Spatial Analysis

M. Fuat Kına

Geospatial data in practical use

- Geospatial data are data for which a specific location is associated with each record.
- Increasing availability of new sources of spatial data.
- In Geographic Information Sciences, there are two data models for how we record the world:
  - Raster data
  - Vector data
- We will focus on vector data
  - Points
  - Lines (connections between points)
  - Polygons (points, connections, and the area covered by the lines)

GeoPandas

- https://geopandas.org/en/stable/
- File formats for geospatial data, such as GeoJSON files, GeoPackage files, or shapefiles, which are specialized in storing spatial data, in addition to traditional tabular data.
- Geospatial data is a usual pandas data, supercharged with geospatial capabilities.
- Geospatial data has an attribute named “geometry”, which includes series like “area”.

Let’s plot some maps!

Spatial Econometrics

- Spatial dependence
- Spatially autoregressive data
  - Spatial weight matrices
- Spillover effects

Spatial dependence

- Spatially-referenced data
  - Observations besides the information of geocodes (GIS)
  - The coordinates of an interior point representing the center: the centroid
- Spatial dependence: Is there any unobserved geographical pattern in the data
  - The variation in the outcome variable that is unexplainable without considering geographical patterns (e.g., neighboring)
- It is quite similar to temporal dependence
- Two potential reasons for the spatial dependence
  - Omitted explanatory variables: X values in neighbors
  - Congestion effects: Y value might be written as a function of Y values in neighbors
- Authors suggest to control for the “spatially lagged dependent variable”

Spatial autoregressive process

- Spatial dependence information is not parsimonious by default. If we add new variables, total amount of distance variables will be N*(N-1), which is unrealistic for a medium-size data.
- The alternative is to use weight matrices that only stores neighboring relations.
- Its size will be N X N, including too many zero points, and represents a generalized spatial lag.
  - First-order neighbors
  - Second-order neighbors
  - …
- Higher-order neighbors resembles social network analysis

$$Wy = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}$$

$$= \begin{pmatrix} y_2 \\ (y_1 + y_3)/2 \\ (y_2 + y_4)/2 \\ (y_3 + y_5)/2 \\ (y_4 + y_6)/2 \\ (y_5 + y_7)/2 \\ y_6 \end{pmatrix}$$

Spatial spillover

- OLS: $y = X\beta + \epsilon$
- OLS with spatially lagged DV: $y = \rho W y + X\beta + \varepsilon$

- Conventional: One unit increase in X value is associated with β unit increase in Y.
- Spatial spillover: One unit increase in neighbors' weighted Y value is associated with ρ unit increase in Y.

Let’s run a spatial regression in python!

- First stage / AITTE: $\mathbf{X_{1,i,t-1} = \pi_0 + \pi_1 Z_{i,1995} + \alpha_n C_{n\ i,t-1} + \eta_{1,k} + \eta_{2,t} + \upsilon_{i,t}}$
- Second stage / LATE: $\mathbf{Y_{i,t} = \beta_0 + \beta_1 X_{1,i,t-1} + \lambda_n C_{n,i,t-1} + \eta_{1,k} + \eta_{2,t} + u_{i,t}}$

The revised version of the second stage equation (LATE) of the baseline models is as follows:

- $Y_{i,t} = \beta_0 + \beta_1 X_{1,i,t-1} + \lambda_n C_{n,i,t-1} + \mathbf{\lambda_0 Y_{i,t-1}} + \mathbf{\lambda_1 W_i X_{1,i,t-1}} + \mathbf{\lambda_2 W_i Y_{i,t-1}} + \eta_{2,t} + u_{i,t}$

- W_i stands for the first order standardized weighting matrix,
- $Y_{i,t-1}$ is Bolsa Familia in a given i^{th} municipality and $t\text{-}1^{th}$ year,
- $W_i Y_{i,t-1}$ is the weighted Bolsa Familia in the neighboring municipalities of the given i^{th} municipality and $t\text{-}1^{th}$ year,
- $W_i X_{1,i,t-1}$ is the weighted cumulative land invasions in the neighboring municipalities of the given i^{th} municipality and $t\text{-}1^{th}$ year.
- $\lambda_0, \lambda_1, \lambda_2$ are coefficients of Durbin terms in the dynamic spatial models.