia

La matriz de vecindades binarias es limitada, ya que únicamente considera la vecindad física, por lo cual no contabiliza la posibilidad de interacción entre regiones alejadas. Por ello, han sido propuestas otras medidas de vecindad alternativas, sustentadas en distancias de diferente tipo y cuya matriz, W , es conocida como la matriz de pesos o contactos espaciales.

Anselin plantea que, en caso de que la unidad espacial sea un sistema urbano, la vecindad puede ser obtenida de la trayectoria más corta en una red o gráfica formada por una conexión de puntos. Por ejemplo, en la figura 5, la distancia más corta entre los puntos es representada por la

línea punteada y la vecindad por el círculo que conecta los puntos y tiene como centroide a la localidad A.

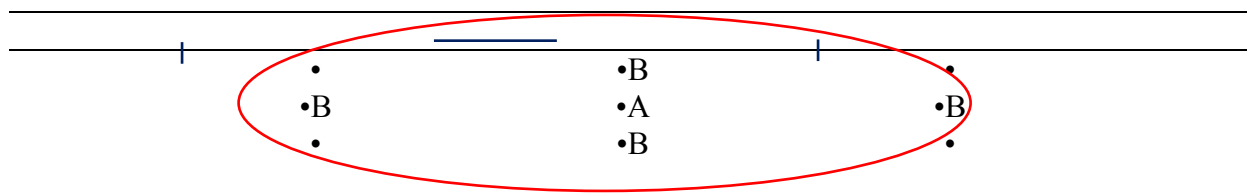


Figura N° 5: Vecindad por distancia más corta

Considerando los centroides como punto de referencia para medir las distancias geográficas, **Fotheringham, Brunsdon y Charlton (2000)** proponen las siguientes medidas de distancias:

- **Localización en el plano cartesiano**

En un sistema cartesiano, la distancia se mide por el teorema de Pitágoras y la localización es por medio de las coordenadas geográficas: latitud y longitud.

- **Distancia Euclidiana**

Con base a las coordenadas de latitud (x) y la longitud (y), la distancia entre los centroides de las localidades i y j.

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

La distancia euclidiana entre dos localidades i y j con coordenadas $(x_i,1, x_i,2)$, $(x_j,1, x_j,2)$, se puede escribir también como:

$$d_E(i, j) = \left[\sum_{k=1}^2 (x_{ik} - x_{jk})^2 \right]^{1/2}$$

La distancia puede ser generalizada a m dimensiones.

$$d_E(i, j) = \left[\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right]^{1/2}$$

- **Métrica de Minkowski**

En el caso de que $p=2$ es la distancia euclidiana, si $p=1$ es la distancia conocida como Manhattan o distancia taxicab.

$$d_E(i, j) = \left[\sum_{k=1}^m (x_{ik} - x_{jk})^p \right]^{1/p}$$

Localización en el globo o superficie de la tierra

En el caso de considerar la superficie de la tierra en lugar del plano cartesiano, se necesita de los cálculos geométricos:

Trigonometría esférica (curvatura de la tierra)

$$S_{ij} = R \arccos[\cos(90^\circ - \Phi_i) \cos(90^\circ - \Phi_j)] + \text{sen}(90^\circ - \Phi_i) \text{sen}(90^\circ - \Phi_j) \cos(\lambda_j - \lambda_i)$$

R es el radio de la tierra, arcoseno (arccos), coseno (cos), seno (sen), la latitud y longitud de la locación i

Son (Φ_i, λ_i)

$$x = R\lambda$$

$$y = R/n (\tan(\pi/4 + \Phi/2))$$

Donde R es el radio de la tierra, ln es el logaritmo natural, tangente (tan), Φ es la latitud y λ es la longitud.

Lambert (proyección a un área cilíndrica)

$$x = R\lambda$$

$$y = R \text{sen}\Phi$$

Construcción de matrices de contigüidad y distancia con OpenGeoDa

Para la construcción de matrices de contigüidad y distancia utilizamos el paquete OpenGeoDa, el cual fue desarrollado por el profesor **Luc Anselin** en el Departamento de Geografía de la Universidad de Illinois, Urbana-Champaign. Dicho paquete tiene la ventaja de ser gratuito y estar construido específicamente para el análisis de datos espaciales y la aplicación de técnicas de regresión espacial.

Para construir las matrices de pesos espaciales de contigüidad y distancias, se requiere que el archivo (PAN_1_ Jaime (Nuevo)1-2) este activo en el paquete OpenGeoDa.

El procedimiento es elegir primero las opciones de **Tools/Weights/Create**

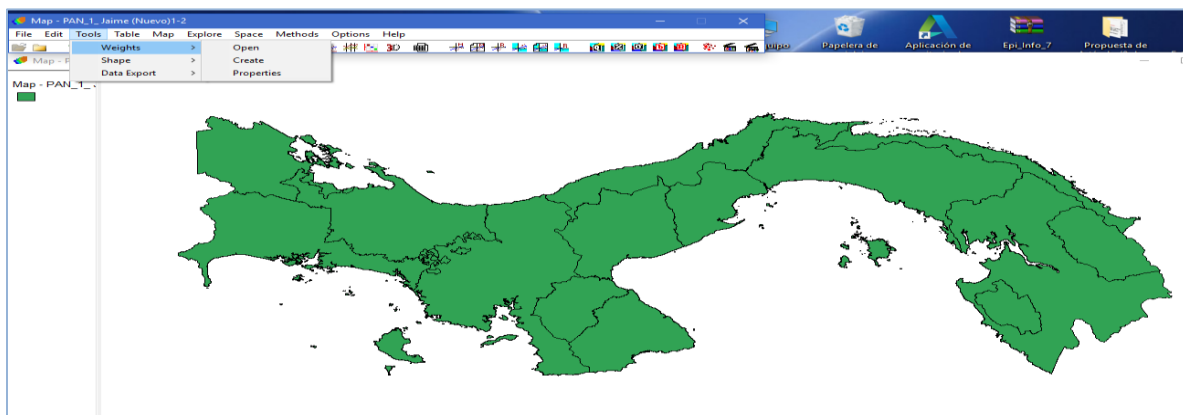


Figura N° 6: El procedimiento es elegir primero las opciones de **Tools/Weights/Create**

Con la elección aparece una ventana con nombre **“Weights File Creation”** y las opciones para construir matrices contigüidad y de distancia. Lo primero que se requiere es indicarle al OpenGeoDa cuál es la variable que identifica a los objetos en la cartografía (ID). En el caso de que no esté incluido en el archivo se requiere incluirla, para ello primero se elegí la opción **“Add ID Variable”** con el nombre sugerido por ID y posterior a ello se elegí en la opción de **Weights File Creation”**. Una vez elegido el ID el OpenGeoDa permite elegir las opciones para la construcción de las matrices.

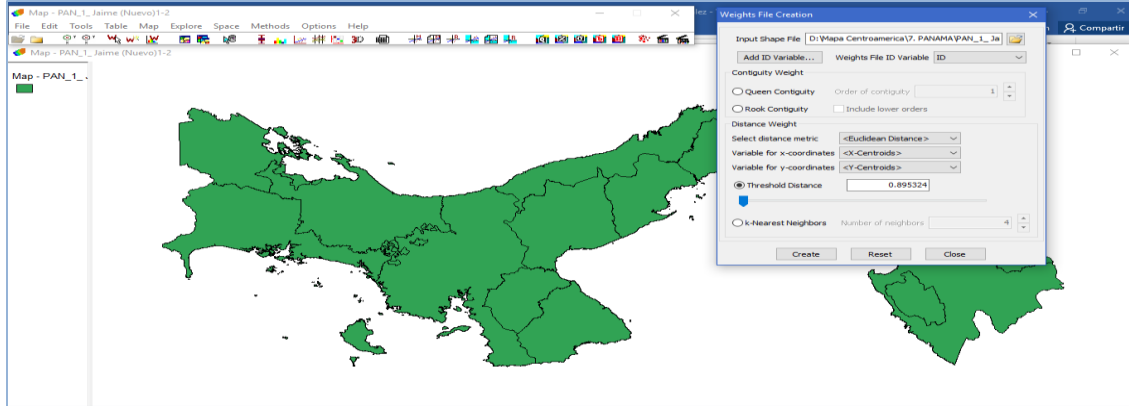


Figura N° 7: Las matrices de contigüidad

Las matrices de contigüidad se guardan asignando un nombre y la extensión gal y las distancias con gwt.

Estadísticos de dependencia espacial

Para la medición de dependencia espacial se han propuesto numerosos estadísticos, uno de los más utilizados es el **índice de Moran (1948)**, que se define en la fórmula siguiente:

$$I = \frac{R}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Donde x_i es la variable cuantitativa en la región i , \bar{x} es su media muestral, w_i son los pesos de la matriz W , R es el tamaño de muestra (Regiones); y,

$$E(I) = \frac{-1}{R - 1}$$

$$V(I) = \frac{RS_4 - S_3S_1(1 - 2R)}{(R - 1)(R - 2)(R - 3)(\sum_i \sum_j w_{ij})^2}$$

$$S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$$

$$S_2 = \sum_i \left(\sum_j w_{ij} + \sum_j w_{ji} \right)^2$$

$$S_3 = \frac{R^1 \sum_i (x_i - \bar{x})^4}{(R^1 \sum_i (x_i - \bar{x})^2)^2}$$

$$S_4 = (R^2 - R + 3)S_1 - RS_2 + 3 \left(\sum_i \sum_j w_{ij} \right)^2$$

$$I^* = \frac{I - E(I)}{\sqrt{V(I)}}$$

El índice de Moran sigue una distribución normal estandarizada en muestras grandes (Vaya y Moreno, 2000), de forma tal que un valor positivo (negativo) significativo del índice $Z(I)$ llevará al rechazo de la hipótesis nula de no autocorrelación espacial y a la aceptación de autocorrelación espacial positiva (negativa).

Es posible graficar la información del índice en un diagrama de dispersión de Moran. Dicho diagrama, presenta en el eje horizontal a la variable xx normalizada y en el eje vertical a la variable multiplicada por la matriz de pesos W , lo cual da lugar al retardo espacial de dicha variable. La visualización de un patrón aleatorio en la gráfica brinda evidencia de la ausencia de auto correlación espacial.

Dependencia espacial

La dependencia temporal, como la correlación serial, es unidireccional (el pasado explica el presente), mientras que la dependencia espacial es multidireccional (una región puede estar afectada no solamente por otra región contigua o vecina sino por otras que la rodean, al igual que ella puede afectar a las otras). Este hecho imposibilita la utilización del operador rezago L , $L^p Y_t = Y_{t-p}$, presente en el contexto temporal, para el análisis de la dependencia espacial. La solución consiste en utilizar la matriz W de efectos espaciales como operador de rezago espacial, que se puede leer como una media ponderada de los valores vecinos y se define como:

$$WY = \sum_{j=1}^N w_{ij} y_j$$

Donde y_j es el valor que toma el atributo medido en la vecindad j , w_{ij} es un ponderador cuya suma es la unidad.

Indicador Local de Asociación Espacial (LISA)

En procesos en los cuales existen patrones de agrupación local o clúster, el índice de Moran no los puede detectar, dado que sólo evalúa la dependencia global de todas las regiones. Como alternativa se han propuesto estadísticos locales, tal es el caso del índice local de Moran que se calcula en cada región o localidad y su definición es la siguiente:

$$I_i = \frac{z_i}{\sum_i z_i^2 / N_j} \sum_i w_{ij} z_j$$

Donde z_i es el valor de la variable correspondiente en la región i , N_j es el conjunto de regiones vecinas a i . Un valor elevado, positivo (negativo) y significativo del estadístico da lugar a la existencia de un clúster alrededor de la región i de valores similares elevados (bajos). Con base en el índice local, I_i , es posible encontrar su contribución al índice global, I , y detectar sus valores extremos lo cual lo convierte en un LISA.

Correlación espacial en el istmo centroamericano con OpenGeoda

Para poder elaborar este análisis se necesita previamente tener activo el archivo states en el paquete OpenGeoda y una matriz de contigüidad tipo Reina (Queen) de orden 1. Para visualizar la dependencia espacial del producto interno bruto utilizaremos la opción del menú de mapas Box-Map, con ella la información del PIB_TOTAL es agrupada en cuartiles y los valores inferiores y superiores son identificados para detectar outliers o valores extremos (**Figura N° 8**).

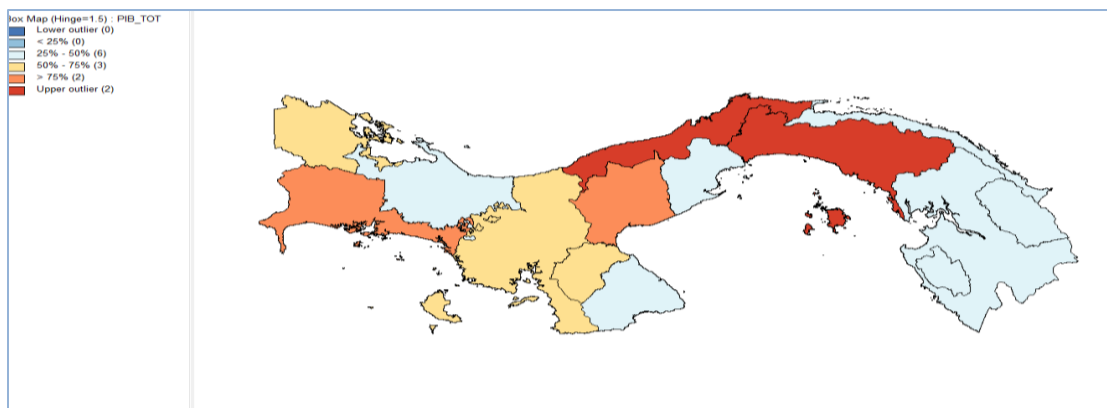


Figura N° 8: Box Map del PIB Total del istmo de Panamá de años 2007-2015

Del mapa anterior se desprende que existe una gran heterogeneidad del istmo de Panamá. En primer lugar, existe una fuerte asociación espacial entre las entidades de ingresos altos y medios altos, las cuales fundamentalmente forman una mancha en el sur. En segundo lugar, se observa que las provincias de PIB_Total (07-15) bajos y medio-bajos también se agrupan formando una mancha que se distribuye fundamentalmente en el centro del istmo de Panamá.

Ahora aplicamos la misma técnica a la PEA para el año 2007-2015, los resultados se muestran en la **figura N° 9**.

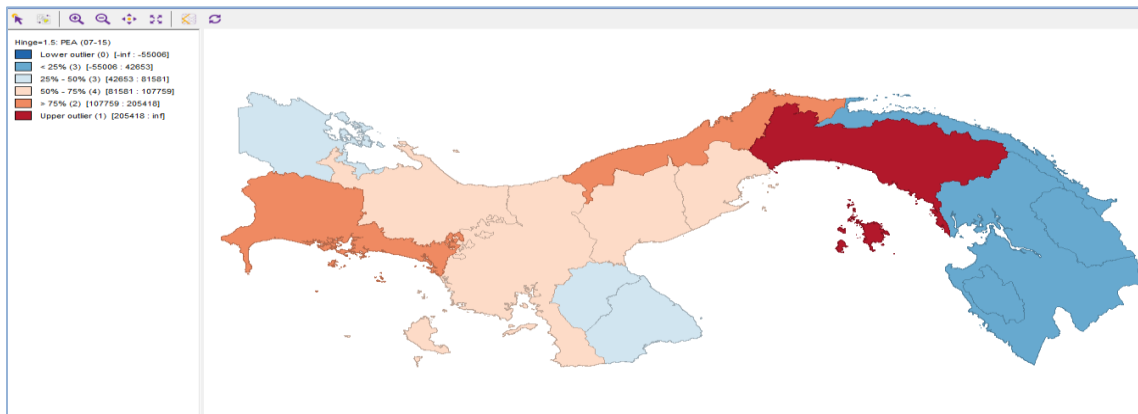


Figura N° 9: Box Map de la Población Económicamente Activa del Istmo de Panamá, años 2007-2015

En este último mapa se observa que la dependencia espacial es notoriamente más elevada que la visualizada antes para la distribución del PIB_Total entre las provincias del istmo de Panamá. Los manchones más oscuros dan cuenta de una fuerte asociación espacial entre las provincias de mayor densidad de población, lo mismo sucede con las manchas más claras que indican asociación entre las entidades con menor densidad de población.

Para evaluar estadísticamente la asociación espacial detectada en los mapas hacemos uso de la opción Space del menú principal del OpenGeoDa y seleccionamos el índice de Moran y su diagrama de dispersión.

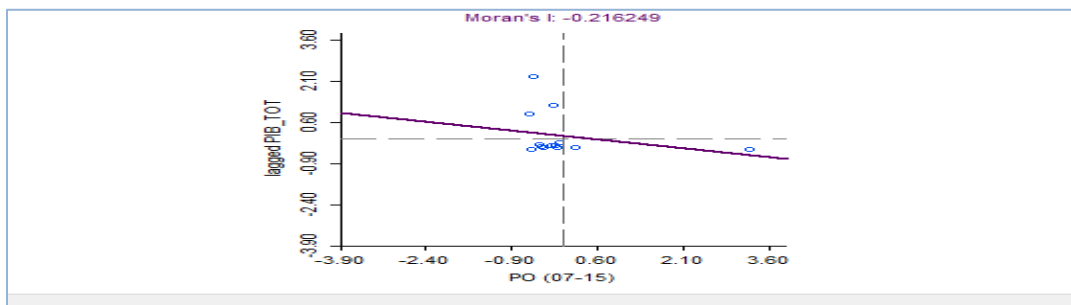


Figura N° 10: Índice de Moran

El índice de Moran, en la figura anterior, es de -0.216249 ello indica una baja correlación negativa en el crecimiento de la PEA a nivel global en el istmo de Panamá.

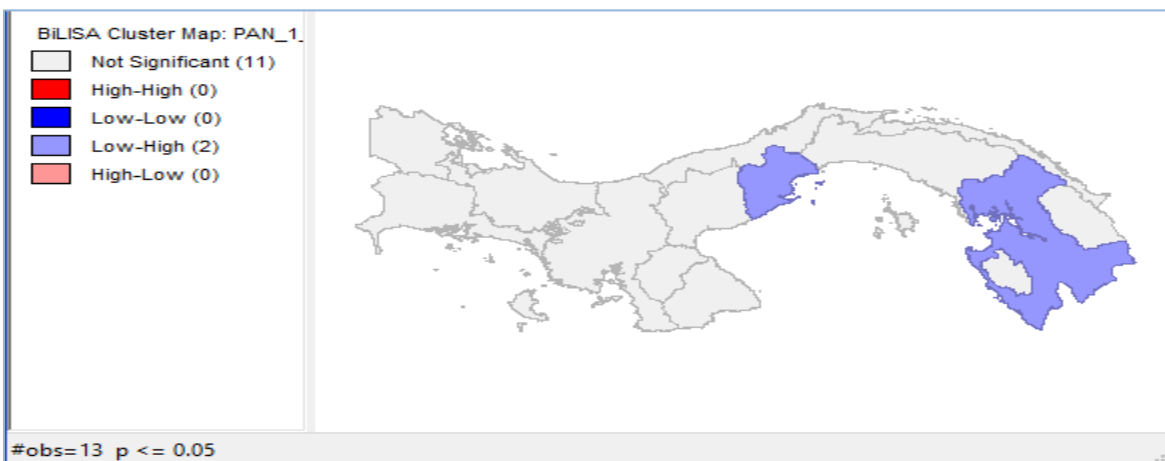


Figura N° 11: LISA del índice de Moran

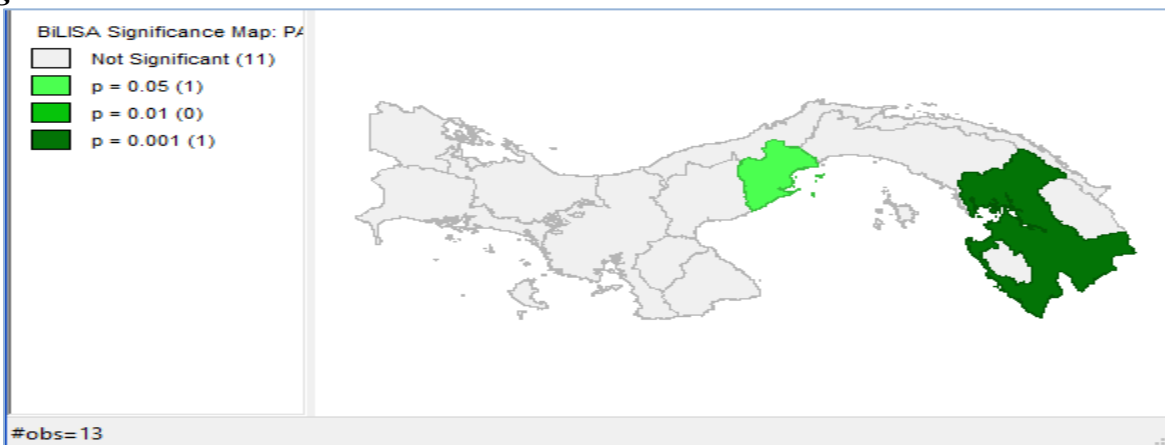


Figura N°12: Clúster significativo de dependencia espacial

El mapa LISA previo muestra las regiones que contribuyen significativamente al índice global de Moran y que conforman entre sí clúster significativos de dependencia espacial en su dinámica de crecimiento. El grupo de entidades de alto crecimiento con vecinos con la misma característica (Low-High): Panamá Oeste.

Modelos espaciales

Confirmada la dependencia espacial de los datos, es necesario especificar un modelo de regresión espacial que tome en cuenta dicha dependencia. Para plantear una especificación

general prototipo, se combinaron las estrategias de **Anselin (1988)**, **Lesage y Pace (2009)** y **Ehorst (2010)** para datos de corte transversal como los que hemos analizado en el modelo 2. El modelo general planteado es:

$$y_i = \rho W_1 y_i + \beta X_i + \theta W_2 x_i + \varepsilon_i$$

$$\varepsilon_i = \lambda W_3 \varepsilon_i + u_i$$

con $u_i \sim N(0, \Omega)$ siendo los elementos diagonales de $\Omega_{ij} = h_i(z\alpha)$ con $h_i > 0$.

Donde y_i es el vector de la variable endógena, X_i es una matriz de variables exógenas y el término de error ε_i que incorpora una estructura de dependencia espacial autorregresiva,

$$W_1, W_2 \text{ y } W_3$$

son matrices de pesos espaciales.

A partir de esta especificación podemos tener cinco casos:

1. Modelo de regresión clásico sin efectos espaciales:

$$\rho = 0, \lambda = 0, \theta = 0$$

$$y_i = \beta X_i + \varepsilon_i$$

$$\varepsilon_i = u_i$$

2. Modelo Autoregresivo:

$$\rho \neq 0, \lambda = 0, \theta = 0$$

$$y_i = \rho W_1 y_i + \beta X_i + \varepsilon_i$$

$$\varepsilon_i = u_i$$

3. Modelo de error espacial

$$\rho = 0, \lambda \neq 0, \theta = 0$$

$$y_i = \beta x_i + \varepsilon_i$$

$$\varepsilon_i = \lambda W_3 \varepsilon_i + u_i$$

Que se puede reescribir en su forma final como

$$y_i = \beta x_i + (I - \lambda W_3)^{-1} u_i$$

4. Modelo Durbin Espacial

La estrategia de Durbin sobre el factor común se aplica al modelo de Retardo Espacial como:

$$y_i = \rho W_1 y_1 + \beta x_i + \theta W_1 x_i + u_i$$

5. Modelo mixto autorregresivo espacial con errores espaciales autorregresivos (SARMA)

$$\rho \neq 0, \lambda \neq 0, \theta = 0$$

$$y_i = \rho W_1 y_i + \beta x_i + (1 - \lambda W_3)^{-1} u_i$$

6. Modelo Error Durbin Espacial

$$\rho = 0, \lambda \neq 0, \theta \neq 0$$

La estrategia Durbin sobre el factor común se aplica al modelo de Error Espacial con los siguientes pasos:

a) De la primera ecuación despejar los errores y sustituir en la segunda

$$y_i - \beta x_i = \lambda W_3 (y_i - \beta x_i) + u_i$$

b) Al despejar y_i , se obtiene

$$y_i = \lambda W_3 y_i + \theta W_3 x_i + u_i$$

Donde $\theta = -\beta\lambda$

Métodos de Estimación

Al igual que en el modelo de regresión clásico, la presencia de autocorrelación espacial dará lugar a que los estimadores de mínimos cuadrados ordinarios sean insesgados, pero ineficientes, por lo cual no se cumple el **teorema de Gauss-Markov**. En los modelos 2, 4, 5 y 6 la especificación considera rezagos autorregresivos de la variable dependiente, en consecuencia, los estimadores de mínimos cuadrados ordinarios serán sesgados e inconsistentes. La estimación del modelo espacial se realiza a través del método de máxima verosimilitud en concordancia con el modelo espacial específico que se seleccione.

De acuerdo a **Lesage y Pace (2009)** la estrategia de estimación de los modelos **Durbin Espacial (SDM)** y **Rezago Espacial (SAR)** por sus siglas en inglés, es la siguiente:

El modelo SDM

$$y = \rho W y + \alpha i_n + X \beta + W X \theta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Donde 0 representa un vector de ceros de $n \times 1$ y i_n un vector de unos $n \times 1$ asociados con el término de la constante α . Este modelo puede ser escrito de forma compacta con

$$Z = [i_n \ X \ W X] \ \gamma \delta = [\alpha \ \beta \ \theta]$$

y entonces definir el caso del modelo SAR cuando

$$Z = [i_n \ X] \ \gamma \delta = [\alpha \ \beta]$$

El modelo SAR

$$y = \rho W y + Z \delta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Si el valor del parámetro rho (ρ) fuera conocido por decir ρ^* , el modelo se puede escribir como

$$y - \rho^* W y = Z \delta + \varepsilon$$

Por lo que se puede resolver el problema de estimación de δ como

$$(I_n - \rho^* W) y = Z \delta + \varepsilon$$

$$\hat{\delta} = (Z' Z)^{-1} Z' (I_n - \rho^* W) y$$

También se encuentra la estimación de la varianza

$$\hat{\sigma}^2 = n^{-1} e(p^*)' e(p^*) \text{ Donde } e(p^*) = y - \rho^* W y - Z \hat{\delta}$$

Donde e son los errores de estimación.

Lo anterior indica que el método de estimación se concentra en el log de verosimilitud con respecto a los parámetros de β y ρ^2 y por tanto la maximización de la verosimilitud se convierte a un problema de optimización univariante en el parámetro ρ .

Propuesta para estimar al mismo tiempo todo:

1. Estimar la función de log-verosimilitud concentrada con respecto a los parámetros β y σ^2 , para obtener soluciones muy cercanas a las condiciones de primer orden junto con rho.

2. Sustituir las estimaciones de β y σ^2 , por lo que la función de log-verosimilitud depende de la muestra de datos y el parámetro desconocido rho.

3. En este punto la función de log-verosimilitud está concentrada con respecto rho, por lo que se usa para encontrar la estimación de máxima verosimilitud $\hat{\rho}$ que será usada a su vez en la estimación de $\hat{\beta}(\hat{\rho})$ y $\hat{\sigma}^2(\hat{\rho})$ en la siguiente vuelta.

La función de verosimilitud para SDM y SAR toma la forma siguiente

$$\ln L = -\left(\frac{n}{2}\right) \ln(\pi\sigma^2) + \ln|I_n - \rho W| - \frac{e'e}{2\sigma^2}$$

$$e = y - \rho W y - Z\hat{\delta}$$

$$\rho \in (\min(\omega)^{-1}, \max(\omega)^{-1})$$

Donde ω es el vector de $n \times 1$ raíces características de la matriz W. Dado que la matriz siempre está construida para tener raíces máximas de 1, entonces $\rho \in (\min(\omega)^{-1}, 1)$ el cual es un subconjunto del empleado en la práctica $\rho \in [0,1)$.

La función de log-verosimilitud concentrada en los valores de $\ln L(\rho)$ se escribe como

$$\ln L(\rho) = k + \ln|I_n - \rho W| - (n/2) \ln(s(\rho))$$

$$S(\rho) = e(\rho)'e(\rho) = e'_0 e_0 - 2\rho e'_0 e_d + \rho e'_d e_d$$

$$e(\rho) = e_0 - \rho e_d$$

$$e_0 = y - Z\delta_0$$

$$e_d = W y - Z\delta_d$$

$$\delta_0 = (Z'Z)^{-1} Z'y$$

$$\delta_d = (Z'Z)^{-1} Z'W y$$

La optimización es con respecto al parámetro rho y una vez estimado $\hat{\rho}$ con máxima verosimilitud se llega a la estimación con máxima verosimilitud de $\hat{\delta}$ y $\hat{\sigma}$

$$\hat{\delta} = \delta_0 - \hat{\rho} \delta_d$$

$$\hat{\sigma}^2 = n^{-1}S(\hat{p})$$

$$\hat{\Omega} = \hat{\sigma}^2[(I_n - \hat{p}W)'(I_n - \hat{p}W)]^{-1}$$

Estimación del modelo de Error Espacial (SEM)

Con una estrategia parecida, se obtiene la solución para SEM

$$y = X\beta + u$$

$$u = \lambda Wu + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

$$\ln L = -\left(\frac{n}{2}\right) \ln(\pi\sigma^2) + \ln|I_n - \lambda W| - \frac{e'e}{2\sigma^2}$$

$$e = (I_n - \lambda W)(y - X\beta)$$

Para un valor dado de λ ,

$$\beta(\lambda) = (X(\lambda)'X(\lambda))^{-1}X(\lambda)'y(\lambda), \text{ donde}$$

$$X(\lambda) = (X - \lambda WX)$$

$$y(\lambda) = (y - \lambda Wy)$$

$$\sigma^2(\lambda) = e(\lambda)'e(\lambda)/n$$

$$e(\lambda) = y(\lambda) - X(\lambda)\beta(\lambda)$$

La función de log-verosimilitud concentrada en los parámetros β y σ^2

$$\ln L(\lambda) = k + \ln|I_n - \lambda W| - (n/2)\ln(S(\lambda))$$

$$S(\lambda) = e(\lambda)'e(\lambda)$$

No es cuadrático, se necesita todo un proceso simultáneo

$$\hat{\beta} = \beta(\hat{\lambda})$$

$$\hat{\sigma}^2 = n^{-1}S(\hat{\lambda})$$

$$\hat{\Omega} = \hat{\sigma}^2[(I_n - \hat{\lambda}W)'(I_n - \hat{\lambda}W)]^{-1}$$

Estrategia de Selección de modelos: de lo particular a lo general

Anselin (2005) propone seguir un proceso de decisión para seleccionar entre el modelo clásico y los modelos espaciales SAR, SEM y SARMA, utilizando la estrategia que se muestra en la figura 10 y los estadísticos de contraste para la prueba de hipótesis de los tipos de dependencia espacial.

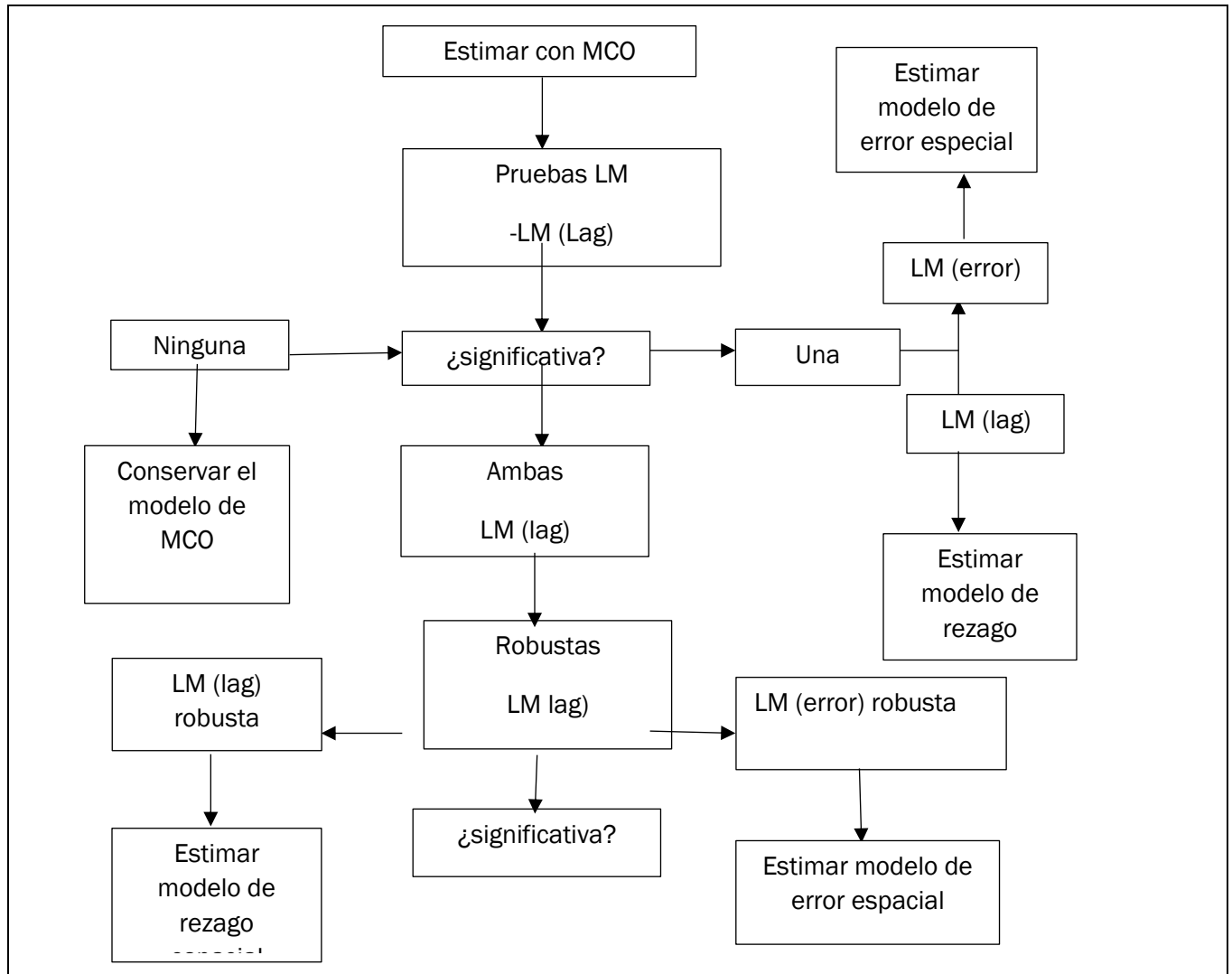


Figura N° 13: Estrategia de selección de modelos: de lo particular a lo general

Contrastes de autocorrelación espacial

Estos contrastes se aplican después de estimar el modelo clásico para analizar la presencia de algún tipo de dependencia espacial. La hipótesis nula es que el tipo de dependencia espacial es igual a cero, contra la hipótesis alternativa de que es diferente de cero.

1. Test I de Moran

Mide el efecto de autocorrelación espacial en los residuos ei en un modelo no-espacial o clásico, sin distinguir estructuras de Rezago o Error Espacial:

$$I = \frac{N \sum_{(2)} w_{ij} e_i e_j}{S_0 \sum_{i=1}^N e_i^2} = \frac{N e' w e}{S_0 e' e}$$

La inferencia se hace con el valor z estandarizado. El primer y segundo momento

$$E[I] = \frac{N \operatorname{tr}(MW)}{S_0 (N - K)}$$

$$E[I]^2 = \frac{\left(\frac{N}{S_0}\right)^2 \operatorname{tr}(MWMW) + \operatorname{tr}(MW)^2 + [\operatorname{tr}(MW)]^2}{(N - K)(N - K + 2)}$$

Se distribuye como una X^2 con un grado de libertad

2. Test LM-EL: Error espacial (robusto)

Se basa en el principio de los multiplicadores de Lagrange y fue propuesto por Burridge (1980):

$$LM - ERR = \frac{\frac{[e' W e]^2}{S^2}}{\operatorname{tr}[W' W + W^2]}$$

Se distribuye como una X^2 con un grado de libertad

3. Test LM-EL: Error espacial (robusto)

El estadístico LM-ERR se ajusta por una mala especificación local de la dependencia espacial, como es el caso de una variable endógena rezagada (Anselin, 1996):

$$LM - EL = \frac{\left[\frac{e' W e}{S^2} - T_1 (R\tilde{J}_{p-\beta})^{-1} \frac{e' W y}{S^2}\right]^2}{T_1 - T_1^2 (R\tilde{J}_{p-\beta})}$$

Con

$$(R\tilde{J}_{p-\beta})^{-1} = \left[T_1 \frac{(WX\beta)' M (WX\beta)}{S^2} \right]^{-1}$$

$$T_1 = \operatorname{tr}(W' W + W^2)$$

Se distribuye como una X^2 con un grado de libertad

4. Test LM-LAG: Rezago Espacial

Por rezago espaciales de la variable endógena (Anselin, 1988):

$$LM - LAG = \frac{\left[\frac{e'Wy}{s^2} \right]^2}{(R\tilde{J}_{p-\beta})}$$

Se distribuye como una X^2 con un grado de libertad

5. Test LM-LE: Rezago Espacial (Robusto)

El estadístico es robusto ante la presencia de dependencia local del error espacial (Anselin, 1988):

$$LM - LE = \frac{\left[\frac{e'Wy}{s^2} - \frac{e'We}{s^2} \right]^2}{(R\tilde{J}_{p-\beta} - T_1)}$$

6. Test SARMA: Rezago y Error Espacial

Es robusto ante la presencia de dependencia local y del error espacial (Anselin, 1988):

$$LM - LE = \frac{\left[\frac{e'Wy}{s^2} - \frac{e'We}{s^2} \right]^2}{(R\tilde{J}_{p-\beta} - T_1)} + \frac{\left[\frac{e'We}{s^2} \right]^2}{T_1}$$

Se distribuye como una X^2 con un grado de libertad

Un modelo espacial para convergencia Beta con OpenGeoDa

La teoría neoclásica del crecimiento, aplicada al crecimiento regional, permite plantear que una región pobre tendrá un acervo de capital per cápita inferior al de una región rica, en consecuencia, tenderá a crecer más rápido pues su productividad marginal del capital es mayor. La consecuencia de esto es que en el largo plazo se puede presentar convergencia entre las regiones pobres y las más ricas.

Para verificar esta hipótesis **Sala i Martin (1999)** presenta un modelo conocido como convergencia beta absoluta y que se define de la siguiente manera:

$$\ln(y_{i,t}) - \ln(y_{i,t-1}) = \alpha - \beta \ln(y_{i,t-1}) + u_{i,t}$$

Donde $(y_{i,t})$ es el logaritmo natural del producto per cápita de la región i en el período actual t , mientras que $(y_{i,t-1})$ es el logaritmo del producto per cápita de la región i en el período inicial $t-1$.

Un coeficiente estimado negativo para la beta será evidencia a favor de la existencia de convergencia absoluta.

Para probar convergencia en las entidades provinciales de Panamá se utiliza la opción Methods/Regression del OpenGeoDa, se incluyen las opciones de resultados, como la del valor z del la I- Moran, se define en la ventana correspondiente la variable dependiente con la tasa de crecimiento entre el PIB_TOTAL entre 2007 y 2015, y la variable explicativa que, para esta ecuación, es la Población Ocupada (PO) y Población desocupada (DES) de 2007-2015. En principio seleccionamos el modelo 1 correspondiente a la regresión clásica, sus resultados se muestran en la figura N° 14 siguiente.

R-squared	: 0.982686	F-statistic	: 283.784
Adjusted R-squared	: 0.979223	Prob(F-statistic)	: 1.55591e-009
Sum squared residual	: 6.56603e+006	Log likelihood	: -103.807
Sigma-square	: 656603	Akaike info criterion	: 213.615
S.E. of regression	: 810.311	Schwarz criterion	: 215.309
Sigma-square ML	: 505079		
S.E of regression ML	: 710.689		

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	-665.917	305.086	-2.18272	0.05400
PO (07-15)	-0.00189551	0.00930496	-0.20371	0.84267
DES (07-15)	0.467773	0.144371	3.24008	0.00887

Figura N° 14: Convergencia beta en el modelo de regresión clásico

Dado el elevado p-valor de beta podemos concluir que no es estadísticamente significativa y se rechaza la hipótesis de convergencia.

Las pruebas de diagnóstico al modelo se muestran en la figura N° 15, los diferentes estadísticos de prueba contrastan la hipótesis nula de no autocorrelación espacial.

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.1643	1.0861	0.27742
Lagrange Multiplier (lag)	1	0.1052	0.74564
Robust LM (lag)	1	0.0509	0.82143
Lagrange Multiplier (error)	1	0.4155	0.51918
Robust LM (error)	1	0.3612	0.54782
Lagrange Multiplier (SARMA)	2	0.4665	0.79197

Figura N° 15: Estadísticos de prueba del modelo de regresión clásico

En los resultados en la Figura N° 15, el índice de Moran presenta un p-valor de 1.08% lo cual permite rechazar la hipótesis nula de no autocorrelación espacial. El LM-lag y el LM-lag robusto presentan la hipótesis alternativa específica de modelo de rezago espacial, mientras que el LM-error y LM-error robusto establecen como hipótesis alternativa al modelo de error espacial. En todos los casos, las versiones robustas se consideran sólo si las versiones estándar son significativas.

En la Figura N° 16 la prueba LM-lag permite rechazar la hipótesis nula, pero el LM robusto no lo confirma, lo mismo sucede con los otros dos estadísticos LM-error, de acuerdo con Anselin, está es una situación patológica en donde el estadístico robusto se invalida debido a la existencia de otro error de especificación, en la forma funcional.

Como alternativa se estimó un modelo de rezago espacial, sus resultados se presentan en la Figura N° 17.

Variable	Coefficient	Std.Error	z-value	Probability
W_PIB_TOT	0.00833129	0.0619904	0.134397	0.89309
CONSTANT	-691.801	330.031	-2.09617	0.03607
PO (07-15)	-0.00177646	0.00820163	-0.216599	0.82852
DES (07-15)	0.466558	0.12685	3.67802	0.00024

Figura N° 17: Convergencia beta en el modelo de rezago espacial

El coeficiente del rezago espacial es estadísticamente significativo, sin embargo, de nueva cuenta, el coeficiente beta no brinda evidencia de convergencia absoluta.

```

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : PAN_1_ Jaime (Nuevo)1-2
TEST          DF      VALUE      PROB
Likelihood Ratio Test      1      0.0198      0.88807
    
```

Figura N° 18: Estadísticos de prueba en el modelo de rezago espacial

El estadístico LR del modelo de rezago espacial se puede apreciar en el Figura N° 18 y permite rechazar la hipótesis nula de modelo de regresión clásica a favor de la alternativa de modelo espacial, por lo menos a un nivel de significancia del 88.8%.

Para completar el análisis se estimó también el modelo de error espacial

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	-672.726	292.377	-2.30088	0.02140
PO (07-15)	-0.001178	0.00769105	-0.153165	0.87827
DES (07-15)	0.457562	0.119705	3.82241	0.00013
LAMBDA	0.224854	0.280217	0.802429	0.42230

Figura N° 18: Convergencia beta en el modelo de error espacial

El coeficiente del error espacial es estadísticamente significativo, pero de nueva cuenta el coeficiente beta no brinda evidencia de convergencia absoluta. Con el estadístico LR del modelo de error espacial también se rechaza la hipótesis nula de modelo de regresión clásica a favor de la alternativa de modelo espacial.

```

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : PAN_1_Jaime (Nuevo)1-2
TEST          DF      VALUE      PROB
Likelihood Ratio Test      1      0.5281      0.46742
    
```

Figura N° 19: Estadísticos de prueba en el modelo de error espacial

Índice de Moran para los residuales del modelo de rezago y error espacial

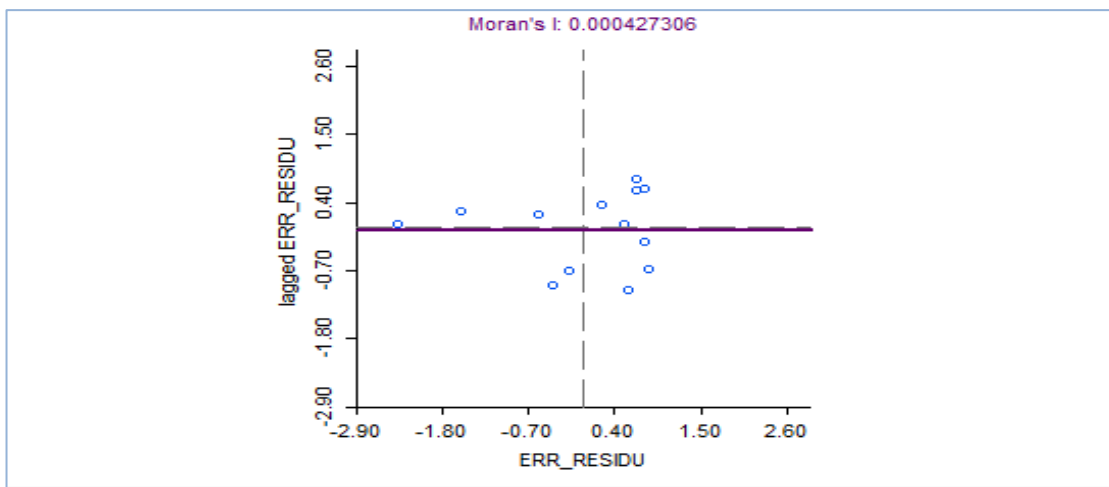


Figura N° 20: Residuales del modelo de rezago

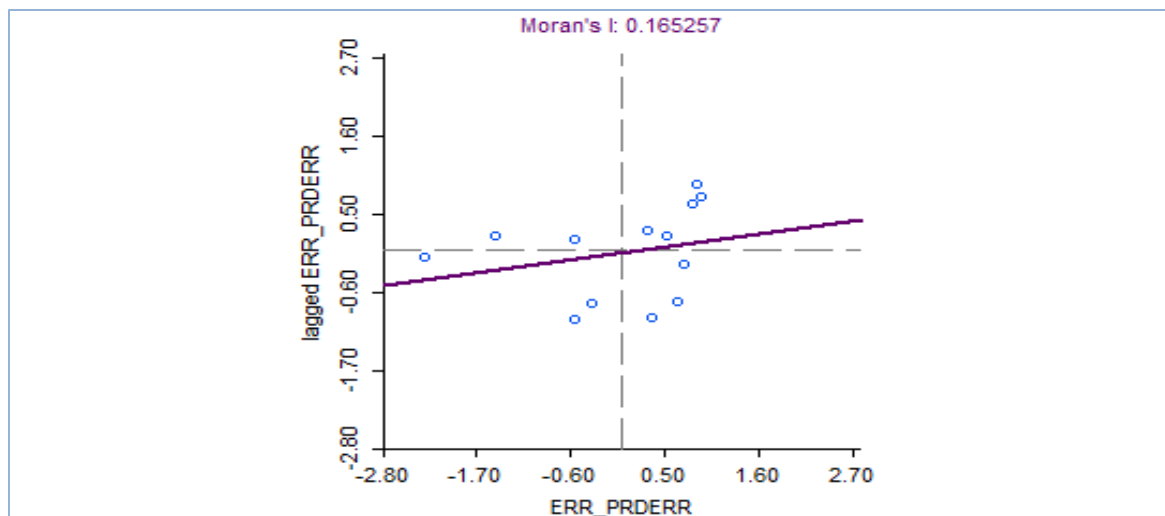


Figura N° 20: Residuales del modelo error espacial

Al evaluar los residuales de los modelos de rezago y de error espacial, se puede confirmar en la figura anterior que los índices son muy pequeños, sus p-valores son de 0.198 y 0.528 respectivamente los cuales no permiten rechazar la hipótesis nula de no autocorrelación espacial. Lo anterior significa que, los modelos de rezago y error espacial consiguieron eliminar la autocorrelación espacial.

Conclusiones

La utilización del software OpenGeoda se convierte en una herramienta de análisis espacial de primera línea, cuando se trata de estudios de variables económicas y que estén relacionadas con variables geográficas. Haciendo uso del análisis por individuo geográfico como unidad jurisdiccional por provincia en Panamá, observamos en este estudio que nuestro país experimenta un crecimiento económico hipertrófico y se evidencia por medio de los mapas y los indicadores espaciales descriptivos. Además, se puede utilizar de igual forma a individuos, empresas, corregimientos, distritos y países como unidades de análisis espacial.

Utilizando la data de valores tanto para el **Producto Interno Bruto Total (PIB_T)**, la **Población Ocupada (PO)** y **Población Desocupada (DES)** enmarcándolo en un análisis de causalidad se pudo establecer modelos de corte espacial y lograr establecer aproximaciones entre las variables. Debemos advertir que la medición del índice de Moran y los modelos de rezago y

de error espacial, constituyen un buen método de predicción econométrica en la investigación inicial

En términos generales el estudiantado y las empresas nacionales pueden ser beneficiadas de los avances que se tenga en el estudio de la Economía Regional y Espacial en Panamá.

Referencias Bibliográficas

Anselin L (2005) *Exploring Spatial Data with GeoDaTM: A Workbook. Centre for Spatially Integrated Social Science*. Disponible en: <http://sal.agecon.uiuc.edu/>

Anselin, L., (1998) **Spatial Econometrics: Methods and Models**. kluwer academic, publishers, dordrecht, the netherlands.

Anselin, L., (2012) **Opengoda 1.2 User's Guide. Spatial Analysis Laboratory (sal)**. department of agricultural and consumer economics, university of illinois, urbana-champaign, il.

Chasco C (2003) *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Comunidad de Madrid. Disponible en: <http://www.madrid.org>

Fotheringham, Brunson y Charlton (2000) **Quantitative Geography: Perspectives on Spatial Data Analysis**.

Haining, Robert (2003) **Spatial Data analysis, Cambridge University Press**,

Lesage, J. y Pace, K. (2009) **Introduction of Spatial econometrics**, Taylor & Francis Group, LLC.

Mendoza, M. A. (1999). **Econometría Espacial: Aplicaciones con Opengoda yr**.

Sala i-Martin-Xavier, **Apuntes de Crecimiento Económico**. Antoni Bosch editor. Barcelona.