Econometria Espacial

Ciencia de Datos y Econometría Aplicada

Ignacio Sarmiento-Barbieri

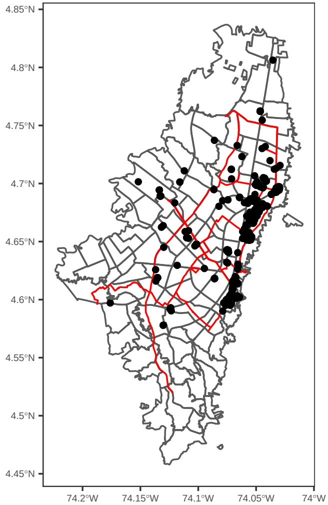
Universidad de los Andes

Agenda

Spatial Econometrics: Motivation

$$y = X\beta + u$$

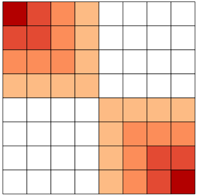
- Independence assumption between observation is no longer valid
- Attributes of observation i may influence the attributes of observation j.
- We will consider various alternatives to model spatial dependence

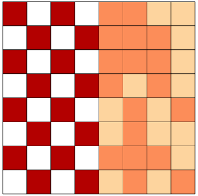


Spatial Econometrics: Motivation

$$y = X\beta + u$$

- Independence assumption between observation is no longer valid
- Attributes of observation i may influence the attributes of observation j.
- Positive Spatial correlation arises when units that are *close* to one another are more similar than units that are far apart
- Similarly spatial heterogeneity arises when some areas present more variability than others





Spatial Econometrics: Closeness

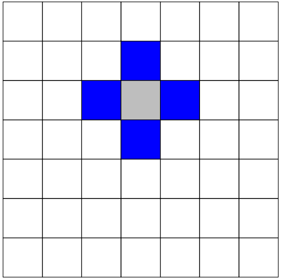
"Everything is related to everything else, but close things are more related than things that are far apart" (Tobler, 1979).

- One of the major differences between standard econometrics and standard spatial econometrics lies, in the fact that, in order to treat spatial data, we need to use two different sets of information

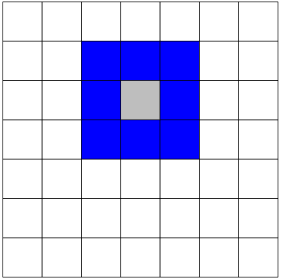
  1. Observed values of the economic variables
  2. Particular location where those variables are observed and to the various links of proximity between all spatial observations

Spatial Econometrics: Closeness

Rook criterion: two units are close to one another if they share a side



Queen criterion: two units are close if they share a side or an edge.



Spatial Econometrics: Weights Matrix

- At the heart of traditional spatial econometrics is the definition of the *weights matrix*:

$$W = \begin{pmatrix} w_{11} & \dots & \dots & w_{n1} \\ \vdots & w_{ij} & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{n_1} & \dots & \dots & w_{nn} \end{pmatrix}_{n \times n} \tag{1}$$

with generic element:

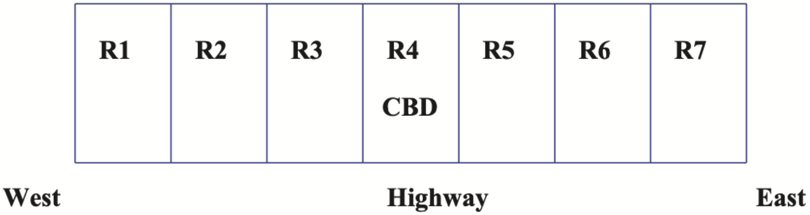
$$w_{ij} = \begin{cases} 1 & if\, j \in N(i) \\ 0 & o.w \end{cases} \tag{2}$$

$N(i)$ being the set of neighbors of location j. By convention, the diagonal elements are set to zero, i.e. $w_{ii} = 0$.

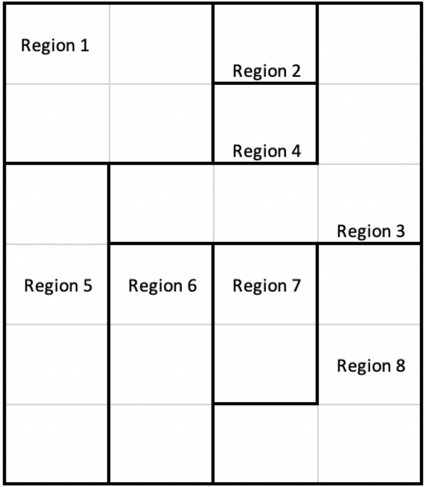
Spatial Econometrics: Weights Matrix

- The specification of the neighboring set ($N(i)$) is quite arbitrary and there's a wide range of suggestions in the literature.
  - Rook criterion
  - Queen criterion
  - Two observations are neighbors if they are within a certain distance, i.e., $j \in N(i)$ if $d_{ij} < d_{max}$ where d is the distance between location i and j.
  - Closest neighbor, ties can be solved randomly
  - More general matrices can also be specified by considering entries of w_{ij} as functions of geographical, economic or social distances between areas rather than simply characterized by dichotomous entries

W matrices are standardized to sum to one in each row



Some Examples of Weights Matrices



Testing for Spatial Dependence

- The problem with ignoring the spatial structure of the data implies that the OLS estimates in the non spatial model may be biased, inconsistent and/or inefficient, depending on what is the true underlying dependence (for more see Anselin and Bera (1998)).

$$e = y - X\hat{\beta}$$

Spatial Lag Model

- We can use the OLS reiduals to test for spatial correlation.
- The most basic one is Moran's I test (1950), a test statistics for the null of uncorrelation among regression residuals.

$$I = \left(\frac{e'We}{e'e} \right) \tag{3}$$

- where $e = y - X\beta$ is a vector of OLS residuals $\beta = (X'X)^{-1}X'y$, W is the row standardized spatial weights matrix
- Moran's I test was originally developed as a two-dimensional analog of Durbin-Watson's test

Spatial Lag Model

- We can think of situations where values observed at one location or region, say observation i, depend on the values of neighboring observations at nearby locations.

$$y_i = \rho_i y_j + X_i \beta + \epsilon_i \tag{4}$$
$$y_j = \rho_j y_i + X_j \beta + \epsilon_j \tag{5}$$

- This situation suggests a simultaneous data generating process, where the value taken by y_i depends on that of y_j and vice versa.

Spatial Lag Model

- Spatial lag dependence in a regression setting can be modeled similar to an autoregressive process in time series. Formally,

$$y = \rho W y + X\beta + u$$

- Wy induces a nonzero correlation with the error term, similar to the presence of an endogenous variable.
- Unlike to time series, Wy_i is always correlated with u
- OLS estimates in the non spatial model will be biased and inconsistent. (Anselin and Bera, 1998)
- In `R` the function `lagsarlm` uses MLE

Spatial Lag Model

The model is then

$$y = \rho W y + X\beta + u$$

with $|\rho| < 1$, we also assume that W is exogenous

If W is row standardized:

- Guarantees $|\rho| < 1$ (Anselin, 1982)
- $[0, 1]$ Weights
- Wy Average of neighboring values
- W is no longer symmetric $\sum_j w_{ij} \neq \sum_i w_{ji}$ (complicates computation)

Spatial Lag Model
Maximum Likelihood Estimator

Note that we can write

$$(I - \rho W)y = X\beta + u$$

- We can think this model as a way to correct for loss of information coming from spatial dependence.
- $(1 - \rho W)y$ is a spatially filtered dependent variable, i.e., the effect of spatial autocorrelation taken out

Spatial Lag Model

- In this case, endogeneity emerges because the spatially lagged value of y is correlated with the stochastic disturbance.
- One solution that emerged in the literature is MLE
- We need an extra assumption, i.e., $u \sim_{iid} N(0, \sigma^2 I)$.

Spatial Lag Model

Maximum Likelihood Estimator

The associated likelihood function is then

$$\mathcal{L}\left(\sigma^2,\rho,y\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n |\sigma^2\Omega|^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(y-(I-\rho W)^{-1}X\beta)'\Omega^{-1}(y-(I-\rho W)^{-1}X\beta)\right\}$$

the log likelihood

$$l\left(\sigma^2,\rho,y\right) = constant - \frac{1}{2} ln|\sigma^2\Omega| - \frac{1}{2\sigma^2}(y-(I-\rho W)^{-1}X\beta)'\Omega^{-1}(y-(I-\rho W)^{-1}X\beta)$$

Spatial Lag Model

Maximum Likelihood Estimator

so returning to the log likelihood we have that the log likelihood is

$$
\begin{aligned}
l\left(\sigma^2, \rho, y\right) = {} & constant - \frac{n}{2} ln\left(\sigma^2\right) + ln\left(\left|(I - \rho W)\right|\right) \\
& - \frac{1}{2\sigma^2}(y - (I - \rho W)^{-1} X\beta)'(I - \rho W)'(I - \rho W)(y - (I - \rho W)^{-1} X\beta)
\end{aligned} \tag{6}
$$

then

$$
\begin{aligned}
l\left(\sigma^2, \rho, y\right) = {} & constant - \frac{n}{2} ln\left(\sigma^2\right) \\
& - \frac{1}{2\sigma^2}((I - \rho W)y - X\beta)'((I - \rho W) - X\beta) \\
& + ln\left(\left|(I - \rho W)\right|\right)
\end{aligned} \tag{7}
$$

Spatial Lag Model

Maximum Likelihood Estimator

- The determinant $|(I-\rho W)|$ is quite complicated because in contrast to the time series, where it is a triangular matrix, here it is a full matrix.
- However, Ord (1975) showed that it can be expressed as a function of the eigenvalues ω_i

$$|(I-\rho W)| = \Pi_{i=1}^{n}(1-\rho\omega_i)$$

So the log likelihood is simplified to

$$\begin{aligned} l\left(\sigma^2,\rho,y\right) = & \; constant - \frac{n}{2} ln\left(\sigma^2\right) \\ & - \frac{1}{2\sigma^2}((I-\rho W)y - X\beta)'((I-\rho W)y - X\beta) \\ & + \sum ln(1-\rho\omega_i) \end{aligned} \tag{8}$$

Spatial Lag Model

Maximum Likelihood Estimator

Applying FOC, the ML estimates for β and σ^2 are:

$$\beta_{MLE} = (X'X)^{-1}X'(I - \rho W)y$$

$$\sigma^2_{MLE} = \frac{1}{n}(y - \rho Xy - X\beta_{MLE})'(y - \rho Xy - X\beta_{MLE})$$

- Conditional on ρ these estimates are simply OLS applied to the spatially filtered dependent variable and explanatory variables X.

Spatial Lag Model

Maximum Likelihood Estimator

- Substituting these in the log likelihood we have a concentrated log-likelihood as a nonlinear function of a single parameter ρ

$$l\left(\rho\right) = -\frac{n}{2} ln\left(\frac{1}{n}(e_0 - \rho e_L)'(e_0 - \rho e_L)\right) + \sum ln(1 - \rho\omega_i) \tag{9}$$

- where e_0 are the residuals in a regression of y on X and
- e_L of a regression of Wy on X.

- This expression can be maximized numerically to obtain the estimators for the unknown parameters ρ.

Interpretation of Parameters

- Consider the following model for the $i-th$ observation

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_r x_{ir} + \cdots + \beta_k x_{ik} \;\; i = 1, \ldots, n$$

- Recall that in OLS we have

$$\beta_1 = \frac{\partial y_i}{\partial x_{i1}}$$

or generically

$$\beta_r = \frac{\partial y_i}{\partial x_{ir}} \;\; \forall i = 1, \ldots, n \;\&\; r = 1, \ldots, k$$

$$\beta_r = \frac{\partial y_i}{\partial x_{jr}} \;\; \forall j \neq i \;\&\; \forall r = 1, \ldots, k$$

- Interpretation is straight forward as long as we take into account units
- In spatial models the interpretation is less immediate and require some clarification

Interpretation of Parameters

- Lets consider the case of a simple Spatial Lag model with a single regressor

$$y_i = \alpha + \beta x_i + \lambda \sum w_{ij} y_j + \epsilon_i \tag{10}$$

with $|\lambda| < 1$, and

$$\beta \neq \frac{\partial y_i}{\partial x_i}$$

$$\frac{\partial y_i}{\partial x_i} = diag(I - \lambda W)^{-1}\beta$$

- The impact depends also on the parameter λ
- The impact is different in each location

Interpretation of Parameters

More generally consider

$$
\begin{aligned}
y &= \lambda W y + X\beta + u \\
&= (I - \lambda W)^{-1} X\beta + (I - \lambda W)^{-1} u
\end{aligned}
$$

Then

$$E(y) = (I - \lambda W)^{-1} X\beta \tag{11}$$

we define

$$S(W) = (I - \lambda W)^{-1} \beta \tag{12}$$

Interpretation of Parameters

Therefore the impact of *each variable* x on y can be described through the partial derivatives $\frac{\partial E(y)}{\partial x}$ which can be arranged in the following matrix:

$$S(W) = \frac{\partial E(y)}{\partial x} = \begin{pmatrix} \frac{\partial E(y_1)}{\partial x_1} & \cdots & \frac{\partial E(y_1)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_i)}{\partial x_1} & \cdots & \frac{\partial E(y_i)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_1} & \cdots & \frac{\partial E(y_n)}{\partial x_n} \end{pmatrix} \tag{13}$$

Interpretation of Parameters

On this basis, LeSage and Pace (2009) suggested the following impact measures that can be calculated for each independent variable X_i included in the model

- *Average Direct Impact*: this measure refers to the impact of changes in the $i-th$ observation of x, which we denote x_i, on y_i. This is the average of all diagonal entries in S

$$
\begin{aligned}
ADI &= \frac{tr(S(W))}{n} \\
&= \frac{1}{n}\sum_{i=1}^{n} S(W)_{ii}
\end{aligned}
\tag{14}
$$

Interpretation of Parameters

- *Average Total Impact To* an observation: this measure is related to the impact produced on one single observation y_i. For each observation this is calculated as the sum of the $i - th$ row of matrix S

$$
\begin{aligned}
ATIT_j &= \frac{\iota' S(W)}{n} \\
&= \frac{1}{n} \sum_{i=1}^{n} S(W)_{ij}
\end{aligned}
\tag{15}
$$

Interpretation of Parameters

- *Average Total Impact From* an observation: this measure is related to the total impact on all other observations y_i . For each observation this is calculated as the sum of the $j-th$ column of matrix S

$$
\begin{aligned}
ATIF_i &= \frac{1}{n} S(W)\iota \\
&= \frac{\sum_{j=1}^{n} S(W)_{ij}}{n}
\end{aligned} \tag{16}
$$

Interpretation of Parameters

- A Global measure of the average impact obtained from the two previous measures.
- It is simply the average of all entries of matrix S

$$ATI = \frac{1}{n}\iota' S(W)\iota = \frac{1}{n}\sum_{i=1}^{n} ATIT_i = \frac{1}{n}\sum_{j=1}^{n} ATIF_i \tag{17}$$

- The numerical values of the summary measures for the two forms of average total impacts are equal.
- The ATIF relates how changes in a single observation j influences all observations.
- In contrast, the ATIT considers how changes in all observations influence a single observation i.

Interpretation of Parameters

- *Average Indirect Impact* obtained as the difference between ATI and ADI

$$AII = ATI - ADI \tag{18}$$

- It is simply the average of all off-diagonal entries of matrix S

Spatial Error Model (SEM)

An OVB motivation

- True GDP

$$y = X\beta + Z\theta$$

- but Z is not observed and $Z \perp X$
- we estimate

$$y = X\beta + \epsilon$$

- if Z has a spatial autoregressive process

$$Z = \rho WZ + r$$

Spatial Error Model (SEM)

An OVB motivation

- Then

$$y = X\beta + (I - \rho W)^{-1}(\theta r)$$

- calling $(\theta r) = u$

$$y = X\beta + (I - \rho W)^{-1}u$$

- β will be unbiased but inefficient

Spatial Error Model (SEM)

The model is now

$$y = X\beta + u$$

with

$$u = \rho X u + \epsilon$$

with $|\rho| < 1$, we also assume that W is exogenous We can estimate this by MLE or FGLS

SARAR

Spatial auto-regressive with additional auto-regressive error structure

Let's consider the following model:

$$y = \rho W y + X\beta + u$$

with

$$u = \lambda W u + \epsilon$$

we assume that W is exogenous

If W is row standardized:

- Guarantees $|\lambda| < 1$ $|\rho| < 1$ (Anselin, 1982)

¿Qué tipo de dependencia espacial tenemos?

El problema: OLS asume independencia. Moran nos dice que hay dependencia espacial, pero...

¿De qué tipo?

1. Lag espacial (SAR)
2. Error espacial (SEM)
3. Ambos (SARAR): Combinación de lag + error

Test LM para Lag Espacial

Lagrange Multiplier test for spatial lag dependence (LMlag)

- H_0: $\rho = 0$ (no hay dependencia en la variable dependiente)
- H_1: $\rho \neq 0$ (hay spillovers: y_i depende de y_j vecinos)

Intuición: Si los residuos de OLS están correlacionados con Wy (lag espacial de y), entonces:

- OLS omite un regresor relevante: el lag espacial
- Necesitamos un modelo SAR

Test LM para Error Espacial

Lagrange Multiplier test for spatial error dependence (LMerr)

- H_0: $\lambda = 0$ (errores independientes)
- H_1: $\lambda \neq 0$ (errores correlacionados espacialmente)

Intuición: Si los residuos de OLS están correlacionados con los residuos de tracts vecinos, entonces:

- Hay factores omitidos con estructura espacial
- Los errores estándar de OLS son incorrectos
- Necesitamos un modelo SEM

Tests Robustos: ¿Y si ambos son significativos?

Problema: LMlag y LMerr pueden ser significativos simultáneamente porque están correlacionados entre sí

Solución: Tests robustos que condicionan en la presencia del otro

- **RLMlag**: Test de lag espacial, condicionando en posible error espacial
  - H_0: $\rho = 0$ dado que λ puede ser $\neq 0$
- **RLMerr**: Test de error espacial, condicionando en posible lag espacial
  - H_0: $\lambda = 0$ dado que ρ puede ser $\neq 0$

Regla: Usar tests robustos para decidir cuál componente espacial es más importante

Regla de Decisión Práctica

Escenario	**LMlag**	**LMerr**	**Modelo**
Solo lag	Sig	No sig	SAR
Solo error	No sig	Sig	SEM
Ambos	Sig	Sig	Ver tests robustos ↓

Si ambos son significativos:

1. Mirar tests **robustos** (RLMlag y RLMerr)
2. Si RLMlag más significativo → SAR
3. Si RLMerr más significativo → SEM
4. Si ambos robustos muy significativos → SARAR