2020 STATA ECONOMETRICS WINTER SCHOOL

Anabela Carneiro[1] João Cerejeira[2] Miguel Portela[2,3] Paulo Guimarães[1,4]

[1]FEP and CEF.UP – U.Porto
[2]NIPE – UMinho
[2]IZA, Bonn
[4]Banco de Portugal
Faculdade de Economia da Universidade do Porto

January 20-24, 2020

What is Spatial Econometrics?

Definitions of Spatial econometrics:

- Jean Paelinck introduced the term Spatial Econometrics in 1974 to designate: a combination of economic theory, mathematical formalization and statistics with: role of spatial interdependence, importance of factors in other places, explicit modelling of space".
- Luc Anselin (1988) defined the Spatial Econometrics as: econometric branch dealing with spatial interaction and spatial structure in cross-sectional models and data panel

(separating spatial dependence and spatial heterogeneity) .

Why is Spatial Econometrics important nowadays?

Theory-driven

- From individual decision to social-spatial interaction.
- Common shocks.
- Peer-effects, contextual effects, neighbourhood effects.

Data-driven

- Geo-referenced information.

Technology

- Geographical Information Systems.
- Capability of statistical software.

Thematic maps:

- Thematic maps represent the spatial distribution of a phenomenon of interest within a given study area.
- Spatial data usually distributed as ESRI shape files. The format uses three files: .shp and .shx files contain the map information while .dbf contains observations on each spatial unit.
- We need to obtain and translate these files into Stata: spshape2dta
- Alternative format: Mapinfointerchange format: mif2dta

Where can we find shapefiles?

Eurostat:
https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/

http://www.gadm.org/country

Example with Stata

We will

1. Find and download a European NUTS shapefile.
2. Translate the downloaded file to Stata format.
3. Merge the translated file with our existing data.
4. Analyze the merged data.

Creating the Stata format shapefile

- download the NUTS 2016 zip file at "https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts#nuts16"

```
spshape2dta NUTS_RG_10M_2016_3035, replace
use NUTS_RG_10M_2016_3035
save map_nuts_europe, replace
```

Merging our data with the Stataformat shapefile

Import Eurostat data with `eurostatuse` command

```
eurostatuse lfst_r_lfu3rt, noflags nolabel long geo()
```

Merge merge our existing data with the Stata-format shapefile

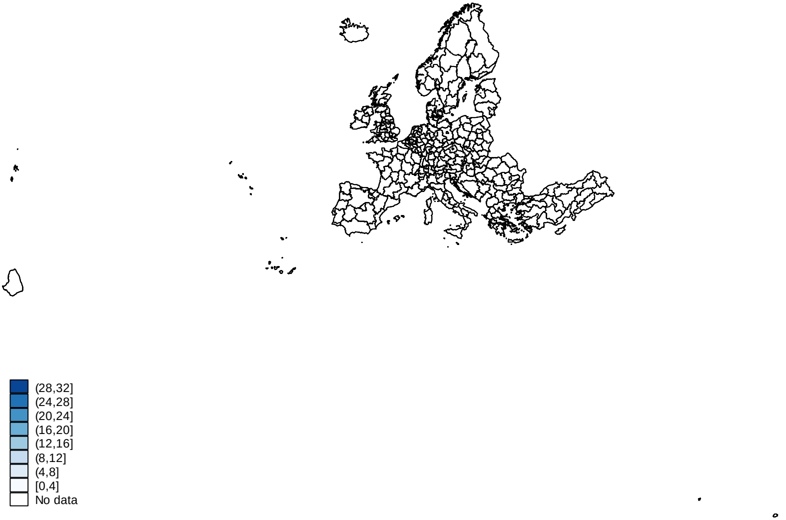
My data and my first map

```
describe
summarize

grmap unemp_eur
```

Unemployment Rate

EU 2016



ESDA: Spatial autocorrelation

Spatial autocorrelation:
"two or more objects that are spatially close tend to be more similar (or more different) to each other with respect a particular characteristic than other that are spatially distant. (spatial clustering)"
How to measure spatial proximity? Spatial weights matrix

Spatial weight matrix (W):

Spatial autocorrelation in a formal way:

$$\begin{bmatrix} y_i \\ y_j \\ y_k \end{bmatrix} = \begin{bmatrix} 0 & \alpha_{ij} & \alpha_{ik} \\ \alpha_{ji} & 0 & \alpha_{jk} \\ \alpha_{ki} & \alpha_{kj} & 0 \end{bmatrix} \begin{bmatrix} y_i \\ y_j \\ y_k \end{bmatrix} + \begin{bmatrix} u_i \\ u_j \\ u_k \end{bmatrix}$$

Strategy of identification:

$$A = \begin{bmatrix} 0 & \alpha_{ij} & \alpha_{ik} \\ \alpha_{ji} & 0 & \alpha_{jk} \\ \alpha_{ki} & \alpha_{kj} & 0 \end{bmatrix} = \rho \begin{bmatrix} 0 & w_{ij} & w_{ik} \\ w_{ji} & 0 & w_{jk} \\ w_{ki} & w_{kj} & 0 \end{bmatrix} = \rho W$$

We transform a non-identified model in other that contains only one parameter: ρ.
W captures 'who is the neighbour of whom': must be EXOGENOUS!

Spatial weight matrix (W)

- $N \times N$ matrix

Different specifications:

- Geographical:
  - Contiguity matrix: neighbours (Rook / Queen; 1st order / 2nd order)
  - Inverse-distance matrix (usually normalized – row)
  - Inverse-distance matrix with threshold
  - K nearest neighbours.
- Socio-economic
  - Flows (asymmetric) . . .
  - Similarity degree in economic dimensions (or social networks).
- Combination of both

Generating W with Stata

`spmatrix` - Stata command to create, import, manipulate, and export W spatial weighting matrices
`spgen` - generates the spatially lagged variable (wrtitten by Kondo

(2017)). Useful command if you have individual data with large N (firms, individuals...)

Univariate spatial tests

Some measures of global spatial autocorrelation and allow us to know its significance.

- Moran's I test (1950):

$$I = \frac{N}{W} \frac{\sum_i \sum_j (y_i - \overline{y}) w_{ij} (y_j - \overline{y})}{\sum_j (y_i - \overline{y})^2},$$

where where N is the number of spatial units indexed by i and j and W is the sum of all w_{ij}. If W is row standardized, then I is the coefficient of a regression of WY on Y.Values of I usually range from -1 to $+1$. Values significantly below $-1/(N-1)$ indicate negative spatial autocorrelation and values significantly above $-1/(N-1)$ indicate positive spatial autocorrelation. For statistical hypothesis testing, Moran's I values can be transformed to $z - scores$.

Univariate spatial tests

- Geary C test (1954):

$$C = \frac{N-1}{2W} \frac{\sum_i \sum_j w_{ij}(y_j - y_j)^2}{\sum_j (y_i - \overline{y})^2},$$

Moran's I is inversely related to Geary's C, but it is not identical. Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation.
The value of Geary's C lies between 0 and some unspecified value greater than 1. Values significantly lower than 1 demonstrate increasing positive spatial autocorrelation, whilst values significantly higher than 1 illustrate increasing negative spatial autocorrelation.

- Getis-Ord G test (1992):

$$G = \frac{\sum_i \sum_{j \neq i} w_{ij} y_j y_j}{\sum_i \sum_{j \neq i} y_j y_j},$$

Null hypotheses of tests: No spatial autocorrelation.

Local indicators of spatial association

A local index of spatial autocorrelation expresses, for each region rigion of a given study area A, the degree of similarity between that region and its neighboring regions with respect to a numeric variable Y
Moran's local index of spatial autocorrelation:

$$I_i = \frac{(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2} \sum_{j=1;j\neq i}^{n} w_{ij}(y_j - \overline{y})$$

Null hypotheses is no spatial autocorrelation and the significance of I_i could be contrasted using normal distribution:

$$z[I_i] = \frac{[I_i - E[I_i]]}{\sqrt{Var[I_i]}}$$

This test allows grouping observations in 4 categories (see scatter Moran): High-High (H-H), Low-Low (L-L), Low-High (L-H) and High-Low (H-L).

Taxonomy of spatial models

General model:

$$\begin{aligned} y &= \lambda Wy + X\beta + WX\gamma + u \\ u &= \rho Wu + \varepsilon \end{aligned}$$

If:

- $\gamma = 0$, $\rho = 0$, $\lambda \neq 0$ $\rightarrow$SLM (Spatial Lag Model)
- $\gamma = 0$, $\rho \neq 0$, $\lambda = 0$ $\rightarrow$SEM (Spatial Error Model)
- $\gamma = 0$, $\rho \neq 0$, $\lambda \neq 0$ $\rightarrow$SARAR (Spatial Autorregressive Model with Autocorrelated errors)
- $\gamma \neq 0$, $\rho = 0$, $\lambda \neq 0$ $\rightarrow$SDM (Spatial Depence Models)

Models with a spatial lag of the dependent variable

$$y_{ue} = \beta_0 + \beta_1 X_{cr} + \beta_2 W y_{ue} + \varepsilon$$

Models with a spatial lag of the independent variable

$$y_{ue} = \beta_0 + \beta_1 X_{cr} + \beta_2 W X_{cr} + \varepsilon$$

Models with spatially autoregressive errors

$$y_{ue} = \beta_0 + \beta_1 X_{cr} + (I - \rho W)^{-1} \varepsilon$$