Modelos espaciales en datos panel

Gustavo A. García

ggarci24@eafit.edu.co

Econometría avanzada II

PhD/Maestría en Economía

Universidad EAFIT

Link slides en formato html

Link slides en formato PDF

En este tema

- Motivación
- Principios básicos en el tratamiento de datos espaciales
- Efectos espaciales
- Heterogeneidad espacial
- Autocorrelación o dependencia espacial
- Análisis exploratorio de datos en el espacio
- Análisis confirmatorio de datos espaciales
- Regresión espacial
- Ejercicio aplicado en R

Lecturas

- Elhorst, J.P. (2010). "Applied Spatial Econometrics: Raising the Bar". *Spatial Economic Analysis*, 5(1):9–28
- Millo, G. y Piras, G. (2012). "splm: Spatial Panel Data Models in R". *Journal of Statistical Software*, 47(1):1–37
- Elhorst, J.P. (2014). *Spatial Econometrics from Cross-Sectional Data to Spatial Panels*, Springer
- LeSage, J. y Pace, R. (2014). "Interpreting spatial econometrics models". En Fischer, M. y Nijkamp, P. (Eds.), *Handbook of Regional Science*, Springer
- Halleck Vega, S. y Elhorst, J.P. (2015). "The SLX model", *Journal of Regional Science*, 55(3):339-363
- Golgher, A. y Voss, P. (2016). "How to interpret the coefficients of spatial models: spillovers, direct and indirect Effects". *Spatial Demography*, 4:175–205
- Belotti, F., Hughes, G. y Mortari, A. (2017). "Spatial panel-data models using Stata", *The Stata Journal*, 17(1):139-180.

Motivación

- Cuando se trabajan con datos de corte transversal suelen aparecer los denominados efectos espaciales: la heterogeneidad y la dependencia espacial
- Heterogeneidad espacial: este efecto aparece cuando se utilizan datos de unidades espaciales muy distintas para explicar un mismo fenómeno $\Longrightarrow$ aparecen problemas como la heteroscedasticidad o la inestabilidad estructural
- Dependencia espacial: o autocorrelación espacial, surge siempre que el valor de una variable en un lugar del espacio está relacionado con su valor en otro u otros lugares del espacio
- Mientras la heterogeneidad espacial puede ser tratada por técnicas estándar de econometría, la dependencia espacial no. Esto se debe a la multidireccionalidad que domina las relaciones de interdependencia entre unidades espaciales
- La econometría espacial, como subdisciplina de la econometría, surge como respuesta para resolver la presencia de efectos espaciales

Motivación

El término econometría espacial fue introducido por Jean Paelinck en los 70. En el libro *Spatial Econometrics*, Paelinck y Klaassen resaltan cinco características del campo en términos de los temas considerados:

- el rol de la interdependencia espacial en los modelos espaciales
- las asimetrías en las relaciones espaciales
- la importancia de factores explicatorios localizados en otros espacios
- diferenciación entre interacciones *ex post* y *ex ante*
- modelación explicita del espacio

Motivación

- R
- Matlab
- GeoDa
- QGis
- ArcGis
- Stata

Motivación

Fuentes de información de datos espaciales:

- SIGOT
- GeoDa
- IDESC-Cali
- Medellín
- IDECA-Bogotá
- Mapas del mundo

Principios básicos en el tratamiento de datos espaciales

Paelinck y Klaassen (1979) destacan cinco principios básicos en el campo de la econometría espacial y el tratamiento de datos de corte transversal en general:

- Interdependencia: todo modelo espacial ha de caracterizarse por su interdependencia, es decir, deben incorporarse relaciones mutuas entre las observaciones de las variables económicas, sociales, demográficas, etc.
- Asimetría: las relaciones espaciales son, en principio, asimétricas
- Alotopía: se ha de buscar a priori "la causa" de un fenómeno espacial en otro lugar
- No linealidad: la no linealidad de soluciones espaciales óptimas *ex-ante* conduce a modelos econométricos *ex-post* que requieren una atención particular en lo que respecta a su especificación, lo cual generalmente será no lineal
- Inclusión de variables topológicas: dado que la vida económica se desarrolla necesariamente en el espacio geográfico, un modelo espacial debe incorporar variables topológicas: coordenadas, distancias, superficies, densidades, etc.

De acuerdo a Paelink y Klaassen (1979), no siempre será posible observar estos cinco principios de construcción de modelos espaciales y probablemente pueden haber otros además de los aquí especificado

Efectos espaciales

Causas de la dependencia espacial:

- la delimitación arbitraria de las unidades espaciales de observación (ejemplo, zonas censales, limites municipales, departamentales...)
- problemas de agregación espacial
- la presencia de externalidades y efectos de desbordamiento

Causa de la heterogeneidad espacial:

- falta de estabilidad en el espacio del comportamiento u otras relaciones bajo estudio
- esto implica que la forma funcional y los parámetros varían con la localización y no son homogéneos en los datos
- puede ocurrir al estimar modelos econométricos con datos de sección cruzada de unidades espaciales diferentes, como regiones ricas y pobres

La heterogeneidad espacial se puede tratar con la econometría estándar que tenga en cuenta la inestabilidad estructural

Heterogeneidad espacial

- Definición: se refiere a la variación en las relaciones en el espacio
- Aspectos de la heterogeneidad espacial:
  - la inestabilidad estructural: falta de estabilidad en el espacio del comportamiento de la variables bajo estudio. La forma funcional y los parámetros de una regresión pueden variar según la localización, por tanto, no son homogéneos en toda la muestra
  - la heterocedasticidad: proviene de la omisión de variables u otras formas de error de especificación que llevan a la aparición de errores de medida
- La heterogeneidad espacial puede tratarse por medio de las técnicas econométricas estándar, en concreto:
  - parámetros variantes, coeficientes aleatorios (Hildreth y Houck, 1968)
  - *Switching regressions* (Quant, 1958)
  - técnicas de filtraje adaptativo espacial (Foster y Gorr, 1983)
  - expansión espacial de parámetros (Casetti, 1972)
  - regresiones ponderadas geográficamente (Fotheringham et al., 1998)

Autocorrelación o dependencia espacial

- Definición: aparece como consecuencia de la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que ocurre en otro lugar

- El valor que toma una variable en una región no viene explicado únicamente por condicionantes internos sino también por el valor de esa misma variable en otras regiones vecinas, incumpliéndose por tanto el supuesto de independencia entre las observaciones muestrales

- La autocorrelación espacial puede ser positiva o negativa

- Positiva: la presencia de un fenómeno determinado en una región lleva a que se extienda ese mismo fenómeno hacia el resto de regiones que la rodean, favoreciendo así la concentración del mismo

- Negativa: cuando la presencia de un fenómeno en una región impida o dificulte su aparición en las regiones vecinas a ella, es decir, cuando unidades geográficas cercanas sean netamente más disímiles entre ellas que entre regiones alejadas en el espacio (tablero de ajedrez)

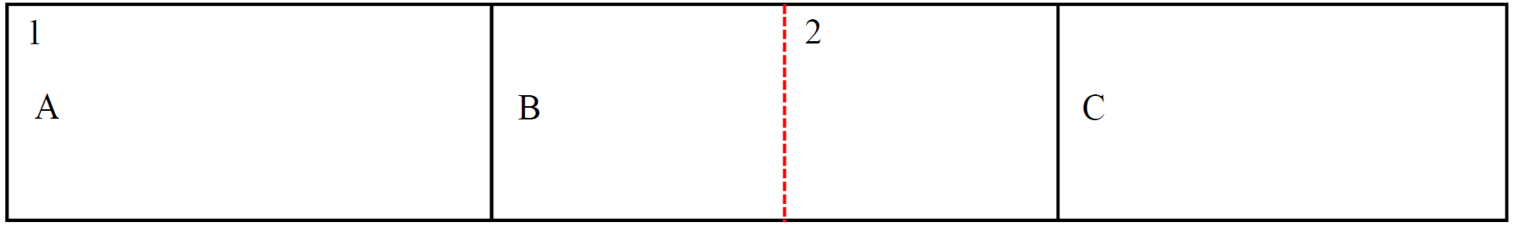
- Cuando la variable analizada se distribuye de forma aleatoria, no existirá autocorrelación espacial

Autocorrelación o dependencia espacial

Causas: la existencia de errores de medida y fenómenos de interacción espacial

Errores de medida: pueden surgir, entre otros aspectos, como consecuencia de una escasa correspondencia entre la extensión espacial del fenómeno económico bajo estudio y las unidades espaciales de observación

Ejemplo:
La correcta delimitación espacial de una variables x corresponde a las áreas ABC
Las observaciones disponibles de x son agregadas a nivel espacial en dos niveles: 1 y 2
Consecuencias: x_1 observada contendrá a x_A y parte de x_B, al tiempo que x_2 contendrá a x_C y parte de x_B



Resultado: x estará correlacionada espacialmente aunque de forma espuria

Interacción espacial: también entendido como efectos de desbordamiento y de jerarquías espaciales. La existencia de efectos desbordamiento de las infraestructuras de transporte o la difusión tecnológica entre economías son ejemplos claros de fenómenos que favorecen la aparición de interdependencias entre unidades espaciales

Autocorrelación o dependencia espacial

Matriz de pesos como instrumento para recoger las interdependencias

- Es posible detectar cierta similitud entre los conceptos de autocorrelación espacial y temporal en la medida en que, en ambos casos, se produce un incumplimiento de la hipótesis de independencia entre las observaciones muestrales

- Diferencias entre estos dos tipos de autocorrelaciones:
  - la dependencia temporal es únicamente unidireccional $\Longrightarrow$ el pasado explica el presente
  - la dependencia espacial es multidireccional $\Longrightarrow$ una región puede no sólo estar afectada por otra región contigua a ella sino por otras muchas que la rodean, al igual que ella puede influir sobre aquéllas

- La multidireccionalidad de la dependencia espacial imposibilita la utilización del operador de retardos L, $L^p x_t = x_{t-p}$, ya que recoge sólo únicamente una relación unidireccional

Autocorrelación o dependencia espacial

Matriz de pesos como instrumento para recoger las interdependencias

- La solución al problema de la multidireccionalidad en el contexto espacial para por la definición de la denominada matriz de pesos espaciales, de retardos o de contactos $\mathbf{W}$:

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & \dots & w_{1N} \\ w_{21} & 0 & \dots & w_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ w_{N1} & w_{N2} & \dots & 0 \end{bmatrix}$$

una matriz cuadrada no estocástica cuyos elementos w_{ij} reflejan la intensidad de la interdependencia existente entre cada par de regiones i y j

- No existe una definición de $\mathbf{W}$ unánimemente aceptada, si bien se ha de cumplir que dichos pesos sean no negativos y finitos (Anselin, 1980)

- De forma habitual se recurre al concepto de congüidad física de primer orden, donde w_{ij} es igual a 1 si las regiones i y j son físicamente adyacentes o a 0 en caso contrario (se asume por definición que $w_{ii} = 0$)

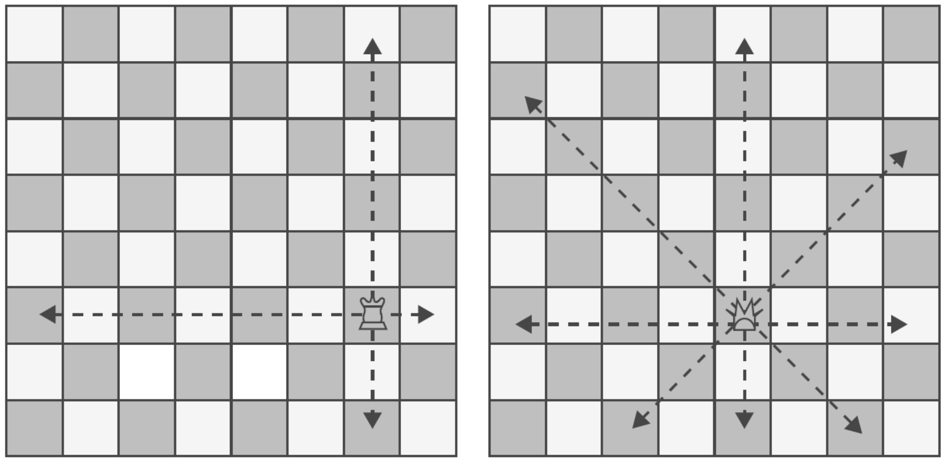
Autocorrelación o dependencia espacial

Matriz de pesos como instrumento para recoger las interdependencias

Existen diversos criterios para la identificación de las regiones vecinas:

Cuadro 1. Criterios de contigüidad física en una cuadrícula regular

Criterio de vecindad	Número total de vecinos	Definición
Criterio lineal	2	Serán vecinas de i las regiones que comparten el lado izquierdo o derecho de i
Criterio torre o *rook*	4	Serán vecinas de i las regiones que comparten algún lado con i
Criterio alfil o *bishop*	4	Serán vecinas de i las regiones que comparten algún vértice con i
Criterio reina o *queen*	8	Serán vecinas de i las regiones que comparten algún lado o vértice con i



Autocorrelación o dependencia espacial

Matriz de pesos como instrumento para recoger las interdependencias

Limitaciones de la matriz $\mathbf{W}$:

- es simétrica, no siendo posible incorporar influencias no recíprocas, violando el segundo de los cinco principios básicos de la econometría espacial
- considera la adyacencia física como único determinante de las interdependencias regionales, descuidando con ello, por ejemplo, posibles influencias mutuas entre regiones, que, aun estando alejadas, mantienen estrechas relaciones comerciales

Autocorrelación o dependencia espacial

Matriz de pesos como instrumento para recoger las interdependencias

Definiciones de $\mathbf{W}$ basadas en la utilización de la distancia entre regiones:

- Cliff y Ord (1973, 1981): $w_{ij} = d_{ij}^{-a}\beta_{ij}^{b}$
  d_{ij}: distancia entre i y j
  β_{ij}: longitud relativa de la frontera común entre i y j con relación al perámetro de i
  a y b: parámetros a estimar

- Dacey (1968): $w_{ij} = \gamma_{ij}\beta_{ij}\alpha_i$
  β_{ij}: igual que antes
  γ_{ij}: es un factor de contgüidad binario
  α_i: es el área de la región i en relación al área total del sistema

- Anselin (1980): matriz inversa de distancias al cuadrado, de manera que la intensidad de la interdependencia entre dos regiones disminuye con la distancia que separa sus respectivos centros

- Bodson y Peeters (1975): función logística que mide la influencia de varios canales de comunicación entre regiones como podrían ser las carreteras, el ferrocarril y otros medios de transporte

$$w_{ij} = \sum_{n=1}^{N} K_N \left\{ \frac{a}{1 + be^{-c_j d_{ij}}} \right\}$$

K_N: la importancia relativa del medio de comunicación n
d_{ij}: la distancia entre dos regiones i y j N: el número de medios de comunicación
a, b y c_j: parámetros a estimar

Autocorrelación o dependencia espacial

Matriz de pesos como instrumento para recoger las interdependencias

Definiciones de $\mathbf{W}$ basadas en la utilización de la distancia entre regiones:

- Case et al. (1993): $\mathbf{W}$ basada en distancias económicas, $w_{ij} = \frac{1}{|x_i - x_j|}$ x_i y x_j: observaciones de características socioeconómicas, tales como la renta per capita
- Vayá et al. (1998a, 1998b) y López-Bazo et al. (1999): $\mathbf{W}$ recoge el grado de intercambio comercial entre regiones analizadas

Otras consideraciones sobre $\mathbf{W}$:

- La matriz de pesos debe ser exógena
- Estandarización de la matriz $\mathbf{W}$
  - se divide cada elemento w_{ij} por la suma total de la fila a la que pertenece, de forma que la suma de cada fila será igual a la unidad

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{j}^{n} w_{ij}}$$

  - la posibilidad de ponderar por igual la influencia total que recibe cada región de sus vecinas, con independencia del número total de vecinos de cada una de ellas, explicaría dicha transformación
  - Anselin (1988) plantea que la estandarización de $\mathbf{W}$ no es siempre adecuada, especialmente cuando ésta se basa en un concepto de distancia dado que, en este caso, la matriz estandarizada carecería de significado

Autocorrelación o dependencia espacial

Matriz de pesos como instrumento para recoger las interdependencias

Un ejemplo de la matriz $\mathbf{W}$:

Matriz $\mathbf{W}$

	1 North East	2 North West	3 Yorksh-Humber	4 East Midlands	5 West Midlands	6 East of England	7 London	8 South East	9 South West	10 Wales	11 Scotland	12 Northern Ireland	Row sum
1 North East	0	1	1	0	0	0	0	0	0	0	1	0	3
2 North West	1	0	1	1	1	0	0	0	0	1	1	0	6
3 Yorkshire and The Humber	1	1	0	1	0	0	0	0	0	0	0	0	3
4 East Midlands	0	1	1	0	1	1	0	1	0	0	0	0	5
5 West Midlands	0	1	0	1	0	0	0	1	1	1	0	0	5
6 East of England	0	0	0	1	0	0	1	1	0	0	0	0	3
7 London	0	0	0	0	0	1	0	1	0	0	0	0	2
8 South East	0	0	0	1	1	1	1	0	1	0	0	0	5
9 South West	0	0	0	0	1	0	0	1	0	1	0	0	3
10 Wales	0	1	0	0	1	0	0	0	1	0	0	0	3
11 Scotland	1	1	0	0	0	0	0	0	0	0	0	0	2
12 Northern Ireland	0	0	0	0	0	0	0	0	0	0	0	0	0

Matriz $\mathbf{W}$ estandarizada ($\mathbf{W}^*$)

	1 North East	2 North West	3 Yorksh-Humber	4 East Midlands	5 West Midlands	6 East of England	7 London	8 South East	9 South West	10 Wales	11 Scotland	12 Northern Ireland	Row sum
1 North East	0	0.3333	0.3333	0	0	0	0	0	0	0	0.3333	0	1
2 North West	0.1667	0	0.1667	0.1667	0.1667	0	0	0	0	0.1667	0.1667	0	1
3 Yorkshire and The Humber	0.3333	0.3333	0	0.3333	0	0	0	0	0	0	0	0	1
4 East Midlands	0	0.2	0.2	0	0.2	0.2	0	0.2	0	0	0	0	1
5 West Midlands	0	0.2	0	0.2	0	0	0	0.2	0.2	0.2	0	0	1
6 East of England	0	0	0	0.3333	0	0	0.3333	0.3333	0	0	0	0	1
7 London	0	0	0	0	0	0.5	0	0.5	0	0	0	0	1
8 South East	0	0	0	0.2	0.2	0.2	0.2	0	0.2	0	0	0	1
9 South West	0	0	0	0	0.3333	0	0	0.3333	0	0.3333	0	0	1
10 Wales	0	0.3333	0	0	0.3333	0	0	0	0.3333	0	0	0	1
11 Scotland	0.5	0.5	0	0	0	0	0	0	0	0	0	0	1
12 Northern Ireland	0	0	0	0	0	0	0	0	0	0	0	0	0

Autocorrelación o dependencia espacial

El operador de retardo espacial

El retardo espacial de una variable resulta del producto de la matriz W y la varaible que se quiere retardar espacialmente:

$$L(y) = \mathbf{W}^* y = \sum_{j}^{n} w^*_{ij} y_j$$

$$\mathbf{W^*} = \begin{vmatrix} 0 & 0.3333 & 0.3333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3333 \\ 0.1667 & 0 & 0.1667 & 0.1667 & 0.1667 & 0 & 0 & 0 & 0 & 0.1667 & 0.1667 \\ 0.3333 & 0.3333 & 0 & 0.3333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0.2 & 0 & 0.2 & 0.2 & 0 & 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0.2 & 0 & 0 & 0 & 0.2 & 0.2 & 0.2 & 0 \\ 0 & 0 & 0 & 0.3333 & 0 & 0 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3333 & 0 & 0 & 0.3333 & 0 & 0.3333 & 0 \\ 0 & 0.3333 & 0 & 0 & 0.3333 & 0 & 0 & 0 & 0.3333 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

y	$\mathbf{W^*}y$
86.2	90.06667
88.6	87.93333
84.7	88.00000
89.2	93.50000
89.1	91.48000
96.8	112.40000
139.7	102.55000
108.3	100.92000
89.8	92.96667
81.5	89.16667
96.9	87.40000

Análisis exploratorio de datos en el espacio

- El análisis exploratorio de datos espaciales (ESDA por sus siglas en inglés) se centra de forma explícita en los efectos espaciales:
  - identificar localizaciones atípicas (*outliers* espaciales)
  - descubrir esquemas de asociación espacial (*cluster* espacial)
  - sugerir diferentes regímenes espaciales u otras formas de inestabilidad espacial

- El centro de este concepto lo ocupa la noción de autocorrelación espacial, es decir, el fenómeno por el cual la similitud locacional (observaciones con proximidad espacial) se une con la similitud de valores (correlación de atributos)

- Dimensiones del ESDA:
  - distinción entre indicadores globales y locales de asociación espacial
  - distinción entre los estadísticos basados en la vecindad y la distancia

Análisis exploratorio de datos en el espacio

Indicadores globales de asociación espacial

- La dependencia espacial se resume en un sólo indicador
- Suelen utilizarse para conocer el rango de interacción espacial en los datos
- Estadísticos: I de Moran y C de Geary

Indicadores locales de asociación espacial (LISA por sus siglas en inglés)

- Un LISA es un indicador que consigue dos objetivos:
    - que el valor del estadístico obtenido para cada observación suministre información acerca de la relevancia de una agrupación espacial de valores similares alrededor de la misma
    - que la suma del valor del estadístico para todas las observaciones sea proporcional a un indicador global de asociación espacial

- Los LISA resultan fáciles de interpretar mediante la visualización en un mapa

Análisis exploratorio de datos en el espacio

Modelos de datos en los cuales la autocorrelación espacial puede ser analizada:

- datos geoestadísticos
  - datos puntuales como una muestra de una distribución continua subyacente
  - se asume que la interacción espacial es una función suave de la distancia entre pares de observaciones

- datos *lattice*
  - una colección fija de localizaciones espaciales discretas (puntos o políginos)
  - la interacción espacial se entiende como una función a pasos según la cual una localización interactúa con un grupo dado de vecinos
  - esta perspectiva es la más comúnmente seguida en la estadísitca espacial y ciencias sociales

Análisis confirmatorio de datos espaciales

- El análisis confirmatorio trata los datos espaciales desde una perspectiva de modelización y está constituido por los distintos métodos de estimación, contrastes de especificación y procedimientos de validación necesarios para implementar modelos multivariantes en los que las observaciones son de corte transversal y están georeferenciados

- Tradicionalmente, el modelo suele estimarse en un primer momento sin incorporar ningún tipo de efecto espacial, de forma que los resultados de la estimación del mismo (y especialmente los residuos) sean el punto de partida de los diagnósticos de dependencia espacial

- Idealmente estos diagnósticos apuntan hacia la dirección correcta en que debe introducirse dicha dependencia espacial en el modelo

- Autocorrelación espacial residual: cuando se deduce la existencia de autocorrelación residual, se reespecifica el término de error con el objetivo de incorporar dicha estructura de dependencia espacial en el mismo

- Autocorrelación espacial sustantiva: en este caso se procede a incorporar la variable dependiente retardada espacialmente como una variable explicativa más en el modelo

Análisis confirmatorio de datos espaciales

- La estimación de tales modelos debe realizarse mediante métodos basados en el principio máximo verosímil o en el método genral de los momentos, entre otros

- Una vez hecha la estimación se utilizan diagnósticos y otros procedimientos de validación a fin de seleccionar el más adecuado

- Este conjunto de estadísticos y métodos crean el cuerpo de lo que se conoce como econometría espacial

Regresión espacial

PGD no espacial

En el caso lineal:

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i$$

$$u_i \sim N(0, \sigma^2), i = 1, \ldots, n$$

Supuestos:

- Los valores observados en la localización i son independientes de aquellos en la localización j
- Los residuales son independientes `$\Longrightarrow E[u_i u_j] = E[u_i]E[u_j] = 0$`

El supuesto de independencia simplifica enormemente el modelo, pero puede ser difícil de justificar en algunos contextos

Regresión espacial

PGD no espacial

Con dos vecinos i y j:

$$y_i = \alpha_j y_j + \mathbf{x}_i \boldsymbol{\beta} + u_i$$

$$y_j = \alpha_i y_i + \mathbf{x}_j \boldsymbol{\beta} + u_j$$

$$u_i \sim N(0, \sigma^2), i = 1$$

$$u_j \sim N(0, \sigma^2), j = 2$$

Supuestos:

- Los valores observados en la localización i dependen de aquellos en la localización j y viceversa
- El PGD es "simultáneo"

Regresión espacial

PGD espacial

Con n observaciones, se puede generalizar:

$$y_i = \rho \sum_{j=1}^{n}$$

$$w_{ij} y_j + \mathbf{x}_i \boldsymbol{\beta} + u_i$$

$$u_i \sim N(0, \sigma^2), i = 1, \ldots, n$$

En notación matricial:

$$\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$$

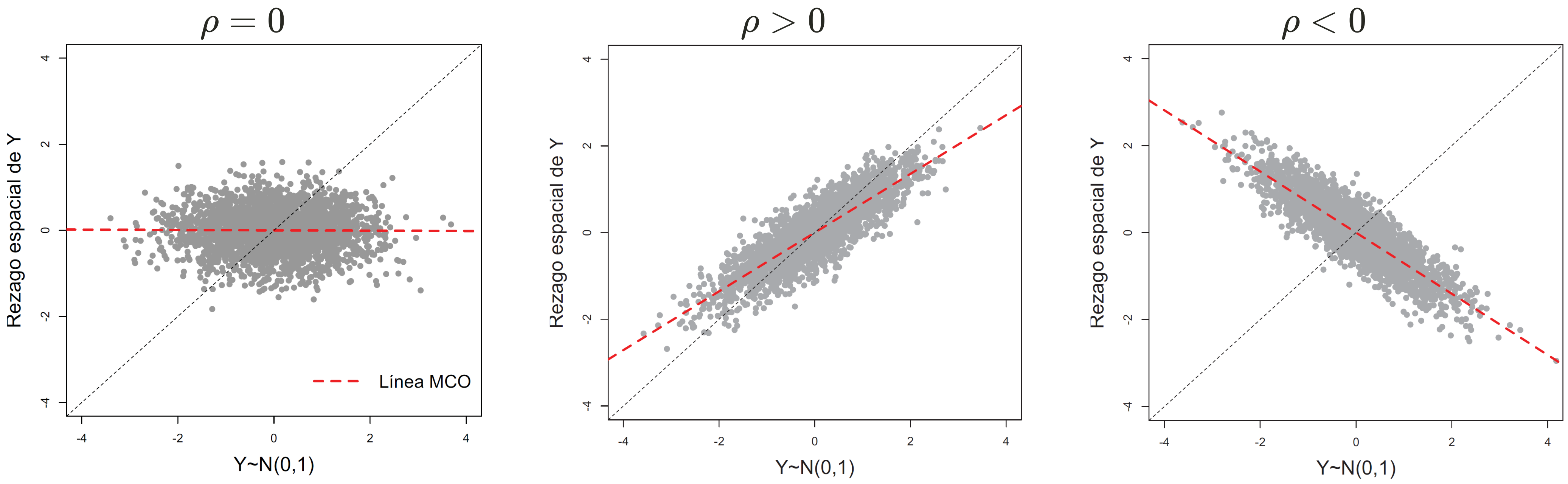
$$\mathbf{u} \sim N(0, \sigma^2 \mathbf{I}_n), i = 1, \ldots, n$$

donde $\mathbf{W}$ es la matriz de pesos espaciales, ρ es un parámetro escalar de autocorrelación espacial e $\mathbf{I}_n$ es una matriz identidad $n \times n$

Regresión espacial

PGD espacial

- Cuando $\rho = 0$, la variable no es espacialmente autocorrelacionada. La información sobre una medida en una localización no nos da información de los valores de las localizaciones vecinas $\Longrightarrow$ independencia espacial
- Cuando $\rho > 0$, la variable es positivamente autocorrelacionada espacialmente. Valores vecinos tienden a ser similares unos a otros $\Longrightarrow$ clustering
- Cuando $\rho < 0$, la variable es negativamente autocorrelacionada espacialmente. Valores vecinos tienden a ser diferentes unos a otros $\Longrightarrow$ segregación



Regresión espacial

Manski (1993) resalta que existen tres diferentes tipos de efectos interactivos que pueden explicar por qué una observación asociada con una localización específica puede depender de otra localización:

1. Efectos de interacción endógenos: donde el comportamiento de una unidad espacial (o sus tomadores de decisiones) en una localización depende de la decisión tomada por otras unidades espaciales
2. Efectos de interacción exógenos: donde el comportamiento de una unidad espacial en una localización depende variables explicativas independientes, de otra unidad espacial
3. Efectos correlacionados: donde similares características no observadas dan como resultado un comportamiento similar

Regresión espacial

El modelo de Manski (1993) o modelo más general (*General Nesting spatial model - GNSM*) en tiene la forma:

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \alpha\boldsymbol{\iota}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{u}$$

$$\mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\epsilon}$$

$\mathbf{WY}$: efectos de interacción endógenos entre las variables dependientes
$\mathbf{WX}$: efectos de interacción exógenos entre las variables independientes
$\mathbf{Wu}$: efectos de interacción entre los términos de error de las diferentes unidades espaciales
ρ: coeficiente espacial autorregresivo
λ: coeficiente espacial de autocorrelación
θ y β: vectores $K \times 1$ de parámetros fijos y desconocidos

El modelo espacio-tiempo se extiende con efectos específicos espaciales y de tiempo, y tiene la forma:

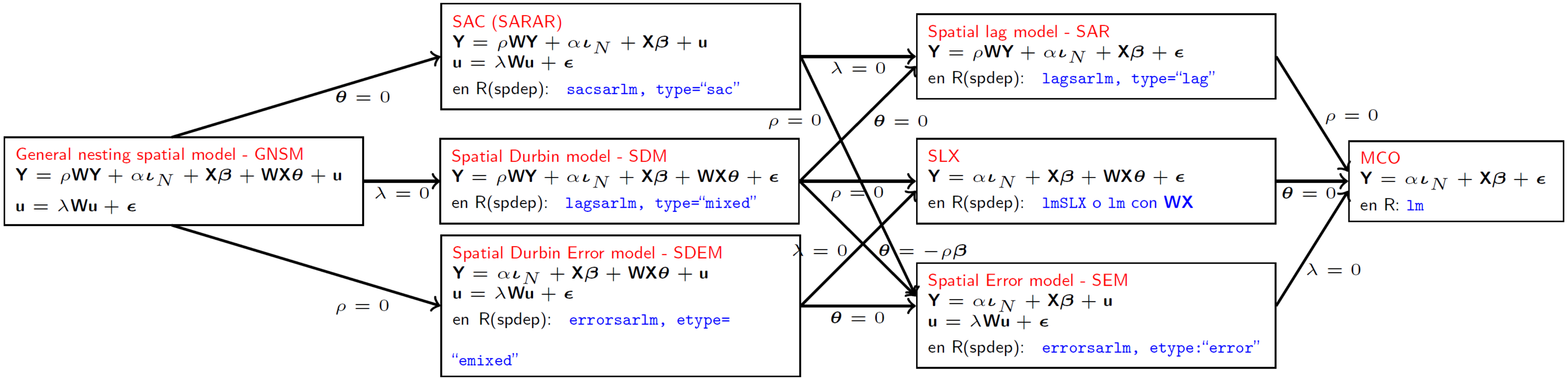
$$\mathbf{Y}_t = \rho\mathbf{W}\mathbf{Y}_t + \alpha\boldsymbol{\iota}_N + \mathbf{X}_t\boldsymbol{\beta} + \mathbf{W}\mathbf{X}_t\boldsymbol{\theta} + \boldsymbol{\mu} + \xi_t\iota_N + \mathbf{u}_t$$

$$\mathbf{u}_t = \lambda\mathbf{W}\mathbf{u}_t + \boldsymbol{\epsilon}_t$$

donde $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$. Los efectos específicos de espacio y tiempo pueden ser tratados como efectos fijos o efectos aleatorios

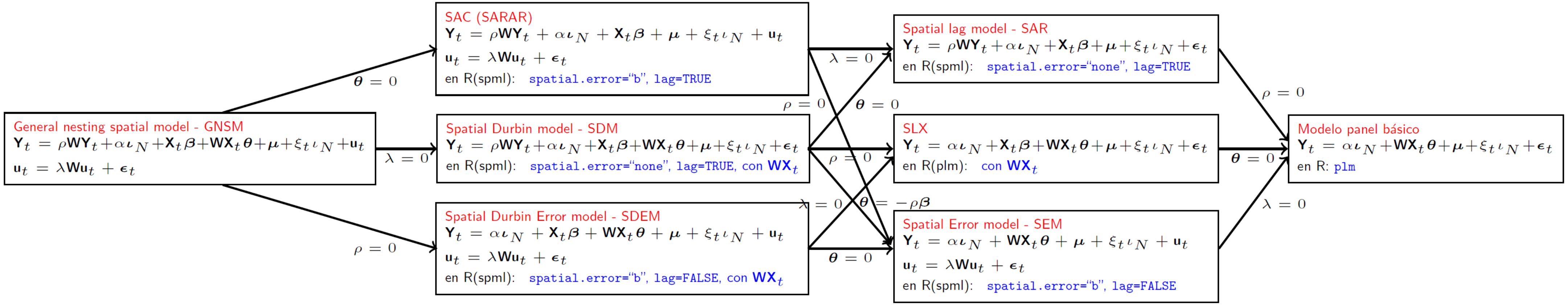
Regresión espacial

El siguiente diagrama resume una familia de 8 modelos econométricos lineales espaciales en sección cruzada:



Regresión espacial

El siguiente diagrama resume una familia de 8 modelos econométricos lineales espaciales en datos panel (con efectos fijos o aleatorios, y con o sin efectos específicos de tiempo):



Regresión espacial

Ineficiencia de los estimadores MCO

- En un contexto de series de tiempo, los estimadores MCO siguen siendo consistentes incluso cuando la variable dependiente rezagada esta presente en el modelo, siempre que el términos de error no muestre correlación serial
- Mientras el estimador puede ser sesgado en muestras pequeñas, éste puede aún ser usado para inferencia asintótica
- En un contexto espacial, estas reglas no se mantienen, independientemente de las propiedades del término de error
- Consideremos el modelo SAR de primer orden en corte transversal (covariables omitidas):

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \boldsymbol{\epsilon}$$

- El estimador MCO para ρ es:

$$\hat{\rho} = \left((\mathbf{W}\mathbf{Y})'(\mathbf{W}\mathbf{Y})\right)^{-1}(\mathbf{W}\mathbf{Y})'\mathbf{Y} = \rho + \left((\mathbf{W}\mathbf{Y})'(\mathbf{W}\mathbf{Y})\right)^{-1}(\mathbf{W}\mathbf{Y})'\boldsymbol{\epsilon}$$

- Similar a series de tiempo, la esperanza del segundo término no es igual a cero

Regresión espacial

Ineficiencia de los estimadores MCO

- Asintóticamente, el estimador MCO será consistente si se cumplen dos condiciones:

$$\text{plim } N^{-1}(\mathbf{WY})'(\mathbf{WY}) = \mathbf{Q} \text{ una matriz finita no singular}$$

$$\text{plim } N^{-1}(\mathbf{WY})'\boldsymbol{\epsilon} = \mathbf{0}$$

- Si bien la primera condición se puede satisfacer con restricciones adecuadas sobre ρ y la estructura de $\mathbf{W}$, la segunda no se mantiene en el caso especial:

$$\text{plim } N^{-1}(\mathbf{WY})'\boldsymbol{\epsilon} = \text{plim } N^{-1}\boldsymbol{\epsilon}'(\mathbf{W})(\mathbf{I}_N - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon} \neq \mathbf{0}$$

- La presencia de $\mathbf{W}$ en la expresión resulta en una forma cuadrática en el término de error
- A menos que $\rho = 0$, el plim no convergerá a cero

Regresión espacial

Métodos de estimación

Tres métodos han sido desarrollados en la literatura para estimar modelos que incluyen efectos de interacción espacial:

- ML: Máxima verosimilitud
- IV/GMM: variables instrumentales o método de los momentos generalizados
- MCMC: *Bayesian Markov Chain Monte Carlo approach*

Regresión espacial

Comparación de modelos

- El esquema de los 7 modelos espaciales parece sugerir que la mejor estrategia para probar por efectos de interacción espacial es iniciar por el modelo general: $GNSM$
- Sin embargo, como Manski (1993) plantea, al menos uno de los $K+2$ efectos de interacción debe ser excluido, ya que de otra forma los parámetros no podrán ser identificados $\Longrightarrow$ De acuerdo a Manski (1993) no existen limitaciones técnicas para estimar el modelo, pero los parámetros estimados no podrán interpretarse de manera adecuada, ya que los efectos endógenos y exógenos no pueden distinguirse entre sí
- Se esta divido entonces en aplicar dos enfoques: general-a-específico o específico-a-general
- LeSage y Pace (2009) argumentan que el modelo Durbin espacial es el mejor punto de partida
- Florax et al. (2003) han encontrado que una expansión de una ecuación de regresión lineal con variables espacialmente rezagadas, condicional a los resultados sobre pruebas de especificación incorrectas, supera el enfoque general-a-específico para encontrar el verdadero proceso generador de datos

Regresión espacial

Comparación de modelos

Partiendo del siguiente modelo de regresión lineal:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

la hipótesis de no autocorrelación espacial puede ser contrastada a partir de los siguientes estadísticos:

- La I de Moran
- Tests estadísticos basados en el principio de multiplicadores de Lagrange (LM):
  - LM-Lag
  - Robust LM-Lag
  - LM-Error
  - Robust LM-Error
  - LM-SARMA
- En los tests LM los dos primeros se refieren al *spatial lag model* como alternativa, los siguientes dos se refieren al *spatial error model* como alternativa y el último se refiere a un modelo con *spatial lag* y *spatial error*

Regresión espacial

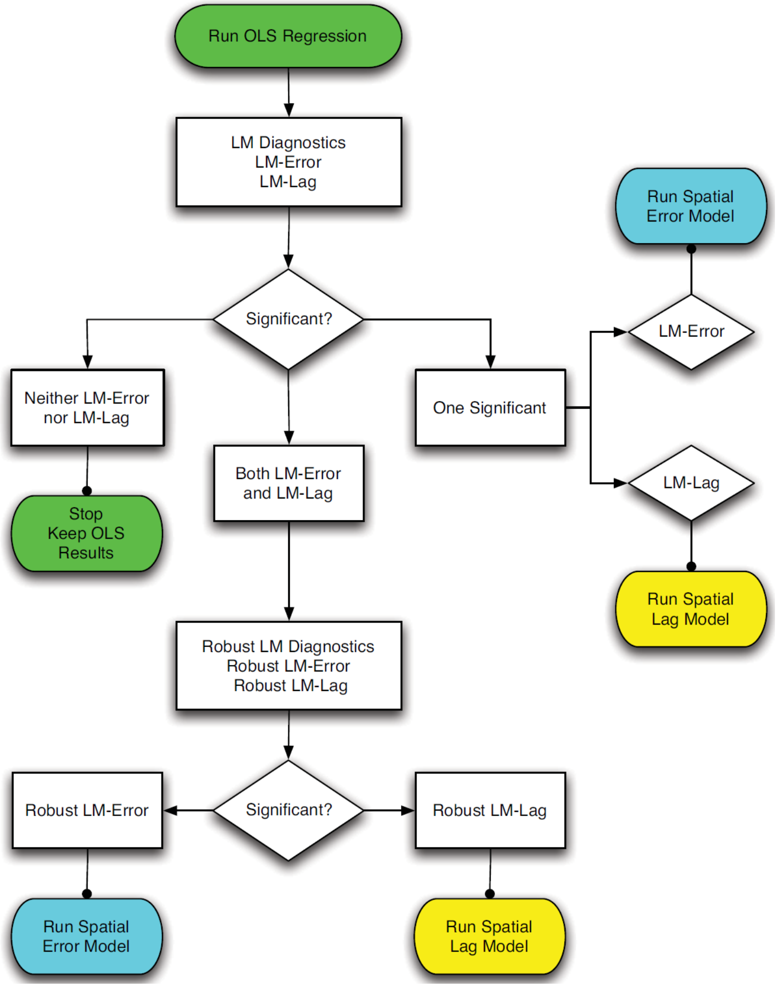
Comparación de modelos

- Todos los tests comparten una misma hipótesis nula: la ausencia de dependencia espacial
- Sin embargo, el test de la I de Moran no tiene una hipótesis alternativa claramente definida. Con lo cual no sirve para discriminar entre la existencia de un esquema de autocorrelación espacial residual o en la variable dependiente
- Por otra parte, tanto la I de Moran como los contrastes basados en el principio de LM requieren de la normalidad del término de perturbación así como la linealidad del modelo de regresión
- Los tests robustos (Robust LM-Lag y Robust LM-Error) prueban la dependencia espacial robusto a la presencia del otro. Es decir, que el Robust LM-Error prueba por dependencia en el error en la posible presencia errónea de una variable endógena retardada espacialmente, y el Robust LM-Lag al revés
- Lo importante a recordar respecto a los tests robustos es que estos sólo deben considerarse cuando la versión estándar (LM-Lag o LM-Error) son estadísticamente significativos
- Los tests robustos ayudan a determinar que tipo de dependencia espacial existe. Si las medidas robustas de lag y error son ambas significativas, la estructura de la autocorrelación espacial estará determinada por el test robusto con mayor valor

Regresión espacial

Comparación de modelos

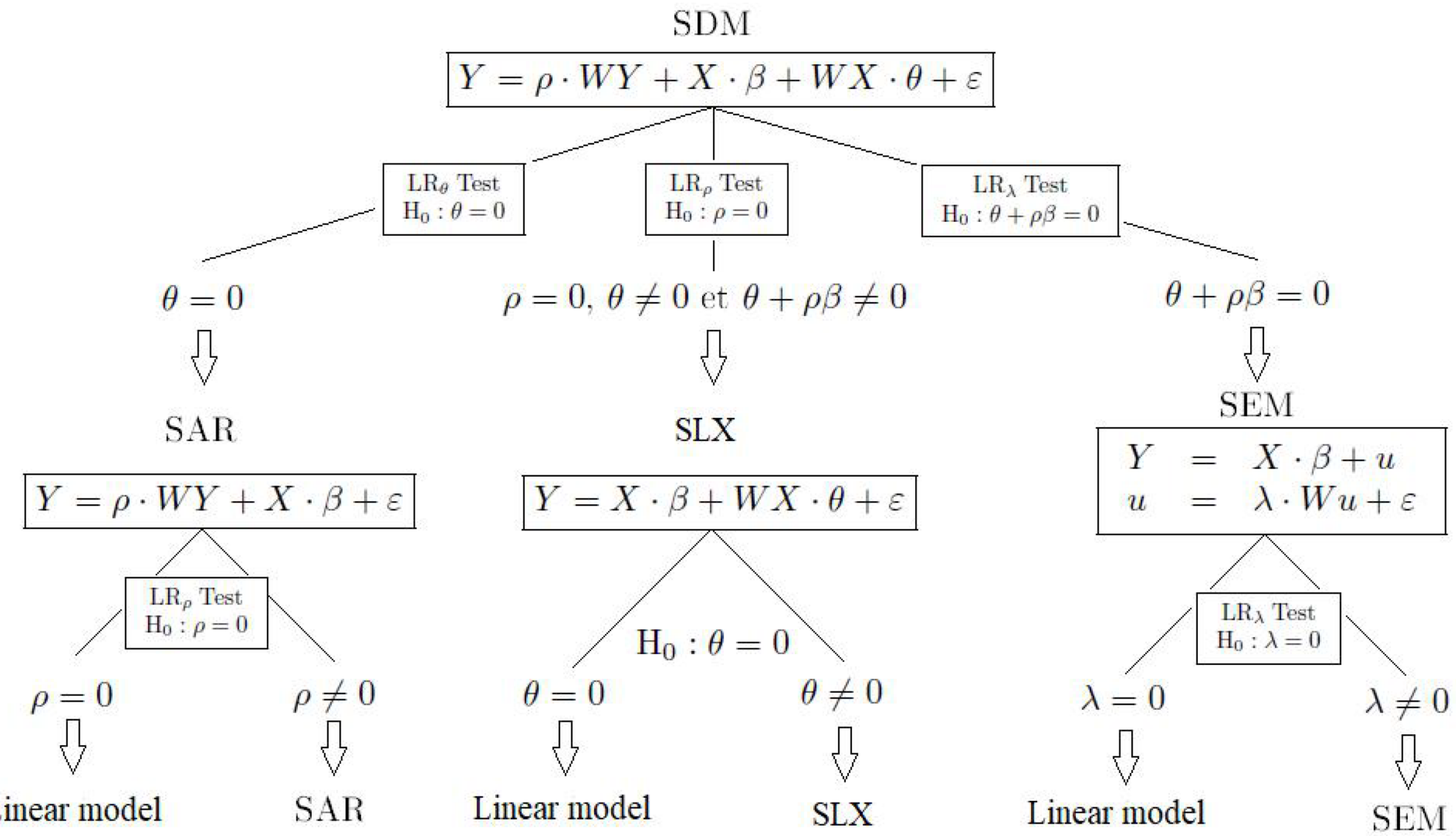
La estrategia simple propuesta por Anselin (2005) y Florax et al (2003) es (*The bottom-up approach*):



Regresión espacial

Comparación de modelos

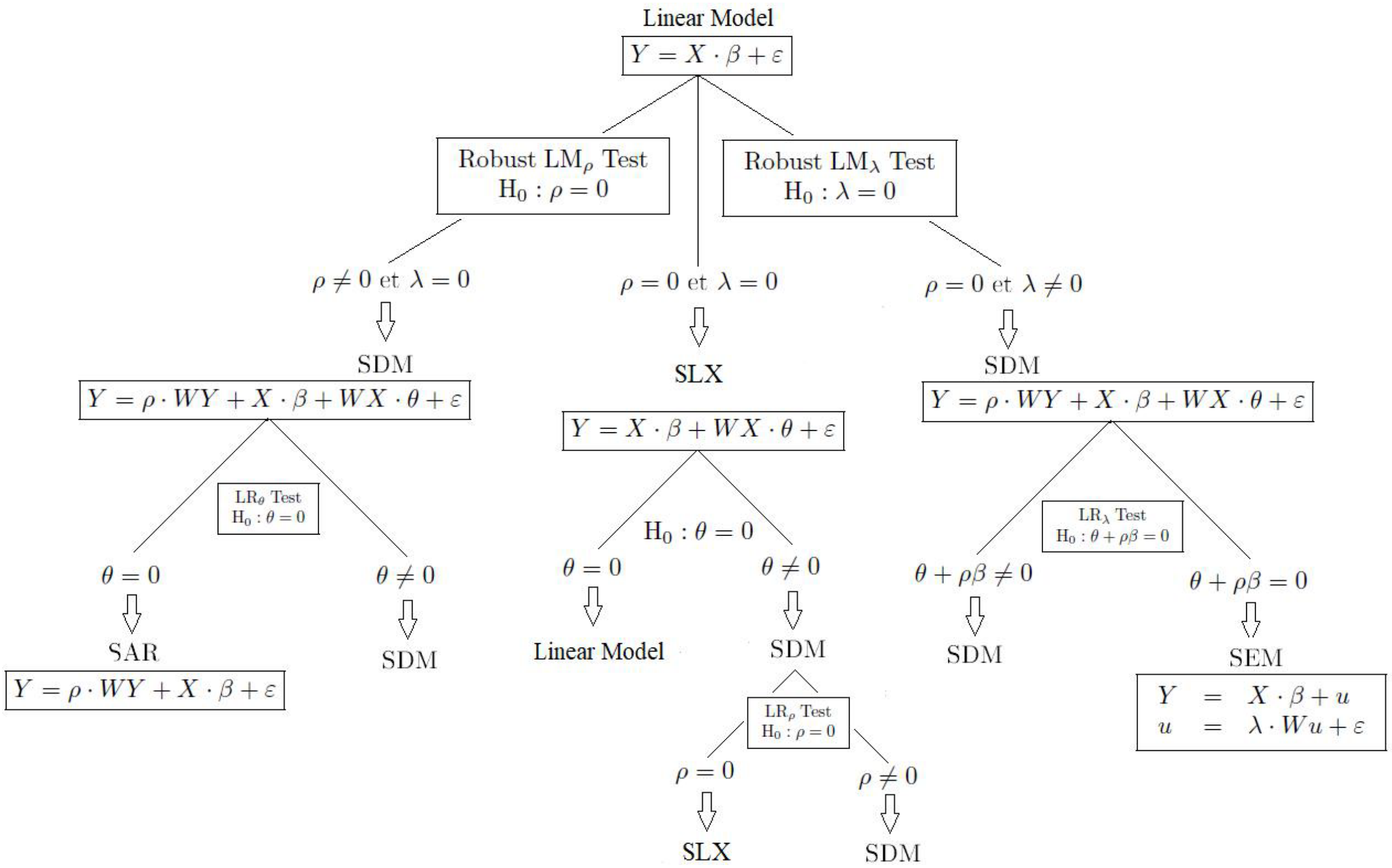
La estrategia de LeSage y Pace (2009) (*The top-down approach*):



Regresión espacial

[Comparación de modelos]

La estrategia de Elhorst (2010):



Regresión espacial

Comparación de modelos

Por tanto, Elhorst (2010) propone el siguiente procedimiento para determinar qué modelo es el candidato más probable para explicar los datos:

1. Estime el modelo por MCO y pruebe si el *spatial lag model - SAR* o el *spatial error model - SEM* es más apropiado para describir los datos. Para esto, se puede utilizar el LM-test propuesto por Anselin (1988) y el robust LM-test propuesto por Anselin et al. (1996)

2. Si el modelo MCO es rechazado en favor del *spatial lag model*, *spatial error model* o en favor de ambos modelos, entonces el modelo Durbin espacial debería ser estimado

3. Si estos modelos son estimados por ML, un LR test puede ser usado para probar las hipótesis:
    - $H_0 : \theta = 0 \implies$ prueba si el modelo Durbin espacial se puede simplificar en el *spatial lag model*
    - $H_0 : \theta + \rho\beta = 0 \implies$ prueba si el modelo Durbin espacial se puede simplificar en el *spatial error model*

        - Si $H_0 : \theta = 0$ y $H_0 : \theta + \rho\beta = 0$ son rechazadas $\implies$ modelo Durbin espacial
        - Si $H_0 : \theta = 0$ no puede ser rechazada $\implies$ *spatial lag model* (siempre que el (robust) LM-test diga lo mismo)
        - Si $H_0 : \theta + \rho\beta = 0$ no puede ser rechazada $\implies$ *spatial error model* (siempre que el (robust) LM-test diga lo mismo)
        - Si el (robust) LM-test a punta a un modelo diferente que el LR test $\implies$ modelo Durbin espacial

4. Si el modelo MCO es estimado y no rechaza en favor del *spatial lag model* y el *spatial error model*, el modelo MCO deberá ser re-estimado incluyendo $\mathbf{WX}$ o una particular selección de estas K variables para probar $H_0 : \theta = 0$

    - Si la hipótesis no puede ser rechazada $\implies$ modelo MCO
    - Si la hipótesis es rechazada $\implies$ modelo Durbin espacial y se debe contrastar $H_0 : \rho = 0$
    - Si ésta última hipótesis es rechazada $\implies$ modelo Durbin espacial
    - Si ésta última hipótesis es no rechazada $\implies$ modelo SLX

Regresión espacial

Efectos directos e indirectos (o *spatial spillover effects*)

- Muchos estudios empíricos usan las estimaciones puntuales de una o más especificaciones de modelos de regresión espacial $(\rho$, θ y/o $\lambda)$ para deducir conclusiones sobre la existencia de spillovers espaciales $\Longrightarrow$ LeSage y Pace (2009, p. 74) plantean y demuestran que esta práctica lleva a erróneas conclusiones

- Re-escribiendo el modelo *GNSM*:

$$\mathbf{Y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta}) + \mathbf{R}$$

donde $\mathbf{R}$ es un término que contiene el intercepto y el término de error, la matriz de derivadas parciales del valor esperado de $\mathbf{Y}$ con respecto a la k-ésima variable explicatoria de $\mathbf{X}$ es:

$$\begin{bmatrix} \frac{\partial E(\mathbf{Y})}{\partial x_{1k}} & . & \frac{\partial E(Y)}{\partial x_{Nk}} \end{bmatrix} = \begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & . & \frac{\partial E(y_1)}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial E(y_N)}{\partial x_{1k}} & . & \frac{\partial E(y_N)}{\partial x_{Nk}} \end{bmatrix} = (\mathbf{I} - \rho\mathbf{W})^{-1} \begin{bmatrix} \beta_k & w_{12}\theta_k & . & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & . & w_{2N}\theta_k \\ . & . & . & . \\ w_{N1}\theta_k & w_{N2}\theta_k & . & \beta_k \end{bmatrix}$$

Regresión espacial

Efectos directos e indirectos (o *spatial spillover effects*)

Esta derivada parcial tiene tres importantes propiedades:

1. Si una variable explicatoria particular en una unidad espacial particular cambia, no sólo cambiará la variable dependiente en esa unidad espacial, sino también cambiará en otras unidades espaciales

    - El primer cambio se llama efecto directo: cada elemento de la diagonal
    - El segundo cambio se llama efecto indirecto: cada elemento por fuera de la diagonal

    Los efectos indirectos no ocurren si $\rho = 0$ y $\theta_k = 0$, ya que los elementos por fuera de la diagonal serían cero

2. Los efectos directos e indirectos son diferentes para diferentes unidades espaciales en la muestra
    - Los efectos directos son diferentes ya que los elementos de la diagonal de la matriz $(\mathbf{I} - \rho\mathbf{W})^{-1}$ son diferentes para cada unidad espacial, siempre que $\rho \neq 0$
    - Los efectos indirectos son diferentes ya que los elementos de la diagonal de la matriz $(\mathbf{I} - \rho\mathbf{W})^{-1}$ y de la matriz $\mathbf{W}$ son diferentes entre unidades espaciales, siempre que $\rho \neq 0$ y/o $\theta_k \neq 0$

3. Los efectos indirectos que ocurren si $\theta_k \neq 0$ son conocidos como efectos locales, como opuesto al efecto indirecto que ocurre si $\rho \neq 0$ que son llamados efectos globales

Regresión espacial

Efectos directos e indirectos (o *spatial spillover effects*)

- Ya que los efectos directos e indirectos son diferentes para diferentes unidades espaciales, la presentación de estos efectos puede ser problemático $\Longrightarrow$ N unidades espaciales y K variables explicatorias, se obtiene K diferentes $N \times N$ matrices de efectos directos e indirectos

- LeSage y Pace (2009) proponen lo siguiente:
    - Efectos directos: promedio de los elementos de la diagonal de la matriz
    - Efectos indirectos: promedio de las filas o las columnas de los elementos por fuera de la diagonal

- Promedio de las filas: el impacto sobre un particular unidad espacial de la variable dependiente ante un cambio en todas las variables explicatorias en todas las unidades espaciales

- Promedio de las columnas: el impacto de cambiar la variable explicatoria en una unidad espacial sobre la variable dependiente de todas las unidades espaciales

- Sin embargo, numéricamente las dos magnitudes son iguales, por lo cual no importa por cual se opte

- Interpretación del efecto indirecto: el impacto de cambiar un particular elemento de una variable exógena sobre la variable dependiente de todas las otras unidades espaciales, lo cual corresponde al promedio del efecto columna

Regresión espacial

Efectos directos e indirectos (o *spatial spillover effects*)

Cuadro 1. Efectos directos e indirectos de diferentes modelos

	Efecto directo	**Efecto indirecto**
MCO/SEM	β_k	0
SAR/SAC	Diagonal de $(\mathbf{I}-\rho\mathbf{W})^{-1}\beta_k$	*Off-diagonal* de $(\mathbf{I}-\rho\mathbf{W})^{-1}\beta_k$
SLX/SDEM	β_k	θ_k
SDM/GNSM	Diagonal de $(\mathbf{I}-\rho\mathbf{W})^{-1}(\beta_k+\mathbf{W}\theta_k)$	*Off-diagonal* de $(\mathbf{I}-\rho\mathbf{W})^{-1}(\beta_k+\mathbf{W}\theta_k)$

Fuente: Halleck Vega y Elhorst (2015)

Ejercicio aplicado en R

En este ejercicio aplicado se va analizar los crímenes violentos en los estados de los Estados Unidos. En los siguientes links se encuentran los datos, el shapefile y el código utilizado en R:

- Datos
- Shapefile
- Código en R