Spatial and Regional Economic Analysis Mini-Course

Author: Vicente Rios

- **Introduction to Spatial Economics**
  - **Spatial heterogeneity (growth example)**
  - **Spatial dependence (fiscal policy example)**
  - **Spatial autocorrelation**
- **Spatial Weight Matrices**
- **Measuring spatial autocorrelation**

Why do we need spatial econometrics?

- Motivations for going spatial
  - **Independence assumption not valid**

    The attributes of observation *i* may influence the attributes of *j*.

    eg: growth in *i* affects growth in *j*

  - **Spatial heterogeneity**

    The magnitude and direction of an effect may vary across space.

    eg: government spending in *i* does not have the same stimulus than in *j*

  - **Accuracy!**

Spatial Heterogeneity

- **Spatial econometrics deals with spatial effects**

  (I) Spatial heterogeneity

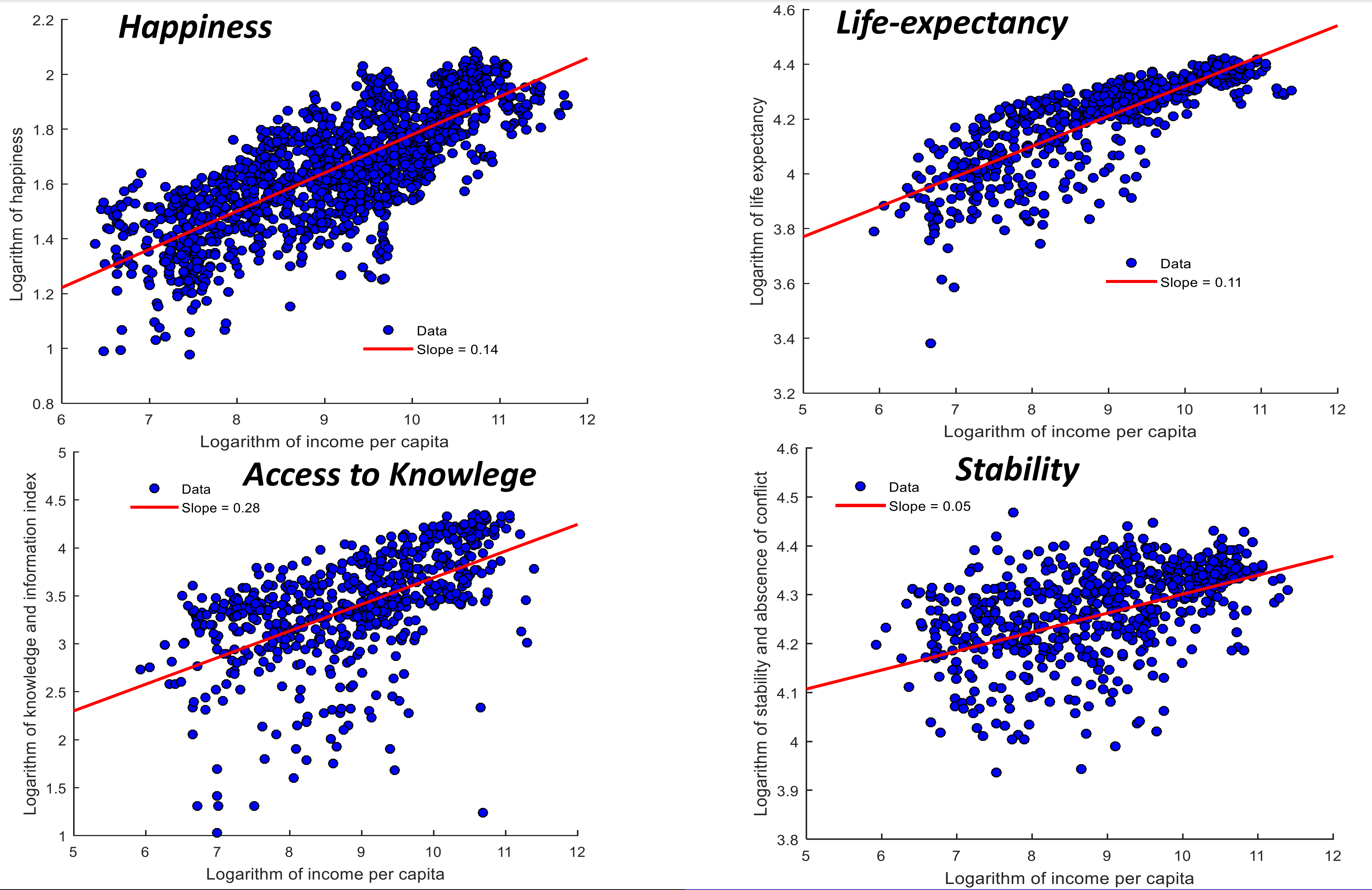
Definition:

Spatial heterogeneity relates to a differentiation of the effects of space over the sample units. Formally, for spatial unit i:

$$y_i = f(x_i)_i + u_i \rightarrow y_i = \beta_i x_i + u_i$$

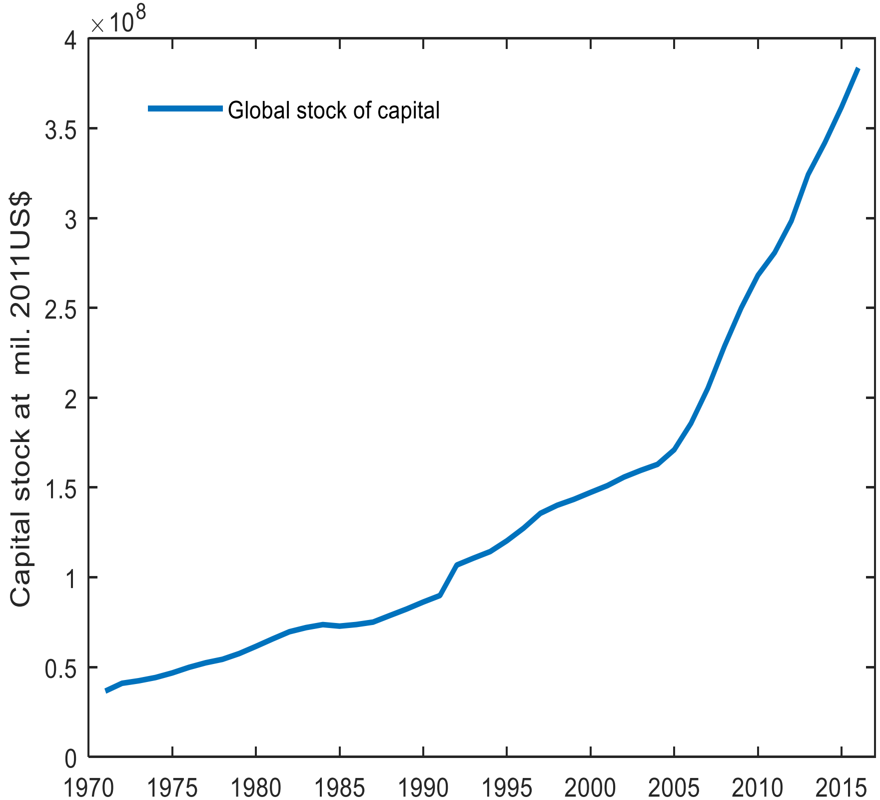
Lack of stability over the geographical space

Spatial Heterogeneity and Regional growth in Europe

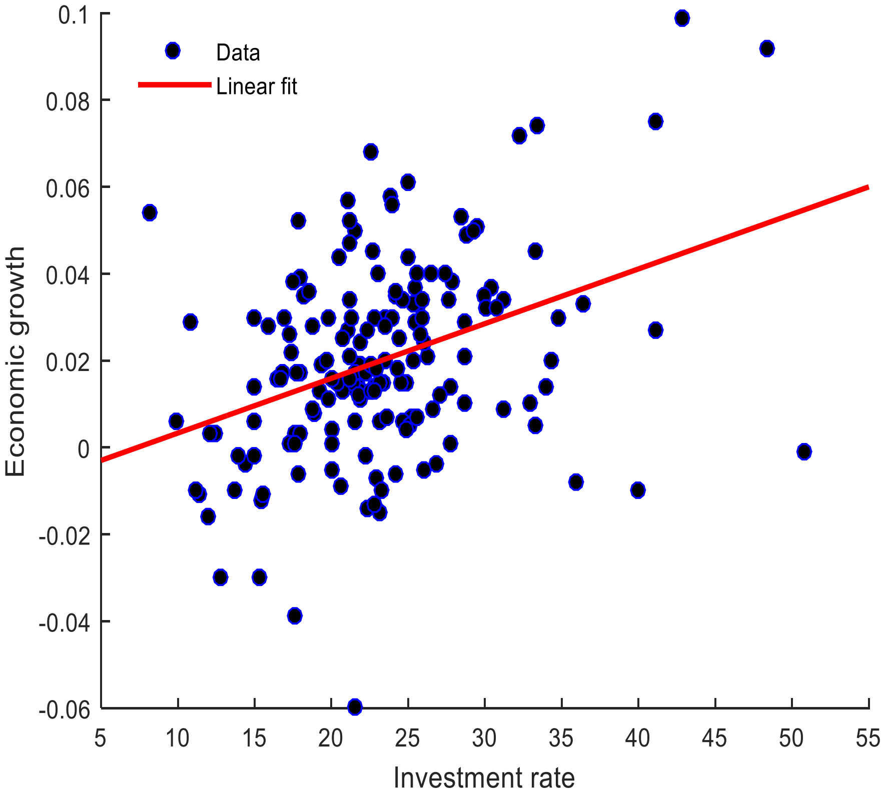


(1) Physical capital and investment

Global capital stock



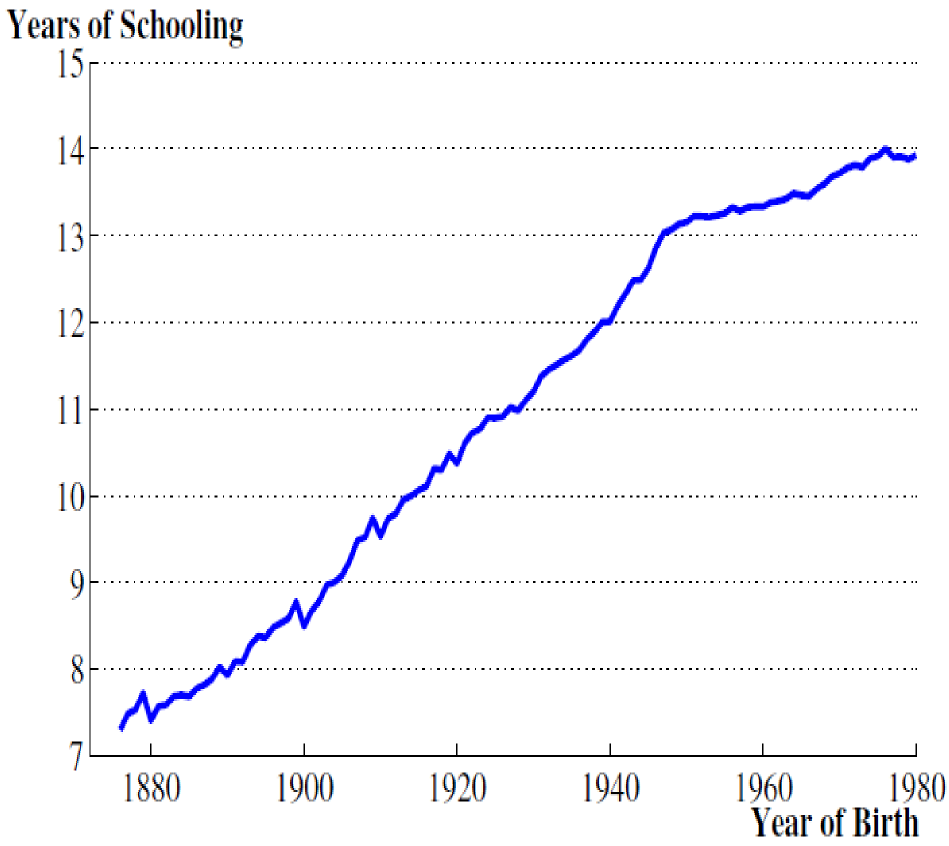
Investment rates and growth



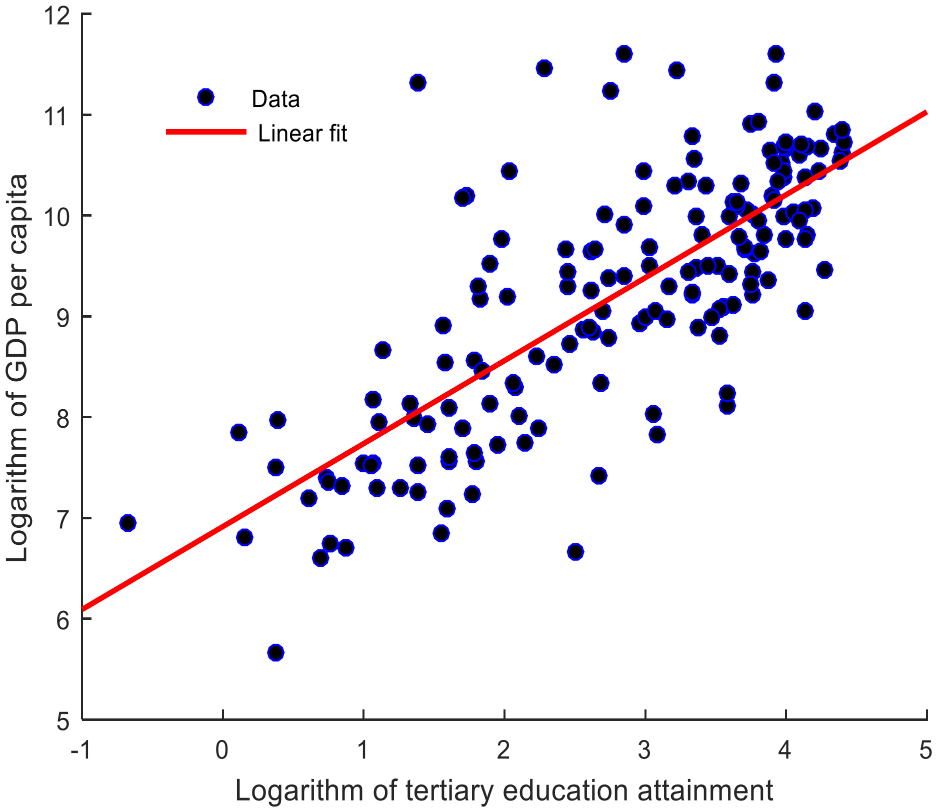
Spatial Heterogeneity and Regional growth in Europe

- (2) Human capital and investment in education

Years of schooling

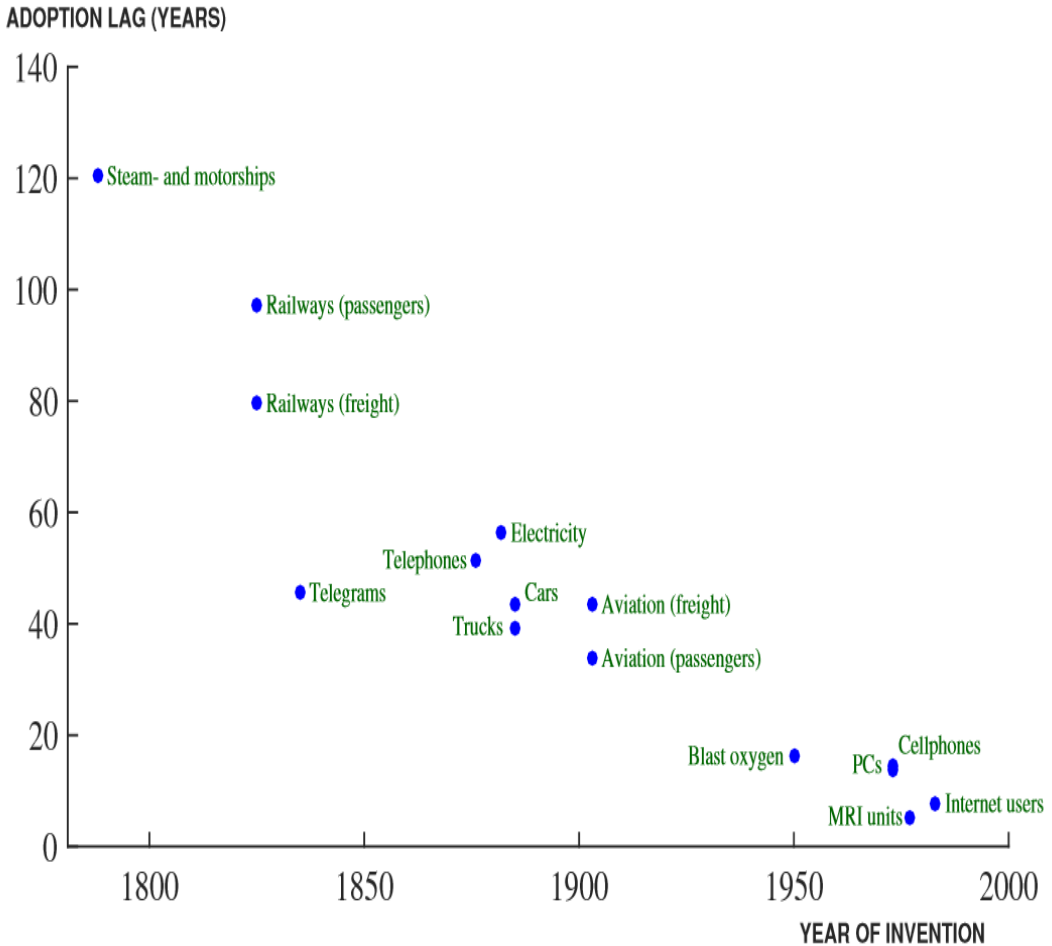


Education and income



- (3) Technological progress

Diffusion and Acceleration of technological progress



Spatial Heterogeneity and Regional growth in Europe

Production: $Y_t = K_t^{\alpha} H_t^{\beta} (A_t L_t)^{1-\alpha-\beta}$

where:

Technology: $A_t = A_0(1+g)^t$

Labor (population): $L_t = L_0(1+n)^t$

Exogenous forces

Physical capital: $K_{t+1} = K_t + I_t^K - \delta_k K_t$

Human capital: $H_{t+1} = H_t + I_t^H - \delta_H H_t$

Endogenous processes

Physical Investment: $I_t^K \equiv s_k\, Y_t$

Human capital Investment: $I_t^H \equiv s_H\, Y_t$

Consumption: $C_t \equiv 1-(s_k + s_H)\, Y_t$

Accounting identities

Parameter values: $\alpha, \beta, s_H, \delta_H, \delta_k \in (0,1)$ and $g, n > 0$

Initial values: $K_0, H_0, L_0, A_0 > 0$

Spatial Heterogeneity and Regional growth in Europe

If Europe has $j=1, \ldots, N$ economies:

For region j, the production function can be writen as:

$$Y_{jt} = K_{jt}^{\alpha} \left(A_{jt} L_{jt}\right)^{1-\alpha-\beta} H_{jt}^{\beta}$$

After some algebraic manipultations and using the equilibrium values of K and H the evolution of per capita income of region j can be expressed as:

$$y_{jt} = A_{jt} \left(\left(\frac{s_{k,j}}{g_j + \delta_k + n_j} \right)^{\frac{\alpha}{1-\alpha-\beta}} \left(\frac{s_{h,j}}{g_j + \delta_h + n_j} \right)^{\frac{\beta}{1-\alpha-\beta}} \right)$$

Taking logarithms:

$$\ln y_{jt} = \ln A_{j0} + gt + \frac{\alpha}{1-\alpha-\beta} \ln \left(\frac{s_{k,j}}{g_j+\delta_k+n_j} \right) + \frac{\beta}{1-\alpha-\beta} \ln \left(\frac{s_{h,j}}{g_j+\delta_h+n_j} \right)$$

which is equivalent to:

$$\ln y_{jt} = \ln A_{j0} + gt + \frac{\alpha}{1-\alpha-\beta} \ln(s_{k,j}) + \frac{\beta}{1-\alpha-\beta} \ln(s_{h,j}) - \frac{\alpha+\beta}{1-\alpha-\beta} \ln(n_j + 0.05)$$

once we asume $\delta_k = \delta_H$ and that $g + \delta = 0.05$.

We can test the if the expected effects hold when looking at the data!!

Spatial Heterogeneity and Regional growth in Europe

Estimates of the Augmented Solow Model	MRW	Updated data	
	1985	1985	2000
$\ln(s_k)$	.69	.65	.96
	(.13)	(.11)	(.13)
$\ln(n+g+\delta)$	-1.73	-1.02	-1.06
	(.41)	(.45)	(.33)
$\ln(s_h)$	.66	.47	.70
	(.07)	(.07)	(.13)
Adj R^2	.78	.65	.60
Implied α	.30	.31	.36
Implied β	.28	.22	.26
No. of observations	98	98	107

→ Positive effect of investment/savings rate

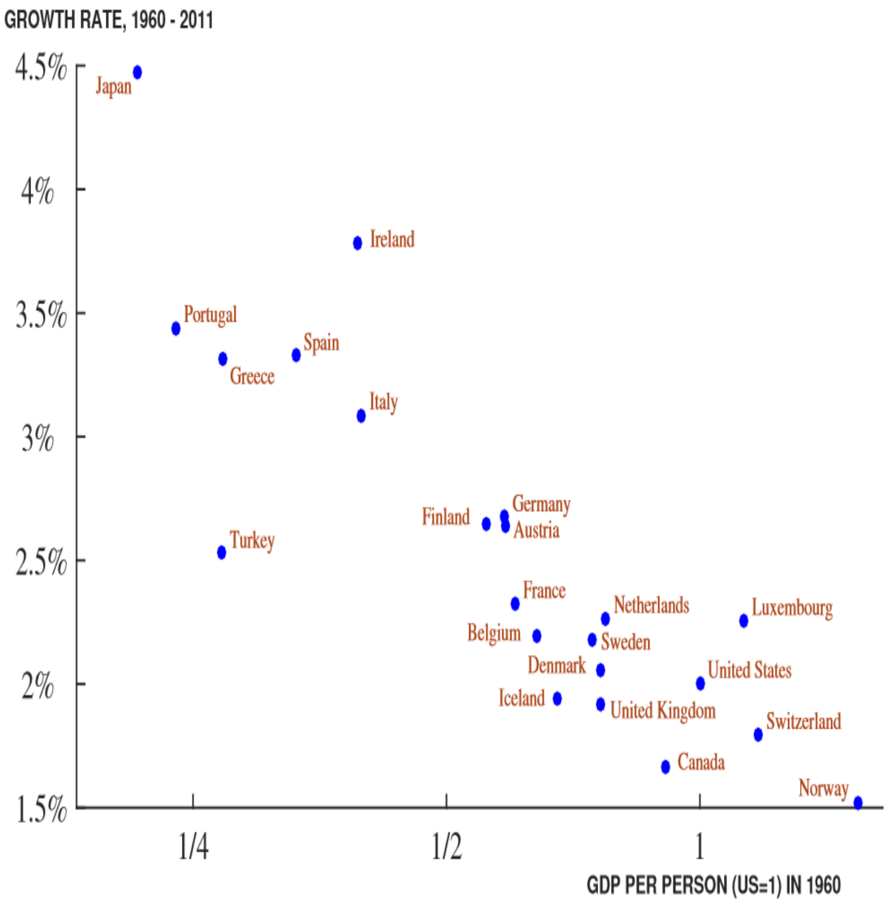
→ Negative effect of population growth and effective depreciation

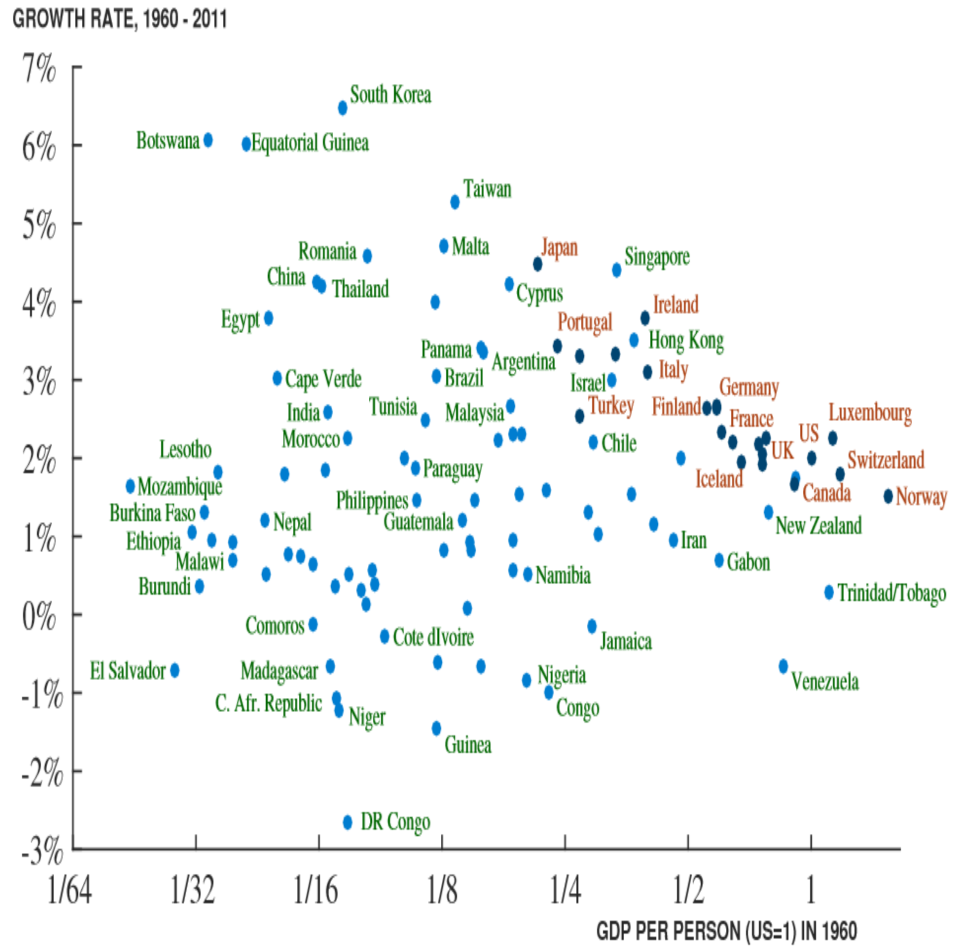
→ Positive effect of human capital investment

→ We can explain 60-75% of income per capita differences based on these three drivers

Spatial Heterogeneity and Regional growth in Europe

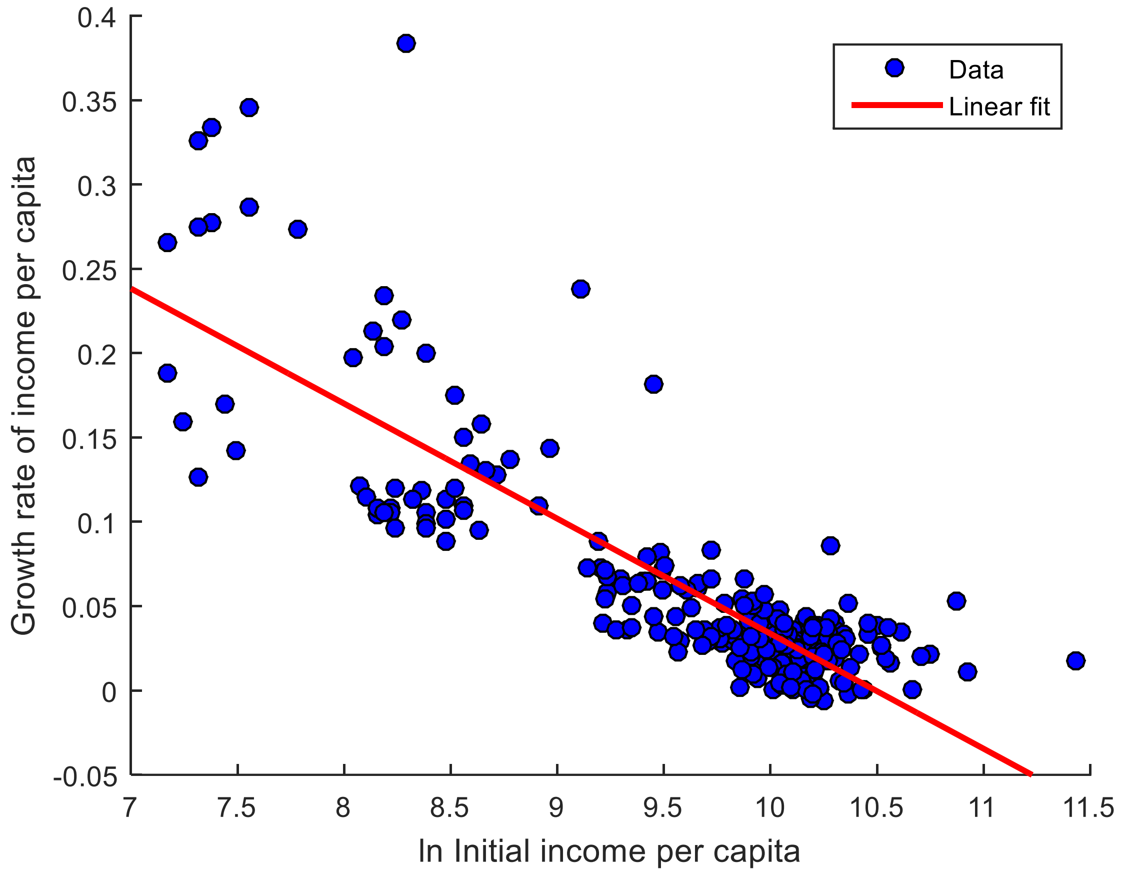
- Key prediction of the model → convergence. This is not in line with global data





Spatial Heterogeneity and Regional growth in Europe

- but the prediction of **regional convergence holds in European regions** (scatter is for period 2000-2008)



Spatial Heterogeneity and Regional growth in Europe

- Application of neoclassical economic framework to guide development policies too simplistic and can be reduced to:

"if you want more economic growth per capita either (i) invest more on physical capital, human capital or (ii) reduce your population growth"

Corolary: to foster investment/savings create good institutional framework, etc…

But recently: **place-based philosophy. We should question the homogeneity** of the effects of the drivers of growth!

Example: What if you do NOT need to invest more in a given capital because of increasing investment there has no returns? That is to say, what if running:

$$gy_i = \beta_{0i} + \beta_{1i} lnY_i + \beta_{2i} I_i + \beta_{3i} H_i + \beta_{3i} n_i + \varepsilon_i$$

tells you $\beta_{2i} > 0$ for some regions and $\beta_{2i} < 0$ for some others?

Should you always recommend increasing physical capital investment?

Spatial Heterogeneity : GWR

A key assumption on spatial modelling is whether or not the parameters are homogenous o heterogeneous (across space, time, etc)

If we are interested in accounting for potential **spatial heterogeneity in parameters** we can use **GWR** (Fotheringham et al., 2002) as this modeling technique allows local variation int he parameters.

GWR has been **used** primarily **for exploratory data analysis**, rather than hypothesis testing

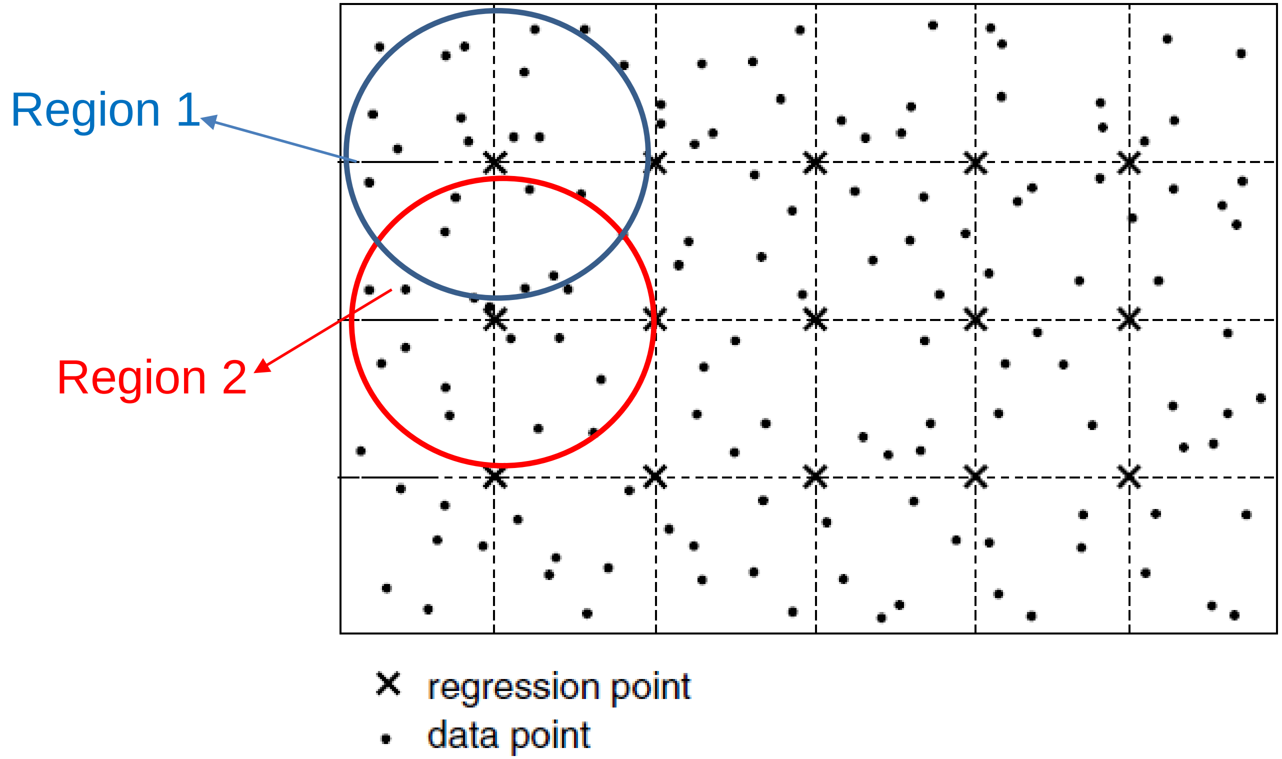
The **basic form of the GW** regression model if we have *m* explanatory variables:

$$y_i = \beta_{0i}(u_i, v_i) + \sum_{k=1}^{m} \beta_{ik}(u_i, v_i) x_{ik} + e_i$$

Where y_i is the dependent variable at location i, x_{ik} is the value of the regressor *k* at location *i*, β_{i0} is the intercept parameter at location *i*, β_{ik} is the local regression coefficient for the *kth* independent variable at location *i* and e_i is the random error term at location *i*. (u_i,, v_i) denotes the coordinates of the *i-th* point in space

Spatial Heterogeneity : GWR

- You can think about **GWR as a *"spatially moving window regression"***
- **A region can be defined as the four cells around each regression point**
- The **regression model is then calibrated on all data that lie within the region** described around a regression point and the process is repeated for all regression points



Spatial Heterogeneity : GWR

- In the spatial **moving window example** presented above:
  - **a region was described around a regression point and** all the data **points within this region** or window were then **used to calibrate** a model.
  - This process was repeated for all regression points

- GWR works in the same way except that:
  - **Each data point is weighted by its distance from the regression point**
  - Hence, **data points closer to the regression point are weighted more** heavily in the local regression than are data points farther away
  - For a given regression point, **the weight of a data point is at a maximum when it shares the same location as the regression point**. This weight decreases continuously as the distance between the two points increases (sometimes the decrease in influence can be abrupt)

Spatial Heterogeneity : GWR

- As data are geographically weighted, **nearer observations have more influence in estimating the local set of regression coefficients** than observations farther away.

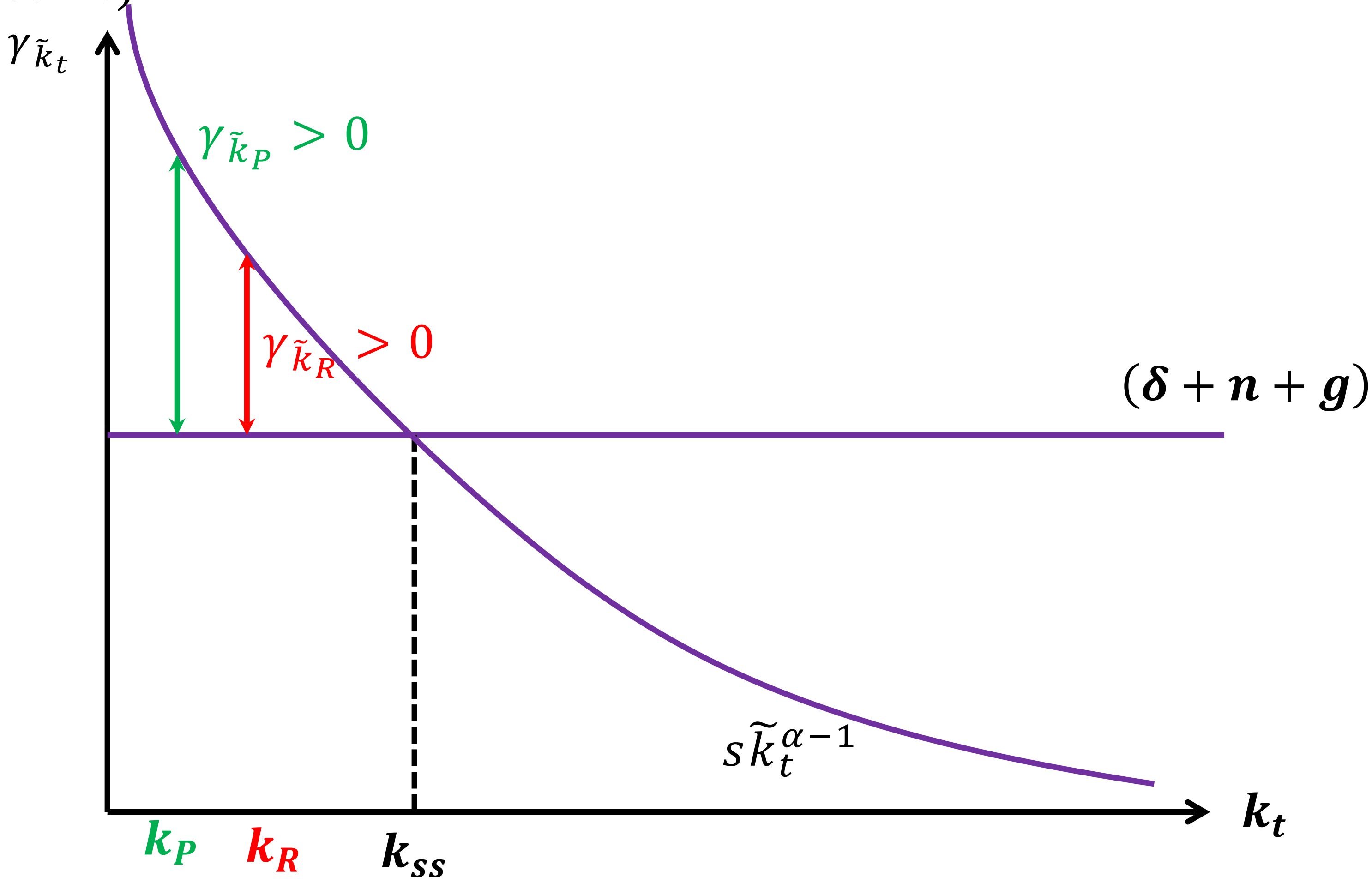
- **The model measures the inherent relationships around each regression point** *i*, where each set of regressors is estimated by a **weighted least squares approach**. The matrix expression for this estimation is:

$$\beta_i = (X'W(u_i, v_i)X)^{-1}\ X'W(u_i, v_i)\mathrm{y}$$

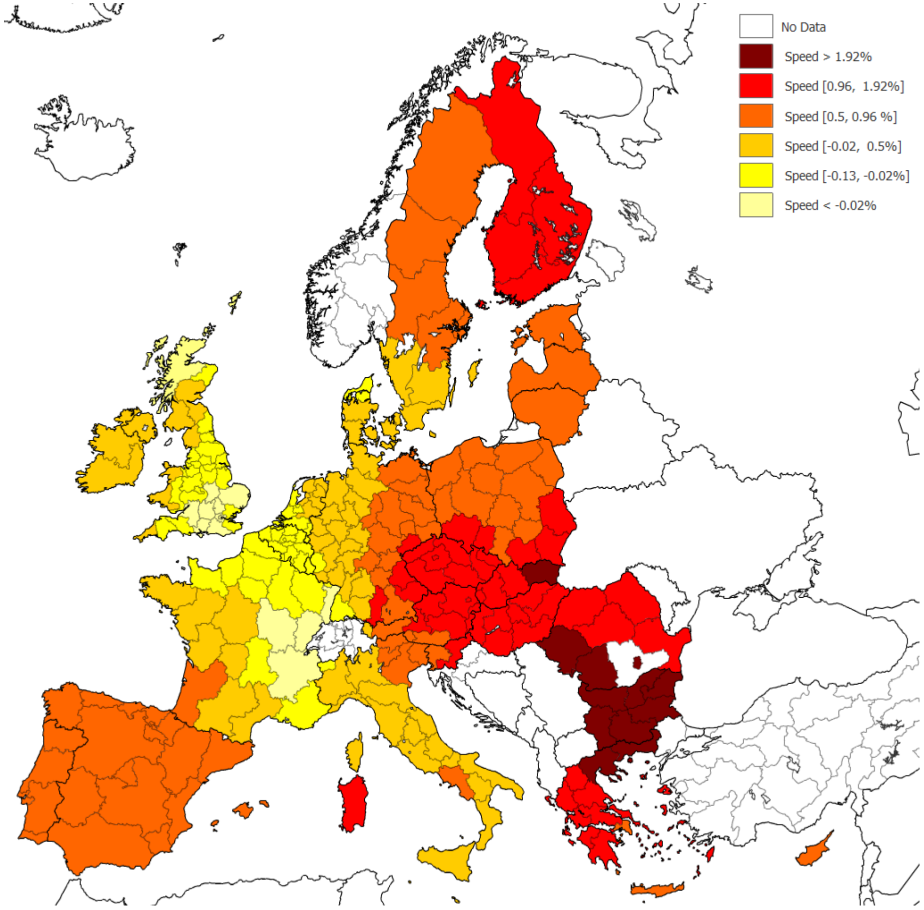
Notice that in this modeling approach $\beta_i = [\beta_{i,0}\ \beta_{i,1}\ \ldots, \beta_{i,m}]'$ is a 1 x k vector, so β is a *n x k* matrix and $W(u_i, v_i)$ is a diagonal *n x n* matrix denoting the geographic observed data for regression point at location (u_i, v_i).

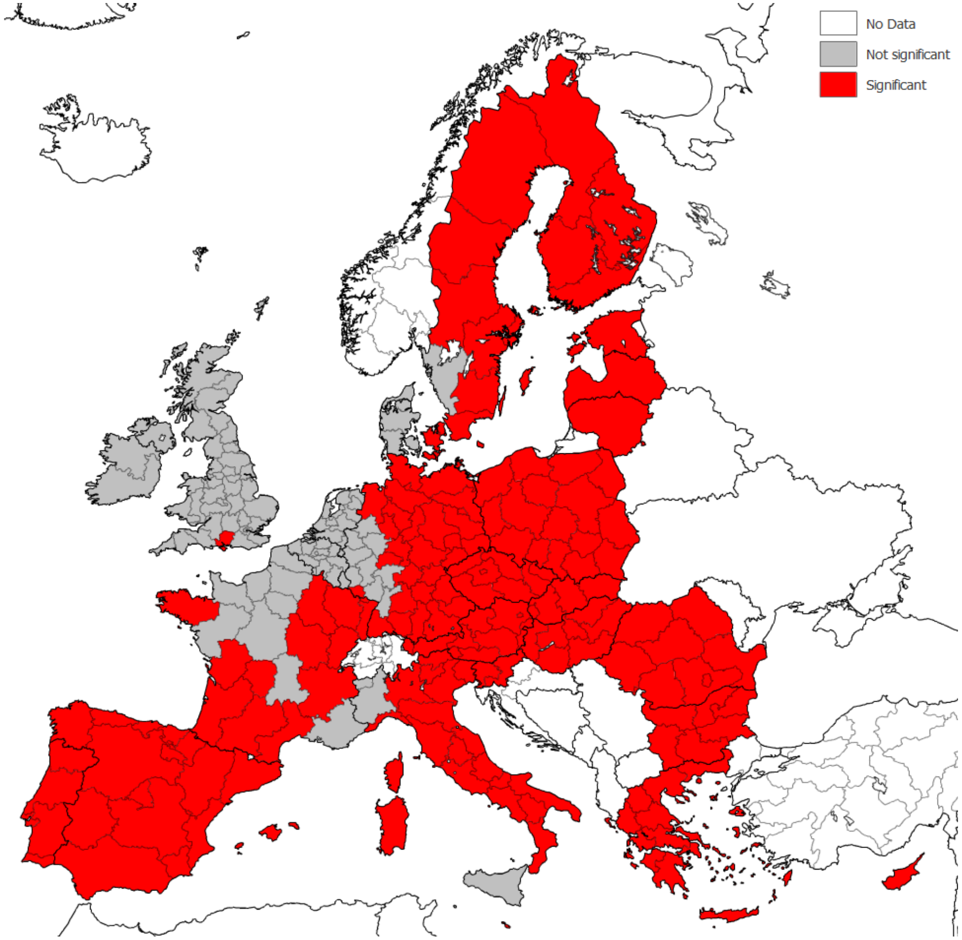
This **W** matrix is determined by some kernel function (i.e, Gaussian, exponential, etc). We will talk later about W!

In the Neoclassical growth model poorer economies grow faster since they are far from their steady state (which produces convergence of income)

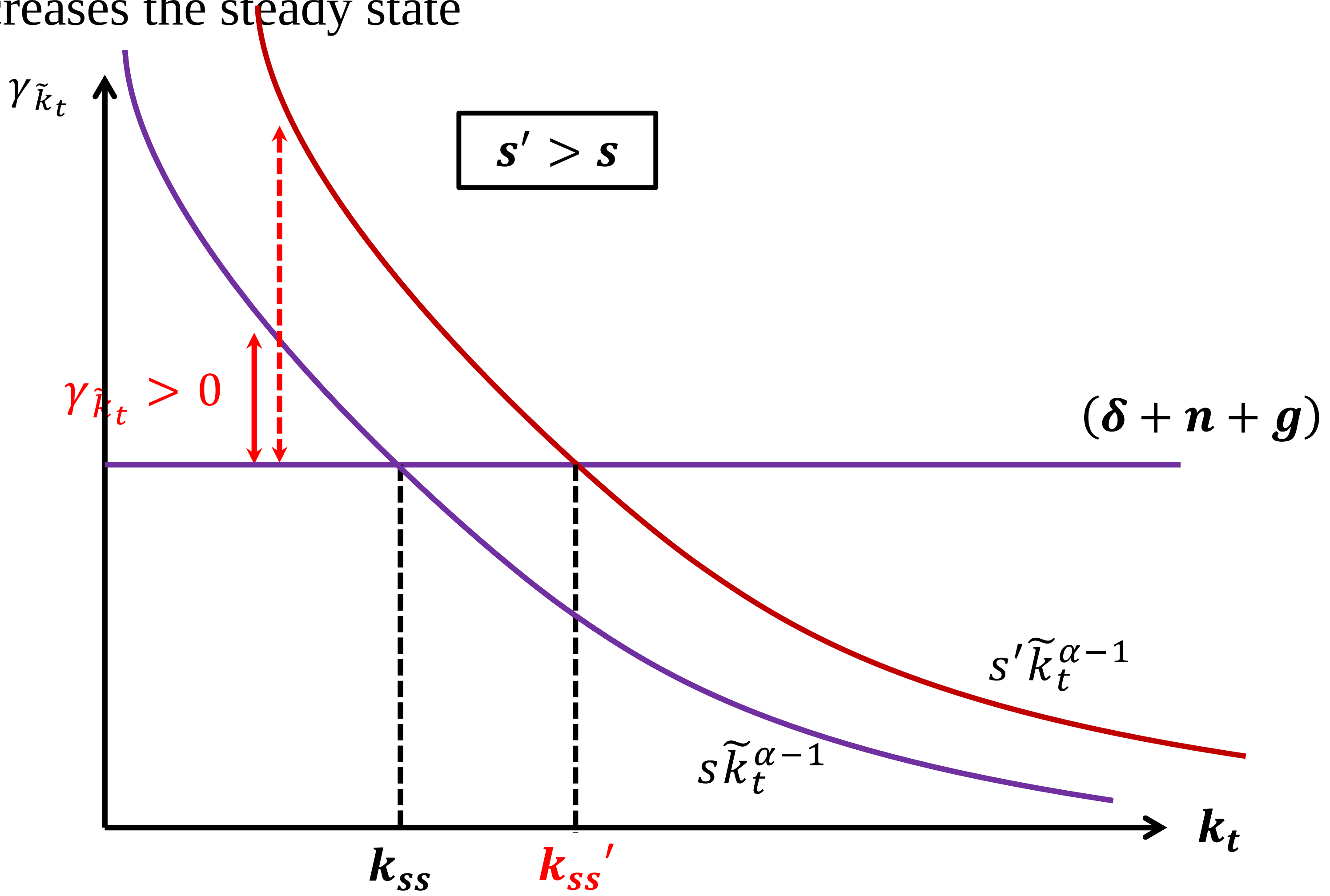


- **The effect of initial income**



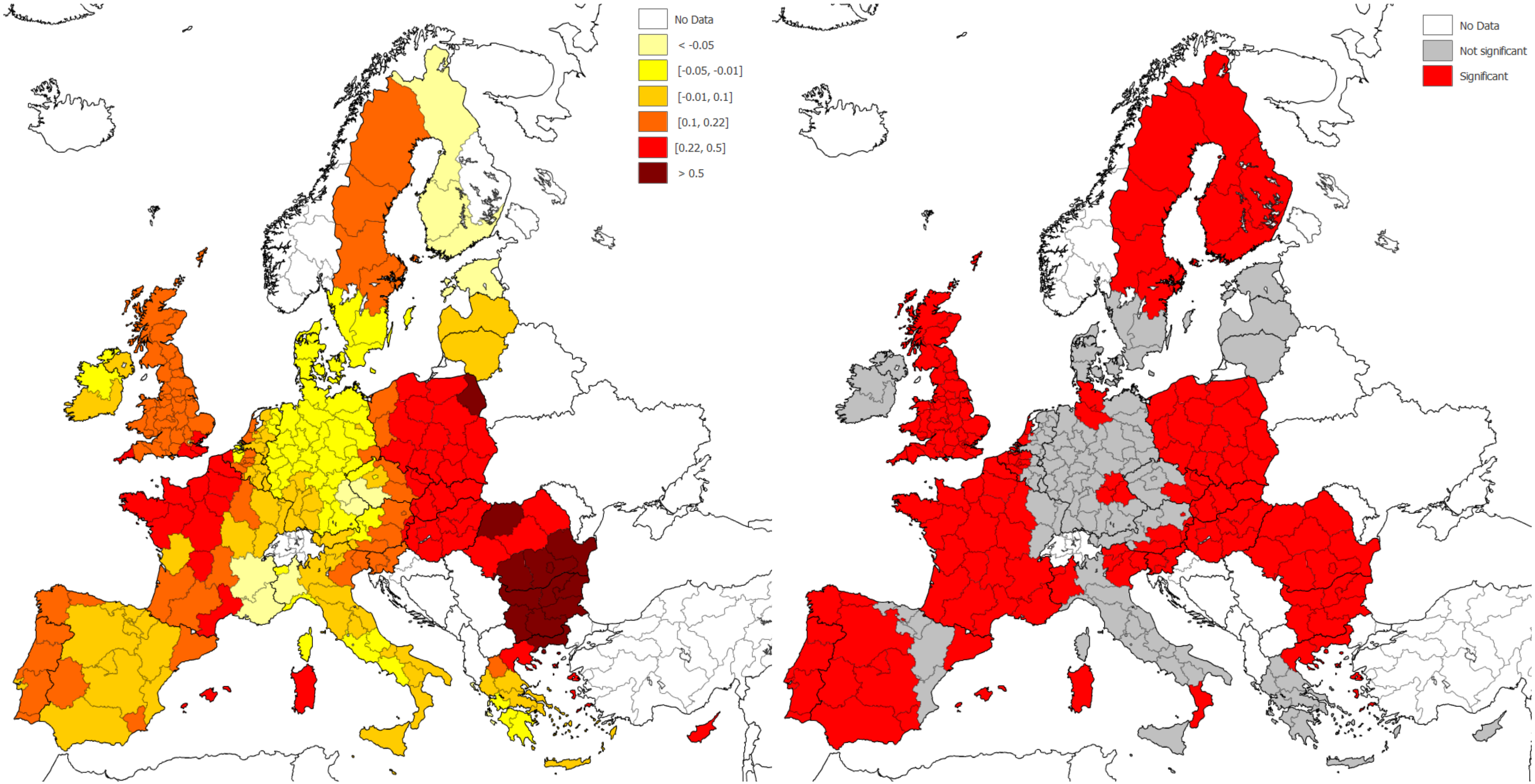


In the Neoclassical growth model, an increase in the investment rate accelerates growth in the transition towards the steady state and increases the steady state



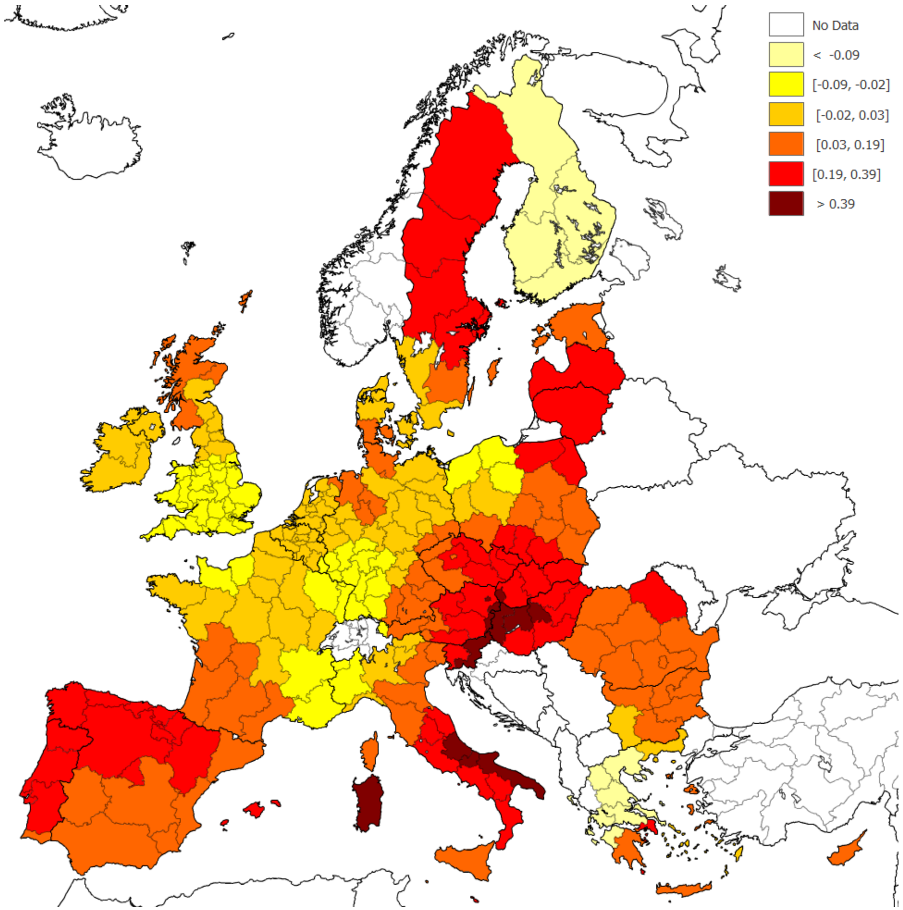
Spatial Heterogeneity and Regional growth in Europe

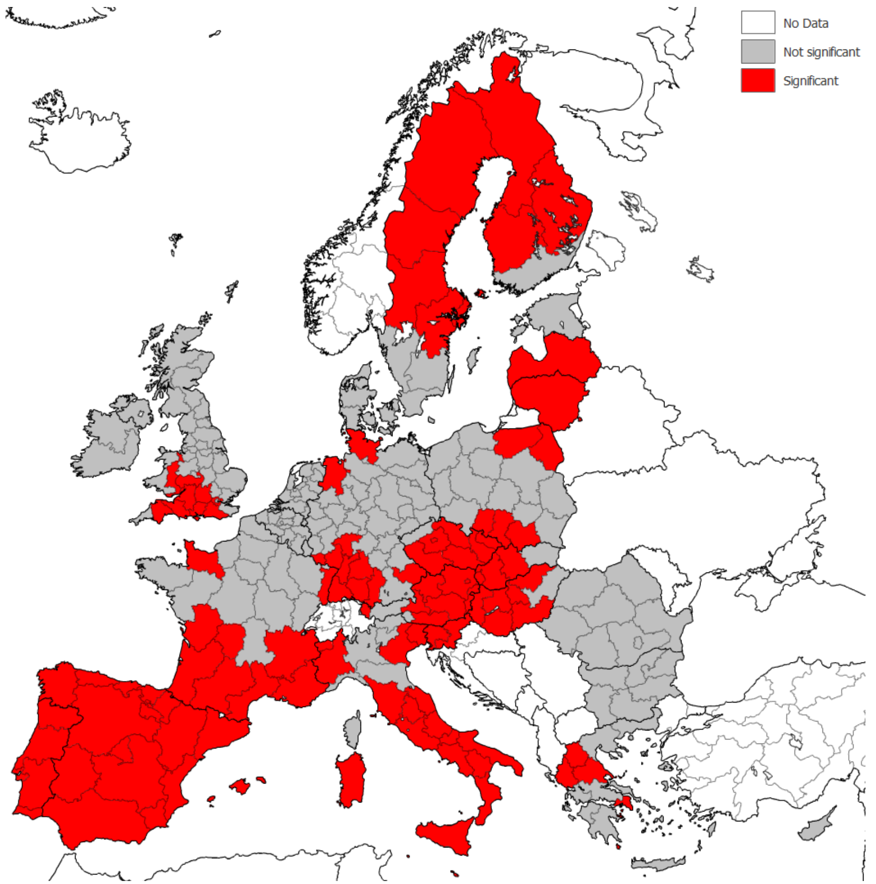
- **The effect of physical capital investment**



Spatial Heterogeneity and Regional growth in Europe

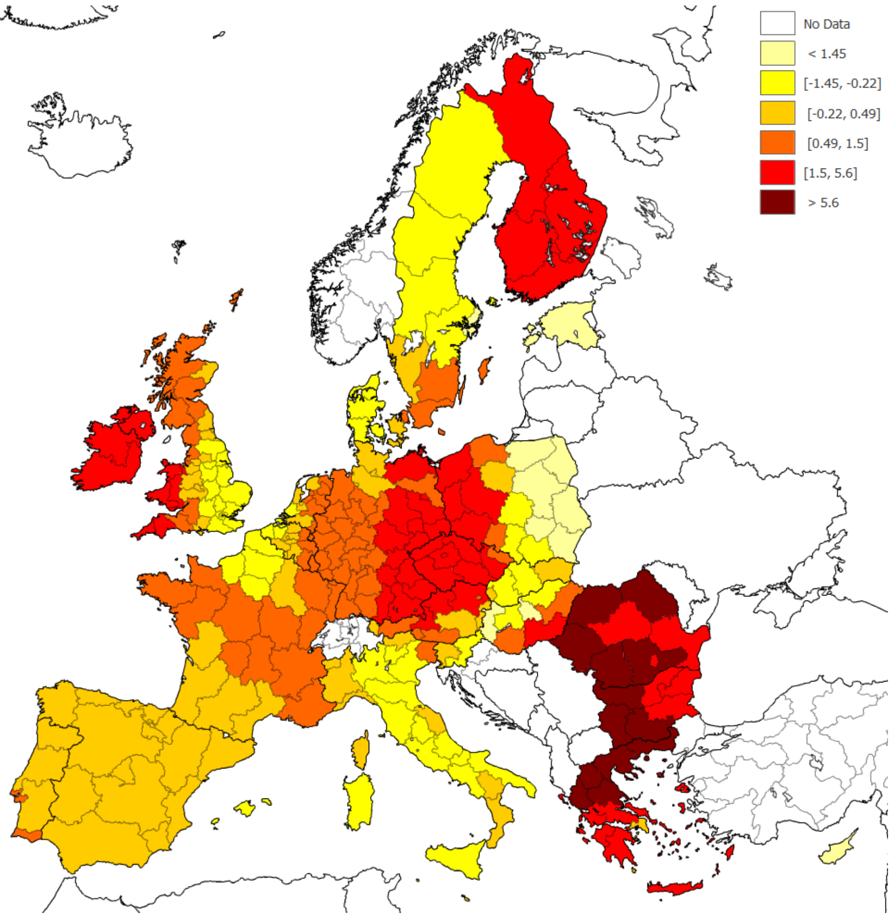
- **The effect of human capital**

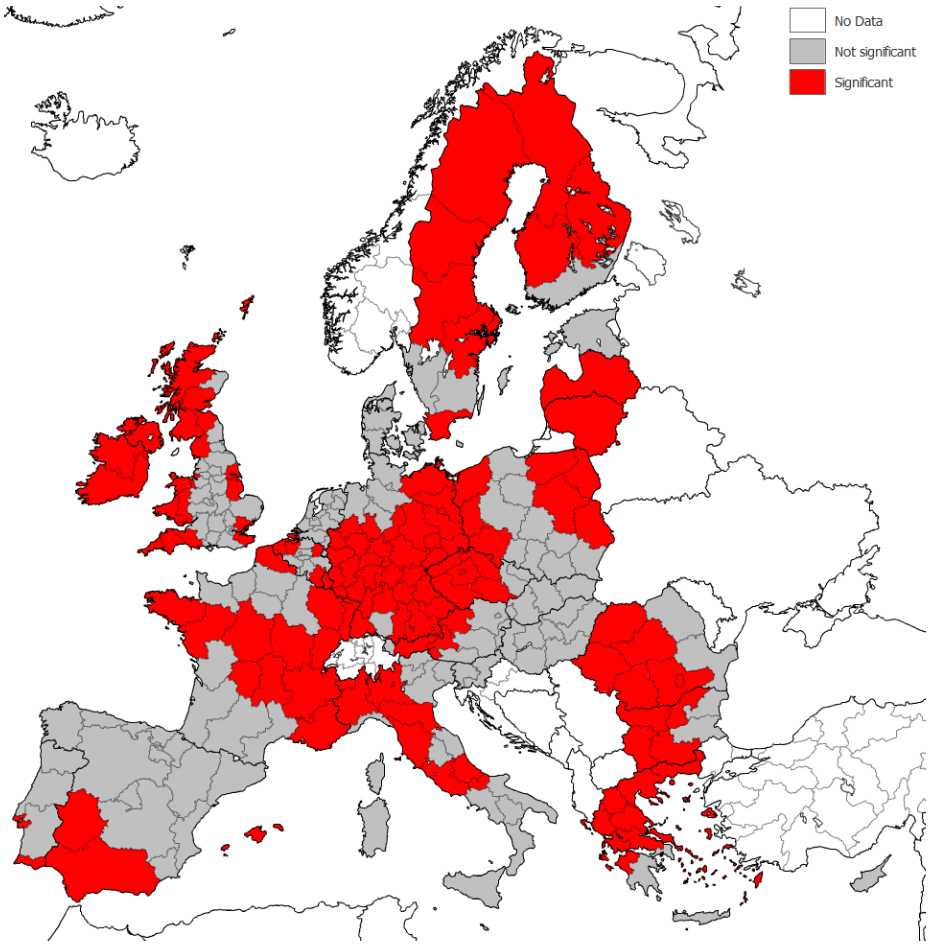




Spatial Heterogeneity and Regional growth in Europe

The effect of population growth





Spatial Heterogeneity and Regional growth in Europe

- As observed GWR produces heterogeneous parameters linked to specific locations
- This might be **useful for policy-making purposes but a causal interpretation on the why there exists such variability is lacking**
- One explanation is that we are missing factors so the variation at some point reflects the omission of variables that are geographically correlated
- If this is a concern, increase the number of X, randomize X and try different combinations
- GWR analysis should be a descriptive 1st step that needs a causal *"add in"* for meaningful analysis. Otherwise interpretation of its results becomes too speculative
- Spatial panel data models with spatially-fixed effects are a much better option to deal with heterogeneity and investigate causal economic relationships

Spatial dependence

- **Spatial econometrics deals with spatial effects**

  (II) Spatial dependence

Definition:

What happens in *i* depends on what happens on *j*. Formally, for spatial unit *i*:

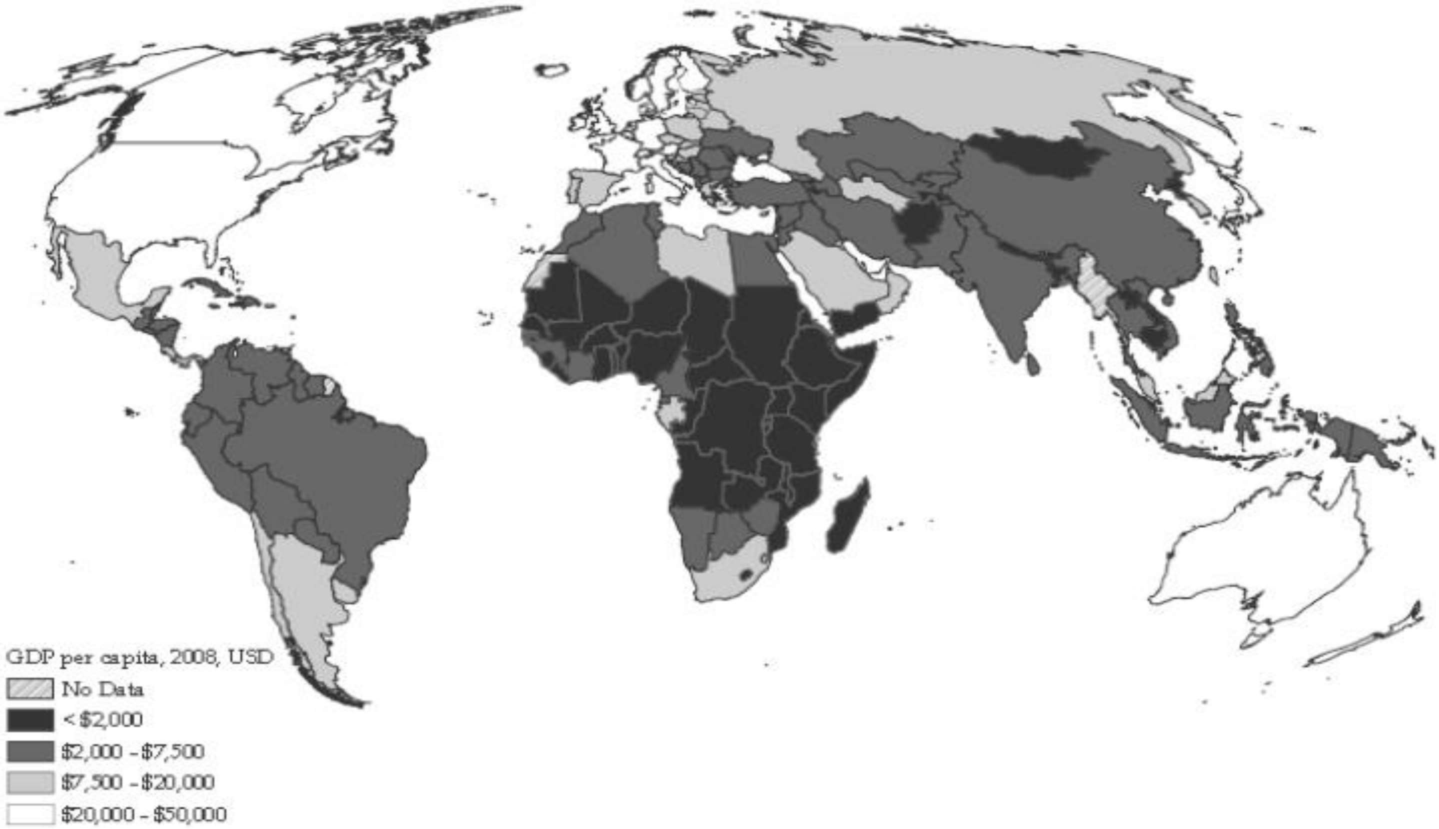
$$y_i = f\left(y_i, y_j\right) + u_i \; \forall \; i \neq j$$

Spatial dependence is a special case of cross-sectional dependence

Correlation structure is derived from a specific ordering, determined by the spatial arrangement in geographic space or in network spacee

Spatial dependence

- GDP per capita

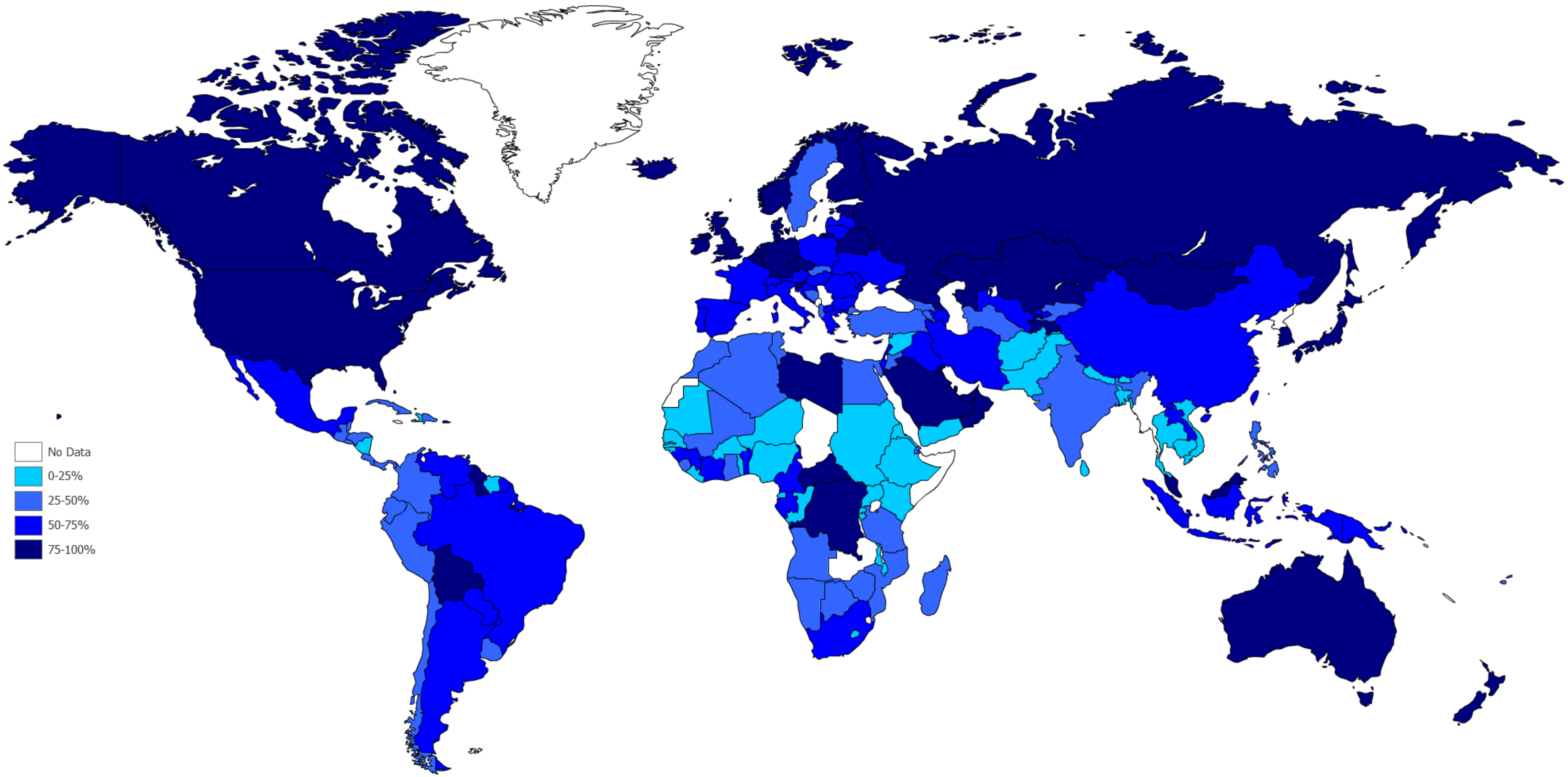


Spatial dependence

- Night Lights

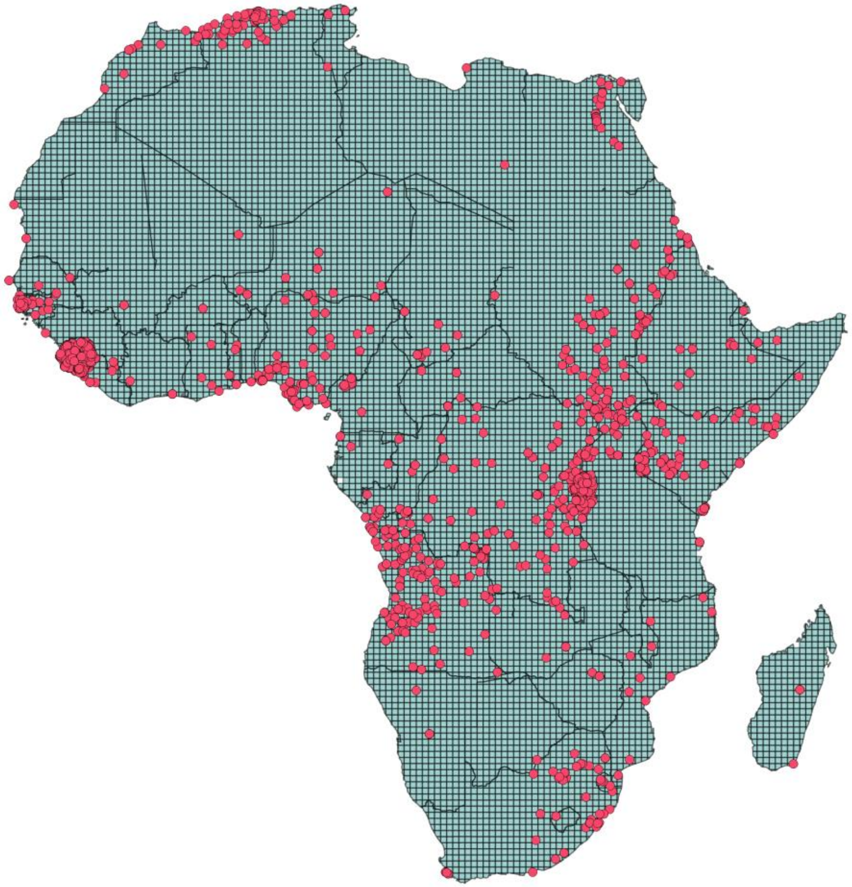
Spatial dependence

- CO2 emissions around the globe



Spatial dependence

- **Conflict events in Africa (with > 25 deaths, revolutions, riots, etc)**



Spatial dependence

Why do we need to model spatial dependence?

- Important aspect when studying spatial units (cities, regions, countries)

  [i] Potential relationships and interactions between them

  [ii]Assumption of economies evolving as isolated entities is not reliable (co-evolution)

- Example of Interdependence: local government spending

  [i] Analyze regions/cities as independent units?

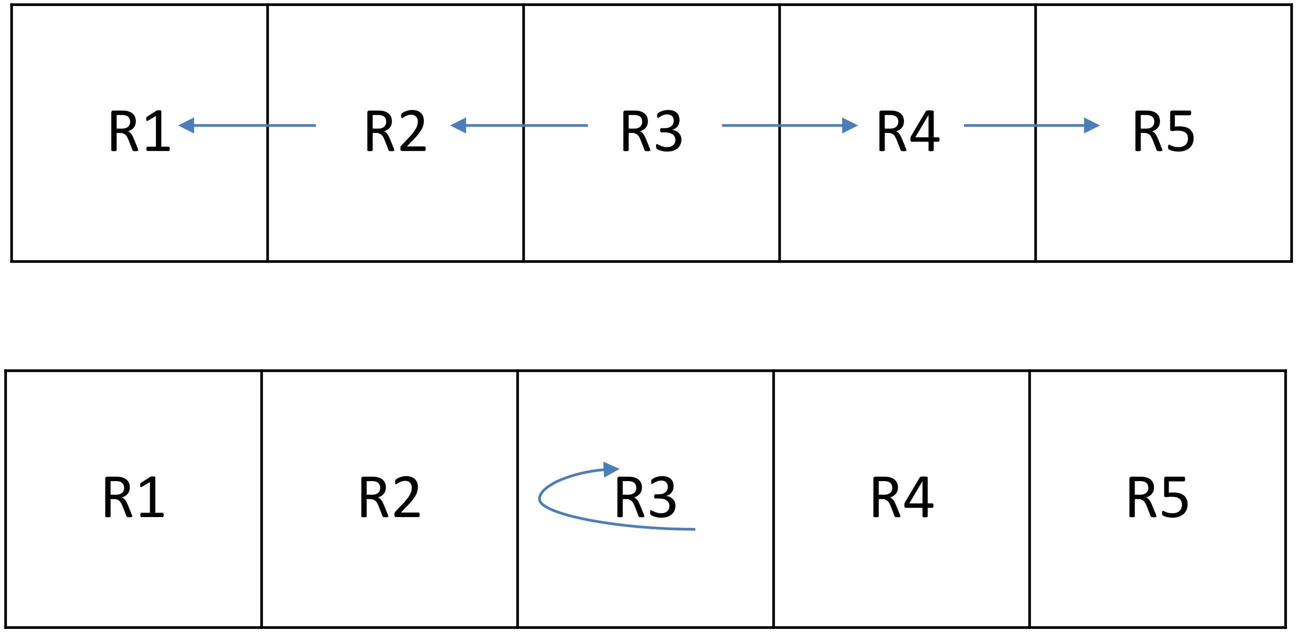
  [ii] No, Regions/cities are spatially/game-theoretically interrelated

  [iii] Existence of fiscal policy spillovers/externalities

  An increase in government spending in Pisa will affect government spending in neighboring cities but the impact will be lower for more distance cities

Spatial dependence

- Fiscal policy spillovers



Spatial dependence: A local fiscal policy example

- Assume we live in a world with 5 cities:

Pisa

Livorno

Lucca

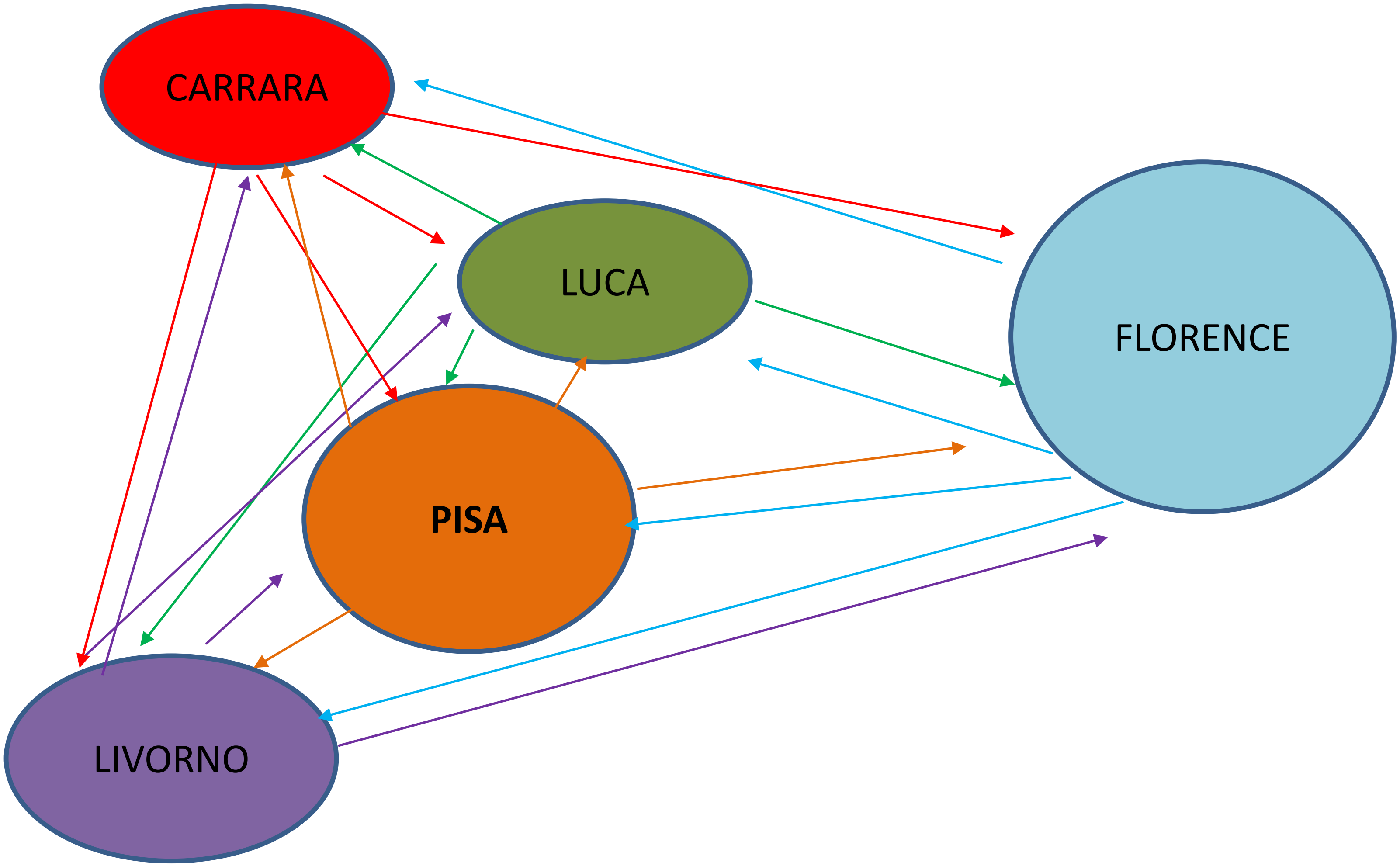
Carrara

Florence

- Each city has a different budget to spend the money of tax-payers in some service (i.e, sport centers, roads, etc)

- Decision on how much to spend in Pisa depends on what the other cities are going to spend, the decision on how much to spend in Livorno depends on what the others are going to spend….

Spatial dependence: A local fiscal policy example



Spatial dependence: A local fiscal policy example

- Assume representative agents in each city i have the following utility function

$$V = v\left(G_i, C_i, \widetilde{G}_j\right)$$

where:

G_i denotes the consumption/expenditure in the public good in *i*

C_i denotes the consumption/expenditure in the private good in *i*

$\widetilde{G}_j$ denotes the consumption/expenditure of the public good in neighboring cities $j \neq i$. Neighbor's expenditure is a weighted average of the expenditure in other cities:

$$\widetilde{G}_j = 1/n \sum_{j=1}^{n} w_{ij} G_j$$

where the weights $w_{ij} \in (0,1)$ are a negative unknown function of geographical distance between *i* and *j*

Spatial dependence: A local fiscal policy example

- Each municipality faces the following budget constraint:

$$pG_i + C_i = Y_i^d + T_i$$

where:

p is the price of the public good and the price of the private good has been normalized to 1 (i.e, numerarie)

Y_i^d is the disposable income after taxes

T_i are the transfers received from upper tier levels of government (regions, central gov)

We are going to model the optimal solution to the problem of an agent that wants to get maximum felicity levels from G and C

Spatial dependence: A local fiscal policy example

We assumed a general concave/monotonic utility function V in G and C

$$V = v\left(C_i, G_i, \widetilde{G}_j\right)$$

To get a numerical closed solution we will assume V has the following form:

$$V = {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta} + \theta \widetilde{G}_j$$

yielding the following Lagrangian:

$$L = {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta} + \theta \widetilde{G}_j + \lambda[Y_i^d + T_i - pG_i + C_i]$$

F.O.C

$[G_i]$:$\lambda p = \beta {C_i}^{\alpha} {G_i}^{\beta-1} {\widetilde{G}_j}^{\delta}$

$[C_i]$: $p = \alpha {C_i}^{\alpha-1} {C_i}^{\beta} {\widetilde{G}_j}^{\delta}$

$[G_j]$: $\theta \ = \delta {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta-1}$

Spatial dependence: A local fiscal policy example

Combining

$[C_i]$: $\lambda = \beta {G_i}^{\alpha} {C_i}^{\beta-1} {\widetilde{G}_j}^{\delta}$

$[G_i]$: $\lambda p = \alpha {G_i}^{\alpha-1} {C_i}^{\beta} {\widetilde{G}_j}^{\delta}$

We get that in the optimum:

$$G_i = \frac{1}{p}\frac{\beta}{\alpha} C_i$$

Now we are going to solve for C in the FOC of G_j to get an expression on *"how spending elsewhere"* affects spending in *i*

We can re-arrange

$$[G_j]:\ \theta\ = \delta {C_i}^{\alpha} {G_i}^{\beta} {\widetilde{G}_j}^{\delta-1}$$

$$C_i = \left[\frac{\theta}{\delta} {G_i}^{\beta} {\widetilde{G}_j}^{\delta-1}\right]^{\frac{1}{\alpha}}$$

Spatial dependence: A local fiscal policy example

If we plug $C_i = \left[\frac{\theta}{\delta} G_i{}^{\beta} \widetilde{G}_j{}^{\delta-1}\right]^{\frac{1}{\alpha}}$ into $G_i = \frac{1}{p}\frac{\beta}{\alpha} C_i$

we obtain:

$$G_i = \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\beta}} \left(\frac{1}{p}\right)^{\frac{\alpha}{\beta}} \left(\frac{\theta}{\delta}\right)^{\frac{1}{\beta}} \widetilde{G}_j{}^{\frac{\beta}{1-\delta}}$$

Therefore, the optimal reaction in the spending decision of a municipality *i* depends on the spending in *j* as long as:

$$\frac{\partial G_i}{\partial \tilde{G}_j} = \frac{\beta}{1-\delta} \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\beta}} \left(\frac{1}{p}\right)^{\frac{\alpha}{\beta}} \left(\frac{\theta}{\delta}\right)^{\frac{1}{\beta}} \widetilde{G}_j{}^{\frac{\beta}{1-\delta}-1} \neq 0$$

From the concavity assumption on C_i and G_i (which is a classic in utility modeling) $\beta, \alpha \in (0,1)$. The sign will depend on θ and δ

Spatial dependence: A local fiscal policy example

- In general, we have three type of spatial interactions among governments (in spending):

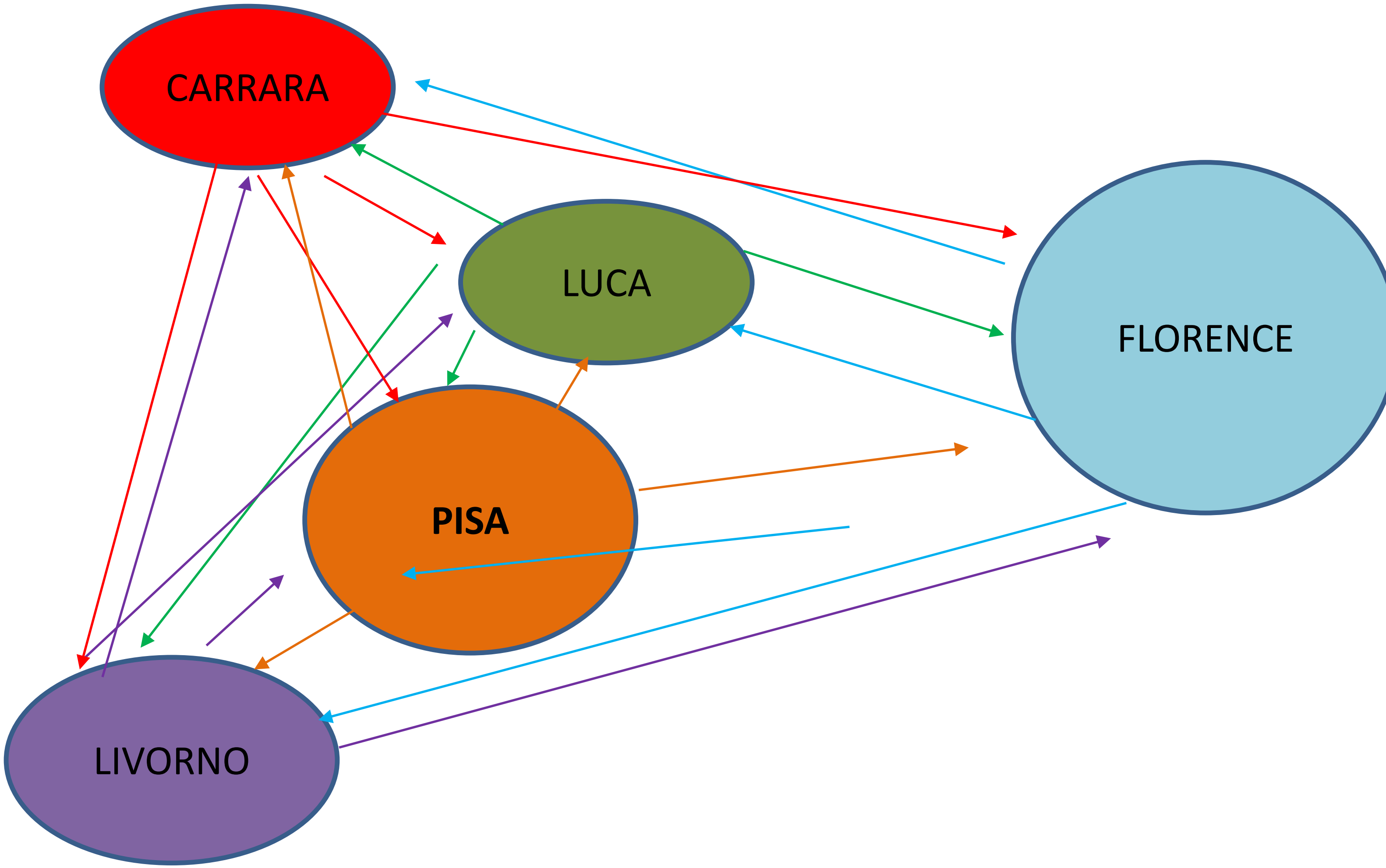
$$\frac{\partial G_i}{\partial \tilde{G}_j} = \rho < 0 \rightarrow \text{Strategic Substitutes (free-riding?)}$$

$$\frac{\partial G_i}{\partial \tilde{G}_j} = \rho > 0 \rightarrow \text{Complementarity (self-reinforcing)}$$

$$\frac{\partial G_i}{\partial \tilde{G}_j} = \rho = 0 \rightarrow \text{Independent (classical non-spatial analysis)}$$

Spatial dependence: A local fiscal policy example

How would you model this econometrically?



Spatial dependence

- Using our previous example, we would like to estimate:

$$y_1 = \beta_{21}y_2 + \beta_{31}y_3 + \beta_{41}y_4 + \beta_{51}y_5 + u_1$$

$$y_2 = \beta_{12}y_1 + \beta_{32}y_3 + \beta_{42}y_4 + \beta_{52}y_5 + u_2$$

$$y_3 = \beta_{13}y_1 + \beta_{23}y_2 + \beta_{43}y_4 + \beta_{53}y_5 + u_3$$

$$y_4 = \beta_{14}y_1 + \beta_{24}y_2 + \beta_{34}y_3 + \beta_{54}y_4 + u_4$$

$$y_5 = \beta_{15}y_1 + \beta_{25}y_2 + \beta_{35}y_3 + \beta_{45}y_4 + u_5$$

where β_{ij} is the effect on government spending *y* of city *i* on city *j*
What is the problem with this modelling strategy?

K > N!!

Under standard econometric modeling, it is imposible to model such interdependence

Spatial dependence

- Spatial econometrics approach to model optimal spending reaction functions

$$y_i = \rho \sum_{j=1}^{n} w_{ij}\, y_j$$

This problem is solvable: 1 parameter and 5 data points.

In the first two scenarios where $\rho \neq 0$, complementary/substitute relationships for each *i* depend on his neighbors and the influence they exert, measured by w_{ij}.

$w_{ij} = f(-d_{ij})$ negative function of distance and informative on how *i* is connected to the rest of the system

However, w_{ij} has to be pre-specified in advance (it is fixed) and usually is not estimated!

Spatial autocorrelation

- **Distance matters**

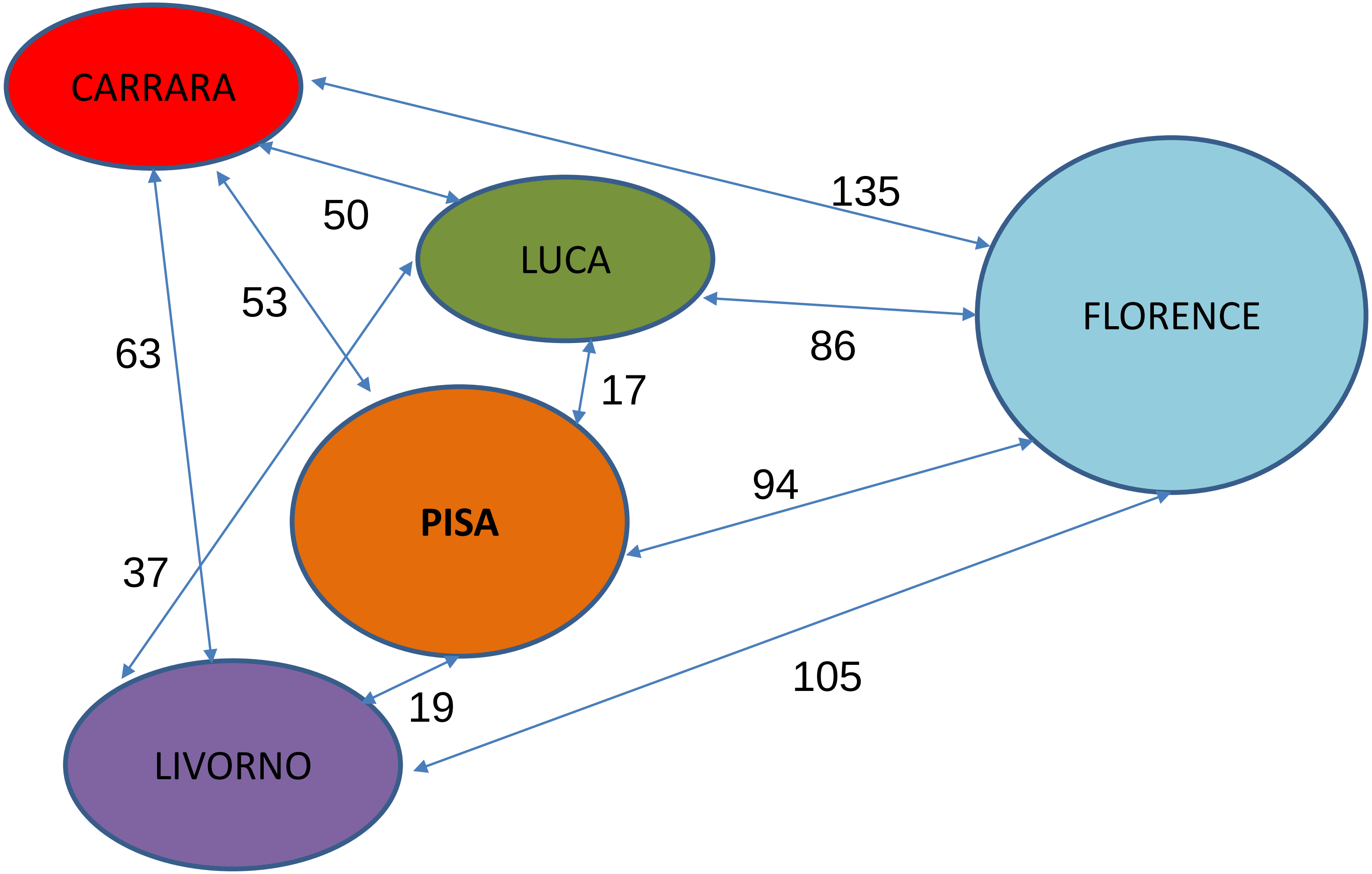
Key Point:

First law of geography of Tobler: *"everything is related to everything else"*, but near things are more related than distant things

This first law is the foundation of the fundamental concept of spatial dependence and spatial autocorrelation

Spatial autocorrelation

Distances between our 5 cities



Spatial autocorrelation

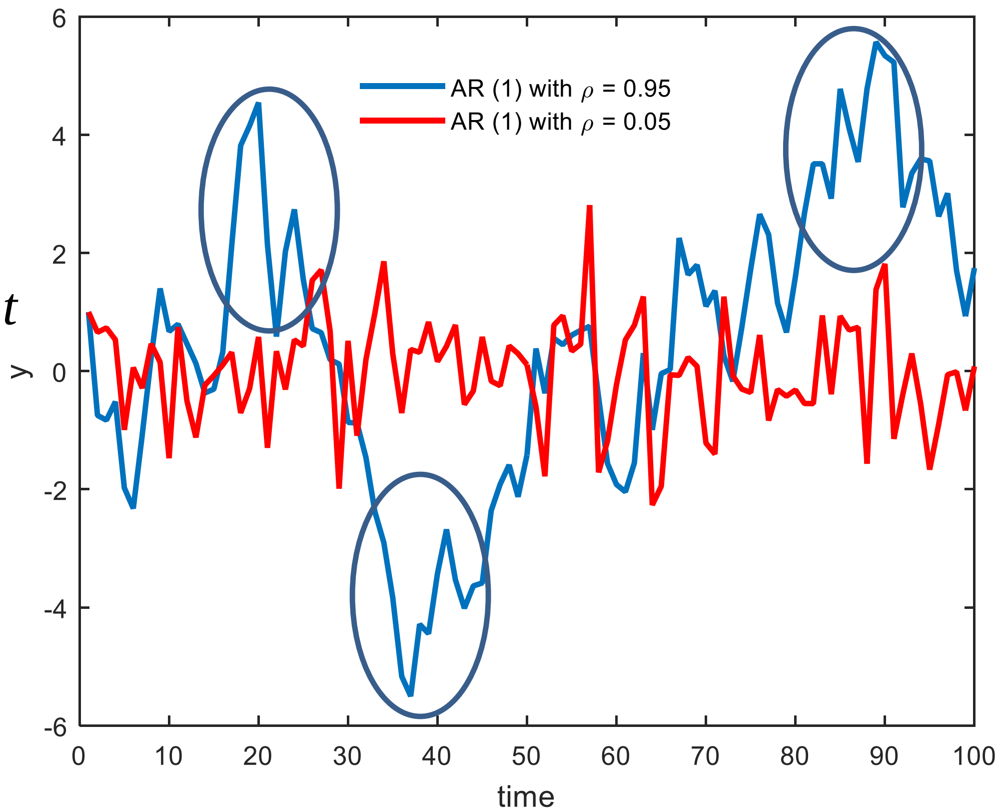
- **Spatial Autocorrelation**
  - Autocorrelation: the correlation of a variables with itself
  - Time series: the value of a variable at time t depends on the value of the same variable at time *t-1*
  - AR(1) model $y_t = \rho y_{t-1} + u_t$

High temporal correlation?

When y is high in period *t-1*, it is very likely that it will be high in *t*

Low temporal correlation?

When y is low in *t-1*, it is difficult to know what would be the value in *t*



Spatial autocorrelation

- **Spatial Autocorrelation**
  - Autocorrelation: the correlation of a variables with itself
  - Time series: the value of a variable at time *t* depends on the value of the same variable at time *t-1*
  - Space: **the value of the variable at point/region i depends on the value of the same variable at point/region j.**

Definition:
Correlation between the same attribute at two (or more) different locations

Value similarity coincides with location similarity

It can be positive and negative

Spatial autocorrelation

- **Definition of positive spatial autocorrelation**

Observations with high (or low) values of a variable tend to be clustered in space

Figure: Positive spatial autocorrelation

1	1	
1	1	

Spatial autocorrelation

- **Definition of negative spatial autocorrelation**

Locations tend to be surrounded by neighbors having very disimilar values

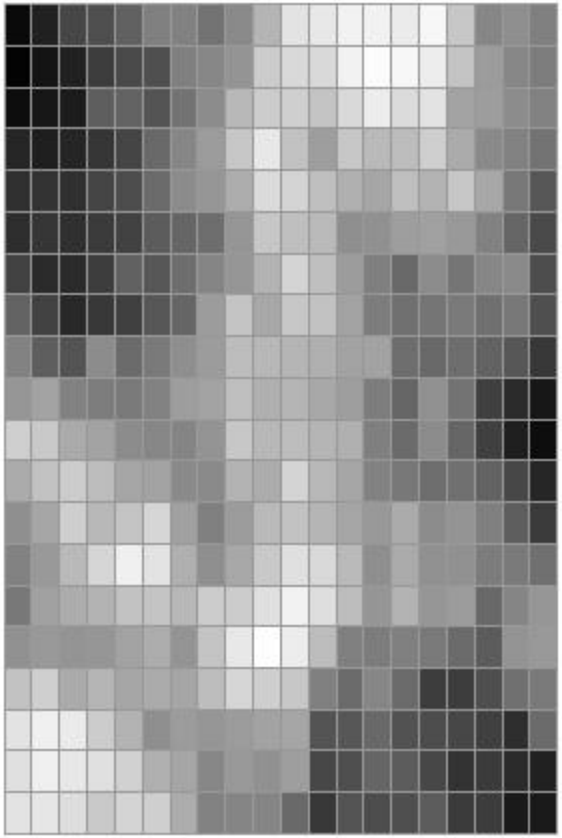
Figure: Negative autocorrelation

1		1
	1	
1		1

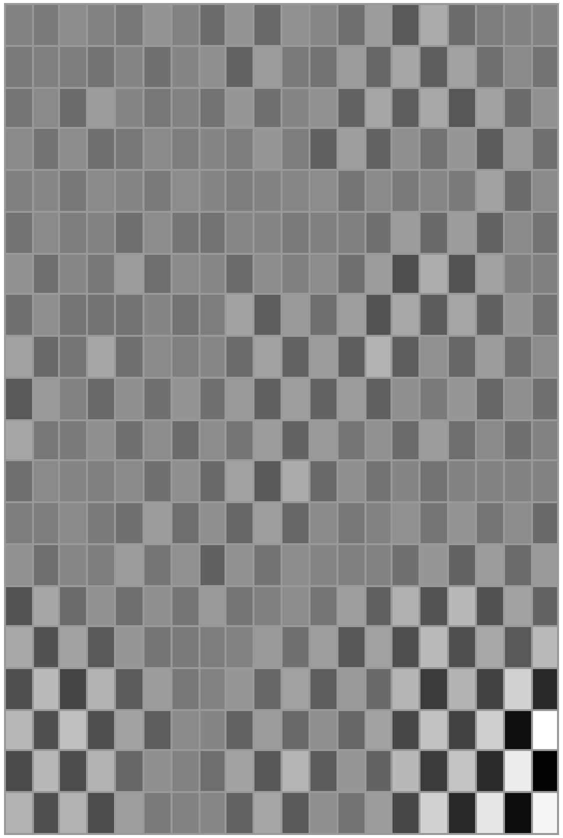
Spatial autocorrelation

- Other examples

Positive spatial autocorrelation



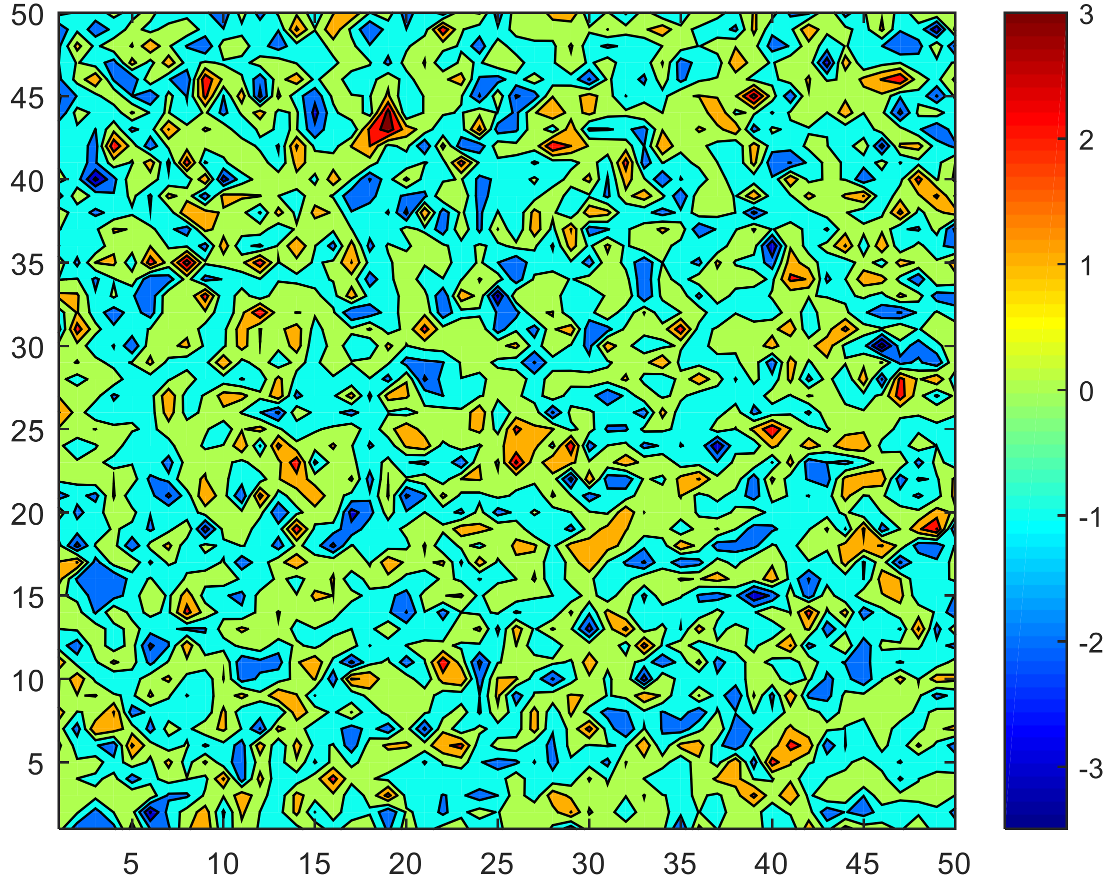
Negative spatial autocorrelation



Spatial autocorrelation

- **Spatial randomness**

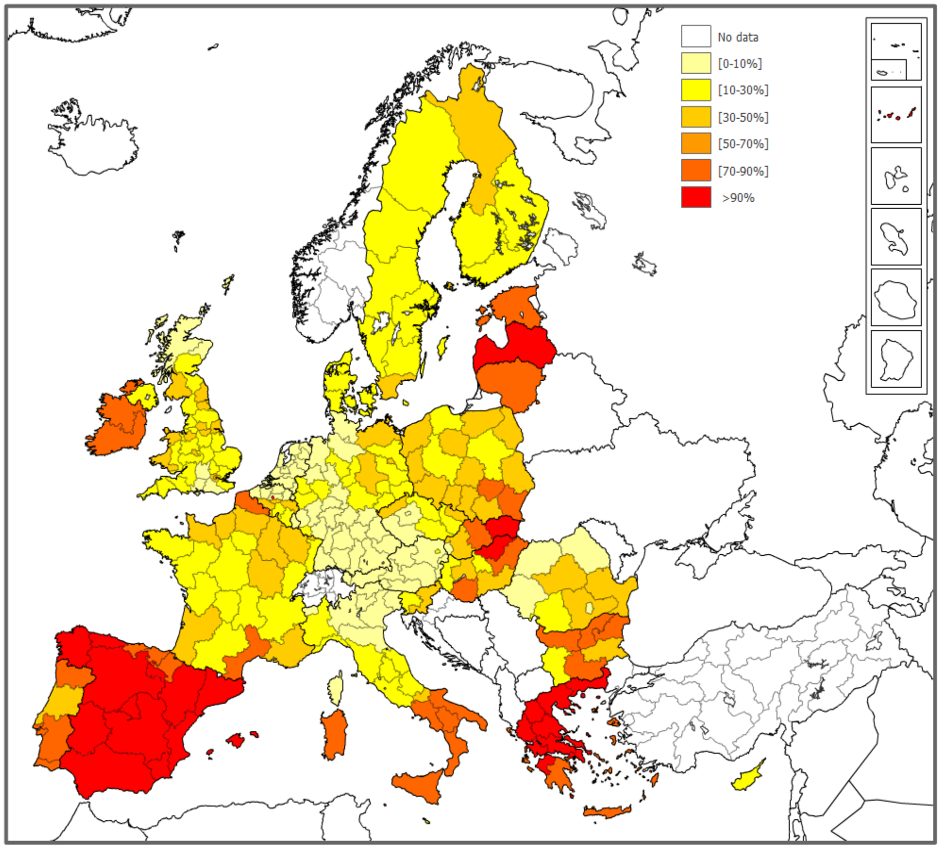
When none of the two situations occur

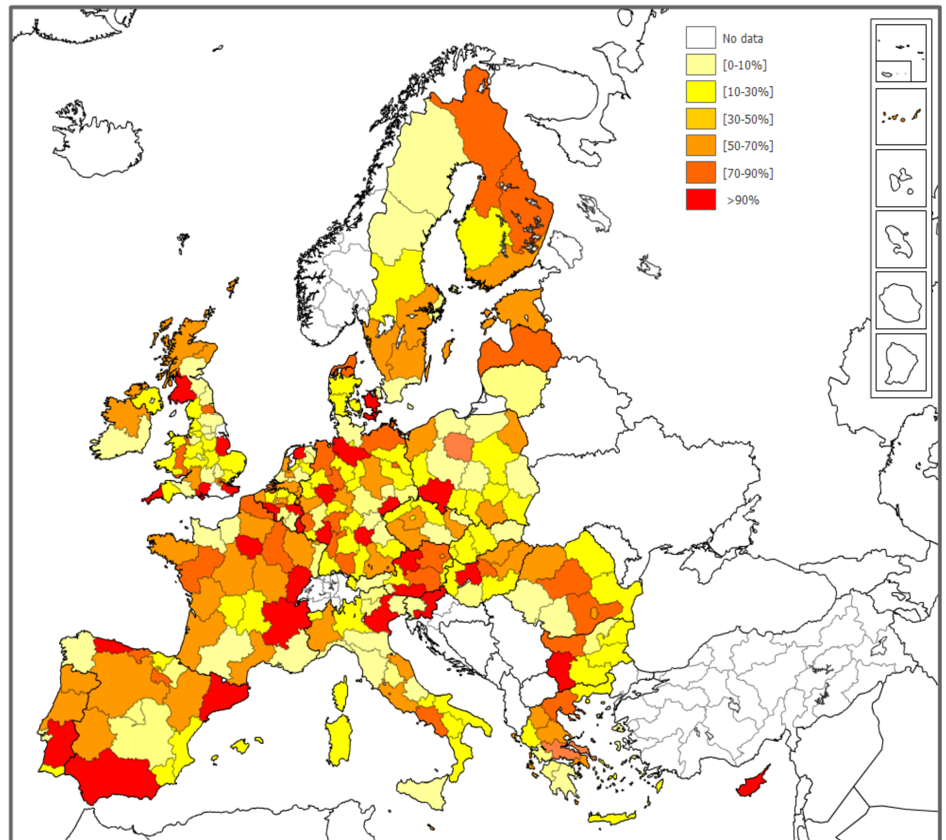


Spatial autocorrelation

Real distribution of unemployment rates in European regions (2011) vs a spatially random distribution with the same parameters

$U \sim (\mu = 9.4, \sigma = 5.25)$





Spatial autocorrelation

- **There are two main sources of spatial autocorrelation** (Anselin, 1988)
  - Measurement errors/Nuisance/Spatially correlated shocks
  - Substantive spatial interactions (like the fiscal policy example before)

Some researchers treat space as nuisance → Spatial filters

The **second one** is of **much more interest**

The essence of regional sciences and new economic geography is that location and distance matters explaining what we observed in real world

What is observed at one point is determined by what happen elsewhere in the system

Spatial Weight Matrix

- **Back to Tobler's first law of geography**

"everything depends on everything else, but closer things more so"

- Important ideas
  - Existence of spatial dependence
  - Structure of spatial dependence
    - Distance decay
    - Closeness = Similarities

Spatial Weight Matrix

One **crucial issue** in spatial econometric is the problem of formally **incorporating spatial dependence into the model**.

- **Questions**

What would be a good criterion to define closeness in space?

How to determine which other units in the system influence the one under consideration?

Spatial Weight Matrix

- The device typically used in spatial analysis is the so-called spatial weight matrix, or simply **W** matrix

- It imposes a structure in terms of what are the neighbors for each location

- Assigns weights that measure the intensity of the relationship among pair of spatial units

Spatial Weight Matrix

- **Definition of W matrix**

Let n be the number of spatial units. The spatial weight matrix, **W**, a n x n positive *symmetric* matrix with element $w_{i,j}$ at location i,j. The values of $w_{i,j}$ or the weights for each pair of locations are assigned by some preset rules which defines the spatial relations among locations. By convention $w_{i,j} = 0$ for the diagonal elements.

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

Entry $w_{i,j} \neq 0$ if i and j are "connected"

The tricky part is how the word "connected" is defined.

Spatial Weight Matrix

- **Two main approaches to define connectivity**
  - Links based on boundaries
  - Links based on distance

Spatial Weight Matrix based on boundary links

- The availability of poligon or lattice data permits the construction of contiguity-based spatial weight matrices. A typical specification of the contiguity relationship in the spatial weight matrix is:

$$w_{i,j} \begin{cases} 1 \text{ if i and j are contiguous} \\ 0 \text{ if i and j not contiguous} \end{cases}$$

- Binary contiguity:
  - Rook criterion (Common border)
  - Bishop criterion (Common vertex)
  - Queen criterion (Either common border or vertex)

Spatial Weight Matrix based on boundary links

Who are the neighbors of region 5?

Criteria: Rook contiguity (border)

1	2	3
4	5	6
7	8	9

1	2	3
4	5	6
7	8	9

Common border: 2, 4, 5, 6

Spatial Weight Matrix based on boundary links

Who are the neighbors of region 5?

Criteria: Bishop contiguity (vertex)

1	2	3
4	5	6
7	8	9

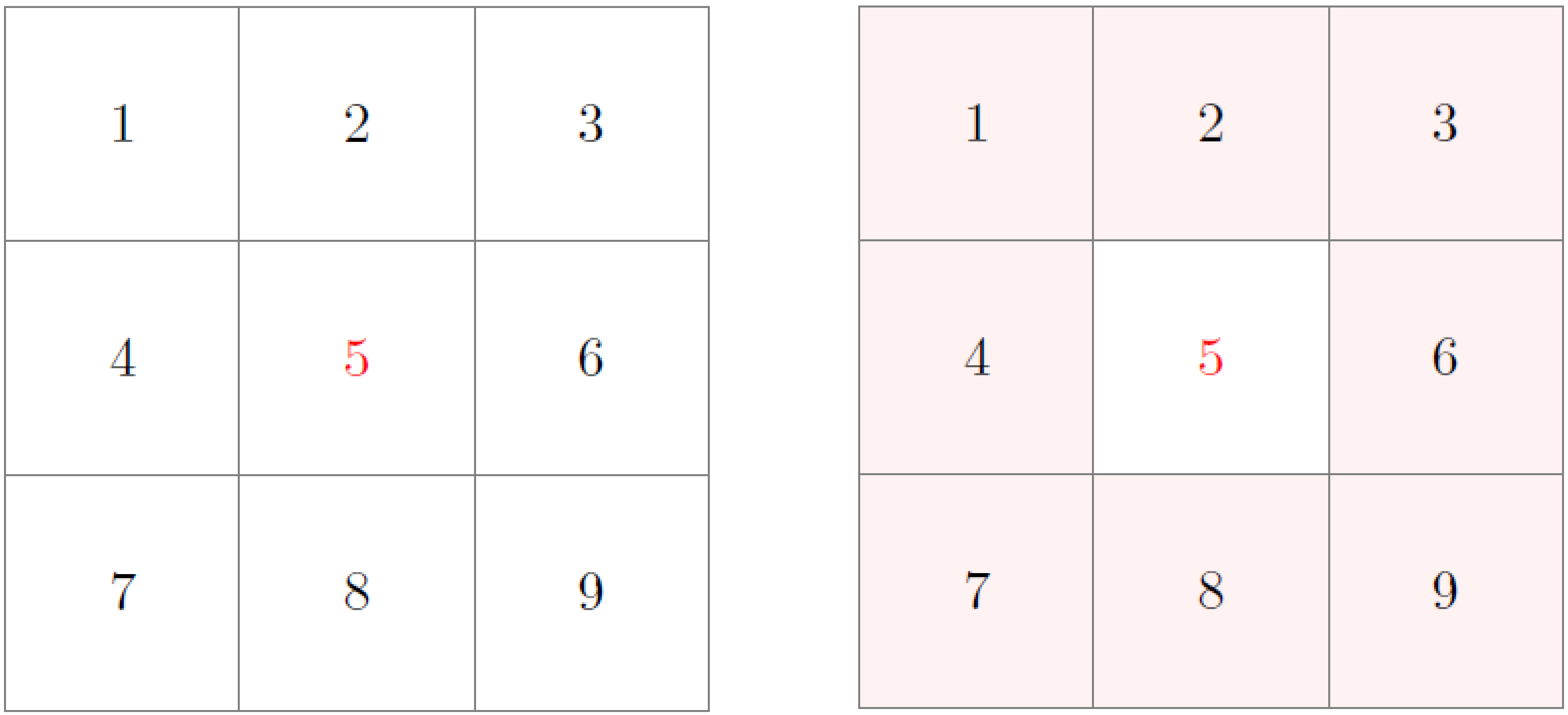
1	2	3
4	5	6
7	8	9

Common vertex: 1, 3, 7, 9

Spatial Weight Matrix based on boundary links

Who are the neighbors of region 5?

Criteria: Queen contiguity

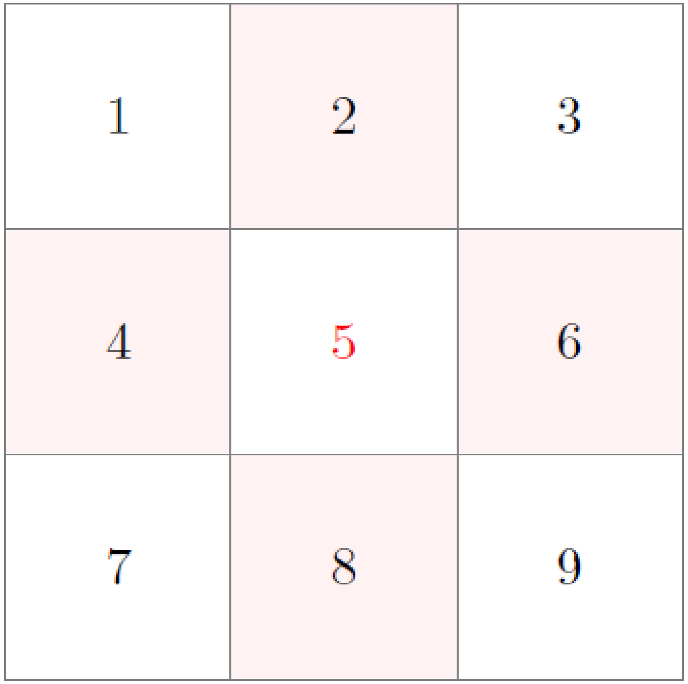


Common border and common vertex: 1, 2, 3, 4, 5, 6, 7, 8 and 9

Spatial Weight Matrix based on boundary links

- Rook contiguity (common border)

- $$\boldsymbol{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$



Spatial Weight Matrix based on boundary links

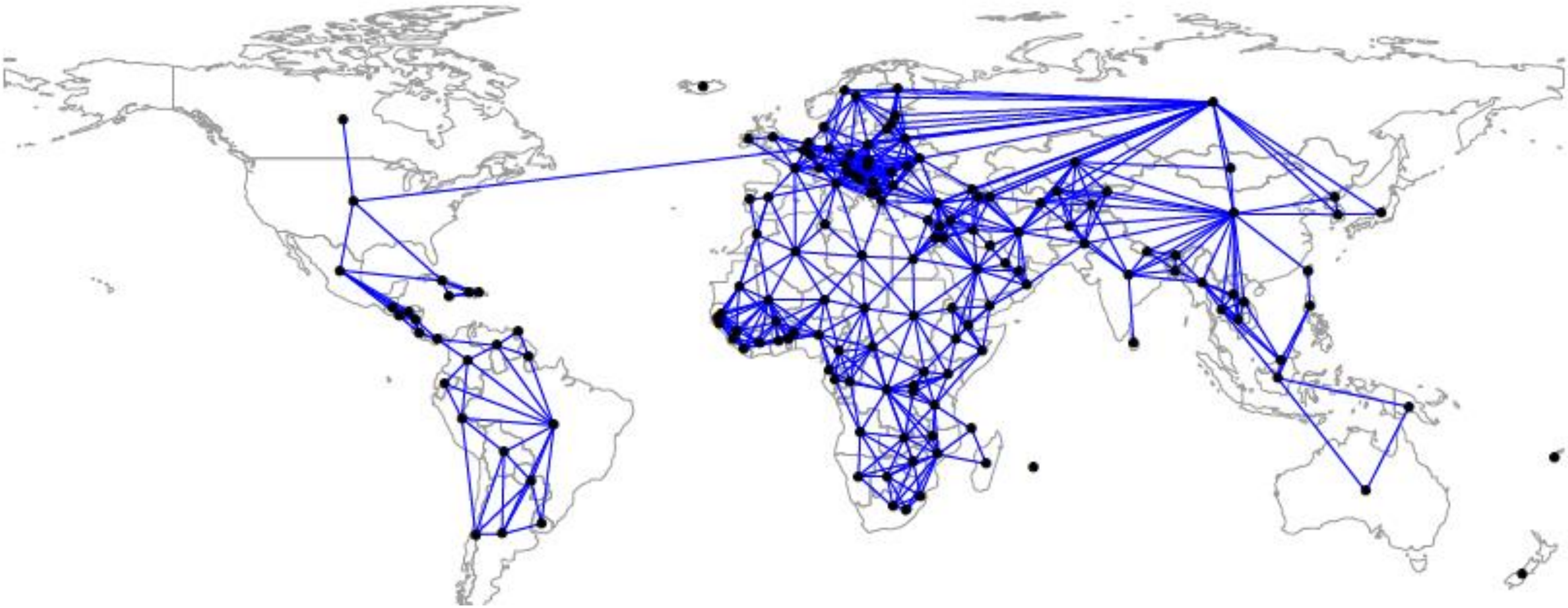
- Bishop contiguity (common vertex)

- $$\boldsymbol{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

1	2	3
4	5	6
7	8	9

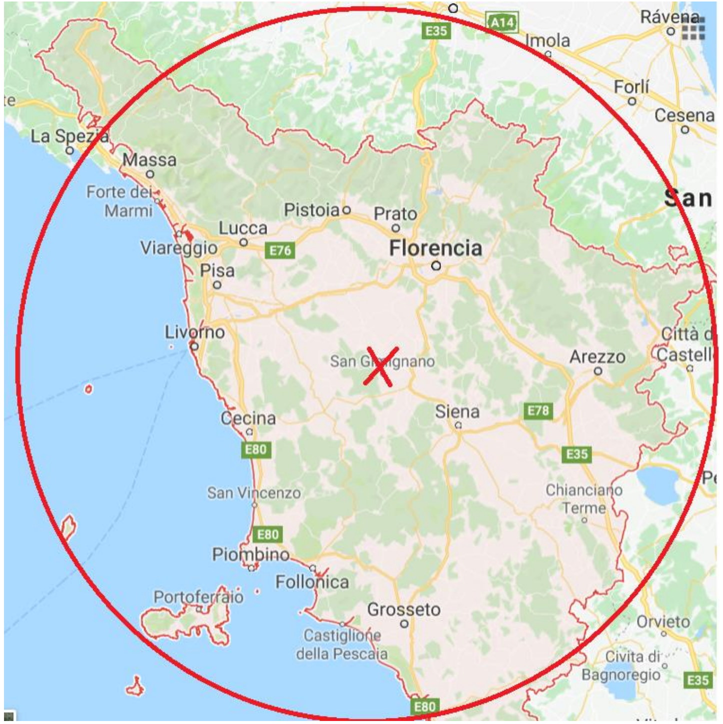
Spatial Weight Matrix based on boundary links

- Contiguity neighbors World **W** (with 500km of snap distance)



Spatial Weight Matrix based on distance links

- Weights may be also defined as a function of the distance between region *i* and *j*, d_{ij}
- d_{ij} is usually computed as the distance between their centroids or other important unit
- Example of centroid:



Spatial Weight Matrix based on distance links

Let x_i and x_j be the longitude and y_i and y_j the latitude coordinates for region i and j respectively:

- Minkowski metric of distance

$$d_{ij}^{p} = \left(\left|x_i - x_j\right|^p + \left|y_i - y_j\right|^p\right)^{1/p}$$

- Euclidean metric of distance (p=2)

$$d_{ij}^{e} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

- Manhattan metric of distance (p=1)

$$d_{ij}^{m} = \left|x_i - x_j\right| \quad + \left|y_i - y_j\right|$$

Spatial Weight Matrix based on distance links

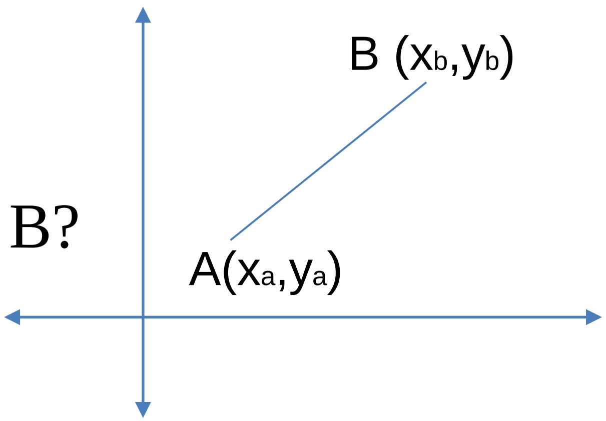
- **Euclidean distance** is the shortest distance in a plane

Numerical example:

A(1,1) and B(3,4)

What is the Euclidean distance between A and B?

Plug the numbers in:



$$d^e_{AB} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

$$d^e_{ij} = \sqrt{(1-3)^2 + (1-4)^2} = \sqrt{4+9} = 3.605$$

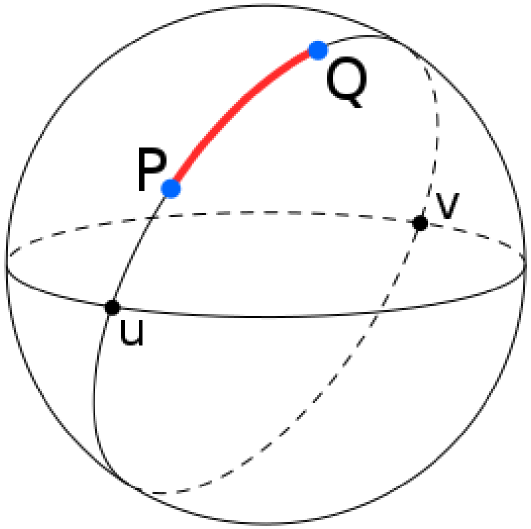
but not necessarily the shortest if you take into account the curvature of the earth!

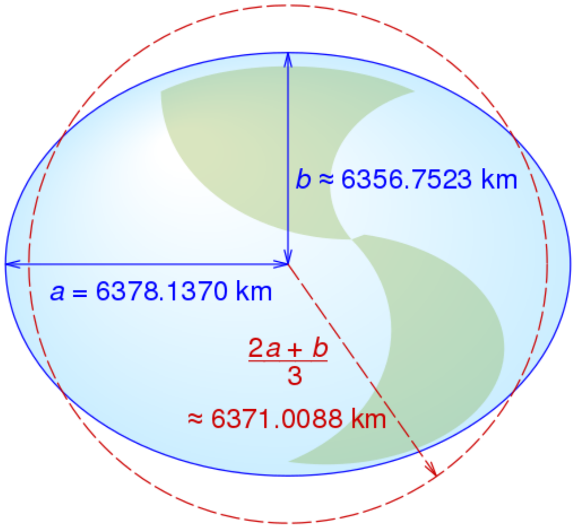
Spatial Weight Matrix based on distance links

- **Great circle distance**
  - Is the shortest distance between two points in a sphere
  - Let x_i and x_j be the longitude and y_i and y_j the latitude coordinates for region i and j respectively:

$$d_{AB}^{e} = r\ x\ arccos^{-1}\left[cos\left|x_i - x_j\right| cosy_i cosy_j + siny_i siny_j\right]$$

Where r is the Earth radius (r = 6371 km)





Spatial Weight Matrix based on distance links

- **K-nearest neighbors**

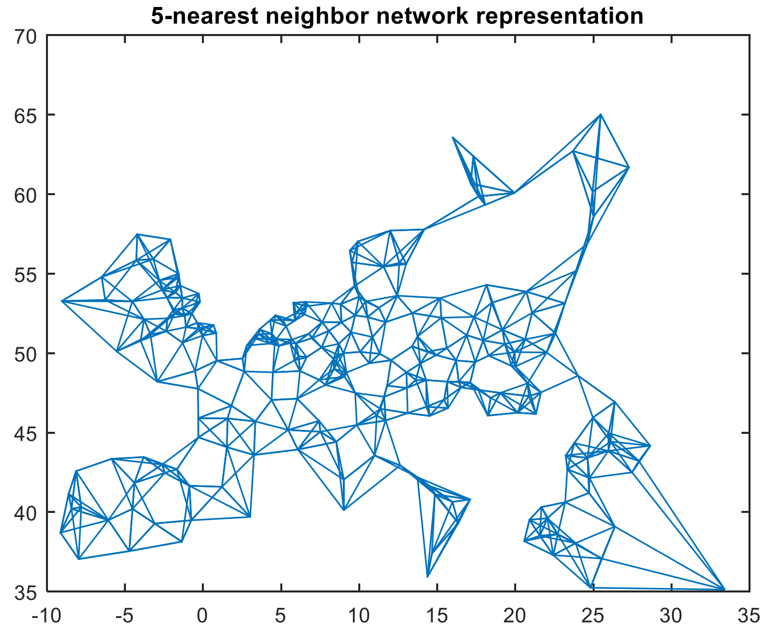
Definition:

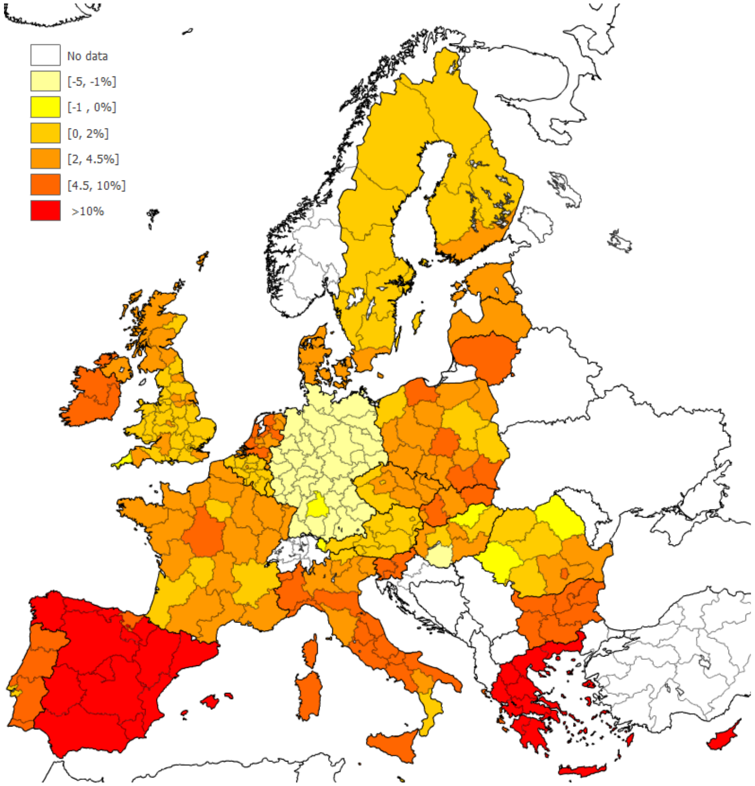
Let centroid distances from each spatial unit *i* to all units $j \neq i$ be ranked as follows: $d_{ij(1)} < d_{ij(2)} < \cdots < d_{ij(n-1)}$. Then for each $k = 1, \ldots, n-1$, the set $N_k(i) \in \{j(1), j(2), \ldots, j(k)\}$ contains the *k* closest units to *i* (where for simplicity we ignore ties). For a given *k* , the *k*-nearest neighbor weight matrix, **W**, then has spatial weights of the form:

$$w_{ij} = \begin{cases} 1 & , \quad j \in N_k(i) \\ 0 & , \quad otherwise \end{cases}$$

Spatial Weight Matrix based on distance links

- **5-nearest neighbor representation of NUTs-2 European regions**





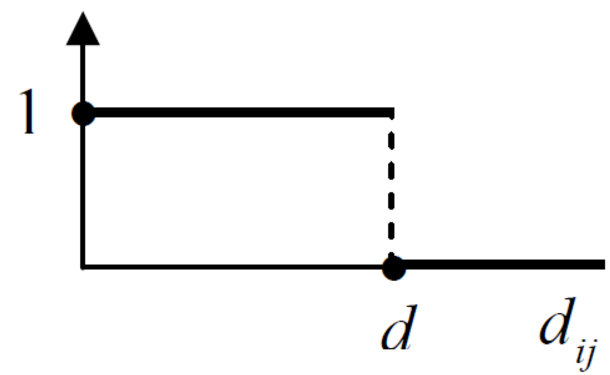
Spatial Weight Matrix based on distance links

- **Radial distance/Distance band matrices**

Definition: If distance itself is an important criterion of spatial influence, and if d denotes a threshold distance (*or bandwidth*) beyond which there is no direct spatial influence between spatial units, then the corresponding radial distance weight matrix, **W** , has spatial weights of the form:

$$w_{ij} = \begin{cases} 1 & , \quad 0 \leq d_{ij} \leq d \\ 0 & , \quad d_{ij} > d \end{cases}$$

for $i \neq j$.



Typically the threshold d is given by some statistic of the distribution of distances in the sample (i.e, median, first quartile, etc)

Spatial Weight Matrix based on distance links

- **Inverse power distance matrices**

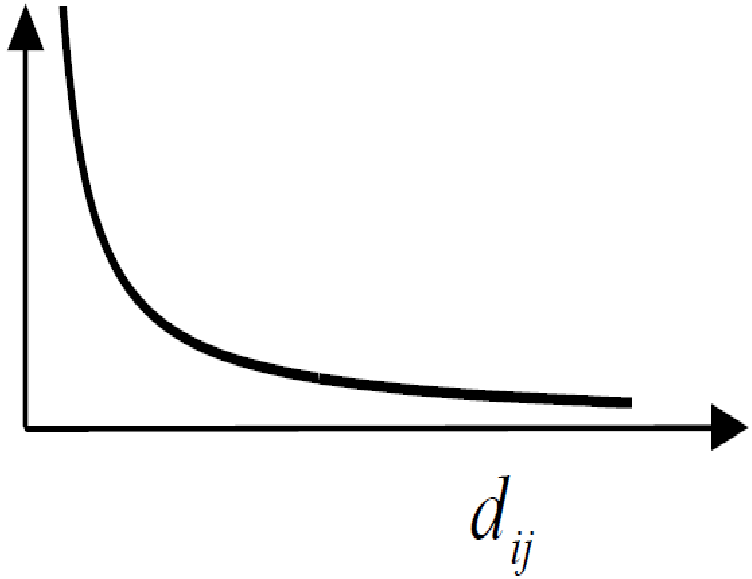
Definition: The **strength of interactions decreases as distance increases**. If there are believed to be diminishing effects, then one standard approach is to assume that **weights are a negative power function of distance** of the form:

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

for $i \neq j$.

Typically, $\alpha = 1$ or $\alpha = 2$ (gravity inspiration)

The higher is α the faster the decrease



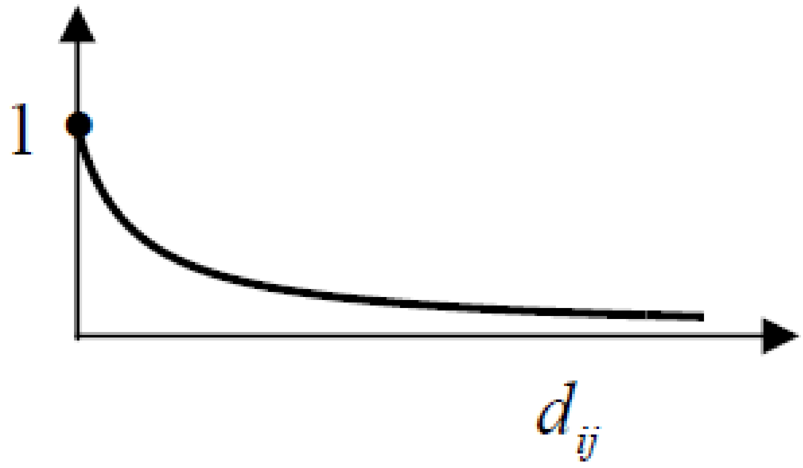
Spatial Weight Matrix based on distance links

- **Negative exponential**
- **Definition:** An alternative to negative power functions are *negative exponential functions* of distance of the form:

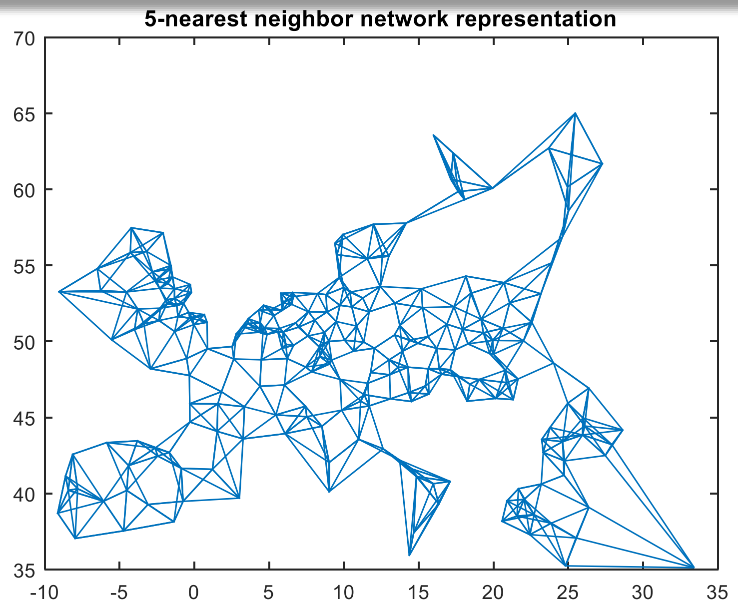
$$w_{ij} = \exp(-\alpha d_{ij})$$

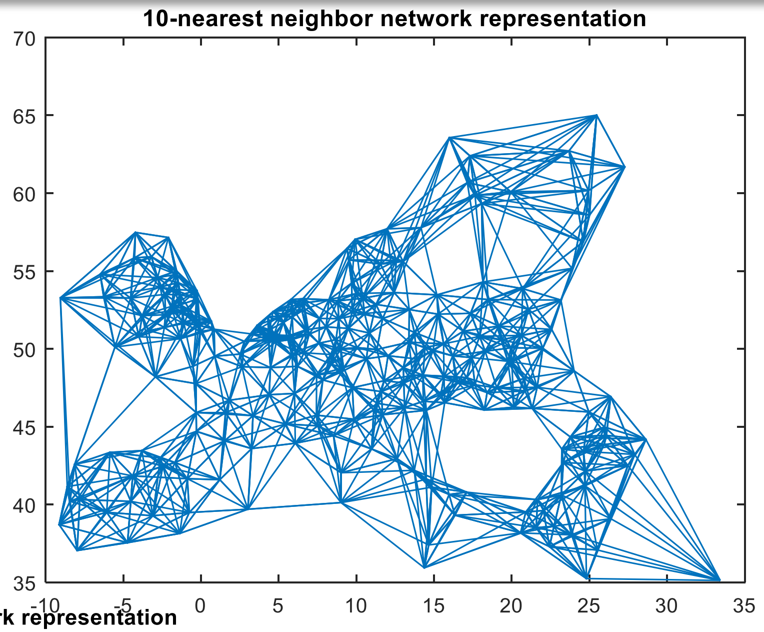
for $i \neq j$.

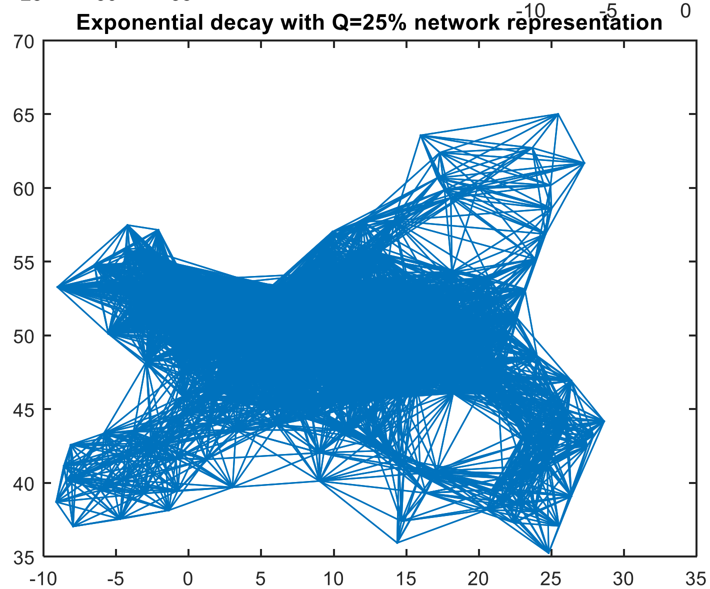
Typically, $\alpha = 1\%$ up to $\alpha = 10\% \rightarrow$ faster decrease than power inverse



Spatial Weight Matrix based on distance links







W normalization

- The row-standardized matrix is also known in the literature as the **row-stochastic matrix**

- Definition: A real n x n matrix **W** is called **Markov** matrix or row-stochastic matrix if:

  $w_{ij} \geq 0$ for $1 \leq \mathrm{i}, \mathrm{j} \leq \mathrm{n}$

  $\sum_{j}^{N} w_{ij} = 1$ for $1 \leq \mathrm{i} \leq \mathrm{n}$

W normalization

- **W's are used to compute weighted averages** in which more weight is placed on nearby observations than on distant observations

- The elements of a row-standardized weight matrix equal:

$$w_{ij}^{s} = \frac{w_{ij}}{\sum_{j}^{N} w_{ij}}$$

- This ensures that **all weights are between 0 and 1** and facilitates the interpretation of operation with weight matrix as an averaging of neighboring values

- Under row-normalization, **the sum of elements of each row add up to 1**

- The row-standardized weight matrix also **ensures that the spatial parameter in many spatial stochastic processes are comparable** between each other

- **It is relative and not absolute distance what matters**

They are **not longer symmetric**

Testing Spatial Autocorrelation

- **Moran's I**

This statistic is given by:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S \sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where $S = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ and w_{ij} is an element of the spatial weight matrix that measures spatial distance or connectivity between regions i and j.

In matrix form:

$$I = \frac{n}{S} \frac{z'Wz}{z'z}$$

where $z = \mathrm{x} - \bar{x}$. If the W matrix is row-standardized, then:

$$I = \frac{z'Wz}{z'z}$$

because S=n.

Values range from -1 (perfect dispersion) to +1 (perfect correlation). A zero value indicates a random spatial pattern

Testing Spatial Autocorrelation

Note that:

$$\hat{\beta}^{OLS} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2}$$

Therefore

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\ (x_i - \bar{x})(x_j - \bar{x})w_{ij}}{S\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

I, is equivalent to the slope of coefficient of a linear regression of the spatial Wx on the observation vector x, measured in deviation from their means.

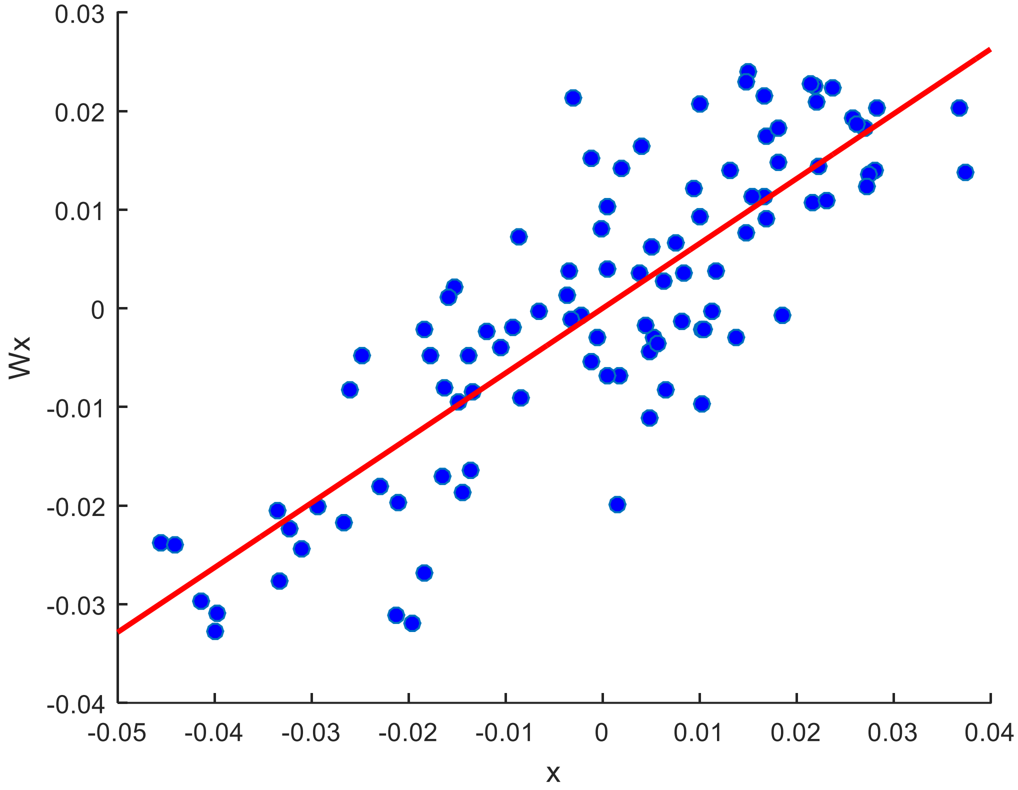
$$\widetilde{WX} = \beta_1\tilde{X}$$

where the tilde represents the variable is in deviation from the mean

Testing Spatial Autocorrelation

$$x = \rho W x + \epsilon \rightarrow x = (I - \rho W)^{-1} \epsilon$$

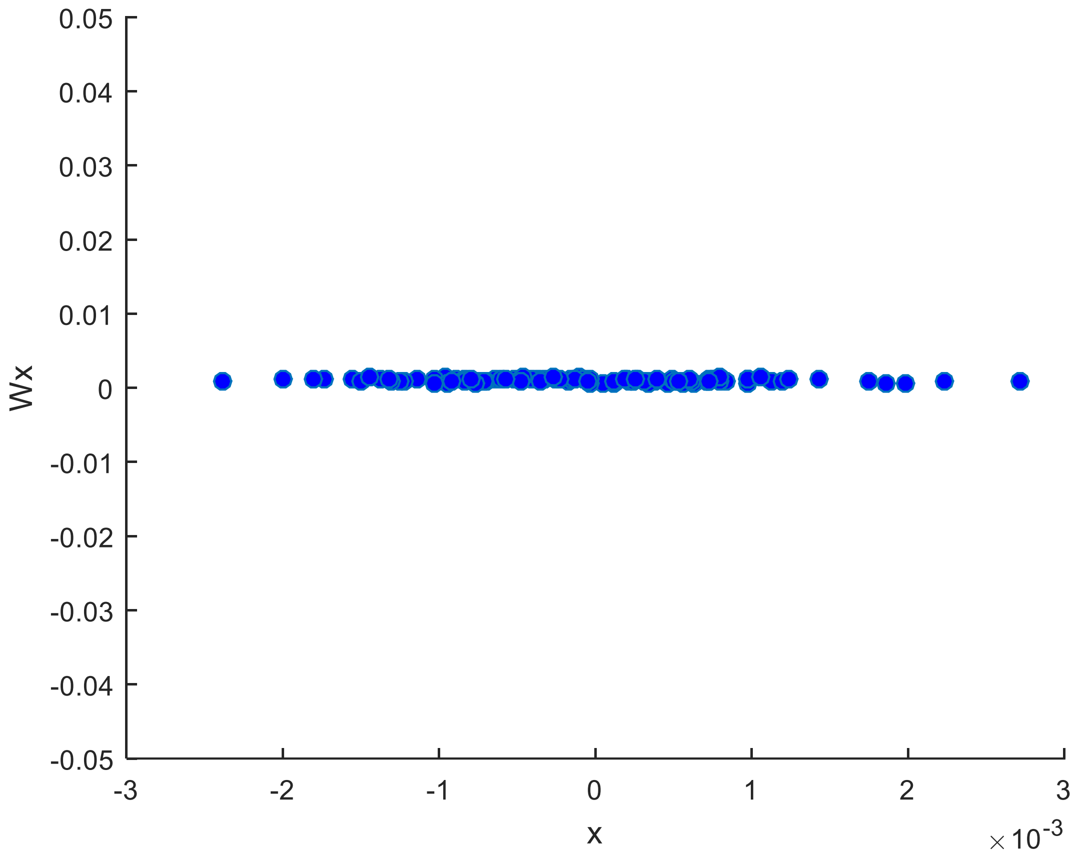
Example with $\rho = 0.99$



Testing Spatial Autocorrelation

$$x = \rho W x + \epsilon \rightarrow x = (I - \rho W)^{-1}\epsilon$$

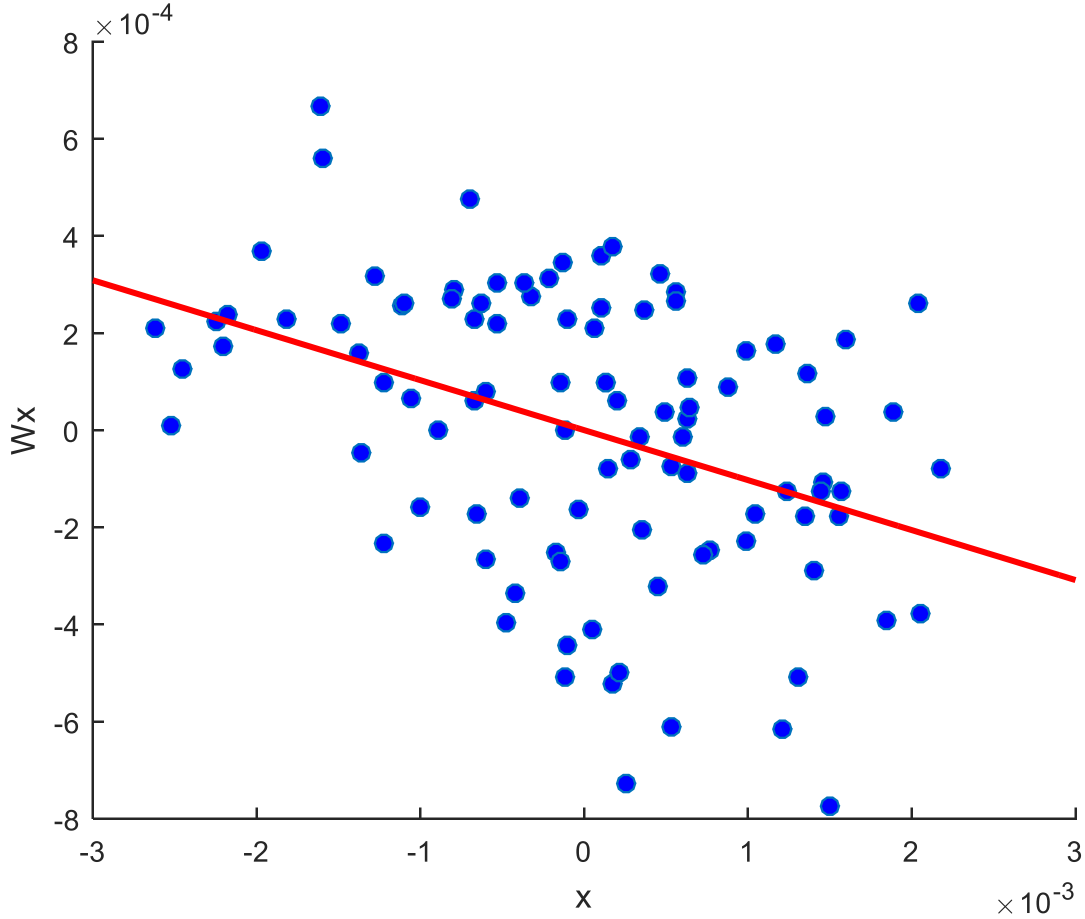
Example with $\rho = 0.00$



Testing Spatial Autocorrelation

$$x = \rho W x + \epsilon \rightarrow x = (I - \rho W)^{-1} \epsilon$$

Example with $\rho = -0.99$



Local Indicators of Spatial Association (LISA)

- **Local Moran statistic** *I(i)* detects local spatial autocorrelation. The *I(i)*'s are indicators of local instability. They decompose the Moran's I into contributions for each location.

According to this property, **Local Moran's I** can be used for two purposes:

  - Indicators of local spatial clusters
  - Diagnostics for outliers in global spatial patterns

- **Local Moran's I statistic**:

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2}$$

Numerator: determines the sign of *I(i)*

+ if both the *i-th* region and neighbors have above or below average values in the geo-referenced variable X and

(-) if the *i-th* region as an above(below) neighboring regions have a below(above) average values of *X*

Denominator: standardization of the cross-product by the variance of the georeferenced variable *X*

Local Indicators of Spatial Association (LISA)

- **Local Moran's I Map**