Lecture 13: Spatial Models (Cont.)

Big Data and Machine Learning for Applied Economics
Econ 4676

Ignacio Sarmiento-Barbieri

Universidad de los Andes

September 22, 2020

Announcements

- Final Project
  - First deadline. Sept 25. Brief zoom hang out, short presentation (5 slides tops). Present idea and basic plan. soft deadline
  - Second deadline. October 25. Show data. hard deadline
  - Final work, December 17. Bonus for "complete papers" hard deadline
- Prof. Tomás Rodríguez Barraquer will jump in and I'll leave

Recap

- Closeness
- Weights matrix
- Examples of weight matrices weights matrix in `R`
- Traditional spatial regressions
- Prediction with spatial model

Agenda

Motivation

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

- Independence assumption between observation is no longer valid
- Attributes of observation i may influence the attributes of observation j.
- Today we will dive into the estimation of spatial lag models.
- Think as a way to model $f(X)$
- Spatial dependence introduces a miss speficication problem

typo

Spatial Lag Model

SAR $\rightarrow y = \lambda W y + \varepsilon$

SLM $\rightarrow y = \lambda W y + X\beta + \varepsilon$

Let's consider the following model:

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

with $\underbrace{|\lambda| < 1}$, we also assume that W is exogenous $\leftarrow$ Anselin

If W is row standardized:

$$W = \begin{pmatrix} 0 & 1 & 0 & \\ 1 & 0 & & 1 \\ 0 & 0 & 0 & \\ & 1 & & 0 \end{pmatrix}$$

- Guarantees $|\lambda| < 1$ (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)

$w^*_{ij} = \frac{w_{ij}}{\sum_{j=1} w_{ij}}$

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

Spatial Lag Model

Maximum Likelihood Estimator

$$y = \underline{\lambda W y} + X\beta + \varepsilon$$

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

Note that we can write

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

$$y = X\beta + \varepsilon$$

$$\underbrace{y^{*}} = X\beta + u$$

Lo Filtrado

- 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

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

In this case, endogeneity emerges because the spatially lagged value of y is correlated with the stochastic disturbance.

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

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

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

$E(u) = 0$

$$= W(I-\lambda W)^{-1}E(uu')$$

$$= \sigma^2 W(I-\lambda W)^{-1} \neq 0$$

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

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

$$= (I - \lambda W)^{-1} X\beta + (I - \lambda W)^{-1} \underbrace{E(u)}_{=0}$$

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

Spatial Lag Model

Maximum Likelihood Estimator

HW

$$
\begin{aligned}
E(yy') &= ((I-\lambda W)^{-1}X\beta + (I-\lambda W)^{-1}u)((I-\lambda W)^{-1}X\beta + (I-\lambda W)^{-1}u)' \\
&= (I-\lambda W)^{-1}X\beta\beta'X'(I-\lambda W')^{-1} + (I-\lambda W)^{-1}u\beta'X'(I-\lambda W')^{-1} \\
&+ (I-\lambda W)^{-1}X\beta u'(I-\lambda W')^{-1} + (I-\lambda W)^{-1}uu'(I-\lambda W')^{-1} \\
&= (I-\lambda W)^{-1}X\beta\beta'X'(I-\lambda W')^{-1} + (I-\lambda W)^{-1}uu'(I-\lambda W')^{-1} \\
&= \underbrace{(I-\lambda W)^{-1}X\beta\beta'X'(I-\lambda W')^{-1}}_{E(y)^2} + \underbrace{(I-\lambda W)^{-1}(I-\lambda W')^{-1}\sigma^2}
\end{aligned}
$$

then

$$
\begin{aligned}
V(y) &= E(yy') - E(y)^2 \\
&= [(I-\lambda W)'(I-\lambda W)]^{-1}\sigma^2 \\
&= \Omega\sigma^2 \qquad (1)
\end{aligned}
$$

Spatial Lag Model

Maximum Likelihood Estimator

The associated likelihood function is then

y - E[y]

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

determinante

the log likelihood

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

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

prop de determinantes (inversa)

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

prop det

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

$$l\left(\sigma^2,\lambda,y\right) = constant - \frac{n}{2}ln\left(\sigma^2\right) + ln\left(\left|(I-\lambda W)\right|\right) - \frac{1}{2\sigma^2}(y-(I-\lambda W)^{-1}X\beta)'(I-\lambda W)'(I-\lambda W)(y-(I-\lambda W)^{-1} \tag{3}$$

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

then

MLE

$$l\left(\sigma^2,\lambda,y\right) = constant - \frac{n}{2}ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\left((I-\lambda W)y - X\beta\right)'\left((I-\lambda W) - X\beta\right) + ln\left(\left|(I-\lambda W)\right|\right) \tag{4}$$

u' u

MCO → Min $u'u$

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

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

a todos

So the log likelihood is simplified to

$$\begin{aligned} l\left(\sigma^2, \lambda, y\right) = {} & constant - \frac{n}{2} ln\left(\sigma^2\right) \\ & - \frac{1}{2\sigma^2}((I - \lambda W)y - X\beta)'((I - \lambda W) - X\beta) \\ & + \sum ln(1 - \lambda\omega_i) \end{aligned} \tag{5}$$

Spatial Lag Model
Maximum Likelihood Estimator

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

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

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

- 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

$$l(\lambda) = -\frac{n}{2} ln \left(\frac{1}{n} (e_0 - \lambda e_L)'(e_0 - \lambda e_L) \right) + \sum ln(1 - \lambda \omega_i) \tag{6}$$

- 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.

Spatial Lag Model

Maximum Likelihood Estimator

- Cons
- Norm
- EF AS

The asymptotic variance follows as the inverse of the information matrix

$$AsyVar\left(\lambda,\beta,\sigma^2\right)=\begin{pmatrix} tr(W_A)^2+tr(W_A'W_A)+\frac{(W_AX\beta)'(W_AX\beta)}{\sigma^2} & \frac{(X'W_AX\beta)'}{\sigma^2} & \frac{tr(W_A)'}{\sigma^2} \\ \frac{(X'W_AX\beta)'}{\sigma^2} & \frac{(X'X)}{\sigma^2} & 0 \\ \frac{tr(W_A)'}{\sigma^2} & 0 & \frac{n}{2\sigma^4} \end{pmatrix}^{-1} \tag{7}$$

- 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 2SLS to eliminate endogeneity.
- Key is to identify proper instruments
  - Need to be uncorrelated with the error term
  - Correlated with WY

$Wy \rightarrow$ endogeno

$E(Wy\,u) \neq 0$

$= \sigma^2(\quad)$

Spatial Lag Model

Two-Stage Least Squares estimators

Consider the following

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

M ω

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

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

hence

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

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

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

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

Now,

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

$$y = M\theta + u$$

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

Spatial Lag Model

Two-Stage Least Squares estimators

The first stage is

$[Wy, X]$ → instrumentos X, WX, W^2X

$$\underline{M} = H\gamma + \eta$$

and

$$\hat{\gamma} = \underbrace{(H'H)^{-1}H'M}$$ OLS

$$\hat{M} = H\hat{\gamma} = P_H M$$

→ $\underbrace{H(H'H)^{-1}H'}_{P_H}M$

and the second stage is

$$y = \hat{M}\theta + u \tag{8}$$

and

$$\begin{aligned} \hat{\theta}_{2SLS} &= (\hat{M}'\hat{M})^{-1}\hat{M}'y \\ &= (M'P_H M)^{-1}M'P_H y \end{aligned} \tag{9}$$

Interpretation of Parameters

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \underset{2}{\beta_1} x_{i2} + \cdots + \underset{r}{\beta_1} x_{ir} + \cdots + \underset{k}{\beta_1} x_{ik} \quad i = \underline{1, \ldots, n}$$

- Recall that in OLS we have

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

or generically

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

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

- 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

$$S_r(W) = (I - \lambda W)^{-1} \beta_r \tag{12}$$

Interpretation of Parameters

$$S_r(W) = (I - \lambda W)^{-1}\beta_r \tag{13}$$

we can write

$$\begin{pmatrix} E(y_1) \\ \vdots \\ E(y_i) \\ \vdots \\ E(y_n) \end{pmatrix} = \sum_{r=1}^{k} \begin{pmatrix} S_r(W)_{11} & S_r(W)_{12} & \dots & \dots & S_r(W)_{1n} \\ S_r(W)_{i1} & \dots & S_r(W)_{ii} & \dots & S_r(W)_{in} \\ S_r(W)_{n1} & \dots & \dots & \dots & S_r(W)_{nn} \end{pmatrix} \begin{pmatrix} x_{1r} \\ \vdots \\ \vdots \\ \vdots \\ x_{nr} \end{pmatrix}$$

Interpretation of Parameters

Then for the $i-th$ observation

$$E(y_i) = \sum_{r=1}^{k} (S_r(W)_{i1}x_{1r} + S_r(W)_{i2}x_{2r} + \cdots + S_r(W)_{ii}x_{ir} + \cdots + S_r(W)_{in}x_{nr})$$

then

$$\frac{\partial E(y_i)}{\partial x_{jr}} = S_r(W)_{ij} \tag{14}$$

and

$$\frac{\partial E(y_i)}{\partial x_{ir}} = S_r(W)_{ii} \tag{15}$$

where $S_r(W)_{ij}$ denotes the $(i,j)-th$ element of the matrix $S_r(W)$

Interpretation of Parameters

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

$$S_r(W) = \frac{\partial E(y)}{\partial x_r} = \begin{pmatrix} \frac{\partial E(y_1)}{\partial x_{1r}} & \dots & \frac{\partial E(y_1)}{\partial x_{nr}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_i)}{\partial x_{1r}} & \dots & \frac{\partial E(y_i)}{\partial x_{nr}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_{1r}} & \dots & \frac{\partial E(y_n)}{\partial x_{nr}} \end{pmatrix} \tag{16}$$

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_r, which we denote x_{ir}, on y_i. This is the average of all diagonal entries in S

$$
\begin{aligned}
ADI &= \frac{tr(S_r(W))}{n} \\
&= \frac{1}{n}\sum_{i=1}^{n} S_r(W)_{ii}
\end{aligned}
\tag{17}
$$

Interpretation of Parameters

- *Average Total Impact To* an observation: this measure is related to the impact produced on one single observation y_i from changing of the $r-th$ independent variable across all other observations. For each observation this is calculated as the sum of the $i-th$ row of matrix S

$$
\begin{aligned}
ATIT_j &= \frac{\iota' S_r(W)}{n} \\
&= \frac{1}{n}\sum_{i=1}^{n} S_r(W)_{ij}
\end{aligned}
\tag{18}
$$

Interpretation of Parameters

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

$$\begin{aligned} ATIF_i &= \frac{1}{n} S_r(W)\iota \\ &= \frac{\sum_{j=1}^{n} S_r(W)_{ij}}{n} \end{aligned} \tag{19}$$

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

$$ATI = \frac{1}{n}\iota' S_r(W)\iota = \frac{1}{n}\sum_{i=1}^{n} ATIT_i = \frac{1}{n}\sum_{j=1}^{n} ATIF_i \tag{20}$$

- 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

$$AII = ATI - ADI \tag{21}$$

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

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

Review & Next Steps

- Today:
  - Details on Spatial Lag Model
  - Interpretation
- Next class: Model assessment and model selection
- Questions? Questions about software?

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.