Co-funded by the Erasmus+ Programme of the European Union

Project funded by
European Commission Erasmus + Programme –Jean Monnet Action
Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Advanced Spatial Nonparametric Techniques - Spatial Dependence -

Margherita Gerolimetto and Stefano Magrini

Department of Economics - Ca'Foscari University of Venice

Pisa 23 October 2017

Spatial Dependence

Margherita Gerolimetto and Stefano Magrini

Spatial autocorrelation (or dependence):

- one of the most important concepts in spatial statistics and spatial econometrics
- derived directly from Tobler's first law of geography according to which everything is related to everything else, but near things are more related than distant things (Tobler, 1979)

in particular:

- positive spatial autocorrelation: similar values tend to group in physical proximity
- negative spatial autocorrelation: dissimilar values tend to group in physical proximity
- absence of spatial autocorrelation (or independence): values are randomly distributed across space

Spatial Effects

- spatial autocorrelation/dependence systematic pattern in the spatial distribution of a variable $\rightarrow$ it does have severe implications for the empirical analysis!!!!!!
  - an observation is partly predictable from neighbouring observations; hence, a series of spatially dependent observations contains less information
  - implications are more complex to deal with than in time series

So, spatial dependence means that the observations at location i depend on other observations at locations $i \neq j$

$$y_i = f(y_j) \qquad i \neq j$$

Possible reasons for this:

- **substantive** spatial dependence due to spatial interactions effects across units (e.g. trade flows, migration, knowledge flows...)
- **nuisance** spatial dependence, due to measurement errors or wrongly defined spatial units

Measuring spatial autocorrelation

- Global measures measure autocorrelation in the whole set of locations
- local measures measure autocorrelation in a spatially delimited subset of locations

As for global autocorrelation, the most commonly used is the **Moran's I** statistic (Moran, 1948)

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

where N is the number of observations (i.e., locations), X_i, X_j is the value of the variable in locations i and j, $\bar{X}$ is the (cross-sectional) mean value of the variable and w_{ij} is a spatial weight assigned to the relation between location i and location j.

In matrix notation,

$$I = \frac{N}{S_0} \frac{Z'WZ}{Z'Z}$$

where Z is a vector (with N observations) with the variable X expressed in deviations from the mean, W is $N \times N$ matrix of spatial weights (usually row-normalised and S_0 is a scalar given by the sum of all elements of W (equal to N when W is row-normalised)

Spatial Models

- Spatial Lag Model $(-1 < \rho < 1)$

$$Y = \rho WY + \beta X + \epsilon$$

- Spatial Error Model $(-1 < \lambda < 1)$

$$Y = \beta X + u$$
$$u = \lambda Wu + \epsilon$$

- Spatial Cross-regressive

$$Y = \beta_1 X + \beta_2 WX + \epsilon$$

... and so on..

Note that W, the spatial weights matrix, plays a crucial role and it must be specified in advance.
The choice of W is a very important part of the specification!

Spatial autocorrelation structure and W matrix

- **Formal** expression of the autocorrelation: it is based on the specification of a spatial process yielding a certain covariance structure.

  This involves choosing carefully W, the matrix of spatial weights (usually row-normalised).
  - spatial weights can be based on the geographic arrangement of the observations or contiguity
  - they also can based on distance decay (inverse or inverse square distance)
- **Direct** representation: each covariance between a pair of observations i and j is specified as a parametrized function f of the distances d_{ij}

  This involves choosing f very carefully in order to guarantee that the resulting variance-covariance matrix is positive semidefinite