

# Spatial econometrics (1)

Applied Econometrics for Spatial Economics

Hans Koster

*Professor of Urban Economics and Real Estate*

- 1. [Introduction](#)
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- **Hans Koster → URBANECONOMICS.NL**
  - Professor of Urban Economics and Real Estate
  - [h.koster@vu.nl](mailto:h.koster@vu.nl)
  - [Lectures](#)
  - Programme director of STREEM ([streem.sbe@vu.nl](mailto:streem.sbe@vu.nl))
- **Materials**
  - All course materials, lecture slides, etc. can be accessed via [www.urbaneconomics.nl/aese](http://www.urbaneconomics.nl/aese)
  - If there is anything unclear, let me know!

- 1. [Introduction](#)
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- This course
  - Learn about advanced tools and techniques important for spatial economics  
→ No theory – an applied course!
- 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

- 1. [Introduction](#)
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- **Today:**
  - 1. **Spatial econometrics**
  - 2. **Discrete choice**
  - 3. **Identification**
- **Tomorrow:**
  - 4. **Hedonic pricing**
  - 5. **Quantitative spatial economics**

- 1. [Introduction](#)
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- Today:
  - 1. **Spatial econometrics**
    - **Spatial data, autocorrelation, spatial regressions**
  - 2. **Discrete choice**
    - **Random utility framework, estimating binary and multinomial regression models**
  - 3. **Identification**
    - **Research design, IV, OLS, RDD, Quasi-experiments**
- Tomorrow:
  - 4. **Hedonic pricing**
    - **Theory and estimation**
  - 5. **Quantitative spatial economics**
    - **General equilibrium models in spatial economics**

- 1. [Introduction](#)
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- 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  $1/x$

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

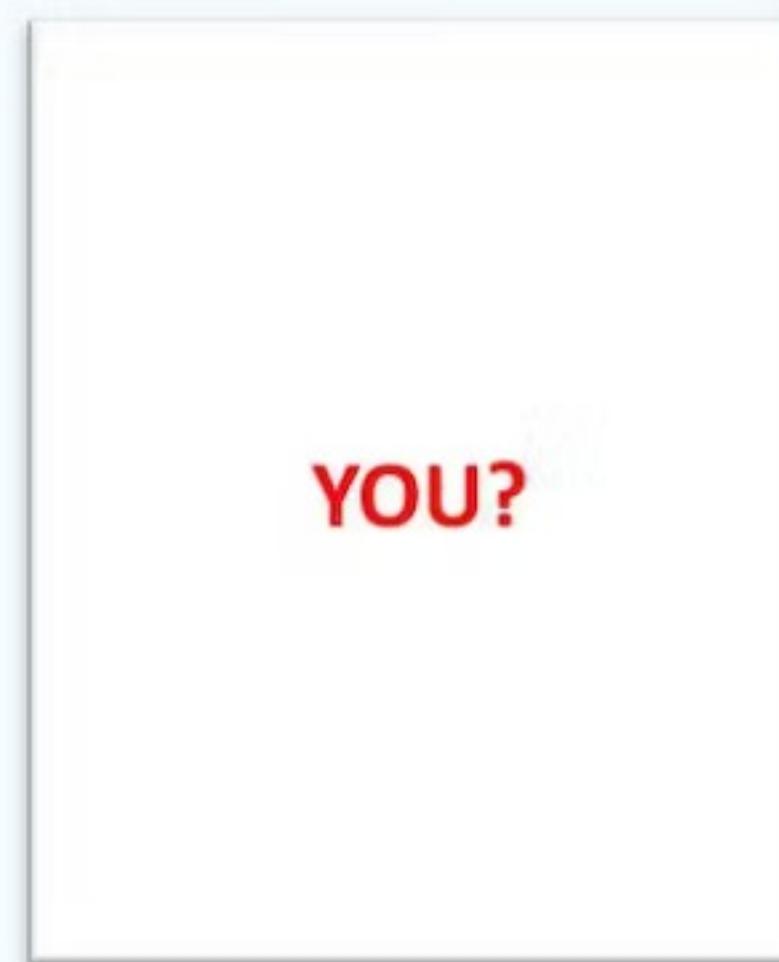
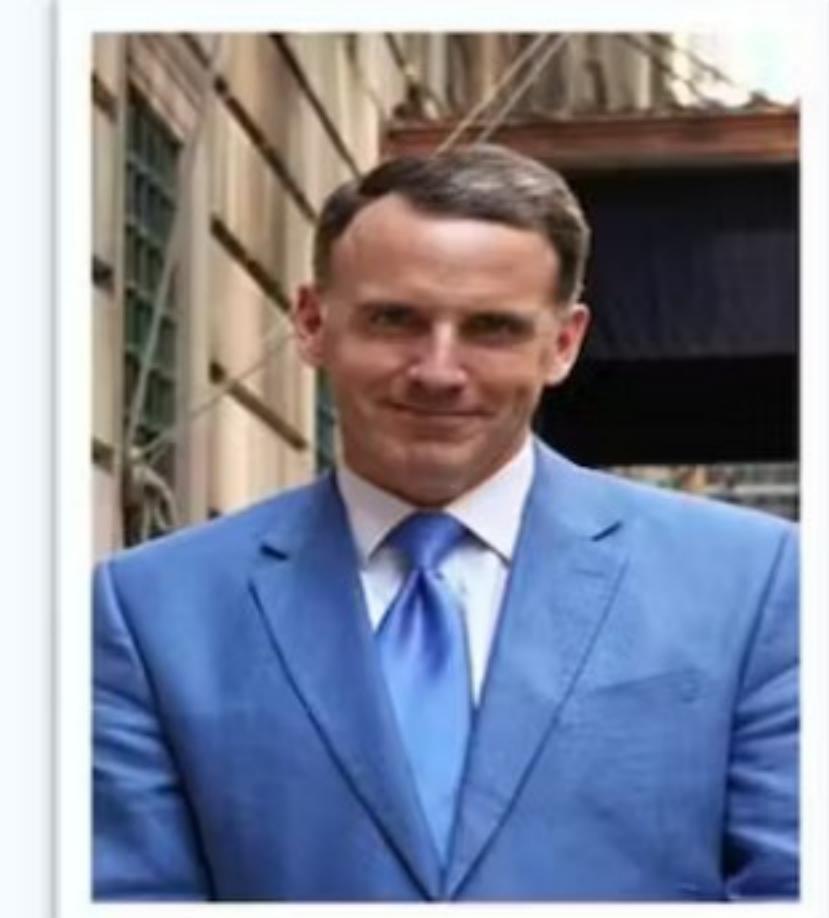
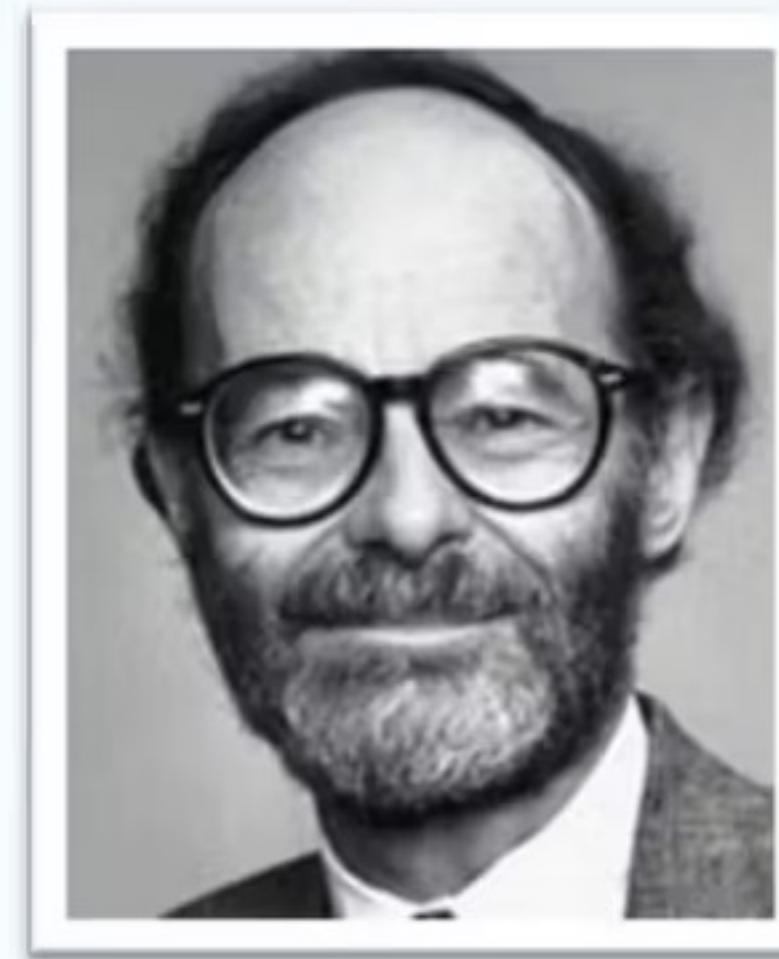
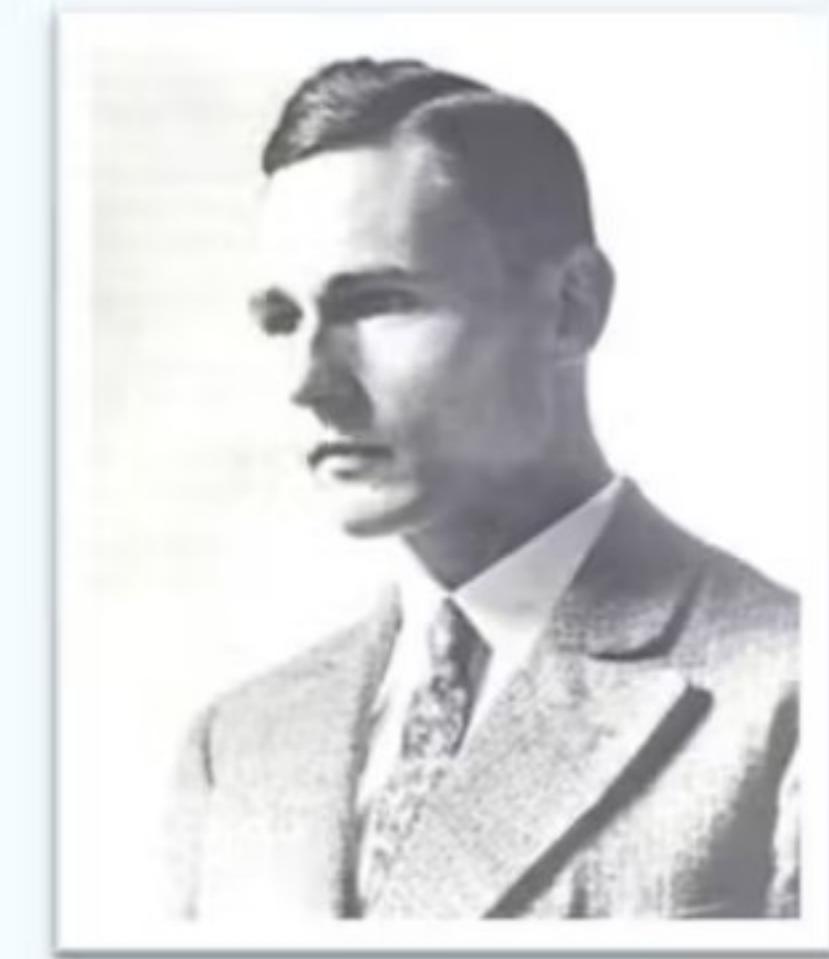
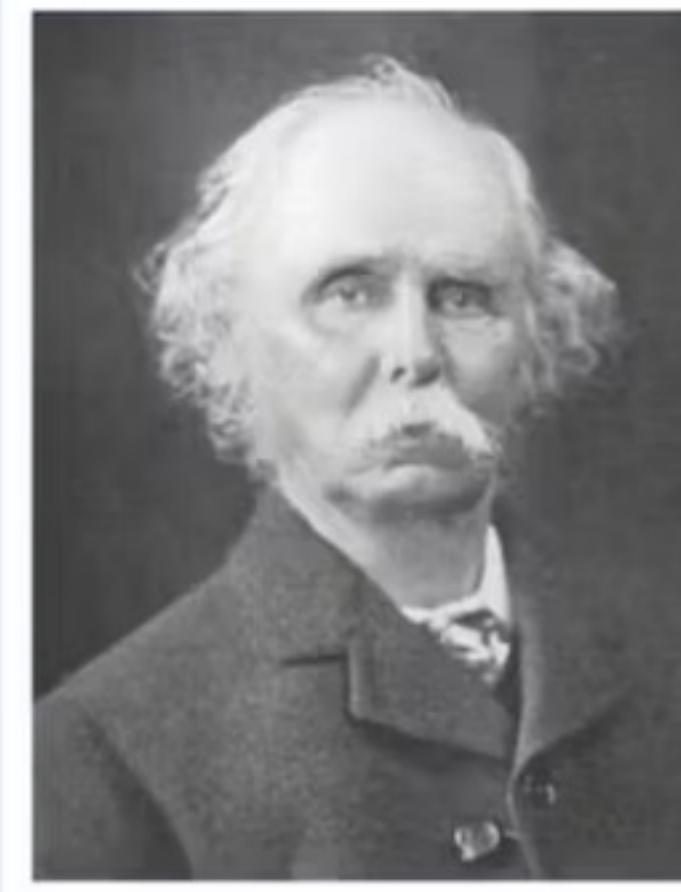
- More details in the appendix of the syllabus

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- 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

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

## ■ Economists and space



- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- **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)*

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

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

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

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

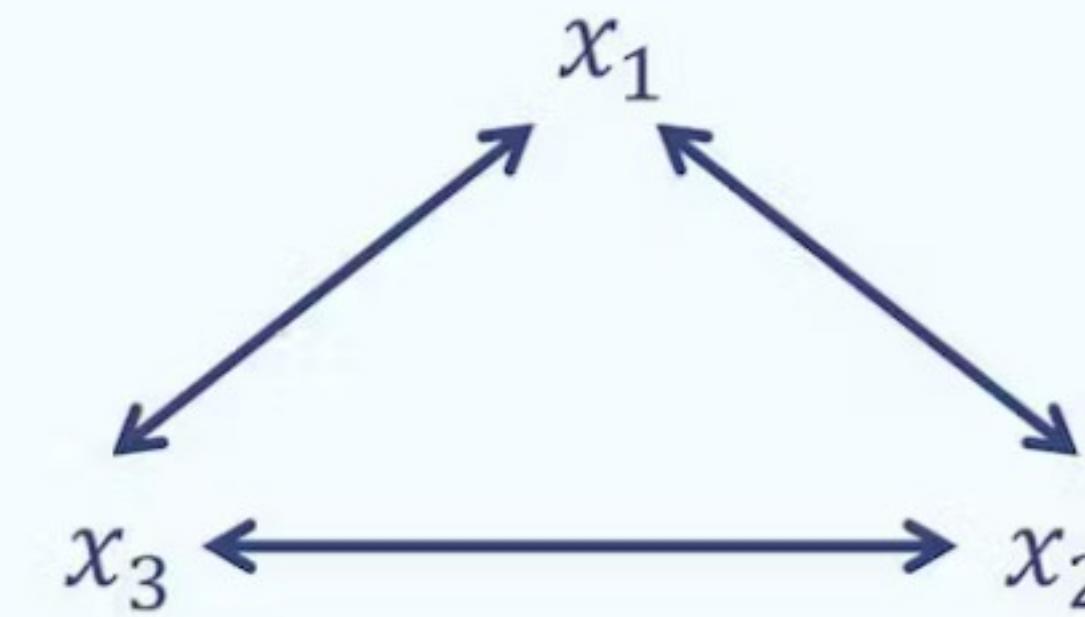


- Linear space (*e.g. beach*) is different
  - No natural origin
  - Reciprocity
  - Unidirectional



- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

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



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

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

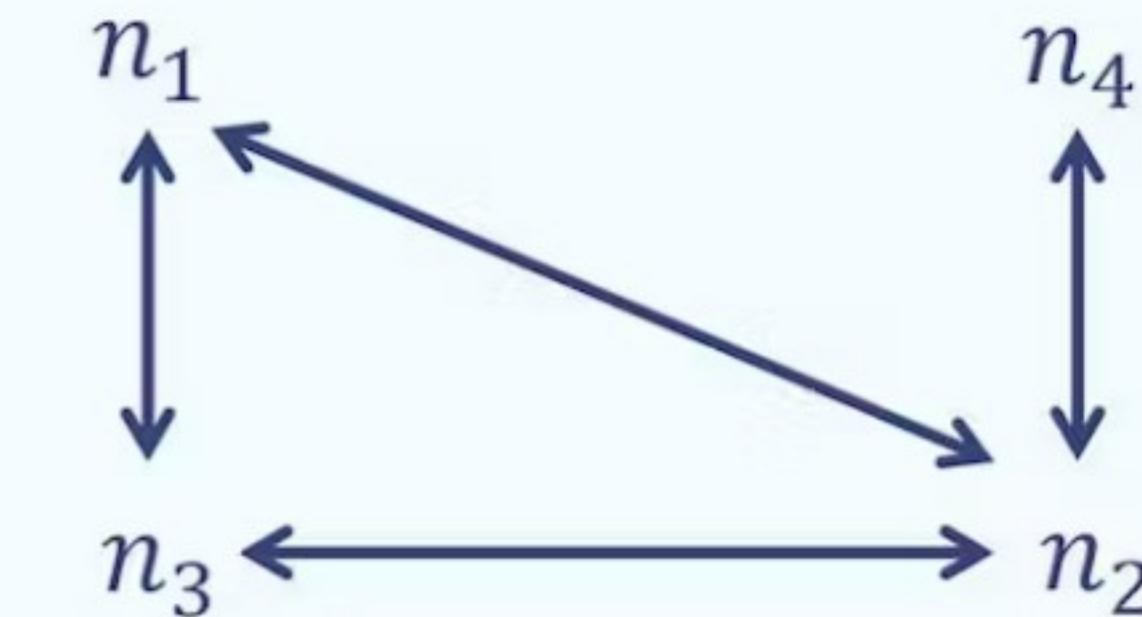
- First, we have to define the spatial structure of the data
- Specified through a spatial weights matrix
- Spatial weights matrix  $W$ :
  - Consists of  $n \times n$  elements
  - Discrete or continuous elements
- How to define weights?
  - Euclidian distance
  - Network distance
  - Spatial interactions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

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- How to define spatial matrices?
- Contiguity matrix
  - Adjacent → 1<sup>st</sup> order contiguous
  - Neighbours of neighbours → 2<sup>nd</sup> order contiguous
- Distance matrix
  - *k*-nearest neighbours
  - Inverse distance weights ( $1/distance$ )
  - Cut-off distance

1. Introduction
  2. Space in economics
  3. Spatial data structure
  4. MAUP
  5. Summary

- Let's provide an example of a contiguity matrix



三〇四

to

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

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- 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

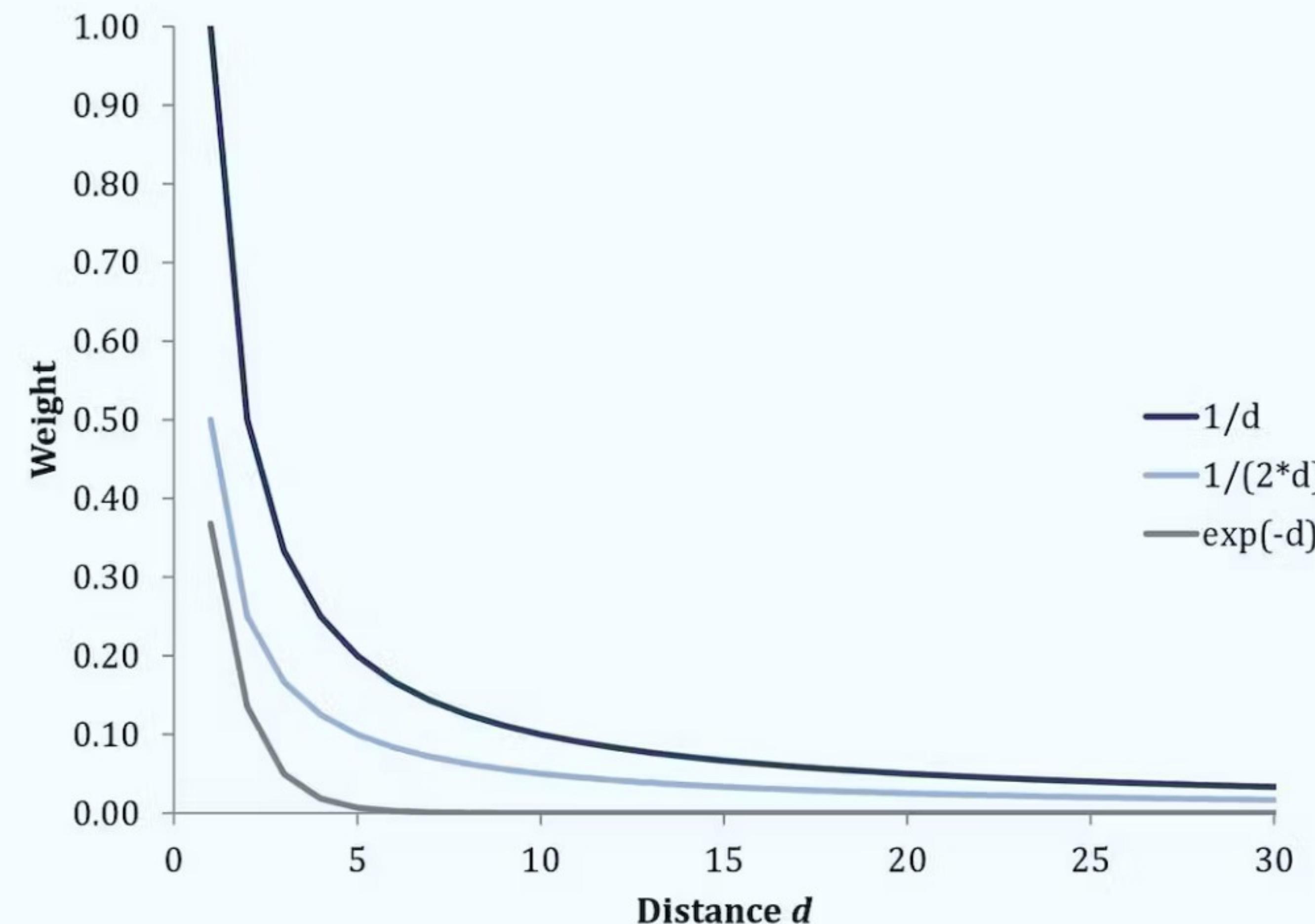
- Interpretation of
  - $\sum_{j=1}^n w_{ij}$ : sum of connections to neighbours
  - $w_{ij}^*$  denotes the share of connections to neighbours

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

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

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

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

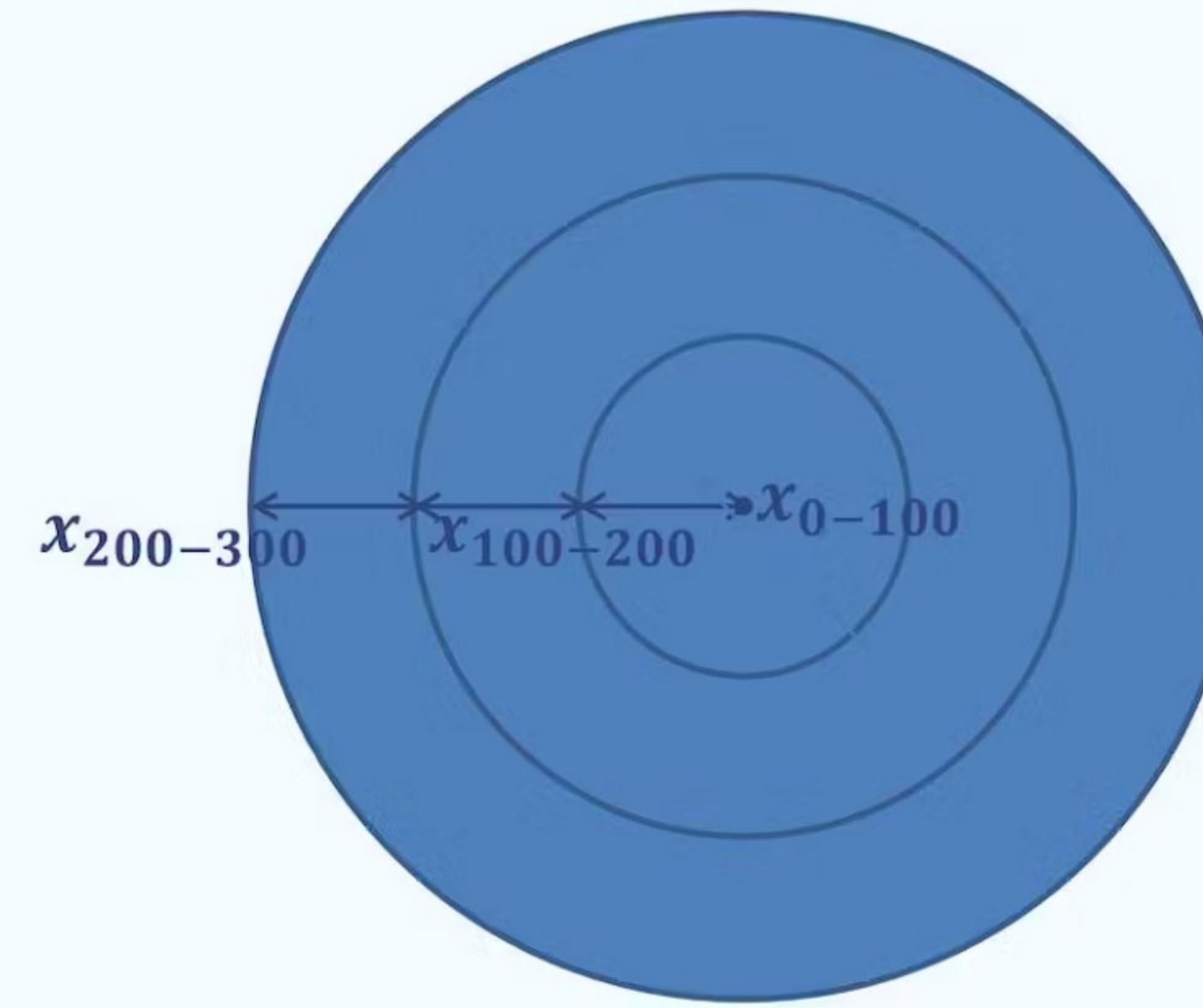


- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- **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

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

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



- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

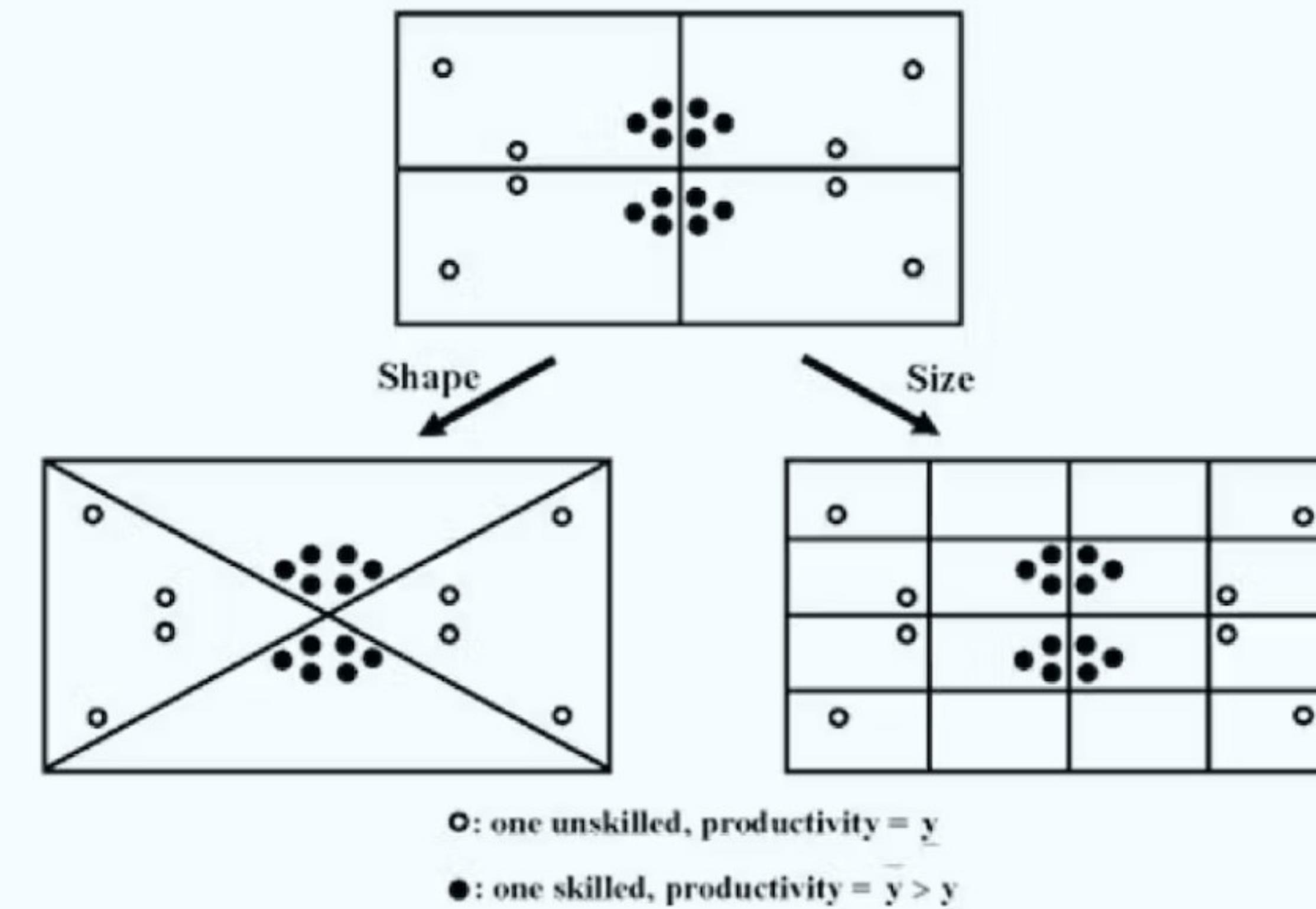
- How to define spatial weight matrix using software
  - SPATWMAT **in STATA**, based on geographic coordinates
  - SPWEIGHT **in STATA**
  - **Geoda**
  - SPATIAL STATISTICS TOOLBOX **in ArcGIS**
  - SPDEP **in R**
- Concentric rings should be calculated manually

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- 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)

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- An illustration:



Briant, Combes and Lafourcade (2010, JUE)

- Aggregation seems to be important!

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- **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)

- 1. Introduction
- 2. Space in economics
- 3. Spatial data structure
- 4. MAUP
- 5. Summary

- **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**

# 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*

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Spatial autocorrelation between values**
  - Implies  $\text{cov}(x_i, x_j) = E[x_i x_j] - E[x_i] \cdot E[x_j] \neq 0$
  - Again,  $j$  refers to other locations
- **Spatial autocorrelation, dependence, clustering**
  - Fuzzy definitions in literature

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- 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

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- Moran's  $I$  is given by:

$$I = \frac{R}{S_0} \times \frac{\tilde{x}' W \tilde{x}}{\tilde{x}' \tilde{x}} \quad (4)$$

where  $R$  is the number of spatial units  
 $S_0$  is the sum of all elements of the spatial weight matrix  
 $W$  is the spatial weight matrix  
 $\tilde{x} = x - \bar{x}$  is a vector with the variable of interest

- Use row-standardised spatial weight matrix  $W$ !
  - So that  $I_S = \frac{\tilde{x}' W \tilde{x}}{\tilde{x}' \tilde{x}}$

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Moran's  $I$**
- **Sidenote:**
  - Please realise that  $W\tilde{x}$  is a vector
    - $I_S = \frac{\tilde{x}'W\tilde{x}}{\tilde{x}'\tilde{x}}$
    - $W\tilde{z}$  is a *vector*
      - $\mathbf{W} \times \tilde{x} = W\tilde{x}$ 
$$\begin{bmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{bmatrix} \times \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$
  - Notation:  $\frac{x'y}{x'x} = x^T y (x^T x)^{-1}$

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Moran's  $I$**
- **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**  $E[I] = -\frac{1}{N-1}$
- **Visualisation**
  - **Moran scatterplot**

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- Moran's  $I$
- How to investigate the statistical significance of (4)?
  - $\frac{I - E[I]}{\sqrt{\text{var}[I]}}$  (5)
  - However,  $\sqrt{\text{var}[I]}$  is difficult to derive
  - $E[I] = -1/(n - 1)$
  - Assume normal distribution of  $I$  to approximate  $\sqrt{\text{var}[I]}$  under  $H_0$
  - Or: bootstrapping/simulation
  - See Cliff and Ord (1973) for more details

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- Moran's *I*
- Also possible: correlation to other variables:

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

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

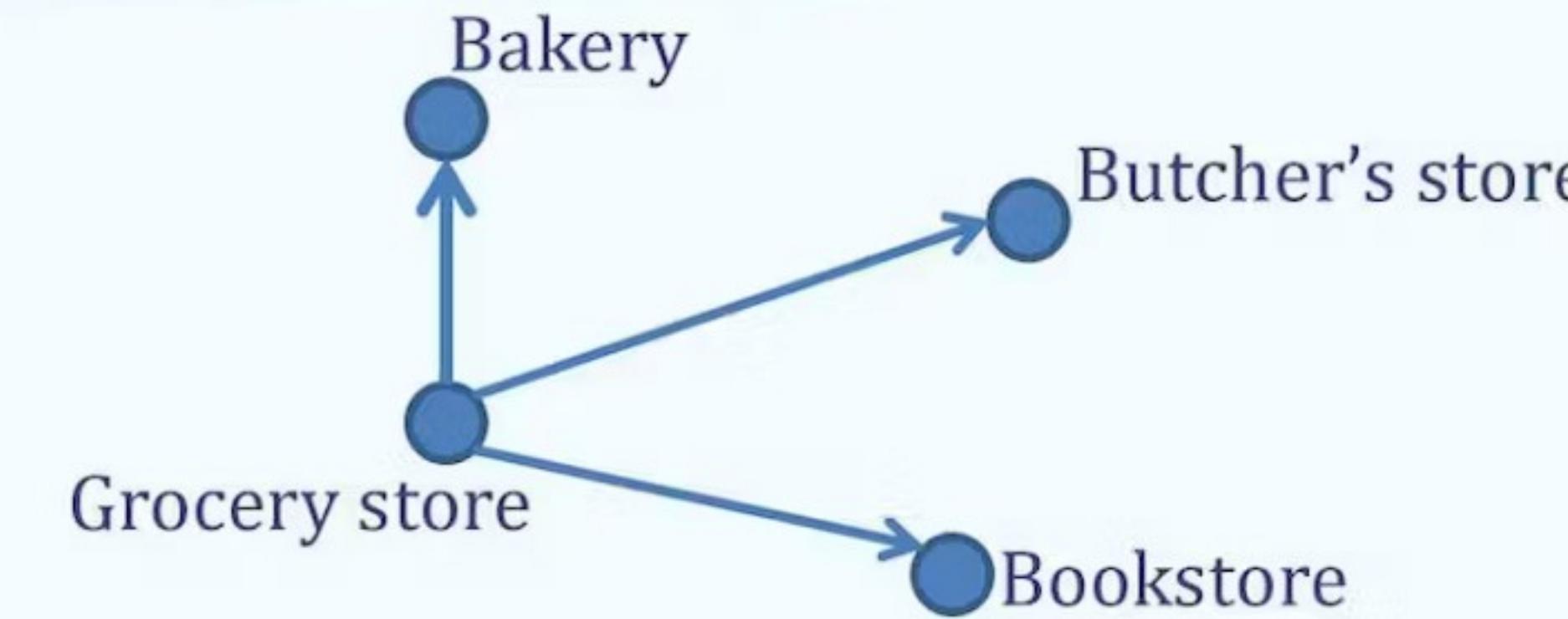
- How to calculate Moran's *I* using software
  - SPAUTOC in STATA
  - SPLAGVAR in STATA
  - SPATIAL STATISTICS TOOLBOX in ArcGIS
- Alternative: Getis and Ord's *G*
  - Most of the time only Moran's *I* is reported

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

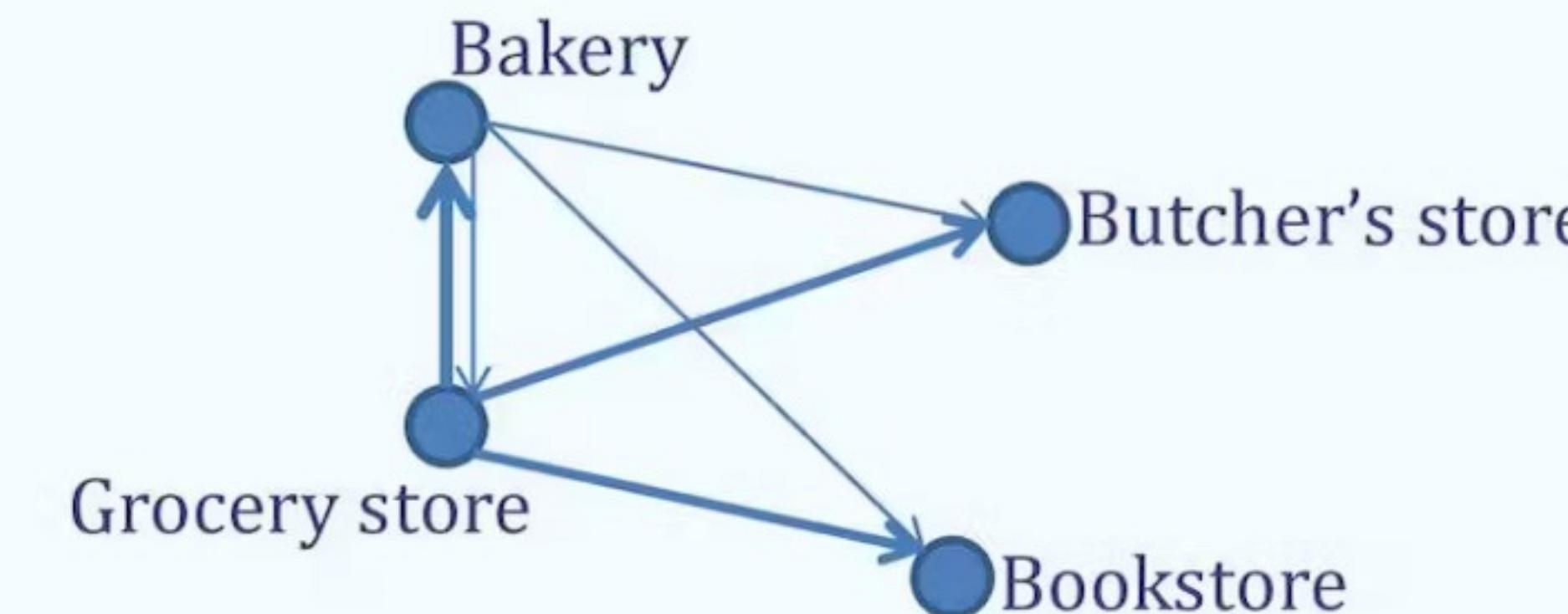
- It is important to make a distinction between *global* and *local* spatial autocorrelation
  - See Anselin (2003) for a discussion
- Global spatial autocorrelation
  - Local shock affects the whole system
- Local spatial autocorrelation
  - Local shock only affects the ‘neighbours’

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

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

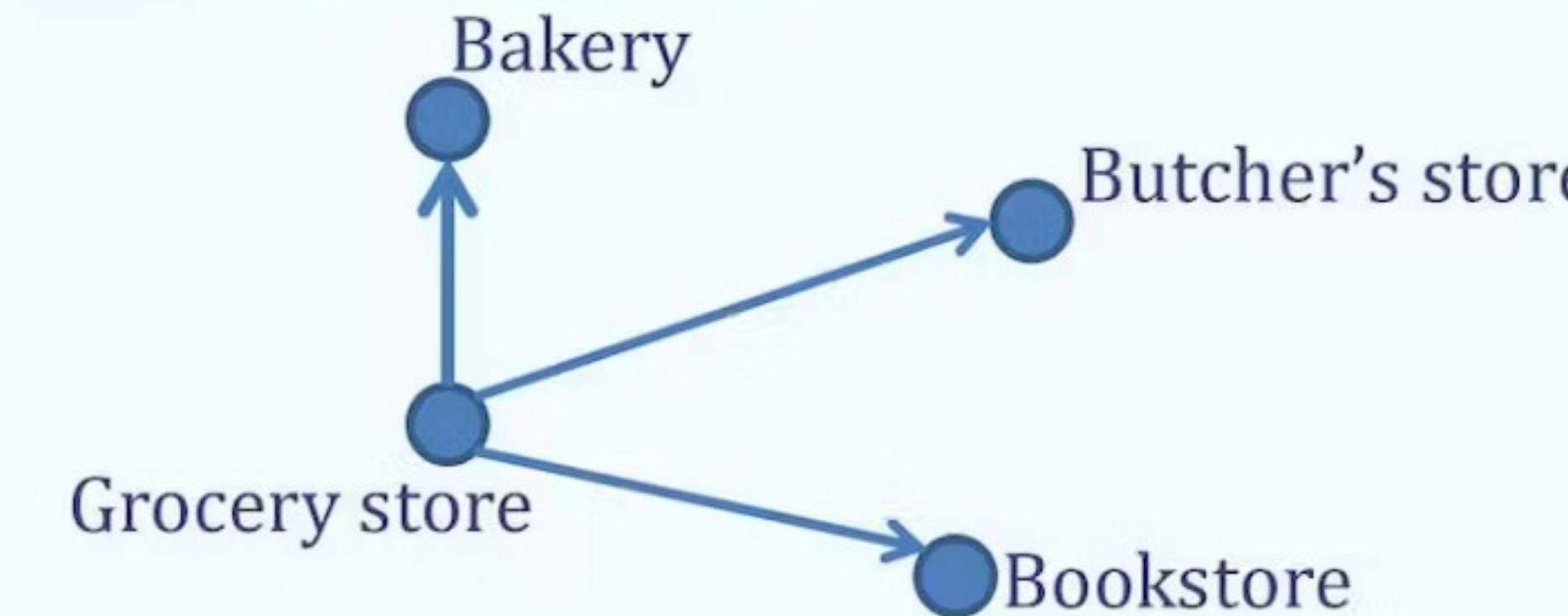


- Global autocorrelation:

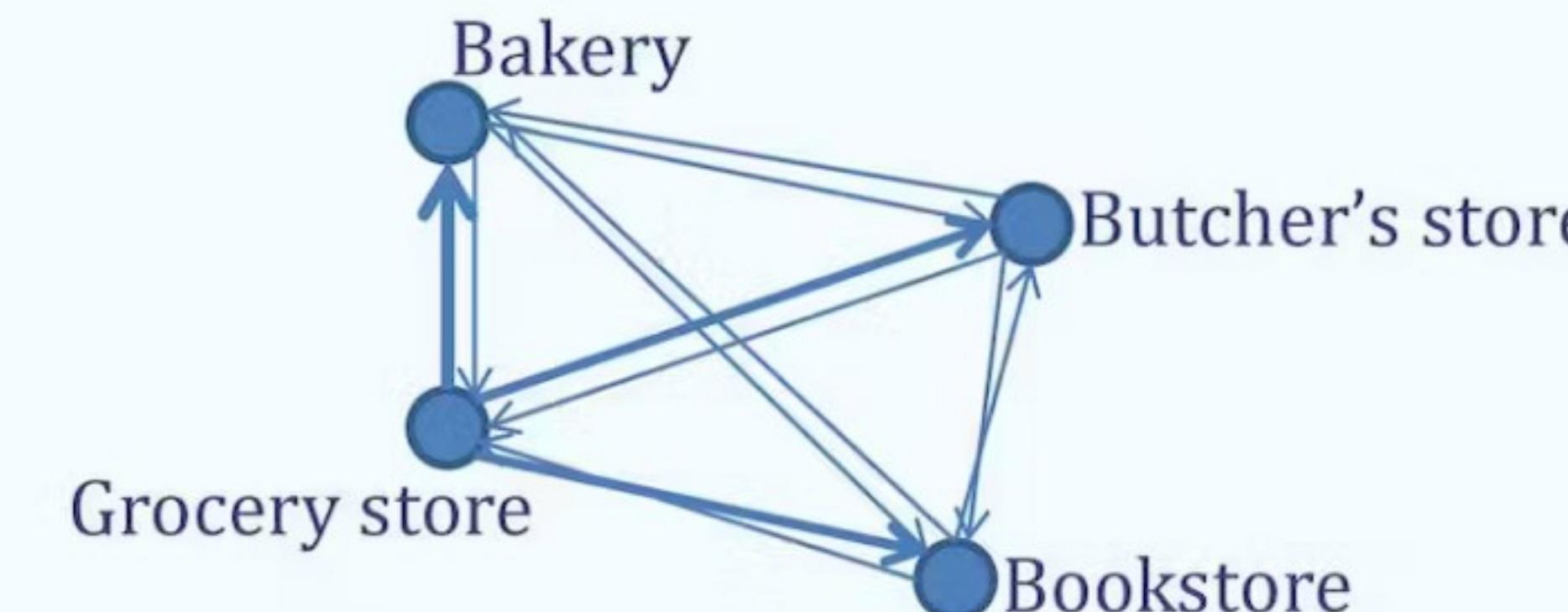


- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

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



- Global autocorrelation:



... spatial multiplier process

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- 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:
  - $[I - \lambda W]^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + \dots$
- $W^2 \rightarrow$  There is an impact of neighbours of neighbours (as defined in  $W$ ) although it is smaller ( $\lambda^2$ )
  - Global autocorrelation
  - Spatial multiplier process
  - In practice: covariance may approach zero after a relatively small number of powers

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- 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

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

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

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Taxonomy:**

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

**with**

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

**W is a row-standardised weight matrix**  
 **$\rho, \gamma, \beta, \lambda$  are parameters to be estimated**

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Spatial lag model**

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

- **Note similarity with time-series models**

- **AR Model**
- $y_t = \rho y_{t-1} + X_t\beta + \mu_t$  (4)

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Spatial lag model**
  - $y = \rho W y + X\beta + \mu$  (3)
- **The outcome variable influences everyone (indirectly)**
  - Global autocorrelation
- **We may write**
$$(I - \rho W)y = X\beta + \epsilon$$
$$y = (I - \rho W)^{-1}(X\beta + \mu) \text{ with}$$
$$(I - \rho W)^{-1} = I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots$$

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

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

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

- **Recall AR-model:**

$$y_t = \rho y_{t-1} + X\beta + \mu_t \quad (4)$$

- **We can estimate this in principle by OLS because  $y_{t-1}$  is not influenced by  $y_t$**

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **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  $W^2X^2$**
  - **Less efficient than ML, but feasible for 'large' datasets**
  - **e.g. Kelejian and Prucha (1998)**

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Spatial cross-regressive model**

- $y = X\beta + \gamma WX + \mu$  (5)
- $\rho = 0, \gamma \neq 0, \lambda = 0$

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **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**

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Spatial error model**

- $y = X\beta + \epsilon$ , with  $\epsilon = \lambda W\epsilon + \mu$  (6)
- $\rho = 0, \gamma = 0, \lambda \neq 0$

- 1. Introduction
- 2. Spatial autocorrelation
- 3. Spatial regressions
- 4. Summary

- **Spatial error model**
  - $y = X\beta + \epsilon$ , with  $\epsilon = \lambda W\epsilon + \mu$  (6)
- **Omitted spatially correlated variables**
  - e.g. Ad-hoc defined boundaries
  - Uncorrelated to  $X$ !
- **Consistent estimation of parameters  $\beta$**
- **But: inefficient**
  - $\epsilon$  are not i.i.d.
  - Standard errors are higher in OLS
  - $\beta$  may be different in ‘small’ samples

# 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
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

- **Spatial lag model**

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

- **Spatial cross-regressive model**

- $y = X\beta + \gamma W X + \mu$  (5)

- **Spatial error model**

- $y = X\beta + \epsilon$ , with  $\epsilon = \lambda W \epsilon + \mu$  (6)

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

- **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. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

## 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...

1. Introduction
2. Spatial regressions
3. Mostly pointless?
4. Summary

## 1. When should you use these models?

- However,
  - Spatial errors and lags may be correlated
  - May also be both present
- Use robust LM tests
  - $LM_{\rho}^*$  adds correction factor for potential spatial error
  - $LM_{\lambda}^*$  adds correction factor for potential spatial lag
  - Complex formulae!

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

### 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!

1. Introduction
2. Spatial regressions
3. Mostly pointless?
4. Summary

### 3. Can we combine these spatial models?

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

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

- **Gibbons and Overman (2012)**
  - “*Mostly pointless spatial econometrics?*”
- **We are interested to identify causal impacts  $\beta$ :**
$$y = X\beta + \mu$$
- **Typical features of spatial data**
  - **Unobserved variables correlated with  $X$**
  - **Omitted variable bias!**
  - **Large datasets**

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

- Tempting to ‘fix’ omitted variable bias by including a spatial lag
- Let’s consider again:  
$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$
- Reduced-form:  
$$\mathbf{y} = \rho \mathbf{W}(\rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}) + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$
  
$$\mathbf{y} = \rho \mathbf{W}(\rho \mathbf{W}(\rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}) + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}) + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$
  
...  
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\rho + \mathbf{W}^2\mathbf{X}\rho^2 + \mathbf{W}^3\mathbf{X}\rho^3 + [...] + \tilde{\boldsymbol{\mu}}$$
- ... The last equation suggests that in the end  $\mathbf{y}$  is just a non-linear function of the  $\mathbf{X}$ -variables

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

- **Reduced-form of spatial lag model  $\approx$  spatial cross-regressive model**
  - It is hard to prove that the spatial lag model is the ‘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

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

- **The spatial lag model *does not* solve the problem of omitted variable bias!**
  - Think of real exogenous sources of variation in  $X$  to identify  $\beta$
  - 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

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

- 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

- 1. Introduction
- 2. Spatial regressions
- 3. Mostly pointless?
- 4. Summary

## Spatial econometrics:

- **Spatial data:**
  - No natural origin, reciprocity, multidirectional
  - Define spatial relationships by the spatial weight matrix
- **Spatial regressions**
  - Spatial lag model
  - Spatial cross-regressive model
  - Spatial error model
  - ... Combine using advanced methods
- **Spatial econometrics are a useful tool, but not a way to identify causal effects**

# Spatial econometrics (3)

Applied Econometrics for Spatial Economics

Hans Koster

*Professor of Urban Economics and Real Estate*