Lecture 13: Spatial Models (Cont.)

Big Data and Machine Learning for Applied Economics
Econ 4676

Ignacio Sarmiento-Barbieri

Universidad de los Andes

September 21, 2021

Recap

- Closeness
- Weights matrix
- Examples of weight matrices weights matrices in `R`
- Example of spatial regression.

Agenda

1. Motivation
2. Spatial Lag Model
   - Maximum Likelihood Estimator
   - Two-Stage Least Squares estimators
3. Interpretation of Parameters
4. Further Readings

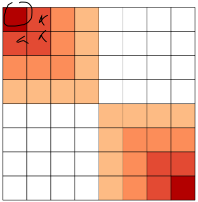
Motivation

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

$$y = X\beta + u$$

- Independence assumption between observation is no longer valid.
- Attributes of observation i may influence the attributes of observation j.
- Spatial dependence introduces a misspecification problem.

– OVB



Motivation

"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. y, X
  2. Particular location where those variables are observed and to the various links of proximity between all spatial observations. $\longrightarrow W$

Spatial Econometrics: Weights Matrix

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

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

$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \end{array}$$

$$w^* = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & \end{pmatrix}$$

- with generic element: $w_{ij} = \begin{cases} 1 & if\ j \in N(i) \\ 0 & o.w \end{cases}$
- $N(i)$ being the set of neighbors of location j. By convention, the diagonal elements are set to zero, i.e. $w_{ii} = 0$.
- Quite often the W matrices are standardized to sum to one in each row $w^*_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$

Spatial Lag Model

Spatial Autoregressive (SAR) Models

SAR $y = \lambda W y + u$

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

Let's consider the following model:

$$y = \lambda Wy + X\beta + u$$

we assume that W is exogenous

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

If W is row standardized:

- Guarantees $|\lambda| < 1$ (Anselin, 1982) * MLE
- $[0, 1]$ Weights
- Wy Average of neighboring values

$$Wy = \frac{\sum_{j \in N(i)} y_j}{\# N(i)}$$

- W is no longer symmetric $\sum_j w_{ij} \neq \sum_i w_{ji}$ (complicates computation)

Spatial Lag Model

Maximum Likelihood Estimator

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

$y - \lambda W y = X\beta + u$

Note that we can write

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

$y^* = X\beta + u$

- We can think this model as a way to correct for loss of information coming from spatial dependence.
- $(1 - \lambda W)y$ 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.

$$E((Wy)u') \neq 0 \tag{2}$$

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

Proof.
Note that $y = (I - \lambda W)^{-1}X\beta + (I - \lambda W)^{-1}u$
Then

$$\begin{aligned}
E((Wy)u') &= E(W(I-\lambda W)^{-1}X\beta u' + W(I-\lambda W)^{-1}uu') && (3)\\
&= W(I-\lambda W)^{-1}X\beta E(u') + W(I-\lambda W)^{-1}E(uu') && (4)\\
&= W(I-\lambda W)^{-1}E(uu') && (5)\\
&= \sigma^2 W(I-\lambda W)^{-1} \neq 0 && (6)
\end{aligned}$$

Spatial Lag Model

Maximum Likelihood Estimator

- One solution that emerged in the literature is MLE.
- We need an extra assumption, i.e., $u \sim_{iid} N(0, \sigma^2 I)$.

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

note that

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

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

$$= \underbrace{(I - \lambda W)^{-1} X\beta}_{\mu}$$

Spatial Lag Model

Maximum Likelihood Estimator

$Var(y) = E(yy') - E(y)^2$

$$E(yy') = ((I-\lambda W)^{-1}X\beta + (I-\lambda W)^{-1}u))((I-\lambda W)^{-1}X\beta + (I-\lambda W)^{-1}u)'$$
$$= (I-\lambda W)^{-1}X\beta\beta'X'(I-\lambda W')^{-1} + (I-\lambda W)^{-1}u\beta'X'(I-\lambda W')^{-1}$$
$$+ (I-\lambda W)^{-1}X\beta u'(I-\lambda W')^{-1} + (I-\lambda W)^{-1}uu'(I-\lambda W')^{-1}$$
$$= (I-\lambda W)^{-1}X\beta\beta'X'(I-\lambda W')^{-1} + (I-\lambda W)^{-1}uu'(I-\lambda W')^{-1}$$
$$= (I-\lambda W)^{-1}X\beta\beta'X'(I-\lambda W')^{-1} + (I-\lambda W)^{-1}(I-\lambda W')^{-1}\sigma^2$$

then

$y \sim N(\mu, \Omega\sigma^2)$

$$V(y) = E(yy') - (E(y))^2$$
$$= [(I-\lambda W)'(I-\lambda W)]^{-1}\sigma^2$$
$$= \Omega\sigma^2 \tag{7}$$

Spatial Lag Model

Maximum Likelihood Estimator

The associated likelihood function is then

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

the log likelihood

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

note that $|\sigma^2\Omega|=\sigma^{2n}|\Omega|$, and that

$$\begin{aligned}|\Omega| &= |[(I-\lambda W)'(I-\lambda W)]^{-1}| \\ &= |(I-\lambda W)^{-1}(I-\lambda W')^{-1}| \\ &= |(I-\lambda W)^{-1}||(I-\lambda W')^{-1}| \\ &= |(I-\lambda W)|^{-2}\end{aligned} \tag{8}$$

Spatial Lag Model

Maximum Likelihood Estimator

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

$$l\left(\sigma^2,\lambda,y\right) = constant - \frac{n}{2}ln\left(\sigma^2\right) + ln\left(|(I-\lambda W)|\right) - \frac{1}{2\sigma^2}\left(y-(I-\lambda W)^{-1}X\beta\right)'(I-\lambda W)'(I-\lambda W)\left(y-(I-\lambda W)^{-1}X\beta\right) \quad (9)$$

then

$$l\left(\sigma^2,\lambda,y\right) = constant - \frac{n}{2}ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\left((I-\lambda W)y - X\beta\right)'\left((I-\lambda W)y - X\beta\right) + ln\left(|(I-\lambda W)|\right) \quad (10)$$

Spatial Lag Model

Maximum Likelihood Estimator

- The determinant $|(I - \lambda W)|$ 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

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

→ Ord Decomp

↳ eigenvalues de la matriz ω

So the log likelihood is simplified to

$$\begin{aligned} l\left(\sigma^2, \lambda, y\right) = constant - \frac{n}{2} ln\left(\sigma^2\right) \\ - \frac{1}{2\sigma^2}((I - \lambda W)y - X\beta)'((I - \lambda W) - X\beta) \\ + \sum ln(1 - \lambda\omega_i) \end{aligned} \tag{11}$$

Spatial Lag Model
Maximum Likelihood Estimator

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

$$\hat{\beta}_{MLE} = (X'X)^{-1}X'(I - \lambda W)y$$

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

- Conditional on λ 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 λ

$Wy \to X$

$$l(\lambda) = -\frac{n}{2} ln \left(\frac{1}{n} (e_0 - \lambda e_L)'(e_0 - \lambda e_L) \right) + \sum ln(1 - \lambda \omega_i) \tag{12}$$

$y \to X$

ω Eigenvalues de W

- 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 λ.
- with λ^*, get $\hat{\beta}_{MLE}$ and $\hat{\sigma}^2_{MLE}$

Spatial Lag Model

Maximum Likelihood Estimator

The asymptotic variance follows as the inverse of the information matrix

$$AsyVar\left(\lambda, \beta, \sigma^2\right) = \begin{pmatrix} tr(W_A)^2 + tr(W_A' W_A) + \frac{(W_A X\beta)'(W_A X\beta)}{\sigma^2} & \frac{(X' W_A X\beta)'}{\sigma^2} & \frac{tr(W_A)'}{\sigma^2} \\ \frac{(X' W_A X\beta)'}{\sigma^2} & \frac{(X'X)}{\sigma^2} & 0 \\ \frac{tr(W_A)'}{\sigma^2} & 0 & \frac{n}{2\sigma^4} \end{pmatrix}^{-1} \tag{13}$$

$tr(W_A)$ $\quad$ $\hat{\sigma}^2$

- where $W_A = W(I - \lambda W)^{-1}$.
- Note that
  - the covariance between β and σ^2 is zero, as in the standard regression model,
  - this is not the case for λ and σ^2.

Spatial Lag Model

Two-Stage Least Squares estimators

$u \sim N(0, \sigma^2 I)$

- An alternative to MLE we can ~~us~~ use 2SLS to eliminate endogeneity.
- Key is to identify proper instruments
  - Need to be uncorrelated with the error term $E(Hu) = 0$
  - Correlated with Wy $\quad Cov(H, Wy) \neq 0$

Spatial Lag Model

Two-Stage Least Squares estimators

Consider the following

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

now, since $|\lambda| < 1$ we can use Neumann series property to expand the inverse matrix as

$$(I - \lambda W)^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + ...$$

$$(I - T)^{-1} = \sum_{k=0}^{\infty} T^k$$

hence

$$E(y) = (I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + ...) X\beta$$

$$= X\beta + \lambda WX\beta + \lambda^2 W^2 X\beta + \lambda^3 W^3 X\beta + ...$$

so we can express $E(y)$ as a function of X, WX, W^2X,...

Spatial Lag Model

Two-Stage Least Squares estimators

We can use the first three elements of the expansion as instruments. Let's define H as the matrix with our instruments

$$H = [X, WX, W^2X]$$

→ Instrumentos

Now,

$$y = \lambda Wy + X\beta + u$$

$$= M\theta + u$$

$M = (\lambda W, X)$

$\theta = \begin{pmatrix} \lambda \\ \beta \end{pmatrix}$

where $M = [Wy, X]$ and $\theta = [\lambda, \beta]$.

Spatial Lag Model

Two-Stage Least Squares estimators

$H \to M$ $\quad P_X = (X'X)^{-1}X'$

- The first stage is: $M = \overset{X}{H}\gamma + \eta$
  - where $\hat{\gamma} = (H'H)^{-1}H'M$
  - and $\hat{M} = H\hat{\gamma} = P_H M$ $\to \hat{M}$
- The second stage is

$\hat{M} \to y$

$$y = \hat{M}\theta + u \tag{14}$$

and

$$\begin{aligned} \boxed{\hat{\theta}_{2SLS}} &= (\hat{M}'\hat{M})^{-1}\hat{M}'y \\ &= (M'P_HM)^{-1}M'P_Hy \end{aligned} \tag{15}$$

$\ell(\lambda) \to \lambda^*$

$\hat{\theta}_{MLE}$, $\hat{\theta}_{2SLS}$ Comparar los resultados

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

Wy

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

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

$$y_i = \alpha + \beta x_i + \lambda \sum w_{ij} y_j + \epsilon_i \tag{16}$$

with $|\lambda| < 1$, and

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

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

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

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_1} \\ \frac{\partial y_2}{\partial x_2} \\ \vdots \\ \frac{\partial y_n}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 1-\lambda w_{11} & & 0 \\ & 1-\lambda w_{22} & \\ 0 & & 1-\lambda w_{nn} \end{pmatrix} \beta$$

- 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

$$E(y) = (I - \lambda W)^{-1} X\beta \tag{17}$$

we define

$$S(W) = \underbrace{(I - \lambda W)^{-1}}_{\text{Efecto [illegible]}} \beta \tag{18}$$

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:

$$S(W) = \frac{\partial E(y)}{\partial x} = \begin{pmatrix} \frac{\partial E(y_1)}{\partial x_1} & \cdots & \frac{\partial E(y_1)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_i)}{\partial x_1} & \cdots & \frac{\partial E(y_i)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_1} & \cdots & \frac{\partial E(y_n)}{\partial x_n} \end{pmatrix} \tag{19}$$

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

$$
\begin{aligned}
ADI &= \frac{tr(S(W))}{n} \\
&= \frac{1}{n}\sum_{i=1}^{n} S(W)_{ii}
\end{aligned} \tag{20}
$$

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

$$
\begin{aligned}
ATIT_j &= \frac{\iota' S(W)}{n} \\
&= \frac{1}{n}\sum_{i=1}^{n} S(W)_{ij}
\end{aligned}
\tag{21}
$$

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

$$
\begin{aligned}
ATIF_i &= \frac{1}{n} S(W)\iota \\
&= \frac{\sum_{j=1}^{n} S(W)_{ij}}{n}
\end{aligned}
\tag{22}
$$

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

$$ATI = \frac{1}{n}\iota' S(W)\iota = \frac{1}{n}\sum_{i=1}^{n} ATIT_i = \frac{1}{n}\sum_{j=1}^{n} ATIF_i \tag{23}$$

- 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

$$AII = ATI - ADI \tag{24}$$

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

Interpretation of Parameters: Example

- We have data on 20 Italian regions on GDP and unemployment.
- We want to estimate the effect of GDP on Unemployment (Okun's Law)

	OLS	Spatial Lag Model
Intercept	10.971***	3.12275***
GDP	-3.326***	-1.13532***
λ	-	0.7476***
ADI	-	-1.542448
AII	-	-2.95571
ATI	-	-4.498159

$$\frac{\partial U}{\partial GDP} = -3$$

$S(w) =$

Review & Next Steps

- Today:
  - Details on Spatial Lag Model
  - Interpretation
- Next class: Prediction, prediction, prediction, ...

Further Readings

- Arbia, G. (2014). A primer for spatial econometrics with applications in R. Palgrave Macmillan. (Chapter 2 and 3)
- Anselin, Luc, & Anil K Bera. 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." Statistics Textbooks and Monographs 155. MARCEL DEKKER AG: 237–90.
- Anselin, L. (1982). A note on small sample properties of estimators in a first-order spatial autoregressive model. Environment and Planning A, 14(8), 1023-1030.
- Tobler, WR. 1979. "Cellular Geography." In Philosophy in Geography, 379–86. Springer.