

Spatial Statistics and Econometrics

Roberto Patuelli
Department of Economics
University of Bologna

EAERE-ETH European Winter School on “Spatial Environmental
and Resource Economics”

Structure

- Basic concepts, definitions, indicators of spatial autocorrelation, exploratory spatial data analysis
- Standard spatial econometric models, nonlinear spatial models?
- Panel spatial econometric models, further (alternative?) specifications

Caveat: I'm not an econometrician... I'm a “user” of spatial methods.

For those interested in going into spatial econometrics in-depth, there are several summer schools around (e.g., SEA's summer school in Rome) with the top spatial econometricians teaching for up to three full weeks

Analysis of Spatial Data

- **WHERE:** in which contexts should we worry about spatial issues regarding our data?
- **WHY:** what are the implications of spatial interaction and in general spatial aspects for statistical/econometric modelling?
- **HOW:** how are spatial issues treated empirically?

Where?

- Spatial (georeferenced) data come in several forms (Cressie 1993)
 - *Geostatistical data* – continuous surface in the bidimensional domain \mathbb{R}^2
 - *Lattice/area (regional) data* – finite (ir)regular set of points in \mathbb{R}^2 or areas that partition \mathbb{R}^2
 - *Point pattern data* – point process that can distinguish between locations having or not having a certain attribute
 - (*Objects* – again point process, like point-pattern data, the set D of points is result of a random process)
- Similar classification by Fischer and Wang (2011) (see next slide)
- Methods often depend on type of data, although they can sometimes be borrowed between classes of data

Where? (2)

Table 1.1 Types of spatial data: conceptual schemes and examples

Type of spatial data	Conceptual scheme		Example	
	Variable scheme	Spatial index	Variable	Space
Point pattern data	Variable (discrete or continuous) is a random variable	Point objects to which the variable is attached are fixed	Trees: diseased or not Hill forts: classified by type	Two-dimensional discrete space Two-dimensional discrete space
Spatially continuous (geostatistical) data	Variable is a continuous valued function of location	Variable is defined everywhere in the (two-dimensional) space	Temperature Atmospheric pollution	Two-dimensional continuous space Two-dimensional continuous space
Area (object) data	Variable (discrete or continuous) is a random variable	Area objects to which the variable is attached are fixed	Gross regional product Crime rates	Two-dimensional discrete space Two-dimensional discrete space
Spatial interaction (flow) data	Variable representing mean interaction frequencies is a random variable	Pairs of locations (points or areas) to which the flow variable is attached	International trade Population migration	Two-dimensional discrete space Two-dimensional discrete space

Where? (3)

- In practical terms (e.g. R programming), we can distinguish between (Bivand et al. 2008):
 - *Point*, a single point location, such as a GPS reading or a geocoded address
 - *Line*, a set of ordered points, connected by straight line segments
 - *Polygon*, an area, marked by one or more enclosing lines, possibly containing holes
 - *Grid*, a collection of points or rectangular cells, organised in a regular lattice
- All spatial data have positional attributes, ‘answering the question “where is it?”’

Why?

- Spatial data are often non independent
 - Violation of assumption of observations coming from independent random variables given in classical statistical theory (sphericity of errors: homoskedasticity and no autocorrelation)
 - Spatial data tend to be positively correlated, with the degree of correlation decreasing over distance
 - In this conditions, OLS is not appropriate anymore
 - F and t tests on regression parameters may lead to wrong conclusions
 - Additionally, the assumption of homoskedasticity may be violated, if, for example, rates from areal data of widely different base population are analysed
- Data support
 - Incompatible data. How to combine data collected on different supports (e.g., different levels of spatial aggregation)?
 - Change of support
 - Combining data towards creating a new variable
 - Modifiable areal units problem (MAUP, Openshaw and Taylor 1979): often data are collected for purely administrative areas which don't have intrinsic geographical meaning. But regression results often depend on the scale of the units (scale problem) and their configuration (aggregation problem)
 - Ecological fallacy (Robinson 1950): making statistical inference on individuals on the basis of aggregate data is flawed

How?

- 1) Exploratory Spatial Data Analysis (ESDA)
 - Extension of Tukey-type data exploration
 - Preliminary data analysis, based in particular on mapping
 - GIS may help summarizing geographic information, finding outliers, manipulating point data, etc
 - Used mostly prior to model building, also to make hypotheses about the data, but new ESDA techniques go directly into the model building phase, showing how variables relate to each other in space

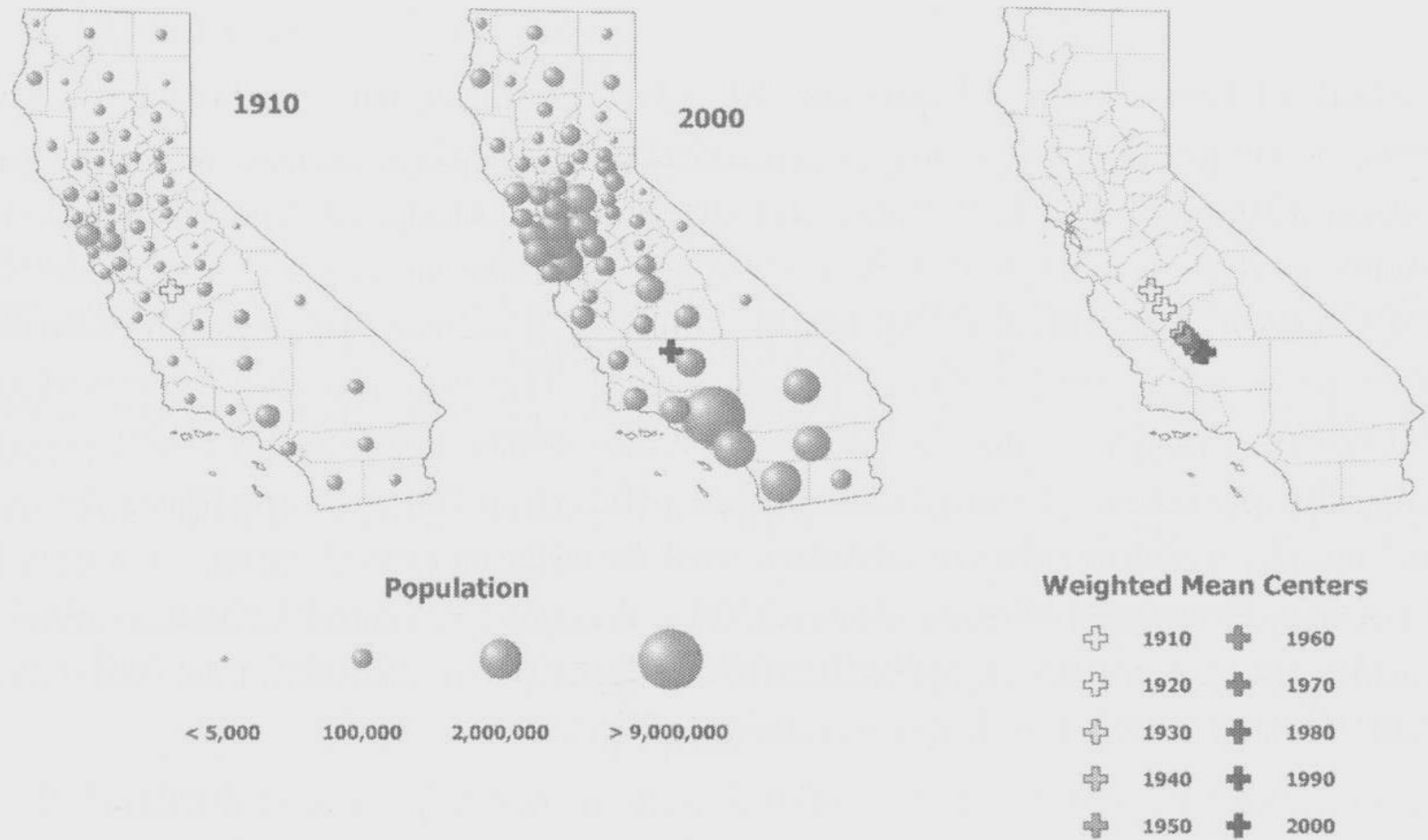


Fig. A.1.2. Weighted mean center of population, by county, 1910 through 2000

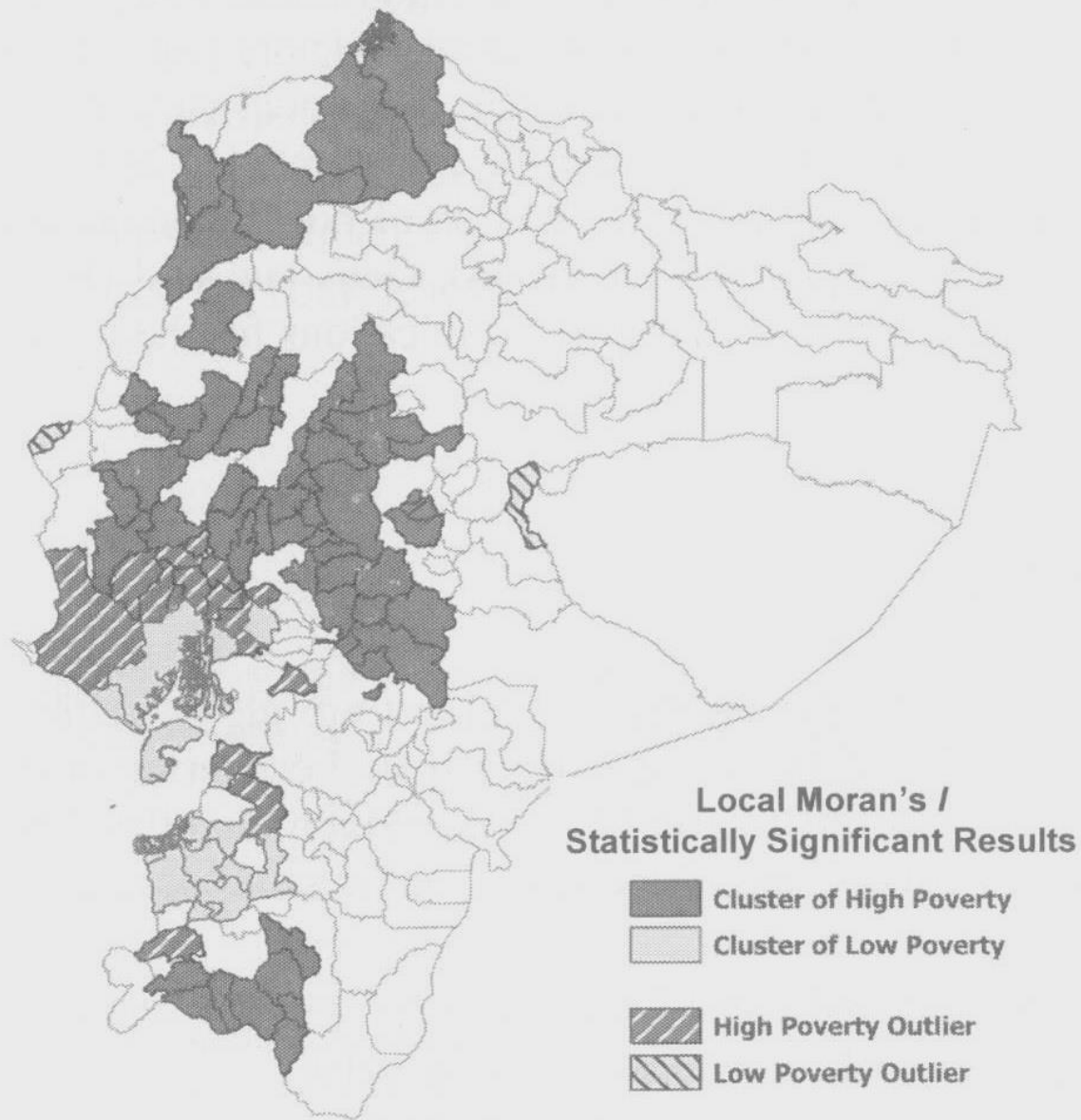
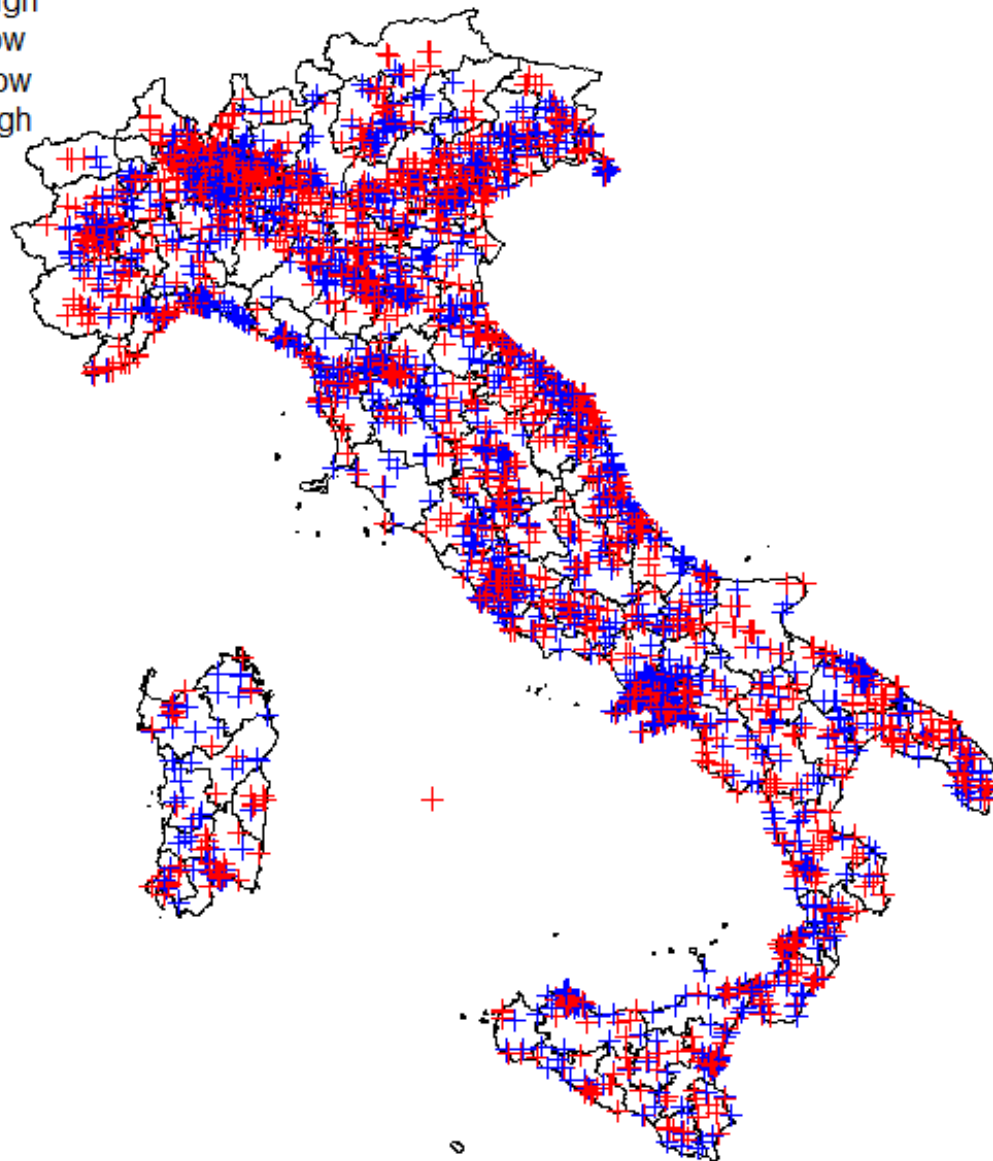
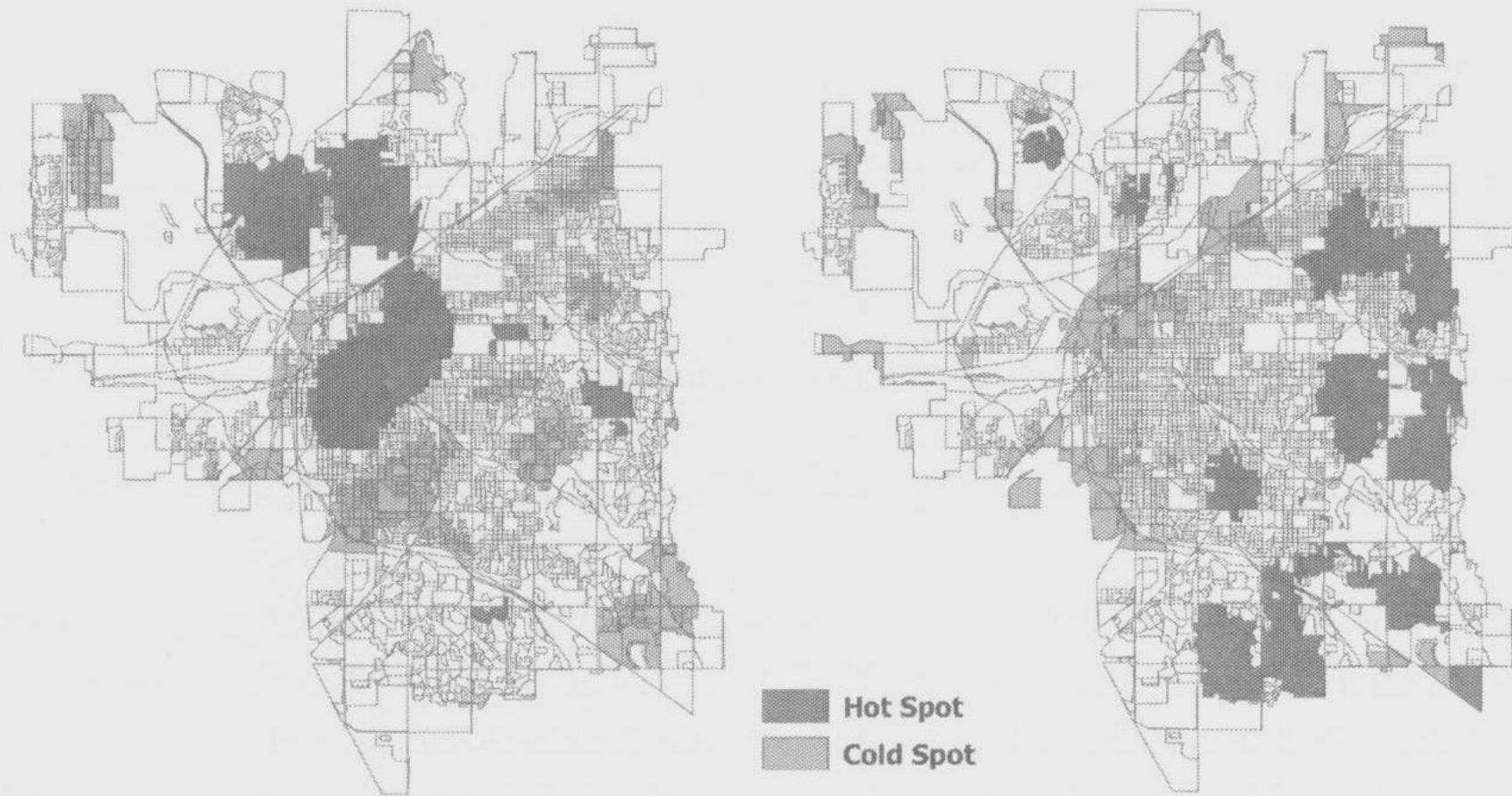


Fig. A.1.6. An analysis of poverty in Ecuador using local Moran's I

- High-High
- Low-Low
- High-Low
- Low-High





Where are there a lot of vandalism incidents?

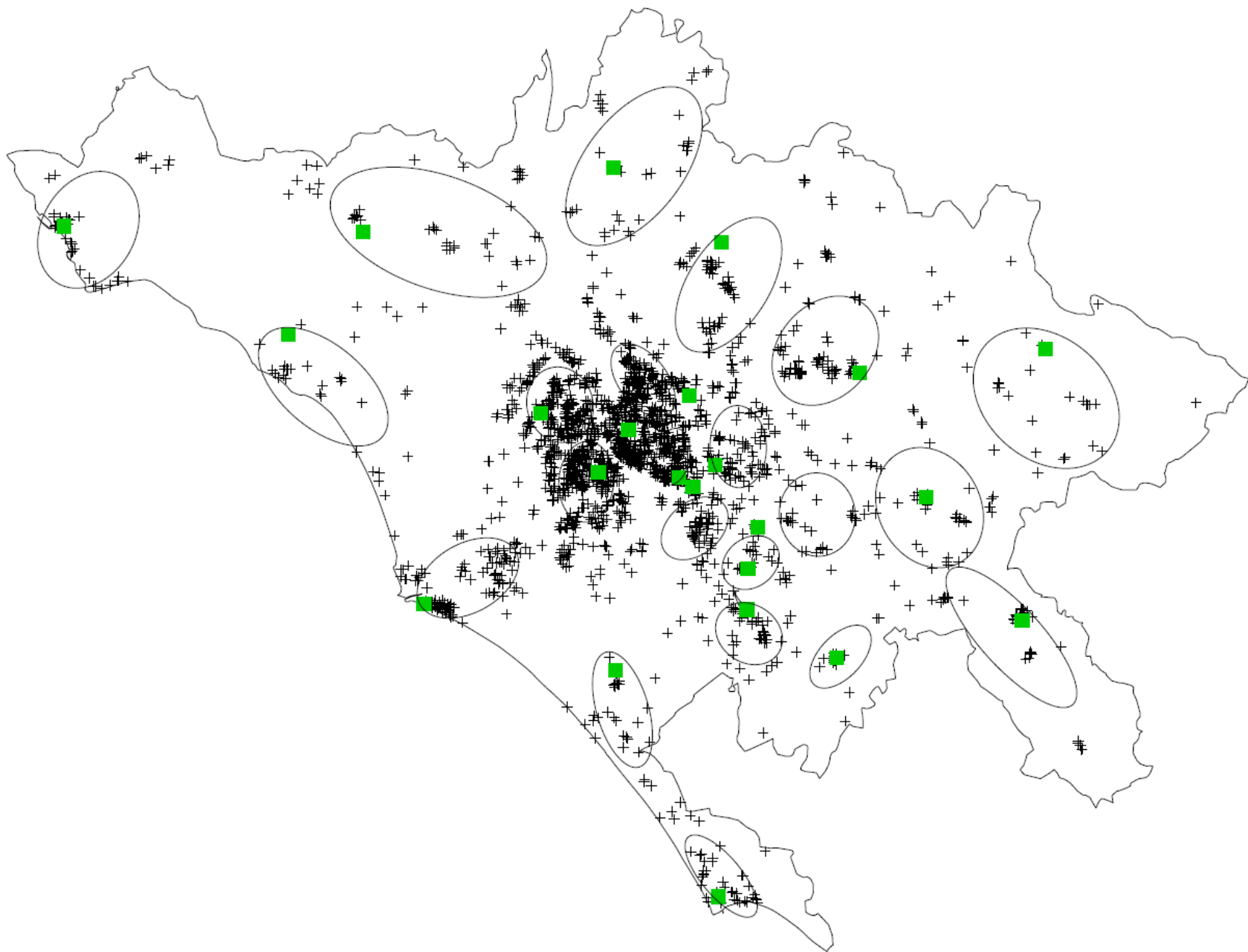
Where is vandalism a higher than expected proportion of all crime incidents?

Fig. A.1.7. An analysis of vandalism hot spots in Lincoln, Nebraska using G_i^*



Fig. A.4.2. Linked box maps, box plot and cartogram, raw and smoothed prostate cancer mortality rates

Ellissi deviazionali (K-Means Clustering): campione e CPI



How? (2)

- 2) Spatial Statistics
 - Started with Whittle, Moran, Geary, Cliff and Ord (late '60s). Also part of ESDA, spatial econometrics and more...
 - Has its *raison d'être* in creating hypotheses and testing map patterns
 - How social/economic/etc. variables pattern on a map and interact with each other?
 - Spatial autocorrelation indices (e.g., Geary, Moran)
 - Creation of spatial weights matrices
 - Spatial filtering (e.g., Getis, Griffith)
 - Spatial cluster analysis (e.g., Ripley's K (*Spatial Statistics*, 1981) – coincidentally, the same guy who later contributed to the birth of R)

How? (3)

- 3) Spatial Econometrics
 - Paelinck and Klaassen (1979), Anselin (1988)
 - Anselin: spatial lag model; spatial error model
 - Need for spatial statistical tests to check assumptions of spatial randomness in regression residuals
 - Moran's I
 - Specification search: Lagrange multiplier tests...
 - Geographically weighted regression (GWR; Fotheringham, Brunson, Charlton) to allow regression parameters to vary over space
 - ... and many more recently developed methods accounting for spatial autocorrelation in econometric techniques (e.g. instrumental variables, GMM methods, nonlinear (GLM) models...)
- 4) Geostatistics (not discussed here)
 - Geostatistical methods most often start from observations at points of single or multiple attributes, and are concerned with their statistical interpolation to a field or continuous surface (e.g. *kriging*) assumed to extend across the whole study area

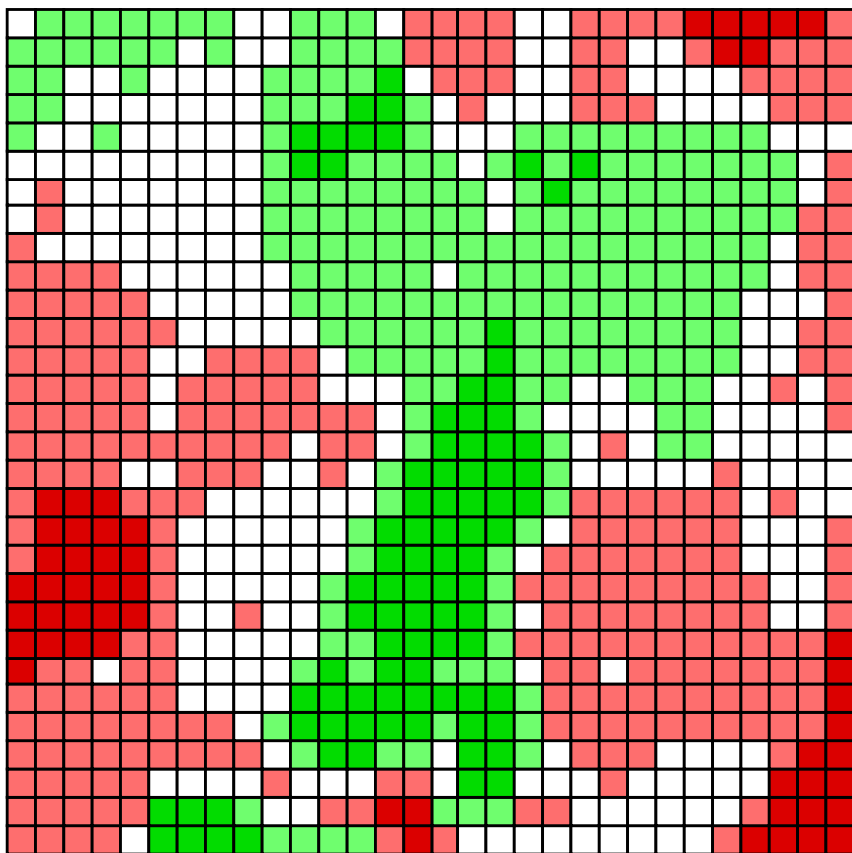
Spatial Autocorrelation

- Definitions
 - ‘It represents the relationship between nearby spatial units, as seen on maps, where each unit is coded with a realization of a single variable’ (Getis 2009, p. 256)
 - ‘Given a set S containing n geographical units, it refers to the relationship between some variable observed in each of the n localities and a measure of geographical proximity defined for all $n(n - 1)$ pairs chosen from S ’ (Hubert et al. 1981, p. 224)

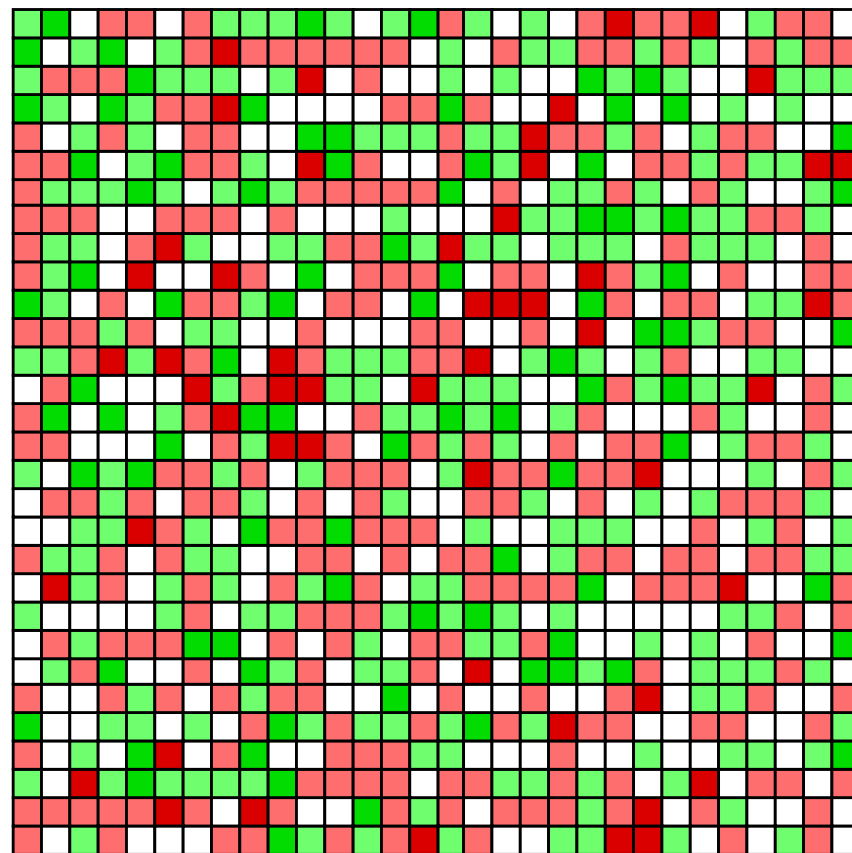
High Peak district biomass index:

ratio of remotely sensed data spectral bands

B_3 and B_4



Spatially autocorrelated



Geographically random

What Is Spatial Dependence?

- Revelli (2003) asks whether the spatial patterns observed in model residuals are a reaction to model misspecification, or if they signal the presence of substantive interaction between observations in space? A similar point is raised by McMillen (2003)
 - “two adjacent supermarkets will compete for trade, and yet their turnover will be a function of general factors such as the distribution of population and accessibility.”
 - “the presence of spatial autocorrelation may be attributable either to trends in the data or to interactions; ... [t]he choice of model must involve the scientific judgement of the investigator and careful testing of the assumptions” (Cliff and Ord, 1981, pp. 141-142)
- One way of testing the assumptions is through changes in the policy context over time, where a behavioural model predicts changes in spatial autocorrelation. If the policy changes, the level of spatial interaction should change too

(borrowed from Roger Bivand)

Spatial Dependence vs Spatial Heterogeneity

- Dependence → Interaction, interdependence
- Heterogeneity → Intrinsic characteristics unevenly distributed over space
- With a cross-section, hard (impossible) to tell whether outcomes arise from interaction or from intrinsic individual characteristics
- Spatial dependence vs spatial heterogeneity
 - Positive spatial autocorrelation → spatial diffusion/spillovers
 - Negative spatial autocorrelation → spatial competition
- Same problem as in social networks: intrinsic individual characteristics or personal interaction?

(borrowed from Daniel Arribas-Bel)

Uses of the Spatial Autocorrelation Concept

- Testing for model misspecification
 - Non-spatially-random residuals indicate misspecification. Moran's I commonly used
- Measuring the strength of spatial effects on a variable
 - Quantifying the spatial effects on both dependent and independent variables
- Testing assumptions of spatial stationarity/heterogeneity
 - E.g., testing assumption that mean and variance do not vary spatially between subgroups
- Identifying spatial clusters
- Quantifying role of distance decay or spatial interaction in spatial autoregressive models
 - Parameters of spatial interaction models (e.g. distance decay) could be obtained through measures of spatial autocorrelation

Uses of the Spatial Autocorrelation Concept (2)

- Understanding the influence of geometry of spatial units on a variable
 - Measures of spatial autocorrelation will change depending on the spatial configuration/spatial scale of units
- Testing hypotheses about spatial relationships...
 - ... between realizations of a single variable. But can also test spatial relations *between* variables! (Wartenberg 1985)
- Weighting the importance of temporal effects...
 - ... by using consecutive (year-by-year) indicators of spatial autocorrelation
- Estimating the effects of a single spatial unit on the others (and vice versa)
 - Based on local indicators of spatial autocorrelation
- Identifying outliers (spatial and non-spatial)
- Designing appropriate spatial samples

Indicators of Spatial Autocorrelation

- The generic cross-product $\Gamma_{ij} = \sum_i \sum_j W_{ij} Y_{ij}$ gives a measure of spatial autocorrelation
 - W is a matrix representing the spatial relations between units
 - Y is a matrix showing the non-spatial relations between units
- If W and Y have similar structure (high-high, low-low values, etc), there is a high degree of spatial autocorrelation
- If either W or Y is random, there will be no spatial autocorrelation
- But this is just a measure, not a test!

Indicators of Spatial Autocorrelation

(2)

- Moran's I (Cliff and Ord 1973)
 - Structured as a Pearson product moment correlation coefficient, plus W
 - Y is a covariance matrix, i.e., the relation between the spatial units is calculated as

$$(y_i - \bar{y})(y_j - \bar{y})$$

- The obtained measure is scaled by

$$\frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]$$

- By convention, $i \neq j$ (or, in \mathbb{R} , $i \neq j$)

Indicators of Spatial Autocorrelation

(3)

- Moran's I (cont.)

- As a result, Moran's I is computed as

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad i \neq j$$

- Its expected value IS NOT ZERO! but $-1/(n - 1)$
- Its variance is calculated differently depending on the assumption (randomness or normality)
- The test takes a different formula for testing regression residuals and depends on the number of independent variables

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \frac{\boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}$$

Indicators of Spatial Autocorrelation

(4)

- Geary's c
 - The sociologists Duncan et al. (1961), in their book *Statistical Geography*, point out that:
 - “Sooner or later in a study of areal variation the investigator runs up against the fact that areal units situated close to each other are more likely to be similar in their characteristics than are areal units which are some distance apart ...” (pp. 128-9)
 - They refer to work on the definition of homogeneous regions, and contrast this with work on regional differentiation. They present Geary's c , which they term the contiguity ratio, because it shows the ratio between variability between neighbours and total variability in a data set (Moran's I is not mentioned)

(borrowed from Roger Bivand)

Indicators of Spatial Autocorrelation

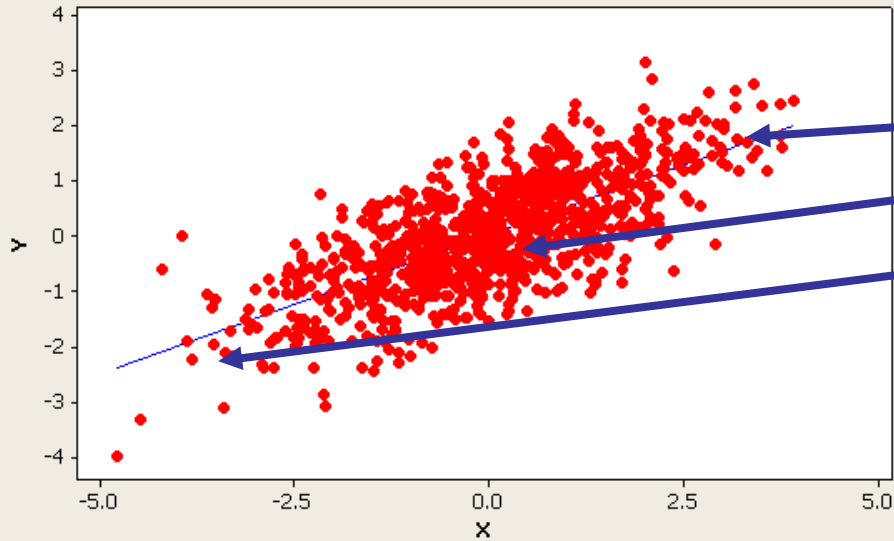
(5)

- Geary's c (cont.)
 - Here the null is that spatial units do not differ from each other, i.e., there is no consistency to their differences
 - Y is made of differences $(y_i - y_j)^2$, and a scale makes the measure normally distributed, and with an expected value of 1
 - < 1 = positive spatial autocorrelation (small differences)
 - > 1 = negative spatial autocorrelation (big differences)
 - Geary's c is therefore negatively correlated to Moran's I

$$c = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n W_{ij} (y_i - y_j)^2}{2W \sum_{i=1}^n (y_i - \bar{y})^2}, \quad i \neq j$$

Describing a scatterplot trend

Scatterplot of Y vs X ($r = 0.717$)



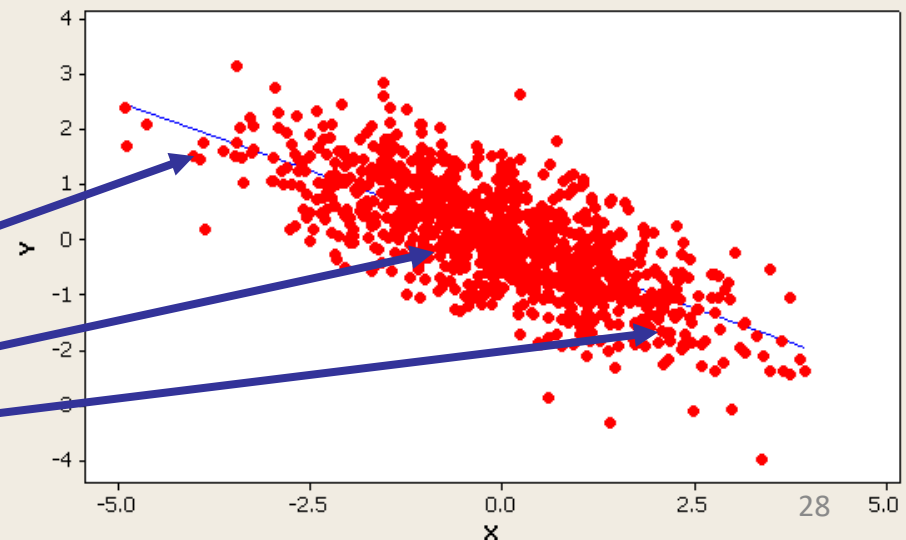
positive relationship:

High Y with High X
& Medium Y with Medium X
& Low Y with Low X

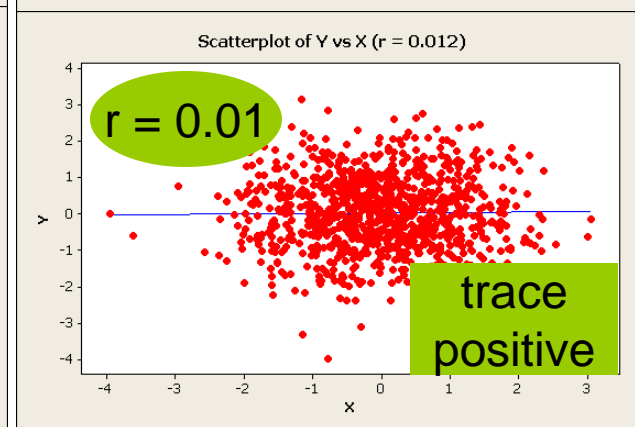
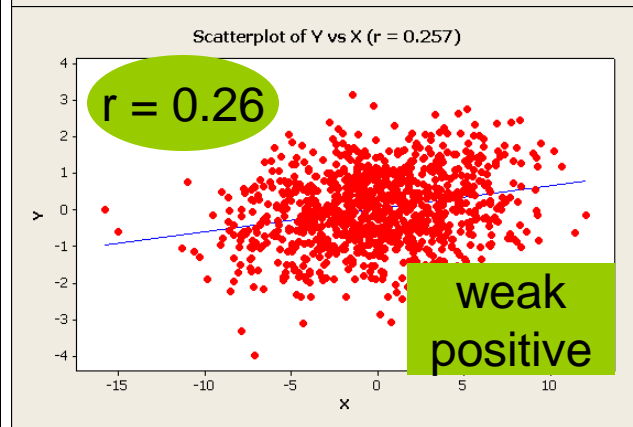
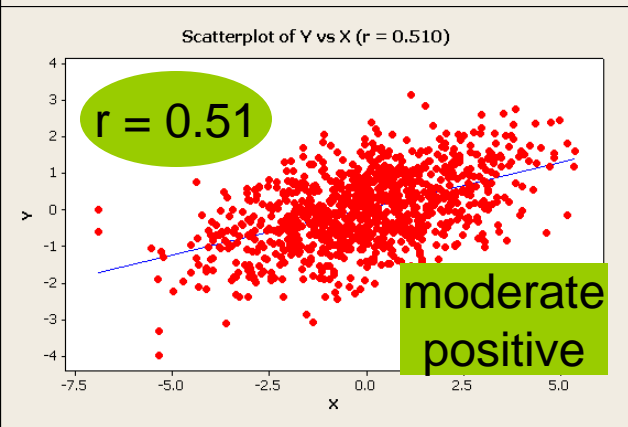
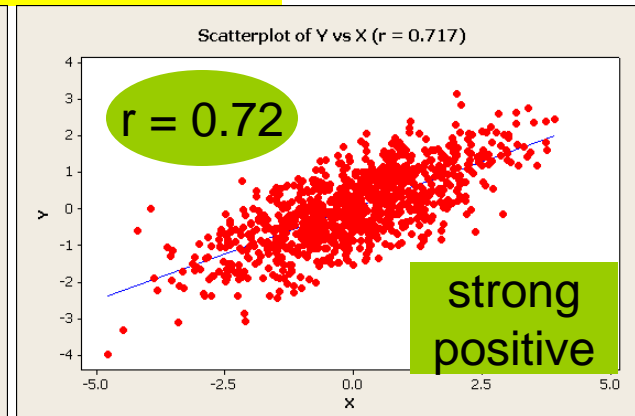
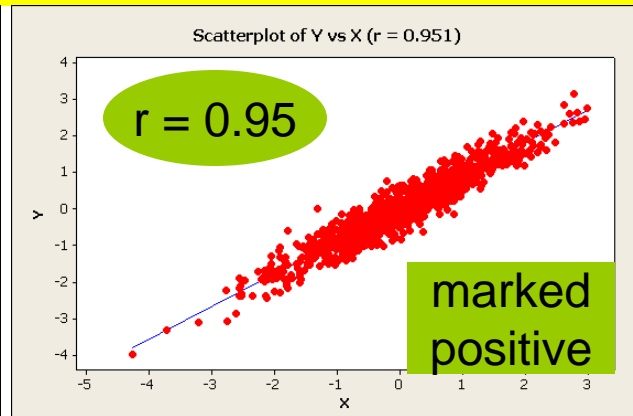
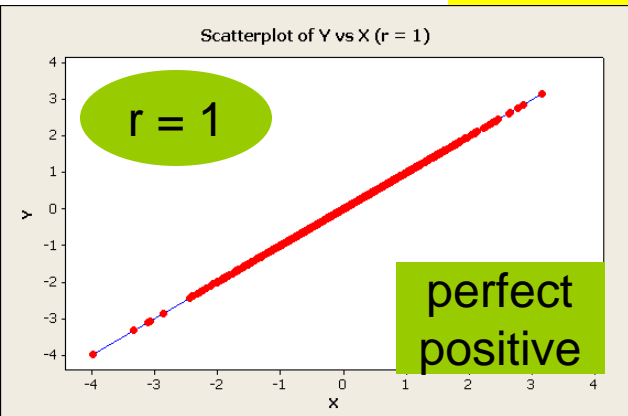
negative relationship:

High Y with Low X
& Medium Y with Medium X
& Low Y with High X

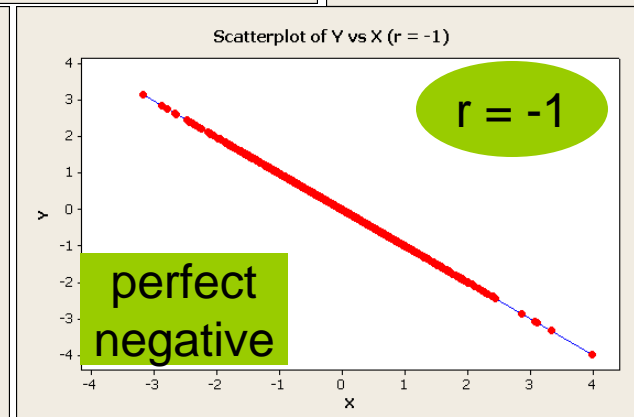
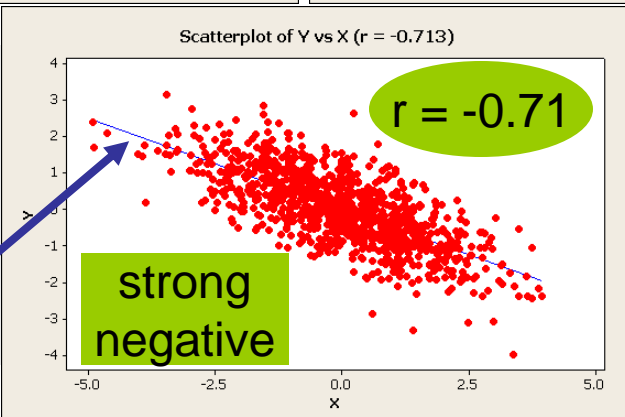
Scatterplot of Y vs X ($r = -0.713$)



Graphic examples

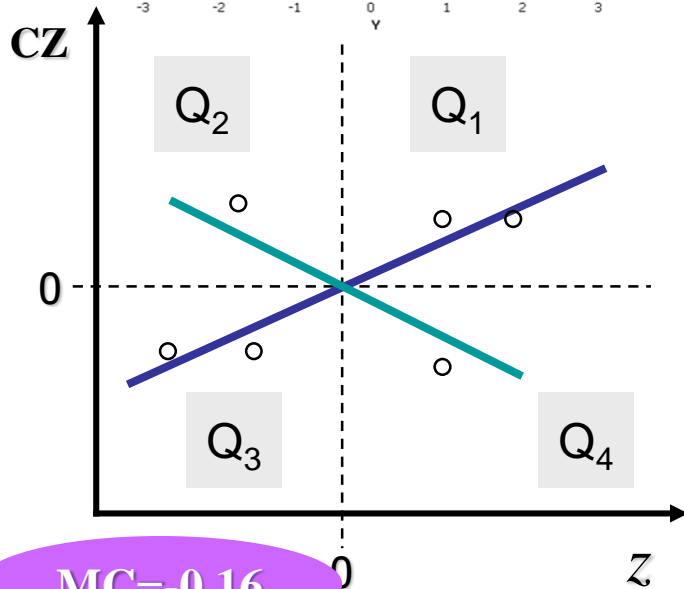
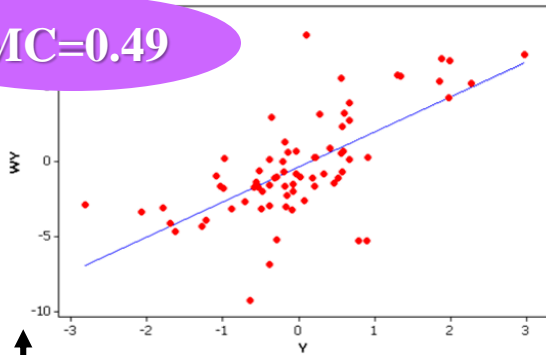


regression trend line in blue

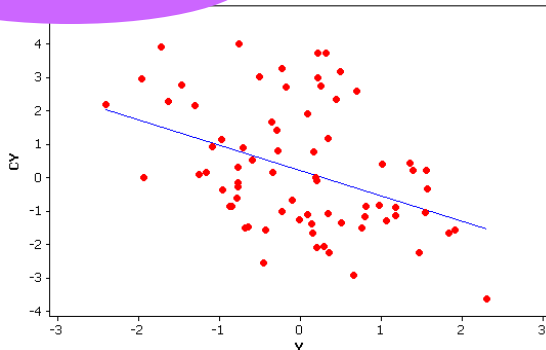


Quadrants of a Moran scatterplot

MC=0.49



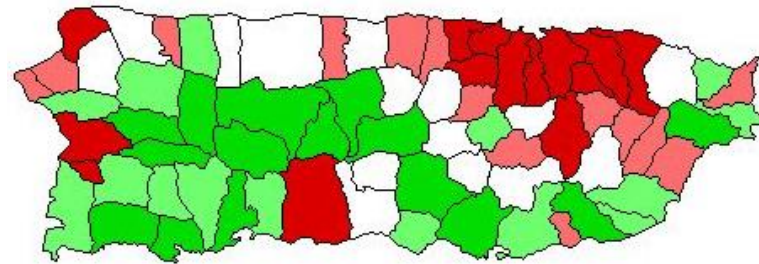
MC=-0.16



Q₁ (values [+], sum of neighboring values [+]): **H-H**

Q₃ (values [-], sum of surrounding values [-]): **L-L**

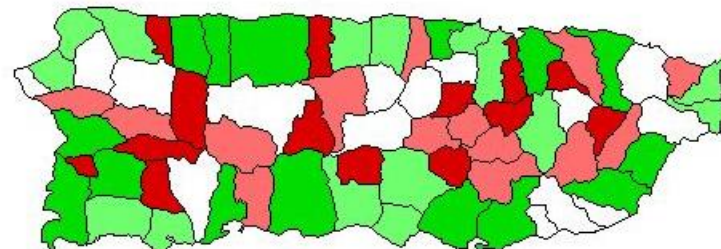
Locations of positive spatial association
 ("I'm similar to my neighbors").



Q₂ (values [+], sum of neighboring values [-]): **H-L**

Q₄ (values [-], sum of neighboring values [+]): **L-H**

Locations of negative spatial association
 ("I'm different from my neighbors").



Local Measures of Spatial Autocorrelation

- Used to investigate certain situational characteristics, i.e., local
- The basis is now: $\Gamma_i = \sum_j W_{ij} Y_{ij}$, $i \neq j$
- Getis and Ord local statistics
 - G_i based on proximity
 - G_i^* based on clustering

$$G_i^*(d) = \frac{\sum_{j=1}^n W_{ij}(d) y_j - W_i^* \bar{y}}{s\{[(nS_{1i}^*) - W_i^{*2}]/(n-1)\}^{1/2}}, \text{ for all } j$$

$$\text{where } W_i^* = W_i + W_{ii} \text{ and } S_{1i}^* = \sum_{j=1}^n W_{ij}^2$$

Local Measures of Spatial Autocorrelation (2)

- The difference between the two measures is in the role of the i th observation, either its influence on the neighbours, or its belonging to a cluster
- G_i^* is related to Moran's I , which may be seen as a weighted average of the local statistics
- Local Indicators of Spatial Association (LISA)
 - Anselin (1995)
 - Aims to decompose the global indicators of Moran and Geary
 - All I_i sum to I
 - Tests available
 - Geary's c also local

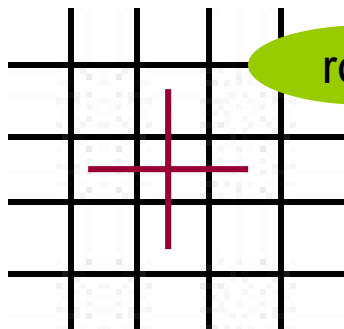
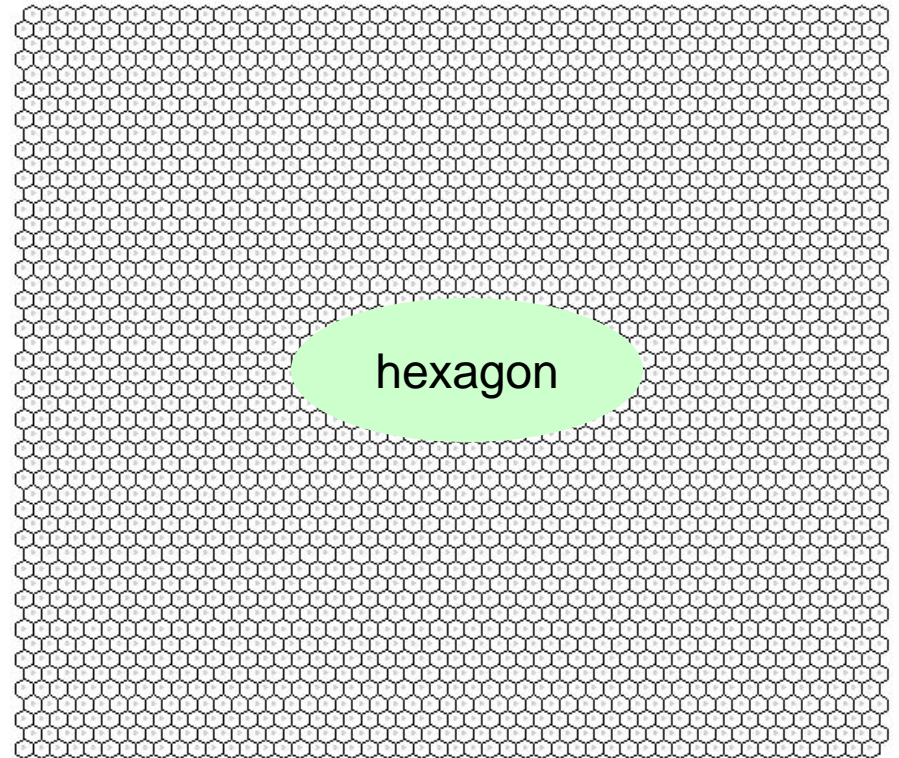
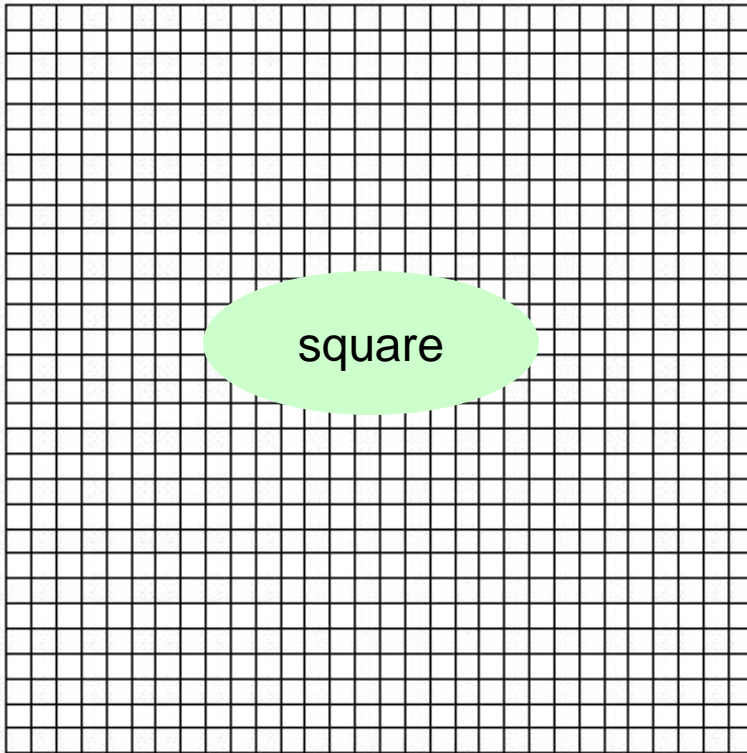
$$I_i = \frac{y_i - \bar{y}}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \sum_{j=1}^n W_{ij} (y_i - y_j),$$

$i \neq j$, for j within d of i

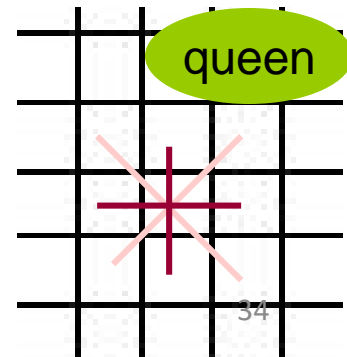
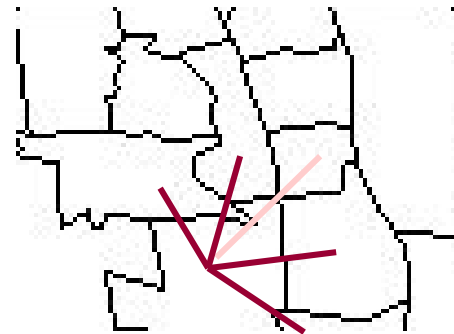
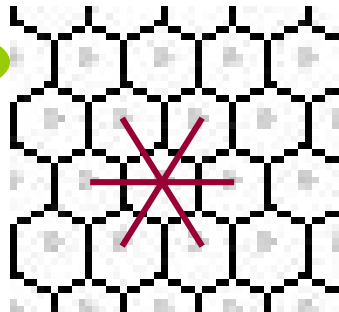
The W Matrix

- But how to construct the W matrix?
- Should it show, because of theoretical considerations, a decay of spatial relations over distance? Many ways to implement that, one of the most common being
$$W_{ij} = d_{ij}^{-\alpha}, \text{ with } \alpha \geq 1$$
- Also the definition of distance may be other than the traditional ones (e.g., relational distance?)
- Other definitions of the W matrix include
 - Contiguous neighbours (queen and rook)
 - Length of shared borders divided by perimeter
 - Bandwidth (distance) to the n th nearest neighbour
 - All centroids within distance d (great circle)
 - Bandwidth distance decay (required for GWR)
 - Gaussian distance decline
 - Derived spatial autocorrelation
 - more... (see for example recent work by Jan Paelinck)

Tessellations (i.e., surface partitionings)

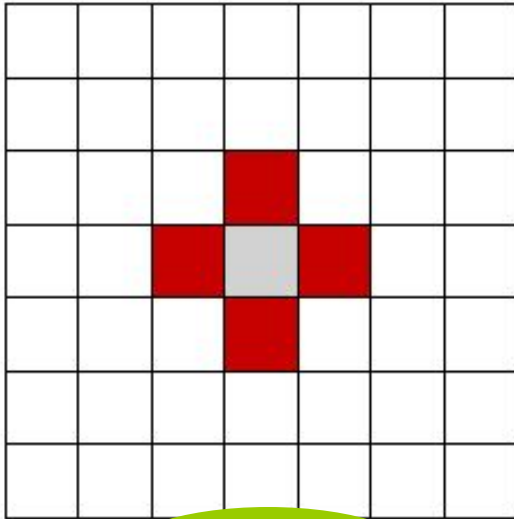


rook

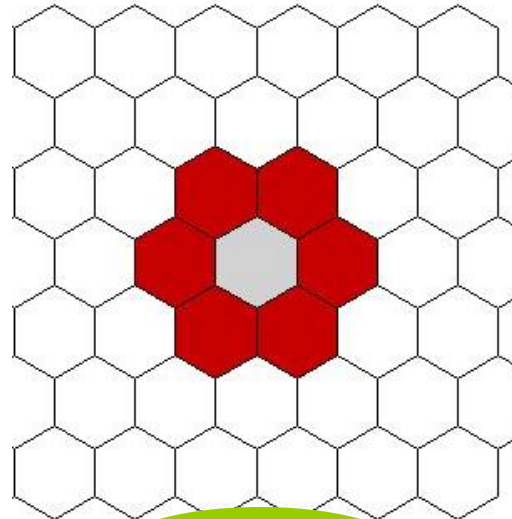


spatial autocorrelation field extents

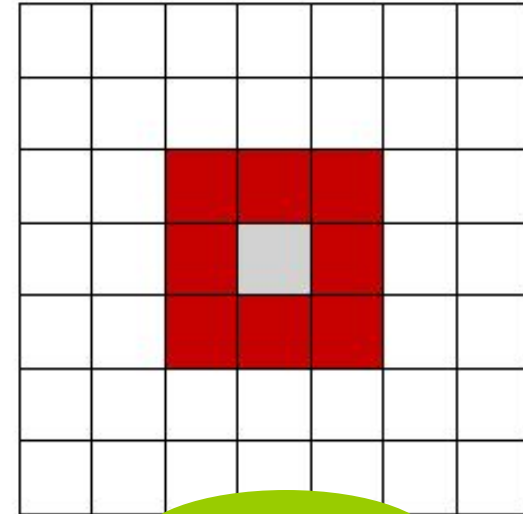
1st
order



rook

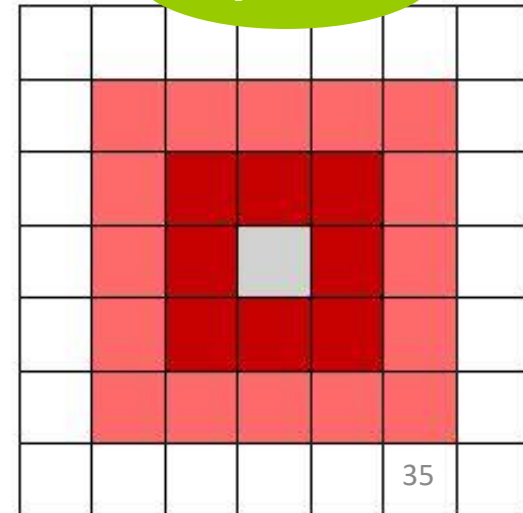
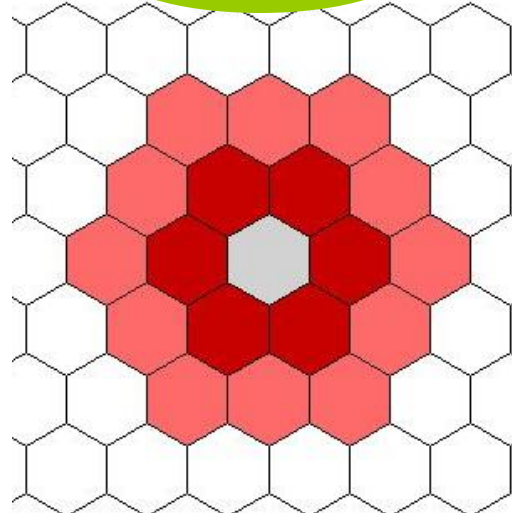
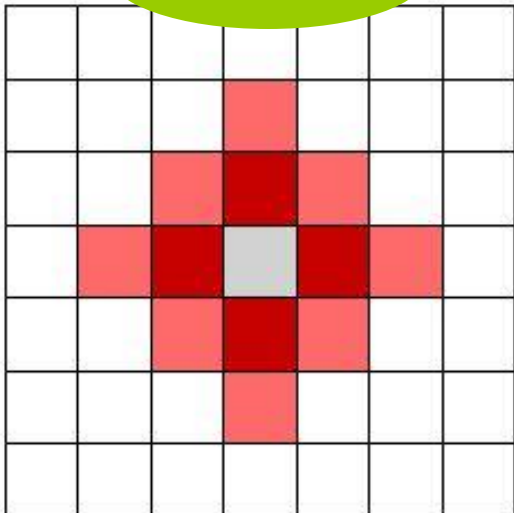


hexagon



queen

2nd
order



The W Matrix (2)

- How to standardize the W matrix?
 - Most commonly used is row-standardization. Each row of W is divided by its marginal sum so that it always sums to one
 - One neighbour: weight = 1
 - Two neighbours: weight = 0.5 each
 - n neighbours: weight = $1/n$ each \rightarrow row-standardization tends to give more weight to regions which are the lone neighbour rather than when there are more
 - Other standardizations: many. Tiefelsdorf (2000) created a family of standardizations, following this expression

$$V_{[q]} = \frac{n}{\sum_{i=1}^n d_i^{q+1}} \cdot D^q \cdot B,$$

- where B is a binary spatial weights matrix, and D^q is a diagonal matrix that contains the d_i^q components belonging to vector $d = B \times \mathbf{1}$
- $q = 0$: C-coding (globally standardized); commonly used in spatial statistics, tends to emphasize spatial objects with a greater linkage degree. Is symmetrical
- $q = -0.5$: S-coding (variance stabilized); tends to even the variation levels of weights
- $q = -1$: W-coding (row-sum standardized); mostly used in spatial econometrics, tends to emphasize the weight of objects with small spatial linkages

Criticisms of Spatial Approaches: Ripley

- During the next decade, we see further contributions from geographers and statisticians, such as Ripley (1981) *Spatial Statistics*; these were at a certain distance from the beginnings of spatial econometrics (Anselin, 2010b)
- As Ripley (1988, p. 2) suggests, the assumptions remove hope that spatial data are a simple extension of time series to a further dimension (or dimensions):
 - the edges of our chosen or imposed study region;
 - how to perform asymptotic calculations and how this doubt impacts the use of likelihood inference;
 - how to handle inter-observational dependencies at multiple scales (both short-range and long-range);
 - stationarity, and discretisation and support.
- Ripley (1988, p. 8) concludes: “(T)he above catalogue of problems may give rather a bleak impression, but this would be incorrect. It is intended rather to show why spatial problems are different and challenging”

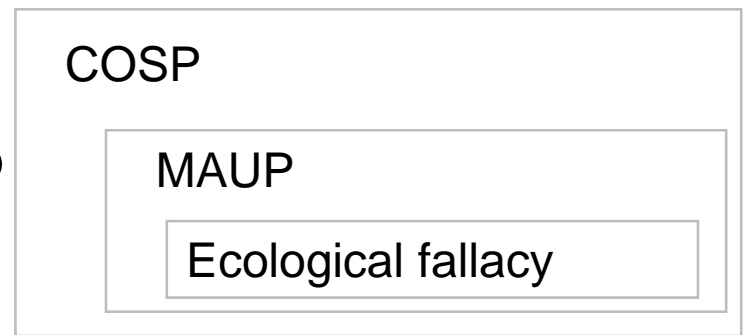
(borrowed from Roger Bivand)

Criticisms of Spatial Approaches: Granger

- Another sceptical voice is that of Granger, thanked specifically for valuable criticism in the acknowledgements to Cliff and Ord (1973). In the cross-sectional setting, possible solutions are “very dependent on the assumption of spatial stationarity. Without this assumption, or something very near it, I can see little hope for model building. ... I am, therefore, very pessimistic about the possibility of model building with purely spatial data” (Granger, 1975)
- Granger’s concluding remark, that: “the quality of the data that geographers, economists or regional scientists have to deal with is too poor for the use of complicated methods to be worthwhile” ignores the need to do one’s best with available data in other arenas of spatial statistics; he subsequently used space-time (panel) data in Giacomini and Granger (2004)

(borrowed from Roger Bivand)

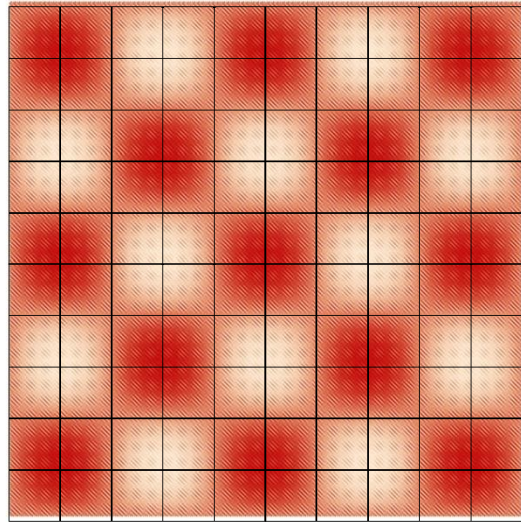
MAUP



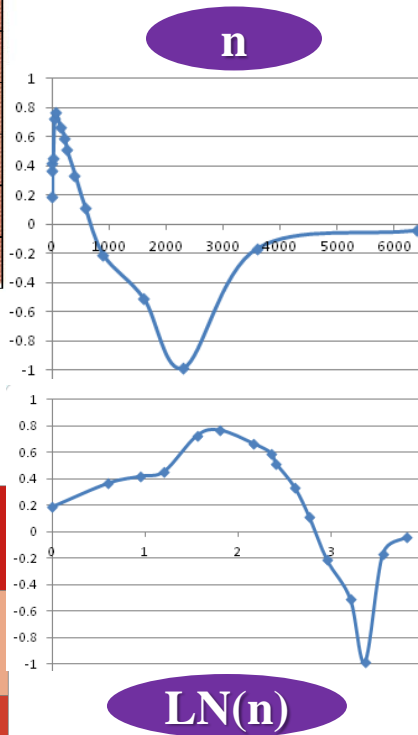
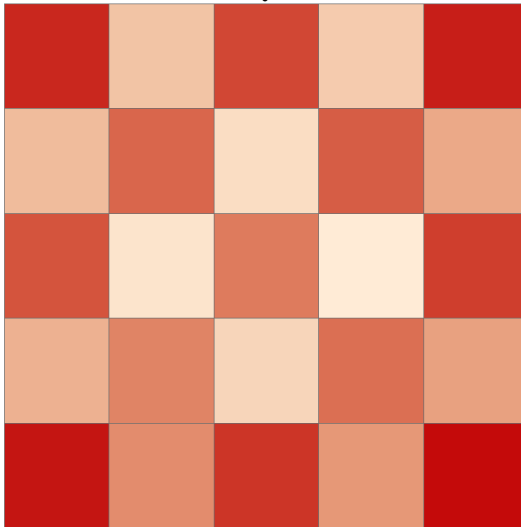
- Gelfand (2010) shows where misaligned spatial data, the modifiable areal unit (MAUP), and change of support (COSP) may take us; see also Stan Openshaw's MAUP CATMOG: <http://qmrg.org.uk/catmog/>, and Haining (2010)
- Briant and Lefourcade (2010) examine whether the size and shape of spatial units jeopardizes estimates in economic geography, finding that size matters more than shape, but that both matter less than model specification
- Menon (2012) uses MAUP to create alternative graph partitions to the functional regions used in analysis
- Arbia and Petrarca (2013) show how MAUP is relevant also for flow data

(again, borrowed from Roger Bivand and modified)

The MAUP: # areal units; shape constant



⋮



dimension	average # pixels	MC
240-by-240	1	0.188
120-by-120	4	0.366
80-by-80	9	0.418
60-by-60	16	0.456
40-by-40	36	0.726
30-by-30	64	0.765
20-by-20	144	0.666
16-by-16	225	0.593
15-by-15	256	0.513
12-by-12	400	0.338
10-by-10	576	0.116
8-by-8	900	-0.214
6-by-6	1,600	-0.507
5-by-5	2,304	-0.982
4-by-4	3,600	-0.164
3-by-3	6,400	-0.043

Ecological Fallacy

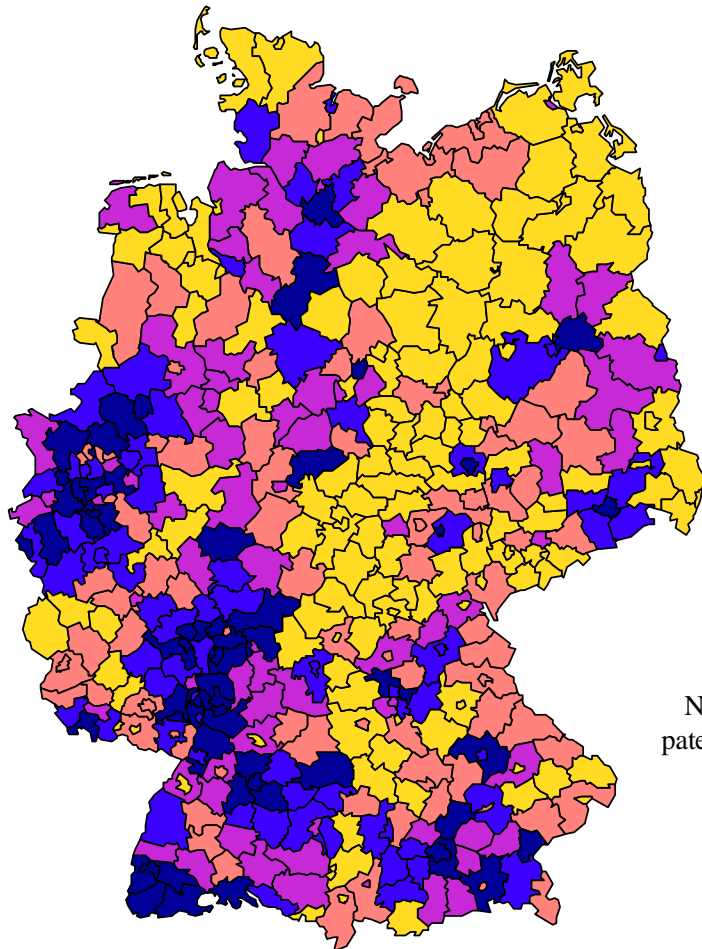
- Wakefield and Lyons (2010) give a survey of the ecological fallacy in connection with spatial aggregation; the point of concern is the extension of aggregated inference to individuals within the aggregates
- They motivate their survey by looking at county asthma disease counts and PM2.5 air pollution; of course, within-county variability in the included variable is challenging, and inferring to the individual is hard
- Haining (2010) also stresses that making statistical inference about individuals based on aggregate data is flawed

(Roger Bivand)

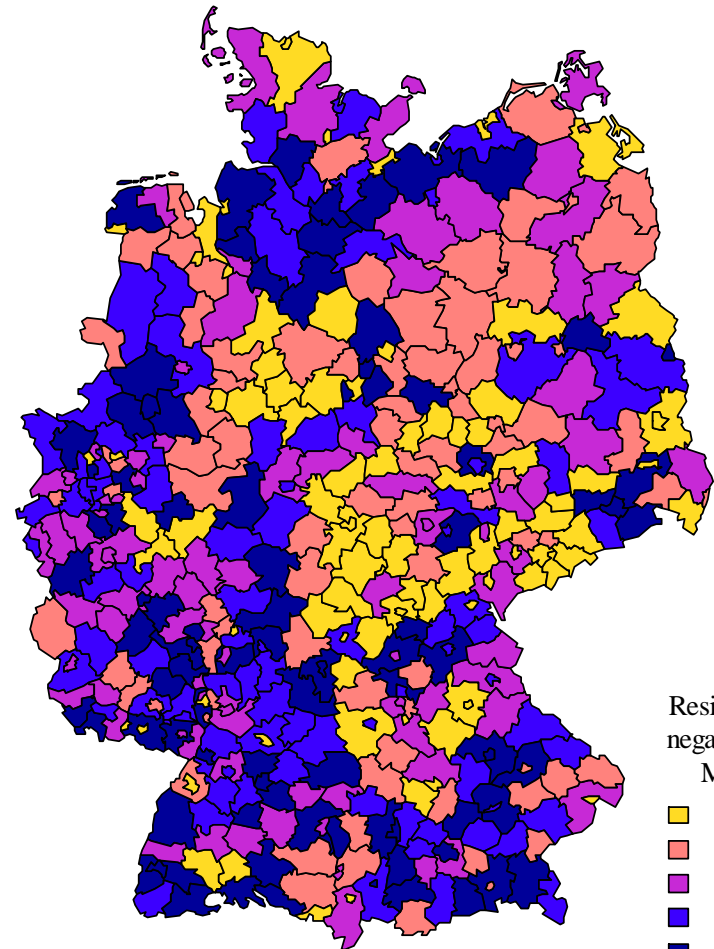
Spatial Autocorrelation and Regressions

- What are the consequences of spatial autocorrelation on regression models?
 - Ideally, all spatial autocorrelation in the dependent variable should just be explained by the model, right?
 - Often, however, regression residuals from models analysing georeferenced data, ARE spatially correlated, indicating *mispecification*
 - Ignoring this mispecification can also lead to bias and/or inefficient parameter estimates
 - A number of econometric models have been suggested to cope with the DGP behind spatially correlated data

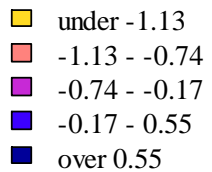
Example: Dependent Variable and Regression Residuals



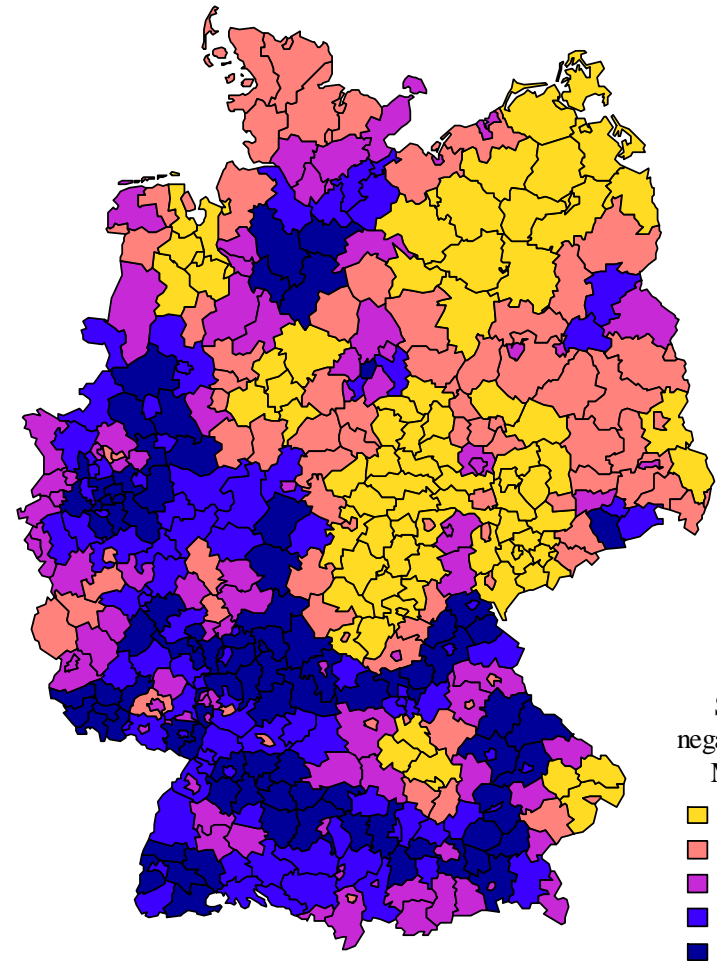
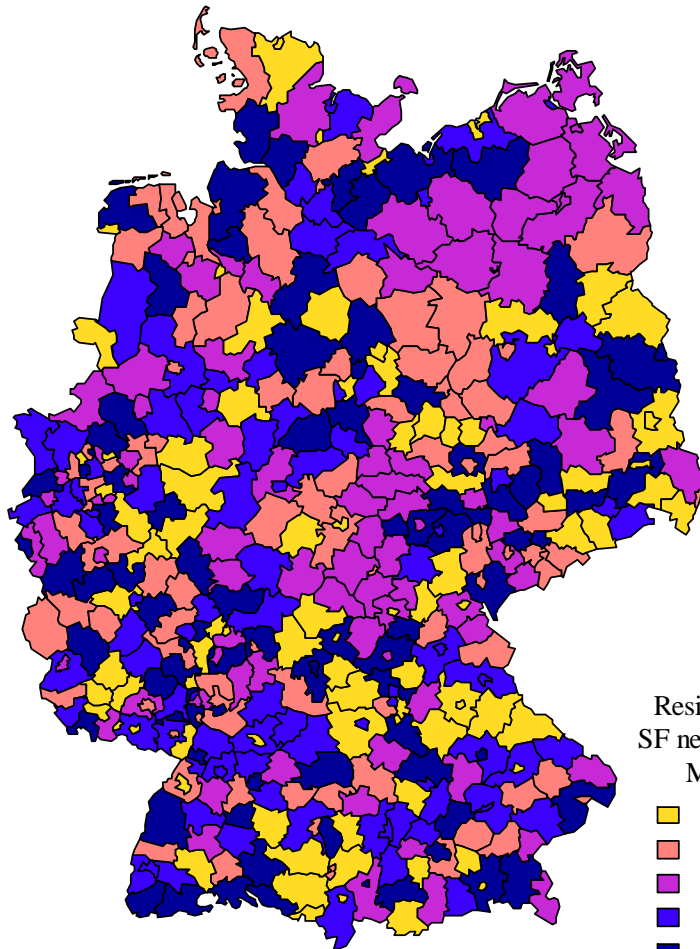
Legend:
Nanomaterials
patent applications
MI = 0.28



Legend:
Residuals from a
negative binomial
MI = 0.21



Example: Spatial Filter Residuals and Spatial Filter



Spatial Autocorrelation and Regressions (2)

- From a historical perspective, merit in starting spatial econometrics is given, also in Anselin (1988), to *Spatial Econometrics*, by Paelinck and Klaassen (1979)
- Reading Paelinck and Klaassen (1979, p. viii), we see that the programme of research into the space economy undertaken at the Netherlands Economic Institute led first to the publication of Paelinck and Nijkamp (1975), and then to Klaassen et al. (1979); all three books were published in the same series and appear to reject the core concerns of economists at the Institute doing research on regionalized national macro-economic models
- Paelinck and his colleagues were aware that an aspatial regionalization of national accounts, of input-output models, or transport models, might prejudice policy advice and outcomes through inadequate and inappropriate calibration
- In the motivation for spatial econometric models given in Paelinck and Klaassen (1979, pp. 2-3), consumption and investment in a region are modelled as depending on income both in the region itself and in its contiguous neighbours, termed a “spatial income-generating model”.

(borrowed from Roger Bivand)

Spatial Autocorrelation and Regressions (3)

- Approaches to spatial econometrics
 - Define a model that flexibly accommodates a range of possible different DGPs (encompassing model), e.g. the spatial Durbin model
 - Resort to economic theory to obtain a theoretical model and justify specification chosen, e.g. spatial spillovers/externalities, dependence of flows in trade data, etc. (see e.g. work by Behrens et al., LeSage, etc.)
 - Use it to accommodate econometrically heteroskedasticity, omitted variable bias, latent variables
 - Let the data speak, estimate different models, and take a probabilistic approach, e.g. by means of Bayesian model averaging?

Spatial Autocorrelation and Regressions (4)

- Can (the need for) spatial econometrics be eliminated?
 - Knowing your data is important: Andersson and Gråsjö (2009) show that ‘a representation of space reflecting the potential of physical interaction between localities by means of accessibility variables on the “right-hand-side” ... captures substantive spatial dependence’ and can be estimated by OLS
 - Similar results by Osland and coauthors, showing that spatial autocorrelation in house pricing models is removed by including distance to CBD or to employment outside of CBD
 - Accessibility matters, and spatial dependence often reflects it

Spatial Autocorrelation and Regressions (5)

- LeSage and Pace (2009, 2010) give a simple example of how a spatial model may emerge
 - Say some regions are all intersected by a highway
 - The commuting time of people living in region 1 will not only depend on density in that region (and distance to work), since people from other regions (partially) use the same road as well
 - Assuming that commuting time of (people in) region 1 is independent of the one of (nearby) region 2 is unrealistic
 - It's then easy to imagine a model with an explicit treatment of spatial dependence, e.g., a so-called spatial lag model (SAR)

$$y_i = \rho \sum_{j=1}^n W_{ij} y_j + \sum_{r=1}^k X_{ir} \beta_r + \varepsilon_i$$

Spatial Autocorrelation and Regressions (6)

- Clearly, if $\rho = 0$, the model scales back to non-spatial, therefore the interest in the statistical significance of ρ
- Grouping for y , the model can be rewritten – in matrix form – as

$$\mathbf{y} = (\mathbf{I}_n - \rho\mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \rho\mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

$$E(\mathbf{y}) = (\mathbf{I}_n - \rho\mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta}$$

- Alternatively, spatial dependence might arise only in the error term, giving way to the so-called spatial error model (SEM)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \text{ where } \mathbf{u} = \rho\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \rho\mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

Spatial Autocorrelation and Regressions (7)

- Other models
 - Spatial Durbin model (WX) (SDM)
 - Manski model (GNS) (complex DGP)
 - Spatial moving average model
 - SARAR(1,1) model (Kelejian and Prucha 1998)
 - Matrix exponential model
 - Spatial models with instrumental variables (Kelejian and Prucha 1999)
 - Spatial simultaneous equations (Kelejian and Prucha 2004; Gebremariam et al. 2011)
 - Spatial simultaneous equations in continuous time (Oud et al. 2012)
 - Getting rid of W ... (Folmer and Oud 2008)
 - more...

Spatial Autocorrelation and Regressions (8)

- Important distinction between SAR and SEM
 - SEM has same expectation as OLS (but in small samples, OLS is inefficient), while SAR and SDM don't, so least squares estimation is not appropriate (biased for both β and ρ)
 - Maximum likelihood estimation (Anselin 1988)
 - Bayesian (MCMC) estimation (LeSage 1997)
 - Others: e.g., two-stage least squares (with instrument for Wy) - GMM, or entropy-based estimation

Spatial Lag Model

- A closer look at the spatial lag model shows that the $(\mathbf{I}_n - \rho\mathbf{W})^{-1}$ inverse can be rewritten as the infinite series expansion $\mathbf{I}_n + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3 + \dots$ leading [with $\text{abs}(\rho) < 1$] to the model

$$\mathbf{y} = (\mathbf{I}_n - \rho\mathbf{W})^{-1} \iota\alpha + (\mathbf{I}_n - \rho\mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

$$\mathbf{y} = \frac{1}{1 - \rho} \iota\alpha + \boldsymbol{\varepsilon} + \rho\mathbf{W}\boldsymbol{\varepsilon} + \rho^2\mathbf{W}^2\boldsymbol{\varepsilon} + \rho^3\mathbf{W}^3\boldsymbol{\varepsilon} + \dots$$

since the series $\iota\alpha + \rho\mathbf{W}\iota\alpha + \rho^2\mathbf{W}^2\iota\alpha + \rho^3\mathbf{W}^3\iota\alpha + \dots$ converges to $(1 - \rho\mathbf{W})^{-1}\iota\alpha$ (because \mathbf{W} is row-standardized, $\mathbf{W}^q\iota = \mathbf{W}\iota = \iota$ for $q \geq 0$)

- \mathbf{W}^2 reflects the neighbours of the neighbours (second order neighbours)
 - Since in space correlation runs multidirectionally, they will include region i itself and therefore \mathbf{W}^2 has positive values on the diagonal. Instead, in time series, the lag operator L is strictly triangular and also L^2 will be keeping zeros on the diagonal. Distinctive aspect of spatial econometrics!
- Because ρ is < 1 in abs. terms, the influence of more far away neighbours is smaller, decaying geometrically

Spatial Lag Model (2)

- In the highway example, if population density doubles in region 2 (R2) of the seven regions (R1-R7), and the spatial lag model suggests fitted values for which $\rho = 0.642$
- Predicted commuting time has risen in EVERY region, less and less the more they are away from R2
- In a non-spatial model, only y_{R2} would increase...
- $(\mathbf{I}_n - \rho\mathbf{W})^{-1}\beta$ acts as a multiplier, applying higher-order effects to explanatory variables X_r
 - As a consequence, the usual interpretation of the regr. coeffs does not hold
 - $\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'(\mathbf{I}_n - \rho\mathbf{W})\mathbf{y}$ and, as a consequence, OLS is (upward) biased and inconsistent
 - The estimated increase in commuting time in R2 is +11.02 (0 elsewhere)

Region	y	y^*	$y^* - y$
R1	42.01	44.58	2.57
R2	37.06	41.06	4.00
R3	29.94	31.39	1.45
R4	26.00	26.54	0.53
R5	29.94	30.14	0.20
R6	37.06	37.14	0.07
R7	42.01	42.06	0.05

Spatial Lag Model (3)

- We may take:
 - +4 mins (increase of commuting time in region 2): as the direct impact from the change in density $\rightarrow \partial y_i / \partial X_{i2}$
 - +4.87 mins (sum of all remaining commuting time increases): as the indirect impact $\rightarrow \partial y_j / \partial X_{i2}$
 - +8.87 mins (total increase): as the overall impact
- If $S_r(\mathbf{W}) = (\mathbf{I}_n - \rho \mathbf{W})^{-1}$, then the DGP of the SAR model can be written as
 - $y = \Sigma_r S_r(\mathbf{W}) X_r + (\mathbf{I}_n - \rho \mathbf{W})^{-1} \varepsilon$, with
 - $E(y) = \Sigma_r S_r(\mathbf{W}) X_r$
- Complication is the derivative of y_i with respect to X_{jr} , since
 - $\partial E(y_i) / \partial X_{jr} = S_r(\mathbf{W})_{ij}$ [i.e. the ij th element of matrix $S_r(\mathbf{W})$]
- This shows how the standard interpretation of regression coefficients as partial derivatives (marginal effects) is not valid for the spatial lag model (OLS is upward biased), because change in an X in one region will affect all (more) regions \rightarrow impact measures!
- $S_r(\mathbf{W})$ can be approximated using traces of powers of the spatial weights matrix as well as analytically

Interpretation of Spatial Models

- *Time-dependence interpretation*: economic agents make decision based on behaviour of other agents in previous periods $\rightarrow \mathbf{W}y_{t-1}$. As in time series, y_{t-1} can be replaced by its counterpart in $(t-2)$, and so on... giving way to a process with increasing powers of \mathbf{W} and ρ and a serially correlated, geometrically decaying spatial dependence
- *Omitted variables interpretation*: omitted variables arise often in regional data because of unmeasured/unmeasurable factors (amenities, highway accessibility, etc.). Take an unobserved variable z following an autoregressive process $z = \rho \mathbf{W}z + r$
 - If the true model is simply $y = x\beta + z\theta = x\beta + (\mathbf{I}_n - \rho \mathbf{W})^{-1}u$, then OLS is consistent [$E(y) = x\beta$] but inefficient
 - But if z and x , as it often happens, are correlated (e.g. $u = x\gamma + v$), then the DGP gets more complicated [$E(y) = x\beta + \dots$ (a function of xy)] and OLS is biased, giving rise to the so-called spatial Durbin model, with both $\mathbf{W}y$ and $\mathbf{W}x$
 - $y = \rho \mathbf{W}y + x(\beta + \gamma) + \mathbf{W}x(-\rho\beta) + v$

Interpretation of Spatial Models (2)

- *Spatial heterogeneity interpretation:* in panel literature it is frequent to allow heterogeneous intercepts, but in a cross-sectional model? This can be done by allowing for a spatially structured random error
 - If α is $(n \times 1)$ and $\alpha = \rho \mathbf{W}\alpha + \varepsilon = (\mathbf{I}_n - \rho \mathbf{W})^{-1} \varepsilon$, then we have the DGP for the spatial error model $\mathbf{y} = \mathbf{X}\beta + (\mathbf{I}_n - \rho \mathbf{W})^{-1} \varepsilon$
 - Again, if α is correlated with \mathbf{X} , it can be shown that we have again the spatial Durbin model
- *Externalities interpretation:* in the real estate market, characteristics of neighbouring houses (\mathbf{WX}) often have an effect on house value, motivating a model of the type $\mathbf{y} = \alpha \mathbf{1} + \mathbf{X}\beta_1 + \mathbf{WX}\beta_2 + \varepsilon$
- *Model uncertainty interpretation:* Bayesian model averaging (not discussed here) [LeSage and Pace (2009) show that e.g. a linear combination of the SAR and SEM models leads again to the SDM]

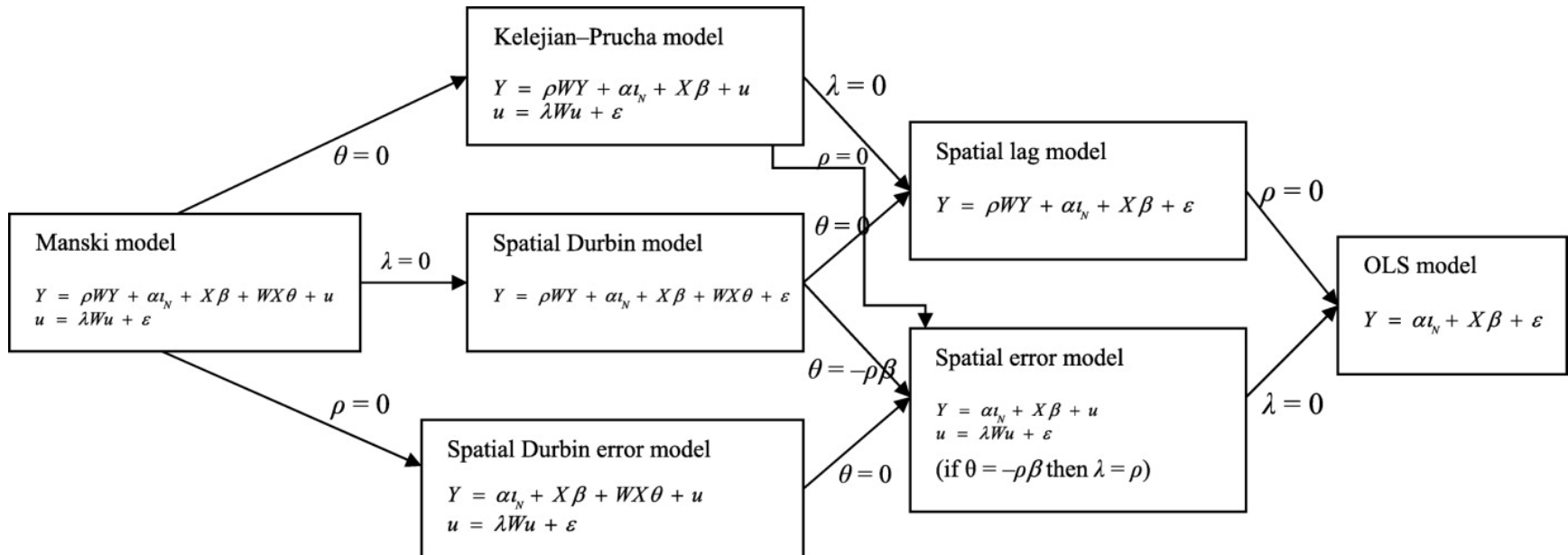
Direct and Indirect Impacts

- As mentioned, in models containing endogenous spatial lags (e.g. SAR, SDM), the derivative of y_i with respect to x_{jr} is not 0, but
 - $\partial y_i / \partial x_{jr} = S_r(\mathbf{W})_{ij}$and the derivative of y_i with respect to x_{ir} is not β_r , but
 - $\partial y_i / \partial x_{ir} = S_r(\mathbf{W})_{ii}$
- \mathbf{W}^2 , implying neighbours of neighbours, will also have non-zero diagonal elements (you're a neighbour to your neighbour) \rightarrow feedback loops. Their extent will depend on
 - the position of the regions in space
 - the degree of connectivity implied by \mathbf{W}
 - the value of ρ , measuring the strength of spatial dependence
 - the value of the parameters β and θ
- Within the $(n \times n)$ matrix $S_r(\mathbf{W})$, diagonal elements will measure direct impacts, and off-diagonal elements the indirect ones
- It is then desirable to compute, from the output of a spatial lag model (or similar), the actual estimated effects from changes in the explanatory variables
- All measures can be obtained from the matrix $S_r(\mathbf{W}) = (\mathbf{I}_n - \rho \mathbf{W})^{-1}(\mathbf{I}_n \beta_r + \mathbf{W} \theta_r)$, where θ is the parameter for \mathbf{WX}

Direct and Indirect Impacts (2)

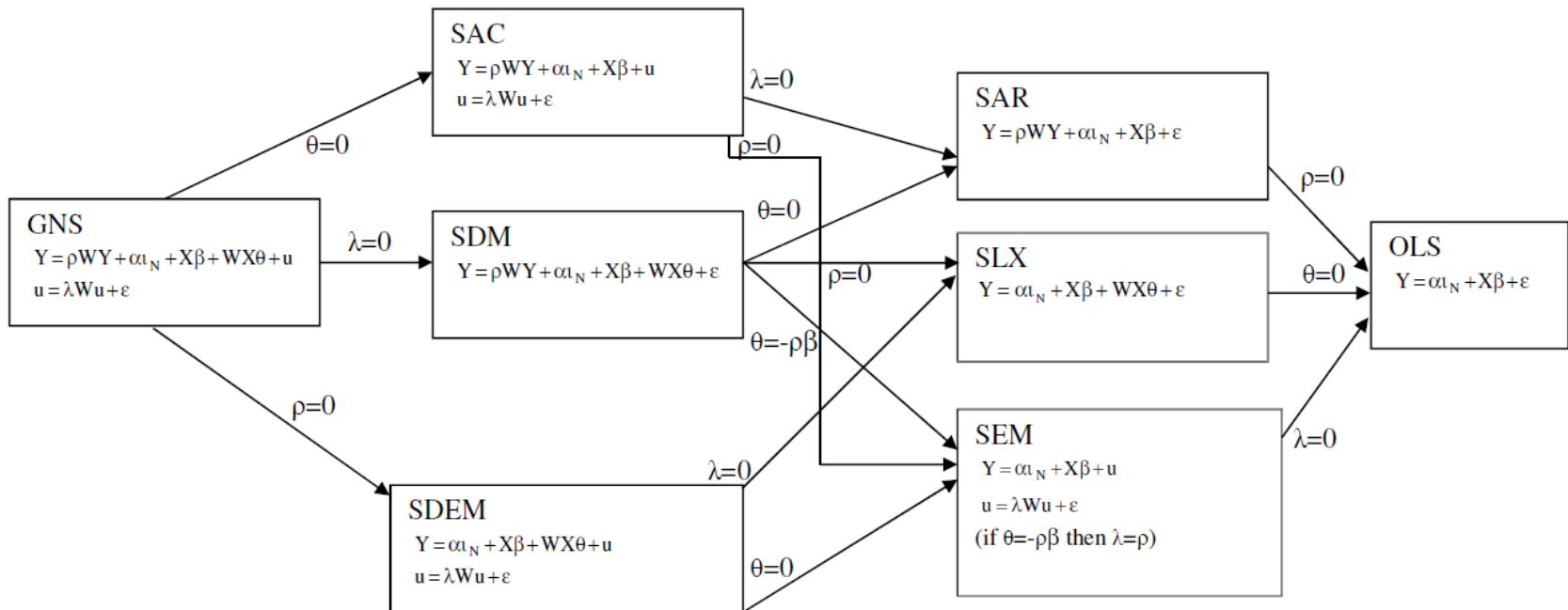
- The following direct/indirect impact measures can be identified:
 - *Average direct impact*: impact of changes in i th observation of x_r on y_i . Obtained as the mean of diagonal elements
 - *Average total impact to an observation*: total impact on individual observation y_i resulting from changing X_r by the same amount across all n observations. Obtained as the mean of row-sums of $S_r(\mathbf{W})$
 - *Average total impact from an observation*: total impact over all y_i from changing X_r in the j th observation. Obtained as the mean of column-sums of $S_r(\mathbf{W})$
 - *Average indirect impact*: difference between average total impact and average direct impact
- The two average total impact measures are numerically equal but allow for different interpretations. They are the average of all derivatives of y_i with respect to x_{jr} , for all i and j
- For the SDM, greater heterogeneity. The presence of $\mathbf{W}\theta_r$ allows spillovers from a change in each explanatory variable to differ (the SAR has a common global multiplier)
- LeSage and Pace (2009) also show how it is possible to partition the effects by order of neighbours

A Taxonomy of Spatial Models



A Taxonomy of Spatial Models (2)

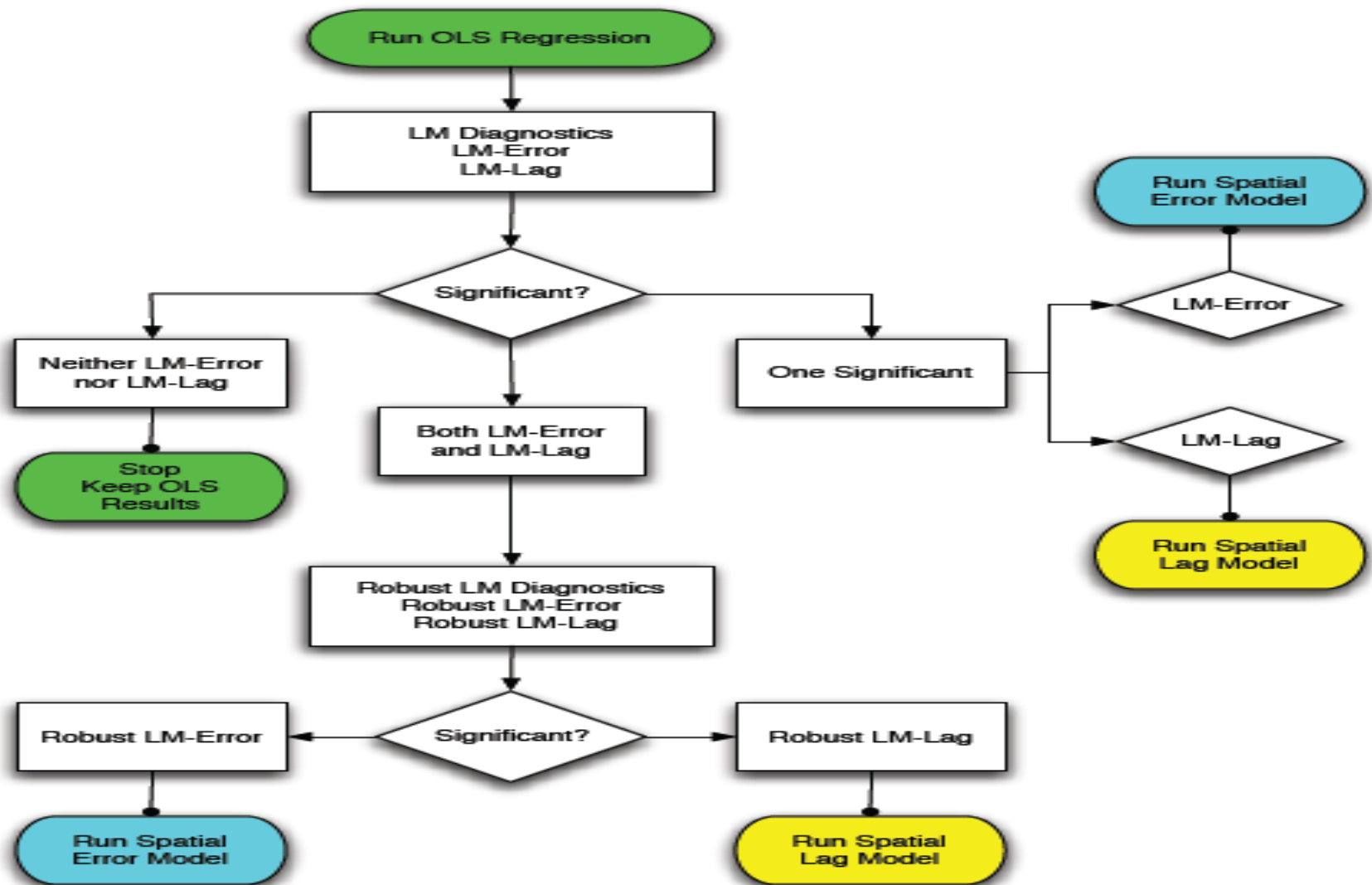
Figure 1. Comparison of different spatial econometric model specifications



Specification Search

- Tests of OLS vs spatial models have been developed in the literature (specific-to-general):
 - Burridge (1980) proposed a Lagrange multiplier test of SEM vs OLS, mathematically related to Moran's I , and which only needs OLS residuals and W
 - Anselin (1988) added the SAR vs OLS test, which like the Burridge test, did not need ML estimation
 - Robust versions of these tests have been given by Anselin et al. (1996)
 - Florax and Folmer (1992), instead, offer tests against erroneously omitted, spatially lagged, explanatory variables
- LeSage and Pace (2009) stress the weakness of such tests, since they ignore the SDM, which nests both spatial lag and spatial error model... and suggest Bayesian model comparison
- It is also possible to use a Hendry-type strategy (general to specific), fitting the most general model first, then testing restrictions by means of LR tests (though they don't extend naturally to non-nested models)
- McMillen (2003) stresses how “tests for spatial autocorrelation also detect functional form misspecification, heteroskedasticity, and the effects of missing variables that are correlated over space”

Anselin's Specification Search Strategy



Some Recent Developments

- Automatic selection of spatial weights matrix (Seya et al. 2013)
 - Model selection based on MCMC
- Standardized and heteroskedasticity-and-non-normality robust LM tests (Baltagi and Yang 2013a,b)
 - Tests which are robust against distributional misspecification and against heterosk. and non-normality
- Spatial autoregressive models for geographically hierarchical data (Dong and Harris 2014)
 - Spatial weights matrix and fixed/random effects specifications to account for data at different (but nested) levels of geographical aggregation
- Spatial quantile regression (McMillen 2012)
 - Proposes a spatial version of quantile regression (short book on Springer Briefs in Regional Science, method shown in R)

Some Recent Developments (2)

- Spatial causality testing (Herrera et al. 2014)
 - Testing for causality in a group of variables with a spatial framework (no time here) by means of symbolic entropy
- Spatial Hausman test (Pace and LeSage 2008)
 - Test on spatial error model (SEM)
- LM tests for panel models (Debarsy and Ertur 2010)
 - LM tests against SAR and SEM in fixed effects panel models
- Spatial model selection in an SUR setting (López et al. 2014)
 - Discussion of general-to-specific and specific-to-general model selection strategies with a spatial SUR estimation framework
- Tests for spatial error dependence in probit models (Amaral et al. 2013)
 - Discussion of general-to-specific and specific-to-general model selection strategies with a spatial SUR estimation framework

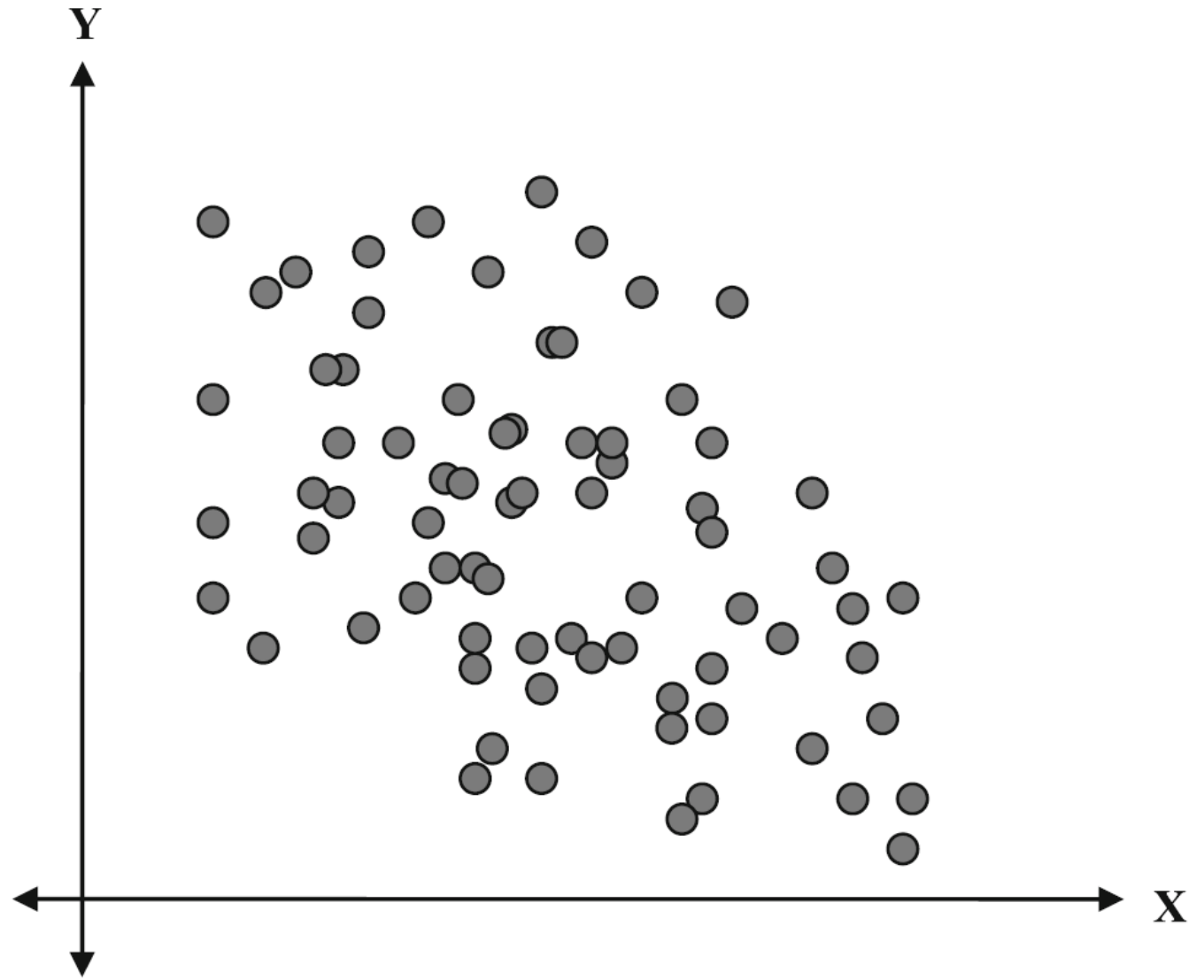


Fig. 1 Spatial structure of observations pooled on only one layer

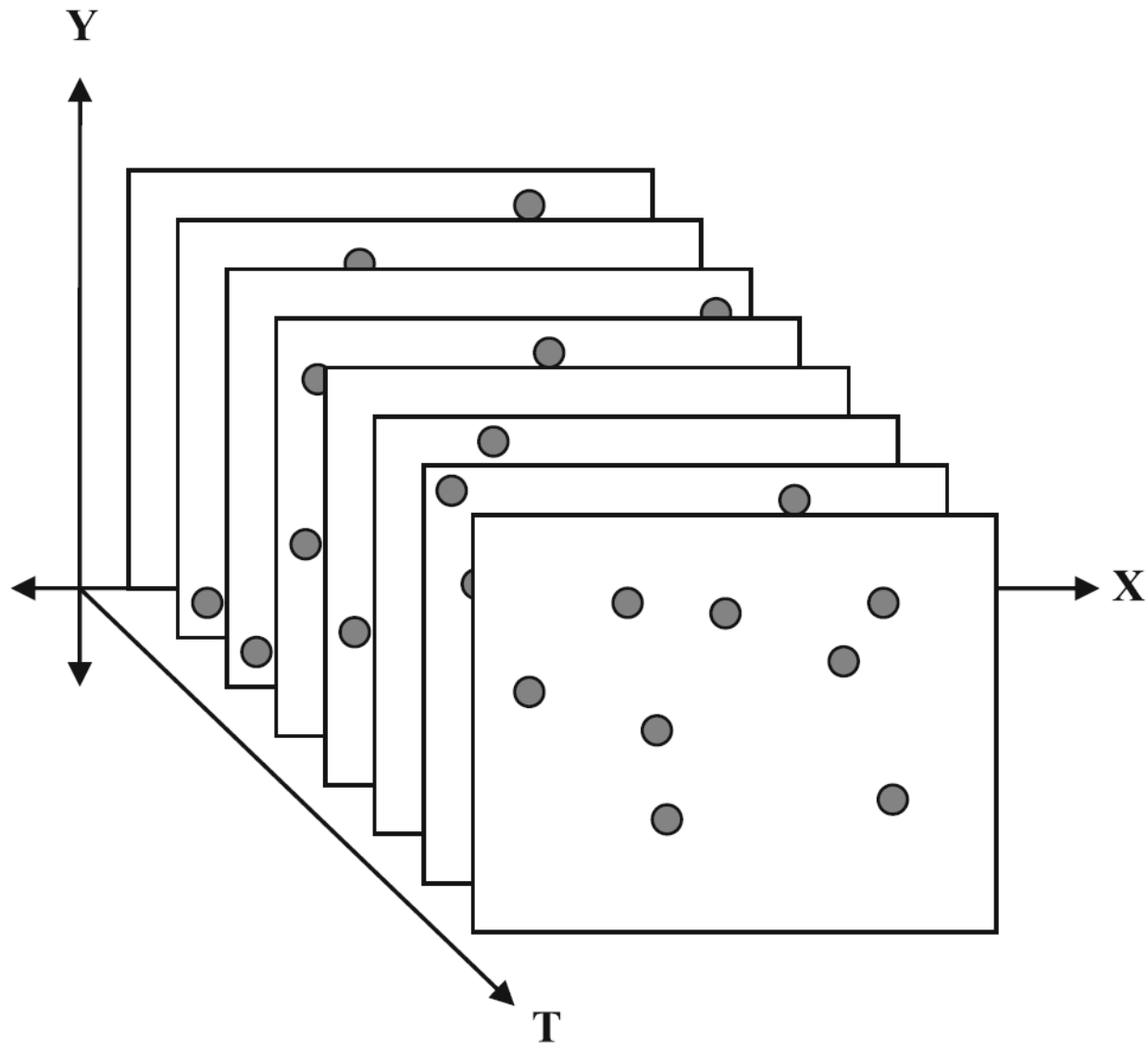


Fig. 2 Spatial observations collected in a discrete time structure (pooled layers)

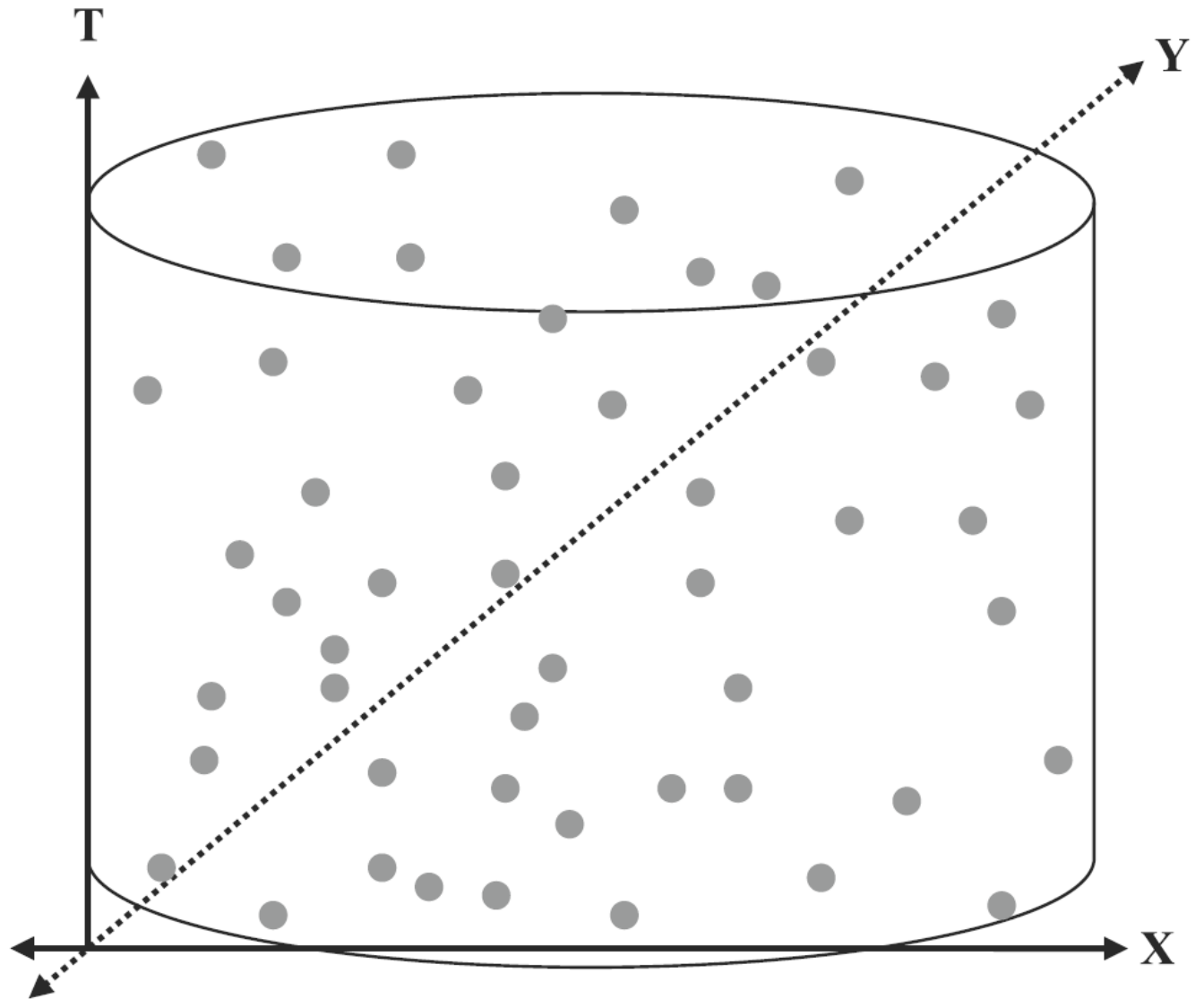


Fig. 3 Spatial observations collected in a continuous time structure

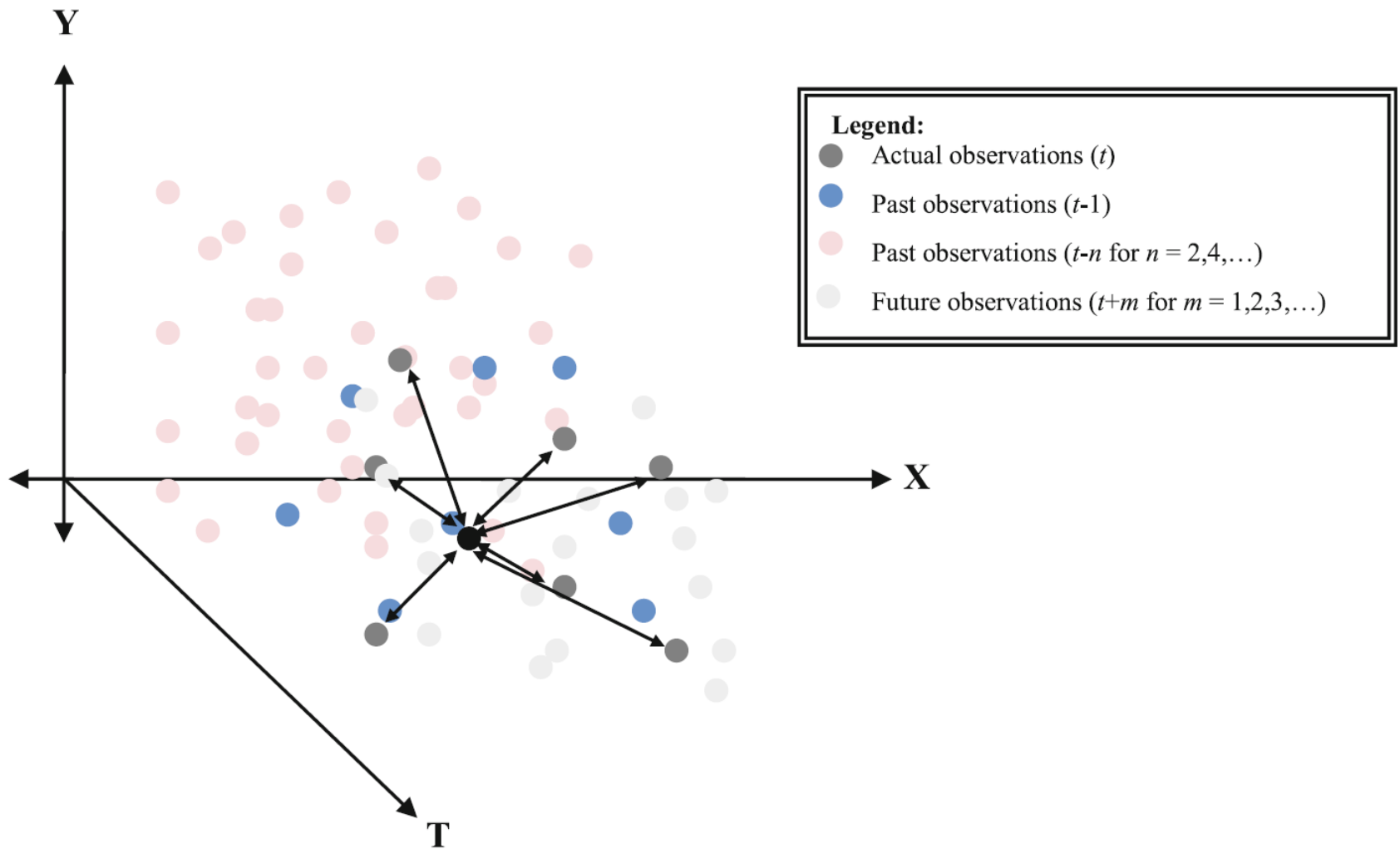


Fig. 4 Spatio-temporal relations for a given observation (*black point*) (discrete time period representation) (color figure online)

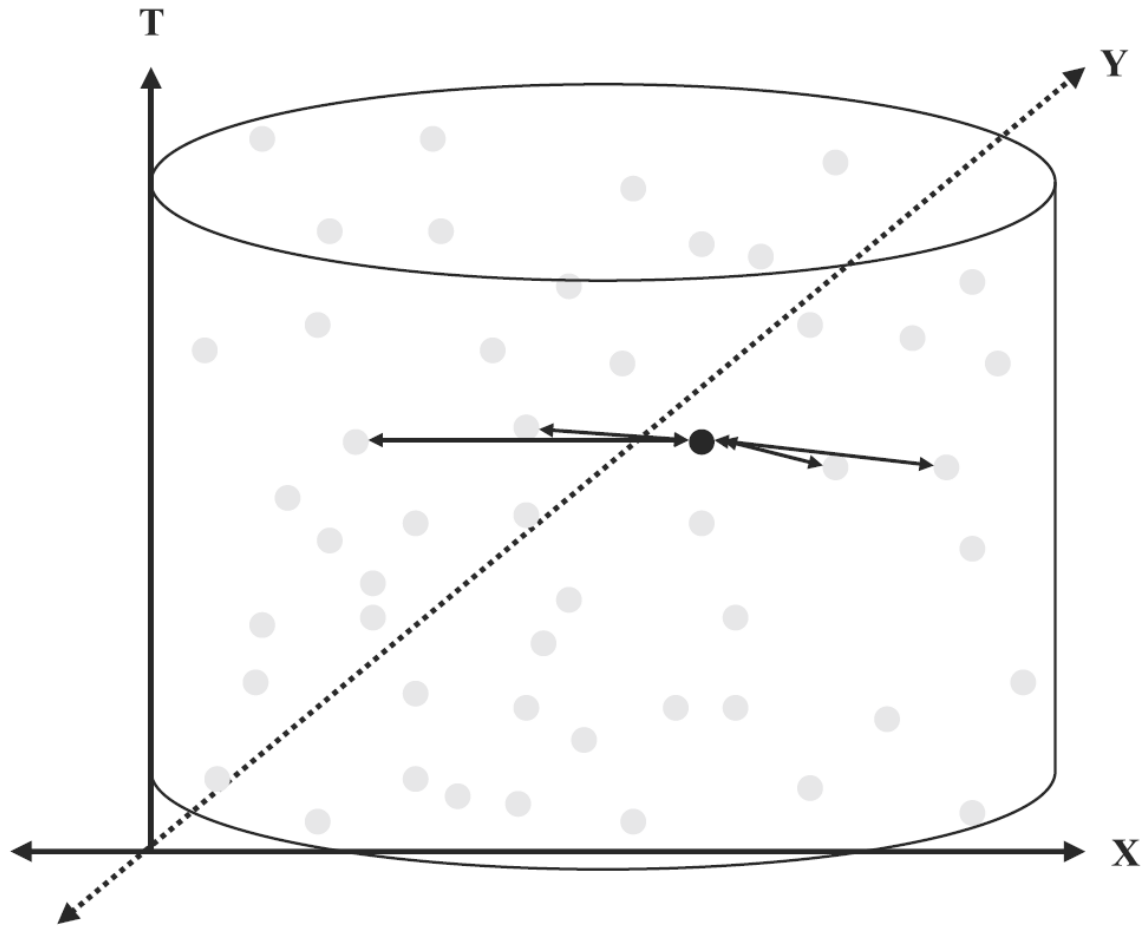


Fig. 5 Spatio-temporal relations for a given observation (*black point*) (continuous time period representation)

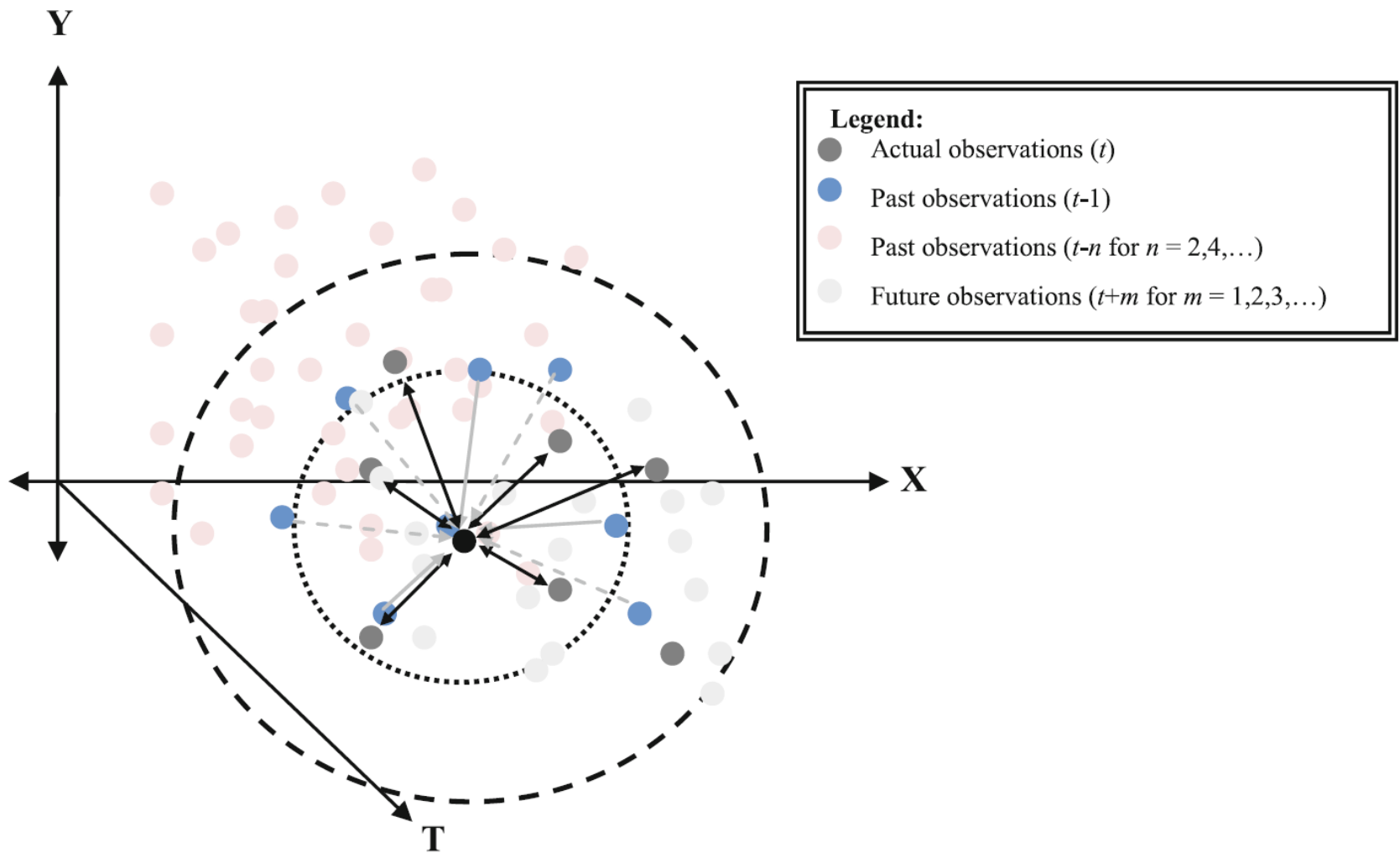


Fig. 6 Possible spatio-temporal relations for a given observation (*black point*) (discrete time period representation) (color figure online)

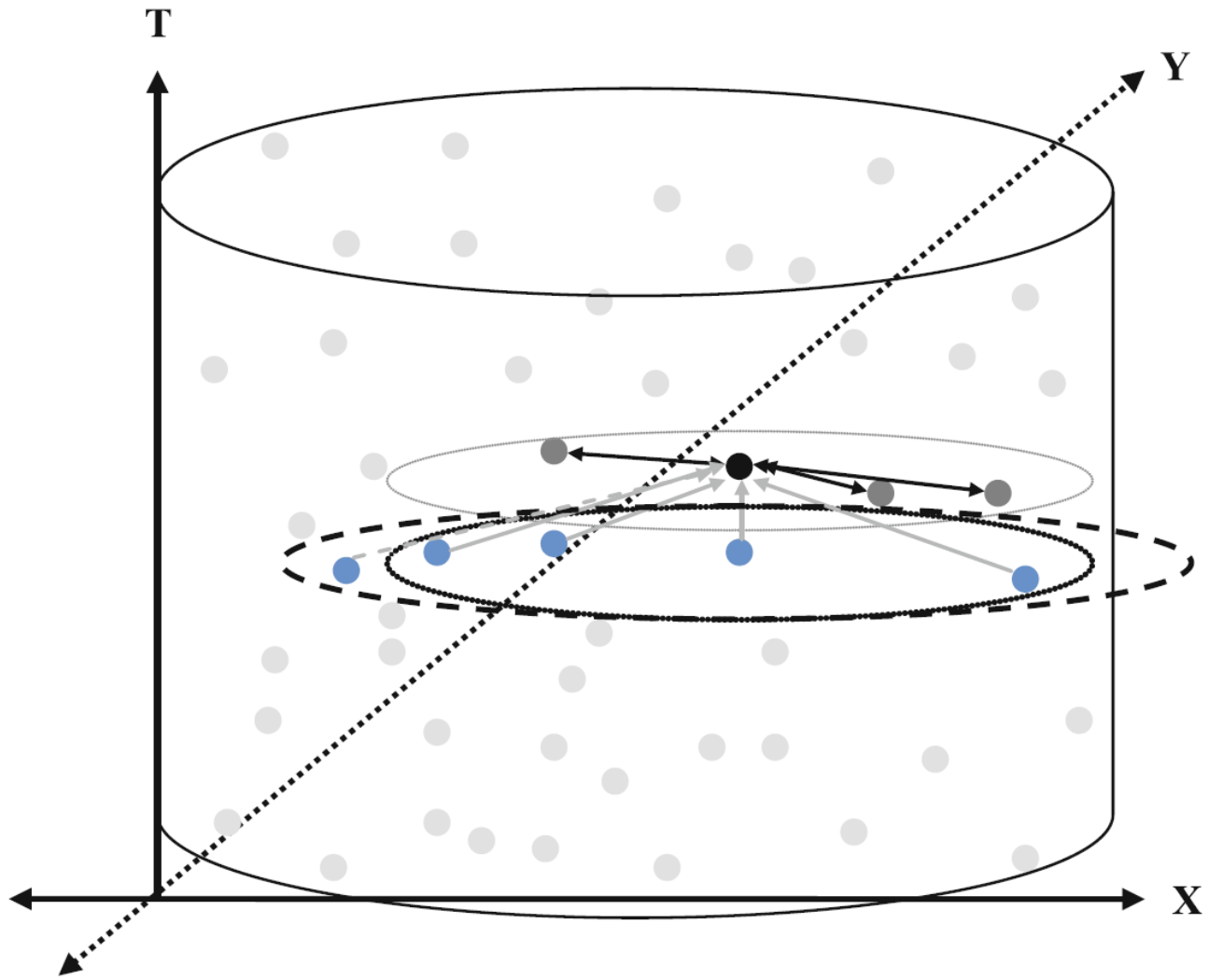


Fig. 7 Possible spatio-temporal relations for a given observation (*black point*) (continuous time period representation) (color figure online)

Types of Panel Data

- Cross-sectional and temporal size:
 - Micro panels – ‘short panels’: N big compared to T
 - Panel Study of Income Dynamics (PSID)
 - Household Budget Continuous Survey (INE-Spain)
 - Macro panels – ‘long panels’: T big compared to N
 - Penn world tables
 - Mini panels: T and N small
 - Grunfeld’s investment data
 - Large panels – ‘huge panels’: N and T big
 - Stock market data
- According to the number of present observations:
 - Balanced, or complete: $n = NT$
 - Non-balanced: $n = \sum_{i=1:N} T_i$

Why Using Panel Data?

- Benefits
 - Usually gathered on micro units (individuals, firms, households). Many variables can be more accurately measured at the micro level, and biases resulting from aggregation may be reduced or eliminated (see e.g. MAUP)
 - Controlling for individual heterogeneity. Not controlling for it involves the risk of obtaining biased results (Moulton 1986, 1987)
 - More informative (i.e. large number of observations; $N \times T$) → more variability → increasing degrees of freedom and reducing collinearity → improving the efficiency of estimates (more accurate inference)
 - They contain information on both time dynamics and characteristics of individuals, which allows one to control for the effects of missing or unobserved variables
 - Better able to identify and measure effects that are simply not detectable in pure cross-sections or time-series data
 - Better able to study adjustment dynamics. Estimation of time-adjustment patterns using time-series data often has relied on arbitrary prior restrictions (such as Koyck or Almon distributed lag models) because time-series observations of current and lagged variables are likely to be highly collinear (Griliches 1967) → Individual heterogeneity reduces collinearity and the need of restrictions
 - Provides micro foundations for aggregate data analysis. Ideal for investigating the ‘homogeneity’ versus ‘heterogeneity’ issue (i.e. ‘representative agent’ assumption often invoked by aggregate analysis)

Why Using Panel Data? (2)

- Limitations
 - Design and data collections problems
 - Problems of coverage (incomplete account of the population of interest), non-response (due to lack of cooperation or interviewer error), recall (respondent not remembering correctly), frequency of interviewing, interview spacing, reference period, etc.
 - Distortions of measurement errors
 - Measurement errors can lead to under-identification
 - Nevertheless, availability of multiple observations for a given individual or at a given time may allow a researcher to make transformations to induce different deductible changes in the estimators, hence to identify an otherwise unidentified model
 - Selectivity problems: Self-selectivity
 - Non response
 - Short time-series dimension. Typically, micro-panels involve annual data covering short time period for each individual → asymptotic arguments rely crucially on the number of individual tending to infinity

Going from Cross-Sectional to Panel Data

- What happens to the models we learned in the previous lecture?
 - Let's add a t subscript (from 1 to T) to the general encompassing spatial model we saw earlier
 - $Y_t = \delta WY_t + \alpha \mathbf{1}_N + X_t \boldsymbol{\beta} + WX_t \boldsymbol{\theta} + \mathbf{u}_t$,
 - where $\mathbf{u}_t = \lambda W\mathbf{u}_t + \boldsymbol{\varepsilon}_t$
 - The restrictions to obtain all more specific models are the same as in cross-sectional models, and it could in theory be estimated likewise
 - BUT: this model(s) does not account for spatial and temporal heterogeneity \rightarrow spatial units are likely to differ in their background variables, usually space-specific and time-invariant (e.g. one unit is on the sea, another at a border)

Going from Cross-Sectional to Panel Data (2)

- The similarity between both generating processes seen before concerns the multidirectional spatial effect that can appear among spatial units within a given time period or interval. For other observations, the effect is no longer multidirectional since temporal dimension is unidirectional
- Thus, the usual strictly spatial representation no longer holds
- In fact, strictly spatial representation may generate spurious relations that can exert influence on the estimated spatial autocorrelation pattern or on spatial autoregressive parameters
- It is possible to account for the unidirectional temporal effect of spatial data located in a given radius of influence to evaluate the possible dynamic (peer) effect, by connecting past observations (dark blue dots) to the actual observations (the solid one-way grey arrows), as well as multidirectional spatial effect for data observed at the same time period, by means of the two-way black arrows
- Dubé and Legros (2013) suggest the use of spatio-temporal weights matrices

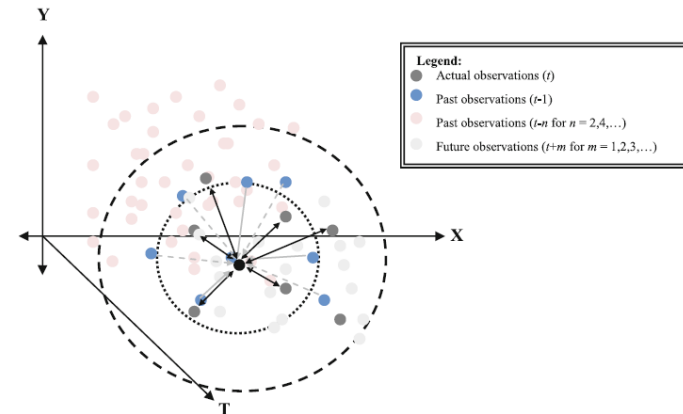


Fig. 6 Possible spatio-temporal relations for a given observation (black point) (discrete time period representation) (color figure online)

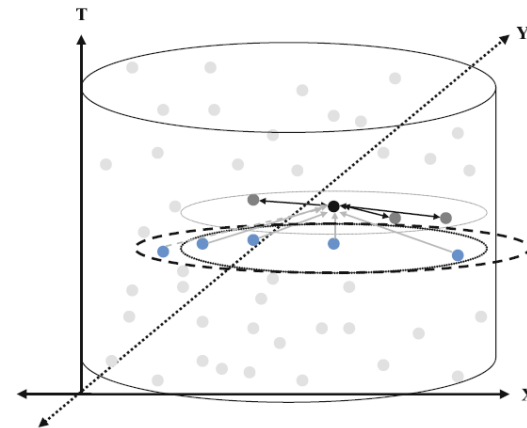


Fig. 7 Possible spatio-temporal relations for a given observation (black point) (continuous time period representation) (color figure online)

Going from Cross-Sectional to Panel Data (3)

- One solution to the aforementioned problem (unobserved spatial units characteristics) is to include a variable intercept μ_i
- Similarly, one might want to include time period effects ξ_t , to control for time-specific but spatially invariant omitted information
- A space-time model would therefore result as
 - $Y_t = \delta WY_t + \alpha \mathbf{1}_N + X_t \boldsymbol{\beta} + W X_t \boldsymbol{\theta} + \boldsymbol{\mu} + \xi_t \mathbf{1}_N + \mathbf{u}_t$,
 - where $\mathbf{u}_t = \lambda W \mathbf{u}_t + \boldsymbol{\varepsilon}_t$, and $\boldsymbol{\mu}$ and ξ_t can be either fixed or random effects

Estimating Spatial Panel Models

- Let's assume W is fixed over time and the panel balanced
 - Estimators can be modified to use time-varying W matrices (e.g. based on socio-economic distance or demographic characteristics)
- We can write the
 - Spatial lag model as: $y_{it} = \delta \sum_j w_{ij} y_{jt} + \mathbf{x}_{it} \boldsymbol{\beta} + \mu_i + \varepsilon_{it}$
 - Spatial error model as: $y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + \mu_i + u_{it}$, where $u_{it} = \lambda \sum_j w_{ij} u_{jt} + \varepsilon_{it}$

Fixed Effects Spatial Lag Model

- Two complications (Anselin et al. 2006)
 - Violation of assumption that $E[(\sum_j w_{ij} y_{jt}) \varepsilon_{it}] = 0$ (simultaneity problem)
 - Spatial dependence of observations at each point in time may affect estimation of fixed effects
- ML estimator developed
 - $\beta = (X^{*T}X)^{-1}X^{*T}[Y^* - \delta(I_T \otimes W)Y^*]$
 - Jacobian can be approximated by numerical methods \rightarrow but estimates of δ and β will change slightly every time (possibility to choose computation method, depending, for example, on amount of data)
 - For large N , determination of variance matrix elements can be computationally impossible \rightarrow approximation by numerical Hessian matrix

Jacobian: matrix of all first-order partial derivatives of a vector-valued function

Hessian: matrix of second-order partial derivatives of a scalar-valued function

Fixed Effects Spatial Error Model

- Again, ML estimator available simply as extension of cross-sectional estimator
- Both spatial lag and spatial error fixed effects models are based on demeaning (Baltagi 2005). Lee and Yu (2010) show that this leads to biased results for some parameter values
 - Without time fixed effects, bias for large N and fixed T
 - With time fixed effects, bias if both N and T are large
- They propose two alternative bias-correction estimators that are NOT based on demeaning

Random Effects Spatial Lag Model

- Identical log-likelihood function to the fixed effects spatial lag model → same estimation procedure [net of a preliminary data transformation (normal for estimating random effects panel models)]

Random Effects Spatial Error Model

- A feasible GLS estimator of β can be derived by assuming a W matrix standardized so that all off-diagonal elements are taken as $1/(N - 1)$ (Baltagi 2006)
- Elhorst (2003) proposes a different procedure based on spectral decomposition
 - Numerically problematic for large N , but for symmetric W (other normalizations than row-standardization) works well in decent amount of time for N up to 4000
- Estimation of variance matrix is also problematic
- Estimation of this model is by far more problematic than the other models

Fixed vs Random Effects

- Abundance of papers in spatial econometrics taking RE as default choice
 - It's a compromise to the all or nothing way of using the cross-sectional component of the data (models with fixed effects only use the time-series component)
 - It's more efficient (no loss of degrees of freedom). And fixed effects need sufficiently large T for consistent estimation of μ_i
 - Allows estimating coefficients for (quasi) time-invariant variables (main reason!)
- Question if RE specification is the right one is often left unanswered. Three conditions for its use
 - Number of spatial units potentially going to infinity
 - Units representative of a larger population (these two conditions controversial for spatial research)
 - Zero correlation between RE and explanatory variables (particularly restrictive)
- When data for all units included in a study area are collected, it's questionable whether they're representative of a larger population
 - Population may be said to be 'sampled exhaustively' (Nerlove and Balestra 1996)
 - 'the individual spatial units have characteristics that actually set them apart from a larger population (Anselin 1988)
 - 'if the sample happens to be the population', specific effects should be fixed (Beenstock and Felsenstein 2007) because each unit represents itself and has not been sampled randomly

Fixed vs Random Effects (2)

- Normally not possible to draw random sets of regions (to infer on population and use RE) because then W cannot be defined and impact of spatial interactions cannot be consistently estimated (you need the neighbouring units to estimate spatial spillovers)
- Therefore, FE generally more appropriate (at least for area-level analysis) than RE
- Hausman's test can be used to test the hypothesis of zero correlation between RE and explanatory variables
 - Lee and Yu (2012) derive a related test for a general spatial panel model (nesting all common ones)
 - Mutl and Pfaffermayr (2011) derive one for when 2SLS are used for estimation instead of ML
 - Debarsy (2012) provides a test for the panel spatial Durbin model

Model Comparison and Testing

- So how are the LM tests we saw earlier for cross-sectional models extended to panel?
- Anselin et al. (2006) develop panel versions of LM tests for δ and λ (but for pooled panel data)
- Classical and robust LM tests based on non-spatial model with or without individual/time fixed effects
- Otherwise, conditional LM tests to test one type of spatial dependence conditional on the other (see Debarsy and Ertur 2010)
- ... based on residuals of either the non-spatial or one spatial model

Model Comparison and Testing (2)

- Evidence in favour of spatial dependence is often weak in presence of time fixed effects → because most variables tend to covariate locally (grow together) over time, following national trends (e.g. participation or unemployment rates)
 - Mathematically, time fixed effects are identical to a spatial error process where W has all elements (including diagonal) set to $1/N$
 - So, if both time fixed effects and a spatial error term are included, the parameter for the latter will automatically fall
 - BUT Lee and Yu (2010) show that ignoring time fixed effects leads to large upward bias in the spatial lag coefficient
- Usual debate over specific-to-general or the opposite. Elhorst (2014) suggests a mixed strategy
 - Estimate non-spatial model to test against spatial lag and spatial error (specific-to-general)
 - If non-spatial is rejected, estimate spatial Durbin model to test whether it can be simplified to spatial lag or spatial error (general to specific) (spatial Durbin error model not considered)

Dynamic Spatial Models

- Ideally needed for cases of
 - Serial dependence between observations on each spatial unit
 - Spatial dependence between observations at each point in time
 - Unobservable spatial and/or time specific effects
 - Endogeneity of regressors lagged in space and/or time
- This class of methods developed very recently, as biases were introduced when mixing serial and spatial dependence
- A generalized space-time dynamic model can be written as
$$\mathbf{Y}_t = \tau \mathbf{Y}_{t-1} + \delta \mathbf{W} \mathbf{Y}_t + \eta \mathbf{W} \mathbf{Y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta}_1 + \mathbf{W} \mathbf{X}_t \boldsymbol{\beta}_2 + \mathbf{X}_{t-1} \boldsymbol{\beta}_3 + \mathbf{W} \mathbf{X}_{t-1} \boldsymbol{\beta}_4 + \mathbf{Z}_t \boldsymbol{\pi} + \mathbf{v}_t$$
, where
 - $\mathbf{v}_t = \rho \mathbf{v}_{t-1} + \lambda \mathbf{W} \mathbf{v}_t + \boldsymbol{\mu} + \boldsymbol{\xi}_t \mathbf{1}_N + \boldsymbol{\varepsilon}_t$
 - $\boldsymbol{\mu} = \kappa \mathbf{W} \boldsymbol{\mu} + \boldsymbol{\zeta}$
 - \mathbf{Z}_t is a matrix of endogenous explanatory variables
 - \mathbf{v}_t reflects the error term specification (serially and spatially autocorrelated)
 - Individual fixed effects are assumed potentially spatially autocorrelated with spatial autocorrelation coefficient κ
- Stationarity conditions involve spatial coefficients as well (not discussed here)

Taxonomy of Dynamic Space-Time Models

- The following model specifications have been considered in the literature
 - $\varepsilon_{t-1} + W\varepsilon_t$
 - $Y_{t-1} + W\varepsilon_t$
 - $Y_{t-1} + WY_t + WY_{t-1} + X_t + WX_t$
 - $Y_{t-1} + WY_t + WY_{t-1} + X_t$, no WX_t
 - $Y_{t-1} + WY_{t-1} + X_t + WX_t$, no WY_t
 - $Y_{t-1} + WY_t + WY_{t-1} + X_t + WX_t$, restriction on coeff. of WY_{t-1}
 - $Y_{t-1} + WY_t + X_t + WX_t$, no WY_{t-1}
- Methods of estimation
 - Bias-correction of the ML or quasi-ML (QML) estimator
 - Instrumental variables or generalized method of moments (IV/GMM)
 - Bayesian MCMC
- Stability condition: $\tau + \delta + \eta < 1$ (otherwise, estimation gets more complicated)
- Effects estimates can be computed for short-term and long-term, as well as the path along which an economy moves to its long-term equilibrium (Debarsy et al. 2012)

Alternative Approach: Spatial Filtering

- Various types of spatial filtering techniques exist. Two well-known ones are the ones by
 - Getis (1995)
 - Griffith (2003)
 - Comparative paper (Getis and Griffith 2002)
 - In spatial filtering, studied variables are split into spatial and non-spatial components
 - Griffith's spatial filtering (SF) is based on the computational formula of Moran's I (MI) statistic. This eigenvector decomposition technique extracts n orthogonal numerical components from a $n \times n$ normalized spatial weights matrix

$$W = (I_n - \iota(\iota'\iota)^{-1}\iota')C(I_n - \iota(\iota'\iota)^{-1}\iota'),$$

Spatial Filtering

- Eigenvectors are extracted in a decreasing order of their partial contribution to MI , the first corresponding to the largest eigenvalue of W
- The set of eigenvecs explaining the spatial pattern in the variable of interest can be found by regressing stepwise the dependent variable on the eigenvecs
 - The linear combination of the eigenvecs selected is the SF
- Advantages
 - The SF works as additional regressors (zero-centred) in a regression model
 - As such, it does not require particular estimation techniques, and can be applied to any functional form, e.g., within a GLM framework, where a linear model is specified ‘behind’ the link function
- The single eigenvectors do not have a straightforward economic interpretation, differently from what happens in PCA (similar technique)
- Recently it was shown by Griffith (2008) that SF can also contribute to explaining spatial heterogeneity in the coefficients. An equivalent to geographically weighted regression (GWR, Brunson et al. 1998) can be computed by interacting the exogenous variables with the eigenvecs