# Spatial Autoregressive Model

Patrick Rickert

11/16/2017

# Why Spatial Autoregressive Models

- One of the assumptions underlying traditional regression models is conditional independence of observations
- Once everything is controlled for, observation $i$ is unaffected by the value of observation $j$
- This is often not a great assumption
- For example, with dyadic analysis of conflict involving multiple actors

# Why Pt 2

- Spatial dependence can be either a parameter of interest to be estimated
- Or it can be a nuisance to be taken care of
- Either way, we want to model it, because not accounting for spatial dependence can lead to biased and inefficient estimates

# Declaration of Independence

- Independence implies $E(\epsilon_i \epsilon_j) = E(\epsilon_i)E(\epsilon_j) = 0$
- For a typical regression framework, this leads to a very familiar:
- $y_i = X_i\beta + \epsilon_i$
- $\epsilon_i \sim N(0, \sigma^2)$

# Declaration of Spatial Dependence

- Suppose however that observations $i$ and $j$ are neighbors that influence one another. The DGP might then be something like
- $y_i = \alpha y_j + X_i\beta + \epsilon_i$
- $y_j = \alpha y_i + X_i\beta + \epsilon_i$
- $\epsilon_i \sim N(0, \sigma^2)$
- $\epsilon_j \sim N(0, \sigma^2)$
- The errors would be correlated, violating independence
- This would require a model with at least $n^2 - n$ parameters to allow for spatial dependence across observations

# Assumptions of the Model

- Continuous dependent variable
- Spatial dependence
- However, you will see that if spatial dependence is not present, the model reduces to the typical OLS model
- Familiar OLS assumptions, aside from independence

# The Model

- The Spatial Autoregressive Model can be expressed by

$$y = \rho W y + X\beta + \epsilon$$
$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where

- $\rho$ describes the degree of spatial dependence
  - Making an average level of spatial dependence hold throughout the entire model is parsimony that keeps us from needing a separate estimate ($\alpha$ in the earlier slide) for all observations
  - Note that $\rho$ is not the correlation coefficient between the spatial lag matrix and the $y$ vector
- $W$ is an $n$ x $n$ spatial weights matrix
  - This makes $Wy$ the spatial lag matrix
- $N(0, \sigma^2 I_n)$ is a zero mean disturbance process with constant variance and no covariance between observations

# The DGP

Exploring the model further gives us a better understanding of our idea about the data generating process

$$
\begin{aligned}
y &= \rho W y + X\beta + \epsilon \\
y - \rho W y &= X\beta + \epsilon \\
(I_n - \rho W) y &= X\beta + \epsilon \\
y &= (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\epsilon
\end{aligned}
$$

Where $I_n$ is an $n$x$n$ identity matrix.
The final expression is the assumed DGP. Thus, each outcome $y_i$ can be expressed as a linear combination of observed characteristics of themselves and of their neighbors.

# Getting to Know the Neighbors


Figure 1: Mexican Borders.

# Adjacency Matrix

-The previous map can be translated into this adjacency matrix

$$\text{Adjacency Matrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

-Note that the diagonals are zero. A place cannot border itself, and further this matrix is multiplied by the matrix of outcome $y$. If the diagonals were anything different than zero, $y$ would be predicting itself.

# Spatial Weights Matrix

- The adjacency matrix can be standardized into this spatial weights matrix, which shows the proportion of neighbor that every observation is.

$$\text{Adjacency Matrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

- Note that these need not be equal. If instead of a dichotomous shared border, the variable was length of border, the US-Mexico border would take up a much higher proportion for Mexico
- The spatial weights matrix becomes the spatial lag matrix when it is multiplied by the vector of outcomes $y$

# The Spatial Dependence Parameter

- $\rho$ is one of three parameters estimated in the SAR model that defines the degree of spatial dependence
- It can be interpreted as the predicted change in the outcome $y$ for a given unit $i$ as the outcomes for all of $i$'s neighbors are exogenously increased by one unit
- It does not imply directionality, only the degree of similarity conditional on covariates

# Estimating The Parameters

The steps for maximum likelihood estimation based on a concentrated likelihood function are given by Anselin (1988)

1. Perform OLS for the model $y = X\beta_0 + \epsilon_0$
2. Perform OLS for the model $Wy = X\beta_L + \epsilon_L$
3. Compute residuals $e_0 = y - X\hat{\beta}_0$ and $e_L = Wy - X\hat{\beta}_L$
4. Given $e_0$ and $e_L$, find the $\rho$ that maximizes the concentrated likelihood function
   $L_c = -(n/2)\ln(\pi) - (n/2)\ln(1/n)(e_0 - \rho e_L) + \ln|I - \rho W|$
5. Given $\rho$ that maximizes $L_C$, compute $\hat{\beta} = (\hat{\beta}_0 - \rho\hat{\beta}_L)$ and $\hat{\sigma_e^2} = (1/n)(e_0 - \rho e_L)'(e_0 - \rho e_L)$

This can be difficult. We're going to let the computer take care of it.

# Interpreting the Coefficients

- You cannot directly interpret the coefficients from an SAR model
- This is because you need to incorporate how changing the value of the neighbors will change observation $i$
- For this we turn to the `impacts` function in R which delivers the the average direct effect, the average indirect effect, and the average total effect

# Direct and Indirect Effects

- The effect size can be broken down into direct, indirect, and total effects, since the effect size isn't determined just by increasing an observation one unit, but also depends on the value of its neighbors
- The average direct effect is the average of the diagonal elements of $(I - \rho W)^{-1}\beta_k$
- The index $k$ refers to one covariate
- Average direct effect: The average change in $y_i$ when some covariate $k$ is increased by one unit just in the observation $i$ for all $i$

# Direct and Indirect Effects

- The average indirect effect is given by the average of the off-diagonal elements of $(I - \rho W)^{-1}\beta_k$
- Average indirect effect: The average change in $y_i$ when some covariate $k$ is increased by one unit in all observations
- The average total effect, then is the average of all elements in the matrix $(I - \rho W)^{-1}\beta_k$
- Average total effect: the average change in $y_i$ when some covariate $k$ is increased by one unit for all observations in the data
- By definition, average total effect $=$ average direct effect $+$ average indirect effect

# An Example

- We are going to run an example that looks at census tract level analysis of violence crime in Chicago (This example is adapted from one prepared by Ignacio Sarmiento-Barbieri (University of Illinois))
- Perhaps we are interested in the role of foreclosures and unemployment on the amount of violent crime in Chicago during the late 2000s financial crisis
- We want to fit a model with census-tract level amount of violent (from the Chicago Data Portal) on the left hand side and foreclosure count (from HUD) and unemployment rate (from BLS) on the right hand side.

# The Example

- The variables are coded where `violent` is the number of violent crimes from January 2007 to December 2008, `estfcsrt` is the estimated forclosure count from 2007 to 2008, and `blsunemp` is the unemployment rate in June 2008.
- We may expect this model to have significant spatial dependence, as violent crime might not stop at the census tract border
- Indeed, we may be interested in whether violent crime exhibits spatial dependence in its own right.

# How To Do It In R

Often, we will have an adjacency matrix that we can turn into weights list object (a `listw` object) that `R` can use as a spatial weight matrix. This can be done with the `mat2listw(neighbours)` command, where neighbors is the adjacency matrix. Another way to create a `listw` object is to use polygon list data (as maps will often be) into a neighbors list using poly2nb, and then to a weights list object using nb2listw.

```
# mat2listw(neighbours)
list.queen<-poly2nb(chi.poly,
                    queen=TRUE)
W<-nb2listw(list.queen, style="W",
            zero.policy=TRUE)
```

# Estimating the Model

- To actually run the model, you need a formula, dataset, and weights list object

```
Formula<-violent~est_fcs_rt+bls_unemp
SARModel<-lagsarlm(Formula, data= chi.poly@data, W)
```

# The Model Output

```
summary(SARModel)
```

```
Call:lagsarlm(formula = Formula, data = chi.poly@data, listw = W)

Residuals:
     Min       1Q   Median       3Q      Max
-519.127  -65.003  -15.226   36.423 1184.193

Type: lag
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -93.7885    41.3162  -2.270  0.02321
est_fcs_rt   15.6822     1.5600  10.053  < 2e-16
bls_unemp     8.8949     5.2447   1.696  0.08989

Rho: 0.49037, LR test value: 141.33, p-value: < 2.22e-16
Asymptotic standard error: 0.039524
    z-value: 12.407, p-value: < 2.22e-16
Wald statistic: 153.93, p-value: < 2.22e-16
```

Figure 2: Partial SAR Output

# Comparison to OLS

Table 1: OLS Results

| | *Dependent variable:* |
|---|---|
| | violent |
| est_fcs_rt | 28.298*** |
| | (1.435) |
| bls_unemp | −0.308 |
| | (5.770) |
| Constant | −18.627 |
| | (45.366) |
| Observations | 897 |
| $R^2$ | 0.314 |
| Adjusted $R^2$ | 0.313 |
| Residual Std. Error | 157.346 (df = 894) |
| F Statistic | 204.699*** (df = 2; 894) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

# But don't forget the direct and indirect effects!

```r
impacts(SARModel, listw = W)
```

```
Impact measures (lag, exact):
               Direct  Indirect    Total
est_fcs_rt 16.434479 14.336896 30.77137
bls_unemp   9.321585  8.131842 17.45343
```

Figure 3: Impact Output

# Diagnostics (or, is this really necessary?)

- How do we know if the model we have estimated is the correct one?
- We can conduct a likelihood ratio tests to test improvement in model fit over the regular linear regression model
- This is the main way to test the value of an SAR model, but there are others
- Additionally, you can check if you modeled the spatial dependence correctly, because it conducts a Lagrange Multiplier test for the residual autocorrelation

# Review of Likelihood Ratio Tests

- Likelihood ratio tests can compare the goodness of fit of nested models. Thus in this context, we are implicitly testing the alternate hypothesis that $\rho \neq 0$ versus the null hypothesis $\rho = 0$ by testing whether a model that includes $\rho$ fits better than a model without.
- The likelihood ratio statistic to test $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^C$ is calculated:

$$\lambda(x) = \frac{\sup_{\theta_0} L(\theta)}{\sup_{\theta} L(\theta)}$$

# Review of Likelihood Ratio Tests

- Wilk's theorem found that $-2\ln(\lambda)$ asymptotically approaches the $\chi^2$ distribution as $n \to \infty$ with degrees of freedom equal to the difference in dimensionality between $\Theta$ and $\Theta_0$ when $H_0$ is true
- This gives us (with some algebra) the familar form $\{\lambda(x) \leq C^*\}$ where we can compare our statistic to the $100(1-\alpha)$ percentile point of the $\chi^2$ distribution
- In reality, this is done in R

# Diagnostics in the Summary Output

```
Rho: 0.49037, LR test value: 141.33, p-value: < 2.22e-16
Asymptotic standard error: 0.039524
    z-value: 12.407, p-value: < 2.22e-16
Wald statistic: 153.93, p-value: < 2.22e-16

Log likelihood: -5738.047 for lag model
ML residual variance (sigma squared): 20200, (sigma: 142.13)
Number of observations: 897
Number of parameters estimated: 5
AIC: 11486, (AIC for lm: 11625)
LM test for residual autocorrelation
test value: 8.1464, p-value: 0.0043146
```

Figure 4: The bottom half of the summary output

# Other Diagnostics (or, is this really necessary?)

- One other ways to test for spatial dependance is using Moran's I
- A two dimensional extension of the Durbin-Watson test

$$I = \frac{e' We}{e' e}$$

  Where $e$ is a vector of OLS residuals
- Compared to the null hypothesis expectation of $\frac{-1}{N-1}$

# Moran's I Test

```r
moran.lm<-lm.morantest(linearModel,
            W, alternative="two.sided")
print(moran.lm)
```

- Model is the OLS model
- W is the listW list
- Alternative specifies $H_a$: greater, less, or two.sided

# Moran I Output

```
        Global Moran I for regression residuals

data:
model: lm(formula = violent ~ est_fcs_rt + bls_unemp, data = chi.poly@data)
weights: W

Moran I statistic standard deviate = 11.785, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:
Observed Moran I      Expectation         Variance
    0.2142252370    -0.0020099108     0.0003366648
```

Figure 5: Moran I

# Summary

- Spatial dependence is a potentially significant problem for a lot of social scientific questions, and yet it is modeled relatively infrequently
- Keep in mind that space can be more than just geography, it can be any shared characteristic that causes dependence.
- One way to model spatial dependence is with the Spatial Autoregressive Model, though there are several models in this class that aim to solve different problems.