Spatial Statistics

EMOS Learning Materials

Service Contract n. 2019.0249 between Eurostat and the University of Pisa, Italy

Course Outline

- The first part of the course will be devoted to describing what are spatial data and which are the different types of spatial data. The spatial autocorrelation indices will be introduced. An R example will be provided using EU-SILC data.
- The second part of the course will be focused on the study of areal data exploring the use and the types of spatial econometric models to study the spatial dependency. Geographically weighted regression and spatial models for categorical data will be also introduced. A case study using R will be proposed.

Course Outline: first part

- Exploratory data analysis
- Types of spatial data
- The spatial autocorrelation indices
- Defining neighborhood and spatial weights
- R example

Spatial statistics

- Spatial statistics is the analysis of statistical observations taking into account the position in which they occur in a given space.
- The main aim of spatial analysis is to understand and explore the connection between the spatial positioning of a phenomenon and its characteristics.
- Using specific statistical methods aiming to establish and quantify the presence of dependence between observations in space.

European Commission

Spatial data

A spatial process in d dimensions is denoted as

$$\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}.$$

- Z denotes the attribute we observe, such as income or employment rate
- $\mathbf{s}$ is the location at wich Z is observed and it is a $(d \times 1)$ vector of coordinates. In two-dimensional space $d = 2$.
- D is possibly a random set in $\mathbb{R}^d$.

European Commission

Typology of spatial data

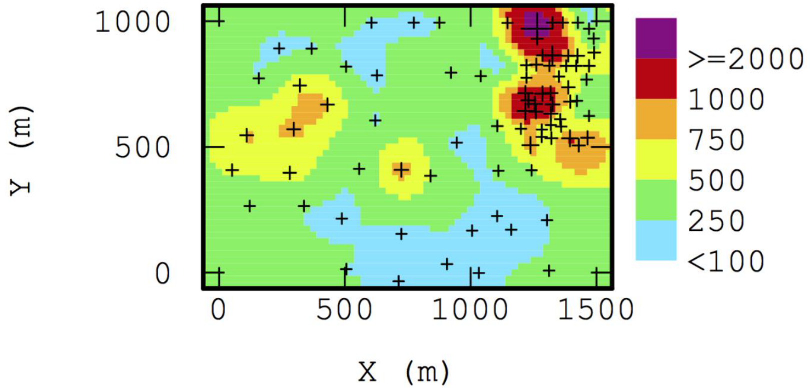
- geostatistical data
- lattice or area (regional) data
- point pattern data

Geostatistical data

$$\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$$

- D is a fixed subset of $\mathbb{R}^d$, with positive d-dimensional volume
- The spatial index $\mathbf{s}$ varies continuosly throughout D
- $Z(\mathbf{s})$ is a random variable at each of the infinite continuous locations $\mathbf{s} \in D$

Predicting pollution



(source: Insee-Eurostat, 2018)

Lattice data

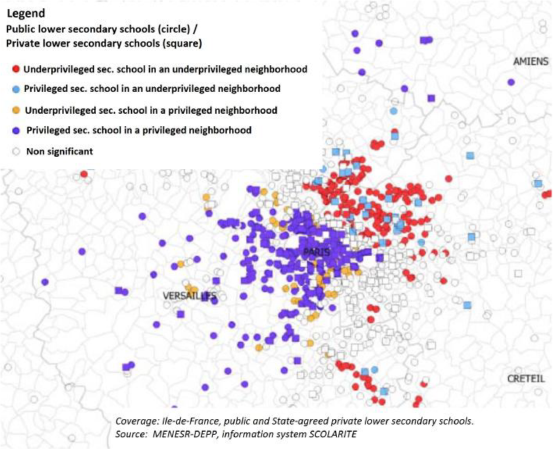
$$\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$$

- lattice data refers to the case where D is a countable collection of spatial sites.
- the neighbors ($N(i)$) of each point are defined

European Commission

Local spatial dependency

Are privileged lower secondary schools always located in a privileged environment?



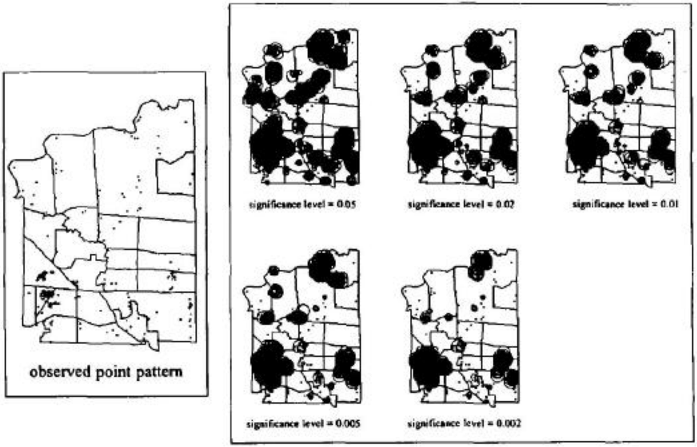
Coverage: Ile-de-France, public and State-agreed private lower secondary schools.
Source: MENESR-DEPP, information system SCOLARITE

(source: Insee-Eurostat, 2018)

Spatial Point patterns

- D is a collection of random events whose realization is called a spatial point pattern
- The data are the locations of the sites themselves, and the collection of all the sites is the event of interest. The data do not consist of realizations of some random variable at a given site.

Cluster Detection



(source: Insee-Eurostat, 2018)

The importance of spatial data

- *Integrating geographical and statistical information offers significant opportunities to maximise the utility of data collected for statistical purposes* (Eurostat website)
- GISCO activities
- GEOSTAT activities. Merging Statistics and Geoinformation grants
- The Global Statistical Geospatial Framework which facilitates the integration of statistical and geospatial information

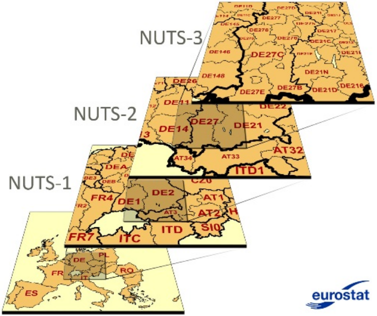
European Commission

Mapping and Geovisualisation

- Exploratory spatial data analysis
- Mapping
- Spatial scales (grid cells covering the European land territory, e.g. 1km grid in Eurostat statistical atlas)
- Administrative units:
  - NUTS 1: major socio-economic regions
  - NUTS 2: basic regions for the application of regional policies
  - NUTS 3: small regions for specific diagnoses

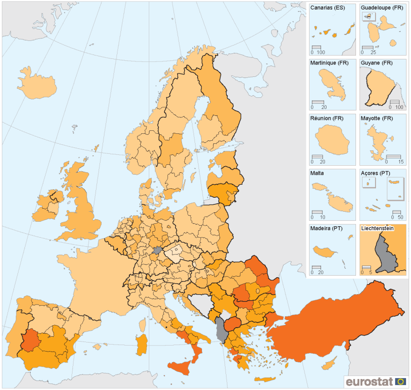
European Commission

Visualization: NUTS



Visualization: example

People at risk of poverty or social exclusion, 2018
(%, by NUTS 2 regions)



EU-27 = 21.6

- < 10
- 10 - < 20
- 20 - < 30
- 30 - < 40
- ≥ 40
- Data not available

Administrative boundaries: © EuroGeographics © UN-FAO © Turkstat
Cartography: Eurostat — GISCO, 04/2020

0 200 400 600 800 km

Note: Poland and Serbia, NUTS 1 regions. Belgium, France, Portugal, the United Kingdom and Turkey: national data. Länsi-Suomi (FI19) and Åland (FI20): the value shown covers both regions. EU-27, Germany and Austria: estimates. The United Kingdom: provisional. Burgenland (AT11): low reliability. Germany, Austria, Montenegro and Turkey: 2017. Iceland: 2016.
Source: Eurostat (online data codes: ilc_peps11 and ilc_peps01)

European Commission

Spatial relation

The first law of geography

Everything is related to everything else, but near things are more related than distant things. (Tobler 1970)

The strength of the spatial relation can be measured by spatial autocorrelation

European Commission

What is Spatial autocorrelation?

- Autocorrelation measures the correlation of a variable with itself.
- It allows to detect regularities in the variable.
- Spatial autocorrelation allows us to describe the degree to which observations (values) at spatial locations are similar to each other

Spatial autocorrelation

- If $Z(\mathbf{s})$ is the attribute Z observed in the plane at spatial location $\mathbf{s} = [x, y]$, the term spatial autocorrelation refers to the correlation between $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$.
- Many statistical analysis are based on the hypothesis of independence of variables, therefore when a variable is spatially auto-correlated, the independence hypothesis is no longer respected.

Spatial autocorrelation

- If there is no autocorrelation the spatial allocation of the observations is random: if two points $\mathbf{s}_i$ and $\mathbf{s}_j$ are close or far has no bearing on the relationship between the values $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$.
- Spatial autocorrelation is positive when similar values tend to be geographically grouped; similar values are close to each other, or cluster, in space
- Spatial autocorrelation is negative when nearby locations are more different than remote locations, or, equivalenty, that similar values are dispersed

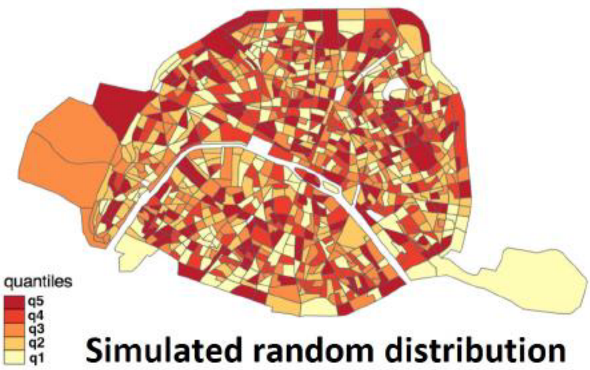
European Commission

Illustration Spatial autocorrelation

- Spatial autocorrelation helps to analyze whether the parameters like income are clustered or uniformly distributed in a certain region.

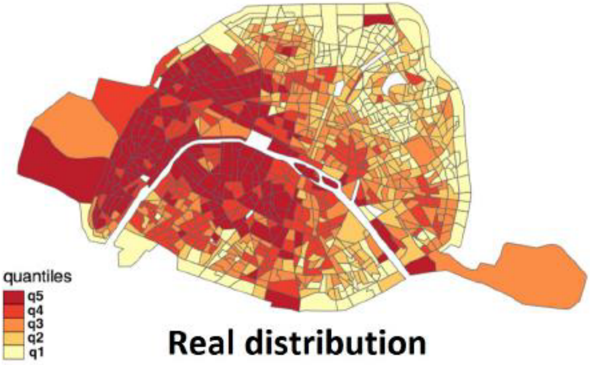
  Example: Evaluate if exist some spatial structure in median income in Paris by IRIS, aggregated units for statistical information. (Example from Insee-Eurostat, 2018)

Illustration Spatial autocorrelation



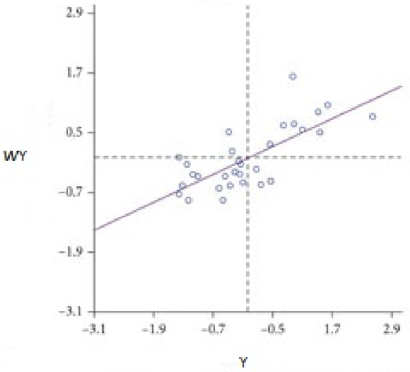
(source: Insee-Eurostat, 2018)

Illustration Spatial autocorrelation

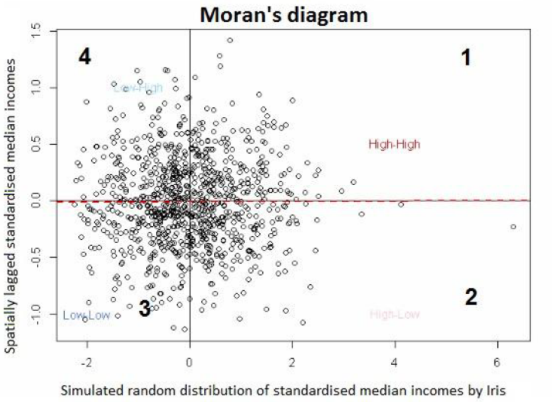


(source: Insee-Eurostat, 2018)

Moran's diagram

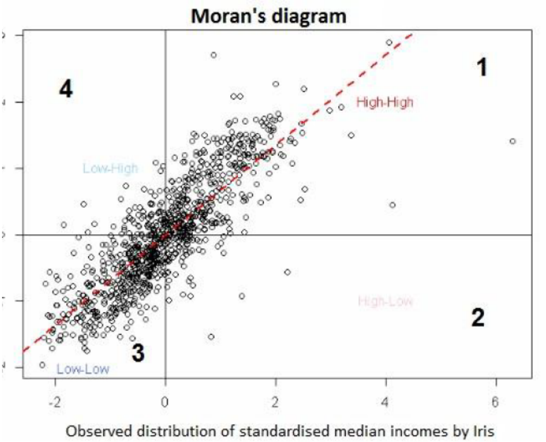


Moran's diagram: Example income in Paris



(source: Insee-Eurostat, 2018)

Moran's diagram: Example income in Paris



(source: Insee-Eurostat, 2018)

Test spatial autocorrelation

- Test if the value of a variable at one location is independent of values of that variable at neighboring locations.
- Could the values taken by the neighbouring observations have been similar (or dissimilar) by mere chance?

$$\begin{cases} H_0 & \text{no spatial autocorrelation} \\ H_1 & \text{spatial autocorrelation} \end{cases}$$

Moran's I

$$I_W = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})}, \quad i \neq j$$

- y_i is the value of a variable for the *i*th observation
- $\bar{y}$ is the sample mean
- w_{ij} is a spatial weight
- Values range from -1 (perfect dispersion) to $+1$ (perfect correlation). A zero value indicates a random spatial pattern.
- Under H_0, $\mathrm{E}[I_W] = -\dfrac{1}{n-1}$

European Commission

Moran's I

The variance of Moran's I is a little more complicated

$$\mathrm{var}(I) = \frac{ns_1 - s_2 s_3}{(n-1)(n-2)(n-3)(\sum_i \sum_j w_{ij})^2}$$

$$s_1 = (n^2 - 3n + 3)(\tfrac{1}{2} \textstyle\sum_i \sum_j (w_{ij} + w_{ij})^2)$$

$$s_2 = \frac{n^{-1} \sum_i (y_i - \bar{x})^4}{(n^{-1} \sum_i (y_i - \bar{x})^2)^2}$$

$$s_3 = \frac{1}{2} \textstyle\sum_j (w_{ij} + w_{ij})^2 - 2n(\tfrac{1}{2} \sum_i \sum_j (w_{ij} + w_{ij})^2) + 6(\sum_i \sum_j w_{ij})^2$$

European Commission

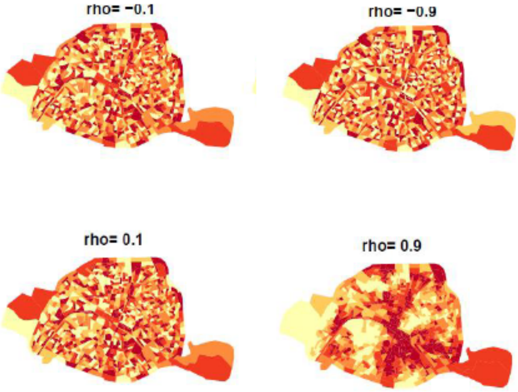
Moran's I

$$\frac{I_W - \mathrm{E}[I_W]}{\sqrt{\mathrm{Var}[I_W]}} \sim N(0,1)$$

Assumption:

- Normality hypothesis
- Randomisation hypothesis

Moran's I: visualization



(source: Insee-Eurostat, 2018)

European Commission

Geary's C

Another index to measure the spatial autocorrelation is the Geary's C

$$C_W = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\sum_{i=1}^{n}(y_i - \bar{y})}$$

- It ranges from 0 (perfect correlation) to 2 (perfect dispersion) where 1 is no spatial autocorrelation. Positive spatial autocorrelation is found with values ranging from 0 to 1 and negative spatial autocorrelation is found between 1 and 2.

European Commission

Geary's C

Test Geary's C

- Under H_0, $\mathrm{E}[I_W] = -\frac{1}{n-1}$

$$\frac{C_W - \mathrm{E}[C_W]}{\sqrt{\mathrm{Var}[C_W]}} \sim N(0, 1)$$

Spatial autocorrelation of categorical variables

- When the variable of interest is not continuous but categorical, the degree of local association is measured by analysing the statistics of the join count
- Binary variable representing two colours, White (B) and Black (N) so that a relation can be called W-W(0-0), B-B(1-1) or W-B(0-1)

European Commission

Spatial autocorrelation of categorical variables

Join count analysis:

- positive spatial autocorrelation (clustering) if the number of W-B joins is significantly lower than what would have resulted with random spatial distribution
- negative spatial autocorrelation (dispersion) if the number of W-B joins is significant greater than what would have resulted with random spatial distribution
- null spatial autocorrelation if the number of W-B joins is approximately the same to what would have occurred with random spatial distribution

Spatial autocorrelation of categorical variables

$$P_B = \frac{n_B}{n} \quad P_W = \frac{n_W}{n} = 1 - P_B$$

- n is the number of observations
- n_W is the number of white observations
- $n_B = n - n_W$ are the black observations

Spatial autocorrelation of categorical variables

In the presence of null spatial autocorrelation, the probabilities of observations of the same colour occurring in two neighbouring cells are

$$P_{BB} = P_B \cdot P_B = P_B^2 \quad P_W = (1 - P_B) \cdot (1 - P_B) = (1 - P_B)^2$$

the probability of observations of different colour occurring in two neighbouring cells is

$$P_{BW} = P_B \cdot (1 - P_B) + (1 - P_B) \cdot P_B = 2P_B \cdot (1 - P_B)$$

Spatial autocorrelation of categorical variables

Expected counts:

$$\mathrm{E}[BB] = \frac{1}{2}\sum_i \sum_j w_{ij} P_B^2$$

$$\mathrm{E}[WW] = \frac{1}{2}\sum_i \sum_j w_{ij}(1 - P_B)^2$$

$$\mathrm{E}[BW] = \frac{1}{2}\sum_i \sum_j w_{ij} 2 P_B (1 - P_B)$$

Assuming a binary connectivity matrix, $\frac{1}{2}\sum_i \sum_j w_{ij}$ is the total number of joins.

European Commission

Spatial autocorrelation of categorical variables

Observed counts:

$$BB = \frac{1}{2} \sum_i \sum_j w_{ij} y_i y_j$$

$$WW = \frac{1}{2} \sum_i \sum_j w_{ij}(1 - y_i)(1 - y_j)$$

$$BW = \frac{1}{2} \sum_i \sum_j w_{ij}(y_i - y_j)^2$$

$y_i = 1$: the observation is black; $y_i = 0$: the observation is white

European Commission

Spatial autocorrelation of categorical variables

A test statistics for the black-white join counts is :

$$Z(BW) = \frac{BW - \mathrm{E}[BW]}{\sqrt{\sigma^2_{BW}}}$$

where σ^2_{BW} is the variance of BW.

The statistic is assumed to be asymptotically normally distributed under the null hypothesis of no spatial autocorrelation

Spatial autocorrelation: recap

- Spatial autocorrelation allows to determine if the overall spatial distribution of the variable of interest was reflective of a geographically random process.
- Measures of spatial autocorrelation are useful beacuse the presence of spatial autocorrelation has important implications for the next statistical analysis
- Determine whether spatial autocorrelation is present in a data set is a critical aspect
- Using the global measures of spatial autocorrelation we obtain a single statistic for the whole data set.

European Commission

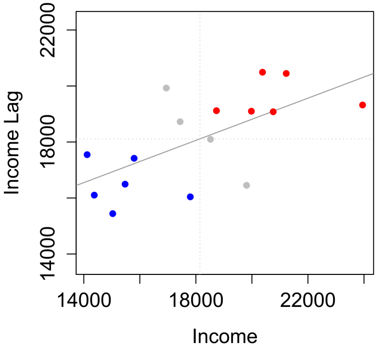
Local Spatial autocorrelation

- Using global spatial autocorrelation statistics (Moran's I and Geary's c) we can reject the null hypothesis of spatial randomness in favour of an alternative of clustering.
- Such clustering does not provide an indication of the location of the clusters.

Local Spatial autocorrelation: example

- Considering the income as our target variable in a given country.
- We are interested in evaluating which are the regions with high (low) income values surrounded by high (low) income regions

Local Spatial autocorrelation: example



Local Spatial autocorrelation

- Global statistics are based on the assumption of a spatial stationary process
- Using (local) measures for identifying local patterns of spatial association and local instabilities in overall spatial association

A first measure: Getis and Ord index

$$G_i = \frac{\sum_j w_{ij} y_j}{\sum_j w_{ij}}$$

- $G_i > 0$ grouping of values higher than average.
- $G_i < 0$ grouping of values lower than average.

This index allows to detect significant groupings of identical values around a particular location (clusters)

Local spatial autocorrelation indicators (LISA)

- These indicators allow for the decomposition of global indicators into the contribution of each individual observation
- They allow to detect clusters (units with similar neighbors) and hotspots (units with dissimilar neighbors)
- They allow to identify spatial non-stationarity zones, which do not follow the global process.

Local Indicators of Spatial Association properties

- For each observation, Local Indicators of Spatial Association (LISA) indicate the intensity of the grouping of similar or values around this observation;
- The sum of local indices on all observations is proportional to the corresponding global index.

European Commission

Local Moran's I

The local Moran's I for unit i is:

$$I_i = (y_i - \bar{y}) \sum_j w_{ij}(y_j - \bar{y})$$

- $I_i > 0$: a grouping of similar values (cluster).
- $I_i < 0$: a unit has neighboring units with dissimilar values

European Commission

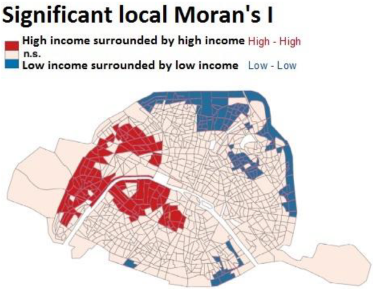
Local Moran's I: example

Values of local Moran's I, on Parisian IRIS

(source: Insee-Eurostat, 2018)

Local Moran's I: example

Significant local Moran's I, on Parisian IRIS



(source: Insee-Eurostat, 2018)

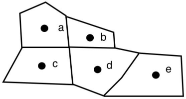
European Commission

Defining neighbours

- To measure the spatial autocorrelation, we need to know the proximity of our observations: how do we define our neighbourhood?
- Need to impose structure on the extent of spatial interaction
- Spatial connectivity (neighborhood) is defined based on the conntections between units in the data: how do we know which points are in our neighbourhood and which ones are not?

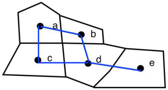
Defining neighbours

Map



Map with areas and their centroids

Specification



Neighbourhood graph

	X	Y
a	x_a	y_a
b	x_b	y_b
c	x_c	y_c
d	x_d	y_d
e	x_e	y_e

Spatial configuration

Codification

	a	b	c	d	e
a	0	1	1	0	0
b	1	0	1	1	0
c	1	1	0	1	0
d	0	1	1	0	1
e	0	0	0	1	0

Neighbourhood matrix
Similarity

cut-off rule

	a	b	c	d	e
a	0	d_{ab}	d_{ac}	d_{ad}	d_{ae}
b	d_{ab}	0	d_{bc}	d_{bd}	d_{be}
c	d_{ca}	d_{cb}	0	d_{cd}	d_{ce}
d	d_{da}	d_{db}	d_{dc}	0	d_{de}
e	d_{ea}	d_{eb}	d_{ec}	d_{ed}	0

Distance matrix
Dissimilarity

(source: Insee-Eurostat, 2018)

Connectivity matrix

- The connectivity, or neighbourhood, matrix C measures how similar observations are
- C is an $n \times n$ binary matrix, $i = \{1, 2, \cdots, n\}$ and $j = \{1, 2, \cdots, n\}$ are the units in the system
  - $c_{ij} = 1$ if two units, i and j, are considered connected (spatially linked)
  - $c_{ij} = 0$ if they are not

European Commission

Spatial neighbours based on contiguity

- Neighborhoods can be defined in a different ways:
  - Contiguity (common boundary): what is a shared boundary?
  - Distance (K-nearest neighbors, distance band)
  - General weights (social distance, distance decay)

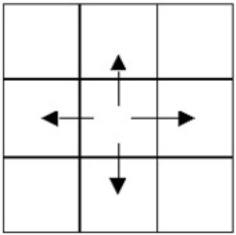
Spatial neighbours based on contiguity

- Spatial neighbours based on contiguity (used to study social and demographic data).
- One way to represent the spatial relationships is through the concept of contiguity. First order contiguous neighbours are defined as areas that have a common boundary.

European Commission

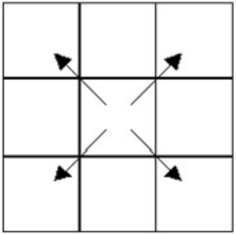
Rook’s contiguity

Cells sharing a common edge are considered contiguous



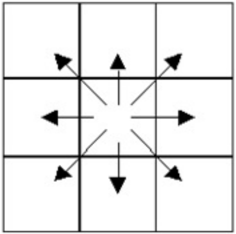
Bishop's contiguity

cells sharing a common vertex are considered contiguous



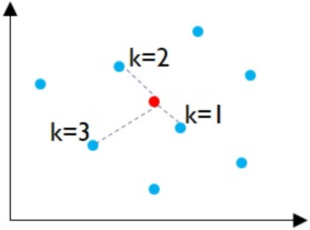
Queen's contiguity

Cells sharing a common edge or common vertex are considered contiguous



Spatial neighbours based on distance

- Neighbourhood graphs based on nearest neighbours
- Finding the k closest observations for each observation (k is an integer)
- x_j is a neighbor of x_i if $xj \in N_k x_i$, where $N_k x_i$ are the k nearest neighbors of x_i



Spatial neighbours based on distance

- Assign neighbors based on a specified distance
- Neighbors of unit x_i defined by interpoint distance:
  - Lower bound: 0
  - Upper bound: $max_{i=1}^{n}(min_{j\neq 1}^{n-1} d(x_i, x_j))$

Spatial weights

- The weights matrix reflects the neighbour definition
- The weights matrix is the "*formal expression of spatial dependency between observations*" (Anselin et al. 1988)
- Translate binary indicators into weights, which will form the elements w_{ij} of matrix W

European Commission

Spatial weights matrix

- W is a $n \times n$ matrix
- $w_{ij} \neq 0$ if i and j are neighbours
- $w_{ij} = 0$ if i and j are not neighbours
- $w_{ii} = 0$
- Usually, the weights matrix is row-standardized, $\sum_{j=1}^{n} w_{ij} = 1$

European Commission

Binary spatial weights matrix

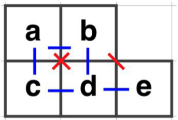
- Commonly, the weight matrix is a binary contiguity matrix

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are spatially linked to each other} \\ 0 & \text{otherwise} \end{cases}$$

- The spatial weights matrices can take into account the distance between the geographical zones

European Commission

Binary weight matrix



ROOK contiguity

	a	b	c	d	e	Sum of the neighbours weights
a	0	1	1	0	0	2
b	1	0	0	1	0	2
c	1	0	0	1	0	2
d	0	1	1	0	1	3
e	0	0	0	1	0	1

QUEEN

	a	b	c	d	e	Sum of the neighbours weights
a	0	1	1	1	0	3
b	1	0	1	1	1	4
c	1	1	0	1	0	3
d	1	1	1	0	1	4
e	0	1	0	1	0	2

Spatial weights matrix

- The spatial weights matrices can take into account the distance between the geographical zones

Spatial weights matrix

- **Power Distance Weights.** Weights are a negative power function of distance of the form

$$w_{ij} = d_{ij}^{-\alpha}$$

- **Exponential Distance Weights.** Weights are negative exponential functions of distance of the form:

$$w_{ij} = \exp(-\alpha d_{ij})$$

- **Double-Power Distance Weights.** For each positive integer k :

$$w_{ij} = \begin{cases} [1 - (d_{ij}/d)^k]^k & 0 \leq d_{ij} \leq d \\ 0 & \text{otherwise} \end{cases}$$

European Commission

What did you learn?

- Three types of spatial data
- Mapping the data can offer a synthesized view of the situation
- Spatial autocorrelation measures the spatial dependence between values of the same variable in different places in space
  - Global and local versions of spatial autocorrelation indices
- Defining the neighborhood is essential for measuring the strength of the spatial relationships between objects
- Attributing weights to neighbors

European Commission

Example in R

- Now
- Mapping data with R (RStudio)
- Measuring and testing spatial autocorrelation in R (RStudio)
- Using EU data

European Commission

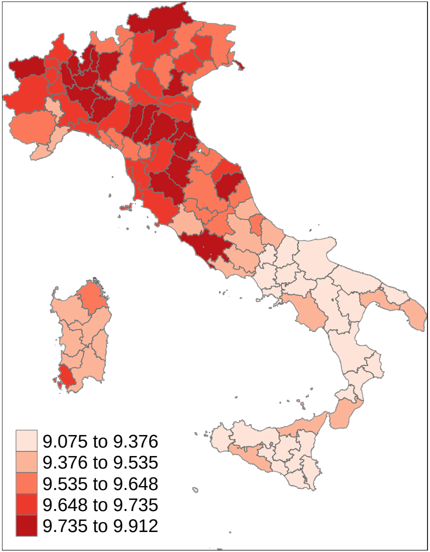
Example in R

```r
# Load shapefile of the Italian provinces
library(rgdal)
setwd("C:/Users/franc/Dropbox/EMOS_LO_fra/Prov01012007")
italy <- readOGR("Prov01012007_WGS84.shp",stringsAsFactors=F)
italy <- italy[order(as.numeric(italy@data$COD_PROV)),]
# Loading data
silcIT3 <- read.csv("silcIT3.csv",sep=";")
# Merge datasets
Silc.Spat <- merge(italy, silcIT3,
                   by.x="COD_PROV", by.y="Prov.Cod")
```

Example in R

```r
# Mapping data
library(tmap)
tm_shape(Silc.Spat) +
  tm_fill("LogIncome",
          palette = "Reds",
          style = "quantile",
          title = "") +
  tm_borders(alpha=.7)  +
  tm_layout("",inner.margins=c(0,0,0,0),
            legend.width=1,
            legend.text.size = 1,
            legend.position = c("left","bottom"),
            legend.bg.color = "white",
            legend.bg.alpha = 1,
            legend.stack = "vertical")
```

Example in R


9.075 to 9.376
9.376 to 9.535
9.535 to 9.648
9.648 to 9.735
9.735 to 9.912

Example in R

```r
# Weight matrix
# To create spatial weight matrices we need to use
# the spdep package
library("spdep")
# neighbour list based on the 'Queen' criteria for
# italian provincens
neigh <- poly2nb(italy,row.names=italy$COD_PROV,queen=TRUE)
# Rook's case neighbors-list
neigh.ROOK <- poly2nb(italy, queen=FALSE)
# Transform the list into an actual matrix W
italy.lw <- nb2listw(neigh, style="W", zero.policy=TRUE)
```

Example in R

```r
# We can also set style="B" for the basic binary coding
italy.sp_w2 <- nb2mat(neigh, glist=NULL, style="B",
                      zero.policy=TRUE)

# Matrix based on distance: k-nearest neighbors criteria
italy.nb <- knn2nb(knearneigh(coordinates(italy),k=2))

# Inverse distance weight matrix
library(fields) # to calculate the distance between two points
distance <- rdist.earth(coordinates(italy),coordinates(italy))
diag(distance) <- 0
distance.inv <- ifelse(distance!=0, 1/distance, distance)
# Standardized inverse weight matrix
distance.inv <- mat2listw(distance.inv, style = "W")
```

Example in R

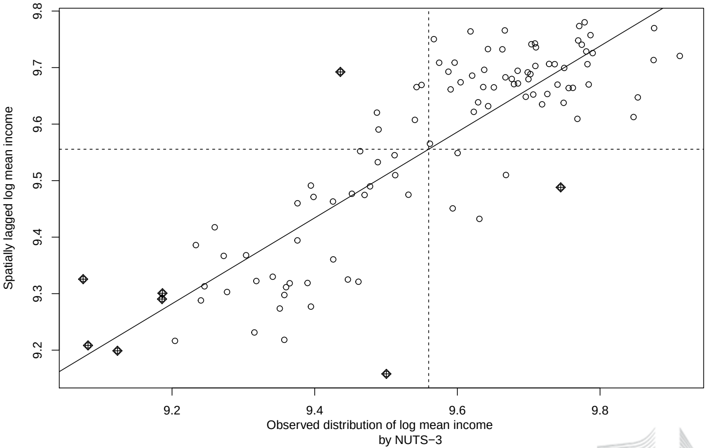
```r
# Moran's I index
Moran.I<-moran.test(x=silcIT3$LogIncome,listw=italy.lw,
                    zero.policy=T)
Moran.I
```

```
##
##  Moran I test under randomisation
##
## data:  silcIT3$LogIncome
## weights: italy.lw
##
## Moran I statistic standard deviate = 11.471, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##       0.760072180      -0.009433962       0.004500357
```

Example in R

```
moran.plot(silcIT3$LogIncome, italy.lw,
           labels=F,
           xlab="Observed distribution of log mean income
                 by NUTS-3",
           ylab="Spatially lagged log mean income")
```

Example in R

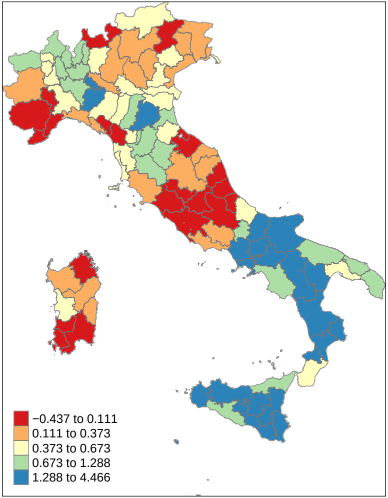


Spatially lagged log mean income
9.2
9.3
9.4
9.5
9.6
9.7
9.8
9.2
9.4
9.6
9.8
Observed distribution of log mean income
by NUTS−3

Example in R

```r
LocalMoran.I <- localmoran(silcIT3$LogIncome, italy.lw,
                           zero.policy=TRUE)
```

Example in R



−0.437 to 0.111
0.111 to 0.373
0.373 to 0.673
0.673 to 1.288
1.288 to 4.466

Course Outline: second part

- Study of areal data
- Spatial econometrics: common models
- Geographically weighted regression
- R example

Spatial regression

- Inadequacies of traditional linear modelling
- Spatial autocorrelations of residuals: inefficiency OLS
- Biased estimator

European Commission

Non-spatial analysis

- spatial data are analyzed using conventional statistical methods (susch as linear regression)
- the geographical coordinates are excluded from the analysis
- the results are independent of the spatial arrangement of the geographical entities

	Variable 1	Variable 2	...	Variable m
unit 1	$attribute_{11}$	$attribute_{21}$	...	$attribute_{1m}$
unit 2	$attribute_{12}$	$attribute_{22}$	...	$attribute_{2m}$
...	...	...	...	...
unit n	$attribute_{1n}$	$attribute_{2n}$	...	$attribute_{nm}$

European Commission

Spatial analysis

- spatial data are analyzed using spatial statistical methods
- the geographical coordinates are included into the the analysis
- the results depend on the spatial arrangement of the geographical entities

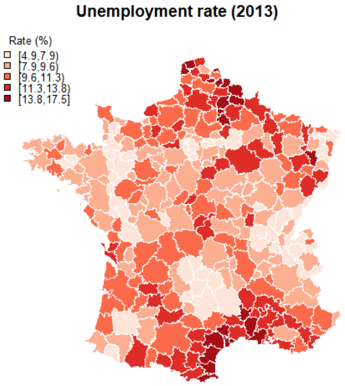
	Lon (X)	Lat (Y)	Variable 1	Variable 2	...	Variable m
unit 1	X_1	Y_1	$attribute_{11}$	$attribute_{21}$	...	$attribute_{1m}$
unit 2	X_2	Y_2	$attribute_{12}$	$attribute_{22}$	...	$attribute_{2m}$
...	...			...	...	...
unit n	X_n	Y_n	$attribute_{1n}$	$attribute_{2n}$	...	$attribute_{nm}$

European Commission

Spatial analysis: example

- Suppose we want to model the unemployment rate in France and we have data for employment zone (example from Insee-Eurostat, 2018)
- After defining a neighbourhood matrix using one of methods previous presented, the data can be mapped and spatial autocorrelation can be evaluated.

Spatial analysis: example



(source: Insee-Eurostat, 2018)

Spatial analysis: example

- The map seems to show a pattern in the variable
- Calculate the p-value of the Moran test we have a near-null p-value indicating that the null hypothesis assuming no spatial autocorrelation should be rejected.
- The descriptive analysis showed that space was not neutral in characterising local unemployment rates.

Spatial analysis: example

- Fitting a standard LM we assume independence between the observations: what happens in area *i* is not in any way related (it is independent) of what happens in area *j*.
- Moran test on residuals of OLS can be help to detect spatially dependence
- In the contest of spatial modelling, some alternatives exist

Spatial models

Classification of Elhorst (2010):

i) endogenous interaction effects, the decision of a spatial unit depends on depend on the decision of its neighbours

ii) exogenous interaction effects, the decision of a spatial unit will depend on the observable characteristics of its neighbours;

iii) correlated effects due to the same unobserved characteristics

$$Y = \rho WY + X\beta + WX\theta + u$$

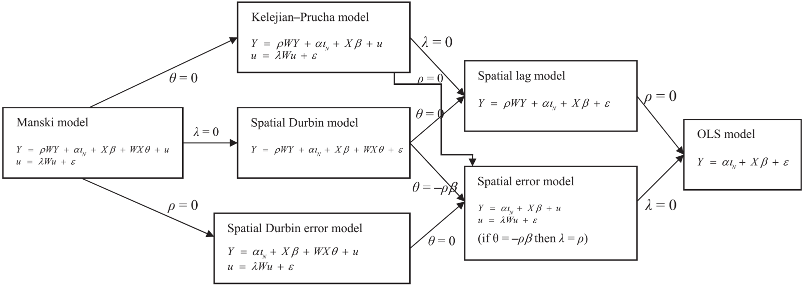
$$u = \lambda Wu + \varepsilon$$

Spatial models

- WY is the endogenous interaction effects among the dependent variables
- WX is the exogenous interaction effects among the explanatory variables
- Wu is the interaction effects among the error terms of the different spatial units
- ρ is the spatial autoregressive coefficient
- λ is the autocorrelation coefficient
- β is the vector of unkown parameters for exogenous explanatory variables
- θ is the vector of unkown parameters for exogenous interaction effects

European Commission

Spatial models



(source: Elhorst, 2010)

Spatial dependence

Starting from the linear regression model (OLS)

$$Y = X\beta + \varepsilon$$

Spatial dependence may be introduced into the model in two major ways

- spatial lag dependence $\rightarrow$ Spatial Lag Models
- spatial error dependence $\rightarrow$ Spatial Error Models

Spatial Lag dependence

- With spatial lag in OLS regression, the assumption of independent observations and the assumption of uncorrelated error terms are violated
- The estimates are biased and inefficient (the size and the sign of the coefficients are not close to the true value and their standard errors are underestimated)
- Spatial lag dependence $\rightarrow$ Spatial Lag Models

Spatial Error dependence

- With spatial error in OLS regression, the assumption of uncorrelated error terms is violated.
- The estimates are inefficient (standard errors are underestimated).
- Spatial error dependence $\rightarrow$ Spatial Error Models

Spatial Lag Models

- It incorporates spatial dependence explicitly by adding a spatially lagged variable y on the right hand side of the regression equation.
- Using this model for modelling the unemployment rate in France, we are saying that the employment rate in a in the neighbouring areas of observation i is an important predictor of employment rate on each individual area i.

European Commission

Spatial Lag Models

The basic spatial lag model is the so-called first order spatial autoregressive (SAR) model:

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \sum_{k=1}^{p} x_{ik} \beta_k + \varepsilon_i$$

- the error term, ε_i, is iid, $\varepsilon_i \sim N(0, \sigma^2 I_n)$
- w_{ij} is the (i, j)th element of the $n \times n$ spatial weights matrix W
- ρ determine the strength of the spatial autoregressive relation between y_i and $\sum_{j=1}^{n} w_{ij} y_j$. If $\rho = 0$ we have a standard regression model

European Commission

Spatial Lag Models: when?

- A spatial lag model and is appropriate when the focus of interest is the assessment of the existence and strength of spatial interaction

European Commission

Spatial Error Models

- The spatial error model treats the spatial autocorrelation as a nuisance that needs to be dealt with.
- For example, adjoining neighbohoods in France may have similar employment rate because because citizens with similar educational level tend to cluster geographically, and level of edcations also predicts employment

Spatial Error Models

In Spatial Errors Models (SEM) the disturbances exhibit spatial dependence.

The most common specification is a spatial autoregressive process of first order:

$$u_i = \lambda \sum_{j=1}^{n} w_{ij} u_j + \varepsilon_i$$

- λ is the autoregressive parameter
- $\varepsilon_i \sim N(0, \sigma^2 I_n)$

European Commission

Spatial Error Models

In matrix notation

$$u = \lambda W u + \varepsilon$$

Assuming $|\lambda| < 1$ and solving for u:

$$u = (I - \lambda W)^{-1} \varepsilon$$

Inserting this in the standard linear regression model we obtain the Spatial Error Model (SEM):

$$Y = X\beta + (I - \lambda W)^{-1} \varepsilon$$

Regression coefficients have the same interpretation of linear models: $\mathrm{E}[Y] = X\beta$

Spatial Error Models: when?

- It is appropriate when the concern is with correcting for the potentially biasing influence of the spatial autocorrelation, due to the use of spatial data (irrespective of whether the model of interest is spatial or not).

Spatial Durbin model

- Spatial dependence in the errors of a standard regression model
- Omitted explanatory variable having non-zero covariance with a variable included in the model

Spatial Durbin Model

The Spatial Durbin Model (SDM) is the SAR model with the insertion of spatially lagged explanatory variables

$$Y = \rho WY + X\beta + WX\theta + \varepsilon$$

The model may be rewritten in reduced form as:

$$Y = (I - \rho W)^{-1}(X\beta + WX\theta + \varepsilon)$$

with $\varepsilon_i \sim N(0, \sigma^2 I_n)$

θ is vector of parameters that measure the marginal impact of the explanatory variables from neighbouring observations on the outcome variable

European Commission

Spatial Durbin model

Let $Z = [XWX]$ and $\delta = [\beta\theta]'$ this model can be written as a SAR model:

$$Y = \rho WY + X\beta + Z\delta + \varepsilon$$

or:

$$Y = (I - \rho W)^{-1} Z\delta (I - \rho W)^{-1} \varepsilon$$

European Commission

Spatial Durbin model

- if $\theta = 0 \rightarrow$ spatial autoregressive model
- estabilish $\theta = -\rho\beta$ (common factor hypothesis) yields the spatial error regression model specification. In this case the SDM: $Y = X\beta + \rho W(Y - X\beta)$

Spatial Durbin model: when?

- Conjunction of two circumstances: spatial dependence in the errors of OLS model and the presence of an omitted covariate that has non-zero covariance with a variable included in the model
- It includes the spatially lagged dependent variable and the explanatory variables, and also the spatially lagged explanatory variables: it is suitable to capture externalities and spillovers arising from different sources
- SDM can mitigate the bias of the OLS estimates when unobservable characteristics, such as the neighborhood prestige, play an influence on the outcome variable

European Commission

Spatial Lag X model

Imposing in the SDM the restriction $\rho = 0$ we obtain the spatially lagged X regression model (SLX)

SLX is model with exogenous interactions. It assumes independence between observations of the outcome variable, but includes features from neighbouring areas in the form of spatially lagged explanatory variables.

Spatial Autoregressive Confused model

- The Kelejian-Prucha model (also referred to Spatial Autoregressive Confused - SAC) represent a mixture of both spatial dependence in the dependent variable reflected in WY and spatial dependence in the disturbances represented by Wu.

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

$$u = \lambda Wu + \varepsilon$$

- β are biased and not convergent when the real model includes exogenous interactions WX
- Interaction effect among error terms: λWu

European Commission

Estimation of Spatial Regression Models

- Ordinary least-squares cannot be used to produce consistent estimates for spatial regression models.
- Maximum likelihood (ML) approach is typically used

Estimation of Spatial Regression Models

The log-likelihood function for the SDM (and SAR) models takes the form

$$\ln L(\rho, \delta, \sigma^2) = -\frac{n}{2}\ln(\pi\sigma^2) + \ln|I - \rho W| - \frac{ee^{'}}{2\sigma^2}$$

- $Z = X$ for SAR model and $Z = XWX$ for SDM
- $e = Y - \rho WY - Z\delta$
- The parameters with respect to which this likelihood has to be maximised are ρ, δ and σ^2

European Commission

Estimation of Spatial Regression Models

A convenient approach is to use the scalar concentrated log-likelihood function value:

$$\ln L_{con}(\rho) = \kappa + \ln|I - \rho W| - \frac{n}{2}\ln[(e_O - e_{\rho L})'(e_O - e_{\rho L})]$$

- $e_0 = Y - Z\delta_O$ and $e_0 = WY - Z\delta_L$, with $\delta_O = (Z'Z)^{-1}Z'Y$ and $\delta_L = (Z'Z)^{-1}Z'WY$
- This approach simplifies the optimising problem by reducing a multivariate optimisation problem to a univariate one.
- Maximising the concentrated log-likelihood function with respect to ρ yields $\rho^{\star}$ that is equal to the maximum likelihood estimate

European Commission

Estimation of Spatial Regression Models

The log-likelihood function for the SEM models takes the form

$$\ln L(\lambda, \beta, \sigma^2) = -\frac{n}{2}\ln(\pi\sigma^2) + \ln|I - \rho W| - \frac{ee^{'}}{2\sigma^2}$$

- $e = (Y - X\beta)(I - \lambda W)$

Estimation of Spatial Regression Models

Estimates of model parameters in SEM can be obtained as:

$$\hat{\beta}_{ML} = [X^{'}(I-\lambda W)^{'}(I-\lambda W)X]^{-1}X^{'}(I-\lambda W)^{'}(I-\lambda W)Y$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{n}[(y-X\hat{\beta}_{ML})-\lambda W(y-X\hat{\beta}_{ML})^{'}][(y-X\hat{\beta}_{ML})-\lambda W(y-X\hat{\beta}_{ML})^{'}]$$

A consistent estimate for λ cannot be obtained from a simple auxiliary regression, but the first order conditions must be solved explicitly by numerical means

European Commission

Interpretation of the parameter estimates

- When $\rho \neq 0$ the interpretation of the parameter in the spatial models is different from a conventional linear regression: the spatial interactions, the variation of an explanatory variable for a given area directly affects its result and indirectly affects the results of all other areas.
- However, the conventional interpretation of linear models is still valid if only the spatial autocorrelation of errors is taken into account (SEM model).
- The reference book for the intepretation model parameters in spatial regression models is LeSage & Pace (2009)

Interpretation of the parameter estimates in OLS

- In OLS a regression coefficient for variable X_k indicates how much Y goes up or down for every one unit increase in X_k when all other variables in the model are fixed. In our example, for the nonspatial model this effect is the same for every county in our dataset

Interpretation of the parameter estimates in OLS

- In standard linear regression model $y = \sum_{k=1}^{p} x_{ik}\beta_k$ the partial derivatives of y_i with respect to x_{ik} have a simple form: $\frac{\partial y_i}{\partial x_{ik}} = \beta_k$ for all i, k and $\frac{\partial y_j}{\partial x_{ik}} = 0$ for $j \neq i$ and all variables x_k
- Changes in observation i on the kth explanatory variable, x_{ik},only influence observation y_i

Interpretation of the parameter estimates

- In the spatial lag model there are two components to how X affect Y.
- X affects Y within each area directly...but we are also including the spatial lag (the measure of Y in the surrounding counties B, C, and D)

European Commission

Interpretation of the parameter estimates

- The spatial lag model includes not only the effect of X in the area A in the level of Y in area A.
- Including the spatial lag (a measure of Y in county B, C and D) we are incorporating as well the effects that X has on Y in areas B, C, and D.
- The effect of a covariate is the sum of two particular effects: a direct, local effect of the covariate in that unit, and an indirect, spillover effect due to the spatial lag.

Interpretation of the parameter estimates

Considering the SAR model:

$$Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon$$

It can be rewritten as:

$$Y = \sum_{k=1}^{p} S_k(W) X_K + V(W)\iota_n \beta_0 + V(W)\varepsilon$$

- $S_k(W) = V(W)(I\beta_k)$
- $V(W) = (I - \rho W)$
- ι_n is a vector of ones and β_0 is the constant term
- For SDM $S_k(W) = V(W)(I\beta_k + W\theta_k)$

European Commission

Interpretation of the parameter estimates

$$S_k(W) = \begin{pmatrix} S_k(W)_{11} & S_k(W)_{12} & \cdots & S_k(W)_{1n} \\ S_k(W)_{21} & S_k(W)_{22} & \cdots & S_k(W)_{2n} \\ \vdots & \vdots & \ddots & \\ S_k(W)_{n1} & S_k(W)_{n2} & \cdots & S_k(W)_{nn} \end{pmatrix}$$

European Commission

Interpretation of the parameter estimates

Considering the determination of a single dependent variable observation y_i:

$$y_i = \sum_{k=1}^{p} [S_k(W)_{i1} X_{1k} + \cdots + S_k(W)_{in} X_{nk}] + V(W)_i \iota_n \beta_0 + V(W)_i \varepsilon$$

- $S_k(W)_{ij}$ is the (i, j)th element of the matrix $S_k(W)$
- $V(W)_i$ is the ith row of $V(W)$

It follows ($j \neq i$):

$$\frac{\partial y_i}{\partial x_{jk}} = S_k(W)_{ij}$$

and

$$\frac{\partial y_i}{\partial x_{ik}} = S_k(W)_{ii}$$

European Commission

Interpretation of the parameter estimates

- Average direct effect ($n^{-1}\text{tr}(S_k)$), the impact of changes in the ith observation of x_k on y_k
- Average total effect: the average of n effects across a zone i due to the modification of a unit of variable X_k in all zones ($n^{-1}\sum_i(\sum_j S_k(W)_{ij})$), or the average of the n effects from modifying a unit of variable X_k in a zone i across all zones ($n^{-1}\sum_j S_k(W)_{ji}$)
- Average indirect effect is the difference between the average total effect and the average direct effect.

European Commission

An applied example

- Relationship between regional total factor productivity (y) as the dependent variable y and regional knowledge stocks (x) as the single explanatory variable (Example from LeSage & Pace, 2009)
- SDM:

$$y = 0.568 + 0.647Wy + 0.111x - 0.016Wx$$

An applied example

	Mean effects	SD
Direct effet	0.120	0.024
Indirect effet	0.172	0.081
Total effet	0.292	0.111

The direct impacts are close to the SDM model coefficient

The difference between the coefficient estimate of 0.111 and thedirect effect estimate of 0.120 equal to 0.009 represents feedback effects that arise as a result of impacts passing through neighboring regions and back to the region itself.

European Commission

An applied example

- A significant indirect impact (spillover) arising from changes in the variable x.
- The total impact estimates can be viewd as elasticities (the model is specified using logged levels y and x). The total effect is 0.292: a 10 percent increase in regional knowledge would result in a 2.9 percent increase in total factor productivity.

European Commission

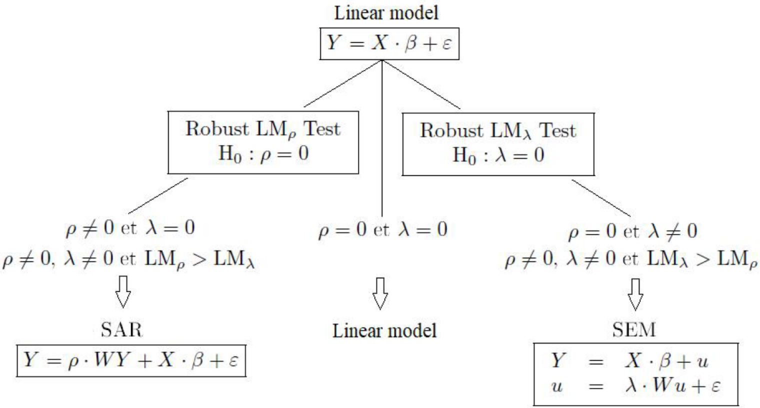
Model selection

An issue that arises in applied practice is the need to select among the different models

- Bottom-up approach
- Top-down approach
- Combined approach

Comparison of different model specifications using likelihood-based testing will be shown in the R example

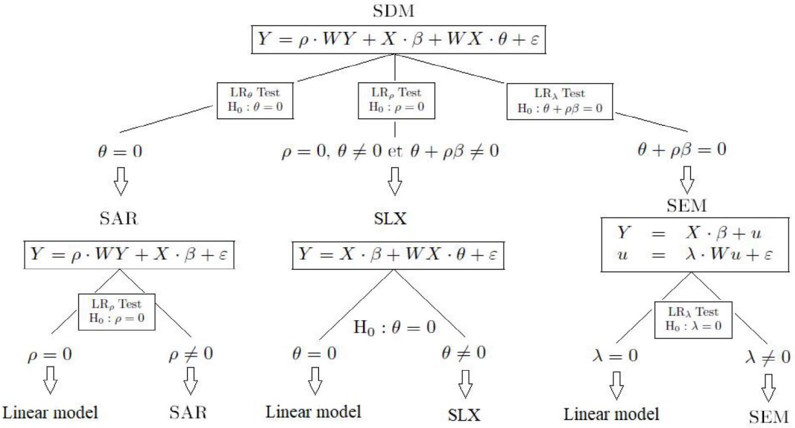
Model selection: Bottom-up approach



(source: Insee-Eurostat, 2018)

European Commission

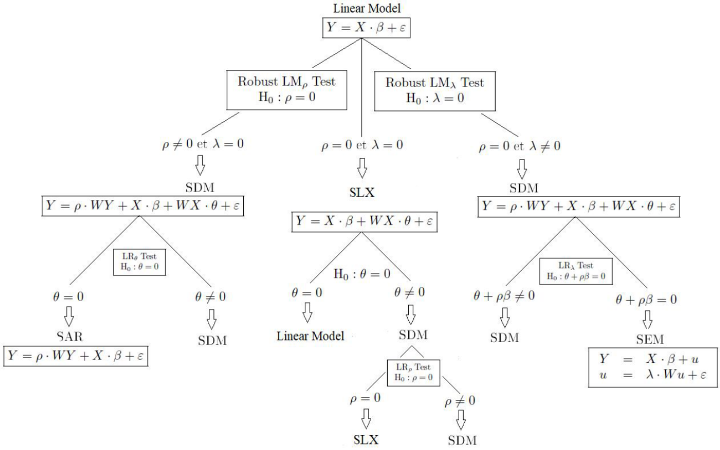
Model selection: Top-down approach



(source: Insee-Eurostat, 2018)

European Commission

Model selection: Combined approach



(source: Insee-Eurostat, 2018)

European Commission

What did you learn?

- Inadequacies of traditional linear modelling in case of spatial dependency between nearby observations
- Spatially endogenous interactions: Spatial Lag Model
- Spatial interactions in the error: Spatial Error Model
- Both spatially endogenous interactions and spatial interactions in the error term as well as exogenous interactions: Spatial Durbin Model

European Commission

Example in R

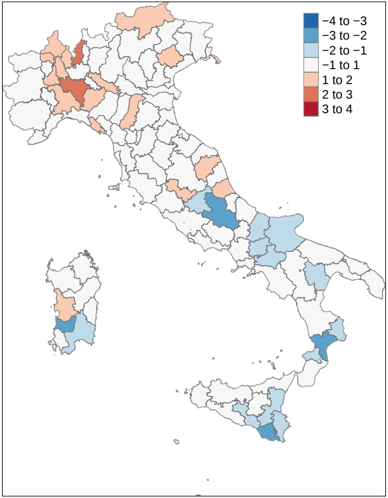
- Example of fitting spatial models with R (RStudio) by using EU data

Example in R

```r
 # Defining the model
model <- LogIncome ~gender+ employment.rate+isced5+owner

# OLS model
mod.ols <- lm(model, data=Silc.Spat)
summary(mod.ols)
# extract residuals to evaluate spatial autocorrelation
res.ols <- residuals(mod.ols)
```

Example in R



Example in R

```
# Moran test to residuals
lm.morantest(mod.ols,italy.lw, zero.policy = T,
             alternative="two.sided")
```

```
##
##  Global Moran I for regression residuals
##
## data:
## model: lm(formula = model, data = Silc.Spat)
## weights: italy.lw
##
## Moran I statistic standard deviate = 5.7922, p-value = 6.947e-09
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation         Variance
##      0.365032457     -0.016449168      0.004337689
```

Example in R

```
# Lagrange Multiplier tests
lm.LMtests(mod.ols,italy.lw,test=c("LMerr","LMlag"))
```

```
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = model, data = Silc.Spat)
## weights: italy.lw
##
## LMerr = 28.561, df = 1, p-value = 9.077e-08
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = model, data = Silc.Spat)
## weights: italy.lw
##
## LMlag = 71.853, df = 1, p-value < 2.2e-16
```

European Commission

Example in R

```r
# Lagrange Multiplier tests
lm.LMtests(mod.ols,italy.lw,test=c("RLMerr","RLMlag"))
```

```
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = model, data = Silc.Spat)
## weights: italy.lw
##
## RLMerr = 1.073, df = 1, p-value = 0.3003
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = model, data = Silc.Spat)
## weights: italy.lw
##
## RLMlag = 44.364, df = 1, p-value = 2.726e-11
```

Example in R

```r
# SAR Model
library(spatialreg)
mod.sar<-lagsarlm(model, data=Silc.Spat, italy.lw)
```

The the spatial autoregressive parameter (ρ = 0.65) is highly significant, as indicated by the p-value of almost 0 of both the asymptotic t-test and the LR test on this parameter.

However, the LM test for residual autocorrelation is significant at 1% (test value: 7.6, p-value: 0.006)

Example in R

```r
# SDM Model
mod.sdm<-lagsarlm(model, data=Silc.Spat, italy.lw,
                  type="mixed")
# SEM model
mod.sem<-errorsarlm(model, data=Silc.Spat, italy.lw)
# Common factor hypothesis test
FC.test<-LR.sarlm(mod.sdm,mod.sem); print(FC.test)
```

```
##
##  Likelihood ratio for spatial linear models
##
## data:
## Likelihood ratio = 27.165, df = 4, p-value = 1.841e-05
## sample estimates:
## Log likelihood of mod.sdm Log likelihood of mod.sem
##                  113.7521                  100.1695
```

Example in R

	Direct		Indirect		Total	
gender	0.163	***	0.459	***	0.533	***
employment.rate	0.125	***	0.399	***	0.443	***
isced5	0.263	***	0.812		0.944	**
owner	0.115	***	0.357		0.396	*

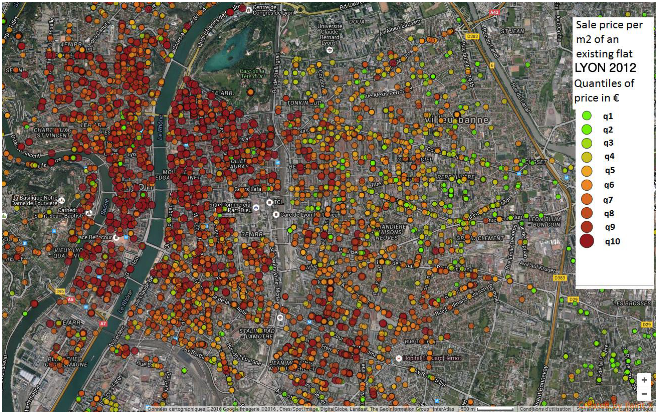
Geographically Weighted Regression: when?

- potential spatial heterogeneity in parameter estimates
- GWR permits the parameter estimates to vary locally
- exploratory technique to detect where non-stationarity is taking place

Geographically Weighted Regression: example

- Hedonic model to study real estate prices in Lyon (Example from Insee-Eurostat, 2018).
- The hedonic model is aimed at isolating the effect of localisation on prices.

Geographically Weighted Regression: example



(source: Insee-Eurostat, 2018)

Geographically Weighted Regression: example

- Positive spatial correlation in the residuals of OLS.
- The hypothesis of spatial stationarity of the relationship between price and characteristic of the property is not valid
- Existence of spatial heterogeneity
- GeographicallyWeighted Regression allows study a model that varies spatially in a continuous way.

European Commission

Geographically Weighted Regression

$$y_i = X\beta_i + \varepsilon_i$$

i is the location at which the local parameters are to be estimated.

$$\beta_i = (X^T W_i X)^{-1} X^T W_i y$$

In order to give a weight to observations decreasing with their distance to the point of interest, the estimation is performed using weighted least squares, the weighting being governed by weight matrix W.

European Commission

Geographically Weighted Regression

- $\mathbf{W}_i$ contains the weight of each observation according to its distance to the point i.
- Observations close to point i have more influence over the estimated parameters at place i than more remote observations.
- The weight of observations decreases with the distance to the point i.

European Commission

Geographically Weighted Regression

Kernel function

- the shape of the kernel
- fixed kernel versus adaptive kernel
- bandwidth size.

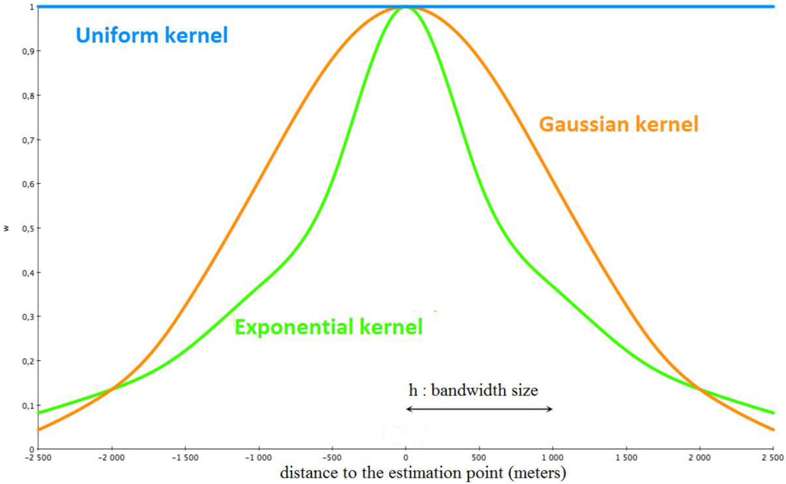
European Commission

Continuous Kernel

The shape of continuous kernel:

- Uniform kernel: $w(d_{ij}) = 1$
- Gaussian kernel: $w(d_{ij}) = \exp(-\frac{1}{2}(\frac{d_{ij}}{h})^2)$
- Exponential kernel: $w(d_{ij}) = \exp(-\frac{1}{2}(\frac{|d_{ij}|}{h}))$

Continuous Kernel



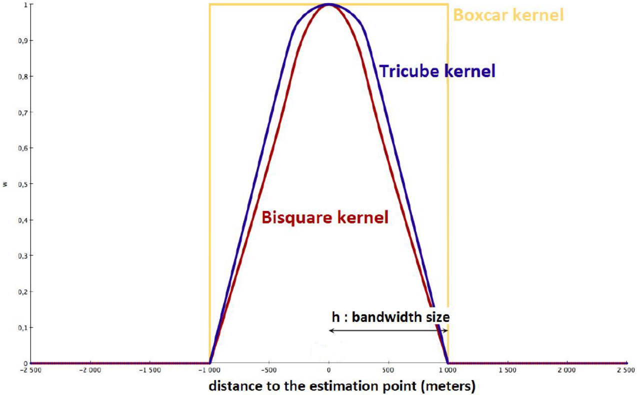
(source: Insee-Eurostat, 2018)

Kernel with compact support

- Box-Car kernel: $w(d_{ij}) = 1$ if $|d_{ij}| < h$, 0 otherwise
- Bi-square kernel: $w(d_{ij}) = (1 - (\frac{d_{ij}}{h})^2)^2$ if $|d_{ij}| < h$, 0 otherwise
- Tri-cube kernel: $w(d_{ij}) = (1 - (\frac{d_{ij}}{h})^3)^3$ if $|d_{ij}| < h$, 0 otherwise

European Commission

Kernel with compact support



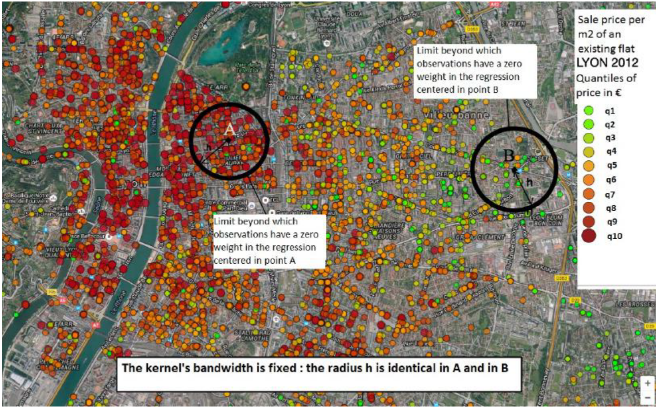
(source: Insee-Eurostat, 2018)

Fixed kernel versus adaptive kernel

- Fixed kernel. The extent of the kernel is determined by the distance to the point of interest. The kernel is identical at any point in space.
- Adaptive kernel. The extent of the kernel is determined by the number of neighbours of the point of interest. The lower the density of the observations, the smaller the kernel

European Commission

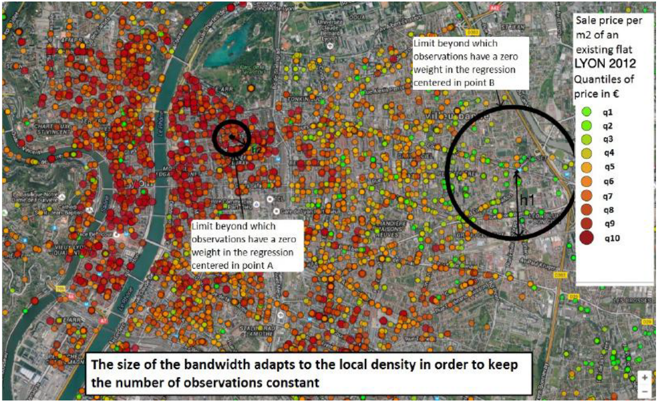
Fixed kernel: example



(source: Insee-Eurostat, 2018)

European Commission

Fixed kernel: example



(source: Insee-Eurostat, 2018)

European Commission

Bandwidth size

- The bandwidth is a distance beyond which the weight of the observations is 0.
- This is the most important parameter because, the choice of the bandwidth h highly influences the results.
  - With larger bandwidth, the number of observations to which the kernel gives a non-zero weight will be higher.
  - When the bandwidth tends towards infinity, the results of the local regression will be similar to those of OLS regression.
- Choose the most suitable bandwidth via cross-validation (more used) or minimizing the modified AIC.

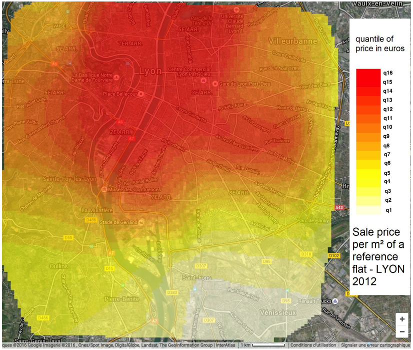
Bandwidth size

$$CV = \sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(h)]^{''}$$

- $\hat{y}_{\neq i}(h)$ is the value of y at point i predicted when calibrating the model with all the observations except y_i.
- Bandwidth h that minimises CV maximises the model's predictive power.

Geographically Weighted Regression: example

Estimate of real estate prices on a square grid of 100m*100m



(source: Insee-Eurostat, 2018)

European Commission

Diagnostic Tools in GWR

- Test the nonstationarity of the coefficients

$$\begin{cases} H_0 & \forall k, \beta_k(u_1, v_1) = \beta_k(u_2, v_2) = \ldots = \beta_k(u_n, v_n) \\ H_1 & \exists k, all \beta_k(u_i, v_i) \text{ are not equal} \end{cases}$$

European Commission

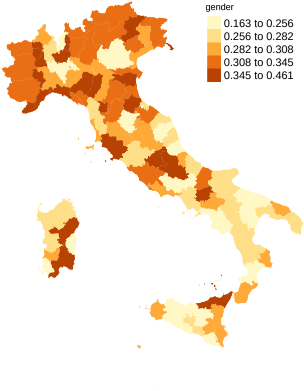
Diagnostic Tools in GWR

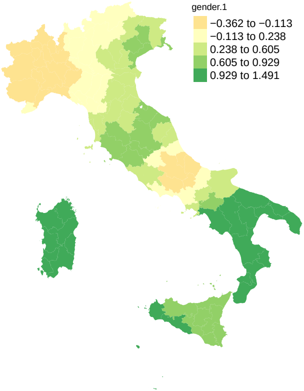
- Multicolinearity
  - scatter plots of regression coefficients for pairs of regression terms
  - Local variance inflation factors (VIFs)
- One method for reducing the colinearity problems is ridge regression

Example in R

```
library(spgwr)
# Get the optimal bandwidth.
GWRbandwidth <- gwr.sel(model, data=Silc.Spat, adapt=T,
                        verbose = F)
# Estimate GWR
mod.gwr <- gwr(model,
               data = Silc.Spat,
               adapt=GWRbandwidth,
               hatmatrix=TRUE,
               se.fit=TRUE)
```

Example in R


gender
0.163 to 0.256
0.256 to 0.282
0.282 to 0.308
0.308 to 0.345
0.345 to 0.461


gender.1
−0.362 to −0.113
−0.113 to 0.238
0.238 to 0.605
0.605 to 0.929
0.929 to 1.491

Spatial autologistic model

- Spatial data with binary response.
- The occurrence of an event $y = 1$ in neighboring units conditions the likelihood that unit i will itself experience the event.
- Spatial autologistic model.

European Commission

Spatial autologistic model

- The autologistic model states the conditional probability p_i that $y_i = 1$, given values y_j at units ($j \neq i$):

$$p_i = P(y_i = 1 | Wy_i) = \frac{exp(\alpha + X\beta + \lambda Wy_i)}{1 + exp(\alpha + X\beta + \lambda Wy_i)}$$

- β is the vector of parameters for exogenous variables,
- λ is a scalar parameter for the spatial lag of y
- W is a connectivity matrix.

Other spatial analysis

- In this course, we consider the analysis of areal data
- Geostatistics is a branch af spatial statistics and includes a set of statistical methods for analysing continuous data
- The variogram is the central tool of geostatistics: it allows to assess whether the data are spatially correlated and to what extent.
- With a suitable model for it is possible to combine it with the data to predict by kriging, which in its simpler forms is one of weighted averaging.
- More on geostatisics in Oliver, M. A., & Webster, R. (2015). Basic steps in geostatistics: the variogram and kriging. Springer

European Commission

Bibliography

Main Textbook

Insee-Eurostat (2018). *Handbook of Spatial Analysis. Theory and practical application with R*. N. 131 October 2018. Insee-Eurostat

Other readings

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer.

Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R*. Springer.

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. *Spatial economic analysis, 5(1)*, 9-28.

Fischer, M.M. & Wang, J. (2011). *Spatial data analysis: models, methods and techniques*. Springer Science & Business Media

LeSage, J. & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman & Hall/CRC.

European Commission