Spatial econometrics (1)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

1. Introduction

- Hans Koster → URBANECONOMICS.NL
  - Professor of Urban Economics and Real Estate
  - h.koster@vu.nl
  - Lectures
  - Programme director of STREEM (streem.sbe@vu.nl)

- Canvas
  - Be aware of the recap materials under *modules*
  - If there is anything unclear, let me know!

1. Introduction

- **A few announcements on the programme**
  - **Enrol with the following link to the MSc STREEM announcement page https://canvas.vu.nl/enroll/FXW8BL**
  - **Thesis support sessions: October 4 and 11, 12:45-13:30**
  - **Drinks on October 28, 15:00, after exam**

1. Introduction

- This course
  - Learn about advanced tools and techniques important for
    - » Urban & Regional,
    - » Real Estate,
    - » Transport and
    - » Environmental Economics

- Do not hesitate to ask questions during the class!

- Notation on slides
  - Most important concept are underlined
  - Questions (via Menti), exercises and applications
    → On red slides

VU VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

- **How to study:**
  → Attend offline lectures and tutorials
  → Work together on assignments on campus
  → Read and study syllabus

1. Introduction

- Tutorials
  - Thomas de Graaff (t.de.graaff@vu.nl, thomasdegraaff.nl)
  - For any questions on the assignments/
  - tutorials please ask Thomas

- Please subscribe to groups on Canvas *before Tuesday 23:59*
  - Groups of 3

1. Introduction

- This week
  - Learn about how to deal with spatial data
  - … and spatial econometrics

- Plan:

Lecture #1:	Spatial data
Lecture #2:	Spatial autocorrelation and regressions
Lecture #3:	Spatial regressions (cont'd)
Assignment:	Open space and school quality

- **Some remarks on matrix notation**
  - **Use bold symbols for vectors**

$$\boldsymbol{x} = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix}$$

  - **Use bold symbols and capitals for matrices**

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

  - **Identity matrix**

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

$\rightarrow \boldsymbol{IX} = \boldsymbol{X}$

  - **Inverse $\boldsymbol{X}^{-1}$ is matrix equivalent of $1/x$**

$\rightarrow \boldsymbol{X}^{-1}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{I}$

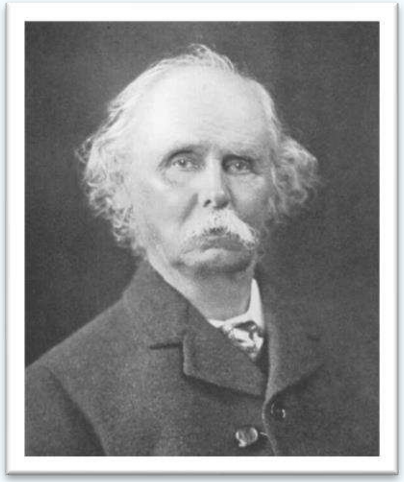
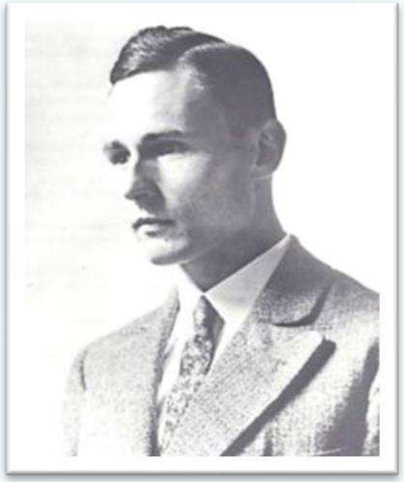
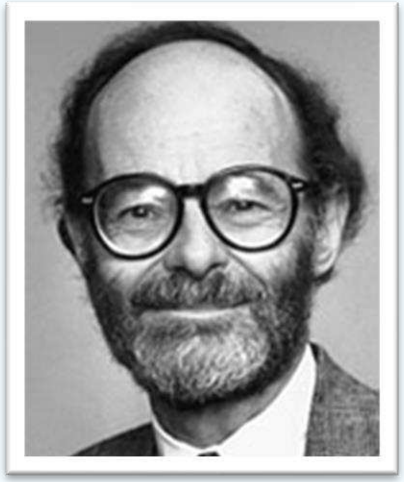
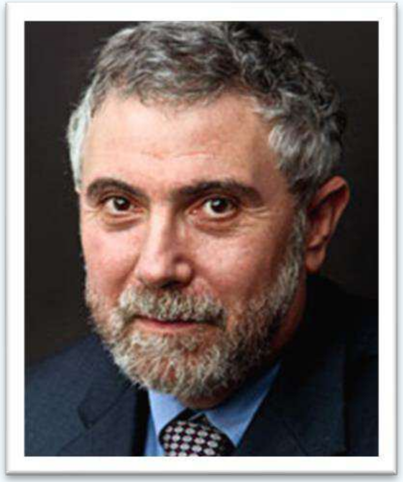
  - **More details in the appendix of the syllabus**

VRIJE UNIVERSITEIT AMSTERDAM

2. Space in economics

- Many economic processes are spatially correlated
  - Tobler's first law of geography
- Most economics models are "topologically invariant"
- New economic fields have emerged
  - Urban economics
  - New economic geography (NEG)
- Synergy with other fields
  - Economic geography
  - Regional science
  - GIS

- **Economists and space**


YOU?

- **Spatial econometrics**

- **40-50s – mainly domain of statisticians**

- **Cliff and Ord (1973): "Spatial autocorrelation"**

- **Paelinck and Klaassen (1979): "Spatial Econometrics"**

- **Rapid growth since Anselin (1988)**

- **New estimators, tests and interpretation**
  - ***e.g.*** **Kelejian and Prucha (1998, 1999, 2004, 2007, 2010)**

2. Space in economics

- **Spatial modelling is becoming increasingly important**
  - **New and geo-referenced data**
  - **Advanced software**
  - ***New methods and regression techniques!***

VRIJE UNIVERSITEIT AMSTERDAM

3. Spatial data structure

- **Time is simple**
  - **Natural origin**
  - **No reciprocity**
  - **Unidirectional**

$$x_{t-3} \longrightarrow x_{t-2} \longrightarrow x_{t-1} \longrightarrow x_t$$

- **Linear space (*e.g. beach*) is different**
  - **No natural origin**
  - **Reciprocity**
  - **Unidirectional**

$$x_1 \longleftrightarrow x_2 \longleftrightarrow x_3 \longleftrightarrow x_4$$

3. Spatial data structure

- **Two-dimensional space becomes even more complex**
  - **No natural origin**
  - **Reciprocity**
  - **Multidirectional**

$$\begin{tikzcd}
& x_1 \arrow[dl, leftrightarrow] \arrow[dr, leftrightarrow] & \\
x_3 \arrow[rr, leftrightarrow] & & x_2
\end{tikzcd}$$

- $i = 1,2,3$ **can refer to point data, areas, grids**

3. Spatial data structure

- First, we have to define the spatial structure of the data
- Specified through a spatial weights matrix
- Spatial weights matrix W:
  - Consists of $n \times n$ elements
  - Discrete or continuous elements
- How to define weights?
  - Euclidian distance
  - Network distance
  - Spatial interactions
  - Social networks

3. Spatial data structure

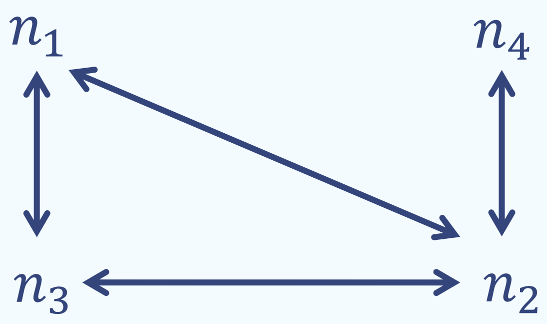
- **How to define spatial matrices?**

- **Contiguity matrix**
  - Adjacent → 1st order contiguous
  - Neighbours of neighbours → 2nd order contiguous

- **Distance matrix**
  - k-nearest neighbours
  - Inverse distance weights ($1/distance$)
  - Cut-off distance

3. Spatial data structure

- **Let's provide an example of a contiguity matrix**



to (columns), from (rows)

W	n_1	n_2	n_3	n_4
n_1	0	1	1	0
n_2	1	0	1	1
n_3	1	1	0	0
n_4	0	1	0	0

VRIJE UNIVERSITEIT AMSTERDAM

- **Matrices can be standardised**
  - **Different principles can be used**
  - **Most common: *row-standardisation*:**

    $$w_{ij}^{*} = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}}$$

    where k are other locations

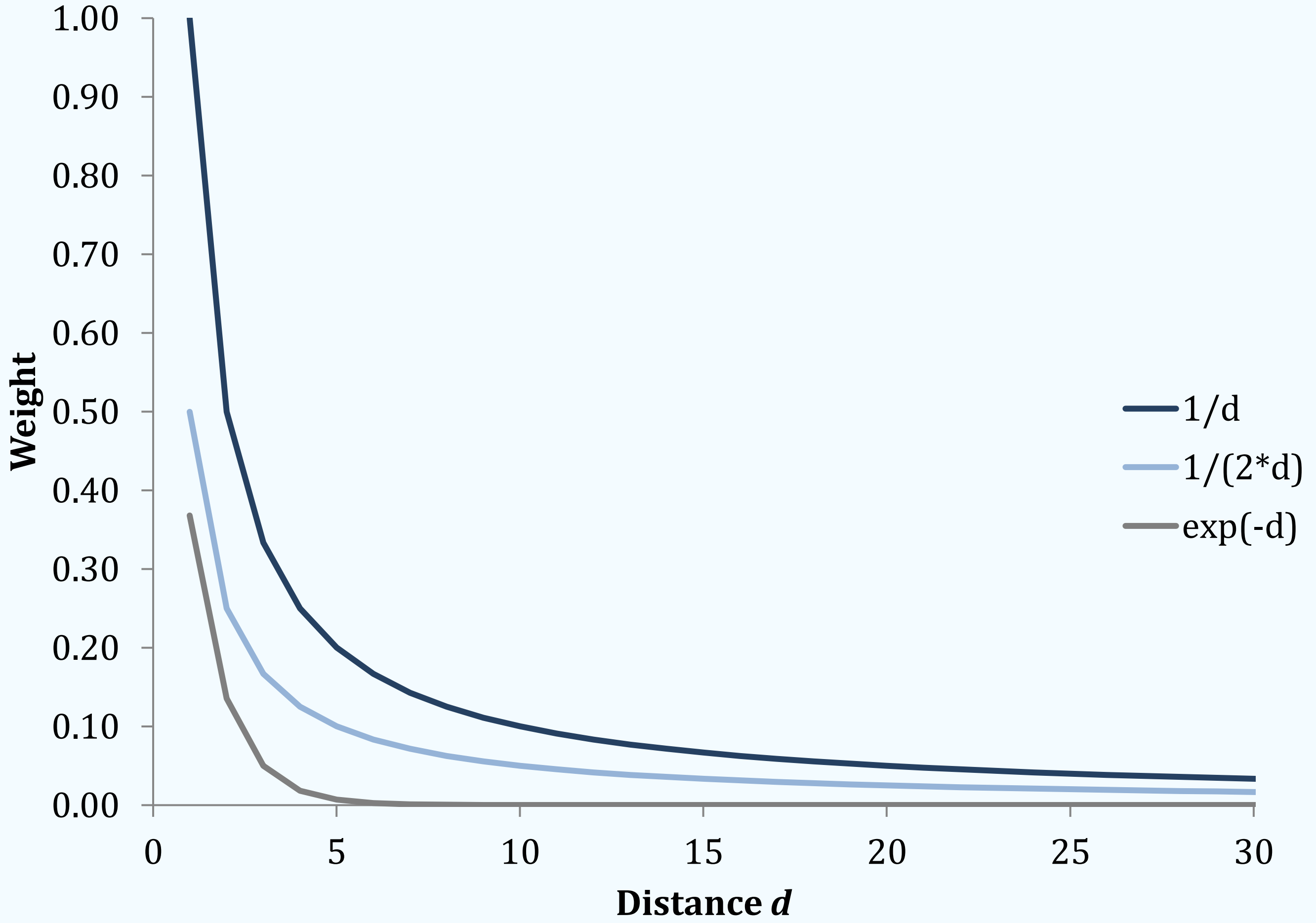
- **Interpretation of**
  - $\sum_{j=1}^{n} w_{ij}$**: sum of connections to neighbours**
  - w_{ij}^{*} **denotes the share of connections to neighbours**

3. Spatial data structure

- Remarks regarding distance weight matrices
  - Check for exogeneity of matrix
  - Connectivity
  - Symmetry
  - Standardisation
  - Distance decay

VU VRIJE UNIVERSITEIT AMSTERDAM

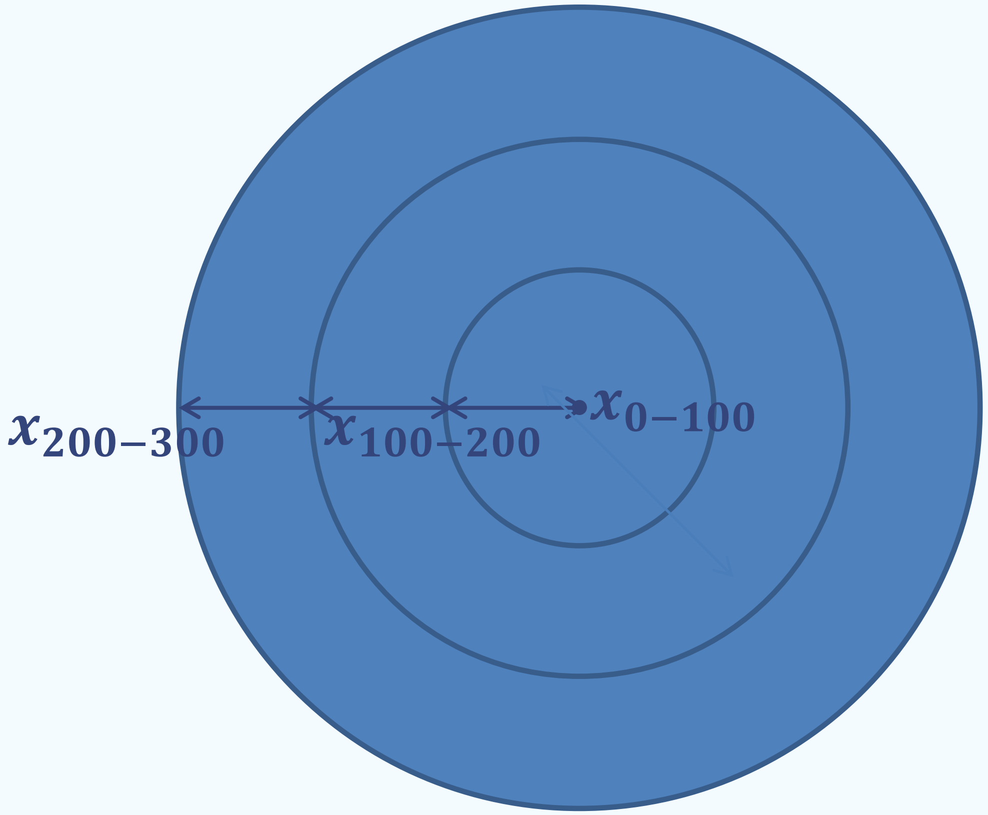
- **Choice of distance decay is arbitrary**
  - Sometimes theory may help
  - May also try to find the optimal decay parameter empirically



3. Spatial data structure

- Choice of distance decay is arbitrary
  - An alternative is to forget about specifying W
  - Alternatively, use different x-variables capturing concentric rings
  - Average of x-variable for different distance bands

- **Choice of distance decay is arbitrary**
  - **e.g.** $y = \alpha x_{0-100} + \beta x_{100-200} + \gamma x_{200-300} + \epsilon$



3. Spatial data structure

- **How to define spatial weight matrix using software**
  - SPATWMAT **in STATA, based on geographic coordinates**
  - SPWEIGHT **in STATA**
  - **Geoda**
  - SPATIAL STATISTICS TOOLBOX **in ArcGIS**
  - SPDEP **in *R***

- **Concentric rings should be calculated manually**

4. Modifiable areal unit problem

- Usually we do not have space-continuous data
  - ‘Dots’ to ‘boxes’

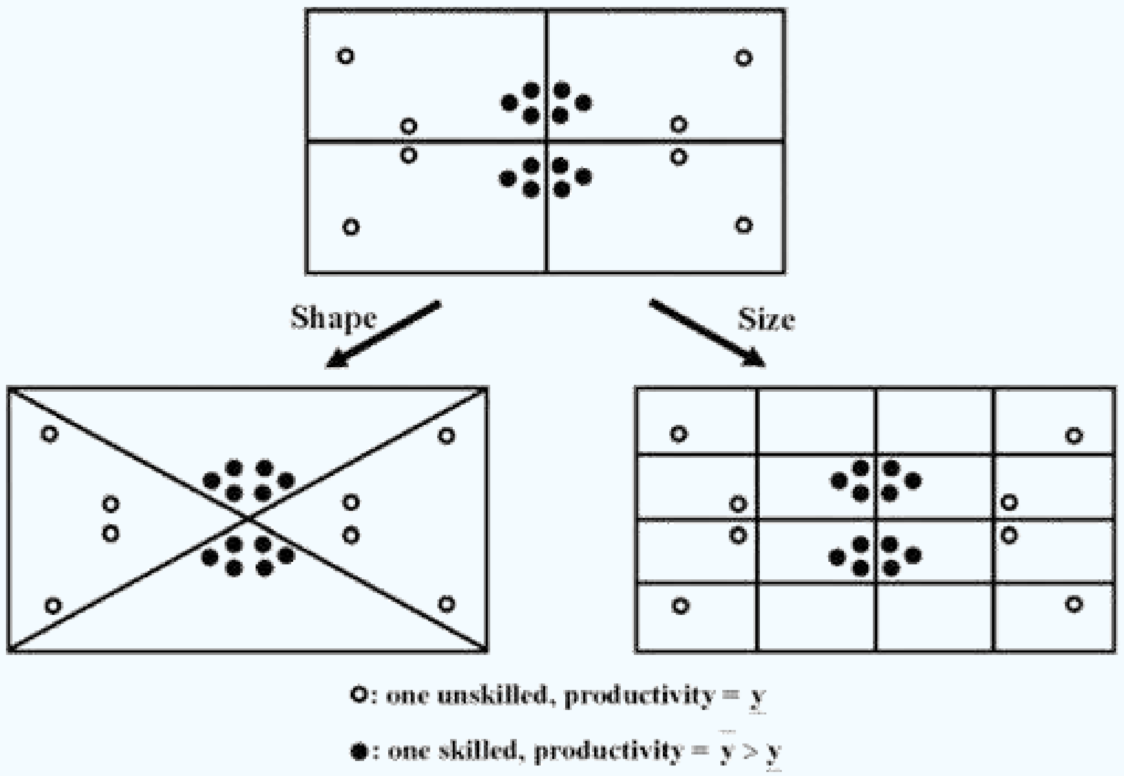
- Data is aggregated at
  - Postcode areas
  - Municipalities
  - Regions
  - Countries

- Problems:
  - Aggregation is often arbitrary
  - Areas are not of the same size

- This may lead to distortions
  - Modifiable areal unit problem (MAUP)

4. Modifiable areal unit problem

- **An illustration:**



Briant, Combes and Lafourcade (2010, JUE)

- **Aggregation seems to be important!**

4. Modifiable areal unit problem

- **Briant et al. (2010) investigate whether choice matters for regression results**

341 Employment Areas (EA)

341 Small squares (SS)

21 Régions (RE)

22 Large squares (LS)

4. Modifiable areal unit problem

- MAUP is of secondary importance
  - If y and x are aggregated in the same way
  - Matters more for larger areas (*e.g.* regions)
  - Use meaningful areas if possible

- Specification issues are much more important

5. Summary

Today:

- 'Space' in economics is becoming more and more important
- Incorporating space in econometric applications is not straightforward
- Important to define the spatial structure of the data
    - Spatial weight matrices
    - Modifiable areal unit problem

VRIJE UNIVERSITEIT AMSTERDAM

5. Summary

Tomorrow:

- Spatial autocorrelation
- Spatial regressions

→ Subscribe to assignment groups before *Tuesday, 23:59*.

Spatial econometrics (1)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

Spatial econometrics (2)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate
VRIJE UNIVERSITEIT AMSTERDAM
tinbergen institute

- **This week**
  - **Learn about how to deal with spatial data**
  - **… and spatial econometrics**

- **Plan:**

Lecture #1:	Spatial data
Lecture #2:	Spatial autocorrelation and regressions
Lecture #3:	Spatial regressions (cont'd)

2. Spatial autocorrelation

- **Spatial autocorrelation between values**
  - **Implies** $\mathrm{cov}(x_i, x_j) = \mathrm{E}[x_i x_j] - \mathrm{E}[x_i] \cdot \mathrm{E}[x_j] \neq 0$
  - **Again,** j **refers to other locations**

- **Spatial autocorrelation, dependence, clustering**
  - **Fuzzy definitions in literature**

VRIJE UNIVERSITEIT AMSTERDAM

2. Spatial autocorrelation

- **How to measure spatial autocorrelation**
  - **Moran's *I***
  - **Focus on one variable x (e.g. crime)**

- **H_0: independence, spatial randomness**
- **H_A: dependence**
  - **On the basis of adjacency, distance, hierarchy**

- **Moran's *I* is given by:**

$$I = \frac{R}{S_0} \times \frac{\tilde{\boldsymbol{x}}' \boldsymbol{W} \tilde{\boldsymbol{x}}}{\tilde{\boldsymbol{x}}' \tilde{\boldsymbol{x}}} \tag{4}$$

where R is the number of spatial units

S_0 is the sum of all elements of the spatial weight matrix

$\boldsymbol{W}$ is the spatial weight matrix

$\tilde{\boldsymbol{x}} = x - \bar{x}$ is a vector with the variable of interest

VU VRIJE UNIVERSITEIT AMSTERDAM

2. Spatial autocorrelation

- **Moran's *I***
- **Use row-standardised spatial weight matrix W!**
- **Recall that** $I_S = \frac{\tilde{x}' W \tilde{x}}{\tilde{x}' \tilde{x}}$ **(standardised** I**)**
  - **Note similarity with OLS:** $\hat{\beta} = \frac{x' y}{x' x}$
  - **Hence:** $W\tilde{x} = \alpha + I\tilde{x} + \epsilon$, **where** $\alpha = 0$
- **Moran's *I* is correlation coefficient (more or less)**
  - $\approx [-1,1]$
  - **But: expectation** $\mathrm{E}[I] = -\frac{1}{N-1}$
- **Visualisation**
  - **Moran scatterplot**

VU VRIJE UNIVERSITEIT AMSTERDAM

2. Spatial autocorrelation

- **Moran's I**
- **How to investigate the statistical significance of (4)?**
  - $\frac{I - \mathrm{E}[I]}{\sqrt{\mathrm{var}[I]}}$ **(5)**
  - **However, $\sqrt{\mathrm{var}[I]}$ is difficult to derive**
  - $\mathrm{E}[I] = -1/(n-1)$
  - **Assume normal distribution of I to approximate $\sqrt{\mathrm{var}[I]}$ under H_0**
  - **Or: bootstrapping/simulation**

  - **See Cliff and Ord (1973) for more details**

2. Spatial autocorrelation

- **Moran's *I***

- **Also possible: correlation to other variables:**

$$I_S = \frac{\tilde{\boldsymbol{x}}' \boldsymbol{W} \tilde{\boldsymbol{z}}}{\tilde{\boldsymbol{x}}' \tilde{\boldsymbol{x}}}$$

VRIJE UNIVERSITEIT AMSTERDAM

2. Spatial autocorrelation

- **How to calculate Moran's *I* using software**
  - SPAUTOC **in STATA**
  - SPLAGVAR **in STATA**
  - SPATIAL STATISTICS TOOLBOX **in ArcGIS**

- **Alternative: Getis and Ord's *G***
  - **Most of the time only Moran's *I* is reported**

2. Spatial autocorrelation

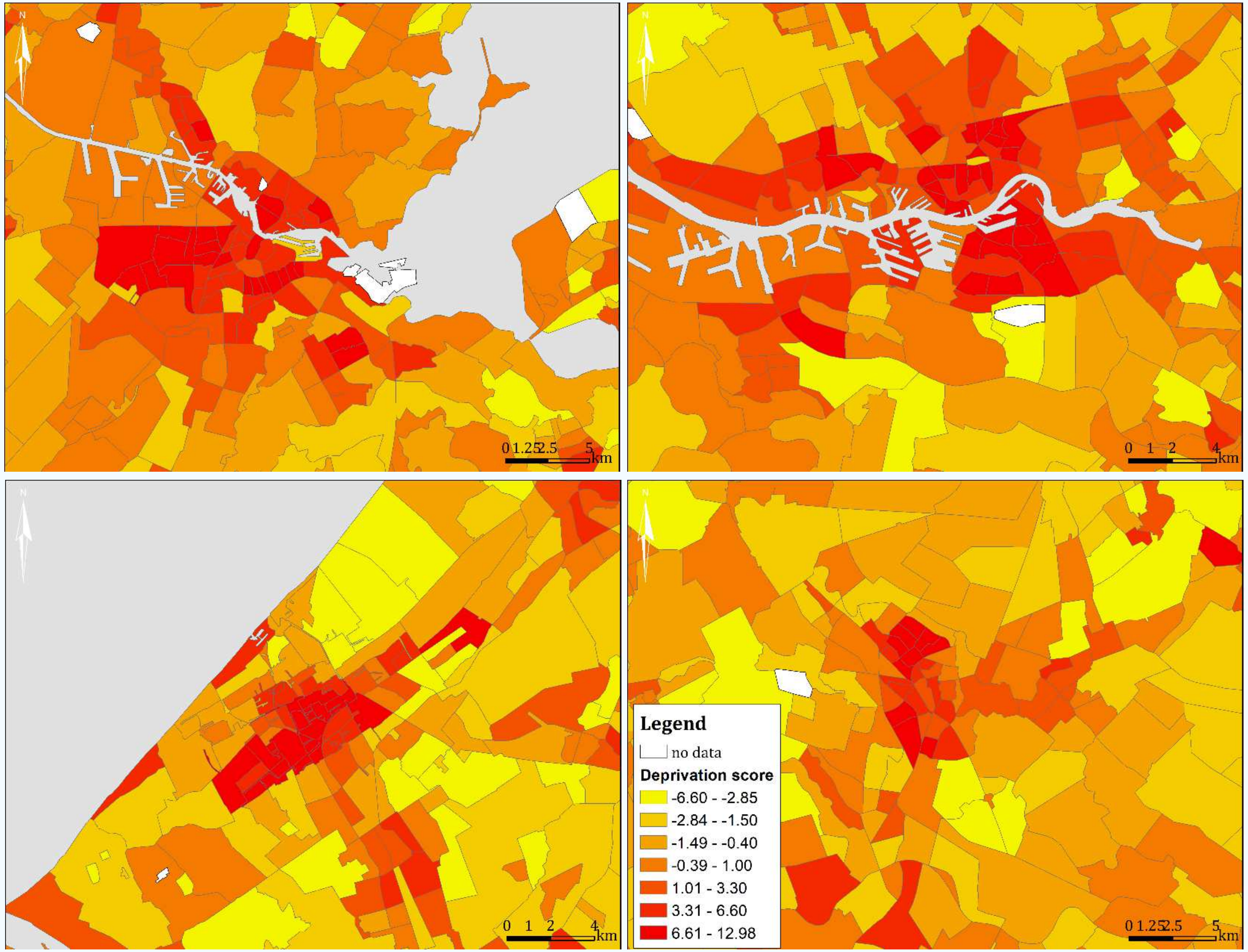
- **Let's try to answer the queston:**
  "Is social deprivation spatially clustered?"

- **How to determine the most deprived neighbourhoods?**

- **Dutch government calculated deprivation *z*-score for each neighbourhood**
  - **Based on housing quality, safety, perception and satisfaction**
  - ***Important:* the 83 most deprived neighbourhoods were selected for an investment of >€1 billion**

2. Spatial autocorrelation

Amsterdam **Rotterdam**



The Hague **Utrecht**

- **Determine spatial autocorrelation**
  1. **Determine distance between all neighbourhoods using centroids**
  2. **Use inverse distance function $w_{ij} = 1/\left(d_{ij}^{\gamma}\right)$ to determine spatial weights in weight matrix**
  3. **Calculate Moran's I: $\mathbf{W}\tilde{\mathbf{z}} = \alpha + I\tilde{\mathbf{z}} + \epsilon$ where $\tilde{\mathbf{z}} = \mathbf{z} - \bar{z}$ and W is a row-standardised weight matrix**
     - ***Note that $\mathbf{W}\tilde{\mathbf{z}}$ is a vector***
  4. **Bootstrap this procedure to estimate standard error (or use software)**

- **Determine spatial autocorrelation. *Note that***
  - **$\mathbf{W\tilde{z}}$ is a *vector***

$$\mathbf{W} \qquad\qquad \times \quad \tilde{\mathbf{z}} \quad = \mathbf{W\tilde{z}}$$

$$\begin{bmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{bmatrix} \times \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \tilde{z}_3 \end{bmatrix} = \begin{bmatrix} \cdots \\ \cdots \\ \cdots \end{bmatrix}$$

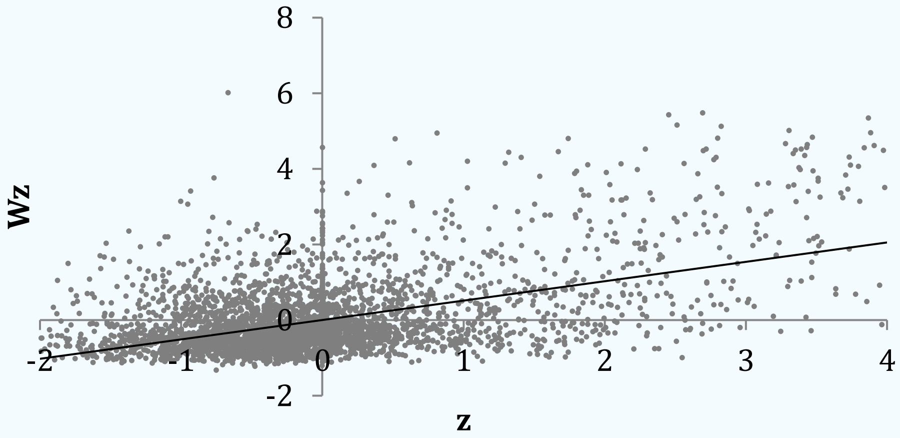
  - **OLS: $\boldsymbol{\beta} = \frac{x'y}{x'x}$, while $I = \frac{\tilde{\mathbf{z}}'W\tilde{\mathbf{z}}}{\tilde{\mathbf{z}}'\tilde{\mathbf{z}}}$**
  - **Notation: $\frac{x'y}{x'x} = x^T y \left(x^T x\right)^{-1}$**

- Determine spatial autocorrelation
  1. Determine distance between all neighbourhoods using centroids
  2. Use inverse distance function $w_{ij} = 1/\left(d_{ij}^{\gamma}\right)$ to determine spatial weights in weight matrix
  3. Calculate Moran's I: $\mathrm{W}\tilde{\mathbf{z}} = \alpha + I\tilde{\mathbf{z}} + \epsilon$ where $\tilde{\boldsymbol{z}} = \boldsymbol{z} - \bar{z}$ and W is a row-standardised weight matrix
     - *Note that* $\mathrm{W}\tilde{\mathbf{z}}$ *is a vector*
  4. Bootstrap this procedure to estimate standard error (or use software)

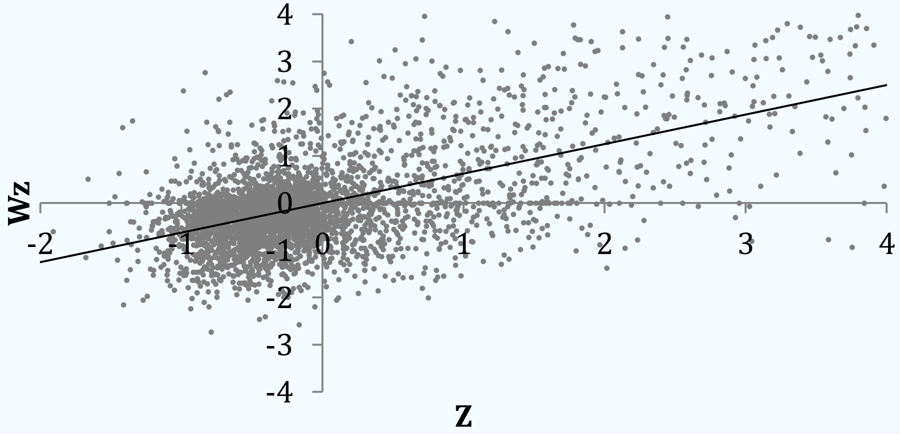
VRIJE UNIVERSITEIT AMSTERDAM

- **Calculate Moran's I**
  - **Using inverse distance function** $w_{ij} = \frac{1}{d_{ij}^{\gamma}}$

I = 0.513* *(γ=1)***



I = 0.625* *(γ=2)***



2. Spatial autocorrelation

- **Spatial correlation in deprivation**
  - **Local phenomenon?**
  - **You do not know *why* scores are autocorrelated…**
  - **No causal relationships!**

VRIJE UNIVERSITEIT AMSTERDAM

2. Spatial autocorrelation

- It is important to make a distinction between *global* and *local* spatial autocorrelation
  - See Anselin (2003) for a discussion

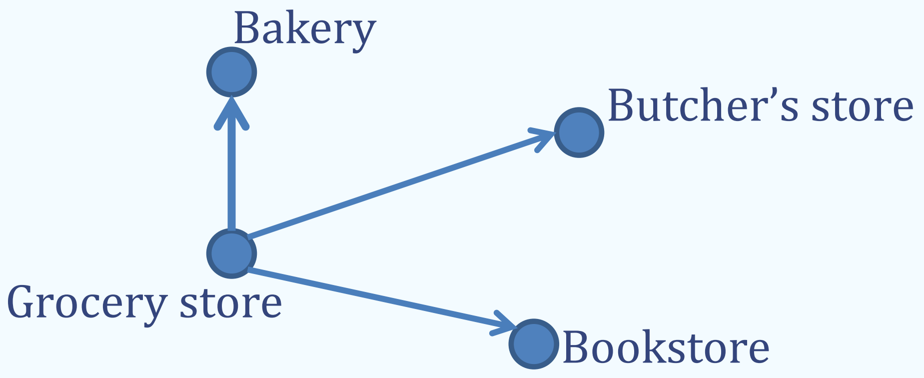
- Global spatial autocorrelation
  - Local shock affects the whole system

- Local spatial autocorrelation
  - Local shock only affects the 'neighbours'

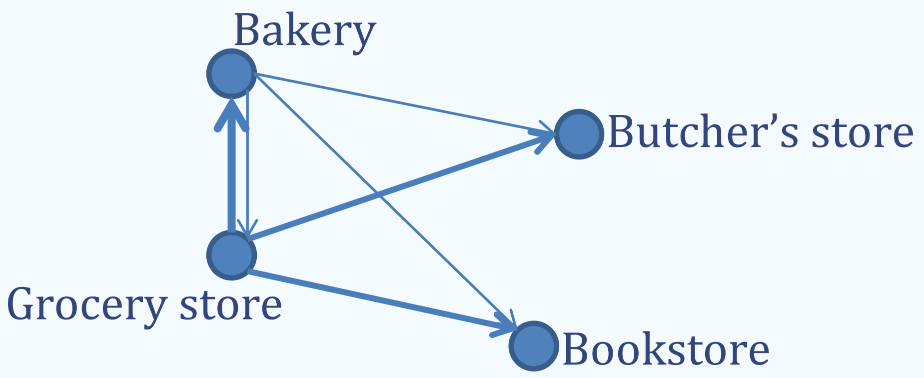
2. Spatial autocorrelation

- **Example: Consider an income increase for grocery store owner**

- **Local autocorrelation:**

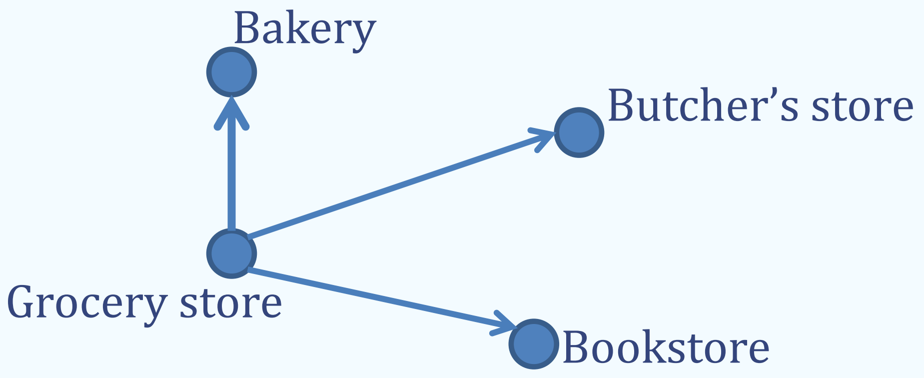


- **Global autocorrelation:**

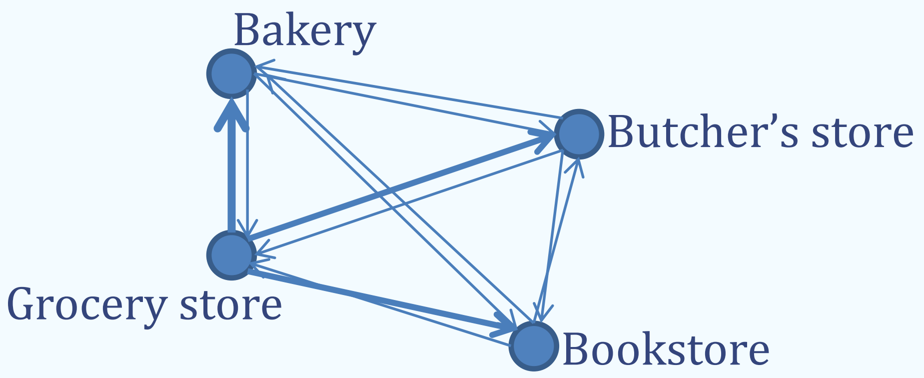


2. Spatial autocorrelation

- **Example: Consider an income increase for grocery store owner**
- **Local autocorrelation:**



- **Global autocorrelation:**



… spatial multiplier process

49

2. Spatial autocorrelation

- Let's define $z = \lambda W z + \mu$
  - Reduced-form of z is $z = [I - \lambda W]^{-1}\mu$
  - With $\lambda < 1$

- A Leontief expansion yields:
  - $[I - \lambda W]^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + \cdots$

- $W^2 \rightarrow$ There is an impact of neighbours of neighbours (as defined in W) although it is smaller (λ^2)
  - Global autocorrelation
  - *Spatial multiplier* process
  - In practice: covariance may approach zero after a relatively small number of powers

2. Spatial autocorrelation

- Let's define $z = \lambda W\mu + \mu$
  - This is already a reduced-form of z
- No impact of behaviour beyond 'bands' of neighbours
  - Dependent on definition of W
  - …Local autocorrelation
- Covariance is zero beyond these bands

2. Spatial autocorrelation

- Local or global autocorrelation?
    - Dependent on application
    - Theory…

- **Taxonomy:**

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \qquad (1)$$

with

$$\boldsymbol{\epsilon} = \lambda \boldsymbol{W}\boldsymbol{\epsilon} + \boldsymbol{\mu} \qquad (2)$$

W is a row-standardised weight matrix

$\rho, \gamma, \beta, \lambda$ **are parameters to be estimated**

VRIJE UNIVERSITEIT AMSTERDAM

- **Taxonomy:**

$$\boxed{\boldsymbol{y}} = \rho \boldsymbol{W}\boldsymbol{y} + \boxed{\boldsymbol{X}\boldsymbol{\beta}} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \qquad (1)$$

with

$$\boldsymbol{\epsilon} = \lambda \boldsymbol{W}\boldsymbol{\epsilon} + \boldsymbol{\mu} \qquad (2)$$

W is a row-standardised weight matrix

$\rho, \gamma, \beta, \lambda$ **are parameters to be estimated**

- **Taxonomy:**

$$\boldsymbol{y} = \boxed{\rho \boldsymbol{W} \boldsymbol{y}} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \qquad (1)$$

with

$$\boldsymbol{\epsilon} = \lambda \boldsymbol{W} \boldsymbol{\epsilon} + \boldsymbol{\mu} \qquad (2)$$

W is a row-standardised weight matrix

$\rho, \gamma, \beta, \lambda$ **are parameters to be estimated**

- **Taxonomy:**

$$y = \rho W y + X\beta + \boxed{WX\gamma} + \epsilon \qquad (1)$$

with

$$\epsilon = \lambda W \epsilon + \mu \qquad (2)$$

W is a row-standardised weight matrix

$\rho, \gamma, \beta, \lambda$ **are parameters to be estimated**

VRIJE UNIVERSITEIT AMSTERDAM

- **Taxonomy:**

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{W} \boldsymbol{X} \boldsymbol{\gamma} + \boldsymbol{\epsilon} \quad (1)$$

with

$$\boldsymbol{\epsilon} = \boxed{\lambda \boldsymbol{W} \boldsymbol{\epsilon}} + \boldsymbol{\mu} \quad (2)$$

W is a row-standardised weight matrix

$\rho, \gamma, \beta, \lambda$ **are parameters to be estimated**

- **Spatial lag model**
  - $\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\mu}$ (3)
  - $\rho \neq 0, \boldsymbol{\gamma} = 0, \lambda = 0$
  - **Spatial dependence in dependent variables**

- **Note similarity with time-series models**
  - **AR Model**
  - $\boldsymbol{y}_t = \rho \boldsymbol{y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\mu}_t$ (4)

- **Spatial lag model**
  - $$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu} \qquad (3)$$

- **The outcome variable influences everyone (indirectly)**
  - **Global autocorrelation**

- **We may write**
  $(\mathbf{I} - \rho \boldsymbol{W})\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
  $\boldsymbol{y} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu})$ **with**
  $(\boldsymbol{I} - \rho \mathbf{W})^{-1} = \boldsymbol{I} + \rho \boldsymbol{W} + \rho^2 \boldsymbol{W}^2 + \rho^3 \boldsymbol{W}^3 + \cdots$

VRIJE UNIVERSITEIT AMSTERDAM

- **Spatial lag model**
  - $\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ (3)

- **We cannot estimate this by OLS because of reverse causality**

- **Recall AR-model:**

  $\boldsymbol{y}_t = \rho \boldsymbol{y}_{t-1} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}_t$ (4)
  - **We can estimate this in principle by OLS because $\boldsymbol{y}_{t-1}$ is not influenced by $\boldsymbol{y}_t$**

- Spatial lag model

- Use maximum likelihood (ML) estimator
  - Selects the set of values of the model parameters that maximizes the likelihood function

- Instrumental variables (IV)
  - Instruments for y may be WX and $W^2\mathrm{X}^2$
  - Less efficient than ML, but feasible for 'large' datasets
  - *e.g.* Kelejian and Prucha (1998)

3. Spatial regressions

- **Spatial cross-regressive model**
  - $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma}\boldsymbol{W}\boldsymbol{X} + \boldsymbol{\mu}$ (5)
  - $\rho = 0, \boldsymbol{\gamma} \neq 0, \lambda = 0$

VRIJE UNIVERSITEIT AMSTERDAM

- **Spatial cross-regressive model**
  - $y = X\beta + \gamma WX + \mu$ (5)

- **Include (transformations) of exogenous variables in the regression**
  - **OLS is fine!**

- **Autocorrelation is local**

- **Spatial error model**
  - $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \text{with } \boldsymbol{\epsilon} = \lambda\mathbf{W}\boldsymbol{\epsilon} + \boldsymbol{\mu}$ (6)
  - $\rho = 0, \boldsymbol{\gamma} = 0, \lambda \neq 0$

- **Spatial error model**
  - $y = X\beta + \epsilon, \text{ with } \epsilon = \lambda W\epsilon + \mu$ (6)

- **Omitted spatially correlated variables**
  - e.g. Ad-hoc defined boundaries
  - Uncorrelated to X!

- **Consistent estimation of parameters β**
- **But: inefficient**
  - ϵ are not i.i.d.
  - Standard errors are higher in OLS
  - β may be different in 'small' samples

VRIJE UNIVERSITEIT AMSTERDAM

- **How to apply these models in practice?**

- SPAUTOREG **in STATA**
- SPATREG **in STATA**
- **GeoDa (free software, also for large datasets)**
- PACE'S SPATIAL STATISTICS TOOLBOX **in MATLAB**

Today:

- Test spatial autocorrelation using Moran's *I*
- Local vs. global spatial autocorrelation
- Incorporate space in regression framework
- Spatial regressions
  - Spatial lag model
  - Spatial cross-regressive model
  - Spatial error model

Tomorrow:

- Spatial regressions (extensions)
- When (not) to use spatial econometrics?

→ **Subscribe to assignment groups before *Tonight, 23:59*.**

Spatial econometrics (2)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

Spatial econometrics (3)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

1. Introduction

- This week
  - Learn about how to deal with spatial data
  - … and spatial econometrics

- Plan:

Lecture #1:	Spatial data
Lecture #2:	Spatial autocorrelation and regressions
Lecture #3:	Spatial regressions (cont'd)
	When (not) to use spatial econometrics?

VRIJE UNIVERSITEIT AMSTERDAM

2. Spatial regressions (continued)

- **Spatial lag model**
  - $\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ (3)
  - $\rho \neq 0, \boldsymbol{\gamma} = 0, \lambda = 0$
  - **Spatial dependence in dependent variables**

- **Spatial cross-regressive model**
  - $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \gamma \boldsymbol{W}\boldsymbol{X} + \boldsymbol{\mu}$ (5)

- **Spatial error model**
  - $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, **with** $\boldsymbol{\epsilon} = \lambda \mathbf{W}\boldsymbol{\epsilon} + \boldsymbol{\mu}$ (6)

VRIJE UNIVERSITEIT AMSTERDAM

2. Spatial regressions (continued)

- **Three issues are on the table**
  1. **When should you use these models?**
  2. **Which of the models should you choose?**
  3. **Can we combine these different spatial models?**

1. When should you use these models?

- **Test for spatial effects**
  - **H_0: No spatial dependence**

- **Estimate standard OLS, $y = X\beta + \epsilon$**
  - **Calculate Moran's I using $\hat{\epsilon}$**
  - $I = \frac{R}{S_0} \times \frac{\hat{\epsilon}' W \hat{\epsilon}}{\hat{\epsilon}' \hat{\epsilon}}$

- **Moran's I does have a rather uninformative alternative hypothesis**
  - **H_A: Spatial dependence…**

VRIJE UNIVERSITEIT AMSTERDAM

1. When should you use these models?

- **Lagrange multiplier tests provide more information**
  - LM_ρ - test for presence of spatial lag
  - LM_λ - test for presence of spatial error

2. Spatial regressions (continued)

1. When should you use these models?

- **Test for spatial lag**
  1. **Run OLS**
  2. **Run LM_ρ-test**

$\mathbf{H_0}$: $\rho = 0$
$\mathbf{H_A}$: $\rho \neq 0$

$$LM_\lambda = \frac{1}{\mathrm{n}J}\left(\frac{\epsilon' W y}{s^2}\right)^2 \sim \chi_1^2 \qquad (9)$$

with $\boldsymbol{J} = [(\boldsymbol{WX\beta})'\boldsymbol{M}(\boldsymbol{WX\beta}) + \boldsymbol{T}s^2]/n\boldsymbol{s}^2$ **and**
$\mathbf{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

1. When should you use these models?

- **Test for spatial error**
  1. **Run OLS**
  2. **Run LM_λ-test**

$\mathbf{H_0}$: $\lambda = 0$
$\mathbf{H_A}$: $\lambda \neq 0$

$$LM_\lambda = \frac{1}{\mathrm{T}}\left(\frac{\epsilon' W \epsilon}{s^2}\right)^2 \sim \chi_1^2 \quad (8)$$

with $\mathrm{T} = \mathrm{tr}\left((\mathbf{W}' + \mathbf{W})W\right)$ **and** $s = \epsilon'\epsilon\,/n$

1. When should you use these models?

- **However,**
  - **Spatial errors and lags may be correlated**
  - **May also be both present**

- **Use *robust* LM tests**
  - LM_ρ^* **adds correction factor for potential spatial error**
  - LM_λ^* **adds correction factor for potential spatial lag**
  - **Complex formulae!**

VRIJE UNIVERSITEIT AMSTERDAM

2. Which of the models should you choose?

- Estimate *robust* LM tests using software

- Common practice
  - Choose and estimate the model for which the statistic is the most significant

2. Which of the models should you choose?

- Estimate *robust* LM tests using software
- Common practice
  - Choose and estimate the model for which the statistic is the most significant
- When LM_{λ}^{*} and LM_{ρ}^{*} are statistically insignificant we may use OLS
- Robust LM-tests are typically provided in STATA output

VRIJE UNIVERSITEIT AMSTERDAM

3. Can we combine these spatial models?

- In practice, both a spatial lag and spatial error may be present

- How to estimate?
  - Use Kelejian and Prucha's GS2SLS method
  - Three-stage procedure!

VU VRIJE UNIVERSITEIT AMSTERDAM

3. Can we combine these spatial models?

- **Complicated stuff!**

- **Let software do the difficult calculations!**
  - SPAUTOREG **in STATA**
  - SPIVREG **in STATA**

VRIJE UNIVERSITEIT AMSTERDAM

3. Mostly pointless spatial econometrics?

- Gibbons and Overman (2012)
  - *"Mostly pointless spatial econometrics?"*
  - Nice read!

- We are interested to identify causal impacts $\boldsymbol{\beta}$:

  $y = X\beta + \mu$

- Typical features of spatial data
  - Unobserved variables correlated with X
  - Omitted variable bias!
  - Large datasets

3. Mostly pointless spatial econometrics?

- **Tempting to 'fix' omitted variable bias by including a spatial lag**

- **Let's consider again:**
  $\boldsymbol{y} = \rho \boldsymbol{W y} + \boldsymbol{X\beta} + \boldsymbol{\mu}$

- **Reduced-form:**
  $\boldsymbol{y} = \rho \boldsymbol{W}(\rho \boldsymbol{W y} + \boldsymbol{X\beta} + \boldsymbol{\mu}) + \boldsymbol{X\beta} + \boldsymbol{\mu}$
  $\boldsymbol{y} = \rho \boldsymbol{W}(\rho \boldsymbol{W}(\rho \boldsymbol{W y} + \boldsymbol{X\beta} + \boldsymbol{\mu}) + \boldsymbol{X\beta} + \boldsymbol{\mu}) + \boldsymbol{X\beta} + \boldsymbol{\mu}$
  …
  $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{WX}\pi_1 + \boldsymbol{W}^2\boldsymbol{X}\pi_2 + \boldsymbol{W}^3\boldsymbol{X}\pi_3 + [\dots] + \tilde{\boldsymbol{\mu}}$

… The last equation suggests that in the end y is just a non-linear function of the X-variables

VRIJE UNIVERSITEIT AMSTERDAM

3. Mostly pointless spatial econometrics?

- Reduced form of spatial lag ≈ spatial cross-regressive model
  - It is hard to prove that spatial lag is 'right' model
  - So, it is hard to distinguish empirically between the two types of models
  - Only when there is a structural (network) model, a spatial lag may be appropriate

3. Mostly pointless spatial econometrics?

- **The spatial lag model *does not* solve the problem of omitted variable bias!**
  - Think of real exogenous sources of variation in X to identify β
  - Use instruments or quasi-experiments
  - More discussion on identification strategies in last week!

- Estimate spatial error model?
  - Spatial datasets are typically large
  - Efficiency issues are *usually* not so important

3. Mostly pointless spatial econometrics?

- **Why then use spatial econometrics?**
  1. **Exploratory tool to investigate spatial autocorrelation**
  2. **Test for spatial dependence and heterogeneity, also in quasi-experiments and when using instruments**
  3. **Investigate whether results are robust to spatial autocorrelation (using different W)**
  4. **Spatial cross-regressive models are often useful and straightforward to interpret**

Today:

- Spatial regressions
  - Spatial lag model
  - Spatial cross-regressive model
  - Spatial error model
- Use robust LM tests to distinguish between different types
  - Or: combine using advanced methods
- Spatial econometrics are a useful tool, *but* not a way to identify causal effects

Next week:

- Discrete choice methods and the random utility framework
- Revealed vs. stated preference data

Spatial econometrics (3)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate