Introduction to Spatial Econometrics

Spatial Econometrics

- Applied work in regional science, urban and spatial economics, and economic geography relies heavily on sample data that are collected with reference to locations measured as points in space.

Spatial Econometrics

- What distinguishes spatial econometrics from traditional econometrics?
- Two problems arise when sample data has a locational component:
- 1) spatial dependence exists between the observations and
- 2) spatial heterogeneity occurs in the relationships we are modeling.

Spatial Dependence

- Spatial dependence in a collection of sample data observations refers to the fact that one observation associated with a location which we might label i depends on other observations at locations j.
- Formally, we might state: $y_i = f(y_j)$; i = 1; …; n
- Note that we allow the dependence to be among several observations, as the index i can take on any value from i = 1; …; n.

- Why would we expect sample data observed at one point in space to be dependent on values observed at other locations?

Measurement Error

- First, data collection of observations associated with spatial units such as zip-codes, counties, states, census tracts and so on, might reflect measurement error.
- This would occur if the administrative boundaries for collecting information do not accurately reflect the nature of the underlying process generating the sample data.

Example

- As an example, consider the case of unemployment rates and labour force measures.
- Because workers are mobile and can cross county or state lines to find employment in neighboring areas, labour force or unemployment rates measured on the basis of where people live could exhibit spatial dependence.

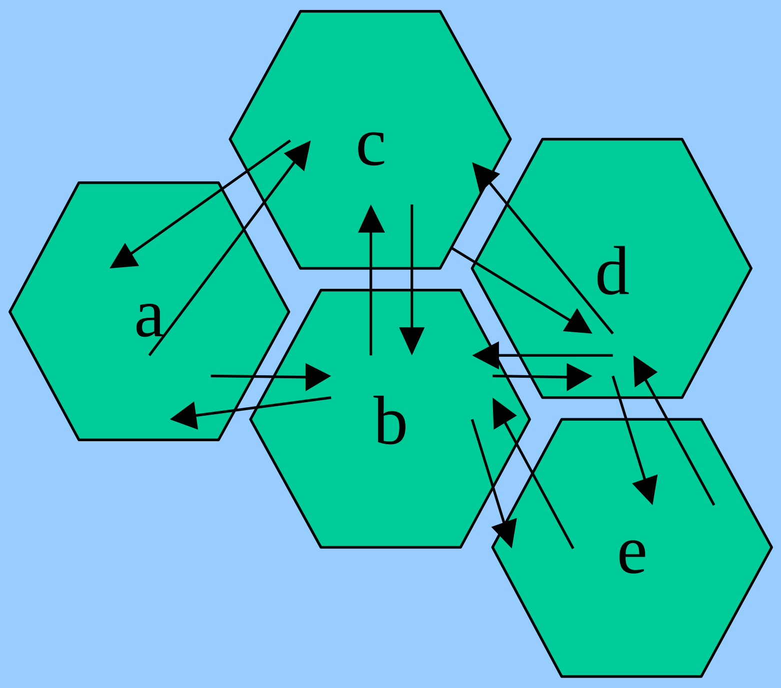
Spatial Interactions

- A second and perhaps more important reason we would expect spatial dependence is that the spatial dimension of socio-demographic, economic or regional activity may truly be an important aspect of a modeling problem.
- Spatial theory is based on the premise that location and distance are important forces at work in human geography and market activity.
- In order to explain economic landscapes we need to understand: spatial interaction and difusion effects, hierarchies of place and spatial spillovers.

Examples

- House prices in neighbouring areas to London are likely to be impacted by house prices in London
- Crime rates are likely to be similar in neighbouring areas
- Employment levels are likely to be similar in neighbouring areas

Spatial Spillovers



Spatial heterogeneity

- The term spatial heterogeneity refers to variation in relationships over space.
- In the most general case we might expect a diferent relationship to hold for every point in space.
- We simply do not have enough sample data information with which to produce estimates for every point in space, a phenomena referred to as a \degrees of freedom" problem.
- To proceed with the analysis we need to provide a specification for variation over space. This specification must be parsimonious, that is, only a handful of parameters can be used.

Questions arise regarding:

1) how sensitive the inferences are to a particular specification regarding spatial variation?
2) is the specification consistent with the sample data information?
3) how do competing specifications perform and what inferences do they provide?

Restrictions on Spatial Variation

- One can also view the specification task as one of placing restrictions on the nature of variation in the relationship over space.
- For example, suppose we classified our spatial observations into urban and rural regions. We could then restrict our analysis to two relationships, one homogeneous across all urban observational units and another for the rural units.

Spatial Data Analysis

Gini Coeficients and Lorenz Curves

- It is possible to construct measure of inequality with respect to spatial distributions of variables through the Lorenz Curves and Gini Coefficients.
- The Lorenz Curve plots cumulative percentages of the total values of a particular variable against cumulative percentages of the number of regions.
- A point on the curve gives the percentage of the regions for a given percentage of the total value of the variable in study.

Gini Coeficients and Lorenz Curves

- The Gini Coefficient gives the area between the Lorenz Curve and the line of absolute equality (45 degrees) as a proportion of the total area under the line of absolute equality
- The Gini Coefficient can be written as

Gini Coeficients

$$G = 1 + \frac{1}{n} - \frac{2}{n_2 \bar{y}} (y_1 + 2y_2 + \cdots + ny_n)$$

where

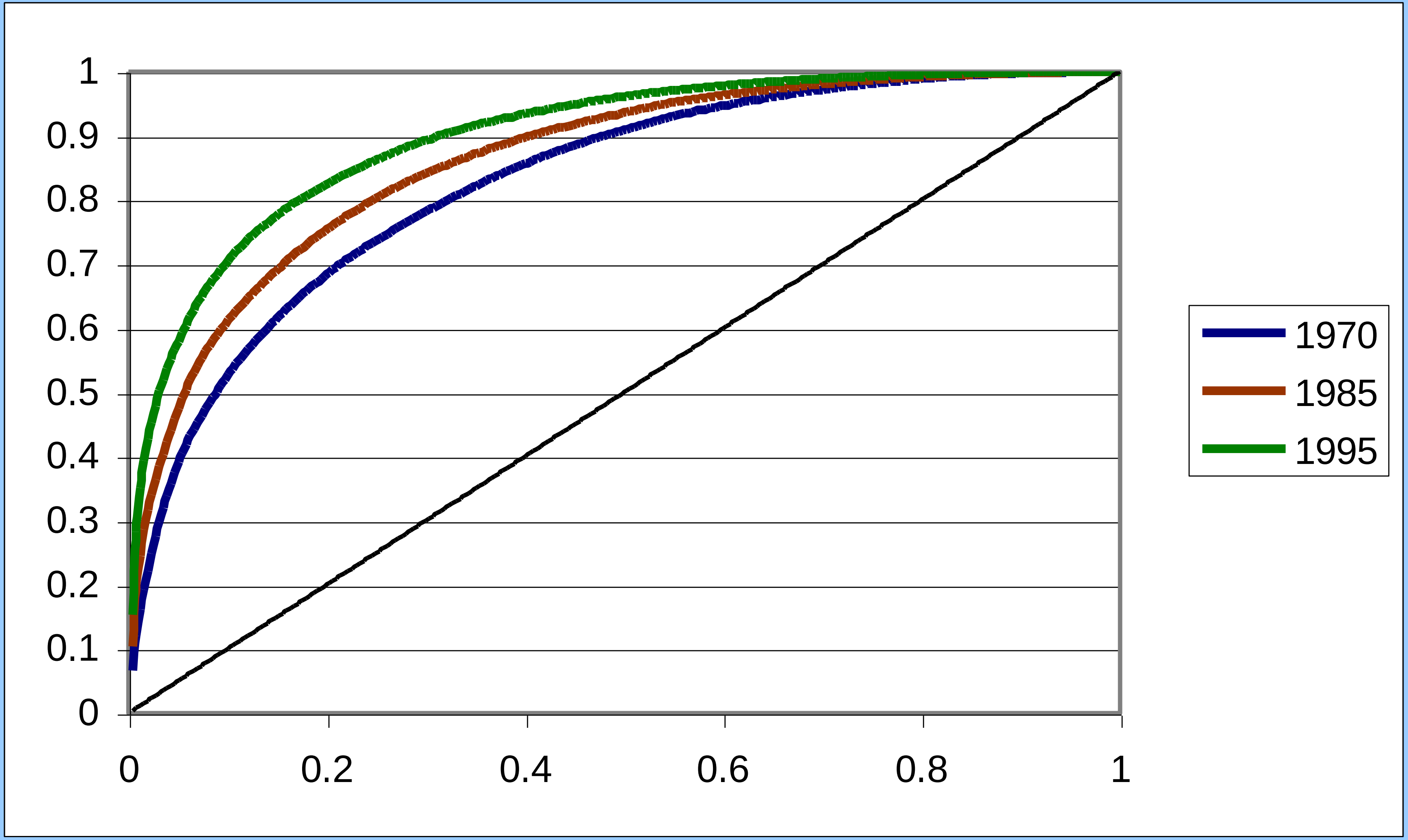
$y_1, y_2, \ldots, y_n$ represent local values in decreasing order of size
is the mean value of the variable in study
n is the number of geographical units

Example: Brazilian Amazon

Table 2. 13. Evolution of Gini Coefficients

	1970	1985	1995
Population	0.2909	0.3277	0.3466
Output	0.2895	0.3393	0.3666
Cleared Land	0.3269	0.3618	0.3977

Diagram 2.6. Cleared Land – Lorenz Curves



The Spatial Weights Matrix (W)

- In order to be able to apply spatial methods we normally construct a so-called Spatial Weight Matrix (*W* matrix henceforth), which is a square matrix of dimension N (N being the number of spatial units in the sample).
- The values in W reflect an *ad-hoc* hypothesis of spatial interaction between the municipalities.
- The diagonal contains zeros, and the off-diagonal elements reflect the spatial proximity between the municipalities.

The Spatial Weights Matrix (W)

$$W = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & 0 & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,n-1} & 0 \end{bmatrix}_{n \times n}$$

Spatial Weights

- Contiguity (1 if 2 units have a common border, 0 otherwise)
- Inverse of distance between any 2 centroids
- Inverse of squared distance between any 2 centroids
- Travel time

Spatial Lags

$$
{}_{n}\begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & 0 & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,n-1} & 0 \end{bmatrix}_{n} \times {}_{1}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n} = {}_{1}\begin{bmatrix} (Wy)_1 \\ (Wy)_2 \\ \vdots \\ (Wy)_n \end{bmatrix}_{n}
$$

$$
Wy = W \times \mathbf{y}
$$

$$
(Wy)_i = \sum_{j=1}^{n} w_{i,j} \times y_j
$$

Normalisation

- A further step in the construction of the *W* matrix is to standardise it so that each row sums to 1.
- Standardising helps with interpretation, since the value for area *j* of the spatial lag, defined as the *j*'th cell of *Wx*, is then the weighted average of the values of the variable *x* in the areas that are 'neighbours' to J.
- Also, using the standardised *W* matrix usefully identifies a parameter value below 1 as being consistent with a 'non-exploding' process while 1 and above leads to complex and little understood consequences for inference and estimation

Example

$$W_{ij}^{*} = \frac{1}{d_{ij}^{2}}$$

$$W_{ij} = \frac{W_{ij}^{*}}{\sum_{j} W_{ij}^{*}}$$

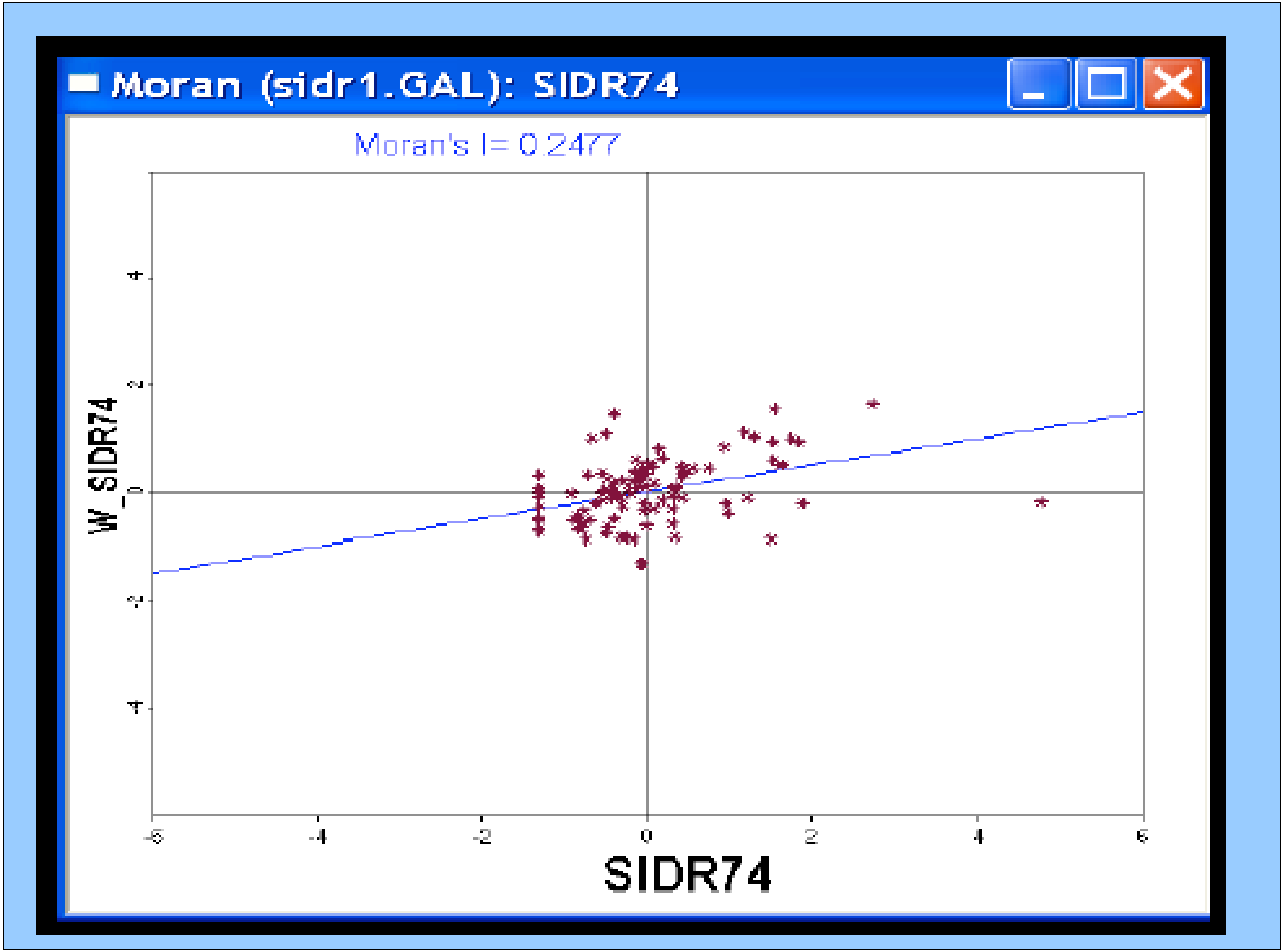
diminishing function of distance
standardised so that each row sums to 1.

Moran Scatterplots

- Linear Spatial Autocorrelation
- linear association between value at i and weighted average of neighbors: Wy and y
- four quadrants

high-high, low-low= spatial clusters

high-low, low-high= spatial outliers



Moran (sidr1.GAL): SIDR74
Moran's I= 0.2477
W_SIDR74
SIDR74
-6
-4
-2
0
2
4
6

Econometric Models

Spatial Dependence

- It is possible to have spatial dependence in 3 components of an econometric model:
- Dependent variable (Y)
- Indepedent variables (X)
- Residuals (e)

Classical Model

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$$

$$\mathbf{u} = + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$$

Spatial Lag Model

$$\mathbf{y} = \rho W_1 \mathbf{y} + X\boldsymbol{\beta} + \mathbf{u}$$

$$\mathbf{u} = \boxed{\qquad\qquad} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$$

Spatial Error Model

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$$

$$\mathbf{u} = \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$$

General Spatial Model

$$\mathbf{y} = \rho W_1 \mathbf{y} + X\boldsymbol{\beta} + \mathbf{u}$$

$$\mathbf{u} = \lambda W_2 \mathbf{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$$

Estimation Issues

- In case of spatial correlation only in the independent variables (X) it is possible to estimate the parameters using OLS. The only thing you need to do is to include WX as covariates
- In case of spatial correlation in the dependent variable (y) or residuals (e), other estimation methods are required.

Estimation Alternatives

- Instrumental Variables (Fingleton)
- Maximum likelihood (Anselin)
- General Methods of Moments (Kelejian and Prucha)

Testing

- In order to correctly specify your econometric model you have to test for spatial autocorrelation in the residuals of your regressions
- Moran’s I
- Others

Moran's I

$$I_t = \frac{n}{S_0} \cdot \frac{\sum_i \sum_j w_{ij}(x_{it} - \mu_t)(x_{jt} - \mu_t)}{\sum_t (x_{it} - \mu_t)^2}$$

Where
x_{it} is the observation in region *i* and year *t*,
μ_t is the mean of the observations across geographic areas in year *t*,
n is the number of areas,
w_{ij} is one element of the spatial weights matrix *W*,
and S_0 is a scaling factor equal to the sum of all elements of *W*.

Software

- SPACESTAT
- GEODA
- MATLAB (Tool Box by James Le Sage)
- STATA (GMM routines by Tim Conley)
- R
- IPEA GEO

References

- Anselin, L. (1988). *Spatial Econometrics : Methods and Models*. Dordrecht: Kluwer.
- James Le Sage: spatial econometrics toolbox (see www.spatial-econometrics.com)
- Fingleton, B. (ed) (2003) *European Regional Growth*, Berlin: Springer Verlag.
- Arbia, G. (2009), *Spatial Econometrics : Methods and Applications*, Springer Verlag
- Gibbons, Stephen, and Henry Overman (2010). Mostly Pointless Spatial Econometrics?, Journal of Regional Science
- Corrado L. & Fingleton B. (2012) Where is the economics in spatial econometrics? Journal of Regional Science