Prof. Dr. Reinhold Kosfeld
University of Kassel

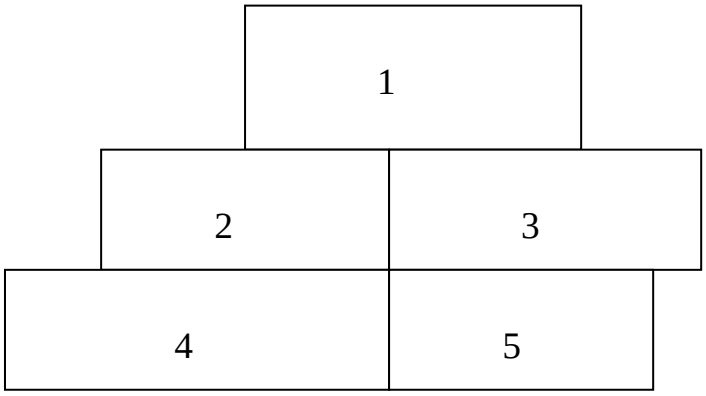
Spatial Econometrics

Assignments with Solutions

Part II – Assignments 11 - 14

General information on a system of regions and data

Five regions are arranged in the following form:



Distances between regional centres (in km):

Region	1	2	3	4	5
1		8	10	20	22
2			16	12	14
3				20	12
4					10

Data on geo-referenced variables:

Region	u	v
1	6	3
2	8	3
3	8	2
4	11	1
5	12	1

u: Unemployment rate
v: Vacancy rate

Beveridge curve:
Relation between the unemployment rate (u) and vacancy rate (v)

Assignment 11 (Local Moran coefficients)

Given are the data on the vacancy rate of five regions,

Region	1	2	3	4	5
Vacancy rate (v)	3	3	2	1	1

and the standardized spatial weight matrix

$$\mathbf{W} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}.$$

a) Compute the local Moran coefficients and interpret them!

● Local Moran coefficient:

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij} \cdot (x_j - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2 / n}$$

Arithmetic mean of the vacancy rate:

$$\bar{x} = \frac{1}{5} \sum_{i=1}^{5} x_i = \frac{1}{5}(3+3+2+1+1) = \frac{10}{5} = 2$$

Variance of the vacancy rate:

$$s_x^2 = \frac{1}{5} \sum_{j=1}^{5} (x_j - \bar{x})^2 = \frac{1}{5}[(3-2)^2 + (3-2)^2 + (2-2)^2 + (1-2)^2 + (1-2)^2$$

$$= \frac{1}{5}[1^2 + 1^2 + 0^2 + (-1)^2 + (-1)^2] = \frac{4}{5} = 0{,}8$$

Global Moran's I for vacancy rate: I = 0.25 (→ Assignment 3b)

Expected value: $E(I_i) = -\frac{w_{i\bullet}}{n-1} = -\frac{1}{n-1} = -\frac{1}{5-1} = -0.25$

Local Moran coeffients for regions 1 to 5:

Region 1

$$I_1 = \frac{1}{0.8} \cdot (x_1 - \bar{x}) \cdot \sum_{j=1}^{5} w_{1j} \cdot (x_j - \bar{x}) = 1.25 \cdot (3-2) \cdot [\frac{1}{2} \cdot (3-2) + \frac{1}{2} \cdot (2-2)]$$

$$= 1.25 \cdot 1 \cdot (\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0) = 1.25 \cdot 1 \cdot 0.5 = 0.625$$

$I_1 = 0.625 > E(I_1) = -0.25$ → positive local spatial autocorrelation
$I_1 = 0.625 > I = 0.25$ → spatial clustering of the vacancy rate (high or low value clustering) around region 1

Region 2

$$I_2 = \frac{1}{0.8} \cdot (x_2 - \bar{x}) \cdot \sum_{j=1}^{5} w_{2j} \cdot (x_j - \bar{x})$$

$$= 1.25 \cdot (3-2) \cdot [\frac{1}{3} \cdot (3-2) + \frac{1}{3} \cdot (2-2) + \frac{1}{3} \cdot (1-2)]$$

$$= 1.25 \cdot 1 \cdot [\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot (-1)] = 1.25 \cdot 1 \cdot 0 = 0$$

$I_2 = 0 > E(I_2) = -0.25$ → positive local spatial autocorrelation
$I_2 = 0 < I = 0.25$ → no spatial clustering of the vacancy rate around region 2

Region 3

$$I_3 = \frac{1}{0.8} \cdot (x_3 - \bar{x}) \cdot \sum_{j=1}^{5} w_{3j} \cdot (x_j - \bar{x})$$

$$= 1.25 \cdot (2-2) \cdot [\frac{1}{3} \cdot (3-2) + \frac{1}{3} \cdot (3-2) + \frac{1}{3} \cdot (1-2)]$$

$$= 1.25 \cdot 0 \cdot [\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot (-1)] = 1.25 \cdot 0 \cdot \frac{1}{3} = 0$$

$I_3 = 0 > E(I_3) = -0.25$ → positive local spatial autocorrelation
$I_3 = 0 < I = 0.25$ → no spatial clustering of the vacancy rate around region 3

Region 4

$$I_4 = \frac{1}{0.8} \cdot (x_4 - \bar{x}) \cdot \sum_{j=1}^{5} w_{4j} \cdot (x_j - \bar{x}) = 1.25 \cdot (1-2) \cdot [\frac{1}{2} \cdot (3-2) + \frac{1}{2} \cdot (1-2)]$$

$$= 1.25 \cdot (-1) \cdot [\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (-1)] = 1.25 \cdot (-1) \cdot 0 = 0$$

$I_4 = 0 > E(I_4) = -0.25$ → positive local spatial autocorrelation
$I_4 = 0 < I = 0.25$ → no spatial clustering of the vacancy rate around region 4

Region 5

$$I_5 = \frac{1}{0.8} \cdot (x_5 - \bar{x}) \cdot \sum_{j=1}^{5} w_{5j} \cdot (x_j - \bar{x}) = 1.25 \cdot (1-2) \cdot [\frac{1}{2} \cdot (2-2) + \frac{1}{2} \cdot (1-2)]$$

$$= 1.25 \cdot (-1) \cdot [\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot (-1)] = 1.25 \cdot (-1) \cdot (-0.5) = 0.625$$

$I_5 = 0.625 > E(I_5) = -0.25$ → positive local spatial autocorrelation
$I_5 = 0.625 > I = 0.25$ → spatial clustering of the vacancy rate (high or low value clustering) around region 5

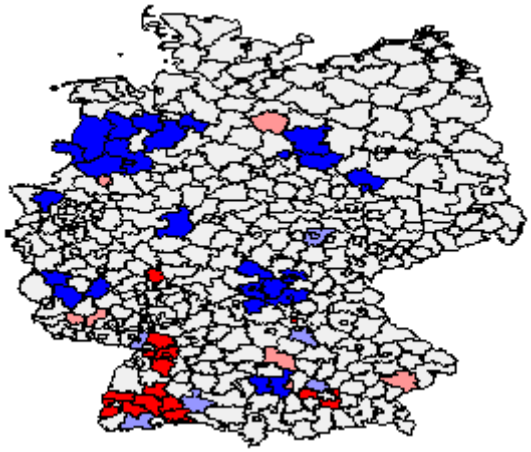
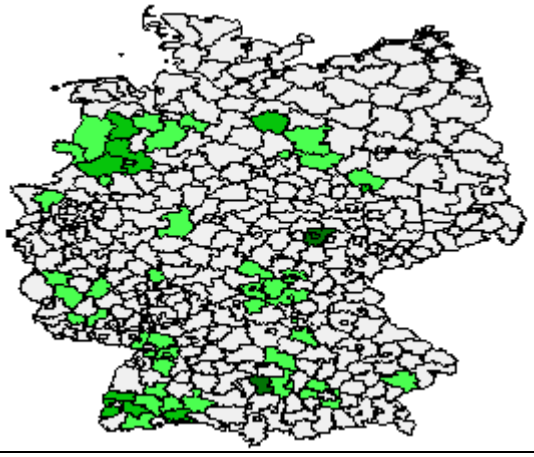
b) Show the relationship between the local Moran cofficients and global Moran's I!

Arithmetic mean of local Moran coefficients I_i's = global Moran's I:

$$\bar{I} = \frac{1}{5} \sum_{i=1}^{5} I_i = \frac{1}{5} \cdot (0.625 + 0 + 0 + 0 + 0.625) = \frac{1}{5} \cdot 1.25 = 0.25 \ \ (= I)$$

Assignment 12 (LISA Cluster and Significance Map)

GeoDa has created the following LISA Cluster and Significance Map for the share of employees of manufacture of electrical and optical instruments with the default settings:

LISA Cluster Map	LISA Significance Map
	

a) Interpret the LISA Cluster Map!

GeoDa creates a LISA Cluster Map for identifying local clusters and spatial outliers of a geo-referenced variable. For each region the Local Moran statistic is computed and tested for significance by randomization. Significant local clusters and spatial outliers are portrayed in different colours.

Spatial clusters are highlighted by bright colours. Hot spots (high-high regions) are coloured bright red. In these districts, positive spatial association arises from the own and neighbouring high values of the attribute variable. Cold spots (low-low regions) are coloured bright blue. Here, positive spatial autocorrelation emerges from the own and neighbouring low values of the variable.

Hot spots of manufacture of electrical and optical instruments are most notably found in Baden-Wuerttemberg. None of the high-high regions are located in north-west, central and east Germany where most of the cold spots are situated. Some cold spots are additionally identified in Bavaria.

Spatial outliers are identified by light colours. Because of their inverse orientation, such areas depict pockets of instationarity. Areas with negative local Moran coefficients originating from own high and low surrounding values are termed high-low outliers. They are coloured light red. Regions with negative local Moran coefficients resulting from own low and high surrounding values are low-high outliers. These territories are coloured light blue.

While high-low outliers are dispersed across West Germany, with one exception low-high outliers are located in the southern states of Baden-Wuerttemberg and Bavaria.

b) What additional information does the LISA Significance Map provide?

The LISA Significance Map provides additional information on the significance level of the identified local clusters and spatial outliers. Nearly all detected spatial outliers are significant at a level of 5% (light green). One district in the north and one district region in the south of Germany at the former inner-German border are significant outliers at a significance level of 1% (medium green) and 0.1% (dark green), respectively.

Hot spots at a significant level of 1% can only be found in southern Baden-Wuerttemberg. All other hot spots are significant at the 5% level. Most of the cold spots are as well significant at this level. A higher significance level is primarily observed for some cold spots in Lower Saxony.

Assignment 13 (Local Getis-Ord G_i and G_i*)

The vacancy rate (v) is unequally distributed across space:

Region	1	2	3	4	5
Vacancy rate (v)	3	3	2	1	1

The distance-based contiguity matrix with a critical distance of d=14.4 reads

$$\mathbf{W}(d=14.4)=\begin{bmatrix}1&1&1&0&0\\1&1&0&1&1\\1&0&1&0&1\\0&1&0&1&1\\0&1&1&1&1\end{bmatrix}.$$

a) Compute the G_i statistics and interpret them!

G_i statistic: $G_i=\dfrac{\sum_{j\neq i} w_{ij}(d)\cdot x_j}{\sum_{j\neq i} x_j}$ with $E(G_i)=\dfrac{\sum_{j\neq i} w_{ij}(d)}{n-1}$

Region 1

$$G_1=\frac{\sum_{j\neq 1} w_{1j}(d=14.4)\cdot x_j}{\sum_{j\neq 1} x_j}=\frac{1\cdot 3+1\cdot 2}{3+2+1+1}=\frac{5}{7}=0.7143$$

$$E(G_1)=\frac{\sum_{j\neq 1} w_{1j}(d=14.4)}{n-1}=\frac{1+1+0+0}{5-1}=\frac{2}{4}=0.5$$

Region 2

$$G_2=\frac{\sum_{j\neq 2} w_{2j}(d=14.4)\cdot x_j}{\sum_{j\neq 2} x_j}=\frac{1\cdot 3+1\cdot 1+1\cdot 1}{3+2+1+1}=\frac{5}{7}=0.7143$$

$$E(G_2)=\frac{\sum_{j\neq 2} w_{2j}(d=14.4)}{n-1}=\frac{1+0+1+1}{5-1}=\frac{3}{4}=0.75$$

Region 3

$$G_3 = \frac{\sum_{j\neq 3} w_{3j}(d=14.4)\cdot x_j}{\sum_{j\neq 3} x_j} = \frac{1\cdot 3+1\cdot 1}{3+3+1+1} = \frac{4}{8} = 0.5$$

$$E(G_3) = \frac{\sum_{j\neq 3} w_{3j}(d=14.4)}{n-1} = \frac{1+0+0+1}{5-1} = \frac{2}{4} = 0.5$$

Region 4

$$G_4 = \frac{\sum_{j\neq 4} w_{4j}(d=14.4)\cdot x_j}{\sum_{j\neq 4} x_j} = \frac{1\cdot 3+1\cdot 1}{3+3+2+1} = \frac{4}{9} = 0.4444$$

$$E(G_4) = \frac{\sum_{j\neq 4} w_{4j}(d=14.4)}{n-1} = \frac{0+1+0+1}{5-1} = \frac{2}{4} = 0.5$$

Region 5

$$G_5 = \frac{\sum_{j\neq 5} w_{5j}(d=14.4)\cdot x_j}{\sum_{j\neq 5} x_j} = \frac{1\cdot 3+1\cdot 2+1\cdot 1}{3+3+2+1} = \frac{6}{9} = 0.6667$$

$$E(G_5) = \frac{\sum_{j\neq 5} w_{5j}(d=14.4)}{n-1} = \frac{0+1+1+1}{5-1} = \frac{3}{4} = 0.75$$

Interpretation:
Tendency of spatial concentration of high values: surrounding of region 1
Tendency of spatial concentration of low values: surrounding of regions 2, 4, 5
No spatial concentration: surrounding of region 3

b) Compute the G_i^* statistics and interpret them!

G_i^* statistic: $G_i^* = \frac{\sum_{j=1}^{n} w_{ij}(d) \cdot x_j}{\sum_{j=1}^{n} x_j}$ with $E(G_i^*) = \frac{\sum_{j=1}^{n} w_{ij}(d)}{n}$

Region 1

$$G_1^* = \frac{\sum_{j=1}^{n} w_{1j}(d=14.4) \cdot x_j}{\sum_{j=1}^{n} x_j} = \frac{1 \cdot 3 + 1 \cdot 3 + 1 \cdot 2}{3+3+2+1+1} = \frac{8}{10} = 0.8$$

$$E(G_1^*) = \frac{\sum_{j=1}^{n} w_{1j}(d=14.4)}{n} = \frac{1+1+1+0+0}{5} = \frac{3}{5} = 0.6$$

Region 2

$$G_2^* = \frac{\sum_{j=1}^{n} w_{2j}(d=14.4) \cdot x_j}{\sum_{j=1}^{n} x_j} = \frac{1 \cdot 3 + 1 \cdot 3 + 1 \cdot 1 + 1 \cdot 1}{3+3+2+1+1} = \frac{8}{10} = 0.8$$

$$E(G_2^*) = \frac{\sum_{j=1}^{n} w_{2j}(d=14.4)}{n} = \frac{1+1+0+1+1}{5} = \frac{4}{5} = 0.8$$

Region 3

$$G_3^* = \frac{\sum_{j=1}^{n} w_{3j}(d=14.4) \cdot x_j}{\sum_{j=1}^{n} x_j} = \frac{1 \cdot 3 + +1 \cdot 2 + 1 \cdot 1}{3+3+2+1+1} = \frac{6}{10} = 0.6$$

$$E(G_3^*) = \frac{\sum_{j=1}^{n} w_{3j}(d=14.4)}{n} = \frac{1+0+1+0+1}{5} = \frac{3}{5} = 0.6$$

Region 4

$$G_4^* = \frac{\sum_{j=1}^{n} w_{4j}(d = 14.4) \cdot x_j}{\sum_{j=1}^{n} x_j} = \frac{1 \cdot 3 + +1 \cdot 1 + 1 \cdot 1}{3 + 3 + 2 + 1 + 1} = \frac{5}{10} = 0.5$$

$$E(G_4^*) = \frac{\sum_{j=1}^{n} w_{4j}(d = 14.4)}{n} = \frac{0 + 1 + 0 + 1 + 1}{5} = \frac{3}{5} = 0.6$$

Region 5

$$G_5^* = \frac{\sum_{j=1}^{n} w_{5j}(d = 14.4) \cdot x_j}{\sum_{j=1}^{n} x_j} = \frac{1 \cdot 3 + +1 \cdot 2 + 1 \cdot 1 + 1 \cdot 1}{3 + 3 + 2 + 1 + 1} = \frac{7}{10} = 0.7$$

$$E(G_5^*) = \frac{\sum_{j=1}^{n} w_{5j}(d = 14.4)}{n} = \frac{0 + 1 + 1 + 1 + 1}{5} = \frac{4}{5} = 0.8$$

Interpretation:
Tendency of spatial concentration of high values: region 1 or surrounding regions
Tendency of spatial concentration of low values: regions 4 and 5 or surrounding regions
No spatial concentration: regions 2 and 3 and surrounding regions

Assignment 14 (OLS regression)

Data on the unemployment rate (u) and the vacancy rate (v) are given for 5 regions.

Region	1	2	3	4	5
Unemployment rate (u)	6	8	8	11	12
Vacancy rate (v)	3	3	2	1	1

a) Estimate the Beveridge curve by ordinary least squares (OLS)!

Dependent variable Y: unemployment rate (u)
Independent variable X: vacancy rate (v)

Vector of the endogenous variable **y**: Observation matrix **X**:

$$\mathrm{y} = \begin{bmatrix} 6 \\ 8 \\ 8 \\ 11 \\ 12 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Matrix product **X'X**, its inverse (**X'X**) -1, matrix product **X'y**:

$$\mathbf{X'X} = \begin{bmatrix} 5 & 10 \\ 10 & 24 \end{bmatrix} \qquad \mathbf{X'y} = \begin{bmatrix} 45 \\ 81 \end{bmatrix}$$

$$(\mathbf{X'X})^{-1} = \frac{1}{|\mathbf{X'X}|}(\mathbf{X'X})^{\mathrm{adj}} = \frac{1}{|\mathbf{X'X}|}\begin{bmatrix} \mathrm{xx}_{22} & -\mathrm{xx}_{21} \\ -\mathrm{xx}_{12} & \mathrm{xx}_{11} \end{bmatrix}$$

xx_{ij}: Element of the *i*th row and *j*th column of **X'X**

$$|\mathbf{X'X}| = \mathrm{xx}_{11} \cdot \mathrm{xx}_{22} - \mathrm{xx}_{12} \cdot \mathrm{xx}_{21} = 5 \cdot 24 - 10^2 = 120 - 100 = 20$$

$$(\mathbf{X'X})^{-1} = \frac{1}{20}\begin{bmatrix} 24 & -10 \\ -10 & 5 \end{bmatrix} = \begin{bmatrix} 1.2 & -0.5 \\ -0.5 & 0.25 \end{bmatrix}$$

OLS estimator of **β**:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} = \begin{bmatrix} 1.2 & -0.5 \\ -0.5 & 0.25 \end{bmatrix}\begin{bmatrix} 45 \\ 81 \end{bmatrix} = \begin{bmatrix} 13.5 \\ -2.25 \end{bmatrix}$$

b) Compute the residual variance and the standard error of regression!

Vector of fitted values $\hat{\mathbf{y}}$:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 13.5 \\ -2.25 \end{bmatrix} = \begin{bmatrix} 6.75 \\ 6.75 \\ 9 \\ 11.25 \\ 11.25 \end{bmatrix}$$

Vector of residuals **e**:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 6 \\ 8 \\ 8 \\ 11 \\ 12 \end{bmatrix} - \begin{bmatrix} 6.75 \\ 6.75 \\ 9 \\ 11.25 \\ 11.25 \end{bmatrix} = \begin{bmatrix} -0.75 \\ 1.25 \\ -1 \\ -0.25 \\ 0.75 \end{bmatrix}$$

Residual variance $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{5-2} = \frac{1}{3}\begin{bmatrix} -0.75 & 1.25 & -1 & -0.25 & 0.75 \end{bmatrix} \begin{bmatrix} -0.75 \\ 1.25 \\ -1 \\ -0.25 \\ 0.75 \end{bmatrix} = \frac{3.75}{3} = 1.25$$

Standard error of regression (SER):

$$\text{SER} = \sqrt{1.25} = 1.1180$$

c) Are the regression coefficients significant (α=0.05)?

- for β_1 (H_0: $\beta_1 = 0$)

OLS estimator for β_1: $\hat{\beta}_1 = 13.5$

Test statistic: $t_1 = \frac{\hat{\beta}_1}{\hat{\sigma} \cdot \sqrt{xx^{11}}} = \frac{13.5}{1.1180 \cdot \sqrt{1.2}} = \frac{13.5}{1.2247} = 11.023$

Critical value (α=0.05, two-sided test): t(3,0.975) = 3.182

Testing decision: (| t_1 |=11.023) > [t(3;0.975)=3.182] => Reject H_0

- for β_2 (H_0: $\beta_2 = 0$)

OLS estimator for β_2: $\hat{\beta}_2 = -2.25$

Test statistic: $t_2 = \frac{\hat{\beta}_2}{\hat{\sigma} \cdot \sqrt{xx^{22}}} = \frac{-2.25}{1.118 \cdot \sqrt{0.25}} = \frac{-2.25}{0.559} = -4.025$

Critical value (α=0.05, two-sided test): t(3,0.975) = 3.182

Testing decision: (| t_2 |=4.025) > [t(3;0.975)=3.182] => Reject H_0

d) Compute the coefficient of determination and interpret it!

Working table:

i	y_i	$\hat{y}_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$
1	6	6.75	-3	9	-2.25	5.0625
2	8	6.75	-1	1	-2.25	5.0625
3	8	9	-1	1	0	0
4	11	11.25	2	4	2.25	5.0625
5	12	11.25	3	9	2.25	5.0625
$\sum$	45	45	0	24	0	20.25

$$\bar{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{1}{5}45 = 9$$

SST = 24, SSE = 20.25, SSR = SST – SSE = 24 – 20.25 = 3.75

$$R^2 = \frac{SSE}{SST} = \frac{20.25}{24} = 0.844$$

or

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{3.75}{24} = 1 - 0.156 = 0.844$$

Interpretation:
84.4% of the variation of regional unemployment rates is explained in a linear regression model by regional vacancy rates.