Spatial Econometrics

Ani Katchova

Spatial Econometrics Overview

- Examples of spatial econometrics models and data
- Spatial weight matrix based on contiguity and distance
- Spatial regression models: spatial lag model and spatial error model
- Spatial dependence test (Moran’s I)

Spatial Econometrics

Spatial econometrics examples

- Real estate economics: House prices depend on the number of bedrooms, bathrooms, etc. House prices also depend on location; prices of houses in the same neighborhood are similar.
- Farmland values: Farmland prices depend on land rent, government payments, etc. but farmland prices are similar if counties are spatially close.
- Precision agriculture: Different rates of nitrogen are applied in a corn field. Corn yields will be different because of the different nitrogen applications but they will be similar if the fields/plots are spatially close.

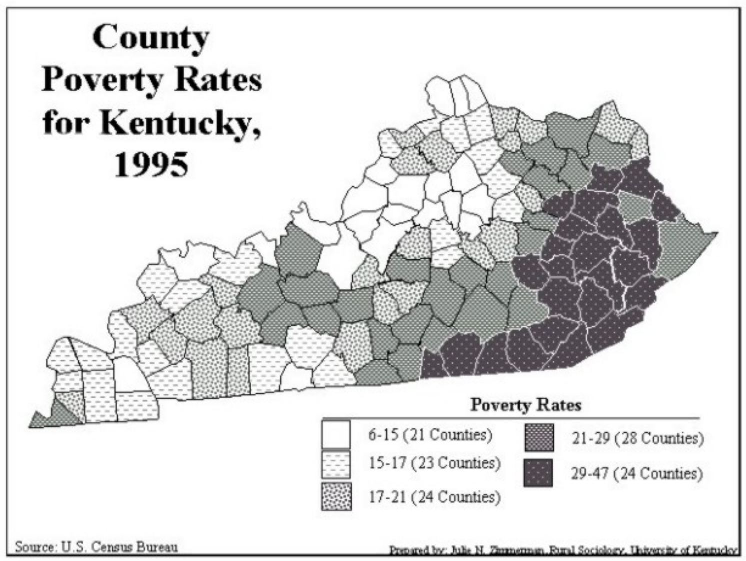
Spatial econometrics

- Spatial econometrics accounts for the presence of spatial effects in regression analysis.
- Spatial econometrics is used in regional science, urban and real estate economics and economic geography.

Spatial econometrics problems

- Spatial econometrics is a special type of econometrics.
- Spatial data needs to be geo-coded for location (coordinates, borders, distance).
- The spatial matrix defines neighbors (observations that are spatially close) and their effects.
- We need to account for the spatial dependence in the regression model.

Spatial data with spatial dependence (spatial autocorrelation)



Spatial weight matrix

- The spatial weight matrix provides the structure of the spatial relationship among observations.
- The spatial weight matrix provides information about which observations are considered neighbors and also how their values are related to each other.
- The spatial weight matrix is defined as W with elements w_{ij} indicating whether observations i and j are spatially close.
- Spatial weight matrices need to be "row-standardized" which means the weights need to sum up to one on each row.
- There are two types of spatial weight matrices based on contiguity and on distance.

Spatial weight matrix based on contiguity

- We need to know if observations are contiguous (share a border and/or a vertex).
- Use GIS ArcView to specify the spatial weights (most software can't recognize contiguity).
- The spatial weight matrix W has elements defined as:

$$w_{ij} = \begin{cases} 1 \; if \; i \; is \; contiguous \; to \; j \\ 0 \; otherwise \end{cases}$$

- If units i and j are neighbors, the spatial weight is 1, otherwise 0.

- Example: spatial weight matrix based on *contiguity* where units 2 and 3 are neighbors and units 3 and 4 are neighbors (but units 2 and 4 are not neighbors).

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

- We can row-standardize the spatial weight matrix by dividing each entry by the total for that row.

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & .5 & 0 & .5 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- Now each row represents the unit's value as a weighted average of the values of its neighbors: y_3=0.5*y_2+0.5*y_4. In general, $y_i = \sum_j w_{ij} y_j$.
- The row standardization is needed because in a weighted average formula, the weights need to sum up to 1.
- We can predict the values for each unit based on the values of its neighbors.

Spatial weight matrix based on distance

Coordinates

- We need to know the location of the observations (X coordinates and Y coordinates (longitude and latitude)) to calculate the distance between observations.
- With county data, we use distance between the centroids (center points) of counties.
- Software can typically calculate the spatial weight matrix if the x and y coordinates are provided for each observation.

Distance

- Let d_{ij} be the distance between observations i and j.
- Assume that there are no spatial effects beyond a certain distance band D.

Spatial weight matrix

- A spatial weight matrix can be constructed based on distance where units within a specified radius have a spatial weight of 1 (they are neighbors), otherwise 0.

$$w_{ij} = \begin{cases} 1 \; if \; the \; distance \; between \; i \; and \; j < D \\ 0 \; otherwise \end{cases}$$

  - The spatial weight matrix will look similar to the one in the previous section – only it will be based on distance, not contiguity.

- Alternative specification: a spatial weight matrix can also be constructed based on distance where units with distance d_{ij} receive a weight that is inversely proportional to the distance between the units and 0 if they are beyond a certain distance band D.

$$w_{ij} = \begin{cases} 1/d_{ij} \; if \; the \; distance \; between \; i \; and \; j < D \\ 0 \; otherwise \end{cases}$$

- Example: spatial weight matrix based on *distance* where units 1 and 2 are neighbors with distance=5 and units 2 and 3 are neighbors with distance=2.

$$W = \begin{bmatrix} 0 & .2 & 0 & 0 \\ .2 & 0 & .5 & 0 \\ 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- The spatial weight matrix is row-standardized by dividing each entry by the total for that row. Now the value for each unit is a weighted average of its neighbors' values.

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 \\ .3 & 0 & .7 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- Create the spatial map for the unit locations.

- The choice of a spatial weight matrix based on contiguity or distance is up to the researcher.
  - Both methods (contiguity and distance) are used in the literature.
  - Example: for farmland value, neighbors that are based on distance may be more appropriate, but for residential housing values, neighbors based on contiguity (or neighborhood block) may be more appropriate.

Spatial weight matrix properties

- The diagonal elements of the spatial matrix are set equal to zero and the non-diagonal elements are non-zero for observations that are spatially close to one another and zero for those that are far away.
- As the distance band goes to zero, the spatial regression results approximate those of OLS (no spatial effects).
- Spatial weight matrices are row standardized, meaning that the row elements sum up to 1. Each unit's value is a weighted average of its neighbors.
- The dimensions of a spatial matrix NxN are based on the sample size N. This is a major issue with large data sets.
- The researcher picks the distance band. If you set your distance band to the maximum distance, then all units will have at least one neighbor.

Spatial regression

- Spatial regression is a regression that accounts for the spatial dependence of the data.
- Spatial dependence is added to a regression in two ways: spatial lag and spatial error.

Spatial lag regression

- The spatial lag model is appropriate when the focus is on the spatial interactions of the dependent variable. In this case we know the structure of the spatial relationship.
  - Example: the price of a house will depend on the prices of neighboring houses.
- Here the dependent variable y has the spatial structure.
- The spatial lag model is a spatial autoregressive model that includes a spatially lagged dependent variable.
- The spatial lag of the variable y is Wy. The dependent variable is a weighted average of its neighbors' values.

- The spatial lag regression is defined as:

$$y = \rho W y + x\beta + e$$

- The spatial lag model reduced form equation is:

$$(I - \rho W) y = x\beta + e$$

- The independent variables are explaining the variation in the dependent variable that is not explained by the neighbors' values.
- The spatial dependence parameter ρ is also estimated.

Spatial error regression

- The spatial error model is appropriate when we are interested in correcting for spatial autocorrelation due to the use of spatial data (irrespective of whether the model of interest is spatial or not). In this case we do not know the structure of the spatial relationship.
- We include spatially correlated errors due to unobservable features or omitted variables associated with location.
    - Example: farmers' technology adoption decisions may be influenced by their neighbors.
- Here the error term e has the spatial structure.
- The regression model is:

$$y = x\beta + e$$

- The errors are spatially correlated:

$$e = \lambda We + u$$

or

$$(I - \lambda W)e = u$$

- The spatial error regression reduced form equation is:

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

- The multipliers in front of the dependent and independent variables are the variation that cannot be explained by the neighbors' values.
- The spatial dependence parameter λ is also estimated.

Spatial dependence test (Moran's *I*)

- Moran's *I* test statistic is used to test if the data have spatial dependence.

$$I = (N/S_0)(e'We/e'e)$$

  - Here S_0 is a standardization factor that corresponds to the sum of weights for the non-zero cross-products: $S_0 = \sum_i \sum_j w_{ij}$
  - For row-standardized weights $S_0 = N$, so $I = e'We/e'e$.