Spatial Pattern in Residuals and Spatial Autocorrelation Tests with Moran's I and local I_i

Outline:

- What is spatial autocorrelation
- Definition of the global and local link matrices
- Definition of coding schemes
- Why autocorrelation in regression residuals?
- Motivating example: Italian Fertility
- General structure of global Moran's I and its distribution under spatial independence.
- Eigen-spectrum of the link matrix and bounds of global Moran's I
- Test under randomization.
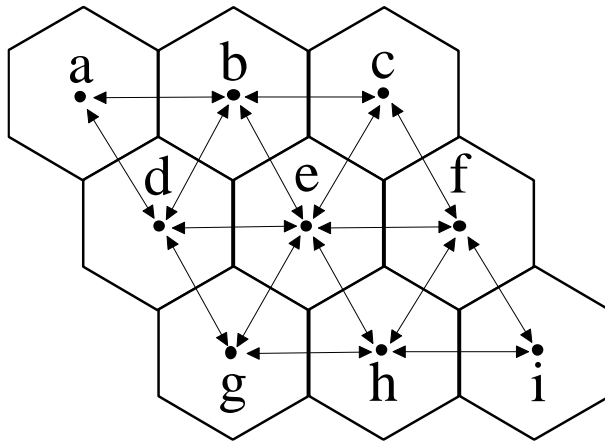- Local Moran's I_i

What is Spatial Autocorrelation

- The concept of an ***internal spatial relationship*** (order) is trickier than in the time-series situation, where the order comes naturally (the future depends on the past)

- In spatial analysis, the relationships among the observations are [a] **multidirectional**, [b] **multilateral** and [c] **not equally spaced**.
  There are also more observations at the **edge of the study area** and factors like **spatial extend of the regions** and **underlying densities** become important.

- Geographical theories provide many **concepts of spatial relationships** within a single variable.
  For instance:

  - simple distances between points or representative points of areas

  - neighborhood relationships between areas (rook's or queen's specification in square tessellations, higher order neighborhood relationship of regions several neighbors apart)

  - traffic flows or migration patterns

  - spatial hierarchies (hub and spokes)

  - diffusion processes etc.

The Spatial Link Matrix

- The spatial connectivity matrix operationalizes the underlying structure of the potential spatial relationships among the observations

- For potential distance relationships we have the distance matrix (known from road atlases, perhaps using spherical distances)

- For potential neighborhood relationships we must use a binary spatial connectivity matrix

Example: Encoding a spatial tessellation as a binary connectivity matrix

	a	b	c	d	e	f	g	h	i
a	**0**	1	0	1	0	0	0	0	0
b	1	**0**	1	1	1	0	0	0	0
c	0	1	**0**	0	1	1	0	0	0
d	1	1	0	**0**	1	0	1	0	0
e	0	1	1	1	**0**	1	1	1	0
f	0	0	1	0	1	**0**	0	1	1
g	0	0	0	1	1	0	**0**	1	0
h	0	0	0	0	1	1	1	**0**	1
i	0	0	0	0	0	1	0	1	**0**

$\Longleftrightarrow$

Spatial tessellation of 9 hexagonal cells with underlying connectivity structure
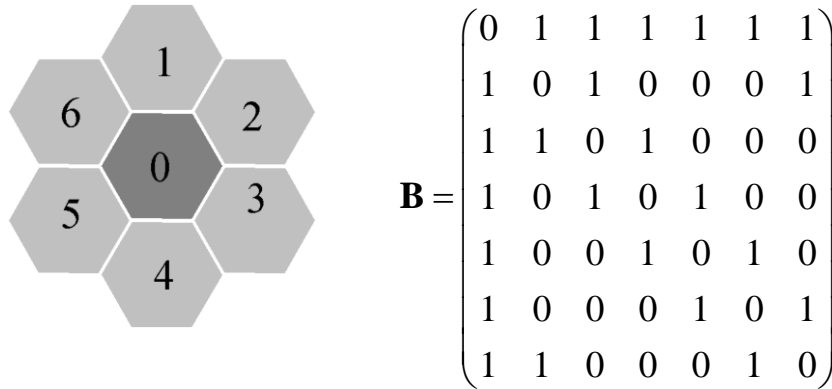
Binary 9×9 spatial connectivity matrix

B

- An element $b_{ij} = 1$ denotes that the tiles *i* and *j* are ***adjacent*** and an element $b_{ij} = 0$ signifies that the tiles *i* and *j* are not common neighbors

- The spatial connectivity matrix $\mathbf{B}$ is ***symmetric***

- A tile is not connected to itself. Thus all ***diagonal*** elements are zero

- For study areas with a ***large number*** *n* of individual regions the generation of the connectivity matrix $\mathbf{B}$ (or distance matrix) must be left to a GIS program. The connectivity matrix has $n \times n$ elements

- Problems occur if we have island and holes in our study area. Usually machine generated connectivity matrices must be polished manually.

- Most of the elements are zero. There are efficient storage modes for sparse matrices. For empirical map patterns an area in the interior has on average 6 neighbors.

Local and Global Spatial Link Matrices

- Each global link matrix can be decomposed into a set of *n* local link matrices. *Vice versa*, a set of local link matrices can be aggregated into a global link matrix.

- Example: The hexagonal spatial tessellation has the global link matrix:

$$
\mathbf{B} = \begin{pmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
$$

The *local link matrices* are:

$$
\mathbf{B}_0 = \begin{pmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\quad
\mathbf{B}_1 = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\cdots
$$

$$
\mathbf{B}_5 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
\quad
\mathbf{B}_6 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
$$

- The local link matrices are star-shaped and associate the satellites to the reference cell *i*.

- The local link matrices satisfy the additive condition:
$$\mathbf{B} = \tfrac{1}{2} \cdot \sum_{i=0}^{6} \mathbf{B}_i$$
The adjustment factor $\tfrac{1}{2}$ is required because each matrix cell in the summation is counted twice.

This decomposition allows to *relate* local spatial statistics to their global counterpart
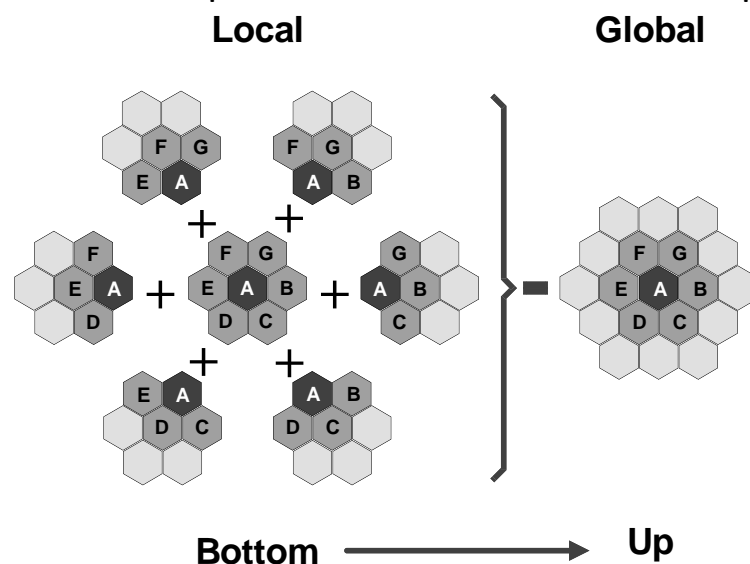
Coding schemes

- The spatial link matrix needs to be standardized in order to control for the increase in the overall linkage a as the number of spatial objects n in the tessellation increases.

- In irregular tessellations each interior cell has on average 6 neighbors (see Euler's Theorem).

- Three coding schemes to standardize the global link matrix are commonly in use.

o Each coding scheme has its specific properties and it is based either on the linkage degree vector.

o Note that edge cells on the boundary usually have fewer neighbors.

- These coding scheme are based on the linkage degree of each cell, which reflect *edge effects* and *spatial heterogeneity*.

- Three coding schemes:

  o The globally standardized C-coding scheme uses the overall linkage degree:

$$\mathbf{C} \equiv \frac{n}{a} \cdot \mathbf{B} = \frac{7}{24} \cdot \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

  Implication: The variance of cells with large linkage degree is larger than for cells with a smaller number of neighbors.

  o The row-sum standardized W-coding scheme uses the degree vector and divides each row to the link matrix by the linkage degree of that row:

$$
\mathbf{W} \equiv \begin{pmatrix}
\frac{1}{d_0} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{d_1} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{d_2} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{d_3} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{d_4} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{d_5} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{d_6}
\end{pmatrix} \cdot \mathbf{B} = \begin{pmatrix}
0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
\frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \\
\frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\
\frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\
\frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0
\end{pmatrix}
$$

This standardized link matrix is no longer symmetric. The row-sums are constant one.

Implication: Cells with a low linkage degree have a larger variance.

o The variance stabilizing S-coding scheme used the square root of the linkage degree and a global scaling constant:

$$
\mathbf{S} \equiv \frac{n}{\sum_{i=0}^{6} \sqrt{d_i}} \cdot \begin{pmatrix}
\frac{1}{\sqrt{d_0}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\sqrt{d_1}} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{d_2}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\sqrt{d_3}} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{\sqrt{d_4}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{d_5}} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{d_6}}
\end{pmatrix} \cdot \mathbf{B} = \frac{7}{12.84} \begin{pmatrix}
0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\
\frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\
\frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\
\frac{1}{\sqrt{3}} & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 \\
\frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0
\end{pmatrix}
$$

o The coded spatial link matrix is asymmetric

Implication: The variance of each spatial object is approximately identical.

- The sums over all cells in a coded spatial link matrix is equal to n.

- These coding scheme and others can be specified an a unified equation defining the **family of coding schemes**:

$$\mathbf{V}_{[q]} = \frac{n}{\sum_{i=1}^{n} d_i^{q+1}} \cdot \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}^q \cdot \mathbf{B}$$

with

$$q = \begin{cases} 0 & C\text{-coding scheme} \\ -1/2 & S\text{-coding scheme} \\ -1 & W\text{-coding scheme} \end{cases}$$

- With the exception of $\mathbf{V}_{[0]}$, i.e., the C-coding scheme, none of the other transformed link matrices are symmetric.

- **Spatial econometrics** prefers the row-sum standardized coding scheme and denotes the standardized link matrix by $\mathbf{W}$. In contrast, geo-statistics usually works with standardized interpoint distance matrices.

Excursion: Spatial Regression Residuals

- In general terms, the residuals can be expressed by the projection matrix

$$\mathbf{M} = \mathbf{I} - \mathbf{X} \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T$$

with

$$\mathbf{e} = \mathbf{M} \cdot \mathbf{y}$$

$$= \left(\mathbf{I} - \mathbf{X} \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \right) \cdot \mathbf{y}$$

$$= \mathbf{y} - \mathbf{X} \cdot \underbrace{\underbrace{(\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y}}_{\mathbf{b}}}_{\hat{\mathbf{y}}}$$

- where X is the matrix of the exogenous variables including the constant unity vector for the regression intercept.

- If we measure only the variation of the dependent variables around its mean then the projection matrix reduces to $\mathbf{M}_{(1)} = \mathbf{I} - \mathbf{1} \cdot \underbrace{(\mathbf{1}^T \cdot \mathbf{1})^{-1}}_{=\frac{1}{n}} \cdot \mathbf{1}^T$

$$\mathbf{M}_{(1)} \equiv \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{pmatrix}.$$

- This gives $\mathbf{e} \equiv \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}$

- Due to the projection matrix the residuals are always slightly correlated, that is,
$Cov(\mathbf{e} \cdot \mathbf{e}^T) = \sigma^2 \cdot \mathbf{M}$.

Why autocorrelation in regression residuals?

- Regression residuals $\mathbf{e}$ capture the unexplained part of the regression model
- They are usually assumed to relate to population disturbances that are $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \cdot \mathbf{I})$. The following reasons lead to spatial autocorrelation
  - (1) Misspecification Rational: if we are missing relevant variables in $\mathbf{X}$ that exhibit a spatial pattern, their spatial pattern will spill over into the regression residuals $\mathbf{e}$.
  - (2) Spatial Process Rational: The spatial objects exhibit some *spatial exchange relationships*, e.g.,
    - interaction flows,
    - competition effects or
    - agglomerative advantages
    - then they will become spatially autocorrelated.
      These exchange relationships cannot be captured by the regression matrix $\mathbf{X}$ but

manifest in a ***covariance*** matrix $\mathbf{\Omega}(\rho)$ with $\mathbf{\varepsilon} \sim N(0, \sigma^2 \cdot \mathbf{\Omega}(\rho))$

where ρ is the ***autocorrelation level*** and measuring the strength of the spatial

process. For

$$\rho = \begin{cases} > 0 & \text{postive autocorrelation} \\ 0 & \text{spatial independence} \\ < 0 & \text{negative autocorrelation} \end{cases}$$

- Moran's I is tailored to work best for Gaussian spatial processes with

$$\mathbf{\Omega}(\rho) = \begin{cases} \text{simultaneous autoregressive spatial process} \\ \text{conditional autoregressive spatial process} \\ \text{moving average spatial process} \end{cases}$$

- All these processes dependent on the ***coded*** spatial link matrix $\mathbf{V}$ among the spatial objects, which reflects the spatial relationships and that is derived from the binary links $\mathbf{B}$.

o <u>(3) Spatial Aggregation Rational:</u> If areal objects are split into parts and these split parts are merged with adjacent areal objects then these aggregated objects share parts of the information that they inherited from the split objects.

This induces spatial autocorrelation among adjacent spatial objects.
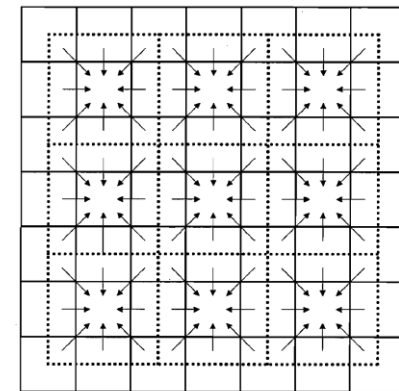
Fig. 4.2. Non-hierarchical aggregation of elemental spatial objects defined by a rectangular 7 × 7 tessellation into a rectangular 3 × 3 tessellation

Implication: One needs to get the spatial scale of analysis right.

Motivating Example: Fertility in Italy

- Several spatial link matrices are conceivable:

Augmented Contiguity Matrix   **Major 424 Migration Links among Provinces**
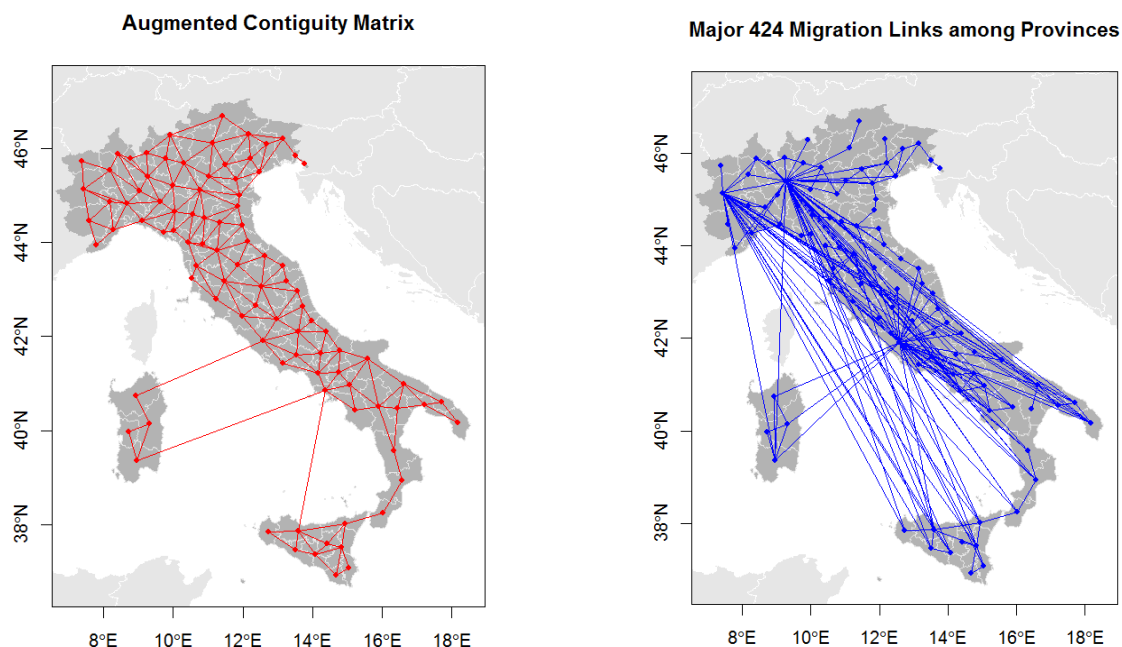
- The spatial relationships can be captured in the spatial link matrix **V**
- The map pattern of the observed dependent variable ***Total Period Fertility*** and the autocorrelation of the regression residuals around the mean is $e_i = y_i - \bar{y}$:

Provinical Total Fertility Rate 1994

Fertility Rate
- under 0.94
- 0.94 - 1
- 1 - 1.04
- 1.04 - 1.12
- 1.12 - 1.19
- 1.19 - 1.35
- 1.35 - 1.51
- over 1.51
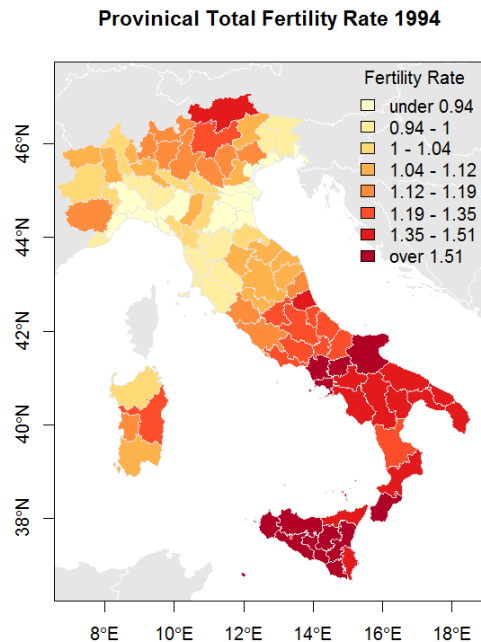
```
Global Moran's I for regression residuals

model: lm(formula = TOTFERTRAT ~ 1, data = prov.df)
       weights: nb2listw(prov.nb, style = "S")


Moran I statistic standard deviate = 12.7804, p-value <
2.2e-16
alternative hypothesis: greater
sample estimates:
Observed Moran's I          Expectation              Variance
     0.853201213          -0.010638298           0.004568551
```

- The applied *regression model* to explain the Total Fertility Rate is

```
lm(formula = TOTFERTRAT ~ FEMMARAGE9 + DIVORCERAT + log(ILLITERRAT) +
    TELEPERFAM, data = prov.df)

Residuals:
     Min       1Q    Median        3Q       Max
-0.19958 -0.05474 -0.01284   0.05272   0.42922

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.78139    0.48606    9.837 6.23e-16 ***
FEMMARAGE9   -0.09647    0.02050   -4.706 9.11e-06 ***
DIVORCERAT   -0.11839    0.05772   -2.051   0.0431 *
```

```
log(ILLITERRAT)   0.03072    0.01707   1.799   0.0753 .
TELEPERFAM       -1.28499    0.18078  -7.108 2.69e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1047 on 90 degrees of freedom
Multiple R-squared: 0.8051,     Adjusted R-squared: 0.7965
F-statistic: 92.96 on 4 and 90 DF,  p-value: < 2.2e-16
```


Quantile-Normal Plot of Base Model
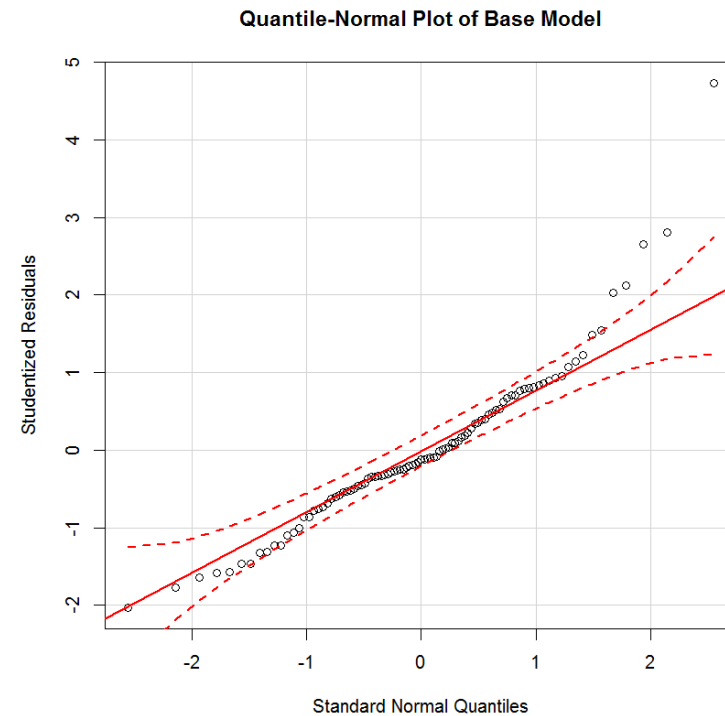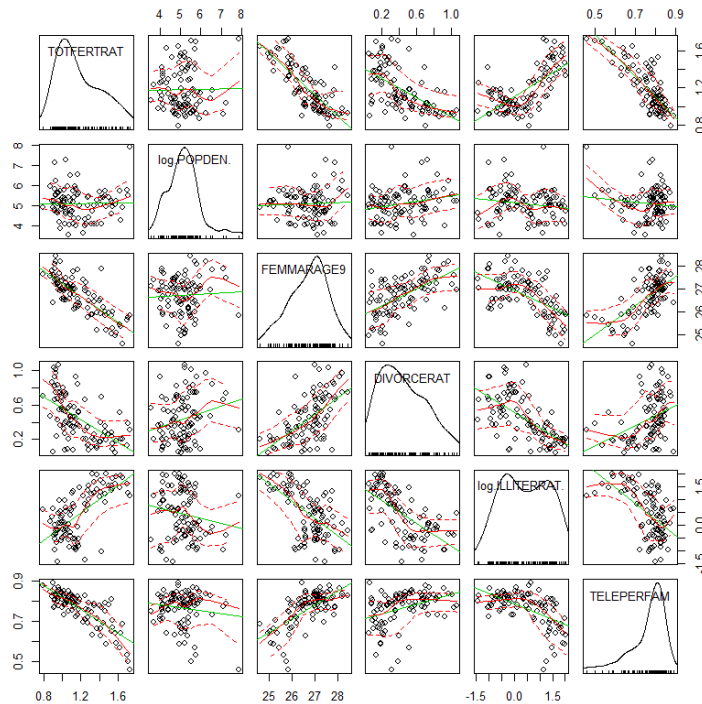
- The map pattern of the residuals $e_i = y_i - \hat{y}_i$ and their autocorrelation level are
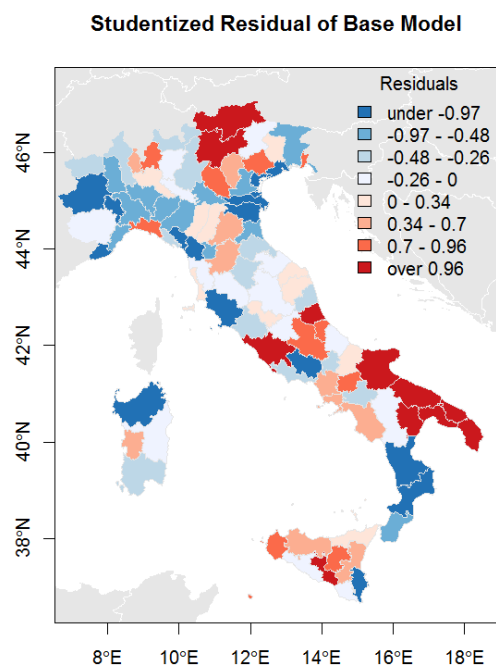
Studentized Residual of Base Model


```
             Global Moran's I for regression residuals

model: lm(formula = TOTFERTRAT ~ FEMMARAGE9 +
        DIVORCERAT + log(ILLITERRAT) + TELEPERFAM, data = prov.df)
        weights: nb2listw(prov.nb, style = "S")

Moran I statistic standard deviate = 4.7288, p-value =1.129e-06
alternative hypothesis: greater
sample estimates:
Observed Moran's I          Expectation              Variance
      0.276027992          -0.032044882           0.004244308
```

- Notes:

  o ***Accounting*** for the ***exogenous variables*** has substantially reduced the autocorrelation level ($\Rightarrow$misspecification perspective).

  o The ***expectation*** and ***variance*** of Moran's I dependent on the regression matrix **X**.

  o Residuals are best mapped by a ***bipolar*** map theme.

General structure of Moran's I and its distribution

The observed value of Moran's I

- Moran's I is defined to measure the strength of spatial autocorrelation in regression residuals $\mathbf{e} = \mathbf{M} \cdot \mathbf{y}$ with

$$\mathbf{e} = \begin{cases} \mathbf{y} - \mathbf{1} \cdot \bar{y} & \text{if } [\mathbf{I} - \mathbf{1} \cdot (\mathbf{1}^T \cdot \mathbf{1})^{-1} \cdot \mathbf{1}^T] \cdot \mathbf{y} \\ \mathbf{y} - \hat{\mathbf{y}} & \text{if } [\mathbf{I} - \mathbf{X} \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T] \cdot \mathbf{y} \end{cases}$$

variance (above first case) · residual (beside second case)

- Its observed value I^{obs} has the formal structure of a ***ratio of quadratic forms*** in the random variable $\mathbf{y}$

$$I^{obs} = \frac{\mathbf{y}^T \cdot \mathbf{M} \cdot \frac{1}{2} \cdot (\mathbf{V} + \mathbf{V}^T) \cdot \mathbf{M} \cdot \mathbf{y}}{\mathbf{y}^T \cdot \underbrace{\mathbf{M} \cdot \mathbf{M}}_{=\mathbf{M}} \cdot \mathbf{y}} = \frac{\mathbf{e}^T \cdot \frac{1}{2} \cdot (\mathbf{V} + \mathbf{V}^T) \cdot \mathbf{e}}{\mathbf{e}^T \cdot \mathbf{e}}$$

The denominator is simply the residual sum of squares: $RSS = \sum_{i=1}^{n} e_i^2 = \mathbf{e}^T \cdot \mathbf{e}$

- Besides the spatial autocorrelation pattern in the dependent ***random variable*** $\mathbf{y}$ — through $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Omega})$ — the observed value of Moran's I also depends on the ***exogenous*** variables $\mathbf{X}$ through the projection matrix $\mathbf{M}$ and the ***exogenous*** spatial link matrix $\mathbf{V}$.

spatial coding scheme also

- The spatial link matrix $\mathbf{V}$ is ***symmetrized*** by the transformation $\frac{1}{2} \cdot (\mathbf{V} + \mathbf{V}^T)$. influence moran's I

While the observed value I^{obs} is invariant under this transformation, the evaluation of its distribution requires a symmetric structure.

- Global Moran's I can be evaluated for different specification so the link matrix $\frac{1}{2} \cdot (\mathbf{V} + \mathbf{V}^T)$ such as <mark>higher order lags</mark> which leads to the <mark>correlogram</mark> or directional correlogram:
- <mark>The number of point pairs depends again on the spatial order of the spatial lag.</mark>
- Example correlograms:
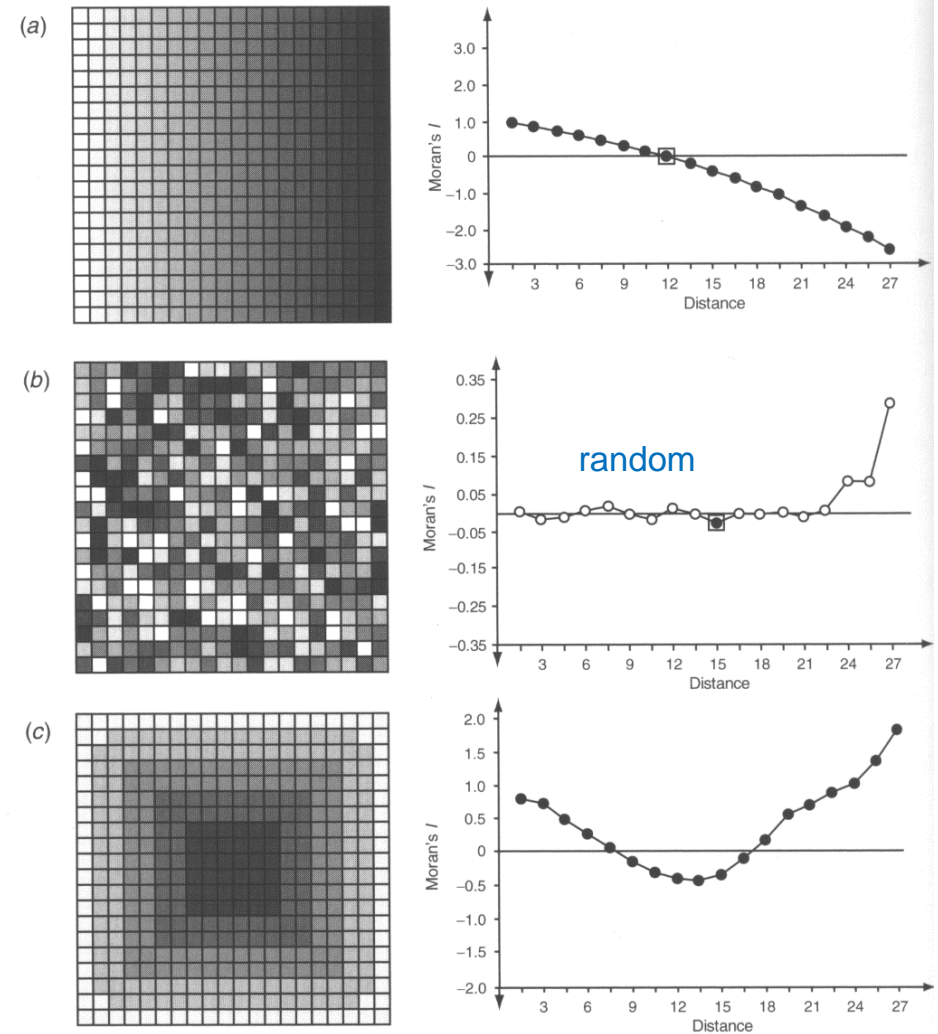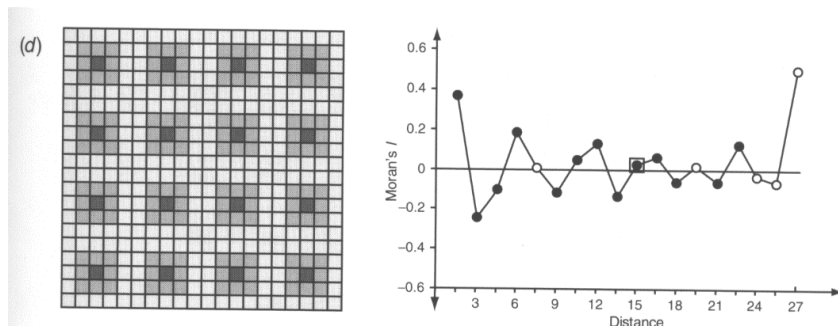
Extreme Values of Moran's I

- The ***largest and smallest possible values*** of Moran's I can be determined by the <mark>largest eigenvalue λ_1</mark> and <mark>smallest eigenvalue λ_n</mark> of the matrix $\mathbf{M} \cdot \frac{1}{2} \cdot (\mathbf{V} + \mathbf{V}^T) \cdot \mathbf{M}.$

- The associate eigenvectors $\mathbf{v}_i$ show a distinctive map pattern. These eigenvectors build the foundation for the <mark>spatial filtering methodology</mark>.   eigenvector may dominated by outlier

- Note that the possible value range of Moran's I does not necessarily need to be $I \in [-1,1]$ such as for thr Pearson's correlation coefficient.

Examples of extreme map patterns for different tessellations:

$I_{\max} = 0.9570$

$I_{\min} = \text{-}1.0304$

$$I_{max} = 0.9775$$

$$I_{min} = -0.5271$$
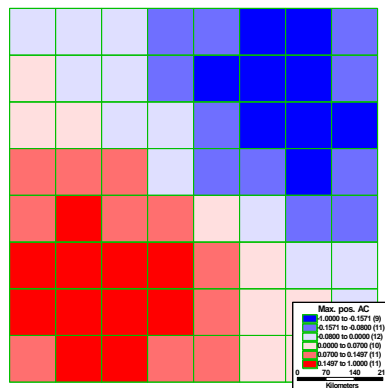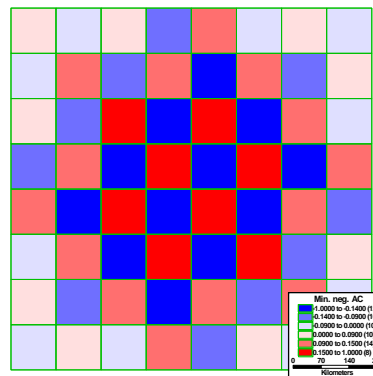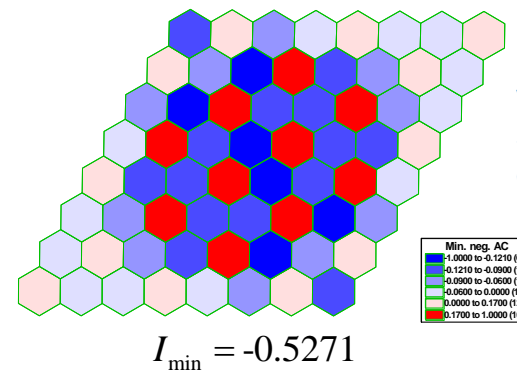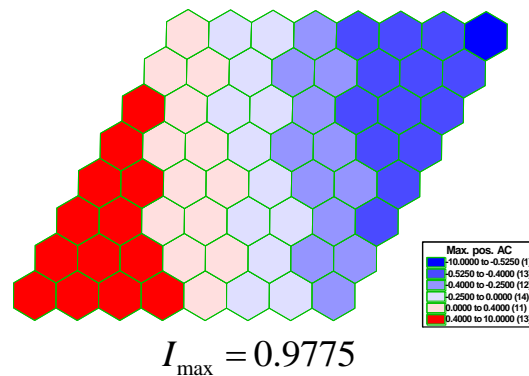
why it is greater than rectangular: since negative value also surround by negative one (share the boundary instead of one point)

- Most ***extreme possible spatial map patterns*** and their autocorrelation levels for the tessellation of the 95 Italian provinces for the *S*-coding scheme:
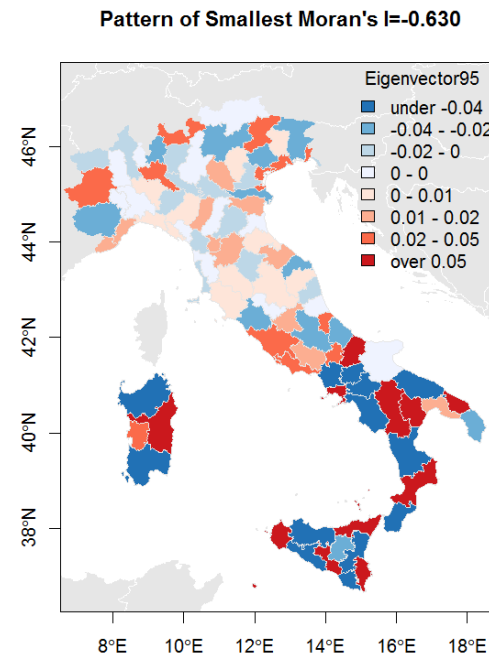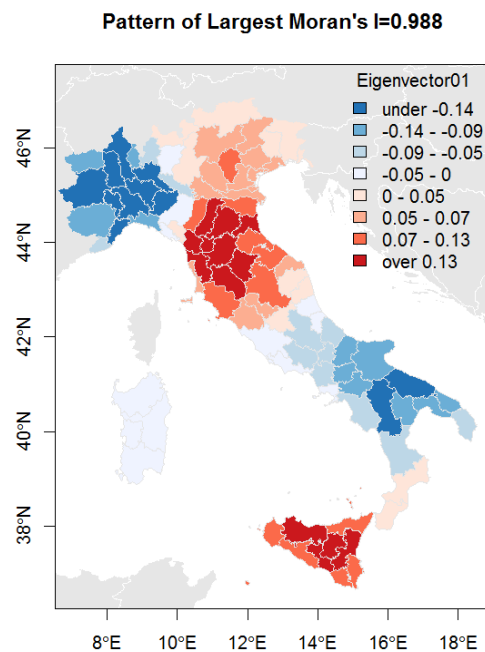

although it seems to have positive spatial autocorrelation, it flips by the interval of surrounding value

Test under the Assumption of Spatial Independence

Exploratory visualization:

- The observed residuals e_i can be plotted against the average values of their neighboring values e_i^{avg} using the "Moran's scatterplot".

  For the row-sum standardized link matrix $\mathbf{V}$ these averages are given by $\mathbf{e}^{avg} = \mathbf{V} \cdot \mathbf{e}$.

average residual of surroundings
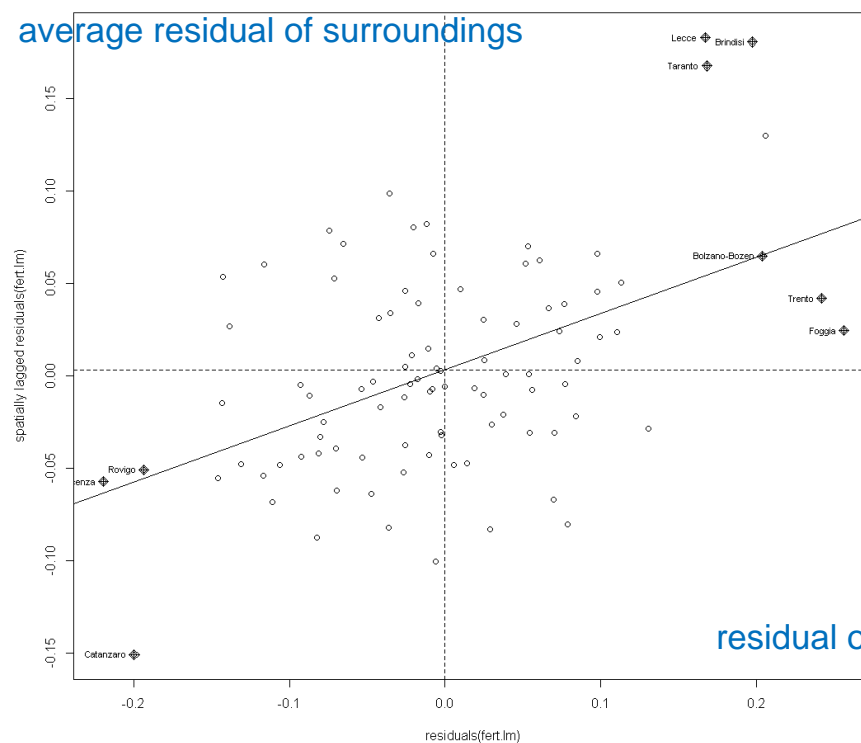
Covariance should be positive

residual of each observation

Under the assumption of $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \cdot \mathbf{I})$:

- Assuming that the population disturbances are independently identically distributed with $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \cdot \mathbf{I})$ the moments of Moran's I can be evaluated directly through the matrix terms:

$$E(I|H_0) = \frac{tr[\mathbf{M} \cdot \mathbf{V}]}{n - K}$$ expectation of moran's I

$$Var(I|H_0) = \frac{tr[\mathbf{M} \cdot \mathbf{V} \cdot \mathbf{M} \cdot \mathbf{V}^T] + tr[\mathbf{M} \cdot \mathbf{V} \cdot \mathbf{M} \cdot \mathbf{V}] + (tr[\mathbf{M} \cdot \mathbf{V}])^2}{(n - K) \cdot (n - K + 2)} - [E(I|H_0)]^2$$ variance

- For sufficiently **large number** of spatial objects n and **"well-behaved"** link matrices $\mathbf{V}$ one can use the normal approximation to evaluate the distribution of I^{obs}:

$$\frac{I^{obs} - E(I|H_0)}{\sqrt{Var(I|H_0)}} \sim N(0,1)$$ similar to t test

- This expression is frequently used in software implementations.
- However, those implementations frequently use incorrectly the projections matrix $\mathbf{M}_{(1)} = \mathbf{I} - \mathbf{1} \cdot (\mathbf{1}^T \cdot \mathbf{1})^{-1} \cdot \mathbf{1}^T$
- To evaluate the exact significance values of an observed value of Moran's I^{obs} numerical integration or the saddle-point approximation need to be employed.

- Rather than looking at tail probabilities to evaluate the significance of I^{obs} one can also use the cumulative distribution function:

- I^{obs} with small cumulative probabilities, say $\Pr(I < I^{obs}) < 0.025$, indicate a significant negatively autocorrelation, whereas,

- I^{obs} with large cumulative probabilities, say $\Pr(I < I^{obs}) > 0.975$, indicate a significant possitive autocorrelation   accumulative function
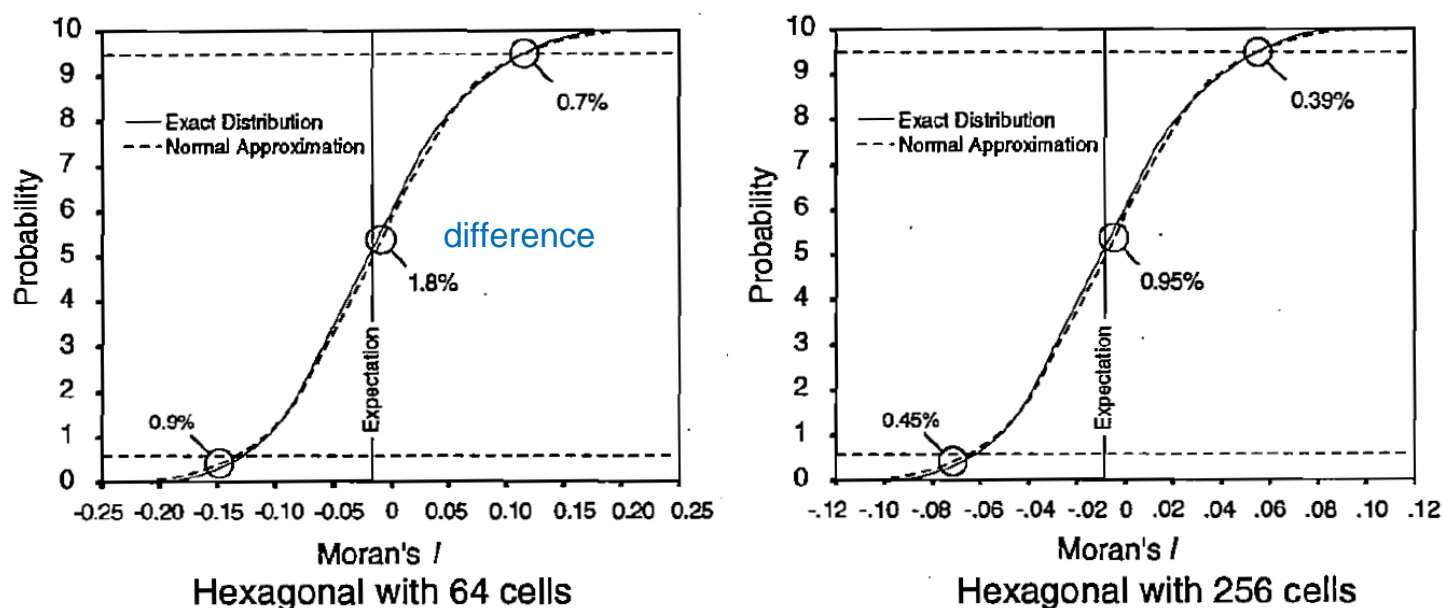
Fig. 8.5. Comparison of Moran's I's cumulative distribution function for hexagons with the normal distribution under the influence of an increasing number of spatial objects

- As the number of cells within a tessellation increases, for global Moran's I the normal approximation (i.e., the $\frac{I^{obs}-E(I|H_0)}{\sqrt{Var(I|H_0)}} \sim N(0,1)$) becomes more and more valid relative to the exact distribution of Moran's I.

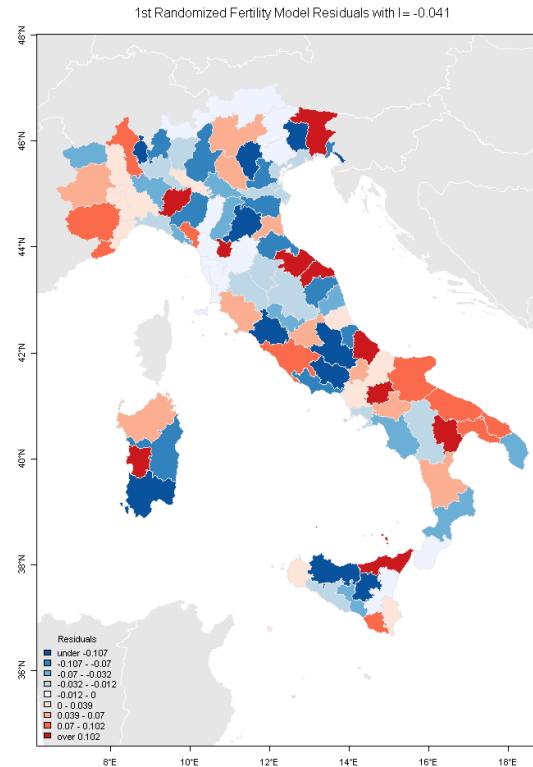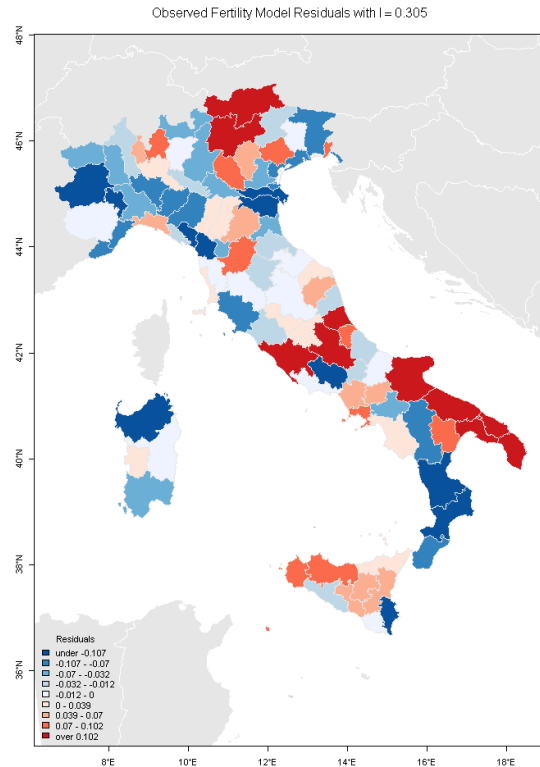Under the assumption of $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \cdot \boldsymbol{\Omega}(\rho))$:

- Again, these conditional distribution functions $F(I|\mathbf{X}, \boldsymbol{\Omega}(\rho))$ of Moran's I can be evaluated by numerical integration.

Under the assumption of $\boldsymbol{\varepsilon} \sim i.i.d$ but with unknown distribution:

- In case the distribution of the residuals is unknown randomization needs to be employed to evaluated the significance of the observed Moran's I^{obs}.
- Underlying idea: data generation process
    - Any observed spatial association of the residuals $\mathbf{e}$ is broken up by randomly distributing these residuals to different areas.
    - This random assignment generates a map pattern under the assumption
$$H_0: \mathbf{e} \sim spatial\ independent$$
    - This random assignment is repeated R-times (say $R = 999$) and for each random pattern r the associated Moran's I^r is calculated.
    - These random Moran's I^r are sorted ascendingly and their distribution is plotted. This is the distribution under the null hypothesis $H_0: \mathbf{e} \sim spatial\ independent$.

- o **If the observed Moran's I^{obs} falls into a tail of this distribution** (either lowest or highest ranking I^r's, **then** the observed pattern exhibits **significant spatial autocorrelation**. Otherwise, it is spatially independent.

Observed Fertility Model Residuals with I = 0.305   1st Randomized Fertility Model Residuals with I = -0.041   2st Randomized Fertility Model Residuals with I = 0.153
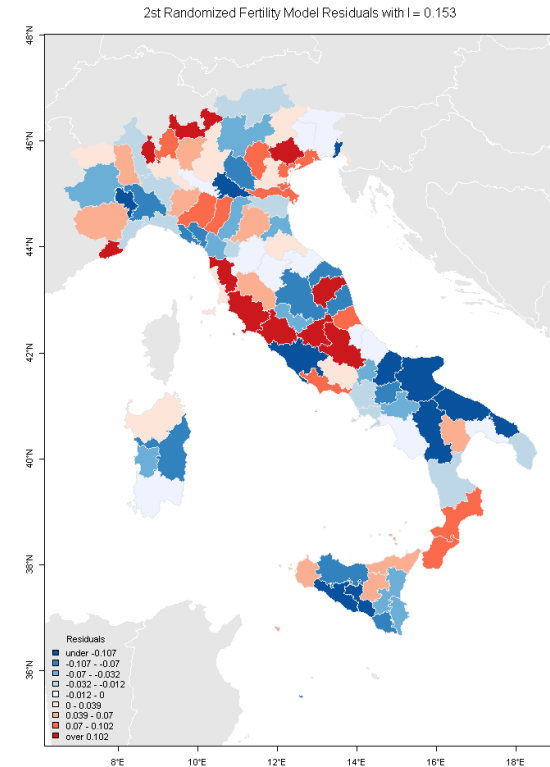
Local Moran's I_i

Technical Aspects

- Local Moran is defined structurally identical to global Moran. However, it uses the local link matrix $\mathbf{B}_i$ and a local scaling constant s_i for standardization:

$$I_i = \frac{\mathbf{e}^T \cdot s_i \cdot \mathbf{B}_i \cdot \mathbf{e}}{\mathbf{e}^T \cdot \mathbf{e}} .$$

local link matrices

- The local scaling constants for the three coding schemes are:

$$s_i = \begin{cases} \dfrac{n^2}{2 \cdot a} & \text{for the } C\text{-coding scheme} \\[2ex] \dfrac{n}{2 \cdot d_i} & \text{for the } W\text{-coding scheme} \\[2ex] \dfrac{n^2}{2 \cdot \sqrt{d_i} \cdot \sum_{j=1}^{n} \sqrt{d_j}} & \text{for the } S\text{-coding scheme} \end{cases}$$

- Due to the additive property of the local link matrices, i.e., $\mathbf{B} = \frac{1}{2} \cdot \sum_{i=1}^{n} \mathbf{B}_i$, the average of the local Moran's I_i's is equal to global Moran I:

global link matrix

$$I = \frac{1}{n} \cdot \sum_{i=1}^{n} I_i$$
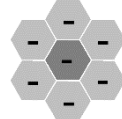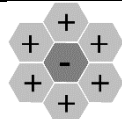
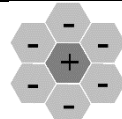$$E(I) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(I_i)$$
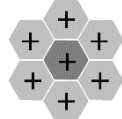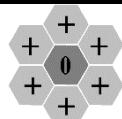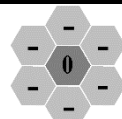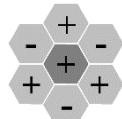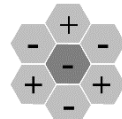
can over 1, those extreme local moran's I areas have huge impact on global, dominate.

- Thus <mark>large positive</mark> and <mark>negative local Moran's I_i contribute significantly to global Moran</mark>.

Interpretative Aspects

- *Example: Forms of Local Spatial Autocorrelation*

Positive local spatial Autocorrelation (spatial cluster)	
Negative local spatial Autocorrelation (spatial outlier, hot or cold spot) cold spot hot spot	
<mark>Local Spatial **Independence**</mark>	

- Local spatial autocorrelation measures the **association** of the residual e_i at the reference cell *i* with some weighted average of the residuals surrounding it.
  The key building block is the numerator which is:

$$e_i \cdot s_i \cdot \sum_{j=1}^{n} b_{ij} \cdot e_j$$

- Thus if either e_i or $\sum_{j=1}^{n} b_{ij} \cdot e_j$ are zero local Moran will be zero and insignificant.

  In the light of residual analysis this makes sense, since residual values around zero denote well-fitted observations that do not require further attention.

- If the terms e_i and $\sum_{j=1}^{n} b_{ij} \cdot e_j$ have opposite signs we call cell *i* a **spatial hot spot**.
 In the cumulative distribution function $\Pr(I_i \leq I_i^{obs}) \leq \alpha / 2$ indicates a significant hot spot

- If the terms e_i and $\sum_{j=1}^{n} b_{ij} \cdot e_j$ have identical signs we call cell *i* a **spatial cluster**.
  In the cumulative distribution function $\Pr(I_i \leq I_i^{obs}) \geq 1 - \alpha / 2$ indicates a significant spatial cluster.

- In order to determine whether we are dealing with a spatial cluster or a hot spot in either negative or
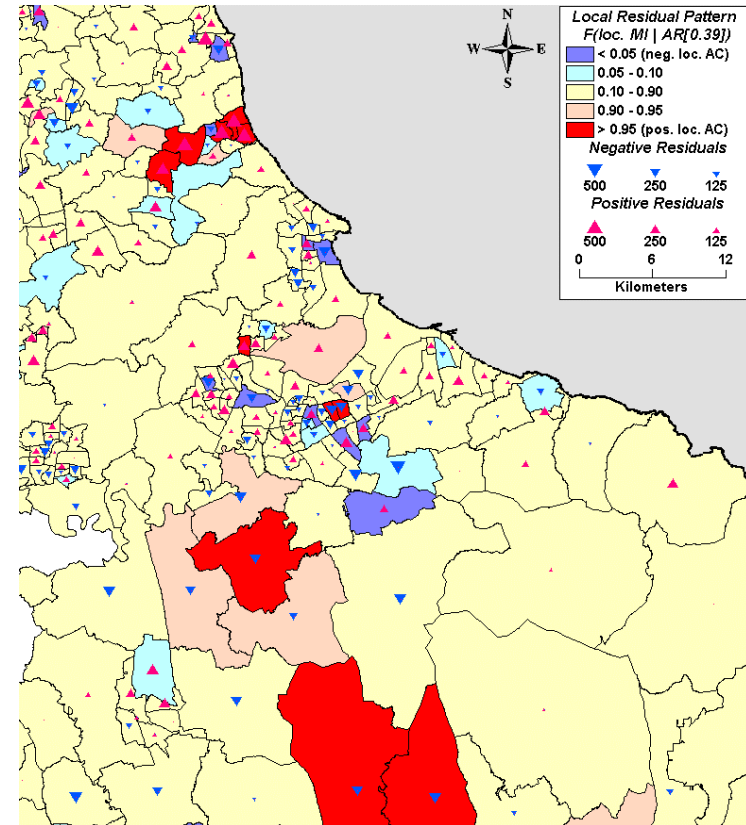
positive residuals at the reference cell *i* the sign of the residuals must be mapped simultaneously with the significance local Moran.

Statistical Aspects

- Local Moran is ***not normal distributed***.

  o It has an excessive kurtosis. Consequently, the normal approximation cannot be used.

  o Randomization approaches are also not feasible because of the low number of observations entering local Moran's I_i

  o However, there is an efficient approximation method, called the Saddlepoint approximation, (it is implemented in the ®-package `spdep`)

- Discussion of statistical problems associated with local statistics:

  o The local Moran's I_i statistics are influenced by edge effects   highly sensitive

  o Testing a set of local Moran's I_i statistics leads to the multiple testing problem and thus the significance level α of rejecting the zero hypothesis erroneously by chance is inaccurate: The true α-error, that at least one statistical test is rejected in error, becomes approximately $\alpha_{adj} = 1 - (1-\alpha)^n$ under the assumption that the individual tests are uncorrelated.
  Consequently, the error probability is increasing substantially.

o However, the local Moran's I_i statistics are correlated among each other even beyond the level of <mark>shared neighbors.</mark>

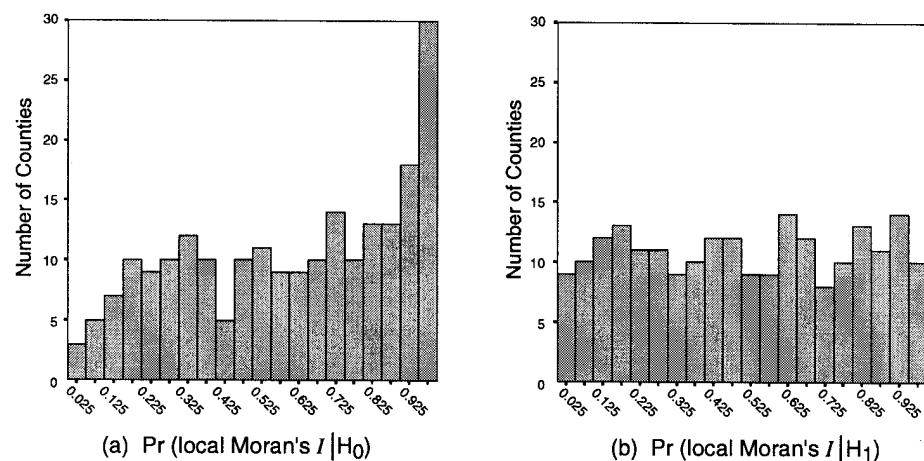Shared cells at lag 1   Shared cell at lag 2
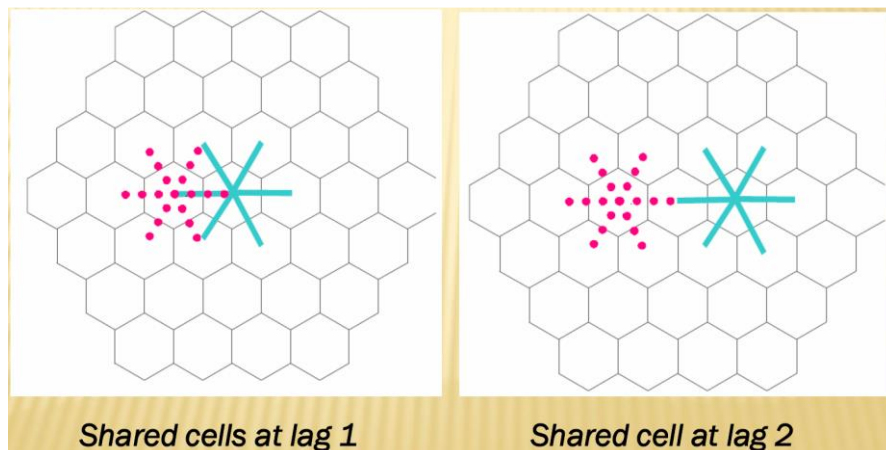
it means local moran's I are not independent with each other,

o Under the assumption of a presence of a global spatial process the local test statistics must be interpreted as variation around the global spatial process
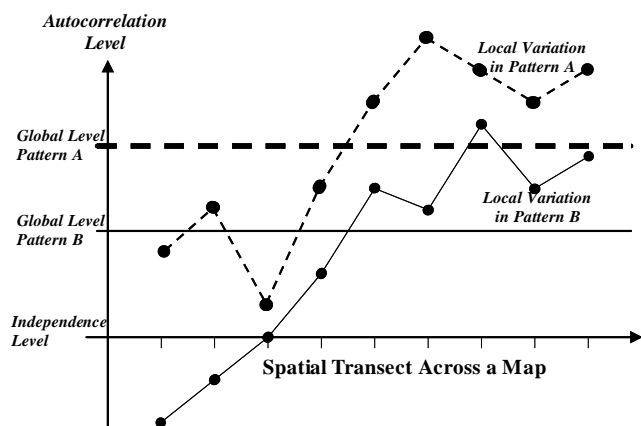
measure local moran's I with the average of each global level instead of absolute 0 line.

Fig. 11.3. Comparison of the distributions of $\Pr(I_i \leq I_{i0} \mid \sigma^2 \cdot \mathbf{I})$ and $\Pr(I_i \leq I_{i0} \mid \Omega(\rho_0 = 0.4197))$ for the 219 counties of the former GDR

- Explore the ®-markdown file **SpatialPatternItaly.rmd**.

The local G_i Statistic

- Another popular local test statistic is <mark>Getis</mark> and Ord's local $G_i(d)$ measure.

  o   It tests for local clusters of above or below the sum of all values.

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d) \cdot x_j}{\sum_{j=1}^{n} x_j} \quad j \neq i$$

cutoff d is hyper parameter here

<mark>with $w_{ij}(d)$ being 1 for all *(i,j)* pairs below the threshold distance *d* and otherwise zero</mark>

o The expectation is $E\left[G_i(d)\right] = \dfrac{W_i}{(n-1)}$ and $W_i = \displaystyle\sum_{j \neq i} w_{ij}(d)$

o Local G_i is only defined for $x_i > 0$ and thus <mark>cannot be applied to regression residuals</mark>

o Local G_i <mark>cannot detect negative autocorrelation</mark>