Praktikum z ekonometrie

VŠE Praha

Tomáš Formánek

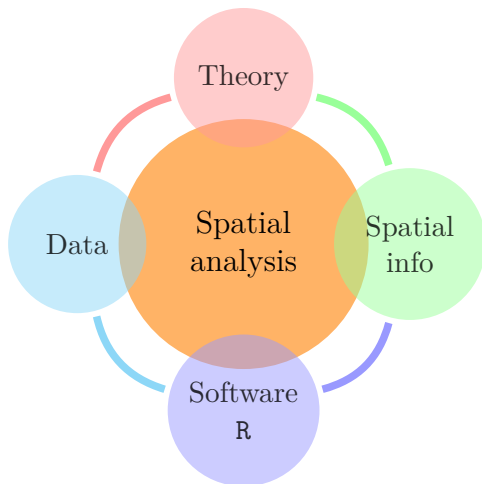# Block 2: Spatial analysis & spatial models – Outline

1. Introduction

2. Spatial stochastic processes

3. Data exploration & descriptive methods
   - Interpolation and Krigging
   - Definition of neighbors
   - Spatial dependency tests
   - Clusters – hotspots and coldspots

4. Spatial regression models (CS data)
   - Parametric spatial models
   - Spatial filtering and semiparametric models
   - Getis' spatial filtering ("nonparametric" spatial models)
   - Moran's eigenvector map filtering (semiparametric spatial models)
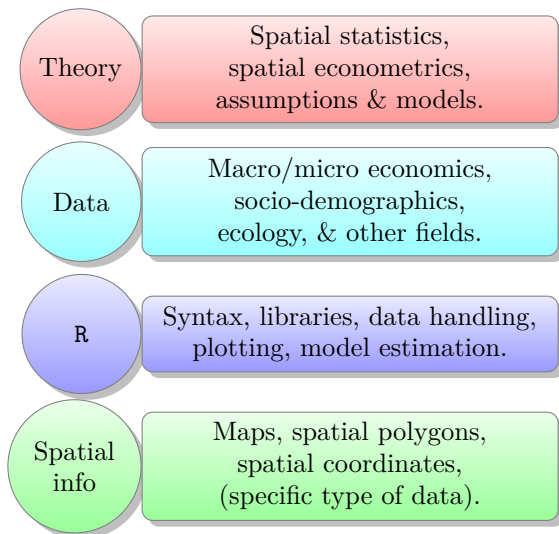
5. Spatial panel models (parametric)

Introduction

Unemployment rate, years 2005 & 2012

Spatial analysis

Spatial analysis

Theory	Spatial statistics, spatial econometrics, assumptions & models.
Data	Macro/micro economics, socio-demographics, ecology, & other fields.
R	Syntax, libraries, data handling, plotting, model estimation.
Spatial info	Maps, spatial polygons, spatial coordinates, (specific type of data).

Introduction - spatial analysis

Methods of quantitative spatial analysis:

- Visualization
  Maps, graphical display

- Data exploration & descriptive methods
  Tools to broadly look at spatial patterns
  Spatial statistics (semivariogram)
  Interpolation (inverse distance weights, krigging)

- Econometric modeling
  Fitting models, testing hypotheses, formalizing spatial dependence,
  discerning spatial effects from other factor (e.g. macroeconomic)

Introduction - spatial analysis

What are spatial data?

- Data that are location specific and that vary in space.

- Referenced by a spatial location s (usually 2D),
  $s = (x; y)$; x is longitude (easting) and y is latitude (northing).

  May also be referenced by a zip code, county or state ID.

- Data that are close together in space (time) are often more alike than those that are far apart.

- Tobler's first law of geography:
  "Everything is related to everything else, but near things are more related than distant things."
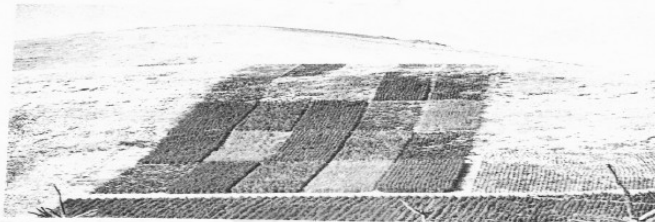
History of spatial analysis: 1854 – London – cholera

Dr. John Snow: Early spatial analysis

- In August 1854, there was a major Cholera outbreak in the Soho neighbourhood of London, UK. There were 127 cholera related deaths around the area.

- At the time, germ theory (microorganisms causing disease) was not generally accepted. Dr. J. Snow was a MD, pioneer of germ theory and a statistician.

- Dr. John Snow spoke to local residents and mapped where cholera cases occurred. As a result of his map, he was able to pinpoint the public water pump on Broad Street as the source of contaminated water causing the cholera outbreak.

- Dr. Snow used statistics to find a relationship between water sources and cholera cases and subsequently found out that the waterworks company supplying water to Broad Street pump was taking water from a sewage polluted area of the Thames river.

Plate 7. *Layout of Bettgelert Experiment.*

History of spatial analysis: 1935 – field experiments

R.A. Fisher: Early spatial analysis

- R.A. Fisher was probably the first to recognize implications of spatial dependency for statistical analysis.

- In his work on design of experiments in agricultural science, he wrote (Fisher, 1935, p. 66):
  "After choosing the area we usually have no guidance beyond the widely verified fact that patches in close proximity are commonly more alike, as judged by the yield of crops, than those which are further apart."

- Observed spatial variability, i.e. field-to-field variability, was largely due to physical properties of the soil and environmental properties of the field. He avoided the confounding of treatment effects with plot effect with the introduction of randomization.

- Fisher's solution was to eliminate spatial dependency bias by localizing the crops under scrutiny into randomly assigned blocks.

Measuring spatial variables

Spatial data: measurements and measurement scales

- **Generally, data vary continuously over space, but are measured only at discrete locations.**
  - To characterize spatial variables, spatial aggregation is necessary.
  - Aggregation of spatial variables may be just another source of bias and potential data miss-manipulation
    Summary values are influenced by shape and scale of spatial units.
    Shape (administrative boundaries) may change over time.
- Scale, consistency and relevance should be carefully considered when collecting and analyzing spatial data

Spatio-temporal data:

- Data that are location specific and repeated in time.
- For each variable, observations have location, time and value.
- Similar methods for analysis, with an added time dimension (choice of sampling frequency, spatial panel data analysis)

Measuring spatial distances

Distances (d) can be defined in a variety of ways, yet the following technical conditions should always apply (invariant to spatial translation, i.e. "shift"):

1. $d(\boldsymbol{s}_i, \boldsymbol{s}_j) = d(\boldsymbol{s}_j, \boldsymbol{s}_i)$
   (symmetry)

2. $d(\boldsymbol{s}_i, \boldsymbol{s}_i) = 0$
   (dist. between a point and itself is zero)

3. $d(\boldsymbol{s}_i, \boldsymbol{s}_j) \leq d(\boldsymbol{s}_i) + d(\boldsymbol{s}_j)$
   (triangle inequality; $d(\boldsymbol{s}_i)$ is the distance from origin)

Measuring spatial distances

Euclidean distances, measured between two point in the "ordinary" Euclidean space. In 2D, the Euclidean distance (L_2 norm) is defined as

$$d(\boldsymbol{s}_i, \boldsymbol{s}_j) = \sqrt{(s_{ix} - s_{jx})^2 + (s_{iy} - s_{jy})^2}\,,$$

where the x and y subscripts handle planar coordinates. For smaller distances, the computational simplicity is attractive.

Great circle distances - for larger distances, planar projection accumulates non-negligible errors. The shortest path between two points on a sphere (given their longitudes and latitudes):

$$d(\boldsymbol{s}_i, \boldsymbol{s}_j) = 2r \arcsin \sqrt{\sin^2\left(\frac{\phi_j - \phi_i}{2}\right) + \cos(\phi_i)\cos(\phi_j)\sin^2\left(\frac{l_j - l_i}{2}\right)}\,,$$

where r is the radius of the sphere, ϕ_1 and ϕ_2 are the latitudes of $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ in radians, l_i and l_j are the longitudes (in radians). Its only an approximation when applied to the Earth, which is not a perfect sphere (correct within a 0.5%; alternative: Vicenty's formulae).

Manhatan distances - L_1 norm, a function on a fixed grid.

Spatial stochastic processes

For a generic location $\boldsymbol{s}$ given by a vector of d coordinates in a d-dimensional Euclidean space, spatial stochastic process ("random field") is often denoted as

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

- Typically, $d = 2$ for most economic and econometric applications, $d = 3$ is often used in fields such as geology or astronomy.

- D is a fixed finite set of N spatial locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N$.

- Individual $\boldsymbol{s}_i$ units are points in space (say, with GPS-based latitude and longitude coordinates). Sometimes, such points can be associated with non-zero surface area elements.

- Much like in time-series analysis, the individual realizations of a spatial stochastic process – random field – are often denoted $z(\boldsymbol{s}_i)$ or, simply, z_i.

Spatial stochastic processes

Stationarity is a common assumption: a spatial process under scrutiny repeats itself over the domain D. If we translate the entire set of coordinates by $\boldsymbol{h}$ – a specific distance in a specified direction, the stochastic process and its features remain unchanged.

Strong stationarity of a random filed: We start with a finite-dimensional distribution:

$$F_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_m}(z_1,\ldots,z_m) = P[Z(\boldsymbol{s}_1) \leq z_1, Z(\boldsymbol{s}_2) \leq z_2,\ldots,Z(\boldsymbol{s}_m) \leq z_m].$$

Strong stationarity $\leftrightarrow$ F is invariant under spatial translation $\boldsymbol{h}$.
Unlike d_{ij} (Euclidean distance between two spatial units $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$), $\boldsymbol{h}$ is an orientated distance "shift" (spatial translation) vector.
For strong stationarity:

$$\begin{aligned}
&P\left[Z(\boldsymbol{s}_1) \leq z_1, Z(\boldsymbol{s}_2) \leq z_2,\ldots,Z(\boldsymbol{s}_m) \leq z_m\right] \\
&= P\left[Z(\boldsymbol{s}_1 + \boldsymbol{h}) \leq z_1, Z(\boldsymbol{s}_2 + \boldsymbol{h}) \leq z_2,\ldots,Z(\boldsymbol{s}_m + \boldsymbol{h}) \leq z_m\right].
\end{aligned}$$

Spatial stochastic processes

Weak stationarity (also called second order stationarity) assumes that the first two moments exist, are invariant (and finite) and covariance only depends on spatial translation (orientated distance) $\boldsymbol{h}$:

$$E[Z(\boldsymbol{s})] = \mu\,,$$
$$\text{var}[Z(\boldsymbol{s})] = \sigma^2\,,$$
$$\text{cov}[Z(\boldsymbol{s} + \boldsymbol{h}), Z(\boldsymbol{s})] = C(\boldsymbol{s} + \boldsymbol{h}, \boldsymbol{s}) = C(\boldsymbol{h})\,.$$

As autocovariance is a function of $\boldsymbol{h}$ only (under weak st.), for any spatial points $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ such that $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}$, we can write:

$$\text{cov}\,[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)] = C(\boldsymbol{s}_i - \boldsymbol{s}_j) = C(\boldsymbol{h})\,.$$

Covariogram $C(\boldsymbol{h})$ is the covariance between two spatial units, separated by $\boldsymbol{h}$. For $\boldsymbol{h} = \boldsymbol{0}$, it simply describes variance:

$$\text{cov}\,[Z(\boldsymbol{s} + \boldsymbol{0}), Z(\boldsymbol{s})] = C(\boldsymbol{0}) = \text{var}\,[Z(\boldsymbol{s})]\,.$$

Under weak dependency, covariance disappears with growing distance:

$$C(\boldsymbol{h}) \to 0 \text{ as } ||\boldsymbol{h}|| \to \infty$$

Spatial stochastic processes

Intrinsic stationarity is less restrictive than weak (second order) stationarity and it is defined in terms of first differences.

A spatial process is intrinsically stationary if the difference between two observed spatial points is weakly stationary:

$$E[Z(\boldsymbol{s} + \boldsymbol{h}) - Z(\boldsymbol{s})] = 0,$$
$$\mathrm{var}[Z(\boldsymbol{s} + \boldsymbol{h}) - Z(\boldsymbol{s})] = 2\gamma(\boldsymbol{h}),$$

where $2\gamma(\boldsymbol{h}) \geq 0$ is the variogram. Generally, $2\gamma(\boldsymbol{h})$ increases with growing oriented distance $\boldsymbol{h}$.

The two types of relaxed stationarity are related: weak stationarity implies intrinsic stationarity but not vice versa. For weakly stationary spatial processes (where $E(Z(\boldsymbol{s} + \boldsymbol{h})) = E(Z(\boldsymbol{s})) = \mu$) the variogram simplifies to:

$$2\gamma(\boldsymbol{h}) = E\left[(Z(\boldsymbol{s} + \boldsymbol{h}) - Z(\boldsymbol{s}))^2\right],$$

i.e. to the expected squared difference between two observed realizations of a spatial stochastic process.

Spatial stochastic processes

Semivariogram is denoted as $\gamma(\boldsymbol{h})$ and it equals to half the variogram.

Since $2\gamma(\boldsymbol{h})$ is calculated as expectation of a square, $\gamma(\boldsymbol{h}) \geq 0$ for both weakly and intrinsically stationary random fields.

Also, at $\boldsymbol{h} = \boldsymbol{0}$, $\gamma(\boldsymbol{0}) = 0$ because

$$E\left[(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i))^2\right] = 0 \text{ for } \forall i\,.$$

Variogram (semivariogram) is a generalization of the covariogram $C(\boldsymbol{h})$ and under weak stationarity, the two functions are related by:

$$\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h})\,.$$

If a stationary stochastic process has no spatial dependency at all (i.e. $C(\boldsymbol{h}) = 0$ for $\boldsymbol{h} \neq \boldsymbol{0}$), the semivariogram is constant: $\gamma(\boldsymbol{h}) = \text{var}[Z(\boldsymbol{s})]$ everywhere, except for $\boldsymbol{h} = \boldsymbol{0}$, where $\gamma(\boldsymbol{0}) = 0$.

Isotropic spatial process may be defined through a semivariogram:

$$\gamma(\boldsymbol{h}) = \gamma(||\boldsymbol{h}||) = \gamma(d).$$

Isotropy means that the semivariogram depends only on the distance d between two points and not on direction.

The lack of isotropy – anisotropy – means the semivariogram depends on direction as well as distance.

To assess and test anisotropy, we can estimate and plot directional semivariograms (shown next).

Empirical semivariogram

To perform empirical analysis of distance-based data correlations, we construct the so called empirical semivariogram. First, we divide the distances observed over the domain D into K conveniently chosen intervals:

$$I_1 = (0, d_1], \ I_2 = (d_1, d_2], \ \ldots, \ I_K = (d_{K-1}, d_K].$$

Here, d_1 is the maximum distance within the I_1 interval and d_K is the maximum distance observed over the field of data.

The intervals can be proportional in terms of distance or in terms of sets of observation pairs allocated to each interval (to adjust for unevenly spaced observations).

Note that distances are determined by d (distance magnitudes) only – here, we do not use the orientated distances $\boldsymbol{h}$.

Spatial stochastic processes

Empirical semivariogram is calculated using the following formula:

$$\hat{\gamma}(d_k) = \frac{1}{2N(d_k)} \sum_{N(d_k)} \left[Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j) \right]^2,$$

where $N(d_k)$ is the number of distinct observation pairs in the interval I_k and $\hat{\gamma}(d_k)$ is the semivariogram estimate for its corresponding group (interval) of distances.

Usually, we fit a convenient parametric function (exponential, spherical, Gaussian, etc.) to the estimated $\hat{\gamma}(d_k)$ values (shown next).

The main goal of empirical semivariogram construction is to estimate and visualize the spatial autocorrelation structure of the observed stochastic process.
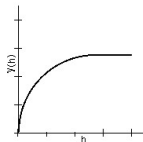
Empirical semivariogram
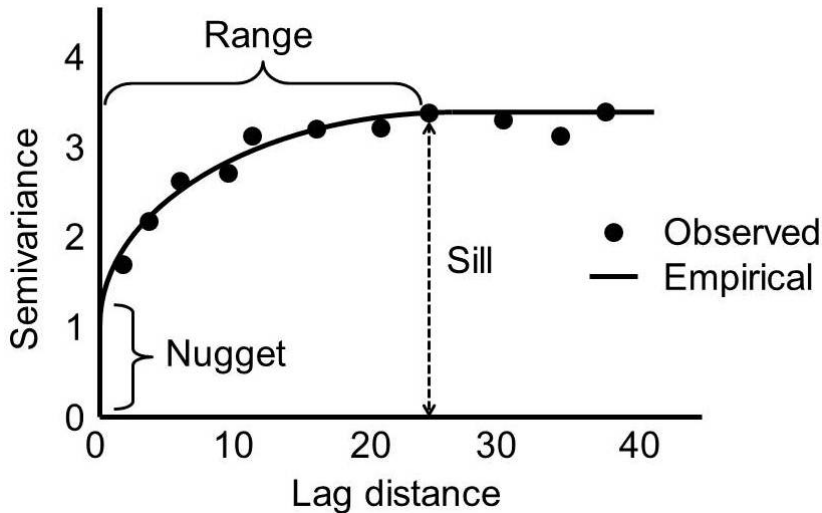
Empirical semivariogram

Three main features of an estimated empirical semivariogram:

- *Nugget* (nugget effect) describes the micro-scale variations or measurement errors in data. Theoretically, at zero distance, $\gamma(0) = 0$. However, two factors play a role here: First, $\gamma(d_1)$ is estimated over the $N(d_1)$ set of pairs, i.e. for the first interval where $d_{ij} \in (0, d_1)$. Second, fitting the empirical semivariogram curve to observed values often causes the non-zero nugget.

- *Sill* amounts to $\lim_{d \to \infty} \gamma(d)$. The sill corresponds to variance of the stochastic field at distances where spatial dependency (which reduces $\gamma(d)$) no longer applies. $\lim_{d \to \infty} \gamma(d) = C(\mathbf{0}) = var[Z(\boldsymbol{s})]$.

- *Range* is the spatial distance (if any) beyond which the data are not autocorrelated. In a way, range describes the strength of spatial structure – based on where the semivariogram "reaches" its asymptote (sill).
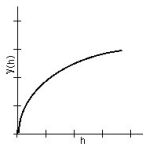
Empirical semivariogram (fitting)

CIRCULAR

$$\gamma(\mathbf{h}) = c_0 + c\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a}\right) + \sqrt{1 - \frac{h^2}{a^2}}\right)$$
$$0 < h \le a$$

$$\gamma(\mathbf{h}) = c_0 + c \qquad h > a$$
$$\gamma(0) = 0$$

Circular semivariance model illustration
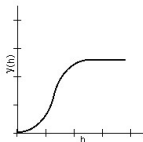
EXPONENTIAL

$$\gamma(\mathbf{h}) = c_0 + c\left(1 - exp\left(\frac{-h}{r}\right)\right) \qquad h > 0$$
$$\gamma(0) = 0$$

Exponential semivariance model illustration

GAUSSIAN
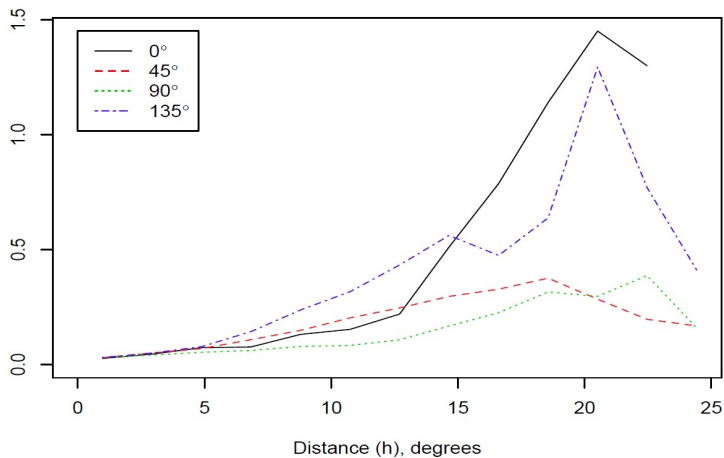

$$\gamma(\mathbf{h}) = c_0 + c\left(1 - exp\left(\frac{-h^2}{r^2}\right)\right) \qquad h > 0$$
$$\gamma(0) = 0$$

Gaussian semivariance model illustration

Empirical semivariogram (directional semivariogram)

Distance (h), degrees

Spatio-temporal stochastic processes

The above discussion can be generalized to accommodate processes that are observed repeatedly over time.

Such observations usually exhibit both spatial and temporal dependency and variability.

Given the frequency and density limitations of empirical measurements in continuous space and time, we model our observations as realizations of a spatio-temporal random function (random field)

$$Z(\boldsymbol{s}, t), \qquad \text{where } (\boldsymbol{s}, t) \in \mathbb{R}^d \times \mathbb{R},$$

the spatio-temporal domain is indexed in space by $\boldsymbol{s} \in \mathbb{R}^d$ and in time by $t \in \mathbb{R}$.

The separation between spatial and time dimensions is substantial, which is reflected in the notation.

Spatio-temporal stochastic processes

Weak and intrinsic stationarity concepts can be easily expanded from spatial to spatio-temporal data.
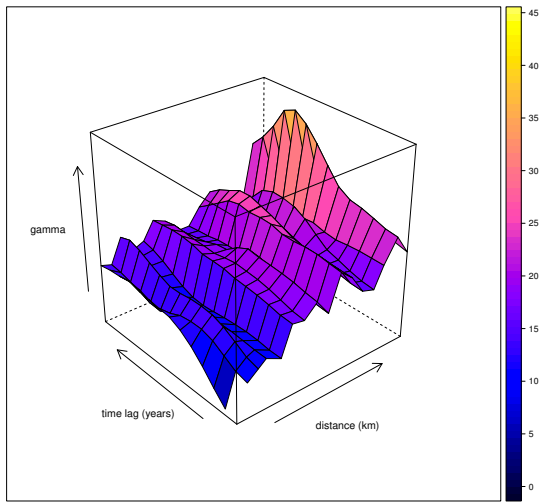
For an intrinsically stationary process $Z(\boldsymbol{s}, t)$, spatio-temporal semivariograms (STSV) is:

$$\gamma(\boldsymbol{h}; t) = \frac{1}{2}\text{var}\left[Z(\boldsymbol{s}_0 + \boldsymbol{h}\,;\, t_0 + t) - Z(\boldsymbol{s}_0; t_0)\right], \qquad (\boldsymbol{h}, t) \in \mathbb{R}^d \times \mathbb{R}.$$

STSV does not depend on the selection of origin $(\boldsymbol{s}_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$ (under intrinsic stationarity).

Also, for intrinsically stationary random fields $Z(\boldsymbol{s}, t)$, the STSV $\gamma(\boldsymbol{h}; t)$ is non-negative and $\gamma(\boldsymbol{0}; 0) = 0$.

Interpolation

Inverse distance weighting (IDW): a deterministic method of interpolation with a known scattered set of points.

Assigned values to unknown points are calculated with a weighted average of the values available at the known points.

To find an interpolated value x_0 at a given point s_0 based on samples $x_i = x(s_i)$ with $i = 1, 2, \ldots, N$ we use a IDW interpolating function:

$$x(s_0) = \begin{cases} \dfrac{\sum_{i=1}^{N} w_i(s_0) x_i}{\sum_{i=1}^{N} w_i(s_0)}, & \text{if } d(s_0, s_i) \neq 0 \text{ for all } i \\ x_i & \text{if } d(s_0, s_i) = 0 \text{ for some } i \end{cases}$$

where $w_i(s_0) = \frac{1}{d(s_0, s_i)^p}$, with s_0, s_i being points on a map, d is their corresponding distance, and p is a positive *real* number (power parameter).

Weight decreases as distance increases from the interpolated points.
High p assigns greater influence to closest values – with sufficiently high p, IDW result may turn into a mosaic of tiles (a Voronoi diagram).
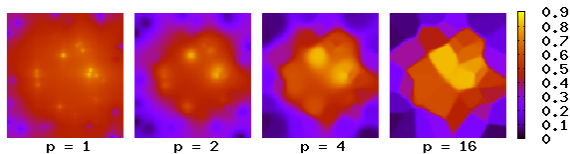
Krigging

- developed by Daniel G. Krige (1919-2013)
  (originally called "weighted moving averages" method)

- Also known as BLUP (best linear unbiased prediction) in spatial analysis and closely related to OLS estimation.

- Interpolates values using the intensity and shape of the empirical (semi)variogram. Results depend on the choice of fitting model (Gaussian, spherical, exponential, etc.).

- Returns observed values at sampling locations.

- Uses neighborhood and/or distance search radius.

- Provides standard errors of interpolated values.

- Multiple approaches and generalizations exist (block krigging).
  http://desktop.arcgis.com/en/arcmap/10.3/tools/
  3d-analyst-toolbox/how-kriging-works.htm

Kriging

- $x(\boldsymbol{s}_0) = \sum_{i=1}^{N} \lambda_i\, x(\boldsymbol{s}_i)$ s.t. $\sum_{i=1}^{N} \lambda_i = 1$.

- In IDW, λ_i depend only on distance to the prediction location.

- With krigging, λ_i depend on a fitted model to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location.

  Basically, krigging weights follow from the estimated empirical semivariogram – see illustration on the next page (note how ordering of the predicted/krigged values may influence results).

Krigging – illustration
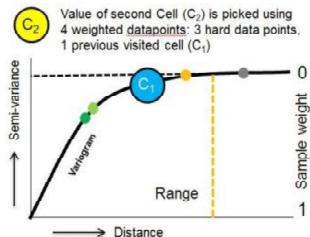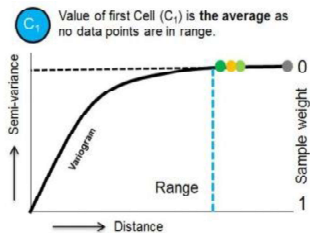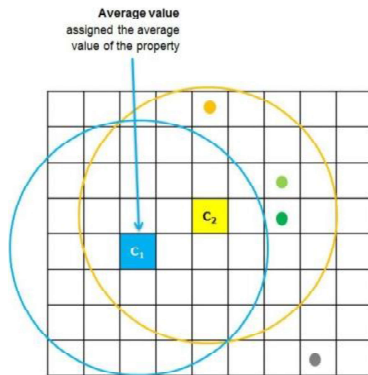
Kriging

Ordinary kriging assumes constant (yet unknown) mean over the neighborhood of s_0.

- Matrix form: $x(s_0) = \boldsymbol{\lambda}' \boldsymbol{x}$ s.t. $\boldsymbol{\lambda}' \mathbf{1} = 1$.
  where $\boldsymbol{\lambda}' = (\lambda_1, \ldots, \lambda_N)$ and $\mathbf{1}$ is a vector of ones.

- With kriging, we solve: $\min E[x(s_0 - \boldsymbol{\lambda}' \boldsymbol{x})^2]$ s.t. $\boldsymbol{\lambda}' \mathbf{1} = 1$.

- Minimum variance of $x(s_0)$: $\sigma^2 = \sum_{i=1}^{N} \lambda_i \gamma(s_i, \ s_0) \ + \ m$

- is obtained when $\sum_{i=1}^{N} \lambda_i \gamma(s_i, \ s_j) \ + \ m = \gamma(s_j, \ s_0)$ for all j.

  Here, γ is the semivariogram, m is an additional LM parameter (mean estimate) that ensures unbiasedness of the estimate. For additional discussion and alternative estimation methods (simple, universal kriging, etc.), see
  https://en.wikipedia.org/wiki/Kriging

Definition of neighbors

- Fotheringham et al (2002): "Spatial dependency is the extent to which the value of an attribute in one location depends on the values of the attribute in nearby locations."

- Different definitions of spatial dependency are possible.

- To discuss spatial dependency, spatial autocorrelation, corresponding tests and spatial econometric models, we need to formalize the concept of **nearby locations – neighbors**

Definition of neighbors

- **Distance-based approach** defines two units as neighbors if their distance does not exceed some ad-hoc predefined threshold: τ.

  - Can generate "islands" (units with zero neighbors), if τ is low compared to minimum distances among unit pairs.
  - Less suited for analysis of areas with uneven geographic density (of measurements).

- **Centroids** are used for measuring distances between units with non-zero areas (e.g. regions)

  - Centroids can be purely geographical, "main" city locations, population-weighted, transportation-weighted (highway/railway), etc.
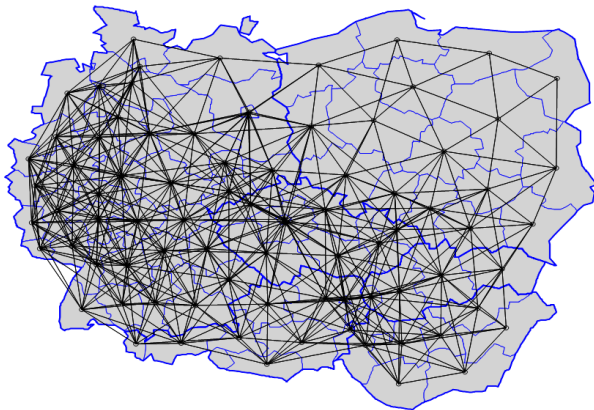
Figure: Plot for distance-based neighbors (NUTS2), maximum neighbor distance threshold at 250 km

Definition of neighbors

- **Contiguity-based approach** spatial units (regions) are neighbors if they share a common border (at least one point).

- **Generalized contiguity approach** is based on the premise that a "second order" neighbor is the neighbor of a first order neighbor (the actual contiguous neighbor).

  With this type of approach, we can define a maximum neighborhood lag (order) to control for the highest accepted number of neighbors traversed (not permitting cycles).

Definition of neighbors

k Nearest neighbors (kNN)

For each spatial unit, we search for a preset number of k nearest units that we define as its neighbors.

- Solves for differences in areal densities (k neighbors are ensured for each unit).

- Usually leads to asymmetric spatial connectivity matrices. Useful for some empirical applications, potentially flawed neighborhood interpretation for others.

- Illustration for $k = 3$ (neighbors only shown for 2 units to illustrate asymmetry):
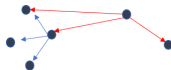
Spatial connectivity matrix (C)

Regardless of the algorithm used for neighbor vs non-neighbor categorization, any spatial structure may be formalized using a connectivity matrix C:

$$C = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \qquad \text{a 4-unit example (symmetric)}$$

$$c_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are neighbors,} \\ 0 & \text{if } i \text{ and } j \text{ are not neighbors.} \end{cases}$$

- Zeros on the diagonal – units are not neighbors to themselves.
- Spatial connectivity matrix interpretation:
  - row/column 1: unit 1 is neigbor to units 2,3,4
  - row/column 2: unit 2 is neigbor to units 1,3 (not 4)

Spatial weights matrix ($\boldsymbol{W}$)

$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \rightarrow \quad \boldsymbol{W} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

- Usually, $\boldsymbol{W}$ is row-standadrized (to unity): $w_{ij} = \frac{c_{ij}}{\sum_{j=1}^{N} c_{ij}}$.

- Connectivity indicators c_{ij} may be generalized prior to standardization. Hence, instead of binary c_{ij}, we may use inverse distances: linear, squared, ... depending on the assumed decay in spatial dependency over distance.

- Validity of spatial prior information (neighborhood definitions, decay patterns) will influence the outcome of subsequent analyses (spatial dependency tests, spatial regression models).

Spatial weights matrix ($\boldsymbol{W}$)

$$\boldsymbol{W} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

- Each row of $\boldsymbol{W}$ "provides" weights for an expected value of an observed spatial variable y_i – weighted averages (fitted values) can be calculated as $\hat{\boldsymbol{y}} = \boldsymbol{W}\boldsymbol{y}$. For example:

$$\hat{y}_1 = \tfrac{1}{3}y_2 + \tfrac{1}{3}y_3 + \tfrac{1}{3}y_4$$
$$\cdots$$
$$\hat{y}_4 = \tfrac{1}{2}y_1 + \tfrac{1}{3}y_3$$

- Observed values of variables can be used to predict corresponding values for neighboring spatial units.

- In this context, $\hat{y}_i$ is often called **spatial lag** of y_i.
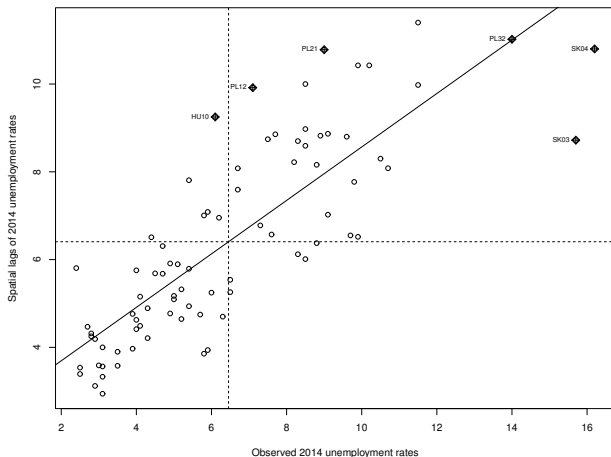
Figure: Moran plot for unemployment rate, 2014, NUTS2 (AT, CZ, DE, HU, PL, SK): observed values vs. spatial lags.

Sample selection in spatial data analysis

- Spatially autocorrelated processes are defined in terms of individual units and their interactions with neighbors.

- Clearly, we can only assess ("use") the impact of neighboring units if such units are part of our sample. Hence, in spatial econometrics, we usually do not draw limited samples from a particular area.

- Instead, we work with data from adjacent units located in unbroken ("complete") study areas.

- Otherwise, C and W matrices would be misleading and we could not consistently estimate spatial interactions and effects.

- Generally speaking, spatial analysis should include the whole geographically defined area/region instead of using random sampling (from a "population" of regions within the relevant area).

Spatial dependency

- Positive spatial autocorrelation occurs if high or low values of a variable cluster in space.

- For negative spatial autocorrelation, spatial units tend to be surrounded by neighbors with very dissimilar observations.

- Spatially random data lack any spatial pattern.

- Sometimes, spatial dependency patterns are easy to discern visually using choropleths.

- However, a formal approach towards evaluation of spatial dependency is often required.

Spatial dependency

Model $\boldsymbol{y} = \lambda \boldsymbol{Wy} + \boldsymbol{\varepsilon}$ example:
(simulated $\boldsymbol{y}$ data, different λ values and *iid* residuals $\boldsymbol{\varepsilon}$)
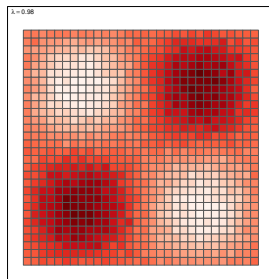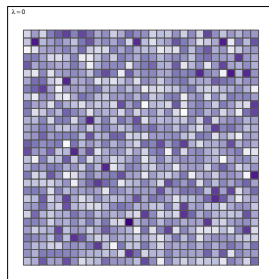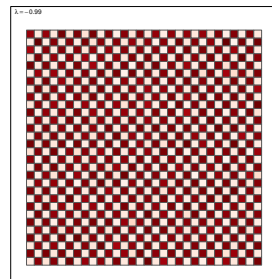
(a)   (b)   (c)

(a) Strong positive spatial dependency in $\boldsymbol{y}$, with $\lambda \to 1$.

(b) Spatially random process: $\lambda = 0 \quad \to \quad \boldsymbol{y} = \boldsymbol{\varepsilon}$.

(c) Strong negative spatial dependency, ("rook" neighbors only).

Moran's I

Measure of global spatial autocorrelation, overall clustering of data:

$$I = \frac{N}{W}\, \boldsymbol{z}'\boldsymbol{W}\boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1},$$

where

N is the number of spatial observations (units) of the variable under scrutiny (say, $\boldsymbol{y}$),

$\boldsymbol{z}$ is the centered form of $\boldsymbol{y}$: $z_i = y_i - \bar{y}$.
(Note a recast: $y_i = \beta_0 + z_i$. If SLRM is expanded by regressors, Moran's I can be applied to regression residuals).

The standardization factor $W = \sum_i \sum_j w_{ij}$ corresponds to the sum of all elements of the spatial weights matrix $\boldsymbol{W}$.
For row-standardized $\boldsymbol{W}$ matrices, $\frac{N}{W} = 1$. Moran's I can be used with $\boldsymbol{C}$ matrices (binary, generalized) as well.

Moran's I

$$I = \frac{N}{W}\, \boldsymbol{z}'\boldsymbol{W}\boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-1},$$

In most empirical circumstances, $I \in [-1, 1]$. Under the null hypothesis of spatial randomness, Moran's I is asymptotically normally distributed with the following first two moments:

$$E(I) = -\frac{1}{N-1} \quad \text{and} \quad \text{var}(I) = \frac{N^2 W_1 - N W_2 + 3W^2}{(N^2 - 1)W^2},$$

where $W_1 = \sum_i \sum_j (w_{ij} + w_{ji})^2$ and $W_2 = \sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2$.

Normality assumption $\rightarrow$ calculate a z-score

$$z = \frac{I - E(I)}{\sqrt{\text{var}(I)}}$$

and test for statistical significance of Moran's I: whether neighboring units are more similar $(I > E(I))$ or more dissimilar $(I < E(I))$ than they would be under null hypothesis of spatial randomness.

Local Moran's I

- Moran's I yields only one statistic that summarizes the nature of spatial dependency in the observed variable – it assumes geographical homogeneity (stationarity) in the data.

- If such assumption does not hold (spatial dependency varies over space), then Moran's I test loses power and the "global" statistic is non-descriptive.

- To address this problem, we can use Local Moran's I statistic (for row-standardized $\boldsymbol{W}$):

$$I_i = \frac{z_i N}{\boldsymbol{z}' \boldsymbol{z}} \boldsymbol{w}_i \boldsymbol{z} \, .$$

- The expected value of Local Moran's I under the null hypothesis of no spatial autocorrelation is: $E(I_i) = -w_i/(N-1)$. Here, w_i is the sum of elements in the i-th row of $\boldsymbol{W}$.

Local Moran's I

$$I_i = \frac{z_i N}{\boldsymbol{z}'\boldsymbol{z}} \boldsymbol{w}_i \boldsymbol{z} \,.$$

- Values of $I_i > E(I_i)$ indicate positive spatial autocorrelation, i.e. that the i-th region is surrounded by regions that, on average, are similar to the i-th region with respect to the observed variable y.
- $I_i < E(I_i)$ would suggest negative spatial autocorrelation.
- Significance of spatial dependency is then evaluated using $var(I_i)$ and the corresponding z-score.

- We may see the global nature of Moran's I from

$$I = \frac{1}{N} \sum_{i=1}^{N} I_i \,.$$

Geary's C

Variance test similar (in principle) to the Durbin-Watson test statistic for residuals' autocorrelation in time-series regressions.

$$C = \frac{N-1}{2W} \frac{\sum_i \sum_j w_{ij}(y_i - y_j)^2}{\sum_i(y_i - \bar{y})^2},$$

where all elements follow from previous slides. Empirical Geary's C values range from 0 to 2. However, occurrences of $C > 2$ are possible.

- Under the null of no spatial autocorrelation, first two moments are:

$$E(C) = 1, \quad \mathrm{var}(C) = \frac{(N-1)(2W_1 + W_2) - 4W^2}{2(N+1)W^2},$$

- Positive spatial dependency: $C < 1$.
  Negative spatial autocorrelation is reflected in $C > 1$.

- z-transformation is asymptotically normally distributed.
  Therefore, $z(C)$ can be used for testing spatial randomness.

Clusters – hotspots and coldspots

Getis' G^*: spatial clusters and hotspot analysis

- Clustering analysis by Getis can only be performed for positively autocorrelated spatial data.

- $G_i^*(\tau) = \frac{\sum_{j=1}^N c_{ij}^* y_j}{\sum_{j=1}^N y_j}$,

  where c_{ij}^* come from amended distance-based (arbitrary τ used) connectivity matrix $\boldsymbol{C}^* = \boldsymbol{C} + \boldsymbol{I}_N$; i.e. y_i observations enter $G_i^*(\tau)$ calculation. Observations of y are assumed to have a natural origin and positive support.

- $G_i^*(\tau)$ is a local stastic, a proportion of the aggregated y_j values that lie within τ of i to the total sum of y_j observations.

- Alternatively, Getis' statistic is calculated using $\boldsymbol{C}$ (not $\boldsymbol{C}^*$) and denoted $G_i(\tau)$

Clusters – hotspots and coldspots

- If we observe high values of y_j within distance τ of unit i, then $G_i^*(\tau)$ would be relatively high compared to its expected value under the null hypothesis of full spatial randomness:

$$E\left[G_i^*(\tau)\right] = \frac{c_i^*}{N},$$

where c_i^* is the sum of elements of i-th row of $\boldsymbol{C}^*$.

- Also, under the H_0 of spatial randomness, we can write

$$\mathrm{var}\left[G_i^*(\tau)\right] = \frac{c_i^*(N - c_i^*)}{N^2(N-1)}\left(\frac{Y_{i2}^*}{(Y_{i1}^*)^2}\right),$$

where $Y_{i1}^* = \frac{\sum_j y_j}{N}$ and $Y_{i2}^* = \frac{\sum_j y_j^2}{N} - (Y_{i1}^*)^2$.

- High positive z-score indicates "hotspot" (cluster of high values) and vice versa. Critical values provided by Getis and Ord. Say, for $N = 100$ and $\alpha = 5\%$, the z-scores would have to exceed ± 3.289 for a statistically significant hot/cold spot.
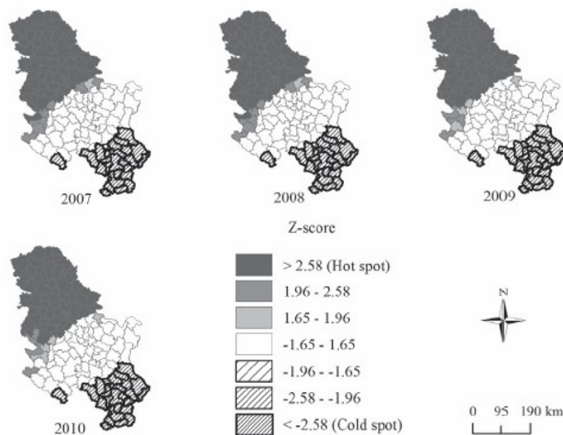
Figure: Spatial clusters (hot and cold spots) of the municipalities in Serbia by the level of average monthly net earning from 2001 to 2010

Spatial regression models (CS data)

- **Parametric models**
  - Spatial dependency is estimated (different functional forms exist).

- **Getis' filter ("nonparametric")**
  - Spatial dependency is removed from observed variables (like s.a. in TS).

- **Semiparametric approach**
  - Observed variables are not transformed, spatial dependency is modelled separately, using map patterns/processes.
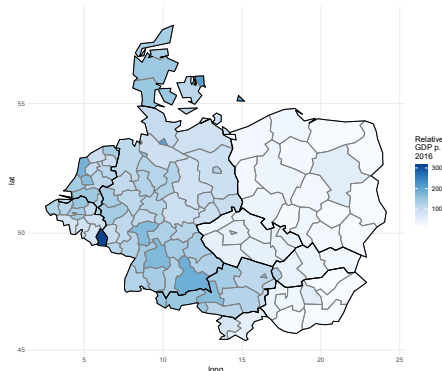
Figure: Relative GDP per capita (100 = EU avg.), 2016, NUTS 2. Source: Eurostat (`nama_10r_lfu3rt`)

Spatial regression models

- Spatial regression models allow us to discern the influence of geographic factors (spatial autocorrelation, spill-over effects) from other relevant factors/variables – say, economic – that may be subject to macroeconomic policy tools.

- Observed data are characterized (indexed) by geographic location, distances between geo-coded units play a crucial role.

Spatial regression models

Spatial regression models – examples

- House prices depend on utility-related variables (area in m^2, number of rooms/bedrooms, elevator, etc.). At the same time, prices in a given area/neighborhood tend to be correlated – usually, positive spatial autocorrelation is present.

- Highway gas-stations' revenues depend on distance from a capital/major city (traffic intensities may be measured – or proxied by distances).
  Can be considered as a special case of one-dimensional spatial dependency (distance measured along the highway route).

- Increased policing activity in a given area (county) may decrease local crime levels while causing increased illegal activities in nearby counties – a negative spatial autocorrelation example.

- Macroeconomic shocks (both positive and negative) may spill more prominently among spatially close economies: we need to account for such effects in econometric models.

Spatial regression models

Spatial regression models – discussion

- Great proportion of spatial effects (spatial dependencies) is closely related to missing variables (significant yet often non-measurable) for a regression model.

- Spatial lag is a proxy variable for (potentially many) unobservable factors. For example, using EU's labor market and/unemployment regression models:
  - Work commuting among regions/districts/states and the problematic consistent measurements of this phenomenon.
  - Language, qualification and administrative barriers on labor market.
  - Aerial distances vs. topology and transportation infrastructure.

- Spatial models may provide a simple and interpretable tool for analysis of macroeconomic (regional) dynamics.

Spatial regression models – main motivations

- *Omitted variables* motivation: discussed on previous slide.
- *Time-dependency*: agents make decisions influenced by the behavior of other agents in previous periods (state authorities may set taxes that reflect policy actions taken by their neighbors in previous periods).
- *Spatial heterogeneity* motivation is largely based on panel data methods and regression models. Spatially close units exhibit more similar individual effects as compared to non-neighboring units.
- *Externalities-based* motivation comes from a well-established economic concept: individuals and regions may be subject to positive/negative consequences of activities exercised by unrelated third parties (heavy traffic, air pollution, etc.).
- *Model uncertainty* motivation: spatial autocorrelation may be used in circumstances where we face uncertainty in terms of specifying a proper data generating process (DGP).

Note that motivations are not mutually exclusive.

Spatial regression models

Generalized nesting spatial (GNS) model

$$y = \lambda Wy + \alpha \iota + X\beta + WX\theta + u \,,$$
$$u = \rho Wu + \varepsilon \,,$$

where Wy is the spatial lag, WX is the spatial lag for regressor matrix X and Wu describes spatial interactions (spatial lag) among disturbance elements. Elements λ, ρ and θ are the spatial parameters, estimated along with α and β.

Under the null hypothesis of no spatial dependency ($\lambda = \rho = 0$ and $\theta = 0$), GNS simplifies to

$$y = \alpha \iota + X\beta + \varepsilon \,,$$

Using apropriate null restrictions, GNS may be transformed to different types of spatial models.

Spatial regression models
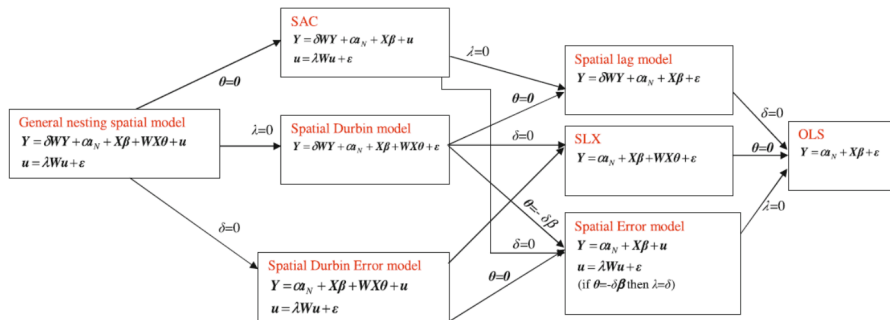
Figure: The relationship between different spatial dependence models for cross-sectional data (source: Elhorst, 2014). Note: this illustration is based on a slightly modified notation of spatial parameters.

Spatial lag model

By assuming spatial interactions in the dependent variable only, ($\boldsymbol{\theta} = \boldsymbol{0}$ and $\rho = 0$), we simplify GNS into a spatial lag model (SLM):

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,.$$

The SLM specification is used commonly throughout empirical literature, e.g. in models describing taxes imposed by governments. Reduced form of SLM can be expressed as:

$$(\boldsymbol{I}_N - \lambda \boldsymbol{W})\boldsymbol{y} = \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,,$$

where $\boldsymbol{I}_N$ is an $(N \times N)$ identity matrix and the RHS regression coefficients explain the variability of individual y_i observations that is not explained spatially. Also, if the inverse to $(\boldsymbol{I}_N - \lambda \boldsymbol{W})$ exists, we can express DGP for $\boldsymbol{y}$ as

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}(\alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\,.$$

Spatial lag model

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,.$$

- For each y_i, the element $SpatialLag(y_i) = \boldsymbol{w}_i\boldsymbol{y}$ enters the equation, where $\boldsymbol{w}_i$ is the i-th row of $\boldsymbol{W}$.

- For model stability, we assume $\lambda \in (-1, 1)$.
  (rule of thumb, see Elhorst, 2014, Spatial Econometrics; From Cross-Sectional Data to Spatial Panels and the discussion provided next).

- Coefficients $\boldsymbol{\beta}$, α and λ are estimated by MLE (if distributional assumptions – e.g. normality of the error term – can be made). Otherwise, GMM may be applied.

- Coefficients $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_k)$ explain (ceteris paribus) variability in $\boldsymbol{y}$ that is not explained spatially.

Spatial Durbin model

If we drop the simplifying assumption $\boldsymbol{\theta} = \mathbf{0}$ from SLM, we get the spatial Durbin model (SDM):

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}\,.$$

For example, if y_i describes household income in a region i, then such income is influenced by incomes (say, wages) in neighboring regions and by both "domestic" and neighboring rates of unemployment, labor force productivities, etc.

By analogy to the SLM case – and given $(\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}$ exists – we may write the DGP as follows:

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}(\alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon})\,.$$

Spatial error model

Spatial error model (SEM) is another frequently used specification of the spatial model. SEM is obtained from the GNS model by assuming $\lambda = 0$ and $\boldsymbol{\theta} = \mathbf{0}$. Hence, spatial interactions take place only among the error terms:

$$\boldsymbol{y} = \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\,,$$
$$\boldsymbol{u} = \rho\boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}\,.$$

Theoretical (macroeconomic) reasoning of the spatial dependency is not required for SEMs – this approach can be used to model a situation where endogenous variables are influenced by exogenous factors that are omitted from the main equation and spatially autocorrelated. Alternatively, unobserved shocks may follow spatial pattern(s).

Spatial model stability conditions

1. Spatial matrix C is a non-negative matrix of known constants with zeros on the diagonal (if this holds for C, it holds for the row-standardized W as well).

2. Spatial weak dependency holds. Two alternative conditions: (a) The row and column sums of C should be uniformly bounded in absolute value as N (the number of observed units) goes to infinity. (b) The row and column sums of C should not diverge to infinity at a speed equal to or faster than the growth of sample size N. Condition (b) is more general (relaxed) and (a) is its special case.

3. Matrices $(I_N - \lambda W)$ and $(I_N - \rho W)$ are non-singular. If the underlying C matrix is symmetric and non-negative, this condition is satisfied whenever λ and ρ lie within the $(1/\kappa_{min}, 1)$ interval, where κ_{min} denotes the smallest (most negative) real eigenvalue of W and 1 is the largest eigenvalue for a row-standardized W.

ML estimation of SLMs and SDMs

- The RHS regressor element $\boldsymbol{Wy}$ in SLM or SDM is correlated with the error term. Hence, OLS estimation of models with spatially lagged endogenous variables is biased and inconsistent.

- ML estimators for such models are consistent (other methods are possible: GMM, 2SLS).

- We use a slightly modified SDM notation to describe ML estimator for both SLMs and SDMs, as their likelihood functions coincide (SLM is a special case of SDM, with $\boldsymbol{\theta} = \boldsymbol{0}$ imposed).

- First, we expand the DGP of SDM by *iid* normality assumption for residuals:

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}(\alpha \boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{WX}\boldsymbol{\theta} + \boldsymbol{\varepsilon}),$$
$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_N),$$

where σ_ε^2 is the variance of $\boldsymbol{\varepsilon}$.

ML estimation of SLMs and SDMs

- In $\boldsymbol{y} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}(\alpha \boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon})$, we use substitution $\boldsymbol{Z} = [\boldsymbol{\iota}, \boldsymbol{X}, \boldsymbol{W}\boldsymbol{X}]$ and $\boldsymbol{\delta} = [\alpha, \boldsymbol{\beta}, \boldsymbol{\theta}]'$.

- We can re-arrange DGP of the SDM/SLM equation as:

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}\boldsymbol{Z}\boldsymbol{\delta} + (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}\boldsymbol{\varepsilon},$$
$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_N).$$

- The above substitution allows us to use a single likelihood function for both SLM and SDM: for SDMs, we use $\boldsymbol{Z} = [\boldsymbol{\iota}, \boldsymbol{X}, \boldsymbol{W}\boldsymbol{X}]$.
  For SLMs, $\boldsymbol{Z} = [\boldsymbol{\iota}, \boldsymbol{X}]$ and analogous amendments are made to the vector of parameters $\boldsymbol{\delta}$.

- Note how our substitution transforms the original SDM model

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \alpha \boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}.$$

  into

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \qquad \rightarrow \qquad \boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{W}\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\delta}.$$

ML estimation of SLMs and SDMs

- The log-likelihood function for the SLM (and SDM) model may be outlined as

$$LL(\lambda, \boldsymbol{\delta}, \sigma_\varepsilon^2) = -\frac{N}{2} \log(\pi \sigma_\varepsilon^2) + \log |\boldsymbol{I}_N - \lambda \boldsymbol{W}| - \frac{\boldsymbol{e}' \boldsymbol{e}}{2\sigma_\varepsilon^2},$$

$$\boldsymbol{e} = \boldsymbol{y} - \lambda \boldsymbol{W} \boldsymbol{y} - \boldsymbol{Z} \boldsymbol{\delta},$$

where N is the number of spatial units, $|\boldsymbol{I}_N - \lambda \boldsymbol{W}|$ is the determinant of this $N \times N$ matrix and $\boldsymbol{e}$ is a vector of residuals.

- Direct estimation (maximization) of the LL function is subject to multiple computational issues. Alternative approach – by means of iterating over concentrated log-likelihood functions – is used.

ML estimation of SEMs

- We add *iid* normality assumption to $\boldsymbol{\varepsilon}$ residuals of the SEM:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}\,,$$
$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma_\varepsilon^2\,\boldsymbol{I}_N)\,,$$

where the intercept term has been incorporated into the $\boldsymbol{X}\boldsymbol{\beta}$ expression for simplicity.

- Now, the log-likelihood function for SEMs has the form

$$LL(\boldsymbol{\beta}, \rho, \sigma_\varepsilon^2) = -\frac{N}{2}\log(\pi\sigma_\varepsilon^2) + \log|\boldsymbol{I}_N - \rho\boldsymbol{W}| - \frac{\boldsymbol{e}'\boldsymbol{e}}{2\sigma_\varepsilon^2}\,,$$
$$\boldsymbol{e} = (\boldsymbol{I}_N - \rho\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\,.$$

- Again, for computational reasons, concentrated log-likelihood functions are calculated iteratively to maximize *LL* function and obtain parameter estimates and corresponding standard errors.

ML estimation & testing of spatial models

- *LL* functions for SDMs (SLMs) and SEMs may be amended to accommodate binomial, count, multinomial and other dependent variables (LeSage, Pace: Introduction to spatial econometrics).

- Likelihood ratio (*LR*) test can be used to evaluate spatial model specification through a set of conveniently chosen restrictions leading to two alternative nested models:

$$LR = 2(\mathcal{L}_{ur} - \mathcal{L}_r) \underset{H_0}{\sim} \chi_q^2,$$

  where $\mathcal{L}_{ur}$ is the maximized *LL* function of the estimated spatial model. The null hypothesis is used to enforce zero restriction on all parameters describing spatial autocorrelation (λ, $\boldsymbol{\theta}$ or ρ – given specification of the unrestricted model) and $\mathcal{L}_r$ is the maximized *LL* of the non-spatial model.

- *LR* tests may be applied more generally (different restrictions may be evaluated).

SLM vs SEM specification tests

- LM-based test for SEM specification evaluates the null hypothesis of no spatial autocorrelation of residuals against the alternative of SEM specification:

$$LM\text{-}SEM = \frac{1}{T} \left(\frac{\hat{\boldsymbol{u}}' \boldsymbol{W} \hat{\boldsymbol{u}}}{\hat{\sigma}^2} \right)^2 \underset{H_0}{\sim} \chi_1^2 \,,$$

where $\hat{\sigma}^2$ is the estimated variance of residuals $\hat{\boldsymbol{u}}$ and $T = \text{tr}(\boldsymbol{W}' \boldsymbol{W} + \boldsymbol{W}^2)$ is trace of the matrix. Note that the *LM-SEM* statistic is just a scaled version of Moran's I.

- The *LM-SEM* test is not robust against SLM specification (spatial dependency in y).

SLM vs SEM specification tests

- The *LM-SLM* statistic is used in OLS-estimated linear models to test H_0 of spatial independence in $\boldsymbol{y}$ against the alternative of its spatial autocorrelation (SLM):

$$LM\text{-}SLM = \frac{1}{N\,J_{\lambda,\beta}} \left(\frac{\boldsymbol{\hat{u}}' \boldsymbol{W} \boldsymbol{y}}{\hat{\sigma}^2} \right)^2 \underset{H_0}{\sim} \chi_1^2 \,,$$

  where the term $J_{\lambda,\beta} = \left[(\boldsymbol{W}\boldsymbol{X}\boldsymbol{\hat{\beta}})' \boldsymbol{M}(\boldsymbol{W}\boldsymbol{X}\boldsymbol{\hat{\beta}}) + T\hat{\sigma}^2 \right] / N\hat{\sigma}^2$ is calculated using the vector of OLS-estimated parameters $\boldsymbol{\hat{\beta}}$ and the "residual maker" (orthogonal projection matrix) $\boldsymbol{M} = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.

- The *LM-SLM* test is not robust against autocorrelated error term (SEM-type).

- The test for a spatial error process that is robust to local presence of a spatial lag is given as:

$$RLM\text{-}SEM = \frac{1}{T - T^2(NJ_{\lambda,\boldsymbol{\beta}})^{-1}} \left(\frac{\hat{\boldsymbol{u}}' \boldsymbol{W} \hat{\boldsymbol{u}}}{\hat{\sigma}^2} - T(NJ_{\lambda,\boldsymbol{\beta}})^{-1} \frac{\hat{\boldsymbol{u}}' \boldsymbol{W} \boldsymbol{y}}{\hat{\sigma}^2} \right)^2 \underset{H_0}{\sim} \chi_1^2,$$

where the subtraction of a correction factor that accounts for the local misspecification (potentially omitted spatial lag process) is visible.

- By *RLM-SEM*, we test the H_0 of no spatial dependency in residuals (OLS-estimated) against the alternative of SEM specification, while controlling for possible local spatial lag (SLM process).

SLM vs SEM specification tests

- Test for a spatial lag process robust to local presence of spatial error autocorrelation is defined as:

$$RLM\text{-}SLM = \frac{1}{N\,J_{\lambda,\beta} - T}\left(\frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\boldsymbol{y}}{\hat{\sigma}^2} - \frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}}{\hat{\sigma}^2}\right)^2 \underset{H_0}{\sim} \chi_1^2 \ .$$

- Heteroskedasticity-robust versions of statistics *LM-SEM* and *LM-SLM* are available.

- Heteroskedasticity-robust versions of the *RLM-SEM* and *RLM-SLM* tests are not easily accessible as accounting for heteroskedasticity leads to highly non-linear expressions.

Marginal effects in spatial models

- In spatial models (SLM, SDM), there are two basic types of marginal effects: direct and indirect effects.

  Sometimes a third effect – total effect – is reported: a sum of the previous two effects.

- In presence of spatial autocorrelation among observed variables (and for SDM), if a given explanatory variable in some i-th unit changes, than not only the dependent variable in the i-th unit is expected to change (direct effect) but also the dependent variables in other units (neighbors of unit i) would change.

  Such effect across spatial units is the indirect effect, sometimes called "spillover".

Marginal effects in spatial models

- Marginal effects can be conveniently illustrated using a slightly modified GNS model:

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}(\boldsymbol{X\beta} + \boldsymbol{WX\theta}) + \boldsymbol{r},$$

where $\boldsymbol{r}$ contains both the intercept and (potentially spatially dependent) error term. Marginal effects for some arbitrary regressor $\boldsymbol{x}_k$ are given by a Jacobian matrix of first derivatives of the expected values of $\boldsymbol{y}$ with respect to the explanatory variable:

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = \left(\frac{\partial E(\boldsymbol{y})}{\partial x_{1k}} \quad \cdots \quad \frac{\partial E(\boldsymbol{y})}{\partial x_{Nk}} \right) = \begin{pmatrix} \dfrac{\partial E(y_1)}{\partial x_{1k}} & \cdots & \dfrac{\partial E(y_1)}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial E(y_N)}{\partial x_{1k}} & \cdots & \dfrac{\partial E(y_N)}{\partial x_{Nk}} \end{pmatrix}.$$

Marginal effects in spatial models

- After some re-arranging, we can write:

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} (\boldsymbol{I}_N \beta_k + \boldsymbol{W} \theta_k).$$

- For convenience and clarity, the RHS may also be re-written as:

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{pmatrix}.$$

- Recall that w_{ij} denotes element of the $\boldsymbol{W}$ matrix and $w_{ij} > 0$ if two spatial units i and j are neighbors (and zero otherwise). β_k and θ_k are parameters of the GNS/SDM, corresponding to the k-th regressor. The RHS is a $N \times N$ matrix.

Marginal effects in spatial models

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{pmatrix}.$$

- Each diagonal element of the partial derivatives matrix represents a direct effect and every off-diagonal element represents an indirect effect.

- Direct effects and indirect effects differ across spatial units. Each element of the RHS matrix might be different.
  - Individual direct effects differ because the diagonal elements of $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ are different for each unit (given $\lambda \neq 0$).
  - Indirect effects are different because off-diagonal element of both $\boldsymbol{W}$ and $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ are different if $\lambda \neq 0$ and/or $\theta_k \neq 0$.

Marginal effects in spatial models

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{pmatrix}.$$

- In absence of spatial autocorrelation of $\boldsymbol{y}$ and $\boldsymbol{x}_k$, i.e. if both $\lambda = 0$ and $\theta_k = 0$, then all off-diagonal elements equal zero. In this case, indirect effects are not present. Also, direct effects are constant (equal to β_k) across all spatial units as $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ simplifies to $\boldsymbol{I}_N$ if $\lambda = 0$.

- The indirect effects that occur if $\theta_k \neq 0$ and $\lambda = 0$ are referred to as **local effects**. The name arises from the fact that such effect only arise from the neighborhood of a given unit. For example, the effect of x_{jk} (k-th regressor for the j-th unit) on y_i is nonzero only if units i and j are neighbors (i.e. $w_{ij} > 0$). For non-neighboring units, x_{jk} has no effect on y_i.

Marginal effects in spatial models

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{pmatrix}.$$

- The indirect effects that occur if $\lambda \neq 0$ and $\theta_k = 0$ are referred to as **global effects**. The name comes from the fact that effects on y_i originate from units that lie within the neighborhood of i as well as from units outside this neighborhood. Mathematically, this is due to the fact that matrix $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ does not contain zero elements (given $\lambda \neq 0$) – even though $\boldsymbol{W}$ does contain (usually many) zero elements.

- For $\lambda \neq 0$ and $\theta_k \neq 0$, both local and global indirect effect are present and they cannot be separated from each other.

Marginal effects in spatial models

$$
\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}
\begin{pmatrix}
\beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\
w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\
\vdots & \vdots & \ddots & \vdots \\
w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k
\end{pmatrix}.
$$

- The presence or absence of spatial autocorrelation ρ in the error term of GNS/SDM has no impact on the marginal effects: as we take the first derivative of $E(\boldsymbol{y})$ with respect to $\boldsymbol{x}_k$, the $\boldsymbol{r} = (\alpha \boldsymbol{\iota} + \rho \boldsymbol{W} \boldsymbol{u} + \boldsymbol{\varepsilon})$ element disappears because it is a "constant".

- Considering the complexity of marginal effects for one regressor, the problem of presenting estimation output from a model with multiple regressors may be severe: even if reliable estimates $\hat{\lambda}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ are available, we have to deal with a $N \times N$ matrix of marginal (direct and indirect) effects for each regressor.

Marginal effects in spatial models

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{pmatrix}.$$

Estimated marginal effects are usually presented in an aggregated form. For each regressor, we usually report two (sometimes three) statistics:

- Summary indicator for direct effects is calculated as the average of all diagonal elements (of the Jacobian for a given regressor $\boldsymbol{x}_k$).

- Indirect effects are reported as the average of all off-diagonal elements.

- The above statistics are usually reported along with their corresponding standard errors and statistical significance indicators (p-values) – see Elhorst (2014) for technical discussion.

Marginal effects in spatial models

$$\frac{\partial E(\boldsymbol{y})}{\partial \boldsymbol{x}_k} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{pmatrix}.$$

- Total effect is just a sum of the direct and indirect impacts. Total standard errors, z scores and significance levels are also calculated by aggregating the underlying direct impacts and spillovers.
- Motivation: in many empirical applications, the direct and indirect effects may come with opposite signs. Therefore, at some higher level of spatial aggregation, direct impacts and spillovers could cancel out. For example, positive direct effects may come at the "price" of equally prominent negative spillovers.
- Total impacts are often reported along with their direct/indirect constituents – even if there are no contradicting signs of direct/indirect impacts.

Marginal effects in spatial models

- Sample output from a SLM (model with a single regressor).

Table: Estimated direct, indirect and total impacts

	Mean	Std. dev	t-statistic
Direct effect	0.586	0.0148	39.6106
Indirect effect	1.084	0.0587	18.4745
Total effect	1.670	0.0735	22.7302

Drawbacks of SLMs/SDMs/SEMs

- In SDM/SLM, the ratio between direct and indirect effect for a regressor $\boldsymbol{x}_k$ is independent of β_k. This is because the β_k coefficients cancel out in the numerator and denominator of such ratio (direct/indirect effects).

- This ratio depends only on the parameter λ and on the $\boldsymbol{W}$ matrix specification.

- Hence, direct/indirect effect ratio is the same for all regressors in a given spatial model.

- Unfortunately, this "behavior" (identical relative strengths of direct and indirect effects for all regressors) seems rather implausible in many types of empirical applications.

Drawbacks of SLMs/SDMs/SEMs

- W matrices cannot be estimated along with model parameters.

- Rather, W needs to be specified prior to model estimation.

- There is little theoretical background for choosing the "right" W matrix specification.

- As W matrix in a SLM/SDM/SEM changes, the estimated parameters tend to change as well. Large changes to W may cause large changes in estimated parameters.

- The variety of available neighborhood definitions and standardization methods implies that researchers usually evaluate several alternative spatial structure settings in order to verify model stability and robustness of the results.

Robustness against changes in spatial structure

1. Start with a relatively sparse W distance matrix, that is generated by using a restrictive (i.e. low) τ. Note that the threshold must be high enough to ensure at least one neighbor for each spatial unit

2. Increase τ by some low amount (eg 10-km in NUTS2 units). Estimate the model and record all relevant information.

3. Repeat step 2 until a some preset maximum τ is reached. Usually, this would happen in one of the following manners: (a) "range" (as in the semivariogram) is reached, (b) τ becomes so large that the assumption of spatial weak dependency no longer applies, (c) We have prior information limiting the plausible range of spatial interactions.

4. Plot the estimated spatial parameters, direct and indirect effects of interest and statistical significance data against the distance thresholds used.

Analogous approach can be used for kNN, contiguity neighbors, etc.

Spatial robustness – SLM illustration

Spatial filtering and semiparametric models
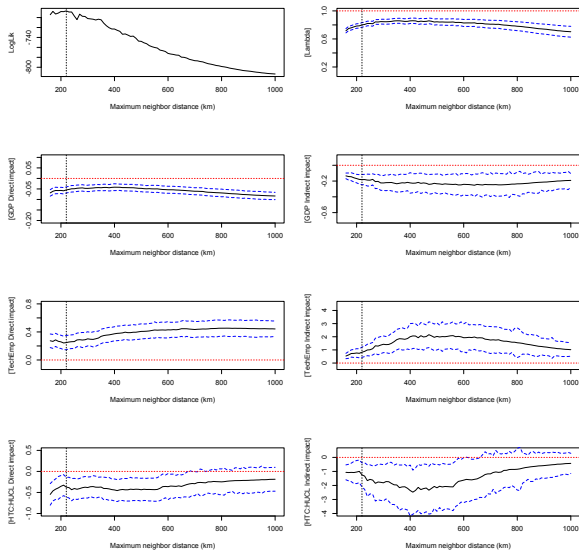
Spatial filtering and semiparametric models

- Parametric methods are potentially not robust in a situation where the model suffers from a simultaneous presence of different sources of misspecification.
  Unaccounted nonlinear relationship among spatially correlated variables, spatially varying relationships (non-stationarity), uncontrolled common factors (spatial and time-related) and other instances of spatial heterogeneity can disrupt spatial dependencies or even manifest themselves as such.

- Spatial filtering may be used to remove global and/or local spatial dependencies among geo-coded variables.

- Spatial filtering does not rely on distributional assumptions and is fairly robust to model misspecification.

Spatial filtering and semiparametric models

- Getis' nonparametric filtering can be used to eliminate spatial autocorrelation from observed data by "spatial demeaning". Based on $G_i(\tau)$ by Getis – only applicable for positive spatial autocorrelation and positive support of the observed spatial variables.

- Moran eigenvector map (MEM) filtering can be used to preserve some level of spatial properties within the model – it is a semiparametric method.

Getis univariate filtering

- Approach based on $G_i(\tau)$ statistic by Getis:

$$G_i(\tau) = \frac{\sum_{j=1}^{N} c_{ij}\, y_j}{\sum_{j=1}^{N} y_j}, \qquad ; \qquad E\left[G_i(\tau)\right] = \frac{c_i}{N},$$

where c_i is the sum of elements of i-th row of $\boldsymbol{C}$
(note that $\boldsymbol{C}$ is used for filtering, not the $\boldsymbol{C}^* = \boldsymbol{C} + \boldsymbol{I}_N$).

- Multiplicative transformation (filtering) of a spatial variable y_i:

$$\ddot{y}_i = \frac{E\left[G_i(\tau)\right]}{G_i(\tau)} \cdot y_i,$$

where $\ddot{y}_i$ is the spatially filtered value of y_i.

Getis univariate filtering

$$\ddot{y}_i = \frac{E\left[G_i(\tau)\right]}{G_i(\tau)} \cdot y_i \,,$$

- Corrects for positive spatial autocorrelation in observed data by counterbalancing the clustering.

- Filtering factor shrinks y_i if the majority of observations y_j within the τ distance of unit i are above average.

- Similarly, y_i is inflated if neighboring observations feature below-average values.

- Simple and intuitive, yet positive support assumption can be a limitation. Also, the process of setting τ is arbitrary (robustness evaluation possible).

- Univariate – to estimate models using filtered data, such as $\ddot{y} = \alpha\iota + \ddot{X}\beta + \varepsilon$, all variables have to be filtered individually.

Moran's eigenvector map (MEM) filtering

Econometric models are partitioned into two elements that are both employed within the spatial regression:

- Exogenous regressors (say, macroeconomic indicators)

- Complementary spatial-filter variate (e.g. a MEM) describing spatial dependency that would otherwise affect model's residuals.

- For MEM-based spatial filtering, all units need to be connected – i.e. there has to be a "path" between any two regions, based on connections (edges) between neighboring units. Hence, for spatial filtering applications, τ value must not be lower than the longest edge of a minimum spanning tree of a graph based on $\boldsymbol{D}$ (where d_{ij} are distances between units). Maximum limit on τ also applies (recall weak dependency assumptions).

Semiparametric model with MEM-filter

1. Use C to construct a double-centered connectivity matrix Ω:

$$\Omega = (I_N - \frac{1}{N}\, \iota_N \iota_N')\, C\, (I_N - \frac{1}{N}\, \iota_N \iota_N'),$$

   where I_N is the identity matrix and ι_N is a column vector of ones. Double-centering standardizes C so that all rows and columns in Ω add up to zero. C may be full rank (not guaranteed), Ω is always rank deficient.

2. Calculate all distinct eigenvectors v_i and eigenvalues κ_i for the double-centered matrix Ω – by solving the characteristic equation system $(\Omega - \kappa\, I_N)\, v = 0$. As Ω is a real symmetric matrix, all distinct eigenvectors are orthogonal and linearly independent (given rank-deficiency of Ω, some eigenvalues and eigenvectors are duplicated).

   MEM – convenient subset of eigenvectors v – can be used as a synthetic explanatory variable in semiparametric regression.

Semiparametric model with MEM-filter

4 Semiparametric model is established using a misspecification paradigm and based on a regression model with spatially autocorrelated disturbances:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\,,$$
$$\boldsymbol{u} = \boldsymbol{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\,,$$

where $\boldsymbol{u}$ are the spatially autocorrelated disturbances that may be decomposed into $\boldsymbol{\varepsilon}$ (white noise) and $\boldsymbol{E}$: a set of missing (unobservable) exogenous variables that follow common spatial dependency pattern ($\boldsymbol{C}$) and $\boldsymbol{\gamma}$ is a vector of parameters.

The $\boldsymbol{E}\boldsymbol{\gamma}$ term is based on approximation and it is not directly comparable with the SEM parametric specification

5 Using a stepwise algorithm described next, the misspecification term $\boldsymbol{E}$ is approximated by MEM – a convenient subset of Morans eigenvectors

$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} \equiv \text{evec}(\boldsymbol{\Omega}) \ ,$$

where $r = \text{rank}(\Omega)$ is the number of distinct eigenvectors of $\boldsymbol{\Omega}$.

We can approximate $\boldsymbol{E}$ by a parsimonious subset of eigenvectors $\boldsymbol{v}$ such that $\boldsymbol{M} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_\ell\}$ where $\ell \leq r$.

Since spatial misspecification term $\boldsymbol{E}$ is correlated with $\boldsymbol{X}$ (at least potentially), the explicit inclusion of proxy element (MEM) $\boldsymbol{M}$ corrects for the bias in estimated coefficients $\hat{\boldsymbol{\beta}}$.

6 After substitution of $\boldsymbol{E}$ by $\boldsymbol{M}$, we get

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{M}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\,,$$

where $\boldsymbol{y}$ is decomposed into a systematic component (featuring $\boldsymbol{X}$), stochastic spatial component and white-noise residuals.

For conveniently specified $\boldsymbol{M}$, the stochastic spatial term $\boldsymbol{M}\hat{\boldsymbol{\gamma}}$ removes a significant portion of the mean squared error (MSE) term, attributable to spatial autocorrelation

Hence, $\boldsymbol{M}\hat{\boldsymbol{\gamma}}$ is often referred to as spatial filter.

Filtering is fairly robust to model specification errors when compared with fully parametric models.

Semiparametric model with MEM-filter

Stepwise algorithm for choosing M:

- We choose M so that the residuals $\hat{\varepsilon}$ become spatially random (independent with respect to the underlying spatial domain).

- We aim to find a parsimonious, i.e. smallest possible subset of eigenvectors v, leading to spatial independence of $\hat{\varepsilon}$.

- Stepwise regression approach is often used.

- Stepwise algorithm often based on a modified Moran's I coefficient: Moran's MC.

Semiparametric model with MEM-filter

Stepwise algorithm for choosing $\boldsymbol{M}$:

- $MC_{\boldsymbol{v}_i} = \frac{N}{\boldsymbol{\iota}_N' \, \boldsymbol{Z} \, \boldsymbol{\iota}_N} \, \boldsymbol{v}_i' \boldsymbol{C} \boldsymbol{v}_i$,

  where $\boldsymbol{Z}$ is a distance-based similarity matrix with $[z_{ij}] = 1 - (\frac{h_{ij}}{\max(h_{ij})})^2$. Individual z_{ij} vary between zero for $h_{ij} = \max(h_{ij})$ and 1 for $h_{ij} = 0$.

- First eigenvector is selected by maximizing $MC_{\boldsymbol{v}_i}$.

- Using such $\boldsymbol{v}_i$ as a starting eigenvector subset for $\boldsymbol{M}$, equation $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{M}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ is estimated and residuals are evaluated for spatial autocorrelation (e.g. using Moran's I).

- If residuals are spatially dependent, new eigenvector is added to $\boldsymbol{M}$ using the same MC criterion, model is re-estimated and spatial autocorrelation in residuals is tested again.

- Eigenvectors are iteratively added to $\boldsymbol{M}$, until spatial autocorrelation in residuals $\hat{\boldsymbol{\varepsilon}}$ falls below a predetermined threshold (e.g. 5 % significance level).

Semiparametric model with MEM-filter

$$y = X\beta + M\gamma + \varepsilon$$

- Eigenvectors in M are orthogonal and follow a strictly decreasing sequence.

- Each eigenvector explains a specific proportion of variance in residuals of the model – the largest proportion of variance is explained by the first eigenvector selected into M, the second largest amount of variance is explained by the second eigenvector, etc.

- This leads to identical M matrices, obtained through forward and backward stepwise selection methods.

- Convenient linear combination of the above eigenvectors (i.e. $M\hat{\gamma}$) is our spatial filter.

- Often, we only take into account positive spatial autocorrelation: only v_i associated with positive κ_i are considered.

Semiparametric model with MEM-filter

$$y = X\beta + M\gamma + \varepsilon$$

Individual v elements of M bear the following interpretation:

- MEMs – spaces spanned by single or multiple v eigenvectors – with associated "large" positive eigenvalues κ_i represent global-scale spatial trends (say, landscape-wide dynamics in observed spatial data).

- Eigenvectors with medium eigenvalues represent medium scale dynamics.

- Eigenvectors with small positive eigenvalues would represent small scale dependencies ("local" patchiness).

MEM - illustrative example
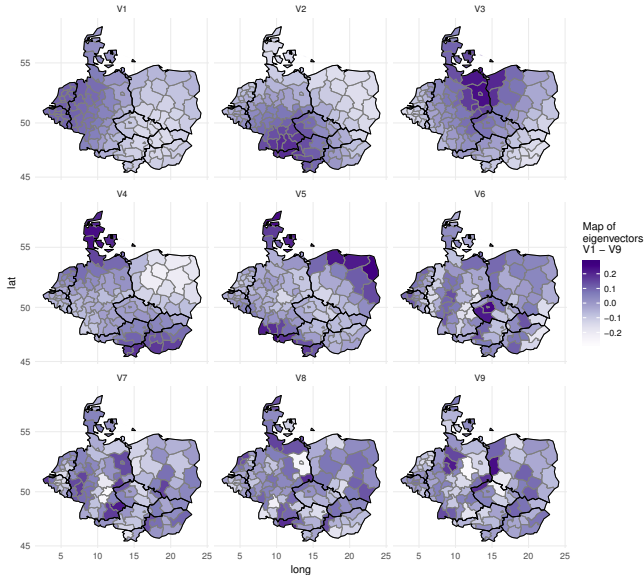
Spatial panel models (parametric)

Spatial panel models

For repeated geo-coded observations, spatial panel models depict interactions among variables across spatial units and over time:

$$\boldsymbol{y} = \lambda\left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right)\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\,,$$
$$\boldsymbol{u} = \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N\right)\boldsymbol{\mu} + \boldsymbol{\varepsilon}\,,$$
$$\boldsymbol{\varepsilon} = \rho\left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right)\boldsymbol{\varepsilon} + \boldsymbol{v}\,,$$

where $\boldsymbol{y}$ is a $NT \times 1$ column vector of dependent variable observations ($i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$ and i is the "fast" index). $\boldsymbol{W}$ is a spatial weights matrix and $\boldsymbol{X}$ is a ($NT \times k$) matrix of exogenous regressors. Elements $\boldsymbol{I}_T$ and $\boldsymbol{I}_N$ are identity matrices and $\boldsymbol{\iota}_T$ is a ($T \times 1$) vector of ones. Operator $\otimes$ is the Kronecker product. Elements of vector $\boldsymbol{\beta}$ as well as λ and ρ are parameters of the model. The disturbance vector $\boldsymbol{u}$ ($NT \times 1$) is a sum of two terms: the unobserved individual effects $\boldsymbol{\mu}$ and spatially autocorrelated innovations $\boldsymbol{\varepsilon}$. The ($N \times 1$) vector $\boldsymbol{\mu}$ holds time-invariant and spatially uncorrelated individual effects. Innovations $\boldsymbol{\varepsilon}$ are spatially autocorrelated with a spatial error autoregressive parameter ρ where $|\rho| < 1$. $\boldsymbol{v}$ is a vector of spatially independent innovations: $v_{it} \sim IID\left(0, \sigma_v^2\right)$.

Spatial panel models

$$\boldsymbol{y} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W} \right) \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u},$$
$$\boldsymbol{u} = \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N \right) \boldsymbol{\mu} + \boldsymbol{\varepsilon},$$
$$\boldsymbol{\varepsilon} = \rho \left(\boldsymbol{I}_T \otimes \boldsymbol{W} \right) \boldsymbol{\varepsilon} + \boldsymbol{v},$$

- Spatial dependency tests (Moran's I) for panel data are available.
- By analogy to the GNS model (for CS data), the above specification is simplified (restricted) before estimation.
- Generally, $\boldsymbol{\beta}$ parameters do not constitute marginal effects – direct/indirect effects have to be calculated (using the same general approach as for CS data).
- With "random effects" (RE) models, we assume that unobserved individual effects $\boldsymbol{\mu}$ are not correlated to other regressors of the model.
- For "fixed effects" (FE) models, some level of correlation between individual effects and other regressors is acceptable.

Spatial panel models

We shall briefly discuss three types of spatial panel models:

- Fixed effects spatial lag model

- Fixed effects spatial error model

- Random effects spatial lag model

Fixed effects spatial lag model

- FE spatial lag model that does not feature spatial autocorrelation in the error term can be written as

$$\boldsymbol{y} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \boldsymbol{y} + \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N\right) \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}\,,$$

where $\boldsymbol{v}$ is a vector of spatially independent and normally distributed innovations.

- First step in FE spatial lag model estimation: eliminating the individual effects:

$$\ddot{\boldsymbol{y}} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \ddot{\boldsymbol{y}} + \ddot{\boldsymbol{X}}\boldsymbol{\beta} + \ddot{\boldsymbol{v}}\,,$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$, etc.

- As individual effects are time-invariant, $\ddot{\mu}_i = \mu_i - \bar{\mu}_i = 0$ and $\boldsymbol{\mu}$ disappears from equation.

Fixed effects spatial lag model

$$\boldsymbol{y} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W} \right) \boldsymbol{y} + \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N \right) \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}$$

- FE spatial lag model is estimated by MLE method. For detailed description, see Milo, Piras (2012)

- Direct/indirect/total effects and their statistical significance is calculated (bootstrapped standard errors) – see Milo, Piras (2012).

- From estimated FE model, we can calculate individual effects:

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} \left[y_{it} - \hat{\lambda} \left(\sum_{j=1}^{N} w_{ij} y_{jt} \right) - \boldsymbol{x}_{it}\hat{\boldsymbol{\beta}} \right] .$$

However, for short panels ($N \gg T$), enough observations for reliable μ_i estimation often do not accumulate.

Fixed effects spatial error model

- Spatial lag is absent in the FE spatial error model:

$$\boldsymbol{y} = (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N)\,\boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,,$$
$$\boldsymbol{\varepsilon} = \rho\,(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{\varepsilon} + \boldsymbol{v}\,.$$

- Model is estimated by MLE (time-demeaning is used).

- Coefficients $\boldsymbol{\beta}$ are the marginal effects here

- Individual effects can be retrieved from the estimated model as

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \boldsymbol{x}_{it}\hat{\boldsymbol{\beta}}\right),$$

and reliability issues for short panels ($N \gg T$) apply.

Random effects spatial lag model

- Individual effects can be viewed as random and independent of regressors – random effect (RE) model and RE estimator provide efficient estimates. RE properties of $\boldsymbol{\mu}$:
  $\mu_i \sim IID\left(0, \sigma_\mu^2\right)$; normality is often assumed.

- Under RE assumptions, GNS panel model can be re-written as:

$$\boldsymbol{y} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u},$$
$$\boldsymbol{u} = \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N\right) \boldsymbol{\mu} + \left(\boldsymbol{I}_T \otimes \boldsymbol{B}_N^{-1}\right) \boldsymbol{v}$$

where $\boldsymbol{B}_N = \left(\boldsymbol{I}_N - \rho \boldsymbol{W}\right)$ and error variance may be outlined as

$$\operatorname{var}\left(\boldsymbol{u}\right) = \boldsymbol{\Omega}_{\boldsymbol{u}} = \sigma_\mu^2 \left(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_N\right) + \sigma_{\boldsymbol{v}}^2 \left[\boldsymbol{I}_T \otimes \left(\boldsymbol{B}_N' \boldsymbol{B}_N\right)^{-1}\right].$$

Random effects spatial lag model

$$y = \lambda \left(I_T \otimes W \right) y + X\beta + u,$$

$$u = \left(\iota_T \otimes I_N \right) \mu + \left(I_T \otimes B_N^{-1} \right) v$$

- Model is estimated by MLE (iterative estimation involving concentrated LL functions), see Milo, Piras (2012).

- Direct/indirect/total effects can be calculated.

- RE vs. FE: Generalized Hausman test can be applied:

$$H = NT \left(\hat{\theta}_{RE} - \hat{\theta}_{FE} \right)' \left(\hat{\Sigma}_{RE} - \hat{\Sigma}_{FE} \right)^{-1} \left(\hat{\theta}_{RE} - \hat{\theta}_{FE} \right) \underset{H_0}{\sim} \chi_k^2,$$

where $\hat{\theta}_{RE}$ and $\hat{\theta}_{FE}$ are the RE and FE-based estimates, $\hat{\Sigma}_{RE}$ and $\hat{\Sigma}_{FE}$ are corresponding variance-covariance matrices. Under H_0 (RE assumptions hold), the test statistic is asymptotically χ^2-distributed with k d.f. (k is the number of regressors).

Spatial analysis and regression models in R

- {spdep}
- {splm}
- {sf}, {sp}, {ggplot2}
- https://cran.r-project.org/web/views/Spatial.html

Unemployment Rate (by State)

Month: Aug 2013 Seasonal Adjustment: Seasonally Adjusted Units: %